JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Molecular Collisions. IV. Nearly Spherical Rigid Body Approximation*†

GEORGE GIOUMOUSIS‡
University of Wisconsin Theoretical Chemistry Laboratory, Madison, Wisconsin and Shell Development Company, Emeryville, California
AND
C. F. CURTISS
University of Wisconsin Theoretical Chemistry Laboratory, Madison, Wisconsin
(Received April 13, 1962)

Rotational transitions in the collisions of diatomic molecules are treated by means of a rigid body approximation. The deviation from sphericity is assumed to be sufficiently small that it may be treated as a perturbation. The differential cross section is calculated for the H_2-HD collision in which the hydrogen remains in its ground state while the hydrogen deuteride goes from its ground state to the first excited rotational state.

1. INTRODUCTION

THE previous papers[1-3] of this series have been concerned with the general treatment of molecular collisions, via quantum mechanics. A rigorous treatment of inelastic collisions was given, mainly in order to be able to take the Pauli principle into account. Then, the special case of diatomic molecules was set up in terms of a set of coupled integral equations, with effective exploitation of the properties of the three-dimensional rotation group.[4] The analysis was carried as far as seemed possible without the introduction of approximations.

The idealization of molecules as rigid bodies has been a frequently applied simplifying assumption, as witness the wide and successful use of the rigid sphere model in the kinetic theory of gases. In quantum mechanics the advantage of the method is that instead of having to solve a differential equation, one has only to fit boundary conditions, which generally is much simpler.

For the collisions of diatomic molecules the proper model is one of rigid cylindrically symmetric bodies. Vibration is not to be considered. It will be assumed that the molecules are nearly spherical, so that the deviation from sphericity can be treated as a perturbation. As an illustrative example, the cross section for the $l = 0$ to $l = 1$ transition for hydrogen deuteride colliding with hydrogen is calculated.

The work follows closely that reported in two previous papers,[1,2] but is independent of the third[3] in this series. They will be referred to as I, II, and III, respectively, and equations from them as, say, (I-35) or (II-3.12).

The nomenclature for the most part follows II, that is, rotations are denoted R, S^a, R^b, etc. However, the difficulties with all of the symbols for the

* This paper is based on a thesis [University of Wisconsin Naval Research Laboratory Rept. WIS-NSF-5 (1955)] submitted by George Gioumousis to the University of Wisconsin in partial fulfillment of the requirements of a PhD degree in theoretical chemistry.
† This work was sponsored in part by a National Science Foundation grant. One of us (GG) held predoctoral fellowships sponsored by the Shell Companies Foundation and by the National Science Foundation.
‡ Present address: Shell Development Company, Emeryville, California.
[1] G. Gioumousis and C. F. Curtiss, J. Chem. Phys. 29, 996 (1958).
[2] G. Gioumousis and C. F. Curtiss, J. Math. Phys. 2, 96 (1961).
[3] G. Gioumousis, J. Math. Phys. 2, 723 (1961).
[4] E. Wigner, *Group Theory* (Academic Press Inc., New York, 1959).

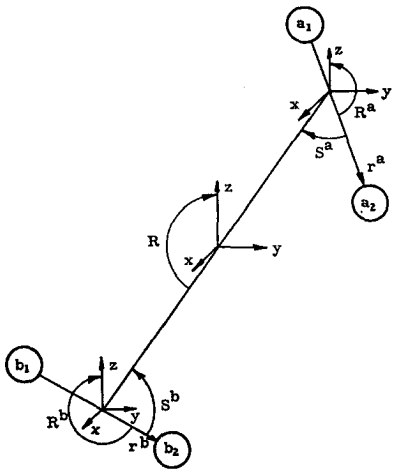


FIG. 1. The coordinate system used. The molecules are "a" and "b," composed of atoms a_1, a_2, b_1, and b_2. The vector $\mathbf{r}^a$ is from a_1 to a_2, $\mathbf{r}^b$ is similarly defined, while $\mathbf{r}$ is from the center of mass of "a" to that of "b." The rotations R, R^a, R^b make $\mathbf{r}$, $\mathbf{r}^a$, $\mathbf{r}^b$, respectively, parallel to the z axis, with the further provision that R put $\mathbf{r}^a$ in the left half of the xz plane. The rotations S^a, S^b make $\mathbf{r}^a$, $\mathbf{r}^b$ parallel to $\mathbf{r}$, and are given by $R^a = S^aR$ and $R^b = S^bR$.

Wigner coefficients have led us to introduce still another. The Wigner $S^{ab}_{c\alpha\beta}$ is impossible for the printer if the indices are sub- or superscripted, let alone both. The three-j symbol, defined by

$$\begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} = \frac{(-1)^{a-b-\gamma}}{(2c+1)^{1/2}} S^{ab}_{c\alpha\beta}\, \delta_{\alpha+\beta+\gamma,0},$$

has the simplest symmetry, but has a redundant index that unduly complicates summation formulas. For this reason, in this paper we use the three-j symbol with the added convention that any one of the three lower indices may be replaced by a dot which will denote the negative sum of the other two. We thus have

$$S^{ab}_{c\alpha\beta} = (2c+1)^{1/2}(-1)^{-a+b-\alpha-\beta}\begin{pmatrix} a & b & c \\ \alpha & \beta & \cdot \end{pmatrix},$$

$$S^{ab}_{c,\alpha,\gamma-\alpha} = (2c+1)^{1/2}(-1)^{-a+b-\gamma}\begin{pmatrix} a & b & c \\ \alpha & \cdot & -\gamma \end{pmatrix}, \quad \text{etc.}$$

For a similar reason, the Racah coefficients are written as six-j symbols.

The coordinate system to be used here is identical to that of II. Briefly, rotations which place a molecular configuration into some standard configuration are used as coordinates. The coordinates are defined in Fig. 1. It should be noted that in the absence of vibration the intermolecular potential is a function only of S^a, S^b, and r.

The shapes of the molecules are most conveniently specified in terms of a "distance-of-closest-approach" function. Let

$$\rho(S^a, S^b) \tag{1.1}$$

be the value of r at relative orientation S^a, S^b at which they just touch; that is, the smallest value of r for which they do not overlap. The molecular potential function may then be considered to be

$$\begin{aligned} V(rS^aS^b) &= 0 \qquad r > \rho(S^aS^b) \\ &= \infty \qquad r \leq \rho(S^aS^b). \end{aligned} \tag{1.2}$$

The statement that the molecules can not overlap means that the wave function is zero for configurations denoting overlapping, and then by continuity must be zero if the molecules just touch. Thus, if the wave function be $\psi(rRS^aS^b)$, the boundary condition is

$$\psi(\rho(S^aS^b)RS^aS^b) = 0 \tag{1.3}$$

identically in R, S^a, and S^b.

However, for the sake of consistency, it is desirable that the same boundary condition be derivable by use of the potential of Eq. (1.2) and the integral equation (II-3.1). Such a potential can only have meaning in terms of a limiting process. Consider for a moment a one-dimensional potential which has a large jump at the point a. Then the wave function has certain values to the left of a, being such that the integral of the product of it, the potential, and the Green's function has the proper value. Now if the height of the step of $V(r)$ be raised, the values of $\psi(r)$ to the left of a must be correspondingly decreased in order that the integral remain finite. In the limit that the step be of infinite height the wave function must be zero at the point a. This is shown in Fig. 2.

Clearly the above argument is valid in more than one dimension. Furthermore, the potential function need not be a step function as long as it does approach a function of the form of Eq. (1.2) in the limit.

2. THE BOUNDARY CONDITION

It is convenient at this point to introduce a more condensed notation. Let q denote the set of quantum numbers $l^am^al^bm^b$, $\bar{q}$ the set $\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b$, and similarly p the set $l^al^bl\lambda$ and $\bar{p}$ the set $\bar{l}^a\bar{l}^b\bar{l}\bar{\lambda}$. Then let $m = m^a + m^b$ and $\bar{m} = \bar{m}^a + \bar{m}^b$, since, for example, it will ocasionally be convenient to sum over m and m^a instead of m^a and m^b.

The form of the function to which the boundary condition is to be applied may be seen by inspection of Eq. (II-3.9). First consider only values of r larger than the maximum of the function ρ. It is evident from the form of the integral operator that the second term is zero, while the integral in the first is a constant $b(\bar{p}\bar{L})$ independent of the value

of r. Thus, the relation[5,6]

$$\psi(\bar{p}\bar{L}; p \mid r) = \delta_{p\bar{p}} j(\bar{\lambda} \mid kr) + b(\bar{p}\bar{L}; p) h(\lambda \mid kr) \quad (2.1)$$

holds, subject to the limitation on the values of r. However, the above function satisfies the differential equation for all configurations such that $V = 0$, and thus is the correct solution for all such configurations.

The boundary condition of Eq. (1.3) may be applied to a wave function possessing sharp angular momentum to give

$$\begin{aligned} 0 &= \psi(\bar{p}\bar{L}\bar{M} \mid \rho(S^a S^b) R S^a S^b) \\ &= \sum_{l^a l^b l \lambda s t} \psi(\bar{p}\bar{L}; l^a l^b l \lambda \mid \rho(S^a S^b)) \\ &\quad \times (2l+1)^{1/2} (2\bar{L}+1)^{1/2} \\ &\quad \times (-1)^{l^a + l^b + l + \lambda} \begin{pmatrix} l^a & l^b & l \\ s & t & \cdot \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ s+t & 0 & \cdot \end{pmatrix} \\ &\quad \times D(l^a 0 s \mid S^a)\, D(l^b 0 t \mid S^b) \\ &\quad \times D(\bar{L}, s+t, \bar{M} \mid R), \end{aligned} \quad (2.2)$$

where p has been written explicitly since it is summed over. Multiplying by $D(\bar{L}\tau\bar{M} \mid R)$ and integrating by R yields

$$\begin{aligned} 0 &= \sum_{l^a l^b l \lambda s} \psi(\bar{p}\bar{L}\bar{M}; l^a l^b l \lambda \mid \rho(S^a S^b)) \\ &\quad \times (2l+1)^{1/2} (2\bar{L}+1)^{1/2} \\ &\quad \times (-1)^{l^a + l^b + l + \lambda} \begin{pmatrix} l^a & l^b & l \\ s & \cdot & -\tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \\ &\quad \times D(l^a 0 s \mid S^a)\, D(l^b, 0, \tau - s \mid S^b). \end{aligned} \quad (2.3)$$

The equation above will be solved by a perturbation method. By Eq. (2.1) the expansion coefficients may be written in the more explicit form

$$\psi(\bar{p}\bar{L}\bar{M}; p \mid \rho) = \delta_{p\bar{p}} j(\bar{\lambda} \mid \bar{k}\rho) + b(p\bar{L}) h(\lambda \mid k\rho). \quad (2.4)$$

The coefficient $b(p\bar{L})$ is to be determined from the boundary condition (2.3). It should be noted that ψ is not a function of $\bar{M}$, while the variable $\bar{p}$ is to be understood. The boundary condition may now be written

$$\begin{aligned} 0 &= \sum_{l^a l^b l \lambda s} [\delta(l^a l^b l \lambda, \bar{l}^a \bar{l}^b \bar{l} \bar{\lambda}) j(\bar{\lambda} \mid \bar{k}\rho(S^a S^b)) \\ &\quad + b(l^a l^b l \lambda \bar{L}) h(\lambda \mid k\rho(S^a S^b))(2l+1)^{1/2} \\ &\quad \times (2\bar{L}+1)^{1/2} (-1)^{l^a + l^b + l + \lambda} \begin{pmatrix} l^a & l^b & l \\ s & \cdot & -\tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \\ &\quad \times D(l^a 0 s \mid S^a)\, D(l^b, 0, \tau - s \mid S^b). \end{aligned} \quad (2.5)$$

[5] Where no confusion is likely to arise, we use k to denote the wave number $k_{l^a l^b}$ [defined by $(2m)^{-1}\hbar^2 k^2 = E - E_{l^a} - E_{l^b}$] and $\bar{k}$ to denote $k_{\bar{l}^a \bar{l}^b}$.

[6] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), pp. 77–79, defines the spherical Bessel functions $j_l(x)$ and $h_l^{(1)}(x)$, which we write $j(l|x)$ and $h(l|x)$ for typographic reasons.

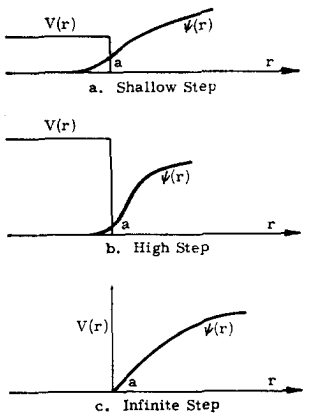


FIG. 2. Behavior of a wave function near the region of a jump in the potential.

Now let the assumption be made that the molecules are so nearly spherical that the deviation may be treated as a perturbation. Clearly, if the molecules were truly spherical the function ρ would be constant, so that it is valid to write

$$\rho(S^a S^b) = \rho_0 + \rho_1(S^a S^b) \quad (2.6)$$

and assert that ρ_0 is the dominant term.

In order to perform the perturbation let a parameter ϵ be introduced in powers of which the functions $h(\lambda \mid \cdots)$ and the coefficients b are to be expanded. Let

$$\rho(S^a S^b) = \rho_0 + \epsilon\rho_1(S^a S^b), \quad (2.6')$$

so that ϵ is a measure of the strength of the perturbation. Then, both the spherical Bessel and Hänkel functions may be written as series in ϵ, i.e.,

$$\begin{aligned} j(\lambda \mid k\rho(S^a S^b)) &= j(\lambda \mid k\rho_0) \\ &\quad + \epsilon k\rho_1(S^a S^b) j'(\lambda \mid k\rho_0) \end{aligned} \quad (2.7)$$

and

$$\begin{aligned} h(\lambda \mid k\rho(S^a S^b)) &= h(\lambda \mid k\rho_0) \\ &\quad + \epsilon k\rho_1(S^a S^b) h'(\lambda \mid k\rho_0), \end{aligned} \quad (2.8)$$

where the prime denotes differentiation. The validity of the first-order perturbation which is intended depends directly on the validity of cutting off the expansions Eqs. (2.7) and (2.8) after the second term.

Let also the coefficient b be expanded in powers of ϵ

$$b(p\bar{L}) = b^{(0)}(p\bar{L}) + \epsilon b^{(1)}(p\bar{L}) + \cdots. \quad (2.9)$$

Then, the boundary condition becomes

$$\begin{aligned} 0 &= \sum_{p s} \{\delta_{p\bar{p}}[j(\bar{\lambda} \mid \bar{k}\rho_0) + \epsilon\bar{k}\rho_1(S^a S^b) j'(\bar{\lambda} \mid \bar{k}\rho_0)] \\ &\quad + [b^{(0)}(p\bar{L}) + \epsilon b^{(1)}(p\bar{L}) + \cdots] \\ &\quad \times [h(\lambda \mid k\rho_0) + \epsilon k\rho_1(S^a S^b) h'(\lambda \mid k\rho_0)]\} \\ &\quad \times (2l+1)^{1/2} (2\bar{L}+1)^{1/2} (-1)^{l^a + l^b + l + \lambda} \\ &\quad \times \begin{pmatrix} l^a & l^b & l \\ s & \cdot & -\tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \\ &\quad \times D(l^a 0 s \mid S^a)\, D(l^b, 0, \tau - s \mid S^b). \end{aligned} \quad (2.10)$$

Collecting the coefficients of the powers of ϵ yields the following two equations for the perturbation up to the first order

$$0 = \sum_{p\bar{p}} \{\delta_{p\bar{p}} j(\bar{\lambda} \mid \bar{k}\rho_0) + h(\lambda \mid k\rho_0) b^{0}(p\bar{L})\} \times (2l+1)^{1/2}(2\bar{L}+1)^{1/2}(-1)^{l^a+l^b+l+\lambda} \times \begin{pmatrix} l^a & l^b & l \\ s & \cdot & \tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \times D(l^a 0 s \mid S^a)\, D(l^b, 0, \tau - s \mid S^b), \tag{2.11}$$

and

$$0 = \sum_{p\bar{p}} \{\delta_{p\bar{p}} \bar{k}\rho_1(S^aS^b) j(\bar{\lambda} \mid \bar{k}\rho_0) + b^{(0)}(p\bar{L}) k\rho_1(S^aS^b) h'(\lambda \mid k\rho_0) + b^{(1)}(p\bar{L}) h(\lambda \mid k\rho_0)\} \times (2l+1)^{1/2}(2\bar{L}+1)^{1/2}(-1)^{l^a+l^b+l+\lambda} \times \begin{pmatrix} l^a & l^b & l \\ s & \cdot & -\tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \times D(l^a 0 s \mid S^a)\, D(l^b, 0, \tau - s \mid S^b). \tag{2.12}$$

The first one, Eq. (2.11), may be solved quite readily by using the orthogonality relations for the Wigner coefficients and the irreducible representation coefficients, yielding the result that the term inside the braces must be zero, or that

$$b^{(0)}(p\bar{L}) = -\delta_{p\bar{p}} \frac{j(\bar{\lambda} \mid \bar{k}\rho_0)}{h(\bar{\lambda} \mid \bar{k}\rho_0)}. \tag{2.13}$$

The equation for $b^{(1)}(p\bar{L})$ is solved in a similar manner. Multiplying Eq. (2.12) by $D(l^{a'}0s' \mid S^a)$, $D(l^{b'}, 0, \tau - s' \mid S^b)$ and then integrating with respect to S^aS^b yields

$$\sum_{l\lambda} b^{(1)}(l^{a'}l^{b'}l\lambda\bar{L}) h(\lambda \mid k\rho_0)(-1)^{l^{a'}+l^{b'}+l+\lambda} \times \begin{pmatrix} l^{a'} & l^{b'} & l \\ s' & \cdot & -\tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} = -\sum_{\bar{s}} \{\bar{k} j'(\lambda' \mid \bar{k}\rho_0) + b^{(0)}(\bar{p}\bar{L}) \bar{k} h'(\lambda \mid \bar{k}\rho_0)\}(2\bar{l}+1)^{1/2}(2\bar{L}+1)^{1/2} \times (-1)^{\bar{l}^a+\bar{l}^b+\bar{l}+\bar{\lambda}} \begin{pmatrix} l^a & \bar{l}^b & \bar{l} \\ s & \cdot & -\tau \end{pmatrix} \begin{pmatrix} \bar{l} & \bar{\lambda} & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \times \int D(l^{a'}0s' \mid S^a)^*\, D(l^{b'}, 0, \tau - s' \mid S^b)^* \times \rho_1(S^aS^b)\, D(\bar{l}^a 0 s \mid S^a) \times D(\bar{l}^b, 0, \tau - s \mid S^b)\, dS^a\, dS^b.$$

Finally, multiplying by

$$(2\bar{L}+1)^{1/2}(2l'+1)^{1/2}(-1)^{\bar{l}+\lambda'+l^{a'}+l^{b'}} \times \begin{pmatrix} \bar{l} & \lambda' & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \begin{pmatrix} l^{a'} & l^{b'} & l' \\ s' & \cdot & -\tau \end{pmatrix}$$

and summing over s' and τ gives the result

$$b^{(1)}(l^al^bl\lambda\bar{L}) = -\frac{\bar{k}}{h(\lambda \mid k\rho_0)} \frac{(2l^a+1)(2l^b+1)}{(8\pi^2)^2} \frac{(2\lambda+1)}{(2\bar{L}+1)} \times \sum_{ss'\tau} (-1)^{l^a+l^b+l+\lambda+\bar{l}^a+\bar{l}^b+\bar{l}+\bar{\lambda}} \times [(2l+1)(2\bar{L}+1)^2(2\bar{l}+1)]^{1/2} \times \begin{pmatrix} l^a & l^b & l \\ s' & \cdot & -\tau \end{pmatrix} \begin{pmatrix} l & \lambda & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s & \cdot & -\tau \end{pmatrix} \begin{pmatrix} \bar{l} & \bar{\lambda} & \bar{L} \\ \tau & 0 & -\tau \end{pmatrix} \times \int D(l^a 0 s' \mid S^a)^* \times D(l^b, 0, \tau - s' \mid S^b)^* \rho_1(S^aS^b)\, D(\bar{l}^a 0 s \mid S^a) \times D(\bar{l}^b, 0, \tau - s \mid S^b)\, dS^a\, dS^b \times \left\{\frac{j'(\bar{\lambda} \mid \bar{k}\rho_0)h(\bar{\lambda} \mid \bar{k}\rho_0) - j(\bar{\lambda} \mid \bar{k}\rho_0)h'(\bar{\lambda} \mid \bar{k}\rho_0)}{h(\bar{\lambda} \mid \bar{k}\rho_0)}\right\}, \tag{2.14}$$

where the primes on $l^al^bl\lambda$ have been dropped.

The numerator of the term inside braces has the value[6]

$$-i/(\bar{k}^2\rho_0^2). \tag{2.15}$$

On the other hand, the integral cannot be evaluated until the form of $\rho_1(S^aS^b)$ is known. It is convenient to expand ρ_1 in a series of representation coefficients in a manner similar to that used previously[2] for the potential

$$\rho_1(S^aS^b) = \sum_{l_1 l_2 \mu_2} c(l_1 l_2 \mu_2)\, D(l_1 0 - \mu_2 \mid S^a) \times D(l_2 0 \mu_2 \mid S^b), \tag{2.16}$$

where the sum does not include the $c(000)$ term since normally any angle-independent terms would be included in ρ_0. Then, the following two integrals appear[4]:

$$\int D(l^a 0 s' \mid S^a)^* \times D(l_1 0 - \mu_2 \mid S^a)\, D(\bar{l}^a 0 s \mid S^a)\, dS^a = 8\pi^2 \begin{pmatrix} l_1 & \bar{l}^a & l^a \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & \bar{l}^a & l^a \\ -\mu_2 & s & \cdot \end{pmatrix} \delta(s', s - \mu_2), \tag{2.17}$$

and

$$\int D(l^b 0 \tau - s' \mid S^b)^* \times D(l_2 0 \mu_2 \mid S^b)\, D(\bar{l}^b 0 \tau - s \mid S^b)\, dS^b = 8\pi^2 \begin{pmatrix} l_2 & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_2 & \bar{l}^b & l^b \\ \mu_2 & \tau - s & \cdot \end{pmatrix} \delta(s', s - \mu_2). \tag{2.18}$$

The final result is

$$b^{(1)}(l^a l^b l\lambda L) = \frac{i(2l^a+1)(2l^b+1)[(2l+1)(2\bar{l}+1)]^{1/2}(2\lambda+1)}{\bar{k}\rho_0^2 h(\bar{\lambda} \mid \bar{k}\rho_0)h(\lambda \mid k\rho_0)} \times (-1)^{l+\lambda+\bar{l}+\bar{\lambda}} \sum_{l_1 l_2 \mu_2 s\tau} (-1)^{l_1+l_2+\tau} \times \begin{pmatrix} l_1 & \bar{l}^a & l^a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & l^a \\ -\mu_2 & s & \cdot \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ \mu_2 & \tau & -s & \cdot \end{pmatrix} \times \begin{pmatrix} l^a & l^b & l \\ s & -\mu_2 & \cdot & -\tau \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s & \cdot & -\tau \end{pmatrix}\begin{pmatrix} l & \lambda & L \\ \tau & 0 & \cdot \end{pmatrix}\begin{pmatrix} \bar{l} & \bar{\lambda} & L \\ \tau & 0 & \cdot \end{pmatrix} \times c(l_1 l_2 \mu_2). \tag{2.19}$$

We have thus obtained the wave function explicitly to first order.

3. CROSS SECTIONS

In the present section we use the results of the previous section to derive the form of the scattering cross sections, first for polarized beams and then for unpolarized beams. It will be shown that the zero-order contribution is simply the cross section for the scattering of rigid spheres with collision diameter ρ_0. The higher-order contributions include a correction to the elastic cross section which is due to the asphericity, as well as the inelastic cross section.

From Eq. (2.1) it may be seen that the part of $\psi(\bar{p}\bar{L}; p \mid r)$ referring to the scattered wave has the asymptotic form

$$b(\bar{p}\bar{L}p)i^{-\lambda-1}\, e^{ikr}/kr$$

whence, by comparison with Eq. (II-4.4), it is seen that

$$f(\bar{p}\bar{L}p) = i^{-\lambda-1}b(\bar{p}\bar{L}p)k^{-1}. \tag{3.1}$$

Now, since b has been written as a series in ϵ, let f be also, that is

$$f(\bar{p}\bar{L}p) = f^{(0)}(\bar{p}\bar{L}p) + \epsilon f^{(1)}(\bar{p}\bar{L}p) + \cdots. \tag{3.2}$$

Equation (2.13) shows that a factor of $\delta_{p\bar{p}}$ occurs in the expression for $f^{(0)}(\bar{p}\bar{L}p)$.

The series for the cross-section coefficients, Eq. (II-4.6), contains f in the form of factors $f(\bar{p}\bar{L}p)f(\bar{p}\bar{L}'p')^*$. Thus, from Eq. (3.2), the coefficients may be written as four terms, $\sigma^{(i)}$, $\sigma^{(ii)}$, $\sigma^{(iii)}$, and $\sigma^{(iv)}$, containing $f^{(0)}f^{(0)*}$, $f^{(1)}f^{(0)*}$, $f^{(0)}f^{(1)*}$, and $f^{(1)}f^{(1)*}$, respectively. From Eq. (II-4.6), and Eq. (2.13) the first of these, $\sigma^{(i)}$, may be written

$$\sigma^{(i)}(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b \mid \lambda_3\lambda_4\mu_4 \mid l^a m^a l^b m^b) = \frac{k}{\bar{k}}\frac{(2\bar{l}^a+1)(2\bar{l}^b+1)}{(2l^a+1)(2l^b+1)} \times \delta(\bar{l}^a\bar{l}^b, l^a l^b)(2\lambda_3+1)(2\lambda_4+1) \times \sum_{\substack{l\lambda l'\lambda' \\ LL'\mu_3}} i^{\lambda-\lambda'}(-1)^{\mu_4+m+\bar{m}}(2\lambda'+1)(2\lambda+1) \times (2L'+1)(2L+1)\begin{pmatrix} l & \lambda & L \\ \bar{m} & \cdot & -\mu_3 \end{pmatrix}\begin{pmatrix} l' & \lambda' & L' \\ \bar{m} & \cdot & \mu_4 & -\mu_3 \end{pmatrix} \times \begin{pmatrix} l & \lambda & L \\ m & \cdot & -\mu_3 \end{pmatrix}\begin{pmatrix} l' & \lambda' & L' \\ m & \cdot & \mu_4 & -\mu_3 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ \bar{m} & -\mu_3 & \cdot & \mu_4 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ \mu_3 & -m & \cdot & -\mu_4 \end{pmatrix} \times \frac{j(\lambda \mid \bar{k}\rho_0)j(\lambda' \mid \bar{k}\rho_0)}{\bar{k}^2 h(\lambda \mid \bar{k}\rho_0)h(\lambda' \mid \bar{k}\rho_0)^*}. \tag{3.3}$$

On examining this equation, it is seen[4] that the sum over L and L' is equal to $\delta_{m\bar{m}}$. Then, in a similar fashion, the sum over l and l' is seen to give $\delta_{m^a\bar{m}^a}$. The presence of the factor $\delta(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b,\ l^a m^a l^b m^b)$ shows that this part of the cross section only refers to elastic scattering, a result which could have been expected since $\sigma^{(i)}$ is a function only of the spherical part of the shape function.

The sum over μ_3 may be written

$$\sum_{\mu_3} [(2\lambda_3+1)(2\lambda_4+1)]^{1/2} \times \begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ m & -\mu_3 & \cdot & \mu_4 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ \mu_3 & -m & \cdot & -\mu_4 \end{pmatrix},$$

which, by use of symmetry and orthogonality, is seen to be equal to $\delta_{\lambda_3\lambda_4}$. From this result it is evident that the index μ_4 occurs in $\sigma^{(i)}$ only in the factor $(-1)^{\mu_4}$. Then, by analogy to the discussion in Sec. 5 of II, the sum over μ_4 in the cross section itself yields simply $D(\lambda_3 \mid RT^{-1})_{00}$. Then, when it is noted that $\bar{l}^a\bar{l}^b$ occur in $\sigma^{(i)}$ only in the wave number $\bar{k} = k(\bar{l}^a\bar{l}^b)$, it is evident that the cross section may be written

$$\sigma^{(i)}(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b T \mid l^a m^a l^b m^b R) = \delta(\bar{l}^a\bar{l}^b\bar{m}^a\bar{m}^b,\ l^a l^b m^a m^b) \times \sum_{\lambda_3} \sigma^{(i)}(\bar{k};\lambda_3)\, D(\lambda_3 \mid RT^{-1})_{00}, \tag{3.4}$$

where

$$\sigma^{(i)}(\bar{k};\lambda_3) = \bar{k}^{-2}\sum_{\lambda\lambda'}(2\lambda+1)(2\lambda'+1)(2\lambda_3+1) \times \begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \frac{j(\lambda \mid \bar{k}\rho_0)j(\lambda' \mid \bar{k}\rho_0)}{h(\lambda \mid \bar{k}\rho_0)h(\lambda' \mid \bar{k}\rho_0)^*}. \tag{3.5}$$

The significance of this may be appreciated when it is noted that Eq. (3.4) is just the usual expression

for the cross section for the scattering of rigid spheres of collision diameter ρ_0.[7]

The next two terms are complex conjugates of each other, so that only the first, $\sigma^{(ii)}$, need be considered. Since the factor $f^{(0)}$ is zero unless $l^a = \bar{l}^a$, $l^b = \bar{l}^b$, its presence introduces a factor of $\delta(l^a l^b, \bar{l}^a \bar{l}^b)$ in $\sigma^{(ii)}$. Similarly, on examining the result of substituting Eq. (3.2) into Eq. (II-4.4), one sees that the series[4]

$$\sum_{L'} (2L'+1)\begin{pmatrix} l' & \lambda' & L' \\ \bar{m} & \cdot & \mu_4 - \mu_3 \end{pmatrix}\begin{pmatrix} l' & \lambda' & L' \\ m & \cdot & \mu_4 - \mu_3 \end{pmatrix} = \delta_{m\bar{m}}$$

and

$$\sum_{l'} (2l'+1)\begin{pmatrix} \bar{l}^a & \bar{l}^b & l' \\ \bar{m}^a & \cdot & -\bar{m} \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l' \\ m^a & \cdot & -\bar{m} \end{pmatrix} = \delta_{m^a\bar{m}^a}$$

lead to a factor of $\delta(m^a m^b, \bar{m}^a \bar{m}^b)$. The end result is that

$$\begin{aligned}
&\sigma^{(ii)}(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b \mid \lambda_3\lambda_4\mu_4 \mid l^a m^a l^b m^b) \\
&= \frac{(-1)^{\mu_4+1}}{\bar{k}^4\rho_0^2}\,\delta(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b,\, l^a m^a l^b m^b) \\
&\times \sum_{\substack{l\bar{l}\lambda\bar{\lambda}\lambda' s \\ L\mu_3 l_1 l_2 \mu_2 \tau}} (2L+1)(2\bar{\lambda}+1)(2\lambda'+1)(2\lambda+1) \\
&\times (2\bar{l}+1)(2l+1)(2\lambda_3+1)(2\lambda_4+1)(2\bar{l}^a+1) \\
&\times (2\bar{l}^b+1)(-1)^{\tau}\begin{pmatrix} \bar{l} & \bar{\lambda} & L \\ \bar{m} & \cdot & -\mu_3 \end{pmatrix}\begin{pmatrix} l & \lambda & L \\ \bar{m} & \cdot & -\mu_3 \end{pmatrix} \\
&\times \begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ \bar{m}^a & \bar{m}^b & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{m}^a & \bar{m}^b & \cdot \end{pmatrix}\begin{pmatrix} \bar{\lambda} & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \bar{\lambda} & \lambda' & \lambda_3 \\ \bar{m} - \mu_3 & \cdot & \mu_4 \end{pmatrix} \\
&\times \begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ \mu_3 - \bar{m} & \cdot & -\mu_4 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & \bar{l}^a \\ 0 & 0 & 0 \end{pmatrix} \\
&\times \begin{pmatrix} l_2 & \bar{l}^b & \bar{l}^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & \bar{l}^a \\ -\mu_2 & s & \cdot \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & \bar{l}^b \\ \mu_2 & \tau - s & \cdot \end{pmatrix} \\
&\times \begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s & \cdot & -\tau \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ s - \mu_2 & \cdot & -\tau \end{pmatrix}\begin{pmatrix} l & \lambda & L \\ \tau & 0 & \cdot \end{pmatrix} \\
&\times \frac{j(\lambda' \mid \bar{k}\rho_0)c(l_1 l_2 \mu_2)}{h(\lambda' \mid \bar{k}\rho_0)^* h(\bar{\lambda} \mid \bar{k}\rho_0)h(\lambda \mid \bar{k}\rho_0)}\,,
\end{aligned} \tag{3.6}$$

and that $\sigma^{(iii)}$ is its complex conjugate. Finally, substituting Eq. (2.19) into Eq. (II-4.4) yields

$$\begin{aligned}
&\sigma^{(iv)}(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^b \mid \lambda_3\lambda_4\mu_4 \mid l^a m^a l^b m^b) \\
&= (-1)^{\mu_4+m+\bar{m}}\frac{(2l^a+1)(2\bar{l}^b+1)(2\bar{l}^a+1)(2\bar{l}^b+1)}{k\bar{k}} \\
&\times \sum_{\substack{\bar{l}\bar{\lambda}l\bar{\lambda}'\bar{\lambda}'\mu_2 \\ l_1 l_2 \mu_2 \tau L \\ l_1' l_2' \mu_2' s' \tau' L'}} \frac{(-1)^{l+l'+\bar{l}+\bar{l}'+\mu_2+\mu_2'+\tau+\tau'}c(l_1 l_2 \mu_2)c(l_1' l_2' \mu_2')}{h(\bar{\lambda} \mid \bar{k}\rho_0)h(\bar{\lambda}' \mid \bar{k}\rho_0)^* h(\lambda \mid k\rho_0)h(\lambda' \mid k\rho_0)^*} \\
&\times (2L+1)^2(2L'+1)^2(2l+1)(2l'+1)(2\bar{l}+1) \\
&\times (2\bar{l}'+1)(2\lambda+1)(2\lambda'+1)(2\bar{\lambda}+1)(2\bar{\lambda}'+1) \\
&\times (2\lambda_4+1)(2\lambda_3+1)\begin{pmatrix} \bar{l} & \bar{\lambda} & L \\ \bar{m} & \cdot & -\mu_3 \end{pmatrix} \\
&\times \begin{pmatrix} \bar{l}' & \bar{\lambda}' & L' \\ \bar{m} & \cdot & -\mu_3 \end{pmatrix}\begin{pmatrix} l & \lambda & L \\ m & \cdot & -\mu_3 \end{pmatrix}\begin{pmatrix} l' & \lambda' & L' \\ m & \cdot & -\mu_3 \end{pmatrix} \\
&\times \begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ \bar{m}^a & \bar{m}^b & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l}' \\ \bar{m}^a & \bar{m}^b & \cdot \end{pmatrix}\begin{pmatrix} l^a & l^b & l \\ m^a & m^b & \cdot \end{pmatrix}\begin{pmatrix} l^a & l^b & l' \\ m^a & m^b & \cdot \end{pmatrix} \\
&\times \begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_4 \\ \mu_3 - m & \cdot & -\mu_4 \end{pmatrix}\begin{pmatrix} \bar{\lambda} & \bar{\lambda}' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix} \\
&\times \begin{pmatrix} \bar{\lambda} & \bar{\lambda}' & \lambda_3 \\ \bar{m} - \mu_3 & \cdot & \mu_4 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & \bar{l}^a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1' & \bar{l}^a & l^a \\ 0 & 0 & 0 \end{pmatrix} \\
&\times \begin{pmatrix} l_2 & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_2' & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & l^a \\ -\mu_2 & s & \cdot \end{pmatrix}\begin{pmatrix} l_1' & \bar{l}^a & l^a \\ -\mu_2' & s' & \cdot \end{pmatrix} \\
&\times \begin{pmatrix} l_2 & \bar{l}^b & l^b \\ \mu_2 & \tau - s & \cdot \end{pmatrix}\begin{pmatrix} l_2' & \bar{l}^b & l^b \\ \mu_2' & \tau' - s' & \cdot \end{pmatrix}\begin{pmatrix} l^a & l^b & l \\ s - \mu_2 & \cdot & -\tau \end{pmatrix} \\
&\times \begin{pmatrix} l^a & l^b & l' \\ s' - \mu_2' & \cdot & -\tau' \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s & \tau - s & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l}' \\ s' & \tau' - s' & \cdot \end{pmatrix} \\
&\times \begin{pmatrix} l & \lambda & L \\ \tau & 0 & \cdot \end{pmatrix}\begin{pmatrix} l' & \lambda' & L' \\ \tau' & 0 & \cdot \end{pmatrix}\begin{pmatrix} \bar{l} & \bar{\lambda} & L \\ \tau & 0 & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}' & \bar{\lambda}' & L' \\ \tau' & 0 & \cdot \end{pmatrix}
\end{aligned} \tag{3.7}$$

for the final term $\sigma^{(iv)}$. This term is the only one which describes inelastic scattering.

The consideration of unpolarized beams does not change the first term, $\sigma^{(i)}$, for it is independent of m already. However, there appears in the second term the series,

$$\sum_{\bar{m}^a} (2l+1)\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{m}^a & \cdot & -\bar{m} \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ \bar{m}^a & \cdot & -\bar{m} \end{pmatrix} = \delta_{l\bar{l}}$$

and then, on changing variables from $\mu_3\bar{m}$ to $r = \bar{m} - \mu_3$, and $\bar{m}$, the series

$$\sum_{\bar{m}} (2\lambda+1)\begin{pmatrix} \bar{l} & \bar{\lambda} & L \\ \bar{m} & r & \cdot \end{pmatrix}\begin{pmatrix} \bar{l} & \lambda & L \\ \bar{m} & r & \cdot \end{pmatrix} = \delta_{\lambda\bar{\lambda}}$$

and

$$\sum_{r} (2\lambda_3+1)\begin{pmatrix} \bar{\lambda} & \lambda' & \lambda_3 \\ r & \cdot & \mu_4 \end{pmatrix}\begin{pmatrix} \bar{\lambda} & \lambda' & \lambda_4 \\ r & \cdot & \mu_4 \end{pmatrix} = \delta_{\lambda_3\lambda_4}.$$

Thus, $\sigma^{(iii)}$ may be written in the form

$$\begin{aligned}
&\sigma^{(iii)}(\bar{l}^a\bar{l}^b \mid \lambda_3\lambda_4\mu_4 \mid l^a l^b) = \delta(\bar{l}^a\bar{l}^b\lambda_3,\, l^a l^b\lambda_4) \\
&\times \frac{(-1)^{\mu_4+1}}{\bar{k}^4\rho_0^2} \sum_{\substack{l\lambda\lambda' L \\ l_1 l_2 \mu_2 s \tau}} (-1)^{\tau}(2\lambda+1)(2\lambda'+1) \\
&\times (2l+1)(2L+1)(2\lambda_3+1)(2\bar{l}^a+1)(2\bar{l}^b+1) \\
&\times \begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}^2\begin{pmatrix} l_1 & \bar{l}^a & \bar{l}^a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & \bar{l}^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & \bar{l}^a \\ -\mu_2 & s & \cdot \end{pmatrix} \\
&\times \begin{pmatrix} l_2 & \bar{l}^b & \bar{l}^b \\ \mu_2 & \tau - s & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ s - m_2 & \cdot & -\tau \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ s & \cdot & -\tau \end{pmatrix} \\
&\times \begin{pmatrix} l & \lambda & L \\ \tau & 0 & \cdot \end{pmatrix}^2 \frac{j(\lambda' \mid \bar{k}\rho_0)c(l_1 l_2 \mu_2)}{h(\lambda' \mid \bar{k}\rho_0) \mid h(\lambda \mid \bar{k}\rho_0)\mid^2}.
\end{aligned} \tag{3.8}$$

[7] L. I. Schiff, reference 6, p. 110.

The sum over L may be performed explicitly

$$\sum_L (2L+1)\begin{pmatrix} l & \lambda & L \\ \tau & 0 & \cdot \end{pmatrix}^2 = 1.$$

Then, there are three indices remaining, l, s, τ, which appear only in Wigner coefficients, and thus possibly may be summed over explicitly. The sum over these indices,

$$\sum_{l s \tau} (-1)^{\tau}(2l+1)\begin{pmatrix} l_1 & \bar{l}^a & \bar{l}^b \\ -\mu_2 & s & \cdot \end{pmatrix} \times \begin{pmatrix} l_2 & \bar{l}^b & \bar{l}^b \\ \mu_2 & \tau & -s & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ s & -\mu_2 & \cdot & -\tau \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & l \\ s & \cdot & -\tau \end{pmatrix},$$

on dropping factors which do not depend on l, s, or τ, and using the defining equation for the Racah coefficients,[4] becomes

$$\sum_{l s \tau} (-1)^{l}(2l+1)\begin{pmatrix} \bar{l}^a & \bar{l}^a & l_1 \\ \mu_2 & -s & s & \cdot \end{pmatrix}\begin{pmatrix} \bar{l}^b & \bar{l}^b & l_2 \\ \cdot & \tau & -s & \mu_2 \end{pmatrix} \times \sum_f (-1)^{\tau}\begin{pmatrix} \bar{l}^a & \bar{l}^a & f \\ \cdot & s & -\mu_2 \end{pmatrix} \times \begin{pmatrix} \bar{l}^b & \bar{l}^b & f \\ \cdot & s & -\tau & -\mu_2 \end{pmatrix}(2f+1)\begin{Bmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{l}^a & \bar{l}^b & f \end{Bmatrix}.$$

However, by orthogonality,[4] the sums over τ and s are $\delta_{l_1 f}$ and $\delta_{l_1 f}$, respectively. The sum over l is then only

$$\sum_l (-1)^l(2l+1)\begin{Bmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{l}^a & \bar{l}^b & f \end{Bmatrix}$$

which, on noting that

$$\begin{Bmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{l}^a & \bar{l}^b & 0 \end{Bmatrix} = \frac{(-1)^{l+\bar{l}^a+\bar{l}^b}}{[(2\bar{l}^a+1)(2\bar{l}^b+1)]^{1/2}}$$

is seen to be proportional to

$$\sum_l (2l+1)\begin{Bmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{l}^a & \bar{l}^b & 0 \end{Bmatrix}\begin{Bmatrix} \bar{l}^a & \bar{l}^b & l \\ \bar{l}^a & \bar{l}^b & f \end{Bmatrix} = \delta_{0f}$$

and thus is zero unless f is zero. However, if f is zero then l_1, l_2, and μ_2 must be zero, so that only the $c(000)$ term would appear in the series of Eq. (3.8). However, $c(000)$ was specifically excluded [see discussion following Eq. (2.16)] from the expansion of $\rho(S^a S^b)$, so that it follows that $\sigma^{(ii)}$ and $\sigma^{(iii)}$ for unpolarized beams are zero. Of course, it is possible that it might be more convenient to take the perturbation in such a way that $c(000)$ is not zero, in which case these two terms would serve as correction terms to the elastic scattering cross section.

The series for $\sigma^{(iv)}$ for unpolarized beams may be found by the substitution of Eq. (2.19) into Eq. (II-5.8). It is convenient to change the notation slightly and write

$$h(\lambda l^a l^b) = i^{\lambda+1}\frac{e^{ik\rho_0}}{k\rho_0}h(\lambda \mid k\rho_0). \tag{3.9}$$

Then the result is

$$\sigma^{(is)}(\bar{l}^a\bar{l}^b T \mid l^a l^b R) = \sum_{\lambda_3} \sigma(\bar{l}^a\bar{l}^b; l^a l^b \mid \lambda_3)\, D(\lambda_3 00 \mid RT^{-1}),$$

where

$$\begin{aligned}\sigma(\bar{l}^a\bar{l}^b; l^a l^b \mid \lambda_3) &= (2l^a+1)(2l^b+1) \\ &\times \frac{k}{\bar{k}} \sum_{\substack{L L' \lambda \lambda' \bar{\lambda} \bar{\lambda}' s s' \tau \tau' \\ l \bar{l} l_1 l_1' l_2 l_2' \mu_2 \mu_2'}} (-1)^{l+\bar{l}}(2L+1)(2L'+1) \\ &\times (2l+1)(2\bar{l}+1)(2\lambda+1)(2\lambda'+1) \\ &\times (2\bar{\lambda}+1)(2\bar{\lambda}'+1)(2\lambda_3+1) \\ &\times \frac{c(l_1 l_2 \mu_2)c(l_1' l_2' \mu_2')^*}{h(\lambda l^a l^b)h(\lambda' l^a l^b)^* h(\bar{\lambda}\bar{l}^a\bar{l}^b)h(\bar{\lambda}'\bar{l}^a\bar{l}^b)^*} \\ &\times \begin{pmatrix} l_1 & \bar{l}^a & l^a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1' & \bar{l}^a & l^a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_2' & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & \bar{l}^a & l^a \\ -\mu_2 & s & \cdot \end{pmatrix} \\ &\times \begin{pmatrix} l_1' & \bar{l}^a & l^a \\ -\mu_2' & s' & \cdot \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ \mu_2 & \tau & -s & \cdot \end{pmatrix}\begin{pmatrix} l_2' & \bar{l}^b & l^b \\ \mu_2' & \tau' & -s' & \cdot \end{pmatrix} \\ &\times \begin{pmatrix} l^a & l^b & l \\ s & -\mu_2 & \cdot & -\tau \end{pmatrix}\begin{pmatrix} l^a & l^b & l \\ s' & -\mu_2' & \cdot & -\tau' \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s & \cdot & -\tau \end{pmatrix} \\ &\times \begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s' & \cdot & -\tau' \end{pmatrix}\begin{pmatrix} l & \lambda & L \\ \tau & 0 & \cdot \end{pmatrix}\begin{pmatrix} l & \lambda' & L' \\ \tau' & 0 & \cdot \end{pmatrix}\begin{pmatrix} \bar{l} & \bar{\lambda} & L \\ \tau & 0 & \cdot \end{pmatrix} \\ &\times \begin{pmatrix} \bar{l} & \bar{\lambda}' & L' \\ \tau' & 0 & \cdot \end{pmatrix}\begin{pmatrix} \bar{\lambda} & \bar{\lambda}' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \begin{Bmatrix} L & \lambda & l \\ L' & \lambda' & \lambda_3 \end{Bmatrix}\begin{Bmatrix} L & \bar{\lambda} & l \\ L' & \bar{\lambda}' & \lambda_3 \end{Bmatrix}. \end{aligned} \tag{3.10}$$

This is the only term in the cross section if the collision is inelastic, and since such are the ones of major interest in this work, will be the one considered in the succeeding sections.

It may be argued that the expression above is little simpler than the corresponding one of Eq. (2.7) for unpolarized beams. The number of indices has been reduced by three, but it is still eighteen, and while the length of a single term has been reduced by some inches, quite a few remain. However, the properties of the Wigner and Racah coefficients are such that not all combinations of the indices lead to nonzero terms in the series. In fact, if all the indices on the left side of Eq. (3.10) are fairly small, the number of terms becomes reasonable. Then it is true that only $\lambda = 0, 1, 2$ are needed for the applications, and for these cases the Racah

functions have been tabulated. The Hänkel functions are not easy to calculate directly, but a recursion relation exists which is well suited to machine calculation.

There yet remains the question of the convergence of the series of Eq. (3.10). The first point to be noted is that it is an infinite series over at most the indices $\lambda\lambda'\bar{\lambda}\bar{\lambda}'$, since the triangle inequalities for the Wigner coefficients restrict all of the other summation indices to a finite range if $\bar{l}^a\bar{l}^b l^a l^b \lambda_3\lambda\lambda'\bar{\lambda}\bar{\lambda}'$ are given. This is readily checked. The convergence then depends on the asymptotic behavior of the $h(\lambda l^a l^b)$ for large λ. From the known formula[8]

$$H_p^{(1)}(z) \sim (iz)^{-1}\Gamma(p)(2^p/z),$$

it is evident that the behavior of $h(\lambda l^a l^b)$ is dominated by a double factorial $(2\lambda + 1)!!$ Since it is unlikely that the increase in the number of terms or that the growth of the Wigner coefficients in each term can match the rate of growth of a double factorial, it is reasonable that the series should converge. While the foregoing is not a rigorous proof, it contains the elements of one, and it should be noted that the proof of convergence can be made rigorous for the special case considered in Sec. 5.

4. SELECTION RULES

The present section is something of a digression from the main line of development, which leads to the calculation of cross sections. It contains a variety of qualitative results, some of which are general selection rules on the expansion coefficients of the distance-of-closest approach function, and some for the cases of quantum statistics and of homonuclear molecules.

The Distance-of-Closest Approach Function $\rho(S^aS^b)$

The function $\rho(S^aS^b)$ must be real (and in fact positive) and must exhibit the symmetry of the molecules under consideration. These requirements lead to corresponding restrictions on the expansion coefficients $c(l_1l_2\mu_2)$. The method of analysis is very similar to that employed for the parallel problem of the expansion of $V(rr^ar^bS^aS^b)$ in Sec. 6 of II,[2] so that only the results will be given here.

That the function $\rho(S^aS^b)$ must be real is expressed by

$$c(l_1l_2 - \mu_2)^* = c(l_1l_2\mu_2). \tag{4.1}$$

There does not seem to be any simple way to express the fact that ρ is positive, but under the present approximation ρ_0 dominates the succeeding terms so that the problem does not arise.

Suppose that "a" is a homonuclear molecule. This is expressed as

$$\rho(R^aR^{-1}, R^bR^{-1}) = \rho(R^{a'}R^{-1}, R^bR^{-1}) \tag{4.2}$$

(where by $R^{a'}$ is meant the rotation diametrically opposite to R^a), since this means that the potential is unchanged if the molecule "a" is turned end for end. The expression of Eq. (4.2) in terms of the series Eq. (2.16) and use of the fact that $D(l\ 0\ m \mid R') = (-1)^l D(l\ 0\ m \mid R)$ leads to the requirement that $c(l_1l_2\mu_2)$ is zero unless l_1 is even. Similarly, if "b" is homonuclear then $c(l_1l_2\mu_2)$ is zero unless l_2 is even. There is no restriction on μ_2 in either case.

If "a" and "b" are identical molecules then the potential must satisfy

$$\rho(R^aR^{-1}, R^bR^{-1}) = \rho(R^bR'^{-1}, R^aR'^{-1}), \tag{4.3}$$

that is, it must be unchanged if R^a and R^b are interchanged and R reversed. This leads to the condition

$$c(l_1l_2\mu_2) = (-1)^{l_1+l_2}c(l_2l_1\mu_2). \tag{4.4}$$

There is another set of restrictions on the $c(l_1l_2\mu_2)$ which limit the number that need be included in the series of Eq. (3.7) or (3.10) for any particular collision. In each series the indices occur in factors

$$\begin{pmatrix} l_1 & l^a & \bar{l}^a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_2 & l^b & \bar{l}^b \\ 0 & 0 & 0 \end{pmatrix}$$

so that l_1 must have the parity of $\bar{l}^a - l^a$ and l_2 that of $\bar{l}^b - l^b$. Furthermore, l_1 and l_2 must satisfy the triangle inequalities

$$|\bar{l}^a - l^a| \leq l_1 \leq \bar{l}^a + l^a \tag{45.a}$$

and

$$|\bar{l}^b - l^b| \leq l_2 \leq \bar{l}^b + l^b. \tag{4.5b}$$

Thus, the range of the l's is always finite, and is restricted to either even or odd. Furthermore, the index μ_2 occurs in such factors as

$$\begin{pmatrix} l_1 & \bar{l}^a & l^a \\ -\mu_2 & s & \cdot \end{pmatrix}\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ \mu_2 & \tau - s & \cdot \end{pmatrix}$$

so that its range is limited to

$$|\mu_2| \leq l_1, \qquad |\mu_2| \leq l_2. \tag{4.5c}$$

For the case where "a" is HD and "b" is H_2 these rules require the l_2 be even and that l^b change only by multiples of two. For the collision $\bar{l}^a = 0$,

[8] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, New York, 1952), Secs. 3.1 and 8.1. We are indebted to Professor A. Erdelyi for pointing this out and thus rescuing us from an exceedingly cumbersome direct proof.

$\bar{l}^b = 0$ to $l^a = 1$, $l^b = 0$ the only coefficient which appears in the series is $c(100)$.

Quantum Statistics

The foregoing theory would be incorrect if indistinguishable particles were being scattered, since the wave function for these must be symmetric or antisymmetric with respect to interchange of particles. The formal theory of the scattering of such particles was considered in I[1]; the relevant result being that while for distinguishable particles the cross section is given [see Eqs. (I-47) and (I-53)] as

$$\sigma(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^bT \mid l^am^al^bm^bR) = (k/\bar{k})\,|f(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^bT \mid l^am^al^bm^bR)|^2$$

for indistinguishable particles it is

$$\sigma(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^bT \mid l^am^al^bm^bR) = (k/\bar{k})\,|f(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^bT \mid l^am^al^bm^bR) + \epsilon f(\bar{l}^a\bar{m}^a\bar{l}^b\bar{m}^bT \mid l^bm^bl^am^aR')|^2,$$

where $\epsilon = +1$ for Bose-Einstein statistics and $\epsilon = -1$ for Fermi-Dirac.[9]

On referring to Eqs. (2.13) and (2.19), which contain the expression of $f(\bar{q}T \mid qR)$ for rigid molecules,[10] it may be seen that $f(\bar{q}T \mid q'R')$ need not be at all related to $f(\bar{q}T \mid qR)$. Thus, if a transition for which

$$\bar{l}^a = 0, \quad \bar{l}^b = 2, \quad l^a = 2, \quad l^b = 0$$

be considered, only the values

$$l_1 = 2, \quad l_2 = 2$$

will enter into the series for $f(\bar{q}T \mid qR)$, while on the other hand, only the values

$$l_1 = 0, \quad l_2 = 0, 2, 4$$

will enter into $f(\bar{q}T \mid q'R')$. Since there is no reason to expect $c(22\mu_2)$ to be related to $c(02\mu_2)$ for all molecules, it is doubtful that $f(\bar{q}T \mid qR)$ and $f(\bar{q}T \mid q'R)$ would be related.

On the other hand, if $\bar{l}^a = \bar{l}^b$ and the collision is inelastic, then precisely the same $c(l_1l_2\mu_2)$ occur in both series, as may be seen from Eq. (3.7). It is then at least possible that $f(\bar{q}T \mid qR)$ and $f(\bar{q}T \mid q'R')$ may be related. Reference to Eqs. (3.7) and (3.1), and use of group theoretical symmetry,[4] shows that $f(\bar{q}T \mid q'R')$ is the sum of a series whose terms are those of $f(\bar{q}T \mid qR)$ multiplied by $(-1)^{\bar{\lambda}}$. Thus, $f(\bar{q}T \mid qR) + \epsilon f(\bar{q}T \mid q'R')$ possesses a factor of $(1 + \epsilon(-1)^{\bar{\lambda}})$, and the cross section, Eq. (3.7), a factor of $(1 + \epsilon(-1)^{\bar{\lambda}})(1 + \epsilon(-1)^{\bar{\lambda}'})$. Thus, $\bar{\lambda}$ and $\bar{\lambda}'$ must be both even or both odd, depending on the statistics. Since a factor $\binom{\bar{\lambda}\bar{\lambda}'\lambda_3}{00\ 0}$ occurs, λ_3 can only be even.

The cross sections for unpolarized beams may be found as before, since these factors do not interfere with the summation over the m's. The result is that the cross section coefficients $\sigma(\bar{l}^a\bar{l}^b; l^al^b \mid \lambda_3)$ are as before except that a factor of

$$(1 + \epsilon(-1)^{\bar{\lambda}})(1 + \epsilon(-1)^{\bar{\lambda}'})$$

is to be inserted in each term of the series Eq. (3.10) for λ_3 even, and that the coefficient is zero for λ_3 odd. One result of this is that the cross section is the same for any angle as for the one diametrically opposite, a not unexpected result.

For this case the cross section is very easy to calculate if the cross section assuming Boltzmann statistics has already been evaluated, for one simply takes every other term in the series. The situation is very similar if instead $l^a = l^b$. There is no such trick method in the general case, but rather one has to evaluate the various series independently.

5. CALCULATION OF $\delta(\bar{l}^a\bar{l}^b; l^al^b \mid \lambda_3)$ FOR SPECIAL CASES

This section will be devoted to the setting up of the general calculational method, to the specialization to a specific case, and to the numerical calculation of this case for H_2-HD collisions.

Recursion Relation for the Hänkel Functions

The most convenient for the present purposes of the functions related to the Hänkel functions is

$$h(ll^al^b) = i^{l+1}\left(\frac{\pi}{2k\rho_0}\right)^{1/2} k\rho_0 e^{-ik\rho_0} H^{(1)}_{l+1/2}(k\rho_0), \qquad (5.1)$$

which may be expressed in terms of real functions as follows:

$$h(ll^al^b) = P_{l+1/2}(k\rho_0) + iQ_{l+1/2}(k\rho_0) \qquad (5.2)$$

where p and Q are functions defined in Jahnke-Emde.[11] These functions, $P_{l+1/2}(x)$ and $Q_{l+1/2}(x)$, are polynomials in x^{-1}. For simplicity, let them be

[9] No confusion should arise between this ϵ and the ϵ introduced in Eq. (2.6′) as the perturbation parameter.

[10] As before, q is an abbreviation for $l^am^al^bm^b$, while R' refers to a direction diametrically opposite to that of R. By q' we will mean q with a and b interchanged, that is, $l^bm^bl^am^a$.

[11] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1945).

TABLE I. The spherical Hänkel functions.

l	$Al(x)$	$Bl(x)$
0	1	0
1	1	x
2	$1 - 3x^2$	$3x$
3	$1 - 15x^2$	$6x - 15x^3$

denoted by $A_l = A_l(1/x)$ and $B_l = B_l(1/x)$. The values for the lowest orders are given in Table I.

Since $[i^{-l-1}h(ll^a l^b)]$ is a multiple of a Hänkel function, it must satisfy the same recursion relation,[9] which, in terms of the A_l and B_l, becomes

$$A_{l+1}(x) = A_{l-1} - (2l + 1)xB_l(x), \qquad (5.3a)$$

$$B_{l+1}(x) = B_{l-1} + (2l + 1)xA_l(x). \qquad (5.3b)$$

Since A_0, A_1, B_0, and B_1 are known from Table I, this relation serves to define all succeeding A_l and B. It is ideally suited to machine calculation.

The Cross Section When $l^b = \bar{l}^b = 0$

It is desired in the next section to calculate $\sigma(0010; \lambda_3)$. However, it is of interest to first find $\sigma(\bar{l}^a 0 l^a 0; \lambda_3)$. The following factors[4] occur in the terms of Eq. (3.10):

$$\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} l_2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \delta_{l_2,0},$$

$$\begin{pmatrix} l_2' & \bar{l}^b & l^b \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} l_2' & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \delta_{l_2',0},$$

$$\begin{pmatrix} l_2 & \bar{l}^b & l^b \\ \mu_2 & \tau & -s & \cdot \end{pmatrix} = \begin{pmatrix} l_2 & 0 & 0 \\ \mu_2 & \tau & -s & \cdot \end{pmatrix} = \delta_{0,l_2}\,\delta_{0,\mu_2}\,\delta_{\tau,s},$$

and

$$\begin{pmatrix} l_2' & \bar{l}^b & l^b \\ \mu_2' & \tau' & -s' & \cdot \end{pmatrix} = \begin{pmatrix} l_2' & 0 & 0 \\ \mu_2' & \tau' & -s' & \cdot \end{pmatrix} = \delta_{0,l_2'}\,\delta_{0,\mu_2'}\,\delta_{\tau',s'}.$$

Clearly, only the coefficients of the form $c(l_1 00)$ enter into the cross section.

It is easy to see that if one molecule, say "b," is spherical then only coefficients of the form $c(l_1 00)$ are nonzero, and that the cross section is of the form

$$\sigma(\bar{l}^a \bar{l}^b l^a l^b; \lambda_3) = \delta(\bar{l}^b l^b)\sigma(\bar{l}^a 0 l^a 0; \lambda_3). \qquad (5.4)$$

Thus, the particular cross section $\sigma(\bar{l}^a 0 l^a 0 : \lambda_3)$ of the previous paragraph, which is for a general shape function as given by Eq. (2.16), is rigorously equal to the cross section for the special shape function with $l_2 = \mu_2 = 0$.

To return to Eq. (3.10), the remaining factors involving l^b and $\bar{l}^b$ are

$$\begin{pmatrix} l^a & l^b & l \\ s & -\mu_2 & \cdot & -\tau \end{pmatrix} = (2l + 1)^{-1/2}(-1)^{l+\tau}\,\delta_{\tau,s-\mu_2},$$

$$\begin{pmatrix} \bar{l}^a & \bar{l}^b & \bar{l} \\ s & \cdot & -\tau \end{pmatrix} = (2\bar{l} + 1)^{-1/2}(-1)^{\bar{l}+\tau}\,\delta_{\tau,s},$$

and two more with some indices primed. The series then is no longer summed over l_2, l_2', μ_2, μ_2', s, s', l, and $\bar{l}$, a gain of eight indices. Substitution of the above into Eq. (3.10) yields

$$\begin{aligned}\sigma(\bar{l}^a 0 l^a 0; \lambda_3) = {} & \frac{k}{\bar{k}} \sum_{\substack{LL'\lambda\lambda'\bar{\lambda}\bar{\lambda}' \\ l_1 l_1' \tau\tau'}} (-1)^{l^a+\bar{l}^b+l_1+l_1'} \\ & \times (2L + 1)(2L' + 1)(2\lambda + 1)(2\lambda' + 1) \\ & \times (2\bar{\lambda} + 1)(2\bar{\lambda}' + 1)(2\lambda_3 + 1) \\ & \times \frac{c(l_1 00)c(l_1' 00)^*}{h(\lambda l^a 0)h(\lambda' l^a 0)^* h(\bar{\lambda}\bar{l}^a 0)h(\bar{\lambda}'\bar{l}^a 0)^*} \\ & \times \begin{pmatrix} l_1 & \bar{l}^a & l^a \\ 0 & \tau & -\tau \end{pmatrix}\begin{pmatrix} l_1' & \bar{l}^a & l^a \\ 0 & \tau' & -\tau' \end{pmatrix}\begin{pmatrix} l^a & \lambda & L \\ \tau & 0 & -\tau \end{pmatrix} \\ & \times \begin{pmatrix} \bar{l}^a & \bar{\lambda} & L \\ \tau & 0 & -\tau \end{pmatrix}\begin{pmatrix} l^a & \lambda' & L' \\ \tau' & 0 & -\tau' \end{pmatrix}\begin{pmatrix} \bar{l}^a & \bar{\lambda}' & L' \\ \tau' & 0 & -\tau' \end{pmatrix} \\ & \times \begin{pmatrix} \bar{\lambda} & \bar{\lambda}' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} L & \lambda & l^a \\ L' & \lambda' & \lambda_3 \end{Bmatrix}\begin{Bmatrix} L & \bar{\lambda} & \bar{l}^a \\ L' & \bar{\lambda}' & \lambda_3 \end{Bmatrix}. \end{aligned} \qquad (5.5)$$

The sum over L, L', τ, and τ' involves only quantities arising from the group theory, and it is believed that it can be expressed in terms of higher-order invariants than the Racah functions. However, it is not necessary to attempt to do so, since it is simple to perform the summation directly for small values of l^a and $\bar{l}^a$.

The Cross Section $\sigma(0010; \lambda_3)$

The present part will be devoted to the calculation of the cross-section coefficients $\sigma(0010; \lambda_3)$ for $\lambda_3 = 0, 1, 2$. It should be noted that Eq. (5.5) is for a general shape function, that is, while the $\rho(S^a S^b)$ that appears there is averaged over orientations of molecule "b," this is not an approximation, but arises from the theory. In a similar fashion, it will be seen that in the cross section here desired only the term $c(100)$ of the shape function will appear. Thus, on the one hand it is possible to calculate the cross section for this collision with only very limited knowledge of the shape function, while on the other hand knowledge of the cross section by no means serves to determine the shape function.

The following factors occur in the series of Eq. (5.5) on substituting $\bar{l}^a = 0$ and $l^a = 1$:

$$\begin{pmatrix} l_1 & \bar{l}^a & l^a \\ 0 & \tau & -\tau \end{pmatrix} = (-1)^{l_1+1}\begin{pmatrix} l_1 & 1 & 0 \\ 0 & -\tau & \tau \end{pmatrix} = -(3)^{-1/2}\,\delta_{l_1,1}\,\delta_{0,\tau}$$

$$\begin{pmatrix} l_1' & \bar{l}^a & l^a \\ 0 & \tau' & -\tau' \end{pmatrix} = -(3)^{-1/2}\,\delta_{l_1,1}\,\delta_{\tau,0}$$

$$\begin{pmatrix} \bar{l}^a & \bar{\lambda} & L \\ \tau & 0 & -\tau \end{pmatrix} = \begin{pmatrix} \bar{\lambda} & L & 0 \\ \tau & 0 & -\tau \end{pmatrix} = (-1)^L(2L+1)^{-1/2}\,\delta_{L,\bar{\lambda}}\,\delta_{0,\tau}$$

and

$$\begin{Bmatrix} L & \bar{\lambda} & \bar{l}^a \\ L' & \bar{\lambda}' & \lambda_3 \end{Bmatrix} = \begin{Bmatrix} L & \bar{\lambda} & 0 \\ L' & \bar{\lambda}' & \lambda_3 \end{Bmatrix} = (2L+1)^{-1/2}(2L'+1)^{-1/2}\,\delta_{L,\bar{\lambda}}\,\delta_{L',\bar{\lambda}'}.$$

Substitution into Eq. (5.5) yields the result

$$\sigma(0010;\lambda_3) = \frac{|c(100)|^2\,k_{10}}{9k_{00}} \times \sum_{LL'\lambda\lambda'} F_{\lambda_3}(L\lambda L'\lambda')H(L\lambda L'\lambda')^{-1}, \tag{5.6}$$

where

$$F_{\lambda_3}(L\lambda L'\lambda') = (-1)^{\lambda_3}(3)(2L+1)(2L'+1)(2\lambda+1) \times (2\lambda'+1)(2\lambda_3+1)\begin{pmatrix} L & \lambda & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} L' & \lambda' & 1 \\ 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} L & L' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & \lambda_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} L & \lambda & 1 \\ L' & \lambda' & \lambda_3 \end{Bmatrix}, \tag{5.7}$$

and

$$H(L\lambda L'\lambda') = h(\lambda 10)h(\lambda' 10)^*h(L00)h(L'00)^*. \tag{5.8}$$

The further reduction of the series is a straightforward but very tedious process, so that only the results will be given here. One case, that of the calculation of $\sigma(0010;1)$, which serves to illustrate the methods to be used, is examined in detail in the Appendix. The first coefficient reduces to two infinite series[12]

$$\sigma(0010;0) = \frac{|c(100)|^2\,k_{10}}{k_{00}}\left[\sum_n \frac{n+1}{H(n+1n\,n+1n)} + \sum_n \frac{n}{H(nn+1\,nn+1)}\right]. \tag{5.9}$$

The second coefficient is conveniently divided into three series

$$\sigma(0010;1) = \frac{|c(100)|^2\,k_{10}}{k_{00}} \times \left[\sum_n \frac{n(n+1)}{(2n+1)}\{H(nn-1\,n+1n)^{-1} + H(n+1n\,nn-1)^{-1}\} + \sum_n \frac{n(n+1)}{(2n+1)}\{H(n-1n\,nn+1)^{-1} + H(nn+1\,n-1n)^{-1}\} + \sum_n \frac{(n+1)}{(2n+1)(2n+3)}\{H(n+1n\,nn+1)^{-1} + H(nn+1\,n+1n)^{-1}\}\right]. \tag{5.10}$$

In a similar fashion the third coefficient may be considered as three series

$$\sigma(0010;2) = \frac{|c(100)|^2\,k_{10}}{k_{00}}\left[45\sum_n \frac{n(n+1)(n+2)}{(2n+1)(2n+3)} \times \{H(n-1n\,n+1\,n+2)^{-1} + H(n+1\,n+2\,n-1n)^{-1} + H(nn-1\,n+2\,n+1)^{-1} + H(n+2\,n+1\,nn-1)^{-1}\} + 30\sum_n \frac{n(n+1)(n+2)}{(2n+1)(2n+3)} \times \{H(nn+1\,nn+1)^{-1} + H(n+1n\,n+1n)^{-1}\} + 90\sum_n \frac{n(n+1)}{(2n-1)(2n+1)(2n+3)} \times \{H(nn+1\,nn-1)^{-1} + H(nn-1\,nn+1)^{-1} + H(n-1nn+1n)^{-1} + H(nn-1\,nn+1)^{-1}\}\right]. \tag{5.11}$$

It should be noted that each pair of braces contains along with a product $H(L\lambda L'\lambda')$ its complex conjugate $H(L'\lambda' L\lambda)$, so the coefficients themselves are real, as they should be.

H_2-HD Collisions

While the calculation of the cross-section coefficients in terms of $c(100)$ and ρ_0 is exact and general, values of the parameters must be chosen in order to go on to numerical calculation. For these preliminary calculations the pair molecular hydrogen-hydrogen deuteride was chosen. Hydrogen, molecule "a," is taken to be spherical with diameter 2.93×10^{-8} cm, and interatomic distance 0.749×10^{-8} cm. Then hydrogen deuteride, mole-

[12] In all cases the summation index ranges over all non-negative integers for which the summand is defined.

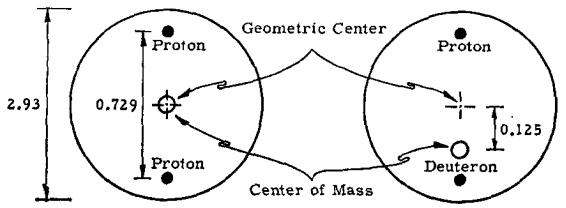


FIG. 3. Geometry of H_2 and HD (distances are in units of 10^{-8} cm).

cule "b," is exactly the same except that the center of mass is displaced from the center of the sphere by an amount which may be calculated to be 0.125×10^{-8} cm, as in Fig. 3.

The first two coefficients of the distance of closest approach function were estimated in the following manner. First ρ_0 was taken to be

$$\rho_0 = 2.93 \times 10^{-8} \text{ cm}. \tag{5.12}$$

Then $c(100)$ was determined by fitting the function at the orientations where it takes on its maximum and its minimum value, to give

$$c(100) = 0.125 \times 10^{-8} \text{ cm}. \tag{5.13}$$

The function was thus taken to be

$$\rho(S^a S^b) = 2.93 \times 10^{-8} + 0.125 \times 10^{-8} \, \mathrm{D}(100 \mid S^a). \tag{5.14}$$

The rotational energy of hydrogen deuteride was taken to be $l(l+1)63.7k$, where k is the Boltzmann constant. Then if $T_0 k$ be the incoming energy in relative coordinates and $T_1 k$ the outgoing, they are related by

$$T_0 = T_1 + 2 \cdot 63.7$$

since we taking the final state to be the $l = 1$ state.

The calculation of the cross section coefficients was performed on the University of Wisconsin Numerical Analysis Laboratory's CPC-II, a digital machine. The calculation was so programmed that the functions $h(\lambda l^a l^b)$ and $h(\bar{\lambda} \bar{l}^a \bar{l}^b)$ were calculated by the recursion relation of Eq. (5.3), and the values λ, λ', L, L' kept in storage at each stage in the calculation. Then the product of four spherical Hänkel functions was formed, its reciprocal taken, and it was multiplied by the factor $F_{\lambda_s}(L\lambda L'\lambda')$. The terms so obtained were caused to print out and were later summed manually, because of the machine's limited storage capacity.

The results of the calculation are given in Fig. 4 and Table II as a function of the initial relative energy $T_0 k$. The cross sections are given in units of 10^{-16} cm^2, or square angstroms. The total transition cross section is the integral over all angles of the differential cross section, thus $4\pi\sigma(0010; 0)$. The positive $\sigma(0010; 1)$ indicates more intense forward scattering than back, while the positive $\sigma(0010: 2)$ increases the intensity fore and aft and decreases it to the sides.

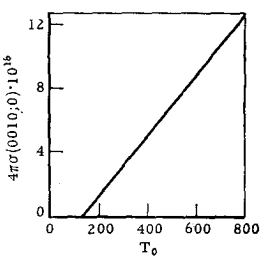


FIG. 4. The total collision cross section for the (0010) transition of hydrogen deuteride and molecular hydrogen.

It is interesting to compare the value of the transition cross section calculated above with the geometric cross section

$$\pi\rho_0^2 = 27 \times 10^{-16} \text{ cm}^2$$

treating the bodies as spheres. The transition cross section is zero below the threshold, while the elastic scattering cross section varies from the above value at high energies to four times that at nearly zero energy. Assuming no other transitions occur, the ratio of the transition cross sections to the total cross section is the fraction of the numbers of collisions which result in a transition. These ratios (neglecting other transitions) are approximately one in one hundred, one in forty, one in ten, and one in three in order of increasing energy.

TABLE II. Cross section for the HD-H_2 collision.

T_0	T_1	$4\pi\sigma(0010; 0)\cdot10^{16}$	$\sigma(0010; 1)\cdot10^{16}$	$\sigma(0010; 2)\cdot10^{16}$
140	12.5	0.169	0.148	0.244
170	42.5	0.679	0.087	1.112
300	172.5	3.091	0.292	5.428
800	672.5	11.519	0.941	21.972

6. DISCUSSION

There are three approximations in the foregoing work. The first is the rigid body model itself, the second, the first-order perturbation in the deviation from sphericity, and the third the truncation of the infinite series of Eq. (3.10). The last is a negligible source of error, since the series converges very rapidly. One may hope that the first approximation

is not unreasonable, since the body of experience with the rigid sphere model has shown it to be extremely valuable for preliminary, order-of-magnitude calculations. On the other hand the perturbation is of an unfamiliar sort, on the boundary condition rather than on some parameter in the potential, so that little of the previous experience with perturbation theory is applicable.

It was our hope, when this work was begun, to relate the results to experiment by applying the cross sections to the calculation of the transport properties of gases under conditions such that inelastic collisions are a significant fraction of the total. Snider[13] gives a good set of references to the theory of such transport properties. Unfortunately, it has not yet been possible to carry the calculations through to that stage, and further Snider has shown that in addition to the cross sections, which are essentially absolute squares of the scattering matrix, there are also needed products of nearby elements. The relation to experiment may be simpler in the study of NMR relaxation in gases. The $\Delta l = 0$, $\Delta m \neq 0$ cross sections for H_2-H_2 collisions are being calculated at the present time for comparison with Bloom's[14] experimental measurements using NMR relaxation.

Finally, there is always the hope that a direct measurement with molecular beams would be possible. For molecules with as widely spaced rotational levels as in H_2 and HD, there is a great difference in velocity between elastically and inelastically scattered molecules. Thus, even a crude velocity selector could differentiate between molecules scattered into the various rotational states.

ACKNOWLEDGMENTS

We wish to acknowledge the assistance of Mrs. Eva Perlman Silversmith in checking the sort of algebra shown in the Appendix and of Mrs. Helen MacDuffee Ward in coding and performing the numerical calculations.

APPENDIX. ALGEBRAIC REDUCTION OF THE SERIES FOR $\sigma(0010; 1)$

The algebraic reductions involved in the calculation of the cross-section coefficient $\sigma(0010; 1)$ will here be carried out in full as an illustration of procedures which may be used in general. The coefficient is given by Eq. (5.6) as a series over four indices. However, the ranges of the indices are such that effectively one has several series over one index. Then for each series the general term, which is expressed in terms of Racah and Wigner coefficients, may be calculated by use of the existing tables.[15]

TABLE III. Allowed values of the indices in $F_1(L\lambda L'\lambda')$.

	L	λ	L'	λ'
i	n	$n+1$	$n+1$	n
ii	n	$n-1$	$n+1$	n
iii	n	$n+1$	$n+1$	$n+2$
iv	n	$n+1$	$n-1$	n
v	n	$n-1$	$n-1$	n
vi	n	$n-1$	$n-1$	$n-2$

From Eqs. (5.6) and (5.7) it may be seen that

$$\sigma(0010; 1) = \frac{|c(100)|^2 k_{10}}{9k_{00}} \sum_{LL'\lambda\lambda'} \frac{F_1(L\lambda L'\lambda')}{H(L\lambda L'\lambda')}, \qquad \text{(A1)}$$

where

$$F_1(L\lambda L'\lambda') = 9(2L+1)(2L'+1)(2\lambda+1) \times (2\lambda'+1)\begin{pmatrix} L & \lambda & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} L' & \lambda' & 1 \\ 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} L & L' & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \lambda & \lambda' & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} L & \lambda & 1 \\ L' & \lambda' & 1 \end{Bmatrix}. \qquad \text{(A2)}$$

Not all values of the indices are allowed, since the Wigner coefficients above are zero unless each of the following relations holds:

$$L - \lambda = \pm 1,$$
$$L' - \lambda' = \pm 1,$$
$$L - L' = \pm 1,$$
$$\lambda - \lambda' = \pm 1.$$

Table III lists the values the other indices can take if L is given the value n.

The coefficients $F_1(L\lambda L'\lambda')$ are invariant under certain permutations of their indices. The use of the symmetries of the Racah functions leads to the result that

$$F_1(L\lambda L'\lambda') = F_1(\lambda L\lambda' L') = F_1(L'\lambda' L\lambda) = F_1(\lambda' L'\lambda L).$$

Then the relation

$$F_1(nn-1\,n-1n) = F_1(n-1n\,nn-1)$$

shows that case (*vi*) in Table III may be expressed in terms of case (*ii*). Similarly from

$$F_1(nn+1\,n+1\,n+2) = F_1(n+1\,nn+2\,n+1),$$
$$F_1(nn+1\,n-1n) = F_1(nn-1\,n+1n),$$

[13] R. F. Snider, J. Chem. Phys. **32**, 1051 (1960).
[14] M. Bloom, Physica **23**, 237 (1957).
[15] L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Revs. Modern Phys. **24**, 249 (1952).

and

$$F_1(nn-1\,n-1\,n-2) = F_1(n-1\,n-2nn-1)$$

cases (*iii*), (*iv*), and (*v*) may be expressed in terms of case (*i*). The cross section coefficient may then be written in the form

$$\begin{aligned}\sigma(0010;1) = {} & \frac{|c(100)|^2\, k_{10}}{9k_{00}} \sum_n \{F_1(nn+1\,n+1n) \\ & \times [H(nn+1\,n+1n)^{-1} + H(n+1n\,nn+1)^{-1}] \\ & + F_1(nn-1\,n+1n)[H(nn-1\,n+1n)^{-1} \\ & + H(n-1n\,nn+1)^{-1} + H(nn+1\,n-1n)^{-1} \\ & + H(n+1n\,nn-1)^{-1}]\} \end{aligned} \tag{A3}$$

if n is replaced by $n-1$ in case (*iii*) and by $n+1$ in cases (*v*) and (*vi*).

Then only two coefficients, $F_1(nn+1n+1n)$ and $F_1(nn-1n+1n)$ need be calculated. For the first coefficient, one notes that[5]

$$\begin{pmatrix} n & n+1 & 1 \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{n+1}{(2n+1)(2n+3)} \tag{A4}$$

and

$$\begin{Bmatrix} n & n+1 & 1 \\ n+1 & n & 1 \end{Bmatrix} = \frac{1}{(n+1)(2n+1)(2n+3)} \tag{A5}$$

so that

$$F_1(nn+1\,n+1n) = \frac{9(n+1)}{(2n+1)(2n+3)}. \tag{A6}$$

Similarly, for the second coefficient one needs in addition to Eq. (A4) the following:

$$\begin{pmatrix} n\,n-1 & 1 \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{n}{(2n-1)(2n+1)} \tag{A7}$$

and

$$\begin{Bmatrix} n & n-1 & 1 \\ n+1 & 1 & 1 \end{Bmatrix} = \frac{1}{2n+1} \tag{A8}$$

to show that

$$F_1(nn-1\,n+1n) = 9\,\frac{n(n+1)}{(2n+1)}. \tag{A9}$$

A substitution of these two results into (A3) yields Eq. (5.10).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Quantization of Fields with Infinite-Dimensional Invariance Groups. III. Generalized Schwinger-Feynman Theory*

BRYCE S. DEWITT
Institute of Field Physics, Department of Physics, University of North Carolina, Chapel Hill, North Carolina
(Received May 18, 1962)

The formal methods of Schwinger and Feynman are applied to nonlinear field theories having elementary vertex functions of arbitrarily high order. In the first half of the paper, familiar theorems are rederived by noncanonical methods. Emphasis is given to purely formal aspects of the theory which may be expected to survive generalization to situations in which standard asymptotic conditions are inapplicable. Since the context in which the field nonlinearities are assumed to appear is that of a non-Abelian infinite-dimensional invariance group, detailed attention is given to the question of a group invariant measure for the Feynman functional integral. It is shown that the physically important part of the measure is *not* determined by the group.

The second half of the paper is devoted to the theory of the propagators and correlation functions which characterize the system when invariant variables are introduced. The existence of a *c-number* action functional Γ which contains a complete description of all quantum processes is proved. The second variational derivatives of this functional constitute the wave operator for the one-particle propagators (including all radiative corrections), and its higher derivatives are the renormalized vertex functions. A description of the renormalization process is easily carried out in terms of Γ. Finally, the implications which its existence has for quantum gravidynamics are discussed. Because it leads to nonlocal covariant equations for a *complex* metric tensor the way is open to transmutations of topology at the quantum level.

INTRODUCTION

CONTINUING a series of investigations[1] on the special problems which arise in the attempt to quantize fields possessing infinite-dimensional invariance groups, we here leave the confines of the "quasi-classical approximation" to which we previously limited ourselves and confront the quantization problem proper. One of the chief troublemaking situations which immediately presents itself is the strong nonlinearity of the dynamical equations which are encountered when the invariance group is non-Abelian. Indeed, in the particularly interesting case of quantum gravidynamics, elementary vertex functions of arbitrarily high order occur. A major purpose of the present paper is to show that in spite of its formidable appearance (and provided no unexpected difficulties arise later in the detailed execution of the renormalization program) this situation can be smoothly incorporated into appropriate generalizations of standard formal procedures.

It is necessary to point out, however, that our derivations will not be entirely deductive. We do not start with a set of postulates from which everything else follows and which guarantee in advance that all difficulties will automatically resolve themselves in some suitable fashion. Rather, we proceed somewhat in the spirit of the transition which led from the old quantum theory to modern quantum mechanics. The "theory" which begins to emerge here has a certain inevitability and inner formal logic about it to which we give full rein.

Much of the material contained in the sections to follow will be familiar to specialists. Its restatement here serves two purposes. First, in order to achieve the greatest possible generality we continue our total boycott of the canonical formalism, and therefore it is useful to show how known theorems follow from a definition of commutators which is based solely on the uncertainty principle and the theory of measurement.[2] Secondly, the broadened context in which the derivations are carried out serves to illumine certain facets of familiar results which do not usually suggest themselves. Although little explicit mention is made of the metric tensor in this paper, the chief new element of context we have in mind is the admission of geometry as an object of quantization, which will ultimately force a generalization of conventional asymptotic conditions so as to take into account the possibility of space-time itself having unusual topology either macroscopically or microscopically. The very gen-

* This research was supported in part by the Department of the Navy, Office of Naval Research, under contract Nonr-855(07) and in part by the Air Force Office of Scientific Research under contract AFOSR 61-72.

[1] B. S. DeWitt, J. Math. Phys. 2, 151 (1961); 3, 625 (1962). These two papers will be referred to as (I) and (II), respectively.

[2] B. S. DeWitt, J. Math. Phys. 3, 619 (1962).

erality of the abstract formalism employed here not only suggests alternative characterizations of boundary conditions for local theories but possibly provides a reasonable framework for the quantization of nonlocal theories as well. The only thing definitely required of the dynamical systems under consideration is that they be describable by means of action functionals S.

In addition to generalizing familiar material, we also obtain some new results. Chief among these is the proof of the existence of a complex nonlocal c-number action functional Γ which contains a complete description of all quantum processes. Loosely speaking, Γ is to the quantum theory what S is to the classical theory, and its introduction constitutes a complete transition from conventional quantum mechanics to an independent quantum theory of fields. This is not to suggest that alternative transitions may not be possible or even desirable, for the present formalism focuses on the fields rather than on the Hilbert space and hence requires the usual infinite renormalizations. Unfortunately, dispersion theoretical techniques are not immediately available here; the usual postulates (locality, causality, Lorentz invariance) on which they are based break down in the face of a quantized metric tensor. On the other hand, recognition of the existence of the functional Γ simplifies the discussion of the renormalization process and, in the case of quantum gravidynamics, has far reaching implications which will be mentioned at the end of the paper.

In Sec. I the Schwinger variational principle for transition amplitudes is derived from the generalized Peierls theory of commutators,[1-3] and T functionals are defined by the familiar device of varying external sources. In Sec. 2 the method of functional Fourier transforms is used to introduce the Feynman functional integral. Because of its relevance to questions of group invariance, unitarity, and factor ordering, attention is given in Sec. 3 to the problem of the "measure" which is to be associated with the functional volume element. The most important conclusion reached is that although group invariance is easily achieved the physically relevant part of the measure is *not* determined by the group. In Sec. 4 the basic theory of propagators and correlation functions, and their relation to T functionals, is outlined. The existence of the functional Γ is proved in Sec. 5 and its relation to the correlation functions is established. The self-energy operator is introduced and the renormalization process is described. Section 6 concludes with a discussion of the possibilities which the existence of the functional Γ opens up in gravitation theory for generalizations to non-Euclidean topology.

The notation is that of (II) with two important exceptions: (1) We use boldface type to distinguish quantum observables from their classical c-number counterparts, lightface being reserved for the latter. (2) The symbols G^{ii}, G^{ij}, etc. now refer to propagators with Feynman boundary conditions and not to the propagation functions appearing in commutators.

1. SCHWINGER VARIATIONAL PRINCIPLE

Consider two observables $\mathbf{A}$ and $\mathbf{B}$ which are such that the local dynamical variables out of which $\mathbf{A}$ is constructed are taken at space-time points all of which lie to the future of the points at which the variables making up $\mathbf{B}$ are taken.[4] This condition will be expressed symbolically in the form

$$\mathbf{A} > \mathbf{B} \quad \text{or} \quad \mathbf{B} < \mathbf{A}, \tag{1.1}$$

which may be read "$\mathbf{A}$ *lies to the future of* $\mathbf{B}$" or "$\mathbf{B}$ *lies to the past of* $\mathbf{A}$." As operators, $\mathbf{A}$ and $\mathbf{B}$ are assumed to be Hermitian and group invariant, and we introduce their eigenvectors $|\mathbf{A}'\rangle$, $|\mathbf{B}'\rangle$, respectively, ignoring for simplicity any additional labels needed for complete characterization of the latter. We then ask how the amplitude $\langle \mathbf{A}' \mid \mathbf{B}' \rangle$ changes under an infinitesimal change $\delta\mathbf{S}$ in the action, which satisfies the condition

$$\mathbf{A} > \delta\mathbf{S} > \mathbf{B}. \tag{1.2}$$

The change in the amplitude arises from changes in the observables $\mathbf{A}$ and $\mathbf{B}$ themselves, which are produced by the alteration in the dynamical system which $\delta\mathbf{S}$ describes. The precise nature of these changes depends on the boundary conditions adopted.[5] The corresponding change in the amplitude, however, is independent of boundary conditions. We shall show this for the particular cases

[3] R. E. Peierls, Proc. Roy. Soc. (London) **A214**, 143 (1952).

[4] We thus extend the usual statement of Schwinger's principle to include quantities having semi-infinite nonlocality. We emphasize here again, as in (I), that the concepts of "past" and "future" continue to be valid in quantum gravidynamics, provided that (a) the permissible eigenvalues of the operator metric all have normal signature and (b) only invariants are admitted as observables so that their relative orientation may be specified intrinsically and not in terms of coordinates x which are only labels.

[5] The dependence of the observables $\mathbf{A}$ and $\mathbf{B}$ on the dynamical variables of the system is here assumed to have an explicit functional form which remains unaltered under the variation.

of retarded and advanced boundary conditions, from which it will follow that the independence holds in general.

In the case of retarded boundary conditions **A** and **B** suffer the changes

$$\delta\mathbf{A} = \delta^{-}\mathbf{A} = D_{\delta\mathbf{S}}\mathbf{A} = (\delta\mathbf{S}, \mathbf{A}) = -i[\delta\mathbf{S}, \mathbf{A}], \quad (1.3)$$

$$\delta\mathbf{B} = 0, \quad (1.4)$$

in which use has been made of the notation of reference 3 and the fact that

$$D_{\mathbf{A}}\delta\mathbf{S} = 0, \quad (1.5)$$

and in which the customary relation between Poisson bracket and commutator has been introduced. The change (1.3) corresponds to the unitary transformation

$$\mathbf{A} + \delta\mathbf{A} = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}, \qquad \mathbf{U} = 1 - i\,\delta\mathbf{S}. \quad (1.6)$$

Hence under retarded boundary conditions the eigenvectors $|\mathbf{A}'\rangle$, $|\mathbf{B}'\rangle$ suffer the changes

$$\delta\,|\mathbf{A}'\rangle = -i\,\delta\mathbf{S}\,|\mathbf{A}'\rangle, \quad (1.7)$$

$$\delta\,|\mathbf{B}'\rangle = 0, \quad (1.8)$$

whence it follows that

$$\delta\langle\mathbf{A}'\,|\mathbf{B}'\rangle = i\langle\mathbf{A}'|\,\delta\mathbf{S}\,|\mathbf{B}'\rangle. \quad (1.9)$$

With advanced boundary conditions, on the other hand, we have

$$\delta\mathbf{A} = 0, \quad (1.10)$$

$$\delta\mathbf{B} = \delta^{+}\mathbf{B} = D_{\mathbf{B}}\delta\mathbf{S} = (\mathbf{B}, \delta\mathbf{S}) = -i[\mathbf{B}, \delta\mathbf{S}], \quad (1.11)$$

in which use has been made of the reciprocity theorem (II, 3.14) and the fact that

$$D_{\delta\mathbf{S}}\mathbf{B} = 0. \quad (1.12)$$

From Eqs. (1.10) and (1.11) it follows that

$$\delta\,|\mathbf{A}'\rangle = 0, \quad (1.13)$$

$$\delta\,|\mathbf{B}'\rangle = i\,\delta\mathbf{S}\,|\mathbf{B}'\rangle, \quad (1.14)$$

which leads again to (1.9). Boundary conditions which differ from either retarded or advanced may be obtained by varying the original operator solution of the dynamical equations of the unaltered system. But such an additional variation, which merely transforms one solution into another, is generated by a canonical (i.e., unitary) transformation which affects both $|\mathbf{A}'\rangle$ and $|\mathbf{B}'\rangle$ equally and hence leaves Eq. (1.9) unchanged. We note also that $\delta\langle\mathbf{A}' \mid \mathbf{B}'\rangle = 0$ whenever $\delta\mathbf{S}$ lies *outside* the time interval bracketed by **A** and **B**.

Let us now for the moment restrict our attention to the case of systems possessing no infinite dimensional invariance groups so that the Hermitian dynamical variables ϕ^i, ψ^i themselves are well-defined operators satisfying definite (anti-)commutation relations. In this case it is a convenience to take the total action in the form

$$\mathbf{S} + J_i\phi^i + J_i\psi^i \quad (1.15)$$

and to consider variations in the J_i, J_i. The latter quantities, which play the role of external sources, constitute a set of adjustable parameters which serve to test the linear response of the system. They are to be regarded as "quasi-classical" quantities (or c numbers), the J_i being real functions of the commuting type and the J_i imaginary functions of the anticommuting type.

Applying the Schwinger variational principle to the action (1.15) we get

$$(\delta/i\,\delta J_i)\langle\infty\,|-\infty\rangle = \langle\infty\,|\phi^i|\,-\infty\rangle, \quad (1.16)$$

$$(\delta/i\,\delta J_i)\langle\infty\,|-\infty\rangle = \langle\infty\,|\psi^i|\,-\infty\rangle, \quad (1.17)$$

where $|-\infty\rangle$ is any eigenvector of any dynamical quantity which lies in the remote past and $|\infty\rangle$ is any eigenvector of any dynamical quantity which lies in the remote future. The functional derivatives with respect to the sources are taken at space-time points corresponding to finite times and are to be understood as *left derivatives*, in contradistinction to functional derivatives with respect to the dynamical variables which are always taken as right derivatives. If we now rewrite the right-hand sides of Eqs. (1.16):

$$\langle\infty\,|\phi^i|\,-\infty\rangle = \sum\langle\infty\,|\,\phi^{i\prime}\rangle\phi^{i\prime}\langle\phi^{i\prime}\,|\,-\infty\rangle, \quad (1.18)$$

$$\langle\infty\,|\psi^i|\,-\infty\rangle = \sum\langle\infty\,|\,\psi^{i\prime}\rangle\psi^{i\prime}\langle\psi^{i\prime}\,|\,-\infty\rangle, \quad (1.19)$$

where the summation is over the eigenvectors $|\phi^{i\prime}\rangle$, $|\psi^{i\prime}\rangle$, respectively, of complete sets of commuting and anticommuting operators[6] of which the ϕ^i and ψ^i are, respectively, members (additional labels corresponding to the other members being omitted for brevity), we obtain, upon again differentiating with respect to the sources and taking note of the occurrence of anticommuting factors [cf. (II, 1.10)],[7]

[6] J. Schwinger, Phys. Rev. **92**, 1283 (1953).

[7] The use of Schwinger's extension of the number system (reference 6) here implies that the "eigenvalues" ψi as well as structures like $\langle\ \infty\ |\ \psi i\ |\ -\ \infty\rangle$ are "c numbers" of the anticommuting type. On the other hand, matrix elements of observables (even degree in the ψ's) and actual probabilities (squares of absolute values) are c numbers of the commuting type.

$$\frac{\delta}{i\,\delta J_j}\frac{\delta}{i\,\delta J_i}\langle\infty\mid -\infty\rangle = \langle\infty\mid T(\phi^j\phi^i)\mid -\infty\rangle, \quad (1.20)$$

$$\frac{\delta}{i\,\delta J_j}\frac{\delta}{i\,\delta J_i}\langle\infty\mid -\infty\rangle = \langle\infty\mid T(\psi^j\phi^i)\mid -\infty\rangle, \quad (1.21)$$

$$\frac{\delta}{i\,\delta J_j}\frac{\delta}{i\,\delta J_1}\langle\infty\mid -\infty\rangle = \langle\infty\mid T(\psi^j\psi^i)\mid -\infty\rangle, \quad (1.22)$$

where

$$T(\phi^j\phi^i) \equiv \theta(j,i)\phi^j\phi^i + \theta(i,j)\phi^i\phi^j, \quad (1.23)$$

$$T(\psi^j\phi^i) \equiv \theta(\mathbf{j},i)\psi^j\phi^i + \theta(i,\mathbf{j})\phi^i\psi^j, \quad (1.24)$$

$$T(\psi^j\psi^i) \equiv \theta(\mathbf{j},\mathbf{i})\psi^j\psi^i - \theta(\mathbf{i},\mathbf{j})\psi^i\psi^j. \quad (1.25)$$

More generally, by repeated functional differentiation with respect to the external sources we obtain the amplitude, between the states $|-\infty\rangle$ and $|\infty\rangle$, of a product of ϕ's and ψ's arranged from right to left in chronological order and taken with a plus sign or a minus sign depending on whether the chronological order of the boldface indices is an even or odd permutation of the order in which the corresponding functional differentiations were performed.

Strictly speaking the chronological ordering operation is defined above only when the times associated with the various indices are all quite distinct. *When some of the times are equal the chronological ordering operation will be defined by the functional differentiation process itself*—a process which will be given definite formal meaning by the developments of the following sections. The very generality of these developments, furthermore, will suggest that analogs of the chronological ordering operation may in this way be defined for nonlocal and even acausal theories.

The results of the functional differentiation process will be called *T functionals*. It will be convenient to employ the abbreviation

$$T^{ij\cdots\mathbf{k}\mathbf{l}\cdots} \equiv T(\phi^i\phi^j\cdots\psi^\mathbf{i}\psi^\mathbf{j}\cdots), \quad (1.26)$$

$$\langle\infty\mid T^{ij\cdots\mathbf{k}\mathbf{l}\cdots}\mid -\infty\rangle \equiv \frac{\delta}{i\,\delta J_i}\frac{\delta}{i\,\delta J_j}\cdots \times\frac{\delta}{i\,\delta J_\mathbf{k}}\frac{\delta}{i\,\delta J_\mathbf{l}}\cdots\langle\infty\mid -\infty\rangle. \quad (1.27)$$

It is evident that each lightfaced index on a T functional commutes with all other indices, while the boldface indices anticommute among themselves.

2. FEYNMAN INTEGRALS

In the statement of Schwinger's variational principle (1.9) no indication was given of how the quantum operator $\delta\mathbf{S}$ is to be constructed from the corresponding classical quantity δS. That is, no procedure was given for handling the factor ordering ambiguity. We now suggest a tentative prescription for resolving this ambiguity. We first note that by a simple translation of the dynamical variables (viz., $\phi^i \to \phi^i + \phi^i_0$, $\psi^\mathbf{i} \to \psi^\mathbf{i} + \psi^\mathbf{i}_0$) any solution of the classical equations $S_{,i} = 0$, $S_{,\mathbf{i}} = 0$ may be chosen as the zero point. Any observable A may then be expanded about this solution in a functional power series of the form

$$A[\phi,\psi] = \sum_{m,n=0}^{\infty}\frac{1}{m!\,(2n)!} \times A^0_{,\mathbf{j}_{2n}\cdots\mathbf{j}_1 i_m\cdots i_1}\phi^{i_1}\cdots\phi^{i_m}\psi^{\mathbf{j}_1}\cdots\psi^{\mathbf{j}_{2n}}, \quad (2.1)$$

which has a certain domain of convergence and is therefore formally summable. The expansion of the action in particular begins with quadratic terms[8]:

$$S = \tfrac{1}{2}S^0_{,ij}\phi^i\phi^j + \tfrac{1}{6}S^0_{,kji}\phi^i\phi^j\phi^k + \cdots + \tfrac{1}{2}S^0_{,\mathbf{j}\mathbf{i}}\psi^\mathbf{i}\psi^\mathbf{j} + \tfrac{1}{2}S^0_{,\mathbf{j}\mathbf{i}k}\phi^k\psi^\mathbf{i}\psi^\mathbf{j} + \cdots + \tfrac{1}{24}S^0_{,\mathbf{l}\mathbf{k}\mathbf{j}\mathbf{i}}\psi^\mathbf{i}\psi^\mathbf{j}\psi^\mathbf{k}\psi^\mathbf{l} + \cdots \quad (2.2)$$

Equation (1.9) will now be rewritten in the unambiguous form

$$\delta\langle\mathbf{A}'\mid\mathbf{B}'\rangle = i\langle\mathbf{A}'\mid T(\delta\mathbf{S})\mid\mathbf{B}'\rangle \quad (2.3)$$

where the T symbol is used to indicate that $\delta\mathbf{S}$ is first to be expanded and then chronologically ordered, term by term. Thus, for an arbitrary observable $\mathbf{A}$,

$$T(\mathbf{A}) \equiv T(A[\phi,\psi]) \equiv \sum_{m,n=0}^{\infty}\frac{1}{m!\,(2n)!} \times A^0_{,\mathbf{j}_{2n}\cdots\mathbf{j}_1 i_m\cdots i_1}T^{i_1\cdots i_m\mathbf{j}_1\cdots\mathbf{j}_{2n}}. \quad (2.4)$$

The remainder of this section will be devoted to the justification of (2.3) in the case of systems possessing no infinite-dimensional invariance groups. Indeed, the T functionals involved have so far been defined only in this case. Consideration of the general case will be postponed to the following sections.

The ultimate basis for the appearance of the T symbol in Eq. (2.3) is the *assumption* that in the quantum theory the operator dynamical equations are themselves to be written in the chronologically

[8] Linear terms are absent since $\phi^i = 0$, $\psi^\mathbf{i} = 0$ is a solution of the dynamical equations, and constant terms are irrelevant. The superscript ^{0} is used in Eqs. (2.1), (2.2), and elsewhere to indicate that the quantity in question (e.g., expansion coefficient) is to be evaluated at the zero point.

ordered forms

$$T(S_{,i}[\phi, \psi]) = -J_i, \tag{2.5}$$

$$T(S_{,\mathfrak{i}}[\phi, \psi]) = -J_{\mathfrak{i}}. \tag{2.6}$$

(External sources have here been included for presently obvious reasons.) A difficulty which immediately presents itself, however, is the fact that a T functional is not generally Hermitian or anti-Hermitian even when the operators out of which it is composed are all Hermitian. In order to avoid later inconsistencies (e.g., violation of unitarity) we must therefore in some cases assume that the "classical" functional S itself has an imaginary part, so that the left-hand sides of Eqs. (2.5) and (2.6), like the right-hand sides, will be Hermitian and anti-Hermitian, respectively. It must be emphasized that this is not to be regarded as a defect of the theory. According to the point of view adopted here the classical theory is only a guide; the requirements of the quantum theory supersede all other considerations.

We now take the matrix elements of Eqs. (2.5) and (2.6) between the states $|-\infty\rangle$ and $|\infty\rangle$ and make use of (1.27), getting

$$S_{,i}[\delta/i\ \delta J]\langle\infty\ |\ -\infty\rangle = -J_i\langle\infty\ |\ -\infty\rangle, \tag{2.7}$$

$$S_{,\mathfrak{i}}[\delta/i\ \delta J]\langle\infty\ |\ -\infty\rangle = -J_{\mathfrak{i}}\langle\infty\ |\ -\infty\rangle, \tag{2.8}$$

in which the functional differential operators on the left are obtained by expanding $S_{,i}$ and $S_{,\mathfrak{i}}$ and replacing the ϕ^i and $\psi^{\mathfrak{i}}$ in these expansions respectively by $\delta/i\delta J_i$ and $\delta/i\delta J_{\mathfrak{i}}$. In order to solve Eqs. (2.7) and (2.8) for the amplitude $\langle\infty\ |\ -\infty\rangle$ as a functional of the J_i, $J_{\mathfrak{i}}$ we introduce its functional Fourier transform $F[\phi, \psi]$ and write[9]

$$\langle\infty\ |\ -\infty\rangle = \int F[\phi, \psi] \times \exp\{i(J_i\phi^i + J_{\mathfrak{i}}\psi^{\mathfrak{i}})\}\ \delta[\phi]\ \delta[\psi], \tag{2.9}$$

where $\delta[\phi]$ and $\delta[\psi]$ are expressed formally by

$$\delta[\phi] \equiv \prod_i d\phi^i, \qquad \delta[\psi] \equiv \prod_{\mathfrak{i}} d\psi^{\mathfrak{i}}. \tag{2.10}$$

Inserting (2.9) into (2.7) and (2.8) we get

$$0 = \int F[\phi, \psi](S_{,i}[\phi, \psi] + J_i) \times \exp\{i(J_i\phi^i + J_{\mathfrak{j}}\psi^{\mathfrak{j}})\}\ \delta[\phi]\ \delta[\psi] = \int F[\phi, \psi]\left(S_{,i}[\phi, \psi] + \frac{\delta}{i\ \delta\phi^i}\right) \times \exp\{i(J_i\phi^i + J_{\mathfrak{j}}\psi^{\mathfrak{j}})\}\ \delta[\phi]\ \delta[\psi], \tag{2.11}$$

and similarly

$$0 = \int F[\phi, \psi]\left(S_{,\mathfrak{i}}[\phi, \psi] + \frac{\delta}{i\ \delta\psi^{\mathfrak{i}}}\right) \times \exp\{i(J_i\phi^i + J_{\mathfrak{j}}\psi^{\mathfrak{j}})\}\ \delta[\phi]\ \delta[\psi]. \tag{2.12}$$

Integrating by parts we obtain the following differential equations for the Fourier transform:

$$(S_{,i}[\phi, \psi] - \delta/i\ \delta\phi^i)F[\phi, \psi] = 0, \tag{2.13}$$

$$(S_{,\mathfrak{i}}[\phi, \psi] - \delta/i\ \delta\psi^{\mathfrak{i}})F[\phi, \psi] = 0, \tag{2.14}$$

of which the general solution is

$$F[\phi, \psi] = Ne^{iS[\phi,\psi]}, \tag{2.15}$$

where N is a normalization constant. Hence finally

$$\langle\infty\ |\ -\infty\rangle = N\int \exp\{i(S[\phi, \psi] + J_i\phi^i + J_{\mathfrak{i}}\psi^{\mathfrak{i}})\} \times \delta[\phi]\ \delta[\psi], \tag{2.16}$$

which is Feynman's well-known integral for the amplitude $\langle\infty\ |\ -\infty\rangle$.

If we apply the identity (1.27) to (2.16) we get

$$\langle\infty\ |T^{ij\cdots\mathfrak{k}\mathfrak{l}\cdots}|\ -\infty\rangle = N\int \phi^i\phi^j \cdots \psi^{\mathfrak{k}}\psi^{\mathfrak{l}} \cdots \times \exp\{i(S + J_a\phi^a + J_{\mathfrak{a}}\psi^{\mathfrak{a}})\}\ \delta[\phi]\ \delta[\psi], \tag{2.17}$$

and, more generally,

$$\langle\infty\ |T(\mathbf{A})|\ -\infty\rangle = N\int A \exp\{i(S + J_i\phi^i + J_{\mathfrak{i}}\psi^{\mathfrak{i}})\} \times \delta[\phi]\ \delta[\psi], \tag{2.18}$$

where A is an arbitrary functional of the ϕ^i, $\psi^{\mathfrak{i}}$ and $\mathbf{A}$ is its quantum form. In particular, setting $J_i = 0$, $J_{\mathfrak{i}} = 0$, we have

$$\delta\langle\infty\ |\ -\infty\rangle = N\int \delta e^{iS}\ \delta[\phi]\ \delta[\psi] = iN\int (\delta S)e^{iS}\ \delta[\phi]\ \delta[\psi], \tag{2.19}$$

which leads immediately to Eq. (2.3) with $|\mathbf{A}'\rangle = |\infty\rangle$

[9] The concepts of functional integration and functional Fourier transformation are legitimately extendable to variables of the anticommuting type. The simplest way to do this is to regard each $\psi^{\mathfrak{i}}$ as the product of an anticommuting constant and a variable of the commuting type, a different anticommuting constant being associated with each value of the index $\mathfrak{i}$. The integration is then performed over the commuting variables.

and $|\mathbf{B}'\rangle = |-\infty\rangle$. Equation (2.3) is indeed included in (2.19), for as long as $\delta\mathbf{S}$ vanishes outside the interval bracketed by $\mathbf{A}$ and $\mathbf{B}$, the states $|\mathbf{A}'\rangle$ and $|\mathbf{B}'\rangle$ may always be characterized by observables of the unaltered system which lie in the remote future and past, respectively.

It will be noted that the dynamical equations (2.5) and (2.6) are immediately regained in matrix form from the identities

$$-iN\int [\exp\{i(S + J_i\phi^i + J_i\psi^i)\}]_{,i} \times \delta[\phi]\,\delta[\psi] \equiv 0, \qquad (2.20)$$

$$-iN\int [\exp\{i(S + J_i\phi^i + J_i\psi^i)\}]_{,i} \times \delta[\phi]\,\delta[\psi] \equiv 0, \qquad (2.21)$$

which follow from Gauss' theorem in functional space. The surface integrals at functional infinity vanish because of the rapidly oscillating behavior of e^{iS} when ϕ^i and ψ^i depart widely from classical solutions.

In general, the integration in (2.16) is limited to a domain occupied by a definite class of functions ϕ^i, ψ^i satisfying well-defined boundary conditions. In certain cases, however, the domain may be taken as unrestricted and the desired boundary conditions may be secured by making appropriate infinitesimal changes in the action functional S. A case of particular importance is that of the "vacuum-to-vacuum" amplitude which is obtained simply by giving all mass parameters appearing in the theory an infinitesimal negative imaginary part, a possibility which is closely related to that of making an analytic continuation of the pertinent Green's functions from a space-time of indefinite metric to one of definite metric with unique Green's functions. From now on we shall restrict our attention to this case, for it is known that all physical processes can be computed from a knowledge of the vacuum-to-vacuum matrix elements of appropriate T functionals.[10] When external sources are present the past and future vacua are generally distinct, and we shall denote them by $|0, -\infty\rangle$ and $|0, \infty\rangle$, respectively. In using this notation, however, we do not mean to imply that either "vacuum" is necessarily unique, except in the case of Lorentz invariant theories. In the case of quantum gravidynamics, in fact, we shall later give reasons for believing that many "vacua" are possible, and one of our aims will be to provide characterizations of such states, different from the conventional one in terms of positive and negative frequencies, which are capable of serving in the broader context of general relativity.

[10] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo cimento **1**, 425 (1955).

3. INVARIANCE GROUPS, INVARIANT MEASURE, AND SUPPLEMENTARY CONDITIONS

When an infinite-dimensional invariance group is present, Eq. (2.18) becomes invalid for two reasons. First, the appearance of external sources J_i and J_i coupled linearly to the dynamical variables violates group invariance. Secondly, no normalization constant N exists. Actually N does not strictly exist even in the absence of invariance groups, but in this case its value is closely related to the spacing of the "lattice" in space-time which is conveniently set up in order to give a proper definition of the volume element $\delta[\phi]\,\delta[\psi]$, and it diverges only in the limit as the lattice spacing tends to zero. When an infinite-dimensional invariance group is present, the functional integral itself diverges, even when the lattice spacing is finite. This is because the integrand, with $J_i = J_i = 0$, remains constant instead of oscillating when the functions ϕ^i, ψ^i range over values which differ from one another only by group transformations.

The first difficulty will be removed in this section simply by eliminating the external sources from discussion and reintroducing them only in the next section where it will be assumed that the dynamical variables are group invariants. The second difficulty may be surmounted by dealing with the ratio

$$\langle\mathbf{A}\rangle \equiv \frac{\langle 0, \infty\,|T(\mathbf{A})|\,0, -\infty\rangle}{\langle 0, \infty\,|\,0, -\infty\rangle} = \frac{\int Ae^{iS}\,\delta[\phi]\,\delta[\psi]}{\int e^{iS}\,\delta[\phi]\,\delta[\psi]} \qquad (3.1)$$

instead of with the matrix elements themselves. It is well known, in fact, that such ratios play a fundamental role in field theory already in the absence of infinite dimensional invariance groups.[11] In Eq. (3.1) a limiting procedure may be imagined, in which the domain of integration is allowed to tend to infinity after the ratio is taken. It is also essential to impose the restriction that A be a group invariant. Otherwise the numerator is not well defined. We shall see that all physical processes can ultimately be computed from a knowledge of such ratios, involving group invariants only.

There is only one additional point which needs to be discussed, but it is by no means a trivial one.

[11] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **37**, 452 (1951).

If a truly group invariant theory is to be achieved it is necessary to demand that the functional volume element, represented symbolically by $\delta[\phi]\ \delta[\psi]$, be itself a group invariant. In general Eq. (2.10) will no longer provide a suitable formal definition of the volume element, and the question of an appropriate "measure" arises.

In order to deal with this question it will first be necessary to review briefly some of the properties of invariance groups in the abstract. In doing this we shall use the language of finite dimensional Lie groups even though the indices appearing in the equations which follow will really be continuous labels. Abstract group elements will be denoted by barred letters $\bar{x}$, $\bar{y}$, $\bar{z}$, and their explicit representations, or "coordinates," in the functional group manifold will be denoted by $\bar{x}^\alpha$, $\bar{y}^\beta$, $\bar{z}^\gamma$, etc. The identity element will be denoted by 1.

The group is determined if the functionals

$$f^\alpha[\bar{x}, \bar{y}] = (\bar{x}\bar{y})^\alpha, \tag{3.2}$$

which fix the multiplication table of the group, are given. These functionals satisfy the necessary identities

$$f^\alpha[\bar{x}, 1] = f^\alpha[1, \bar{x}] = \bar{x}^\alpha, \tag{3.3}$$

$$f^\alpha[\bar{x}, \bar{x}^{-1}] = f^\alpha[\bar{x}^{-1}, \bar{x}] = 1^\alpha, \tag{3.4}$$

$$f^\alpha[\bar{x}, \bar{y}\bar{z}] = f^\alpha[\bar{x}\bar{y}, \bar{z}]. \tag{3.5}$$

By differentiating these identities repeatedly with respect to $\bar{z}^\alpha$, $\bar{y}^\beta$, $\bar{z}^\gamma$, and setting various group elements equal to 1, one obtains the following[12]:

$$l^\alpha_{\beta,\delta} l^\delta_\gamma - l^\alpha_{\gamma,\delta} l^\delta_\beta = l^\alpha_\delta c^\delta_{\beta\gamma}, \tag{3.6}$$

$$r^\alpha_{\beta,\delta} r^\delta_\gamma - r^\alpha_{\gamma,\delta} r^\delta_\beta = -r^\alpha_\delta c^\delta_{\beta\gamma}, \tag{3.7}$$

$$c^\delta_{\alpha\epsilon} c^\epsilon_{\beta\gamma} + c^\delta_{\beta\epsilon} c^\epsilon_{\gamma\alpha} + c^\delta_{\gamma\epsilon} c^\epsilon_{\alpha\beta} = 0, \tag{3.8}$$

where l^α_β and r^α_β are the *auxiliary functionals* of the group and the $c^\alpha_{\beta\gamma}$ are the *structure constants*:

$$l^\alpha_\beta[\bar{x}] \equiv (\delta f^\alpha[\bar{y}, \bar{x}]/\delta \bar{y}^\beta)_{\bar{y}=1}, \tag{3.9}$$

$$r^\alpha_\beta[\bar{x}] \equiv (\delta f^\alpha[\bar{x}, \bar{y}]/\delta \bar{y}^\beta)_{\bar{y}=1}, \tag{3.10}$$

$$c^\alpha_{\beta\gamma} \equiv (\delta^2 f^\alpha[\bar{x}, \bar{y}]/\delta\bar{x}^\beta\ \delta\bar{y}^\gamma - \delta^2 f^\alpha[\bar{x}, \bar{y}]/\delta\bar{x}^\gamma\ \delta\bar{y}^\beta)_{\bar{x}=\bar{y}=1}. \tag{3.11}$$

The auxiliary functionals alone can be shown to determine the group.[12] Furthermore, with the introduction of *canonical "coordinates"* based on the one-parameter Abelian subgroups generated by infinitesimal group transformations $1 + \delta\xi$, the auxiliary functionals can in turn be shown to be determined locally by the structure constants. Canonical "coordinates" are characterized by the conditions

$$l^\alpha_\beta[\bar{x}]\bar{x}^\beta = r^\alpha_\beta[\bar{x}]\bar{x}^\beta = \bar{x}^\alpha, \qquad 1^\alpha = 0. \tag{3.12}$$

When these conditions are satisfied the auxiliary functionals are given uniquely by

$$l[\bar{x}] = \frac{c\cdot\bar{x}}{\exp(c\cdot\bar{x}) - 1}, \qquad r[\bar{x}] = \frac{c\cdot\bar{x}}{1 - \exp(-c\cdot\bar{x})}$$

where the symbols on the left, without indices, are to be understood as representing the continuous matrices having the l^α_β and r^α_β as elements. Here

$$c\cdot\bar{x} \equiv c_\alpha \bar{x}^\alpha, \tag{3.14}$$

where the c_α are the matrices formed from the structure constants by treating their first and last indices as matrix indices. With this abbreviated notation Eq. (3.8) may be rewritten in the compact form

$$[c_\alpha, c_\beta] = c_\gamma c^\gamma_{\alpha\beta}, \tag{3.15}$$

which reveals the c_α as the generators of a linear representation of the group, known as the *adjoint representation*, which may also be shown to represent the group of inner automorphisms.[13] The matrix $D[\bar{x}]$ of the adjoint representation which represents a finite group transformation is given by

$$D = l^{-1}r = e^{c\cdot\bar{x}}, \tag{3.16}$$

the intermediate form holding in an arbitrary "coordinate" system, and the final form holding only in canonical "coordinates." Among the corollaries of the above identities we may cite the following:

$$(\mathrm{Tr}\, c_\gamma)c^\gamma_{\alpha\beta} = 0, \tag{3.17}$$

$$(\mathrm{Tr}\, c_\beta)r^\beta_\alpha = (\mathrm{Tr}\, c_\beta)l^\beta_\alpha = \mathrm{Tr}\, c_\alpha \text{ in canonical "coordinates,"} \tag{3.18}$$

$$\det l = \det r \text{ in any "coordinate" system if and only if } \mathrm{Tr}\, c_\alpha = 0 \text{ for all } \alpha. \tag{3.19}$$

The group manifold possesses two natural volume elements,

$$\delta_l[\bar{x}] \equiv \det l^{-1} \prod_\alpha d\bar{x}^\alpha, \qquad \delta_r[\bar{x}] \equiv \det r^{-1} \prod_\alpha d\bar{x}^\alpha, \tag{3.20}$$

which satisfy the invariance conditions

$$\delta_l[\bar{x}\bar{y}] = \delta_l[\bar{x}], \quad \delta_r[\bar{y}\bar{x}] = \delta_r[\bar{x}], \quad \bar{y} \text{ fixed}, \tag{3.21}$$

[12] See, for example, L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, New Jersey, 1946).

[13] The adjoint representation of the general coordinate transformation group is that which is provided by a contravariant vector field.

$$\delta_l[\bar{x}^{-1}] = \delta_r[\bar{x}], \qquad \delta_r[\bar{x}^{-1}] = \delta_l[\bar{x}], \tag{3.22}$$

and hence

$$\int_R f[\bar{x}\bar{y}]\ \delta_l[\bar{x}] = \int_{R\bar{y}} f[\bar{x}]\ \delta_l[\bar{x}], \tag{3.23}$$

$$\int_R f[\bar{y}\bar{x}]\ \delta_r[\bar{x}] = \int_{\bar{y}R} f[\bar{x}]\ \delta_r[\bar{x}], \tag{3.24}$$

$$\int_R f[\bar{x}^{-1}]\ \delta_l[\bar{x}] = \int_{R^{-1}} f[\bar{x}]\ \delta_r[\bar{x}], \tag{3.25}$$

$$\int_R f[\bar{x}^{-1}]\ \delta_r[\bar{x}] = \int_{R^{-1}} f[\bar{x}]\ \delta_l[\bar{x}], \tag{3.26}$$

If the group is compact[14] and $f[\bar{x}]$ is bounded the integrals may be extended over the entire group. In this case condition (3.19) may be shown to hold and the two volume elements become identical. In general, however, they are not identical. In fact, with infinite-dimensional groups Tr c_α may not exist, and the determinants det l, det r may not even be formally definable.[15]

Returning now to the consideration of the dynamical system we remark that each group element $\bar{x}$ defines a transformation of the dynamical variables which is expressible in the general form

$$\phi'^i = \Phi^i[\bar{x}, \phi, \psi], \tag{3.27}$$

$$\psi'^{\mathfrak{i}} = \Psi^{\mathfrak{i}}[\bar{x}, \phi, \psi]. \tag{3.28}$$

The functionals Φ^i, $\Psi^{\mathfrak{i}}$ must satisfy the identities

$$\Phi^i[1, \phi, \psi] = \phi^i, \qquad \Psi^{\mathfrak{i}}[1, \phi, \psi] = \psi^{\mathfrak{i}}, \tag{3.29}$$

$$\left.\begin{aligned}\Phi^i[\bar{x}\bar{y}, \phi, \psi] &= \Phi^i[\bar{x}, \Phi[\bar{y}, \phi, \psi], \Psi[\bar{y}, \phi, \psi]],\\ \Psi^{\mathfrak{i}}[\bar{x}\bar{y}, \phi, \psi] &= \Psi^{\mathfrak{i}}[\bar{x}, \Phi[\bar{y}, \phi, \psi], \Psi[\bar{y}, \phi, \psi]],\end{aligned}\right\} \tag{3.30}$$

which, upon differentiation, yield

$$\left.\begin{aligned}\Phi^i_{,\beta}[\bar{x}, \phi, \psi] l^\beta_\alpha[\bar{x}] &= R^i_\alpha[\Phi[\bar{x}, \phi, \psi], \Psi[\bar{x}, \phi, \psi]],\\ \Psi^{\mathfrak{i}}_{,\beta}[\bar{x}, \phi, \psi] l^\beta_\alpha[\bar{x}] &= R^{\mathfrak{i}}_\alpha[\Phi[\bar{x}, \phi, \psi], \Psi[\bar{x}, \phi, \psi]],\end{aligned}\right\} \tag{3.31}$$

$$\left.\begin{aligned}\Phi^i_{,\beta}[\bar{x}, \phi, \psi] r^\beta_\alpha[\bar{x}] &= \Phi^i_{,j}[\bar{x}, \phi, \psi] R^j_\alpha[\phi, \psi]\\ &\quad + \Phi^i_{,\mathfrak{j}}[\bar{x}, \phi, \psi] R^{\mathfrak{j}}_\alpha[\phi, \psi],\\ \Psi^{\mathfrak{i}}_{,\beta}[\bar{x}, \phi, \psi] r^\beta_\alpha[\bar{x}] &= \Psi^{\mathfrak{i}}_{,j}[\bar{x}, \phi, \psi] R^j_\alpha[\phi, \psi]\\ &\quad + \Psi^{\mathfrak{i}}_{,\mathfrak{j}}[\bar{x}, \phi, \psi] R^{\mathfrak{j}}_\alpha[\phi, \psi],\end{aligned}\right\} \tag{3.32}$$

where the R^i_α, $R^{\mathfrak{i}}_\alpha$ are the functionals characterizing infinitesimal transformations [cf. II, Eq. (1.1)]:

$$\left.\begin{aligned}R^i_\alpha[\phi, \psi] &= (\delta\Phi^i[\bar{x}, \phi, \psi]/\delta\bar{x}^\alpha)_{\bar{x}=1},\\ R^{\mathfrak{i}}_\alpha[\phi, \psi] &= (\delta\Psi^{\mathfrak{i}}[\bar{x}, \phi, \psi]/\delta\bar{x}^\alpha)_{\bar{x}=1}.\end{aligned}\right\} \tag{3.33}$$

Differentiating Eqs. (3.31) and making use of Eq. (3.6) one obtains

$$\left.\begin{aligned}R^i_{\alpha,j}R^j_\beta + R^i_{\alpha,\mathfrak{j}}R^{\mathfrak{j}}_\beta - R^i_{\beta,j}R^j_\alpha - R^i_{\beta,\mathfrak{j}}R^{\mathfrak{j}}_\alpha &= R^i_\gamma c^\gamma_{\alpha\beta},\\ R^{\mathfrak{i}}_{\alpha,j}R^j_\beta + R^{\mathfrak{i}}_{\alpha,\mathfrak{j}}R^{\mathfrak{j}}_\beta - R^{\mathfrak{i}}_{\beta,j}R^j_\alpha - R^{\mathfrak{i}}_{\beta,\mathfrak{j}}R^{\mathfrak{j}}_\alpha &= R^{\mathfrak{i}}_\gamma c^\gamma_{\alpha\beta}.\end{aligned}\right\} \tag{3.34}$$

When the representation which the ϕ^i and $\psi^{\mathfrak{i}}$ provide is linear, Eqs. (3.34) reduce to Eqs. (1.5) and (1.6) of (II). As in (II) we shall confine our attention to this case. The functional space of the dynamical variables then becomes the direct product of two independent subspaces, one for the ϕ^i and the other for the $\psi^{\mathfrak{i}}$.

The invariance group defines a family of natural metrics in each of these subspaces. A "natural" metric is one which admits the group itself as a group of motions. Denoting the components of such a metric by g_{ij} and $g_{\mathfrak{i}\mathfrak{j}}$, respectively, for the two subspaces, we must have

$$g_{ij,k}R^k_\alpha + g_{kj}R^k_{\alpha,i} + g_{ik}R^k_{\alpha,j} = 0, \tag{3.35}$$

$$g_{\mathfrak{i}\mathfrak{j},\mathfrak{k}}R^{\mathfrak{k}}_\alpha + g_{\mathfrak{k}\mathfrak{j}}R^{\mathfrak{k}}_{\alpha,\mathfrak{i}} + g_{\mathfrak{i}\mathfrak{k}}R^{\mathfrak{k}}_{\alpha,\mathfrak{j}} = 0, \tag{3.36}$$

which are the conditions that the displacements $\delta\phi^i$, $\delta\psi^{\mathfrak{i}}$ in the dynamical variables produced by any infinitesimal group transformation shall be Killing vectors in these spaces. Equations (3.35) and (3.36) are also the necessary and sufficient conditions that the metrics g_{ij} and $g_{\mathfrak{i}\mathfrak{j}}$ transform, under group transformations, contragrediently to the homogeneous parts (taken twice) of the representations provided by ϕ^i and $\psi^{\mathfrak{i}}$, respectively. The integrability of these equations is assured by the identities (1.5) and (1.6) of (II); their solutions are unique, however, only in the case of transitive representations, which are empty of dynamical interest.

The group invariant volume element associated with the metrics g_{ij}, $g_{\mathfrak{i}\mathfrak{j}}$ is given by

$$\delta[\phi]\ \delta[\psi] = \Delta[\phi, \psi] \prod_i d\phi^i \prod_{\mathfrak{i}} d\psi^{\mathfrak{i}}, \tag{3.37}$$

---

[14] In the case of the Yang-Mills field [see (I)] the invariance group may be regarded as compact if the associated finite dimensional Lie group is compact. The coordinate transformation group is not compact.

[15] Tr c^α is undefinable in the case of the coordinate transformation group. If the abstract group elements $\bar{x}$ are represented by the functions $\bar{x}^\mu(x)$ which describe the coordinate transformation $x^\mu \to \bar{x}^\mu$, then det r^{-1} can be defined formally as

$$\det r^{-1}[\bar{x}] = \Pi_x \frac{\partial(x)}{\partial(\bar{x})},$$

where $\partial(x)/\partial(\bar{x})$ is the Jacobian of the transformation, but it appears impossible to define det l^{-1} for this group.

where the "measure" Δ is defined formally by

$$\Delta[\phi, \psi] \equiv [\det(g_{ii}) \det(g_{ij})]^{1/2} \tag{3.38}$$

and satisfies the identity

$$(\Delta R^i_\alpha)_{,i} + \Delta_{,i} R^i_\alpha + \Delta R^i_{\alpha,i} = 0. \tag{3.39}$$

Δ is determined by this identity only up to an arbitrary group invariant factor. At this stage no argument exists which can delimit the measure further.[16]

The metrical structure of the functional space of the dynamical variables can be visualized schematically by referring to Fig. 1. Each dynamical configuration ϕ^i, ψ^i corresponds to a point ϕ, ψ in the functional space. This point is not assumed to correspond to a solution of the dynamical equations, for it must represent an arbitrary field history in the Feynman integral. Starting from this point one can generate a group invariant subspace by performing all possible group transformations on the ϕ^i, ψ^i. The points of this subspace provide a transitive representation of the group. In the cases of physical interest the dimensionality of this subspace is identical with that of the group, the transitive representation which it provides being equivalent to that which is provided by the multiplication table of the group itself. (There seems to be no *a priori* reason why this should be so in general, however.)

The group decomposes the functional space into invariant subspaces, each of which corresponds to a physically distinct field history. Each invariant subspace possesses a unique natural metric, and the functional Δ is completely determined within it up to a multiplicative factor. The multiplicative factor may, however, vary in a completely arbitrary way from one invariant subspace to another.

Each invariant subspace may be labeled by choosing a representative point ϕ_0, ψ_0 within it. The manifold of representative points is conveniently taken as a set satisfying conditions of the form

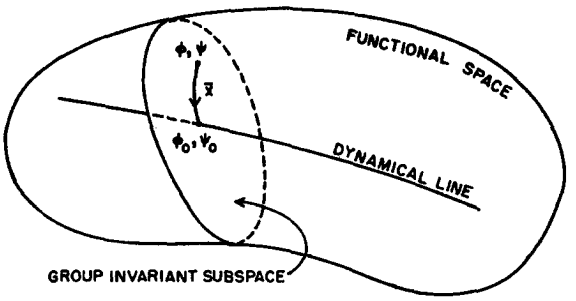


FIG. 1. Decomposition of the functional space of dynamical variables into group invariant subspaces.

$$R_\alpha[\phi_0, \psi_0] = 0, \tag{3.40}$$

where the R_α are *arbitrary* independent local functionals equal in number to the dimensionality of the group. Although the difference between this dimensionality and that of the full functional space is transfinite we shall, in referring to Fig. 1, speak of the manifold of representative points as the *dynamical line*. The only restriction on the dynamical line is that it must intersect each invariant subspace in exactly one point.

Conditions of the form (3.40) are referred to as *supplementary conditions*. Because of the hyperbolic character of space-time it is usually argued that supplementary conditions do *not* define a unique representative point ϕ_0, ψ_0, i.e., that the dynamical line intersects each invariant subspace in not one but an infinity of points. In support of this view it is pointed out that the supplementary conditions remain unchanged under infinitesimal group transformations for which the parameters $\delta\xi^\alpha$ satisfy the equation

$$F_{\alpha\beta}[\phi_0, \psi_0]\ \delta\xi^\beta = 0, \tag{3.41}$$

where

$$F_{\alpha\beta} \equiv R_{\alpha,i} R^i_\beta + R_{\alpha,i} R^i_\beta. \tag{3.42}$$

Since $F_{\alpha\beta}$ is usually a wave operator of normal type, Eq. (3.41) has an infinity of solutions.

Within the framework of the Feynman theory, however, this multi-valuedness disappears. It is important to note that in the definition of the functional integrals *two* limiting procedures are implied. Not only must a lattice be set up, with a spacing which tends to zero, but also this lattice must at any stage be of finite size, tending to infinity only as an idealization at the end. Thus the fields in the functional integration vary only within a bounded region of space-time, which we shall call the *experimental region* in

[16] In the case of the Yang-Mills field $R^i_{\alpha,i} = \mathrm{Tr}\, c_\alpha$, and Δ may be set equal to a constant (e. g., unity) if the associated finite dimensional Lie group is semi-simple. This so-called "linear measure" can also be employed in quantum gravidynamics if the covariant metric is chosen as the fundamental field variable. Any nonsingular structure of the form

$$g^{1/2}[a(g^{\mu\sigma}g^{\nu\tau} + g^{\mu\tau}g^{\nu\sigma}) + bg^{\mu\nu}g^{\sigma\tau}]\delta(x, x')$$

provides a group-invariant metric for the space of the $g_{\mu\nu}$. Such structures, when regarded as $(10 \times \infty^4) \times (10 \times \infty^4)$ matrices, have determinants which are independent of $g_{\mu\nu}$. It is to be noted that the criteria of this section reject the measure

$$\Delta = \Pi_x\, g^{-5/2}(x),\ g \equiv -\det(g_{\mu\nu}),$$

which has been proposed by Misner. [C. W. Misner, Revs. Modern Phys. 29, 497 (1957).]

recognition of the fact that it is determined in practice by the limits of some physical apparatus, the meter readings of which the theory is designed to predict.[17] Outside of this region the fields are held at fixed values which may for convenience be chosen so as to satisfy the classical field equations there, as well as the supplementary conditions (3.40), and which, by appropriate adjustment of the zero point, may be taken as $\phi^i = 0, \psi^i = 0$. Fields satisfying these conditions will be said to lie in the *experimental subspace.*

The representative point ϕ_0, ψ_0 may now be uniquely fixed by requiring that the fields ϕ_0^i, ψ_0^i vanish everywhere outside the future light cone of the experimental region. This imposes a retarded or "outgoing wave" condition on the parameters $\delta\xi^\alpha$ of Eq. (3.41), with the result that the only permissible solution is $\delta\xi^\alpha = 0$.[18] Under these restrictions the quantities ϕ_0^i, ψ_0^i become true group invariants.

We shall demonstrate this invariance in detail by obtaining an equation for the unique group element $\bar{x}[\phi, \psi]$ which transforms each point ϕ, ψ in the experimental subspace into its corresponding representative point. It is to be noted that the representative point itself need not lie in the experimental subspace since the functions ϕ_0^i, ψ_0^i are not required to vanish at space-time points which lie outside the experimental region and inside the forward light cone of this region. No practical generality is lost if we limit our attention to the case considered in (II), in which the group representation provided by the ψ^i is linear homogeneous while that provided by the ϕ^i is linear inhomogeneous, with the zero point chosen as above.[19] The supplementary conditions are then expressible in terms of the ϕ_0^i alone,

$$R_\alpha[\phi_0] = 0, \qquad F_{\alpha\beta} = R_{\alpha,i}R^i_\beta, \tag{3.43}$$

and the group element $\bar{x}$ depends only on the ϕ^i. We must have

$$\left.\begin{aligned} \phi_0^i[\phi] &= \Phi^i[\bar{x}[\phi],\ \phi], \\ \psi_0^i[\phi,\ \psi] &= \Psi^i[\bar{x}[\phi],\ \psi], \end{aligned}\right\} \tag{3.44}$$

and the condition to be satisfied is

$$R_\alpha[\Phi[\bar{x}[\phi],\ \phi]] \equiv 0. \tag{3.45}$$

Differentiating this equation with respect to ϕ^i and making use of (3.31), we get

$$F_{\alpha\gamma}[\Phi[\bar{x}[\phi],\ \phi]]l^{-1}{}^\gamma_\beta[\bar{x}[\phi]]\bar{x}^\beta_{,i}[\phi] + R_{\alpha,j}[\Phi[\bar{x}[\phi],\ \phi]]\Phi^j_{,i}[\bar{x}[\phi],\ \phi] = 0, \tag{3.46}$$

which yields, in virtue of the outgoing wave boundary conditions,

$$\bar{x}^\alpha_{,i}[\phi] = l^\alpha_\beta[\bar{x}[\phi]]G^{-\beta\gamma}[\Phi[\bar{x}[\phi],\ \phi]] \times R_{\gamma,j}[\Phi[\bar{x}[\phi],\ \phi]]\Phi^j_{,i}[\bar{x}[\phi],\ \phi], \tag{3.47}$$

where $G^{-\alpha\beta}$ is the retarded Green's function of the operator $F_{\alpha\beta}$. Equations (3.47) constitute a set of simultaneous first-order functional differential equations which must be integrated subject to the boundary condition: $\bar{x}^\alpha[\phi] = 1^\alpha$ when ϕ lies in the experimental subspace and $R_\alpha[\phi] = 0$. It is straightforward to show the integrability of these equations, by differentiating them with respect to ϕ^j and verifying that they imply the necessary symmetry of the result in the indices i and j.

Making use of (3.31), (3.32), (3.44), and (3.47), we have

$$\begin{aligned} \phi^i_{0,j}R^j_\alpha &= (\Phi^i_{,\beta}\bar{x}^\beta_{,j} + \Phi^i_{,j})R^j_\alpha \\ &= (\Phi^i_{,\beta}l^\beta_\gamma G^{-\gamma\delta}R_{\delta,j}\Phi^j_{,\epsilon} + \Phi^i_{,\epsilon})r^\epsilon_\alpha \\ &= (\Phi^i_{,\beta}l^\beta_\gamma G^{-\gamma\delta}R_{\delta,j}R^j_\zeta l^{-1}{}^\zeta_\epsilon + \Phi^i_{,\epsilon})r^\epsilon_\alpha = 0, \end{aligned} \tag{3.48}$$

$$\begin{aligned} \psi^i_{0,j}R^j_\alpha + \psi^i_{0,j}R^j_\alpha &= \psi^i_{,\beta}\bar{x}^\beta_{,j}R^j_\alpha + \psi^i_{,j}R^j_\alpha \\ &= (\psi^i_{,\beta}l^\beta_\gamma G^{-\gamma\delta}R_{\delta,j}\Phi^j_{,\epsilon} + \psi^i_{,\epsilon})r^\epsilon_\alpha = 0, \end{aligned} \tag{3.49}$$

which reveals the group invariance of ϕ_0^i and ψ_0^i in explicit form. The group transformation law for $\bar{x}[\phi]$ is readily inferred by noting that when ϕ^i suffers the infinitesimal transformation

$$\phi^i \to \Phi^i[1 + \delta\xi, \phi],$$

$\bar{x}[\phi]$ must transform according to

$$\bar{x}[\phi] \to \bar{x}[\phi](1 - \delta\xi),$$

and hence

$$\delta\bar{x}^\alpha = -r^\alpha_\beta[\bar{x}]\ \delta\xi^\beta. \tag{3.50}$$

This can also be verified directly with the aid of (3.31), (3.32), and (3.47):

---

[17] It is not essential to the argument that the experimental region be here given an intrinsic (i. e., coordinate invariant) specification. Such a specification is always automatically accomplished when the apparatus itself is included in the action functional.

[18] Compare with V. Fock, *The Theory of Space Time and Gravitation* (Pergammon Press, New York, 1959). $\delta\xi^\alpha$ is always assumed to vanish at infinity, thus eliminating any remaining finite dimensional groups of transformations which might otherwise be possible (e.g., Lorentz group).

[19] In (II) the zero point was chosen to correspond to flat empty space-time. The restriction to this particular solution of the classical field equations is not essential, and for the sake of greater generality we abandon it here. We note that a shift of the zero point does not destroy the linearity of group transformation laws.

$$\delta\bar{x}^\alpha = \bar{x}^\alpha_{,i} R^i_\beta \, \delta\xi^\beta$$
$$= l^\alpha_\gamma G^{-\gamma\delta} R_{\delta,i} R^i_\epsilon l^{-1\epsilon}_\iota r^\iota_\beta \, \delta\xi^\beta = -r^\alpha_\beta \, \delta\xi^\beta. \quad (3.51)$$

Because of the supplementary condition (3.43) the ϕ^i_0 are not all independent. Out of them we may, however, construct a complete independent set ϕ^A_0 and then perform the transformation $\phi^i, \psi^i \to \bar{x}^\alpha, \phi^A_0, \psi^i_0$. The new variables are "natural" variables for the dynamical system. The cross sections $\phi^A_0 = \text{const}$, $\psi^i_0 = \text{const}$ are the invariant subspaces. Since each invariant subspace transforms into itself, as Eq. (3.50) explicitly shows, the group invariant measure for the functional space is separable in these variables, and it is not hard to show that it takes the form

$$\Delta_0[\bar{x}, \phi_0, \psi_0] = \Delta_0[\phi_0, \psi_0] \det l^{-1}[\bar{x}], \quad (3.52)$$

where $\Delta_0[\phi_0, \psi_0]$ is completely arbitrary.

The natural measure of the group makes its appearance here as a necessary consequence of the introduction of the group parameters themselves as "dynamical variables." That it should be $\det l^{-1}$ rather than $\det r^{-1}$ which appears is a consequence merely of the particular convention employed for defining $\bar{x}[\phi]$. Under the transformation $\bar{x} \to \bar{x}^{-1}$ we get $\det r^{-1}$ instead.

For the measure (3.52), Eq. (3.39) reduces, in virtue of (3.50), to

$$(\det l^{-1} r^\beta_\alpha)_{,\alpha} = 0. \quad (3.53)$$

This identity, which holds for any group, is readily verified in canonical "coordinates." Using (3.7), (3.16), and (3.18) we get

$$(\det l^{-1} r^\beta_\alpha)_{,\beta} = (e^{\operatorname{Tr} c \cdot \bar{x}} \det r^{-1} r^\beta_\alpha)_{,\beta}$$
$$= e^{\operatorname{Tr} c \cdot \bar{x}} \det r^{-1} [(\operatorname{Tr} c_\beta) r^\beta_\alpha - r^{-1\gamma}_\delta r^\delta_{\gamma,\beta} r^\beta_\alpha + r^\beta_{\alpha,\beta}]$$
$$= \det l^{-1} (\operatorname{Tr} c_\alpha + r^{-1\gamma}_\delta r^\delta_\epsilon c^\epsilon_{\gamma\alpha} - r^{-1\gamma}_\delta r^\delta_{\alpha,\beta} r^\beta_\gamma + r^\beta_{\alpha,\beta})$$
$$= 0, \quad (3.54)$$

thus confirming the correctness of (3.52). If we now introduce the functional Jacobian $\delta[\bar{x}, \phi_0, \psi_0]/\delta[\phi, \psi]$ we may write the group invariant measure for the original variables ϕ^i, ψ^i in the form

$$\Delta[\phi, \psi] = \Delta_0[\phi_0, \psi_0] \det l^{-1}[\bar{x}] \, \delta[\bar{x}, \phi_0, \psi_0]/\delta[\phi, \psi]. \quad (3.55)$$

Using (3.48), (3.49), and (3.51) it is straightforward to show that this functional satisfies Eq. (3.39):

$$(\Delta R^i_\alpha)_{,i} + \Delta_{,i} R^i_\alpha + \Delta R^i_{\alpha,i}$$
$$= \Delta[-l^{-1\beta}_\gamma l^\gamma_{\beta,\delta} \bar{x}^\delta_{,i} R^i_\alpha + (\delta\phi^j/\delta\bar{x}^\beta) \bar{x}^\beta_{,ji} R^i_\alpha$$
$$+ (\delta\phi^j/\delta\phi^k_0)(\phi^k_{0,i} R^i_\alpha)_{,j}$$
$$+ (\delta\psi^j/\delta\psi^k_0)(\psi^k_{0,i} R^i_\alpha + \psi^k_{0,i} R^i_\alpha) + R^i_{\alpha,i}]$$
$$= \Delta[l^{-1\beta}_\gamma l^\gamma_{\beta,\delta} r^\delta_\alpha + (\delta\phi^j/\delta\bar{x}^\beta)(\bar{x}^\beta_{,i} R^i_\alpha)_{,j}$$
$$+ (\delta\phi^j/\delta\phi^k_0)(\phi^k_{0,i} R^i_\alpha)_{,j}$$
$$+ (\delta\psi^j/\delta\psi^k_0)(\psi^k_{0,i} R^i_\alpha + \psi^k_{0,i} R^i_\alpha)_{,j}]$$
$$= \Delta(l^{-1\beta}_\gamma l^\gamma_{\beta,\delta} r^\delta_\alpha - r^\beta_{\alpha,\beta})$$
$$= -\Delta \det l (\det l^{-1} r^\beta_\alpha)_{,\beta} = 0. \quad (3.56)$$

We are now in a position to reformulate the quantization prescription expressed by Eq. (3.1). Let us denote the experimental subspace by R. In terms of the variables $\bar{x}^\alpha, \phi^i_0, \psi^i_0$ it may be expressed in the form

$$R = \prod_{\phi_0, \psi_0 \in R_0} \mathfrak{R}[\phi_0, \psi_0], \quad (3.57)$$

where the symbol $\prod$ denotes the set-theoretic direct product, R_0 is the projection of R on the dynamical line, and $\mathfrak{R}[\phi_0, \psi_0]$ is the intersection of R with the invariant subspace labeled by ϕ^i_0, ψ^i_0. Since A and S are group invariants, and hence independent of $\bar{x}^\alpha$, we may write

$$\langle \mathbf{A} \rangle = \lim_{R \to \infty} \frac{\int_R A e^{iS} \, \delta[\phi] \, \delta[\psi]}{\int_R e^{iS} \, \delta[\phi] \, \delta[\psi]}$$
$$= \lim_{\substack{\mathfrak{R}[\phi_0, \psi_0] \to \text{group} \\ R_0 \to \infty}} \frac{\int_{R_0} \delta[\phi_0] \, \delta[\psi_0] \int_{\mathfrak{R}[\phi_0, \psi_0]} \delta_l[\bar{x}] A e^{iS}}{\int_{R_0} \delta[\phi_0] \, \delta[\psi_0] \int_{\mathfrak{R}[\phi_0, \psi_0]} \delta_l[\bar{x}] e^{iS}}$$
$$= \lim_{R_0 \to \infty} \frac{\int_{R_0} A e^{iS} \, \delta'[\phi_0] \, \delta'[\psi_0]}{\int_{R0} e^{iS} \, \delta'[\phi_0] \, \delta'[\psi_0]}, \quad (3.58)$$

where

$$\delta[\phi_0] \, \delta[\psi_0] \equiv \Delta_0[\phi_0, \psi_0] \prod_A d\phi^A_0 \prod_i d\psi^i_0, \quad (3.59)$$

$$\delta'[\phi_0] \, \delta'[\psi_0] \equiv \Delta'_0[\phi_0, \psi_0] \prod_A d\phi^A_0 \prod_i d\psi^i_0, \quad (3.60)$$

$$\Delta'_0[\phi_0, \psi_0] = \lim \frac{\int_{\mathfrak{R}[\phi_0, \psi_0]} \delta_l[\bar{x}]}{\int_{\mathfrak{R}[0,0]} \delta_l[\bar{x}]} \Delta_0[\phi_0, \psi_0]. \quad (3.61)$$

Thus the Feynman integrals reduce to integrals on the dynamical line. Furthermore since the functional $\Delta_0[\phi_0, \psi_0]$ is so far completely arbitrary we lose no generality by assuming that the original experi-

mental subspace R has uniform cross section in the invariant subspaces, so that the weighting factor in (3.61) reduces to unity.

These results are evidently independent of the choice of the dynamical line, i.e., of the particular choice of supplementary conditions. Indeed, it is easy to show that a change in the supplementary conditions leads to new representative points ϕ_0, ψ_0 which are obtainable from the old by a group transformation. The requisite group parameters are functionals of the ϕ_0^i, ψ_0^i alone, and hence the new representative variables are functionals solely of the old, as is to be expected from the fact that the new variables, like the old, constitute a complete set of invariants. Thus if we consider infinitesimal changes δR_α in the functionals $R_\alpha[\phi_0]$ and write

$$\phi_0'^i = \phi_0^i + \delta\phi_0^i, \qquad \delta\phi_0^i = R_\alpha^i[\phi_0]\ \delta\xi^\alpha, \tag{3.62}$$

$$\psi_0'^i = \psi_0^i + \delta\psi_0^i, \qquad \delta\psi_0^i = R_\alpha^i[\psi_0]\ \delta\xi^\alpha, \tag{3.63}$$

$$\begin{aligned} 0 &= R_\alpha[\phi_0 + \delta\phi_0] + \delta R_\alpha[\phi_0 + \delta\phi_0] \\ &= F_{\alpha\beta}[\phi_0]\ \delta\xi^\beta + \delta R_\alpha[\phi_0], \end{aligned} \tag{3.64}$$

we obtain

$$\delta\xi^\alpha = G^{-\alpha\beta}[\phi_0]\ \delta R_\beta[\phi_0], \tag{3.65}$$

$$\delta\phi_0^i = R_\alpha^i[\phi_0]G^{-\alpha\beta}[\phi_0]\ \delta R_\beta[\phi_0], \tag{3.66}$$

$$\delta\psi_0^i = R_\alpha^i[\psi_0]G^{-\alpha\beta}[\phi_0]\ \delta R_\beta[\phi_0]. \tag{3.67}$$

Group transformations of this kind are often regarded as posing serious problems for the theory because in operator form the parameters (3.65) become q numbers. Here, however, no operator difficulties arise because we are working with the variables of the Feynman integrands, which are c numbers. In passing from one set of supplementary conditions to another we have only to remember to include the Jacobian of the transformation ϕ_0^A, $\psi_0^i \to \phi_0'^A$, $\psi_0'^i$ in the new measure.

This nevertheless does not mean that the Feynman quantization prescription by itself solves such problems as the factor ordering ambiguity, even in principle. This and a multitude of other serious problems, such as unitarity, renormalization, etc., are actually hidden in the question of what choice to make for the measure $\Delta_0[\phi_0, \psi_0]$. The formal theory itself offers no immediate guide to this question. That is to say, *the measure is not directly determined by the group*. It must be determined by the physics. The physics is described by the variables ϕ_0^A, ψ_0^i, which are independent of the group; the variables $\bar{x}^\alpha$ are irrelevant in this description. The only significance which the group has is that it provides (in normal cases) the means by which the physics can be described in terms of a *local* field theory. This significance should not be minimized, but there is no reason to suppose that it is decisive in the question of the measure.

It should finally be mentioned that the introduction of invariant variables ϕ_0^A, ψ_0^i is not merely of theoretical importance but is also of practical utility. In the case of local field theories the supplementary conditions are usually chosen to be linear differential conditions. In momentum space these become algebraic conditions which permit independent invariant variables to be found without difficulty. Momentum space is well suited for perturbation calculations, and the use of the invariant variables is a convenient means for keeping such calculations manifestly group invariant. Actually, perturbation calculations have seldom been carried out in this way. The supplementary conditions are usually employed to *restore* group invariance to a theory which has been previously mutilated by the addition of a non-invariant term of the form $\frac{1}{2}g^{\alpha\beta}R_\alpha R_\beta$ to the action, which necessitates a nonphysical enlargement of the Hilbert space.[20] The invariant variable method, on the other hand, introduces no nonphysical elements. Its utility will be illustrated in a future paper through application to specific examples.

From now on we shall work only with the invariant variables. Although these variables are necessarily nonlocal when the local form of the theory involves an infinite dimensional invariance group, this causes no difficulty for the application of the formal theory. T functionals can be defined in momentum space just as well as in coordinate space by the device of varying external sources. From the point of view of the theory of functionals a Fourier transformation is merely a particular linear transformation.[21]

4. CORRELATION FUNCTIONS AND PROPAGATORS

If the invariance group is removed from the scene by the method indicated in the preceding section the discussion may be resumed on the basis of the procedures developed in Secs. 1 and 2. We may, without loss of generality in fact, revert to the

[20] The equivalence of the invariant variable method to the conventional approach has been demonstrated in detail within the framework of the Feynman quantization scheme by H. van Dam (to be published). The author is indebted to Dr. van Dam for many enlightening discussions on the problem of group invariance.

[21] If the topology of space-time is non-Euclidean, alternative linear transformations will be required.

assumptions of these earlier sections, namely that ϕ^i and ψ^i are themselves invariant, it only being necessary to recognize that the indices i and i may now refer to a continuum other than space-time (e.g., momentum space).

An essential point of distinction between the classical and quantum theories, and one which makes the "rigorous" discussions of the present paper so much more difficult than the quasi-classical considerations of (I) and (II), is the fact that in the quantum theory the mean value of a product is not generally equal to the product of the individual mean values. The type of "mean value" which is convenient to use in quantum field theory is the Schwinger average given by Eq. (3.1), and the distinction referred to may be fully described by the following hierarchy of *correlation functions*:

$$G^{ij\cdots \mathbf{k}\mathbf{l}\cdots} \equiv \frac{\delta}{\delta J_i}\frac{\delta}{\delta J_j}\cdots\frac{\delta}{\delta J_{\mathbf{k}}}\frac{\delta}{\delta J_{\mathbf{l}}}G, \tag{4.1}$$

$$G \equiv -i \ln \langle 0, \infty \mid 0, -\infty\rangle. \tag{4.2}$$

The correlation functions of the two lowest orders play a special role in the theory. The second-order functions G^{ii}, G^{ij}, $G^{\mathbf{i}\mathbf{i}}$, $G^{\mathbf{i}\mathbf{j}}$ are called *propagators* for the individual field quanta, and the first-order functions G^i, $G^{\mathbf{i}}$ are conveniently given the special notational designation

$$G^i = \langle \phi^i \rangle = \frac{\langle 0, \infty \mid \phi^i \mid 0, -\infty\rangle}{\langle 0, \infty \mid 0, -\infty\rangle} \equiv \phi^i, \tag{4.3}$$

$$G^{\mathbf{i}} = \langle \psi^{\mathbf{i}} \rangle = \frac{\langle 0, \infty \mid \psi^{\mathbf{i}} \mid 0, -\infty\rangle}{\langle 0, \infty \mid 0, -\infty\rangle} \equiv \psi^{\mathbf{i}}. \tag{4.4}$$

Although the classical symbols ϕ^i, $\psi^{\mathbf{i}}$ have already been used in the Feynman integrals, their reintroduction here to denote also the mean values $\langle\phi^i\rangle$, $\langle\psi^{\mathbf{i}}\rangle$ is deliberate, for it turns out, as we shall see in the next section, that the quantum ϕ^i, $\psi^{\mathbf{i}}$, like the classical ϕ^i, $\psi^{\mathbf{i}}$, are solutions of a set of *c-number* equations derivable from a variational principle. In the present case the variational principle is based on a c-number action functional Γ, of which the propagators G^{ii}, $G^{\mathbf{i}\mathbf{j}}$, etc., are Green's functions. The difference between the classical and quantum variational principles consists in the fact that the quantum ϕ^i, $\psi^{\mathbf{i}}$ are *complex* even when the classical ϕ^i, $\psi^{\mathbf{i}}$ are real, and in the fact that Γ leads to *nonlocal* equations even when S (in the absence of infinite dimensional invariance groups) yields local ones.

The functions $G^{i\mathbf{i}}$, $G^{i\mathbf{j}}$ etc., describe the linear response of the quantized system and, in virtue of Eqs. (4.3), (4.4), and the relation

$$\begin{bmatrix} G^{ii} & G^{i\mathbf{j}} \\ G^{\mathbf{i}i} & G^{\mathbf{i}\mathbf{j}} \end{bmatrix} = \begin{bmatrix} \delta/\delta J_i \\ \delta/\delta J_{\mathbf{i}} \end{bmatrix} (\phi^i \psi^{\mathbf{j}}), \tag{4.5}$$

evidently satisfy the boundary conditions of Feynman propagators. They also satisfy the symmetry relations

$$G^{ii} = G^{ii}, \qquad G^{i\mathbf{j}} = G^{\mathbf{j}i}, \qquad G^{\mathbf{i}\mathbf{j}} = -G^{\mathbf{j}\mathbf{i}}. \tag{4.6}$$

Vacuum-to-vacuum matrix elements of T functionals may be expressed in terms of the correlation functions by the following device. We introduce a set of parameters ξ_i, $\eta_{\mathbf{i}}$ having the same commutation properties as the sources J_i, $J_{\mathbf{i}}$ and write

$$\begin{aligned}
&\sum_{m,n=0}^{\infty} \frac{1}{m!\,n!}\, \eta_{\mathbf{j}_n}\cdots\eta_{\mathbf{j}_1}\xi_{i_m}\cdots\xi_{i_1} \\
&\qquad \times \langle 0, \infty \mid T^{i_1\cdots i_m \mathbf{j}_1\cdots\mathbf{j}_n} \mid 0, -\infty\rangle \\
&\quad = \exp\left(\xi_i \frac{\delta}{i\,\delta J_i} + \eta_{\mathbf{i}}\frac{\delta}{i\,\delta J_{\mathbf{i}}}\right)\langle 0, \infty \mid 0, -\infty\rangle \\
&\quad = (e^{iG})_{J_i\to J_i - i\xi_i,\, J_{\mathbf{i}}\to J_{\mathbf{i}} - i\eta_{\mathbf{i}}} \\
&\quad = \exp\left[i \sum_{m,n=0}^{\infty} \frac{(-i)^{m+n}}{m!\,n!}\, \eta_{\mathbf{j}_n}\cdots\eta_{\mathbf{j}_1}\xi_{i_m}\cdots\xi_{i_1} \right. \\
&\qquad \left. \times\, G^{i_1\cdots i_m \mathbf{j}_1\cdots\mathbf{j}_n} \right].
\end{aligned} \tag{4.7}$$

If we divide this equation by the vacuum-to-vacuum amplitude $\langle 0, \infty \mid 0, -\infty\rangle$ we get

$$\begin{aligned}
&\sum_{m,n=0}^{\infty} \frac{1}{m!\,n!}\, \eta_{\mathbf{j}_n}\cdots\eta_{\mathbf{j}_1}\xi_{i_m}\cdots\xi_{i_1}\langle \phi^{i_1}\cdots\phi^{i_m}\psi^{\mathbf{j}_1}\cdots\psi^{\mathbf{j}_n}\rangle \\
&\quad = \exp\left[\xi_i\phi^i + \eta_{\mathbf{i}}\psi^{\mathbf{i}} + i \sum_{m+n\geq 2} \frac{(-i)^{m+n}}{m!\,n!} \right. \\
&\qquad \left. \times\, \eta_{\mathbf{j}_n}\cdots\eta_{\mathbf{j}_1}\xi_{i_m}\cdots\xi_{i_1} G^{i_1\cdots i_m \mathbf{j}_1\cdots\mathbf{j}_n} \right].
\end{aligned} \tag{4.8}$$

By picking out terms of equal degree in the ξ's and η's on both sides of this equation, we obtain expressions for the Schwinger averages of T functionals in terms of the correlation functions. For example,

$$\left.\begin{aligned}
\langle \phi^i\phi^j\rangle &= -iG^{ij} + \phi^i\phi^j, \\
\langle \phi^i\phi^j\phi^k\rangle &= (-i)^2 G^{ijk} - iG^{ij}\phi^k - iG^{jk}\phi^i \\
&\quad - iG^{ki}\phi^j + \phi^i\phi^j\phi^k, \quad \text{etc.},
\end{aligned}\right\} \tag{4.9}$$

with corresponding expressions in which the light-faced indices are replaced in all possible ways by boldface ones.

For the Schwinger average of the chronologically ordered form of an arbitrary operator **A** we have

$$\langle \mathbf{A} \rangle = \sum_{m,n=0}^{\infty} \frac{1}{m!\,n!} A^{0}_{,j_n \cdots j_1 i_m \cdots i_1} \times \langle \phi^{i_1} \cdots \phi^{i_m} \psi^{j_1} \cdots \psi^{j_n} \rangle$$
$$= \Big(\bar{A} \sum_{m,n=0}^{\infty} \frac{1}{m!\,n!} \frac{\overleftarrow{\delta}}{\delta \bar{\psi}^{j_n}} \cdots \frac{\overleftarrow{\delta}}{\delta \bar{\psi}^{j_1}} \frac{\overleftarrow{\delta}}{\delta \bar{\phi}^{i_m}} \cdots \times \frac{\overleftarrow{\delta}}{\delta \bar{\phi}^{i_1}} \langle \phi^{i_1} \cdots \phi^{i_m} \psi^{j_1} \cdots \psi^{j_n} \rangle \Big)_{\bar{\phi},\bar{\psi}=0}$$
$$= \Big(\bar{A} \exp \Big[i \sum_{m+n \geq 2} \frac{(-i)^{m+n}}{m!\,n!} \frac{\overleftarrow{\delta}}{\delta \bar{\psi}^{j_n}} \cdots \frac{\overleftarrow{\delta}}{\delta \bar{\psi}^{j_1}} \times \frac{\overleftarrow{\delta}}{\delta \bar{\phi}^{i_m}} \cdots \frac{\overleftarrow{\delta}}{\delta \bar{\phi}^{i_1}} G^{i_1 \cdots i_m j_1 \cdots j_n} \Big] \Big)_{\bar{\phi}=\phi,\bar{\psi}=\psi}, \tag{4.10}$$

where $\bar{A}$ is an abbreviation for $A[\bar{\phi}, \bar{\psi}]$ and the derivatives with respect to the barred variables are to be understood as right derivatives which act to the *left* in the order in which they appear from left to right. Equation (4.10) becomes clearer in expanded form. For example, if **A** depends only on the ϕ^i, then

$$\langle \mathbf{A} \rangle = A + \frac{(-i)}{2!} A_{,ji} G^{ij} + \frac{(-i)^2}{3!} A_{,kji} G^{ijk} + \frac{(-i)^3}{4!} A_{,lkji} (G^{ijkl} + 3iG^{ij}G^{kl}) + \cdots, \tag{4.11}$$

which expresses the Schwinger average in terms of the correlation functions of order two and higher, together with the "classical" quantities A, $A_{,ij}$, etc. evaluated with $\phi^i = \langle \phi^i \rangle$. The case in which **A** depends also on the ψ^i is obtained by a straightforward inclusion of boldface as well as lightfaced indices.

With the aid of Eq. (4.10) it is not difficult to show that

$$\left.\begin{aligned} \frac{\delta}{\delta J_i} \langle \mathbf{A} \rangle &= \mathfrak{A}_j G^{ji} + \mathfrak{A}_j G^{ji}, \\ \frac{\delta}{\delta J_i} \langle \mathbf{A} \rangle &= \pm \mathfrak{A}_j G^{ji} \pm \mathfrak{A}_j G^{ji}, \end{aligned}\right\} \tag{4.12}$$

where

$$\left.\begin{aligned} \mathfrak{A}_j &\equiv \tfrac{1}{2} \sum_{m,n=0}^{\infty} 2^{\delta_{n0}} \frac{(-i)^{m+n}}{(m+1)!\,n!} \langle \mathbf{A}_{,j_n \cdots j_1 j i_m \cdots i_1} \rangle \times \frac{\delta}{\delta J_{i_1}} \cdots \frac{\delta}{\delta J_{i_m}} \frac{\delta}{\delta J_{j_1}} \cdots \frac{\delta}{\delta J_{j_n}}, \\ \mathfrak{A}_j &\equiv \tfrac{1}{2} \sum_{m,n=0}^{\infty} 2^{\delta_{m0}} \frac{(-i)^{m+n}}{m!\,(n+1)!} \langle \mathbf{A}_{,j j_n \cdots j_1 i_m \cdots i_1} \rangle \times \frac{\delta}{\delta J_{i_1}} \cdots \frac{\delta}{\delta J_{i_m}} \frac{\delta}{\delta J_{j_1}} \cdots \frac{\delta}{\delta J_{j_n}}, \end{aligned}\right\} \tag{4.13}$$

and where the upper or lower sign is taken according as **A** is of the commuting or anticommuting type. The proof of this useful lemma will be left to the reader.

5. THE FUNCTIONAL Γ. RENORMALIZATION

The functionals of the preceding section may be regarded as depending either on the ϕ^i, ψ^i or on the J_i, J_i. The transformation coefficients from one set of variables to the other are the propagators (4.5). The explicit functional relation between the two sets of variables is determined by the Schwinger averages of Eqs. (2.5) and (2.6):

$$\langle \mathbf{S}_{,i} \rangle = -J_i, \qquad \langle \mathbf{S}_{,i} \rangle = -J_i. \tag{5.1}$$

Under a variation in the independent variables we have

$$\delta(\langle \mathbf{S}_{,i} \rangle \quad \langle \mathbf{S}_{,i} \rangle) = (\delta J_i \ \delta J_j) \begin{bmatrix} \delta/\delta J_i \\ \delta/\delta J_j \end{bmatrix} (\langle \mathbf{S}_{,i} \rangle \quad \langle \mathbf{S}_{,i} \rangle)$$
$$= (\delta\phi^k \ \delta\psi^k) \begin{bmatrix} \langle \mathbf{S}_{,i} \rangle_{,k} & \langle \mathbf{S}_{,i} \rangle_{,k} \\ -\langle \mathbf{S}_{,i} \rangle_{,k} & \langle \mathbf{S}_{,i} \rangle_{,k} \end{bmatrix}$$
$$= (\delta J_i \ \delta J_j) \begin{bmatrix} G^{ik} & -G^{ik} \\ G^{jk} & -G^{jk} \end{bmatrix} \begin{bmatrix} \langle \mathbf{S}_{,i} \rangle_{,k} & \langle \mathbf{S}_{,i} \rangle_{,k} \\ \langle \mathbf{S}_{,i} \rangle_{,k} & -\langle \mathbf{S}_{,i} \rangle_{,k} \end{bmatrix} \tag{5.2}$$

and hence, in virtue of Eqs. (5.1),

$$\begin{bmatrix} G^{ik} & -G^{ik} \\ G^{jk} & -G^{jk} \end{bmatrix} \begin{bmatrix} \langle \mathbf{S}_{,i} \rangle_{,k} & \langle \mathbf{S}_{,i} \rangle_{,k} \\ \langle \mathbf{S}_{,i} \rangle_{,k} & -\langle \mathbf{S}_{,i} \rangle_{,k} \end{bmatrix} = \begin{bmatrix} -\delta^i_i & 0 \\ 0 & -\delta^j_i \end{bmatrix}. \tag{5.3}$$

The matrix formed from the $\langle \mathbf{S}_{,i} \rangle_{,k}$, etc., is thus seen to be the negative inverse of the matrix formed from the propagators. Although the former matrix is generally nonlocal (even when no infinite dimensional invariance groups are present in the original formulation of the theory) it is nevertheless equal in zeroth approximation to the matrix

$$\begin{bmatrix} S_{,ki} & S_{,ki} \\ S_{,ki} & S_{,ki} \end{bmatrix}$$

and has therefore the general characteristics of a wave operator. It is consequently the *unique* inverse of the propagator matrix. Furthermore, from arguments of the type already introduced in (II) one may infer that Eq. (5.3) implies also the equation

$$\begin{bmatrix} \langle \mathbf{S}_{,k} \rangle_{,i} & \langle \mathbf{S}_{,k} \rangle_{,i} \\ \langle \mathbf{S}_{,k} \rangle_{,i} & -\langle \mathbf{S}_{,k} \rangle_{,i} \end{bmatrix} \begin{bmatrix} G^{kj} & -G^{kj} \\ G^{kj} & -G^{kj} \end{bmatrix} = \begin{bmatrix} -\delta^i_i & 0 \\ 0 & -\delta^j_i \end{bmatrix}, \tag{5.4}$$

just as in the case of finite matrices. From Eqs. (5.3) and (5.4) and the symmetry relations (4.6) the following identities may therefore be deduced:

$$\langle \mathbf{S}_{,i}\rangle_{,k} = \langle \mathbf{S}_{,k}\rangle_{,i}, \qquad \langle \mathbf{S}_{,i}\rangle_{,\mathbf{k}} = \langle \mathbf{S}_{,\mathbf{k}}\rangle_{,i}, \qquad \langle \mathbf{S}_{,\mathbf{i}}\rangle_{,\mathbf{k}} = -\langle \mathbf{S}_{,\mathbf{k}}\rangle_{,\mathbf{i}}. \tag{5.5}$$

But these identities imply the existence of a functional Γ such that

$$\langle \mathbf{S}_{,i}\rangle = \Gamma_{,i}, \qquad \langle \mathbf{S}_{,\mathbf{i}}\rangle = \Gamma_{,\mathbf{i}}. \tag{5.6}$$

In terms of this functional the basic equations of the theory become

$$\Gamma_{,i} = -J_i, \qquad \Gamma_{,\mathbf{i}} = -J_{\mathbf{i}}, \tag{5.7}$$

$$\begin{pmatrix} \Gamma_{,ik} & \Gamma_{,i\mathbf{k}} \\ \Gamma_{,\mathbf{i}k} & \Gamma_{,\mathbf{i}\mathbf{k}} \end{pmatrix} \begin{pmatrix} G^{kj} & -G^{k\mathbf{j}} \\ G^{\mathbf{k}j} & -G^{\mathbf{k}\mathbf{j}} \end{pmatrix} = \begin{pmatrix} G^{ik} & -G^{i\mathbf{k}} \\ G^{\mathbf{j}k} & -G^{\mathbf{j}\mathbf{k}} \end{pmatrix} \begin{pmatrix} \Gamma_{,ki} & \Gamma_{,k\mathbf{i}} \\ \Gamma_{,\mathbf{k}i} & \Gamma_{,\mathbf{k}\mathbf{i}} \end{pmatrix} = \begin{pmatrix} -\delta_i^j & 0 \\ 0 & -\delta_{\mathbf{i}}^{\mathbf{j}} \end{pmatrix}. \tag{5.8}$$

The higher-order correlation functions may be re-expressed in terms of propagators and derivatives of the functional Γ by repeated use of the identity

$$\delta \begin{pmatrix} G^{ij} & -G^{i\mathbf{j}} \\ G^{\mathbf{i}j} & -G^{\mathbf{i}\mathbf{j}} \end{pmatrix} = \begin{pmatrix} G^{ik} & -G^{i\mathbf{k}} \\ G^{\mathbf{i}k} & -G^{\mathbf{i}\mathbf{k}} \end{pmatrix} \times \delta \begin{pmatrix} \Gamma_{,kl} & \Gamma_{,k\mathbf{l}} \\ \Gamma_{,\mathbf{k}l} & \Gamma_{,\mathbf{k}\mathbf{l}} \end{pmatrix} \begin{pmatrix} G^{lj} & -G^{l\mathbf{j}} \\ G^{\mathbf{l}j} & -G^{\mathbf{l}\mathbf{j}} \end{pmatrix}. \tag{5.9}$$

Before showing how this comes about we remark that although the identity (5.9) always holds in the case of finite matrices it does *not* hold for all Green's functions in a space-time of indefinite metric. That it does hold for the Feynman propagators, at least in a space-time which is asymptotically flat, is a consequence of the special boundary conditions on these propagators (i.e., positive frequencies in the remote future, negative frequencies in the remote past) which are preserved by Eq. (5.9). The Feynman propagators are, in fact, the *unique* Green's functions which satisfy *both* Eq. (5.9) and the symmetry conditions (4.6).[22] This means that in formal manipulations the Feynman propagators always follow the rules of finite matrix theory, a fact which stems from the close relation (via analytic continuation) which the Feynman propagators have to the unique Green's functions of manifolds with definite metric. But it also suggests an inverted view of the Feynman propagators. It should be possible to *demand* that these propagators satisfy both (4.6) and (5.9) and hence to define them uniquely in contexts in which the concepts of positive and negative frequencies and a unique vacuum are inappropriate, e.g., in the case of space-times having nonstatic asymptotic curvature or space-like cross sections of non-Euclidean and even dynamically changing topology.

If we now expand the left- and right-hand sides of Eq. (5.9) we get

$$\begin{aligned} &\delta J_k \begin{pmatrix} G^{ijk} & -G^{ij\mathbf{k}} \\ G^{\mathbf{i}jk} & -G^{\mathbf{i}j\mathbf{k}} \end{pmatrix} + \delta J_{\mathbf{k}} \begin{pmatrix} G^{ij\mathbf{k}} & G^{ij\mathbf{k}} \\ -G^{\mathbf{i}j\mathbf{k}} & -G^{\mathbf{i}j\mathbf{k}} \end{pmatrix} \\ &= \begin{pmatrix} G^{il} & -G^{i\mathbf{l}} \\ G^{\mathbf{i}l} & -G^{\mathbf{i}\mathbf{l}} \end{pmatrix} \left[\delta\phi^n \begin{pmatrix} \Gamma_{,lmn} & \Gamma_{,lm\mathbf{n}} \\ \Gamma_{,l\mathbf{m}n} & \Gamma_{,l\mathbf{m}\mathbf{n}} \end{pmatrix} + \delta\psi^{\mathbf{n}} \begin{pmatrix} -\Gamma_{,lmn} & \Gamma_{,lm\mathbf{n}} \\ \Gamma_{,l\mathbf{m}n} & -\Gamma_{,l\mathbf{m}\mathbf{n}} \end{pmatrix} \right] \begin{pmatrix} G^{mj} & -G^{m\mathbf{j}} \\ G^{\mathbf{m}j} & -G^{\mathbf{m}\mathbf{j}} \end{pmatrix} \\ &= \left\{ \delta J_k \begin{pmatrix} G^{il} & -G^{i\mathbf{l}} \\ G^{\mathbf{i}l} & -G^{\mathbf{i}\mathbf{l}} \end{pmatrix} \left[G^{kn} \begin{pmatrix} \Gamma_{,lmn} & \Gamma_{,lm\mathbf{n}} \\ \Gamma_{,l\mathbf{m}n} & \Gamma_{,l\mathbf{m}\mathbf{n}} \end{pmatrix} + G^{k\mathbf{n}} \begin{pmatrix} -\Gamma_{,lmn} & \Gamma_{,lm\mathbf{n}} \\ \Gamma_{,l\mathbf{m}n} & -\Gamma_{,l\mathbf{m}\mathbf{n}} \end{pmatrix} \right] \right. \\ &\quad + \delta J_{\mathbf{k}} \begin{pmatrix} G^{il} & G^{i\mathbf{l}} \\ -G^{\mathbf{i}l} & -G^{\mathbf{i}\mathbf{l}} \end{pmatrix} \left[G^{\mathbf{k}n} \begin{pmatrix} \Gamma_{,lmn} & \Gamma_{,lm\mathbf{n}} \\ \Gamma_{,l\mathbf{m}n} & \Gamma_{,l\mathbf{m}\mathbf{n}} \end{pmatrix} \right. \\ &\quad \left. \left. + G^{\mathbf{k}\mathbf{n}} \begin{pmatrix} -\Gamma_{,lmn} & \Gamma_{,lm\mathbf{n}} \\ \Gamma_{,l\mathbf{m}n} & -\Gamma_{,l\mathbf{m}\mathbf{n}} \end{pmatrix} \right] \right\} \\ &\quad \times \begin{pmatrix} G^{im} & -G^{i\mathbf{m}} \\ G^{\mathbf{i}m} & G^{\mathbf{i}\mathbf{m}} \end{pmatrix}, \end{aligned} \tag{5.10}$$

from which expressions for the third-order correlation functions may be directly extracted. Equation (5.10) and the equations which follow from it through repeated differentiation with respect to the sources all have simple graphical representations. Ignoring signs and anticommutativity we may represent the propagators by lines and the derivatives of Γ by vertices or forks having prongs equal in number to the number of functional differentiations. The correlation functions are then represented by diagrams in which lines are joined together at vertices in the same ways that the propagators in the explicit expressions are coupled to derivatives of Γ by dummy indices. It is easy to see that differentiation with respect to a source corresponds to the insertion of an external line in all possible ways into a given diagram (including

[22] Other Green's functions may satisfy one or the other of these conditions but not both. For example, the retarded and advanced Green's functions satisfy (5.9) but not (4.6), whereas the average of the retarded and advanced Green's functions satisfies (4.6) but not (5.9).

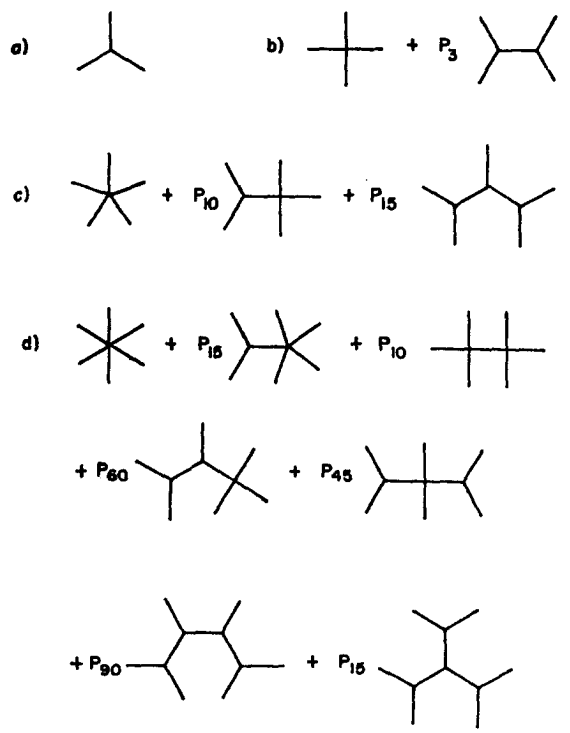


FIG. 2. Diagrammatic representation of correlation functions: (a) third order, (b) fourth order, (c) fifth order, (d) sixth order. The symbol P indicates that the indices associated with the external lines are to be permuted just sufficiently to yield complete symmetry. The numerical subscript indicates the number of permutations required in each case.

at pre-existing vertices). Similarly, differentiation with respect to a variable ϕ^i or ψ^i corresponds to the insertion of a vertex prong in all possible ways. Iteration of this rule leads to Fig. 2 which depicts the correlation functions of orders 3, 4, 5, and 6. It is seen that each correlation function is expressible as the sum of all *simply connected*[23] diagrams having a fixed number of external lines, the indices attached to the latter being permuted just sufficiently to yield the appropriate symmetry.

When the anticommuting sources J_i vanish it is possible to make a diagrammatic distinction between fermion and boson propagators, the former being conventionally represented by solid lines and the latter by dotted or wavy lines. This is because Γ, like S, is necessarily of even degree in the ψ^i so that $\Gamma_{,ij}$, G^{ij}, ψ^i, and, in fact, all anticommuting quantities vanish when $J_i = 0$. The fermion lines must then join together in pairs at every vertex, forming, throughout a given diagram, a network of continuous lines which can be divided into two groups: (1) those which terminate only in external lines, and (2) those which close upon themselves within the interior of the diagram.

The functional derivatives of Γ of order three and higher are called *vertex functions*. These are the *full* irreducible vertex functions of the theory. The so called *bare* vertex functions are the corresponding derivatives of the original classical action S.

The c-number theory based on Γ is seen to be a nonlocal theory having a form or initial "direction" which is set, via the correspondence principle, by a limiting theory based on S. The classical field theory may be regarded as a kind of "tangent theory," the relevance of which as a starting point for the rigorous quantum theory depends on two crucial conditions: (1) the *renormalizability* of the quantum theory, and (2) the applicability of perturbation theory so that Γ differs from S only by a small amount which describes pure quantum effects (e.g., vacuum polarization).

It is possible to obtain a number of interesting expansions by various formal manipulations of the quantities thus far introduced. We mention only one of these here: If we differentiate Eqs. (5.1) and make use of the lemma (4.12), (4.13), we obtain

$$\begin{bmatrix} \mathfrak{F}_{ik} & \mathfrak{F}_{ik} \\ \mathfrak{F}_{ik} & \mathfrak{F}_{ik} \end{bmatrix} \begin{bmatrix} G^{kj} & -G^{kj} \\ G^{kj} & -G^{kj} \end{bmatrix} = \begin{bmatrix} -\delta_i^j & 0 \\ 0 & -\delta_i^j \end{bmatrix}, \tag{5.11}$$

where

$$\left.\begin{aligned}
\mathfrak{F}_{ij} &\equiv \tfrac{1}{2} \sum_{m,n=0}^{\infty} 2^{\delta_{n0}} \frac{(-1)^{m+n}}{(m+1)!n!} \\
&\quad \times \langle \mathbf{S}_{,ij\mathbf{j}_n \cdots \mathbf{j}_1 \mathbf{i}_m \cdots \mathbf{i}_1} \rangle \frac{\delta}{\delta J_{i_1}} \cdots \frac{\delta}{\delta J_{i_m}} \frac{\delta}{\delta J_{\mathbf{j}_1}} \cdots \frac{\delta}{\delta J_{\mathbf{j}_n}}, \\
\mathfrak{F}_{ij} &\equiv \tfrac{1}{2} \sum_{m,n=0}^{\infty} 2^{\delta_{m0}} \frac{(-i)^{m+n}}{m!(n+1)!} \\
&\quad \times \langle \mathbf{S}_{,ij\mathbf{j}_n \cdots \mathbf{j}_1 \mathbf{i}_m \cdots \mathbf{i}_1} \rangle \frac{\delta}{\delta J_{i_1}} \cdots \frac{\delta}{\delta J_{i_m}} \frac{\delta}{\delta J_{\mathbf{j}_1}} \cdots \frac{\delta}{\delta J_{\mathbf{j}_n}}, \\
\mathfrak{F}_{ij} &\equiv \tfrac{1}{2} \sum_{m,n=0}^{\infty} 2^{\delta_{n0}} \frac{(-i)^{m+n}}{(m+1)!n!} \\
&\quad \times \langle \mathbf{S}_{,ij\mathbf{j}_n \cdots \mathbf{j}_1 \mathbf{i}_m \cdots \mathbf{i}_1} \rangle \frac{\delta}{\delta J_{i_1}} \cdots \frac{\delta}{\delta J_{i_m}} \frac{\delta}{\delta J_{\mathbf{j}_1}} \cdots \frac{\delta}{\delta J_{\mathbf{j}_n}}, \\
\mathfrak{F}_{ij} &\equiv \tfrac{1}{2} \sum_{m,n=0}^{\infty} 2^{\delta_{m0}} \frac{(-i)^{m+n}}{m!(n+1)!} \\
&\quad \times \langle \mathbf{S}_{,ij\mathbf{j}_n \cdots \mathbf{j}_1 \mathbf{i}_m \cdots \mathbf{i}_1} \rangle \frac{\delta}{\delta J_{i_1}} \cdots \frac{\delta}{\delta J_{i_m}} \frac{\delta}{\delta J_{\mathbf{j}_1}} \cdots \frac{\delta}{\delta J_{\mathbf{j}_n}}.
\end{aligned}\right\} \tag{5.12}$$

This result may then be used to write

[23] A simply connected diagram is one which has no disconnected parts but which is divided into two disconnected parts by cutting any line.

$$\begin{aligned}
\begin{bmatrix} \langle \mathbf{S}_{,ij} \rangle & \langle \mathbf{S}_{,i\mathbf{j}} \rangle \\ \langle \mathbf{S}_{,\mathbf{i}j} \rangle & \langle \mathbf{S}_{,\mathbf{ij}} \rangle \end{bmatrix} &= \begin{bmatrix} \mathfrak{F}_{ij} & \mathfrak{F}_{i\mathbf{j}} \\ \mathfrak{F}_{\mathbf{i}j} & \mathfrak{F}_{\mathbf{ij}} \end{bmatrix} \cdot 1 \\
&= -\begin{bmatrix} \mathfrak{F}_{ik} & \mathfrak{F}_{i\mathbf{k}} \\ \mathfrak{F}_{\mathbf{i}k} & \mathfrak{F}_{\mathbf{ik}} \end{bmatrix} \left[\begin{bmatrix} G^{kl} & -G^{k\mathbf{l}} \\ G^{\mathbf{k}l} & -G^{\mathbf{kl}} \end{bmatrix} \begin{bmatrix} \Gamma_{,lj} & \Gamma_{,l\mathbf{j}} \\ \Gamma_{,\mathbf{l}j} & \Gamma_{,\mathbf{lj}} \end{bmatrix} \right] \\
&= -\left[\begin{bmatrix} \mathfrak{F}_{ik} & \mathfrak{F}_{i\mathbf{k}} \\ \mathfrak{F}_{\mathbf{i}k} & \mathfrak{F}_{\mathbf{ik}} \end{bmatrix} \begin{bmatrix} G^{kl} & -G^{k\mathbf{l}} \\ G^{\mathbf{k}l} & -G^{\mathbf{kl}} \end{bmatrix} \right] \begin{bmatrix} \Gamma_{,lj} & \Gamma_{,l\mathbf{j}} \\ \Gamma_{,\mathbf{l}j} & \Gamma_{,\mathbf{lj}} \end{bmatrix} \\
&\quad + \left[\begin{bmatrix} \mathfrak{F}_{ik} - \langle \mathbf{S}_{,ik} \rangle & \mathfrak{F}_{i\mathbf{k}} - \langle \mathbf{S}_{,i\mathbf{k}} \rangle \\ \mathfrak{F}_{\mathbf{i}k} - \langle \mathbf{S}_{,\mathbf{i}k} \rangle & \mathfrak{F}_{\mathbf{ik}} - \langle \mathbf{S}_{,\mathbf{ik}} \rangle \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} G^{kl} & -G^{k\mathbf{l}} \\ G^{\mathbf{k}l} & -G^{\mathbf{kl}} \end{bmatrix} \right] \begin{bmatrix} \Gamma_{,lj} & \Gamma_{,l\mathbf{j}} \\ \Gamma_{,\mathbf{l}j} & \Gamma_{,\mathbf{lj}} \end{bmatrix}, \qquad (5.13)
\end{aligned}$$

whence

$$\begin{aligned}
\begin{bmatrix} \Gamma_{,ij} & \Gamma_{,i\mathbf{j}} \\ \Gamma_{,\mathbf{i}j} & \Gamma_{,\mathbf{ij}} \end{bmatrix} &= \begin{bmatrix} \langle \mathbf{S}_{,ij} \rangle & \langle \mathbf{S}_{,i\mathbf{j}} \rangle \\ \langle \mathbf{S}_{,\mathbf{i}j} \rangle & \langle \mathbf{S}_{,\mathbf{ij}} \rangle \end{bmatrix} \\
&\quad + \begin{bmatrix} X^k_i & -X^{\mathbf{k}}_i \\ X^k_{\mathbf{i}} & -X^{\mathbf{k}}_{\mathbf{i}} \end{bmatrix} \begin{bmatrix} \Gamma_{,kj} & \Gamma_{,k\mathbf{j}} \\ \Gamma_{,\mathbf{k}j} & \Gamma_{,\mathbf{kj}} \end{bmatrix}, \qquad (5.14)
\end{aligned}$$

where

$$\begin{aligned}
X^i_i &\equiv i \sum_{m+n\geq 2} \frac{(-i)^{m+n}}{m!\,n!} \\
&\quad \times \langle \mathbf{S}_{,ij_n \cdots j_1 i_m \cdots i_1} \rangle G^{i_1 \cdots i_m j_1 \cdots j_n i}, \qquad (5.15)
\end{aligned}$$

and where $X^i_{\mathbf{i}}$, $X^{\mathbf{i}}_i$, $X^{\mathbf{i}}_{\mathbf{i}}$ are defined by the same expression with the lightfaced indices i and/or j rewritten in boldface. If no fermion fields are present (5.14) reduces to

$$\begin{aligned}
\Gamma_{,ij} &= \langle \mathbf{S}_{,ij} \rangle + \frac{(-i)}{2!} \langle \mathbf{S}_{,ikl} \rangle G^{klm} \Gamma_{,mj} \\
&\quad + \frac{(-i)^2}{3!} \langle \mathbf{S}_{,iklm} \rangle G^{klmn} \Gamma_{,nj} + \cdots . \qquad (5.16)
\end{aligned}$$

Equation (5.14) is an integral equation which may be solved by iteration to obtain the $\Gamma_{,ij}$. $\Gamma_{,i\mathbf{j}}$, etc., in terms of the $\langle \mathbf{S}_{,ij} \rangle$, $\langle \mathbf{S}_{,ij\mathbf{k}} \rangle$, etc., and hence, through application of Eq. (4.10), in terms of the functional derivatives $S_{,ij}$, $S_{,ij\mathbf{k}}$, etc., of the original classical action.

It is important to note that because the X's contain the correlation functions as well as the $S_{,ij}$, $S_{,ij\mathbf{k}}$, etc. Equation (5.14) has a hybrid character involving both the full vertex functions and the bare vertex functions. Such a situation is actually an inconvenience in the practical application of the theory, when infinite renormalizations must be performed, since it is related to the difficulty of overlapping divergences. This difficulty can be avoided by first getting rid of the bare vertex functions, and the discussion of the whole renormalization program is then greatly simplified. Fortunately the procedure for doing this is straightforward. Instead of starting from Eq. (5.14) we go back to Eq. (4.10), which we note may be expressed in the form

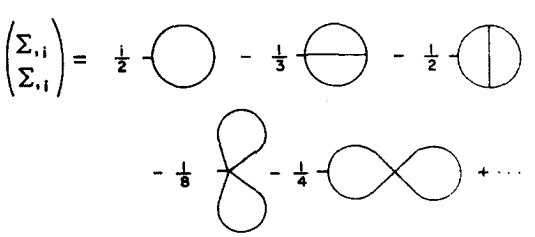


FIG. 3. Diagrammatic representation of the first derivative of the self-energy functional (radiative corrections to the classical field equations).

$$\langle \mathbf{A} \rangle = (1 + \mathfrak{D}) A, \qquad (5.17)$$

where $\mathfrak{D}$ is a sum of functional differential operators having combinations of the correlation functions as coefficients. The advantage of this equation is that it may (formally at least) be inverted, allowing us to write in particular

$$\left.\begin{aligned} S_{,i} &= (1 + \mathfrak{D})^{-1} \Gamma_{,i}, \\ S_{,\mathbf{i}} &= (1 + \mathfrak{D})^{-1} \Gamma_{,\mathbf{i}}. \end{aligned}\right\} \qquad (5.18)$$

Everything on the right of these equations is immediately expressible in terms of propagators and full vertex functions only.

Expanding the inverse $(1 + \mathfrak{D})^{-1}$ by the binomial theorem, expressing the correlation functions in terms of the vertex functions as in Fig. 2, and performing the required functional differentiations, one obtains equations of the form

$$\left.\begin{aligned} S_{,i} &= \Gamma_{,i} + \Sigma_i \\ S_{,\mathbf{i}} &= \Gamma_{,\mathbf{i}} + \Sigma_{\mathbf{i}} \end{aligned}\right\} \qquad (5.19)$$

where Σ_i and $\Sigma_{\mathbf{i}}$ can be represented as a sum of diagrams the first few terms of which are pictured in Fig. 3. It is clear that Σ_i and $\Sigma_{\mathbf{i}}$ must themselves be derivatives of some formal functional Σ:

$$\Sigma_i = \Sigma_{,i}, \qquad \Sigma_{\mathbf{i}} = \Sigma_{,\mathbf{i}} \qquad (5.20)$$

so that

$$\Gamma = S - \Sigma. \qquad (5.21)$$

Indeed, remembering that functional differentiation with respect to ϕ^i and $\psi^{\mathbf{i}}$ corresponds to the insertion of a vertex prong in all possible ways into a given diagram, one may verify this fact directly from the

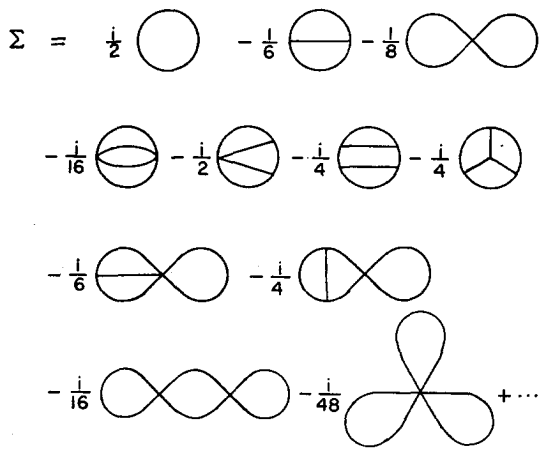


FIG. 4. Diagrammatic representation of the self-energy functional.

diagrams. The first few terms in the expansion of Σ itself are pictured in Fig. 4.[24] The author has not succeeded in obtaining any simple closed expression for the general term of this expansion, which is unfortunate since the expansion sums up the basic formal structure of all quantum field theories. However, this does not prevent us from reaching important conclusions about the theory.

The continuous matrix

$$\begin{bmatrix} \Sigma_{,ij} & \Sigma_{,ij} \\ \Sigma_{,ij} & \Sigma_{,ij} \end{bmatrix}$$

is known as the *mass operator* or *self-energy operator*, and we shall conform to this convention by calling Σ the *self-energy functional* even though it describes all radiative corrections to the classical theory and not merely the mass corrections. Actually the mass corrections are the least important because they are subtracted. To see how the renormalization program works we consider the equation

$$\begin{bmatrix} \Gamma_{,ij} & \Gamma_{,ij} \\ \Gamma_{,ij} & \Gamma_{,ij} \end{bmatrix} = \begin{bmatrix} S_{,ij} & S_{,ij} \\ S_{,ij} & S_{,ij} \end{bmatrix} - \begin{bmatrix} \Sigma_{,ij} & \Sigma_{,ij} \\ \Sigma_{,ij} & \Sigma_{,ij} \end{bmatrix}, \quad (5.22)$$

the first three terms of which are pictured in Fig. 5. This is a nonlinear functional differential equation in Γ, with the second derivatives of S forming an inhomogeneous term. In lowest approximation we may use bare vertex functions and the classical propagators in the expansion of the self-energy operator. This then gives us a first approximation to the matrix (5.22), which may be functionally differentiated to obtain corrections to the bare vertex functions and inverted to obtain corrections to the classical propagators. The corrected quantities may then be used in the expansion of the self-energy operator and the whole process repeated. The functional differentiations may either be performed on the explicit expressions obtained, or diagrammatically. A wide variety of iteration or approximation procedures may be in principle employed.

The validity of such procedures depends, of course, on their convergence, about which essentially nothing is known. The advantage of the present scheme, however, is that renormalization may be performed equally easily with any of these procedures. At any level of approximation divergences will generally appear when summations (integrations) are performed in the individual terms of the self-energy expansion. The renormalization program consists simply of throwing these divergences away, which is equivalent to assuming that corresponding "counter terms" are already present in the classical action S. The matrix (5.22) itself must always be kept finite so that only renormalized propagators and vertex functions are used in the evaluation of the self-energy operator at any approximation stage. The net result of this procedure may be summarized by the equation

$$\Gamma = S_R - [\Sigma], \quad (5.23)$$

where S_R denotes the renormalized or "observed"

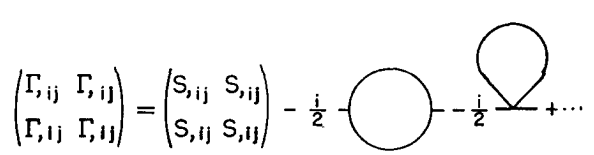


FIG. 5. Diagrammatic representation of Eq. (5.22).

[24] When fermion fields are absent the first term on the right of Fig. 4 represents the logarithm of a formal determinant: $-(i/2)\ln \det(\Gamma_{,ij})$. Its derivative is $(i/2)G^{jk}\Gamma_{,kji}$, which is depicted in the first term on the right of Fig. 3. The occurrence of this determinant is interesting because it suggests a relation between the lowest order self-energy diagrams (to which it gives rise in Fig. 5) and the choice of *measure* in the Feynman functional integral. It may be shown that if the measure is chosen to be

$$\Delta = [\det(\Gamma_{,ij})]^{1/2} \approx [\det(S_{,ij})]^{1/2}$$

then the left side of Eq. (2.5) is formally Hermitian up to the third order in a perturbation expansion, a result which is particularly suggestive in view of the similarity of this measure to the well-known Van Vleck determinant of WKB theory [see J. H. Van Vleck, Proc. Natl. Acad. Sci. U. S. 14, 178 (1928); also B. S. DeWitt, Revs. Modern Phys. 29, 377 (1957)]. Two things are wrong with this measure, however: (1) It fails to maintain Hermiticity in fourth order, (2) If elevated into the exponent of the Feynman integral it is equivalent to the addition of $-(i/2)\ln \det(\Gamma_{,ij})$ to the classical action and hence to the complete cancellation of lowest order radiative corrections, in contradiction to experiment.

classical action and the square brackets indicate that only the finite part of the self-energy functional is to be taken.

Of course there is some ambiguity about which is the "finite part" and which is the "divergent part" of a given integral, and the success of the renormalization program depends on (1) the possibility of separating the divergent parts in an invariant manner, and (2) giving them simple physical interpretations in terms of renormalizations of fundamental constants such as mass, charge, gravitation constant, etc. The latter condition appears to be necessary to insure that the renormalization process does not destroy basic features already present in the classical theory. In particular Γ must describe an essentially local theory in the long wavelength limit. Furthermore the field equations of this theory must, in the small amplitude limit, be of no higher than the second differential order if difficulties with "ghosts," negative probabilities, and lack of unitarity are to be avoided. The problem of unitarity is a very difficult one in the general case and we shall have nothing very useful to say about it here. It is possible that the process of field measurement itself[25] and the question of what variables are to be chosen as fundamental in the description of the field will in the end have to be studied in order to resolve this problem.

6. Γ AND TOPOLOGY. GENERAL DISCUSSION

It was pointed out in Sec. 2 that all physical processes can be computed from a knowledge of the vacuum-to-vacuum matrix elements of appropriate T functionals. These matrix elements can be obtained by multiplying corresponding Schwinger averages by the amplitude $\langle 0, \infty \mid 0, -\infty \rangle$. Since the Schwinger averages can be expressed in terms of the correlation functions, and these in turn can be expressed in terms of the propagators and vertex functions, we see that *the functional* Γ *by itself provides us with a complete quantum theory* provided we can express $\langle 0, \infty \mid 0, -\infty \rangle$ in terms of Γ. We now show how to do this.

From Eqs. (4.1), (4.3), and (4.4) we have

$$\begin{bmatrix} \phi^i \\ \psi^i \end{bmatrix} = \begin{bmatrix} \delta G/\delta J_i \\ \delta G/\delta J_i \end{bmatrix} = \begin{bmatrix} G^{ij} & -G^{ij} \\ G^{ij} & -G^{ij} \end{bmatrix} \begin{bmatrix} G_{,j} \\ G_{,j} \end{bmatrix} \tag{6.1}$$

and hence

$$\begin{bmatrix} G_{,i} \\ G_{,i} \end{bmatrix} = -\begin{bmatrix} \Gamma_{,ij} & \Gamma_{,ij} \\ \Gamma_{,ij} & \Gamma_{,ij} \end{bmatrix} \begin{bmatrix} \phi^j \\ \psi^j \end{bmatrix} = \begin{bmatrix} \Gamma_{,i} \\ \Gamma_{,i} \end{bmatrix} - \begin{bmatrix} (\Gamma_{,j}\phi^j + \Gamma_{,j}\psi^j)_{,i} \\ (\Gamma_{,j}\phi^j + \Gamma_{,j}\psi^j)_{,i} \end{bmatrix}, \tag{6.2}$$

which may be integrated from the zero point and combined with Eq. (4.2) to give[26]

$$\langle 0, \infty \mid 0, -\infty \rangle = e^{iG}, \tag{6.3}$$

$$G = \Gamma - \Gamma_{,i}\phi^k - \Gamma_{,i}\psi^i - \Gamma_0. \tag{6.4}$$

The zero point is conventionally chosen to be that solution of the classical field equations which corresponds to flat empty space-time.[19] This choice is not absolutely essential, and in some instances it may be desirable to generalize it. The quantum theory is basically a theory of small fluctuations about an *arbitrary* solution of the classical field equations, a fact which we have already emphasized in the arbitrariness of our choice of zero point for series expansions.[19] For many purposes we may equally well regard such an arbitrary "background" solution as "the vacuum."

Strictly speaking, since quantum effects (e.g., scattering of light by light) persist at large amplitudes even in the classical long-wavelength limit, the "background" field should be chosen as a solution of the *quantum* equations (5.7) with sources absent. Now these equations, unlike their classical counterparts $S_{,i} = 0$, $S_{,i} = 0$, have left-hand sides which become complex at thresholds for real particle production.[27] However, macroscopic wave solutions of these equations may be chosen having imaginary components which are negligible and in some instances (e.g., smooth cosmological solutions in gravidynamics) even absent. The theory of the preceding sections may then be applied with this generalized concept of "vacuum," the Feynman

---

[25] For a preliminary analysis of the measurability of the quantized gravitational field see B. S. DeWitt, *Gravitation, An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, to be published), Chap. 8.

[26] In the case of a linear boson field, where

$$S = \Gamma = \tfrac{1}{2}K_{ij}(\phi^i + \phi_0{}^i)(\phi^j + \phi_0{}^j),$$

$$K_{ij}\phi_0{}^j = 0, \qquad \phi^k = G^{ij}J_j, \qquad K_{ik}G^{kj} = -\delta_i{}^j,$$

K^{ij} being a constant symmetric kernel, we obtain the familiar result

$$G = -\tfrac{1}{2}K_{ij}\phi^i\phi^j = \tfrac{1}{2}G_{ij}G^{ij}J_j,$$

$$|\langle 0, \infty \mid 0, -\infty \rangle|^2 = e^{-\operatorname{Im} J_i G^{ij} J_j} = e^{-\Sigma_\alpha |J_\alpha|^2}, \qquad J_\alpha \equiv f_\alpha{}^{i*} J_i,$$

where the $f_\alpha{}^i$ are the orthogonal functions of Eqs. (6.5), (6.6), and (6.7).

[27] For massless fields this threshold occurs at zero frequency, and the complications of the infrared catastrophe have to be taken into account.

propagator being *defined* as that Green's function of the operator (5.22) which uniquely satisfies both (4.6) and (5.9).

It is of interest, in particular, to see how S-matrix theory is developed within such a context. The procedure is a straightforward generalization of the methods of Lehmann, Symanzik, and Zimmermann.[10] We give merely the results, restricting ourselves for simplicity to the case of boson fields only: A system of orthogonal complex functions f_α^i is introduced satisfying the equations

$$\Gamma_{,ij}f_\alpha^j = 0, \tag{6.5}$$

$$\text{Im } G^{ij} = \tfrac{1}{2}\sum_\alpha (f_\alpha^i f_\alpha^{j*} + f_\alpha^j f_\alpha^{i*}), \tag{6.6}$$

and, in configuration space,[28]

$$-i\int_\Sigma d\Sigma_\mu \int d^4x' \int d^4x'' f_{i'j''}^{\mu} f_\alpha^{i'*} f_\beta^{j''} = \delta_{\alpha\beta}, \tag{6.7}$$

where the $f_{i'j''}^{\mu}$ are the analogs for the operator $\Gamma_{,ij}$ of the functions introduced in Eq. (3.2) of (II) in connection with the Cauchy problem. The f_α^i serve to define a representation in terms of initial and final states, in which the S-matrix elements themselves are given by

$$S_{\alpha\beta} = \sum_\lambda S'_{\alpha/\lambda\beta/\lambda}, \tag{6.8}$$

$$\begin{aligned} S'_{\alpha\beta} = (-i)^{m+n} f_{\alpha_1}^{i_1*} \cdots f_{\alpha_m}^{i_m*} \Gamma_{,i_1k_1} \cdots \Gamma_{,i_mk_m} \\ \times \langle \phi^{k_1} \cdots \phi^{k_m} \phi^{l_1} \cdots \phi^{l_n} \rangle \\ \times \Gamma_{,l_1j_1} \cdots \Gamma_{,l_nj_n} f_{\beta_1}^{j_1} \cdots f_{\beta_n}^{j_n}, \end{aligned} \tag{6.9}$$

the notation of Eq. (6.8) being that of Nishijima.[29] For example, the amplitude for two-particle scattering (omitting the no-scattering amplitudes and the amplitudes for particle production and absorption by the sources) becomes

$$\begin{aligned} R_{\alpha_1\alpha_2\beta_1\beta_2} = f_{\alpha_1}^{i*} f_{\alpha_2}^{j*} (\Gamma_{,ijkl} + \Gamma_{,ijm}G^{mn}\Gamma_{,nkl} \\ + \Gamma_{,ikm}G^{mn}\Gamma_{,njl} + \Gamma_{,ilm}G^{mn}\Gamma_{,njk}) f_{\beta_1}^k f_{\beta_2}^l . \end{aligned} \tag{6.10}$$

In the context of gravitation theory the admissibility of arbitrary macroscopic background fields immediately implies also the admissibility of arbitrary topology in the large. Perhaps more important than this, however, are the possibilities which the functional Γ opens up in regard to topology in the small. Let us suppose that after having computed Γ in terms of invariant variables we transform back to configuration space by reintroducing the invariance group and the original dynamical variables of the theory. This we can easily do since the Γ language is a c-number language. Since Γ is an invariant the field equations $\Gamma_{,i} = 0$ for the original variables (in the absence of sources) are covariant equations. In the case of pure gravitation theory these equations have a local part identical with Einstein's equations and a nonlocal part, coming from the self-energy functional, which describes quantum corrections important at high energies and small distances. Now it is well known that Einstein's equations admit of solutions for which space-like cross sections have non-Euclidean topology, giving under certain conditions a particle-like or "wormhole" structure to space itself.[30] There are, however, strong reasons for believing that classical wormholes are immutable, i.e., that the topology of 3-space cannot change as long as the metric tensor satisfies Einstein's equations and retains its normal signature. On the other hand, with the equations $\Gamma_{,i} = 0$ the situation is quite different. In the first place the metric tensor in the Γ language is *complex* (it satisfies Feynman boundary conditions), and hence the question of signature loses its rigid classical significance. Secondly the Γ equations are nonlocal at small distances ($\sim 10^{-32}$ cm) and hence afford greater possibilities for real dynamical changes in 3-space topology.

If no wormholes are initially present, of course, none are to be expected to develop in the course of time—*provided* no field quanta (represented as small-amplitude superpositions of the complex functions f_α^i) are present either. The same should also be true for single wormholes. But in the case of two or more colliding wormholes it is likely that the real metric which describes them in the remote past acquires complex components (corresponding to final gravitons) in the course of time, while the wormholes themselves either bounce inelastically or, under suitable phase relationships, annihilate one another.

Since possibilities of this kind were not anticipated in the original formulation of the theory several questions immediately arise. For example, how is the Feynman functional integral to be understood under these circumstances? The choice of numerical range for the continuous labels i, $\mathbf{i}$, which is essential for the definition of the functional volume element, depends critically on the topology assumed

[28] The existence of such functions, with a positive definite Kronecker delta on the right of Eq. (6.7), depends on the negligibility of the imaginary part of $\Gamma_{,ij}$ and on the positive definiteness of the symmetric real matrix $\text{Im}G_{ij}$. When these conditions are not satisfied negative probabilities and failure of unitarity occur.

[29] K. Nishijima, Phys. Rev. **111**, 995 (1958).

[30] C. W. Misner and J. A. Wheeler, Ann. Phys. (New York) **2**, 525 (1957).

for space-time. Now that many topologies are possible, how are we to interpret them? Must we sum over topologies as well as histories in the Feynman integral, thus giving rise to a kind of *third quantization*, with amplitude for each topology sequence?[31]

Of course these questions become empty if, as is quite possible, field nonlocality occurs by some other presently unaccounted for mechanism already at a higher level (i.e., anywhere between 10^{-32} and 10^{-13} cm). Furthermore, although it would seem very interesting to attempt at least a lowest order calculation of the radiative corrections to the classical field equations, and then to see what effect these corrections have on the Schwarzschild solution and on the two-wormhole problem, serious questions would unfortunately have to be raised in regard to the significance of such a calculation, even assuming the underlying theory to be correct.

In the first place there is a rather uncomfortable situation in regard to the divergences. Owing to the special nature of gravitational coupling, whereby the insertion of a bare vertex prong at any point in any diagram leaves the degree of divergence of the diagram unaffected (p^2 behavior of the vertex, $1/p^2$ behavior of the extra bare propagator) all the full vertex functions have the same asymptotic momentum-space behavior as the operator $\Gamma_{,ij}$, and hence all the self-energy diagrams diverge at least quartically. If the asymptotic behavior of $\Gamma_{,ij}$ is *assumed* to be such as to maintain consistency in a perturbation calculation to any finite order, it is easy to see that all diagrams diverge *exactly* quartically, and that this asymptotic behavior must be p^4 (or $p^4 \ln |p^2|$) after subtraction of divergences. Unfortunately the assumption is a dubious one because there is no evidence at all that the renormalized perturbation series converges nor, *a fortiori*, that the sum of the series has the same asymptotic behavior as the individual terms. The trouble is that there exists no small dimensionless coupling constant in the theory of the pure gravitational field, nor in the theory of the combined Einstein-Maxwell field. When natural units are used in which $\hbar = c = 16\pi G = 1$, no physical constants are left, and all diagrams begin to diverge equally at 10^{-32} cm, and have roughly equal magnitudes after renormalization.

Moreover, if the asymptotic behavior really is p^4 then we are faced with the unpleasant fact that one counter term beyond the classical Einsteinian term is needed to effect the renormalization of every diagram.[32] Such a counter term corresponds to an initial classical theory having field equations of the *fourth* differential order, with corresponding complications in regard to unitarity. Of course, one might argue that this counter term precisely cancels a preexisting one in S, leaving no term of the unwanted type in the "observed" classical action S_R, and indeed such a proposal has some merit. The p^4 behavior would then mean that the singularities of the propagators on the light cone have been softened without obvious violation of unitarity, and that although divergences are not eliminated, quantum gravidynamics does partially live up to its initial expectations of providing a natural cutoff in momentum space. It is worth pointing out that, unlike the situation in other field theories, this would not imply a violation of Lehmann's theorem,[33] for one of the assumptions of that theorem, namely, the existence of an energy-momentum 4-vector which generates infinitesimal displacements of local quantities with respect to the (nonintrinsic) coordinates x^μ, fails to hold in gravidynamics.

[31] As Wheeler has emphasized [J. A. Wheeler, Ann. Phys. (New York) **2**, 604 (1957)] the wormholes envisaged here, having the excessively large mass ($\sim 3 \times 10^{-6}$g) associated with 10^{-33}cm, are not to be confused with observed elementary particles, and hence the quantum theory of gravitation does not immediately obviate the necessity of introducing extra fields *ab initio* to represent matter. In this connection it is worth observing that only a small fraction of all topologies have a simple particle-like character. Moreover, enthusiasm for topological dynamics is dampened by the observation that topological flexibility can actually undermine the experimentally observed dimensionality of macroscopic space-time. This is most simply seen by thinking of a close-spaced rectangular lattice of solid rods in three dimensions, which is dipped in a rubber cement solution and then allowed to dry. If the rods are somehow dissolved by immersion in another solution, leaving only the thin rubber coating, we have a model of a mathematically two-dimensional space, the gravity waves (membrane vibrations) of which require three dimensions for any sort of simple physical description. This sort of disaster can obviously proceed *ad dimensionem infinitum*.

[32] In general three counter terms are needed: (1) a quartically divergent cosmological term to compensate the zero point vacuum energy; (2) a quadratically divergent Einsteinian term, which renormalizes the gravitation constant; (3) a term quadratic in the Riemann tensor involving two independent logarithmically divergent constants which have no counter parts in the classical theory. [Compare with R. Utiyama and B. S. DeWitt, J. Math. Phys. **3** 608, (1962). Also lectures given by R. P. Feynman at the New York meeting of the American Physical Society in January 1961.]

[33] H. Lehmann, Nuovo cimento **11**, 342 (1954).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

On the Gauge Covariance of Quantum Electrodynamics*

I. BIALYNICKI-BIRULA
Institute of Theoretical Physics, Warsaw University, Warsaw, Poland
(February 20, 1962)

The gauge covariant formulation of quantum electrodynamics given by Zumino is further investigated. The method used is that of functional integration. This allows for a slight generalization of Zumino's results and provides a direct relation between the gauge condition in classical and quantum theories. All results are derived without any reference to the canonical formalism. The same procedure can be applied to the Yang-Mills field with or without a mass, leading in both cases to a renormalizable theory. The Landau gauge is studied in some detail and it is shown that in the perturbation expansion of the propagators no terms violating the gauge invariance appear. Finally, a new interpretation of the generalized Ward identity is proposed.

1. INTRODUCTION

An exhaustive discussion of the gauge properties of Green's functions was given recently in a very interesting paper by Zumino.[1] With the use of the functional differentiation method[2] he exhibited the transformation properties of Green's functions under various kinds of gauge transformations. In our opinion, however, there is one point that needs further clarification. It is the connection between the gauge conditions in classical and quantum electrodynamics. Such a connection is provided in this paper. Also, a slight generalization of Zumino's results is given and his rather heuristic procedure is justified. Finally, the formulation of quantum electrodynamics in the Landau gauge[3] is studied in some detail. The set of Schwinger's equations for the electron and photon propagators is derived and it is shown that there are no terms violating the gauge invariance.

To classify the gauge conditions in quantum electrodynamics Zumino introduced[4] a vector operator a_μ and a gauge function Λ. The vector a_μ was introduced first in a special form (in the Coulomb gauge) and then the resulting equations for the functional Z were postulated to be valid for a large class of a_μ's.

In the present paper a different method of obtaining the set of gauge covariant equations for propagators is proposed. It is based on the technique of integration over all fields. This method seems to be most suitable in dealing with theories invariant under continuous groups of transformations. With its help one can quantize not only electrodynamics but also the Yang-Mills field and general relativity in the linear approximation.[5] The Yang-Mills field with a mass can be quantized in practically the same manner as quantum electrodynamics. It will lead to a renormalizable theory.[6]

2. CLASSICAL ELECTRODYNAMICS

In classical electrodynamics there is a freedom in choosing the vector potential A_μ to represent a given field $f_{\mu\nu}$. This freedom allows to perform a gauge transformation,

$$'A_\mu = A_\mu + \chi_{,\mu}, \tag{1}$$

without affecting the field $f_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu}$. In some cases it is convenient to restrict the freedom by imposing certain conditions on the potentials. These gauge conditions will be assumed to be linear in A_μ. The most general linear gauge condition can be written in the symbolic form

$$b_\mu A^\mu = \kappa, \tag{2}$$

which stands for

$$\int b_\mu(x, x')\, dx' A^\mu(x') = \kappa(x). \tag{3}$$

All gauge conditions having use in electrodynamics can be cast into this form. There is one condition which must be satisfied by the vector operator b_μ to

* Preliminary accounts of this work were given in the Bull. Acad. Polon. Sci. Cl. III **10**, 225 (1962).

[1] B. Zumino, J. Math. Phys. **1**, 1 (1960). In the following this paper will be referred to as A.

[2] The same method has been applied independently but much less successfully by the present author [I. Bialynicki-Birula, Nuovo cimento **17**, 961 (1960)].

[3] As is shown in this paper the Landau gauge is essentially the Lorentz gauge of the classical theory.

[4] The additional function M introduced by Zumino will be discussed later.

[5] One can quantize formally even the exact version of general relativity. Resulting equations will be similar to those proposed by Klauder [J. R. Klauder, Nuovo cimento **19**, 1059 (1961); (to be published)]. Such a procedure is open, however, to many doubts [C. W. Misner, Revs. Modern Phys. **29**, 497 (1957)].

[6] The renormalizability follows from the absence of the $\mu^{-2}\,\partial_\mu\partial_\nu$ term in the propagator. This term is replaced by $\Box^{-1}\,\partial_\mu\partial_\nu$.

guarantee that every field can be represented by the potential obeying (2). This condition is that the operator $b_\mu\partial^\mu$ has an inverse. Otherwise it would not be possible, in general, to find a gauge transformation (1) which transforms A_μ into $'A_\mu$ satisfying (2). Assuming the existence of the operator $(b_\mu\ \partial^\mu)^{-1}$, the condition (2) can be written in the form

$$a_\mu A^\mu + \Lambda = 0, \tag{4}$$

where

$$a_\mu(x, x') = -\int (b, \partial^\nu)^{-1}(x, x'')\, dx'' b_\mu(x'', x') \tag{5}$$

so that a_μ satisfies the condition

$$a_\mu(x, x')\ \partial^\mu = -\delta(x - x'). \tag{6}$$

With such a normalization of the gauge condition two potentials related through the transformation (1) satisfy condition (4) with two Λ's differing just by the gauge function χ, i.e.,

$$'\Lambda = \Lambda + \chi. \tag{7}$$

An important class of gauge conditions obtains when a_μ is chosen in the form

$$a_\mu = -\frac{n_\mu(n\cdot\partial) - \rho\partial_\mu}{(n\cdot\partial)^2 - \rho\Box}, \tag{8}$$

where n_μ is a unit vector lying in the forward light cone and ρ is a real parameter. The three most frequently used gauge conditions

$$A^0 = 0, \quad \text{div } \mathbf{A} = 0, \quad \text{and} \quad A^\mu{}_{,\mu} = 0 \tag{9}$$

imposed in an arbitrary coordinate frame belong to the class (8) with ρ taking on values 0, 1, and ∞.

3. QUANTUM ELECTRODYNAMICS

The first complete gauge covariant version of quantum electrodynamics was given by Zumino in A and very little will be added here as far as the form of the Green's functions in different gauges is concerned. We shall try, however, to clarify the role of the classical gauge condition in the quantum theory.

The vector a_μ was introduced in A at first "for the sake of concise notation" in the Coulomb gauge and the generalization[7] to other cases was made rather arbitrarily without referring to any gauge conditions for the potentials. It will be shown below that this vector a_μ is substantially the same as the one introduced in Sec. 2 in classical theory. To this end the generating functional Z will be written as an integral over all fields,

$$Z\{\mathfrak{z}_\mu, \Lambda, \eta, \bar\eta\} = N^{-1}\int \delta_g A_\mu\ \delta\psi\ \delta\bar\psi$$

$$\times \exp\left[i\int (\mathfrak{L} + \bar\psi\eta + \bar\eta\psi + \mathfrak{z}_\mu A^\mu)\, dx\right], \tag{10}$$

where

$$\mathfrak{L} = -\tfrac{1}{4}f_{\mu\nu}f^{\mu\nu} + \bar\psi(i\,\partial - m)\psi + j_\mu A^\mu.$$

This expression differs from what has been customarily assumed[8] in having the integration over all potentials replaced by the integration over potentials restricted by the gauge condition (4). This restriction, indicated in (10) by the index g, makes Z a functional of Λ. Such a procedure seems to be in agreement with the concept of the Feynman integral over all histories. Every history should be counted once and this corresponds to integrating over all potentials with the additional condition (4). There are several objections which might be raised in connection with the integrals over all fields. It will be, however, quite sufficient for our purposes to treat them quite formally (in the spirit of Symanzik's paper[9]) without going more deeply into the question of their evaluation, etc.

In what follows only three, very general, properties of such integrals will be needed.

I. The integrals over all fields allow for an integration by parts, i.e.,

$$\int \delta\lambda G\frac{\delta F}{\delta\lambda} = -\int \delta\lambda\frac{\delta G}{\delta\lambda}F. \tag{11}$$

II. The functional differentiation can be interchanged with the functional integration, for example,

$$\frac{1}{i}\frac{\delta}{\delta\varphi(x)}\int \delta\lambda F\{\lambda\}\exp\left[i\int\varphi\lambda\, dx\right]$$

$$= \int \delta\lambda\,\lambda(x)F\{\lambda\}\exp\left[i\int\varphi\lambda\, dx\right]. \tag{12}$$

III. The origin in the space of all functions can be freely shifted i.e.

$$\int \delta\lambda G\{\lambda\} = \int \delta\lambda G\{\lambda + \kappa\}. \tag{13}$$

The symbol $\delta_g A_\mu$ introduced in (10) can be given a precise meaning with the help of the functional δ function,

[7] This part of Zumino's considerations we had in mind calling his procedure heuristic in the Introduction.

[8] B. Laurent, Nuovo cimento 4, 1445 (1956). N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959). A. Visconti and H. Umezawa, Compt. rend. 252, 1910 (1961).

[9] K. Symanzik, Z. Naturforsch. 9, 809 (1954).

$$\delta(a_\mu A^\mu + \Lambda) = \int \delta\lambda \exp\left[i \int \lambda(a_\mu A^\mu + \Lambda)\, dx\right], \quad (14)$$

where the factors $(2\pi)^{-1/2}$ have been absorbed[9] into the symbol $\delta\lambda$. The functional Z can be written then as an integral over *all* fields A_μ and an additional field λ,

$$Z\{\mathfrak{z}_\mu, \Lambda, \eta, \bar\eta\} = N^{-1} \int \delta A_\mu\, \delta\lambda\, \delta\psi\, \delta\bar\psi \times \exp\left[i \int (\mathfrak{L} + \bar\psi\eta + \bar\eta\psi + \mathfrak{z}_\mu A^\mu + \lambda a_\mu A^\mu + \lambda\Lambda)\, dx\right]. \quad (15)$$

It is now a matter of very simple calculations, making use of the properties I and II, to show that Z obeys the following set of equations in the functional derivatives.

$$\left[\partial^\nu\left(\partial_\nu \frac{1}{i}\frac{\delta}{\delta\mathfrak{z}^\mu} - \partial_\mu \frac{1}{i}\frac{\delta}{\delta\mathfrak{z}^\nu}\right) + \mathfrak{z}_\mu + e\frac{\delta}{\delta\eta}\gamma_\mu\frac{\delta}{\delta\bar\eta}\right]Z = -\tilde a_\mu \frac{1}{i}\frac{\delta}{\delta\Lambda}Z, \quad (16)$$

$$\left[-i\,\partial + m + e\gamma^\mu \frac{1}{i}\frac{\delta}{\delta\mathfrak{z}^\mu}\right]\frac{1}{i}\frac{\delta}{\delta\bar\eta}Z = \eta Z, \quad (17)$$

$$\frac{1}{i}\frac{\delta}{\delta\eta}\left[+i\,\partial + m + e\gamma^\mu \frac{1}{i}\frac{\delta}{\delta\mathfrak{z}^\mu}\right]Z = \bar\eta Z, \quad (18)$$

$$\left(a^\mu \frac{1}{i}\frac{\delta}{\delta\mathfrak{z}^\mu} + \Lambda\right)Z = 0, \quad (19)$$

where $\tilde a_\mu$ stands for the transpose of a_μ, i.e.,

$$(\tilde a_\mu f)(x') \equiv \int f(x) a_\mu(x, x')\, dx. \quad (20)$$

As a result of (6) the transposed operator $\tilde a_\mu$ satisfies the condition

$$\partial_\mu \tilde a^\mu = 1. \quad (21)$$

This condition, together with Eqs. (16), (17) and (18), leads to the additional equation,

$$\frac{1}{i}\frac{\delta}{\delta\Lambda}Z = \left(-\mathfrak{z}^\mu{}_{,\mu} + e\frac{\delta}{\delta\eta}\eta - e\bar\eta\frac{\delta}{\delta\bar\eta}\right)Z, \quad (22)$$

which could be used to eliminate $\delta Z/\delta\Lambda$ from Eq. (16). Eqs. (16) to (19) coincide with Zumino's equations only for a restricted class of a_μ's, namely for $a_\mu = -\tilde a_\mu$. Although Zumino did not mention this condition his examples do satisfy it. The property III of functional integrals makes it possible to obtain a more explicit form of the integral (10),

$$Z\{\mathfrak{z}_\mu, \Lambda, \eta, \bar\eta\} = N_1^{-1} \int \delta\psi\, \delta\bar\psi \times \exp\left\{i \int [\bar\psi(i\,\partial - m)\psi + \bar\psi\eta + \bar\eta\psi - \partial^\mu(\mathfrak{z}_\mu + j_\mu)\Lambda]\, dx\right\} \times \exp\left[\frac{i}{2}\int (\delta^\nu_\mu - \tilde a_\mu\,\partial^\nu)(\mathfrak{z}_\nu + j_\nu) \times D_F(\delta^\mu_\lambda - \tilde a^\mu\,\partial_\lambda)(\mathfrak{z}^\lambda + j^\lambda)\, dx\, dx'\right]. \quad (23)$$

In the preceding discussion no mention was made of the function M introduced by Zumino in A. It can, however, be easily included into our formalism by replacing the δ functional (14) by a more general functional Φ,

$$\Phi(a_\mu A^\mu + \Lambda) = \int \delta\lambda \exp\left[-(i/2)\int \lambda M \lambda\, dx\, dx'\right] \times \exp\left[i\int \lambda(a_\mu A^\mu + \Lambda)\, dx\right]. \quad (24)$$

The functional Φ represents a model of the δ functional. The symmetric kernel $M(x, x')$ describes the spread of the quantity $a_\mu A^\mu$ around the value Λ. The replacement of (14) by (24) in (15) leads to

$$Z\{\mathfrak{z}_\mu, \Lambda, M, \eta, \bar\eta\} = N^{-1} \int \delta A_\mu\, \delta\lambda\, \delta\psi\, \delta\bar\psi \times \exp\left(-(i/2)\int \lambda M \lambda\, dx\, dx'\right) \times \exp\left[i\int (\mathfrak{L} + \bar\psi\eta + \bar\eta\psi + \mathfrak{z}_\mu A^\mu + \lambda a_\mu A^\mu + \lambda\Lambda)\, dx\right]. \quad (25)$$

After the integration over A_μ and λ one obtains

$$Z\{\mathfrak{z}_\mu, \Lambda, M, \eta, \bar\eta\} = N_2^{-1} \int \delta\psi\, \delta\bar\psi \times \exp\left\{i\int [\bar\psi(i\,\partial - m)\psi + \bar\psi\eta + \bar\eta\psi - \partial^\mu(\mathfrak{z}_\mu + j_\mu)\Lambda]\, dx\right\} \times \exp\left\{\frac{i}{2}\int [(\delta^\nu_\mu - \tilde a_\mu\,\partial^\nu)(\mathfrak{z}_\nu + j_\nu) \times D_F(\delta^\nu_\lambda - \tilde a^\mu\,\partial_\nu)(\mathfrak{z}^\lambda + j^\lambda) - \partial^\mu(\mathfrak{z}_\mu + j_\mu) M\, \partial^\nu(\mathfrak{z}_\nu + j_\nu)]\, dx\, dx'\right\}. \quad (26)$$

This is in full agreement with the formula (A6) of A. Expression (25) allows us to write a set of equations for Z in the presence of M. Eqs. (16), (17), and (18) remain unchanged and Eq. (19) takes on the form

$$\left[M\frac{1}{i}\frac{\delta}{\delta\Lambda}+a^{\mu}\frac{1}{i}\frac{\delta}{\delta\mathfrak{J}^{\mu}}+\Lambda\right]Z=0. \tag{27}$$

The generalization to nonvanishing M's has no interpretation in terms of classical fields and potentials. It corresponds to the introduction of a new field λ with the $\frac{1}{2}\lambda M\lambda$ playing role of a nonlocal Lagrangian. From this point of view the Landau or the radiation gauge are much more natural than the Feynman gauge.

Having at our disposal the closed expression (25) for Z we can easily derive the formula connecting two Z's in different gauges,

$$\begin{aligned}Z\{\mathfrak{J},\Lambda+\delta\Lambda,M+\delta M,a_{\mu}+\delta a_{\mu},\eta,\bar{\eta}\}&\\ =\exp\left(i\int\delta\Lambda\frac{1}{i}\frac{\delta}{\delta\Lambda}\,dx\right)&\\ \times\exp\left(-\frac{i}{2}\int\frac{1}{i}\frac{\delta}{\delta\Lambda}\,\delta M\,\frac{1}{i}\frac{\delta}{\delta\Lambda}\,dx\,dx'\right)&\\ \times\exp\left(\frac{i}{2}\int\frac{1}{i}\frac{\delta}{\delta\Lambda}\,\delta a_{\mu}\frac{1}{i}\frac{\delta}{\delta\mathfrak{J}_{\mu}}\,dx\right)&\\ \times Z\{\mathfrak{J}_{\mu},\Lambda,M,a_{\mu},\eta,\bar{\eta}\}.&\end{aligned} \tag{28}$$

This formula holds for finite changes of Λ, a_{μ}, and M and we can derive from it the transformation properties of all Green's functions under arbitrary changes of Λ, a_{μ}, and M.

4. EQUATIONS FOR ONE ELECTRON AND ONE PHOTON PROPAGATORS

To obtain an explicitly Lorentz invariant form of quantum electrodynamics we choose in this section a_{μ} to be a Lorentz invariant operator,

$$a_{\mu}=-\partial_{\mu}(\Box-i\epsilon)^{-1}. \tag{29}$$

We put also $M=0$ to be as close as possible to the classical electrodynamics. As shown in the Appendix, the electron and the photon propagators and the vertex function satisfy the following set of integral equations.

$$\begin{aligned}D_{x}G(x-y)&=\delta(x-y)-ie^{2}\gamma^{\mu}\int G(x-x')\,dx'\\ &\times\Gamma^{\nu}(x',y',z)\,dy'G(y'-y)\,dz\mathfrak{G}_{\mu\nu}(x-z),\end{aligned} \tag{30}$$

$$\Box\mathfrak{G}_{\mu\nu}(z-z')=-(g_{\mu\nu}-\partial_{\mu}\partial_{\nu}\Box^{-1})\,\delta(z-z')$$

$$\begin{aligned}&+ie^{2}\,\mathrm{Tr}\int\gamma_{\mu}G(z-x)\,dx\Gamma^{\lambda}(x,y,z'')\,dy\\ &\times G(y-z)\,dz''\mathfrak{G}_{\lambda\nu}(z''-z')\end{aligned} \tag{31}$$

$$i\,\partial_{\nu}\Gamma^{\nu}(x,y,z)=\delta(z-x)G^{-1}(z-y)-\delta(z-y)G^{-1}(x-z). \tag{32}$$

These equations differ from those derived by Schwinger[10] in yielding automatically a purely transverse photon propagator,

$$\partial^{\mu}\mathfrak{G}_{\mu\nu}=0=\partial^{\nu}\mathfrak{G}_{\mu\nu}. \tag{33}$$

This transversality condition had usually to be added from outside and contradicts the results obtained in the perturbation theory. Nothing of that kind will occur in our formalism. The photon self-mass, appearing in the perturbation theory does not violate the condition (33) and can be eliminated by the standard renormalization procedure.

There is one more property of our formulation which seems to be of some importance. Due to the property (33) of $\mathfrak{G}_{\mu\nu}$, the longitudinal part of the vertex function *does not* enter into equations for the propagators. This can be easily seen in the simplest case of equations for the electron and photon propagators and can be shown to be valid in the general case. The *only* equation involving the longitudinal part of the vertex function is the generalized Ward identity (32) and, therefore, this identity may serve as a definition of this part. Computing the physical properties of any system containing electrons and photons one is never forced to consider the longitudinal part of Γ_{μ}. It was to be expected since this part describes the interaction of an electron with unphysical photons.

ACKNOWLEDGMENTS

The author is indebted to Dr. R. Gajewski, Dr. A. Trautman and Dr. W. Tulczyjew for helpful discussions.

APPENDIX

We adopt the usual definitions of the electron and the photon propagators in the presence of external sources $\mathfrak{J}_{\mu}$ and Λ,

$$G(x,y)=L^{-1}\frac{1}{i}\frac{\delta^{2}Z}{\delta\bar{\eta}(x)\,\delta\eta(y)}\bigg|_{\eta=0=\bar{\eta}}, \tag{A1}$$

$$\mathfrak{G}_{\mu\nu}(z,z')=\frac{\delta\mathfrak{A}_{\mu}(z)}{\delta\mathfrak{J}^{\nu}(z')}=\frac{\delta\mathfrak{A}_{\nu}(z')}{\delta\mathfrak{J}^{\mu}(z)}, \tag{A2}$$

[10] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **37**, 452, 455 (1951).

where

$$L = Z\{\mathfrak{J}_\mu, \Lambda, \eta = 0, \bar{\eta} = 0\} \tag{A3}$$

and

$$\mathfrak{A}_\mu(z) = \frac{1}{i} L^{-1} \frac{\delta L}{\delta \mathfrak{J}^\mu(z)}. \tag{A4}$$

It is convenient to introduce also the expectation value of λ in analogy to the expectation value of A_μ. It will be denoted by κ and defined as

$$\kappa(z) = L^{-1} \frac{1}{i} \frac{\delta L}{\delta \Lambda(z)}. \tag{A5}$$

All functionals of $\mathfrak{J}_\mu$ and Λ can be expressed in terms of A_μ and κ. The derivatives with respect to $\mathfrak{J}_\mu$ and Λ are related to those with respect to $\mathfrak{A}_\mu$ and κ through the formulas

$$\frac{\delta}{\delta \mathfrak{J}^\mu(z)} = \int \mathcal{G}_{\mu\lambda}(z, z')\, dz' \frac{\delta}{\delta \mathfrak{A}_\lambda(z')} + \partial_\mu \frac{\delta}{\delta x(z)}, \tag{A6}$$

$$\frac{\delta}{\delta \Lambda(z)} = -\partial_\lambda \frac{\delta}{\delta \mathfrak{A}_\lambda(z)}. \tag{A7}$$

Equations (16), (22), and (19) lead to the following equations for $\mathfrak{A}_\mu$

$$\partial^\nu(\partial_\nu \mathfrak{A}_\mu(z) - \partial_\mu \mathfrak{A}_\nu(z)) + (\delta^\nu_\mu - \tilde{a}_\mu \partial^\nu)\mathfrak{J}_\nu(z) - ie \operatorname{Tr} \gamma_\mu G(z, z) = 0, \tag{A8}$$

$$a_\mu \mathfrak{A}^\mu(z) + \Lambda(z) = 0. \tag{A9}$$

Equation (17) differentiated with respect to η gives the equation for G in the form

$$[-i\,\partial + m + e\gamma^\mu \mathfrak{A}_\mu(x)]G(x, y) = \delta(x - y) - e\gamma^\mu \frac{1}{i} \frac{\delta}{\delta \mathfrak{J}^\mu(x)} G(x, y). \tag{A10}$$

To obtain from (A8) to (A10) the Eqs. (30), (31), and (33), we have to differentiate (A8) and (A9) with respect to $\mathfrak{J}_\nu$, make use of the formulas

$$\Gamma_\mu(x, y, z) = \frac{1}{e} \frac{\delta}{\delta \mathfrak{A}_\mu(z)} G^{-1}(x, y), \qquad \frac{\delta}{\delta \kappa(z)} G(x, y) = 0, \tag{A11}$$

and finally put $\mathfrak{A}_\mu = 0 = \kappa$. The generalized Ward identity (32) directly follows from Eq. (22), differentiated with respect to η and $\bar{\eta}$, and from (A7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Modified Mandelstam Representation for Heavy Particles

JAMAL N. ISLAM
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received May 18, 1962)

The fourth-order Feynman amplitude ceases to satisfy the Mandelstam representation when the external masses are sufficiently large. A representation which replaces the Mandelstam one is found in the cases where the four mass invariants are equal in pairs. The physical interpretation is briefly discussed.

1. INTRODUCTION

WE shall be concerned with the partial Feynman amplitude associated with the fourth-order diagram shown in Fig. 1. It is known[1,2] that this amplitude ceases to satisfy the Mandelstam representation for a certain set of values of the external masses. If we set

$$y_{ij} = -(p_{ij}^2 - m_i^2 - m_j^2)/2m_i m_j$$

with $p_{13} = p_{12} + p_{23}$ and $p_{24} = p_{23} + p_{34}$, the stability conditions on the external and internal masses are given by

$$-1 < y_{12}, y_{23}, y_{34}, y_{41} < +1$$

so that for these y's we may put $y_{ij} = \cos \theta_{ij}$, $0 < \theta_{ij} < \pi$. The condition for the validity of the Mandelstam representation is then given by

$$\theta_{12} + \theta_{23} + \theta_{34} + \theta_{41} \leq 2\pi.$$

When the above inequality ceases to be valid, the

[1] S. Mandelstam, Phys. Rev. 115, 1741 (1959).
[2] J. Tarski, J. Math. Phys. 1, 149 (1960).

where

$$L = Z\{\mathfrak{J}_\mu, \Lambda, \eta = 0, \bar{\eta} = 0\} \tag{A3}$$

and

$$\mathfrak{A}_\mu(z) = \frac{1}{i} L^{-1} \frac{\delta L}{\delta \mathfrak{J}^\mu(z)}. \tag{A4}$$

It is convenient to introduce also the expectation value of λ in analogy to the expectation value of A_μ. It will be denoted by κ and defined as

$$\kappa(z) = L^{-1} \frac{1}{i} \frac{\delta L}{\delta \Lambda(z)}. \tag{A5}$$

All functionals of $\mathfrak{J}_\mu$ and Λ can be expressed in terms of A_μ and κ. The derivatives with respect to $\mathfrak{J}_\mu$ and Λ are related to those with respect to $\mathfrak{A}_\mu$ and κ through the formulas

$$\frac{\delta}{\delta \mathfrak{J}^\mu(z)} = \int \mathcal{G}_{\mu\lambda}(z, z')\, dz' \frac{\delta}{\delta \mathfrak{A}_\lambda(z')} + \partial_\mu \frac{\delta}{\delta x(z)}, \tag{A6}$$

$$\frac{\delta}{\delta \Lambda(z)} = -\partial_\lambda \frac{\delta}{\delta \mathfrak{A}_\lambda(z)}. \tag{A7}$$

Equations (16), (22), and (19) lead to the following equations for $\mathfrak{A}_\mu$

$$\partial^\nu(\partial_\nu \mathfrak{A}_\mu(z) - \partial_\mu \mathfrak{A}_\nu(z)) + (\delta^\nu_\mu - \tilde{a}_\mu \partial^\nu)\mathfrak{J}_\nu(z) - ie \operatorname{Tr} \gamma_\mu G(z, z) = 0, \tag{A8}$$

$$a_\mu \mathfrak{A}^\mu(z) + \Lambda(z) = 0. \tag{A9}$$

Equation (17) differentiated with respect to η gives the equation for G in the form

$$[-i\,\partial + m + e\gamma^\mu \mathfrak{A}_\mu(x)]G(x, y) = \delta(x - y) - e\gamma^\mu \frac{1}{i} \frac{\delta}{\delta \mathfrak{J}^\mu(x)} G(x, y). \tag{A10}$$

To obtain from (A8) to (A10) the Eqs. (30), (31), and (33), we have to differentiate (A8) and (A9) with respect to $\mathfrak{J}_\nu$, make use of the formulas

$$\Gamma_\mu(x, y, z) = \frac{1}{e} \frac{\delta}{\delta \mathfrak{A}_\mu(z)} G^{-1}(x, y), \qquad \frac{\delta}{\delta \kappa(z)} G(x, y) = 0, \tag{A11}$$

and finally put $\mathfrak{A}_\mu = 0 = \kappa$. The generalized Ward identity (32) directly follows from Eq. (22), differentiated with respect to η and $\bar{\eta}$, and from (A7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Modified Mandelstam Representation for Heavy Particles

JAMAL N. ISLAM
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received May 18, 1962)

The fourth-order Feynman amplitude ceases to satisfy the Mandelstam representation when the external masses are sufficiently large. A representation which replaces the Mandelstam one is found in the cases where the four mass invariants are equal in pairs. The physical interpretation is briefly discussed.

1. INTRODUCTION

WE shall be concerned with the partial Feynman amplitude associated with the fourth-order diagram shown in Fig. 1. It is known[1,2] that this amplitude ceases to satisfy the Mandelstam representation for a certain set of values of the external masses. If we set

$$y_{ij} = -(p_{ij}^2 - m_i^2 - m_j^2)/2m_i m_j$$

with $p_{13} = p_{12} + p_{23}$ and $p_{24} = p_{23} + p_{34}$, the stability conditions on the external and internal masses are given by

$$-1 < y_{12}, y_{23}, y_{34}, y_{41} < +1$$

so that for these y's we may put $y_{ij} = \cos \theta_{ij}$, $0 < \theta_{ij} < \pi$. The condition for the validity of the Mandelstam representation is then given by

$$\theta_{12} + \theta_{23} + \theta_{34} + \theta_{41} \leq 2\pi.$$

When the above inequality ceases to be valid, the

[1] S. Mandelstam, Phys. Rev. 115, 1741 (1959).
[2] J. Tarski, J. Math. Phys. 1, 149 (1960).

amplitude $A(y_{13}, y_{24})$ has singularities for complex values of y_{13}, y_{24} in the physical sheets for these variables and this precludes the possibility of a representation such as Mandelstam's. Our objective is to find a representation for $A(y_{13}, y_{24})$ when it has complex singularities, in the special case when the four invariants y_{12}, y_{23}, y_{34}, y_{41} are set equal in pairs. We find this representation with the aid of the Bergman-Weil formula.

2. THE BERGMAN-WEIL FORMULA[3]

This formula is a generalization of Cauchy's formula and gives the values of a function of several complex variables in its domain of analyticity in terms of its values on a certain subset of the boundary of this domain, when the boundary consists of pieces of analytic hypersurfaces. An analytic hypersurface is a surface of the form $F(z_k, r) = 0$, where F is analytic in the complex variables z_k and r is a real parameter varying over a certain interval. For the case of two complex variables, which is the relevant one here, we have, explicitly

$$A(z_1, z_2) = \iint d\zeta_1\, d\zeta_2 A(\zeta_1, \zeta_2)\, |q(z_1, z_2, \zeta_1, \zeta_2)|.$$

Here the integration is over the distinguished boundary of the domain of analyticity, which is defined as the two dimensional intersections of the analytic hypersurfaces that form the boundary of this domain taken two at a time. The q factor in the integrand is a 2×2 determinant of certain functions q^{kl} defined as follows: We associate two functions q^{k1} and q^{k2} functions of z_1, z_2, ζ_1, and ζ_2 with the kth analytic hypersurface such that these functions are analytic for z_1, z_2 lying inside the domain of analyticity and for ζ_1, ζ_2 lying in the kth analytic hypersurface which forms a part of the boundary

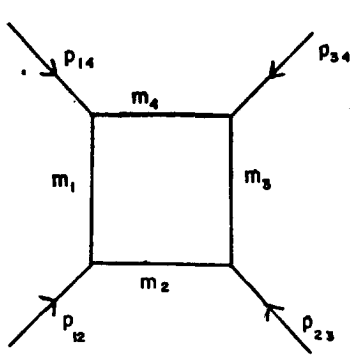


FIG. 1. Fourth-order Feynman diagram for scattering process.

[3] See, for example, A. S. Wightman, *Les Houches Lecture Notes: Dispersion Relations and Elementary Particles*, edited by C. R. De Witt and C. R. Omnes (John Wiley & Sons, Inc., New York, 1959). Also G. Källén and J. Toll, Helv. Phys. Acta 33, 753 (1960).

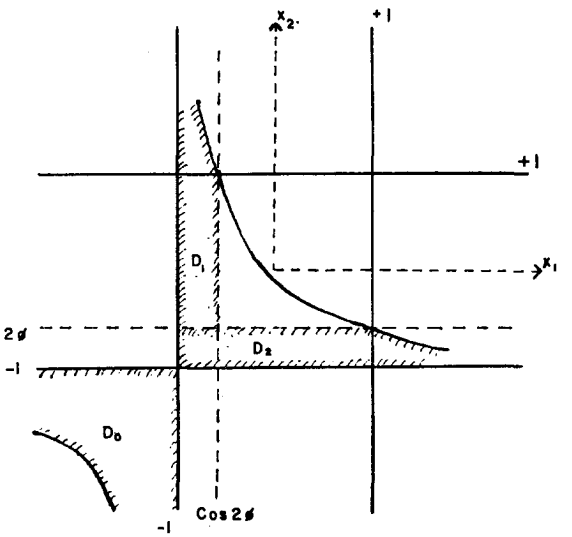


FIG. 2. The real plane when the mass invariants are equal.

and such that the following identity is satisfied

$$(\zeta_1 - z_1)q^{k1} + (\zeta_2 - z_2)q^{k2} = 1.$$

The q factor for the intersection of the kth and lth analytic hypersurfaces is then the 2×2 determinant of the q's associated with these two surfaces. Lastly, $A(\zeta_1, \zeta_2)$ is the value of the function as ζ_1, ζ_2 approach the distinguished boundary in some suitable manner.

3. SINGULARITIES OF $A(y_{13}, y_{24})$

The analytic properties of the amplitude associated with the above diagram have been worked out in detail by Tarski[2] for general values of the internal and external masses.

From continuity one can readily deduce these properties for degenerate cases. We shall first consider the case in which the four mass invariants are equal. We set

$$\theta_{12} = \theta_{23} = \theta_{34} = \theta_{41} = \phi$$

and choose the masses such that $\phi > \pi/2$, so that the amplitude has complex singularities. The amplitude in this case has anomalous threshold branch points at $y_{24} = \cos 2\phi$, $y_{13} = \cos 2\phi$ in addition to the normal threshold ones at $y_{24} = -1$, $y_{13} = -1$. It is also singular on the surface

$$(z_1 + 1)(z_2 + 1) = 4\cos^2\phi.$$

(We set $y_{24} = z_1$, $y_{13} = z_2$). This is essentially the complex surface joining the two branches of the hyperbola (Fig. 2).

$$(x_1 + 1)(x_2 + 1) = 4\cos^2\phi.$$

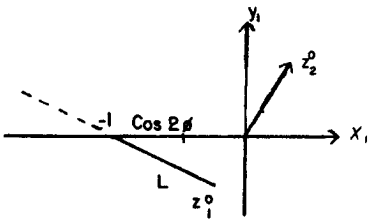


FIG. 3. The z_1 plane for a fixed complex value of z_2.

Thus $A(z_1, z_2)$ has singularities in the physical sheets for z_1 and z_2 which in this case are defined by

$$-\pi < \arg(z_1 - \cos 2\phi) < \pi$$

$$-\pi < \arg(z_2 - \cos 2\phi) < \pi.$$

We can now enumerate the analytic hypersurfaces that form the boundary of the domain of analyticity of $A(z_1, z_2)$. These are just the Z_1 cut, the z_2 cut, and the cut

$$\Delta(z_1, z_2) \equiv (z_1 + 1)(z_2 + 1) - 4\cos^2\phi = \rho, \quad \rho \leq 0.$$

(We call this the Δ cut.) It is necessary to introduce this cut as the singularity associated with the surface. $\Delta = 0$ is known to be of the square root type. This cut can of course be introduced arbitrarily but as we shall see the above choice is a convenient one. We note that since this cut introduces a two-sheeted surface in the z_1 plane for a particular value of z_2, the definition of the physical sheet is ambiguous, but the other singularities in the physical sheet are unambiguously determined. We merely choose a particular one of these sheets.

4. THE BERGMAN-WEIL INTEGRAL FOR $A(z_1, z_2)$.

We first determine the q's for the three cuts. For the z_1 and z_2 cuts we take the sets $[1/(\zeta_1 - z_1), 0]$ and $[0, 1/(\zeta_2 - z_2)]$, respectively, as in Cauchy's formula.

For the Δ cut we see that $\Delta(z_1, z_2) \neq \rho$, $\rho \leq 0$ as long as z_1, z_2 lie inside the domain of analyticity. We may write

$$1 \equiv \frac{\{\Delta(z_1, z_2) - \rho\} - \{\Delta(\zeta_1, \zeta_2) - \rho\}}{\Delta(z_1, z_2) - \rho}$$

$$\equiv \frac{\Delta(z_1, z_2) - \Delta(\zeta_1, \zeta_2)}{\Delta(z_1, z_2) - \rho}$$

since $\Delta(\zeta_1, \zeta_2) = \rho$ for ζ_1, ζ_2 lying in the Δ cut. We now choose Q_1, Q_2 such that

$$(\zeta_1 - z_1)Q_1 + (\zeta_2 - z_2)Q_2 \equiv \Delta(z_1, z_2) - \Delta(\zeta_1, \zeta_2).$$

Then the q's for the Δ cut are given by

$$q^{\Delta 1} = Q_1/[\Delta(z_1, z_2) - \rho], \quad q^{\Delta 2} = Q_2/[\Delta(z_1, z_2) - \rho],$$

since then

$$(\zeta_1 - z_1)q^{\Delta 1} + (\zeta_2 - z_2)q^{\Delta 2} = 1.$$

We find

$$Q_1 = -\tfrac{1}{2}(\zeta_2 + z_2 + z) \qquad Q_2 = -\tfrac{1}{2}(\zeta_1 + z_1 + 2).$$

The ρ may be replaced by $\Delta(\zeta_1, \zeta_2)$. We thus have finally

$$q^{\Delta 1} = -\frac{(\zeta_2 + z_2 + 2)}{2(z_1 z_2 + z_1 + z_2 - \zeta_1\zeta_2 - \zeta_1 - \zeta_2)},$$

$$q^{\Delta 2} = -\frac{(\zeta_1 + z_1 + 2)}{2(z_1 z_2 + z_1 + z_2 - \zeta_1\zeta_2 - \zeta_1 - \zeta_2)}.$$

We may remark that for a general analytic hypersurface the q's are not unique. For the Δ cut the above q's seem to be the simplest ones.

Next we determine the distinguished boundary of our domain. It consists of the intersections of the three cuts taken two at a time. There are thus three contributions to the Bergman-Weil integral, coming from z_1 cut $\cap$ z_2 cut, z_1 cut $\cap$ Δ cut, and z_2 cut $\cap$ Δ cut, respectively. The first of these contributions corresponds to the Mandelstam representation. The second intersection one would expect to be the region in the real plane between the branches of the hyperbola and left of the line $x_1 = \cos 2\phi$. But this region has a two-dimensional intersection with the z_2 cut, viz. the portion bounded by the lines $x_1 = \cos 2\phi$, $x_2 = \cos 2\phi$ and the lower branch of the hyperbola. We thus get a two-dimensional region common to the three cuts whereas in general three analytic hypersurfaces (in two complex variables) should have a one-dimensional intersection. This discrepancy is got over by opening out the z_1 and z_2 cuts slightly.

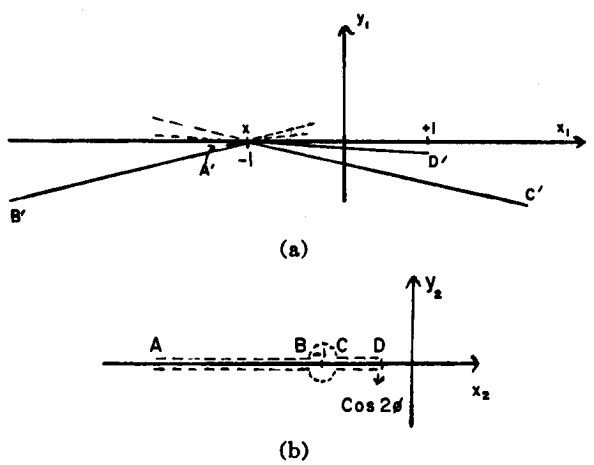


FIG. 4. Plot of the Δ cut on the z_1 plane as z_2 varies along the z_2 cut.

We consider in detail the intersection between the z_2 cut and the Δ cut. This is done conveniently by plotting the intersection in the z_1 plane. We note first that for a particular value of z_2, say z_2^0, the z_1 plane is as shown in Fig. 3. The Δ cut starts from z_1^0 given by

$$(z_1^0 + 1)(z_2^0 + 1) = 4\cos^2\phi$$

and goes through the z_1 cut at $z_1 = -1$ so that only the portion L in the physical sheet is relevant. We now plot z_2 cut $\cap$ Δ cut in the z_1 plane. As z_2 varies from $-\infty$ to $-1-\delta$ (δ small > 0) along $y_2 = \epsilon$, (ϵ small > 0), i.e., from A to B (Fig. 4) the relevant portion of the intersection plotted in the z_1 plane varies from A'X to B'X. The intersection is the surface swept out by the portion A'X going to B'X and gives merely the region D_0 of the real x_1x_2 plane as $\epsilon \to 0$, approaching the z_1 real axis from below. As z_2 goes round the little semicircle from B to C (we take the semicircle for convenience), the intersection on the z_1 plane swerves round from B'X to C'X, and as z_2 goes from C to D we get the surface swept out by C'X going to D'X, which in the limit $\epsilon \to 0$ is just the region D_2 of the real plane, approaching the z_1 real axis from below. For z_2 going round the branch point back to $-\infty$ slightly below the cut, we get the same regions except that these approach the real z_1 axis from above. This gives the complete intersection between the z_2 cut and the Δ cut. In an analogous manner we obtain the third intersection. We can now write down the various contributions to the Bergman-Weil integral. Taking account of the way ζ_1 and ζ_2 approach the various parts of the distinguished boundary, one gets the following representation for $A(z_1, z_2)$:

$$\begin{aligned}A(z_1, z_2) = {} & \iint_{-\infty}^{\cos 2\phi} \frac{d\zeta_1\, d\zeta_2 \rho(\zeta_1\zeta_2)}{(\zeta_1 - z_1)(\zeta_2 - z_2)} \\ & + \iint_{D_0} \left\{ \frac{\sigma_1(\zeta_1\zeta_2)(\zeta_1 + z_1 + 2)}{2F(z, \zeta)(\zeta_1 - z_1)} \right. \\ & \left. + \frac{\sigma_2(\zeta_1\zeta_2)(\zeta_2 + z_2 + 2)}{2F(z, \zeta)(\zeta_2 - z_2)} \right\} d\zeta_1\, d\zeta_2 \\ & + \iint_{D_1} \frac{\sigma(\zeta_1\zeta_2)(\zeta_1 + z_1 + 2)}{2F(z, \zeta)(\zeta_1 - z_1)}\, d\zeta_1\, d\zeta_2 \\ & + \iint_{D_2} \frac{\sigma_2(\zeta_1\zeta_2)(\zeta_2 + z_2 + 2)}{2F(z, \zeta)(\zeta_2 - z_2)}\, d\zeta_1\, d\zeta_2,\end{aligned}$$

where

$$\rho(\zeta_1\zeta_2) = \lim \{A(\zeta_1^+, \zeta_2^+) - A(\zeta_1^+, \zeta_2^-) - A(\zeta_1^-, \zeta_2^+) + A(\zeta_1^-, \zeta_2^-)\}, \qquad \zeta^* = \mathrm{Re}\,\zeta \pm i\epsilon,$$

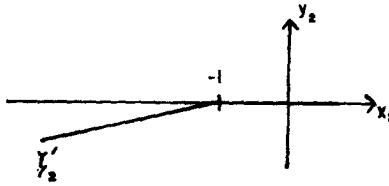


FIG. 5. The z_2 plane for $\zeta_1 = \zeta_1^0 + i\epsilon$, where $\zeta_1^0 < -1$ and ϵ small > 0.

$$\sigma_1(\zeta\zeta) = \lim \{A(\zeta_1^+, \zeta_2^-) - A(\zeta_1^+, \zeta_2^+) - A(\zeta_1^-, \zeta_2^+) + A(\zeta_1^-, \zeta_2^+)\},$$

and $F(z, \zeta) = (z_1z_2 + z_1 + z_2 - \zeta_1\zeta_2 - \zeta_1 - \zeta_2)$. The limit $\{A(\zeta_1^+, \zeta_2^-) - A(\zeta_1^+, \zeta_2^+)\}$ represents the discontinuity of $A(\zeta_1, \zeta_2)$ across the Δ cut, the first superscripts implying that ζ_1 is to approach the ζ_1 cut from above, in which case the Δ cut in the ζ_2 plane lies below the ζ_2 cut. The other limits are defined similarly. σ_2 is obtained from σ_1 by interchanging ζ_1 and ζ_2. $\rho(\zeta_1, \zeta_2)$ corresponds to the Mandelstam spectral function. The first integral may, of course, be over a smaller region.

5. DISPERSION RELATION FOR $A(z_1, z_2)$

For a particular value of $z_2 = z_2^0$, we obtain the following Cauchy representation for $A(z_1, z_2)$ (see Fig. 3):

$$A(z_1, z_2^0) = \left\{ \int_{-\infty}^{\cos 2\phi} + \int_L \right\} \frac{A(\zeta_1^+, z_2^0) - A(\zeta_1^-, z_2^0)}{(\zeta_1 - z_1)}\, d\zeta_1.$$

Thus for a particular value of z_2, we should be able to deduce the above relation from the Berman-Weil representation for $A(z_1, z_2)$. We proceed to obtain this relation.

Let $\zeta_1 = \zeta_1^+ = \zeta_1^0 + i\epsilon$ where $-\infty < \zeta_1^0 < -1$. The Δ cut in the z_2 plane then lies just below the z_2 cut as shown in Fig. 5. starting from ζ_2' given by

$$(\zeta_1^+ + 1)(\zeta_2' + 1) = 4\cos^2\phi.$$

From Cauchy we get, using the same notation for limits as above:

$$\begin{aligned}A(\zeta^+, z^0) = {} & \int_{-\infty}^{\cos 2\phi} \frac{\{A(\zeta_1^+, \zeta_2^+) - A(\zeta_1^+, \zeta_2^-)\}}{(\zeta_2 - z_2^0)}\, d\zeta_2 \\ & + \int_{-1}^{\zeta_2'} \frac{A(\zeta_1^+, \zeta_2^-) - A(\zeta^+, \zeta^-)}{(\zeta_2 - z_2^0)}\, d\zeta_2.\end{aligned}$$

We get similar relations for $-1 < \zeta_1^0 < \cos 2\phi$ and for $\zeta_1^- = \zeta_1^0 - i\epsilon$. We now write the first term of the Bergman-Weil representation in the form

$$\int_{-\infty}^{\cos 2\phi} \frac{d\zeta_1}{(\zeta_1 - z_1)} \left[\int_{-\infty}^{\cos 2\phi} d\zeta_2 \frac{A(\zeta_1^+, \zeta_2^+) - A(\zeta_1^+, \zeta_2^-)}{(\zeta_2 - z_2^0)} - \int_{-\infty}^{\cos 2\phi} \frac{A(\zeta_1^-, \zeta_1^+) - A(\zeta_1^-, \zeta_2^-)}{(\zeta_2 - z_2^0)}\, d\zeta_2 \right]$$

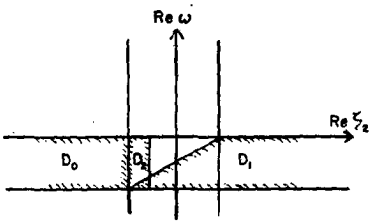


FIG. 6. The real plane of the transformed variables ω and ζ_2.

and substitute for the two terms in the square bracket from the above relations. We get finally for the first term of the Bergman-Weil integral the expression

$$\int_{-\infty}^{\cos 2\phi} \frac{A(\zeta_1^+, z_2^0) - A(\zeta_1^-, z_2^0)}{(\zeta_1 - z_1)}\, d\zeta_1 + \left\{\iint_{D_0} + \iint_{D_1}\right\} \frac{\sigma_1(\zeta_1, \zeta_2)\, d\zeta_1\, d\zeta_2}{(\zeta_1 - z_1)(\zeta_2 - z_2^0)}.$$

Substituting this into the Bergman-Weil integral we get, since

$$\frac{1}{(\zeta_1 - z_1)(\zeta_2 - z_2)} + \frac{(\zeta_1 + z_1 + 2)}{2F(z, \zeta)(\zeta_1 - z_1)} = -\frac{(\zeta_2 + z_2 + 2)}{2F(z, \zeta)(\zeta_2 - z_2)},$$

$$\begin{aligned} A(z_1 z_2^0) = {} & \int_{-\infty}^{\cos 2\phi} d\zeta_1 \frac{A(\zeta_1^+, z_2^0) - A(\zeta_1^-, z_2^0)}{(\zeta_1 - z_1)} \\ & + \iint_{D_0} \frac{\{\sigma_2(\zeta_1, \zeta_2) - \sigma_1(\zeta_1, \zeta_2)\}(\zeta_2 + z_2^0 + 2)}{2F(z, \zeta)(\zeta_2 - z_2^0)} \\ & - \iint_{D_1} d\zeta_1\, d\zeta_2 \frac{\sigma_1(\zeta_1, \zeta_2)(\zeta_2 + z_2^0 + 2)}{2F(z, \zeta)(\zeta_2 - z_2^0)} \\ & + \iint_{D_2} d\zeta_1\, d\zeta_2 \frac{\sigma_2(\zeta_1, \zeta_2)(\zeta_2 + z_2^0 + 2)}{2F(z, \zeta)(\zeta_2 - z_2^0)}. \end{aligned}$$

For the last three terms, we change the integration variables (ζ_1, ζ_2) to (ω, ζ_2) where

$$\omega = (\zeta_1 + 1)(\zeta_2 + 1) - 4\cos^2\phi.$$

The regions of integration in the real (ω, ζ_2) plane are as shown in Fig. 6. Then, noting that the Jacobian of the transformation is given by $1/(\zeta_2 + 1)$ we get for the last three integrals, with $F(z, \zeta) = \omega' - \omega$, $\omega' = (z_1 + 1)(z_2 + 1) - 4\cos^2\phi$,

$$\begin{aligned} & \left\{\iint_{D_0} \{\sigma_2(\zeta_1(\omega), \zeta_2) - \sigma_1(\zeta_1(\omega), \zeta_2)\} \right. \\ & \qquad \left. - \iint_{D_1} \sigma_1(\zeta_1(\omega), \zeta_2) + \iint_{D_2} \sigma_2(\zeta_1(\omega), \zeta_2)\right\} \\ & \qquad \times \frac{(\zeta_2 + z_2^0 + 2)\, d\omega\, d\zeta_2}{2(\omega' - \omega)(\zeta_2 + 1)(\zeta_2 - z_2^0)} \\ & = \int_{-4\cos^2\phi}^{0} \frac{d\omega}{(\omega' - \omega)} \left[\left\{\int_{-\infty}^{-1} (\sigma_2(\zeta_1(\omega), \zeta_2) \right.\right. \\ & \qquad - \sigma_1(\zeta_1(\omega), \zeta_2)) + \int_{\omega/(2\cos^2\phi)+1}^{\infty} (-\sigma_1(\zeta_1(\omega), \zeta_2)) \\ & \qquad \left.\left. + \int_{-1}^{\cos 2\phi} \sigma_2(\zeta_1(\omega), \zeta_2)\right\} \frac{(\zeta_2 + z_2^0 + 2)}{2(\zeta_1 + 1)(\zeta_2 - z_2^0)}\, d\zeta_2\right] \end{aligned}$$

Now the expressions $\sigma(\zeta_1(\omega), \zeta_2) - \sigma(\zeta_1(\omega), \zeta_2)$, etc., contain terms in a symmetric way for which the limits ω^+ and ω^- are to be taken. For a fixed value of $\omega^+ = \omega^0 + i\epsilon$, $-4\cos^2\phi < \omega^0 < 0$, the singularities of $A(\zeta_1(\omega), \zeta_2)$ in the ζ_2 plane are as shown in Fig. 7 where the two extra cuts are in fact parts of the ζ_1 cut for this fixed value of ω. Thus the expression in the square brackets above is the Cauchy integral for $A(\zeta_1(\omega), \zeta_2)$ in the ζ_2 variable for a fixed value of ω with of course the kernel $(\zeta_2 + z_2^0 + 2)/(\zeta_2 + 1)$. The integration over the ζ_2 cut is given by the first part of the first integral and the third integral in the square brackets above, whereas the integration over S and T arise from the second part of the first integral and the second integral respectively. We may thus carry out the integration in the square brackets—for this we merely replace ζ_2 by z_2^0. We note that the kernel $(\zeta_2 + z_2^0 + 2)/2(\zeta_2 + 1)$ cancels out. We obtain finally

$$A(z_1, z_2^0) = \int_{-\infty}^{\cos 2\phi} d\zeta_1 \frac{A(\zeta_1^+, z_2^0) - A(\zeta_1^-, z_2^0)}{(\zeta_1 - z_1)} + \int_{-4\cos^2\phi}^{0} d\omega \frac{A\left(\frac{\omega^+ + 4\cos^2\phi}{z_2^0 + 1} - 1, z_2^0\right) - A\left(\frac{\omega^- + 4\cos^2\phi}{z_2^0 + 1} - 1, z_2^0\right)}{\omega - \omega'}$$

We put $(\omega + 4\cos^2\phi)/(z_2^0 + 1) - 1 = \zeta_1$. Then $d\omega/(\omega - \omega') = d\zeta_1/(\zeta_1 - z_1)$ and the limits are -1 and $4\cos^2\phi/(z_2^0 + 1) - 1 = z_1^0$. Thus

$$A(z_1, z_2^0) = \int_{-\infty}^{\cos 2\phi} d\zeta_1 \frac{A(\zeta_1^+, z_2^0) - A(\zeta_1^-, z_2^0)}{(\zeta_1 - z_1)} + \int_{-1}^{z_1^0} \frac{A(\zeta_1^+, z_2^0) - A(\zeta_1^-, z_2^0)}{(\zeta_1 - z_1)}\, d\zeta_1$$

This expression is the same as the one we obtained previously for $A(z_1, z_2^0)$.

It seems a bit surprising at first that the Bergman-Weil integral for $A(z_1, z_2)$ should contain integrations over the real plane only. This becomes somewhat clear if one considers analytic continuation in the external masses.[4] Let us consider the dispersion

[4] See S. Mandelstam, Phys. Rev. Letters 4, 84 (1960).

relation for $A(z_1, z_2)$ for $z_2 = z_2^0$ (complex) in the situation where there are no complex singularities. It is given by

$$A(z_1, z_2^0) = \int_{-\infty}^{-1} \frac{A_1(\zeta_1, z_2^0)}{\zeta_1 - z_1} d\zeta_1.$$

We continue this relation analytically in y, where $y = y_{12} = y_{23} = y_{34} = y_{41}$. For this we include a small imaginary part in y. When $y > 0$, the singularities of $A(z_1, z_2^0)$ are in the unphysical sheet and as y varies from a value k, $0 < k < 1$ to $-k$, the anomalous threshold and leading singularities move into the physical sheet as shown in Fig. 8, deforming the contour of the above integral with it so that finally we get the dispersion relation we obtained previously. In an analogous manner one can continue the Mandelstam representation, deforming the two dimensional hypercontour as the complex singularities appear in the physical sheets for the two variables. In the first part of the Mandelstam representation, viz.,

$$\iint_{-\infty} d\zeta_1\, d\zeta_2 \frac{A(\zeta_1^+ \zeta_2^+, y)}{(\zeta_1 - z_1)(\zeta_2 - z_2)},$$

we see that it involves the values of A near the real $\zeta_1\zeta_2$ plane so that the corresponding leading singularities are on the unphysical sheet near the real plane. As y goes from k to $-k$, with a small imaginary part, these singularities "graze" along the real plane, go round -1 and back to their original position, deforming the hypercontour in the process, so that the deformed part lies also near the real plane. A detailed consideration shows that we get precisely the regions D_0, D_1 and D_2 for the deformed part. It is not straightforward, however, to get the exact representation by this process.

6. UNEQUAL MASS INVARIANTS

We now consider the cases in which the mass invariants are equal in pairs. The case $\theta_{12} = \theta_{34} = \phi$ and $\theta_{23} = \theta_{41} = \psi$ is entirely analogous to the previous one, since here the Landau curve reduces to the surfaces

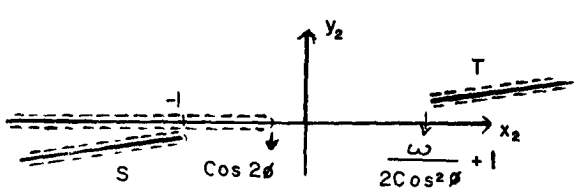


FIG. 7. The z_2 plane for a fixed value of ω.

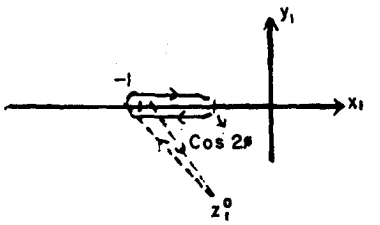


FIG. 8. Path of the singularities in the z_1 plane as the mass invariants increase.

and
$$(z_1 + 1)(z_2 + 1) = (\cos\phi + \cos\psi)^2,$$
$$(z_1 - 1)(z_2 - 1) = (\cos\phi - \cos\psi)^2,$$

of which only the former is singular when $\phi + \psi > \pi$. We consider next the case $\theta_{12} = \theta_{23} = \phi$ and $\theta_{34} = \theta_{41} = \psi$. We describe briefly how in this case the singularities arise in the situation we are interested in. This is contained implicitly in Tarski's paper.

Let $\phi, \psi < \pi/2$ and $\psi > \phi$. There are then no anomalous thresholds or complex singularities in the physical sheet. The Landau curve is as shown in Fig. 9, being given by the line $x_2 = 1$ and the curve

$$(z_2 + 1)(z_1^2 - 1) - 4z_1 \cos\phi \cos\psi + 2\cos^2\phi + 2\cos^2\psi = 0.$$

Only the branch Γ_1 is singular in the limit $(x_1 \pm i\epsilon, k_2 \pm i\epsilon)$, so that the attached complex surface is not singular. The tangents to Γ_2 and Γ_3 are $\cos 2\phi$ and $\cos 2\psi$, respectively. We now increase ϕ and ψ such that $\psi > \phi$. As ψ exceeds $\pi/2$, the line $z_2 = \cos 2\psi$ goes to -1 and recedes, thereby becoming singular and becoming tangent to Γ_1 instead of Γ_3. Next we have $\phi + \psi = \pi$. Here the lines $z_2 = \cos 2\phi$, $z_2 = \cos 2\psi$, and $z_1 = -1$, $z_1 = \cos(\phi + \psi)$ coincide, while Γ_1 and Γ_2 together form a branch of a hyperbola and the line $z_1 = -1$. Beyond this point the line $z_1 = \cos(\phi + \psi)$ becomes singular, the tangents to Γ_1 and Γ_2 are exchanged and we have complex singularities in the physical sheet—being essentially the complex surface joining the portion of Γ_1 with negative slope to the similar portion of Γ_2. As ϕ crosses $\pi/2$ we have again the situation depicted in the figure. Here in addition $z_2 = \cos 2\phi$ is singular and so is the complex surface joining Γ_1 to the portion of Γ_3 with negative slope. We now find a representation for $A(z_1, z_2)$ in this situation—we can compare it readily with the previous case by setting $\phi = \psi$.

If we solve for z_1 the equation to the Landau curve we get

$$z_1 = \frac{2\cos\phi\cos\psi \pm [(z_2 - \cos 2\phi)(z_2 - \cos 2\psi)]^{1/2}}{(z_2 + 1)}.$$

For any z_2 only one of these values of z_1 gives a

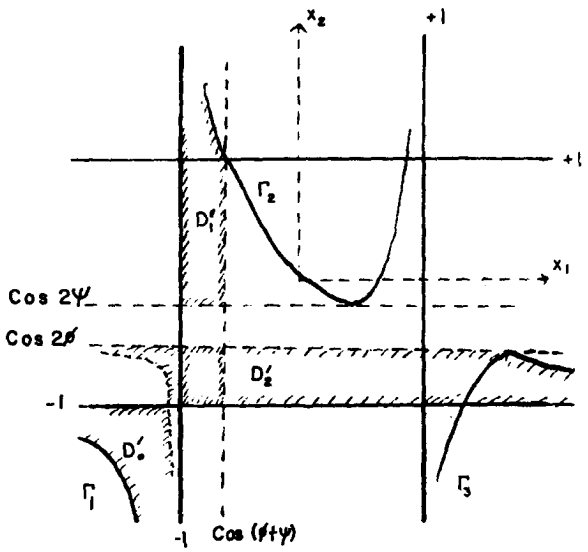


FIG. 9. The Landau curve when $\theta_{12} = \theta_{23}$, $\theta_{34} = \theta_{41}$.

singularity. It can be verified that the sign which gives $-(z_2 - \cos 2\phi)$ in the limit $\phi \to \psi$ is the one yielding a singularity. Thus the relevant portion of the surface may be taken as

$$z_1(z_2 + 1) + \{(z_2 - \cos 2\phi)(z_2 - \cos 2\psi)\}^{1/2} - 2\cos\phi\cos\psi \equiv \Delta'(z_1, z_2) = 0.$$

We note that in the above expression for z_1 if we set $\phi = \psi$ we get the singular surface as

$$z_1 = 4\cos^2\phi/(z_2 + 1) - 1,$$

and the nonsingular one as $z_1 = 1$. This was so in the case considered first.

We now determine the q's for the new surface—taking the usual sets for the z_1 and z_2 cuts. As before we introduce the cut

$$\Delta'(z_1, z_2) = \rho, \qquad \rho \leq 0.$$

To determine the q's we have merely to find Q_1 and Q_2 such that

$$\Delta'(z_1, z_2) - \Delta'(\zeta_1, \zeta_2) \equiv (\zeta_1 - z_1)Q_1 + (\zeta_2 - z_2)Q_2 .$$

Here (ζ_1, ζ_2) lies in the cut and (z_1, z_2) does not. We find $Q_1 = -\frac{1}{2}(\zeta_2 + z_2 + z)$ and,

$$Q_2 = -\frac{1}{2}\left\{\zeta_1 + z_1 + \frac{2(\zeta_2 + z_2 - \cos 2\phi - \cos 2\psi)}{N}\right\},$$

where

$$N = [(z_2 - \cos 2\phi)(z_2 - \cos 2\psi)]^{1/2} + [(\zeta_2 - \cos 2\phi)(\zeta_2 - \cos 2\psi)]^{1/2}.$$

The q's may then be taken as

$$q^{\Delta' i} = Q_i/[\Delta'(z_1, z_2) - \Delta'(\zeta_1, \zeta_2)], \qquad i = 1, 2.$$

We note that these q's reduce to those of the first case when $\phi = \psi$.

Next we determine the distinguished boundary. We note first that the part of the real plane for which $\Delta'(z_1, z_2)$ is negative consists of the region between Γ_1 and parts of Γ_2 and Γ_3 with negative slope, excluding the strip $z_2 = \cos 2\phi, \cos 2\psi$, where $\Delta'(z_1, z_2)$ is imaginary. We thus expect the contribution to the distinguished boundary from the Δ' cut to be contained in this region, with perhaps additional complex surfaces. We determine first the intersection between the portion of the z_2 cut below $z_2 = \cos 2\phi$ and the Δ' cut. For $z_2 = x_2^0 + i\epsilon$, $x_2^0 < \cos 2\phi$, the Δ' cut lies slightly below or above the real z_1 axis according as $\epsilon > 0$ or < 0. This is to be expected as Im z_1 and Im z_2 must have opposite signs for a singularity. As before only the portion of the Δ' cut in the physical sheet is relevant. It can be verified that for any $z_2 = x_2^0 + i\epsilon$, $x_2^0 < \cos 2\psi$, the Δ' cut passes through the z_1 cut, though not through $z_1 = -1$ as in the previous case. The exact form of this part of the distinguished boundary is obtained by determining the point, for any given x_2^0, at which the Δ' cut crosses the z_1 cut in the limit $\epsilon \to 0$. This can be done readily. For $z_2 = x_2 + i\epsilon$, the Δ' cut is given by

$$z_1(x_2 + i\epsilon + 1) + [(x_2 + i\epsilon - \cos 2\phi) \times (x_2 + i\epsilon - \cos 2\psi)]^{1/2} - 2\cos\phi\cos\psi = \rho.$$

Expanding the square root term in ϵ, setting Im $z_1 = 0$ and subsequently equating the imaginary parts we obtain

$$2x_1[(x_2 - \cos 2\phi)(x_2 - \cos 2\psi)]^{1/2} + 2x_2 - \cos 2\phi - \cos 2\psi = 0.$$

Thus this part of the distinguished boundary consists of the region in the real plane bounded by a part of the above curve (the dotted curve in Fig. 9), the relevant portion of the Landau curve and the lines $z_2 = \cos 2\phi, -1$, i.e., the regions D_0' and D_2' in Fig. 9. We note that for $\phi = \psi$ the above curve reduces to $x_1 = -1$ so that D_0' and D_2' reduce to D_0 and D_2 of the previous case. The intersection

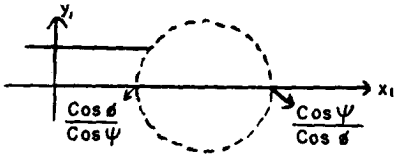


FIG. 10. Plot of the Δ' cut in the z_1 plane for $\cos 2\phi \leq z_2 \leq \cos 2\psi$.

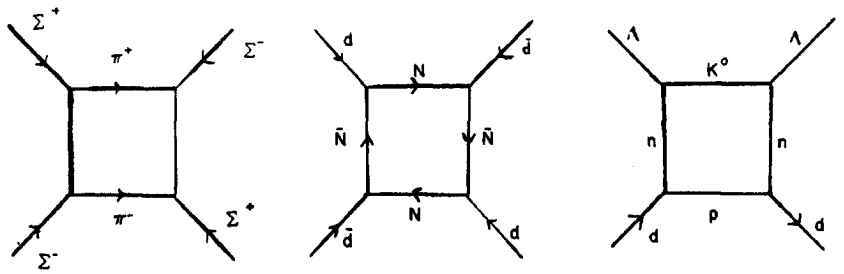


FIG. 11. Diagrams satisfying the modified Mandelstam representation.

between the Δ' cut and the portion $\cos 2\phi \leq z_2 \leq \cos 2\psi$ of the z_2 cut does not lie in the real plane as we get complex values of z_1 for real z_2 (see Fig. 10). As z_2 goes from $\cos 2\phi$ to $\cos 2\psi$ slightly above the cut, z_1 goes from $\cos\psi/\cos\phi$ to $\cos\phi/\cos\psi$ along the lower semicircle and for z_2 going back below the cut z_1 traces the upper semicircle. The Δ' cut is here parallel to the z_1 cut in the limit $\epsilon \to 0$, starting from a z_1 on the circle and going towards $-\infty$. The distinguished boundary is thus the complex surface swept out by this cut as z_2 goes from $\cos 2\phi$ to $\cos 2\psi$ and back, i.e., as z_1 goes round the circle. We call this surface Σ.

For the intersection between the z_1 cut and the Δ' cut we note that for a given z_1 the Δ' cut in the z_2 plane passes through $z_2 = -1$. A detailed consideration shows that this part of the distinguished boundary lies in the real plane consisting of the region bounded by Γ_2, $z_1 = -1$, $\cos(\phi + \psi)$ and $z_2 = -1$, excluding the rectangular bit $z_2 = \cos 2\phi$, $\cos 2\psi$ (the region D_1' in Fig. 9) and the region bounded by Γ_1, $z_1 = -1$, $z_2 = -1$, i.e., a region similar to D_0 of the first case. These regions reduce to D_1 and D_0 of the previous case when $\phi = \psi$.

We can now write down the Bergman-Weil integral for $A(z_1, z_2)$:

$$A(z_1, z_2) = \iint_{-\infty}^{\cos 2\psi \cos(\phi+\psi)} \frac{\rho'(\zeta_1, \zeta_2)}{(\zeta_1 - z_1)(\zeta_2 - z_2)} d\zeta_1\, d\zeta_2 + \iint_{D_1'+D_0} \frac{\sigma'(\zeta_1, \zeta_2) q^{\Delta' 2}}{(\zeta_1 - z_1)} d\zeta_1\, d\zeta_2 + \iint_{D_2'+D_0'+\Sigma} \frac{\tau'(\zeta_1, \zeta_2) q^{\Delta' 1}}{(\zeta_2 - z_2)} d\zeta_1\, d\zeta_2.$$

The third portion of the last integral is over the complex surface Σ, ρ', σ', τ' are the various discontinuity functions obtained by considering the manner in which ζ_1, ζ_2 approach the various parts of the distinguished boundary.

A similar consideration to the above one may be applied to the situation where $\phi < \pi/2$ and $\phi + \psi > \pi$. The case $\theta_{12} = \theta_{14}$ and $\theta_{23} = \theta_{34}$ is analogous to the above one with z_1 and z_2 interchanged.

7. PHYSICAL INTERPRETATION

In Fig. 11 are shown three of the diagrams which do not satisfy the Mandelstam representation. The first two belong to the first case and the third one belongs to the second case considered above (neglecting spins, etc.). One may ask for possible physical interpretation of the additional terms appearing in the representation. The physical significance of anomalous thresholds is well known. In the case of the third-order vertex function, for example, the presence of an anomalous threshold gives an extra term in the dispersion relation for this function which can be interpreted, in the nonrelativistic limit, as "long-range" contributions due to bound structure effects.[5] The terms other than the Mandelstam one appearing in the above representation could thus be interpreted as a manifestation of the compound structure of the particles involved in the scattering. It is not clear if and how this connection can be made precise. It seems to be of some interest to see what the dispersion relation for a fixed physical value of one of the variables looks like. In the case of equal external and equal internal

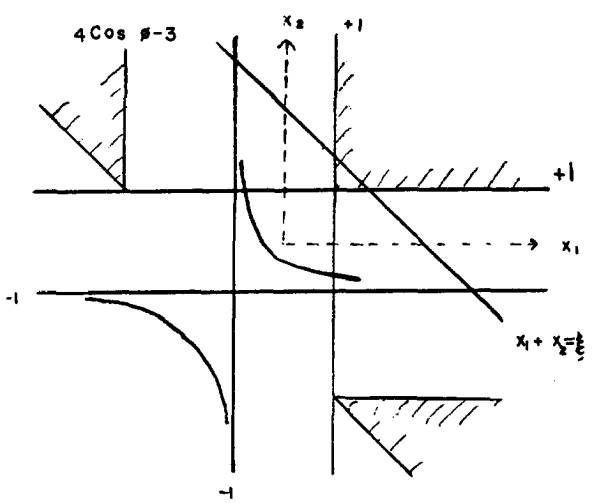


FIG. 12. The physical region in the equal mass case.

[5] R. Oehme, Nuovo cimento 13, 778 (1959).

masses, the physical regions are as shown in Fig. 12 (the shaded regions) so that for a fixed value of the momentum transfer in the physical region we have the following dispersion relation:

$$A(z_1 t) = \int_{-\infty}^{-1} \frac{A_1(\zeta_1 t)}{(\zeta_1 - z_1)}\, d\zeta_1 + \int_{-1}^{\cos 2\phi} d\zeta_1 \frac{A_2(\zeta_1 t)}{(\zeta_1 - z_1)} + \int_{-1}^{g(t)} d\zeta_1 \frac{A_3(\zeta_1 t)}{(\zeta_1 - z_1)},$$

where $g(t) = 4\cos^2\phi/(t+1) - 1$ so that $-1 < g(t) < \cos 2\phi$ for $t > 1$. Here A_1, A_2, and A_3 are the relevant discontinuity functions. We thus get two extra contributions, rather similar to the case of the vertex function, except that the latter has no analogue of the third integral. One gets a similar dispersion relation in the case $\theta_{12} = \theta_{23}$, $\theta_{34} = \theta_{41}$, except that the lower limit of the third integral is also a function of t and lies below -1. We also write down a dispersion relation for a fixed value of the energy of the crossed channel, which in the notation of Sec. 1 is given by $u = (p_{12} + p_{34})^2$. For the second diagram in Fig. 11 this would be the energy of the channel representing deuteron–deuteron scattering. The condition $u =$ constant is equivalent to $z_1 + z_2 =$ constant, the physical values of u being given by $z_1 + z_2 \geq 2$. For a fixed value ξ of $(z_1 + z_2)$ in this region we have the following dispersion relation in z_1:

$$A(z_1, \xi - z_1) = \left\{ \int_{-\infty}^{-1} A_1 + \int_{\xi+1}^{\infty} A_1' + \int_{-1}^{\cos 2\phi} A_2 + \int_{\xi - \cos 2\phi}^{\xi+1} A_2' + \int_{-1}^{a} A_3 + \int_{b}^{\xi+1} A_3' \right\} \frac{d\zeta_1}{(\zeta_1 - z_1)},$$

where a, b are the roots of

$$x^2 - x\xi - \xi - 1 + 4\cos^2\phi = 0$$

and are in fact the points at which the Δ cut starts in the physical sheet. We have $-1 < a, b < +1$. The first, second, and third pairs of terms come from the normal and anomalous thresholds and the Δ cut, respectively, the A's being the relevant discontinuities. For $\phi \to \pi$ (i.e. for deuteron–deuteron scattering) and $\xi \to 2$ (onset of the physical region) we have $a \to 1-$ and $b \to 1+$ so that the integrals extend over the entire real axis.

Note added in proof. While this paper was being typed we received a paper by Fronsdal, Mahanthappa, and Norton[6] with work very similar to the first five sections above.

ACKNOWLEDGMENTS

I am grateful to Dr. J. C. Taylor for suggesting the problem and for help and encouragement. I would like to thank Cambridge University for a research maintenance grant.

[6] C. Fronsdal, K. Mahanthappa, and R. Norton, Phys. Rev. 127, 1847 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Axiomatic Perturbation Theory for Retarded Functions*

H. M. FRIED†
Institute for Advanced Study, Princeton, New Jersey
(Received April 12, 1962)

A construction of the retarded n-point functions of perturbation theory is given within the Lehmann, Symanzik, and Zimmermann framework and without the specification of an interaction Lagrangian. An intermediate-state expansion of retarded functionals is employed to define a systematic set of equations representing approximations to the (integral) unitarity conditions; the requirement of symmetry of the $n-1$ retarded coordinates of an n-point retarded function enters in an essential way. The class of solutions to these equations contains the renormalized perturbation theory retarded functions corresponding to local renormalizable Lagrangian interactions, as well as more singular functions corresponding to nonrenormalizable interactions; if the latter are excluded all the n-point functions may be successively determined to all orders in the renormalized coupling constants. The construction is explicitly performed for the first radiative corrections to the 2- and 3-point functions of a self-interacting neutral scalar boson field, yielding the finite renormalized results of perturbation theory. Similar but slightly singular results are quoted for the π-π scattering amplitude.

I. INTRODUCTION

THE first attempt to compute the renormalized perturbation expansions of field theory directly from the axioms and without a Lagrangian was was made by Lehmann, Symanzik, and Zimmermann[1] (LSZ), who indicated a general method to be used for the construction of the time-ordered functions (equations of system A). More recently, this problem has been considered by Nishijima,[2] who showed that the connected time-ordered functions may be defined in perturbation theory with the aid of parametric dispersion relations.[3] Nishijima also showed that symmetry difficulties apparently prevent the construction of the retarded functions directly from unitarity (with the exception of the special case of quantum electrodynamics, to which the Ward-Takahashi restrictions are applicable).

It will be demonstrated here that it is just the symmetry requirements on the retarded coordinates of every retarded n-point function which permit perturbative solutions to the unitarity equations to be unambiguously determined. However, the method of construction must be carried through step by step for sequences of retarded functions. A systematic approximation scheme to the (integral) unitarity conditions is employed to generate a set of integral equations for the retarded functions. This expansion, essentially an intermediate-state approximation to unitarity, is defined independently of the existence of coupling constants or any perturbative phraseology; but the class of solutions to the equations so obtained contains the renormalized n-point functions of perturbation theory. In effect, and as in references 1 and 2, the approximate unitarity equations replace the coupling between different Green's functions conventionally obtained in a Lagrangian theory. These equations also possess other more singular solutions corresponding to nonrenormalizable Lagrangian theories, which are to be excluded when calculating successively higher-order amplitudes. An explicit construction of the first radiative corrections to the 2- and 3-point retarded functions is carried out and shown to yield the finite, renormalized results of Lagrangian perturbation theory for the simplest case of a self-interacting neutral scalar boson field. However, the π-π scattering amplitude constructed in this manner requires one subtraction.

The basic axiomatic field theory postulates, fundamental to the LSZ formulation followed here, will be adopted and collectively called assumptions A. To these postulates will be added an additional assumption, condition B, which will serve to exclude nonrenormalizable interactions in the sense discussed above. The functional notation and formalism of Symanzik[4] will be adhered to except for some trivial modifications. We therefore con-

* This research was supported in whole or in part by the U. S. Air Force under Grant No. AF-AFOSR-16-19 monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

† Present address: Courant Institute of Mathematical Sciences, New York University, New York 3, New York.

[1] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo cimento 1, 205 (1955).

[2] K. Nishijima, Phys. Rev. 119, 485 (1960).

[3] These relations have been shown to follow from the axioms only for the special case of decay processes, and must therefore be introduced into the formalism as a separate and essential postulate. See H. M. Fried and D. L. Pursey, Phys. Rev. 124, 1281 (1961).

[4] K. Symanzik, J. Math. Phys. 1, 249 (1960).

sider a field operator $A(x)$, corresponding to a stable particle of mass m, which possesses an asymptotic limit (in the weak sense) $A_{in}(x)$, and construct the time-ordered operator

$$T = \left(\exp\left[i\int dx j(x)A(x)\right]\right)_+,$$

where functional differentiation $\delta T/\delta j(x)$ is denoted by T_x. The process of amputation of a given coordinate will be defined by

$$T_x = \delta T/\delta j(\mathbf{x}) = K_x T_x,$$

where $K_x = m^2 - \Box_x$. For a general n-point function this definition agrees with Symanzik's only on the mass shell of the corresponding Fourier momenta, but it is convenient to retain the same terminology. If S represents the unitary S matrix of the theory, the asymptotic condition takes the form

$$ST = :\exp(A_{in}\,\delta/\delta\mathbf{j}):\langle T\rangle,$$

where the double dots denote a Wick product, the bracket means vacuum expectation value, and the symbol $A_{in}\,\delta/\delta\mathbf{j}$ stands for the operator

$$\int du A_{in}(u)\frac{\delta}{\delta j(\mathbf{u})}.$$

The retarded functional operator may be defined as $R_x = -iT^+T_x$, and satisfies a similar asymptotic condition,

$$R_x = :\exp(A_{in}\,\delta/\delta\mathbf{j}):\langle R_x\rangle. \tag{1}$$

Inversion of the defining equation for R_x provides a relation useful in obtaining the time-ordered functions from the retarded functions,

$$T_x = iTR_x, \tag{2}$$

or

$$\langle T_x\rangle = i\langle T\rangle e^D\langle R_x\rangle, \tag{3}$$

where the symbol D represents the operator:

$$i\frac{\overleftarrow{\delta}}{\delta\mathbf{j}}\Delta^{(+)}\frac{\overrightarrow{\delta}}{\delta\mathbf{j}} \equiv i\int du\,dv\frac{\overleftarrow{\delta}}{\delta j(\mathbf{u})}\Delta^{(+)}(u-v)\frac{\overrightarrow{\delta}}{\delta j(\mathbf{v})};$$

this succinct way of representing the infinite sum over all the positive energy intermediate states implicit in Eq. (2) provides a convenient method of expressing the functional expansions to follow. The $(n+1)$-point functions $\langle T_{x,y_1\cdots y_n}\rangle_0$ and $\langle R_{x,y_1\cdots y_n}\rangle_0$ are obtained by functional differentiation of $\langle T_x\rangle$ and $\langle R_x\rangle$; the subscript 0 indicates that the source $j(z)$ has been set equal to zero after the differentiations have been performed. The time-ordered function is symmetric in all $n+1$ indices; the retarded function is symmetric in all ny_α indices, none of which refer to a time coordinate later than x^0.

The basic statement of unitarity for the retarded functions is

$$R_{x,y} = i\theta(xy)[R_x, R_y], \tag{4a}$$

or

$$\langle R_{x,y}\rangle = i\theta(xy)[\langle R_x\rangle, e^D, \langle R_y\rangle], \tag{4b}$$

where $\theta(xy) = \theta(x^0 - y^0)$ is the positive unit step function vanishing for negative values of its argument, and the notation $[A, e^D, B] = Ae^DB - Be^DA$ has been used. A distinction should be made between this relation, here called integral unitarity,[5] and the conventional unitarity conditions obtained by calculating functional derivatives of the difference $\langle R_{x,y}\rangle - \langle R_{y,x}\rangle$. The corresponding integral unitarity condition for the general function $\langle R_{x,y_1\cdots y_n}\rangle_0$ can be obtained by functional differentiation of Eq. (4b); symmetry of the ny_α coordinates may be demonstrated by rewriting the result of such differentiation of Eq. (4a) in the form[6]

$$R_{x,y_1\cdots y_n} = i^n\sum_P \theta(xy_1)\theta(y_1,y_2)\cdots\theta(y_{n-1}y_n) \times[\cdots[[R_x, R_{y_1}], R_{y_2}]\cdots], R_{y_n}], \tag{5}$$

where $\sum_P$ represents the sum over all permutations of the y_α coordinates. The functional expansions given below may be carried out starting from the vacuum expectation value of Eq. (5) rather than (4), but the forms obtained are unnecessarily complicated.

II. LAGRANGIAN PERTURBATION THEORY

Before proceeding with the details of the method it is worthwhile to indicate the motivation for the expansion contemplated and to define an appropriate notation. All the various ways of defining perturbation expansions in a Lagrangian theory have as their common and essential feature the systematic enumeration, in powers of a coupling constant, of the many-particle virtual structure of any process. It is convenient to make this explicit by introducing a functional way of grouping successive orders of n-point functions; this may best be described by considering the simplest sort of nontrivial Lagrangian interaction, $L' = (g/3!)[A(x)]^3$,

[5] In momentum space, Eq. (4) takes the form of an integral over an absorptive type commutator; hence the adjective.

[6] For the 3-point operator, $n = 2$, this follows with the aid of the Jacobi identity and the relation $\theta(xy)\,\theta(xz) = \theta(xy) \times \theta(yz) + \theta(xz)\,\theta(zy)$. The general proof follows by induction on n.

which leads to the field equation

$$K_x A(x) = (g/2)[A(x)]^2. \tag{6}$$

It is of no consequence to this discussion that such a Lagrangian theory does not exist, or that the renormalization constants and counter terms which should be present if Eq. (6) is to refer to a renormalized field have been omitted; all that is of interest here is the form of each of the resulting perturbative retarded n-point functions. If the conventional canonical commutation relations between the field and its time derivative are employed, Eq. (6) may be converted to the functional field equation

$$R_x = j(x) + (g/2)R_x^2 - (g/2)\langle R_x^2\rangle_{j=0}, \tag{7}$$

where the vacuum expectation value of the right-hand side of Eq. (6) has been subtracted off for later convenience. The vacuum expectation value of Eq. (7) is then

$$\langle R_x\rangle = j(x) + (g/2)\langle R_x\rangle e^D\langle R_x\rangle - (g/2)[\langle R_x\rangle e^D\langle R_x\rangle]_{j=0}, \tag{8}$$

which is compatible with the proper behavior as the source vanishes: $\langle R_x\rangle_0 = \langle A(x)\rangle = 0$.

Expansion of the operator e^D generates the full structure and complexity of the theory. This suggests considering an equation for a related functional $\langle R_x^{(0)}\rangle$,

$$\langle R_x^{(0)}\rangle = j(x) + (g/2)\langle R_x^{(0)}\rangle^2, \tag{9}$$

where

$$\langle R_x^{(0)}\rangle_0 = 0. \tag{10}$$

Equation (9) cannot be solved in full generality, but the expansion of $\langle R_x^{(0)}\rangle$ in powers of the source can easily be obtained. A single functional differentiation yields

$$\langle R_{x,y}^{(0)}\rangle = \delta(x-y) + g\langle R_x^{(0)}\rangle\langle R_{x,y}^{(0)}\rangle,$$

which may be written in integral form as

$$\langle R_{x,y}^{(0)}\rangle = [K_x - g\langle R_x^{(0)}\rangle]^{-1}\,\delta(x-y), \tag{11}$$

where $K_x^{-1}\,\delta(x-y) \equiv \Delta_R(x-y)$. From Eq. (11), all the n-point functions of $\langle R_x^{(0)}\rangle$ may be read off by simple functional differentiation combined with Eq. (10); viz.:

$$\langle R_{x,y}^{(0)}\rangle_0 = \Delta_R(x-y),$$

$$\langle R_{x,yz}^{(0)}\rangle_0 = [K_x - g\langle R_x^{(0)}\rangle]^{-1} g\langle R_{x,z}^{(0)}\rangle \times [K_x - g\langle R_x^{(0)}\rangle]^{-1}\,\delta(x-y)|_{j=0}$$

$$= g\int dx'\,\Delta_R(x-x')\,\Delta_R(x'-y)\,\Delta_R(x'-z), \text{ etc.}$$

It is evident that these are just the lowest order connected n-point functions of perturbation theory, the "Born" terms containing no virtual structure; this is, of course, not surprising, since the passage from Eq. (8) to Eq. (9) consisted of the neglect of just such structure.

The Born functional $\langle R_x^{(0)}\rangle$, an approximation to the complete functional $\langle R_x\rangle$, will be called the functional of index zero, and higher-order approximations defined in the following way. The n-point functions of $\langle R_x^{(1)}\rangle$, the functional of index one, shall contain the first radiative corrections (of relative order g^2) to each of the n-point functions of $\langle R_x^{(0)}\rangle$; each n-point function of $\langle R_x^{(2)}\rangle$, the functional of index two, shall represent the second radiative correction to the corresponding Born amplitude (relative order g^4), etc. Thus,

$$\langle R_x\rangle \equiv \sum_{j=0}^{\infty}\langle R_x^{(j)}\rangle,$$

where symbolically, $\langle R_x^{(j+1)}\rangle \sim g^2\,\langle R_x^{(j)}\rangle$; further, for any fixed index j,

$$0[\langle R_{x,y_1\cdots y_{n+1}}^{(j)}\rangle_0] = g0[\langle R_{x,y_1\cdots y_n}^{(j)}\rangle_0].$$

The requirements $\langle R_x^{(j)}\rangle_0 = 0$ will be adopted, in agreement with the form of Eq. (8).

This grouping of the perturbative n-point functions into functionals of different indices is especially convenient, for it permits a functional method of approximating Eq. (8). Insertion of the operator D^m between a pair of functionals and multiplication by the coupling constant produces a quantity

$$g\langle R_{x,y_1\cdots y_\alpha}^{(i)}\rangle\, D^m\langle R_{x,z_1\cdots z_\beta}^{(j)}\rangle|_{j\to 0}$$

of order g^{2m} relative to what would be obtained if the operator D were not present; in this latter case ($m = 0$) the order of the product is $g \times g^{2i+\alpha-1} \times g^{2j+\beta-1}$. Since the order of the n-point function $\langle R_{x,y_1\cdots y_\alpha z_1\cdots z_\beta}^{(i+j+m)}\rangle_0$ is $g^{2(i+j+m)+\alpha+\beta-1}$, it follows that the perturbation expansion in powers of g can be described in terms of the above functional groupings and the expansion of the e^D factor in powers of D. If Eq. (8) is rewritten in the form

$$\langle R_x^\lambda\rangle = j(x) + (g/2)\langle R_x^\lambda\rangle e^{\lambda D}\langle R_x^\lambda\rangle - (g/2)[\langle R_x^\lambda\rangle e^{\lambda D}\langle R_x^\lambda\rangle]_{j=0}, \tag{12}$$

where

$$\langle R_x^\lambda\rangle \equiv \sum_{j=0}^{\infty}\lambda^j\langle R_x^{(j)}\rangle,$$

then the expansion of Eq. (12) in powers of λ is equivalent to the perturbation expansion of every

n-point function. More generally, this intermediate-state functional expansion may be formally defined independently of the coupling constant, but the solutions of the resulting approximate field equations will just generate the perturbation expansions (containing divergences) in powers of g, with the functional groupings described above. With trivial modifications, these statements are independent of the particular interaction Lagrangian employed.

III. AXIOMATIC PERTURBATION THEORY

Without further reference to the nature of the expected results, an intermediate-state functional expansion is defined for the axiomatic functional by rewriting Eq. (4b) in the form

$$\langle R^{\lambda}_{x,y}\rangle = (i/\lambda)\theta(xy)[\langle R^{\lambda}_{x}\rangle, e^{\lambda D}, \langle R^{\lambda}_{y}\rangle] \tag{13}$$

and expanding both sides of Eq. (13) in powers of λ. Again, $\langle R^{\lambda}_{x}\rangle \equiv \sum_{j=0}^{\infty} \lambda^{j}\langle R^{(j)}_{x}\rangle$, and as before, the only duty of the parameter λ is to act as a guide in equating functionals of various indices; at the end of the calculation $\lambda = 1$. All the n-point functions so defined will be understood to possess the conventional translational and rotational invariance properties, and to be connected, real and retarded; it remains to be shown that they may be constructed symmetric in their retarded coordinates.

Upon calculating an arbitrary number of functional derivatives of Eq. (13), the general integral equation which results from this expansion is

$$\begin{aligned}\langle R^{(j)}_{x,y_1\cdots y_n}\rangle_0 &= \theta(xy_1)Q^{(j)}(x, y_1 \cdots y_n)\\ &+ \theta(xy_1)\int du[\langle R^{(j)}_{x,uy_2\cdots y_n}\rangle_0\, \Delta(y_1 - u)\\ &+ \Delta(u - x)\langle R^{(j)}_{y_1,uy_2\cdots y_n}\rangle_0],\end{aligned} \tag{14}$$

where the inhomogeneous $Q^{(j)}$ terms are constructed from retarded functions of index less than or equal to j; those functions of index j which do appear in $Q^{(j)}$ will always be $(m + 1)$-point functions with $m < n$. If Eq. (14) can be solved for $\langle R^{(j)}_{x,y_1\cdots y_n}\rangle_0$ in terms of $Q^{(j)}(x, y_1 \cdots y_n)$, the expansion procedure is then meaningful, since any n-point function of a given index can be constructed from functionals of lower index. Actually, only the fully amputated functions are of physical interest (except for $n = 2$), and it is these which will be obtained from relations of the form of Eq. (14).

For the time sequence $x^0 > y_1^0$ it follows from Eq. (14) that

$$\langle R^{(j)}_{\mathbf{x},\mathbf{y}_1\cdots\mathbf{y}_n}\rangle_0 = Q^{(j)}(\mathbf{x}, \mathbf{y}_1 \cdots \mathbf{y}_n), \tag{15}$$

and therefore, taking into account retardedness,

$$\begin{aligned}\langle R^{(j)}_{\mathbf{x},\mathbf{y}_1\cdots\mathbf{y}_n}\rangle_0 &= \theta(xy_1)Q^{(j)}(\mathbf{x}, \mathbf{y}_1 \cdots \mathbf{y}_n)\\ &+ \delta(x - y_1)f^{(j)}(x, y_2 \cdots y_n).\end{aligned} \tag{16}$$

It is at this point that more singular functions are excluded from the right-hand side of Eq. (16), since, as pointed out by LSZ and Nishijima, such terms may be expected to lead to divergences in higher orders. This restriction will be required so often that it merits an explicit statement: Condition B. No singularities worse than δ functions are to be inserted in any amputated n-point function. All spurious contact terms are also excluded. [The definition of a spurious contact term will be given in Sec. IV (c).]

It must now be demonstrated that the $\langle R^{(j)}_{\mathbf{x},\mathbf{y}_1\cdots\mathbf{y}_n}\rangle_0$ given by Eq. (16), in terms of the assumed known $Q^{(j)}$, are symmetric in all retarded coordinates. An alternate way of writing the right-hand side of Eq. (13) is to define the operator

$$R^{\lambda}_{x} \equiv :\exp(\lambda^{1/2}A_{\mathrm{in}}\,\delta/\delta\mathrm{j}):\langle R^{\lambda}_{x}\rangle,$$

in which case Eq. (13) may be expressed as

$$R^{\lambda}_{x,y} = (i/\lambda)\theta(xy)[R^{\lambda}_{x}, R^{\lambda}_{y}]. \tag{17}$$

No particular physical significance need be attached to the operator R^{λ}_{x}; it merely provides a convenient method of demonstrating the symmetry which appears upon calculating further functional derivtives,

$$\begin{aligned}R^{\lambda}_{x,y_1\cdots y_n} = \frac{i^n}{\lambda^n}\sum_{P}\theta(xy_1)\theta(y_1y_2)\cdots\theta(y_{n-1}y_n)\\ \times[\cdots[[R^{\lambda}_{x}, R^{\lambda}_{y_1}], R^{\lambda}_{y_2}], \cdots], R^{\lambda}_{y_n}],\end{aligned} \tag{18}$$

$$\begin{aligned}\langle R^{\lambda}_{x,y_1\cdots y_n}\rangle = \frac{i^n}{\lambda^n}\sum_{P}\theta(xy_1)\theta(y_1y_2)\cdots\theta(y_{n-1}y_n)\\ \times[\cdots[\langle R^{\lambda}_{x}\rangle, e^{\lambda D}, \langle R^{\lambda}_{y_1}\rangle], \cdots], e^{\lambda D}, \langle R^{\lambda}_{y_n}\rangle].\end{aligned} \tag{19}$$

Equation (18) follows from (17) in the same manner as (5) follows from (4a); the right-hand side of (19) is identically the same as the corresponding quantity obtained from functional differentiation of (13). For the retarded time sequence $x^0 > y_1^0, y_2^0, \cdots y_n^0$ the right-hand side of (19) is symmetric in all the retarded coordinates, and since amputation on all coordinates and the expansion in powers of λ does not change this situation, it follows that $Q^{(j)}(\mathbf{x}, \mathbf{y}_1 \cdots \mathbf{y}_n)$ is symmetric in all the retarded coordinates for $x^0 >$ all y_a^0. However, the only condition used in passing from (14) to (15) was that $x^0 > y_1^0$; no reference was made there to the time sequence of x^0 and the other retarded co-

ordinates. Because of this there can and will always appear in the amputated $Q^{(j)}$ contact terms proportional to $\delta(x - y_\alpha)$, $n \geq \alpha \geq 2$ [in addition to terms proportional to $\delta(y_\alpha - y_\beta)$]; these contact terms are correctly specified if $Q^{(j)}$ is constructed, as assumed, from n-point functions which are themselves symmetric in all retarded coordinates. The only terms which are missing from the amputated $Q^{(j)}$ of Eq. (15) are those proportional to $\delta(x - y_1)$, and the additional term in (16) expresses just this omission. An $\langle R^{(j)}_{x,y_1\cdots y_n}\rangle_0$ completely symmetric in all retarded coordinates can therefore be constructed by the simple expedient of choosing[7] the unknown $f^{(j)}(x, y_2 \cdots y_n)$ such that all the contact terms of form $\delta(x - y_\alpha)$ are symmetric in all the y_α;[8] the nonsingular terms of the amputated $Q^{(j)}$ will carry the appropriate $\theta(xy_\alpha)$ factors and are already symmetric. Finally, if condition B has been consistently invoked and no singularities worse than δ functions have been introduced into any of the retarded functions comprising $Q^{(j)}$, then no such singularities will appear in the amputated $Q^{(j)}$.

With the understanding of the above symmetry requirements on every n-point function of every index, equations for the functionals of different indices may be written down from the λ expansion of Eq. (13),

$$\langle R^{(0)}_{x,y}\rangle = i\theta(xy)[\langle R^{(0)}_x\rangle, D, \langle R^{(0)}_y\rangle], \tag{20}$$

$$\langle R^{(1)}_{x,y}\rangle = i\theta(xy)\{[\langle R^{(0)}_x\rangle, D^2/2!, \langle R^{(0)}_y\rangle] + [\langle R^{(1)}_x\rangle, D, \langle R^{(0)}_y\rangle] + [\langle R^{(0)}_x\rangle, D, \langle R^{(1)}_y\rangle]\}, \text{ etc.} \tag{21}$$

The functionals of index zero are of particular interest since they serve to introduce appropriate renormalized coupling constants; these will be considered in the next section, and the simplest radiative corrections to them treated in Sec. V.

[7] The only remaining unspecified terms are pure connected contact terms of form $\delta(x\text{-}y_1)\ \delta(y_1\text{-}y_2)\cdots\delta(y_{n-1}\text{-}y_n)$; discussion of this point is deferred to Sect. IV (c). Such terms first appear in Eq. (27).

[8] This can always be done if the coefficients of each $\delta(x - y_\alpha)$, $n \geq \alpha \geq 2$, are symmetric in $y_1 y_2 \cdots y_{\alpha-1} y_{\alpha+1} \cdots y_n$. This will always be the case because the $\delta(x - y_\alpha)$ terms arise from the application of K_x to the particular $\theta(xy_\alpha)$ factors of Eq. (19), and the coefficients of these factors are symmetric in all the other y coordinates. Another way of seeing this last point is to consider the fully amputated λ expansion of the right-hand side of (19), which must be symmetric in all y_α,

$$F(x, y_1 \cdots y_n) + \sum_{\alpha=1}^{n} \delta(x - y_\alpha) f^{(\alpha)}(x, y_1 \cdots y_{\alpha-1} y_{\alpha+1} \cdots y_n).$$

Here F represents the nonsingular (in x) fully retarded terms already shown to be symmetric in all y_α; hence the sum of all the singular terms must also be symmetric in all y_α, and this requires that each $f^{(\alpha)}(x, y_1 \cdots y_{\alpha-1} y_{\alpha+1} \cdots y_n)$ be independent of the superscript α and be symmetric in all $n-1$ y coordinates.

IV. FUNCTIONS OF INDEX ZERO

(a) The 2-Point Function

Setting the source equal to zero in Eq. (20) leads to the equation

$$\langle R^{(0)}_{x,y}\rangle_0 = \theta(xy) \int du\, dv\, \Delta(v - u)\langle R^{(0)}_{x,u}\rangle_0 \langle R^{(0)}_{y,v}\rangle_0. \tag{22}$$

The existence of a stable particle of mass m implies that the Fourier transform of the exact $\langle R_{x,y}\rangle_0$ has a pole of unit residue on the mass shell. It is apparent that the Fourier transforms of the zero index functions under the integral in Eq. (22) are evaluated on the mass shell, and these may each be assigned the residue Z, at the moment an arbitrary number. In configuration space this means the replacement of the $\langle R^{(0)}_{x,u}\rangle_0$ factor by $Z\ \delta(x - u)$, and similarly for $\langle R^{(0)}_{y,v}\rangle_0$. Equation (22) then has the immediate solution

$$\langle R^{(0)}_{x,y}\rangle_0 = -Z^2\theta(xy)\ \Delta(x - y) = Z^2\ \Delta_R(x - y).$$

But operation with the Klein-Gordon operator implies that $Z = Z^2$, or $Z = 0$ or 1. The only nonzero solution of (22) is therefore

$$\langle R^{(0)}_{x,y}\rangle_0 = \Delta_R(x - y), \tag{23}$$

which implies that the Fourier transforms of all the remaining 2-point functions are not to have a mass shell pole,

$$\int dv\ \Delta^{(*)}(u - v)\langle R^{(j)}_{x,v}\rangle_0 = 0, \qquad j \geq 1. \tag{24}$$

(b) The 3-Point Function

A single functional differentiation of Eq. (20) yields, in the limit of zero source,

$$\langle R^{(0)}_{x,y_1y_2}\rangle_0 = \theta(xy_1) \int du[\langle R^{(0)}_{x,y_2u}\rangle_0\ \Delta(y_1 - u) + \Delta(u - x)\langle R^{(0)}_{y_1,y_2u}\rangle_0], \tag{25}$$

where (23) has been used. This has the form of (14) with $Q^{(0)}(x, y_1y_2) = 0$; Eqs. (22) and (25) are the only ones for which the corresponding Q terms vanish (unless the 3-point function is itself zero). For the time sequence $x^0 > y_1^0$, it follows from (25) that

$$\langle R^{(0)}_{x,y_1y_2}\rangle_0 = 0, \tag{26}$$

and the most general acceptable solution of (26), symmetric in the retarded coordinates, is given by

$$\langle R^{(0)}_{x,y_1,y_2}\rangle_0 = g\ \delta(x - y_1)\ \delta(x - y_2), \tag{27}$$

where g is an arbitrary constant (with the dimensions of mass) and condition B has been invoked. It is not difficult to show that the somewhat stronger statement

$$\langle R^{(0)}_{x,y_1y_2}\rangle_0 = g\,\Delta_R(x-y_1)\,\Delta_R(x-y_2) \qquad (28)$$

also follows from (25), condition B, and the symmetry restrictions. The constant g will represent the renormalized charge of the theory, and may be defined by equating it to the complete vertex function with all three invariants on the mass shell; this then implies that contact terms (with finite coefficients) of the form of Eq. (27) will be required in the corresponding equations for all three point functions of higher index, such that the vertex functions constructed from the latter shall each vanish on the mass shell of all invariants.

Once Eq. (27) and condition B have been written down, the content of the entire intermediate-state expansion procedure becomes identical to that of the renormalized perturbation expansions. For the same argument used in Sec. II to relate functions of different orders in g may be applied to the λ expansion of Eq. (13); that is, the quantity

$$\langle R^{(i)}_{x,y_1\cdots y_l}\rangle\, D^m\langle R^{(j)}_{y,z_1\cdots z_k}\rangle|_{j\to 0}$$

will be of order $g^{2i+l-1}\cdot g^{2m}\cdot g^{2j+k-1}$, which is the order of $\langle R^{(i+j+m-1)}_{x,y,y_1\cdots y_l z_1\cdots z_k}\rangle_0$. The axiomatic λ expansion is therefore equivalent to an expansion of all n-point functions in powers of the renormalized coupling constant.

(c) The 4-Point Function

In the limit of zero source, a second functional differentiation of Eq. (20) yields

$$\langle R^{(0)}_{x,y_1y_2y_3}\rangle_0 = \theta(xy_1)Q^{(0)}(x, y_1y_2y_3) + \theta(xy_1)\int du[\langle R^{(0)}_{x,y_2y_3u}\rangle_0\,\Delta(y_1-u) + \Delta(u-x)\langle R^{(0)}_{y_1,y_2y_3u}\rangle_0], \qquad (29)$$

where

$$Q^{(0)}(x, y_1y_2y_3) = \int du\,dv\,\Delta(v-u) \times \{\langle R^{(0)}_{x,y_2u}\rangle_0\langle R^{(0)}_{y_1,y_3v}\rangle_0 + \langle R^{(0)}_{x,y_3u}\rangle_0\langle R^{(0)}_{y_1,y_2v}\rangle_0\}, \qquad (30)$$

and Eq. (23) has been used. For the time sequence $x^0 > y_1^0$,

$$\langle R^{(0)}_{x,y_1y_2y_3}\rangle_0 = Q^{(0)}(\mathbf{x}, \mathbf{y}_1\mathbf{y}_2\mathbf{y}_3),$$

and again following the method of Sec. III,

$$\langle R^{(0)}_{x,y_1y_2y_3}\rangle_0 = \theta(xy_1)Q^{(0)}(\mathbf{x}, \mathbf{y}_1\mathbf{y}_2\mathbf{y}_3) + \delta(x-y_1)f^{(0)}(y_1, y_2y_3). \qquad (31)$$

Substitution of Eq. (27) into the amputated form of (30) yields

$$Q^{(0)}(\mathbf{x}, \mathbf{y}_1\mathbf{y}_2\mathbf{y}_3) = -g^2\,\Delta(x-y_1) \times [\delta(x-y_2)\,\delta(y_1-y_3) + \delta(x-y_3)\,\delta(y_1-y_2)],$$

and Eq. (31) then becomes

$$\langle R^{(0)}_{x,y_1y_2y_3}\rangle_0 = g^2\,\Delta_R(x-y_1)[\delta(x-y_2)\,\delta(y_1-y_3) + \delta(x-y_3)\,\delta(y_1-y_2)] + \delta(x-y_1)f^{(0)}(y_1, y_2y_3). \qquad (32)$$

Clearly, the correct choice for $f^{(0)}$ is

$$f^{(0)}(y_1, y_2y_3) = g^2\,\Delta_R(y_1-y_2)\,\delta(y_2-y_3),$$

which leads to a 4-point function completely symmetric in all retarded coordinates; this is the same result as that obtained directly from Eq. (11).

To the right-hand side of (32) may be added the connected contact term

$$G\,\delta(x-y_1)\,\delta(y_1-y_2)\,\delta(y_2-y_3),$$

which corresponds to the quartic interaction generated in Lagrangian theory by an interaction term $(G/4!)[A(x)]^4$; for simplicity this is neglected here (see Sec. VI). The possibility of including such pure contact terms arises in the calculation of every amputated n-point function ($n > 2$). From conventional renormalization theory it is known that interaction Lagrangians of form $(\Gamma/l!)[A(x)]^l$ are not renormalizable for $l \geq 5$, and therefore, as expressed by the complete condition B, the corresponding spurious contact terms are to be excluded from all amputated n-point functions ($n \geq 5$).

With Eq. (27) and the restrictions of symmetry, retardedness and condition B, it is evident from these examples that the λ-expansion procedure generates a zero index functional identical to the Born functional of Lagrangian perturbation theory. This has been explicitly verified for the 5-point function, and it is intuitively clear that it is true for the entire functional.

V. FUNCTIONS OF INDEX ONE

The simplest radiative corrections to the propagator and retarded vertex will be calculated in this section, in order to show that no infinite quantities enter into their equations, and to demonstrate the symmetrization procedure of Sec. III for a nontrivial amplitude.

(a) The 2-Point Function

Setting the source equal to zero in Eq. (21) leads to the equation

$$\langle R^{(1)}_{x,y}\rangle_0 = \theta(xy)Q^{(1)}(x,y),$$

$$Q^{(1)}(x,y) = -\frac{i}{2}\int du_1\, du_2\, dv_1\, dv_2 \times [\Delta^{(+)}(u_1 - v_1)\,\Delta^{(+)}(u_2 - v_2) - \Delta^{(-)}(u_1 - v_1)\,\Delta^{(-)}(u_2 - v_2)] \times \langle R^{(0)}_{x,u_1u_2}\rangle_0\langle R^{(0)}_{y,v_1v_2}\rangle, \quad (33)$$

where the homogeneous term of the integral equation vanishes because of (24); it will be obvious that the latter condition is satisfied by the $\langle R^{(1)}_{x,y}\rangle_0$ given by (33). From Eqs. (27) and (28), the form of the zero index 3-point function needed in (33) is uniquely determined; substitution into (33) yields the result

$$\langle R^{(1)}_{x,y}\rangle_0 = \frac{-ig^2}{2}\,\theta(xy)\int du\, dv\, \Delta_R(x-u)\,\Delta_R(y-v) \times [\Delta^{(+)}(u-v)\,\Delta^{(+)}(u-v) - \Delta^{(-)}(u-v)\,\Delta^{(-)}(u-v)]. \quad (34)$$

Inserting representations for θ, Δ_R and $\Delta^{(*)}$ into (34), the momentum space propagator defined by

$$\langle R_{x,y}\rangle_0 = (2\pi)^{-4}\int dp\, r(p)e^{ip\cdot(x-y)}$$

then has a radiative correction of order g^2

$$r^{(1)}(p) = \frac{g^2}{32\pi^2}\int_{4m^2}^{\infty} d\kappa^2[\kappa^2 + p^2 - i\epsilon(p)]^{-1} \times [\kappa^2 - m^2]^{-2}[1 - 4m^2/\kappa^2]^{1/2},$$

which is the renormalized perturbation theory result.

(b) The 3-Point Function

A single functional differentiation of Eq. (21) yields the equation

$$\langle R^{(1)}_{x,yz}\rangle_0 = \theta(xy)Q^{(1)}(x,yz) + \theta(xy)\int du \times \{\langle R^{(1)}_{x,zu}\rangle_0\,\Delta(y-u) + \Delta(u-x)\langle R^{(1)}_{y,zu}\rangle_0\} \quad (35)$$

where

$$Q^{(1)}(x,yz) = -\frac{i}{2}\int du_1\, du_2\, dv_1\, dv_2 \times [\Delta^{(+)}(u_1 - v_1)\,\Delta^{(+)}(u_2 - v_2) - \Delta^{(-)}(u_1 - v_1)\,\Delta^{(-)}(u_2 - v_2)] \times [\langle R^{(0)}_{y,v_1v_2}\rangle_0\langle R^{(0)}_{x,zu_1u_2}\rangle_0 + \langle R^{(0)}_{x,u_1u_2}\rangle_0\langle R^{(0)}_{y,zv_1v_2}\rangle_0], \quad (36)$$

and Eqs. (23) and (24) have again been used. Proceeding in the standard way, the 3- and 4-point Born functions are inserted into the fully amputated form of (36) to compute $Q^{(1)}(\mathbf{x}, \mathbf{yz})$, the latter is multiplied by a factor $\theta(xy)$, and to this combination is added an unknown contact term proportional to $\delta(x-y)$; the result is

$$\begin{aligned}\langle R^{(1)}_{\mathbf{x},\mathbf{yz}}\rangle_0 = {} & -ig^3\{\theta(xy)\,\Delta_R(x-z)[\Delta^{(+)}(x-y)\,\Delta^{(+)}(z-y) - \Delta^{(-)}(x-y)\,\Delta^{(-)}(z-y)] \\ & + \theta(xy)\,\Delta_R(y-z)[\Delta^{(+)}(x-z)\,\Delta^{(+)}(x-y) - \Delta^{(-)}(x-z)\,\Delta^{(-)}(x-y)] \\ & + \delta(x-z)\cdot\theta(zy)\cdot\tfrac{1}{2}\int du\,\Delta_R(z-u)[\Delta^{(+)}(u-y)\,\Delta^{(+)}(u-y) - \Delta^{(-)}(u-y)\,\Delta^{(-)}(u-y)] \\ & + \delta(y-z)\cdot\theta(xy)\cdot\tfrac{1}{2}\int du\,\Delta_R(z-u)[\Delta^{(+)}(x-u)\,\Delta^{(+)}(x-u) - \Delta^{(-)}(x-u)\,\Delta^{(-)}(x-u)]\} \\ & + \delta(x-y)f^{(1)}(y,z) + gZ_{(1)}\,\delta(x-y)\,\delta(y-z). \end{aligned} \quad (37)$$

A little algebra shows that the first and second terms of Eq. (37) may be rewritten in the form

$$\begin{aligned} -ig^3\{ & \Delta_R(x-y)\theta(yz)[\Delta^{(+)}(x-z)\,\Delta^{(+)}(y-z) - \Delta^{(-)}(x-z)\,\Delta^{(-)}(y-z)] \\ & + \Delta_R(x-z)\theta(zy)[\Delta^{(+)}(x-y)\,\Delta^{(+)}(z-y) - \Delta^{(-)}(x-y)\,\Delta^{(-)}(z-y)]\}, \end{aligned} \quad (38)$$

and, as expected, are symmetric in the retarded coordinates. The function $f^{(1)}$ is determined by the requirement of symmetry to be

$$f^{(1)}(y,z) = -\frac{i}{2}\,g^3\theta(yz)\int du\,\Delta_R(y-u) \times [\Delta^{(+)}(u-z)\,\Delta^{(+)}(u-z) - \Delta^{(-)}(u-z)\,\Delta^{(-)}(u-z)]. \quad (39)$$

The constant $Z_{(1)}$ is determined by the (arbitrary) condition that the mass shell vertex function of order g^3 shall vanish; there will be a corresponding $Z_{(i)}$ factor to all orders of the 3-point function. It should perhaps be again emphasized that such pure contact terms are permitted only for (and in each order of) the 3- and 4-point functions; for n-point functions with $n \geq 5$ such contact terms become spurious and are excluded by condition B.

For the simple case considered here of a single (vertex) interaction, such terms should be consistently excluded from the 4-point functions.

It is interesting to compare this result with that of the corresponding Lagrangian calculation, which is most simply obtained by the λ expansion of Eq. (12). To avoid confusion, the retarded functions derived from the field equation will be distinguished by the notation $\langle \tilde{R}_x \rangle$. A simple calculation yields

$$\begin{aligned}\langle \tilde{R}^{(1)}_{x,yz} \rangle_0 = \frac{i}{2} g^3 \Big\{ & \Delta_R(x-z)\, \Delta_R(x-y)[\Delta^{(+)}(z-y) - \Delta^{(-)}(z-y)] \\ & + \Delta_R(x-y)\, \Delta_R(y-z)[\Delta^{(+)}(x-z) - \Delta^{(-)}(x-z)] \\ & + \Delta_R(x-z)\, \Delta_R(z-y)[\Delta^{(+)}(x-y) - \Delta^{(-)}(x-y)] \\ & + \delta(x-y) \int du\, \Delta_R(x-u)\, \Delta_R(u-z)[\Delta^{(+)}(z-u) - \Delta^{(-)}(z-u)] \\ & + \delta(x-z) \int du\, \Delta_R(x-u)\, \Delta_R(u-y)[\Delta^{(+)}(y-u) - \Delta^{(-)}(y-u)] \\ & + \delta(y-z) \int du\, \Delta_R(x-u)\, \Delta_R(u-y)[\Delta^{(+)}(x-u) - \Delta^{(-)}(x-u)] \Big\}. \end{aligned} \tag{40}$$

The first three terms of Eq. (40) are symmetric in y and z, and it is easy to show that they are identical to the contribution of Eq. (38); no divergences occur in the sum of the Fourier transforms of these terms since they correspond to the triangle diagram which is finite in this theory. The contact terms of Eq. (40) differ from those of the axiomatic result by the presence of θ functions under the integrals rather than standing outside of them; this indicates a divergence most simply seen in the Fourier transforms. If the amputated 3-point function in momentum space is defined by

$$r(\mathbf{q};\mathbf{pk}) = (2\pi)^{-4} \int dx\, dy\, dz \times \exp\,[i(q\cdot x + p\cdot y + k\cdot z)]\langle R_{x,yz}\rangle_0,$$

then the contact terms of (40) lead to the contributions $\tilde{r}^{(1)}_c$,

$$\begin{aligned}\tilde{r}^{(1)}_c(\mathbf{q};\mathbf{pk}) = {} & \frac{g^3}{32\pi^2}\, \delta(q+p+k) \int_{4m^2}^{\infty} d\kappa^2 \Big[1 - \frac{4m^2}{\kappa^2}\Big]^{1/2} \{[k^2 + m^2 - i\epsilon(k)]^{-1}[\kappa^2 + k^2 - i\epsilon(k)]^{-1} \\ & + [p^2 + m^2 - i\epsilon(p)]^{-1}[\kappa^2 + p^2 - i\epsilon(p)]^{-1} + [q^2 + m^2 + i\epsilon(q)]^{-1}[\kappa^2 + q^2 + i\epsilon(q)]^{-1}\}, \end{aligned} \tag{41}$$

which represent the diagrams obtained by insertion of a bubble into each of the q, p, k lines. The divergence of these integrals is due to the omission of a mass renormalization counter term in the original field Eq. (6). If the renormalization is performed by subtracting from each integral of (41) its mass shell value, there results

$$\begin{aligned}\tilde{r}^{(1)}_{c,R}(\mathbf{q};\mathbf{p}k) = {} & \frac{-g^3}{32\pi^2} \int_{4m^2}^{\infty} d\kappa^2 \Big[1 - \frac{4m^2}{\kappa^2}\Big]^{1/2} \\ & \times [\kappa^2 - m^2]^{-1}\{[\kappa^2 + k^2 - i\epsilon(k)]^{-1} + [\kappa^2 + p^2 - i\epsilon(p)]^{-1} + [\kappa^2 + q^2 + i\epsilon(q)]^{-1}\}, \end{aligned} \tag{42}$$

which is precisely the contribution obtained by directly calculating the Fourier transforms of the axiomatic contact terms.

VI. DISCUSSION

Analogous calculations have been carried through for the retarded perturbative π-π scattering amplitudes (neglecting isotopics) to order G^3; here, G represents the renormalized π-π coupling constant first appearing in the amputated 4-point Born function,

$$\langle R^{(0)}_{x,y_1y_2y_3}\rangle_0 = G\, \delta(x-y_1)\, \delta(y_1-y_2)\, \delta(y_2-y_3). \tag{43}$$

The first radiative correction to this amplitude may be calculated with the aid of (43),

$$\langle R^{(1)}_{x,y_1y_2y_3}\rangle_0 = -(i/2)\,G^2\{\delta(x-y_1)\,\delta(y_2-y_3)\theta(y_1y_2)[\Delta^{(+)}(y_1-y_2)\,\Delta^{(+)}(y_1-y_2)-\Delta^{(-)}(y_1-y_2)\,\Delta^{(-)}(y_1-y_2)]$$
$$+\,\delta(x-y_2)\,\delta(y_1-y_3)\theta(y_2y_3)[\Delta^{(+)}(y_2-y_3)\,\Delta^{(+)}(y_2-y_3)-\Delta^{(-)}(y_2-y_3)\,\Delta^{(-)}(y_2-y_3)]$$
$$+\,\delta(x-y_3)\,\delta(y_2-y_1)\theta(y_3y_1)[\Delta^{(+)}(y_3-y_1)\,\Delta^{(+)}(y_3-y_1)-\Delta^{(-)}(y_3-y_1)\,\Delta^{(-)}(y_3-y_1)]\}$$
$$+\,Z'_{(1)}\,\delta(x-y_1)\,\delta(y_1-y_2)\,\delta(y_2-y_3), \tag{44}$$

but in contrast to the previous situation, the constant $Z'_{(1)}$ must appear and be assigned an infinite value in order to specify the equal-time dependence of the right-hand side of (44). If the momentum space amplitude is defined by

$$r(\mathbf{q};\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3) = (2\pi)^{-4}\int dx\,dy_1\,dy_2\,dy_3 \times \exp\,[i(q\cdot x + p_iy_1 + p_2\cdot y_2 + p_3\cdot y_3)]\langle R_{x,y_1y_2y_3}\rangle_0$$

then the Fourier transform of (44) is given by

$$r^{(1)}(\mathbf{q};\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3) = \frac{G^2}{32\pi^2}\,\delta(q+p_1+p_2+p_3)\int_{4m^2}^{\infty} d\kappa^2\left[1-\frac{4m^2}{\kappa^2}\right]^{1/2}$$
$$\times\{[(p_1+p_2)^2+\kappa^2-i\epsilon(p_1+p_2)]^{-1}+[(p_1+p_3)^2+\kappa^2-i\epsilon(p_1+p_3)]^{-1}$$
$$+\,[(p_2+p_3)^2+\kappa^2-i\epsilon(p_2+p_3)]^{-1}\}+Z'_{(1)}\cdot\delta(q+p_1+p_2+p_3). \tag{45}$$

Choosing the constant $Z'_{(1)}$ such that as a function of the invariant variables s, t, u, and with all four momenta on the mass shell,

$$r(s,t,u)|_{s-t-u=4/3m^2} \equiv G.$$

then the condition

$$r^{(1)}(s,t,u)|_{s-t-u=4/3m^2} = 0$$

serves to determine $Z'_{(1)}$, which, of course, disappears from the final result,

$$r^{(1)}(q;p_1p_2p_3) = -\frac{G^2}{32\pi^2}\,\delta(q+p_1+p_2+p_3)\int_{4m^2}^{\infty}\frac{d\kappa^2[1-4m^2/\kappa^2]^{1/2}}{[\kappa^2-\frac{4}{3}m^2]}\left\{\frac{(p_1+p_2)^2+\frac{4}{3}m^2}{[(p_1+p_2)^2+\kappa^2-i\epsilon(p_1+p_2)]}\right.$$
$$\left.+\,\frac{(p_1+p_3)^2+\frac{4}{3}m^2}{[(p_1+p_3)^2+\kappa^2-i\epsilon(p_1+p_3)]}+\frac{(p_2+p_3)^2+\frac{4}{3}m^2}{[(p_2+p_3)^2+\kappa^2-i\epsilon(p_2+p_3)]}\right\}.$$

Two further statements should be made concerning the relation of this perturbation construction to those of other methods. The first remark is to note that, with the aid of Eqs. (3) or (4), any connected time-ordered n-point function of order g^k may be constructed from retarded functions of order less than or equal to k. As a method of computing renormalized time-ordered functions, this procedure is straight-forward in principle but cumbersome in practice; one quickly appreciates the relative simplicity of conventional graphical and renormalization techniques. As a method of computing double graphs, the axiomatic construction is simpler than that of the corresponding Lagrangian calculations.

The second remark concerns the apparent difficulty of proving the convergence of the renormalized Fourier transforms of all retarded amplitudes of arbitrary order. While the entire construction may easily be transcribed into momentum space (at the price of having certain symmetry statements obscured), more than the conventional power counting theorem is required to demonstrate convergence. This is essentially due to the need of retaining the knowledge of which momenta correspond to specific retarded configuration space coordinates, and this information cannot be specified by counting powers of momenta alone.

It should also be remarked that a proof of the Lorentz invariance of every amplitude would be very desirable. It has been assumed that all radiative corrections constructed in the manner described here will turn out to be relativistically invariant. The essential point in question is the requirement that all θ functions, of whatever argument, must eventually multiply a function which vanishes for spacelike separation of the same argument; e.g., $\theta(x)\,\Delta(x)$ is invariant but $\theta(x)\,\Delta_{(1)}(x)$ is not. All the expressions obtained above possess this property, as do all the terms (some of considerable complexity) obtained to order G^3 for the π-π scattering amplitude.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. R. Stora and Professor A. Wightman for helpful and encouraging discussions, and to thank Professor J. R. Oppenheimer for extending the hospitality of The Institute for Advanced Study.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

On the Restricted Lorentz Group and Groups Homomorphically Related to It*

A. J. MACFARLANE
Department of Physics and Astronomy, University of Rochester, Rochester, New York

A study is made of the real restricted Lorentz group, L, and of its relationship
(a) to the group, SL(2C), of complex unimodular two-dimensional matrices, and
(b) to the group, O_3, of orthogonal transformations in a complex space of three dimensions.
The discussion of case (a) is an improved version of the treatment by Wightman. Its notable features are, firstly, that it gives important formulas in new concise forms and their proofs in an elegant and economical manner, and, secondly, that it deals with the nontrivial matter of proving the internal consistency of the formalism. To illustrate the practical utility of the theory, the product of two nonparallel pure Lorentz transformations is studied. In the discussion of case (b), explicit formulas realizing the isomorphism of O_3 and L are obtained. These formulas are new and have been applied, for illustrative purposes, to the derivation of the transformation properties under L of the electromagnetic field vectors, regarded as a complex three-vector $(\mathbf{E} + i\mathbf{H})$. A result analagous to the factorization of the general element of L into a spatial rotation and a pure Lorentz transformation, and to the polar decomposition of the general element of SL(2C), is derived for O_3. Insight into the relationship of O_3 to L is provided by considering the unimodular matrix description of the complex Lorentz group, and the contrasting specializations of it that lead to the unimodular matrix descriptions of its subgroups, O_3 and L.

1. INTRODUCTION

IN this paper, a detailed discussion is given of the theory of the transformations that constitute the real restricted[1] Lorentz group, L, and its relationship to the group, SL(2C), of complex unimodular two-dimensional matrices, and to the group, O_3, of orthogonal transformations in a complex space of three dimensions.

The small part of the present work that refers to Lorentz transformations themselves is principally of the nature of a review of well-known matter.[2] It is given in the interests of completeness and for convenience of reference in the study of the relationship of SL(2C) and O_3 to L. The only new material involved is a set of identities, Eq. (54), below, satisfied by the transformation coefficients L^μ_ν of a proper Lorentz transformation, and part of the discussion of the product of two pure Lorentz transformations in nonparallel directions.

The homomorphism of SL(2C) onto L has been noted and already studied by many authors.[3-6] The present discussion however is quite new, although based on that of Wightman,[2] over which it offers several important advantages. Firstly, it is consistent with the usual notations of tensor calculus in Minkowski space; secondly, it presents formulas in their most concise form and their proofs in an elegant and economical manner; and thirdly, it gives a proof of the internal consistency of the formalism. In the provision of such a proof the identities mentioned in the previous paragraph are essential. The new formulation of the theory of the homomorphism has emerged chiefly from the realization[7] that the Pauli matrix four-vector $\tau^\mu = (\tau^0, \tau^k)$ does not by itself provide an adequate algebraic basis for the discussion of two-dimensional unimodular matrices. It is necessary to introduce also the auxiliary matrix four-vector ρ^μ

$$\rho^\mu = \zeta(\tau^\mu)^*\zeta^{-1} \tag{1}$$

* Research supported in part by the U. S. Atomic Energy Commission.

[1] The various adjectives restricted, proper, etc., used in connection with Lorentz transformations, are all defined below.

[2] See, for example, the books by Fock [V. A. Fock, *The Theory of Space, Time and Gravitation*, translated by N. Kemmer (Pergamon Press, New York, 1959)], and Møller [C. Møller, *The Theory of Relativity* (Oxford University Press, New York, 1952)], or else the lectures by Tolhoek [H. A. Tolhoek, Part II of lectures given at CERN in 1959 on the Representations of the Lorentz Group in Quantum Mechanics (unpublished)], and Wightman [A. S. Wightman, lectures given at Les Houches in 1960 on "Invariance in Relativistic Quantum Mechanics," in *Dispersion Relations and Elementary Particles*, edited by C. de Witt and R. Omnès (John Wiley & Sons, New York, 1960)]. The two sets of lectures provide a comprehensive and, to some extent, complementary discussion of Lorentz transformations. The lectures by Tolhoek are based on lectures given by Wigner [E. P. Wigner, lectures given at Leyden in 1957 on Relativistic Invariance in Quantum Mechanics, report by R. M. F. Houtappel (unpublished)].

[3] B. L. van der Waerden, *Die Gruppentheoretische Methode in der Quantenmechanik*, (Verlag Julius Springer, Berlin, 1932), p. 57.

[4] V. Bargmann, Ann. Math. 48, 568 (1947).

[5] M. A. Neumark, Am. Math. Soc. Trans. 6, 379 (1957).

[6] A. S. Wightman, reference 2.

[7] This point has previously been made by Brown [L. M. Brown, Phys. Rev. 111, 957 (1958)], who also uses the quantities ρ^μ, as defined by Eq. (1).

with $\zeta = i\tau^2$ and the asterisk denoting complex conjugation. Once various algebraic relationships satisfied by the quantities τ^μ and ρ^μ have been derived, investigation of the homomorphisms and the formulas that realize it explicitly is readily undertaken. It is quite interesting to notice that specialization of the formalism to the subgroup, R_3, of spatial rotations of L and to the unitary subgroup, SU(2C), of SL(2C) leads to an extremely compact exposition of the essential formulas of the quaternion theory[8,9] of rotations.

The isomorphism of O_3 to L has also been noted by several authors,[10-13] but no investigation of formulas that realize it explicitly has been undertaken before. These formulas are obtained here by relating the transformations of O_3 to those of SL(2C) and hence to those of L. They are then applied, for purposes of illustration, to the derivation of the transformation properties under L of the electromagnetic field vectors by building from them the complex three-vector $\mathbf{E} + i\mathbf{H}$. Familiar results[14] are thus reproduced in quite an interesting manner. Further insight into the relationship of O_3 and L is obtained by consideration of the complex Lorentz group, which contains them both as subgroups. For the unimodular matrix description of their transformations can be obtained by appropriate and contrasting specializations of the unimodular matrix description of complex Lorentz transformations. The discussion of complex Lorentz transformations necessary to clarify these remarks is a reformulation of the work of Wightman,[3] which offers the same advantages over his discussion as were mentioned in the real case.

One aspect of the relationship between the three groups, L, SL(2C), and O_3, that is worthy of notice is that for each of the two latter groups one has a decomposition of the general element analogous to the unique resolution of the general element of L into a spatial rotation and a pure Lorentz transformation. In the case of SL(2C), one has the well-known polar decomposition of a unimodular 2×2 matrix into a unitary 2×2 matrix and a Hermitian positive 2×2 matrix, which, as stressed by Wightman, is extremely important in practical work on Lorentz transformations. The analogous result for O_3 given below by Eq. (137) is likewise a general matrix theorem,[15] and, though of less interest practically, is important for the completeness of the theory.

The present work ends with a discussion of the (somewhat artificial) transition from the 2×2 unimodular description of Lorentz transformations to the 4×4 (Dirac) matrix description that allows the representation of inversions to be included. This is aimed at providing a link between the matter discussed here and the discussion given in text books on field theory.[16,17] It further focuses attention on the fact that the homomorphism of SL(2C) onto L and the associated work on two component spinor theory can be discussed in terms of the usual operations of matrix algebra. As a corollary to this remark, one sees that use of the methods of spinor calculus[18,19] leads to unnecessary algebraic complication.

Concluding the introduction, a brief sketch of the contexts of the ensuing sections of the paper is given. Section 2 deals carefully with matters of notation and with the properties of the completely antisymmetric tensors $\epsilon^{\mu\nu\kappa\lambda}$ and ϵ^{ijk} in Minkowski and ordinary three-dimensional space. Section 3 gives the algebra of the matrices τ^μ and ρ^μ. Section 4 gives a variety of usual remarks on Lorentz transformation as well as the identities (54) mentioned above. Section 5 deals with the homomorphism of SL(2C) onto L, Sec. 6 with spatial rotations and pure Lorentz transformations. Section 7 contains the work on O_3, and Sec. 8 discusses inversions and the four-dimensional (Dirac) matrix description of Lorentz transformations.

2. NOTATION

Let $v^\mu = (v^\circ, \mathbf{v}) = (v^\circ, v^k)$ and $v_\mu = (v^\circ, -\mathbf{v})$ denote the contravariant and covariant components of a four-vector in Minkowski space with metric tensor $g^{\mu\nu} = g_{\mu\nu}$ such that

[8] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1950), Chap. 4.
[9] H. C. Corben, P. Stehle, *Classical Mechanics* (John Wiley & Sons, New York, 1960), 2nd ed., Appendix IV.
[10] A. Einstein, W. Mayer, Sitzber preuss. Akad. Wiss. Physik-math-Kl. p. 522 (1932).
[11] E. Cartan, *Lecons sur la Théorie des Spineurs, Actualités Sci. et Ind.* (Hermann & Cie Paris, 1938), Nos. 643 and 701, Secs. 82, 160.
[12] G. Racah, Nuovo cimento Suppl. **14**, 75 (1959).
[13] P. Roman, *Theory of Elementary Particles* (North-Holland Publishing Company, Amsterdam, 1960), p. 60.
[14] See Sec. 24 of Fock's book.[2]
[15] F. R. Gantmacher, *Applications of Theory of Matrices*, translated by J. L. Brenner (Interscience Publishers, Inc., New York, 1959), p. 4.
[16] S. S. Schweber, *Introduction of Quantum Field Theory* (Row, Peterson & Company, Evanston, 1961).
[17] N. N. Bogoliubov, V. Shirkov, *Theory of Quantized Fields*, translated by G. M. Volkoff (Interscience Publishers, Inc., New York, 1958).
[18] E. M. Corson, *Tensors, Spinors and Relativistic Wave-Equations* (Blackie and Son Limited, London, 1953).
[19] W. L. Bade, H. Jehle, Revs. Modern Phys. **25**, 714 (1953).

$$g^{\mu\nu} = 0, \qquad \mu \neq \nu,$$
$$g^{00} = -g^{11} = -g^{22} = -g^{33} = 1. \tag{2}$$

Lower-case Greek letters are used for indices which may take on the values 0, 1, 2, and 3, and summation over these values is implied for any Greek letter appearing in any multiplicative expression once as a subscript and once as a superscript. Lower-case Latin letters from i onwards are used for indices that only take on the values 1, 2, and 3. Subsequent equations are arranged[20] so that Latin indices appear as superscripts only, and then summation over the values 1, 2, and 3 is implied for repeated ones. Thus, for the scalar product of two four-vectors u and v one has

$$\begin{aligned} u\cdot v = u_\mu v^\mu &= g^{\mu\nu}u_\mu v_\nu \\ &= u^\circ v^\circ - u^k v^k = u^\circ v^\circ - \mathbf{u}\cdot\mathbf{v}. \end{aligned} \tag{3}$$

The Levi-Civita symbol $\epsilon^{\mu\nu\kappa\lambda}$ is a completely antisymmetric tensor defined as follows:

$\epsilon^{\mu\nu\kappa\lambda} = 0$, unless its indices take on distinct values,

$\epsilon^{\mu\nu\kappa\lambda} = S$, when they do, where S is the signature of the permutation that carries the set of distinct values $\mu\nu\kappa\lambda$ into 1230. Thus, $\epsilon^{1230} = -\epsilon_{1230} = 1$. For later use, the following identities involving the Levi-Civita symbol are needed

$$\epsilon^{\mu\nu\kappa\lambda}\epsilon_{\mu\nu\kappa\alpha} = -3!\ \delta^\lambda{}_\alpha, \tag{4}$$
$$\epsilon^{\mu\nu\kappa\lambda}\epsilon_{\mu\nu\alpha\beta} = -2!\,(\delta^\kappa{}_\alpha\,\delta^\lambda{}_\beta - \delta^\kappa{}_\beta\,\delta^\lambda{}_\alpha), \tag{5}$$
$$\begin{aligned} \epsilon^{\mu\nu\kappa\lambda}\epsilon_{\mu\alpha\beta\gamma} = -1!\,(&\delta^\nu{}_\alpha\,\delta^\kappa{}_\beta\,\delta^\lambda{}_\gamma + \delta^\nu{}_\gamma\,\delta^\kappa{}_\alpha\,\delta^\lambda{}_\beta \\ &+ \delta^\nu{}_\beta\,\delta^\kappa{}_\gamma\,\delta^\lambda{}_\alpha - \delta^\nu{}_\alpha\,\delta^\kappa{}_\gamma\,\delta^\lambda{}_\beta - \delta^\nu{}_\gamma\,\delta^\kappa{}_\beta\,\delta^\lambda{}_\alpha \\ &- \delta^\nu{}_\beta\,\delta^\kappa{}_\alpha\,\delta^\lambda{}_\gamma), \end{aligned} \tag{6}$$

where $\delta^\mu{}_\nu = g^\mu{}_\nu = g_\nu{}^\mu$ is a Kronecker delta. These results can be verified directly.

The Levi-Civita symbol of ordinary space is ϵ^{ikl}, defined similarly to $\epsilon^{\mu\nu\kappa\lambda}$, with $\epsilon^{123} = 1$. For it, one notes the results

$$\epsilon^{ikl}\epsilon^{ikn} = 2\,\delta^{ln}, \tag{7}$$
$$\epsilon^{ikl}\epsilon^{imn} = (\delta^{km}\,\delta^{ln} - \delta^{kn}\,\delta^{lm}). \tag{8}$$

Also

$$(\mathbf{u}\wedge\mathbf{v})^k = \epsilon^{klm}u^l v^m.$$

Next, the subject of matrix notation is taken up, noting first that $\tilde{M}$, M^*, $M^\dagger$, M^{-1}, respectively, denote the transpose, complex conjugate, Hermitian conjugate, inverse of the matrix M. In agreement with the convention of denoting the contravariant and covariant components of a four-vector in Minkowski space, respectively, by a column and a row matrix of four elements, time component first, one may associate with a mixed second-rank tensor $T^\mu{}_\nu$ a matrix T, according to

$$T = \begin{bmatrix} T^0{}_0 & T^0{}_1 & T^0{}_2 & T^0{}_3 \\ T^1{}_0 & T^1{}_1 & T^1{}_2 & T^1{}_3 \\ T^2{}_0 & T^2{}_1 & T^2{}_2 & T^2{}_3 \\ T^3{}_0 & T^3{}_1 & T^3{}_2 & T^3{}_3 \end{bmatrix}. \tag{9}$$

The elements of the transposed matrix $\tilde{T}$ are related to those of T by

$$\tilde{T}^\mu{}_\nu = T^\nu{}_\mu. \tag{10}$$

for $\det T$, one has the expression

$$\det T\,\epsilon_{\alpha\beta\gamma\delta} = \epsilon_{\mu\nu\kappa\lambda}T^\mu{}_\alpha T^\nu{}_\beta T^\kappa{}_\gamma T^\lambda{}_\delta, \tag{11}$$

and, since $TT^{-1} = 1$, one can write also

$$\det T\,\epsilon_{\alpha\beta\gamma\delta}(T^{-1})^\gamma{}_\kappa(T^{-1})^\delta{}_\lambda = \epsilon_{\mu\nu\kappa\lambda}T^\mu{}_\alpha T^\nu{}_\beta. \tag{12}$$

In ordinary space, one associates with a tensor R^{ik} of second rank a matrix R, according to

$$R = \begin{bmatrix} R^{11} & R^{12} & R^{13} \\ R^{21} & R^{22} & R^{23} \\ R^{31} & R^{32} & R^{33} \end{bmatrix}, \tag{13}$$

with $\det R$ given by

$$\det R\,\epsilon^{ikl} = \epsilon^{pqr}R^{pi}R^{qk}R^{nl}. \tag{14}$$

Using $RR^{-1} = 1$, this gives

$$\det R\,\epsilon^{ikl}(R^{-1})^{lr} = \epsilon^{pqr}R^{pi}R^{qk}. \tag{15}$$

If $T^\mu{}_\nu$ is a tensor with $T^{oo} = 1$, $T^{ok} = T^{ko} = 0$, one can associate with its spatial components a matrix R defined by

$$R^{ik} = T^i{}_k, \tag{16}$$

which agrees with the association for the unit case

$$\delta^{ik} = -g^{ik} = g^i{}_k = \delta^i{}_k. \tag{17}$$

This last point has been mentioned with a particular view to the restriction of Lorentz transformation formulas to spatial rotation formulas. For, as in the case of several other points noted in this section, this is a place where errors of sign can creep in.[21]

3. ALGEBRAIC PROPERTIES OF MATRIX VECTORS τ^μ AND ρ^μ.

One begins with the properties of the familiar

[20] With a few exceptions such as Eqs. (16), (17) below.

[21] For example, the equations in reference 18 that correspond to Eqs. (4)–(6) contain an error of sign on one side.

Pauli matrices τ^k which are a set of three 2×2 matrices defined by the law of multiplication

$$\tau^i\tau^k = \delta^{ik} + i\epsilon^{ikl}\tau^l. \tag{18}$$

In the usual representation, one has

$$\tau^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \tau^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \tau^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{19}$$

From (18), one gets

$$\tau^i\tau^k + \tau^k\tau^i = 2\,\delta^{ik}, \tag{20}$$

$$\tau^i\tau^k\tau^l = \delta^{ik}\tau^l - \delta^{il}\tau^k + \delta^{kl}\tau^i + i\epsilon^{ikl}, \tag{21}$$

and, from (18) and (21), one gets

$$\mathrm{Tr}\,(\tau^i\tau^k) = 2\,\delta^{ik}, \tag{22}$$

$$\mathrm{Tr}\,(\tau^i\tau^k\tau^l) = 2i\epsilon^{ikl}. \tag{23}$$

Using (18), one can simplify the product of four Pauli matrices and hence prove

$$\mathrm{Tr}\,(\tau^i\tau^k\tau^l\tau^m) = 2(\delta^{ik}\,\delta^{lm} - \delta^{il}\,\delta^{kn} + \delta^{im}\,\delta^{kl}). \tag{24}$$

When one attempts to pass from the above discussion to the properties of the quantities $\tau^\mu = (\tau^0, \tau^k)$, where τ^0 is the unit matrix of two rows and columns, one soon sees that a concise statement in terms of the quantities $g^{\mu\nu}$, $\epsilon^{\mu\nu\kappa\lambda}$ cannot be obtained. Accordingly, one introduces also ρ^μ

$$\rho^\mu = \zeta(\tau^\mu)^*\zeta^{-1} = (\tau^0, -\tau) \tag{25}$$

where[22] $\zeta = i\tau^2 = -\zeta^{-1}$. Now Eq. (18) can be extended to give the law of multiplication

$$\tau^\mu\rho^\nu = g^{\mu\nu} + \tfrac{1}{2}i\epsilon^{\mu\nu\kappa\lambda}\tau_\kappa\rho_\lambda \tag{26}$$

or

$$\rho^\mu\tau^\nu = g^{\mu\nu} - \tfrac{1}{2}i\epsilon^{\mu\nu\kappa\lambda}\rho_\kappa\tau_\lambda. \tag{27}$$

Equations (26) and (27) can be generated, the one from the other, by means of the replacements $\tau \rightarrow \rho$, $\rho \rightarrow \tau$, $i \rightarrow -i$. Indeed, it follows from (25) that one can proceed from any identity to an equally valid one by this means. This should be remembered in connection with the identities (28) to (34) proved below.

From (26), one can easily prove

$$\tau^\mu\rho^\nu + \tau^\nu\rho^\mu = 2g^{\mu\nu}, \tag{28}$$

$$\tau^\mu\rho^\nu - \tau^\nu\rho^\mu = -i\epsilon^{\mu\nu\kappa\lambda}\tau_\kappa\rho_\lambda. \tag{29}$$

One may also extend (21) to give

$$\tau^\mu\rho^\nu\tau^\kappa = g^{\mu\nu}\tau^\kappa - g^{\mu\kappa}\tau^\nu + g^{\nu\kappa}\tau^\mu - i\epsilon^{\mu\nu\kappa\lambda}\tau_\lambda, \tag{30}$$

and apply (26) to it to prove

$$\begin{aligned}\tau^\mu\rho^\nu\tau^\kappa\rho^\lambda = {} & g^{\mu\nu}g^{\kappa\lambda} - g^{\mu\kappa}g^{\nu\lambda} + g^{\mu\lambda}g^{\nu\kappa} - i\epsilon^{\mu\nu\kappa\lambda} \\ & + \tfrac{1}{2}i(g^{\mu\nu}\epsilon^{\kappa\lambda\alpha\beta} - g^{\mu\kappa}\epsilon^{\nu\lambda\alpha\beta} + g^{\mu\lambda}\epsilon^{\nu\kappa\alpha\beta} \\ & + g^{\kappa\lambda}\epsilon^{\mu\nu\alpha\beta} - g^{\nu\lambda}\epsilon^{\mu\kappa\alpha\beta} + g^{\nu\kappa}\epsilon^{\mu\lambda\alpha\beta})\tau_\alpha\rho_\beta.\end{aligned} \tag{31}$$

From (25) and (31), one next obtains the results

$$\mathrm{Tr}\,(\tau^\mu\rho^\nu) = 2g^{\mu\nu}, \tag{32}$$

$$\begin{aligned}&\mathrm{Tr}\,(\tau^\mu\rho^\nu\tau^\kappa\rho^\lambda) \\ &\quad = 2(g^{\mu\nu}g^{\kappa\lambda} - g^{\mu\kappa}g^{\nu\lambda} + g^{\mu\lambda}g^{\nu\kappa} - i\epsilon^{\mu\nu\kappa\lambda}).\end{aligned} \tag{33}$$

The remaining identity to be proved for later use is one of a different type. It is

$$(\tau^\mu)_{ab}(\rho_\mu)_{cd} = 2\,\delta_{ad}\,\delta_{bc}, \tag{34}$$

with a, b, c, and d ordinary matrix row and column labels. Equation (34) may be directly proved in the representation (19). It may be presented alternatively in the bizarre but useful form

$$(\tau^i)_{ab}(\tau^i)_{cd} = 2\,\delta_{ad}\,\delta_{bc} - \delta_{ab}\,\delta_{cd}. \tag{35}$$

4. LORENTZ TRANSFORMATIONS

A Lorentz transformation is defined to be a real linear transformation of the type

$$x^\mu \rightarrow x'^\mu = L^\mu{}_\nu x^\nu, \tag{36}$$

in Minkowski space, which leaves invariant the quadratic form

$$g^{\mu\nu}x_\mu x_\nu.$$

The transformation coefficients thus satisfy

$$g_{\mu\nu}L^\mu{}_\kappa L^\nu{}_\lambda = g_{\kappa\lambda}, \tag{37}$$

or

$$L^\mu{}_\kappa L_\nu{}^\kappa = \delta^\mu{}_\nu. \tag{38}$$

In matrix notation, one writes (36) as

$$x \rightarrow x' = L\cdot x, \tag{39}$$

with L satisfying

$$\tilde{L}I_s L = I_s, \tag{40}$$

and I_s is given by

$$I_s = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{41}$$

From (40) one sees that the set of all Lorentz trans-

[22] The matrix ζ is identical to the matrix that plays the role of metric spinor in spinor calculus. See reference 9.

formations form a group: the full Lorentz group. In particular, if L is a Lorentz transformation so is L^{-1}

$$LL^{-1} = 1,$$
$$L^{\mu}{}_{\kappa}(L^{-1})^{\kappa}{}_{\nu} = \delta^{\mu}{}_{\nu}. \tag{42}$$

From (38) and (42), one can see that

$$(L^{-1})^{\kappa}{}_{\nu} = L_{\nu}{}^{\kappa}, \tag{43}$$

or

$$(L^{-1})^{\kappa\nu} = L^{\nu\kappa}. \tag{44}$$

Equation (40) is equivalent to ten independent conditions on the coefficients $L^{\mu}{}_{\nu}$

$$\begin{aligned} L^{oo}L^{oo} - L^{ok}L^{ok} &= 1, \\ L^{oo}L^{oo} - L^{io}L^{ik} &= 0, \\ L^{oi}L^{ok} - L^{li}L^{lk} &= -\delta^{ik}. \end{aligned} \tag{45}$$

Similarly, from the fact that L^{-1} is also a Lorentz transformation, one also has, using (44)

$$\begin{aligned} L^{oo}L^{oo} - L^{ko}L^{ko} &= 1, \\ L^{oo}L^{ko} - L^{oi}L^{ki} &= 0, \\ L^{io}L^{ko} - L^{il}L^{kl} &= -\delta^{ik}. \end{aligned} \tag{46}$$

From (40), it follows that[23]

$$\det L = \pm 1 \tag{47}$$

and, from the first line of (45), (46), it follows that

$$(L^{oo})^2 \geq 1,$$

that is,

$$L^{oo} \geq 1, \quad \text{or} \quad L^{oo} \leq -1. \tag{48}$$

Lorentz transformations with $\det L = +1$ (-1) are named proper (improper); those with $L^{oo} \geq 1$ (≤ -1) are named orthochronous (nonorthochronous). From (47) and (48), one sees that the set of all Lorentz transformations is divided into four disjoint continuously connected subsets characterized by

	I	II	III	IV
$\det L$	1	-1	-1	1
L^{oo}	≤ 1	≤ 1	≤ -1	≤ -1.

(49)

The elements of subset I constitute the restricted or proper orthochronous Lorentz group.[24] In subsequent work the letter L will always refer to an element of the restricted Lorentz group, unless indication to the contrary is given; the adjective restricted will, further, often be left out. In agreement with this, the elements of subsets II through IV can be generated from the elements L of subset I in the forms

$$I_s L, \qquad I_t L, \qquad I_s I_t L \tag{50}$$

where I_s, given by (41), is the matrix describing space inversion

$$x^{\mu} \to x'^{\mu}, \qquad x'^{0} = x^{0}, \qquad \mathbf{x}' = -\mathbf{x}, \tag{51}$$

and $I_t = -I_s$ describes time inversion.

Next, the conditions (40) on the $L^{\mu}{}_{\nu}$ are shown to be replaceable by two distinct sets of conditions, of which one holds for proper and the other for improper Lorentz transformations. One starts with the expansion of the determinant of L in the form (12), and uses Eq. (43), which is fully equivalent to (40), to write this in the form

$$\epsilon_{\mu\nu\kappa\lambda}L^{\mu}{}_{\alpha}L'^{\nu}{}_{\beta} = \pm\epsilon_{\alpha\beta\gamma\delta}L_{\kappa}{}^{\gamma}L_{\lambda}{}^{\delta}, \tag{52}$$

with the upper and lower signs for the cases $\det L = +1$, -1, respectively. On multiplying both sides of (52) by $-\epsilon^{\alpha\sigma\kappa\lambda}$ and using Eqs. (5) and (6) to simplify the left and right sides, one is led, after some algebra, to

$$\begin{aligned} 2(\mathrm{Tr}\ L)L^{\sigma\beta} - 2(LL)^{\sigma\beta} = \pm g^{\sigma\beta}\{(\mathrm{Tr}\ L)^2 - \mathrm{Tr}\ (LL)\} \\ \mp 2(\mathrm{Tr}\ L)L^{\beta\sigma} \pm 2(LL)^{\beta\sigma}. \end{aligned} \tag{53}$$

For the case of proper Lorentz transformations, Eq. (53) rearranged to read as

$$\begin{aligned} 2(\mathrm{Tr}\ L)(L^{\beta\sigma} + L^{\sigma\beta}) = g^{\sigma\beta}\{(\mathrm{Tr}\ L)^2 - \mathrm{Tr}\ (LL)\} \\ + 2(LL)^{\sigma\beta} + 2(LL)^{\beta\sigma}. \end{aligned} \tag{54}$$

This identity is an essential step in the proof of the consistency of the formalism developed in the next section.

Further discussion of the transformations of the restricted Lorentz group is deferred until its relationship to SL(2C) has been investigated.

5. DESCRIPTION OF LORENTZ TRANSFORMATIONS BY COMPLEX UNIMODULAR 2 × 2 MATRICES

With each point x^{μ} of Minkowski space, one may associate a 2×2 matrix X by setting

[23] Tolhoek[2] proves the theorem: Of every set of four linearly independent four-vectors that are mutually orthogonal, three are space-like and one time-like. Regarding the four columns (or rows) of L as a set of four four-vectors, Eq. (47) implies their linear independence, and then Eq. (45) [or Eq. (46)] contains the rest of the information stated in the theorem.

[24] The only nontrivial part of the proof of this statement is the proof that the product of two orthochronous Lorentz transformations is again orthochronous. This is proved in the lectures cited in footnote 2.

$$X = x_\mu \tau^\mu \tag{55}$$

If x^μ is real, X is Hermitian and conversely. Also

$$\det X = x_\mu x^\mu. \tag{56}$$

From (55), one obtains the inverse relationship

$$x^\mu = \tfrac{1}{2} \operatorname{Tr} (\rho^\mu X) \tag{57}$$

by using (32).

If A is a complex unimodular (i.e., $\det A = 1$) matrix of two rows and columns, it is easily seen that the transformation

$$X \to X' = AXA^\dagger \tag{58}$$

preserves the Hermiticity and the determinant of X, which suggests that it is intimately related to the Lorentz transformation

$$x^\mu \to x'^\mu = L^\mu{}_\nu x^\nu. \tag{59}$$

It is the aim of this section to investigate this relationship fully.

From (58), one gets

$$L^\mu{}_\nu \tau_\mu = A \tau_\nu A^\dagger$$

which leads, using (32), to the explicit relationship of L to A

$$L^\mu{}_\nu = \tfrac{1}{2} \operatorname{Tr} (\rho^\mu A \tau_\nu A^\dagger) \equiv L(A)^\mu{}_\nu. \tag{60}$$

If one can now prove that $L(A)$ is real, has determinant $+1$, and $L(A)^{00} \geq 1$, and that $L(A_1)L(A_2) = L(A_1A_2)$, then it follows that (60) realizes a 2 : 1 homomorphism

$$\pm A \leftrightarrow L(A) \tag{61}$$

of the group, SL(2C), of complex unimodular 2×2 matrices onto the restricted Lorentz group L. That $L(A)$ as given by (60) does indeed satisfy these requirements is now proved.

Using the general matrix results

$$(\operatorname{Tr} M)^* = \operatorname{Tr} M^\dagger, \qquad \operatorname{Tr} (MM') = \operatorname{Tr} (M'M),$$

one easily shows that

$$L(A)^{*\mu}{}_\nu = L(A)^\mu{}_\nu.$$

In order to evaluate $\det L(A)$, one uses the representation (19) to give Eq. (60) in the form

$$2\begin{bmatrix} L^0{}_0 & L^0{}_1 & L^0{}_2 & L^0{}_3 \\ L^1{}_0 & L^0{}_2 & L^0{}_1 & L^1{}_3 \\ L^2{}_0 & L^2{}_1 & L^2{}_2 & L^2{}_3 \\ L^3{}_0 & L^3{}_1 & L^3{}_2 & L^3{}_3 \end{bmatrix} = \begin{bmatrix} -a & -b & -c & -d \\ c & d & a & b \\ -ic & -id & ia & ib \\ a & b & -c & -d \end{bmatrix}$$

$$\times \begin{bmatrix} -a^* & b^* & -ib^* & a^* \\ -b^* & a^* & ia^* & -b^* \\ -c^* & d^* & -id^* & c^* \\ -d^* & c^* & ic^* & -d^* \end{bmatrix} \tag{62}$$

with $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Thede terminants of the matrices on the right can be evaluated easily by Laplace's rule. One thereby obtains

$$16 \det L = [4i(\det A)^2][-4i(\det A^\dagger)^2] = 16,$$

as required.

If one writes

$$A = a^0 \tau^0 - \mathbf{a} \cdot \boldsymbol{\tau} = a^\mu \tau_\mu$$

then the conditions $\det A = \det A^\dagger = 1$ are equivalent to

$$a_\mu a^\mu = a^*{}_\mu a^{*\mu} = 1,$$

so that

$$L(A)^{00} = \tfrac{1}{2} \operatorname{Tr} (AA^\dagger) = a_\mu a^{*\mu} \geq 1,$$

also as required. Finally, one has

$$\begin{aligned} L(A_1)^\mu{}_\nu L(A_2)^\nu{}_\lambda &= \tfrac{1}{2} \operatorname{Tr} (\rho^\mu A_1 \tau_\nu A_1^\dagger) \tfrac{1}{2} \operatorname{Tr} (\rho^\nu A_2 \tau_\lambda A_2^\dagger) \\ &= \tfrac{1}{2} \operatorname{Tr} (\rho^\mu A_1 A_2 \tau_\lambda A_2^\dagger A_1^\dagger) \\ &= L(A_1A_2)^\mu_\lambda, \end{aligned} \tag{63}$$

where Eq. (34) has been used to obtain the second line. One immediate consequence of (63) is

$$L(A)^{-1} = L(A^{-1}). \tag{64}$$

The next part of the discussion concerns the inversion of Eq. (60) to give A in terms of $L(A)$. To obtain the required formula, one starts out from (59)

$$L^\mu{}_\nu \tau_\mu = A \tau_\nu A^\dagger$$

and

$$L_\mu{}^\nu \tau_\nu = A^{-1} \tau_\mu A^{\dagger -1}, \tag{65}$$

obtained from it with the help of (43). One now uses (34) to derive from them the results

$$L^\mu{}_\nu \tau_\mu \rho^\nu = A \tau_\nu A^\dagger \rho^\nu = 2A(\operatorname{Tr} A^\dagger), \tag{66}$$

$$L_\mu{}^\nu \tau_\nu \rho^\mu = A^{-1} \tau_\mu A^{\dagger -1} \rho^\mu = 2A^{-1}(\operatorname{Tr} A^{\dagger -1}). \tag{67}$$

For unimodular A, $\operatorname{Tr} A^\dagger = \operatorname{Tr} A^{\dagger -1}$, so that multiplication of (66) and (67) gives

$$4(\operatorname{Tr} A^\dagger)^2 = L_{\mu\nu} L_{\kappa\lambda} \tau^\mu \rho^\nu \tau^\lambda \rho^\kappa. \tag{68}$$

With the aid of Eq. (31), one proves that the right

side of (68) is indeed a multiple of the unit matrix. Hence, (66) and (68) can be combined to give

$$\pm A = L_{\mu\nu}\tau^{\mu}\rho^{\nu}/[L_{\alpha\beta}L_{\gamma\delta}\tau^{\alpha}\rho^{\beta}\tau^{\delta}\rho^{\gamma}]^{1/2}, \tag{69}$$

$$= \frac{[(\mathrm{Tr}\, L) + (L^{ok} - L^{ko} - i\epsilon^{ijk}L^{ij})\tau^{k}]}{[4 + (\mathrm{Tr}\, L)^2 - \mathrm{Tr}\,(LL) + i\epsilon^{\mu\nu\kappa\lambda}L_{\mu\nu}L_{\kappa\lambda}]^{1/2}}. \tag{70}$$

The formula for A in terms of L has not been given in the concise form (69) before, although Joos[25] has given form (70). Further, the derivation given here is to be contrasted with the laborious derivation of form (70) sketched by Wightman.[2] Finally, verification that Eqs. (69) and (70) do indeed satisfy Eq. (60) is no trivial task (as appears to be implied by Wightman[2]). It is effected as follows. From (67), one obtains

$$L_{\mu\nu}\rho^{\nu}\tau^{\mu} = 2\zeta A^{*-1}\zeta^{-1}(\mathrm{Tr}\, A^{-1}) = 2A^{\dagger}(\mathrm{Tr}\, A), \tag{71}$$

on using the result

$$\zeta\, \tilde{A}\zeta^{-1} = A^{-1}. \tag{72}$$

From (60), one derives with the aid of (34) the result

$$\mathrm{Tr}\, L = (\mathrm{Tr}\, A)(\mathrm{Tr}\, A^{\dagger}). \tag{73}$$

One now inserts (66) and (71) into (60) and uses (73) to give

$$4L^{\mu\lambda}(\mathrm{Tr}\, L) = \tfrac{1}{2}\,\mathrm{Tr}\,[\rho^{\mu}\tau^{\nu}\rho^{\kappa}\tau^{\lambda}\rho^{\alpha}\tau^{\beta}]L_{\nu\kappa}L_{\beta\alpha}. \tag{74}$$

The sixfold can be evaluated without much difficulty from the results of Sec. 3. To see that the imaginary part of the right side of Eq. (74) vanishes, one requires the identity

$$\epsilon^{\mu\alpha\beta\gamma}L_{\gamma}{}^{\nu}L_{\beta\alpha} = \epsilon^{\nu\alpha\beta\gamma}L^{\mu}{}_{\gamma}L_{\beta\alpha},$$

which follows, for proper Lorentz transformations, from Eq. (52). Then the right side can be simplified, by use of identities (54), to $4L^{\mu\lambda}$ (Tr L), completing the consistency proof.

For the work of Sec. 8, one notes that Eq. (65)

$$L_{\mu}{}^{\nu}\tau_{\nu} = A^{-1}\tau_{\mu}A^{\dagger-1}$$

can be converted into

$$L_{\mu}{}^{\nu}\rho_{\nu} = \zeta A^{*-1}\zeta^{-1}\rho_{\mu}\zeta\tilde{A}^{-1}\zeta^{-1} = A^{\dagger}\rho_{\mu}A, \tag{75}$$

with the aid of (72).

6. ROTATIONS AND PURE LORENTZ TRANSFORMATIONS

Two important subsets of the restricted Lorentz transformations are the spatial rotations and the pure Lorentz transformations. Spatial rotations are considered first.

A spatial rotation of angle θ in a positive sense[26] about the unit spatial vector $\mathbf{n}$ is given by

$$\begin{aligned} &x^{\mu} \to x'^{\mu},\\ &x'^{0} = x^{0},\\ &\mathbf{x}' = \mathbf{x}\cos\theta + \mathbf{x}\cdot\mathbf{n}\mathbf{n}(1 - \cos\theta) + \mathbf{n}\wedge\mathbf{x}\sin\theta. \end{aligned} \tag{76}$$

One may also write this as

$$x^{\mu} \to x'^{\mu} = L^{\mu}{}_{\nu}x^{\nu}$$

with

$$\begin{aligned} &L^{oo} = 1, \qquad L^{ok} = L^{ko} = 0,\\ &-L^{jk} = \cos\theta\,\delta^{jk} + (1 - \cos\theta)n^{j}n^{k} - \sin\theta\epsilon^{jkl}n^{l}, \end{aligned} \tag{77}$$

or else simply as

$$x^{j} \to x'^{j} = R^{jk}x^{k}$$

with[27]

$$R^{jk} = -L^{jk}. \tag{78}$$

Introducing the matrices T^{j} which satisfy the commutation relations

$$[T^{j}, T^{k}] = i\epsilon^{jkl}T^{l}, \tag{79}$$

and

$$T^{j}T^{k}T^{l} + T^{l}T^{k}T^{j} = \delta^{jk}T^{l} + \delta^{kl}T^{j}, \tag{80}$$

and, which can be represented by

$$(T^{j})^{ik} = i\epsilon^{ijk}, \tag{81}$$

one can write the matrix R as given by (78) with (77) in the form

$$R = \exp\,[i\theta(\mathbf{n}\cdot\mathbf{T})]. \tag{82}$$

To verify this statement, one must note that, as a consequence of (80), one has

$$(\mathbf{n}\cdot\mathbf{T})^{3} = (\mathbf{n}\cdot\mathbf{T}) \tag{83}$$

and use identity (8) to simplify the term proportional to $(\mathbf{n}\cdot\mathbf{T})^{2}$

If an element L_R of L describes a spatial rotation, it must commute with I, as given by (41) and hence, from (40) satisfy

$$\tilde{L}_R L_R = 1. \tag{84}$$

One can show, using (10), that the matrix L given

[25] H. Joos, "Bemerkungen zur Phase-Shift Analysis auf Grunde der Darstellungstheorie der inhomogenen Lorentz gruppe," Oberwolfach 1959 (unpublished).

[26] If i, j, and k are unit vectors in the (positive) directions of the right-handed set of spatial coordinate axes x^1, x^2, and x^3, a rotation of amount θ in a positive sense about k must carry i over to (i cos θ + j sin θ).

[27] In agreement with the concluding remarks of Sec. 2.

by (77) does, indeed, satisfy (84). Conversely, if an element of L satisfies (84), it describes a spatial rotation.

One can see from the discussion of Sec. 5, that the elements of SL(2C), which correspond to spatial rotations are unitary, and conversely. Indeed the element of A of SL(2C), which corresponds to the spatial rotation (77) is given by (70) in the form

$$A = \cos \tfrac{1}{2}\theta - i \sin \tfrac{1}{2}\theta \tau \cdot \mathbf{n} \equiv A(\mathbf{n}, \theta), \tag{85}$$

which is evidently unitary. It is of course very well known that the spatial rotations form a group, R_3, a subgroup of L, to which the unitary subgroup, SU(2C), of SL(2C) is 2 : 1 homomorphic. Formulas, which describe the homomorphism, are contained within the work of Sec. 5. Nevertheless, it is convenient to spell out in detail a separate derivation of them, using an approach similar in spirit to that already described. What is thereby obtained is a more compact derivation of the essential results of the quaternion theory[8,9] of rotations than has hitherto been presented. A more compelling reason for its inclusion here is that it can be taken over, in the next section, with little modification into the theory of orthogonal transformations (complex rotations) in a space of three complex dimensions.

A spatial rotation is a real linear transformation in three-dimensional Euclidean space of the type

$$y^i \to y'^i = R^{ij} y^j, \tag{86}$$

which leaves invariant $y^2 = y^k y^k$. The transformation coefficients thus satisfy

$$R^{ij} R^{ik} = \delta^{jk}, \tag{87}$$

i.e., the matrix R satisfies

$$\tilde{R} R = 1, \tag{88}$$

from which it follows that $\det R = \pm 1$. Matrices R with $\det R = +1$ (-1) describe proper (improper) rotations. Improper rotations are not of interest here and the term spatial rotation is used without qualification (as in the early part of the section) for the $\det R = +1$ case. For this case, one gets, using (13) and (88),

$$\epsilon^{jkl} R^{lr} = \epsilon^{pqr} R^{pj} R^{qk}. \tag{89}$$

Operating on Eq. (89) with ϵ^{jmr} and using Eq. (8) to simplify, leads to

$$(\operatorname{Tr} R)\, \delta^{mk} - R^{mk} = (\operatorname{Tr} R) R^{km} - R^{ks} R^{sm} \tag{90}$$

and, hence by taking traces, to

$$2 \operatorname{Tr} R = (\operatorname{Tr} R)^2 - \operatorname{Tr} (RR). \tag{91}$$

Now, one associates with each point y^i of ordinary space a matrix Y

$$Y = y^i \tau^i \tag{92}$$

with inverse relation

$$y^i = \tfrac{1}{2} \operatorname{Tr} (\tau^i Y), \tag{93}$$

and considers the transformation

$$Y \to Y' = A Y A^{-1}, \tag{94}$$

with A unitary and unimodular. Equation (94) leads to

$$R^{ik} \tau^i = A \tau^k A^{-1} \tag{95}$$

and, hence, to

$$R^{ij} = \tfrac{1}{2} \operatorname{Tr} (\tau^i A \tau^j A^{-1}) \equiv R(A)^{ij}. \tag{96}$$

One can easily prove that $R(A)$ is real, has determinant $+1$, and that

$$R(A_1) R(A_2) = R(A_1 A_2) \tag{97}$$

so that (96) explicitly gives the 2 : 1 homomorphism

$$\pm A \leftrightarrow R(A) \tag{98}$$

of SU(2C) onto R_3. From (95), one obtains, using Eq. (35),

$$\begin{aligned} R^{ij} \tau^i \tau^j &= A \tau^j A^{-1} \tau^j \\ &= 2A(\operatorname{Tr} A^{-1}) - A A^{-1}. \end{aligned} \tag{99}$$

Now, a unitary unimodular matrix A can be written in the form

$$A = a^0 + i a^k \tau^k \tag{100}$$

with a^0, a^k real, so that

$$\operatorname{Tr} A = 2a^0 = \operatorname{Tr} A^{-1}$$

follows. Hence (96) with the help of (35) gives

$$\begin{aligned} \operatorname{Tr} R &= (\operatorname{Tr} A)(\operatorname{Tr} A^{-1}) - \tfrac{1}{2} \operatorname{Tr} (A A^{-1}) \\ &= (\operatorname{Tr} A)^2 - 1. \end{aligned} \tag{101}$$

The inversion of (96) now follows from Eqs. (99) and (101) in the form

$$\pm A = \frac{1 + R^{ij} \tau^i \tau^j}{[4(1 + \operatorname{Tr} R)]^{1/2}}. \tag{102}$$

The check that A, as given by (102), satisfies (96) depends on the use of identities (90) and (91).

The subject of pure Lorentz transformations is taken up next. A pure Lorentz transformation of velocity v in the positive direction of the unit spatial vector $\mathbf{n}$ can be given in the form

$$x^\mu \to x'^\mu$$
$$x'^0 = x^0 \cosh\chi + \mathbf{x}\cdot\mathbf{n} \sinh\chi \tag{103}$$
$$\mathbf{x}' = \mathbf{x} - \mathbf{x}\cdot\mathbf{nn}(1 - \cosh\chi) + x^0\mathbf{n} \sinh\chi,$$

with $\cosh\chi = 1/\beta$, $\sinh\chi = v/\beta$, $\tanh\chi = v$, $\beta = (1 - v^2)^{1/2}$. This may be put in the form

$$x^\mu \to x'^\mu = L^\mu{}_\nu x^\nu$$

with

$$L^{00} = \cosh\chi, \qquad L^{k0} = -L^{0k} = n^k \sinh\chi, \tag{104}$$
$$L^{ij} = g^{ij} + n^{ij}(1 - \cosh\chi).$$

From (104) one sees that a matrix L_p of L, which describes a pure Lorentz transformation, satisfies $\tilde{L}_p = L$. The converse is not so, however. The following theorem is proved in Tolhoek's[2] lectures: A matrix L_p of the restricted Lorentz group L, which satisfies $\tilde{L}_p = L_p$, describes either a pure Lorentz transformation or else a pure Lorentz transformation followed by a rotation of angle π about the direction of the pure Lorentz transformation. The lack of a clearcut criterion for recognizing a pure Lorentz transformation can be irksome in practical work on Lorentz transformations. From Eq. (70), one finds that the matrix A which corresponds to the pure Lorentz transformation (104) is given by

$$A = \cosh\tfrac{1}{2}\chi - \sinh\tfrac{1}{2}\chi\boldsymbol{\tau}\cdot\mathbf{n} \equiv A(\chi, \mathbf{n}). \tag{105}$$

The notation puts the "amount" χ before the direction $\mathbf{n}$ to distinguish a pure Lorentz transformation from a rotation [cf., Eq. (85)]. Matrices of the form (105) are Hermitian. From (85), one obtains $A(\mathbf{n}, \pi)$, which describes a rotation of angle π about the direction $\mathbf{n}$ of the pure Lorentz transformation (104), in the form

$$A(\mathbf{n}, \pi) = -i\boldsymbol{\tau}\cdot\mathbf{n}. \tag{106}$$

Since $A(\chi, \mathbf{n})A(\mathbf{n}, \pi)$ is not Hermitian, one concludes that A is Hermitian if and only if it corresponds to a pure Lorentz transformation—a result which makes the unimodular matrix description of Lorentz transformations useful practically.

The above discussion of spatial rotations and pure accelerations allows a complete characterization of the elements of L. This is because the general element of L can be expressed uniquely in the form

$$L = L_R L_p \tag{107}$$

with L_R, L_p, respectively, describing a spatial rotation and a pure Lorentz transformation. For a proof of this statement the reader is referred to the book by Fock.[2] The corresponding statement for the general element A of SL(2C) is that it can be uniquely expressed as the product

$$A = UH \tag{108}$$

of a unitary matrix and a positive[28] Hermitian one. To prove this, one notes that the matrix $A^\dagger A$ is positive Hermitian, and sets $A^\dagger A = A(2\chi, \mathbf{n})$. If one now takes

$$H = A(\chi, \mathbf{n}), \tag{109}$$

then the fact that

$$U = AA(-\chi, \mathbf{n}) \tag{110}$$

satisfies

$$U^\dagger U = A(-\chi, \mathbf{n})A(2\chi, \mathbf{n})A(-\chi, \mathbf{n}) = 1 \tag{111}$$

completes the proof. If $L_R = L(U)$ and $L_p = L(H)$, then it follows that $L = L_R L_p = L(U)L(H) = L(UH) = L(A)$.

Some comment on the fact that positive H is required in the statement (108) has to be made. It corresponds to the indicated bracketing in the following

$$L = L_R L_p \leftrightarrow \pm A = (\pm U)H \tag{112}$$

which is allowed because of the two valuedness of the homomorphisms of SL(2C) to L and SU(2C) to R_3. The fact that one has arbitrariness of sign for spatial rotations and restricted Lorentz transformations reflects the fact that these constitute doubly-connected topological groups,[29] with the groups SU(2C) and SL(2C) isomorphic to their covering groups. Likewise the nonarbitrariness of sign in the case of pure Lorentz transformations reflects the fact that these transformations are a simply connected set.

In order to exhibit the practical utility of the unimodular matrix description of Lorentz transformations, the case of the product of two nonparallel pure Lorentz transformations is to be examined. Thus, it is sought to write

$$A = A(\chi_1, \mathbf{n}_1)A(\chi_2, \mathbf{n}_2) \tag{113}$$

in the form[30]

$$A = A(\mathbf{n}, \theta)A(-\chi_3, \mathbf{n}_3). \tag{114}$$

[28] Positive here means "with positive eigenvalues." Thus, $A(\chi, \mathbf{n})$ as given by (105) is positive Hermitian, while $-A(\chi, \mathbf{n})$ is not.

[29] For discussion of the topological properties of R_3 and L, see the lectures of Tolhoek.[2] For the concept of covering group, see also the book of Pontrjagin [L. S. Pontrjagin, *Topologische Gruppe* (B. G. Teubner, Leipzig, 1958)].

[30] The minus sign is present to allow subsequent formulas to be presented in a symmetric form.

The fact that $A(\mathbf{n}, \theta)$ turns out, below, not to be the unit matrix gives a further indication that the set of all pure Lorentz transformations do not possess the group property. To find χ_3 and $\mathbf{n}_3$, one calculates $A^\dagger A = A(-2\chi_3, \mathbf{n}_3)$, easily obtaining

$$\cosh \chi_3 = \cosh \chi_1 \cosh \chi_2 + \sinh \chi_1 \sinh \chi_2 \mathbf{n}_1 \cdot \mathbf{n}_2, \tag{115}$$

and

$$-\sinh \chi_3 \mathbf{n}_3 = \sinh \chi_1 \mathbf{n}_1 + \sinh \chi_2 \cosh \chi_1 \mathbf{n}_2 + \sinh \chi_1 (\cosh \chi_2 - 1) \mathbf{n}_1 \cdot \mathbf{n}_2 \mathbf{n}_2. \tag{116}$$

Introducing $\mathbf{v}_i = v_i \mathbf{n}_i$,[31] $\beta_i = (1 - v_i^2)^{1/2}$, $v_i = \tanh \chi_i$, one can put (115) and (116) into the form

$$\beta_1 \beta_2 = \beta_3 (1 + \mathbf{v}_1 \cdot \mathbf{v}_2) \tag{117}$$

$$-\mathbf{v}_3 (1 + \mathbf{v}_1 \cdot \mathbf{v}_2) = \mathbf{v}_1 + \mathbf{v}_2 + [(1 - \beta_2)/v_2^2] \mathbf{v}_2 \wedge (\mathbf{v}_2 \wedge \mathbf{v}_1). \tag{118}$$

Equations (116) and (117) agree with Eqs. (16.11) and (16.07) of Fock's book.[2] These equations however are derived differently, and his Eq. (16.07) is not in so neat a form as (118). In the case when $\mathbf{v}_1$ and $\mathbf{v}_2$ are parallel, Eq. (118) reduces to the well-known Einstein addition theorem for parallel velocities

$$-\mathbf{v}_3 = (\mathbf{v}_1 + \mathbf{v}_2)/(1 + \mathbf{v}_1 \cdot \mathbf{v}_2). \tag{119}$$

From (118) and (117), one easily finds

$$\beta_2 \beta_3 = \beta_1 (1 + \mathbf{v}_2 \cdot \mathbf{v}_3), \tag{120}$$

$$\beta_3 \beta_1 = \beta_2 (1 + \mathbf{v}_3 \cdot \mathbf{v}_1). \tag{121}$$

Introducing $\mathbf{n}_1 \cdot \mathbf{n}_2 = \cos \phi_3$, etc., one can write (117), (120), and (121) in the form

$$\frac{\sinh \chi_1}{\sin \phi_1} = \frac{\sinh \chi_2}{\sin \phi_2} = \frac{\sinh \chi_3}{\sin \phi_3}. \tag{122}$$

The resemblance of (115), beside which one can place the analogous pair of formulas with 1, 2, 3 permuted cyclically, and (122) to the cosine and sine rule for a spherical triangle is very marked. Various other formulas may be proved in analogy to the usual results of spherical trigonometry,[32] e.g.,

$$\cos \phi_3 = \cos \phi_1 \cos \phi_2 - \sin \phi_1 \sin \phi_2 \cosh \chi_3. \tag{123}$$

For further remarks on the subject and references to some early papers by Sommerfeld, the reader is referred to a paper by Wick.[33]

Finally, one calculates $A(\mathbf{n}, \theta)$ finding that $\mathbf{n}$ is parallel to $\mathbf{n}_1 \wedge \mathbf{n}_2$ and that θ is given by

$$\cos \frac{\theta}{2} = \frac{1 + \cosh \chi_1 + \cosh \chi_2 + \cosh \chi_3}{4 (\cosh \chi_1/2)(\cosh \chi_2/2)(\cosh \chi_3/2)}. \tag{124}$$

7. THE GROUP O_3

Orthogonal transformations in a complex space of three dimensions are often referred to as complex rotations. Their relationship to the transformations of the real restricted Lorentz group can be exhibited in the following manner.

From a pair of orthogonal four-vector p^μ and q^μ one builds the quantity

$$PQ^{-1}, \tag{125}$$

where

$$P = p_\mu \tau^\mu, \qquad Q = q_\mu \tau^\mu. \tag{126}$$

Under Lorentz transformation $L = L(A)$, one knows that

$$P \to P' = APA^\dagger$$

$$Q \to Q' = AQA^\dagger$$

or

$$Q^{-1} \to Q'^{-1} = A^{\dagger -1} Q^{-1} A^{-1},$$

so that

$$PQ^{-1} \to P'Q'^{-1} = APQ^{-1}A^{-1}. \tag{127}$$

If one writes PQ^{-1} in the form

$$PQ^{-1} = \tau^k z^k = \boldsymbol{\tau} \cdot \mathbf{z}, \tag{128}$$

with

$$\mathbf{z} = p^0 \mathbf{q} - q^0 \mathbf{p} - i \mathbf{p} \wedge \mathbf{q}, \tag{129}$$

and[34]

$$\mathbf{z}^2 = z^k z^k = -p^2 q^2, \tag{130}$$

then (127) reads as

$$\boldsymbol{\tau} \cdot \mathbf{z} \to \boldsymbol{\tau} \cdot \mathbf{z}' = A \boldsymbol{\tau} \cdot \mathbf{z} A^{-1}. \tag{131}$$

Equation (131) can be seen to describe transformations of the complex three-vector $\mathbf{z}$ which preserve its norm (130), i.e., rotations in a complex space of three dimensions. Setting

$$z'^i = \mathcal{R}^{ik} z^k, \tag{132}$$

one is able to take over the work of Sec. 6 on real rotations even though now $\mathcal{R}$ is not real and A is not unitary. In particular

[31] No summation.

[32] I. Todhunter, G. Leathem, *Spherical Trigonometry* (MacMillan and Company, Inc., London, 1901).

[33] G. C. Wick, Ann. Phys. (New York) **18**, 65 (1962).

[34] Since p and q are orthogonal, either both are space-like in which case $\mathbf{z}^2$ is negative, or else one is time-like and the other space-like, in which case $\mathbf{z}^2$ is positive.

$$\mathfrak{R}^{ij} = \tfrac{1}{2}\,\mathrm{Tr}\,(\tau^i A \tau^j A^{-1}) \equiv \mathfrak{R}(A)^{ij} \tag{133}$$

and

$$\pm A = \frac{1 + \tau^i \tau^j \mathfrak{R}^{ij}}{[4(1 + \mathrm{Tr}\,\mathfrak{R})]^{1/2}}. \tag{134}$$

The relationship of the group O_3 to L can be deduced directly. Equation (133) shows that there is a 2 : 1 homomorphism $\pm A \leftrightarrow \mathfrak{R}(A)$ of SL(2C) onto O_3, so that, from the work of Sec. 5, it follows that O_3 is simply isomorphic to L. Explicit formulas for this isomorphism arise from the insertion of (69) or (70) into (133).

When A is unitary and hence corresponds to a spatial rotation in Minkowski space, $\mathfrak{R}$ evidently reduces to the real matrix R describing this rotation. The real and imaginary parts of $\mathbf{z}$ transform separately as real three-vectors. For A Hermitian, $\mathfrak{R}$ will not be real, so that the real and imaginary parts of $\mathbf{z}$ are mixed by a pure Lorentz transformation in Minkowski space. Inserting

$$A = A(\chi, \mathbf{n})$$

into (133) gives, with the aid of Eqs. (22) to (24), the result

$$\mathfrak{R}^{ij} = \delta^{ij} \cosh\chi + n^i n^j (1 - \cosh\chi) + i\epsilon^{ijl} n^l \sinh\chi. \tag{135}$$

In terms of the matrices T^i of Eqs. (79)–(81), one can write this as

$$\mathfrak{R} = \exp\,[-\chi(\mathbf{n}\cdot\mathbf{T})]. \tag{136}$$

To obtain the relationship of the general element $\mathfrak{R}$ of O_3 to the general element of L, one writes the latter in the form

$$L = L_R L_p$$

with L_R and L_p, respectively, describing a spatial rotation θ about $\mathbf{n}$, and a pure Lorentz transformation χ along $\mathbf{n}'$. Then one gets

$$\mathfrak{R}[A(\mathbf{n}, \theta)A(\chi, \mathbf{n}')] = \mathfrak{R}[A(\mathbf{n}, \theta)]\mathfrak{R}[A(\chi, \mathbf{n}')]$$

$$= \mathfrak{R}[A(\mathbf{n}, \theta)] \exp\,[-\chi(\mathbf{n}'\cdot\mathbf{T})] \tag{137}$$

$$= \exp\,[i\theta(\mathbf{n}\cdot\mathbf{T})] \exp\,[-\chi(\mathbf{n}'\cdot\mathbf{T})]. \tag{138}$$

The form (137) of this result corresponds in the case of O_3 to the results $L = L_R L_p$ and $A = UH$ previously noted for L and SL(2C). As in the case of SL(2C), it illustrates a general matrix theorem, one which Gantmacher[15] states as follows. A complex orthogonal matrix $\mathfrak{R}$ can be uniquely resolved into the form

$$\mathfrak{R} = Re^K, \tag{139}$$

where R is real orthogonal and K Hermitian pure imaginary. That $K = -\chi(\mathbf{n}'\cdot\mathbf{T})$ is Hermitian pure imaginary in the representation (81) of $\mathbf{T}$ is easily seen to be the case.

As an illustration of the formalism just developed, one may derive the transformation properties with respect to L of the electromagnetic field-vectors, $\mathbf{E}$ and $\mathbf{H}$. Those vectors are related to a four-vector potential A^μ by the equations

$$\mathbf{E} = \boldsymbol{\partial} A^\circ - \partial^\circ \mathbf{A} \tag{140}$$

$$\mathbf{H} = \boldsymbol{\partial} \wedge \mathbf{A} \tag{141}$$

with

$$\partial^\mu = (\partial^\circ, \boldsymbol{\partial}) = \partial/\partial x_\mu.$$

Since the condition

$$\partial_\mu A^\mu = 0 \tag{142}$$

is imposed on A^μ, one can see that ∂^μ and A^μ play the role of p^μ and q^μ in the general discussion, and that

$$\mathbf{z} = -(\mathbf{E} + i\mathbf{H}). \tag{143}$$

Thus, it follows, from the fact that the (complex) norm of $\mathbf{z}$ is preserved by complex rotations, that the quantities

$$\mathbf{E}^2 - \mathbf{H}^2, \qquad \mathbf{E}\cdot\mathbf{H} \tag{144}$$

are scalar with respect to L. With $\mathbf{z}$ as given by (144), $\mathfrak{R}$ as given by (135), one finds from the real and imaginary parts of Eq. (132), that $\mathbf{E}, \mathbf{H} \rightarrow \mathbf{E}', \mathbf{H}'$ under pure Lorentz transformation $v = \tanh\chi$ along $\mathbf{n}$ with

$$\mathbf{E}' = (1/\beta)[\mathbf{E} + \mathbf{v} \wedge \mathbf{H} - (\mathbf{E}\cdot\mathbf{v}\mathbf{v}/v^2)] + \mathbf{E}\cdot\mathbf{v}\mathbf{v}/v^2, \tag{145}$$

$$\mathbf{H}' = (1/\beta)[\mathbf{H} - \mathbf{v} \wedge \mathbf{E} - (\mathbf{H}\cdot\mathbf{v}\mathbf{v}/v^2)] + \mathbf{H}\cdot\mathbf{v}\mathbf{v}/v^2, \tag{146}$$

and $\mathbf{v} = v\mathbf{n}$. These results agree with those obtained in Fock's book[2] [Eqs. (24.37), (24.38)] by rather different methods.

The relationship of O_3 to L is next investigated by viewing these groups as subgroups of the complex Lorentz group. The discussion is made in terms of a reformulation, using the methods of Secs. 3 and 5, of the work of Wightman[2] on the 2×2 unimodular matrix description of the latter group. A complex Lorentz transformation is a linear transformation of the type

$$z^\mu \rightarrow z'^\mu = \mathfrak{L}^\mu{}_\nu z^\nu \tag{147}$$

of the points z^μ of a complex space with metric (2),

which leaves invariant the quadratic form

$$g^{\mu\nu}z_\mu z_\nu.$$

The related matrices $\mathfrak{L}$ satisfy

$$\tilde{\mathfrak{L}} I_s \mathfrak{L} = I_s \tag{148}$$

with I_s as before, Eq. (41). The values ± 1 are both allowed for $\det \mathfrak{L}$, but only the case $+1$ is of interest here.

With each point z^μ, one associates a matrix

$$Z = z_\mu \tau^\mu, \tag{149}$$

and considers transformations of the form

$$Z \to Z' = AZ\tilde{B} \tag{150}$$

with A and B independent unimodular matrices. They preserve $g^{\mu\nu}z_\mu z_\nu$ and hence provide a mapping of the complex transformations (147). Equation (150) gives

$$\mathfrak{L}^{\mu\nu}\tau_\mu = A\tau^\nu\tilde{B} \tag{151}$$

and hence

$$\mathfrak{L}^{\mu\nu} = \tfrac{1}{2}\,\mathrm{Tr}\,(\rho^\mu A \tau^\nu \tilde{B}) = \mathfrak{L}(A, B)^{\mu\nu}. \tag{152}$$

As in Sec. 5, one proves that $\det \mathfrak{L} = +1$ and that

$$\mathfrak{L}(A_1, B_1)\mathfrak{L}(A_2, B_2) = \mathfrak{L}(A_1A_2, B_1B_2).$$

To express A and B in terms of $\mathfrak{L}$, one uses Eq. (151) and the result

$$\mathfrak{L}^{\mu\nu}\tau_\nu = A^{-1}\tau^\mu\tilde{B}^{-1} \tag{153}$$

which follows from it, using $(\mathfrak{L}^{-1})^{\mu\nu} = \mathfrak{L}^{\nu\mu}$. With the help of Eq. (34), Eqs. (151) and (155) lead to

$$\mathfrak{L}_{\mu\nu}\tau^\mu\rho^\nu = 2A(\mathrm{Tr}\,\tilde{B}), \tag{154}$$

$$\mathfrak{L}_{\mu\nu}\rho^\nu\tau^\mu = 2\tilde{B}(\mathrm{Tr}\,A), \tag{155}$$

$$\mathfrak{L}_{\mu\nu}\tau^\nu\rho^\mu = 2A^{-1}(\mathrm{Tr}\,\tilde{B}), \tag{156}$$

$$\mathfrak{L}_{\mu\nu}\rho^\mu\tau^\nu = 2\tilde{B}^{-1}(\mathrm{Tr}\,A), \tag{157}$$

the results $\mathrm{Tr}\,\tilde{B}^{-1} = \mathrm{Tr}\,\tilde{B}$, $\mathrm{Tr}\,A^{-1} = \mathrm{Tr}\,A$, having been used in (156) and (157), respectively. Combining these equations in pairs gives the desired results

$$\pm A = \mathfrak{L}_{\mu\nu}\tau^\mu\rho^\nu/[\mathfrak{L}_{\alpha\beta}\mathfrak{L}_{\gamma\delta}\tau^\alpha\rho^\beta\tau^\delta\rho^\gamma]^{1/2} \equiv \pm A(\mathfrak{L}), \tag{158}$$

$$\pm\tilde{B} = \mathfrak{L}_{\mu\nu}\rho^\nu\tau^\mu/[\mathfrak{L}_{\alpha\beta}\mathfrak{L}_{\gamma\delta}\rho^\alpha\tau^\beta\rho^\delta\tau^\gamma]^{1/2} \equiv \pm\tilde{B}(\mathfrak{L}). \tag{159}$$

Equations (158) and (159) bear the same relationship to the results given by Wightman[2] as Eq. (69) does to Eq. (70), above.

One notes that

$$A(\mathfrak{L})^\dagger = \tilde{B}(\mathfrak{L}^*) \tag{160}$$

follows from (158) and (159), so that real $\mathfrak{L}$ implies $A^\dagger = \tilde{B}$ as required to reproduce the formalism of Sec. 5. If one sets $\mathfrak{L}^{oo}=1$, $\mathfrak{L}^{ok}=\mathfrak{L}^{ko}=0$, $\mathfrak{L}^{ij}=-\mathfrak{R}^{ij}$, and $\tilde{B} = A^{-1}$, one can likewise reproduce the formalism for O_3 developed earlier in this section. One observes how the different restrictions, necessary to give the (isomorphic) unimodular description of L and O_3, operate

$$\tilde{B} \to A^\dagger, \qquad \tilde{B} \to A^{-1}$$

and become equivalent in the subsequent restriction to the group R_3 of spatial rotations, which are described by unitary matrices A.

The reader is referred to the lectures of Wightman[2] for the use to which he puts the theory of complex Lorentz transformations, described above.

8. TWO- AND FOUR-COMPONENT SPINORS

The formalism of Secs. 3 and 5 is intimately connected with the calculus of two-component spinors as devised by van der Waarden[35,3] and systematically discussed in references 18 and 19. It is the view of the present author that this theory can be adequately presented within the ordinary methods of matrix algebra and that one thereby avoids, or rather subordinates, the algebraic complications associated with spinor indices. Here two-component spinors are introduced and the passage to four-component spinors, which makes possible the representation of inversions, is effected. This serves to relate present work to the description of Lorentz transformations in terms of Dirac matrices as given in text books[16,17] on field theory.

Let ϕ be a two-component (column)[36] spinor, which under $L = L(A)$ transforms according to

$$\phi \to \phi' = A\phi. \tag{161}$$

If one wrote ϕ as ϕ_x, say, with x a spinor index, then one would write ϕ^*, χ, χ^* as $\phi_{\dot{x}}$, ϕ^x, $\phi^{\dot{x}}$, with

$$\chi = \zeta\phi, \qquad \phi = -\zeta\chi \tag{162}$$

since ζ is exactly the matrix usually used as metric spinor to raise and lower spinor indices. Under $L(A)$, the spinors ϕ^*, χ, and χ^* transform according to

$$\phi^* \to \phi'^* = A^*\phi^*, \tag{163}$$

$$\chi \to \chi' = \tilde{A}^{-1}\chi, \tag{164}$$

$$\chi^* \to \chi'^* = A^{\dagger-1}\chi^*. \tag{165}$$

[35] B. L. van der Waerden, Nachr. kgl. Ges. Wiss. Göttingen 100, (1929).

[36] It will always be evident from context when a spinor is to be regarded as a row of a column matrix.

Since A is not unitary, it is evident, from (161) and (165), that the spinors ϕ and χ^* transform according to inequivalent (irreducible) representations of the restricted Lorentz group. It is further true that one cannot describe a representation of the full Lorentz group in terms of either a spinor transforming like ϕ or a spinor transforming like χ^*. Since the elements of the subsets II to IV of Lorentz transformations can be generated in the form (50) from the elements $L(A)$ of the restricted group, one need only, to obtain a representation of the full group, adjoin (in somewhat *ad hoc* manner) the following transformation laws to the laws (161), (163)–(165). Under space inversion I_s

$$\phi \rightarrow \pm i\chi^*, \qquad \chi^* \rightarrow \pm i\phi, \tag{166}$$

under time inversion I_t

$$\phi \rightarrow \pm\chi^*, \qquad \chi^* \rightarrow \mp\phi, \tag{167}$$

and under the combined inversion $I_{st} = I_s I_t$

$$\phi \rightarrow \pm i\phi, \qquad \chi^* \rightarrow \mp i\chi^*. \tag{168}$$

Two remarks are necessary in conjunction with (166)–(168). Firstly, the presence of the factor i in (166), (168) is not an essential ingredient of the transformation law: A perfectly satisfactory representation of the inversions is obtained without it. It is present simply to allow agreement with the work of Schweber.[16] Secondly, the signs within each of the Eqs. (166)–(168) are coupled, but one may associate upper and lower lines of (166) with upper and lower lines of (167) and (168) in eight different ways, thus obtaining eight nonequivalent representations of the inversions. This is associated with the fact that the full Lorentz group has eight distinct covering groups. The reader is referred to the papers by Shirokov[37] and to the lectures by Wightman[2] for a discussion of the topic. In what follows, for simplicity, let attention be confined to the case of upper signs throughout Eqs. (166)–(168).

Next, the formation of various tensorial quantities from two-component spinors is briefly considered. Under general element L of the full Lorentz group, scalar, pseudoscalar, vector, and axial vector are defined according to

$$S \rightarrow S' = S \tag{169}$$

$$P \rightarrow P' = \epsilon(L^{oo})\epsilon(\det L)P \tag{170}$$

$$V^\mu \rightarrow V'^\mu = L^\mu{}_\nu V^\nu \tag{171}$$

$$A^\mu \rightarrow A'^\mu = \epsilon(L^{oo})\epsilon(\det L)L^\mu{}_\nu A^\nu, \tag{172}$$

respectively, where $\epsilon(x) = +1(-1)$ for $x > 0$ (< 0). In particular, one notes that under I_s

$$\begin{aligned} P &\rightarrow P' = -P \\ A^\mu &\rightarrow A'^\mu = (-A^o, A^k), \end{aligned} \tag{173}$$

under I_t

$$\begin{aligned} P &\rightarrow P' = P \\ A^\mu &\rightarrow A'^\mu = (-A^o, A^k), \end{aligned} \tag{174}$$

and under I_{st}

$$\begin{aligned} P &\rightarrow P' = -P \\ A^\mu &\rightarrow A'^\mu = A^\mu. \end{aligned} \tag{175}$$

If ϕ, ϕ' are spinors with law (161) of transformation under the restricted Lorentz transformation $L(A)$, and χ, χ' are spinors related to them by (162), one may readily verify that the quantities

$$\phi^*\chi'^* \pm \chi\phi' \tag{176}$$

are, respectively, scalar and pseudoscalar in the sense of (169) and (170). Likewise the quantities

$$\phi^*\rho^\mu\phi' \pm \chi\tau^\mu\chi'^* \tag{177}$$

are the components, respectively, of a vector and an axial vector in the sense of Eqs. (171) and (172). Under the element $L = L(A)$ of the restricted group, one sees that

$$\begin{aligned} \phi^*\rho^\mu\phi' \rightarrow \phi^*A^\dagger\rho^\mu A\phi' &= L^\mu{}_\nu\phi^*\rho^\nu\phi', \\ \chi\tau^\mu\chi'^* \rightarrow \chi A^{-1}\tau^\mu A^{\dagger-1}\chi'^* &= L^\mu{}_\nu\chi\tau^\nu\chi'^*, \end{aligned} \tag{178}$$

where Eqs. (75) and (65) have been employed. Under the element $I_s L = I_s L(A)$ of subset II, one finds that

$$\begin{aligned} \phi \underset{L}{\rightarrow} A\phi \underset{I_s}{\rightarrow} i\zeta(A\phi)^* &= iA^{\dagger-1}\chi^*, \\ \chi \underset{L}{\rightarrow} A^{\dagger-1}\chi^* \underset{I_s}{\rightarrow} -i\zeta(A^{\dagger-1}\chi^*)^* &= iA\phi, \end{aligned} \tag{179}$$

and uses

$$\begin{aligned} [I_s L(A)]^\mu{}_\nu\rho^\nu &= I_s{}^\mu{}_\lambda A^\dagger\rho^\lambda A = A^\dagger\tau^\mu A, \\ [I_s L(A)]^\mu{}_\nu\tau^\nu &= I_s{}^\mu{}_\lambda A^{-1}\tau^\lambda A^{\dagger-1} = A^{-1}\rho^\mu A^{\dagger-1}, \end{aligned}$$

to show that

$$\begin{aligned} \phi^*\rho^\mu\phi' \rightarrow \chi A^{-1}\rho^\mu A^{\dagger-1}\chi'^* &= (I_s L)^\mu{}_\nu\chi\tau^\nu\chi'^*, \\ \chi\tau^\mu\chi'^* \rightarrow \phi^*A^\dagger\tau^\mu A\phi' &= (I_s L)^\mu{}_\nu\phi^*\rho^\nu\phi'. \end{aligned} \tag{180}$$

Equations (178), (180), and similar Eqs. for elements of subsets III and IV confirm the statement made regarding the quantities (177).

It is immediate to pass from the above work to

[37] Iu. M. Shirokov, Nuclear Phys. **15**, 1 and 13 (1960).

the four-component formalism. In agreement with Corson[18] and Shirokov[38] one forms a four-component spinor Ψ according to

$$\Psi = \begin{bmatrix} \phi \\ \chi^* \end{bmatrix}. \tag{181}$$

Under restricted Lorentz transformation $L(A)$, Ψ transforms according to

$$\Psi \to \Psi' = S(A)\Psi, \tag{182}$$

with $S(A)$, from (161) and (165), given by

$$S(A) = \begin{bmatrix} A & 0 \\ 0 & A^{\dagger -1} \end{bmatrix} \tag{183}$$

Then, if one introduces Dirac matrices γ^μ in the representation

$$\gamma^0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \gamma^k = \begin{bmatrix} 0 & \tau^k \\ -\tau^k & 0 \end{bmatrix},$$
$$\gamma^5 = -\gamma^1\gamma^2\gamma^3\gamma^0 = i\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{184}$$

satisfying

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}, \tag{185}$$

the results (75) and (65)

$$A^\dagger\rho^\mu A = L(A)^\mu{}_\nu\rho^\nu$$
$$A^{-1}\tau^\mu A^{\dagger -1} = L(A)^\mu{}_\nu\tau^\nu$$

can be united to read as

$$S(A)^\dagger(\gamma^0\gamma^\mu)S(A) = L(A)^\mu{}_\nu(\gamma^0\gamma^\nu),$$

or

$$S(A)^{-1}\gamma^\mu S(A) = L(A)^\mu{}_\nu\gamma^\nu. \tag{186}$$

In the representation (184), Eqs. (166)–(168) become

$$\Psi \to \pm i\gamma^0\Psi, \tag{187}$$
$$\Psi \to \pm i\gamma^0\gamma^5\Psi, \tag{188}$$
$$\Psi \to \pm\gamma^5\Psi, \tag{189}$$

and the scalar, pseudoscalar, vector, and axial vectors of Eqs. (176) and (177) become

$$\Psi^*\gamma^0\Psi', \quad \Psi^*i\gamma^0\gamma^5\Psi', \tag{190}$$
$$\Psi^*\gamma^0\gamma^\mu\Psi', \quad \Psi^*i\gamma^0\gamma^5\gamma^\mu\Psi'. \tag{191}$$

Further, for $A = A(\mathbf{n}, \theta)$ and $A = A(\chi, \mathbf{n})$ as given by Eqs. (85) and (105), $S(A)$ takes on the respective forms

$$S(\mathbf{n}, \theta) = \cos\tfrac{1}{2}\theta + \tfrac{1}{2}\sin\tfrac{1}{2}\theta\epsilon^{ijk}\gamma^i\gamma^j n^k, \tag{192}$$
$$S(\chi, \mathbf{n}) = \cosh\tfrac{1}{2}\chi + \sinh\tfrac{1}{2}\chi\gamma^0\gamma^k n^k. \tag{193}$$

Equations (182) and (186) to (193) constitute a four-spinor irreducible representation of the full Lorentz group for a general set of Dirac matrices satisfying Eq. (185), and agree exactly with the corresponding equations (in Secs. 4c and 4h) of reference 16. Use of $S(A)$ in the form (183) is, naturally, only proper in the representation (184) of the Dirac matrices.

[38] Iu. M. Shirokov, Soviet Phys.—JETP 6, 664 (1958).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

A Criterion for Singularities in Perturbation Theory

FRANCIS R. HALPERN
University of California, San Diego, La Jolla, California
(Received March 26, 1962)

A criterion is proposed to distinguish between the singular and nonsingular portions of the Landau surface on the physical sheet. The set of diagrams under consideration are those containing a single loop. A proof of the Mandelstam representation for the four-point function is given based on this criterion.

THERE are several difficulties involved in an investigation of the analytic structure of the S matrix from the point of view of perturbation theory. The discussion of apparently simple graphs can be quite involved and can require separate investigation of many special cases.[1] The number of graphs to be studied increases very rapidly with the number of both internal and external lines. In the spirit of Landau it seems wiser to use perturbation theory as a guide to conjecturing the analytic structure of the "real" S matrix rather than to disentangle the full details of perturbation theory itself.[2] The usefulness of the Mandelstam representation,[3] despite the existence of perturbation theoretic counter examples,[4] serves to emphasize this point. To this end it seems reasonable to provisionally assume that the n-point function has the analyticity of the n-sided polygon diagram. This approach has been followed in the preliminary approaches to the 5-point function.[5] This hypothesis disposes of the internal lines. In the remainder of this paper an attempt is made to give a unified and concise treatment of the polygonal diagram that shall be called the n-point function.

The n-point function F that is the analytic expression for the graph of Fig. 1. After the conventional manipulations it can be reduced to the form

$$F_n(x_{ij}) = C \int_0^1 \cdots \int_0^1 \times \frac{du_1 \cdots du_n\, \delta(1 - u_1 - \cdots - u_n) P(u_i)}{[\sum u_i^2 + 2 \sum u_s u_i x_{ij}]^{n-2}},$$

the u's are related to the Feynmann α's by

$$u_i = (\alpha_i/m_i)/\sum_s (\alpha_s/m_s)$$

and x_{ij} is given by

$$x_{ij} = [m_i^2 + m_j^2 - (\sum p_s)^2]/2m_i m_j$$

where the summation is carried out over all the external momenta between i and j (both directions around the polygon are equivalent). The factor $P(u_i)$ comes from the Jacobian of the transformation from α to u, and is a polynomial in u. Its role in determining the analyticity of F will be ignored.

If $n \geq 6$, all the x_{ij} are not independent but there exist a variety of geometrical constraints between them. The n-point function in these cases has only contracted singularities.[6] This phenomenon is closely related to the result that the leading singularities of the 5-point function are poles and not branch cuts.[7] In the ensuing analysis the x_{ij} will be assumed independent for convenience, but only the analysis of the 3-, 4-, and 5-point functions will be given.

The singularities of F_n are known to lie on a set of $[n(n-1)/2] - 1$ dimensional manifolds (Landau surface) specified by the vanishing of the determinant Δ_n, whose diagonal elements are unity and whose ijth element is x_{ij} for $i \neq j$, or by the vanishing of any of the principal minors of Δ_n. In fact each point of these manifolds is singular on some sheet of F_n.[7]

There is a set of 2^{n-1} interchanges of sign among the x_{ij} that leave Δ_n invariant. Suppose that all the elements that have a subscript i (either first or second) are changed in sign. This is the same as changing the sign of the ith row and the ith column, the diagonal element 1 would have its sign changed twice or equivalently not at all but it carries no subscript. This transformation may be carried out for all n values of the subscript. It may be applied successively to two, three, or more indices. If all these sign changes are counted, there is a total of 2^n sets of values of x_{ij} related by sign changes

[1] J. Tarski, J. Math. Phys. 1, 149 (1960).
[2] L. D. Landau, Nuclear Phys. 13, 181 (1959).
[3] S. Mandelstam, Phys. Rev. 112, 1344 (1958); 115, 1741 (1959); 115, 1752 (1959).
[4] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. 2, 656 (1961).
[5] L. Cook and J. Tarski, J. Math. Phys. 3, 1 (1962).
[6] L. Brown, Nuovo cimento 22, 178 (1961); P. V. Landshoff, Nuclear Phys. 20, 129 (1960).
[7] R. E. Cutkosky, J. Math. Phys. 1, 429 (1960).

that leave Δ_n unaltered. However, not all of these are distinct. If, for example, the sign of all rows and columns is changed, it is equivalent to no sign changes. Similarly there are always two ways to achieve any sign permutation and so there are only 2^{n-1} distinct changes. In particular any root $\bar{x}_{ij}$ such that $\Delta_n(\bar{x}_{ij}) = 0$ gives rise to 2^{n-1} related points on the Landau surface.

The equation $\Delta_n = 0$ is a quadratic in any of the x_{ij}

$$\Delta_n = A_{ij}x_{ij}^2 + B_{ij}x_{ij} + C_{ij} = 0$$

(not summed on i and j),

where the coefficients A_{ij}, B_{ij}, and C_{ij} depend on all the variables but x_{ij}. According to a theorem of Tarski[1]

$$B_{ij}^2 - A_{ij}C_{ij} = L_iL_j,$$

where L_i is the $n-1$ rowed principal minor formed by crossing out the ith row and ith column of Δ.

If $x_{ij} = \bar{x}_{ij}$ is a root of $\Delta = 0$ then the $n(n-1)/2$ equations

$$\bar{x}_{ij} = [-\bar{B}_{ij} \pm (\bar{L}_i\bar{L}_j)^{1/2}]/\bar{A}_{ij} \qquad (1)$$

must also be true. The quantities $\bar{A}$, $\bar{B}$, and $\bar{L}$ are the quantities A, B, and L evaluated at $x = \bar{x}$. It is clear that both values of the sign in (1) cannot be correct. It has been shown that if $\bar{x}_{ij}$ is a root of $\Delta = 0$ then other roots can be found permuting the signs of some rows and columns of Δ. Under these permutations the principal minors, L and A are invariant for the same reason that Δ itself is invariant. The B's may change signs. Thus by permuting the signs of the x_{ij} the two possible signs of the radicals are exhibited.

For complex values of the variables x, it will be more convenient to define each of the roots $(L_i)^{1/2}$ as a two valued function $(L_i)^{1/2}(S)$. The variable (S) in the expression $(L_i)^{1/2}(S_i)$ is a dichotomic variable whose values will be taken as plus $(+)$ and minus $(-)$. These values of S_i will specify the sign of the real part of the square root and the cuts will be taken along L_i real and negative. In (1) the negative sign is suppressed and the variables S_i allow for the double-valued character of the square root. There is an ambiguity in the set of equations (1) since every S_i may be changed and the equations will still be satisfied. At a point $\bar{x}_{ij}$ such that $\Delta_n(\bar{x}_{ij}) = 0$, the values of the S_i in $(L_i)^{1/2}(S_i)$ may be chosen to be the same in each of the $(n-1)$ places that this radical occurs. Thus the point $\bar{x}_{ij}$ is characterized by a set of values

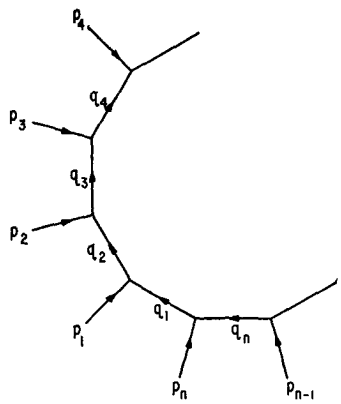


FIG. 1. The graph for the n-point function.

of the S_i. More properly, because of the sign ambiguity at each point $\bar{x}_{ij}$, all of the S_i are grouped into two classes; the members of either class may be chosen plus and then the others will be minus, or vice versa.

To exhibit the possibility of choosing S_i to be the same each time that it occurs first choose S_1 to be $+$ and the other S_i to be fixed by the equations for x_{12} through x_{1n}. It must now be shown that the remaining equations are consistent with this assignment of signs. Consider the expression

$$(\bar{A}_{1i}\bar{x}_{1i} + \bar{B}_{1i})(\bar{A}_{1j}\bar{x}_{1j} + \bar{B}_{1j})(\bar{A}_{ij}\bar{x}_{ij} + \bar{B}_{ij}) = \bar{L}_1(\bar{L}_i)^{1/2}(\bar{L}_j)^{1/2}(\bar{L}_i)^{1/2}(\bar{L}_j)^{1/2} \qquad (2)$$

that arises by solving (1) for the square root and multiplying together the results for $1i$, $1j$, and ij. It is clear that $\bar{L}_1$ comes out of the radical with a plus sign since in both the expression for $\bar{x}_{1i}$ and $\bar{x}_{1j}$ it was specified that $(\bar{L}_1)^{1/2}$, should be on its first sheet. The first $(L_i)^{1/2}$ and $(L_j)^{1/2}$ that occur are those that arise from the $\bar{x}_{1i}$ and the $\bar{x}_{ij}$ equations and their sheets are specified by consistency with the convention for $(L_1)^{1/2}$. It must be shown that the second pair $(L_i)^{1/2}$, $(L_j)^{1/2}$ can be chosen so that they are on the same sheets, respectively, as the first pair. This will be true if the right-hand side of (2) is $+\bar{L}_1\bar{L}_i\bar{L}_j$ and not $-\bar{L}_1\bar{L}_i\bar{L}_j$. Since there is a discontinuous change from one sign to the other it will be sufficient to show that the plus sign obtains at one point P to show that it is generally valid. By continuity there is a neighborhood of P in which (2) holds with a plus sign, but since both sides are polynomials it is an identity and holds everywhere.

A useful point to choose is $\bar{x}_{ij} = -1/(n-1)$. For this choice The $\bar{A}$, $\bar{B}$, $\bar{C}$, and $\bar{L}$ are independent of i and j and are given by

$$A = -2(n-1)^{2-n}n^{n-3}$$

$$B = (n-2)(n-1)^{1-n}n - 3$$

$$C = 2(n-1)^{(1-n)}n^{n-3}$$

$$L = (n-1)^{1-n}n^{n-2}.$$

It can easily be seen that (2) is satisfied with the radicals chosen consistently from a single sheet.

Thus at each point on the Landau surface if $(L_1)^{1/2}$ is put on its first sheet the other $(n-1)(L_i)^{1/2}$ are fixed on either their first or second sheets. The condition can be stated in a more homogeneous way. At each point on the Landau surface some s of $(\bar{L})^{1/2}$ are on the same sheet and the other $(n-s)(L)^{1/2}$ are on the other sheet. It is irrelevant which of the two sheets each group is on, but rather the set of L's which are grouped together. There are a total 2^{n-1} such groupings of the $(L)^{1/2}$'s. Each point $\bar{x}$ of the Landau surface corresponds to one and only one such configuration. The boundary between configurations occurs when an L vanishes or along the cut that connects the surface $L_i = 0$ to infinity. The surfaces $L_i = 0$ are the Landau surfaces for the $n-1$ point function in some subset of the variables, and the cuts for these $n-1$ point functions will be taken in the same way as those for the L_i are taken.

To complete the proof, it should be shown that none of the cuts of F associated with the $n-2$-point, $\cdots$ 2-point singularities divide any of the 2^{n-1} regions into two or more parts. The convention for taking all the cuts will be Im $\{L_i\} = 0$ and Re $\{L_i\} < 0$. These cuts join the points $L_i = \Delta_{n-1} = 0$ to infinity and reduce to the usual normal threshold cuts for the two point function.

In the region $|\text{Re}\ \{x_{ij}\}| < 1$ and Im $\{x_{ij}\} = 0$ there is a family of closed surfaces S_n consisting of the points of the Landau surface that are closest to the origin in the following sense. A point P is a point of S_n if the coordinates x_{ij} of P satisfy the Landau equation and if the straight line joining P to the origin does not intersect the Landau surface. For example, the surface S_2 is a hypercube with sides given by $x_{ij} = \pm 1$. The important property of this family of surfaces is that S_n lies inside or touches S_{n-1}. That is if a straight line is drawn from the origin in any direction it passes through a point of S_n before or at the same time it reaches S_{n-1}. Since Δ_n equals one at the origin independently of the value of n the assertion may be proven by showing that $\Delta_n \leq 0$ at a point of S_{n-1}. This follows readily since on S_{n-1} one of the radicals L_i in (1) will vanish and Δ_n is then a perfect square. If Δ_n is written this way its coefficient is negative. Thus by continuity there must be a point of S_n closer to the origin than the point of S_{n-1}.

The points of contact between S_{n-1} and S_n separate S_n into the 2^{n-1} regions. This portion of the Landau surface is separated by the singularity surface S_{n-1} rather than the branch cuts coming from S_{n-1}.

No cuts of the $(n-2)$-point or lower function can intersect S_n and the points of S_{n-2} or lower surfaces simultaneously points of S_n are a manifold of sufficiently low dimensionality (since they must simultaneously belong to at least S_n, S_{n-1}, and S_{n-2}) that they cannot separate the points of S_n.

S_n lies closer to the origin than all the cuts of the lower functions and if it is possible to get to the end of a $(n-2)$-point or lower cut there is a path around it. The present proof is incomplete because it has not yet been possible to show the existence of a path from an arbitrary point of the Landau surface to a point of S_n.

Let the values of the u_i for which $\partial D/\partial u_i = 0$, at the point $\bar{x}_{ij}$ of the Landau surface be $\bar{u}_i$. At the point $\bar{x}'_{ij}$ where the sign of all terms with an index k are changed the u have the values $\bar{u}'_i = \bar{u}_i$ for $i \neq k$ and $\bar{u}'_k = -\bar{u}_k$. Since it is necessary to have all the x's real and positive to have a singularity on the physical sheet for those singularities nearest to the Euclidean region it follows that only one of the 2^{n-1} regions on the Landau surface is singular on the physical sheet. Since the regions are bounded by the $(n-1)$-point function branch cuts an entire region is singular if one point of it is.

The point $\bar{x}_{ij} = -1/(n-1)$ is singular on the physical sheet. The corresponding u's are $1/n$ and they are real and positive. The denominator D is positive for values of x_{ij} in the neighborhood of $x_{ij} = -1/(n-1)$. For simplicity take all the x's to be equal. Then the denominator D is given by

$$D = \sum u_i^2 + x \sum u_i u_j.$$

To establish that this point is singular it is sufficient to observe that D is positive if x is greater than $-1/(n-1)$. Thus the hypercontour of integration is undistorted until the point $\bar{x}_{ij} = -1/(n-1)$ is reached. To show that D is positive, observe that it is a quadratic form in the u's and that the eigenvalues are $1-x$ repeated $n-1$ times and a simple eigenvalue at $1+nx-x$. The multiple eigenvalue becomes negative for positive values of x where D is positive. The simple eigenvalue vanishes at $\bar{x}$ and is negative for more negative values of x. At the point $\bar{x}_{ij} = -1/(n-1)$ all the radicals are taken on the same sheet. Thus all those points of

the Landau surface that have all the radicals taken on the same sheet will be singular on the physical sheet.

As an illustration of the utility of these notions some applications are made to the three-, four-, and five-point functions.

In the case of the three-point function there are three variables x_{12}, x_{13}, and x_{23}. It is convenient to write them as

$$x_{12} = \cos\alpha, \qquad x_{13} = \cos\beta, \qquad x_{23} = \cos\gamma.$$

In terms of α, β, and γ, the equation of the Landau surface may be written

$$\begin{aligned}[\cos(\alpha+\beta+\gamma)-1][\cos(-\alpha+\beta-\gamma)-1]\\ \times[\cos(\alpha-\beta+\gamma)-1]\\ \times[\cos(-\alpha+\beta+\gamma)-1]=0 \qquad (3)\end{aligned}$$

The use of elementary trigonometric identities reduces (3) to

$$\begin{aligned}(1-\cos^2\alpha-\cos^2\beta-\cos^2\gamma\\ +2\cos\alpha\cos\beta\cos\gamma)^2=0, \qquad (4)\end{aligned}$$

which is the Landau equation.

There are four possible relations between α, β and γ each of which will serve to make one of the four factors of (3) vanish.

These are

$$\alpha+\beta+\gamma=2n\pi \qquad \alpha-\beta+\gamma=2n\pi$$
$$\alpha+\beta-\gamma=2n\pi \qquad -\alpha+\beta+\gamma=2n\pi.$$

The conventional results follow if the solutions for which the real parts of the variables α, β and γ are restricted to lie between 0 and π. The region for which $\alpha+\beta+\gamma=2\pi$ is the region that is singular on the physical sheet.[2] The three solutions of (4) are

$$\begin{aligned}-\cos\alpha+\cos\beta\cos\gamma&=(1-\cos^2\beta)^{1/2}(1-\cos^2\gamma)^{1/2}\\ -\cos\beta+\cos\alpha\cos\gamma&=(1-\cos^2\alpha)^{1/2}(1-\cos^2\gamma)^{1/2}\\ -\cos\gamma+\cos\alpha\cos\beta&=(1-\cos^2\alpha)^{1/2}(1-\cos^2\beta)^{1/2}.\end{aligned} \qquad (5)$$

The choice of signs in (5) is dictated by the convention established in (1). If $\alpha+\beta+\gamma=2\pi$ it must be shown that the three radicals are all to be taken on the same sheet. In the first equation of (5) substitute $\alpha=2\pi-\beta-\gamma$. Then $\cos\alpha$ becomes $\cos\beta\cos\gamma-\sin\beta\sin\gamma$ and a consistent choice of signs in that $(1-\cos^2\beta)^{1/2}=\sin\beta$ not $-\sin\beta$ and similarly $(1-\cos^2\gamma)^{1/2}=\sin\gamma$. By symmetry it is clear that $(1-\cos^2\alpha)^{1/2}=\sin\alpha$ will give a consistent set of solutions for the three Eqs. (5). Now if α, β, and γ have real parts between 0 and π the real parts of $\sin\alpha$, $\sin\beta$, and $\sin\gamma$ are all positive and all three radicals have been taken on the same sheet.

If in the four-point function x_{12}, x_{23}, x_{34}, and x_{14} are chosen real and less than one in absolute value the analyticity in x_{13} and x_{24} which play the role of the usual scalar invariants s and t may be discussed. The appropriate three-point cuts are given by

$$\begin{aligned}L_1&:\ \mathrm{Im}\,\{x_{24}\}=0\\ &\quad 1-x_{23}^2-x_{24}^2-x_{34}^2+2x_{23}x_{24}x_{34}<0,\\ L_2&:\ \mathrm{Im}\,\{x_{13}\}=0\\ &\quad 1-x_{13}^2-x_{14}^2-x_{34}^2+2x_{13}x_{14}x_{34}<0,\\ L_3&:\ \mathrm{Im}\,\{x_{24}\}=0\\ &\quad 1-x_{12}^2-x_{14}^2-x_{24}^2+2x_{12}x_{14}x_{24}<0,\\ L_4&:\ \mathrm{Im}\,\{x_{13}\}=0\\ &\quad 1-x_{12}^2-x_{13}^2-x_{23}^2+2x_{12}x_{13}x_{23}<0.\end{aligned} \qquad (6)$$

There is a considerable overlapping of the cuts on the Landau surface. If the conditions L_1 and L_3 are both satisfied then L_2 and L_4 will also be satisfied by virtue of the restriction imposed by the Landau equation and, vice versa, the satisfaction of L_2 and L_4 implies the simultaneous satisfaction of L_1 and L_3.

If a path is chosen in the Landau surface that penetrates this fourfold cut, there is no change in the character of the surface after it passes through the cut since all the radicals have changed their sheets simultaneously. Only the real values of x_{24} such that one but not both of L_1 or L_3 are satisfied (similarly for x_{13} and L_2 and L_4) give cuts across which the surface may change from singular to nonsingular. The ends of these cuts closest to the origin are on the surface S_4 with $x_{12}x_{23}x_{34}$ and x_{14} fixed, this surface becomes a two-dimensional curve (Γ_5 in reference 1), and if there are singularities on S_4 then these will appear on either side of the various cuts and there will be complex singularities. If S_4 is free of singularities then the only singularities will be in the fourfold cut. This is the same conclusion as Tarski.[1]

In treating the five-point function it is useful to parameterize the equation $\Delta_5=0$ in the following way.[8] If $\Delta_5=0$ then the matrix with the same elements as Δ_5 has an eigenvector belonging to the characteristic value 0. This vector is of course just

[8] F. R. Halpern, Phys. **127**, 1819 (1962).

the vector formed from the integration variables u_i in the definition of F_5. If the masses $x_{i\,i+1}$ and x_{15} are fixed in the five-point function, it is possible to regard the dynamical variables x_{13}, x_{14}, x_{24}, x_{25}, and x_{35} as unknowns and solve for them in terms of the 5 masses and $5u_i$. This solution is

$$x_{13} = (-u_1^2 - u_2^2 - u_3^2 + u_4^2 + u_5^2 + 2u_4u_5 - 2u_1u_2x_{12} - 2u_2u_3x_{23})/2u_1u_3 \qquad (7)$$

and the four obvious permutations of this equation. This parameterization is homogeneous in the u's, and thus contains only four free variables. The only forbidden values of the u_i are zero. Other than this, any real or complex values may be used. The value $u_i = 0$ corresponds to an intersection between a four-point singularity $L_i = 0$ and the five-point surface.

An immediate consequence of the fact that the dynamical variables have a rational parameterization is that the L_i must have the form $L_i = f_i^2 g$, where g is independent of i. This follows since if this parameterization is substituted in Eq. (1) the only irrationality is $(L_iL_j)^{1/2}$ and consequently it must be possible to take this square root. This will only be possible if L_i has the suggested form. The function g turns out to be quite complicated but f_i is simply u_i.[8]

To determine what values of u_i correspond to singular points of the Landau surface it is necessary to determine the sheet of the radical. It may be shown by direct calculation that $(u_i^2 g)^{1/2}$ should be taken to be $u_i g^{1/2}$ and the sign of the real part of this expression must be found. Let the phase of g be θ and of the phase u_i be ϕ_i. Then the phase of the radical $(L_i)^{1/2}$ is $\phi_i + \theta/2$ and the corresponding complex numbers must all be either in the right or left half-planes for a singularity on the physical sheet. If the phase of each u is increased by α, the phase of g decreases by 2α, this follows from the homogeneity of all expressions. In the particular case when the phase of g is 0, all the u's must lie in either the left or right half-planes. Changing the phases by α rigidly rotates this configuration but maintains all the u's in a half-plane. Thus a necessary condition for a singularity is that all the u_i differ in phase by less than π.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

On Spinors in n Dimensions*

ABRAHAM PAIS†
Brookhaven National Laboratory, Upton, Long Island, New York
(Received June 29, 1962)

The matrix which enters in the charge conjugation transformation of the usual spinors in 4-space is an invariant matrix and is skew symmetric. It is shown that there exists such an invariant matrix C for any number of dimensions (and independent of the number of time like dimensions). Its symmetry properties depend on the dimension number n *modulo* 8. With the help of the C matrix one can construct, for $n = 1, 2, 7, 8$ mod 8, an n-dimensional invariant bilinear in the components of a *single* n-dimensional spinor. Some examples are given for $n = 2, 3, 7$. A bilinear baryon invariant is formed for a theory with high symmetry. Its existence is closely related to the triality property of 8-space.

I. INTRODUCTION

SPINORS Ψ in n dimensions are multicomponent objects which transform in a specific way [see Eqs. (13) and (18) below] under transformations which leave invariant a quadratic form $E^n = \sum_{i=1}^{n} \epsilon_i X_i^2$, where $\epsilon_i = \pm 1$. The signature of E^n [given by $(n - m) + m$, where m is the number of minus signs or time like dimensions in E^n] will be immaterial unless otherwise stated. For $n = 2\nu$ or $2\nu + 1$, Ψ has 2^ν spinor components. In general each component may be a function of the X_i. In this note we shall only be concerned with the case that there is no such dependence (constant spinors).

We call Dirac matrices in E^n a set of n matrices Γ_α, $\alpha = 1, \cdots, n$ which satisfy

$$\Gamma_\alpha \Gamma_\beta + \Gamma_\beta \Gamma_\alpha = 2\,\delta_{\alpha\beta} \cdot I, \qquad \alpha, \beta = 1, \cdots, n, \quad (1)$$

where I is the unit matrix. For $n = 2\nu$ and $2\nu + 1$ these relations can be satisfied by Γ's which are $2^\nu \times 2^\nu$ matrices.

It is well known[1] that the Γ_α can be expressed as direct products (also called Kronecker products or tensor products) of the Pauli spin matrices

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2)$$

This is quite familiar in the case $n = 4$ where, for example, we can take

$$\Gamma_1 = \sigma_x^{(1)} \times \sigma_x^{(1)}, \quad \Gamma_2 = \sigma_x^{(1)} \times \sigma_y^{(2)},$$
$$\Gamma_3 = \sigma_x^{(1)} \times \sigma_z^{(2)}, \quad \Gamma_4 = \sigma_y^{(1)}. \quad (3)$$

The superscripts (1), (2) refer to two distinct sets of Pauli matrices and the notation $\times$ denotes direct product. For example to get Γ_2, first write down a matrix σ_x, but consider its elements as 2×2 null and unit matrices, respectively. Next multiply each of these four matrices by the 2×2 matrix $\sigma_y^{(2)}$, so that

$$\Gamma_2 = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix}.$$

The direct multiplication process can be repeated. For example to obtain $\Gamma_2 \times \sigma_z^{(3)}$, consider the elements 0 and $\pm i$ of Γ_2 as 2×2 null matrices and $\pm i$ times 2×2 unit matrices, respectively, and multiply each of these matrices with σ_z. For $n = 2\nu$ or $2\nu + 1$ we shall need direct products of ν matrices of the kinds $\sigma_x^{(i)}$, $\sigma_y^{(i)}$, $\sigma_z^{(i)}$, $1^{(i)}$, $i = 1, \cdots, \nu$, where $1^{(i)}$ is the ith 2×2 unit matrix.

This construction procedure for Dirac matrices makes it considerably simpler to derive and understand some results due to Cartan.[2] In his book on spinors a more geometrical reasoning is used which, I believe, makes the derivations for general n unduly cumbersome. There are several interesting theorems on bilinear spinor covariants which depend in a non-trivial way on n *modulo* 8. These can all be obtained by asking, in the language of the physicists, the following question. What is the n-dimensional generalization of the familiar charge conjugation matrix C?

Briefly, C is a unitary, skew symmetric matrix and is invariant under the full Lorentz group.[3] In any given representation, C considered as a unitary transformation sends any of the n Dirac

* Work supported, in part, by the U. S. Atomic Energy Commission.
† Permanent address: The Institute for Advanced Study, Princeton, New Jersey.
[1] R. Brauer and H. Weyl, Am. J. Math. 57, 447 (1935).
[2] E. Cartan, *Leçons sur la theorie des spineurs* (Hermann and Cie, Paris, 1938), 2 vols. Actualités Scientifiques, Nos. 643, 701.
[3] For a discussion of the properties of C, see e. g., A. Pais and R. Jost, Phys. Rev. 87, 871 (1952).

matrices into its transposed. In the next section we construct a unitary and invariant matrix C for any dimension. The interesting thing is the way the symmetry of C depends on the dimension number, see Eqs. (9) and (25) below. In Sec. III we give some illustrations for two-, three-, and seven-dimensional rotation groups.

II. THE C MATRIX

(A) Even Dimensions, $E^{2\nu}$

Lemma. There exists a representation in which all Γ_α are Hermitian, half of them are symmetric (real), the other half antisymmetric (imaginary).

One way of achieving this is as follows:

$$\left.\begin{aligned}\Gamma_1 &= \Lambda_\nu,\\ \Gamma_{2m} &= \Lambda_{\nu-m}\times\sigma_y^{(\nu-m+1)}\times 1^{(\nu-m+2)}\times\cdots\times 1^{(\nu)},\\ \Gamma_{2m+1} &= \Lambda_{\nu-m}\times\sigma_z^{(\nu-m+1)}\times 1^{(\nu-m+2)}\times\cdots\times 1^{(\nu)},\\ \Gamma_{2\nu} &= \sigma_y^{(1)}\times 1^{(2)}\cdots\times 1^{(\nu)},\end{aligned}\right\}\quad m=1,\cdots,\nu-1 \tag{4}$$

where

$$\Lambda_m = \sigma_x^{(1)}\times\sigma_x^{(2)}\times\cdots\times\sigma_x^{(m)}. \tag{5}$$

For $n = 2$ this reduces to Eq. (3). It is easy to see that Eq. (1) is satisfied. As the σ's are Hermitian, so are the Γ's. Furthermore

$$\Gamma_{2\alpha-1}^t = \Gamma_{2\alpha-1},\qquad \Gamma_{2\alpha}^t = -\Gamma_{2\alpha}, \tag{6}$$

where t denotes transposition. Equation (6) is true because the transposed of a direct product equals the direct product of the transposed factors.

We now define the "charge conjugation matrix" C by

$$C = \prod_{\alpha=1}^{\nu}\Gamma_{2\alpha-1}. \tag{7}$$

C has three main properties. First,

$$C^\dagger C = 1. \tag{8}$$

Secondly, from Eqs. (1) and (6)

$$C^t = (-1)^{(\nu/2)(\nu-1)}C = \begin{cases} C, & \nu = 0,\ 1 \bmod 4,\\ -C, & \nu = 2,\ 3 \bmod 4.\end{cases} \tag{9}$$

Finally, it follows from Eqs. (1) and (7) that $C\Gamma_\alpha = (-1)^{\nu+\epsilon}\Gamma_\alpha C$ where $\epsilon = 0(1)$ for α = even (odd). Hence using Eq. (6)

$$C\Gamma_\alpha = (-1)^{\nu+1}\Gamma_\alpha^t C. \tag{10}$$

Thus for $\nu = 2$, Eqs. (8), (9), and (10) show that C has the familiar properties.[3]

Next consider the behavior of C under the following kinds of transformations.

(a) *Change of* Γ *Representation*

$$\Gamma'_\mu = S^{-1}\Gamma_\mu S. \tag{11}$$

The corresponding C' is defined so as to preserve Eq. (10) so that

$$C' = S^tCS. \tag{12}$$

It follows that C' again satisfies Eqs. (8) and (9).

(b) *Invariance of C under Rotations*

An infinitesimal rotation on the 2^ν-component spinor Ψ is given by

$$\Psi' = S\Psi,\quad S = 1 + \tfrac{1}{4}\epsilon_{\rho\sigma}\Gamma_\rho\Gamma_\sigma,\quad \epsilon_{\rho\sigma} = -\epsilon_{\sigma\rho}. \tag{13}$$

Under this *specific* S-transformation $C' = C$, from Eqs. (10) and (12). Note that the reality properties of $\epsilon_{\rho\sigma}$ do not enter the discussion here, so that these results are independent of the signature of the metric.

(c) *Invariance of C under Reflections*

At this point it is helpful to introduce the generalization of the customary γ_5 matrix, namely,

$$\Gamma_{2\nu+1} = (-i)^\nu\prod_{\alpha=1}^{2\nu}\Gamma_\alpha. \tag{14}$$

In the special representation of Eqs. (4) and (5) we have

$$\Gamma_{2\nu+1} = \sigma_z^{(1)} = \begin{bmatrix} I_{\nu-1} & 0\\ 0 & -I_{\nu-1}\end{bmatrix}, \tag{15}$$

where $I_{\nu-1}$ is the $2^{\nu-1}$-dimensional unit matrix. Thus

$$\Gamma_{2\nu+1}^t = \Gamma_{2\nu+1}, \tag{16}$$

while from Eqs. (6), (10), and (14) we obtain

$$C\Gamma_{2\nu+1} = (-1)^\nu\Gamma_{2\nu+1}^t C. \tag{17}$$

Consider the coordinate reflection in $E^{2\nu}$: $X'_\alpha = -X_\alpha$, $X'_\beta = X_\beta$, $\beta \neq \alpha$, $\alpha = 1$ or $2\cdots$ or 2ν. There is an ambiguity in the definition of the behavior of Ψ

under reflections. We may put

$$\Psi' = S_\alpha \Psi, \qquad S_\alpha = i\Gamma_\alpha \Gamma_{2\nu+1}. \tag{18}$$

With this choice, C is also invariant under reflections, as follows from Eqs. (12) and (17). Thus Eqs. (8), (9), and (10) are invariant and independent of the choice of representation.

Bilinear covariants. We have at once that the quantity T, defined by

$$T = \Psi^t C \Psi, \tag{19}$$

is a scalar with respect to the full n-dimensional rotation group (i.e., including reflections). For $T' = \Psi^t S^t C S \Psi = T$. However, it is evident that $T \equiv 0$ if[4] $C^t = -C$. Therefore

$$T \not\equiv 0 \quad \text{if} \quad \nu = 0, 1 \bmod 4, \tag{20}$$

that is, in 2, 8, $\cdots$ dimensions. The extension to tensors of higher rank is trivial. Thus $\Psi^t C(\Sigma \pm \Gamma_{\mu_1} \cdots \Gamma_{\mu_p})\Psi$ is a tensor[5] of rank p which is $\not\equiv 0$ for $\nu(\nu-1) + 2p(\nu+1) + p(p-1) = 0 \bmod 4$. In particular, for $p = 2\nu$ we get a nonvanishing "pseudoscalar" for $\nu(\nu+1) = 0 \bmod 4$.

Bilinear covariants excluding reflections. It follows from Eq. (15) that in the special representation Eq. (14) the quantities

$$\Psi_\pm = [(1 \pm \Gamma_{2\nu+1})/2]\Psi \tag{21}$$

have, in general, $2^{\nu-1}$ nonvanishing components. These are the so-called semispinors of first and second kind. Define

$$T_\pm = \Psi_\pm^t C \Psi_\pm. \tag{22}$$

These are scalars with respect to the restricted group where reflections are excluded. From Eqs. (16) and (17), $T_+ \not\equiv 0$ for $\nu = 0 \bmod 4$. Likewise $T_-' = \Psi_-^t C \Psi_- \not\equiv 0$ for $\nu = 1 \bmod 4$. The construction of higher rank tensors in terms of semi-spinors is obvious.

(B) Odd Dimensions, $E^{2\nu+1}$

According to Eq. (15), $\Gamma_{2\nu+1}^2 = 1$, and furthermore $\Gamma_{2\nu+1}$ anticommutes with all Γ_α, $\alpha = 1, \cdots, 2\nu$. For $2\nu+1$ dimensions a representation of the Dirac matrices is therefore given by Eqs. (4) and (15). Define

$$C = \prod_{\alpha=1}^{\nu+1} \Gamma_{2\alpha-1}. \tag{23}$$

One has

$$C^\dagger C = 1, \tag{24}$$

$$C^t = (-1)^{(\nu/2)(\nu+1)} C, \tag{25}$$

$$C\Gamma_\alpha = (-1)^\nu \Gamma_\alpha^t C. \tag{26}$$

The transformation Eq. (12) holds true here too. So does the invariance of C under rotations. We need not consider reflections. We have

$$T = \Psi^t C \Psi \not\equiv 0, \qquad \nu = 0, 3 \bmod 4, \tag{27}$$

and of course T is a scalar. Tensors of higher rank are discussed as under (A).

Expressions like Eqs. (19), (22), and (27) can of course be generalized to $\Phi^t C \Psi$, where Φ, Ψ are two distinct spinors. For $\Phi \not\equiv \Psi$ the corresponding scalars and other covariants clearly are in general nonzero for any dimension.

III. APPLICATIONS

(1) For orientation, consider the familiar instance $n = 3$. We are in case (B), $\nu = 1$ and according to the recipe of Eq. (4), $\Gamma_1 = \sigma_x$, $\Gamma_2 = \sigma_y$, $\Gamma_3 = \sigma_z$, so from Eq. (23), $C = -i\sigma_y$. Consider two spinors $\Phi^t = [\alpha(1), \beta(1)]$ and $\Psi^t = [\alpha(2), \beta(2)]$. α and β denote the eigenstates for spin up and down, respectively. The "arguments" 1 and 2 simply denote that we deal with two distinct spinors. The scalar $\Phi^t C \Psi = -\alpha(1)\beta(2) + \alpha(2)\beta(1)$. This is of course the two-particle singlet state. Now put "1 = 2," that is, take the scalar $\Phi^t C \Phi$. This vanishes identically.

Thus Eqs. (20) and (27) say that only for dimensions $n = 1, 2, 7, 8 \bmod 8$ is it possible to form a nonvanishing singlet state bilinear in one single spinor.

(2) As the simplest example for a nonvanishing invariant take $n = 2$. We have case (A), $\nu = 1$. Thus $\Gamma_1 = \sigma_x$, $\Gamma_2 = \sigma_y$, $C = \sigma_x$. Therefore $\Phi^t C \Psi = \alpha(1)\beta(2) + \alpha(2)\beta(1)$. This quantity is indeed a scalar for rotations in the xy plane, as it is the z component of the spin 1-vector in 3-space. And $\Phi^t C \Phi$ indeed does not vanish.

(3) A less trivial example of a nonzero scalar is provided by the seven-dimensional charge space formalism of baryon meson interactions.[6] Here the coupling $\bar{\Psi}\Gamma_\mu \Psi \Phi_\mu$ is considered with

$$\Psi^t = \left(p, n, \Xi^0, \Xi^-, \Sigma^+, \frac{\Lambda - \Sigma^0}{\sqrt{2}}, \frac{\Lambda + \Sigma^0}{\sqrt{2}}, \Sigma^-\right) \tag{28}$$

[4] The components of Ψ are supposed to commute.

[5] $\Sigma +$ denotes the alternating sum over permutations of $(1, \cdots, p)$.

[6] First given by J. Tiomno, Nuovo cimento **6**, 69 (1957).

$$\Phi_\mu = \left(\frac{K^- + K^+}{\sqrt{2}}, \frac{K^- - K^+}{i\sqrt{2}}, \frac{\bar{K}^0 + K^0}{\sqrt{2}}, \frac{\bar{K}^0 - K^0}{i\sqrt{2}}, \frac{\pi^- + \pi^+}{\sqrt{2}}, \frac{\pi^- - \pi^+}{i\sqrt{2}}, \pi_0\right), \quad (29)$$

$$\Gamma_\mu = \sigma_x^{(1)} \times \sigma_y^{(2)} \times 1^{(3)}, \quad \sigma_x^{(1)} \times \sigma_y^{(2)} \times 1^{(3)},$$
$$\sigma_x^{(1)} \times \sigma_z^{(2)} \times 1^{(3)}, \quad \sigma_y^{(1)} \times 1^{(2)} \times 1^{(3)},$$
$$\sigma_z^{(1)} \times 1^{(2)} \times \sigma_x^{(3)}, \quad \sigma_z^{(1)} \times 1^{(2)} \times \sigma_y^{(3)},$$
$$\sigma_z^{(1)} \times 1^{(2)} \times \sigma_z^{(3)}. \quad (30)$$

Ψ is an 8-component spinor with respect to charge space. Each of its components is a 4-spinor with respect to Lorentz space. We shall denote by Ψ_A, $A = 1, 2, 3$ or 4 the four quantities obtained by taking the same Ath Lorentz space component of each charge space component. Each Ψ_A transforms as a spinor under the seven-dimensional rotation group.

The Γ representation of Eq. (30) is not like Eqs. (4) and (15) but it does satisfy Eqs. (6) and (16). We can therefore apply the formalism of case (B) with $\nu = 3$. Thus $C = -1^{(1)} \times \sigma_y^{(2)} \times \sigma_y^{(3)}$. Define T_{AB} by

$$T_{AB} = \Psi_A' C \Psi_B = \Lambda_A \Lambda_B - (\Sigma_A^+ \Sigma_B^- + \Sigma_A^0 \Sigma_B^0 + \Sigma_A^- \Sigma_B^+) - (p_A \Xi_B^- - n_A \Xi_B^0) - (\Xi_A^- p_B - \Xi_A^0 n_B). \quad (31)$$

For all $A, B = 1, \cdots, 4$, T_{AB} is a nonvanishing scalar with respect to the 7-group. The same is true for $\bar{T}_{AB} \equiv \Psi_A' C \Psi_B$.

T_{AB} is of course also a scalar with respect to any subgroup of the 7-group, but then it breaks up in separate parts invariant with respect to that subgroup. An obvious example is the isotopic spin with respect to which T_{AB} clearly consists of four additive scalar parts. A less trivial case is the exceptional group $G2$ with respect to which $\Lambda_A \Lambda_B$ and $T_{AB} - \Lambda_A \Lambda_B$ are separate scalars. For example the finite angle rotations of $G2$ explicitly given by Behrends and Sirlin[7] are seen to leave $\Lambda_A \Lambda_B$ and $T_{AB} - \Lambda_A \Lambda_B$ separately invariant.

There begins to emerge a certain three-way equivalence for the baryon, the antibaryon (each in a given spin state) and the meson. For each there exists a bilinear invariant, namely, T_{AA}, $\bar{T}_{AA}$ and $\phi_\mu^2 = \bar{K}K + K\bar{K} + \pi^2$, respectively. Let us introduce an additional meson state σ with zero spin and hypercharge which forms an octet with π and K. Now all three quantities are 8-component. The structures now before us are just the ones which play a role in the principle of triality,[8] unique for dimension eight. Briefly stated, triality is a certain substitutional invariance (involving the alternating group on three things) between vectors, semispinors of the first kind and semispinors of the second kind in an 8-space which leaves invariant a trilinear form. In our case this form is $\bar{\Psi}\Gamma_\mu \Psi \phi_\mu + \bar{\Psi}\Psi\sigma$. Note that such a substitutional invariance is conceivable only in an 8-space, because only for that dimension do vectors and semispinors of either kind all have the same number of components. Triality invariance is very closely connected with the octonion formulation of interactions given elsewhere.[9] If applicable, the principle of triality would state that the baryon, the antibaryon and the meson have identical properties of higher symmetry and that the interaction should be such that these three objects are interchangeable (with respect to their intrinsic symmetries) as is the case for their trilinear invariant. In all this we have not insisted on the spatial transformation properties such as the spin zero character of the meson.

In all the foregoing, the signature of the metric played no role. This only enters when we have to consider simultaneously a spinor and its adjoint. Let the directions $h, k, \cdots$ be timelike. Then the adjoint of Ψ is $\bar{\Psi} = \Psi^\dagger \Gamma_h \Gamma_k \cdots$.

It is well known that familiar charge conjugation can only be formulated consistently in a quantized theory. Likewise the desired change of sign of the 4-current is also closely related to the 3 + 1 signature of the metric.

Consider as a last example a 7-space with signature 6 + 1. Let "4" be the timelike direction. Form the current $\bar{\Psi}\Gamma_\mu\Psi$, $\bar{\Psi} = \Psi^\dagger \Gamma_4$. Perform now a charge conjugation $\Psi' = C^{-1}\bar{\Psi}^t$. Taking the adjoint of this gives $\bar{\Psi}' = -\Psi' C$. [We are in case (B) with $\nu = 3$.] Note how this minus sign is closely connected with the oddness of the number of timelike directions. Now let $\Psi^\dagger$ and Ψ anticommute. Then $(\bar{\Psi}\Gamma_\mu\Psi)' = -(\bar{\Psi}\Gamma_\mu\Psi)$.

This property is isomorphic to G conjugation. This is seen as follows. A representation of the Γ_μ is provided by $\Gamma_\mu = \gamma_\mu$, μ = 1-4, $\Gamma_{5,6,7} = \gamma_5 \tau_{1,2,3}$. Here the γ_μ are the usual four Dirac matrices and τ_i is the isotopic spin vector. Take γ_1 and γ_3 real, γ_2 and γ_4 imaginary, and apply the general formulas. This gives $C_7 = \gamma_1\gamma_3 \cdot -i\tau_2 = C_4 \cdot -i\tau_2$. C_7 is the over-all charge conjugation matrix, C_4 is the corre-

[7] R. Behrends and A. Sirlin, Phys. Rev. 121, 324 (1961), Table I.

[8] Reference 2, Vol. 2, p. 53.

[9] A. Pais, Phys. Rev. Letters 7, 291 (1961).

sponding matrix for Lorentz space. Thus C_7 is the G-conjugation operator. The 1–4 components of $\Psi\Gamma_\mu\Psi$ are the baryon current and the 5–7 components are the π-meson source. Together, these form a *vector* in a 7-space.

The enlarged current structure is just due to the fact that π is pseudoscalar.[10] If we use the Γ's of Eq. (30) and denote them now as $\Gamma_\alpha^{(7)}$ we get again such a structure, namely, $\Gamma_a = (\gamma_\mu, \gamma_5\Gamma_\alpha^{(7)})$. The Γ_a again satisfy Eq. (1) and $\Psi\Gamma_a\Psi$ is again a *vector*, just because π and K are pseudoscalar. This current consists of baryon current, π and K sources. There exists a corresponding conjugation, enlarging G conjugation (with $K \to -K$) just as G conjugation enlarges charge conjugation. It would be interesting if, along with the enlarged conjugation one could also enlarge the gauge principle.

[10] However, a scalar π with scalar coupling would also be odd under G.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Integral Representations for Production Amplitudes in Perturbation Theory

NOBORU NAKANISHI
Institute for Advanced Study, Princeton, New Jersey
(Received February 9, 1962)

The integral representation proposed previously in the two-particle scattering is extended to the case of one-particle-production amplitudes. Five integral representations are found whose main parts consist of only six terms. Their support properties are investigated for the equal-mass case in every order of perturbation theory. It is shown that lowest-order graphs are by no means the representatives of analyticity in perturbation theory even in the equal-mass case.

I. INTRODUCTION

IN previous works[1,2] we have proposed an integral representation for the two-particle scattering amplitude, which is

$$\int_0^\infty d\alpha \int_0^1 dz \frac{\rho_{12}(\alpha, z)}{\alpha - zs - (1-z)t - i\epsilon} + \int_0^\infty d\beta \int_0^1 dz \frac{\rho_{23}(\beta, z)}{\beta - zt - (1-z)u - i\epsilon} + \int_0^\infty d\gamma \int_0^1 dz \frac{\rho_{31}(\gamma, z)}{\gamma - zu - (1-z)s - i\epsilon} + \int_0^\infty ds' \frac{\rho_1(s')}{s' - s - i\epsilon} + \int_0^\infty dt' \frac{\rho_2(t')}{t' - t - i\epsilon} + \int_0^\infty du' \frac{\rho_3(u')}{u' - u - i\epsilon} + \text{const} \qquad (1.1)$$

n unsubtracted form, where s, t, u are the invariant quares of energy and momentum transfers. This representation has the following properties:

(a) *It can be proved in every order of perturbation theory* for many practical cases. It remains valid even if there appear some scattering-type anomalous thresholds.

(b) If the Mandelstam representation[3] (its validity is not clear yet) is correct, (1.1) can be derived from it.

(c) There are some graphs in which $\alpha \geq 0$, $\beta \geq 0$, and $\gamma \geq 0$ are not assured. But if a single dispersion relation holds, (1.1) always holds for every graph.

(d) Under appropriate support properties, single dispersion relations for s, t, and u (with a limited momentum transfer) and the *partial-wave dispersion relation* are derived from (1.1).

(e) The Mandelstam bounding curves[4] are derived from the support properties of (1.1) for some practical cases.

(f) Contrary to the Mandelstam representation, *this representation can be generalized to production amplitudes.*

The present paper concerns the last item, namely, we investigate concrete forms of integral representa-

[1] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 337 (1961).
[2] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 927 (1961).
[3] S. Mandelstam, Phys. Rev. **112**, 1344 (1958).
[4] S. Mandelstam, Phys. Rev. **115**, 1752 (1959).

sponding matrix for Lorentz space. Thus C_7 is the G-conjugation operator. The 1–4 components of $\Psi\Gamma_\mu\Psi$ are the baryon current and the 5–7 components are the π-meson source. Together, these form a *vector* in a 7-space.

The enlarged current structure is just due to the fact that π is pseudoscalar.[10] If we use the Γ's of Eq. (30) and denote them now as $\Gamma_\alpha^{(7)}$ we get again such a structure, namely, $\Gamma_a = (\gamma_\mu, \gamma_5\Gamma_\alpha^{(7)})$. The Γ_a again satisfy Eq. (1) and $\Psi\Gamma_a\Psi$ is again a *vector*, just because π and K are pseudoscalar. This current consists of baryon current, π and K sources. There exists a corresponding conjugation, enlarging G conjugation (with $K \to -K$) just as G conjugation enlarges charge conjugation. It would be interesting if, along with the enlarged conjugation one could also enlarge the gauge principle.

[10] However, a scalar π with scalar coupling would also be odd under G.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Integral Representations for Production Amplitudes in Perturbation Theory

NOBORU NAKANISHI
Institute for Advanced Study, Princeton, New Jersey
(Received February 9, 1962)

The integral representation proposed previously in the two-particle scattering is extended to the case of one-particle-production amplitudes. Five integral representations are found whose main parts consist of only six terms. Their support properties are investigated for the equal-mass case in every order of perturbation theory. It is shown that lowest-order graphs are by no means the representatives of analyticity in perturbation theory even in the equal-mass case.

I. INTRODUCTION

IN previous works[1,2] we have proposed an integral representation for the two-particle scattering amplitude, which is

$$\int_0^\infty d\alpha \int_0^1 dz \frac{\rho_{12}(\alpha, z)}{\alpha - zs - (1-z)t - i\epsilon} + \int_0^\infty d\beta \int_0^1 dz \frac{\rho_{23}(\beta, z)}{\beta - zt - (1-z)u - i\epsilon} + \int_0^\infty d\gamma \int_0^1 dz \frac{\rho_{31}(\gamma, z)}{\gamma - zu - (1-z)s - i\epsilon} + \int_0^\infty ds' \frac{\rho_1(s')}{s' - s - i\epsilon} + \int_0^\infty dt' \frac{\rho_2(t')}{t' - t - i\epsilon} + \int_0^\infty du' \frac{\rho_3(u')}{u' - u - i\epsilon} + \text{const} \qquad (1.1)$$

n unsubtracted form, where s, t, u are the invariant quares of energy and momentum transfers. This representation has the following properties:

(a) *It can be proved in every order of perturbation theory* for many practical cases. It remains valid even if there appear some scattering-type anomalous thresholds.

(b) If the Mandelstam representation[3] (its validity is not clear yet) is correct, (1.1) can be derived from it.

(c) There are some graphs in which $\alpha \geq 0$, $\beta \geq 0$, and $\gamma \geq 0$ are not assured. But if a single dispersion relation holds, (1.1) always holds for every graph.

(d) Under appropriate support properties, single dispersion relations for s, t, and u (with a limited momentum transfer) and the *partial-wave dispersion relation* are derived from (1.1).

(e) The Mandelstam bounding curves[4] are derived from the support properties of (1.1) for some practical cases.

(f) Contrary to the Mandelstam representation, *this representation can be generalized to production amplitudes.*

The present paper concerns the last item, namely, we investigate concrete forms of integral representa-

[1] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 337 (1961).
[2] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 927 (1961).
[3] S. Mandelstam, Phys. Rev. **112**, 1344 (1958).
[4] S. Mandelstam, Phys. Rev. **115**, 1752 (1959).

tions for the production amplitude having five external lines. We denote the external lines by A, B, C, D, E, and the corresponding external momenta by k_I ($I = A, B, C, D, E$). Then 15 invariant squares are defined by

$$s_I = k_I^2, \qquad (I = A, B, C, D, E),$$
$$s_{IJ} = (k_I + k_J)^2, \qquad (I \neq J), \tag{1.2}$$

where $k^2 = k_0^2 - \mathbf{k}^2$.[5] The first five squares s_I are fixed on mass shells. The remaining ten squares s_{IJ} are of course not independent. Five mutually independent identities hold between them because of conservation law,

$$\sum_{I=A}^{E} k_I = 0. \tag{1.3}$$

We should therefore select five squares among ten. The number of such combinations is ${}_{10}C_5 = 252$. Five squares are, however, not always linearly independent. We shall show in next section that the number of the combinations of independent squares is 162. Therefore the main part of our integral representation would consist of 162 terms. But, fortunately, we may not take such too many terms, because their supports mutually overlap. In Sec. III we shall select appropriate sets R of combinations such that a production amplitude corresponding to any graph can be represented unambiguously in terms of the sum of integrals,

$$\int_{-\infty}^{\infty} d\alpha \int_0^1 dz_1 \cdots \int_0^1 dz_5\, \delta\Big(1 - \sum_{i=1}^{5} z_i\Big) \times \frac{\rho(\alpha, z_1, \cdots, z_5)}{\alpha - \sum_{i=1}^{5} z_i s_i - i\epsilon}, \tag{1.4}$$

over the combinations $\{s_1, s_2, s_3, s_4, s_5\}$ belonging to R (s_i stands for any one of s_{IJ}). We shall find that such a set consists of only 6 combinations, that is to say, the main part of our integral representation contains only six terms. In Sec. IV we shall investigate support properties of (1.4) in the equal-mass case, and show that the weight function $\rho(\alpha, z_i)$ vanishes unless α is larger than some positive value. Section V is devoted to the consideration on additional terms which correspond to the single dispersion terms in (1.1). Some discussions are made on lowest-order graphs in the final section.

Throughout this paper the subtraction problem is not considered.

II. COMBINATIONS OF INDEPENDENT SQUARES

From (1.2) and (1.3) we immediately obtain five identities

$$G_I \equiv \sum_{J \neq I} s_{IJ} - s_I - \sum_{J=A}^{E} s_J = 0 \qquad (I = A, B, C, D, E), \tag{2.1}$$

which are mutually independent. All other identities are obtained as linear combinations of (2.1). For instance,

$$(G_I + G_J + G_K - G_L - G_M)/2 = s_{IJ} + s_{IK} + s_{JK} - s_{LM} - s_I - s_J - s_K = 0, \tag{2.2}$$

where (I, J, K, L, M) is a permutation of (A, B, C, D, E).

Theorem. The identities which involve five or less squares s_{IJ} are (2.1) and (2.2) only.

Proof. An arbitrary identity is written as

$$\sum_{I=A}^{E} a_I G_I = 0, \tag{2.3}$$

a_I being constants. In (2.3) the coefficient of s_{IJ} is $a_I + a_J$. According to the condition stated in the theorem, five coefficients among them must vanish. We classify the types of the combinations of the remaining five coefficients into the following six kinds:

(i)	$a_I + a_J$,	$a_I + a_K$,	$a_I + a_L$,	$a_I + a_M$,	$a_J + a_K$,	30,
(ii)	$a_I + a_J$,	$a_I + a_K$,	$a_I + a_L$,	$a_J + a_K$,	$a_J + a_L$,	30,
(iii)	$a_I + a_J$,	$a_I + a_K$,	$a_I + a_L$,	$a_J + a_K$,	$a_J + a_M$,	60,
(iv)	$a_I + a_J$,	$a_I + a_K$,	$a_I + a_L$,	$a_J + a_K$,	$a_L + a_M$,	60,
(v)	$a_I + a_J$,	$a_I + a_K$,	$a_I + a_L$,	$a_J + a_M$,	$a_K + a_M$,	60,
(vi)	$a_I + a_J$,	$a_I + a_K$,	$a_J + a_L$,	$a_K + a_M$,	$a_L + a_M$,	12.

Here the number of the combinations belonging to each type is written at the right end.

For type (i) the condition is

$$a_J + a_L = a_J + a_M = a_K + a_L = a_K + a_M = a_L + a_M = 0, \tag{2.4}$$

[5] Note the metric is different from that used in previous articles.

which has a nontrivial solution

$$a_I \neq 0, \qquad a_J = a_L = a_K = a_M = 0. \tag{2.5}$$

Likewise for type (iv) we get a nontrivial solution,

$$a_I = a_J = a_K = -a_L = -a_M \neq 0. \tag{2.6}$$

(2.5) and (2.6) correspond to (2.1) and (2.2), respectively. For the other four types we have no nontrivial solutions, Q.E.D.

From the above theorem we see that the combinations of five independent squares are as follows:

$$[A] \quad \{s_{IJ}, s_{IK}, s_{IL}, s_{JK}, s_{JL}\} \quad 30,$$
$$[B] \quad \{s_{IJ}, s_{IK}, s_{IL}, s_{JK}, s_{JM}\} \quad 60,$$
$$[C] \quad \{s_{IJ}, s_{IK}, s_{IL}, s_{JM}, s_{KM}\} \quad 60,$$
$$[D] \quad \{s_{IJ}, s_{IK}, s_{JL}, s_{KM}, s_{LM}\} \quad 12.$$

Here types $[A]$, $[B]$, $[C]$, $[D]$, respectively, correspond to the nonexistence of the solution in types (ii), (iii), (v), (vi). Thus we see that *the total number of the combinations of five independent squares is 162.*

III. INTEGRAL REPRESENTATION

When all particles are scalar, the production amplitude corresponding to a Feynman graph is proportional to[1,6]

$$\int_0^1 dx_1 \cdots \int_0^1 dx_N \frac{\delta(1 - \sum_{i=1}^N x_i)}{U^2(V - i\epsilon)^{N-2n}}, \tag{3.1}$$

with

$$V = \sum_{i=1}^{N} x_i m_i^2 - \sum_{I=A}^{E} \zeta_I s_I - \sum_{I \neq J} \zeta_{IJ} s_{IJ}, \tag{3.2}$$

where m_i is an internal mass, and U, ζ_I, and ζ_{IJ} are nonnegative definite functions of Feynman parameters x_i.

When we take five independent squares shown in the end of the last section, the other five squares are expressed by their linear combinations. Then V is expressed as a linear function of five squares s_{IJ}. We are interested in the coefficients ζ'_{IJ} in this expression.

$[A]$ $\{s_{IJ}, s_{IK}, s_{IL}, s_{JK}, s_{JL}\}$:

$$\begin{aligned}
s_{IM} &= -s_{IJ} - s_{IK} - s_{IL} + s_I + s_0,\\
s_{JM} &= -s_{IJ} - s_{JK} - s_{JL} + s_J + s_0,\\
s_{KL} &= -s_{IJ} - s_{IK} - s_{IL} - s_{JK} - s_{JL} - s_M + 2s_0,\\
s_{KM} &= s_{IJ} + s_{IL} + s_{JL} - s_I - s_J - s_L,\\
s_{LM} &= s_{IJ} + s_{IK} + s_{JK} - s_I - s_J - s_K
\end{aligned} \tag{3.3}$$

[6] Y. Nambu, Nuovo cimento **6**, 1064 (1957); K. Symanzik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958).

with $s_0 \equiv \sum_{I=A}^{E} s_I$. Hence the coefficients are as follows:

$$\begin{aligned}
\zeta'_{IJ} &= \zeta_{IJ} + \zeta_{KM} + \zeta_{LM} - \zeta_{IM} - \zeta_{JM} - \zeta_{KL},\\
\zeta'_{IK} &= \zeta_{IK} + \zeta_{LM} - \zeta_{IM} - \zeta_{KL},\\
\zeta'_{IL} &= \zeta_{IL} + \zeta_{KM} - \zeta_{IM} - \zeta_{KL},\\
\zeta'_{JK} &= \zeta_{JK} + \zeta_{LM} - \zeta_{JM} - \zeta_{KL},\\
\zeta'_{JL} &= \zeta_{JL} + \zeta_{KM} - \zeta_{JM} - \zeta_{KL}.
\end{aligned} \tag{3.4}$$

$[B]$ $\{s_{IJ}, s_{IK}, s_{IL}, s_{JK}, s_{JM}\}$:

$$\begin{aligned}
\zeta'_{IJ} &= \zeta_{IJ} + \zeta_{LM} - \zeta_{IM} - \zeta_{JL},\\
\zeta'_{IK} &= \zeta_{IK} + \zeta_{LM} - \zeta_{IM} - \zeta_{KL},\\
\zeta'_{IL} &= \zeta_{IL} + \zeta_{KM} - \zeta_{IM} - \zeta_{KL},\\
\zeta'_{JK} &= \zeta_{JK} + \zeta_{LM} - \zeta_{JL} - \zeta_{KM},\\
\zeta'_{JM} &= \zeta_{JM} + \zeta_{KL} - \zeta_{JL} - \zeta_{KM}.
\end{aligned} \tag{3.5}$$

$[C]$ $\{s_{IJ}, s_{IK}, s_{IL}, s_{JM}, s_{KM}\}$:

$$\begin{aligned}
\zeta'_{IJ} &= \zeta_{IJ} + \zeta_{LM} - \zeta_{IM} - \zeta_{JL},\\
\zeta'_{IK} &= \zeta_{IK} + \zeta_{LM} - \zeta_{IM} - \zeta_{KL},\\
\zeta'_{IL} &= \zeta_{IL} + \zeta_{JK} + \zeta_{LM} - \zeta_{IM} - \zeta_{JL} - \zeta_{KL},\\
\zeta'_{JM} &= \zeta_{JM} + \zeta_{KL} - \zeta_{JK} - \zeta_{LM},\\
\zeta'_{KM} &= \zeta_{KM} + \zeta_{JL} - \zeta_{JK} - \zeta_{LM},
\end{aligned} \tag{3.6}$$

$[D]$ $\{s_{IJ}, s_{IK}, s_{JL}, s_{KM}, s_{LM}\}$:

$$\begin{aligned}
\zeta'_{IJ} &= \zeta_{IJ} + \zeta_{KL} - \zeta_{IL} - \zeta_{JK},\\
\zeta'_{IK} &= \zeta_{IK} + \zeta_{JM} - \zeta_{IM} - \zeta_{JK},\\
\zeta'_{JL} &= \zeta_{JL} + \zeta_{IM} - \zeta_{IL} - \zeta_{JM},\\
\zeta'_{KM} &= \zeta_{KM} + \zeta_{IL} - \zeta_{IM} - \zeta_{KL},\\
\zeta'_{LM} &= \zeta_{LM} + \zeta_{JK} - \zeta_{JM} - \zeta_{KL}.
\end{aligned} \tag{3.7}$$

Now in order to rewrite (3.1) as the form of (1.4), we must require for these five coefficients to be nonnegative. This requirement imposes some inequalities upon ζ_{IJ}. For example, for a combination $\{s_{IJ}, s_{IK}, s_{IL}, s_{JK}, s_{JL}\}$ the amplitude (3.1) can be expressed as the integral (1.4) only when

$$\begin{aligned}
\zeta_{IJ} + \zeta_{KM} + \zeta_{LM} &\geq \zeta_{KL} + \zeta_{IM} + \zeta_{JM},\\
\zeta_{IK} + \zeta_{LM} &\geq \zeta_{KL} + \zeta_{IM},\\
\zeta_{IL} + \zeta_{KM} &\geq \zeta_{KL} + \zeta_{IM},\\
\zeta_{JK} + \zeta_{LM} &\geq \zeta_{KL} + \zeta_{JM},\\
\zeta_{JL} + \zeta_{KM} &\geq \zeta_{KL} + \zeta_{JM}.
\end{aligned} \tag{3.8}$$

Therefore, in order to cover all possibilities we must consider a set R of combinations. We denote the set of all the combinations of type $[J]$ by $R[J]$ $(J = A, B, C, D)$. Then we shall see in the Appendix that $R[A]$ covers all possibilities fivefold, and $R[B]$ does fourfold; on the other hand $R[C]$ or $R[D]$ does not cover all possibilities, and the former covers some parts twofold.

We have thus seen that the desired integral representation cannot be symmetric with respect to the five external lines. In order to cover all possibilities without overlapping (except for boundary points), it is necessary to treat at least one external line asymmetrically.

By putting

$$2\zeta_{IJ}^{M} \equiv 2\zeta_{IJ} - \zeta_{IM} - \zeta_{JM} + \zeta_{KM} + \zeta_{LM}, \tag{3.9}$$

(3.8) is rewritten simply as

$$\min\,[\zeta_{IJ}^{M}, \zeta_{IK}^{M}, \zeta_{IL}^{M}, \zeta_{JK}^{M}, \zeta_{JL}^{M}] \geq \zeta_{KL}^{M}. \tag{3.10}$$

Hence, for example, if we take the set of six combinations

$$\{s_{AC}, s_{AD}, s_{BC}, s_{BD}, s_{CD}\},$$
$$\{s_{AB}, s_{AD}, s_{BC}, s_{BD}, s_{CD}\},$$
$$\{s_{AB}, s_{AC}, s_{BC}, s_{BD}, s_{CD}\},$$
$$\{s_{AB}, s_{AC}, s_{AD}, s_{BD}, s_{CD}\},$$
$$\{s_{AB}, s_{AC}, s_{AD}, s_{BC}, s_{CD}\},$$
$$\{s_{AB}, s_{AC}, s_{AD}, s_{BC}, s_{BD}\} \tag{3.11}$$

as R, it covers all possibilities without overlapping, because these combinations correspond to the cases $\zeta^{E} = \zeta_{AB}^{E}$, $\zeta^{E} = \zeta_{AC}^{E}$, $\zeta^{E} = \zeta_{AD}^{E}$, $\zeta^{E} = \zeta_{BC}^{E}$, $\zeta^{E} = \zeta_{BD}^{E}$, $\zeta^{E} = \zeta_{CD}^{E}$, respectively, where

$$\zeta^{E} \equiv \min\,[\zeta_{AB}^{E}, \zeta_{AC}^{E}, \zeta_{AD}^{E}, \zeta_{BC}^{E}, \zeta_{BD}^{E}, \zeta_{CD}^{E}]. \tag{3.12}$$

In the above we may take any one of A, B, C, D instead of E as the asymmetrically treated external line. *We have thus obtained five integral representations which consist of only six terms*:

$$\int_{-\infty}^{\infty} d\alpha \int_{0}^{1} dz_1 \cdots \int_{0}^{1} dz_6 \frac{\delta(1 - \sum_{i=1}^{6} z_i)\{\sum_{i=1}^{6} \delta(z_i)\}\rho(\alpha, z_1, \cdots, z_6)}{\alpha - z_1 s_{IJ} - z_2 s_{IK} - z_3 s_{IL} - z_4 s_{JK} - z_5 s_{JL} - z_6 s_{KL} - i\epsilon}. \tag{3.13}$$

It may be noteworthy that an identity

$$(G_I + G_J + G_K + G_L - G_M)/2 = s_{IJ} + s_{IK} + s_{IL} + s_{JK} + s_{JL} + s_{KL} - 2s_0 + s_M = 0, \tag{3.14}$$

which is similar to the identity of the two-particle scattering case, holds between the six squares appearing in (3.13).

IV. SUPPORT PROPERTIES IN THE EQUAL-MASS CASE

It is a very important problem to find support properties of the weight functions $\rho(\alpha, z_i)$. We expect that for practical cases they vanish unless $\alpha > 0$. For simplicity we will show this only for the equal-mass case, which is completely symmetric with respect to all the external lines.

For definiteness we consider a term (1.4) in which s_1, s_2, s_3, s_4, s_5 are identified with $s_{AB}, s_{AC}, s_{AD}, s_{BC}, s_{BD}$. The denominator function V is rewritten as

$$V = \sum_{i=1}^{N} x_i\mu^2 - \sum_{I=A}^{E} \zeta_I'\,\mu^2 - \zeta_{AB}'s_{AB} - \zeta_{AC}'s_{AC} - \zeta_{AD}'s_{AD} - \zeta_{BC}'s_{BC} - \zeta_{BD}'s_{BD}, \tag{4.1}$$

with

$$\begin{aligned}
\zeta_A' &\equiv \zeta_A + 2\zeta_{CD} + 2\zeta_{AE} + \zeta_{BE} - \zeta_{CE} - \zeta_{DE},\\
\zeta_B' &\equiv \zeta_B + 2\zeta_{CD} + \zeta_{AE} + 2\zeta_{BE} - \zeta_{CE} - \zeta_{DE},\\
\zeta_C' &\equiv \zeta_C + 2\zeta_{CD} + \zeta_{AE} + \zeta_{BE} - \zeta_{DE},\\
\zeta_D' &\equiv \zeta_D + 2\zeta_{CD} + \zeta_{AE} + \zeta_{BE} - \zeta_{CE},\\
\zeta_E' &\equiv \zeta_E + \zeta_{CD} + \zeta_{AE} + \zeta_{BE},\\
\zeta_{AB}' &\equiv \zeta_{AB} - \zeta_{CD} - \zeta_{AE} - \zeta_{BE} + \zeta_{CE} + \zeta_{DE} \geq 0,\\
\zeta_{AC}' &\equiv \zeta_{AC} - \zeta_{CD} - \zeta_{AE} + \zeta_{DE} \geq 0,\\
\zeta_{AD}' &\equiv \zeta_{AD} - \zeta_{CD} - \zeta_{AE} + \zeta_{CE} \geq 0,\\
\zeta_{BC}' &\equiv \zeta_{BC} - \zeta_{CD} - \zeta_{BE} + \zeta_{DE} \geq 0,\\
\zeta_{BD}' &\equiv \zeta_{BD} - \zeta_{CD} - \zeta_{BE} + \zeta_{CE} \geq 0.
\end{aligned} \tag{4.2}$$

The integration variables are defined by

$$\alpha = \left(\sum_{i=1}^{N} x_i\mu^2 - \sum_{I=A}^{E} \zeta_I'\,\mu^2\right)\Big/\xi,$$
$$z_i = \zeta_i/\xi, \quad (i = AB, AC, AD, BC, BD) \tag{4.3}$$

with

$$\xi \equiv \zeta_{AB}' + \zeta_{AC}' + \zeta_{AD}' + \zeta_{BC}' + \zeta_{BD}'. \tag{4.4}$$

Now, in order to get a lower bound of α we make use of the result obtained in a previous work.[7] Consider an arbitrary graph of the two-particle

[7] N. Nakanishi, Suppl. Progr. Theoret. Phys. (Kyoto) No. 18, 1 (1961).

scattering which contains no one-particle intermediate states. Let k_1 and k_2 be such momenta that

$$k_1^2 = k_2^2 = \mu^2, \qquad k_1 k_2 = 0. \tag{4.5}$$

Then it was proved that

$$V \geq 0 \tag{4.6}$$

when the four external momenta are given by $k_1 + k_2$, $k_1 - k_2$, $-k_1 + k_2$, $-k_1 - k_2$.

In the present case, if one of the five external momenta is put equal to zero, we can make use of the above result. For example, if we put

$$k_A = k_1 + k_2, \quad k_B = -k_1 - k_2, \quad k_C = k_1 - k_2, \quad k_D = -k_1 + k_2, \quad k_E = 0, \tag{4.7}$$

then we have an inequality

$$\sum x_i\mu^2 - (\zeta'_A + \zeta'_B + \zeta'_C + \zeta'_D)\cdot 2\mu^2 - (\zeta'_{AC} + \zeta'_{AD} + \zeta'_{BC} + \zeta'_{BD})\cdot 4\mu^2 \geq 0. \tag{4.8}$$

Likewise we get other fourteen inequalities by identifying k_A, k_B, k_C, k_D, k_E to $k_1 + k_2$, $-k_1 - k_2$, $k_1 - k_2$, $-k_1 + k_2$ in various ways. Adding these 15 inequalities, we obtain

$$\sum_{i=1}^{N} x_i\mu^2 - \tfrac{8}{5}\sum_{I=A}^{E} \zeta'_I\,\mu^2 - \tfrac{12}{5}\xi\mu^2 \geq 0. \tag{4.9}$$

Unlike the case of the two-particle scattering, $\sum_{I=A}^{E} \zeta'_I$ is not nonnegative definite. So we can deduce only

$$\alpha \geq \tfrac{3}{2}\mu^2 \tag{4.10}$$

from (4.9).

If we add the 12 inequalities which correspond to $k_E \neq 0$ among the above 15 ones, we have

$$\sum_{i=1}^{N} x_i\mu^2 - \tfrac{3}{2}\left(\sum_{I=A}^{E} \zeta'_I + \tfrac{1}{3}\zeta'_E\right)\mu^2 - \tfrac{7}{8}\xi\mu^2 \geq 0 \tag{4.11}$$

instead of (4.9). Since $\zeta'_E \geq 0$ [see (4.2)], (4.11) leads to

$$\alpha \geq \tfrac{14}{9}\mu^2, \tag{4.12}$$

that is a little better than (4.10).

V. ADDITIONAL TERMS

In (1.1) the three single dispersion terms are the contributions from the graphs which consist of two vertex parts. These terms can be included in the main part if the weight functions contain a term proportional to $\delta(z)$. But it is usually convenient to write them separately as in (1.1).

In case of the production amplitude such additional terms appear from the graphs shown in Fig. 1.

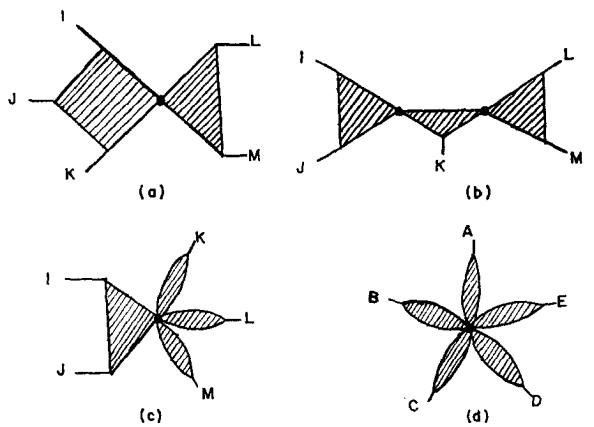


FIG. 1. Graphs corresponding to additional terms of the integral representation.

A. Three-Variable Terms

In Fig. 1(a) there are no intermediate states corresponding to IL, IM, JL, JM, KL, KM. Therefore,

$$\zeta_{IL} = \zeta_{IM} = \zeta_{JL} = \zeta_{JM} = \zeta_{KL} = \zeta_{KM} = 0. \tag{5.1}$$

So, using the identity (2.2) we get

$$V = \sum x_i m_i^2 - (\zeta_I - \zeta_{LM})s_I - (\zeta_J - \zeta_{LM})s_J - (\zeta_K - \zeta_{LM})s_K - \zeta_L s_L - \zeta_M s_M - (\zeta_{IJ} + \zeta_{LM})s_{IJ} - (\zeta_{IK} + \zeta_{LM})s_{IK} - (\zeta_{JK} + \zeta_{LM})s_{JK}. \tag{5.2}$$

Therefore, the contributions from Fig. 1(a) are represented as a sum of ten integrals

$$\int_{-\infty}^{\infty} d\alpha \int_0^1 dz_1 \int_0^1 dz_2 \int_0^1 dz_3\, \delta\left(1 - \sum_{i=1}^{3} z_i\right) \times \frac{\sigma(\alpha, z_i)}{\alpha - z_1 s_{IJ} - z_2 s_{IK} - z_3 s_{JK} - i\epsilon}. \tag{5.3}$$

If no one-particle line is present, in the equal-mass case we get

$$\alpha \geq \tfrac{5}{3}\mu^2 \tag{5.4}$$

by the method stated in the last section (add the 9 inequalities which correspond to $k_L \neq 0$ and $k_M \neq 0$). The pole terms are $1/(\mu^2 - s_{LM} - i\epsilon)$ multiplied by the three main terms of (1.1) in which s, t, u are replaced by s_{IJ}, s_{IK}, s_{JK}. Their support properties are given in previous works.[1,2]

B. Two-Variable Terms

In Fig. 1(b) we have

$$\zeta_{IK} = \zeta_{JK} = 0 \tag{5.5}$$

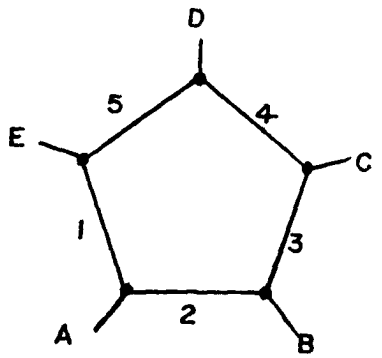


FIG. 2. Lowest-order graph of the production amplitude.

in addition to (5.1). Hence the integral representation is a sum of 15 integrals

$$\int_{-\infty}^{\infty} d\alpha \int_0^1 dz \frac{\varphi(\alpha, z)}{\alpha - zs_{IJ} - (1-z)s_{LM} - i\epsilon}. \tag{5.6}$$

For nonpole contributions, we obtain

$$\alpha \geq \tfrac{12}{5}\mu^2 \tag{5.7}$$

from (4.9) in the equal-mass case. If one-particle lines are present, we have 30 single-pole terms

$$\frac{1}{\mu^2 - s_{IJ} - i\epsilon} \int_{4\mu^2}^{\infty} d\alpha \frac{\varphi'(\alpha)}{\alpha - s_{LM} - i\epsilon}, \tag{5.8}$$

and 15 double-pole terms

$$\frac{\text{const}}{(\mu^2 - s_{IJ} - i\epsilon)(\mu^2 - s_{LM} - i\epsilon)}. \tag{5.9}$$

C. One-Variable Terms

For the graphs shown in Fig. 1(c) we have 10 single dispersion terms

$$\frac{1}{\mu^2 - s_{IJ} - i\epsilon} + \int_{4\mu^2}^{\infty} d\alpha \frac{\psi(\alpha)}{\alpha - s_{IJ} - i\epsilon} \tag{5.10}$$

in the equal-mass case.

D. Constant Term

A constant term can appear from the graphs shown in Fig. 1(d). If renormalizable interactions only are taken, this term will not appear.

VI. DISCUSSION

We have obtained integral representations for the production amplitude whose main parts consist of only six terms. As the simplest example we will consider a nontrivial lowest-order graph shown in Fig. 2.

The denominator function for this graph is

$$\begin{aligned} V = {} & \sum_{i=1}^{5} x_i m_i^2 - x_1x_2s_A - x_2x_3s_B - x_3x_4s_C \\ & - x_4x_5s_D - x_5x_1s_E - x_1x_2s_{AB} - x_2x_4s_{BC} \\ & - x_3x_5s_{CD} - x_4x_1s_{DE} - x_5x_2s_{AE}. \end{aligned} \tag{6.1}$$

If s_{AB}, s_{BC}, s_{CD}, s_{DE}, s_{AE} are adopted as independent squares, the contribution from this graph will be represented by (1.4) with

$$\alpha = \frac{\sum x_i m_i^2 - x_1x_2s_A - x_2x_3s_B - x_3x_4s_C - x_4x_5s_D - x_5x_1s_E}{x_1x_3 + x_2x_4 + x_3x_5 + x_4x_1 + x_5x_2}. \tag{6.2}$$

Even in the equal-mass case, $\alpha \geq (2\mu)^2$ does not hold, or more precisely,

$$\min \alpha = (2 + \sqrt{3})\mu^2. \tag{6.3}$$

Now, the combination $\{s_{AB}, s_{BC}, s_{CD}, s_{DE}, s_{AE}\}$ belongs to $R[D]$ of Sec. III. We have stated there that the 12 combinations of $R[D]$ cannot cover all possibilities. This means that any superposition of these lowest-order terms is not sufficient to represent general terms [for example, the term corresponding to Fig. 3(d)].[8]

On the other hand, our integral representations can naturally describe the term corresponding to Fig. 2. Indeed, (6.1) is rewritten as

$$\begin{aligned} V = {} & \sum_{i=1}^{5} x_i m_i^2 - (x_1x_2 - x_4x_1)s_A \\ & - (x_2x_3 - x_4x_1 - x_5x_2)s_B \\ & - (x_3x_4 - x_4x_1 - x_5x_2)s_C \\ & - (x_4x_5 - x_5x_2)s_D - x_5x_1s_E \\ & - (x_1x_3 + x_4x_1)s_{AB} - x_4x_1s_{AC} \\ & - (x_2x_4 + x_4x_1 + x_5x_2)s_{BC} \\ & - x_5x_2s_{BD} - (x_3x_5 + x_5x_2)s_{CD}, \end{aligned} \tag{6.4}$$

in which the coefficients of s_{AB}, s_{AC}, s_{BC}, s_{BD}, s_{CD} are nonnegative.

Some authors[9] conjectured that lowest-order graphs (i.e., one-loop graphs) would be the representatives of analyticity of general terms. It is evident from the above consideration, however, that *this conjecture is not true even in the equal-mass case.*

[8] Added note: This reasoning may be not completely rigorous because the uniqueness of the weight functions in our integral representation (3.13) is not proved yet.

[9] R. J. Eden, Phys. Rev. 119, 1763 (1960). L. F. Cook, Jr. and J. Tarski, J. Math. Phys. 3, 1 (1961).

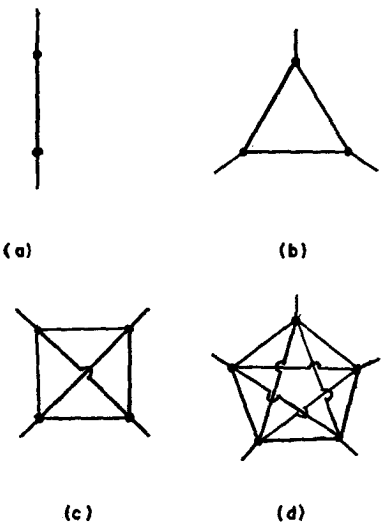


FIG. 3. Simplest symmetric graphs.

Some special counterexamples exist also for the two-particle scattering.[1,10]

So we conjecture that the representative graphs will be those shown in Fig. 3 rather than lowest-order ones, namely, any term in perturbation theory will be represented by an internal-mass superposition of the terms corresponding to Fig. 3. This conjecture gives the Umezawa-Kamefuchi-Källén-Lehmann representation for the one-body propagator. For the vertex function it will be consistent with the Källén-Toll[11] integral representation because the f_0 term is nonperturbational. No counterexamples against it are found for the scattering or production amplitudes.

ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Professor J. Robert Oppenheimer for the hospitality extended to him at the Institute for Advanced Study and to the National Science Foundation for financial support.

APPENDIX

We will analyze the conditions $\zeta' \geq 0$ in (3.4)–(3.7). For this purpose it is convenient to introduce five quantities

$$\theta_M \equiv \min\,(\zeta_{IJ} + \zeta_{KL},\ \zeta_{IK} + \zeta_{JL},\ \zeta_{IL} + \zeta_{JK}), \quad \text{(A1)}$$

where (I, J, K, L, M) is a permutation of (A, B, C, D, E) as in the text. Any possibility must exclusively belong to only one of the $243 (= 3^5)$ classes which are classified into the following seven types:

(a)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IJ} + \zeta_{KM}$, $\theta_K = \zeta_{IJ} + \zeta_{LM}$, $\theta_J = \zeta_{IM} + \zeta_{KL}$, $\theta_I = \zeta_{JM} + \zeta_{KL}$;	15.
(b)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IJ} + \zeta_{KM}$, $\theta_K = \zeta_{IJ} + \zeta_{LM}$, $\theta_J = \zeta_{IM} + \zeta_{KL}$, $\theta_I = \zeta_{JL} + \zeta_{KM}$;	60.
(c)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IJ} + \zeta_{KM}$, $\theta_K = \zeta_{IL} + \zeta_{JM}$, $\theta_J = \zeta_{IL} + \zeta_{KM}$, $\theta_I = \zeta_{JM} + \zeta_{KL}$;	12.
(d)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IJ} + \zeta_{KM}$, $\theta_K = \zeta_{IL} + \zeta_{JM}$, $\theta_J = \zeta_{IL} + \zeta_{KM}$, $\theta_I = \zeta_{JK} + \zeta_{LM}$;	60.
(e)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IJ} + \zeta_{KM}$, $\theta_K = \zeta_{IL} + \zeta_{JM}$, $\theta_J = \zeta_{IM} + \zeta_{KL}$, $\theta_I = \zeta_{JK} + \zeta_{LM}$;	60.
(f)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IJ} + \zeta_{KM}$, $\theta_K = \zeta_{IL} + \zeta_{JM}$, $\theta_J = \zeta_{IK} + \zeta_{LM}$, $\theta_I = \zeta_{JK} + \zeta_{LM}$;	30.
(g)	$\theta_M = \zeta_{IJ} + \zeta_{KL}$, $\theta_L = \zeta_{IK} + \zeta_{JM}$, $\theta_K = \zeta_{IM} + \zeta_{JL}$, $\theta_J = \zeta_{IL} + \zeta_{KM}$, $\theta_I = \zeta_{JK} + \zeta_{LM}$;	6.

Here the number of the classes belonging to each type is written at the right end.

In the above, types (d), (e), (f), (g) actually do not occur. For example, consider type (d). On

[10] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. 2, 656 (1961).
[11] G. Källén and J. Toll, Helv. Phys. Acta 33, 753 (1960).

account of (A1) we have

$$\zeta_{IJ} + \zeta_{KL} \leq \zeta_{IL} + \zeta_{JK},$$
$$\zeta_{IL} + \zeta_{JM} \leq \zeta_{IJ} + \zeta_{LM}, \tag{A2}$$
$$\zeta_{JK} + \zeta_{LM} \leq \zeta_{JM} + \zeta_{KL},$$

which lead to inconsistency except for boundary points. Types (e), (f), (g) are likewise self-inconsistent. Therefore, we have only to consider types (a), (b), (c), which are indeed nontrivial. For the sake of later use, we write examples of these types in the following:

(a)
$$\zeta_{AB} + \zeta_{CD} \leq \min[\zeta_{AC} + \zeta_{BD}, \zeta_{AD} + \zeta_{BC}],$$
$$\zeta_{AB} + \zeta_{CE} \leq \min[\zeta_{AC} + \zeta_{BE}, \zeta_{AE} + \zeta_{BC}],$$
$$\zeta_{AB} + \zeta_{DE} \leq \min[\zeta_{AD} + \zeta_{BE}, \zeta_{AE} + \zeta_{BD}],$$
$$\zeta_{AE} + \zeta_{CD} \leq \min[\zeta_{AC} + \zeta_{DE}, \zeta_{AD} + \zeta_{CE}],$$
$$\zeta_{BE} + \zeta_{CD} \leq \min[\zeta_{BC} + \zeta_{DE}, \zeta_{BD} + \zeta_{CE}]; \tag{A3}$$

(b)
$$\zeta_{AB} + \zeta_{CD} \leq \min[\zeta_{AC} + \zeta_{BD}, \zeta_{AD} + \zeta_{BC}],$$
$$\zeta_{AB} + \zeta_{CE} \leq \min[\zeta_{AC} + \zeta_{BE}, \zeta_{AE} + \zeta_{BC}],$$
$$\zeta_{AB} + \zeta_{DE} \leq \min[\zeta_{AD} + \zeta_{BE}, \zeta_{AE} + \zeta_{BD}],$$
$$\zeta_{AE} + \zeta_{CD} \leq \min[\zeta_{AC} + \zeta_{DE}, \zeta_{AD} + \zeta_{CE}],$$
$$\zeta_{BD} + \zeta_{CE} \leq \min[\zeta_{BC} + \zeta_{DE}, \zeta_{BE} + \zeta_{CD}]; \tag{A4}$$

(c)
$$\zeta_{AB} + \zeta_{CD} \leq \min[\zeta_{AC} + \zeta_{BD}, \zeta_{AD} + \zeta_{BC}],$$
$$\zeta_{AB} + \zeta_{CE} \leq \min[\zeta_{AC} + \zeta_{BE}, \zeta_{AE} + \zeta_{BC}],$$
$$\zeta_{AD} + \zeta_{BE} \leq \min[\zeta_{AB} + \zeta_{DE}, \zeta_{AE} + \zeta_{BD}],$$
$$\zeta_{AD} + \zeta_{CE} \leq \min[\zeta_{AC} + \zeta_{DE}, \zeta_{AE} + \zeta_{CD}],$$
$$\zeta_{BE} + \zeta_{CD} \leq \min[\zeta_{BC} + \zeta_{DE}, \zeta_{BD} + \zeta_{CE}]. \tag{A5}$$

On the other hand, as is easily seen, the conditions $\zeta' \geq 0$ in (3.4)–(3.7) are rewritten as follows:

[A]
$$\zeta_{IM} + \zeta_{KL} \leq \min[\zeta_{IK} + \zeta_{LM}, \zeta_{IL} + \zeta_{KM}],$$
$$\zeta_{JM} + \zeta_{KL} \leq \min[\zeta_{JK} + \zeta_{LM}, \zeta_{JL} + \zeta_{KM}],$$
$$\zeta_{KL} + \zeta_{IM} + \zeta_{JM} \leq \zeta_{IJ} + \zeta_{KM} + \zeta_{LM}. \tag{A6}$$

[B]
$$\zeta_{IM} + \zeta_{JL} \leq \min[\zeta_{IJ} + \zeta_{LM}, \zeta_{IL} + \zeta_{JM}],$$
$$\zeta_{IM} + \zeta_{KL} \leq \min[\zeta_{IL} + \zeta_{KM}, \zeta_{JK} + \zeta_{LM}],$$
$$\zeta_{JL} + \zeta_{KM} \leq \min[\zeta_{JK} + \zeta_{LM}, \zeta_{JM} + \zeta_{KL}]. \tag{A7}$$

[C]
$$\zeta_{IM} + \zeta_{JK} \leq \min[\zeta_{IJ} + \zeta_{KM}, \zeta_{IK} + \zeta_{JM}],$$
$$\zeta_{IM} + \zeta_{JL} \leq \min[\zeta_{IJ} + \zeta_{LM}, \zeta_{IL} + \zeta_{JM}],$$
$$\zeta_{IM} + \zeta_{KL} \leq \min[\zeta_{IK} + \zeta_{LM}, \zeta_{IL} + \zeta_{KM}],$$
$$\zeta_{JK} + \zeta_{LM} \leq \min[\zeta_{JM} + \zeta_{KL}, \zeta_{JL} + \zeta_{KM}],$$
$$\zeta_{IM} + \zeta_{JL} + \zeta_{KL} \leq \zeta_{JK} + \zeta_{IL} + \zeta_{LM}. \tag{A8}$$

[D]
$$\zeta_{IL} + \zeta_{JK} \leq \min[\zeta_{IJ} + \zeta_{KL}, \zeta_{IK} + \zeta_{JL}],$$
$$\zeta_{IM} + \zeta_{JK} \leq \min[\zeta_{IK} + \zeta_{JM}, \zeta_{IJ} + \zeta_{KM}],$$
$$\zeta_{IL} + \zeta_{JM} \leq \min[\zeta_{IM} + \zeta_{JL}, \zeta_{IJ} + \zeta_{LM}],$$
$$\zeta_{IM} + \zeta_{KL} \leq \min[\zeta_{IL} + \zeta_{KM}, \zeta_{IK} + \zeta_{LM}],$$
$$\zeta_{JM} + \zeta_{KL} \leq \min[\zeta_{JK} + \zeta_{LM}, \zeta_{JL} + \zeta_{KM}]. \tag{A9}$$

Comparing (A6)–(A9) with (A3)–(A5), we obtain the result that $R[A]$, $R[B]$, $R[C]$, $R[D]$ cover each class of types (a), (b), (c) with the following multiplicities:

R	(a)	(b)	(c)
$R[A]$	5	5	5
$R[B]$	4	4	5
$R[C]$	2	2	0
$R[D]$	0	0	1

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Atomic Many-Body Problem. II. The Matrix Components of the Hamiltonian with Respect to Correlated Wave Functions

LEVENTE SZÁSZ
Solid-State and Molecular Theory Group, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received April 18, 1962)

The wave function for an atom with N electrons in arbitrary configuration will be written in the form

$$\Psi = \Psi_0 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(ij),$$

where Ψ_0 is a Slater determinant and $f(ij)$ is the antisymmetrized product of $(N-2)$ one-electron spin-orbitals and one 2-electron function $\Phi(ij/12)$. The correlation between the two spin-orbitals φ_i and φ_j can be taken into account by introducing r_{12} (the interelectronic distance) explicitly into the 2-electron function Φ. The purpose of the paper is to analyze the structure of the matrix components of the Hamiltonian with respect to the wave function given above. Starting from exact, general formulas for the matrix components it will be shown that, all integrals which occur in the diagonal, as well as in the nondiagonal matrix components can be reduced to six basic integrals which are 2- , 3- , and 4-electron integrals, containing interelectronic distances. It will be indicated that, five of the six basic integrals can be calculated in closed form whereas one of them, (an exchange integral) can be given only in the form of an infinite series.

1. FORMULATION OF THE PROBLEM

THE general theory of correlated wave functions was developed in two previous papers by this writer.[1,2] The purpose of this theory is to develop a method for the calculation of electronic wave functions which may depend explicitly on the distances between the electrons. In A we have suggested a variational procedure which enables us to calculate all 2-electron correlations in a system of electrons; in I we extended the theory to include 3-electron, $\cdots$ N-electron correlations.

In the present paper we restrict ourselves to the discussion of the 2-electron approximation.[3] The basic idea of this approximation is to set up the approximate solution for the Schrödinger equation in the following form[4]:

$$\Psi = \Psi_0 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(ij), \tag{1.1}$$

where Ψ_0 is a Slater determinant and $f(ij)$ is a correlated wave function of second order that is defined as follows[5]:

$$f(ij) = \frac{1}{(N!)^{1/2}} \tilde{A}\{\varphi_1(q_1) \cdots \varphi_{i-1}(q_{i-1})\varphi_{i+1}(q_{i+1}) \cdots \varphi_{j-1}(q_{j-1})\varphi_{j+1}(q_{j+1}) \cdots \varphi_N(q_N)\Phi(ij \mid q_i q_j)\}, \tag{1.2}$$

where $\varphi_1 \cdots \varphi_N$ are one-electron spin-orbitals,[6] $\Phi(ij \mid q_1 q_2)$ is a 2-electron function,[7] q_i stands for the spatial and spin coordinates of the ith electron, and $\tilde{A}$ is an antisymmetry operator which makes the function (1.2) antisymmetric with respect to all coordinates after Φ has been antisymmetrized separately. For Φ we put

$$\Phi(ij \mid q_1 q_2) = [\varphi_i(q_1)\varphi_j(q_2) - \varphi_i(q_2)\varphi_j(q_1)] \times \sum_{\alpha\beta\gamma} c(ij\alpha\beta\gamma)(r_1^\alpha r_2^\beta + r_1^\beta r_2^\alpha) r_{12}^\gamma, \tag{1.3}$$

where r_1 and r_2 are the distances of the electrons from the nucleus, r_{12} is the distance between the two electrons, $c(ij\alpha\beta\gamma)$ is a variational parameter, and α, β, γ are positive integers or zero.

It is known that if we put (1.3) into (1.2) and the resulting expression into the Schrödinger energy integral, the variation of the energy integral with respect to the variational parameters $c(ij\alpha\beta\gamma)$ leads to a secular equation. The elements of the matrix

[1] L. Szász, Z. Naturforsch. **15a**, 909 (1960); hereafter referred to as A.

[2] L. Szász, Phys. Rev. **126**, 169 (1962); hereafter referred to as I. A list of references related to the theory of correlated wave functions can be found in this paper. The most important references are the following: V. Fock, M. Wesselow, and M. Petrashen, J. Exptl. Theoret. Phys. (U.S.S.R.) **10**, 723 (1940); M. Wesselow, M. Petrashen, and A. Kritschagina, *ibid.* **10**, 857 (1940); M. Wesselow and M. Petrashen, *ibid.* **10**, 1172 (1940); A. P. Jucys, *ibid.* **23**, 357, 371 (1952); L. Szász, Z. Naturforsch. **14a**, 1014 (1959); J. Chem. Phys. **35**, 1072 (1961).

[3] This phrase was introduced in I. See Secs. 2 and 5.

[4] See Eqs. (5), (10), and (19) in A; Eq. (5.6) in I.

[5] The definition of a correlated wave function of nth order $(2 \leq n \leq N)$ was given by Eq. (2.8) of I. This is a special case of that definition for $n = 2$.

[6] Several different notations will be used to denote the one-electron spin-orbitals. We note here that $\varphi_i \equiv \varphi_i(q)$; $\varphi_i(q_k) \equiv \varphi_i(k)$; $\varphi_i(q_k) \equiv (\varphi_i|q_k) \equiv \varphi(i|k)$.

[7] We shall use also the condensed notation $\Phi(ij|q_1q_2) \equiv \Phi(ij|12)$.

of the secular equation are the matrix components of the Hamiltonian with respect to correlated wave functions. In the present article and in a forthcoming paper,[8] our purpose is to develop the mathematical technique for the calculation of the matrix components of the Hamiltonian, with respect to correlated wave functions. We consider correlated wave functions of the form (1.2) with 2-electron functions which have the form given by (1.3). In the present paper our goal is to derive some theorems on the structure of the matrix components. It will be shown that all matrix components can be reduced to certain simple basic integrals. In a subsequent communication we shall establish the mathematical technique for the calculation of these basic integrals. The results of this paper as well as those of the forthcoming one are valid for atoms with an arbitrary number of electrons and for any configuration.

2. THE DIAGONAL MATRIX COMPONENTS OF THE HAMILTONIAN

Let us consider the Hamiltonian

$$H = \sum_{i=1}^{N}\left(-\frac{1}{2}\Delta_i - \frac{z}{r_i}\right) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{r_{ij}}. \tag{2.1}$$

Formulas for the diagonal matrix components of the various parts of H with respect to a function of type (1.2) were derived in A.[9] According to the formulas (51) to (53) of A, these components are given by the following expressions:

$$\begin{aligned}\int f^*(ij)\Big[\sum_{k=1}^{N}(-\tfrac{1}{2}\Delta_k)\Big]f(ij)\,dq &= \int \Phi^*(ij\mid 12)[-\tfrac{1}{2}\Delta_1]\Phi(ij\mid 12)\,dq_1\,dq_2 \\ &+ \frac{1}{2}\int |\Phi(ij\mid 12)|^2\,dq_1\,dq_2 \\ &\times\Big\{\sum_{\substack{s=1\\(s\neq ij)}}^{N}\int \varphi_s^*(1)[-\tfrac{1}{2}\Delta_1]\varphi_s(1)\,dq_1\Big\},\end{aligned} \tag{2.2}$$

$$\begin{aligned}\int f^*(ij)\Big[\sum_{k=1}^{N}\Big(-\frac{z}{r_k}\Big) + \frac{1}{2}\sum_{k,l=1}^{N}\frac{1}{r_{kl}}\Big]f(ij)\,dq &= \frac{1}{2}\int |\Phi(ij\mid 12)|^2\frac{1}{r_{12}}\,dq_1\,dq_2 \\ &+ \int |\Phi(ij\mid 12)|^2\Big(-\frac{z}{r_1}\Big)\,dq_1\,dq_2 \\ &+ \sum_{\substack{s=1\\(s\neq ij)}}^{N}\int |\varphi_s(2)|^2\frac{1}{r_{12}}|\Phi(ij\mid 13)|^2\,dq_1\,dq_2\,dq_3 \\ &- \sum_{\substack{s=1\\(s\neq i,j)}}^{N}\int \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}\Phi^*(ij\mid 13) \\ &\times \Phi(ij\mid 23)\,dq_1\,dq_2\,dq_3 \\ &+ \frac{1}{2}\int |\Phi(ij\mid 12)|^2\,dq_1\,dq_2 \\ &\times\Big\{\sum_{\substack{s=1\\(s\neq ij)}}^{N}\int \varphi_s^*(1)\Big(-\frac{z}{r_1}\Big)\varphi_s(1)\,dq_1 \\ &+ \frac{1}{2}\sum_{\substack{s,t=1\\(s,t\neq ij)}}^{N}\int\Big[\frac{|\varphi_s(1)|^2\,|\varphi_t(2)|^2}{r_{12}} \\ &- \frac{\varphi_s(1)\varphi_s^*(2)\varphi_t^*(1)\varphi_t(2)}{r_{12}}\Big]dq_1\,dq_2\Big\},\end{aligned} \tag{2.3}$$

and

$$\int |f(ij)|^2\,dq = \frac{1}{2}\int |\Phi(ij\mid 1,2)|^2\,dq_1\,dq_2. \tag{2.4}$$

In the formulas above, dq_i means integration with respect to the spatial coordinates $x_iy_iz_i$ and summation with respect to the spin coordinate σ_i; dq means the integration and summation with respect to *all* coordinates $q_1, q_2, \cdots q_N$. $\Phi(ij\mid 12)$ is a 2-electron function, orthogonalized to all one-electron spin-orbitals $\varphi_1, \varphi_2, \cdots \varphi_N$ except φ_i and φ_j, and it is defined in the following way:

$$\begin{aligned}\Phi(ij\mid 12) = [1 &- \Omega(ij\mid 1) - \Omega(ij\mid 2) \\ &+ \Omega(ij\mid 1)\Omega(ij\mid 2)]\Phi(ij\mid 12),\end{aligned} \tag{2.5}$$

where $\Omega(ij\mid 1)$ is the *orthogonality projection operator*[10] and is defined as

$$\Omega(ij\mid 1)F(1) = \sum_{\substack{s=1\\(s\neq ij)}}^{N}\varphi_s(1)\int \varphi_s^*(2)F(2)\,dq_2 \tag{2.6}$$

(F = arbitrary space-spin function). It was shown in A,[11] that the orthogonalization (2.5) does not change the total wave function (1.2), i.e.,

$$\begin{aligned}f(ij) &= \frac{1}{(N!)^{1/2}}\tilde{A}\{\varphi_i(1)\cdots\quad\cdots\varphi_N(N)\Phi(ij\mid ij)\} \\ &= \frac{1}{(N!)^{1/2}}\tilde{A}\{\varphi_i(1)\cdots\varphi_N(N)\Phi(ij\mid ij)\},\end{aligned} \tag{2.7}$$

where Φ is defined by (2.5).

[8] The investigations presented in this paper and in the forthcoming paper mentioned above will be reported shortly in the Quarterly Progress Report of the Solid-State and Molecular Theory Group of the Massachusetts Institute of Technology. Copies of that report are available upon request.

[9] See reference 1, Sec. 3.

[10] See the discussion in I Sec. 3. In the symbol $\Omega(ij|1)$ the indices (i, j) indicate that the orbitals φ_i and φ_j are excluded from the summation in the kernel of the operator [see (2.6)] and 1 indicates that the operator operates on the coordinate q_1.

[11] See Sec. 2 in A and Sec. 3 in I.

3. ANALYSIS OF THE DIAGONAL MATRIX COMPONENT

Let us consider the first, second, third, and fifth expressions of the matrix component (2.3). We are going to prove the following theorem:

Theorem 1. Let $\chi_A(q)$ be either one of the one-electron spin-orbitals in the function (1.1) or the function which we obtain by applying the operator of the kinetic energy to the one-electron spin-orbital, i.e., $-(\frac{1}{2})\,\Delta\varphi_A(q)$.

Let us introduce the notation

$$\rho_{AB}(1) \equiv \rho_{AB}(q_1) \equiv \chi_A^*(q_1)\chi_B(q_1).$$

Assuming that Φ has the form given by (1.3) all integrals which occur in the matrix component (2.3) with the exception of the fourth (exchange) integral, can be expressed in terms of the following four basic integrals:

$$I_1 = \int \rho_{AB}(1)\rho_{CD}(2) r_1^{\kappa} r_2^{\lambda} r_{12}^{\delta}\, dq_1\, dq_2 \tag{3.1}$$

$$I_2 = \int \rho_{AB}(1)\rho_{CD}(2)\rho_{EF}(3) r_1^{\kappa} r_2^{\lambda} r_3^{\mu} r_{12}^{\delta} r_{32}^{\epsilon} \times dq_1\, dq_2\, dq_3 \tag{3.2}$$

$$I_3 = \int \rho_{AB}(1)\rho_{CD}(2)\rho_{EF}(3)\rho_{GH}(4) r_1^{\kappa} r_2^{\lambda} r_3^{\mu} r_4^{\nu} r_{12}^{\delta} r_{32}^{\epsilon} r_{42}^{\rho} \times dq_1\, dq_2\, dq_3\, dq_4 \tag{3.3}$$

$$I_4 = \int \rho_{AB}(1)\rho_{CD}(2)\rho_{EF}(3)\rho_{GH}(4) r_1^{\kappa} r_2^{\lambda} r_3^{\mu} r_4^{\nu} r_{12}^{\delta} r_{32}^{\epsilon} r_{41}^{\rho} \times dq_1\, dq_2\, dq_3\, dq_4. \tag{3.4}$$

Proof. Let us introduce the notation

$$\varphi_{ij}^{\kappa\lambda\mu}(1, 2) \equiv \varphi_i(1)\varphi_j(2) r_1^{\kappa} r_2^{\lambda} r_{12}^{\mu}. \tag{3.5}$$

According to (1.3) the 2-electron function Φ may be written in the following form:

$$\Phi(ij \mid 1, 2) = \sum_{\alpha\beta\gamma} c(ij\alpha\beta\gamma) \times [\varphi_{ij}^{\alpha\beta\gamma}(1, 2) + \varphi_{ij}^{\beta\alpha\gamma}(1, 2) - \varphi_{ji}^{\alpha\beta\gamma}(1, 2) - \varphi_{ji}^{\beta\alpha\gamma}(1, 2)]. \tag{3.6}$$

Consequently, in order to prove our theorem, it is enough to consider the corresponding matrix components with respect to the function (3.5) instead of the matrix components with respect to (3.6). Therefore, let us consider the following matrix component which corresponds to the first integral of (2.3):

$$\mathfrak{I}_1 = \frac{1}{2}\int \{[1 - \Omega^*(1) - \Omega^*(2) + \Omega^*(1)\Omega^*(2)] \times \varphi_{uv}^{\alpha\beta\gamma *}(1, 2)\}(1/r_{12})\{[1 - \Omega(1) - \Omega(2) + \Omega(1)\Omega(2)]\varphi_{ij}^{\kappa\lambda\mu}(1, 2)\}\, dq_1\, dq_2, \tag{3.7}$$

where, in order to keep the discussion as general as possible, the orthogonality operator is defined as follows

$$\Omega(1)F(1) = \sum_s \varphi_s(1) \int \varphi_s^*(2)F(2)\, dq_2, \tag{3.8}$$

where we have not specified the limits of the summation in the kernel of the operator [$F(1)$ is an arbitrary space-spin function].

In the expression (3.7) we shall have first the integral containing the product $\varphi_{uv}^{\alpha\beta\gamma *}\varphi_{ij}^{\kappa\lambda\mu}$. Evidently, that integral will be a special case of I_1. Then we shall have an expression where to the left of $1/r_{12}$ we shall have $\varphi_{uv}^{\alpha\beta\gamma}(1, 2)$, to the right of $1/r_{12}$ $\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1, 2)$. Taking into account the definition of the orthogonality operator and incorporating $1/r_{12}$ into $\varphi_{uv}^{\alpha\beta\gamma}$ we obtain the expression

$$\sum_s \int \varphi_u^*(1)\varphi_v^*(2) r_1^{\alpha} r_2^{\beta} r_{12}^{\gamma-1} \varphi_s(1)\varphi_s^*(3)\varphi_i(3)\varphi_j(2) r_3^{\kappa} r_2^{\lambda} r_{32}^{\mu} \times dq_1\, dq_2\, dq_3, \tag{3.8}$$

which is clearly a special case of I_2. We obtain the same type of integral from the term containing the product $\varphi_{uv}^{\alpha\beta\gamma *}(1, 2)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1, 2)$. Next we shall have an expression where to the left of $1/r_{12}$ we shall have $\varphi_{uv}^{\alpha\beta\gamma *}(1, 2)$, to the right of $(1/r_{12})$ the function $\Omega(1)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1, 2)$. Taking into account the definition of the orthogonality operator we obtain

$$\Omega(1)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1, 2) = \sum_{s,t} \varphi_s(1)\varphi_t(2) A_{st}, \tag{3.9}$$

where

$$A_{st} = \int \varphi_s^*(1)\varphi_t^*(2)\varphi_i(1)\varphi_j(2) r_1^{\kappa} r_2^{\lambda} r_{12}^{\mu}\, dq_1\, dq_2 \tag{3.10}$$

which is an integral of type I_1. Incorporating $1/r_{12}$ into $\varphi_{uv}^{\alpha\beta\gamma *}(1, 2)$ we find that the integral containing the product $\varphi_{uv}^{\alpha\beta\gamma *}(1, 2)\Omega(1)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1, 2)$ will be built from integrals of type I_1.

Next let us consider the term where we have the product $\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma *}(1, 2)\Omega(1)\varphi_{ij}^{\kappa\mu\lambda}(1, 2)$. Taking into account the definition of the orthogonality operator we obtain the expression

$$\sum_{s,t} \int \varphi_s^*(1)\varphi_s(3)\varphi_u^*(3)\varphi_v^*(2) r_3^{\alpha} r_2^{\beta} r_{32}^{\gamma}(1/r_{12})\varphi_t(1)\varphi_t^*(4) \times \varphi_i(4)\varphi_j(2) r_4^{\kappa} r_2^{\lambda} r_{42}^{\mu}\, dq_1\, dq_2\, dq_3\, dq_4, \tag{3.11}$$

which is a special case of I_3. Similarly, it is easy to show that the term which contains the product

$$[\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma *}(1, 2)][\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1, 2)] \tag{3.12}$$

leads to the integral I_4. Next let us consider the term

with

$$[\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)][\Omega(1)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1,2)].$$

Taking into account (3.9) we obtain the following expression:

$$\sum_{a,b,c}\int \varphi_a^*(1)\varphi_a(3)\varphi_u^*(3)\varphi_\tau^*(2)r_{31}^{\alpha}r_{21}^{\beta}r_{32}^{\gamma}\varphi_b(1)\varphi_c(2) \times (1/r_{12})A_{bc}, \quad (3.13)$$

where A_{bc} is given by (3.10). The above expression has the form of the basic integral I_2.

Let us consider now that term in (3.7) in which to the left of $1/r_{12}$ we have $\Omega^*(2)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)$, and to the right of $1/r_{12}$ the function $\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1,2)$. Obviously, that integral leads to an expression similar to (3.11), i.e., it leads to the integral I_3. Then we shall have an integral with the product

$$[\Omega^*(2)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)][\Omega(1)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1,2)].$$

Simple manipulations show, that this integral leads to an expression like (3.13) which is built from products of type $I_2 \cdot I_1$. Finally, we shall have in the expression (3.7) the integral

$$\int [\Omega^*(1)\Omega^*(2)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)] \times (1/r_{12})[\Omega(1)\Omega(2)\varphi_{ij}^{\kappa\lambda\mu}(1,2)]\,dq_1\,dq_2. \quad (3.14)$$

Taking into account (3.9) and (3.10) it is easy to see that (3.14) will be built from products of 3 integrals each having the form of the basic integral I_1. Theorem 1 is now proved with respect to the first expression of (2.3).

Let us consider now the following integral which corresponds to the second integral of the matrix component (2.3):

$$\mathfrak{I}_2 = \int \{[1-\Omega^*(1)][1-\Omega^*(2)]\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)\}(-z/r_1) \times \{[1-\Omega(1)][1-\Omega(2)]\varphi_{ij}^{\kappa\lambda\mu}(1,2)\}\,dq_1\,dq_2. \quad (3.15)$$

Taking into account that $[1-\Omega(2)]$ commutes with $(1/r_1)$ and that it is a Hermitian projection operator we obtain

$$\mathfrak{I}_2 = \int \{[1-\Omega^*(1)][1-\Omega^*(2)]\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)\}(-z/r_1) \times \{[1-\Omega(1)]\varphi_{ij}^{\kappa\lambda\mu}(1,2)\}\,dq_1\,dq_2. \quad (3.16)$$

Let us consider first those terms in which to the right of z/r_1 we have $\varphi_{ij}^{\kappa\lambda\mu}(1,2)$. The $1/r_1$ can be incorporated into $\varphi_{ij}^{\kappa\lambda\mu}(1,2)$ by lowering the index κ by one:

$$(1/r_1)\varphi_{ij}^{\kappa\lambda\mu}(1,2) = \varphi_{ij}^{\kappa-1,\lambda\mu}(1,2). \quad (3.17)$$

The 4 integrals which contain this term, belong to the same category as the integrals which we obtained from those terms of (3.7) in which to the right of $1/r_{12}$ we had $\varphi_{ij}^{\kappa\lambda\mu}(1,2)$. Consequently, it is not necessary to discuss these integrals here. Next let us consider those terms of (3.16) in which to the right of z/r_1 we shall have $\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1,2)$. The first of these will contain to the left of z/r_1 the function $\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)$. We can incorporate $1/r_1$ into $\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)$ by lowering the index α by one, and then the resulting expression will be a special case of I_2. Next consider the integral

$$\int \left[\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)\right]\left(-\frac{z}{r_1}\right)\left[\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1,2)\right] dq_1\,dq_2. \quad (3.18)$$

According to the definition of the orthogonality operator we obtain the expression

$$\sum_{a,b}\int \varphi_a^*(1)\varphi_a(3)\varphi_u^*(3)\varphi_\tau^*(2)r_{31}^{\alpha}r_{21}^{\beta}r_{32}^{\gamma}(-z/r_1) \times \varphi_b(1)\varphi_b^*(4)\varphi_i(4)\varphi_j(2)r_{41}^{\kappa}r_{21}^{\lambda}r_{42}^{\mu}\,dq_1\,dq_2\,dq_3\,dq_4. \quad (3.19)$$

The above integral can be divided into two parts, the first part being the integral with respect to q_1, the second being the integral with respect to q_2, q_3, q_4. Obviously, the integral over q_1 will be the special case of I_1 with $\lambda = 0$, $\delta = 0$, $\kappa = -1$. The integral over $(q_2q_3q_4)$ is a special case of I_2.

Next we shall have a term containing the product

$$[\Omega^*(2)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)][\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1,2)]. \quad (3.20)$$

The expression which we obtain may be written as

$$\sum_{a,b}\int \varphi_a^*(2)\varphi_a(3)\varphi_u^*(1)\varphi_\tau^*(3)r_{13}^{\alpha}r_{23}^{\beta}r_{13}^{\gamma}(-z/r_1)\varphi_b(1)\varphi_b^*(4) \times \varphi_i(4)\varphi_j(2)r_{41}^{\kappa}r_{21}^{\lambda}r_{42}^{\mu}\,dq_1\,dq_2\,dq_3\,dq_4. \quad (3.21)$$

Every integral of the above double sum may be expressed as the product of two integrals, one over (q_1q_3), the other over q_2 and q_4. Both are special cases of I_1. The last term of (3.16) will be an integral, where to the left of (z/r_1) we shall have $\Omega^*(1)\Omega^*(2)\varphi_{uv}^{\alpha\beta\gamma^*}(1,2)$ and to the right of (z/r_1) we shall have $\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1,2)$. Taking into account (3.9), this term may be written in the following form:

$$\sum_{abc}\int \varphi_a(1)\varphi_b(2)A_{ab}\left(-\frac{z}{r_1}\right)\varphi_c(1)\varphi_c^*(3)\varphi_i(3)\varphi_j(2) \times r_{31}^{\kappa}r_{21}^{\lambda}r_{32}^{\mu}\,dq_1\,dq_2\,dq_3, \quad (3.22)$$

where A_{ab} is given by (3.10). The integral (3.22) can be expressed as the product of two integrals separat-

ing the integration with respect to q_1 from the integration with respect to q_2 and q_3. The integral over q_1 will be a special case of I_1 with $\lambda = \delta = 0$, $\kappa = -1$; the integral over q_2 and q_3 is again a special case of I_1. Theorem 1 is therefore proved also with respect to the second expression of (2.3).

Let us consider now the following integral which corresponds to the third expression of the matrix component (2.3):

$$\mathfrak{I}_3 = \int \sum_{s=1}^{N} \varphi_s(2)\varphi_s^*(2) \frac{1}{r_{12}} \times \{[1 - \Omega^*(1)][1 - \Omega^*(3)]\varphi_{uv}^{\alpha\beta\gamma^*}(1, 2)\} \times \{[1 - \Omega(1)][1 - \Omega(3)]\varphi_{ij}^{\kappa\lambda\mu}(1, 3)\} \times dq_1\, dq_2\, dq_3. \tag{3.23}$$

We observe that the Hartree potential

$$\int \sum_{s=1}^{N} |\varphi_s(2)|^2 \frac{1}{r_{12}} dq_2 \tag{3.24}$$

is a function of $\mathbf{r}_1$ only. Taking into account that $[1 - \Omega(3)]$ commutes with the Hartree potential (3.24) and that it is a Hermitian projection operator, we obtain for the integral (3.23) the following expression:

$$\mathfrak{I}_3 = \int \{[1 - \Omega^*(1) - \Omega^*(3) + \Omega^*(1)\Omega^*(3)] \times \varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)\}\left[\sum_{s=1}^{N} |\varphi_s(2)|^2 \frac{1}{r_{12}}\right] \times \{[1 - \Omega(1)]\varphi_{ij}^{\kappa\lambda\mu}(1, 3)\}\, dq_1\, dq_2\, dq_3. \tag{3.25}$$

Let us consider first those terms of (3.25) in which to the right of the Hartree potential $\sum_{(s)} |\varphi_s(2)|^2\ (1/r_{12})$ we shall have $\varphi_{ij}^{\kappa\lambda\mu}(1, 3)$ without the operator $\Omega(1)$. In the first of these four integrals we shall have the product

$$\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)\varphi_{ij}^{\kappa\lambda\mu}(1, 3) = \varphi_u^*(1)\varphi_v^*(3)\varphi_{ij}^{\kappa+\alpha,\beta+\lambda,\gamma+\mu}(1, 3)$$

which, combined with the Hartree potential will lead to an integral of type I_2. Next we shall have the expression

$$\int [\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)] \times \sum_{s=1}^{N} \varphi_s(2)\varphi_s^*(2) \frac{1}{r_{12}} \varphi_{ij}^{\kappa\lambda\mu}(1, 3)\, dq_1\, dq_2\, dq_3 = \sum_{s=1}^{N} \sum_{a} \int \varphi_a^*(1)\varphi_a(4)\varphi_u^*(4)\varphi_v^*(3)r_4^{\alpha}r_3^{\beta}r_{43}^{\gamma}\varphi_s(2)\varphi_s^*(2) \frac{1}{r_{12}} \times \varphi_i(1)\varphi_j(3)r_1^{\kappa}r_3^{\lambda}r_{13}^{\mu}\, dq_1\, dq_2\, dq_3\, dq_4 \tag{3.26}$$

which is a special case of I_4. Similarly, it is easy to see that the next integral containing $\Omega^*(3)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)$ leads to an expression which is a special case of I_3. Finally, consider that term which contains $\Omega^*(1)\Omega^*(3)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)$. According to (3.9) this function is a sum of products of two one-electron spin-orbitals multiplied by a constant which is an integral like I_1. The two one-electron functions together with $\varphi_{ij}^{\kappa\lambda\mu}(1, 3)$ and with the Hartree potential will lead to an integral of the type I_2.

Next let us consider those four integrals in the expression (3.25) in which to the right of the Hartree potential we have $\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1, 3)$. The first of these will contain to the left of the Hartree potential the function $\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)$. A similar integral already occurred above when we considered the term which contained the product $[\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)]\varphi_{ij}^{\kappa\lambda\mu}(1, 3)$. The next term will be an integral in which we shall have $\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)$ as well as $\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1, 3)$ besides the Hartree potential. We shall obtain an r_{43}^{γ} from the first of these expressions, an r_{53}^{μ} from the second, and the Hartree potential contains r_{12}^{-1}. Therefore what we obtain is a sum of products, each product having the form $I_2 \cdot I_1$. The next term will contain the product

$$[\Omega^*(3)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)] \sum_{s=1}^{N} |\varphi_s(2)|^2\ (1/r_{12})[\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1, 3)]. \tag{3.27}$$

Here we obtain an r_{14}^{γ} from the first orthogonality operator, an r_{53}^{μ} from the second, and we also have the r_{12}^{-1}. Evidently, the product $r_{14}^{\gamma}r_{12}^{-1}$ gives rise to an integral like I_2 and the r_{53}^{μ} leads to an integral like I_1. Finally the last term of $\mathfrak{I}_3$ contains the product

$$[\Omega^*(1)\Omega^*(3)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)] \sum_{s=1}^{N} |\varphi_s(2)|^2\ (1/r_{12}) \times [\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1, 3)]. \tag{3.28}$$

According to (3.9), the function in the first bracket may be written in the form

$$\sum_{m,n} \varphi_m^*(1)\varphi_n^*(3)A_{mn}. \tag{3.29}$$

The orthogonality operator $\Omega(1)$ in the expression (3.28) produces an r_{43}^{μ} and in the Hartree potential we have r_{12}^{-1}. Consequently, the expression (3.28) will be a sum of terms, each term being the product of three integrals of the type I_1. Theorem 1 is thereby proved also with respect to the expression $\mathfrak{I}_3$.

In order to complete the proof of Theorem 1 let us consider the fifth expression of the matrix component (2.3). That expression contains the overlap integral and a Hartree-Fock-type energy ex-

pression. It is known that the Hartree-Fock-type energy expression may be expressed in terms of the basic integral I_1. Let us consider the following integral which corresponds to the overlap integral:

$$\begin{aligned}\mathfrak{I}_5 = \int \{&[1 - \Omega^*(1) - \Omega^*(2) \\ &+ \Omega^*(1)\Omega^*(2)]\varphi_{uv}^{\alpha\beta\gamma^*}(1, 2)\} \\ &\times \{[1 - \Omega(1) - \Omega(2) \\ &+ \Omega(1)\Omega(2)]\varphi_{ij}^{\kappa\lambda\mu}(1, 2)\}\, dq_1\, dq_2. \end{aligned} \tag{3.30}$$

Taking into account that the orthogonality operator in the square bracket is a Hermitian projection operator we obtain

$$\begin{aligned}\mathfrak{I}_5 = \int \{&[1 - \Omega^*(1) - \Omega^*(2) + \Omega^*(1)\Omega^*(2)] \\ &\times \varphi_{uv}^{\alpha\beta\gamma^*}(1, 2)\}\varphi_{ij}^{\kappa\lambda\mu}(1, 2)\, dq_1\, dq_2. \end{aligned} \tag{3.31}$$

An expression of the type (3.31) was already discussed when we analyzed (3.7). We are referring to that part of (3.7) in which to the right of $1/r_{12}$ we had the function $\varphi_{ij}^{\kappa\lambda\mu}(1, 2)$. That part of (3.7) may be written in the following form:

$$\begin{aligned}\frac{1}{2}\int \{&[1 - \Omega^*(1) - \Omega^*(2) + \Omega^*(1)\Omega^*(2)] \\ &\times \varphi_{uv}^{\alpha\beta\gamma^*}(1, 2)\}\varphi_{ij}^{\kappa\lambda(\mu-1)}(1, 2)\, dq_1\, dq_2 \end{aligned} \tag{3.32}$$

which has the same structure as (3.31). Consequently, (3.31) can also be reduced to the basic integrals (3.1)–(3.4).

This completes the proof of Theorem 1.

Theorem 2. Let us consider the fourth (exchange) integral of the matrix component (2.3). That integral can be reduced to five basic integrals I_1, I_2, $\cdots$, I_5 where I_1 to I_4 were given by (3.1)–(3.4) and I_5 is defined as

$$I_5 = \int \rho_{AB}(1)\rho_{CD}(2)\rho_{EF}(3)r_1^{\kappa}r_2^{\lambda}r_3^{\mu}r_{13}^{\delta}r_{23}^{\epsilon}r_{12}^{-1}\, dq_1\, dq_2\, dq_3, \tag{3.33}$$

where $\rho_{AB}, \cdots, \rho_{EF}$ were defined in connection with the definition of the integrals I_1 to I_4.

Proof. Let us consider the following integral which corresponds to the exchange integral of the matrix component (2.3):

$$\begin{aligned}\mathfrak{I}_4 = \int &\sum_{s=1}^{N} \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}} \\ &\times \{[1 - \Omega^*(1)][1 - \Omega^*(3)]\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)\} \\ &\times \{[1 - \Omega(2)][1 - \Omega(3)]\varphi_{ij}^{\kappa\lambda\mu}(2, 3)\} \\ &\times dq_1\, dq_2\, dq_3. \end{aligned} \tag{3.34}$$

Taking into account the fact that $[1 - \Omega(3)]$ commutes with $\varphi_s(1)\varphi_s^*(2)(1/r_{12})$ and that it is a Hermitian projection operator we obtain

$$\begin{aligned}\mathfrak{I}_4 = \int &\sum_{s=1}^{N} \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}} \\ &\times \{[1 - \Omega^*(1)][1 - \Omega^*(3)]\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)\} \\ &\times \{[1 - \Omega(2)]\varphi_{ij}^{\kappa\lambda\mu}(2, 3)\}\, dq_1\, dq_2\, dq_3. \end{aligned} \tag{3.35}$$

Let us consider first those four terms in which $\varphi_{ij}^{\kappa\lambda\mu}$ occurs without the orthogonality operator:

$$\mathfrak{I}_4^1 = \int \sum_s \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)\varphi_{ij}^{\kappa\lambda\mu}(2, 3)\, dq_{123}, \tag{3.36}$$

$$\mathfrak{I}_4^2 = \int \sum_s \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}[\Omega^*(1)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)] \times \varphi_{ij}^{\kappa\lambda\mu}(2, 3)\, dq_{123}, \tag{3.37}$$

$$\mathfrak{I}_4^3 = \int \sum_s \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}[\Omega^*(3)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)] \times \varphi_{ij}^{\kappa\lambda\mu}(2, 3)\, dq_{123}, \tag{3.38}$$

$$\mathfrak{I}_4^4 = \int \sum_s \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}[\Omega^*(1)\Omega^*(3)\varphi_{uv}^{\alpha\beta\gamma^*}(1, 3)] \times \varphi_{ij}^{\kappa\lambda\mu}(2, 3)\, dq_{123} \tag{3.39}$$

$$(dq_{123} \equiv dq_1\, dq_2\, dq_3).$$

By proving Theorem 1, we have seen that with each integral under consideration the arrangement of the interelectronic distances determined the basic integral to which they could be reduced. Taking into account (3.5) we realize that each integral of the sum in $\mathfrak{I}_4^1$ will contain the product

$$(r_{12}^{-1}\quad r_{13}^{\gamma}\quad r_{23}^{\mu}),$$

from which it follows that each integral which occurs in $\mathfrak{I}_4^1$ will be a special case of (3.33). Making use of the definition of the orthogonality operator we find that $\mathfrak{I}_4^2$ to $\mathfrak{I}_4^4$ will contain the following arrangements of the interelectronic distances:

$\mathfrak{I}_4^2$ contains $(r_{12}^{-1}\quad r_{43}^{\gamma}\quad r_{23}^{\mu})$;

$\mathfrak{I}_4^3$ contains $(r_{12}^{-1}\quad r_{14}^{\gamma}\quad r_{23}^{\mu})$;

$\mathfrak{I}_4^4$ contains $(r_{12}^{-1}\quad r_{43}^{\gamma}\quad r_{23}^{\mu})$;

from which it follows that these integrals may be reduced in the following way:

(1) The integrals which occur in $\mathfrak{I}_4^2$ and $\mathfrak{I}_4^3$ can be reduced to the basic integral I_4.

(2) The integrals which occur in $\mathfrak{I}_4^4$ can be reduced to the basic integrals I_1 and I_2.

Next let us consider those four terms of (3.35) in

which $\varphi_{ij}^{\kappa\lambda\mu}(2, 3)$ occurs multiplied by $\Omega(2)$. Writing out the orthogonality operator in detail we have

$$\mathfrak{I}_4' = -\int \sum_{s=1}^{N} \varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}} \times \{[1 - \Omega^*(1) - \Omega^*(3) + \Omega^*(1)\Omega^*(3)]\varphi_{uv}^{\alpha\beta\gamma *}(1, 3)\} \times \{\sum_a \varphi_a(2)\varphi_a^*(4)\varphi_{ij}^{\kappa\lambda\mu}(43)\}\, dq_{1234}. \quad (3.40)$$

Comparing this expression with (3.25), we realize that that part of (3.25) in which the function $\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1, 3)$ occurs is very similar to (3.40). Let us denote the corresponding part of $\mathfrak{I}_3$ by $\mathfrak{I}_3'$:

$$\mathfrak{I}_3' = \int \sum_{s=1}^{N} \varphi_s(2)\varphi_s^*(2)\frac{1}{r_{12}} \times \{[1 - \Omega^*(1) - \Omega^*(3) + \Omega^*(1)\Omega^*(3)]\varphi_{uv}^{\alpha\beta\gamma *}(1, 3)\} \times \{\sum_a \varphi_a(1)\varphi_a^*(4)\varphi_{ij}^{\kappa\lambda\mu}(4, 3)\}\, dq_{1234}. \quad (3.41)$$

We obtain $\mathfrak{I}_3'$ from $\mathfrak{I}_4'$ by exchanging the indices of $\varphi_s(1)$ and $\varphi_a(2)$. Consequently, $\mathfrak{I}_4'$ will be the composition of the same basic integrals as $\mathfrak{I}_3'$. Since we have shown above that $\mathfrak{I}_3'$ can be reduced to the four basic integrals I_1 to I_4, it is proved now that the same is true for $\mathfrak{I}_4'$.

This completes the proof of Theorem 2.

Theorem 3. The normalization integral (2.4) can be expressed in terms of the basic integrals I_1 and I_2.

Proof. We met with this type of integral when we analyzed the matrix component (2.3). The normalization integral occurred in the fifth expression of that matrix component. We have shown there that the overlap integral can be reduced to the basic integrals I_1 and I_2.

Theorem 4. Consider the matrix component of the kinetic energy operator, which was given by Eq. (2.2). The integrals which occur in the expression (2.2) are special cases of the basic integrals I_1, I_2, and I_6 where I_1 and I_2 were given by (3.1) and (3.2) and I_6 is defined as follows:

$$I_6 = \int \varphi_A^*(1)\varphi_B^*(2)\varphi_C(2)r_1^\sigma r_2^\omega r_{12}^\tau(-\tfrac{1}{2}\Delta_1) \times [\varphi_D(1)\varphi_E^*(3)\varphi_F(3)r_1^\tau r_3^\delta r_{13}^\varepsilon]\, dq_1\, dq_2\, dq_3. \quad (3.42)$$

Proof. Consider the first integral of the matrix component (2.2):

$$\mathfrak{I}_6 = \int \{[1 - \Omega^*(ij \mid 1)] \times [1 - \Omega^*(ij \mid 2)]\Phi^*(ij \mid 1, 2)\}[-\tfrac{1}{2}\Delta_1] \times \{[1 - \Omega(ij \mid 1)][1 - \Omega(ij \mid 2)]\Phi(ij \mid 1, 2)\} \times dq_1\, dq_2. \quad (3.43)$$

Taking into account that $[1 - \Omega(ij \mid 2)]$ commutes with $-(\frac{1}{2})\Delta_1$ and that it is a Hermitian projection operator we obtain

$$\mathfrak{I}_6 = \int \{[1 - \Omega^*(ij \mid 1)] \times [1 - \Omega^*(ij \mid 2)]\Phi^*(ij \mid 1, 2)\}[-\tfrac{1}{2}\Delta_1] \times \{[1 - \Omega(ij \mid 1)]\Phi(ij \mid 1, 2)\}\, dq_1\, dq_2. \quad (3.44)$$

The reduction of (3.44) to the basic integrals I_1, I_2, and I_6 will take place in several steps. First let us consider the integral

$$\mathfrak{I}_6^1 = \int \Phi^*(ij \mid 1, 2)[-\tfrac{1}{2}\Delta_1]\Phi(ij \mid 1, 2)\, dq_1\, dq_2. \quad (3.45)$$

According to (1.3) Φ may be written in the following form:

$$\Phi(ij \mid 1, 2) = \mu(ij \mid 1, 2)\Lambda(1, 2), \quad (3.46)$$

where

$$\mu(ij \mid 1, 2) = [\varphi_i(1)\varphi_j(2) - \varphi_i(2)\varphi_j(1)] \quad (3.47)$$

and

$$\Lambda(1, 2) = \sum_{\alpha\beta\gamma} c(ij\alpha\beta\gamma)(r_1^\alpha r_2^\beta + r_1^\beta r_2^\alpha)r_{12}^\gamma. \quad (3.48)$$

Putting (3.46) into (3.45) and transforming the integral according to Hellmann,[12] we obtain

$$\mathfrak{I}_6^1 = \int [\Lambda(1, 2)]^2\mu(ij \mid 1, 2) \times \{[-\tfrac{1}{2}\Delta_1]\mu(ij \mid 1, 2)\}\, dq_1\, dq_2 + \frac{1}{2}\int |\mu(ij \mid 1, 2)|^2\, [\nabla_1\Lambda(1, 2)]^2\, dq_1\, dq_2. \quad (3.49)$$

Putting (3.49) into (3.44) we obtain

$$\mathfrak{I}_6 = \int [\Lambda(1, 2)]^2\mu(ij \mid 1, 2) \times \{[-\tfrac{1}{2}\Delta_1]\mu(ij \mid 1, 2)\}\, dq_1\, dq_2 + \frac{1}{2}\int |\mu(ij \mid 1, 2)|^2\, [\nabla_1\Lambda(1, 2)]^2\, dq_1\, dq_2 + \int \{[-\Omega^*(ij \mid 1) - \Omega^*(ij \mid 2) + \Omega^*(ij \mid 1)\Omega^*(ij \mid 2)]\Phi^*(ij \mid 1, 2)\} \times [-\tfrac{1}{2}\Delta_1]\{[1 - \Omega(ij \mid 1)]\Phi(ij \mid 1, 2)\}\, dq_1\, dq_2 - \int \Phi^*(ij \mid 1, 2)[-\tfrac{1}{2}\Delta_1] \times [\Omega(1)\Phi(ij \mid 1, 2)]\, dq_1\, dq_2. \quad (3.50)$$

[12] H. Hellmann, *Einführung in die Quantenchemie* (Franz Deuticke, Leipzig, 1937), p. 73. See also C. C. J. Roothaan and A. W. Weiss, Revs. Modern Phys. 32, 194 (1960).

Since Φ and Λ may be written in the form given by (3.5) and (3.48), respectively, it is evident that in order to prove Theorem 4 it will be sufficient to consider instead of (3.50) the following integral

$$\begin{aligned}\mathfrak{J}_6' &= \int r_1^{\alpha+\kappa} r_2^{\beta+\lambda} r_{12}^{\gamma+\mu} \varphi_u^*(1)\varphi_v^*(2) \\ &\times \{[-\tfrac{1}{2}\Delta_1]\varphi_i(1)\varphi_j(2)\}\, dq_{12} \\ &+ \frac{1}{2}\int \varphi_u^*(1)\varphi_v^*(2)\varphi_i(1)\varphi_j(2) \\ &\times [\nabla_1(r_1^\alpha r_{12}^\gamma)][\nabla_1(r_1^\kappa r_{12}^\mu)] r_2^{\beta+\lambda}\, dq_{12} \\ &+ \int \{[-\Omega^*(1) - \Omega^*(2) + \Omega^*(1)\Omega^*(2)]\varphi_{uv}^{\alpha\beta\gamma *}(1,2)\} \\ &\times \{[-\tfrac{1}{2}\Delta_1][1 - \Omega(1)]\varphi_{ij}^{\kappa\lambda\mu}(1,2)\}\, dq_{12} \\ &- \int \varphi_{uv}^{\alpha\beta\gamma *}(1,2)\{[-\tfrac{1}{2}\Delta_1][\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1,2)]\}\, dq_{12} \\ &\quad (dq_{12} = dq_1\, dq_2). \end{aligned} \tag{3.51}$$

Introducing the notation

$$\chi_A(q) \equiv -\tfrac{1}{2}\Delta\varphi_A(q), \tag{3.52}$$

the first of the integrals of $\mathfrak{J}_6'$ may be written in the following form:

$$\int \varphi_u^*(1)\chi_i(1)\varphi_v^*(2)\varphi_j(2) r_1^{\alpha+\kappa} r_2^{\beta+\lambda} r_{12}^{\gamma+\mu}\, dq_{12}, \tag{3.53}$$

which is a special case of the basic integral I_1. Next, consider the second integral of (3.51). We obtain, with elementary manipulations,

$$\begin{aligned}&[\nabla_1(r_1^\alpha r_{12}^\gamma)][\nabla_1(r_1^\kappa r_{12}^\mu)] \\ &\quad = [\gamma\mu + \tfrac{1}{2}(\alpha\mu + \kappa\gamma)] r_1^{\alpha+\kappa} r_{12}^{\gamma+\mu-2} \\ &\qquad + [\alpha\kappa + \tfrac{1}{2}(\alpha\mu + \kappa\gamma)] r_1^{\alpha+\kappa-2} r_{12}^{\gamma+\mu} \\ &\qquad - [\tfrac{1}{2}(\alpha\mu + \kappa\gamma)] r_1^{\alpha+\kappa-2} r_2^2 r_{12}^{\gamma+\mu-2}. \end{aligned} \tag{3.54}$$

On putting (3.54) into the second term of (3.51) we obtain a sum of integrals each of which is a special case of I_1.

We turn now our attention to the third expression of (3.51). First, let us consider those terms in which the $\varphi_{ij}^{\kappa\lambda\mu}$ occurs without the orthogonality operator. Making use of the Hermitian property of the Laplacian we obtain

$$\begin{aligned}\mathfrak{J}_6'' &= \int \{[-\Omega^*(1) - \Omega^*(2) \\ &\quad + \Omega^*(1)\Omega^*(2)]\varphi_{uv}^{\alpha\beta\gamma *}(1,2)\} \\ &\quad \times \{[-\tfrac{1}{2}\Delta_1]\varphi_{ij}^{\kappa\lambda\mu}(1,2)\}\, dq_{12} \\ &= \int \varphi_{ij}^{\kappa\lambda\mu *}(1,2)\{[-\tfrac{1}{2}\Delta_1][-\Omega(1) - \Omega(2) \\ &\quad + \Omega(1)\Omega(2)]\varphi_{uv}^{\alpha\beta\gamma}(1,2)\}\, dq_{12}. \end{aligned} \tag{3.55}$$

We observe that, according to the definition of the orthogonality operator $\Omega(1)$ [Eq. (2.6)], the functions

$$\Omega(1)\varphi_{uv}^{\alpha\beta\gamma}(1,2) \tag{3.56}$$

and

$$\Omega(1)\Omega(2)\varphi_{uv}^{\alpha\beta\gamma}(1,2) \tag{3.57}$$

depend on the coordinate $\mathbf{r}_1$ only through the one-electron spin-orbital $\varphi_a(q_1)$ in the kernel of the operator $\Omega(1)$. The application of the Laplacian changes each of these functions into the corresponding $\chi_a(q_1)$ [according to the definition (3.52)]. After that, those integrals which contain (3.56) and (3.57) will have the same form as the second and fourth integral of the expression (3.32), except that now, in each integral, one of the one-electron spin-orbitals will be replaced by the corresponding χ_a function. Consequently, the integrals which contain (3.56) and (3.57) will be reducible to the basic integrals I_1 and I_2, respectively.

Consider now the second term of (3.55). Writing out in detail the orthogonality operator we obtain the expression

$$\begin{aligned}&\int \varphi_i^*(1)\varphi_j^*(2) r_1^\kappa r_2^\lambda r_{12}^\mu \{[-\tfrac{1}{2}\Delta_1][\sum_a \varphi_a(2)\varphi_a^*(3) \\ &\qquad \times \varphi_u(1)\varphi_v(3) r_1^\alpha r_3^\beta r_{13}^\gamma]\}\, dq_{123}. \end{aligned} \tag{3.58}$$

Comparison with (3.42) shows that each of the integrals of the above sum will be a special case of the basic expression I_6.

Next we consider those terms of the third expression of (3.51) in which $\varphi_{ij}^{\kappa\lambda\mu}(1,2)$ occurs multiplied by $\Omega(1)$. Here we observe again that

$$\Omega(1)\varphi_{ij}^{\kappa\lambda\mu}(1,2) \tag{3.59}$$

depends on the coordinate $\mathbf{r}_1$ only through the one-electron spin-orbitals in the kernel of the orthogonality operator. The application of the Laplacian transforms these into the corresponding χ_a functions. After that, it is easy to see that these integrals will be reducible to the basic integrals I_1 and I_2.

Finally, let us consider the last expression of the matrix component (3.51). In that expression we again have the function (3.59). Therefore the same arguments, which we have presented in the previous paragraph also apply to this integral.

This completes the proof of Theorem 4.

4. THE NONDIAGONAL MATRIX COMPONENTS OF THE HAMILTONIAN

There are two different types of nondiagonal matrix components which occur if we use the wave function (1.1). We shall have components with one different orbital index:

$$\int f^*(ij)Hf(jl)\,dq \tag{4.1}$$

and with two different orbital indices

$$\int f^*(ij)Hf(kl)\,dq. \tag{4.2}$$

For both types explicit formulas were derived in A. Here we first quote these formulas and then in the next section investigate them. Let us define the 3-electron functions $\Phi(ijl \mid 1, 2, 3)$ and $\bar{\Phi}(ijl \mid 1, 2, 3)$ as follows:

$$\Phi(ijl \mid 1,2,3) \equiv \bar{\Phi}(ij \mid 1,2)\varphi_l(3) - \bar{\Phi}(ij \mid 1,3)\varphi_l(2) + \bar{\Phi}(ij \mid 2,3)\varphi_l(1), \tag{4.3}$$

$$\bar{\Phi}(ijl \mid 1,2,3) \equiv \bar{\Phi}(jl \mid 1,2)\varphi_i(3) - \bar{\Phi}(jl \mid 1,3)\varphi_i(2) + \bar{\Phi}(jl \mid 2,3)\varphi_i(1), \tag{4.4}$$

where $\bar{\Phi}(ij \mid 1, 2)$ and $\bar{\Phi}(jl \mid 1, 2)$ are defined according to (2.5), i.e., they are orthogonal to all orbitals except to "their own" (except to those to which the orbital indices of the 2-electron functions are referring). According to the formulas (62)–(64) of A the matrix component (4.1) is given by the following formulas:

$$\begin{aligned}
&\int f^*(ij)\Big[\sum_{m=1}^{N}(-\tfrac{1}{2}\Delta_m)\Big]f(jl)\,dq \\
&= \frac{1}{2}\int \Phi^*(ijl \mid 123)[-\tfrac{1}{2}\Delta_1]\bar{\Phi}(ijl \mid 123)\,dq_{123} \\
&\quad + \frac{1}{3!}\int \Phi^*(ijl \mid 123)\bar{\Phi}(ijl \mid 123)\,dq_{123} \\
&\quad \times\Big\{\sum_{\substack{s=1\\(s\neq ijl)}}^{N}\int \varphi_s(1)[-\tfrac{1}{2}\Delta_1]\varphi_s(1)\,dq_1\Big\},
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
&\int f^*(ij)\Big[\sum_{m=1}^{N}\Big(-\frac{z}{r_m}\Big)+\frac{1}{2}\sum_{m,n,=1}^{N}\frac{1}{r_{mn}}\Big]f(jl)\,dq \\
&= \frac{1}{2}\int \Phi^*(ijl \mid 123)\bar{\Phi}(ijl \mid 123)\frac{1}{r_{12}}\,dq_{123} \\
&\quad + \frac{1}{2}\int \Phi^*(ijl \mid 123)\bar{\Phi}(ijl \mid 123)\Big[-\frac{z}{r_1}\Big]\,dq_{123} \\
&\quad + \frac{1}{2}\int \sum_{\substack{s=1\\(s\neq ijl)}}^{N}|\varphi_s(2)|^2\frac{1}{r_{12}}\Phi^*(ijl \mid 134) \\
&\quad \times \Phi(ijl \mid 134)\,dq_{1234} \\
&\quad - \frac{1}{2}\int \sum_{\substack{s=1\\(s\neq ijl)}}^{N}\varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}\Phi^*(ijl \mid 134) \\
&\quad \times \Phi(ijl \mid 234)\,dq_{1234} \\
&\quad + \frac{1}{3!}\int \Phi^*(ijl \mid 123)\bar{\Phi}(ijl \mid 123)\,dq_{123} \\
&\quad \times\Big\{\sum_{\substack{s=1\\(s\neq ijl)}}^{N}\int \varphi_s^*(1)\Big[-\frac{z}{r_1}\Big]\varphi_s(1)\,dq_1 \\
&\quad + \frac{1}{2}\sum_{\substack{s=1\\(s,t\neq ijl)}}^{N}\sum_{t=1}^{N}\int\Big[\frac{|\varphi_s(1)|^2\,|\varphi_t(2)|^2}{r_{12}} \\
&\quad - \frac{\varphi_s(1)\varphi_s^*(2)\varphi_t^*(1)\varphi_t(2)}{r_{12}}\Big]\,dq_{12}\Big\}.
\end{aligned} \tag{4.6}$$

$$\int f^*(ij)f(jl)\,dq = \frac{1}{3!}\int \Phi^*(ijl \mid 123)\bar{\Phi}(ijl \mid 123)\,dq_{123}, \tag{4.7}$$

where dq means integration with respect to all N coordinates $q_1 \cdots q_N$.

We next turn our attention to (4.2). Let us define the 4-electron functions $\Phi(ijkl \mid 1234)$ and $\bar{\Phi}(ijkl \mid 1234)$ as follows:

$$\begin{aligned}
\Phi(ijkl \mid 1234) &= \bar{\Phi}(ij \mid 12)\mu(kl \mid 34) - \bar{\Phi}(ij \mid 13)\mu(kl \mid 24) + \bar{\Phi}(ij \mid 14)\mu(kl \mid 23) \\
&\quad + \bar{\Phi}(ij \mid 23)\mu(kl \mid 14) - \bar{\Phi}(ij \mid 24)\mu(kl \mid 13) + \bar{\Phi}(ij \mid 34)\mu(kl \mid 12),
\end{aligned} \tag{4.8}$$

and

$$\begin{aligned}
\bar{\Phi}(ijkl \mid 1234) &= \mu(ij \mid 12)\bar{\Phi}(kl \mid 34) - \mu(ij \mid 13)\bar{\Phi}(kl \mid 24) + \mu(ij \mid 14)\bar{\Phi}(kl \mid 23) \\
&\quad + \mu(ij \mid 23)\bar{\Phi}(kl \mid 14) - \mu(ij \mid 24)\bar{\Phi}(kl \mid 13) + \mu(ij \mid 34)\bar{\Phi}(kl \mid 12),
\end{aligned} \tag{4.9}$$

where $\bar{\Phi}(ij \mid 12)$ and $\bar{\Phi}(kl \mid 12)$ are defined according to (2.5) and

$$\mu(ab \mid 12) \equiv \begin{vmatrix}\varphi_a(1) & \varphi_a(2)\\ \varphi_b(1) & \varphi_b(2)\end{vmatrix}. \tag{4.10}$$

According to formulas (73)to (75) of A, the matrix component (4.2) is given by the following formulas

$$\begin{aligned}
&\int f^*(ij)\Big[\sum_{m=1}^{N}(-\tfrac{1}{2}\Delta_m)\Big]f(kl)\,dq \\
&= \frac{1}{3!}\int \Phi^*(ijkl \mid 1234)[-\tfrac{1}{2}\Delta_1]\bar{\Phi}(ijkl \mid 1234)\,dq_{1234}
\end{aligned}$$

$$+\frac{1}{4!}\int \Phi^*(ijkl \mid 1234)\bar{\Phi}(ijkl \mid 1234)\, dq_{1234}$$

$$\times\left\{\sum_{\substack{s=1\\(s\neq ijkl)}}^{N}\int \varphi_s^*(1)[-\tfrac{1}{2}\Delta_1]\varphi_s(1)\, dq_1\right\}. \qquad (4.11)$$

$$\int f^*(ij)\left[\sum_{m=1}^{N}\left(-\frac{z}{r_m}\right)+\frac{1}{2}\sum_{m,n=1}^{N}\frac{1}{r_{mn}}\right]f(kl)\, dq$$

$$=\frac{1}{4}\int \Phi^*(ijkl \mid 1234)\bar{\Phi}(ijkl \mid 1234)\frac{1}{r_{12}}\, dq_{1234}$$

$$+\frac{1}{3!}\int \Phi^*(ijkl \mid 1234)\bar{\Phi}(ijkl \mid 1234)\left[-\frac{z}{r_1}\right] dq_{1234}$$

$$+\frac{1}{3!}\int \sum_{\substack{s=1\\(s\neq ijkl)}}^{N}|\varphi_s(2)|^2\frac{1}{r_{12}}\Phi^*(ijkl \mid 1345)$$

$$\times\bar{\Phi}(ijkl \mid 1345)\, dq_{12345}$$

$$-\frac{1}{3!}\int \sum_{\substack{s=1\\(s\neq ijkl)}}^{N}\varphi_s(1)\varphi_s^*(2)\frac{1}{r_{12}}\Phi^*(ijkl \mid 1345)$$

$$\times\bar{\Phi}(ijkl \mid 2345)\, dq_{12345}$$

$$+\frac{1}{4!}\int \Phi^*(ijkl \mid 1234)\bar{\Phi}(ijkl \mid 1234)\, dq_{1234}$$

$$\times\left\{\sum_{\substack{s=1\\(s\neq ijkl)}}^{N}\int \varphi_s^*(1)\left[-\frac{z}{r_1}\right]\varphi_s(1)\, dq_1+\frac{1}{2}\sum_{\substack{s=1\\(s,t\neq ijkl)}}^{N}\sum_{t=1}^{N}\right.$$

$$\times\int\left[\frac{|\varphi_s(1)|^2\,|\varphi_t(2)|^2}{r_{12}}\right.$$

$$\left.\left.-\frac{\varphi_s(1)\varphi_s^*(2)\varphi_t^*(1)\varphi_t(2)}{r_{12}}\right]\right.$$

$$\left.\times\, dq_1\, dq_2\right\}, \qquad (4.12)$$

$$\int f^*(ij)f(kl)\, dq$$

$$=\frac{1}{4!}\int \Phi^*(ijkl \mid 1234)\bar{\Phi}(ijkl \mid 1234)\, dq_{1234}. \qquad (4.13)$$

5. ANALYSIS OF THE NONDIAGONAL MATRIX COMPONENTS

The formulas of the preceding section give the nondiagonal matrix components for atoms with any number of electrons, and for any configuration. However, these expressions are rather complicated for calculations of wave functions for larger atoms. It is therefore worthwhile to study the possibility of reducing the complexity of these expressions. As was mentioned in Sec. 2 the two-electron functions $\bar{\Phi}(ij \mid 12)$ are orthogonalized to all orbitals φ_s ($s = 1, 2, \cdots N$), except to φ_i and φ_j, by means of the orthogonality operator [Eqs. (2.5) and (2.6)]. It was shown in A and in I that this orthogonalization does not change the total wave function $f(ij)$ and therefore it is irrelevant for the physical situation which is described by the total wave function (1.1). It was shown recently[13] that second-order perturbation theory leads to the function (1.1), and it was suggested that the 2-electron function $\Phi(ij \mid 12)$ should be orthogonalized also to the orbitals φ_i and φ_j. However, in this case, the flexibility of $f(ij)$ becomes restricted. Let us discuss here the consequences of this approximation.

Let us recall here some results from A and I. We have shown there that the function

$$\bar{\Phi}(ij \mid 12) = [1-\Omega(ij \mid 1)][1-\Omega(ij \mid 2)]\Phi(ij \mid 12), \qquad (5.1)$$

where the operator $\Omega(ij \mid 1)$ is defined by (2.6), satisfies the orthogonality condition

$$\int \bar{\Phi}(ij \mid 12)\varphi_s^*(1)\, dq_1 \equiv 0, \quad \begin{bmatrix} s = 1, 2, \cdots N \\ s \neq ij \end{bmatrix}, \qquad (5.2)$$

regardless of the form of $\Phi(ij \mid 12)$. It was shown also that this orthogonalization does not change the total wave function $f(ij)$.

Theorem 5. The function

$$\tilde{\Phi}(ij \mid 1, 2) = [1-\Omega_T(1)][1-\Omega_T(2)]\Phi(ij \mid 1, 2), \qquad (5.3)$$

where

$$\Omega_T(1)F(1) = \sum_{s=1}^{N}\varphi_s(1)\int \varphi_s^*(2)F(2)\, dq_2, \qquad (5.4)$$

satisfies the orthogonality condition

$$\int \tilde{\Phi}(ij \mid 12)\varphi_s^*(1)\, dq_1 \equiv 0 \quad (s = 1, 2, \cdots N) \qquad (5.5)$$

regardless of the form of $\Phi(ij \mid 12)$.

Proof. We note that the only difference between $\Omega(ij \mid 1)$ and $\Omega_T(1)$ is that in $\Omega_T(1)$, φ_i and φ_j are added to the summation in the kernel of the operator Ω_T. In A and in I we have proved that (5.1) satisfies the condition (5.2). From the proof given there follows the proof of Theorem 5 in a trivial way.

Definition. We define an n-fold substitution configuration with respect to the Slater determinant Ψ_0 in the following way. Let $\varphi_1 \cdots \varphi_N, \varphi_{N+1} \cdots$ be a complete set of one-electron spin-orbitals. We call

[13] O. Sinanoglu, Proc. Roy. Soc. (London) **A260**, 379 (1961).

the first N orbitals from which the Slater determinant Ψ_0 is built the "basic set." An n-fold substitution configuration $1 \leq n \leq N$ is defined as

$$\begin{aligned}\mathfrak{F}(k_i k_j \cdots k_m) &= (N!)^{-1/2} \det [\varphi(1 \mid 1) \\ &\cdots \varphi(i-1 \mid i-1)\varphi(k_i \mid i)\varphi(i+1 \mid i+1) \\ &\cdots \varphi(m-1 \mid m-1)\varphi(k_m \mid m)\varphi(m+1 \mid m+1) \\ &\cdots \varphi(N \mid N)]. \end{aligned} \tag{5.6}$$

Obviously (5.6) may be obtained from

$$\Psi_0 = (N!)^{-1/2} \det [\varphi_1 \cdots \varphi_N] \tag{5.7}$$

by replacing the n orbitals $\varphi(i \mid q), \varphi(j \mid q), \cdots \varphi(m \mid q)$ by the orbitals $\varphi(k_i \mid q), \varphi(k_j \mid q), \cdots \varphi(k_m \mid q)$, where the latter are chosen from the complete set *excluding* the basic set.

Theorem 6. The function (1.1) contains besides the Slater determinant Ψ_0 all single- and double-substitution configurations.

Proof. Let us expand $\Phi(ij \mid 12)$ in terms of products built from the complete set

$$\Phi(ij \mid 12) = \sum_{k_i k_j=1}^{\infty} c(k_i k_j)\varphi(k_i \mid 1)\varphi(k_j \mid 2). \tag{5.8}$$

Multiplication of both sides with $\frac{1}{2}(1 - P_{12})$, where P_{12} interchanges q_1 and q_2, yields

$$\Phi(ij \mid 12) = \tfrac{1}{2} \sum_{k_i k_j=1}^{\infty} c(k_i k_j)\mu(k_i k_j \mid 12), \tag{5.9}$$

where μ was defined by (4.10). On putting (5.9) into (1.2), we realize that $\varphi(k_i \mid q)$ or $\varphi(k_j \mid q)$ can not be one of the orbitals $\varphi_1 \cdots \varphi_{i-1}\varphi_{i+1} \cdots \varphi_{j-1}\varphi_{j+1} \cdots \varphi_N$ since then we would have a determinant with two identical rows. We obtain therefore, taking into account that $c(ab) = -c(ba)$ and $\mu(ab \mid 12) = -\mu(ba \mid 12)$, the following formula:

$$\begin{aligned}\Phi(ij \mid 12) &= c(ij)\mu(ij \mid 12) + \sum_{k_j=N+1}^{\infty} c(ik_j)\mu(ik_j \mid 12) \\ &+ \sum_{k_i=N+1}^{\infty} c(k_i j)\mu(k_i j \mid 12) \\ &+ \tfrac{1}{2} \sum_{k_i k_j=N+1}^{\infty} c(k_i k_j)\mu(k_i k_j \mid 12) \end{aligned} \tag{5.10}$$

On putting (5.10) into (1.2) and the resulting expression in (1.1) and using the notation (5.6) we obtain

$$\begin{aligned}\Psi = \Psi_0 &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} c(ij)\Psi_0 \\ &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k_i=N+1}^{\infty} c(ik_i)\mathfrak{F}^{(1)}(k_i) \\ &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k_i=N+1}^{\infty} c(k_i j)\mathfrak{F}^{(1)}(k_i) \\ &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{2} \sum_{k_i k_j=N+1}^{\infty} c(k_i k_j)\mathfrak{F}^{(2)}(k_i k_j). \end{aligned} \tag{5.11}$$

Let us consider in the third and fourth expression that function in which the Ath orbital is substituted, i.e., $\mathfrak{F}^{(1)}(k_A)$, $(A = 1, 2, \cdots N)$. For the coefficient of this function we obtain from the third expression

$$\sum_{k_A} \sum_{i=1}^{A-1} c(ik_A) \tag{5.12a}$$

and from the fourth expression we obtain

$$\sum_{k_A} \sum_{j=A+1}^{N} c(k_A j). \tag{5.12b}$$

Therefore, the sum of the third and fourth expression may be written in the following form:

$$\sum_{A=1}^{N} \sum_{k_A=N+1}^{\infty} a(k_A)\mathfrak{F}^{(1)}(k_A), \tag{5.13}$$

where

$$a(k_A) = \sum_{i=1}^{A-1} c(ik_A) + \sum_{i=A+1}^{N} c(k_A i). \tag{5.14}$$

Introducing the notations $1 + \sum_{(i,j)} c(ij) \equiv a_0$, $c(AB) \equiv a(AB)$, and taking into account that $a(AB) = -a(BA)$ and $\mathfrak{F}^{(2)}(AB) = -\mathfrak{F}^{(2)}(BA)$ we obtain

$$\begin{aligned}\Psi &= a_0\Psi_0 + \sum_{A=1}^{N} \sum_{k_A=N+1}^{\infty} a(k_A)\mathfrak{F}^{(1)}(k_A) \\ &+ \sum_{A=1}^{N} \sum_{B=A+1}^{N} \sum_{k_A=N+1}^{\infty} \sum_{k_B=k_A+1}^{\infty} a(k_A k_B)\mathfrak{F}^{(2)}(k_A k_B). \end{aligned} \tag{5.15}$$

The first term on the right side is the Hartree-Fock function, the second and third contain all single- and double-substitution configurations. Theorem 6 is therefore proved.

Theorem 7. The orthogonalization of the two-electron function $\Phi(ij \mid 12)$ with respect to φ_i and φ_j removes all single substitution configurations from the total wave function Ψ.

Proof. As we have seen above, the orthogonalization is carried out by means of the orthogonality operator. Let Ω_i be the orthogonality operator with respect to the orbital φ_i, and let us consider

$$\begin{aligned}\Phi(ij \mid 12) &= [1 - \Omega_i(1) - \Omega_j(1)] \\ &\times [1 - \Omega_i(2) - \Omega_j(2)]\Phi(ij \mid 12), \end{aligned} \tag{5.16}$$

where

$$\Omega_i(1)f(1) = \varphi_i(1)\int \varphi_i^*(2)f(2)\,dq_2. \tag{5.17}$$

Writing Φ in the form given by (5.9) we obtain

$$\hat{\Phi}(ij \mid 12) = \tfrac{1}{2}\sum_{k_ik_j=N+1}^{\infty} c(k_ik_j)\mu(k_ik_j \mid 12) \tag{5.18}$$

Comparison with (5.10) shows that those terms which yielded the single substitution configurations disappeared from $\hat{\Phi}$. This proves Theorem 7.

The next theorem will show what will be the effect of orthogonalizing $\Phi(ij \mid 12)$ to φ_i and φ_j on the nondiagonal matrix components.

Theorem 8. If the orthogonality condition (5.5) is introduced, the nondiagonal matrix components (4.1) and (4.2) contain only those integrals which occur in the diagonal matrix components (2.2)–(2.4). In other words, in this case, all integrals which occur in the nondiagonal matrix components can be reduced to the 6 basic integrals (3.1)–(3.4), (3.33), and (3.42).

Proof. Let us first consider the expressions (4.5)–(4.7). Those expressions were derived by assuming that the two-electron functions satisfy the orthogonality condition (5.2). If we introduce the stronger condition (5.5) the only change in the formulas (4.5)–(4.7) will be that the 2-electron function $\tilde{\Phi}$ must be replaced by $\tilde{\Phi}$. Let us replace $\tilde{\Phi}$ by $\tilde{\Phi}$ in the 3-electron functions (4.3) and (4.4) and let us denote the resulting expressions by Λ and $\bar{\Lambda}$. Since the operators which occur in the matrix components (4.5)–(4.7) are 2 electron, 1 electron, and constant operators let us integrate in (4.5)–(4.7) with respect to q_3, (q_3q_2), and $(q_3q_2q_1)$, respectively. We obtain

$$\begin{aligned}
&\int \Lambda^*(ijl \mid 123)\bar{\Lambda}(ijl \mid 123)\,dq_3 \\
&= \int \tilde{\Phi}^*(ij \mid 13)\tilde{\Phi}(jl \mid 13)\varphi_i^*(2)\varphi_i(2)\,dq_3 \\
&\quad - \int \tilde{\Phi}^*(ij \mid 13)\tilde{\Phi}(jl \mid 23)\varphi_i^*(2)\varphi_i(1)\,dq_3 \\
&\quad - \int \tilde{\Phi}^*(ij \mid 23)\tilde{\Phi}(jl \mid 13)\varphi_i^*(1)\varphi_i(2)\,dq_3 \\
&\quad + \int \tilde{\Phi}^*(ij \mid 23)\tilde{\Phi}(jl \mid 23)\varphi_i^*(1)\varphi_i(1)\,dq_3.
\end{aligned} \tag{5.19}$$

$$\begin{aligned}
&\int \Lambda^*(ijl \mid 123)\bar{\Lambda}(ijl \mid 123)\,dq_2\,dq_3 \\
&= \int \tilde{\Phi}^*(ij \mid 23)\tilde{\Phi}(jl \mid 23)\varphi_i^*(1)\varphi_i(1)\,dq_2\,dq_3,
\end{aligned} \tag{5.20}$$

$$\int \Lambda^*(ijl \mid 123)\bar{\Lambda}(ijl \mid 123)\,dq_1\,dq_2\,dq_3 = 0. \tag{5.21}$$

Putting (5.19)–(5.21) into (4.5)–(4.7) we obtain

$$\begin{aligned}
&\int \tilde{f}^*(ij)\Big[\sum_{m=1}^{N}(-\tfrac{1}{2}\Delta_m)\Big]\tilde{f}(jl)\,dq \\
&\quad = \frac{1}{2}\int \tilde{\Phi}^*(ij \mid 23)\tilde{\Phi}(jl \mid 2\,3)\,dq_2\,dq_3 \\
&\qquad \times \int \varphi_i^*(1)[-\tfrac{1}{2}\Delta_1]\varphi_i(1)\,dq_1,
\end{aligned} \tag{5.22}$$

$$\begin{aligned}
&\int \tilde{f}^*(ij)\Big[\sum_{m=1}^{N}\Big(-\frac{z}{r_m}\Big) + \frac{1}{2}\sum_{m,n=1}^{N}\frac{1}{r_{mn}}\Big]\tilde{f}(jl)\,dq \\
&= \int \tilde{\Phi}^*(ij \mid 13)\tilde{\Phi}(jl \mid 13)\varphi_i^*(2)\varphi_i(2)\frac{1}{r_{12}}\,dq_{123} \\
&\quad - \int \tilde{\Phi}^*(ij \mid 13)\tilde{\Phi}(jl \mid 23)\varphi_i^*(2)\varphi_i(1)\frac{1}{r_{12}}\,dq_{123} \\
&\quad + \frac{1}{2}\int \tilde{\Phi}^*(ij \mid 23)\tilde{\Phi}(jl \mid 23)\,dq_{23} \\
&\quad \times \Big\{\int \varphi_i^*(1)\Big[-\frac{z}{r_1}\Big]\varphi_i(1)\,dq_1 \\
&\quad + \sum_{\substack{s=1\\(s\neq i\,j\,l)}}^{N}\int \frac{|\varphi_s(2)|^2\,\varphi_i^*(1)\varphi_i(1)}{r_{12}}\,dq_{12} \\
&\quad - \sum_{\substack{s=1\\(s\neq i\,j\,l)}}^{N}\int \frac{\varphi_s(1)\varphi_i^*(2)\varphi_i^*(1)\varphi_i(2)}{r_{12}}\,dq_{12}\Big\}.
\end{aligned} \tag{5.23}$$

$$\int \tilde{f}^*(ij)\tilde{f}(jl)\,dq = 0, \tag{5.24}$$

where $\tilde{f}(ij)$ denotes the total wave function containing $\tilde{\Phi}$, i.e., denotes the total wave function in which the flexibility of Φ is restricted by orthogonalizing it to φ_i and φ_j.

We turn now our attention to the nondiagonal component (4.2). We replace Φ by $\tilde{\Phi}$ in the 4-electron functions (4.8) and (4.9) and denote the resulting expressions by Λ and $\bar{\Lambda}$. Taking into account the orthogonality condition (5.5) and the orthogonality of the one-electron spin-orbitals, we obtain

$$\begin{aligned}
&\int \Lambda^*(ijkl \mid 1234)\bar{\Lambda}(ijkl \mid 1234)\,dq_3\,dq_4 \\
&\quad = \int \tilde{\Phi}^*(ij \mid 34)\tilde{\Phi}(kl \mid 34)\,dq_{34} \\
&\qquad \times \mu^*(ij \mid 12)\mu(kl \mid 12),
\end{aligned} \tag{5.25}$$

$$\int \Lambda^*(ijkl \mid 1234)\bar{\Lambda}(ijkl \mid 1234)\,dq_2\,dq_3\,dq_4 = 0. \tag{5.26}$$

where μ was defined by (4.10). By means of (5.25) and (5.26) we obtain from (4.11)–(4.13)

$$\int \bar{f}^*(ij)\Big[\sum_{m=1}^{N}(-\tfrac{1}{2}\Delta_m)\Big]\bar{f}(kl)\,dq = 0, \tag{5.27}$$

$$\int \bar{f}^*(ij)\Big[\sum_{m=1}^{N}\Big(-\frac{z}{r_m}\Big)+\frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N}\frac{1}{r_{mn}}\Big]\bar{f}(kl)\,dq$$
$$= \frac{1}{4}\int \bar{\Phi}^*(ij \mid 34)\bar{\Phi}(kl \mid 34)\,dq_{34}$$
$$\times \int \frac{\mu^*(ij \mid 12)\mu(kl \mid 1,2)}{r_{12}}\,dq_{12}, \tag{5.28}$$

$$\int \bar{f}^*(ij)\bar{f}(kl)\,dq = 0. \tag{5.29}$$

We can now prove Theorem 8 by simply comparing the expressions (5.22)–(5.24) and (5.27)–(5.29) with the formulas for the diagonal matrix components (2.2)–(2.4). We realize that: (1) (5.22) has the same form as the second integral in (2.2); (2) the first two integrals of (5.23) have the same form as the third and fourth expressions of (2.3), and the remaining terms of (5.23) have the same form as the last term in the expression (2.3); (3) the integral (5.28) has again the same form as the last term of (2.3). This completes the proof of Theorem 8.

6. DISCUSSION

We have shown in this paper that the matrix components of the Hamiltonian with respect to correlated wave functions of second order have the following properties:

(1) The diagonal matrix components can be reduced to the basic integrals I_1–I_6.

(2) If the 2-electron functions $\Phi(ij \mid 12)$ are orthogonalized to the orbitals φ_i and φ_j, the nondiagonal matrix components can also be reduced to the basic integrals I_1–I_6.

We have shown in A and in I that the total wave function (1.2) does not change if we orthogonalize the 2-electron function $\Phi(ij \mid 12)$ to all orbitals $\varphi_1 \cdots \varphi_N$ except φ_i and φ_j. However, as was shown in the present paper, if $\Phi(ij \mid 12)$ is orthogonalized also to φ_i and φ_j, the flexibility of the total wave function becomes restricted. In order to investigate the effect of the "(ij) orthogonalization" on the total wave function, we have established Theorem 7. The statement of Theorem 7 was that the (ij) orthogonalization removes all single substitution configurations from the 2-electron function $\Phi(ij \mid 12)$.

Evidently, it is not possible at the present time to understand precisely the meaning of Theorem 7. As was pointed out in I, we can set up the 2-electron function Φ in two different ways. We can set up Φ either as a finite set of (2×2) Slater determinants or as a Hylleraas-type function containing r_{12} [Formula (1.3)]. Theorem 7 shows that, in the first case, there can be no single substitution configuration among the Slater determinants. There is some evidence,[14] that in certain cases the effect of the single-substitution configurations is negligible compared with the terms which represent many-electron correlations. However, there is no general theorem which would show that the single-substitution configurations are always negligible. The situation is even more complicated if Φ is a Hylleraas-type function. It is not known at the present time what will be the effect of the "(ij) orthogonalization" in this case. We may summarize the situation by by saying that the "(ij) orthogonalization" greatly simplifies the nondiagonal matrix components and therefore greatly simplifies the calculations; but the flexibility of the wave function becomes restricted. *We emphasize, that full, "ab initio" calculations must be based upon the wave function (1.1) without the "(ij) orthogonalization."* The nondiagonal matrix components for this case are given in Sec. 4.

In the forthcoming part of this series we shall investigate the basic integrals I_1 to I_6. We shall show, that the integrals I_1 to I_4 and I_6 can be given in closed form, whereas I_5 will be given in the form of an infinite series. For the integrals I_1 to I_4 and I_6 general expressions will be derived, and general expressions will be also derived for the terms of the series of I_5. The expressions which will be given for the integrals will be valid for *arbitrary* correlation factors, i.e., for arbitrary values of the parameters $\kappa, \cdots \rho$ which characterize the correlation part of the integrals. For the practical calculation of the integrals we have developed a method, which will enable us to calculate them regardless of the form of the radial part of the one-electron spin-orbitals which occur in the integrals.

What we mean by this is the following. Determinantal wave functions are available now for a large number of atoms and ions. The radial functions, which represent the radial parts of the one-electron spin-orbitals which occur in the determinantal functions, belong to three main types.

For light atoms, one can obtain a very good approximation to the empirical energy by using simple analytical expressions for the radial parts of the one-electron spin-orbitals. For heavier atoms, how-

[14] R. E. Watson, Phys. Rev. **119**, 170 (1960).

ever, good approximations to the empirical energy can be obtained only by solving the Hartree-Fock equations. The Hartree-Fock equations can be solved numerically, or one can obtain (approximate) solutions by analytical methods. In the first case the radial functions will be given in the form of tables; in the second case they will be given as linear combinations of a large number of analytical functions. Thus, the following three types of radial functions may occur in the basic integrals:

(1) Simple analytical functions,
(2) Numerical Hartree-Fock functions,
(3) Analytical (Roothaan-type) Hartree-Fock functions.

The method which we have developed for the calculation of the basic integrals will apply to all three cases. In other words, we shall be able to calculate these integrals, regardless of whether the radial functions belong to the first, second, or third group given above. Consequently, we shall be able to calculate the basic integrals for light atoms (where the radial functions are simple analytical expressions) as well as for atoms with large number of electrons (where the radial function are numerical Hartree-Fock, or Roothaan-type functions). This means that the method of correlated wave functions which we have been developing in this series, will not be restricted to the calculation of wave functions for light atoms, but *we shall be able to apply the theory to the calculation of wave functions for atoms with any number of electrons.*

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Hard Core Produced by Orthogonality Constraints*

S. TANI AND D. A. UHLENBROCK
Physics Department, New York University, University Heights, New York
(Received May 28, 1962)

It is shown that the effect of a hard core as an external force can be reproduced by requirements of orthogonality alone. A fictitious orthonormal set of functions is introduced, which form a complete set in the domain occupied by the hard core. The actual wave function is constructed as the eigenfunction of a modified kinetic energy operator, and is required to be orthogonal to all members of the fictitious set. The limit of infinitely many constraints is carefully discussed. It turns out to be impossible to regard the hard core effect as equivalent to some kind of "potential" acting on free waves (for which the Hamiltonian would be well defined). However, the eigenfunction is well defined in the limit of infinitely many constraints, and is independent of the way in which the fictitious set is chosen.

1. INTRODUCTION

THE presence of a bound state affects the behavior of the phase shift in the same channel in a definite way. We shall denote the phase shift as a function of momentum p by $\eta(p)$. For the nonrelativistic potential scattering problem, Levinson has shown that

$$\eta(0) - \eta(\infty) = N\pi, \tag{1}$$

where N is the number of bound states. This theorem and other related problems have been studied by various authors.[1] The spectrum of scattering states is continuous, but it makes sense to interpret the theorem as if an analogy to a linear space of finite dimension is valid. As the eigenstates of the total Hamiltonian, the scattering states are orthogonal to each bound state; therefore, the formation of each bound state reduces, so to speak, the dimension of the space of scattering states by one, and Levinson's theorem reflects this fact. Two comments can be made at this point. First, the above interpretation can be made plausible by considering the formation of a bound state as the attractive potential becomes deeper in a system enclosed in a finite box. The lowest positive energy level drops to a negative one at a certain depth of the potential. A negative energy level is affected only weakly by the size of the box; and it tends to the level of a bound state in the limit of an infinite box. Secondly, it has been shown in reference I that the Hilbert space orthogonal to all bound states is sufficient to discuss the scattering,[2] and that the constraint of orthogonality readily leads to Levinson's theorem, Eq. (1).

It is the purpose of this paper to apply the above idea to the derivation of the phase shift caused by a hard core. The use of orthogonality constraints imposed on the relevant Hilbert space is basic in this approach. If there is a hard core, the region inside the core is not available for the physical system. Roughly speaking, the Hilbert space in the presence of a hard core is narrower than when the hard core is absent. The Hilbert space is made narrower when an orthogonality constraint is imposed on it. We introduce a fictitious complete set of functions in the region occupied by the core, and require that a physical wave function be orthogonal to all of these functions. Then the wave function will vanish inside the core, since it cannot have any square integrable component there. Under the orthogonality constraints, the kinetic energy operator will be modified in the way which has been discussed in reference I, and the phase shift is obtained for each eigenfunction of the modified kinetic energy. The phase shift which comes from the orthogonality alone turns out to be able to reproduce the hard core phase shift. The phase shift of the same nature in the bound-state problem explains Levinson's

* The research reported in this paper was supported by the National Science Foundation.

[1] N. Levinson, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **25**, No. 9 (1949); P. Swan, Proc. Roy. Soc. (London) **A228**, 10 (1955); J. M. Jauch, Helv. Phys. Acta **30**, 143 (1957); R. Haag, Nuovo cimento **5**, 203 (1957); A. Martin, N. *ibid.* **7**, 607 (1958); M. Ida, Progr. Theoret. Phys. (Kyoto) **21**, 625 (1959); S. Tani, Phys. Rev. **117**, 252 (1960); the last paper will be referred to as I in the text.

[2] The following conjecture is most likely to be valid: The Born series does not converge over a certain energy range near zero, if there is a bound state. However, if the projection to the space orthogonal to the bound state is made and one starts with the eigenfunction of the modified Hamiltonian, the Born series becomes applicable again. This can be confirmed in the example of a separable potential

$$(k\,|V|\,k') = -\lambda k(M^2 + k^2)^{-1/2}\,k'(M^2 + k'^2)^{-1/2},$$

where λ is the strength parameter. A bound state exists for $\lambda\pi/4M^3 > 1$. The projection separates the part of the wave function to which the Born series cannot apply. (See Appendix I in reference I.)

theorem. Thus, an analog to Levinson's theorem exists for the hard core phase shift where N in Eq. (1) tends to infinity.

The limit of infinitely many constraints actually needs a close investigation. We start with a system of free waves. When the number of constraints is finite, the Hamiltonian of the system is given by the sum of the kinetic energy and a nonlocal potential. In the limit of infinitely many constraints, a part of the potential diverges although the eigenfunction is well defined. This only means that it is difficult to regard the effect of a hard core as equivalent to a scattering potential, yet the eigenfunction of the system is well defined. If our eigenfunctions are used as the basic, instead of free waves, the Hamiltonian becomes well defined. It is satisfactory that the completeness of the set of our eigenfunctions is different from that for free waves such that it is taken into account that there is no particle inside the core.

The use of orthogonality constraints and the derivation of the Hamiltonian are discussed in Sec. 2. The phase shift is dealt with in Sec. 3. The way the limit of infinitely many constraints is taken is discussed in detail in Sec. 4.

2. ORTHOGONALITY CONSTRAINTS AND HAMILTONIAN

The radius of the core will be denoted by a. Since the actual wave function does not penetrate into the hard core domain, it should be independent of the particular choice of the fictitious functions in the core domain, otherwise the result depends on the mathematical trick used, a situation unsatisfactory from the physical point of view. A set of eigenfunctions is defined if the Schrödinger equation in the domain $a \geq r \geq 0$ is set up. One may introduce a finite local potential and impose an arbitrary boundary condition at $r = a$. At $r = 0$ we require the wave function to vanish. One set defined for one potential is equivalent to another set for another potential (under the fixed boundary condition), in the sense that the same square integrable function can be expanded with use of either set. Then it ought to be shown that any set of eigenfunctions for an arbitrary potential will give the same result in producing the hard core effects. By the same token, it ought to be shown that any boundary condition will give the same result. However we have to begin with the orthogonality to a finite number of functions and later take the limit of infinitely many constraints. Unless the uniform convergence to the limit is guaranteed, it may happen that the limit depends on the special choice of the set of functions. This happens in writing down the Hamiltonian, yet in other respects the limit gives reasonable results. We postpone the discussion concerning the choice of a complete set until Sec. 4 and start with a particular set to show that our idea works in many respects.

Our discussions will be restricted to the S wave, the generalization to other cases being obvious. The radial wave function will be dealt with as in a one-dimensional problem in the domain $0 \leq r \leq \infty$ by considering $\Psi = rU$, where U is the radial wave function in the original sense. In the absence of a hard core, the eigenfunction of the kinetic energy is given by the free wave

$$\Psi_k(r) = (2/\pi)^{1/2} \sin(kr), \qquad r \geq 0. \tag{2}$$

It is convenient to employ the second quantization in momentum space in order to deal with some transformations. We assume Bose statistics, but the same conclusions follow with use of Fermi statistics. Choosing the unit of energy such that the kinetic energy is k^2 when the momentum is k, the Hamiltonian is given by

$$H_0 = \int_0^\infty dk\, k^2 a_k^\dagger a_k. \tag{3}$$

The operator $a_k^\dagger$ creates a particle in the state Ψ_k, (2).

The following fictitious set of functions will be chosen inside $r = a$:

$$g_n(r) = \begin{cases} (2/a)^{1/2} \sin(k_n r), & 0 \leq r < a, \\ 0, & r > a, \end{cases}$$

$$k_n = (\pi/a)n, \qquad (n = 1, 2, \cdots). \tag{4}$$

This is an eigenfunction of the free wave Schrödinger equation under the boundary condition

$$g_n(a) = 0. \tag{5}$$

In momentum space we have

$$f_n(k) = (\Psi_k, g_n) = c_n k_n \sin(ak)/(k^2 - k_n^2), \tag{6}$$

where

$$c_n = (-1)^n 2(\pi a)^{-1/2}. \tag{6'}$$

The projection operator onto the f_n will be denoted by Λ_n. Its matrix element is

$$(k\,|\Lambda_n|\,k') = f_n(k) f_n(k'). \tag{7}$$

When the orthogonality to the first N members of the set is introduced, we make use of the notation

$$\Lambda^{(N)} = \sum_{n=1}^{N} \Lambda_n. \tag{8}$$

Then $\Lambda^{(N)}$ is the projection operator onto the subspace spanned by the first N members of the set. The modified Hamiltonian is given by

$$H^{(N)} = \int_0^\infty dk\, dk' \times (k\,|(1-\Lambda^{(N)})W(1-\Lambda^{(N)})|\,k')a_k^+ a_{k'}, \tag{9}$$

where W denotes the diagonal matrix defined by

$$(k\,|W|\,k') = k^2\,\delta(k-k'). \tag{10}$$

We introduce a matrix $V^{(N)}$ by

$$(k\,|V^{(N)}|\,k') = -(k^2+k'^2)(k\,|\Lambda^{(N)}|\,k') + (k\,|\bar{\Lambda}^{(N)}|\,k'), \tag{11}$$

where we have put

$$(k\,|\bar{\Lambda}^{(N)}|\,k') = \int dk''(k\,|\Lambda^{(N)}|\,k'')(k''\,|\Lambda^{(N)}|\,k')k''^2 = \sum_{m,n=1}^{N}\left(\int_0^\infty f_m(k'')f_n(k'')k''^2\,dk''\right)f_m(k)f_n(k'). \tag{12}$$

Then the Hamiltonian $H^{(N)}$ can be written as

$$H^{(N)} = H_0 + V^{(N)}, \tag{13}$$

where

$$V^{(N)} = \int dk\,dk'\,(k\,|V^{(N)}|\,k')a_k^+ a_{k'}. \tag{13'}$$

An eigenfunction of $H^{(N)}$ is a scattering state[3] in the field of the nonlocal potential $V^{(N)}$. The phase shift caused by $V^{(N)}$ is the orthogonality phase shift studied in reference I. It should be noted that the Hamiltonian $H^{(N)}$ is not equivalent to H_0, although their spectra are the same; there is no unitary transformation which transforms H_0 into $H^{(N)}$.

In order to understand the change in "dimensionality" of the Hilbert space, it is instructive to follow the transition from the system of free waves into the system restricted by the orthogonality more closely. We follow steps similar to those of reference I. In order to handle a normalizable wave packet we introduce into the system of free waves some redundant variables b_n which satisfy the commutation relations

$$[b_n, b_m^+] = \delta_{nm} \qquad (n, m = 1, 2, \cdots, N). \tag{14}$$

They are redundant, since they do not appear in the Hamiltonian H_0. Furthermore, the following subsidiary conditions are imposed on a physical state vector Φ of the modified Hilbert space:

$$b_n^+ b_n \Phi = 0, \qquad (n = 1, 2, \cdots, N). \tag{15}$$

It is obvious that this system is completely equivalent to the system of free waves. Next, we perform the unitary transformation

$$\Phi \rightarrow \Phi' = U^{-1}\Phi, \tag{16}$$

where

$$U = \prod_{n=1}^{N} U_n,$$

$$U_n = \exp\left[\frac{\pi}{2} i\left(b_n^+ \int f_n(k)a_k\,dk + \text{conj}\right)\right]. \tag{17}$$

In the new representation the number of particles with the wave function $f_n(k)$ is given by the value of the number operator $b_n^+ b_n$. The transformed Hamiltonian is given by

$$H' = U^{-1}HU = H_a + H_{ab} + H_b. \tag{18}$$

The part H_a is concerned exclusively with the a operators, and is given by $H^{(N)}$, (13). The part H_b contains only the b operators and is of the form

$$H_b = \sum_{n=1}^{N} k_n^2 b_n^+ b_n. \tag{19}$$

The part H_{ab} is the cross term and is given by

$$H_{ab} = i\sum_{n=1}^{N}\int dk\, a_k^+ \times\left(k^2 - \int f_n(k')^2 k'^2\,dk'\right)f_n(k)b_n + \text{conj}. \tag{20}$$

The transformed subsidiary condition reads

$$\int a_k^+ f(k)\,dk \int a_{k'} f(k')\,dk'\Phi' = 0, \qquad (n = 1, 2, \cdots, N). \tag{21}$$

A particle described by a b operator is confined inside $r = a$ and should be neglected in the presence of a hard core. This means we have to truncate the Hamiltonian (18). In the presence of H_{ab}, the subspace spanned by the eigenfunctions of H_a cannot be disconnected from the subspace spanned by the eigenfunctions of the H_b. In contradistinction to the bound state problem, the cross term H_{ab} cannot be made to vanish by a proper choice of the

[3] There is also an $f_n(k)$ which is an eigenfunction with the eigenvalue zero; they are suppressed by the subsidiary conditions (21).

f_n.[4] Here we must discard H_{ab} as one of necessary devices in going over from the system of free waves to the system with a hard core. Obviously we must discard H_b also. Therefore we are led to consider the eigenfunctions of the Hamiltonian $H^{(N)}$, (13), subject to the subsidiary conditions, (21). Now we have seen that $H^{(N)}$ cannot be obtained from H_0 by a unitary transformation alone. It is to be noted that a wave function $f_n(k)$ is an eigenfunction of the Hamiltonian $H^{(N)}$ with the eigenvalue zero. But a state with such a wave function is forbidden by the subsidiary conditions, (21). A permissible state must be orthogonal to the first N members of the complete set. The Hilbert space is made narrower by suppressing the subspace pertaining to b particles after the transformation (16).

3. PHASE SHIFT AND EIGENFUNCTION

Suppose a one-particle eigenstate of $H^{(N)}$ with energy p^2 is given by

$$\Phi_p = \int dk\, a_k^+ h^{(N)}(k, p)\Phi_0, \tag{22}$$

where Φ_0 is the vacuum state. The Schrödinger equation for $h^{(N)}$ is given by

$$\int dk'(k\,|(1 - \Lambda^{(N)})W(1 - \Lambda^{(N)})|\,k')h^{(N)}(k', p) = p^2 h^{(N)}(k, p). \tag{23}$$

That $h^{(N)}$ is orthogonal to the f_n can be shown as follows: taking the scalar product with $f_n(k)$ $(n = 1, 2, \cdots, N)$ from the left we have

$$\int dk\, dk'\, f_n(k)(k\,|(1 - \Lambda^{(N)})W(1 - \Lambda^{(N)})|\,k') \times h^{(N)}(k', p) = 0 = p^2 \int f_n(k)h^{(N)}(k, p). \tag{24}$$

The second statement follows from a property of the projection operator Λ_n or $\Lambda^{(N)}$, namely,

$$\int dk\, f_n(k)(k\,|\Lambda^{(N)}|\,k') = f_n(k'). \tag{25}$$

Therefore, when p does not vanish

$$\int dk\, f_n(k)h^{(N)}(k, p) = 0, \tag{26}$$

or

$$\int (k'\,|\Lambda^{(N)}|\,k)h^{(N)}(k, p) = 0. \tag{26'}$$

Consequently, the Schrödinger equation can be simplified to

$$(k^2 - p^2)h^{(N)}(k, p) - \sum_{n=1}^{N} f_n(k) \int dk'\, f_n(k')k'^2 h^{(N)}(k', p) = 0. \tag{27}$$

The solution of (27) under the standing wave boundary condition can be put in the form

$$h^{(N)}(k, p) = \delta(k - p) + (k^2 - p^2)^{-1} \sum_n f_n(k)X_n(p). \tag{28}$$

From (26) we have a set of linear equations to determine $X_n(p)$

$$\sum_{n=1}^{N} G_{mn}(p)X_n(p) + f_m(p) = 0, \qquad (m = 1, 2, \cdots, N). \tag{29}$$

where we have put

$$G_{mn}(p) = \int dk\, f_m(k)f_n(k)(k^2 - p^2)^{-1}. \tag{30}$$

Substituting (6) into (30), the explicit form for $G_{mn}(p)$ is

$$G_{mn}(p) = (-1)^{m+n} \frac{k_n}{k_n^2 - p^2} \frac{k_m}{k_m^2 - p^2} \frac{\sin(2ap)}{ap} + \delta_{mn} \frac{1}{k_m^2 - p^2}. \tag{31}$$

Equation (31) applies for all p and there is no singularity with respect to p. The coefficient $X_n(p)$ in (28) is

$$X_n(p) = -D^{(N)}(k_1, \cdots, k_N)^{-1}\, D_n^{(N)}(k_1, \cdots, k_N) \tag{32}$$

by solving (29). In (32) $D^{(N)}$ stands for the determinant

$$D^{(N)}(k_1, \cdots, k_N) = \det\,[G_{mn}(p)], \tag{33}$$

and $D_n^{(N)}$ means the determinant of the matrix, which obtains from $[G_{mn}(p)]$ by replacing its nth column by the one-row matrix $[f_n(p)]$. It is easy to evaluate $D^{(N)}$ and $D_n^{(N)}$; we find

$$D^{(N)}(k_1, \cdots, k_N) = \prod_{n=1}^{N} (k_n^2 - p^2)^{-1} \times \left[1 + \sum_{n=1}^{N} k_n^2(k_n^2 - p^2)^{-1} \sin(2pa)/pa\right] \tag{34}$$

[4] It can be shown that H_{ab} vanishes identically only when $f_n(k)$ is an eigenfunction of the original Hamiltonian. A wave packet confined to the domain $r \leq a$ cannot be an eigenfunction of the original Hamiltonian H_0.

and

$$D_n^{(N)}(k_1, \cdots, k_N) = -\prod_{n=1}^{N} (k_n^2 - p^2)^{-1}(-1)^n 2(\pi a)^{1/2} k_n \sin(ap). \quad (35)$$

Therefore the coefficient $X_n(p)$ is

$$X_n(p) = (-1)^n 2k_n(\pi a)^{1/2}\Big[1 + \sum_{n=1}^{N} k_n^2(k_n^2 - p^2)^{-1} \times \sin(2pa)/pa\Big]^{-1} \sin(ap). \quad (36)$$

With the $X_n(p)$ thus determined, both the phase shift and the wave function can be calculated explicitly.

The formula for the phase shift $\eta(p)$ is

$$\tan \eta^{(N)}(p) = \frac{\pi}{2p} \sum_{n=1}^{N} f_n(p) X_n(p), \quad (37)$$

which follows from (28) by the standard method.[5] Explicitly we have

$$\tan \eta^{(N)}(p) = -2 \sin^2(pa)\Gamma^{(N)}(p) \times [pa + \sin(2pa)\Gamma^{(N)}(p)]^{-1}, \quad (38)$$

where we have defined

$$\Gamma^{(N)}(p) = \sum_{n=1}^{N} k_n^2(k_n^2 - p^2)^{-1} = \sum_{n=1}^{N} (n\pi)^2[(n\pi)^2 - (pa)^2]^{-1}. \quad (39)$$

From the theory of functions we have

$$\sum_{n=1}^{\infty} [(pa - n\pi)^{-1} + (pa + n\pi)^{-1}] + (pa)^{-1} = \cot(pa),$$

or

$$\sum_{n=1}^{\infty} [(pa)^2 - (n\pi)^2]^{-1} = [\cot(pa) - (pa)^{-1}]/2pa. \quad (40')$$

Therefore, we may write $\Gamma^{(N)}(p)$ as:

$$\Gamma^{(N)}(p) = N - \sum_{n=1}^{N} (pa)^2[(pa)^2 - (n\pi)^2]^{-1} = N + \tfrac{1}{2} + \sum_{n=N+1} (pa)^2[(pa)^2 - (n\pi)^2]^{-1} - \tfrac{1}{2}pa \cot(pa) = \Delta^{(N)}(p) - \tfrac{1}{2}pa \cot(pa), \quad (41)$$

with the abbreviation

$$\Delta^{(N)}(p) = N + \tfrac{1}{2} + \sum_{n=N+1}^{\infty} (pa)^2[(pa)^2 - (n\pi)^2]^{-1}. \quad (42)$$

Using (41), we can rewrite (38) as

$$\tan \eta^{(N)}(p) = \frac{pa \cos(pa) - 2\,\Delta^{(N)}(p) \sin(pa)}{pa \sin(pa) + 2\,\Delta^{(N)}(p) \cos(pa)} = \tan(\chi^{(N)}(p) - pa), \quad (43)$$

where $\chi^{(N)}$ is defined by

$$\tan \chi^{(N)}(p) = pa/2\,\Delta^{(N)}(p). \quad (44)$$

We see from

$$(d/dp)\,\Delta^{(N)}(p) = -2pa^2 \sum_{n=N+1}^{\infty} (n\pi)^2[(pa)^2 - (n\pi)^2]^{-2} \leq 0 \quad (45)$$

that $\Delta^{(N)}(p)$ is a monotonically decreasing function having simple poles at

$$p = P_l = \pi(N + l)/a, \quad (l = 1, 2, \cdots),$$

such that

$$\lim_{\epsilon \to 0} \Delta^{(N)}(P_l \pm \epsilon) = \pm\infty, \quad \epsilon > 0. \quad (46)$$

Between two adjacent poles at P_l and P_{l+1} there is only one zero, say at $p = Q_l$ $(l = 1, 2, \cdots)$. The first zero Q_0 is located between 0 and $\pi(N + 1)/a$. At $p = Q_l$ the tangent of $\chi^{(N)}(p)$ becomes singular as

$$\lim_{\epsilon \to 0} \tan \chi^{(N)}(Q_l \mp \epsilon) = \pm\infty, \quad \epsilon > 0. \quad (47)$$

At $p = P_l$ the tangent of $\chi^{(N)}(p)$ has a zero; from (46) we obtain for small ϵ

$$\tan \chi^{(N)}(P_l - \epsilon) < 0 < \tan \chi^{(N)}(P_l + \epsilon). \quad (48)$$

Therefore, if we set

$$\chi^{(N)}(0) = 0, \quad (49)$$

which is in conformity with (44), it follows that

$$\chi^{(N)}(P_l) = l\pi, \quad (l = 1, 2, \cdots), \quad (50)$$

$$\chi^{(N)}(Q_l) = (l + \tfrac{1}{2})\pi, \quad (l = 0, 1, 2, \cdots). \quad (51)$$

In conformity with (43) and (49), we set

$$\eta^{(N)}(0) = 0. \quad (52)$$

Then it follows from (43) and (50), that

$$\eta^{(N)}(P_l) = -N\pi, \quad P_l = (N + l)\pi/a, \quad (l = 1, 2, \cdots). \quad (53)$$

The behavior of $\eta^{(N)}(p)$ for large p is such that it

[5] See, for example, W. Kohn, Phys. Rev. 84, 495 (1951); especially Sec. II.

fluctuates around $-N\pi$, but the deviation from $-N\pi$ becomes smaller as p becomes larger. Indeed, we obtain

$$\eta^{(N)}(0) - \eta^{(N)}(\infty) = N\pi, \tag{54}$$

a result analogous to Levinson's theorem. From (39) we conclude

$$|\Gamma^{(N)}(p)| < N(N\pi)^2[(pa)^2 - (N\pi)^2]^{-1} \quad \text{for} \quad pa > N\pi. \tag{55}$$

If this is substituted in (38), we see that

$$|\tan \eta^{(N)}(p)| < \frac{2\sin^2(pa)}{pa} \frac{N(N\pi)^2}{(pa)^2 - (N\pi)^2} \times \left[1 - \frac{N(N\pi)^2}{pa[(pa)^2 - (N\pi)^2]}\right]^{-1}. \tag{56}$$

Therefore,

$$\eta^{(N)}(p) = -N\pi + O[(N/pa)^2] \quad \text{for} \quad p \gg N\pi/a \tag{57}$$

or

$$\eta^{(N)}(\infty) = -N\pi. \tag{57'}$$

For small p ($p \ll N\pi/a$), $\eta^{(N)}(p)$ is well approximated by $-pa$. As can be shown easily from (42) and (44), $\chi^{(N)}(p)$ at $p = N^{1/2}\pi/a$ is of the order of $\pi/2N^{1/2}$ for large N. Therefore, for finite p, we have

$$\lim_{N\to\infty} \chi^{(N)}(p) = 0, \tag{58}$$

and

$$\lim_{N\to\infty} \eta^{(N)}(p) = -pa. \tag{59}$$

Thus the hard core phase shift is produced in the limit of infinitely many orthogonality constraints.

Next we turn to the behavior of the wave function. In momentum space it is given by (28) with $X_h(p)$ from (36). In configuration space the Fourier transformation yields

$$\psi^{(N)}(r;p) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty dk \sin krh^{(N)}(k;p)$$

$$= \left(\frac{2}{\pi}\right)^{1/2} \left\{\sin pr + \frac{\sin pa}{1 + \Gamma^{(N)}(p)\sin(2pa)/pa} \times \sum_{n=1}^{N} \frac{4k_n^2}{\pi a} \int_0^\infty \frac{\sin(ak)\sin(rk)}{(k-k_n)^2(k-p)^2}\, dk\right\}. \tag{60}$$

In the last equality, use has been made of (6), (6'), (36), and (39). The last integral is evaluated as

$$\frac{2}{\pi}\int_0^\infty dk \frac{\sin(ak)\sin(rk)}{(k^2-k_n^2)(k^2-p^2)}$$

$$= [p(p^2-k_n^2)]^{-1}\sin(ap)\cos(pr), \quad r \geq a, \tag{61}$$

and

$$= (p^2-k_n^2)^{-1}[p^{-1}\cos(ap)\sin(pr) - (-1)^n k_n^{-1}\sin(k_n r)], \quad r < a. \tag{62}$$

Therefore, outside the core $r \geq a$, the wave function follows by substitution of (61) into (60):

$$\psi^{(N)}(r;p) = (2/\pi)^{1/2}\left\{\sin(pr) - \frac{2\sin^2(pa)\Gamma^{(N)}(p)}{ap + \sin(2pa)\Gamma^{(N)}(p)}\cos(pr)\right\}$$

$$= (2/\pi)^{1/2}\{\sin(pr) + \tan\eta^{(n)}\cos(pr)\}$$

$$= (2/\pi)^{1/2}\sin(pr + \eta^{(N)})/\cos\eta^{(N)}, \quad r \geq a, \tag{63}$$

with the use of (38) and (39) in passing to the second equality. As expected the wave function is a free wave with the phase shift $\eta^{(N)}$. Since $\eta^{(N)}$ tends to the hard core phase shift as N tends to infinity, the eigenfunction is reasonable outside the core and approaches the right result as N tends to infinity. For $r \leq a$, substituting (62) into (60), we have

$$\psi^{(N)}(r;p) = (2/\pi)^{1/2}[1 + \sin(2pa)\Gamma^{(N)}(p)/pa]^{-1} \times \left\{\sin(pr) - \sum_{n=1}^{N} \frac{2(-1)^n k_n \sin(ap)\sin(k_n r)}{a(p^2-k_n^2)}\right\}. \tag{64}$$

We recall that $\sin(pr)$ can be expanded into a Fourier series for $r \leq a$

$$\sin(pr) = \sum_{n=1}^{\infty} \frac{2}{a}\sin(k_n r)\int_0^a \sin(pr')\sin(k_n r')\, dr'$$

$$= \sum_{n=1}^{\infty} \frac{2}{a}(-1)^n k_n \sin(ap)(p^2-k_n^2)^{-1}\sin(k_n r). \tag{65}$$

Hence,

$$\psi^{(N)}(r;p) = (2/\pi)^{1/2}[1 + \sin(2pa)\Gamma^{(N)}(p)/pa]^{-1} \times \sum_{n=N+1}^{\infty} \frac{2}{a}(-1)^n k_n \sin(ap)(p^2-k_n^2)^{-1}\sin(k_n r), \tag{66}$$

namely, the first N members of the complete set are missing in the expansion. The norm of $\psi^{(N)}(r;p)$ is small when p is small compared to $N\pi/a$. Actually

$$\int_0^a [\psi^{(N)}(r;p)]^2\, dr \propto \sum_{n=N+1}^{\infty} k_n^2(p^2-k_n^2)^{-2} \sim \sum_{n=N+1}^{\infty} k_n^{-2} \sim a^2/\pi^2 N, \quad p \ll N\pi/a, \quad N \gg 1. \tag{67}$$

Therefore, for a finite p, we establish

$$\lim_{N\to\infty} \psi^{(N)}(r;p) = 0, \quad r < a, \tag{68}$$

i.e., the actual wave function does not penetrate into the hard core domain $r \leq a$.

Before concluding this section, we would like to establish the completeness of the set of eigenfunctions (28). They are normalized when multiplied by $\cos \eta^{(N)}(p)$ [cf. (63)] and will be denoted by $\bar{h}^{(N)}(k;p)$

$$\bar{h}^{(N)}(k;p) = \cos \eta^{(N)}(p)\ \delta(p-k) + (k^2-p^2)^{-1} \times \sum_{n=1}^{N} f_n(k) X_n(p) \cos \eta^{(N)}(p)$$
$$= \cos \eta^{(N)}(p)\ \delta(p-k) + (k^2-p^2)^{-1} \times \sum_{n=1}^{N} \frac{4k_n^2 \sin(ak) \sin \chi^{(N)}(p)}{(k^2-k_n^2)a\pi} \qquad (69)$$

which follows from (6), (6'), (36), (39), (41), (43), and (44). In deriving the last statement of (69), we employed

$$\sin \chi^{(N)}(p) = \pm pa\{(pa)^2 + 4[\Delta^{(N)}(p)]^2\}^{-1/2}, \qquad (70)$$

where the upper sign holds for

$$pa < (N+1)\pi, \qquad (N+2m)\pi < pa < (N+2m+1)\pi, \qquad (m = 1, 2, \cdots),$$

while the lower sign holds for

$$(N+2m-1)\pi < pa < (N+2m)\pi, \qquad (m = 1, 2, \cdots)$$

as implied by (44) and the arguments which follow there. To show the completeness of the $\bar{h}^{(N)}$ we evaluate the integral

$$\int_0^\infty dp\ \bar{h}^{(N)}(k;p)\bar{h}^{(N)}(k';p) = \cos^2 \eta^{(N)}(k)\ \delta(k-k')$$
$$+ (k^2-k'^2)^{-1} \sum_{n=1}^{N} \frac{4k_n^2}{a\pi(k^2-k_n^2)} \sin(ak) \sin \chi^{(N)}(k') \cos \eta^{(N)}(k')$$
$$+ (k'^2-k^2)^{-1} \sum_{n=1}^{N} \frac{4k_n^2}{a\pi(k'^2-k_n^2)} \sin(ak') \sin \chi^{(N)}(k) \cos \eta^{(N)}(k)$$
$$+ (4/a\pi)^2 \sum_{n=1}^{N} k_n^2(k^2-k_n^2)^{-1} \sum_{m=1}^{N} k_m^2(k'^2-k_m^2)^{-1} \sin(ak) \sin(ak') \int_0^\infty \frac{\sin^2 \chi^{(N)}(p)}{(p^2-k^2)(p^2-k'^2)}\, dp. \qquad (71)$$

The integral appearing in the last term is evaluated as follows

$$\int_0^\infty \frac{\sin^2 \chi^{(N)}(p)\, dp}{(p^2-k^2)(p^2-k'^2)} = \int_0^\infty \pi^2\, \delta(p^2-k^2)\, \delta(p^2-k'^2) \sin^2 \chi^{(N)}(p)\, dp$$
$$+ (k^2-k'^2)^{-1} \int_0^\infty (p^2-k^2)^{-1} \sin^2 \chi^{(N)}(p)\, dp + (k'^2-k^2)^{-1} \int_0^\infty (p^2-k'^2)^{-1} \sin^2 \chi^{(N)}(p)\, dp$$
$$= (\pi/2)^2 k^{-2} \sin^2 \chi^{(N)}(k)\ \delta(k-k') + (k^2-k'^2)^{-1}\pi/4k \sin 2\chi^{(N)}(k) + (k'^2-k^2)^{-1}\pi/4k' \sin 2\chi^{(N)}(k'), \qquad (72)$$

where we employed a formula concerning products of principal values.[6] Now note that

$$2 \sum_{n=1}^{N} \frac{k_n^2}{k^2-k_n^2} \sin(ak) = -2\Gamma^{(N)}(k) \sin(ak) = ka \cos(ak) - 2\,\Delta^{(N)}(k) \sin(ak)$$
$$= \pm\{(pa)^2 + 4[\Delta^{(N)}(k)]^2\}^{1/2} \sin \eta^{(N)}(k), \qquad (73)$$

which follows from (39), (41), (43), (44), and (70). When (73) and (72) are substituted into (71), we find with the help of (70)

$$\int_0^\infty dp\ \bar{h}^{(N)}(k;p)\bar{h}^{(N)}(k';p) = [\cos^2 \eta^{(N)}(k) + \sin^2 \eta^{(N)}(k)]\ \delta(k-k')$$
$$+ (k^2-k'^2)^{-1} \sum_n \frac{4k_n^2}{a\pi(k^2-k_n^2)} \sin(ak)[\sin \chi^{(N)}(k') \cos \eta^{(N)}(k') - \cos \chi^{(N)}(k') \sin \eta^{(N)}(k')]$$
$$+ (k'^2-k^2)^{-1} \sum_n \frac{4k_n^2}{a\pi(k'^2-k_n^2)} \sin(ak')[\sin \chi^{(N)}(k) \cos \eta^{(N)}(k) - \cos \chi^{(N)}(k) \sin \eta^{(N)}(k)]. \qquad (74)$$

[6] Equation (3.17), Ref. I; S. Tani, J. Math. Phys. 2, 198 (1961).

Hence in view of (43),

$$\int_0^\infty dp\bar{h}^{(N)}(k;p)\bar{h}^{(N)}(k';p)$$
$$= \delta(k-k') - \sum_{n=1}^{N} 4k_n^2[a\pi(k^2-k_n^2)$$
$$\times (k'^2-k_n^2)]^{-1}\sin(ak)\sin(ak')$$
$$= \delta(k-k') - \sum_{n=1}^{N} f_n(k)f_n(k')$$
$$= (k\,|1-\Lambda^{(N)}|\,k'). \tag{75}$$

This shows that the set of eigenfunctions $\bar{h}^{(N)}(k;p)$ is complete in the Hilbert space restricted by the orthogonality to the N wave packets $f_n(k)$. Therefore the substitution, in which the auxiliary variables (14) are used,

$$a_k \to \int_0^\infty dp\bar{h}^{(N)}(k;p)a_p + \sum_{n=1}^{N} f_n(k)b_n \tag{76}$$

is a unitary transformation.

The unitary transformation (76) transforms the Hamiltonian $H^{(N)}$, Eq. (9), into

$$H^{(N)} \to \int_0^\infty dpp^2a_p^+a_p. \tag{77}$$

Thus, the Hamiltonian is diagonal after the transformation. In the new representation the operator a_p^+ creates a particle with the wave function $\bar{h}^{(N)}(k;p)$ in the free wave representation. Although its commutation relation

$$[a_p, a_{p'}^+] = \delta(p-p')$$

is the same as that for a free particle, the Hilbert space is restricted by the orthogonality constraints. We shall come back to the discussion of the commutation relation among field operators in Sec. 5.

4. CHOICE OF FICTITIOUS SET

In the preceding sections it has been shown that a system with a hard core can be produced by orthogonality constraints in the limits of infinitely many constraints. A particular choice of the fictitious set $g_n(r)$, Eq. (4), or $f_n(k)$, Eq. (6), has been made for the boundary condition (5). In case we use a different boundary condition

$$g_n'(a)/g_n(a) = \lambda, \tag{78}$$

with an arbitrary finite λ, the corresponding set of orthonormal functions will be

$$g_n(r) = \begin{cases} [2k_n^2[ak_n^2 - \lambda\sin^2(k_na)]^{-1}]^{1/2}\sin(k_nr), & 0 \le r \le a, \\ 0, & r > a. \end{cases}$$

Here k_n is determined as a positive root of the eigenvalue condition:

$$k_n\cos(k_na) = \lambda\sin(k_na). \tag{79}$$

The Fourier transform f_n of g_n is

$$f_n(k) = 2k_n[\pi(ak_n^2 - \lambda\sin^2 k_na)]^{-1/2}(k_n^2-k^2)^{-1}$$
$$\times \sin(k_na)[k\cos(ka) - \lambda\sin(ka)]. \tag{80}$$

We omit the detail but assert that all the arguments presented in the last section can be repeated with use of $f_n(k)$ given by (80) and lead to the same conclusion about the phase shift and the wave function in the limit $N \to \infty$. The results of the last sections are therefore independent of the choice of the boundary condition, (5) or (78), in the limit of infinitely many constraints. When the orthonormal set is chosen as (80), we find

$$G_{mn}(p) = \phi_m(p)\phi_n(p)\{(2p)^{-1}[\lambda\sin(pa) - p\cos(pa)]$$
$$\times [\lambda\cos(pa) + p\sin(pa)]$$
$$- [a(k_n^2+\lambda^2) - \lambda](p^2-k_n^2)\,\delta_{mn}/4k_n^2\}, \tag{81}$$

with

$$\phi_m(p) = 2k_m\sin(k_ma)(p^2-k_m^2)$$
$$\times [ak_m^2 - \lambda\sin^2(k_ma)]^{-1/2} \tag{82}$$

in place of (30). Instead of (36), we have

$$X_m(p) = -[\phi_m(p)]^{-1}\frac{4k_m^2(p^2-k_m^2)\pi^{-1/2}}{a(k_m^2+\lambda^2)-\lambda}$$
$$\times [p\cos(pa) - \lambda\sin(pa)]$$
$$\times \{1 + (2p)^{-1}[p\cos(pa) - \lambda\sin(pa)]$$
$$\times [p\sin(pa) + \lambda\cos(pa)]B^{(N)}(p)\}^{-1}, \tag{83}$$

abbreviating

$$B^{(N)}(p) = \sum_{n=1}^{N}(p^2-k_n^2)^{-1}4k_n^2[a(k_n^2+\lambda^2) - \lambda]^{-1}, \tag{84}$$

which plays a similar role to $\Gamma^{(N)}(p)$, (39). The phase shift is

$$\tan\eta^{(N)} = \pi(2p)^{-1}\sum X_n(p)f_n(p)$$
$$= \frac{[p\cos(pa) - \lambda\sin(pa)]^2B^{(N)}(p)}{2p + [p\cos(pa) - \lambda\sin(pa)][p\sin(pa) + \lambda\cos(pa)]B^{(N)}(p)} \tag{85}$$

instead of (38); in (85) we have used (80) and (83). In place of (40′), we get

$$\lim_{N\to\infty} B^{(N)}(p) = 2\sum_{n=1}^{\infty} 2k_n^2[a(k_n^2+\lambda^2)-\lambda]^{-1}$$
$$= \sin(pa)[\lambda\sin(pa) - p\cos(pa)]^{-1}. \quad (86)$$

Therefore again

$$\lim_{N\to\infty} \eta^{(N)} = -pa$$

according to those arguments in the last section.

It is to be noted that in the development of the arguments above we have needed only

(1) the orthonormal property of the f_n

$$\int dk\, f_n(k)f_m(k) = \delta_{mn},$$

and,

(2) the existence of $G_{mn}(p)$, (30) or (81), which follows from the square integrability of f_n,

when the number of constraints is finite.

As for the limit of infinitely many constraints, the existence of

$$\lim_{N\to\infty}\sum_{n=1}^{N}(p^2-k_n^2)^{-1},$$

which is guaranteed by (40′) or

$$\lim_{N\to\infty}\sum_{n=1}^{N} k_n^2[a(k_n^2+\lambda^2)-\lambda]^{-1}(p^2-k_n^2)^{-1},$$

which is guaranteed by (86), is essential. We actually never use the quantity

$$\int dk\, f_n(k)f_m(k)k^2,$$

although it appears in the Hamiltonian $H^{(n)}$, (9). It is noticeable that the projection operator $\Lambda^{(N)}$ is independent of the boundary condition of the fictitious set in the limit $N\to\infty$. It follows from (40′) that

$$\lim_{N\to\infty}(k\,|\Lambda^{(N)}|\,k') = \sum_{n=1}^{N} f_n(k)f_m(k')$$
$$= \pi^{-1}[(k-k')^{-1}\sin((k-k')a)$$
$$+ (k+k')^{-1}\sin((k+k')a)], \quad (87)$$

when (6) is used; when (80) is used, (87) follows from

$$\sum_{n=1}^{\infty} 2k_n^2\sin^2(k_na)(k^2-k_n^2)^{-1}[ak_n^2-\lambda\sin^2(k_na)]^{-1}$$
$$= \sin(ka)[\lambda\sin(ka) - k\cos(pa)]^{-1}. \quad (88)$$

Therefore, the first term of the "potential" $V^{(N)}$ in (11) becomes independent of the boundary condition in the limit $N\to\infty$. Note that (87) is the Fourier transform of the projection operator in the configuration space

$$(r\,|\Lambda|\,r') = \delta(r-r')\theta(a-r) \quad (89)$$

where $\theta(x)$ is the unit step function

$$\theta(x) = \begin{cases} 1, & x>0, \\ 0, & x<0. \end{cases}$$

It seems reasonable to conjecture that all statements up to this point will apply equally well for an arbitrary set of functions which form a complete set in the region $0\le r<a$.

We have to point out an unpleasant situation in explicitly writing down the Hamiltonian (13). The second term of the "potential" $V^{(N)}$, (11), is divergent in the limit $N\to\infty$, when the set defined by (6) is used; in case we use the set defined by (80), $\bar{\Lambda}^{(n)}$ diverges even for a finite N. If

$$\lim_{N\to\infty}\bar{\Lambda}^{(N)}$$

were a well-defined quantity, it would be given by

$$\lim_{N\to\infty}(k\,|\bar{\Lambda}^{(N)}|\,k') = \int_0^{\infty} dk''\,(k\,|\Lambda|\,k'')k''^2(k''\,|\Lambda|\,k')$$

with use of the well-defined limit of $\Lambda^{(N)}$, given by (87). For any two functions $F_1(k'')$ and $F_2(k'')$, which are Fourier transforms of functions with a discontinuity at $r=a$ in configuration space, the integral

$$\int dk''\,F_1(k'')k''^2F_2(k'')$$

diverges even when both F_1 and F_2 are square integrable. This renders the definition of

$$\lim_{N\to\infty}\bar{\Lambda}^{(N)}$$

impossible in our problem. Therefore, we conclude that it is impossible to regard the effect of a hard core as equivalent to a potential acting on a free wave. The divergent term corresponds to a surface potential of infinite height at $r=a$; the matrix element is correspondingly proportional to

$$\sin(ak)\sin(ak');$$

and the term which is proportional to

$$k\cos(ak)\sin(ak') + k'\sin(ak)\cos(ak')$$

is finite but not unique, depending on how $\bar{\Lambda}$ is

evaluated; [the result depends on whether (87), or either (6) or (80) is used if the limit $N \to \infty$ is taken at the end]. As has been pointed out in (77), the Hamiltonian is well defined when reference is made to the eigenfunction of $H^{(N)}$ instead of free waves. Therefore, if we first define the eigenfunction $h^{(N)}(k; p)$ by (27), disregarding the term $\bar{\Lambda}^{(N)}$ in the Hamiltonian, we can construct a system with a hard core in the limit of infinitely many constraints, and the result if independent of the particular choice of the fictitious set.

5. CONCLUDING REMARK

The advantage of the present approach lies in the fact that it is possible to generalize it to relativistic scattering and to multichannel scattering. Therefore it may be useful in a phenomenological approach to such a problem as the nucleon–nucleon scattering. The procedure to follow, in general, is exhibited in Sec. 2. The theoretical point that the introduction of hard core is not "adiabatic" is very clear in the present approach. The "soft" core for which the number of orthogonality constraints remains finite is different from a repulsive square well potential with a large but finite height. In the latter case one has

$$\bar{\eta}(0) - \bar{\eta}(\infty) = 0,$$

while in the former we found

$$\eta(0) - \bar{\eta}(\infty) = N\pi$$

indicating the nonadiabatic nature of the orthogonality constraints.

The remark concerning the "dimensionality of the Hilbert space" is relevant to elementary particle physics. It is reasonable to introduce a quantized field for each of the stable particles in order to construct an asymptotic state for a many-particle problem. But the Hilbert space thus constructed may turn out to be too wide; in that case the introduction of suitable constraints can be regarded as equivalent to a "core" interaction, soft or hard.

The conventional method of introducing a hard core is to introduce a repulsive square well potential and make it infinitely high in the limit. Apparently the Hamiltonian is not well defined in the limit. The situation has not been improved in our approach as far as this point is concerned. However, we have obtained a satisfactory result as for the completeness of the set of eigenfunctions. This has an important implication. Transparent discussions have been given by Siegert[7] of the commutation relations among field operators in the presence of a hard core. Here they are simplified corresponding to the fact that the hard core effect is dealt with as an external force, yet they are different from those in the absence of a hard core. The field operator in configuration space is defined by

$$\psi(r) = \lim_{N\to\infty} \int dk\, dp \left(\frac{2}{\pi}\right)^{1/2} \sin(kr)\bar{h}^{(N)}(k; p)a_p \qquad (90)$$

which is obtained by making a Fourier transform of the operator which appeared in (76). From (75), (89), and (90) follows the commutation relation

$$[\psi(r), \psi^+(r')] = \delta(r - r')[1 - \theta(a - r)]. \qquad (91)$$

Equation (91) is the right result representing the fact that no particle can be found inside the core; the commutation relation for a finite number of orthogonality constraints tends uniformly to the limit (91) as $N \to \infty$. If one uses the set of eigenfunctions in the field of a repulsive square well potential as the basis of second quantization, one will obtain

$$[\psi(r), \psi^+(r')] = \delta(r - r')$$

which does not tend to the limit of infinite height uniformly.

[7] A. J. F. Siegert, Phys. Rev. 116, 1057 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Studies in Perturbation Theory. V. Some Aspects on the Exact Self-Consistent Field Theory*

PER-OLOV LÖWDIN
Quantum Chemistry Group, University of Uppsala, Uppsala, Sweden and Quantum Theory Project, University of Florida, Gainseville, Florida
(Received April 9, 1962)

The independent-particle model in the theory of many-particle systems is studied by means of the self-consistent-field (SCF) idea. After a review of the characteristic features of the Hartree and Hartree-Fock schemes, the extension of the SCF method developed by Brueckner is further refined by introducing the exact reaction operator containing all correlation effects. This operator is here simply defined by means of the partitioning technique, and, if the SCF potentials are derived from this operator, one obtains a formalism which is completely analogous to the Hartree scheme but which still renders the exact energy and the exact wave function. An elementary derivation of the linked-cluster theorem is given, and finally the inclusion of various symmetry properties is discussed.

1. INTRODUCTION

TODAY, there are very good reasons for believing that modern quantum mechanics provides the essential tool for the theoretical understanding of the fundamental properties of the structure of matter: atoms, molecules, crystals, and atomic nuclei. The Schrödinger and Dirac equations represent in this connection the quintessence of more than a century of ingenious experimental work put in interplay with theoretical concepts and ideas. For simple systems containing one or two particles, the quantum theory has been quite successful in giving results in complete agreement with experience, whereas, for many-particle systems, the success has mainly been of a more qualitative nature.

For many-particle systems, there have actually been considerable mathematical difficulties in solving the wave equations, and there is a certain conflict between the mathematical complexity of the problem and the simplicity of the corresponding physical and chemical ideas. It is certainly true that the basic concepts in atomic, molecular, and solid-state theory as well as in nuclear theory are of a quantum-mechanical nature, but it is still far from possible to calculate all quantities of interest with a sufficient accuracy from the wave equations. So far, most treatments have usually been based on some simplifying "model" of the system under consideration, and the independent-particle model has played a dominant role in this connection. Many fundamental physical concepts are to a large extent also based on this scheme.

According to the independent-particle model, each particle in a many-particle system moves in the outer field acting on the system and the "average" field of all the other particles. This fruitful idea originates from Bohr,[1] and it has later been applied to the electronic structure of atoms, molecules, and crystals in form of the so-called self-consistent-field (SCF) schemes developed by Hartree[2] and others.[3] These schemes were successfully used for studying the electronic clouds of the atoms and their shell structure, for investigating the mobile π electrons of the conjugated organic compounds in organic chemistry, and for treating the band structure of crystals. In this connection, the Hartree and Hartree-Fock schemes were considered as the ultimate theoretical tool for understanding how the independent-particle model would work in a many-particle system. It seemed natural to assume that the qualitative and to a certain extent also quantitative success of the model would depend on the fact that the interactions between the electrons were comparatively weak, and that the correlation effects could be considered as a small perturbation.

This aspect on the independent-particle model was changed entirely with the discovery that it apparently worked extremely well also for the atomic nuclei in the so-called nuclear shell model. Here, the explanation could hardly be that the forces were weak, and it seemed necessary to find an extension of the independent-particle model which would work also for strong interactions between the particles. Such a generalization has actually been developed by Brueckner and his

* The research reported in this paper was sponsored in part by the King Gustaf VI Adolf's 70-Years Fund for Swedish Culture, Knut and Alice Wallenberg's Foundation, The Swedish Natural Science Research Council, and in part by the Chemistry Research Branch, ARL, AFRD, of the Air Research and Development Command, U. S. Air Force, through its European Office.

[1] N. Bohr, Proc. Phys. Soc. (London) **35**, 296 (1923).
[2] D. R. Hartree, Proc. Cambridge Phil. Soc. **24**, 89 (1928).
[3] V. Fock, Z. Physik **61**, 126 (1930); J. C. Slater, Phys. Rev. **35**, 210 (1930); P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930); **27**, 240 (1931).

collaborators[4] in a form which is closely analogous to the Hartree scheme. In the new scheme, however, the "average" potentials are derived from a scattering or reaction operator in such a way that the correlation between any two particles is exactly included, whereas the correlation between three and more particles is neglected. This so-called Brueckner approximation works exceedingly well in nuclear matter, where the forces are of short-range nature.

In a many-electron system, the Coulomb forces are more of a long-range type, and this makes it necessary to include also the correlation between three and more electrons. How far one should go is ultimately a question of order of magnitude and depends on the accuracy desired but, in principle, it would, of course, be of importance to have *all* correlation effects included. The purpose of this note is to show that one can extend the line of development which goes from Hartree to Brueckner still further, and that it is possible to relate the exact formal solution of the many-particle Schrödinger equation to the independent-particle model by means of a self-consistent-field scheme containing "average" potentials. These are derived from the exact reaction operator in the same way as the Hartree potentials were derived from the interaction part of the Hamiltonian. Since the exact SCF theory is completely analogous to the Hartree scheme, we will start with a brief review of the latter.

2. SELF-CONSISTENT-FIELD THEORIES

Let us consider the stationary states of a many-particle system which are described by the solutions to the Schrödinger equation $H\Psi = E\Psi$, where the Hamiltonian has the form

$$H_{op} = H_{(0)} + \sum_i H_i + \frac{1}{2!}\sum_{ij}{}' H_{ij} + \frac{1}{3!}\sum_{ijk}{}' H_{ijk} + \cdots . \qquad (1)$$

For completeness, we have here included also many-particle interactions,[5] the treatment of which will be very instructive. The prime on a summation sign indicates that two indices must never be equal, and we will sometimes also use the alternative forms

$$\frac{1}{2!}\sum_{ij}{}' = \sum_{i<j}, \qquad \frac{1}{3!}\sum_{ijk}{}' = \sum_{i<j<k}. \qquad (2)$$

The term $H_{(0)}$ is a constant which may be of importance from the point of view of convergence[6] but, since it has no influence on the interactions between the particles, it will here be temporarily omitted.

We will start by reviewing some of the features which are common for all self-consistent-field (SCF) schemes. Let us divide the Hamiltonian (1) into two parts $H = H_0 + V$, where

$$H_0 = \sum_i (H_i + u_i), \qquad (3)$$

$$V = -\sum_i u_i + \sum_{i<j} H_{ij} + \sum_{i<j<k} H_{ijk} + \cdots ,$$

and u_i are one-particle potentials at our disposal. The eigenvalue problem connected with H_0 is separable and, for the unperturbed eigenfunction, one obtains

$$\varphi_0 = \psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)\cdots\psi_N(\mathbf{x}_N), \qquad (4)$$

where

$$(H_i + u_i)\psi_i = \epsilon_i\psi_i, \qquad E_0 = \sum_i \epsilon_i. \qquad (5)$$

Here $H_{eff}(i) = H_i + u_i$ is characterized as the *effective Hamiltonian* for the particle i, and $\mathbf{x}_i = (\mathbf{r}_i, \zeta_i, \tau_i)$ is the complete coordinate for this particle describing its situation in ordinary space and in the spaces associated with spin, isotopic spin, etc. The one-particle functions $\psi_i(\mathbf{x}_i)$ are called spin orbitals, and we will assume them to be normalized so that $(\psi_i \mid \psi_i) = 1$, whereas the eigenvalues ϵ_i are denoted as one-particle or spin-orbital energies.

Hartree scheme. At first, we will leave the antisymmetry requirement aside. In the so-called Hartree scheme,[2,3] the total wave function Ψ is actually approximated by the simple product (4). For the corresponding total energy, one obtains

$$\langle\varphi_0 \mid H_{op} \mid \varphi_0\rangle = \sum_i (\psi_i \mid H_i \mid \psi_i) + \frac{1}{2!}\sum_{ij}{}' (\psi_i\psi_j \mid H_{ij} \mid \psi_i\psi_j) + \frac{1}{3!}\sum_{ijk}{}' (\psi_i\psi_j\psi_k \mid H_{ijk} \mid \psi_i\psi_j\psi_k) + \cdots , \qquad (6)$$

where we have used brackets of the type $\langle\ \rangle$ for integrals over all coordinates and brackets of the type $(\ )$ for integrals over one, two, three, or more coordinates. The best one-particle functions ψ_i are now determined by the variation principle $\delta\langle H\rangle = 0$. According to (6), we get directly

[4] K. A. Brueckner, C. A. Levinson, and H. M. Mahmoud, Phys. Rev. 95, 217 (1954); K. A. Brueckner, *ibid.* 96, 508 (1954); 97, 1353 (1955); 100, 36 (1955); K. A. Brueckner and C. A. Levinson, *ibid.* 97, 1344 (1955); H. A. Bethe, *ibid.* 103, 1353 (1956); J. Goldstone, Proc. Roy. Soc. (London) A239, 267 (1957); H. A. Bethe and J. Goldstone, *ibid.* A238, 551 (1957); L. S. Rodberg, Ann. Phys. (New York) 2, 199 (1957); and several other papers.

[5] Compare also P. O. Löwdin, Phys. Rev. 97, 1490 (1955).

[6] For instance, the Coulomb repulsion between the nuclei in a solid; see P. O. Löwdin, Advances in Phys. 5, 1 (1956), p. 11 f.

$$\delta\langle H\rangle = \sum_i (\delta\psi_i\,|H_i + \sum_{j\neq i}(\psi_i\,|H_{ij}|\,\psi_j) + \frac{1}{2!}\sum_{j,k\neq i}{}'(\psi_i\psi_k\,|H_{ijk}|\,\psi_i\psi_k) + \cdots|\,\psi_i) + \text{complex conjugate term} = 0, \quad (7)$$

where the operator in the matrix elements defines the effective Hamiltonian $(H_i + u_i)$. Together with the variation of the normalization condition $(\psi_i\,|\,\psi_i) = 1$, this leads to the Hartree equations $(H_i+u_i)\psi_i = \epsilon_i\psi_i$ of type (5), where the potentials are given by the relations:

$$\begin{aligned} u_i &= u_i^{(2)} + u_i^{(3)} + \cdots; \\ u_i^{(2)} &= \sum_{j\neq i}(\psi_j\,|H_{ij}|\,\psi_j), \\ u_i^{(3)} &= \frac{1}{2!}\sum_{j,k\neq i}{}'(\psi_j\psi_k\,|H_{ijk}|\,\psi_j\psi_k), \cdots. \end{aligned} \quad (8)$$

Here each term has an upper index which indicates the order of the interaction term in the Hamiltonian, from which it has been derived. By using the Hartree potentials (8), the total energy (6) can now be written in the form

$$\begin{aligned} &\langle\varphi_0\,|H_{op}|\,\varphi_0\rangle \\ &= \sum_i(\psi_i\,|H_i + \tfrac{1}{2}u_i^{(2)} + \tfrac{1}{3}u_i^{(3)} + \cdots|\,\psi_i) \\ &= \left\langle\varphi_0\left|\sum_i\left(H_i + \tfrac{1}{2}u_i^{(2)} + \tfrac{1}{3}u_i^{(3)} + \cdots + \frac{1}{N}u_i^{(N)}\right)\right|\varphi_0\right\rangle. \end{aligned} \quad (9)$$

We note that $\langle H\rangle$ is not identical to E_0: The factor $(1/p)$ connected with the potential $u_i^{(p)}$ prevents, actually, the p-body interaction from being counted p times, as it would be in the sum $E_0 = \sum_i \epsilon_i$. Of course, it had been much simpler, particularly from semiempirical points of view, if all these coefficients would have been equal to 1, but this occurs only if the total Hamiltonian (1) does not contain any interactions between the particles.

In addition to the basic wave function φ_0, we will consider the "singly excited" functions $\varphi_{s.e.}$, which are obtained from φ_0 by replacing one, and only one, of the functions ψ_k by another $\bar{\psi}_k$ which is orthogonal to the former, so that $(\bar{\psi}_k\,|\,\psi_k) = 0$. By using (3) and (8), one obtains directly

$$\langle\varphi_{s.e.}\,|V|\,\varphi_0\rangle = -(\bar{\psi}_k\,|u_k^{(2)} + u_k^{(3)} + \cdots|\,\psi_k) + \sum_{i\neq k}(\psi_i\bar{\psi}_k\,|H_{ik}|\,\psi_i\psi_k) + \frac{1}{2!}\sum_{i,j\neq k}{}'(\psi_i\psi_j\bar{\psi}_k\,|H_{ijk}|\,\psi_i\psi_j\psi_k) + \cdots = 0, \quad (10)$$

which is a form of Brillouin's theorem[7] saying that all matrix elements of the perturbation V between the basic function φ_0 and all singly excited functions will vanish identically. We note that this theorem is a consequence of the definition of the Hartree potentials, and vice versa. Since $V = H_{op} - H_0$ and $H_0\varphi_0 = E_0\varphi_0$, one gets further

$$\langle\varphi_{s.e.}\,|H_{op}|\,\varphi_0\rangle = 0. \quad (11)$$

We note that this relation does not prevent the singly excited functions to appear in addition to φ_0 in the configurational expansion of the total wave function, since they may come in through couplings with functions which are at least doubly excited with respect to φ_0.

Hartree-Fock scheme. Let us now consider a system of fermions obeying the Pauli exclusion principle and characterized by wave functions fulfilling the antisymmetry requirement. In the so-called Hartree-Fock scheme,[3] the total wave function Ψ is approximated by the "antisymmetric component" of the simple product (4), i.e., by the Slater determinant

$$D_0 = (N!)^{-1/2}\det\{\psi_1, \psi_2, \cdots, \psi_N\}. \quad (12)$$

Since such a determinant is invariant under linear transformations of the set $\psi_1, \psi_2, \cdots, \psi_N$, the basic set may be assumed to be orthonormalized so that $(\psi_i\,|\,\psi_j) = \delta_{ij}$. For the total energy $\langle D_0\,|H|\,D_0\rangle$, one obtains the same expression as in (6) but with $H_i, H_{ij}, H_{ijk}, \cdots$ replaced by the operators

$$\begin{aligned} \bar{H}_i &= H_i, \\ \bar{H}_{ij} &= H_{ij}(1 - P_{ij}), \\ \bar{H}_{ijk} &= H_{ijk}\sum_p(-1)^p P_{ijk}, \cdots, \end{aligned} \quad (13)$$

where $P_{ij}, P_{ijk}, \cdots$ are permutations working on the coordinates $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \cdots$. The Hartree-Fock potentials are then given by (8) with the same replacements, and formulas (9), (10), and (11) are valid with similar modifications, i.e., Brillouin's theorem takes now the form

$$\langle D_{s.e.}\,|H_{op}|\,D_0\rangle = 0. \quad (14)$$

The Hartree-Fock scheme is, of course, complicated by the occurrence of the permutations in (13) and the corresponding exchange potentials in $u_i^{(p)}$. There is at least one simplification in comparison to the Hartree scheme, however, since one may now omit the restrictions $j \neq i$, $j \neq k \neq i, \cdots$ in the definitions (8) depending on the fact that the terms for $j = i$, $j, k = i, \cdots$ etc. calculated for the operators (13) give vanishing contributions. One obtains

[7] L. Brillouin, Actualités sci. et ind. No. 71 (1933); No. 159 (1934); C. Møller and M. S. Plesset, Phys. Rev. 46, 618 (1934).

$$H_{eff}(i) = H_i + \sum_j (\psi_j \,|H_{ij}(1 - P_{ij})|\, \psi_j)$$

$$+ \frac{1}{2!} \sum_{jk} (\psi_j\psi_k \,|H_{ijk} \sum (-1)^p P_{ijk}|\, \psi_j\psi_k) + \cdots , \tag{15}$$

i.e., the effective Hamiltonians are one and the same for all the spin-orbitals ψ_i. Introducing the Fock-Dirac density matrix

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k=1}^{N} \psi_k(\mathbf{x}_1)\psi_k^*(\mathbf{x}_2), \tag{16}$$

which is the fundamental invariant[6] of the scheme, one gets further

$$H_{eff}(1) = H_1 + \int dx_2 H_{12}(1 - P_{12})\rho(\mathbf{x}_2, \mathbf{x}_2')_{x_2'=x_2}$$

$$+ \frac{1}{2!} \int dx_2\, dx_3 H_{123}$$

$$\times (1 - P_{12} - P_{13} - P_{23} + P_{123} + P_{132})$$

$$\times \rho(\mathbf{x}_2, \mathbf{x}_2')\rho(\mathbf{x}_3, \mathbf{x}_3')_{x_i'=x_i} + \cdots , \tag{17}$$

where the permutations are not supposed to work on the primed coordinates $\mathbf{x}_2'$, $\mathbf{x}_3'$, $\cdots$ which are put equal to $\mathbf{x}_2$, $\mathbf{x}_3$, $\cdots$ before the integrations are carried out. The permutations give rise to exchange potentials of a nonlocal nature, so the quantities u_i are no longer simple functions of $\mathbf{x}_i$ but have operator character.

Self-consistent-field procedure. Both the Hartree and the Hartree-Fock equations represent systems of nonlinear integro-differential equations connected with eigenvalue problems which are solved by the "self-consistent-field" procedure.[2,3] One starts out from estimates of the spin-orbitals $\psi_1, \psi_2, \cdots, \psi_N$, evaluates the potentials u_i according to (8), solves the eigenvalue problems $(H_i + u_i)\psi_i = \epsilon_i\psi_i$, and obtains new functions $\psi_1, \psi_2, \cdots \psi_N$, which provides the starting functions for a new cycle:

$$\{\psi_i\} \rightarrow u_i \rightarrow H_{eff} \rightarrow \{\psi_i\} \tag{18}$$

The procedure is continued until it becomes "self-consistent," i.e., no further changes occur in the significant figures when the cycle is repeated. The iteration process may in certain cases also be divergent, but the procedure may anyway be used to derive the solution.[8] Instead of starting from the set $\{\psi_i\}$, one can start from the potentials u_i representing the "average" fields involved, which has given the method its name.

The eigenvalue problem (5) has in the atomic case[9] been solved by numerical integration, and this approach has also been applied to crystals in the cellular method[10] and in the augmented plane wave method.[11] Ritz's expansion method[12] was first applied to molecules,[13] but this technique has later proven to be very useful also in connections with atoms and crystals.

3. EXACT SELF-CONSISTENT-FIELD THEORY

Reaction operator. The extension of the Hartree scheme developed by Brueckner and his collaborators is based on the use of the scattering or reaction operator.[4] Deviating somewhat from the conventional approach, we will here derive this concept by means of the so-called partitioning technique, which was originally a tool for the numerical solution of secular equations of high orders.[14] Since this method has been described in greater detail in a preceding paper,[15] we will here limit ourselves to a few remarks.

Let us start by giving an alternative derivation of the fundamental formulas in the projection operator formalism which corresponds to ∞-order perturbation theory. Let O be a projection operator which selects a certain subspace (a) of order g, so that $O^2 = O$, $O^\dagger = O$, Tr $(O) = g$. Let further $P = 1 - O$ be the projection operator which selects its "orthogonal complement" (b), and which satisfies the relations $P^2 = P$, $P^\dagger = P$, $OP = PO = 0$.

Starting by considering a nondegenerate level E, we will choose $g = 1$. Let Φ be an arbitrary trial function or "model" function with a nonvanishing projection $O\Phi = \varphi$, which we will normalize so that $\langle\varphi \mid \varphi\rangle = 1$, i.e., $\langle\Phi\,|O|\,\Phi\rangle = 1$. For the eigenfunction Ψ satisfying the eigenvalue relation $(H - E)\Psi = 0$, one has now the identity:

[8] See, for instance, P. O. Löwdin, Technical Note No. 11, Uppsala Quantum Chemistry Group, 1958 (unpublished), particularly the Appendix.

[9] For a survey of the atomic SCF calculations, see D. R. Hartree, Repts. Progr. in Phys. **11**, 113 (1948); *Calculation of Atomic Structures* (John Wiley & Sons, Inc., New York, 1957); R. S. Knox, *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic Press Inc., New York, 1957), Vol. 4, p. 413; P. O. Löwdin, *Proceedings of the Robert A. Welch Foundation Conference on Chemical Research, 1958*, edited by W. O. Milligan (Robert A. Welch Foundation, Houston, Texas, 1959), p. 5.

[10] E. Wigner and F. Seitz, Phys. Rev. **43**, 804 (1933); **46**, 509 (1934).

[11] J. C. Slater, Phys. Rev. **51**, 846 (1937); **92**, 603 (1953).

[12] W. Ritz, J. reine angew. Math. **135**, 1 (1909).

[13] C. A. Coulson, Proc. Cambridge Phil. Soc. **34**, 204 (1938); C. C. J. Roothaan, Revs. Modern Phys. **23**, 69 (1951).

[14] P. O. Löwdin, Technical Note No. 11, Uppsala Quantum Chemistry Group (1958) (unpublished); Advances in Chem. Phys. **2**, 207 (1959).

[15] P. O. Löwdin, J. Math. Phys. **3**, 969 (1962).

$$\Psi = (O + P)\Psi = \varphi + PK^{-1}K\Psi$$
$$= \varphi + PK^{-1}[K + P(H - E)(O + P)]\Psi$$
$$= \varphi + PK^{-1}PH\varphi$$
$$+ PK^{-1}[K - P(E - H)P]\Psi. \quad (19)$$

Here K is an arbitrary nonsingular operator which will be chosen in a convenient way. One cannot simply put $K = P(E - H)P$ to get rid of the last term, since this operator is actually singular and has no inverse, but one can easily circumvent this difficulty. Introducing the definitions

$$K = \alpha \cdot O + P(E - H)P, \qquad T = PK^{-1}P, \quad (20)$$

where α is an arbitrary number, we note that K is nonsingular for $\alpha \neq 0$ and that one has the relation

$$[\alpha \cdot O + P(E - H)P]K^{-1} = 1. \quad (21)$$

Multiplying this equation to the left and right by O and P, we obtain for $\alpha \neq 0$ the following four relations:

$$OK^{-1}O = \alpha^{-1}O, \qquad OK^{-1}P = 0,$$
$$PK^{-1}O = 0, \qquad P(E - H)PK^{-1}P = P \ . \quad (22)$$

Since $\partial T/\partial\alpha = -PK^{-1}OK^{-1}P \equiv 0$, the operator T is independent of the value of α. Since the quantity $OK^{-1}O$ does not occur at all in the eigenvalue theory, it is tempting to put $\alpha = 0$, but this leads actually to an improper notation, which is nevertheless sometimes used in the theory. The fundamental operator T is characterized by the relations

$$OT = TO = 0, \qquad P(E - H)T = P, \quad (23)$$

and, in the following, we will use the symbolic notation

$$T = P/(E - H), \quad (24)$$

but remember that it corresponds to the full definition (20) with $\alpha \neq 0$. Substituting our choice (20) of K into (19) and using (22), we obtain finally

$$\Psi = \varphi + TH\varphi = (O + THO)\Phi, \quad (25)$$

showing that the eigenfunction Ψ may be derived from any trial function Φ having $O\Phi \neq 0$ by means of the operator

$$\Omega = O + THO. \quad (26)$$

One has $(H - E)\Psi = (H - E)\Omega\Phi = 0$ for all trial functions Φ, which gives $(H - E)\Omega \equiv 0$. The operator Ω is actually characterized by the three relations

$$H\Omega = E\Omega, \qquad \Omega^2 = \Omega, \qquad \mathrm{Tr}\,(\Omega) = 1, \quad (27)$$

and it may hence be characterized as an idempotent eigenoperator to H associated with the eigenvalue E. It is not a normal projection operator, and its importance comes actually from its connection with ∞-order perturbation theory.[15] Using the relation $O(H - E)\Omega = O(H + HTH - E)O = 0$, one gets for the energy the condition

$$OEO = O(H + HTH)O, \quad (28)$$

and, after multiplication to the left by Φ^* and to the right by Φ and integration, this leads to the energy formula

$$E = \langle\varphi\,|H + H[P/(E - H)]H|\,\varphi\rangle. \quad (29)$$

The eigenfunction (25) is characterized by the normalization condition $\langle\varphi \mid \Psi\rangle = 1$ and, for the actual normalization integral, one obtains

$$\langle\Psi \mid \Psi\rangle = \langle\varphi\,|1 + HT^2H|\,\varphi\rangle. \quad (30)$$

Let us now consider the special case when $H = H_0 + V$, where V is a weak or strong perturbation. In this connection, it is convenient to assume that O is the eigenoperator to the unperturbed Hamiltonian H_0 associated with the level E_0 under consideration, so that $H_0O = OH_0 = E_0O$. This means that O will project out the unperturbed eigenfunction φ_0. In this approach, one needs only a *single* eigenfunction to H_0 and not the complete set associated with the entire spectrum, and the "orthogonal complement" to φ_0 is here simply described by orthogonalizing any complete set towards φ_0. Putting $H = H_0 + V$ into formulas (26) and (28), and using the relation $OP = PO = 0$, we obtain

$$\Omega = (1 + TV)O, \quad (31)$$

$$OEO = O(E_0 + V + VTV)O. \quad (32)$$

Of particular interest is here the operator

$$t = V + VTV, \quad (33)$$

which is called the *reaction operator* associated with the perturbation V, the unperturbed Hamiltonian H_0, and the specific state under consideration. From (32) follows the relation

$$E = E_0 + \langle\varphi_0\,|t|\,\varphi_0\rangle, \quad (34)$$

showing that the expectation value of the reaction operator t with respect to the unperturbed state gives the true energy shift. Substitution of (34) into (33) gives finally

$$t = V + V\frac{P}{(E_0 - H_0) - (V - \langle t\rangle_0)}V, \quad (35)$$

which is the fundamental formula for the reaction operator here. For some aspects concerning the evaluation of this operator, particularly the connection with the Schrödinger perturbation theory, we will refer to a preceding paper.[15] Here we will

study t from somewhat different points of view in connection with the generalization of the SCF idea.

The operator $W = 1 + TV$ has been called the "wave operator," and we note the three relations

$$\Psi = W\varphi_0, \qquad t = VW, \qquad t\varphi_0 = V\Psi, \tag{36}$$

which have sometimes been used to define the reaction operator. It is worthwhile to observe that, in order to derive the eigenfunction Ψ and its energy E, one needs only the "components" PtO and OtO, respectively, whereas the components OtP and PtP do not occur at all in the theory. This fact will later be used to simplify the treatment of the reaction operator.

So far, we have only considered a nondegenerate level E, but the detailed discussion in a preceding paper[15] shows that exactly the same approach can be used also for a degenerate level E of order g. In this case, it is convenient to choose the projection operator O to be of the same order, and one obtains then the same fundamental formulas as before with a reinterpretation of the operator $P = 1 - O$. Hence, it is not necessary to treat the degenerate case separately here.

Exact SCF theory. Utilizing the reaction operator, we will now study the perturbation problem connected with the independent-particle model. In this case, H_0 and V are defined by (2) and (3), respectively, and the unperturbed wave function φ_0 is represented by the Hartree product (4). Introducing the notation

$$H_{\text{int}} = \sum_{i<j} H_{ij} + \sum_{i<j<k} H_{ijk} + \cdots, \tag{37}$$

for the interaction part of the Hamiltonian and the abbreviation $u = \sum_i u_i$, we have

$$H_0 = \sum_i (H_i + u_i), \qquad V = -u + H_{\text{int}}. \tag{38}$$

Substituting the expression for V into (35), one finds that it is convenient to introduce another operator τ by the relation

$$t = -u + \tau. \tag{39}$$

The idea is to separate out the one-particle term in t and to let τ represent the part of the reaction operator which involves two, three, and more particles, Let us write τ in the form

$$\tau = \tau^{(1)} + \tau^{(2)} + \tau^{(3)} + \cdots + \tau^{(N)} = \sum_i \tau_i + \frac{1}{2!}\sum_{ij}{}' \tau_{ij} + \frac{1}{3!}\sum_{ijk}{}' \tau_{ijk} + \cdots, \tag{40}$$

where $\tau^{(p)}$ represents the part of the reaction operator involving p particles. For the sake of completeness, we have here included also a one-particle term $\tau^{(1)}$, which we later will show gives a vanishing contribution. According to (34) and (39), the total energy, E, can now be written

$$E = \langle \varphi_0 \,|H_0 + t|\, \varphi_0\rangle = \langle \varphi_0 \,|\sum_i H_i + \tau|\, \varphi_0\rangle. \tag{41}$$

This expression is, in principle, exact and hence it cannot be improved by variation. However, in order to establish a connection with the Hartree scheme, we will now remove the coupling between φ_0 and τ and consider τ momentarily as a fixed operator. The right-hand member of (41) is then no longer invariant under variations of φ_0, and the best function φ_0 is again determined by the condition that this quantity should be stationary, which leads to equations of the type

$$(H_i + u_i)\psi_i = \epsilon_i \psi_i \tag{42}$$

with potentials u_i defined by the relations

$$\begin{aligned} u_i &= u_i^{(1)} + u_i^{(2)} + u_i^{(3)} + \cdots + u_i^{(N)}, \\ u_i^{(1)} &= \tau_i, \\ u_i^{(2)} &= \sum_{j\neq i} (\psi_j \,|\tau_{ij}|\, \psi_j), \\ u_i^{(3)} &= \frac{1}{2!}\sum_{j,k\neq i}{}' (\psi_j\psi_k \,|\tau_{ijk}|\, \psi_j\psi_k), \cdots, \end{aligned} \tag{43}$$

i.e., exactly the same relations as (8) but with the interaction terms in the Hamiltonian replaced by the operator τ. This gives finally the energy formula

$$E = \left\langle \varphi_0 \left| \sum_i \left(H_i + u_i^{(1)} + \tfrac{1}{2}u_i^{(2)} + \tfrac{1}{3}u_i^{(3)} + \cdots \frac{1}{N} u_i^{(N)} \right) \right| \varphi_0 \right\rangle, \tag{44}$$

which is completely analogous to relation (9) in the Hartree scheme; we note that the contribution from $u_i^{(1)}$ is actually vanishing, as will be shown below.

The SCF potentials u_i are here considerably more complicated than in the Hartree scheme, but the energy (44) is also the true energy including all correlation effects. For the iterative SCF procedure, one has the following "flow diagram":

$$\begin{array}{ccccccc} & & V & \rightarrow & t & \rightarrow & \tau \\ & \nearrow & & & & & \searrow \\ u_i & \rightarrow & (H_i + u_i) & \rightarrow & \{\psi_i\} & \rightarrow & u_i \\ \uparrow & & & & & & \downarrow \end{array} \tag{45}$$

Each cycle is here much more complicated than the corresponding Hartree cycle (18), since it involves the calculation of the reaction operator t. This step corresponds actually to an exact solution of the many-particle Schrödinger equation, which it ought to be sufficient to carry out only once. There exists hence probably a shortcut in the procedure, perhaps

by means of the first-order density matrix, and research on this point is in progress. Below we will show that the formal expression for the reaction operator may also be essentially simplified by using the conditions (43).

In this treatment, we have assumed that the operator τ could be written in the form (40) with the terms arranged after the number of particles involved. Since the operator τ defined by (35) and (39) has a rather complicated character, it is evident that the transformation to the form (40) may be rather cumbersome and requires a careful mathematical analysis. In this connection, it is worthwhile to observe that the formulation of the SCF-scheme actually does not depend on the explicit knowledge of the expansion (40). This comes from the fact that, even if a p-body reaction usually cannot be reduced to lower orders, it can never lead to wrong results if it happens to be considered as being of higher order $(p+1)$, $(p+2)$, $\cdots N$. One would hence obtain a correct result, even if all terms in τ would be considered to be of the highest order possible, i.e., $p = N$, and the treatment would then not involve the expansion (40). Applying (43) to the case of $p = N$, we obtain

$$u_i' = (\psi_1\psi_2 \cdots \psi_{i-1}\psi_{i+1} \cdots \psi_N \times |\tau|\ \psi_1\psi_2 \cdots \psi_{i-1}\psi_{i+1} \cdots \psi_N), \qquad (46)$$

and, except for additive constants, these potentials are identical with those previously defined by (43). For the total energy, one has

$$E = \left\langle \varphi_0 \left| \sum_i \left(H_i + \frac{1}{N} u_i' \right) \right| \varphi_0 \right\rangle, \qquad (47)$$

which expression goes over into (44) if τ is expanded according to (40).

Brillouin-Brueckner theorem. The SCF potentials u_i have here been derived under the somewhat artificial assumption that the operator τ was momentarily kept fixed when the variation principle was applied to the expression (41). We will now give a reinterpretation of the conditions (43) which has a more natural form and which has deep-going consequences for the theory. Following the development in the Hartree scheme, we will in addition to the unperturbed φ_0 consider also all functions $\varphi_{s.e.}$ which are *singly excited* with respect to φ_0. Using (39) and (43), we obtain

$$\begin{aligned}\langle \varphi_{s.e.} \mid t \mid \varphi_0 \rangle &= \langle \psi_1 \cdots \bar{\psi}_k \cdots \psi_N \\ &\quad \times \Big| - \sum_i u_i + \tau \Big| \psi_1 \cdots \psi_k \cdots \psi_N \rangle \\ &= -(\bar{\psi}_k \,|u_k|\, \psi_k) + (\bar{\psi}_k \,|\tau_k|\, \psi_k) \\ &\quad + \sum_{i \neq k} (\psi_i \bar{\psi}_k \,|\tau_{ik}|\, \psi_i \psi_k) \\ &\quad + \frac{1}{2!} \sum_{i,j \neq k}{}' (\psi_i \psi_j \bar{\psi}_k \,|\tau_{ijk}|\, \psi_i \psi_j \psi_k) + \cdots \\ &= (\bar{\psi}_k \mid - u_k + u_k^{(1)} + u_k^{(2)} + u_k^{(3)} + \cdots \\ &\quad + u_k^{(N)} \mid \psi_k) = 0, \end{aligned} \qquad (48)$$

in complete analogy with (10). This gives the relation

$$\boxed{\langle \varphi_{s.e.} \mid t \mid \varphi_0 \rangle = 0,} \qquad (49)$$

which we will call the "Brillouin-Brueckner theorem" in the exact SCF theory. We note that (49) may be used to define the SCF potentials u_i, and later we will show that this approach leads to a simplified formulation of the theory.

In order to study the consequences of (49) in greater detail, it is now convenient to partition the total Hilbert space for the N-particle functions into three parts: one connected with φ_0 and described by the projection operator O, one connected with all *singly excited* functions φ_1 and characterized by the projection operator P_1, and the remaining part containing all functions which are *at least doubly excited* with respect to φ_0 and characterized by the projection operator Q.

One may describe such a partitioning by means of the eigenfunctions to H_0, but this diagonal representation is often complicated by the fact that part of the eigenfunctions belong to the continuous part of the eigenvalue spectrum. It is then often simpler to proceed in the following way: let us start from the one-particle spaces associated with the N electronic coordinates, and let us, for each coordinate $\mathbf{x}_k$, to the function $\psi_k(=\psi_{k1})$ actually occurring in φ_0 add a suitable discrete "orthogonal complement" $\psi_{k2}, \psi_{k3}, \psi_{k4}, \cdots$ which is sufficient to make the set complete, for instance, by orthonormalizing a given complete set with respect to the function ψ_k. A p-fold excited function is now defined by means of the Hartree product formed by substituting p functions ψ_i in φ_0 by p functions taken from their respective orthogonal complements $\psi_{i2}, \psi_{i3}, \psi_{i4}, \cdots$, and all these functions describe together a subspace which will be characterized by the projection operator P_p. One has the resolution $1 = O + P$, $P = P_1 + P_2 + P_3 + \cdots + P_N$, and further

$$P = P_1 + Q, \qquad Q = P_2 + P_3 + \cdots + P_N. \qquad (50)$$

We note that the operators O, P_1, and Q are essentially defined by φ_0 but are independent of any

specific choice of the "orthogonal complement" $\psi_{N+1}, \psi_{N+2}, \cdots$.

One can now write the Brillouin-Brueckner theorem (49) in the operator form $P_1tO = 0$, indicating that the matrix elements of the left-hand member with respect to any trial functions will vanish identically. Since $P = P_1 + Q$, this leads to the relation

$$PtO = QtO, \tag{51}$$

which will now be used to simplify the reaction operator. Since $u = \sum_i u_i$ is a one-particle operator, one has further

$$QuO = OuQ = 0, \tag{52}$$

depending on the fact that, in forming the corresponding matrix elements, each integral will contain at least one orthogonality integral.

For the further analysis, it is convenient to introduce temporarily the operator

$$T_0 = P[\alpha \cdot O + P(E - H_0)P]^{-1}P, \tag{53}$$

which actually belongs to the Brillouin-type perturbation theory.[15] The operator T_0 is completely analogous to T defined by (20) but has H_0 instead of H. Since T_0 is entirely independent of the parameter $\alpha \neq 0$, we will use the symbolic notation:

$$T_0 = P/(E - H_0). \tag{54}$$

From the fact that O is an eigenoperator to H_0 follows $P_1(E - H_0)O = 0$, $Q(E - H_0)O = 0$, and, from the eigenvalue properties of H_0, one can further conclude that

$$P_1(E - H_0)Q = 0, \tag{55}$$

for arbitrary values of E. This means that the operators $(E - H_0)$ and $(E - H_0)^{-1}$ are already partly diagonalized:

$$E - H_0 = \begin{array}{c} \begin{matrix} O & P_1 & Q \end{matrix} \\ \left[\begin{array}{c|c|c} & 0 & 0 \\ \hline 0 & & 0 \\ \hline 0 & 0 & \end{array}\right] \end{array};$$

$$(E - H_0)^{-1} = \begin{array}{c} \begin{matrix} O & P_1 & Q \end{matrix} \\ \left[\begin{array}{c|c|c} & 0 & 0 \\ \hline 0 & & 0 \\ \hline 0 & 0 & \end{array}\right] \end{array}. \tag{56}$$

The diagonal submatrices in $(E - H_0)^{-1}$ are in order the inverse of the diagonal submatrices in $(E - H_0)$; we note that the (OO) matrix has only one element, whereas the (P_1P_1) matrix and the (QQ) matrix are both infinite. This gives the important relation

$$P_1T_0Q = 0, \tag{57}$$

which is also easily proven directly. Hence we obtain $T_0Q = PT_0Q = QT_0Q = Q[\alpha \cdot O + Q(E - H_0)Q]^{-1}Q$, and for this operator we will use the symbolic notation

$$T_0^Q = Q/(E - H_0), \tag{58}$$

so that now $T_0Q = T_0^Q$. We note that this transformation is specific for T_0 and that it is not valid for T, since the operator $(E - H)$ usually has a nonvanishing (P_1Q) component. In order to handle the operator T, we will proceed in a different way.

Using the identity $(A - B)^{-1} = A^{-1} + A^{-1}B(A - B)^{-1}$ and putting $A = \alpha \cdot O + P(E - H_0)P$ and $B = PVP$, we obtain the relations

$$T = T_0 + T_0VT = T_0(1 + VT), \tag{59}$$

$$TV = T_0(V + VTV) = T_0t, \tag{60}$$

$$t = V + VT_0t, \tag{61}$$

of which the last is the Lippmann-Schwinger integral equation; for details and references, see an earlier paper.[15] Multiplying (61) to the right by O and using (51) and (58), we obtain

$$tO = (V + VT_0^Qt)O. \tag{62}$$

This equation gives by iteration

$$\begin{aligned} T_0^QtO &= T_0^QVO + T_0^QVT_0^QtO \\ &= (T_0^Q + T_0^QVT_0^Q + T_0^QVT_0^QVT_0^Q + \cdots)VO \qquad (63) \\ &= \frac{Q}{(E - H_0) - V}VO \\ &= \frac{Q}{E - H}H_{\text{int}}O = T_QH_{\text{int}}O, \qquad (63) \end{aligned}$$

which leads to the fundamental transformation formula

$$TVO = [Q/(E - H_0)]tO = [Q/(E - H)]H_{\text{int}}O. \tag{64}$$

This relation is here derived by means of an infinite power series expansion, but we note that it can also be proven directly. The operator of essential interest is here given by the symbolic notation

$$T_Q = Q/(E - H), \tag{65}$$

but it should, of course, be strictly defined by the relation $T_Q = Q[\alpha \cdot O + Q(E - H)Q]^{-1}Q$ which is analogous to (20). According to (31), we get as a

first consequence of (64) that

$$\Omega = \{1 + [Q/(E - H)]H_{\text{int}}\}O, \tag{66}$$

and

$$\boxed{\Psi = \varphi_0 + [Q/(E - H)]H_{\text{int}}\varphi_0.} \tag{67}$$

The result implies that $P_1\Omega = 0$, i.e., if the exact wave function is expanded in terms of Hartree products, *all singly excited states with respect to* φ_0 *are missing*. This property of the wave function is a direct consequence of the Brillouin-Brueckner theorem (49), i.e., of the definitions (43) of the SCF potentials u_i.

For a two-particle system ($N = 2$), one can now show that the first SCF orbital ψ_1 associated with the ground state must be identical to the first natural orbital.[16] This follows directly from the fact that, in (67), there are no singly excited states, i.e., the expansion is diagonalized with respect to the first orbital, and a comparison with the so-called natural expansion proves then the theorem. The connection between the SCF orbitals and the natural orbitals for many-particle systems is now also being studied.

Simplification of the reaction operator. In this section, we will use the Brillouin-Brueckner theorem (49) to simplify the formal expression for the reaction operator. Let us start by introducing the *modified reaction operator* κ defined by the relation

$$\kappa = H_{\text{int}} + V[Q/(E - H)]H_{\text{int}}. \tag{68}$$

According to (64), one has $tO = (V + VTV)O = (V + VT_QH_{\text{int}})O = (-u + \kappa)O$ and further

$$OtO = O(-u + \kappa)O, \tag{69}$$

$$P_1tO = P_1(-u + \kappa)O, \tag{70}$$

which relations show that κ may be used instead of τ in calculating the total energy and the SCF potentials. This is a considerable simplification, particularly since the expression (68) for κ depends explicitly on the potentials u_i only through the linear term and implicitly through the occurrence of the operator Q. The operator κ does not contain any one-particle terms and, instead of (43), we have now

$$\begin{aligned} u_i &= u_i^{(2)} + u_i^{(3)} + \cdots + u_i^{(N)}, \\ u_i^{(2)} &= \sum_{j \neq i} (\psi_i \, |\kappa_{ij}| \, \psi_i), \\ u_i^{(3)} &= \frac{1}{2!} \sum_{j,k \neq i}{}' (\psi_i\psi_k \, |\kappa_{ijk}| \, \psi_i\psi_k). \end{aligned} \tag{71}$$

According to (41), the total energy takes then the form

$$E = E_0 + \langle\varphi_0 \, |-u + \kappa| \, \varphi_0\rangle = \langle\varphi_0 \, |\sum_i H_i + \kappa| \, \varphi_0\rangle, \tag{72}$$

and these expressions may be used for evaluating the denominator $(E - H)$ in (68). For the iterative element contained in calculating κ according to (68), we will refer to a previous paper.[15]

Using (67), (68), (70), and (72), we can now summarize our results in the following formulas for the wave function Ψ and the energy E:

$$\Psi = \{1 + [Q/(E - H)]H_{\text{int}}\}\varphi_0, \tag{73}$$

$$E = \left\langle \varphi_0 \left| \sum_i H_i + H_{\text{int}} + H_{\text{int}} \frac{Q}{E - H} H_{\text{int}} \right| \varphi_0 \right\rangle, \tag{74}$$

whereas the SCF potentials u_i are defined by the relations

$$\left\langle \varphi_{\text{s.e.}} \left| -\sum_i u_i + H_{\text{int}} + V \frac{Q}{E - H} H_{\text{int}} \right| \varphi_0 \right\rangle = 0. \tag{75}$$

These equations are so simple and condensed that one wonders whether they could be derived in a more direct way, and that this is the case will actually be shown in Sec. 4.

Linked-cluster expansion. The key problem in the application of the SCF theory is apparently the evaluation of the modified reaction operator and, according to (68) and (72), one has

$$\kappa = H_{\text{int}} + V \frac{Q}{(E_0 - H_0) - (V - \langle t\rangle_0)} H_{\text{int}}, \tag{76}$$

where $\langle t\rangle_0 = \langle -u + \kappa\rangle_0$. This problem may be approached in several different ways.[15] Expansion of the inverse in power series gives

$$\kappa = H_{\text{int}} + VR_0^Q \sum_{k=0}^{\infty} [(V - \langle t\rangle_0)R_0^Q]^k H_{\text{int}}, \tag{77}$$

with the notation

$$R_0^Q = Q/(E_0 - H_0), \tag{78}$$

and, if this expansion is systematized after powers of H_{int}, one obtains a result closely associated with the Schrödinger perturbation theory. In connection with many-particle systems, one has found that considerable simplifications can be achieved by means of the so-called linked-cluster theorem. It was pointed out by Bardeen that the formal expansions in the many-particle theory contained certain terms of a rather strange character, and it

[16] P. O. Löwdin, and H. Shull, Phys. Rev. **101**, 1730 (1956).

was shown by Brueckner[17] that all these "unlinked" terms would actually cancel depending on certain algebraic identities. The general proof of this theorem was given by Goldstone[18] by means of field-theoretical methods.

Here we will make only some brief comments in this connection concerning certain energy contributions which one can easily show give vanishing contributions, and which can hence be omitted. The proof is built on a simple separability theorem for the many-particle eigenvalue problem, and no reference to field-theoretical considerations is necessary. For this purpose, it is convenient to write the interaction term in the Hamiltonian in the form

$$H_{\text{int}}(\lambda) = \frac{1}{2!}\sum_{ij}{}' \lambda_{ij}H_{ij} + \frac{1}{3!}\sum_{ijk}{}' \lambda_{ijk}H_{ijk} + \cdots, \tag{79}$$

where the quantities λ_{ij}, λ_{ijk}, $\cdots$ are interaction parameters at our disposal which have to be put equal to 1 in the final results. Let us start by considering the Schrödinger-type perturbation theory, let us leave the question of self-consistency aside, and let u_i be arbitrary fixed potentials, so that $H_0 = \sum_i (H_i + u_i)$ and $V = -u + H_{\text{int}}$. The energy E is given by (34) with a reaction operator t of the form (35). Using a power series expansion of the inverse, one obtains

$$t = V + VR_0 \sum_{k=0}^{\infty} [(V - \langle t\rangle_0)R_0]^k V, \tag{80}$$

where

$$R_0 = P/(E_0 - H_0), \tag{81}$$

or more strictly $R_0 = P[\alpha\cdot O + P(E_0 - H_0)P]^{-1}P$. It is then possible to systematize the results after powers of V, leading to the well-known Schrödinger-type formulas,[15] and we will now use the parameters λ to further classify the terms within each order.

Let us divide the particles 1, 2, $\cdots$ N into groups (a), (b), (c), $\cdots$, and let λ_{aa}, λ_{bb}, λ_{cc} $\cdots$ denote the sets of interaction parameters λ_{ij}, λ_{ijk}, $\cdots$ involving only the particles within such a group. We will further let λ_{ab}, λ_{ac}, λ_{bc}, λ_{abc}, $\cdots$ denote the sets of interaction parameters connecting particles within different groups. A term in the energy shift $\langle\varphi_0 |t| \varphi_0\rangle$ which contains the interaction parameters of at least two groups (a) and (b) as factors:

$$(\lambda_{aa})(\lambda_{bb}) \tag{82}$$

without containing simultaneously some factors from the set λ_{ab} will be called an *unlinked term*, and the energy contribution from all terms of the same type will be denoted by

$$E(\lambda_{aa}, \lambda_{bb}) \tag{83}$$

indicating what λ factors are involved. This energy contribution is apparently independent of the value of the interaction parameters in the set λ_{ab}. If one temporarily puts $\lambda_{ab} = 0$, the total Hamiltonian takes the simplified form

$$H = H_a + H_b, \tag{84}$$

and the corresponding eigenvalue problem is then partly separable, so that

$$\Psi = \Psi_a\cdot\Psi_b, \tag{85}$$

$$E = E_a + E_b, \tag{86}$$

where Ψ_a and Ψ_b are usually many-particle functions satisfying the relations

$$H_a\Psi_a = E_a\Psi_a, \qquad H_b\Psi_b = E_b\Psi_b, \tag{87}$$

where E_a and E_b depends only on λ_{aa} and λ_{bb}, respectively. Apparently there are no product terms of the type (82), and, for the unlinked terms under consideration, one has consequently

$$\lim_{\lambda_{ab}=0} E(\lambda_{aa}, \lambda_{bb}) = 0. \tag{88}$$

However, since the value of $E(\lambda_{aa}, \lambda_{bb})$ is independent of the value of the parameters λ_{ab}, one obtains

$$\boxed{E(\lambda_{aa}, \lambda_{bb}) = 0,} \tag{89}$$

i.e., all unlinked energy terms of the same type will cancel identically. The result is, of course, also true for all various orders in λ separately.

In treating a specific term in the λ expansion of the energy shift $\langle\varphi_0 |t| \varphi_0\rangle$, one has now to investigate the λ factor to see whether it is possible to split off at least one group of particles (b) which are not linked to the others (a) through parameters λ_{ab}. If this is the case, the term is unlinked and belongs to a group of terms which will cancel each other identically; the term can hence be omitted. It is evident that the energy shift $\langle\varphi_0 |t| \varphi_0\rangle$ can be systematically reduced in this way, so that it finally contains contributions only from "linked terms," where the particles connected with the λ factors cannot be separated into groups. This is the content of the linked-cluster theorem in the many-particle perturbation theory.

[17] K. A. Brueckner and C. A. Levinson, Phys. Rev. **97**, 1344 (1955); K. A. Brueckner, *ibid.* **100**, 36 (1955); for a survey, see the various papers in *The Many-Body Problem: L'Ecole d'été de physique théorique, Les Houches* 1958, edited by C. DeWitt (John Wiley & Sons, Inc., New York, 1959).

[18] J. Goldstone, Proc. Roy. Soc. (London) **A239**, 267 (1957).

Some examples of unlinked λ factors are listed below:

$$\lambda_{ij}\lambda_{kl}, \quad \lambda_{ij}\lambda_{klm}, \quad \lambda_{ij}^2\lambda_{kl}^2, \cdots, \tag{90}$$

where we have assumed that different indices denote different particles. The following combinations

$$\lambda_{ij}\lambda_{jk}, \quad \lambda_{ij}^2, \quad \lambda_{ij}\lambda_{jkl}, \quad \lambda_{ij}\lambda_{jk}\lambda_{ki}, \cdots \tag{91}$$

are all examples of linked λ factors. In our study we have, so far, used only the separability of the Hamiltonian (84), and we have not considered the explicit form of the matrix elements whatsoever. In treating the linked terms, further simplifications are often possible by observing that only those interactions which will let the particles return to the "ground state" φ_0 will give nonvanishing contributions.

Here we have considered the expansion of the energy shift $\langle\varphi_0 |t| \varphi_0\rangle$ in terms of a power series in the interaction parameters λ, but even other forms are possible.[15] In this connection, it is worthwhile to note that the linked-cluster theorem may be expressed in the somewhat more general form

$$\lim_{\lambda_{ab}=0} E = E_a(\lambda_{aa}) + E_b(\lambda_{bb}), \tag{92}$$

i.e., if the interaction parameters λ_{ab} connecting two groups (a) and (b) are put equal to zero, the total energy should be additive according to (86).

The treatment of the wave function Ψ is more complicated, since it depends on whether one could find some convenient form for $\ln \Psi$. According to (85), one has

$$\lim_{\lambda_{ab}=0} \ln \Psi = \ln \Psi_a(\lambda_{aa}) + \ln \Psi_b(\lambda_{bb}), \tag{93}$$

which gives an additivity theorem analogous to (92). In case it is possible to express $\ln \Psi$ in terms of a power series in λ, this series will hence contain contributions only from linked λ factors. In the same way, one can conclude that the logarithm of the normalization integral is characterized by a linked-cluster theorem.

The many-particle theory is in a certain dilemma, since the formal expressions obtained contain apparently a large number of irrelevant terms which give vanishing contributions in the expansions but which are still hard to eliminate in the original expressions. It is hence important to refine the theory in this respect, and valuable results have been obtained by Brueckner and his collaborators.[4,17] One can understand some of the difficulties involved by considering an arbitrary operator

$$\Omega_{op} = \sum_i \Omega_i + \frac{1}{2!}\sum_{ij}{}' \omega_{ij}\Omega_{ij} + \frac{1}{3!}\sum_{ijk}{}' \omega_{ijk}\Omega_{ijk} + \cdots, \tag{94}$$

where the quantities ω_{ij}, ω_{ijk}, $\cdots$ are interaction parameters at our disposal which in the final results are put equal to 1. Putting $\omega_{ab} = \omega_{abc} = \cdots = 0$, one obtains $\Omega' = \Omega_a + \Omega_b$ and, if further $\lambda_{ab} = \lambda_{abc} = \cdots = 0$, application of (85) gives

$$\lim_{\substack{\lambda_{ab}=0 \\ \omega_{ab}=0}} \langle\Omega_{op}\rangle = \frac{\langle\Psi_a |\Omega_a| \Psi_a\rangle}{\langle\Psi_a | \Psi_a\rangle} + \frac{\langle\Psi_b |\Omega_b| \Psi_b\rangle}{\langle\Psi_b | \Psi_b\rangle}, \tag{95}$$

i.e., an additivity theorem. In treating a specific term in the $(\lambda\omega)$ expansion of $\langle\Omega_{op}\rangle$, one has now to investigate the combined λ and ω factor to see whether it is possible to split off at least one group of particles (b) which are not linked to the others through either λ_{ab} or ω_{ab} or many-particle interactions. If this is the case the term is unlinked and, using (95), one can prove that the total contribution to $\langle\Omega_{op}\rangle$ from all unlinked terms of the same type vanishes identically. Groups of particles which are unlinked with respect to λ may now be linked together through ω_{ij}, ω_{ijk}, $\cdots$ etc., so terms which are irrelevant in calculating the energy may become of importance in treating other physical quantities, provided these are at least of two-body nature.

In conclusion, we will return to the self-consistent-field problem. In each SCF cycle (45), the potentials u_i are considered as fixed, and one can then apply the results obtained in this section and simplify the expansions by means of the linked-cluster theorem. Even in the final stage, when self-consistency is reached, one can use the same type of arguments, but one has now to observe that the potentials u_i and the one-particle functions ψ_i actually also depend on the interaction parameters λ_{ij}, λ_{ijk}, $\cdots$ which influences the λ expansion of the total energy and leads to a modification of the linked-cluster theorem.

4. REFORMULATION OF THE SCF THEORY

Formulas (74) and (75) for the eigenfunction Ψ and the energy E, respectively, look rather different from the original partitioning formulas (31) and (32) for $V = -u + H_{int}$, and we have actually gone a rather long way to show how the subspace $P = P_1 + Q$ in the operator T could be reduced to the subspace Q by a proper choice of the SCF potentials u_i. The same result can now be obtained in a much simpler way, which also gives some new aspects on the entire SCF theory.

It has previously been shown[15] that, if the total Hilbert space connected with the N-particle problem is divided into two subspaces (a) and (b), the Schrödinger equation may be written in the matrix form

$$\begin{pmatrix} \mathbf{H}_{aa} & \mathbf{H}_{ab} \\ \mathbf{H}_{ba} & \mathbf{H}_{bb} \end{pmatrix} \begin{pmatrix} \mathbf{C}_a \\ \mathbf{C}_b \end{pmatrix} = E \begin{pmatrix} \mathbf{C}_a \\ \mathbf{C}_b \end{pmatrix} \tag{96}$$

which corresponds to two equations. Solving $\mathbf{C}_b$ from the second equation, one obtains

$$\mathbf{C}_b = (E \cdot \mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1} \mathbf{H}_{ba} \mathbf{C}_a, \tag{97}$$

and substitution of this expression into the first equation gives

$$\bar{\mathbf{H}}_{aa} \mathbf{C}_a = E \mathbf{C}_a, \tag{98}$$

with

$$\bar{\mathbf{H}}_{aa} = \mathbf{H}_{aa} + \mathbf{H}_{ab}(E \cdot \mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1} \mathbf{H}_{ba}. \tag{99}$$

We note that both subspaces (a) and (b) could be infinite. Let us now start out from the resolution $1 = O + P_1 + Q$, and let the subspaces (a) and (b) be characterized by the projection operators $\bar{O} = O + P_1$ and Q, respectively. This partitioning turns out to be more feasible than the previous one, where we instead used O and $P = P_1 + Q$. Writing (99) in operator form, we obtain

$$\bar{H} = O\left(H + H \frac{Q}{E - H} H\right) O. \tag{100}$$

The operator $\bar{H}$ would give a simple expression of the energy and the wave function, if it could be partly diagonalized, so that

$$H = \left[\begin{array}{c|c|c} & & \\ \hline & & \\ \hline & & \end{array}\right] \begin{matrix} O \\ P_1 \\ Q \end{matrix} \rightarrow \bar{H} = \left[\begin{array}{c|c} & 0 \\ \hline 0 & \end{array}\right] \begin{matrix} O & P_1 \end{matrix}. \tag{101}$$

The requirement $P_1 \bar{H} O = 0$ immediately gives

$$P_1\left(-u + H_{\text{int}} + V \frac{Q}{E - H} H_{\text{int}}\right) O = 0, \tag{102}$$

where we have used (52) to simplify the expression. This relation is apparently equivalent to (75), i.e., to the definition (71) of the SCF potentials u_i in terms of the modified reaction operator κ defined by (68). The total energy E is now given by the isolated diagonal element $O\bar{H}O$, and one obtains

$$E = \left\langle \varphi_0 \middle| \sum_i H_i + H_{\text{int}} + H_{\text{int}} \frac{Q}{E - H} H_{\text{int}} \middle| \varphi_0 \right\rangle, \tag{103}$$

i.e., formula (74), whereas the wave function Ψ corresponding to the vector $\mathbf{C}$ having the components $\mathbf{C}_a$ and $\mathbf{C}_b$ according to (97) and (101) takes the form

$$\Psi = \{1 + [Q/(E - H)] H_{\text{int}}\} \varphi_0, \tag{104}$$

which is identical to (73).

In this way, we have obtained a simple and transparent derivation of the fundamental formulas in the exact SCF theory. The characteristic feature of this theory is apparently that it tries to eliminate the subspace P_1 connected with the singly excited states entirely from the expressions for the wave function and the energy, and this is achieved by choosing the SCF potentials u_i properly, i.e., so that $P_1 \bar{H} O = 0$ according to (102).

5. SYMMETRY PROPERTIES

In discussing many-particle theory and correlation effects, the symmetry properties of the system are certainly highly important.[19] It is well known that, in the theory of fermions, the antisymmetry requirement connected with Pauli's exclusion principle eliminates essentially the correlation error connected with particles having parallel spins, and the total correlation error with respect to the Hartree scheme is then diminished by about 50%. The remaining correlation error in the Hartree-Fock scheme is apparently connected with particles having antiparallel spins and, even for electrons, it is quite large if one uses the idea of doubly filled orbitals. The method of pairing fermions with antiparallel spins in one and the same orbital is connected with the fact that it renders a simple way of constructing total wave functions having a pure total spin. It has been pointed out by the author[20] that the correlation error associated with antiparallel spins could be essentially diminished, if one used "different orbitals for different spins" and then constructed pure spin functions by considering the spin components which are uniquely defined by the Slater determinant D_0. In this *extended Hartree-Fock scheme*, the total wave function is hence approximated by a spin projection O of the determinant D_0, i.e., OD_0. Some applications to electronic systems have shown that about 85% of the correlation error associated with electrons having antiparallel spins could be removed in this way, so that actually only about

[19] For a discussion of the connection between symmetry and correlation, see P. O. Löwdin, Phys. Rev. **97**, 1474, 1490, 1509 (1955); Advances in Chem. Phys. **2**, 207 (1959).

[20] P. O. Löwdin, *Nikko Symposium, Molecular Physics*, edited by M. Kotani, (Maruzen Company, Ltd., Tokyo, 1954), p. 13. Phys. Rev. **97**, 1509 (1955); *Proceedings of the 10th Solvay Conference* (Gauthier-Villars, Paris, 1955), p. 71; Revs. Modern Phys. **32**, 328 (1960).

1/12 of the original correlation error has to be accounted for by real many-particle theory. Hence it seems highly desirable to try to incorporate the symmetry properties in the SCF theories.

Antisymmetry requirement. In order to study the antisymmetry property of the fermions, it is convenient to introduce the antisymmetry projection operator

$$O_{AS} = (N!)^{-1} \sum_p (-1)^p P, \tag{105}$$

where P is a permutation of the coordinates $\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_N$. The total Hamiltonian defined by (1) is assumed to be symmetric in all coordinates, so that $HP = PH$, and the operators P are then normal constants of motion. If one solves the Schrödinger equation $H\Psi = E\Psi$, the eigenfunctions Ψ will automatically belong to one of the symmetry classes associated with the permutations, and they will be symmetric, antisymmetric, etc. In the theory of fermions obeying the exclusion principle, only antisymmetric wave functions correspond to the states occurring in nature, and these wave functions will here be selected by the auxiliary condition

$$O_{AS}\Psi = \Psi \tag{106}$$

i.e., they should be eigenfunctions to O_{AS} associated with the eigenvalue 1. For eigenfunctions of other symmetry types, one has $O_{AS}\Psi \equiv 0$. Since $O_{AS}H = HO_{AS}$, the operator (105) represents a normal constant of motion, which for fermions has the value 1. O_{AS} is a projection operator which selects the "antisymmetric" subspace out of the total Hilbert space, and we note that this subspace is not only orthogonal but also *noninteracting* with the remaining part of the Hilbert space, i.e.,

$$O_{AS}H(1 - O_{AS}) = 0. \tag{107}$$

In studying the Hamiltonian $H = H_0 + V$, we will assume that also $H_0 = \sum_i (H_i + u_i)$ is symmetric in all coordinates so that $H_0O_{AS} = O_{AS}H_0$. As unperturbed wave function, one may then, instead of the Hartree product (4), choose the Slater determinant

$$D_0 = (N!)^{1/2}O_{AS}\varphi_0. \tag{108}$$

In the exact SCF theory, one can now proceed as before by means of the partitioning technique, observing that we can actually confine our interest to the antisymmetric subspace, since the Hamiltonian H is already partly diagonalized. The operator P in the reaction operator t defined by (35) does then no longer represent the total orthogonal complement to O but the orthogonal complement with respect to the antisymmetric subspace. In analogy with (13), the operator τ will now be modified through the antisymmetry so that $\tau_{ii} = 0$, $\tau_{iik} = \cdots = 0$, etc., which means that one can remove the restrictions on the summations in (43) and base the entire study on the fundamental invariant ρ defined by (16). In this respect, the introduction of the exchange terms will simplify also the exact SCF theory. We note that the potentials $u_i = u(\mathbf{x}_i)$ will be *the same* for all particles, which implies that $H_0 = \sum_i (H_i + u_i)$ will actually fulfill the basic symmetry requirement.

Other symmetry properties. Let us assume that there exists some other symmetry property which is represented by a normal constant of motion Λ commuting with H_{0p} and with O_{AS}, say the total spin $(S^2; S_z)$. Let us describe the total Hilbert space by a complete set $\{f_l\}$. By introducing the set of projection operators O_Λ associated with Λ, one can split the antisymmetric basis $\{O_{AS}f_l\}$ into a series of subsets $\{O_\Lambda O_{AS}f_l\}$, one for each eigenvalue to Λ. We can now confine our interest to *one* of these subspaces at a time, since each one of them is entirely independent of the other, being not only orthogonal but also noninteracting with respect to H_{0p}, and Λ. Within this subspace, we can now start from the unperturbed function $O_\Lambda D_0$ and carry out a partitioning procedure, evaluate the reaction operator t, and construct an exact SCF theory based on the fundamental invariant ρ. This is apparently a generalization of the extended Hartree-Fock scheme discussed above to an exact form. It has already been emphasized in connection with the electrons that the main part of the correlation error affecting the original Hartree scheme is removed by an inclusion of the symmetry requirements through the projection technique, and only a comparatively small part of the correlation error has then to be treated by true many-particle theory, i.e., by a study of the reaction operator.

6. DISCUSSION

The purpose of this paper is to give some common aspects on the SCF theories, and to study the line of development which goes from Hartree, via Brueckner, to the exact SCF theory by means of the partitioning technique. The formal framework is essentially the same, and the degree of accuracy depends on how one has approximated the modified reaction operator κ defined by (68), and how well one has eliminated the infinite subspace P_1 associated

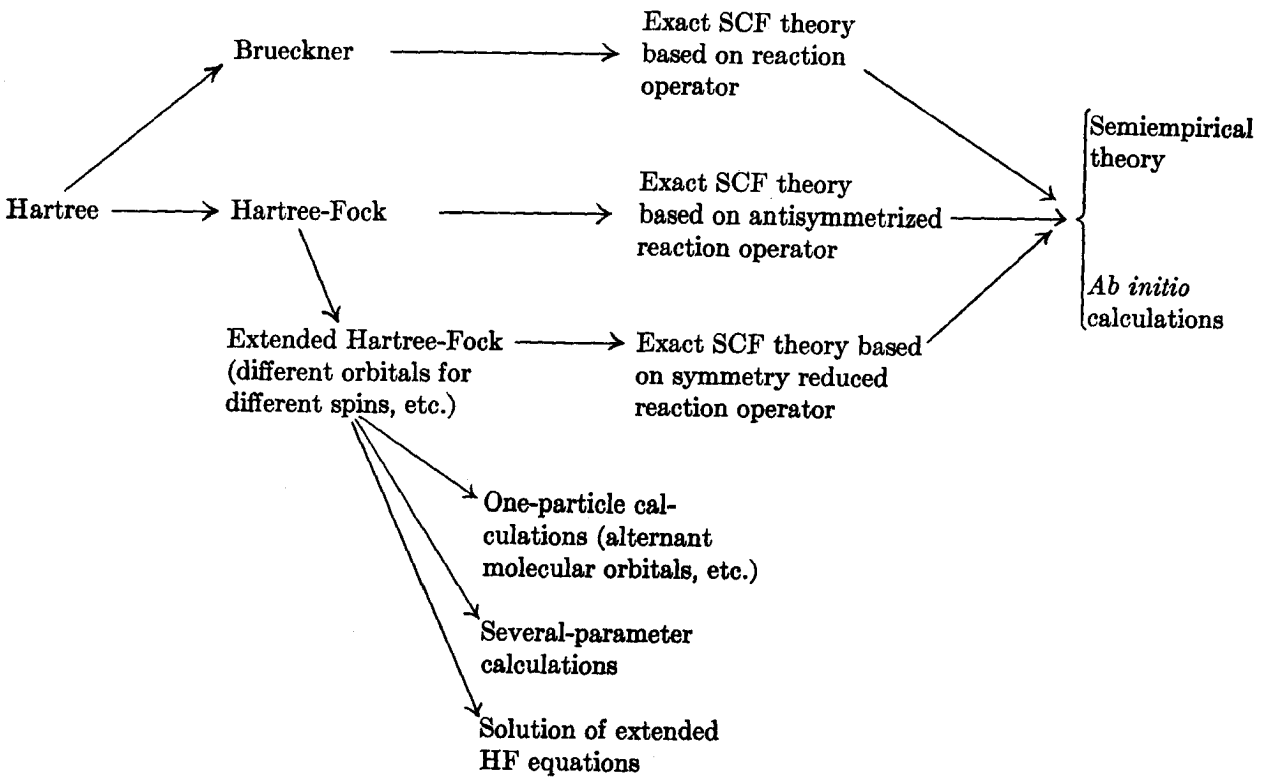


FIG. 1. Schematic survey of the various SCF schemes which may be utilized in connection with the development of many-particle theory.

with the singly excited states according to (102). One has for the various schemes:

Hartree:

$$\kappa \approx H_{\text{int}},$$

Brueckner:

$$\kappa \approx \frac{1}{2!}\sum_{ij}{}' \kappa_{ij}; \tag{109}$$

Exact SCF theory:

$$\kappa \approx \frac{1}{2!}\sum_{ij}{}' \kappa_{ij} + \frac{1}{3!}\sum_{ijk}{}' \kappa_{ijk} + \cdots .$$

The Brueckner scheme seems to be particularly appropriate for nuclear matter, where the forces between the particles are of short-range nature, whereas the treatment of many-electron systems with more long-range forces may require inclusion also of the many-particle correlation terms in κ. In Fig. 1, there is a diagram of the development in this field with some aspects also for the future.

In the introduction, it was emphasized that the independent-particle model had been successfully applied to many systems: for studying the electronic clouds of the atoms, for investigating the mobile π electrons of the conjugated organic compounds (the so-called Hückel scheme), for treating the band theory of solids and the nuclear shell model. It is now clear that the ultimate basis for this approach is not necessarily given by the Hartree and Hartree-Fock schemes, since refined theories of a similar type are now available in the form of the Brueckner scheme and the exact SCF method. The improved methods may be used either for *ab initio* calculations or for constructing semiempirical theories, and one can certainly expect a fruitful development in both connections.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

On a Functional Averaging Method in the Classical Ensemble Theory

A. GALINDO*†
Courant Institute of Mathematical Sciences,
New York University, New York and *Junta de Energia Nuclear, Madrid, Spain*
(Received February 20, 1962)

Albertoni, Bocchieri, and Loinger (ABL) have given a general prescription to calculate the average $E(F)$ of functionals $F[\rho]$ defined on the space $\mathfrak{L}$ of the initial states (Liouville density functions) $\rho(x)$ of an arbitrary (finite) dynamical system (S, Σ, μ, H). Then they have proved that $E(B_H) = 0$ for any $B \in \Sigma$, where

$$B_H[\rho] \equiv \lim_{T\to\infty} 1/T \times \int_0^T \left| \int_B \rho(x, t)\, d\mu(x) - \mu^{-1}(S)\mu(B) \right|^2 dt$$

[ABL theorem]. On the other hand, these authors claim that E gives the same weight to each $\rho \in \mathfrak{L}$, and therefore, they contend that their theorem is sufficient to justify the classical statistical mechanics.

In this note, however, the following statements are proved: (A) If (S, Σ, μ, H) is ergodic with discrete spectrum, then every probability measure ν in $\mathfrak{L}$ for which $B_H[\rho] = 0$, ν-almost everywhere (for any $B \in \Sigma$) must be concentrated at e [$e(x) \equiv \mu^{-1}(S)$ on S]. (B) Given (S, Σ, μ, H_o) with $H_o = 0$, every (normally defined) functional average $A(F)$ such that $A(B_{H_o}) = 0$ for any $B \in \Sigma$ must define a probability measure ν in $\mathfrak{L}$ such that $\nu(\{e\}) = 1$. From (B) it will follow that: (1) The ABL average $E(F)$ *does not* give the same weight to each $\rho \in \mathfrak{L}$; actually, E defines a probability measure in $\mathfrak{L}$ *concentrated* at e; consequently, no significant nontrivial conclusion can be naturally drawn from the ABL theorem regarding the foundations of the classical statistical mechanics; (2) it is impossible to restore in general the validity of the ABL theorem by choosing a different normal functional averaging method. To illustrate (2) a new (nonconcentrated) functional average $E^*(F)$ is introduced and it is proved that $E^*(B_H) \neq 0$ for every nonweakly mixing system (S, Σ, μ, H) and some $B \in H$.

INTRODUCTION

1

LET (S, Σ, μ) denote a space S with a measure[1] μ defined on a σ-ring Σ of subsets $B \subset S$, and such that μ is nonnegative, totally finite, separable, nonatomic, and $\mu(S) > 0$.

Assumption 1: All the basic measure spaces (S, Σ, μ) considered throughout all this paper will be tacitly assumed to satisfy the above requirements.

An element $\rho \in \mathfrak{K} \equiv L^2(S, \Sigma, \mu)$ is said to be a Liouville density function if and only if $\rho(x) \geq 0$ and $\int_S \rho\, d\mu = 1$; the subset of $\mathfrak{K}$ consisting of such functions ρ will be denoted by $\mathfrak{L}$.

Any (μ-preserving, invertible, and t-measurable) flow $\mathfrak{I}_t$ in S induces a one-parameter continuous unitary group U_t in $\mathfrak{K}$ which leaves $\mathfrak{L}$ invariant.[2-4] Let H be the infinitesimal generator of U_t; then[5]

$$U_t = \int e^{it\lambda}\, dE_\lambda, \tag{1.1}$$

where $\{E_\lambda\}$ is the (continuous on the right) spectral resolution of the identity associated to H. A simple application of von Neumann's mean ergodic theorem and the formula

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T |f(t)|^2\, dt = \sum_\lambda \left| \lim_{T\to\infty} \frac{1}{T} \int_0^T f(t) e^{-i\lambda t}\, dt \right|^2 \tag{1.2}$$

[valid whenever $f(t)$ is positive definite][2] yields

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T |(g_1, U_t g_2) - \mu(S)(g_1, e)(e, g_2)|^2\, dt = \sum_\lambda |(g_{1,\lambda}, g_2)|^2, \tag{1.3}$$

where

---

* The author is an exchange visitor under the International Atomic Energy Agency Training Program sponsored by the National Academy of Sciences.

† On leave of absence from the Junta de Energía Nuclear, Madrid, Spain.

[1] P. R. Halmos, *Measure Theory* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1950). An extended real-valued measure μ is called totally finite if $\mu(S)$ is finite. Given a totally finite nonnegative measure μ, let N denote the subset of Σ consisting of all μ-null sets; the set Σ/N is a metric space under the metric $d(E, F) = \mu(E\text{-}F) + \mu(F\text{-}E)$, where $E, F \in \Sigma/N$. Whenever this metric space is separable, one says that μ is separable. This definition can be readily extended to more general measures. Finally, a measure space (S, Σ, μ) is called nonatomic if there exists no element $A \in \Sigma$ such that $\mu(A) \neq 0$ and either $\mu(B) = 0$, or $\mu(B) = \mu(A)$ for any $B \in \Sigma$, $B \subseteq A$.

[2] E. Hopf, *Ergodentheorie* (Verlag Julius Springer, Berlin, Germany, 1937).

[3] P. R. Halmos, *Ergodic Theory* (Chicago University Notes, 1955).

[4] K. Jacobs, *Neuere Methoden und Ergebnisse der Ergodentheorie* (Springer-Verlag, Berlin, Germany 1960).

[5] F. Riesz and B. Sz-Nagy, *Lécons d'analyse fonctionelle* (Akademiai Kiado, Budapest, 1955), Chap. X.

$g_1, g_2 \in \mathfrak{K}$, $g_{1,\lambda} \equiv (E_\lambda - E_{\lambda-0})g_1 \quad (\lambda \neq 0)$,

$$g_{1,0} \equiv (E_0 - E_{0-0})g_1 - \mu^{-1}(S)(e, g_1)e,$$

and $e(x) \equiv \mu^{-1}(S)$ on S. [In both (1.2) and (1.3), $\sum_\lambda$ contains at most countably many terms different from zero[2]; besides, the expressions for $g_{1,\lambda}$ show that only those λ's belonging to the point spectrum of H may contribute to $\sum_\lambda$.]

The flow $\mathfrak{I}_t$ is said to be weakly mixing[2-4,6] whenever the member on the left of (1.3) vanishes for arbitrary $g_1, g_2 \in \mathfrak{K}$; therefore[2] (1.3) implies that in order for $\mathfrak{I}_t$ to be weakly mixing it is necessary and sufficient that H have no eigenvalue other than $\lambda = 0$, and that this be simple (first mixing theorem).

Many important questions concerning the classical systems can be suitably posed and analyzed within the abstract framework of a measure space with a distinguished flow (for instance, an energy shell with the Liouville measure and the flow representing the time evolution of the system). Therefore it is natural (and suggestive as well) to say that every (S, Σ, μ, H) represents (or defines) a finite $[\equiv \mu(S) < \infty]$ dynamical system. In addition, this will be called ergodic, weakly mixing, etc., whenever the associated flow $\mathfrak{I}_t$ is ergodic, weakly mixing, etc., respectively.

Given (S, Σ, μ, H) and $B \in \Sigma$, let us define on $\mathfrak{L}$ the following functional:

$$B_H[\rho] \equiv \lim_{T\to\infty} \frac{1}{T} \times \int_0^T dt \left| \int_B \rho(x,t)\, d\mu(x) - \mu^{-1}(S)\mu(B) \right|^2, \qquad (1.4)$$

where $\rho(x, t) \equiv (U_t\rho)(x)$. As $\int_B \rho(x, t)\, d\mu(x) = (\varphi_B, U_t\rho)$ and $\mu^{-1}(S)\mu(B) = \mu(S)(\varphi_B, e)(e, \rho)$, where φ_B is the characteristic function (ch. f.) of B, (1.3) shows that

$$B_H[\rho] = \sum_\lambda |(\varphi_{B,\lambda}, \rho)|^2 = (\rho, H_B\rho), \qquad (1.5)$$

H_B being the nonnegative self-adjoint integral operator in $\mathfrak{K}$ with kernel

$$H_B(x, y) = \sum_\lambda \varphi^*_{B,\lambda}(x)\varphi_{B,\lambda}(y). \qquad (1.6)$$

It is clear from (1.6) that H_B belongs to the trace class[7] and that

$$\mathrm{Tr}\, H_B = \sum_\lambda ||\varphi_{B,\lambda}||^2. \qquad (1.7)$$

2

Let (S, Σ, μ) be a measure space (see assumption 1, in Sec. 1). The subset $\mathfrak{L} \subset \mathfrak{K} \equiv L^2(S, \Sigma, \mu)$, consisting of the Liouville density functions, is a nonlinear topological space (with the topology relative to $\mathfrak{K}$); given a partition π: $S = B_1 \cup B_2 \cup \cdots \cup B_n$ $[B_i \cap B_j = \phi$ for $i \neq j$, $B_i \in \Sigma$, $\mu(B_i) > 0]$, let us associate to each $\rho \in \mathfrak{L}$ its projection

$$\rho_\pi \equiv \sum_1^n \mu^{-1}(B_i)(\varphi_{B_i}, \rho)\varphi_{B_i}. \qquad (2.1)$$

Clearly:

$$\rho_\pi \in \mathfrak{L}, \qquad \lim_\pi \rho_\pi = \rho \qquad (2.2)$$

(as the partitions form a directed set this limit will be understood as a directed limit). On the other hand,

$$P_\pi : \rho_\pi \to ((\varphi_{B_1}, \rho), \cdots, (\varphi_{B_n}, \rho)) \qquad (2.3)$$

is a (homeomorphic) map of $\mathfrak{L}_\pi$ *onto* the hypertriangle

$$T_\pi \equiv \left\{(\xi_1, \cdots, \xi_n) \in R^n : \xi_i \geq 0, \sum_1^n \xi_i = 1\right\} \qquad (2.4)$$

$$(\mathfrak{L}_\pi \equiv \{\rho \in \mathfrak{L} : \rho_\pi = \rho\}).$$

Therefore, we can easily extend to $\mathfrak{L}$, without any essential alteration, some of the simplest basic ideas and methods of the functional integration[8,9] in linear topological spaces.

A functional $F[\rho]$ will be said to be a π-cylinder functional (π-c. f.) iff $F[\rho] = F[\rho_\pi]$ for every $\rho \in \mathfrak{L}$; a subset $\mathfrak{L}' \subset \mathfrak{L}$ whose ch. f. is a π-c. f. will be called a π-cylinder set (π-c. s.). When π is left unspecified, we shall briefly say that F is a c. f. and $\mathfrak{L}'$ a c. s. Suppose now given for each π a nonnegative Radon measure[10] $\bar{\nu}_\pi$ in T_π such that $\bar{\nu}_\pi(T_\pi) = 1$, and define

$$\nu_\pi(P_\pi^{-1}C) \equiv \bar{\nu}_\pi(C) \qquad (2.5)$$

[6] The definition of weak mixing given in references 3 and 4 applies to the case of a discrete flow consisting of the powers of a fixed μ-preserving invertible transformation $\mathfrak{I}$. This definition and the one given in reference 2 share the property that both imply that any transformation in the flow has a simple point spectrum consisting only of the number 1. See for instance P. R. Halmos, *Lectures on Ergodic Theory* (The Mathematical Society of Japan, 1956), p. 39.

[7] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, Germany, 1960), Chap. III. A bounded linear operator A in a separable Hilbert space $\mathfrak{H}$ is said to belong to the trace class if $\Sigma_j(\varphi_j, [A]\varphi_j) < \infty$ for a fixed orthonormal basis $\{\varphi_j\}$ in $\mathfrak{H}$, where $[A] \equiv (A^*A)^{\frac{1}{2}}$. It is then shown that $\Sigma_j(\varphi_j, A\varphi_j)$ is also finite. $\Sigma_j(\varphi_j, [A]\varphi_j)$ and $\Sigma_j(\varphi_j, A\varphi_j)$ are independent of the basis chosen. The trace of A is defined as $\mathrm{Tr}\, A \equiv \Sigma_j(\varphi_j, A\varphi_j)$.

[8] K. O. Friedrichs, H. N. Shapiro, J. Schwartz, B. Wendroff, and T. Seidman, *Integration of Functionals* (New York University Notes, 1957), Chaps. I and III–VIII.

[9] I. M. Gel'fand and N. Ya. Vilenkin, *Obobščennye Funkcii* (Gosudarstvennyi Izdatel'stvo Fizicheskii-Matematichekii Literatura, Moscow, 1961), Vol. 4, Chap. IV.

for every Borel set $C \subset T_\pi$. It is clear that for each π, Eq. (2.5) defines a probability measure ν_π on the σ-ring, Σ_π consisting of all the π-Borel cylinder sets (π-B. c. s.) $P_\pi^{-1}C$. Since a π—B. c. s. $P_\pi^{-1}C$ may also be a c. s. [and therefore a B. c. s.] with respect to a different partition π', it is plain that in order for (2.5) [with π arbitrary] to define uniquely an additive set function ν on the ring $\bigcup_\pi \Sigma_\pi$ it is necessary and sufficient that the following compatibility condition,[8,9]

$$\nu_\pi(P_\pi^{-1}C) = \nu_{\pi'}(P_{\pi'}^{-1}C), \tag{2.6}$$

hold whenever $P_\pi^{-1}C$ is also a π'-B. c. s. It can be similarly proved that (2.6) holds if and only if

$$\int \Phi_\pi \, d\bar{\nu}_\pi = \int \Phi_{\pi'} \, d\bar{\nu}_{\pi'}, \tag{2.7}$$

where $\Phi_\pi(P_\pi\rho_\pi) \equiv F[\rho_\pi]$, $\Phi_{\pi'}(P_{\pi'}\rho_{\pi'}) \equiv F[\rho_{\pi'}]$, $F[\rho]$ being an arbitrary continuous c. f. with respect to both π and π'. Whenever (2.6) [or (2.7)] holds, it is possible therefore to *define* uniquely the integral $\int F$ as $\int \Phi_\pi \, d\bar{\nu}_\pi$. This integral obviously satisfies all the requirements of an expectation value (or functional average) in the (linear) space $\mathfrak{F}$ of all the continuous c. f.'s; besides, if $\{F_n\}$ is a sequence of continuous π-c. f.'s (π independent of n) such that $F_n \to 0$ pointwise, then $\int F_n \to 0$ (an obvious consequence of the definition of $\int F$ and the properties[10] of the Radon measure $\bar{\nu}_\pi$ in T_π). These functional averages (in $\mathfrak{F}$) so introduced will be called *normal* and only they will be considered in this paper. [It is not difficult to prove that any functional average $A : F \to A(F)$ ($F \in \mathfrak{F}$) such that $A(F_n) \to 0$ whenever F_n ($n = 1, 2, \cdots$) are π-c. f.'s (π fixed) and $F_n \to 0$ pointwise, must be normal.]

Finally, suppose that A is a normal functional average: given any continuous functional F, we shall define $A(F)$ as

$$A(F) \equiv \lim_\pi A(F_\pi) \qquad (F_\pi[\rho] \equiv F[\rho_\pi]) \tag{2.8}$$

provided that this limit exists and is finite (clearly $F_\pi \in \mathfrak{F}$). To see that this definition is consistent with the original A for c. f.'s, it suffices to note that $A(F_\pi) = A(F_{\pi'})$, whenever F is a π-c. f. and π' is finer than π. [The extension of A via (2.8) to the linear space $\bar{\mathfrak{F}}_A$ consisting of those continuous functionals for which (2.8) is defined is also a functional average which will be referred to as the natural extension of the normal functional average A in $\mathfrak{F}$.]

[10] N. Bourbaki, *Eléments de mathématique, Livre VI, Intégration* (Hermann & Cie., Paris, France, 1952), Chap. III. Let X, $\mathcal{G}$ be, respectively, a compact topological space and the linear space of the continuous real-valued functions defined on X. Suppose $\mathcal{G}$ endowed with its uniform topology. Any continuous complex-valued linear functional $I(g)$ defined on $\mathcal{G}$ is called a Radon measure.

3

Albertoni, Bocchieri, and Loinger[11] (ABL) have introduced a functional average, which we shall denote by E, for a certain class $\mathcal{E}$ of functionals $F[\rho]$ defined on $\mathfrak{L}$.

Concerning E, the author of the present paper has proved[12]: (i) $\mathfrak{F} \subset \mathcal{E}$; (ii) the restriction E_0 of E to $\mathfrak{F}$ is a normal functional average such that $E_0(F) = F[e]$ for every $F \in \mathfrak{F}$ [in other words, E_0 is associated (see Sec. 2) to a set $\{\bar{\nu}_\pi\}$ of probability measures $\bar{\nu}_\pi$ in T_π satisfying $\bar{\nu}_\pi(\{P_\pi e\}) = 1$ or, equivalently, $\bar{\nu}_\pi$ is a Dirac-delta measure concentrated at $P_\pi e$][13]. (iii) No probability measure σ exists in $\mathfrak{L}$ such that $\int F \, d\sigma = E(F)$ for every $F \in \mathcal{E}$ [the proof of (iii) is essentially based on this consideration: the functional $G[\rho] \equiv ||\rho||^2$ belongs to $\mathcal{E}$, and $E(G) = 2$; on the other hand, there exists a sequence $\{G_n\} \subset \mathcal{E}$ such that $G_n \uparrow G$ pointwise, while $E(G_n) = G_n[e] \uparrow G[e] = 1 \neq E(G)$].

It is clear that (ii) contradicts the ABL contention[11] that E gives the same weight to each $\rho \in \mathfrak{L}$; on the other hand, (iii) answers the question left open in reference 11 as to whether E is associated to a probability measure in $\mathfrak{L}$.

By making use of E, ABL have proved a theorem which may be stated as follows: For an arbitrary finite dynamical system (S, Σ, μ, H), the relation

$$E(B_H) = 0 \tag{3.1}$$

holds for every $B \in \Sigma$. [See (1.4) for the definition of B_H.]

A stronger form of this theorem has been recently given by Lomont[14]; he has proved that

$$E(B_{H,t}) = 0 \tag{3.2}$$

for every $B \in \Sigma$ and t, where

$$B_{H,t}[\rho] \equiv \left| \int_B \rho(x, t) \, d\mu(x) - \mu^{-1}(S)\mu(B) \right|^2. \tag{3.3}$$

[Both results (3.1) and (3.2) follow at once from this more general statement[12]: Every monomial

$$M[\rho] \equiv \int M(x_1, \cdots, x_n)\rho(x_1) \cdots \times \rho(x_n) \, d\mu(x_1) \cdots d\mu(x_n) \tag{3.4}$$

[11] S. Albertoni, P. Bocchieri, and A. Loinger, J. Math. Phys. 1, 244 (1960).

[12] A. Galindo, Communication presented at the II Reunión de Matemáticos Españoles, Zaragoza, Spain, November 1961 (to be published).

[13] Throughout this paper we will say that a probability measure ν is concentrated at a fixed point e when $\nu(\{e\}) = 1$.

[14] J. S. Lomont, J. Math. Phys. 2, 858 (1961).

with

$$\int |M(x_1, \cdots, x_n)|^2 \, d\mu(x_1) \cdots d\mu(x_n) < \infty$$

belongs to $\bar{\mathfrak{F}}_E$, and the following relation

$$E(M) = E_0(M) = M[e] \tag{3.5}$$

holds.]

The aim of the present paper is to show that the concentration phenomenon indicated in (ii) cannot be avoided by choosing a different technique of functional averaging while maintaining the general validity of the ABL theorem. More precisely, we shall establish the following statements:

(A) If (S, Σ, μ, H) is ergodic with discrete spectrum, and ν is a probability measure in $\mathfrak{L}$, then $\nu(\{\rho \in \mathfrak{L} : B_H[\rho] \neq 0\}) = 0$ for every $B \in \Sigma$ if and only if $\nu(\{e\}) = 1$.

Consequently, it is impossible, in general, to find a nontrivial probability measure ν such that $B_H[\rho] = 0$ ν-almost everywhere for every $B \in \Sigma$. (Nontrivial $\equiv$ nonconcentrated at e; note that $B_H[e]$ is always zero.)

(B) Given (S, Σ, μ, H_0) with $H_0 = 0$, every normal functional average A in $\mathfrak{F}$ such that $A(B_{H_0}) = 0$ for each $B \in \Sigma$ must define in $\mathfrak{L}$ a probability measure, ν, concentrated at e (i.e., the additive set function ν on $\bigcup_\tau \Sigma_\tau$ associated to A (see 2) must be countably additive and its extension to the minimal σ-ring containing $\bigcup_\tau \Sigma_\tau$ must satisfy $\nu(\{e\}) = 1$).

The ABL theorem implies that $E_0(B_{H_0}) = 0$; therefore (B) [together with the fact that E_0 is normal[11]] provides an alternative proof of the "concentration" of E_0, since, first, the definition of E_0 does not depend upon H, and secondly, if (S, Σ, μ, H) is a finite dynamical system, (S, Σ, μ, H_0) is also one (and a trivial one, indeed). As we anticipated elsewhere,[11] this result clearly indicates, in our view, that no significant nontrivial conclusions can be reasonably drawn from the ABL theorem regarding the foundations of the classical statistical mechanics.

On the other hand, (ii) and (B) imply that the normal functional average E_0 introduced by ABL can be uniquely characterized by (3.1) [note (3.5)]. Therefore, the general validity of the ABL theorem cannot be restored by choosing a different (and more suitable) normal functional averaging method.

Finally, and to illustrate (B), we shall exhibit a new normal functional average E^* associated to measures $\bar{\nu}_\tau$ with positive density functions and we will verify that $E^*(B_H) \neq 0$ for some $B \in \Sigma$ whenever (S, Σ, μ, H) is not a weakly mixing system.

PROOFS OF STATEMENTS (A) AND (B)

4

Proof (A). Let (S, Σ, μ, H) be ergodic with discrete spectrum; as μ is separable (assumption 1, Sec. 1) there exists a countable family, say $\{B_n\}$, of elements of Σ such that $(\varphi_{B_n}, g) = 0$ $(n = 1, 2, \cdots, g \in \mathfrak{H})$ implies[1] $g = 0$. Therefore, if $\{f_{\lambda_n}\}$ denotes the (complete orthonormal) system of (simple) eigenfunctions of H $(Hf_{\lambda_n} = \lambda_n f_{\lambda_n})$, it is clear that, for an arbitrarily given m, there exists some integer, say m_1, such that $\varphi_{m_1, m} \equiv (f_{\lambda_m}, \varphi_{B_{m_1}}) f_{\lambda_m} \neq 0$; in other words, the system $\{\varphi_{m_1, m}\}$ $(m = 1, 2, \cdots)$ is orthogonal and complete. Consequently, in order for $\rho \in \mathfrak{L}$ to satisfy

$$(B_n)_H[\rho] = 0 \qquad (n = 1, 2, \cdots), \tag{4.1}$$

it is necessary and sufficient that $\rho = e$ [see 1 for the definitions of $\mathfrak{L}$ and $(B_n)_H$, and take also into account the spectral properties of H].

Suppose now that ν is a probability measure in $\mathfrak{L}$ such that: Every $B_H[\rho]$ is ν-integrable and $\int B_H \, d\nu = 0$ [or equivalently, since $B_H \geq 0$, that $\nu(\{\rho \in \mathfrak{L} : B_H[\rho] \neq 0) = 0$ for every $B \in \Sigma$]. The above result implies that

$$\{e\} = \cap_n \{\rho \in \mathfrak{L} : (B_n)_H[\rho] = 0\} \tag{4.2}$$

and therefore,

$$\mathfrak{L} - \{e\} = \cup_n \{\rho \in \mathfrak{L} : (B_n)_H[\rho] \neq 0\}. \tag{4.3}$$

Consequently

$$\nu(\mathfrak{L} - \{e\}) \leq \sum_n \nu(\{\rho \in \mathfrak{L} : (B_n)_H[\rho] \neq 0\}) = 0, \tag{4.4}$$

and hence $\nu(\{e\}) = 1$.

This proves the "only if" part of (A); the "if" part is trivial, since $B_H[e] = 0$ for every $B \in \Sigma$. Therefore, (A) is established, Q.E.D.

Remark 1. That in some cases a stronger statement may actually hold is revealed by the following significant example:

Take S as the 2-dimensional torus[2] $x_i \pmod 1$ $i = 1, 2$, μ as the ordinary Lebesgue measure in the unit square, and $\mathfrak{I}_t : (x_1, x_2) \to (x_1 + t, x_2 + at)$, where a is irrational. It is well known[2] that $\mathfrak{I}_t$ is ergodic with discrete spectrum. The eigenfunctions f_{k_1, k_2} of H are $\exp\{2\pi i(k_1 x_1 + k_2 x_2)\}$ (k_1, k_2 are arbitrary integers) with (simple) eigenvalues $2\pi(k_1 + ak_2)$. Let $B_0 \equiv \{(x_1, x_2) \in S : 0 \leq x_1 \leq a_1 < 1, 0 \leq x_2 \leq a_2 < 1\}$ with a_1, a_2 (fixed) irrationals. Then $(f_{k_1, k_2}, \varphi_{B_0}) \neq 0$ for every f_{k_1, k_2}, and therefore $(B_0)_H[\rho] = 0$ iff $\rho = e$. Consequently, in order for a probability measure ν in $\mathfrak{L}$ to satisfy

$\int (B_0)_H \, d\nu = 0$, it is necessary and sufficient that $\nu(\{e\}) = 1$.

5

Proof of (B). Let (S, Σ, μ, H_0) be a (trivial) finite dynamical system with $H_0 = 0$. Given a partition $\pi : S = B_1 \cup \cdots \cup B_n$ ($B_i \in \Sigma$, $B_i \cap B_j = \phi$ for $i \neq j$, $\mu(B_i) > 0$), it is obvious [see (2.1)] that the fulfillment of the following relations:

$$(\varphi_{B_i} - \mu^{-1}(S)(e, \varphi_{B_i})e, \rho_\pi) = 0, \qquad i = 1, 2, \cdots, n, \tag{5.1}$$

implies $\rho_\pi = e$; and hence

$$(B_i)_{H_0}[\rho] = 0, \qquad i = 1, 2, \cdots, n, \tag{5.2}$$

iff $\rho_\pi = e$, since

$$(B_i)_{H_0}[\rho] = |(\varphi_{B_i} - \mu^{-1}(S)(e, \varphi_{B_i})e, \rho)|^2 = (B_i)_{H_0}[\rho_\pi] \tag{5.3}$$

[see (1.1), (1.3), and (1.5), and note that $H_0 = 0$, by hypothesis].

Consequently, in order for a probability measure $\bar{\nu}_\pi$ in T_π [(2.4)] to be such that

$$\int (\mathfrak{B}_i)_{H_0} \, d\bar{\nu}_\pi = 0, \qquad i = 1, 2, \cdots, n, \tag{5.4}$$

where $(\mathfrak{B}_i)_{H_0}(P_\pi \rho_\pi) \equiv (B_i)_{H_0}[\rho_\pi]$, it is necessary and sufficient that $\bar{\nu}_\pi(\{P_\pi e\}) = 1$.

Suppose now that A is a normal functional average in $\mathfrak{F}$ such that $A(B_{H_0}) = 0$ for every $B \in \Sigma$. As B_{H_0} is a continuous c. f. [see (5.3)], we have:

$$\int (\mathfrak{B})_{H_0} \, d\bar{\nu}_\pi = A(B_{H_0}) = 0, \tag{5.5}$$

where π is any partition containing B, and $\bar{\nu}_\pi$ the corresponding measure in T_π associated to A [see 2]. But B is arbitrary; hence (5.4) is fulfilled and therefore $\bar{\nu}_\pi(\{P_\pi e\}) = 1$; it is now a trivial matter to verify that the additive set function ν defined on $\bigcup_\pi \Sigma_\pi$ by $\{\nu_\pi\}$ [see (2.5)] is actually countably additive and that its extension to a probability measure is concentrated at e, Q.E.D.

NEW AVERAGING METHOD

6

Let (S, Σ, μ) be a measure space and let us define in T_π a probability Radon measure $\bar{\nu}_\pi$ as follows:

$$\bar{\nu}_\pi(C) \equiv \int_C \prod_1^n [\Gamma(\mu(B_i))]^{-1} \xi_i^{\mu(B_i)-1} \, d\sigma_\pi(\xi), \tag{6.1}$$

where

$$\int_{T_\pi} \Phi(\xi_1, \cdots, \xi_n) \, d\sigma_\pi(\xi) \equiv \int_0^1 d\xi_1 \int_0^{1-\xi_1} d\xi_2 \cdots \int_0^{1-\xi_1-\cdots-\xi_{n-2}} d\xi_{n-1}, \times \Phi(\xi_1, \cdots, \xi_{n-1}, 1 - \xi_1 - \cdots - \xi_{n-1}) \tag{6.2}$$

and C is an arbitrary Borel set in T_π.

By using the formula

$$\frac{1}{\Gamma(\alpha_1 + \alpha_2)} \xi^{\alpha_1+\alpha_2-1} = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \times \int_0^\xi (\xi - \eta)^{\alpha_1-1} \eta^{\alpha_2-1} \, d\eta \quad (\alpha_1, \alpha_2 > 0), \tag{6.3}$$

it is easy to show that

$$\bar{\nu}_\pi(C) = \bar{\nu}_{\pi'}(C'), \tag{6.4}$$

whenever $P_\pi^{-1}C = P_{\pi'}^{-1}C'$; therefore, the measures $\{\nu_\pi\}$ associated to $\{\bar{\nu}_\pi\}$ [(2.5)] are compatible. Let us denote by E^* the normal functional average defined by $\{\bar{\nu}_\pi\}$.

We maintain that every functional

$$K[\rho] \equiv (\rho, \bar{K}\rho), \tag{6.5}$$

where $\bar{K}$ is self-adjoint, nonnegative and in the trace class, belongs to $\bar{\mathfrak{F}}_{E^*}$.

In fact, if $\bar{K}$ satisfies these requirements, it is known[7] that there exists an orthonormal set $\{f_n\} \subset \mathfrak{K} \equiv L^2(S, \Sigma, \mu)$ such that $\bar{K}$ is an integral operator with kernel

$$\bar{K}(x, y) = \sum \lambda_n f_n^*(x) f_n(y), \qquad \lambda_n \geq 0, \tag{6.6}$$

and

$$\operatorname{Tr} \bar{K} = \sum \lambda_n \qquad (< \infty). \tag{6.7}$$

On the other hand, a simple application of the formula

$$\int_{T_\pi} \prod_1^n \xi_i^{\beta_i - 1} \, d\sigma_\pi(\xi) = \left[\Gamma\left(\sum_1^n \beta_i\right)\right]^{-1} \prod_1^n \Gamma(\beta_i), \qquad (\beta_i > 0) \tag{6.8}$$

yields

$$E^*(K_\pi) = \tfrac{1}{2}\left(K[e] + \sum_1^n \mu^{-1}(B_i) \times \int_{B_i \times B_i} \bar{K}(x, y) \, d\mu(x) \, d\mu(y)\right). \tag{6.9}$$

Writing $\psi_i \equiv \mu^{-1/2}(B_i)\varphi_{B_i}$, (clearly $||\psi_i|| = 1$), we

have

$$E^*(K_\pi) - \tfrac{1}{2}K[e] = \sum_i \mu^{-1}(B_i) \times \int_{B_i \times B_i} \tilde{K}(x, y)\, d\mu(x)\, d\mu(y) = \sum_1^n (\psi_i, \tilde{K}\psi_i). \tag{6.10}$$

If π' is finer than π, an easy computation shows that

$$\sum_1^n (\psi_i, \tilde{K}\psi_i) \leq \sum_1^{n'} (\psi_i', \tilde{K}\psi_i'), \tag{6.11}$$

where the member on the right, according to (6.10), equals $E^*(K_{\pi'}) - \frac{1}{2}K[e]$. Therefore, the directed set $\{E^*(K_\pi)\}$ is nondecreasing and bounded as well, since

$$E^*(K_\pi) \leq \tfrac{1}{2}(K[e] + \operatorname{Tr} \tilde{K}) \tag{6.12}$$

[recall that $\operatorname{Tr} \tilde{K} = \Sigma(g_n, \tilde{K}g_n)$ for any complete orthonormal system $\{g_n\} \subset \mathfrak{H}$].

Consequently $\lim_\pi E^*(K_\pi) (\equiv E^*(K))$ exists and is finite, Q.E.D.

Besides, we claim that

$$E^*(K) = \tfrac{1}{2}(K[e] + \operatorname{Tr} \tilde{K}). \tag{6.13}$$

In fact, it follows from (6.10) and (6.11) that

$$E^*(K_\pi) = \tfrac{1}{2}(K[e] + \operatorname{Tr} \tilde{K}_\pi), \tag{6.14}$$

where

$$\tilde{K}_\pi \equiv Q_\pi \tilde{K} Q_\pi,$$

Q_π being the projector

$$Q_\pi(x, y) \equiv \sum_1^n \psi_i^*(x)\psi_i(y).$$

But $Q_\pi \to_\pi 1$; since, on the other hand, given $\epsilon > 0$,

$$\operatorname{Tr} \tilde{K} - \sum_1^m (f_i, \tilde{K}f_i) < \epsilon \quad \text{for} \quad m \geq m(\epsilon)$$

[see (6.7)], it is obvious that

$$\operatorname{Tr} \tilde{K} - \sum_1^m (f_i, \tilde{K}_\pi f_i) < 2\epsilon \quad \text{for} \quad m \geq m(\epsilon), \quad \pi \geq \pi(\epsilon, m).$$

As

$$\sum_1^m (f_i, \tilde{K}_\pi f_i) \leq \operatorname{Tr} \tilde{K}_\pi \leq \operatorname{Tr} \tilde{K},$$

it follows

$$\operatorname{Tr} \tilde{K} - \operatorname{Tr} \tilde{K}_\pi < 2\epsilon \quad \text{for} \quad \pi \geq \pi(\epsilon),$$

and hence

$$\lim_\pi \operatorname{Tr} \tilde{K}_\pi = \operatorname{Tr} \tilde{K}, \qquad \text{Q.E.D.}$$

Remark 2. Suppose A is any operator in the trace class; the nonnegative self-adjoint operators $A_\pm \equiv \frac{1}{2}([A + A^*] \pm (A + A^*))$, and $A_\pm' \equiv \frac{1}{4}([i(A - A^*)] \pm i(A - A^*))$, where $[L] \equiv (L^*L)^{1/2}$, belong[7] also to the trace class and $A = A_+ - A_- - i(A_+' - A_-')$, $\operatorname{Tr} A = \operatorname{Tr} A_+ - \operatorname{Tr} A_- - i(\operatorname{Tr} A_+' - \operatorname{Tr} A_-')$. Therefore, if $\bar{A}[\rho] \equiv (\rho, A\rho)$, then $E^*(\bar{A}) = \frac{1}{2}(\bar{A}[e] + \operatorname{Tr} A)$ as it follows from (6.13) and the linearity of E^*.

Suppose finally that $\mathfrak{I}_t$ is a flow in (S, Σ, μ), and H the infinitesimal generator of U_t (see 1). The functional $B_H[\rho]$ is of the form $(\rho, H_B\rho)$, with H_B in the trace class [(1.5), (1.6), and (1.7)], and therefore,

$$E^*(B_H) = \tfrac{1}{2} \operatorname{Tr} H_B \tag{6.15}$$

(note that $B_H[e] = 0$). Let $\{B_n\}$ be a family of elements of Σ such that φ_{B_n}, $n = 1, 2, \cdots$, span a linear manifold dense in $\mathfrak{H}$ [its existence[1] is ensured by the separability of (S, Σ, μ)], and suppose that (S, Σ, μ, H) is not weakly mixing; then it is plain that $H_{B_n} \neq 0$ for some n, say for $n = n_0$, and hence

$$E^*((B_{n_0})_H) = \tfrac{1}{2} \operatorname{Tr} H_{B_{n_0}} \neq 0 \tag{6.16}$$

Q.E.D.

ACKNOWLEDGMENTS

The author wishes to thank Professor Friedrichs for the kind hospitality extended to him at the Courant Institute of Mathematical Sciences. Thanks are also due Professor R. Salinas, Professor Rejto, and Professor Scarfiello for many interesting discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

A Brownian-Motion Model for the Eigenvalues of a Random Matrix

FREEMAN J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received June 22, 1962)

A new type of Coulomb gas is defined, consisting of n point charges executing Brownian motions under the influence of their mutual electrostatic repulsions. It is proved that this gas gives an exact mathematical description of the behavior of the eigenvalues of an $(n \times n)$ Hermitian matrix, when the elements of the matrix execute independent Brownian motions without mutual interaction. By a suitable choice of initial conditions, the Brownian motion leads to an ensemble of random matrices which is a good statistical model for the Hamiltonian of a complex system possessing approximate conservation laws. The development with time of the Coulomb gas represents the statistical behavior of the eigenvalues of a complex system as the strength of conservation-destroying interactions is gradually increased. A "virial theorem" is proved for the Brownian-motion gas, and various properties of the stationary Coulomb gas are deduced as corollaries.

I. STATIONARY MATRIX ENSEMBLES

THIS introductory section recapitulates known facts about ensembles of matrices. Consider an $(n \times n)$ square matrix M whose elements are of the form

$$M_{ij} = \sum_{\alpha=0}^{\beta-1} M_{ij\alpha} e_\alpha, \qquad i, j = 1, \cdots, n. \tag{1}$$

The $M_{ij\alpha}$ are real coefficients, and the e_α ($\alpha = 0, 1, \cdots, \beta - 1$) are units of a certain algebra Φ. The three possible choices for Φ are: (i) $\Phi = R$, the algebra of real numbers, for which $\beta = 1$, $e_0 = 1$; (ii) $\Phi = C$, the complex numbers, for which $\beta = 2$, $e_0 = 1$, $e_1 = i$; (iii) $\Phi = Q$, the algebra of real quaternions, for which $\beta = 4$, $e_0 = 1$, $e_1 = i$, $e_2 = j$, $e_3 = k$. We assume that M is Hermitian, which means $M = M^D$, where M^D is the matrix dual to M with respect to Φ. The coefficients of M^D are derived from those of M by the relations

$$M^D_{ij0} = M_{ij0}, \tag{2}$$

$$M^D_{ij\alpha} = -M_{ji\alpha} \quad \text{for} \quad \alpha \neq 0. \tag{3}$$

The number of independent coefficients for a Hermitian M is

$$N = n + \tfrac{1}{2}n(n-1)\beta. \tag{4}$$

It is convenient to label these independent coefficients M_μ, where μ is a single index running from 1 to N and replacing the triple index $(ij\alpha)$. We write

$$d_\mu = d_{ij\alpha} = 2 - \delta_{ij}. \tag{5}$$

Then

$$\text{spur } M^2 = \sum_{ij\alpha} M^2_{ij\alpha} = \sum_\mu d_\mu M^2_\mu. \tag{6}$$

The eigenvalues of M are n real numbers $(x_1, \cdots, x_n)$.

The properties of random matrices have often been discussed[1] on the basis of the so-called Gaussian ensemble E_G. E_G consists of the set of all Hermitian matrices M satisfying Eq. (1), with a probability distribution assigned as follows. The probability for finding each coefficient within a given small interval $[M_\mu, M_\mu + dM_\mu]$ is

$$P(M_1, \cdots, M_N)\, dM_1 \cdots dM_N, \tag{7}$$

$$P(M_1, \cdots, M_N) = c \exp\left[-\beta\, (\text{spur } M^2)/2a^2\right], \tag{8}$$

where c and a are constants. In E_G, the M_μ are independent Gaussian random variables, each having mean value zero and variance

$$\langle M^2_\mu \rangle = [a^2/2\beta] g_\mu, \tag{9}$$

$$g_\mu = g_{ij\alpha} = 1 + \delta_{ij}. \tag{10}$$

This ensemble has been found useful in many investigations of the statistical behavior of random matrices of large order.

The main theorem concerning the ensemble E_G is the following.[2]

Theorem I. If M is chosen at random in E_G, the probability for finding an eigenvalue within each of the small intervals $[x_i, x_i + dx_i]$ is

$$F(x_1, \cdots, x_n)\, dx_1 \cdots dx_n, \tag{11}$$

$$F(x_1, \cdots, x_n) = C\Big[\prod_{i<j} |x_i - x_j|^\beta\Big] \times \exp\Big[-\beta\Big(\sum_i x_i^2\Big)/2a^2\Big], \tag{12}$$

[1] E. P. Wigner, *Proceedings of the 4th Canadian Mathematics Congress* (Toronto University Press, Toronto, Canada, 1959), p. 174; C. E. Porter and N. Rosenzweig, Suomalaisen Tiedeakat. Toimituksia A VI, No. 44 (1960); Phys. Rev. **120**, 1698 (1960).

[2] A proof of this theorem for $\beta = 1$ is given by Porter and Rosenzweig (reference 1). Their argument can easily be extended to the cases $\beta = 2, 4$.

where C is a normalization constant depending on a, n, and β.

When $\beta = 1$, the distribution function (12) is known as the Wishart distribution.[3]

For any value of β, Eq. (12) can be written

$$F = C \exp[-\beta W], \tag{13}$$

$$W(x_1, \cdots, x_n) = -\sum_{i<j} \ln|x_i - x_j| + \sum_i (x_i^2/2a^2). \tag{14}$$

The distribution (12) is identical with the probability distribution of the positions of n point charges, free to move on the infinite line $[-\infty < x < \infty]$ under the influence of forces derived from the potential energy (14) according to the laws of classical mechanics, in a state of thermodynamical equilibrium at a temperature given by

$$kT = \beta^{-1}. \tag{15}$$

This system of point charges in thermodynamical equilibrium is called the Coulomb gas model[4] corresponding to the ensemble E_a.

In all previous discussions of random matrices, the matrix ensembles and the Coulomb gas models were assumed to be stationary. No time-dependence of the distribution was allowed, and the velocities of the charges x_i played no role in the calculations of thermodynamic properties of the Coulomb gas.

II. THEORY OF BROWNIAN MOTION

The idea of the present paper is to generalize the notion of matrix ensemble in such a way that the Coulomb gas model acquires a meaning, not only as a static model in timeless thermodynamical equilibrium, but as a dynamical system which may be in an arbitrary nonequilibrium state changing with time. The word "time" in this paper will always refer to a fictitious time which is a property of the mathematical model and has nothing to do with real physical time.

When one tries to interpret the Coulomb gas as a dynamical system, one naturally thinks of it first as an ordinary conservative system in which the charges move as Newtonian particles and exchange energy with one another only through the electric forces arising from the potential (14). One has then to give a meaning to the velocity of each particle, and to regulate the behavior of the matrix M in such a way that the eigenvalues x_i possess the normal Newtonian property of inertia. There does not seem to be any reasonable way of doing this. The program of interpreting the Coulomb gas as a conservative Newtonian system has therefore failed.

After considerable and fruitless efforts to develop a Newtonian theory of ensembles, we discovered that the correct procedure is quite different and much simpler. The x_i should be interpreted as positions of particles in Brownian motion.[5] This means that the particles do not have well-defined velocities, nor do they possess inertia. Instead, they feel frictional forces resisting their motion. The gas is not a conservative system, since it is constantly exchanging energy with its surroundings through these frictional forces. The potential (14) still operates on the particles in the following way. The particle at x_i feels an external electric force

$$E(x_i) = -\frac{\partial W}{\partial x_i} = \sum_{i \neq j} \left[\frac{1}{x_i - x_j}\right] - \frac{x_i}{a^2}, \tag{16}$$

in addition to the local frictional force and the constantly fluctuating forces which give rise to the Brownian motion.

A precise description of the Brownian motion of the Coulomb gas is the following.[6] Let the positions of the particles be $[x_1, \cdots, x_n]$ at time t. Then the positions at time $(t + \delta t)$ are $[x_1 + \delta x_1, \cdots, x_n + \delta x_n]$, where the δx_i are random variables. To the first order in the small quantity δt, the variables δx_i are independent and have first and second moments given by

$$f\langle \delta x_i \rangle = E(x_i)\, \delta t, \tag{17}$$

$$f\langle (\delta x_i)^2 \rangle = 2kT\, \delta t, \tag{18}$$

all higher moments being zero to this order. The constant f is the friction coefficient which fixes the rate of diffusion, $E(x_i)$ is the external force (16), and kT is the temperature in energy units.

An equivalent description of the motion is obtained by considering the time-dependent probability density $F(x_1, \cdots, x_n; t)$ for finding the particles at the positions x_i at time t. In consequence of Eqs. (17) and (18), F satisfies the Smoluchowski equation

[3] J. Wishart, Biometrika 20, 32 (1928).
[4] F. J. Dyson, J. Math. Phys. 3, 140 (1962).
[5] A convenient summary of the theory of Brownian motion is contained in G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. 36, 823 (1930), and in M. C. Wang and G. E. Uhlenbeck, Revs. Modern Phys. 17, 323 (1945). These two papers are reprinted in *Noise and Stochastic Processes*, edited by N. Wax (Dover Publications, New York, 1954).
[6] See Wang and Uhlenbeck, reference 5, Sec. 8.

$$f\frac{\partial F}{\partial t} = \sum_i \left[kT\frac{\partial^2 F}{\partial x_i^2} - \frac{\partial}{\partial x_i}(E(x_i)F)\right]. \quad (19)$$

Equation (19) describes the development with time of the Coulomb gas. Starting from an arbitrary initial probability distribution F at $t = t_0$, a unique solution of Eq. (19) will exist for all $t > t_0$. Any such solution we will call a *time-dependent Coulomb gas model.*

It is easy to verify that Eq. (19) implies Eqs. (17) and (18), so that the descriptions of the motion by Eqs. (17) and (18) and by Eq. (19) are equivalent. Also, there exists a unique solution of Eq. (19) which is independent of time, and this time-independent solution is given by Eqs. (13) and (14). So the stationary Coulomb gas is a special case of the more general time-dependent model.

A Brownian-motion model can also be constructed for the matrix M of which the x_i are the eigenvalues. Suppose that the coefficients of the matrix have the values $[M_1, \cdots, M_N]$ at time t, and the values $[M_1 + \delta M_1, \cdots, M_N + \delta M_N]$ at time $(t + \delta t)$. A Brownian motion of M is defined by requiring that each δM_μ is a random variable with the moments

$$f\langle \delta M_\mu \rangle = -[M_\mu/a^2]\,\delta t, \quad (20)$$

$$f\langle (\delta M_\mu)^2 \rangle = g_\mu kT\,\delta t, \quad (21)$$

with g_μ defined by Eq. (10). This is a Brownian motion of the simplest type, the various components M_μ being completely uncoupled, and each being subject only to a fixed simple harmonic force. The Smoluchowski equation corresponding to Eqs. (20) and (21) is

$$f\frac{\partial P}{\partial t} = \sum_\mu \left[\frac{1}{2} g_\mu kT\frac{\partial^2 P}{\partial M_\mu^2} + \frac{1}{a^2}\frac{\partial}{\partial M_\mu}(M_\mu P)\right], \quad (22)$$

where $P(M_1, \cdots, M_N; t)$ is the time-dependent probability density of the M_μ. The solution of Eq. (22) corresponding to a given initial condition $M = M'$ at $t = 0$ is known explicitly.[7] It is

$$P(M; t) = c[1 - q^2]^{-N/2}$$

$$\times \exp\{-\text{spur}\,(M - qM')^2/[2a^2kT(1 - q^2)]\}, \quad (23)$$

$$q = \exp[-t/a^2 f]. \quad (24)$$

The solution shows that the Brownian process is invariant under symmetry-preserving unitary transformations of the matrix M; in fact the awkward-looking factor g_μ in Eq. (21) is put in just in order to assure this invariance. When $t \to \infty$, $q \to 0$, and the distribution (23) tends to the stationary form (8), which is the unique time-independent solution of Eq. (22).

We are now ready to state our main result.

Theorem II. When the matrix M executes a Brownian motion according to the simple harmonic law (20), (21), starting from any initial condition whatever, its eigenvalues $(x_1, \cdots, x_n)$ execute a Brownian motion obeying the equations of motion (17), (18), (19) of the time-dependent Coulomb gas.

Note that the temperature kT is still related to the basic algebra Φ by Eq. (15).

To prove the theorem, we need only show that Eqs. (17) and (18) follow from Eqs. (20) and (21). Suppose then that Eqs. (20) and (21) hold. We have seen that the process described by Eqs. (20) and (21) is independent of the representation of M. Therefore we may choose the representation so that M is diagonal at the time t. The instantaneous values of the $M_{ij\alpha}$ at time t are then

$$M_{jj0} = x_j, \qquad j = 1, \cdots, n, \quad (25)$$

with all other components zero. At the later time $(t + \delta t)$, the matrix $(M + \delta M)$ is no longer diagonal, and its eigenvalues $(x_j + \delta x_j)$ must be calculated by perturbation theory. We have, to second order in δM,

$$\delta x_j = \delta M_{jj0} + \sum_{i \neq j} \sum_{\alpha=0}^{\beta-1} \frac{(\delta M_{ij\alpha})^2}{(x_j - x_i)}. \quad (26)$$

Higher terms in the perturbation series will give no contribution to first order in δt. When we take the expectation value on each side of Eq. (26), using Eqs. (20), (21), (15), and (16), the result is Eq. (17). When we take the expectation value of $(\delta x_j)^2$, only the first term on the right of Eq. (26) contributes to order δt, and this term gives Eq. (18) by virtue of Eq. (21). The theorem is thus proved.

The proof of Theorem II incidentally provides a new proof of the old Theorem I. The new proof is simpler than the standard proof[2] of Theorem I, and is in some respects more illuminating. The new proof shows how the repulsive Coulomb potential (14), pushing apart each pair of eigenvalues, arises directly from the perturbation formula (26). It has long been known that perturbations generally split levels which are degenerate in an unperturbed system. We now see that this splitting effect of perturbations is quantitatively identical with the repulsive force of the Coulomb gas model.

Theorem II is a much stronger statement than Theorem I. It shows that the electric force (16) acting upon the eigenvalue x_j has a concrete meaning

[7] Except for a misprint, this is Eq. (15) of Uhlenbeck and Ornstein (reference 5).

for any matrix M whatever, not only for an ensemble of matrices in stationary thermal equilibrium. The force $E(x_i)$ is precisely proportional to the mean rate of drift of x_i which occurs when the matrix M is subjected to a random perturbation.

III. UNITARY BROWNIAN MOTION MODEL

In Sec. II we constructed a Brownian motion model for a random Hermitian matrix. We now construct a similar model for a random matrix U which is unitary and self-dual with respect to Φ. Such a matrix has elements of the form

$$U_{ij} = \sum_{\alpha=0}^{\beta-1} U_{ij\alpha} e_\alpha, \qquad i, j = 1, \cdots, n, \tag{27}$$

where the $U_{ij\alpha}$ are now complex coefficients but still satisfy the same symmetry condition $U = U^D$. The eigenvalues of U are n complex numbers $[\exp(i\theta_j)]$, $j = 1, \cdots, n$, all lying on the unit circle.

An isotropic and representation-independent Brownian motion of U is defined in the following way. Every unitary self-dual U can be represented in the form

$$U = VV^D, \tag{28}$$

with V unitary. A permissible small change in U is then given by

$$\delta U = iV\, \delta M\, V^D, \tag{29}$$

where δM is an infinitesimal Hermitian self-dual matrix. Suppose that U satisfies Eq. (28) at the time t. We assume that U moves by Brownian motion to the position $(U + \delta U)$ at time $(t + \delta t)$, where δU is given by Eq. (29), and the matrix δM has real coefficients δM_μ which are independent random variables with the moments

$$\langle \delta M_\mu \rangle = 0, \tag{30}$$

$$f\langle (\delta M_\mu)^2 \rangle = g_\mu kT\, \delta t. \tag{31}$$

This Brownian motion of U is a pure diffusion without any restoring force, since the harmonic force term which appeared in Eq. (20) has been omitted from Eq. (30). The rate of diffusion given by Eq. (21) remains the same as before.

The process of diffusion will spread the probability distribution of U more and more evenly as time goes on. The unique time-independent configuration for U is one in which the probability density is constant (in the sense of invariant group measures) over the whole space of unitary self-dual matrices. This uniform distribution of U is the stationary ensemble E_β defined in an earlier paper.[4]

Now we consider the effect of the Brownian motion of U on the eigenvalues $[\exp(i\theta_j)]$. We may use a representation in which U is diagonal at time t. The perturbation-theory formula analogous to Eq. (26) is then

$$\delta\theta_i = \delta M_{ii0} + \sum_{i \neq j} \sum_{\alpha=0}^{\beta-1} \{(\delta M_{ij\alpha})^2 \tfrac{1}{2} \cot [\tfrac{1}{2}(\theta_i - \theta_j)]\}. \tag{32}$$

Equations (30) and (31) then imply that the angles θ_i execute a Brownian motion with

$$f\langle \delta\theta_i \rangle = E(\theta_i)\, \delta t, \tag{33}$$

$$f\langle (\delta\theta_i)^2 \rangle = 2kT\, \delta t, \tag{34}$$

$$E(\theta_i) = \sum_{i \neq j} \tfrac{1}{2} \cot [\tfrac{1}{2}(\theta_i - \theta_j)]. \tag{35}$$

This force $E(\theta_i)$ is exactly the component, tangential to the circle, of the electric field produced at $\exp(i\theta_i)$ by unit charges at all the other points $\exp(i\theta_j)$ at which U has eigenvalues. Thus

$$E(\theta_i) = -(\partial W/\partial\theta_i), \tag{36}$$

$$W = -\sum_{i<j} \ln |\exp(i\theta_i) - \exp(i\theta_j)|. \tag{37}$$

So the eigenvalue angles θ_i behave like a gas of n unit charges on the unit circle, executing Brownian motions with repulsive Coulomb forces derived from the potential (37). Every free diffusion of the matrix U gives rise to a corresponding Brownian motion of the Coulomb gas formed by its eigenvalues on the unit circle.

In particular, the uniform probability distribution of U corresponds to the unique stationary probability density,

$$F(\theta_1, \cdots, \theta_n) = c \exp[-\beta W] = c \prod_{i<j} |\exp(i\theta_i) - \exp(i\theta_j)|^\beta, \tag{38}$$

for the eigenvalue angles. This is again a new and simple proof of an old result.[4]

IV. APPLICATION TO SYSTEMS WITH SEMI-CONSERVED QUANTUM NUMBERS

The stationary random-matrix models which have been studied in the past have always had an all-or-nothing character. That is to say, they represented situations in which a certain set of quantities (for example, total spin, charge, or isotopic spin) was exactly conserved, while no other integral of the motion existed even approximately. The total set

of states of the system could then be divided into subsets, each subset corresponding to a particular set of values for the conserved quantities. The energy levels of states belonging to different subsets would be completely uncorrelated, while the level distributions within each subset would be described by a separate random-matrix model.

The time-dependent Brownian motion models discussed in Sec. II provide a basis for a statistical theory of energy levels in systems possessing approximate conservation laws. Such systems occur in practice more frequently than those of the all-or-nothing type. For example, in complex atomic spectra it is usually the case that either an LS or a JJ coupling scheme is approximately valid, so that it is incorrect to treat all matrix elements of the Hamiltonian as having the same *a priori* probability distribution. We will now illustrate with a simple model how such situations can be quantitatively described in terms of the Brownian-motion picture.

Suppose that a system has an approximately conserved two-valued quantum number g, taking the values 1 and 2. Suppose that the manifolds of quantum states with $g = 1, 2$ have dimensions n_1, n_2, respectively. We write the Hamiltonian H in a representation with g diagonal. Then H splits into four blocks

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}. \tag{39}$$

For simplicity we suppose that we are in the case $\Phi = R$, $\beta = 1$ of Sec. I, so that the matrix H is real and symmetric. A reasonable statistical hypothesis for the elements of H is then the following. Each element is an independent random variable, having a Gaussian distribution with mean value zero and variance given by

$$\langle (H_{11,ij})^2 \rangle = \tfrac{1}{2}(1 + \delta_{ij})a^2\xi, \tag{40}$$

$$\langle (H_{22,ij})^2 \rangle = \tfrac{1}{2}(1 + \delta_{ij})a^2\eta, \tag{41}$$

$$\langle (H_{12,ij})^2 \rangle = \tfrac{1}{2}a^2\epsilon. \tag{42}$$

The parameter ϵ measures the ratio between the strengths of non-g-conserving and g-conserving interactions. It is convenient to choose ξ and η given by

$$\xi = 1 + [n_2/(1 + n_1)](1 - \epsilon), \tag{43}$$

$$\eta = 1 + [n_1/(1 + n_2)](1 - \epsilon). \tag{44}$$

Then Eqs. (40) to (42), with $\epsilon = 1$, describe the situation in which g is completely unconserved; the probability distribution of H is then identical with the stationary ensemble E_G in which all the elements of H are treated alike. When $\epsilon = 0$, Eqs. (40) to (42) describe two independent Gaussian ensembles E_{G1}, E_{G2}, referring to the states with $g = 1$ and $g = 2$ separately; this is the case of exact g conservation. The advantage of the choice (43) and (44) for ξ and η is that it makes the variance of any eigenvalue

$$\langle x_i^2 \rangle = \tfrac{1}{2}a^2(n_1 + n_2 + 1) \tag{45}$$

independent of ϵ and the same for E_{G1}, E_{G2}, E_G. Thus the over-all spread of the eigenvalue distributions does not vary as the parameter ϵ is changed.

We denote by $E(\epsilon)$ the ensemble of matrices H whose elements are Gaussian variables satisfying Eqs. (40) to (44). It is easy to verify by referring to Eqs. (23) and (24) that $E(\epsilon)$ is identical with the time-dependent Brownian motion model defined by Eqs. (20) and (21), provided we choose

$$kT = 1, \quad \epsilon = 1 - q^2 = 1 - \exp[-2t/a^2f]. \tag{46}$$

The initial condition at $t = \epsilon = 0$ is

$$E(0) = E_{G1} \times E_{G2}, \tag{47}$$

while in the limit $t \to \infty$ we have

$$E(\epsilon) \to E(1) = E_G. \tag{48}$$

The distribution of eigenvalues for a matrix in $E(\epsilon)$ is determined by Theorem II. We have thus proved

Theorem III. Let H be a matrix chosen at random in the ensemble $E(\epsilon)$. Its eigenvalues $(x_1, \cdots, x_n)$ are then distributed according to the time-dependent Coulomb gas model defined by Eq. (19), taken at the time

$$t = -\tfrac{1}{2}a^2f[\ln(1 - \epsilon)], \tag{49}$$

the initial condition at $t = 0$ being a superposition of two uncorrelated stationary Coulomb gases containing n_1 and n_2 charges, respectively.

The statistical distribution of eigenvalues of a system with an approximate conservation law is thus determined in principle. The eigenvalue distribution is the solution of the Smoluchowski equation (19) with the initial condition that F should be at $t = 0$ a product of two Wishart distributions. This solution has a very simple physical interpretation. The situation with exact g conservation is represented by two superposed noninteracting stationary Coulomb gases. At the instant $t = 0$ a repulsive Coulomb interaction between the charges in one gas and those in the other is suddenly switched on. The resulting nonstationary Coulomb gas is

allowed to adjust itself to the sudden change in the forces by Brownian motion for a time given by Eq. (49). At the end of this time the gas will represent the eigenvalue distribution for the situation with partial g conservation.

It is an interesting and deep mathematical problem to describe accurately the approach of the Coulomb gas to equilibrium as t or ϵ increases. Since we were careful to choose the parameters ξ and η to make the over-all shape of the charge distribution independent of t, the approach to equilibrium involves only local adjustments. We make the conjecture that the approach to equilibrium proceeds exponentially with a time scale which is of the order of

$$\tau = [fS/E], \tag{50}$$

where S is an average level-spacing and E an average Coulomb force between nearest-neighbor charges. This τ is the time taken for a single level to respond to the changed interaction between itself and its neighbors. In the ensemble E_G, Eq. (45) gives

$$S \cong a(n_1 + n_2)^{-1/2}, \qquad E \cong S^{-1}, \tag{51}$$

and therefore

$$\tau \cong [fa^2/(n_1 + n_2)]. \tag{52}$$

Comparing this with Eq. (49), we see that the eigenvalues already approach their final distribution when ϵ is of the order

$$\epsilon \sim (n_1 + n_2)^{-1}. \tag{53}$$

So for matrices whose order $n = n_1 + n_2$ is large, the g-violating matrix elements $H_{12,ij}$ need only to be of the order of $(n^{-1/2})$ times the g-conserving elements in order to bring the eigenvalue distribution close to the state of complete g-nonconservation. In random matrices of large order, a conservation law must be almost exact in order to produce a noticeable effect on eigenvalue distributions.

A different type of system with partially conserved quantum numbers can arise in the following way. Suppose that there are two almost-conserved quantities g, g' which do not commute with each other. Suppose that g takes the values 1, 2, while g' has matrix elements linking states with $g = 1$ to states with $g = 2$. A familiar example of such a situation occurs when g and g' are two different components of angular momentum. The Hamiltonian H again splits into four blocks as in Eq. (39). However, the states of the system must now become degenerate in pairs in the limit of exact g and g' conservation. A reasonable statistical hypothesis for H is that its elements are independent Gaussian variables satisfying the Brownian motion conditions (20) and (21), but now starting from the initial condition

$$H = \begin{bmatrix} H' & 0 \\ 0 & H' \end{bmatrix} \tag{54}$$

at $t = 0$, with H' distributed according to the Gaussian ensemble E_{G1}. In this case we must have $n_1 = n_2 = \frac{1}{2}n$. The distribution of eigenvalues in the system with partial g and g' conservation will be obtained from the time-dependent Coulomb gas as before, but now the initial condition has the positions of the charges coinciding in pairs. Instead of two noninteracting Coulomb gases, we have in the initial state a single Coulomb gas of $\frac{1}{2}n$ charges, each charge being suddenly replaced by two independently moving charges at the instant when the Brownian motion begins. The approach to equilibrium should again occur in a time of the order of Eq. (52).

Many other statistical ensembles, describing random matrices with approximate conservation laws, could be constructed from the time-dependent Brownian motion model with suitable initial conditions. In some cases it will be appropriate to use ensembles with $\beta = 2$ or 4 instead of $\beta = 1$, depending on the type of group representation which the symmetry of the problem requires.[8] We will not attempt here any systematic discussion of the possible alternatives. The two simple examples which we described above show well enough the general principles to be followed.

V. PROPERTIES OF THE BROWNIAN MOTION MODEL

In studying the approach to equilibrium of the time-dependent Coulomb gas defined by Eqs. (16) to (18), we have derived a few general properties of the gas which may be worthy of record. Let $G = G(x_1, \cdots, x_n)$ be any function of the positions of the charges, not depending explicitly on time. Then $\langle G \rangle$, the ensemble average of G, varies with time according to the equation

$$f\frac{d}{dt}\langle G\rangle = -\sum_i \left\langle \frac{\partial W}{\partial x_i}\frac{\partial G}{\partial x_i}\right\rangle + kT\sum_i \left\langle \frac{\partial^2 G}{\partial x_i^2}\right\rangle, \tag{55}$$

with W given by Eq. (14).

Take in this equation

$$R = \sum_k \langle x_k^2 \rangle \tag{56}$$

for $\langle G \rangle$. The equation becomes

$$\tfrac{1}{2}a^2 f\dot{R} = R_\infty - R, \tag{57}$$

[8] F. J. Dyson, J. Math. Phys. 3, 1199 (1962), following paper.

with

$$R_\infty = a^2[\tfrac{1}{2}(n^2 - n) + nkT]. \tag{58}$$

Equation (57) has the solution

$$R = R_0 q^2 + R_\infty(1 - q^2), \tag{59}$$

where q is given by Eq. (24) and R_0 is the value of R at $t = 0$. Equation (59) shows rigorously that at least the ensemble average R approaches its equilibrium value R_∞ with exponential speed as $t \to \infty$.

Next take $G = W$ in Eq. (55). After some algebra we find

$$f(d\langle W\rangle/dt) = (kT - 1) \sum_{i\neq j} \langle (x_j - x_i)^{-2}\rangle + (n^2 - n + nkT)a^{-2} - \sum_j \langle x_j^2\rangle a^{-4}. \tag{60}$$

For the stationary Coulomb gas at temperature T, the left side of Eq. (60) vanishes, and Eq. (58) may be used on the right. We thus find a "virial theorem" for the stationary gas,

$$\sum_{i\neq j} \langle (x_j - x_i)^{-2}\rangle = \frac{n^2 - n}{2a^2(1 - kT)}. \tag{61}$$

The probability density of eigenvalues becomes proportional to $|x_j - x_i|^\beta$ when two eigenvalues (x_i, x_j) come close together. The ensemble average of $(x_j - x_i)^{-2}$ is therefore defined only for $\beta > 1$, and Eqs. (60), (61) hold only for $kT < 1$.

We are especially interested in the case $kT = 1$, which requires a passage to the limit in Eq. (60). As $kT \to 1$, we have for any fixed value of Δ

$$\lim (kT - 1) \int_{-\Delta}^{\Delta} |y|^{\beta-2}\, dy = \lim (kT - 1)(\beta - 1)^{-1} 2\Delta^{\beta-1} = -2. \tag{62}$$

We obtain the correct limit in Eq. (60) if we replace

$$(kT - 1)(x_j - x_i)^{-2}$$

by

$$-2\,|x_j - x_i|^{-1}\,\delta(x_j - x_i), \tag{63}$$

which has a well-defined meaning as an ensemble average when $kT = 1$, since the probability density then contains a factor $|x_j - x_i|$. Equation (60) thus becomes, in the limit,

$$f(d\langle W\rangle/dt) = -2 \sum_{i\neq j} \langle |x_j - x_i|^{-1}\,\delta(x_j - x_i)\rangle + n^2a^{-2} - \sum_j \langle x_j^2\rangle a^{-4}, \qquad kT = 1. \tag{64}$$

The corresponding "virial theorem" is

$$\sum_{i\neq j} \langle |x_j - x_i|^{-1}\,\delta(x_j - x_i)\rangle = \frac{n^2 - n}{4a^2}, \qquad kT = 1, \tag{65}$$

for the stationary gas.

Equation (64) suggests very forcibly the following picture of the approach of the gas to equilibrium. The first term on the right is a "collision term" measuring the frequency with which two charges come into coincidence. This term is mainly sensitive to the local (microscopic) configuration of the gas particles. By means of this term the gas will come into local thermodynamic equilibrium in a microscopic time scale

$$t \sim fa^2n^{-1}. \tag{66}$$

After local equilibrium is established, the gas will still not be in a stationary state, because the third term on the right of Eq. (64) will not in general have its stationary value. The gas must adjust itself by macroscopic motion on the time scale

$$t \sim fa^2, \tag{67}$$

until the over-all charge distribution reaches its stationary shape.

We chose the parameters ξ and η in Eqs. (40) to (42) so that the Coulomb gas representing the eigenvalues of a system with semiconserved quantum numbers should have the stationary macroscopic shape from the beginning. In this case the entire process of reaching equilibrium should occur with the microscopic time scale (52).

Of course the picture of the gas coming into equilibrium in two well-separated stages, with microscopic and macroscopic time scales, is only suggested by Eq. (64) with the help of physical intuition. A rigorous proof that this picture is accurate would require a much deeper mathematical analysis.

The equations of this section all have analogs for the time-dependent Coulomb gas on the unit circle, with the Brownian motions defined by Eqs. (33) to (35). The potential energy W is now given by Eq. (37). The macroscopic mass center of the gas is

$$R = n^{-1} \sum_k \langle \exp(i\theta_k)\rangle, \tag{68}$$

and the analog of Eq. (57) is

$$f\dot{R} = -[\tfrac{1}{2}(n - 1) + kT]R. \tag{69}$$

The analogs to Eqs. (60) and (64) are[9]

[9] Here use is made of the identity
$$\cot a \cot b + \cot b \cot c + \cot c \cot a = 1,$$
which holds when $a + b + c = 0$.

$$f(d\langle W\rangle/dt) = \tfrac{1}{12}(n^3 - n) + (kT - 1)\sum_{i\neq j} \times \langle|\exp(i\theta_j) - \exp(i\theta_i)|^{-2}\rangle, \quad kT < 1, \tag{70}$$

$$f(d\langle W\rangle/dt) = \tfrac{1}{12}(n^3 - n) - 2\sum_{i\neq j} \times \langle|\theta_j - \theta_i|^{-1}\,\delta(\theta_j - \theta_i)\rangle, \quad kT = 1. \tag{71}$$

These give virial theorems analogous to Eqs. (61) and (65). We state the results only for the limiting case $n \to \infty$ which is most important in practice.

Let x_j, $(-\infty < j < +\infty)$, be the positions of charges in an infinite Coulomb gas in thermodynamic equilibrium at temperature T. Let D be the mean spacing between nearest neighbors. The virial theorems are

$$\sum_{j\neq 0}\langle(x_j - x_0)^{-2}\rangle = \frac{\pi^2}{3D^2(1 - kT)}, \quad kT < 1, \tag{72}$$

$$\sum_{j\neq 0}\langle|x_j - x_0|^{-1}\,\delta(x_j - x_0)\rangle = \frac{\pi^2}{6D^2}, \quad kT = 1. \tag{73}$$

Equation (73) is a known result, giving the slope of the level-spacing distribution function at zero spacing.[10] The above derivation of it seems to be the simplest yet found.

ACKNOWLEDGMENT

The author is indebted to Dr. A. Lenard, whose lecture on the statistical behavior of electric fields in a different kind of Coulomb gas model[11] gave the initial impetus for this work.

[10] M. L. Mehta, Nuclear Phys. 18, 395 (1960).

[11] A. Lenard, J. Math. Phys. 2, 682 (1961); S. F. Edwards and A. Lenard, *ibid.* 3, 778 (1962). The Edwards-Lenard paper describes a Brownian motion model which has some similarity to ours; however, their model differs fundamentally from ours in identifying the fictitious time variable t with the space coordinate x.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

The Threefold Way. Algebraic Structure of Symmetry Groups and Ensembles in Quantum Mechanics

FREEMAN J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received June 22, 1962)

Using mathematical tools developed by Hermann Weyl, the Wigner classification of group-representations and co-representations is clarified and extended. The three types of representation, and the three types of co-representation, are shown to be directly related to the three types of division algebra with real coefficients, namely, the real numbers, complex numbers, and quaternions. The author's theory of matrix ensembles, in which again three possible types were found, is shown to be in exact correspondence with the Wigner classification of co-representations. In particular, it is proved that the most general kind of matrix ensemble, defined with a symmetry group which may be completely arbitrary, reduces to a direct product of independent irreducible ensembles each of which belongs to one of the three known types.

I. INTRODUCTION

THE purpose of this paper is to bring together and unify three trends of thought which have grown up independently during the last thirty years. These are (i) the classification by Wigner[1] of representations of groups which include time-inversion, (ii) Weyl's general theory of matric algebras and their commutator algebras,[2] and (iii) the study of ensembles of random matrices, begun by Wigner[3] and continued by various other physicists.[4] It will be shown that these three theories are all variations upon a single mathematical theme. It is not surprising that the three theories should turn out to be closely related, since they all took their origin from the work of the great algebraists Frobenius and Schur at the beginning of the twentieth century.[5]

Our way is threefold in another and deeper sense. In each of the three theories which we aim to unify, there appears a triple alternative, a choice between three mutually exclusive possibilities. (i) The irreducible representations of a group by unitary matrices fall into three classes, which are called potentially real, complex, and pseudoreal.[6] Another, and quite independent, threefold choice exists for representations of a group by unitary and antiunitary matrices. Wigner[7] calls such representations co-representations, and he classifies them into types I, II, and III. (ii) The classical groups studied by Weyl are of three types, namely orthogonal, unitary, and symplectic. (iii) The present author[4] found three distinct kinds of ensembles of random matrices, to which he attached the same three names as are given to the classical groups. In the previous discussion of matrix ensembles,[4] the question whether all irreducible ensembles belong to one of these three types was not raised. This question will here be answered in the affirmative.

The recurrence of the threefold choice in all these contexts gave the first hint that a unified mathematical treatment of group representations, commutator algebras, and ensembles should be possible. It was Bargmann who pointed out to the author[8] that the root of the matter is to be found in the classical theorem of Frobenius.[9]

Frobenius' Theorem. Over the real number field

[1] E. P. Wigner, Nachr. Akad. Wiss. Göttingen, Math. physik. Kl., 546 (1932). See also, E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959), English edition, Chaps 24 and 26.

[2] H. Weyl, *The Classical Groups, Their Invariants and Representations* (Princeton University Press, Princeton, New Jersey, 1939). Chapter 3 of this book contains the essential theorems on which all of our arguments hang. For Weyl's treatment of semilinear representations, see Duke Math. J. 3, 200 (1937).

[3] E. P. Wigner, Ann. Math. 53, 36 (1951); 62, 548 (1955); 65, 203 (1957); 67, 325 (1958).

[4] F. J. Dyson, J. Math. Phys. 3, 140, 157, and 166 (1962). This series of three papers includes references to earlier work by others in the same field. Paper IV in the series is being written in collaboration with Dr. M. L. Mehta and will be published later. The present paper should logically be considered to be number zero in the series, since it provides an improved mathematical and logical foundation for the rest of the series. Since Roman numerals contain no symbol for zero, we preferred to publish the present paper under a separate title.

[5] A sketch of the historical development is to be found in the section headed "Remembrance of Things Past" in Weyl's book (reference 2), p. 27.

[6] Chapter 24 of Wigner's book (reference 1). This classification was discovered by A. Loewy, Trans. Am. Math. Soc. 4, 171 (1903). See also G. Frobenius and I. Schur, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl. 186 (1906).

[7] Chapter 26 of Wigner's book (reference 1).

[8] V. Bargmann (private communication).

[9] G. Frobenius, J. reine u. angew. Math. 84, 59 (1878); L. E. Dickson, *Linear Algebras* (Cambridge University Press, New York, 1914), p. 10.

there exist precisely three associative[10] *division algebras, namely the real numbers, the complex numbers, and the real quaternions.*

Once this is understood, the further development of the theory is extremely simple. All that is necessary is to apply the general theorems of Weyl[2] to the special case in which the ground field of the matric algebras is the field of real numbers.

Probably all these connections would have been clarified long ago, if quantum physicists had not been hampered by a prejudice in favor of complex and against real numbers. It has been generally believed that only the complex numbers could legitimately be used as the ground field in discussing quantum-mechanical operators. Over the complex field, Frobenius' theorem is of course not valid; the only division algebra over the complex field is formed by the complex numbers themselves. However, Frobenius' theorem is relevant precisely because the appropriate ground field for much of quantum mechanics is real rather than complex. Specifically, as soon as anti-unitary operators such as time inversion are included, it is simpler and more natural to work with a real ground field than to follow Weyl[2] in studying semilinear operators over the complex field.

Physicists have known for a long time that in practice, when invariance under time-inversion is in question, complex phases are no longer arbitrary and undetermined coefficients may be taken to be real. Physicists are, in fact, like M. Jourdain talking prose, using the real numbers for their ground field without knowing it. One purpose of this paper is to make the use of the real ground field in quantum mechanics official and undisguised.[11] No change in the physical content of the theory is thereby implied. Only it may be easier for students to understand what they are doing if the mathematical formalism is brought into closer correspondence with physical practice.

A final by-product of the work described in this paper is that it defines an area of quantum mechanics within which quaternions play a natural and essential role. Several attempts have been made in the past[12] to construct a radically new version of quantum mechanics in which complex numbers are from the beginning replaced by quaternions. Our analysis has nothing to do with these attempts. Proceeding in a modest and conservative spirit, we merely show that quaternions form the appropriate algebraic basis for a description of nature whenever we have to deal either with pseudoreal group representations or with co-representations of Wigner's type II. The context in which quaternions arose historically, in a study of the three-dimensional rotation group, can now be seen to be an extremely special case of this general principle. Every group which admits pseudoreal representations equally admits a natural description in terms of real quaternions.

II. GROUP ALGEBRA AND COMMUTATOR ALGEBRA

The starting point of our analysis is a group G which is supposed to be a symmetry-group for some quantum-mechanical system. For example, G could be a rotation group, or an isotopic-spin group, or a time-inversion group, or all of these in combination. The quantum-mechanical states belong to a linear vector space H_e of finite dimension n over the field C of complex numbers. An element g of G is represented in H_e by an operator $\Lambda(g)$ which is either unitary or antiunitary. Physically, the antiunitary $\Lambda(g)$ will correspond to operations g which involve time-inversion. We make the convention that the letter g may denote any element of G, the letter u denotes an element for which $\Lambda(u)$ is unitary, and the letter a denotes an element for which $\Lambda(a)$ is antiunitary. The set of u forms a subgroup G_1 of G. We assume that G contains some antiunitary elements a. Then G_1 is an invariant subgroup of G with index 2. The a form a set G_2 which is the unique co-set of G_1 in G.

The $\Lambda(a)$ are not matrices over the field of complex numbers. The notion of group representation can be enlarged, following Weyl[2] and Wigner,[7] so as to include such semilinear operations. However, we find it simpler and more fruitful to represent

[10] The restriction to associative algebras is forced by the fact that the rule of matrix multiplication is associative. In all applications of group theory to quantum mechanics we identify the operation of multiplication with ordinary matrix multiplication. It is well-known that a fourth division algebra over the real number field exists, namely the algebra of octonions, if multiplication is allowed to be nonassociative. It is interesting to speculate upon possible physical interpretations of the octonion algebra [see A. Pais, Phys. Rev. Letters 7, 291, 1961]. We have tried, and failed, to find a natural way to fit octonions into the mathematical framework developed in this paper.

[11] The general formalism of quantum mechanics over a real ground field has been worked out by E. C. G. Stueckelberg, Helv. Phys. Acta 32, 254 (1959); 33, 727 (1960). Two further papers by Stueckelberg and collaborators have been circulated as preprints and will appear in Helv. Phys. Acta. These papers have many points of contact with the present work. For a brief summary of Stueckelberg's conclusions, see also the paper of Finkelstein *et al.* (reference 12).

[12] G. Birkhoff and J. von Neumann, Ann. Math. 37, 823 (1936). E. J. Schremp, Phys. Rev. 99, 1603 (1955); 113, 936 (1959). D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. 3, 207 (1962).

the $\Lambda(a)$ by true matrices over the field R of real numbers. We define the correspondence

$$\Lambda(g) \leftrightarrow M(g) \tag{1}$$

in the following way. $M(g)$ is a $[2n \times 2n]$ matrix with real elements. Each (2×2) block in $M(u)$ is derived from a single element of the $[n \times n]$ complex matrix $\Lambda(u)$ by the replacement

$$\alpha + i\beta \leftrightarrow \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}. \tag{2}$$

Each $\Lambda(a)$ is of the form

$$\Lambda(a) = U(a)j, \tag{3}$$

where $U(a)$ is unitary and j is the operation of complex conjugation. The $M(a)$ are defined by making the substitution (2) in $U(a)$ together with the replacement

$$j = I_n \times \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4}$$

The space H_R in which the matrices $M(g)$ operate is a real $2n$-dimensional vector space. Each vector in H_R is composed of the real and imaginary parts of the components of the corresponding vector in H_c. It is convenient to consider the symbol

$$i = I_n \times \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{5}$$

also as a matrix operating in H_R.

The $M(g)$ now form a true $2n$-dimensional representation of the group G over the field R. The distinction between unitary and antiunitary elements of G is provided by the commutation rules

$$M(u)i = iM(u), \tag{6}$$

$$M(a)i = -iM(a). \tag{7}$$

All the $M(g)$ are orthogonal matrices.

A matric algebra over R is defined as a set of matrices which is closed under the three operations of addition, matrix multiplication, and multiplication by scalar coefficients in R. Three such algebras will now be introduced:

A generated by the $M(u)$, (8)

B generated by the $M(u)$ and i, (9)

D generated by the $M(u)$, $M(a)$, and i. (10)

The commutator algebra of a given algebra K is defined as the set of matrices which commute with all matrices in K. The commutator algebra is itself a matric algebra over R. In particular we define

$$X = \text{commutator algebra of } A, \tag{11}$$

$$Y = \text{commutator algebra of } B, \tag{12}$$

$$Z = \text{commutator algebra of } D. \tag{13}$$

The inclusion relations

$$A \subset B \subset D \tag{14}$$

immediately imply

$$X \supset Y \supset Z. \tag{15}$$

The algebra A is given the name "group algebra of G_1 over R." In an obvious sense, B is identical with the group algebra of G_1 over C. The algebra D is not a group algebra over C in the ordinary sense, but it may be considered to be the group algebra of G over C. However, it is important that we have defined each of A, B, D as algebras with coefficients in R.

We next introduce some convenient notations, following Weyl.[2] If K is any algebra and m a positive integer, we denote by mK the algebra of matrices consisting of m identical blocks,

$$\begin{bmatrix} M & O & O & \cdots \\ O & M & O & \\ \vdots & & \ddots & \\ O & \cdots & & M \end{bmatrix}, \tag{16}$$

with M in K. Symbolically, we may write this as an outer product,

$$mK = I_m \times K. \tag{17}$$

We denote by $[K]_m$ the algebra of all matrices consisting of m^2 blocks,

$$\begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1m} \\ M_{21} & M_{22} & \cdots & M_{2m} \\ \vdots & \vdots & & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mm} \end{bmatrix}, \tag{18}$$

with each M_{ij} independently a matrix in K. In particular, when $K = R$ is the algebra of scalars, R_m is the algebra of all real matrices of degree m.

Two algebras A, A' are said to be equivalent ($A \sim A'$) if there exists a fixed nonsingular matrix N such that every matrix M of A is related to the corresponding M' in A' by

$$M = NM'N^{-1}. \tag{19}$$

If A is any algebra, the inverse algebra $\bar{A}$ is obtained

from A by inverting the order of factors in all products, thus

$$\tilde{M} = \tilde{M}_1\tilde{M}_2 \tag{20}$$

if and only if

$$M = M_2M_1. \tag{21}$$

Finally, a division algebra is defined to be an algebra in which every nonzero element M has a reciprocal M^{-1}.

With these notations and definitions, we are in a position to state the main theorem of Weyl.[13]

Weyl's Theorem. Let K be any group algebra over R, and L its commutator algebra. Then K and L are simultaneously equivalent to the canonical forms

$$K \sim \sum_j s_j[E_j]_{t_j}, \qquad L \sim \sum_j t_j[\tilde{E}_j]_{s_j}. \tag{22}$$

The summations here represent direct sums over diagonal blocks of matrices. Each value of j corresponds to one inequivalent irreducible representation of the group Γ which generates K over R. For each j, E_j is a division algebra, and s_j, t_j are positive integers. The matrix block corresponding to index j has degree

$$d_j = s_jt_je_j, \tag{23}$$

where e_j is the degree of E_j.

The following remarks may be made concerning this theorem.

Remark 1. The relation between the algebras K and L is symmetrical. Thus K is also the commutator algebra of L.

Remark 2. When the sums (22) reduce to a single term, the algebras K and L are called simple. In this case the suffixes j may be dropped.

Remark 3. When K is generated by an irreducible representation of Γ, K is simple and the integer s is equal to unity. In this case

$$K \sim E_t, \qquad L \sim t\tilde{E}. \tag{24}$$

Remark 4. By Frobenius' theorem (see Sec. I), the possible division algebras over R are three in number, and are denoted by R, C, and Q. R has degree 1, and is generated by the scalar $I_1 = 1$. C has degree 2 and is generated by

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{25}$$

Note that this e_2 is not necessarily identical with the original imaginary unit i defined by Eq. (5). Since C is commutative, $C = \tilde{C}$. The quaternion division algebra Q has degree 4 and is generated by

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\tau_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix},$$

$$\tau_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{26}$$

$$\tau_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

The inverse algebra $\tilde{Q}$ is then generated by

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\tau_1' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\tau_2' = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \tag{27}$$

$$\tau_3' = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

This particular representation of Q and $\tilde{Q}$ by real

[13] This is theorem (3.5B) on p. 95 of Weyl's book (reference 2), combined with the theorem that every group ring is fully reducible (p. 101 of the same book).

matrices is called the regular representation. It has the property that all matrices in Q commute with all matrices in $\tilde{Q}$. Thus Q and $\tilde{Q}$ are commutator algebras of each other, as required by the theorem.

Remark 5. When K is a simple algebra, the division algebra E is uniquely fixed and must be either R, C, or Q. In these three cases we say that the representation of Γ by K is of type R, type C, or type Q, respectively.

Remark 6. We shall apply Weyl's theorem to the algebras A, B, and D defined at the beginning of this section. In the case of A, the group Γ is identical with G_1. In the case of B, the group Γ is the direct product of G_1 with the Abelian group Γ_4 generated by (I, i). In the case of D, the group Γ is the product of G with Γ_4, the commutation rules between G and Γ_4 being given by Eqs. (6) and (7). Each of A, B, D is thus a group algebra over R in the ordinary sense, although only B is a group algebra over C.

The following lemma[14] is important in determining the structural relations between the algebras B and D.

Lemma. Let $M_1(g)$, $M_2(g)$ be two inequivalent irreducible components of the algebra D. Then the subalgebras $M_1(u)$, $M_2(u)$ in B are inequivalent, and no irreducible component of $M_1(u)$ can be equivalent to any irreducible component of $M_2(u)$.

To prove the lemma, we assume that $M_1(g)$ and $M_2(g)$ are inequivalent and that $M_1(u)$ and $M_2(u)$ have two equivalent irreducible components. There then exists a matrix P in the algebra Y, linking the two inequivalent blocks M_1 and M_2 of the algebra D, but commuting with the algebra B. This P satisfies

$$Pi = iP, \qquad PM_1(a_1a_2^{-1}) = M_2(a_1a_2^{-1})P, \tag{28}$$

for any two antiunitary elements a_1, a_2 in G. Therefore,

$$[M_2(a_1)]^{-1}PM_1(a_1) = [M_2(a_2)]^{-1}PM_1(a_2) = W, \tag{29}$$

where W is a matrix independent of a_1, a_2. Hence

$$PM_1(a) = M_2(a)W \tag{30}$$

for all a in G. Since Eq. (30) also holds with a replaced by a^{-1}, we have

$$M_2(a)P = WM_1(a). \tag{31}$$

Therefore,

$$(P + W)M_1(a) = M_2(a)(P + W) \tag{32}$$

for all a in G, and this implies

$$(P + W)M_1(g) = M_2(g)(P + W). \tag{33}$$

Since $M_1(g)$ and $M_2(g)$ are supposed irreducible and inequivalent, Schur's lemma[15] now implies

$$P + W = 0. \tag{34}$$

But then Eq. (30) becomes

$$PM_1(a) = -M_2(a)P. \tag{35}$$

Equations (28) and (35) together give

$$iPM_1(g) = M_2(g)iP \tag{36}$$

for all g in G, and therefore by Schur's lemma again

$$iP = 0.$$

Thus the operator P cannot exist, and the lemma is proved.

Remark 7. An equivalent statement of this lemma is as follows. Let the algebras Y and Z be written in the canonical form of Weyl's theorem as direct sums of diagonal blocks,

$$Y = \sum_k Y_k, \qquad Z = \sum_j Z_j, \tag{37}$$

where the Z_j are inequivalent simple algebras and likewise the Y_k. The lemma states that each Y_k is confined to a single block containing precisely one Z_j. This means that the structural relation between Y and Z is completely determined by considering the separate blocks Z_j.

III. WIGNER'S CLASSIFICATION OF IRREDUCIBLE REPRESENTATIONS

In this section we shall establish the connection between Weyl's theory of group algebras (Sec. II) and the classification of group representations by Wigner.[1]

A few preliminary observations must first be made. The equivalence relations in Eq. (22) refer to a transformation to canonical forms by a similarity relation (19) in which N may be an arbitrary real nonsingular matrix. According to a standard argument,[16] since the algebra K is generated by an orthogonal group representation, the transformation matrix N may be chosen to be orthogonal. Next we show that, when the algebra K is one of the trio A, B, or D, the matrix N may be chosen so as to commute with i. The operator i belongs to B and D,

[14] This lemma could probably be deduced as a special case from the general theorems of A. H. Clifford, Ann. Math. 38, 533 (1937), concerning the connections between representations of groups and subgroups. However, it seemed simpler to give a direct and elementary proof of the lemma without appeal to Clifford's work.

[15] See Wigner (reference 1), p. 75, Theorem 2.

[16] See Wigner (reference 1), p. 78.

and to the commutator algebras X and Y of A and B. So in each of the three cases, i belongs either to K or to L. When the transition to canonical forms is made, i is transformed into some matrix i' which has nonzero elements only within the blocks where the canonical form of K or L exists. The transformed i' still satisfies

$$[i']^2 = -I. \tag{38}$$

It is therefore possible to transform i' back into the standard form i by a real orthogonal transformation working within each diagonal block separately. As a result, we have an orthogonal matrix N which transforms K and L into the canonical forms (22) and transforms i into i. This N then commutes with i.

When N is chosen to be orthogonal and to commute with i, N is identical with a unitary transformation of the original complex vector space H_0. Thus the canonical forms (22) are obtained by a change in the representation of state vectors, according to the usual terminology of quantum mechanics. It is convenient for us to choose N to be a transformation of this special kind. When this is done, the division algebras C and Q will not in general appear in the particular representations (25) and (26). For the quantum-mechanical applications it is useful to have i in the standard form (5), whereas there is no strong reason to prefer the representations (25), (26) of C and Q to other equivalent representations.

Let now G_1 be a group composed of unitary operators only. For the moment we are not concerned with the antiunitary part of G, and so we consider the algebras A, B, X, Y only. Suppose that the operators $\Lambda(u)$ form an irreducible representation of G_1 over C. Since C is the only division algebra over C, the forms of the algebras B and Y are completely determined by Weyl's theorem

$$B = (C)_n, \qquad Y = nC. \tag{39}$$

Equation (39) is in fact merely a statement of Schur's lemma.[15] Also, it follows from the definitions that every matrix in X which commutes with i belongs to Y.

The order (number of linearly independent elements) of the algebra B is $2n^2$. According to Eqs. (8) and (9), the order of A is $2n^2$ if i belongs to A, n^2 if i does not belong to A. Weyl's theorem then gives precisely three possible canonical forms for the algebras A, X, as follows:

$$A = 2R_n, \qquad X = nR_2, \tag{40}$$

$$A = C_n, \qquad X = nC, \tag{41}$$

$$A = Q_m, \qquad X = m\bar{Q}, \tag{42}$$

where we have written $m = \frac{1}{2}n$. In all three cases the operator i belongs to X, and the order of X is 2 or 4.

Wigner's classification of irreducible representations $\Lambda(u)$ is the following. Let $\Lambda^*(u)$ be the representation formed by taking the complex conjugate in each element of $\Lambda(u)$. If

$$\Lambda^*(u) = M\Lambda(u)M^{-1}, \quad \text{all} \quad u, \tag{43}$$

with M unitary and symmetric, then $\Lambda(u)$ is "potentially real." If Eq. (43) holds with M unitary and antisymmetric, then $\Lambda(u)$ is "pseudoreal." If $\Lambda^*(u)$ is not equivalent to $\Lambda(u)$, then $\Lambda(u)$ is "complex." We write M as usual as a $(2n \times 2n)$ real matrix, and define

$$P = jM \tag{44}$$

with j given by Eq. (4). Then Eq. (43) holds if and only if the matrix P belongs to the commutator algebra X. Therefore an equivalent statement of Wigner's classification is this. If X contains an antiunitary operator P with

$$P^2 = I, \tag{45}$$

then $\Lambda(g)$ is potentially real. If X contains an antiunitary operator P with

$$P^2 = -I, \tag{46}$$

then $\Lambda(g)$ is pseudoreal. If X contains no antiunitary operator P, then $\Lambda(g)$ is complex. An inspection of the canonical forms (40), (41), (42) then yields the following theorem.

Equivalence Theorem I. Let $\Lambda(u)$ be an irreducible representation over C of a unitary group G_1. Let the algebra A be defined by Eq. (8) with real coefficients. Then

(i) *If A is of type R, $\Lambda(u)$ is potentially real;*
(ii) *If A is of type C, $\Lambda(u)$ is complex;*
(iii) *If A is of type Q, $\Lambda(u)$ is pseudoreal.*

In each case the converse is also true.

The following remarks are corollaries of Theorem I.

Remark 1. The matrices $M(u)$ form a real representation of the group G_1. This representation is irreducible over R when A is of type C or Q. It reduces to two equivalent irreducible components when A is of type R.

Remark 2. It is well known[6] that $\Lambda(u)$ is potentially

real if and only if it is equivalent to a representation composed entirely of real matrices. We now can make another statement of the same kind. The irreducible representation $\Lambda(u)$ is pseudoreal if and only if it is equivalent to a representation composed entirely of matrices whose elements are quaternions with real coefficients.

Remark 3. It is well-known (see Wigner's book,[1] p. 289) that the irreducible representations of the 3-dimensional rotation group are potentially real for integer spin, pseudoreal for half-integer spin. From remark 2 it then follows that the integer-spin representations may be taken to be real, and the half-integer-spin representations may be written in terms of real quaternion matrices.

We now turn our attention to the full group G including antiunitary operators. We shall be concerned with the algebras B, D and their commutators Y, Z. An irreducible co-representation of G is a set of matrices $M(u)$, $M(a)$ such that the algebra D is irreducible over R. According to Weyl's theorem there are then three possibilities for the canonical forms of D and Z.

$$D = R_{2n}, \qquad Z = (2n)R, \tag{47}$$

$$D = C_n, \qquad Z = nC, \tag{48}$$

$$D = Q_m, \qquad Z = m\bar{Q}, \quad m = \tfrac{1}{2}n. \tag{49}$$

The algebra B may now be reducible, but its irreducible components must be of the form sC_1. Also, by Eqs. (9) and (10), the order of D must be exactly twice that of B. Equations (47), (48), and (49) then imply that the order of B is $2n^2$, n^2, $\frac{1}{2}n^2$ in the three cases. The only possibilities are

$$B = C_n, \qquad Y = nC, \tag{50}$$

$$B = C_m + C_m, \qquad Y = mC + mC, \tag{51}$$

$$B = 2C_m, \qquad Y = mC_2, \tag{52}$$

and these correspond precisely to the three alternatives (47) to (49).

Wigner's classification of irreducible co-representations[7] is the following. The co-representation is type I if its unitary part is irreducible. It is type II if its unitary part reduces to two equivalent irreducible components. It is type III if its unitary part reduces to two inequivalent irreducible components. Now when the co-representation generates the algebra D, the unitary part of it generates the algebra B. An inspection of Eqs. (50) to (52) shows that these three alternatives correspond to the Wigner types I, III, II, respectively.

Equivalence Theorem II. Let $\Lambda(g)$ be an irreducible co-representation over C of a group G including antiunitary operations. Let the algebra D be defined by Eq. (10) with real coefficients. Then

(i) *If D is of type R, $\Lambda(g)$ is of Wigner type I,*
(ii) *If D is of type C, $\Lambda(g)$ is of Wigner type III,*
(iii) *If D is of type Q, $\Lambda(g)$ is of Wigner type II.*

In each case the converse is also true.

Remark 4. If follows from this theorem that an irreducible co-representation is of type II if and only if it can be expressed in terms of matrices whose elements are real quaternions.

Remark 5. According to Eqs. (47) to (52), the algebra Y has always precisely double the order of the algebra Z. Also, it is known that Y contains the matrix i, which commutes with Z but does not belong to Z. Therefore, in the case here considered (D being irreducible and Z a simple algebra), Y is precisely the direct product of Z with the algebra generated by $(1, i)$.

Remark 6. The statement that Y is the direct product of Z with $(1, i)$ has been established for the case of Z simple. However, by virtue of the lemma of Sec. II (see remark 7 following the lemma) the same relation between Y and Z holds in the general case.

Remark 7. The lemma of Sec. II can be stated very concisely as a statement about co-representations: inequivalent irreducible co-representations of G contain inequivalent irreducible representations of G_1.

IV. FURTHER ANALYSIS OF THE WIGNER CLASSIFICATION

The equivalence Theorems I and II are so alike in form that one might suppose them to be two statements of the same triple alternative. We shall show that in fact the precise opposite is true. The two triple alternatives are entirely independent. Within the same irreducible co-representation of G, any one of the three types of algebra D may occur in combination with any one of the three types of algebra A.

To study the relation between the two theorems, we fix a particular irreducible co-representation $\Lambda(g)$ of G and investigate the possible structure of the six algebras A, B, D, X, Y, Z in combination. Since $\Lambda(g)$ is irreducible, the possible structures for D, Z, B, Y are described by Eqs. (47) to (52). The representation $\Lambda(u)$ of G_1 is however not necessarily irreducible. Theorem I, and the three alternatives given by Eqs. (40) to (42), apply directly only to the irreducible components of $\Lambda(u)$.

When D is of type R, then $\Lambda(u)$ is irreducible and Eqs. (40) to (42) apply unchanged. When D is of type Q, then, according to Eq. (52), $\Lambda(u)$ splits into two identical irreducible components, or symbolically $\Lambda(u) = 2\Lambda'(u)$. In this case Eqs. (40) to (42) apply to $\Lambda'(u)$. When D is of type C, then Eq. (51) holds, and so $\Lambda(u)$ splits into two irreducible components inequivalent over C,

$$\Lambda(u) = \Lambda_1(u) + \Lambda_2(u). \tag{53}$$

The real representation $M(u)$ of G_1 splits correspondingly into two components

$$M(u) = M_1(u) + M_2(u). \tag{54}$$

Equations (40) to (42) apply to Λ_1 and Λ_2 separately. However, we shall prove that the algebra A is necessarily of the same type (R, C, or Q) for the representations Λ_1 and Λ_2. Thus one of Eqs. (40) to (42) applies to both components of $\Lambda(u)$.

Let a be any one of the antiunitary operators in G. The transformation

$$u \to V(u) = a^{-1}ua \tag{55}$$

is an automorphism V of the unitary group G_1. The representations

$$\Lambda_V(u) = \Lambda(V(u)), \qquad M_V(u) = M(V(u)) \tag{56}$$

differ from $\Lambda(u)$ and $M(u)$ only by a relabeling of the elements of G_1. Thus $M_V(u)$ and $M(u)$ generate isomorphic group algebras. Moreover, Eq. (54) implies

$$M_V(u) = [M(a)]^{-1}M(u)M(a) \equiv M(u), \tag{57}$$

where the equivalence is over R and not over C. Suppose now that D is of type C and Eq. (54) holds. Then Eq. (57) means either

$$M_{1V}(u) = [M(a)]^{-1}M_1(u)M(a),$$
$$M_{2V}(u) = [M(a)]^{-1}M_2(u)M(a), \tag{58}$$

or

$$M_{1V}(u) = [M(a)]^{-1}M_2(u)M(a),$$
$$M_{2V}(u) = [M(a)]^{-1}M_1(u)M(a). \tag{59}$$

Because the algebra D generated by $M(u)$, $M(a)$, and i is irreducible, Eq. (58) cannot hold. Therefore Eq. (59) must hold and

$$M_{1V}(u) \equiv M_2(u), \qquad M_{2V}(u) \equiv M_1(u). \tag{60}$$

The algebra A generated by $M_1(u)$ is therefore necessarily of the same type as that generated by $M_2(u)$.

We may thus classify irreducible co-representations of G into nine possible cases, which we denote by RR, RC, RQ, CR, $\cdots$, QQ. Case CR, for example, means that algebra D is of type C while algebra A is of type R, i.e., we have a co-representation of Wigner type III whose unitary part splits into two irreducible inequivalent representations each of which is potentially real.

Using Eqs. (40) to (42) we can write down the possible forms of the algebras A and X in each of the nine cases:

$$\text{case } RR, \quad A = 2R_n, \quad X = nR_2, \tag{61}$$
$$\text{case } RC, \quad A = C_n, \quad X = nC, \tag{62}$$
$$\text{case } RQ, \quad A = Q_m, \quad X = m\tilde{Q}, \tag{63}$$
$$\text{case } CR, \quad A = 2R_m + 2R_m, \quad X = mR_2 + mR_2, \tag{64}$$
$$\text{case } CC1, \quad A = C_m + C_m, \quad X = mC + mC, \tag{65}$$
$$\text{case } CC2, \quad A = 2C_m, \quad X = mC_2, \tag{66}$$
$$\text{case } CQ, \quad A = Q_p + Q_p, \quad X = p\tilde{Q} + p\tilde{Q}, \tag{67}$$
$$\text{case } QR, \quad A = 4R_m, \quad X = mR_4, \tag{68}$$
$$\text{case } QC, \quad A = 2C_m, \quad X = mC_2, \tag{69}$$
$$\text{case } QQ, \quad A = 2Q_p, \quad X = p\tilde{Q}_2, \tag{70}$$

For convenience we wrote here $m = \frac{1}{2}n$, $p = \frac{1}{4}n$. The forms of A and X are uniquely fixed in all cases except CC. Case CC divides into two alternatives $CC1$ and $CC2$. Case $CC1$ holds when the representations $M_1(u)$ and $M_2(u)$ are inequivalent over R; case $CC2$ holds when M_1 and M_2 are equivalent over R.

The results (61) to (70) follow immediately from Eqs. (40) to (42) when D is of type R or Q. However, when D is of type C some further argument is needed. Suppose then that D is of type C, so that Eqs. (48) and (51) hold, and the representation $M(u)$ splits according to Eq. (54). When M_1 and M_2 are inequivalent over R, every matrix commuting with the $M(u)$ must commute separately with $M_1(u)$ and $M_2(u)$. The algebra X is then the direct sum of the commutator algebras of M_1 and M_2. Therefore for M_1 and M_2 inequivalent, Eq. (64), (65), or (67) holds according as A is of type R, C, or Q.

It remains to consider the case in which D is of type C while M_1 and M_2 are equivalent over R. There is then a real matrix L which commutes with all the $M(u)$ but does not commute with $M_1(u)$, $M_2(u)$ separately. This L satisfies

$$M_1(u) = L^{-1}M_2(u)L, \quad M_2(u) = L^{-1}M_1(u)L. \tag{71}$$

Since $\Lambda_1(u)$ and $\Lambda_2(u)$ are inequivalent over C, L must anticommute with i. Now suppose if possible that A were of type R or Q. Then there would exist also a matrix L' in X, anticommuting with i and commuting with each of $M_1(u)$, $M_2(u)$ separately. The product $U = LL'$ would be a matrix commuting with i and also satisfying Eq. (71). This is impossible since Λ_1 and Λ_2 are inequivalent over C. We have thus proved that, if D is of type C and M_1 and M_2 are equivalent, A is also necessarily of type C. There exists then only the case $CC2$ with A and X given by Eq. (66).

We next discuss a special situation in which the above enumeration of possibilities simplifies considerably. We say that the group G is "factorizable" if the automorphism V given by Eq. (55) is an inner automorphism of G_1. Suppose that G is factorizable. Then there exists an element w in G_1 such that

$$V(u) = a^{-1}ua = w^{-1}uw, \quad \text{all} \quad u \quad \text{in} \quad G_1. \tag{72}$$

Then there exists an antiunitary operator

$$T = aw^{-1} \tag{73}$$

in G which commutes with all elements of G_1. Conversely, if such T exists, then $V(u)$ is an inner automorphism for any choice of the antiunitary operator a in Eq. (55). In many physical applications, when such an operator T exists it is convenient to give it the name "time-inversion operator." In any representation $M(g)$ of G, the antiunitary matrix $M(T)$ belongs to the algebra X.

We now classify the possible types of irreducible co-representation of a factorizable group G. Many cases can be immediately eliminated. First, the matrix $M(T)$ belongs to X but does not belong to Y since it anticommutes with i. Therefore $X \neq Y$ for a factorizable group. Hence, by comparing Eqs. (50) to (52) with Eqs. (62), (65), and (69), the cases RC, $CC1$, and QC are excluded. Next, suppose that D is of type C. Then Eq. (72) gives

$$M_V(u) = [M(w)]^{-1}M(u)M(w), \tag{74}$$

with $M(w)$ unitary. Since $\Lambda_1(u)$ and $\Lambda_2(u)$ are inequivalent over C, Eq. (74) implies

$$M_{1V}(u) = [M(w)]^{-1}M_1(u)M(w),$$

$$M_{2V}(u) = [M(w)]^{-1}M_2(u)M(w).$$

This together with Eq. (60) shows that $M_1(u)$ and $M_2(u)$ are equivalent. We proved earlier that cases CR, $CC1$, and CQ are then impossible.

The surviving cases for a factorizable group G are RR, RQ, QR, QQ, and $CC2$.

The operator $[M(T)]^2$ commutes with all $M(g)$ and with i, and it is also equal to $M(u)$ with $u = T^2$. Thus $[M(T)]^2$ belongs to both the algebras D and Z. By Eqs. (47) to (49), the common part of D and Z is $(2n)R$ when D is of type R or Q, and is nC when D is of type C. Since $[M(T)]^2$ is a real orthogonal matrix, it must be a scalar

$$[M(T)]^2 = \epsilon = \pm 1, \tag{75}$$

in any of the four cases RR, RQ, QR, QQ. However, Eq. (75) need not hold in case $CC2$.

We determine lastly which cases go with the plus sign and which with the minus sign in Eq. (75). When G is factorizable and Eq. (75) holds, the algebra D is a direct product of the commuting algebras A and W, where W is the algebra of order 4 generated by $[I, i, M(T), iM(T)]$. The structure of $W = (D/A)$ is then determined as follows:

$$\text{case } RR, \quad A = 2R_n, \quad D = R_{2n}, \quad W \sim R_2, \tag{76}$$

$$\text{case } RQ, \quad A = Q_m, \quad D = R_{2n}, \quad W \sim \tilde{Q}, \tag{77}$$

$$\text{case } QR, \quad A = 4R_m, \quad D = Q_m, \quad W \sim Q, \tag{78}$$

$$\text{case } QQ, \quad A = 2Q_p, \quad D = Q_m, \quad W \sim R_2. \tag{79}$$

The sign of ϵ in Eq. (75) is plus when W is of type R_2, minus when W is of type Q. These results will now be summarized in a theorem.

Theorem III. Let $M(g)$ be an irreducible co-representation of a factorizable group G, in which $M(T)$ is anti-unitary and commutes with all the $M(g)$. Then the following three possibilities alone exist:

(i) *case RR or QQ with* $[M(T)]^2 = +1$,
(ii) *case RQ or QR with* $[M(T)]^2 = -1$,
(iii) *case $CC2$ with* $[M(T)]^2 = \cos\alpha + e \sin\alpha$,

where α may be any real angle, and e is an element of the algebra A with $e^2 = -1$.

Remark 1. It is noteworthy that the sign of $[M(T)]^2$ is determined neither by the Wigner type of the co-representation $M(g)$, nor by the reality type of the unitary subrepresentation $M(u)$, but only by these two types in combination. Thus $[M(T)]^2 = +1$ corresponds to Wigner type I and potentially real, or to Wigner type II and pseudoreal; $[M(T)]^2 = -1$ corresponds to Wigner type II and potentially real, or to Wigner type I and pseudoreal.

Remark 2. In the majority of applications of the theorem, T will be identified with the physical operation of time inversion. In these circumstances $[M(T)]^2 = +1$ for co-representations with integer spin, and $[M(T)]^2 = -1$ for co-representations with

half-integer spin.[7] Therefore cases RR and QQ occur only with integer spin, cases RQ and QR only with half-integer spin. Case $CC2$ may occur with either integer or half-integer spin.

V. EXAMPLES

The classification theory of Secs. III and IV would be empty if one could not produce examples to show that each of the enumerated possibilities can actually occur. We list here one example of each of the ten possibilities (61) to (70). The first five examples are factorizable and illustrate Theorem III. The last five are nonfactorizable.

To simplify the notations we write (2×2) matrices in terms of the standard basis

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad e_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$
$$e_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{80}$$

The quaternion units are defined by Eqs. (26) and (27). The order of the co-representation is the order of the real matrices $M(g)$; this is twice the dimension of the complex vector space H_c.

Example 1. Case RR. Order 2.

G_1 contains identity I only. $G = [I, T]$, $T^2 = I$.

$$M(I) = I_2, \quad M(i) = e_2,$$
$$M(T) = e_3, \quad [M(T)]^2 = I_2.$$

Example 2. Case QR. Order 4.

G_1 generated by $[I, T^2]$, $G = [I, T, T^2, T^3]$ with $T^4 = I$.

$$M(I) = I_4, \quad M(i) = \tau_2,$$
$$M(T) = \tau_1, \quad [M(T)]^2 = -I_4.$$

Example 3. Case RQ. Order 4.

G is generated by the 3-dimensional rotation group O_3, with the time-inversion operator T commuting with O_3. Representation $M(u)$ is with spin $\frac{1}{2}$.

$$M(\vec{n}, \phi) = \exp[\tfrac{1}{2}\phi\vec{n}\cdot\vec{\tau}],$$
$$M(i) = \tau_2', \quad M(T) = \tau_1', \quad [M(T)]^2 = -I_4.$$

Example 4. Case QQ. Order 8.

Same group as example 3. Representation $M(u)$ has two spin-$\frac{1}{2}$ components which are interchanged by the T operator.

$$M(\vec{n}, \phi) = \exp[\tfrac{1}{2}\phi\vec{n}\cdot\vec{\tau}] \times I_2, \quad M(i) = \tau_2' \times I_2,$$
$$M(T) = \tau_1' \times e_2, \quad [M(T)]^2 = I_8.$$

Example 5. Case $CC2$. Order 4.

G is generated by the 2-dimensional rotation group 0_2 together with an operator T commuting with 0_2. The operator T is a combination of time-inversion with space reflection. The phase angle α is a fixed parameter.

$$M(\phi) = \cos\phi[I_2 \times I_2] + \sin\phi[e_3 \times e_2],$$
$$M(i) = I_2 \times e_2.$$
$$M(T) = \cos\tfrac{1}{2}\alpha[e_1 \times e_3] + \sin\tfrac{1}{2}\alpha[e_2 \times e_1],$$
$$[M(T)]^2 = \cos\alpha[I_2 \times I_2] + \sin\alpha[e_3 \times e_2] = M(\alpha).$$

Example 6. Case RC. Order 2.

G is generated by the 2-dimensional rotation group 0_2 with an operator T not commuting with 0_2. T is now time-inversion without space reflection.

$$M(\phi) = (\cos\phi)I_2 + (\sin\phi)e_2,$$
$$M(i) = e_2, \quad M(T) = e_3.$$

Example 7. Case QC. Order 4.

Same group as example 6.

$$M(\phi) = (\cos\phi)I_4 + (\sin\phi)\tau_2,$$
$$M(i) = \tau_2, \quad M(T) = \tau_1.$$

Example 8. Case CR. Order 4.

G_1 is a 4-element group generated by the reflections R_x and R_y in two perpendicular planes. $G = [G_1, TG_1]$, $T^2 = I$, where T is a combination of time-inversion with a reflection in the plane $x = y$.

$$M(I) = I_2 \times I_2, \quad M(R_x) = e_3 \times I_2,$$
$$M(R_y) = -e_3 \times I_2,$$
$$M(i) = I_2 \times e_2, \quad M(T) = e_1 \times e_3.$$

Example 9. Case CQ. Order 8.

G_1 is a direct product $[0_3 \times 0_3]$ of two 3-dimensional rotation groups. $G = [G_1, TG_1]$, where T interchanges the two groups.

$$M(\vec{m}, \phi; \vec{n}, \psi) = [\tfrac{1}{2}(I_2 + e_3)] \times \exp[\tfrac{1}{2}\phi\vec{m}\cdot\tau]$$
$$+ [\tfrac{1}{2}(I_2 - e_3)] \times \exp[\tfrac{1}{2}\psi\vec{n}\cdot\tau],$$
$$M(i) = I_2 \times \tau_2', \quad M(T) = e_1 \times \tau_1'.$$

Example 10. Case $CC1$. Order 4.

$G_1 = [0_2 \times 0_2]$, $G = [G_1, TG_1]$ where T interchanges the two 0_2 groups.

$$M(\phi, \psi) = [\tfrac{1}{2}(I_2 + e_3)] \times [\cos\phi I_2 + \sin\phi e_2]$$
$$+ [\tfrac{1}{2}(I_2 - e_3)] \times [\cos\psi I_2 + \sin\psi e_2],$$

$$M(i) = I_2 \times e_2, \qquad M(T) = e_1 \times e_3.$$

The most interesting of these examples are numbers 4, 5, 8. They have some features which are nontrivial and appear to be novel. We leave to the reader the exercise of verifying that in each case the commutator algebras X, Y, Z have the structure described in Eqs. (47) to (52), (61) to (70).

VI. ALGEBRAIC CHARACTERIZATION OF REPRESENTATION TYPES

In this section we conclude the study of representation types by proving a generalized version of a classical theorem of Frobenius and Schur. Let $M(g)$ be a representation of a group G, irreducible over some ground field Φ with characteristic zero. We suppose that the group G is either finite or compact, and that the matrices $M(g)$ have a finite order d. If $f(g)$ is any function of the group element g, the average of $f(g)$ over G is defined by

$$\mathrm{av}_g f(g) = h^{-1} \sum_g f(g), \tag{81}$$

or by

$$\mathrm{av}_g f(g) = \nu^{-1} \int f(g)\, d\mu(g), \tag{82}$$

where h is the order of G when G is finite, and where ν is the volume of G in the invariant group measure $d\mu(g)$ when G is compact. We consider the fourth-rank tensor

$$P_{ij,kl} = \mathrm{av}_g [M_{ij}(g^{-1}) M_{kl}(g)]. \tag{83}$$

Let K be the group algebra generated by the $M(g)$ with coefficients in Φ.

The structure of K is given by Eq. (24), since Weyl's theorem holds in any field with characteristic zero. The commutator algebra L of K has the structure

$$L = I_t \times \bar{E}, \tag{84}$$

where $\bar{E}$ is an irreducible division algebra of order e over Φ, and $d = te$.

The *type* of the representation $M(g)$ is specified by the division algebra $\bar{E}$. For example, when $\Phi = R$ is the field of real numbers, there are three types of representation corresponding to $\bar{E} = R, C$, or Q. The tensor $P_{ij,kl}$ is useful in classifying representations by virtue of the following theorem.

Theorem IV. The tensor $P_{ij,kl}$ depends only on the integer t and on the algebra $\bar{E}$, and is otherwise independent of the group G and of the representation $M(g)$.

Thus $P_{ij,kl}$ is characteristic of the type of the representation $M(g)$.

To prove the theorem, let g' be any element of G. Then

$$\begin{aligned} P_{ij,kl} &= \mathrm{av}_g [M_{ij}(g'^{-1} g^{-1}) M_{kl}(gg')] \\ &= \sum_{mn} M_{im}(g'^{-1}) P_{mj,kn} M_{nl}(g'). \end{aligned} \tag{85}$$

Thus $P_{ij,kl}$, considered as a matrix in the indices (i, l), commutes with all $M(g')$ and belongs to the commutator algebra L. Similarly, $P_{ij,kl}$ belongs to L when considered as a matrix in (k, j). Let e^μ, $\mu = 1, \cdots, e$, be a linearly independent basis for the algebra $\bar{E}$. Then Eq. (84) gives

$$tP_{ij,kl} = [(I_t)_{il}(I_t)_{kj}] \times \sum_{\mu\nu} c_{\mu\nu}(e^\mu)_{il}(e^\nu)_{kj}, \tag{86}$$

where the $c_{\mu\nu}$ are coefficients in Φ.

Now consider the sum

$$(\overset{\lambda}{\sum})_{jk} = \sum_{il} P_{ij,kl}[I_t \times e^\lambda]_{li}. \tag{87}$$

On the one hand, by Eq. (86),

$$(\overset{\lambda}{\sum})_{jk} = (I_t)_{kj} \times \sum_{\mu\nu} s_{\lambda\mu} c_{\mu\nu}(e^\nu)_{kj}, \tag{88}$$

where

$$s_{\lambda\mu} = \mathrm{spur}\,[e^\lambda e^\mu]. \tag{89}$$

On the other hand, by Eq. (83), since e^λ commutes with all the $M(g)$,

$$\begin{aligned} (\overset{\lambda}{\sum})_{jk} &= \mathrm{av}_g [M(g)(I_t \times e^\lambda) M(g^{-1})]_{kj} \\ &= (I_t \times e^\lambda)_{kj}. \end{aligned} \tag{90}$$

Comparison of Eqs. (88) and (90) shows that

$$\sum_\mu s_{\lambda\mu} c_{\mu\nu} = \delta_{\lambda\nu}, \tag{91}$$

so that the matrix $c_{\mu\nu}$ is the inverse of the matrix $s_{\lambda\mu}$. The coefficients $c_{\mu\nu}$ are thus uniquely determined by $\bar{E}$, and Eq. (86) establishes the truth of Theorem IV.

We shall be interested in applying Theorem IV to cases in which the matrices $M(g)$ are *orthogonal*. So we assume

$$M_{ij}(g^{-1}) = M_{ji}(g). \tag{92}$$

The algebra K then contains the transposed of every matrix in K, and L has the same property. We can therefore choose the basis elements e^λ of the algebra $\bar{E}$ to be either symmetric or antisymmetric. Suppose that the number of symmetric e^λ is q, and the number of antisymmetric e^λ is q'. The invariant

$$P = \sum_{ij} P_{ij,ij} \tag{93}$$

provides a simple criterion for the type of the representation $M(g)$.

Theorem V. When $M(g)$ is an orthogonal irreducible representation of G over a field Φ of characteristic zero,

$$P = \text{av}_g \text{ spur } [M(g^2)] = q - q'. \tag{94}$$

The first part of Eq. (94) follows at once from the definition of P and Eq. (92). To prove the second part, we suppose the e^λ chosen so that

$$(e^\lambda)_{kj} = \eta_\lambda (e^\lambda)_{jk}, \tag{95}$$

with each η_λ equal to ± 1. Then Eqs. (86), (89), and (91) give

$$\begin{aligned} P &= \sum_{\mu\nu ij} c_{\mu\nu} (e^\mu)_{ij} \eta_\nu (e^\nu)_{ji} \\ &= \sum_{\mu\nu} c_{\mu\nu} s_{\nu\mu} \eta_\nu \\ &= \sum_\nu \eta_\nu = q - q'. \end{aligned} \tag{96}$$

Remark 1. Suppose that Φ is the field of real numbers. Then Theorem V gives the following characterization of the type of the representation $M(g)$:

$$P = +1 \text{ for } M(g) \text{ of type } R, \tag{97}$$

$$P = \ \ 0 \text{ for } M(g) \text{ of type } C, \tag{98}$$

$$P = -2 \text{ for } M(g) \text{ of type } Q. \tag{99}$$

Remark 2. We apply remark 1 to the situation discussed in Theorem I of Sec. III. Let $\Lambda(u)$ be an irreducible representation over C of a unitary group G_1. Then the corresponding real representation $M(u)$ splits into two equivalent irreducible representations $M'(u)$ when $M(u)$ is of type R, while $M(u)$ is irreducible when it is of type C or Q. The correspondence between $\Lambda(u)$ and $M(u)$ gives

$$\text{spur } M(u) = 2 \text{ Re spur } \Lambda(u), \tag{100}$$

and therefore the quantity

$$\Pi = \text{av}_u [\text{spur } \Lambda(u^2)] \tag{101}$$

is equal to $\frac{1}{2}P$. This P is given by Eqs. (98), (99) when $M(u)$ is of type C, Q, but is equal to $(+2)$ when $M(u)$ is of type R since Eq. (97) then refers to the irreducible component $M'(u)$. So we derive the classical criterion of Frobenius and Schur[17] for the type of an irreducible unitary representation:

$$\Pi = +1 \text{ for } \Lambda(u) \text{ potentially real}, \tag{102}$$

$$\Pi = \ \ 0 \text{ for } \Lambda(u) \text{ complex}, \tag{103}$$

$$\Pi = -1 \text{ for } \Lambda(u) \text{ pseudoreal}. \tag{104}$$

[17] G. Frobenius and I. Schur, reference 6.

Remark 3. We apply remark 1 to the situation discussed in Theorem II of Sec. III. Let $\Lambda(g)$ be an irreducible co-representation over C of a group G. According to remark 6 of Sec. II, the group algebra D is generated over R not by the group G itself but by an extended group Γ. The representation of Γ which generates D consists of the matrices

$$M(u), \quad iM(u), \quad M(a), \quad iM(a), \tag{105}$$

which are all real and orthogonal. When Theorem V is applied to the group Γ, the contributions from $M(u)$, $iM(u)$ to P cancel each other, while the contributions from $M(a)$, $iM(a)$ are equal. Thus

$$P = \tfrac{1}{2} \text{ av}_a \text{ spur } [M(a^2)], \tag{106}$$

averaged over the antiunitary part only of G. If $\Lambda(u)$ is the unitary part of the co-representation, $\Lambda(u)$ is irreducible when D is of type R, while $\Lambda(u)$ has two irreducible components when D is of type C or Q. In any case we let $\Lambda'(u)$ be one of the (one or two) irreducible components of $\Lambda(u)$, and we write

$$\Pi' = \text{av}_a [\text{spur } \Lambda'(a^2)]. \tag{107}$$

By Eq. (100), this Π' is equal to P when D is of type R, and is equal to $\frac{1}{2}P$ when D is of type C or Q. The criterion of Eqs. (97)–(99) then becomes

$$\Pi' = +1 \text{ for } \Lambda(g) \text{ of Wigner type I}, \tag{108}$$

$$\Pi' = \ \ 0 \text{ for } \Lambda(g) \text{ of Wigner type III}, \tag{109}$$

$$\Pi' = -1 \text{ for } \Lambda(g) \text{ of Wigner type II}. \tag{110}$$

This elegant analog to the Frobenius-Schur criterion was discovered by Bargmann.[18]

VII. THEORY OF MATRIX ENSEMBLES

In this section we deal with the problem for which the theory of Sec. II was specifically introduced, namely the classification of ensembles of matrices with given symmetry properties. An ensemble is a set of objects with an assigned probability distribution. We shall define the probability distributions later; it is necessary first of all to study the classification of sets of matrices invariant under some symmetry-group G.

As in Sec. II, we suppose that the matrices S which we are studying operate in a complex vector space H_c of finite dimension n. We are given a representation of the group G in H_c, consisting of unitary operators $\Lambda(u)$ and antiunitary operators $\Lambda(a)$. The matrices S are supposed to be invariant under G, but this notion of invariance already introduces an ambiguity. There is a choice between two

[18] V. Bargmann (private communication).

definitions of invariance. We say that S is "formally invariant under G" if

$$S\Lambda(g) = \Lambda(g)S, \quad \text{all} \quad g \quad \text{in} \quad G. \tag{111}$$

Formal invariance means that S is unchanged by any of the transformations

$$S \rightarrow \Lambda(g)S[\Lambda(g)]^{-1}, \tag{112}$$

whether g be unitary or antiunitary. We say that S is "physically invariant under G" if for every pair of vectors (ϕ, ψ) in H_c

$$(\phi, S\psi) = (\Lambda(u)\phi, S\Lambda(u)\psi) = (\Lambda(a)\psi, S\Lambda(a)\phi). \tag{113}$$

Note that the initial and final state vectors are interchanged in Eq. (113) in the case of antiunitary elements of G. The effect of Eq. (113) is that we have instead of Eq. (111)

$$S\Lambda(u) = \Lambda(u)S, \qquad S\Lambda(a) = \Lambda(a)S^{+}, \tag{114}$$

where S^{+} means the Hermitian conjugate of S.

The two types of invariance are relevant in different circumstances. If S is, for example, a unitary operator describing a change in the representation of states, then formal invariance under G is a meaningful requirement, signifying that this change in representation does not disturb the symmetry relations of the states under the operations of G. If S is an operator characterizing a physical system, for example a scattering matrix, then the antiunitary operations of G are associated with a reversal of the physical roles of initial and final states; in this case physical invariance of S is the physically meaningful requirement, signifying that the system to which S belongs is invariant under the operations of G in the usual dynamical sense. The two definitions of invariance under G become equivalent only when the matrix S is Hermitian, for example when S is the Hamiltonian of a system.

It is convenient to transcribe the matrix S into a real $(2n \times 2n)$ matrix operating in the real vector space H_R according to Eq. (2). The real form of S then satisfies

$$Si = iS, \tag{115}$$

with i defined by Eq. (5). For S to be invariant under the unitary subgroup G_1 (in either sense) it is necessary and sufficient that

$$SM(u) = M(u)S, \qquad u \quad \text{in} \quad G_1, \tag{116}$$

where the matrices $M(g)$ are the representation of G defined in Sec. II. The condition for S to be formally invariant under G is

$$SM(a) = M(a)S, \qquad a \quad \text{in} \quad G, \tag{117}$$

in addition to Eq. (116). The condition for S to be physically invariant under G is Eq. (116) and

$$SM(a) = M(a)S^{T}, \qquad a \quad \text{in} \quad G, \tag{118}$$

where S^{T} means the transpose of S.

From Eqs. (115), (116) we see that the set Y of matrices in H_R invariant under G_1 is identical with the commutator algebra Y defined by Eqs. (9) and (12). From Eqs. (115)–(117), the set Z of matrices formally invariant under G is identical with the commutator algebra Z defined by Eqs. (10) and (13). We define the set W to consist of those matrices which are both formally and physically invariant under G. Then W is the set of all symmetric matrices in Z. Lastly, we define V to be the set of matrices physically invariant under G. Then we shall prove

Theorem VI. For S to be in V, it is necessary and sufficient that

$$S = S_1 + iS_2, \tag{119}$$

where S_1 and S_2 are matrices in W.

The sufficiency follows immediately from the relations

$$i^{T} = -i, \qquad iM(a) = -M(a)i. \tag{120}$$

To prove the necessity, we observe that all the matrices $M(g)$ are orthogonal, and thus

$$M(g^{-1}) = [M(g)]^{T}. \tag{121}$$

Hence S^{T} belongs to V whenever S does, and we may then write

$$S = S' + S'', \tag{122}$$

where S' is symmetric and S'' antisymmetric, and both S', S'' belong to V. The matrices $S_1 = S'$ and $S_2 = -iS''$ now satisfy both Eq. (117) and Eq. (118), and therefore belong to W.

The results of Sec. II, and in particular Weyl's theorem, provide us with a complete structural analysis of the sets V, W, Y, Z. We use Frobenius' theorem (Sec. I) in order to replace the division algebras E_i of Weyl's theorem by the standard trio R, C, and Q. The integers t_i of Weyl's theorem are now irrelevant since they contribute to the structure of the group algebra D but not to the commutator algebra Z. We thus state the main result of the theory of matrix ensembles as follows.

Theorem VII. The set Z of matrices in H_R formally invariant under G is a direct product of irreducible components, one component Z_i corresponding to each inequivalent irreducible co-representation of G contained in the given co-representation $\Lambda(g)$. Each com-

ponent Z_i may be written as the set of all square matrices of order s_i with elements in an algebra Φ_i. Each Φ_i is either R, the algebra of real numbers, or C, the algebra of complex numbers, or Q, the algebra of real quaternions.

Remark 1. The structure of the set Z_i depends on the Wigner type of the corresponding co-representation of G in the manner specified by equivalence Theorem II. The reality type of the unitary part of the representation, specified by equivalence Theorem I, is here entirely irrelevant, except insofar as the Wigner type and the reality type may be correlated for factorizable groups G according to Theorem III.

According to remark 7 at the end of Sec. II, the sets V, W, Y are direct products of independent components, one corresponding to each component Z_i of Z. To avoid unnecessary repetition, we describe the structure of V, W, Y corresponding to a single component of Z. Thus in the following theorems we assume that Z is irreducible, which means that all irreducible co-representations contained in $\Lambda(g)$ are equivalent. From this special case the general case is easily derived by writing Z_i for Z and taking a direct product over j.

We have seen, in Remarks 5 and 6 of Sec. III, that the algebra Y is generated by i and Z. The matrix i commutes with Z, and therefore commutes with the algebras Φ_i. Hence we may form a new algebra Φ_i^c by adding the independent unit i to Φ_i.

Theorem VIII. When Z is irreducible, the set Y of matrices in H_R invariant under G_1 may be written as the set of all square matrices of order s with elements in an algebra Φ^c, derived from Φ by allowing each element of Φ to have complex instead of real coefficients.

Remark 2. When $\Phi = R$, Φ^c is the algebra of ordinary complex numbers. When $\Phi = C$, Φ^c is the algebra of complex-complex numbers with two commuting imaginary units; in this case Φ^c is reducible and has the structure

$$\Phi^c \sim C + C. \tag{123}$$

When $\Phi = Q$, Φ^c is the algebra of complex quaternions, which is equivalent to an algebra of complex (2×2) matrices,

$$\Phi^c \sim C_2. \tag{124}$$

The algebra W consists of matrices which are symmetric when written in expanded form in H_R. When S is written, as in Theorems VII and VIII, as a smaller matrix with elements in Φ, the condition of symmetry becomes a condition of Φ duality, as follows. We define the Φ conjugate of a number in Φ to be the number obtained by reversing the signs of the coefficient of e_2 (in the case $\Phi = C$) or of the coefficients of τ_1, τ_2, τ_3 (in the case $\Phi = Q$). We define the Φ dual of a matrix to be the transposed matrix with each element Φ conjugated. Since the units e_2, τ_1, τ_2, τ_3 when written in expanded form are antisymmetric, a matrix which is symmetric in expanded form becomes Φ self-dual when written with elements in Φ.

Theorem IX. When Z is irreducible, the set W of matrices in H_R invariant under G in both physical and formal senses may be written as the set of all square self-dual matrices of order s with elements in $\Phi = R$, C, or Q.

The Φ conjugate of an element of Φ^c is obtained by changing the signs of the coefficients of the Φ units, leaving the unit i unchanged. So from Theorems VI and IX follows immediately the result:

Theorem X. When Z is irreducible, the set V of matrices in H_R invariant under G in the physical sense may be written as the set of all square self-dual matrices of order s with elements in Φ^c.

Remark 3. We now finally make contact with the theory of matrix ensembles developed earlier by the author.[4] Let V_U be the subset of unitary matrices in V. Then Theorem X states that, for the most general symmetry group G and the most general quantum-mechanical representation of G, the set V_U is a direct product of independent components, each of which is identical with one of the three ensemble-spaces T_1, T_2, and T_4 defined in reference 4. The cases T_1, T_2, T_4 correspond, respectively, to $\Phi = R$, C, Q. The spaces T_1, T_2, and T_4 were originally obtained by considering special groups G of a very simple kind. It is satisfactory to find that the same three spaces, and no others, occur in all possible circumstances.

The reason for choosing V_U as the space in which to construct an ensemble is that no natural definition of uniform probability appears to exist in V. For the same reason we study the subset Z_U of unitary matrices in Z. The following theorem follows from Theorem VII together with well-known properties of the classical groups.[2]

Theorem XI. The set Z_U of unitary matrices in H_R formally invariant under G is a direct product of irreducible components, each of which is a simple classical group. When $\Phi_i = R$, C, or Q, the corresponding component of Z_U is an orthogonal, unitary, or symplectic group of dimension s_i.

In the same way we define the unitary subset Y_U of Y. The components of Y_U are

$$Y_V = U(s), \quad U(s) \times U(s), \quad U(2s), \tag{125}$$

corresponding to

$$Z_V = 0(s), \quad U(s), \quad Sp(2s). \tag{126}$$

The unitary space V_V is not a group. But it can be represented conveniently in terms of the groups Y_V, Z_V in the following way. A matrix S belongs to V_V if and only if it can be expressed as a Φ-symmetric product

$$S = UU^D, \quad U \text{ in } Y_V, \tag{127}$$

where D denotes Φ dual. All matrices U' of the form

$$U' = UU_1, \quad U_1 \text{ in } Z_V, \tag{128}$$

correspond to the same S by Eq. (127), and every U' corresponding to S is of the form (128). Thus each matrix S in V_V corresponds to a unique co-set of the subgroup Z_V in the group Y_V. We have thus proved

Theorem XII. The set V_V of unitary matrices in V is abstractly equivalent to the homogeneous space (Y_V/Z_V), the quotient of the group Y_V by its subgroup Z_V.

Having defined the spaces Z_V and V_V, we are now in a position to define the corresponding invariant matrix ensembles. The ensemble E^P of unitary matrices formally invariant under G is defined as the space Z_V with probability distribution given by the invariant group measure in Z_V. Since Z_V is a direct product of simple classical groups, the group measure in Z_V is merely the product of the invariant measures in the irreducible components of Z_V. The ensemble E^P of unitary matrices physically invariant under G is defined as the space V_V with measure given according to Theorem XII by

$$d\mu(V_V) = [d\mu(Y_V)/d\mu(Z_V)]. \tag{129}$$

Here $d\mu(Y_V)$ and $d\mu(Z_V)$ are the invariant group measures in Y_V and Z_V, and the quotient measure is defined in the obvious way. Alternatively, the quotient measure may be uniquely defined as the measure in V_V which is invariant under all automorphisms

$$S \to USU^D, \quad U \text{ in } Y_V, \tag{130}$$

of V_V into itself. The ensemble E^P is a direct product of irreducible components, each of which is identical with one of the three types E_1, E_2, E_4 which were studied in reference 4.

Two other types of ensemble naturally suggest themselves for study, composed of Hermitian and anti-Hermitian matrices, respectively. A matrix S of V_V is of the form

$$S = \exp[iH], \quad H \text{ in } W, \tag{131}$$

while a matrix S of Z_V is of the form

$$S = \exp[A], \quad A \text{ in } Z_A, \tag{132}$$

where Z_A is the subset of Z containing anti-Hermitian matrices. Thus W and Z_A are the spaces of infinitesimal generators for V_V and Z_V, respectively.

We define the Hermitian Gaussian ensemble E^H as the space W of matrices H with the probability distribution

$$d\mu(H) = C \exp[-(\text{spur } H^2)/4a^2] \prod dH^a_{ij}, \tag{133}$$

where c, a are constants and the product extends over all the independent real coefficients of the elements of H in the algebra Φ. The anti-Hermitian Gaussian ensemble E^H is defined as the space Z_A with probability distribution

$$d\mu(A) = C \exp[+(\text{spur } A^2)/4a^2] \prod dA^a_{ij}. \tag{134}$$

These ensembles have an algebraic structure precisely analogous to that of E^P and E^F, respectively. They divide into irreducible components each of which is of one of the three types R, C, or Q. In particular, E^H is the natural ensemble to use in describing the statistical properties of the Hamiltonian H of a system known to be physically invariant under the group G.

The physical motivation for considering ensembles of matrices with probability distributions defined in these various ways has been discussed by Wigner[19] and by the author.[4] In the case of the ensembles E^F and E^P, consisting of unitary matrices, the existence of a natural uniform measure provides an intuitively plausible definition of "equal *a priori* probability." In the case of the ensembles E^H and E^A, consisting of Hermitian and anti-Hermitian matrices, the choice of a Gaussian probability distribution is mainly a matter of mathematical convenience. Rosenzweig[20] has argued that one should use in preference to Eq. (133) a "microcanonical ensemble" with the exponential replaced by a delta function

$$\delta[\text{spur } (H^2) - \sigma^2].$$

The algebraic structure of E^H and E^A would of course not be affected by such a change.

In any physical situation to which the ensembles E^P or E^H are relevant, we have a system specified by a unitary operator S or by a Hermitian H.

[19] E. P. Wigner, *Proceedings of the 4th Canadian Mathematics Congress* (University of Toronto Press, Toronto, Canada, 1959), p. 174.

[20] N. Rosenzweig, Bull. Am. Phys. Soc. 7, 91 (1962).

Since the system is invariant under G, every stationary state is associated with a particular irreducible co-representation of G. Each irreducible co-representation fixes the values of a certain set of quantum numbers (spin, parity, isotopic spin, etc.) which are attached to the energy levels belonging to that co-representation. The fact that the ensemble E^P or E^H is a direct product of irreducible components means that the energy levels belonging to different sets of quantum numbers are statistically uncorrelated. Thus the statistical properties of energy levels are entirely determined by the behavior of the individual level-series, each associated with one set of quantum numbers. A single level-series is described by an irreducible ensemble. The final result of our analysis may then be stated as follows: When we consider a single series of energy levels of a complex system, having definite values for all quantum numbers of the symmetry-group G, the statistical behavior of these levels follows one of three possible laws, corresponding to the three types of irreducible ensemble E^P or E^H.

VIII. EIGENVALUE DISTRIBUTIONS

In this section we list without proof the joint probability distributions of the eigenvalues of matrices belonging to the irreducible ensembles E^P, E^A, E^P, E^H. In each case the integer s is the dimension of the algebra Z over the field Φ which may be R, C, or Q. The constant c will not be the same each time it appears.

1. E^P. *Ensemble of Unitary Matrices Formally Invariant under G*

(α) $\Phi = R$, $Z_U = O(s)$. In this case Z_U (the orthogonal group) splits into two disconnected parts, consisting of matrices with determinant Δ equal to $+1$ and -1, respectively. There are thus four distinct eigenvalue distributions to be listed.

(i) $s = 2n$, $\Delta = 1$, eigenvalues $\exp(\pm i\theta_j)$,

$$P(\theta_1, \cdots, \theta_n) = c \prod_{i<j} [\cos\theta_i - \cos\theta_j]^2. \quad (135)$$

(ii) $s = 2n$, $\Delta = -1$, eigenvalues ± 1, $\exp(\pm i\theta_j)$,

$$P\ \theta_1, \cdots, \theta_{n-1}) = c \prod_j (1 - \cos^2\theta_j) \times \prod_{i<j} [\cos\theta_i - \cos\theta_j]^2. \quad (136)$$

(iii) $s = 2n + 1$, $\Delta = 1$, eigenvalues $+1$, $\exp(\pm i\theta_j)$,

$$P(\theta_1, \cdots, \theta_n) = c \prod_j (1 - \cos\theta_j) \times \prod_{i<j} [\cos\theta_i - \cos\theta_j]^2. \quad (137)$$

(iv) $s = 2n + 1$, $\Delta = -1$, eigenvalues -1, $\exp(\pm i\theta_j)$,

$$P(\theta_1, \cdots, \theta_n) = c \prod_j (1 + \cos\theta_j) \times \prod_{i<j} [\cos\theta_i - \cos\theta_j]^2. \quad (138)$$

(β) $\Phi = C$, $Z_U = U(s)$, eigenvalues $\exp(i\theta_j)$.

$$P(\theta_1, \cdots, \theta_s) = c \prod_{i<j} |\exp(i\theta_i) - \exp(i\theta_j)|^2. \quad (139)$$

(γ) $\Phi = Q$, $Z_U = \mathrm{Sp}(2s)$, eigenvalues $\exp(\pm i\theta_j)$.

$$P(\theta_1, \cdots, \theta_s) = c \prod_j (1 - \cos^2\theta_j) \times \prod_{i<j} [\cos\theta_i - \cos\theta_j]^2. \quad (140)$$

2. E^A. *Gaussian Ensemble of Anti-Hermitian Matrices Formally Invariant under G*

(α) $\Phi = R$, matrices real and antisymmetric.

(i) $s = 2n$, eigenvalues $\pm iE_j$,

$$P(E_1, \cdots, E_n) = c[\prod_{i<j} (E_i^2 - E_j^2)^2] \times \exp[-\sum_j E_j^2/2a^2]. \quad (141)$$

(ii) $s = 2n + 1$, eigenvalues 0, $\pm iE_j$,

$$P(E_1, \cdots, E_n) = c[\prod_j E_j^2][\prod_{i<j} (E_i^2 - E_j^2)^2] \times \exp[-\sum_j E_j^2/2a^2]. \quad (142)$$

(β) $\Phi = C$. Eigenvalues iE_j.

$$P(E_1, \cdots, E_s) = c[\prod_{i<j} (E_i - E_j)^2] \times \exp[-\sum_j E_j^2/4a^2]. \quad (143)$$

(γ) $\Phi = Q$. Eigenvalues $\pm iE_j$.

$$P(E_1, \cdots, E_s) = c[\prod_j E_j^2][\prod_{i<j} (E_i^2 - E_j^2)^2] \times \exp[-\sum_j E_j^2/2a^2]. \quad (144)$$

3. E^P. *Ensemble of Unitary Matrices Physically Invariant under G*

Eigenvalues $\exp(i\theta_j)$, each doubly degenerate in the case $\Phi = Q$.

$$P(\theta_1, \cdots, \theta_s) = c \prod_{i<j} |\exp(i\theta_i) - \exp(i\theta_j)|^\beta, \quad (145)$$

with $\beta = 1, 2, 4$ for $\Phi = R, C, Q$, respectively.

4. E^H. Gaussian Ensemble of Hermitian Matrices Invariant (in either sense) under G

Eigenvalues E_i, each doubly degenerate in the case $\Phi = Q$.

$$P(E_1, \cdots, E_n) = c[\prod_{i<j} |E_i - E_j|^\beta] \times \exp[-\sum_i E_i^2/4a^2], \quad (146)$$

with $\beta = 1, 2, 4$ for $\Phi = R, C, Q$.

Proofs of Eqs. (135) to (140) are to be found in Chapter 7 of Weyl's book.[2] Equations (141) to (144) can be deduced as limiting cases of Eqs. (135) to (140) when all angles θ_j are small. Similarly Eq. (146) can be deduced from Eq. (145). The proof of Eq. (145) has been given by the author.[4]

The statistical properties of the eigenvalues resulting from each of these ensembles can be studied by following the method used by the author[4] for the case of Eq. (145). The eigenvalue distribution in each ensemble has an exact mathematical analog in the form of a classical Coulomb gas.

We briefly describe the Coulomb gas analogs to E^P and E^A when $\Phi = R$ or Q. In E^P the numbers

$$x_j = \cos \theta_j \quad (147)$$

are considered to be positions of unit charges, constrained to move on the segment $[-1 \leq x \leq +1]$, which may be imagined to be a straight conducting wire of length 2. Every two charges repel each other with the potential

$$W(x_i - x_j) = -\ln |x_i - x_j|. \quad (148)$$

In addition there are fixed charges of q_+ units at $x = +1$ and of q_- units at $x = -1$. When $\Phi = R$, the angles θ_j are rotation angles of a random rotation in the orthogonal group $O(s)$. The values of q_+, q_- are

(i) $s = 2n$, $\Delta = 1$; $q_+ = q_- = -\frac{1}{4}$.

(ii) $s = 2n$, $\Delta = -1$; $q_+ = q_- = +\frac{1}{4}$.

(iii) $s = 2n + 1$, $\Delta = 1$; $q_+ = +\frac{1}{4}$, $q_- = -\frac{1}{4}$.

(iv) $s = 2n + 1$, $\Delta = -1$; $q_+ = -\frac{1}{4}$, $q_- = +\frac{1}{4}$.

When $\Phi = Q$, the angles θ_j are rotation angles of a random matrix in the symplectic group $\mathrm{Sp}(2s)$. In this case $q_+ = q_- = +\frac{1}{4}$. The temperature of the gas is the same for $\Phi = R$ or Q, namely, $T = \frac{1}{2}$.

The Gaussian antisymmetric ensembles E^A for $\Phi = R$ or Q have a Coulomb analog composed of unit charges with positions

$$x_j = E_j^2, \quad (149)$$

constrained to move on the semi-infinite straight wire $0 \leq x < \infty$. The repulsion between charges is again given by Eq. (148), and $T = \frac{1}{2}$ as before. There is a fixed charge of q units at $x = 0$, where

$$q = -\tfrac{1}{4} \text{ when } \Phi = R, \quad s = 2n,$$

$q = +\frac{1}{4}$ when $\Phi = R$, $s = 2n + 1$, or when $\Phi = Q$. In addition to the Coulomb forces, each charge x_j is subject to a constant downward force produced by a "gravitational potential"

$$V(x) = [x/4a^2]. \quad (150)$$

When $\Phi = C$, the ensembles E^P and E^A become identical with E^P and E^H, for which the Coulomb analogs have been described previously.[4]

The whole of the previous analysis[4] of level distributions, based on the ensembles E^P, can be repeated with minor modifications for the other ensembles E^P, E^A, E^H. However, there is one basic difference between the physical ensembles E^P, E^H on the one hand and the formal ensembles E^P, E^A on the other.

Theorem XIII. Consider an irreducible ensemble of matrices over the field Φ, with order $s \to \infty$. In E^P or E^A, the local statistical behavior of eigenvalues is described by an infinite Coulomb gas with temperature $T = \frac{1}{2}$ independent of Φ. In E^P or E^H the local behavior of eigenvalues is described by an infinite Coulomb gas with temperature $T = 1, \frac{1}{2}, \frac{1}{4}$ corresponding to $\Phi = R, C, Q$.

The most striking qualitative feature of the physical ensembles E^P, E^H is that the strength of the repulsion between neighboring energy levels depends on the Wigner type of the co-representation to which these levels belong. This feature is absent in the formal ensembles E^P, E^A.

ACKNOWLEDGMENTS

The author is grateful above all to Dr. V. Bargmann for suggesting to him this line of inquiry, and for help on specific points as noted in the text. He also profited greatly from conversations with Dr. E. Wigner and Dr. A. Grossmann.

Evaluation of the Quantum-Mechanical Ring Sum with Boltzmann Statistics*

HUGH E. DEWITT
Lawrence Radiation Laboratory, University of California, Livermore, California
(Received December 9, 1961, revised manuscript received July 26, 1962)

A procedure is given for the evaluation of the quantum-mechanical ring sum at finite temperature. The method is used for the evaluation of the quantum corrections to the classical Debye-Hückel free energy for an electron gas obeying Boltzmann statistics. The ring sum is shown to be of the form $(\beta e^2/\pi\lambda_D)P(\gamma)$, where $\gamma = \lambda\!\!\!^{-}/\lambda_D$, $\lambda\!\!\!^{-} = \hbar/(2mkT)^{1/2}$, and λ_D is the Debye screening length. The quantum effects for finite γ are due only to the operation of the uncertainty principle. The function $P(\gamma)$ decreases monotonically from the classical value $\pi/3$, and the form is shown to be

$$P(\gamma) = (\pi/3)[1 + \sum_{n=2} a_n\gamma^{2(n-1)}]^{3/2} - \sum_{n=2} b_n\gamma^{2n-3}.$$

The coefficients a_n and b_n are evaluated exactly for small n and asymptotically for large n. The two series converge for $\gamma^2 < \gamma_c^2 = 2.042\cdots$.

For $\gamma \gg \gamma_c$ the function $P(\gamma)$ is also evaluated as an asymptotic expansion in inverse powers of $\gamma^{1/2}$. Thus the low-temperature correlation energy of distinguishable electrons is obtained in random phase approximation. At zero temperature, the result is the same as the correlation energy obtained by Foldy for charged bosons. The correlation pressure in this approximation is negative and diverges as $\rho^{1/4}$ at high density.

I. INTRODUCTION

It is well known in the classical theory of the electron gas and in the theory of electrolytic solutions that the sum of the ring interactions gives the most important correction to the ideal gas free energy in the low density limit, namely, the Debye-Hückel term.[1] Similarly, the same group of terms gives the Gell-Mann and Brueckner expression for the correlation energy of the electron gas at high density and zero temperature.[2] A general expression for the sum of the ring interactions valid at all temperatures has been derived by several authors using different formalisms for the perturbation expansion of the quantum-mechanical partition function.[3-5] In published work, this general formula has been used only to obtain the two limits, namely, at very high temperatures and at zero temperature. The purpose of this and succeeding papers is to produce analytic evaluations of the general ring sum formula which will bridge the enormous gap between the two known limits.

The quantum effects in the ring sum are due to quantum statistics and the uncertainty principle. At intermediate temperatures the two effects are inextricably intertwined in the rather complicated ring sum formula. In order to reduce the complexity somewhat in this paper, we will evaluate the ring sum with Boltzmann statistics. Thus, the quantum corrections to the classical Debye-Hückel result will be due only to the uncertainty principle. These quantum corrections are often referred to as diffraction effects.

The fundamental lengths of the electron gas are: the classical interaction length βe^2, the Debye screening length $\lambda_D = (4\pi\beta e^2/\rho)^{-1/2}$, and the thermal de Broglie wavelength $\lambda\!\!\!^{-} = \hbar(\beta/2m)^{1/2}$ with $\beta = 1/kT$. Out of these three lengths, two independent dimensionless parameters may be formed. The classical parameter is $\Lambda = \beta e^2/\lambda_D = 2\pi^{1/2}e^3\beta^{3/2}\rho^{1/2}$. The second parameter measures quantum effects; it is the ratio $\gamma = \lambda\!\!\!^{-}/\lambda_D = (2\pi)^{1/2}\hbar e\beta\rho^{1/2}m^{-1/2}$. The Helmholtz free energy for a gas of N electrons in volume V containing a continuous neutralizing positive charge background with $\rho = \lim N/V$ as $N \to \infty$ and $V \to \infty$ is

$$\beta F = \beta F_0 - NS(\Lambda, \gamma). \tag{1}$$

$S(\Lambda, \gamma)$ is the contribution from the sum of all multiply connected diagrams representing the perturbation expansion of the canonical ensemble.[6] The

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] J. E. Mayer, J. Chem. Phys. 18, 1426 (1950).

[2] M. Gell-Mann and K. A. Brueckner, Phys. Rev. 106, 364 (1957).

[3] E. W. Montroll and J. C. Ward, Phys. Fluids 1, 55 (1958).

[4] A. Sakakura, "The Connected Diagram Expansion of the Grand Partition Function and the Statistical Mechanics of the Electron Gas," Ph. D. thesis (1960) at the University of Colorado (unpublished).

[5] H. E. DeWitt, J. Nuclear Energy, Part C: Plasma Physics 2, 27 (1961).

[6] Multiply connected has the same meaning as irreducible in the Mayer cluster expansion. For an account of the irreducible or multiply connected diagram expansion for equation of state of a classical nonideal gas see T. E. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), Chap. 5. For a more recent derivation of the perturbation expansion of the Helmholtz free energy including quantum statistics see R. Brout and F. Englert, Phys. Rev. 120, 1519 (1960). If Boltzmann statistics are used in the Brout-Englert method, then all singly connected diagrams vanish.

ring diagrams are the simplest terms in this sum. They represent n particles which interact n times with momentum $\hbar\mathbf{k}$ transferred from one particle to the next at each interaction. For the Coulomb interaction the nth ring term is the most divergent part of the nth virial coefficient.

In the classical limit ($\hbar = 0$ and hence $\gamma = 0$), the form of S when $\Lambda \ll 1$ for the electron gas is

$$S(\Lambda, 0) = \Lambda/3 + \Lambda^2/6\,(\ln 3\Lambda - 2C + 11/6) \cdots. \quad (2)$$

The $\Lambda/3$ in Eq. (2) is the classical limit of the ring sum, the Debye-Hückel term. The higher terms in Λ may be obtained by using the Meeron nodal expansion.[7] The explicit result for the $\Lambda^2 \ln \Lambda$ term was given by Abe.[8] In a later paper the evaluation of quantum corrections to these higher-order terms will be discussed. Our task in this paper is to evaluate the function of γ multiplying the classical Debye term.

II. THE QUANTUM-MECHANICAL RING SUM

The general form of the quantum-mechanical ring sum as derived by summing the ring diagram part of nth-order perturbation theory from $n = 2$ to ∞ is

$$NS_{\text{ring}} = \frac{1}{2}\int \frac{V\,d^3k}{(2\pi)^3} \sum_{t=-\infty}^{\infty} \{4\pi\alpha(\mathbf{k}, 2\pi it) - \ln\,[1 + 4\pi\alpha(\mathbf{k}, 2\pi it)]\}, \quad (3)$$

where

$$4\pi\alpha(\mathbf{k}, 2\pi it) = u(k)\lambda_t(k), \quad (4)$$

$u(k)$ is the Fourier transform of the pair potential. For $u(r) = e^2/r$ it is

$$u(k) = (1/V)\int d^3r\, \exp(i\mathbf{k}\cdot\mathbf{r})e^2/r = 4\pi e^2/Vk^2.$$

$\lambda_t(k)$ is the tth Fourier component of the pair interaction propagator. Because of the periodicity property of the propagator $G(k, \beta - \tau) = G(k, \tau)$, the expansion is a Fourier series. For Boltzmann statistics this propagator is

$$G(\mathbf{k}, \beta' - \beta'') = N(\beta/2\pi m)^{3/2} \times \int d^3p\, \exp -\{\beta p^2/2m + (\beta' - \beta'') \times [(\mathbf{p} + \hbar\mathbf{k})^2 - p^2]/2m\} = Ne^{-\kappa^2 v(1-v)}, \quad (5)$$

where $v = (\beta' - \beta'')/\beta$, and $\kappa = \lambda k$. The Fourier components are defined as

$$\lambda_t(k) = \int_0^\beta d\tau G(k, \tau)e^{2\pi i t\tau/\beta}.$$

If we define $\lambda_t(k) = N\beta L(\lambda^2k^2, 2\pi it)$, where L is a dimensionless function, the expansion of the Boltzmann form of the propagator is

$$e^{-\kappa^2 v(1-v)} = \sum_{t=-\infty}^{\infty} e^{2\pi i t v} L_t(\kappa^2), \quad (6)$$

$$L_t(\kappa^2) = \int_0^1 dv\, \exp\,[-\kappa^2 v(1 - v) + 2\pi itv],$$

$$= (-1)^t(2/\kappa)e^{-\kappa^2/4}\int_0^{\kappa/2} ds\, e^{s^2}\cos(2\pi ts/\kappa). \quad (7)$$

For a multicomponent gas composed of several types of charged particles and subject to the electrical neutrality condition, $\sum_i z_iN_i$, the same form as Eq. (3) holds for the ring sum, but in place of (4) we have

$$4\pi\alpha(\mathbf{k}, 2\pi it) = u(k)\sum_i z_i^2\lambda_{i,t}(k) = \beta u(k)\sum_i z_i^2N_iL(\lambda_i^2k^2, 2\pi it). \quad (8)$$

In the classical limit $\hbar = 0$, one sees from Eq. (7) that $L_0 = 1$ and $L_{t\neq 0} = 0$, so that for the Coulomb potential Eq. (8) reduces to

$$4\pi\alpha(\mathbf{k}, 0) = (4\pi\beta e^2\sum_i z_i^2\rho_i)/k^2 = 1/\lambda_D^2k^2, \quad (9)$$

where Eq. (9) defines the usual multicomponent Debye screening length. Consequently, the ring sum, Eq. (3), reduces to

$$S_{\text{ring}} = \left(\frac{1}{4\pi^2\rho}\right)\int_0^\infty k^2\,dk \times \left\{\frac{1}{(k\lambda_D)^2} - \ln\left[1 + \frac{1}{(k\lambda_D)^2}\right]\right\}. \quad (10)$$

The integral in Eq. (10) is $(\pi/3)\lambda_D^{-3}$, and since $(4\pi\rho\lambda_D^3)^{-1} = \Lambda$, we obtain the Debye-Hückel result $S_{\text{ring}} = \Lambda/3$.

Another form of the ring sum should be mentioned because of its importance in the physical interpretation of the expression (3), and because it shows the equivalence of the random phase approximation with the approximation of retaining only the ring diagrams in the perturbation expansion of the free energy. The summation in Eq. (3) may be written as

$$\frac{1}{4\pi i}\int_C dw\, \coth(w/2) \times \{4\pi\alpha(\mathbf{k}, w) - \ln\,[1 + 4\pi\alpha(\mathbf{k}, w)]\}, \quad (11)$$

[7] E. Meeron, Phys. Fluids 1, 139 (1958); E. Meeron and E. R. Rodemich, *ibid.* 1, 246 (1958).

[8] R. Abe, Progr. Theoret. Phys. (Kyoto) 22, 213 (1959).

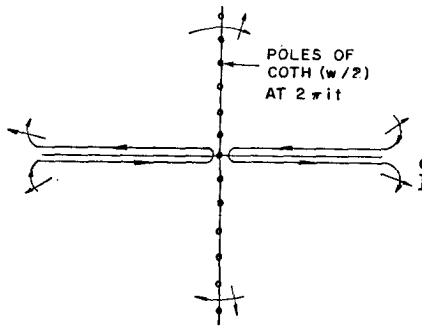


FIG. 1. Contour of integration for Eq. (12).

where the contour C encloses the entire imaginary axis of the complex variable $w = \eta + i\epsilon$. The singularities of the integrand are the simple poles of coth $(w/2)$ at $w = 2\pi it$. The sum of the residues from these poles is the result in Eq. (3). The function $L(\kappa^2, w)$ may be shown to be discontinuous across the real axis, so that we have

$$\alpha^{\pm} = \lim_{\epsilon \to 0} \alpha(\mathbf{k}, \eta \pm i|\epsilon|) = \alpha_R(\mathbf{k}, \eta) \pm i\alpha_I(\mathbf{k}, \eta).$$

Thus, when the contour C is deformed as shown in Fig. 1, Eq. (11) may be written as an integral along the real axis:

$$\begin{aligned} \sum_t &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\eta \coth(\eta/2) \\ &\quad \times 2\,\mathrm{Im}\{4\pi\alpha^+ - \ln(1 + 4\pi\alpha^+)\}, \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\eta \coth(\eta/2) \\ &\quad \times \left\{4\pi\alpha_I - \tan^{-1}\left(\frac{4\pi\alpha_I}{1 + 4\pi\alpha_R}\right)\right\}. \end{aligned} \tag{12}$$

The integration variable η in Eq. (12) should be thought of as $\beta\hbar\omega$ where ω is the oscillation frequency of a density fluctuation in the medium.

The quantity $4\pi\alpha(\mathbf{k}, \beta\hbar\omega)$ is the generalized, temperature-dependent polarizability in the dynamic dielectric constant of the medium, $\epsilon(\mathbf{k}, w) = 1 + 4\pi\alpha(\mathbf{k}, \beta\hbar\omega)$. Englert and Brout have developed a dielectric formulation of the plasma free energy in the framework of the random phase approximation.[9] They obtained Eq. (3) by calculating from the average of the square of a density fluctuation defined by $\rho_q = \sum_p a_{p+q}^+ a_p$ with $\mathbf{q} = \hbar\mathbf{k}$. From their formulation it is clear that Eq. (12) is a special case of the fluctuation dissipation theorem.

With the interpretation of $4\pi\alpha(\mathbf{k}, w)$ as the polarizability of the medium, one sees that the classical limit form of the free energy depends only on the static dielectric constant. Quantum effects enter both as a modification of the static dielectric constant, i.e., for $t = 0$, and in the appearance of contributions from $t \neq 0$ in Eq. (3). The $t \neq 0$ contributions correspond to the imaginary frequencies $\omega_t = 2\pi it/\hbar\beta$, for which the dynamic polarizability is nonzero when $\hbar \neq 0$.

Equations (3) and (12) are two alternative forms to use in the analytical evaluation of $S_{\mathrm{ring}}(\Lambda, \gamma)$. The dielectric formulation, Eq. (12), is perhaps the easier of the two forms for physical interpretation. However, for the purpose of obtaining an analytic evaluation of S_{ring}, Eq. (3) is most suitable. For later convenience we rewrite Eq. (3) as

$$\begin{aligned} S_{\mathrm{ring}} &= (\Lambda/\pi)P(\gamma) \\ &= (\Lambda/\pi)\left[P_0(\gamma) + 2\sum_{t=1}^{\infty} P_t(\gamma)\right], \\ P_t(\gamma) &= \int_0^{\infty} x^2\,dx\{x^{-2}L_t(\gamma^2x^2) \\ &\quad - \ln[1 + x^{-2}L_t(\gamma^2x^2)]\}, \end{aligned} \tag{13}$$

where the integration variable is $x = k\lambda_D$. As noted from Eq. (10) the classical limit of $P(\gamma)$ is $P_0(0) = \pi/3$.

III. PROPERTIES OF THE FOURIER COMPONENTS OF THE PROPAGATOR

In order to perform the integration giving $P_t(\gamma)$ as defined by Eq. (13), several properties of the $L_t(\kappa^2)$ functions are needed. For $t = 0$ the function given by Eq. (7) may be expressed in terms of Erfi $(a) = \int_0^a ds\, e^{s^2}$ and it has the following series expansion convergent for all κ:

$$\begin{aligned} L_0(\kappa^2) &= (2/\kappa)e^{-\kappa^2/4}\,\mathrm{Erfi}\,(\kappa/2), \\ &= \sum_{n=0} \frac{(-1)^n(\kappa^2/2)^n}{(2n+1)!!}. \end{aligned} \tag{14}$$

For real $\kappa \ll 1$ the following asymptotic expansion is valid:

$$L_0(\kappa^2) = \frac{1}{(\kappa^2/2)} + \frac{1}{(\kappa^2/2)^2} + \cdots + \frac{(2m-3)!!}{(\kappa^2/2)^m}. \tag{15}$$

Similar results for $L_t(\kappa^2)$ when $t \neq 0$ are easily obtained with the help of the differential equation satisfied by $L_t(\kappa^2)$:

$$\kappa\frac{dL_t}{d\kappa} = 1 - (1 + \tfrac{1}{2}\kappa^2)L_t - \frac{2(\pi t)^2}{\kappa^2}L_t, \tag{16}$$

which may be verified by differentiating the integral definition Eq. (7) and integrating by parts. The

[9] F. Englert and R. Brout, Phys. Rev. 120, 1085 (1960).

series expansion of $L_t(\kappa^2)$ obtained by solving Eq. (16) is

$$L_t(\kappa^2) = \frac{1}{(\pi t)^2}(\kappa^2/2) - \frac{1\cdot 3}{(\pi t)^4}(\kappa^2/2)^2 + \left[\frac{1\cdot 3\cdot 5}{(\pi t)^6} - \frac{1}{(\pi t)^4}\right](\kappa^2/2)^3 - \cdots. \quad (17)$$

The coefficient of $(\kappa^2/2)^n$ in Eq. (17) is

$$(-1)^{n+1}\sum_{s=0}\frac{(-1)^s(n-s-1)!}{s!\,(n-2s-1)!} \times \frac{(2n-2s-1)!!}{(2s+1)!!}\frac{1}{(\pi t)^{2n-2s}}.$$

Similarly, the asymptotic expansion for large real κ is found to be

$$L_t(\kappa^2) = \frac{1}{(\kappa^2/2)} + \frac{1}{(\kappa^2/2)^2} + \frac{1\cdot 3 - (\pi t)^2}{(\kappa^2/2)^3} + \cdots + \frac{d_m}{(\kappa^2/2)^m}, \quad (18)$$

where the coefficients d_m satisfy the recursion relation

$$d_{m+1} = (2m-1)\,d_m - (\pi t)^2\,d_{m-1}.$$

As functions of a complex variable $z = \kappa + iy$, the $L_t(z^2)$ behave very differently on the imaginary axis. With $z = iy$ the L_t function becomes $L_0(-y^2) = (2/y)e^{y^2/4}\,\mathrm{Erf}\,(y/2)$, an increasing exponential for $y \gg 1$. Using the second form of Eq. (7), the asymptotic form of $L_t(-y^2)$ for $t \neq 0$ when $y \gg 1$ is found to be

$$L_t(-y^2) = -(\pi^{1/2}/y)\exp[y^2/4 - (\pi t)^2/y^2]. \quad (19)$$

Two sum rules satisfied by the L_t functions should be mentioned. From the Fourier expansion of the pair propagator Eq. (6), we have for $v = 0$

$$\sum_t L_t(\kappa^2) = 1. \quad (20)$$

Also from the relation

$$\int_0^\beta d\beta' \int_0^{\beta'} d\beta'' G^2(\kappa, \beta' - \beta'') = \tfrac{1}{2}\sum_t \lambda_t^2,$$

we find for Boltzmann statistics

$$\sum_t L_t^2(\kappa^2) = \int_0^1 dv e^{-2\kappa^2 v(1-v)} = L_0(2\kappa^2). \quad (21)$$

The relation Eq. (21) is a consequence of the separability of the kinetic energy of two particles into center-of-mass and relative-motion kinetic energies, i.e., $p_1^2/2m + p_2^2/2m = P^2/4m + p^2/m$ where $\mathbf{P} = (\mathbf{p}_1 + \mathbf{p}_2)/2$, $\mathbf{p} = \mathbf{p}_1 - \mathbf{p}_2$. There is no simple counterpart to Eq. (21) when quantum statistics apply.

The $L_t(\kappa^2)$ functions may be reasonably well approximated on the real axis with such forms as

$$L_0 \cong 1/(1 + a\kappa^2) \quad (22a)$$

$$L_t \cong 2\kappa^2/[\kappa^4 + (2\pi t)^2]. \quad (22b)$$

The constant a must be $1/6$ in order to satisfy (20) and (21) to $O(\kappa^2)$. These approximate forms are, of course, grossly wrong off the real axis. They have pole singularities in the complex plane whereas the $L_t(\kappa^2)$ are entire functions.

In order to evaluate $P_t(\gamma)$ for large γ and $t \neq 0$ the following form of $L_t(\kappa^2)$ as an expansion in powers of $1/\pi t$ is useful:

$$\begin{aligned} L_t(\kappa^2) &= 2\sum_{r=0}\frac{1}{(2r)!(2\pi t)^{2r+1}}\left\{u^{2r}\left(\frac{d}{du}\right)^{4r}\frac{u^{4r+1}}{1+u^2} - \frac{u^{2r+1}}{(2r+1)(2\pi t)}\left(\frac{d}{du}\right)^{4r+2}\frac{u^{4r+3}}{1+u^2}\right\} \\ &= u\left\{\frac{1}{\pi t}\frac{1}{(1+u^2)} + \frac{1}{(\pi t)^2}\frac{u^3-3u}{(1+u^2)^3} + \frac{3}{(\pi t)^3}\frac{u^6-10u^4+5u^2}{(1+u^2)^5} + \cdots\right\}, \end{aligned} \quad (23)$$

where $u = \kappa^2/2\pi t$. Equation (23) is obtained from the series expansion of the $L_t(\kappa^2)$ function, Eq. (17). It reduces to Eq. (17) for $u \ll 1$ and to the asymptotic expansion, Eq. (18), for $u \gg 1$. The first term of Eq. (23) is the approximate form Eq. (22b).

The first term of $P_t(\gamma)$ as defined by Eq. (13) is $O(\gamma^{-1})$ and the coefficient may be evaluated exactly by using the integral representation of $L_t(\kappa^2)$, Eq. (7). The result is

$$\begin{aligned} \int_0^\infty d\kappa L_t(\kappa^2) &= \int_0^\infty d\kappa \int_0^1 dv \exp[-\kappa^2 v(1-v) + 2\pi i t v] \\ &= \frac{\pi^{1/2}}{2}\int_0^\infty \frac{dv e^{2\pi i t v}}{[v(1-v)]^{1/2}} \\ &= (-1)^t(\pi^{3/2}/2)J_0(\pi t) \\ &= \frac{1}{2}\left(\frac{\pi}{t}\right)^{1/2}\left\{1 - \frac{1}{8\pi t} - \frac{1^2\cdot 3^2}{2!(8\pi t)^2} + \cdots\right\}. \end{aligned} \quad (24)$$

If one evaluates the above definite integral using Eq. (23), the asymptotic expansion of the Bessel function given in the final form of Eq. (24) is obtained directly.

IV. METHOD OF INTEGRATION

We are now ready to consider the evaluation of the integrals P_t in Eq. (12). When $t \neq 0$ we note

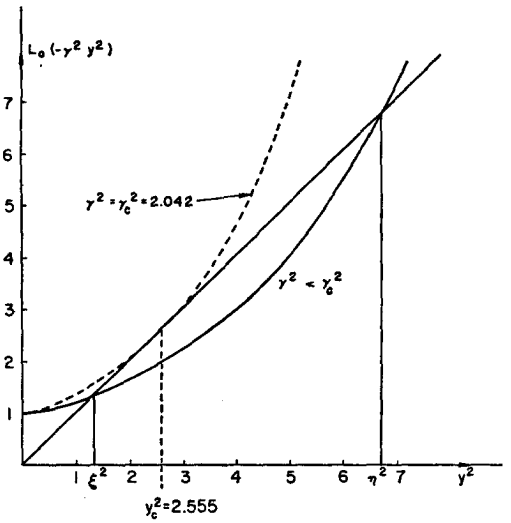


FIG. 2. Graphical solution of Eq. (23).

that $x^{-2}L_t(\gamma^2x^2)$ is finite at $x = 0$, and has the value $1/2(\pi t)^2$. Thus, it is possible to expand the logarithm in powers of $x^{-2}L_t$ and integrate term by term. P_0 is more troublesome, however, because $x^{-2}L_0$ is infinite at $x = 0$. There are several ways to perform integrals of the type P_0 with any function $f(x^2)$ in place of $L_0(\gamma^2x^2)$ subject only to the requirement that $f(x^2)$ vanishes at ∞ on the real axis. The simplest method is an appropriately chosen contour. The choice of the contour is dictated by the location of the roots of the argument of the logarithm in the integrand of P_0, namely, the roots of

$$1 + z^{-2}L_0(\gamma^2z^2) = 0 \tag{25a}$$

with $z = x + iy$. It may be shown that this transcendental equation has only two roots. These roots are pure imaginary when γ is small, so that with $z = iy$ we must solve:

$$y^2 = L_0(-\gamma^2y^2) = (2/\gamma y)\exp(\gamma^2y^2/4)\,\mathrm{Erf}\,(\gamma y/2). \tag{25b}$$

The graphical solution obtained by plotting both sides of Eq. (23) is shown in Fig. 2.

The smaller root ξ^2 may be calculated analytically in powers of γ^2 with the Lagrange inversion formula; it is

$$\xi^2 = \sum_{n=1}^{\infty} a_n\gamma_1^{2(n-1)} = \sum_{n=1}^{\infty}\frac{\gamma^{2(n-1)}}{n!}\left(\frac{d}{du}\right)^{n-1}L_0(-u)^n|_{u=0}. \tag{26}$$

No such expansion is possible for the upper root η^2, which is the solution of $\eta^2 = (\pi^{1/2}/\gamma\eta)\exp(\gamma\eta)^2/4$ when $\gamma \ll 1$. The approximate value is

$$(\gamma\eta)^2 = 4\ln 1/\gamma^2 - O(\ln\ln 1/\gamma^2). \tag{27}$$

In the classical limit the roots are $\xi^2 = 1$ and $\eta^2 = \infty$. As γ becomes larger, the two roots move toward each other and finally coalesce when $\gamma^2 = \gamma_c^2 = 2.042\cdots$. This occurs when $y^2 = y_c^2 = 2.555\cdots$, the point at which $L_0(-\gamma^2y^2)$ is tangent to the straight line at 45° as shown in Fig. 2. For $\gamma > \gamma_c$, the two roots move off the real axis and have the form $\pm\xi_R + i\xi_I$. The asymptotic form of the roots for $\gamma \gg \gamma_c$ may be found by using the asymptotic expansion Eq. (18) for $L_0(z^2)$. This expansion is valid for $z = Re^{i\theta}$ with $-\pi/4 < \theta < \pi/4$ and $3\pi/4 < \theta < 5\pi/4$. The roots are asymptotic to the underside of the rays at $\pi/4$ and $3\pi/4$, and are found to be

$$z = (2\gamma^2)^{1/4}(\pm 1 + i)/\sqrt{2}.$$

The evaluation of $P_0(\gamma)$ is accomplished with the use of

$$\tfrac{1}{2}\int_c dzz^2\{z^{-2}L_0(\gamma^2z^2) - \ln[1 + z^{-2}L_0(\gamma^2t^2)]\}$$

around the contour shown in Fig. 3. The branch cuts are made from 0 to $i\xi$, and from $i\eta$ to ∞. The upper horizontal piece of the contour may be taken anywhere between $i\xi$ and $i\eta$ when $\gamma < \gamma_c$. As the contour goes around the point $i\xi$, the argument of the logarithm has a phase change of $2\pi i$, and consequently the contour segment from 0 to $i\xi$ gives

$$-\tfrac{1}{2}2\pi i\int_{i\xi}^{0} d(iy)(iy)^2 = (\pi/3)\xi^3.$$

In the classical limit, $\gamma = 0$; this is the Debye-Hückel result.

The remaining contribution to P_0 comes from the straight line portion of the contour at height iy. Along this segment it is justified to expand the logarithm in powers of $z^{-2}L_0(\gamma^2z^2)$ since $|z^{-2}L_0(\gamma^2z^2)| < 1$.

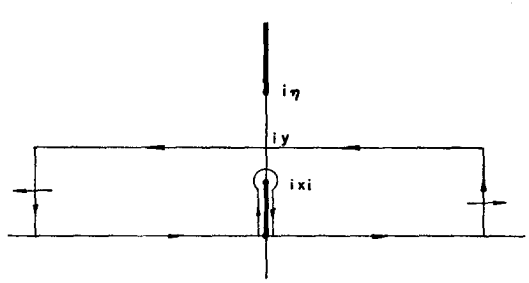


FIG. 3. Contour of integration for the evaluation of $P_0(\gamma)$ when $\gamma < \gamma_c$.

We obtain

$$P_0(\gamma) = \int_0^\infty dx L_0(\gamma^2 x^2) + (\pi/3)\xi^3 - \sum_{n=1} \frac{(-1)^n \gamma^{2(n-1)}}{n} \frac{1}{2} \int_{-\infty+iy}^{\infty+iy} dz \frac{L_0^n(\gamma^2 z^2)}{z^{2n-2}}.$$

Integration by parts gives

$$\int_{-\infty+iy}^{\infty+iy} dz \frac{L_0^n(\gamma^2 z^2)}{z^{2n-2}} = \frac{1}{(2n-3)!} \int_{-\infty}^{\infty} \frac{dx}{(x+iy)} \left(\frac{d}{dx}\right)^{2n-3} L_0^n(\gamma^2(x+iy)^2).$$

This step is justified for any finite y since $L_0(\gamma^2 z^2) \to 0$ as $x \to \pm i\infty$. Finally, the straight-line portion of the contour at height iy may be pushed down to the real axis and we obtain

$$P_0(\gamma) = (\pi/3)\xi^3 + \sum_{n=2} \frac{(-1)^n \gamma^{2n-3}}{n(2n-3)!} \times \int_0^\infty \frac{d\kappa}{\kappa} \left(\frac{d}{d\kappa}\right)^{2n-3} L_0^n(\kappa^2). \quad (28)$$

For $t \neq 0$ the logarithm in P_t may be expanded in powers of $x^{-2}L_t(\gamma^2 x^2)$, and after integrating each term by parts $2n-3$ times one obtains the same form as Eq. (26), but without the $(\pi/3)\xi^3$, i.e.,

$$P_t(\gamma) = \sum_{n=2} b_{nt} \gamma^{2n-3},$$

with

$$b_{nt} = \frac{(-1)^n}{n(2n-3)!} \int_0^\infty \frac{d\kappa}{\kappa} \left(\frac{d}{d\kappa}\right)^{2n-3} I_t^n(\kappa^2). \quad (29)$$

Consequently

$$P(\gamma) = (\pi/3)\xi^3 + \sum_{n=2} b_n \gamma^{2n-3}, \quad (30)$$

with

$$b_n = \sum_t b_{nt}.$$

It should be noted that the method used here for the evaluation of $P_0(\gamma)$ is not limited to Boltzmann statistics. The final results Eqs. (26) and (28) apply also to the ring sum with quantum statistics. In place of the $L_t(\kappa^2)$ functions one must use the more complicated expressions for the Fourier components of the pair interaction propagator with statistics included. The evaluation of the ring sum with quantum statistics will be considered in a later paper.

The integrand of $P_0(\gamma)$ is an analytic function of γ^2. The function $P_0(\gamma)$ obtained by integration is not analytic in γ^2 since the result, Eq. (28), has both odd and even powers of γ. The odd powers of γ must be interpreted as $(\gamma^2)^{1/2}$. The portion of $P_0(\gamma)$ analytic in γ^2, namely, $(\pi/3)\xi^3$, obviously comes from the lower of the two roots of Eq. (25b). The nonanalytic portion of $P_0(\gamma)$ comes from the horizontal piece of the contour at height iy above the real axis. It may be suspected that this portion, $\sum b_{n0}\gamma^{2n-3}$ in Eq. (26), has something to do with the upper root at $z = i\eta$. In order to clarify this point, one notes that the contour used for the integration of $P_0(\gamma)$ works also when $L_0(\gamma^2 x^2)$ is replaced by any function $f(\gamma^2 x^2)$, entire or not, which vanishes as $x \to \infty$. For example, if we use the approximate form Eq. (22a), i.e., $f(\gamma^2 x^2) = 1/(1+\gamma^2 x^2)$, we find

$$\int_0^\infty x^2\, dx\{x^{-2} f(\gamma^2 x^2) - \ln[1 + x^{-2} f(\gamma^2 x^2)]\}$$
$$= (\pi/3)\xi_1^3(\gamma^2) + \sum_{n=2} \frac{(-1)^n \gamma^{2n-3}}{n(2n-3)!} \times \int_0^\infty \frac{d\kappa}{\kappa} \left(\frac{d}{d\kappa}\right)^{2n-3} f^n(\kappa^2),$$
$$= (\pi/3)\{\xi_1^3(\gamma^2) + \xi_2^3(\gamma^2) - 1/\gamma^3 + 3/2\gamma\}, \quad (31)$$

where ξ_1^2 and ξ_2^2 are, respectively, the lower and upper roots of $y^2 = f(-\gamma^2 y^2)$, namely,

$$(1/2\gamma^2)[1 \pm (1 - 4\gamma^2)^{1/2}].$$

The second line of Eq. (29) is the result obtained with the contour in Fig. 3, and the third line is the exact result. The series in γ^{2n-3} in the second line is seen to be an expansion of $\xi_2^3(\gamma^2)$ beginning with the $O(\gamma)$ term.

Similarly $P_t(\gamma)$ for $t \neq 0$ may be evaluated approximately by using the approximate form Eq. (22b) for $L_t(\gamma^2 x^2)$. The result is

$$P_t(\gamma) \cong \frac{1}{\gamma^3}\left\{\frac{\pi^{1/2}\gamma^2}{2t^{1/2}} - \frac{2^{1/2}\pi}{3}[(2\gamma^2 + 4\pi^2 t^2)^{3/4} - (2\pi t)^{3/2}]\right\}$$
$$= \frac{\pi}{2^{3/4}} \sum_{n=2} \frac{(-1)^n \gamma^{2n-3}}{(2^{1/2}\pi t)^{2n-3/2}} \frac{(n-\frac{7}{4})(n-\frac{11}{4})\cdots\frac{1}{4}}{n!} \quad (32)$$

One sees from the second form of Eq. (32) that the series expansion of $P_t(\gamma)$ converges for $\gamma < 2^{1/2}\pi t$. The estimate of b_{nt} obtained from Eq. (32), however, is poor for small t because the approximation for Eq. (22b) underestimates $L_t(\gamma^2 x^2)$ in the region $\gamma x \sim \pi t$. For $\gamma \gg 2^{1/2}\pi t$ the first form of Eq. (32) is fairly good. The coefficient of the $O(\gamma^{-1})$ term is correct to $O(t^{-1/2})$ as may be seen by comparison

with the exact result Eq. (24). The coefficient of the $O(\gamma^{-3/2})$ term is exact for all values of t.

V. EXACT RESULTS FOR BOLTZMANN STATISTICS

The coefficients a_n defined by Eq. (26) for the expansion of $\xi^2(\gamma^2)$ are relatively easy to evaluate for small n since they are obtained by differentiation. The coefficients b_{nt} in Eqs. (29) and (30) are more difficult to evaluate since integration is required. Exact results for a_n and b_n may be calculated for small n, but for large n only asymptotic expressions can be obtained.

The expression for a_n is obtained with the use of the integral definition of the $L_0(\kappa^2)$ function, Eq. (7). The result is

$$\begin{aligned} a_n &= \frac{1}{n!}\left(\frac{d}{du}\right)^{n-1} L_0^n(-u)|_{u=0}, \\ &= \frac{1}{n!}\left(\frac{d}{du}\right)^{n-1} \int_0^1 \cdots \int dv_1 \cdots dv_n \\ &\times \exp u[v_1(1-v_1) + \cdots + v_n(1-v_n)], \qquad (33) \\ &= \frac{n^{n-1}}{2^{2(n-1)}n!} \int_0^1 \cdots \int ds_1 \cdots ds_n \\ &\times [1 - (s_1^2 + \cdots + s_n^2)/n]^{n-1}. \end{aligned}$$

The final form of a_n is obtained with the change of variable $v = s/2 + \frac{1}{2}$. The multiple integral in this final form is over the volume of an n-dimensional cube. No exact answer is known, but in the Appendix it is shown that the multiple integral is asymptotic to θ_0^n where

$$\begin{aligned} \theta_0 &= A_0(1-\phi) = 0.7178 \cdots, \\ \phi &= A_1/A_0 - \ln(1 - A_1/A_0) = 0.0389 \cdots, \qquad (34) \\ A_m &= \int_0^1 ds s^{2m} e^{-s^2}. \end{aligned}$$

Using this result, Eq. (30), and the Stirling approximation for $n!$ gives

$$a_n = \frac{4}{(2\pi)^{1/2}} \frac{1}{n^{3/2}} \left(\frac{e\theta_0}{4}\right)^n \qquad (35)$$

for large n. Including the exact values for the first few a_n, the result for $\xi^2(\gamma^2)$ from Eq. (24) is

$$\begin{aligned} \xi^2(\gamma^2) = 1 &+ \frac{1}{6}\gamma^2 + \frac{2}{45}\gamma^4 + \frac{13}{480}\gamma^6 + \cdots \\ &+ \frac{4}{(2\pi)^{1/2}} \frac{\gamma^{2(n-1)}}{n^{3/2}\gamma_c^{2n}} + \cdots, \qquad (36) \end{aligned}$$

where $\gamma_c^2 = 4/e\theta_0 = 2.042 \cdots$. The series expansion Eq. (36) for ξ^2 is thus convergent for $\gamma^2 < \gamma_c^2$.

In a similar fashion the coefficient b_{nt} from Eq. (29) is found to be

$$\begin{aligned} b_{nt} &= \frac{(-1)^{n+1}}{n(2n-3)!} \int_0^\infty \frac{d\kappa}{\kappa} \left(\frac{d}{d\kappa}\right)^{2n-3} \\ &\times \int_0^1 \cdots \int dv_1 \cdots dv_n \\ &\times \exp\left[2\pi i t \sum_{i=1}^n v_i - \kappa^2 \sum_{i=1}^n v_i(1-v_i)\right] \qquad (37) \\ &= (-1)^{nt} \frac{\pi^{1/2}}{2} \frac{2^n(n-2)!n^{n-3/2}}{n(2n-3)!} \\ &\times \int_{-1}^1 \cdots \int \frac{ds_1 \cdots ds_n}{2^n} \\ &\times \exp(\pi i t \sum s_i)\left[1 - \frac{\sum s_i^2}{n}\right]^{n-3/2}. \end{aligned}$$

The second form of Eq. (37) is obtained by interchanging the order of the integrations and using

$$\int_0^\infty \frac{d\kappa}{\kappa}\left(\frac{d}{d\kappa}\right)^{2n-3} e^{-a\kappa^2} = (-1)^{n+1}\pi^{1/2}2^{2(n-2)}(n-2)!a^{n-3/2}.$$

To obtain b_n we sum over all t in the first form of Eq. (37) and use the periodic delta function

$$\sum_{t=-\infty}^{\infty} \exp(2\pi i t \sum v_i) = \delta(r - \sum v_i),$$

where r can, in general, be any integer. In the unit volume hypercube over which the multiple integration in Eq. (37) is to be performed, the possible values of r are $1, 2, \cdots, n-1$. Hence for b_n we find

$$\begin{aligned} b_n &= \sum_t b_{nt} \\ &= \frac{\pi^{1/2}}{2} \frac{2^n(n-2)!}{n(2n-3)!} \sum_{r=1}^{n-1} \cdots \int_0^1 ds_1 \cdots ds_n \\ &\times \delta(r - n/2 - \tfrac{1}{2}\sum s_i)[n - \sum s_i^2]^{n-3/2}. \qquad (38) \end{aligned}$$

The multiple integral in Eq. (37) for b_{nt} is the same type that occurred in the a_n as given by Eq. (37). The difference in the exponents, $n-1$ in Eq. (33) and $n - \frac{3}{2}$ in Eq. (37), is of no consequence when $n \to \infty$. Using the estimate of the multiple integral in the Appendix the result for b_{n0} is found to be

$$\begin{aligned} b_{n0} &= \frac{\pi^{1/2}}{2} \frac{(n-2)!n^{n-5/2}\theta_0^n}{(2n-3)!} \\ &\approx \frac{2^{3/2}\pi^{1/2}}{n^{3/2}} \frac{1}{\gamma_c^{2n}} = \pi a_n. \end{aligned} \qquad (39)$$

For $t \neq 0$ the method of the Appendix gives $(-1)^{n(t+1)}(2/e\pi^2t^2)^n$ for the multiple integral and

is valid when $n^{1/2}/\pi t > 1$; hence the approximate result for small t is

$$b_{nt} = (-1)^n 2^{3/2}\pi^{1/2}/n^{3/2}(2^{1/2}\pi t)^{2n}. \tag{40}$$

When $n^{1/2}/\pi t < 1$, then b_{nt} may be calculated by using the expansion, Eq. (23), for $L_t(\kappa^2)$ with the result:

$$\begin{aligned} b_{nt} &= \frac{(-1)^n}{2^{n+1/2}\pi^{2n-3/2}t^{2n-3/2}} \frac{\Gamma(\frac{3}{4})\Gamma(n-\frac{3}{4})}{n!} \\ &\quad \times \left\{1 - \frac{1}{\pi t}\frac{(3n-2)}{(n+1)}\frac{\Gamma(\frac{5}{4})\Gamma(n-\frac{1}{4})}{\Gamma(\frac{3}{4})\Gamma(n-\frac{3}{4})}\cdots\right\} \\ &= \frac{(-1)^n\Gamma(\frac{3}{4})}{2^{5/4}n^{7/4}} \frac{1}{(2^{1/2}\pi t)^{2n-3/2}} \\ &\quad \times \left\{1 - \frac{3\Gamma(\frac{5}{4})}{\Gamma(\frac{3}{4})}\frac{n^{1/2}}{\pi t}\cdots\right\}. \end{aligned} \tag{41}$$

It is clear from Eqs. (40) and (41) that for large n the contribution from $t \neq 0$ vanishes. Thus the nonstatic part in the ring sum, i.e., for $\omega_t = 2\pi it/\hbar\beta \neq 0$, affects only the coefficients b_n for small n.

Exact results have been obtained for $n = 2$ and 3. The evaluation of b_2 is very easy, essentially because of the sum rule for $L^2(\kappa^2)$, Eq. (21). The result is

$$b_2 = -\tfrac{1}{2}\int_0^\infty \frac{dk}{\kappa}\frac{d}{dk}L_0(2\kappa^2) = \frac{2^{1/2}\pi^{3/2}}{16} = 0.491. \tag{42}$$

The value of b_{20} obtained from Eq. (37) is

$$b_{20} = (\pi^{3/2}/24)(5 - 2\sqrt{2}) = 0.543.$$

Thus, for $n = 2$ the contribution from nonstatic terms, i.e., all b_{2t} with $t \neq 0$, is about 10.5%. The result for b_3 was obtained after a great deal of elementary and tedious integration; it is

$$\begin{aligned} b_3 &= \frac{4\pi^{1/2}}{9}\int_0^1\!\!\int\!\!\int dv_1\, dv_2\, dv_3\, \delta(1 - v_1 - v_2 - v_3) \\ &\quad \times [v_1(1-v_1) + v_2(1-v_2) + v_3(1-v_3)]^{3/2}, \\ &= 2^{7/2}\pi^{3/2}/3^5 = 0.259. \end{aligned} \tag{43}$$

It appears that as a consequence of the complicated geometry of the hypercube, the labor required to evaluate b_n exactly increases with n much faster than exponentially.

Combining the results of this section gives the complete result for $P(\gamma)$ as

$$\begin{aligned} P(\gamma) &= (\pi/3) \\ &\times \left[1 + \tfrac{1}{6}\gamma^2 + \cdots + \frac{2^{3/2}}{(2\pi)^{1/2}}\frac{\gamma^{2(n-1)}}{n^{3/2}\gamma_c^{2n}} + \cdots\right]^{3/2} \\ &- (8\pi)^{1/2}\left[\frac{\pi}{2^5}\gamma + \frac{2\pi}{3^5}\gamma^3 + \cdots + \frac{\gamma^{2(n-1)}}{n^{3/2}\gamma_c^{2n}} + \cdots\right]. \end{aligned} \tag{44}$$

The function $P(\gamma)$ decreases monotonically from $\pi/3$ as γ increases. The expression (39) for $P(\gamma)$ is valid for $\gamma < \gamma_c$ for which the two series converge. At $\gamma = \gamma_c$ the two roots ξ^2 and η^2, which determine the value of $P_0(\gamma)$, coalesce and the contour integration method fails.

The coefficients b_2 and a_2 have previously been obtained by Sakakura in his thesis[4] and more recently by Stephen.[10] The function $P(\gamma)$ decreases monotonically from $\pi/3$ as γ increases. Equation (44) is valid when $\gamma < \gamma_c$ for which the two series converge. At $\gamma = \gamma_c$ the two roots, ξ^2 and η^2, which determine the value of $P_0(\gamma)$, coalesce and, thereafter, become complex. The contour integration method then fails, and the form of $P(\gamma)$ for large γ must be obtained by a different method which is described in the next section.

The methods used to obtain Eq. (44) for the single-component gas may be easily extended to the ring sum for the multicomponent gas. The coefficients a_n and b_n become functions of the ratios of the particle masses. Consider a two-component gas with electrons of mass m_e and charge z_e, and ions of mass m_i and charge z_i. Electrical neutrality requires that $z_e\rho_e + z_i\rho_i = 0$. The polarizability of this plasma using Eq. (8) is

$$4\pi\alpha = [z_e^2\alpha_e L_t(\gamma_e^2x^2) + z_i^2\alpha_i L_t(\gamma_i^2x^2)]/\langle z^2\rangle\kappa^2, \tag{45}$$

where

$$\begin{aligned} x &= k\lambda_D, \\ \langle z^2\rangle &= z_e^2\alpha_e + z_i^2\alpha_i, \\ \alpha_e &= \rho_e/\rho, \qquad \gamma_e = \lambda\!\!\!^{-}_e/\lambda_D, \\ \alpha_i &= \rho_i/\rho, \qquad \gamma_i = \lambda\!\!\!^{-}_i/\lambda_D. \end{aligned}$$

The particle masses occur only in the thermal wavelengths, $\lambda\!\!\!^{-}_e$ and $\lambda\!\!\!^{-}_i$. The multicomponent form of Eq. (13) when evaluated is

$$\begin{aligned} P(\gamma_e, \gamma_i) &= (\pi/3)\left[1 + \sum_{n=2} a_n(m_e/m_i)\gamma_e^{2(n-1)}\right]^{3/2} \\ &\quad - \sum_{n=2} b_n(m_e/m_i)\gamma_e^{2n-3}, \end{aligned} \tag{46}$$

[10] M. J. Stephen, Proc. Roy. Soc. (London) **A265**, 215 (1962).

with

$$a_n(m_e/m_i) = \frac{1}{n!}\left(\frac{d}{du}\right)^{n-1} \times \left[\frac{z_e^2\alpha_e L_0(-u) + z_i^2\alpha_i L_0[-(m_e/m_i)u]}{\langle z^2\rangle}\right]^n\Bigg|_{u=0}$$

$$b_n(m_e/m_i) = \frac{(-1)^n}{n(2n-3)!}\int_0^\infty \frac{d\kappa}{\kappa}\left(\frac{d}{d\kappa}\right)^{2n-3} \times \sum_t \left[\frac{z_e^2\alpha_e L_t(\kappa^2) + z_i^2\alpha_i L_t[(m_e/m_i)\kappa^2]}{\langle z^2\rangle}\right]^n.$$

The explicit results for a_2 and b_2 are

$$a_2(m_e/m_i) = (1/6)[z_e^4\alpha_e^2 + z_e^2 z_i^2\alpha_e\alpha_i(1 + m_e/m_i) + z_i^4\alpha_i^2 m_e/m_i]/\langle z^2\rangle^2,$$
$$b_2(m_e/m_i) = 2^{-7/2}\pi^{3/2}[z_e^4\alpha_e^2 + 2^{1/2}z_e^2 z_i^2\alpha_e\alpha_i \times [\tfrac{1}{2}(1 + m_e/m_i)^{1/2} + z_i^4\alpha_i^2(m_e/m_i)^{1/2}]/\langle z^2\rangle^2. \quad (47)$$

VI. THE ASYMPTOTIC EXPANSION OF $P(\gamma)$ FOR $\gamma \gg \gamma_c$

When γ is very small most of the ring sum contribution to the free energy comes from the static polarizability, i.e., $\omega_t = 0$. As increases the contribution from the nonstatic part, $\omega_t \neq 0$, also increases and becomes dominant when $\gamma \gg \gamma_c$. The asymptotic form of $P(\gamma)$ for large γ may be obtained correctly by summing the approximate form of $P_t(\gamma)$, Eq. (32), over all t. However, in order to obtain the asymptotic expansion in powers of γ and hence the low-temperature form of F_{ring} in powers of kT, we shall use the Mellin transform method suggested by Iwata[11] for the evaluation of $P_t(\gamma)$ and finally $P(\gamma)$.

In terms of the variable $u = \kappa^2/2\pi t$ Eq. (13) for $P_t(\gamma)$ may be rewritten as

$$P_t(\gamma) = \frac{(2\pi t)^{3/2}}{2\gamma^3}\int_0^\infty u^{1/2}\,du \times \left\{\frac{\gamma^2 L_t}{2\pi t u} - \ln\left(1 + \frac{\gamma^2 L_t}{2\pi t u}\right)\right\}$$
$$= \frac{(2\pi t)^{3/2}}{2\gamma^3}\int_C \frac{(-)\,ds}{2\pi i}\frac{\pi}{s\sin\pi s} \times \int_0^\infty u^{1/2}\,du\left(\frac{\gamma^2 L_t}{2\pi t u}\right)^s. \quad (48)$$

The contour C for the s integration runs from $\sigma - i\infty$ to $\sigma + i\infty$ with $1 < \sigma < 2$. If the s contour is closed to the right then one obtains the series expansion of $P_t(\gamma)$ in powers of γ appropriate for $\gamma < 2^{1/2}\pi t$, and if the contour is closed to the left one obtains the asymptotic expansion of $P_t(\gamma)$ in inverse powers of γ appropriate for $\gamma > 2^{1/2}\pi t$. By using Eq. (23) for $L_t(\kappa^2)$ expanded in powers of $1/\pi t$ the u integration in the second form of Eq. (48) may be performed to give

$$\frac{1}{2}\left(\frac{\gamma^2}{2\pi^2 t^2}\right)\left\{\frac{\Gamma(\frac{3}{4})\Gamma(s - \frac{3}{4})}{\Gamma(s)} - \frac{s(3s-2)}{\pi t}\frac{\Gamma(\frac{5}{4})\Gamma(s - \frac{1}{4})}{\Gamma(s+2)} + \cdots\right\}. \quad (49)$$

With s equal to positive integer values, one obtains Eq. (41) for b_{nt} from Eq. (49).

The summation over integers t may now be done exactly in terms of the zeta function, $\sum_{t=1} t^{-s} = \zeta(z)$, to give

$$P_{t\neq 0}(\gamma) = 2\sum_{t=1} P_t(\gamma) = \frac{1}{\gamma^3}\int_c \frac{(-)\,ds}{2\pi i}\frac{\pi}{s\sin\pi s}\frac{\gamma^{2s}}{2^{s-1/2}\pi^{2s-3/2}} \times \left\{\frac{\Gamma(\frac{3}{4})\Gamma(s-\frac{3}{4})\zeta(2s-\frac{3}{2})}{\Gamma(s)} - \frac{s(3s-2)\Gamma(\frac{5}{4})\Gamma(s-\frac{1}{4})\zeta(2s-\frac{1}{2})}{\pi\Gamma(s+2)} + \frac{s(9s^3+24s^3-58s+7)}{2\pi^2} \times \frac{\Gamma(\frac{7}{4})\Gamma(s+\frac{1}{4})\zeta(2s+\frac{1}{2})}{\Gamma(s+4)} + \cdots\right\}. \quad (50)$$

Because of the appearance of the singularity of the zeta function at $s = \frac{5}{4}$, the s contour must now cross the real axis in the interval $\frac{5}{4} < \sigma < 2$. The integrand of Eq. (50) has simple poles at $s = \frac{5}{4}, 1, \frac{3}{4}, \frac{1}{4}, 0, -\frac{1}{4}, -\frac{3}{4}, \cdots$, and consequently the asymptotic expansion has the form

$$P_{t\neq 0} = \frac{c_1}{\gamma^{1/2}} + \frac{c_2'}{\gamma} + \frac{c_3'}{\gamma^{3/2}} + \frac{c_5'}{\gamma^{5/2}} + \cdots,$$

where the primes on the coefficients c_2', c_3', $\cdots$ indicate the absence of the contribution from $t = 0$. The asymptotic expansion of $P_0(\gamma)$ may be found by writing $L_0(\kappa^2)$ in a form analogous to Eq. (23) for $L_t(\kappa^2)$, namely,

$$L_0(\kappa^2) = \frac{1}{1 + z^2/2} + \frac{z^2}{(1 + z^2/2)^3} + \frac{2z^4 - 2z^2/3}{(1 + z^2/2)^5} + \cdots$$

from which we obtain

[11] G. Iwata, Progr. Theoret. Phys. (Kyoto) 24, 1118 (1960).

$$P_0(\gamma) = \frac{1}{\gamma^3}\int_c \frac{(-)\,ds}{2\pi i}\frac{\pi}{s\sin\pi s}\frac{\gamma^{2s}}{2^{s-1/2}} \times\left\{\frac{\Gamma(\frac{3}{2}-s)\Gamma(2s-\frac{3}{2})}{\Gamma(s)} + \frac{2s\Gamma(\frac{5}{2}-s)\Gamma(2s-\frac{1}{2})}{\Gamma(s+2)}+\cdots\right\}. \quad (51)$$

From the location of the simple poles in the integrand of Eq. (51) at $s = 1, \frac{3}{4}, \frac{1}{4}, 0, -\frac{1}{4}, \cdots$ one sees the form of the asymptotic expansion of $P_0(\gamma)$, and by calculating the residues at these poles the result is obtained as

$$P_0(\gamma) = \frac{c_2''}{\gamma}+\frac{c_3''}{\gamma^{3/2}}+\frac{c_5''}{\gamma^{5/2}}+\cdots = \frac{\pi^{3/2}/2}{\gamma}-\frac{2^{5/4}\pi/3}{\gamma^{3/2}}-\frac{\pi/2^{1/4}}{\gamma^{5/2}}+\cdots. \quad (52)$$

The residue at $s = 1$ giving c_2'' is given by Eq. (24).

The coefficients c_1, c_2', c_3', etc., are obtained by calculating the residues of the integrand of Eq. (50). We will first consider the residue at $s = 1$ which may be obtained exactly with the help of Eq. (24):

$$c_2' = \pi^{1/2}\left[\zeta(\tfrac{1}{2})-\frac{\zeta(\frac{3}{2})}{8\pi}-\cdots\right] = \pi^{1/2}\zeta(\tfrac{1}{2})+\pi^{3/2}\times\sum_{t=1}^{\infty}(-1)^t[J_0(\pi t)-(-1)^t/\pi t^{1/2}]. \quad (53)$$

The t summation in c_2' may be converted to a definite integral by using Eq. (24), the first term of the $L_t(\kappa^2)$ expansion in Eq. (23), and the sum rule Eq. (20). One finds

$$\pi^{3/2}\sum_{t=1}(-1)^t[J_0(\pi t)-(-1)^t/\pi t^{1/2}] = \int_0^\infty d\kappa\left\{\sum_{t=-\infty}^{\infty}L_t(\kappa^2)-L_0(\kappa^2)-2\sum_{t=1}\frac{2\kappa^2}{\kappa^4+(2\pi t)^2}\right\} = -\pi^{3/2}/2+\int_0^1 d\kappa\{1-(\coth\kappa^2/2-2/\kappa^2)\} = -\pi^{3/2}/2-\pi^{1/2}\zeta(\tfrac{1}{2}).$$

Hence c_2' is exactly

$$c_2' = -\pi^{3/2}/2 \quad (54)$$

and the complete coefficient of the γ^{-1} term of $P(\gamma)$, $c_2 = c_2' + c_2''$, is identically zero.

The asymptotic expansion of $P(\gamma)$ is obtained by adding the results for $P_0(\gamma)$ and $P_{t\neq 0}(\gamma)$ and has the form

$$P(\gamma) = \frac{c_1}{\gamma^{1/2}}+\frac{c_3}{\gamma^{3/2}}+\frac{c_5}{\gamma^{5/2}}+\cdots. \quad (55)$$

The coefficient c_1 is obtained as the residue of the first term in the brackets of the integrand of Eq. (50). The coefficients $c_3 = c_3' + c_3''$, etc., have contributions from terms in the integrands of both Eqs. (50) and (51). The results are

$$c_1 = \frac{2^{3/4}}{5}\frac{\Gamma(\frac{3}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{5}{4})} = 0.806$$

$$c_3 = \frac{2^{5/4}}{3}\frac{3}{16}\frac{\Gamma(\frac{5}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{11}{4})} = 0.158$$

$$c_5 = \frac{2^{7/4}}{1}\frac{375}{512}\frac{\Gamma(\frac{7}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{17}{4})} = 0.485. \quad (56)$$

From the form of the integrands of Eqs. (50) and (51) it is clear that there are no terms of order γ^{-2m} in the expansion of $P(\gamma)$. From the fact that c_2 for the γ^{-1} term is zero, it is conjectured that all the coefficients of terms of order $\gamma^{-(2m+1)}$ are zero, but a proof has not been found.

The asymptotic expansion of $P(\gamma)$ for large γ is of interest for low temperature, particularly if there is a density region for which the ring sum gives the dominant contribution to the free energy as it does in the case of Fermi statistics. Our result Eqs. (55) and (56) would then be analogous to the correlation energy and heat capacity for the high-density electron gas at zero temperature as obtained by Gell-Mann and Brueckner.[2] The free energy may be expressed in terms of the standard parameter r_s defined by

$$(\rho a_0^3)^{-1} = \tfrac{4}{3}\pi r_s^3,$$

where a_0 is the Bohr radius. The dimensionless parameters Λ and γ in terms of r_s are

$$\Lambda = 2\cdot 6^{1/2}(\mathrm{Ryd}/kT)^{3/2}r_s^{-3/2}, \qquad \gamma = 6^{1/2}(\mathrm{Ryd}/kT)r_s^{-3/2} \quad (57)$$

and the ring contribution to the free energy becomes

$$F_{\mathrm{ring}}/\mathrm{Ryd} = -N(kT/\mathrm{Ryd})(\Lambda/\pi)P(\gamma) = -N\frac{2\cdot 6^{1/4}}{\pi}\left\{c_1 r_s^{-3/4}+\frac{c_3}{6^{1/2}}\left(\frac{kT}{\mathrm{Ryd}}\right)r_s^{3/4}+\frac{c_5}{6}\left(\frac{kT}{\mathrm{Ryd}}\right)^2 r_s^{9/4}+\cdots\right\}. \quad (58)$$

The correlation energy obtained from Eq. (58) with the relation $E = (\partial/\partial\beta)(\beta F)$ is

$$E_{\mathrm{ring}}/\mathrm{Ryd} = N\{-0.803 r_s^{-3/4}+0.081(kT/\mathrm{Ryd})^2 r_s^{9/4}\cdots\}, \quad (59)$$

and the heat capacity is

$$C = C_0 + C_{\text{ring}}$$
$$= Nk\{\tfrac{3}{2} + 0.162(kT/\text{Ryd})r_s^{9/4} \cdots\}. \quad (60)$$

The first term of Eq. (59) has also been obtained by Stephen for the gas of distinguishable electrons.[12] It should be noted that these results predict that the correlation energy diverges in the high-density limit, $r_s \to 0$, whereas the heat capacity contribution from the ring sum goes to zero.

The pressure obtained from Eq. (58) with the relation $PV = \rho(\partial/\partial\rho)(\beta F)$ is

$$PV = (P_0 + P_{\text{ring}})V = N\left\{kT - \text{Ryd}\,\frac{6^{1/4}}{2\pi}\left[c_1 r_s^{-3/4} - \frac{c_3}{6^{1/2}}\left(\frac{kT}{\text{Ryd}}\right)r_s^{3/4} - \frac{c_5}{2}\left(\frac{kT}{\text{Ryd}}\right)^2 r_s^{9/4} \cdots\right]\right\}$$
$$= N\text{Ryd}\{-0.201 r_s^{-3/4} + (kT/\text{Ryd})(1 + 0.0161 r_s^{3/4}) - 0.060(kT/\text{Ryd})^2 r_s^{9/4} \cdots\}. \quad (61)$$

The $r_s^{-3/4}$ term in Eq. (61) is dominant and consequently this pressure expression is necessarily negative under the condition of validity of the asymptotic expansion of $P(\gamma)$, namely, $\gamma \gg \gamma_c$. At zero temperature, the pressure is

$$P/\rho = \tfrac{1}{4}E_{\text{ring}}/N = -0.201 r_s^{-3/4}\,\text{Ryd}. \quad (62)$$

Thus if the ring sum approximation, i.e., the random phase approximation, is valid for Boltzmann statistics (and also for Bose statistics) at high density and zero temperature as it is for Fermi statistics, then the gas of distinguishable charged particles should collapse.

VII. CONCLUDING REMARKS

The main result of this article is Eq. (44) in which the diffraction corrections to the classical ring sum are exhibited as an expansion in powers of γ. The first thing to notice about this result is that the diffraction corrections involve both even and odd powers, hence even and odd powers of $\hbar$. This result apparently contradicts the form of the Wigner-Kirkwood (WK) expansion[13] which is commonly used for the calculation of diffraction corrections at high temperature to the equation of state of nonideal gases.[14] According to the WK expansion, the diffraction corrections appear to involve only even powers of Planck's constant. It was pointed out in a previous paper, however, that for potentials with a $1/r$ singularity at the origin, the nonanalytic form $(\hbar^2)^{1/2}$ is to be expected.[15] The reason is that for the $1/r$ potential, the parameter of smallness for the expansion is $(\lambda\!\!\!^{-}/\beta e^2)^2$ which depends linearly on the temperature. Since the parameter of smallness diverges at high temperature, the WK expansion breaks down.

It was also pointed out in the previous paper that the quantum-mechanical ring sum for the Coulomb potential is rather similar in form to the second-order perturbation term of the second virial coefficient for the screened Coulomb potential, $(e^2/r)\exp(-r/r_0)$. The classical form of this second-order term comparable to S_{ring} is

$$\rho B_{22} = (1/2!)\rho \int d^3r(-\beta u)^2.$$

The result with quantum corrections [Eq. (32) of Foldy[16]] is

$$\rho B_{22} = (\pi/2)\rho r_0(\beta e^2)^2 F(\gamma_1), \quad (63)$$
$$F(\gamma_1) = [1 + \gamma_1 e^{\gamma_1^2/4}\,\text{Erf}\,(\gamma_1/2)] - (\pi^{1/2}/2)\gamma_1 e^{\gamma_1^2/4},$$

with $\gamma_1 = \lambda\!\!\!^{-}/r_0$, $\lambda\!\!\!^{-} = \hbar(\beta/2\mu)^{1/2}$, and $\mu = m/2$. The function $F(\gamma_1)$ for the quantum corrections to second-order perturbation theory should be compared with $(3/\pi)P(\gamma)$ for the quantum corrections in the ring sum. In both cases the functions may be written as two power series, one with even powers of γ_1 and γ, respectively, and the other with odd powers of γ_1 and γ. Thus, the function in square brackets in Eq. (63) has the expansion:

$$1 + \frac{1}{2}\gamma_1^2 + \frac{1}{12}\gamma_1^4 + \cdots + \frac{(\gamma_1^2/2)^n}{2(2n+1)!!} + \cdots,$$

which should be compared with the expansion of $\xi^3(\gamma^2)$ from Eq. (36). The comparison of the quantum corrections to ρB_{22} with those in S_{ring} has been pointed out because second-order perturbation theory is sufficiently simple that the quantum corrections, $F(\gamma_1)$, to ρB_{22}, could be obtained in terms of known functions. The quantum corrections, $(3/\pi)P(\gamma)$, to S_{ring}, however, could only be evaluated as the series expansion, Eq. (44).

The functions $F(\gamma_1)$ and $(3/\pi)P(\gamma)$ which give the quantum corrections to ρB_{22} and S_{ring} due to the uncertainty principle are rather similar in form

[12] M. J. Stephan [Proc. Phys. Soc. (London) 79, 994 (1962)], suggested the use of the Mellin transformation for the evaluation of the ring sum at zero temperature. Since he used the approximation Eq. (22b) for $L_l(k^2)$ and approximated the t summation by an integral he could obtain only the first term of the asymptotic expansion of $P(\gamma)$.

[13] E. P. Wigner, Phys. Rev. 40, 747 (1932); J. G. Kirkwood, *ibid.* 44, 31 (1933).

[14] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), Chap. 6.

[15] H. E. DeWitt, J. Math. Phys. 3, 1003 (1962).

[16] L. L. Foldy, Phys. Rev. 124, 649 (1961).

at least for small γ_1 and γ. In the classical limit one is considering charges localized at points. With $\hbar \neq 0$, however, the wave functions of the particles describe packets that extend a distance $\lambda = \hbar/\langle p \rangle = \hbar/(2mkT)^{1/2}$ around each charge center. This spreading of the point charges over a finite region causes the functions $F(\gamma_1)$ and $(3/\pi)P(\gamma)$ to be monotonically decreasing as the temperature is decreased. The differences between $F(\gamma_1)$ and $(3/\pi)P(\gamma)$ become apparent for large γ_1 and γ. The power series expansions for the two pieces of $F(\gamma_1)$ have radii of convergence of ∞, whereas the power series expansions of the two series for $(3/\pi)P(\gamma)$ converge only for $\gamma < \gamma_c$. For $\gamma_1 \gg 1$ the asymptotic form of $F(\gamma_1)$ is $2/\gamma_1^2$, while the asymptotic form of $(3/\pi)P(\gamma)$ is $(3c_1/\pi)/\gamma^{1/2}$. In both cases these asymptotic forms lead to nonzero contributions to the free energy at zero temperature.

The ring sum contribution to the free energy of distinguishable charged particles at low temperature is of interest for comparison with similar results for charged bosons. At zero temperature charged particles obeying either Boltzmann or Bose statistics must have the same ground state energy since all particles are in the lowest momentum state. Thus the $r_s^{-3/4}$ term of Eq. (58) may be compared with recent results for the correlation energy of charged bosons. The excitation spectra and hence the heat capacities for the two kinds of statistics are presumably quite different. Foldy, by using a method due to Bogoliubov, has obtained the following result for the ground state energy of high density bosons[16]:

$$E/N \text{ Ryd} = -0.803 r_s^{-3/4} + 0.424$$

and more recently Girardeau using a variational method has obtained Foldy's result, but with an additional term which diverges at high density, $\frac{1}{8} \ln (1/r_s)$.[17] Since the $r_s^{-3/4}$ term obtained by Foldy and Girardeau is identical with our result, Eqs. (58) and (59) at zero temperature, it is clear that their approximation in the use of the Bogoliubov method and the variational method respectively is equivalent to the ring diagram sum. The $\ln r_s$ and the constant term obtained by these authors cannot possibly come from the ring sum, however. One may then ask what diagrams in the low temperature expansion of the partition function would give a contribution of order $\ln r_s$ + const to the ground state energy. At high temperature and low density the correct choice is the first term of the Meeron nodal expansion, i.e., the watermelon diagrams evaluated by Abe to give the $\Lambda^2 \ln \Lambda$ in Eq. (2). Assuming the Meeron nodal expansion to be valid also at low temperature and high density, the same choice of diagrams would have to reduce to the asymptotic form $(\Lambda^2/\gamma^2)[a \ln (\Lambda/\gamma^{1/2}) + b]$ in order to give the $\ln r_s$ + const obtained by Girardeau.

[17] M. Girardeau, Phys. Rev. (to be published).

ACKNOWLEDGMENT

The author wishes to thank Professor R. S. Lehman for several helpful discussions on the mathematical analysis involved in this work.

APPENDIX. EVALUATION OF THE MULTIPLE INTEGRAL IN a_n AND b_n

For the evaluation of the coefficients a_n and b_n, we need the asymptotic value for large n of the multiple integral:

$$\begin{aligned} I_{nt} &= (1/2^n) \int_{-1}^{1} \cdots \int ds_1 \cdots ds_n \\ &\quad \times \exp(\pi i t \sum s_i)[1 - \sum_i s_i^2/n]^n, \\ &= \int_0^1 \cdots \int ds_1 \cdots ds_n \prod_i^n \cos \pi t s_i \\ &\quad \times [1 - \sum_i s_i^2/n]^n. \end{aligned} \tag{A1}$$

For a_n we need only $t = 0$ and the exponent of the square bracket in the integrand is $n - 1$; for b_n we need all values of t and the exponent is $n - \frac{3}{2}$. In the limit of large n, however, the exponents $n - 1$ and $n - \frac{3}{2}$ may be replaced with n.

In order to obtain the asymptotic value of I_{nt}, one hopes that the integrand will factor into n products as $n \to \infty$, by using

$$\lim_{n\to\infty} (1 - \sum_i s_i^2/n)^n = \prod_i^n e^{-s_i^2}. \tag{A2}$$

If this factorization were correct, then (A1) for $t = 0$ would be

$$\left\{ \int_0^1 ds\, e^{-s^2} \right\}^n = \theta_0^n, \tag{A3}$$

with $\theta_0 = \text{Erf}(1) = 0.74683$. Unfortunately, this evaluation is only a first approximation to θ_0, because the exponential approximation Eq. (A2) fails in one of the 2^n corners of the hypercube over which the multiple integration in Eq. (A1) is to be performed. As one approaches the corner at $(1, 1, \cdots, 1)$ the integrand becomes $(1 - n/n)^n = 0$, whereas the exponential approximation, Eq. (A2),

gives e^{-n}. Thus, the estimate of θ_0 as equal to Erf (1) is too large.

The exponential approximation for large but finite n may be improved by multiplying by a correction function, $f(x, n)$, defined by

$$(1 - x/n)^n = e^{-x} f(x, n),$$

where

$$\begin{aligned} f(x, n) &= e^x(1 - x/n)^n, \\ &= e^{x + n \ln (1 - x/n)}, \\ &= e^{-[x^2/2n + x^3/3n^2 + \cdots]}, \\ &= 1 - (x^2/2n + x^3/3n^2 + \cdots) \\ &\quad + (1/2!)(x^2/2n + \cdots)^2 - \cdots. \end{aligned} \tag{A4}$$

Thus by using Eq. (A4), I_{n_0} may be written as

$$\begin{aligned} I_{n0} &= \int_0^1 \cdots \int ds_1 \cdots ds_n \\ &\quad \times \exp\left(-\sum s_i^2\right) f\left(\sum s_i^2, n\right), \\ &= \int_0^1 \cdots \int ds_1 \cdots ds_n \prod_i^n \exp\left(-s_i^2\right) \\ &\quad \times \{1 - (\sum s_i^2)^2/2n + \cdots\}. \end{aligned} \tag{A5}$$

The multiple integrals over the terms of the expansion of $f(\sum s_i^2, n)$ may be written as products of single integrals involving powers of

$$A_m = \int_0^1 ds\, s^{2m} e^{-s^2}.$$

In order to obtain this expansion, we need to have the powers of $\sum s_i^2$ multiplied out. Thus for $(\sum s_i^2)^2$ we have

$$(\sum s_i^2)^2 = \sum_{i,j} s_i^2 s_j^2 + \sum_i s_i^4.$$

This expression has $n(n - 1)$ terms of the type $s_i^2 s_j^2$ and n terms of s_i^4; thus we write

$$(\sum s_i^2)^2 = n(n - 1) s_1^2 s_2^2 + n s_1^4.$$

Similarly, for higher powers of $\sum s_i^2$ one has

$$\begin{aligned} (\sum s_i^2)^r &= n_r s_1^2 \cdots s_r^2 \\ &\quad + \tfrac{1}{2} r(r - 1) n_{r-1} s_1^4 s_2^2 \cdots s_{r-1}^2 + \cdots, \end{aligned} \tag{A6}$$

where $n_r = n(n - 1) \cdots (n - r + 1) = n!/(n - r)!$. A good estimate of I_{n0} is obtained if only the $n_r s_1^2 \cdots s_r^2$ term in (A6) is used for the powers of $\sum s_i^2$ in the expansion of $f(\sum s_i^2, n)$. The next term in Eq. (A6) gives corrections smaller by $1/n$, $1/n^2$, etc. After some laborious algebra using (A6) in (A5) one finds

$$\begin{aligned} I_{n0} &= A_0^n \left\{1 - \frac{n_2}{n}\phi + \frac{n_4}{2!n^2}\phi^2 - \frac{n_6}{3!n^3}\phi^3 + \cdots\right\}, \\ \phi &= -[A_1/A_0 + \ln(1 - A_1/A_0)] \\ &= 0.03893. \end{aligned} \tag{A7}$$

Equation (A7) would be the desired form, namely, $I_{n0} = \theta_0^n$, if the quantity in braces were the binomial expansion of $(1 - \phi)^n$. Actually, coefficients of ϕ^s in Eq. (A7) are

$$\begin{aligned} \frac{n_{2s}}{s!\, n^s} &= \frac{n_s}{s!} \frac{(n - s - 1) \cdots (n - 2s)}{n^s}, \\ &\cong n_s/s!, \qquad s \ll n. \end{aligned} \tag{A8}$$

Thus, with the approximation of Eq. (A8) we have

$$I_{n0} = \theta_0^n = A_0^n(1 - \phi)^n, \tag{A9}$$

with $\theta_0 = A_0(1 - \phi) = 0.71776$. This numerical value of θ_0 gives the value of the radius of convergence of $\xi^2 = \sum_n a_n \gamma^{2(n-1)}$ as $\gamma_c^2 = 4/e\theta_0 = 2.050$. The numerical value of γ_c^2 found by solving the transcendental equations for the tangent point in Fig. 2 is $\gamma_c^2 = 2.042$. Probably the estimate of θ_0 could be improved by taking more terms from (A6) and using them in the expansion of $f(\sum s_i^2, n)$.

The procedure used to evaluate I_{n0} is invalid for I_{nt} when $n^{1/2}/\pi t < 1$, i.e., for small t, as may be seen from the exact expansion of b_{nt}, Eq. (41). One notes that because of the $e^{\pi i t s_i}$ factors in the integrand, the value of I_{nt} is very much less than I_{n0}. A quick estimate using Eq. (A2) gives

$$I_{nt} = \theta_t^n = \left\{\int_0^1 ds e^{-s^2} \cos \pi t s\right\}^n \tag{A10}$$

If the limit of integration in Eq. (A10) were 0 to ∞, then the estimate of θ_t would be $(\sqrt{\pi}/2) \exp -(\pi t)^2/4$. Because of the finite integration interval, however, the correct estimate to order $(\pi t)^{-2}$ is

$$\theta_t = \frac{2(-1)^{t+1}}{e(\pi t)^2} \tag{A11}$$

which leads to Eq. (40) for b_{nt}. Apparently when $\pi t/n^{1/2}$ is large, the approximation of Eq. (A8) fails.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Quantum Statistics and the Boltzmann Equation*

ROBERT M. LEWIS
Institute of Mathematical Sciences, New York University, New York, New York
(Received September 20, 1961)

The system of hierarchy equations for the reduced density operators of quantum statistical mechanics is replaced by a single functional differential equation for a generating functional. A formal solution of the initial value problem for the latter equation is obtained, leading to series expansions of the reduced density operators. These expansions are used to obtain an improved derivation of the quantum-mechanical Boltzmann equation.

1. INTRODUCTION

THE statistical-mechanical treatment of a quantum-mechanical many-particle system usually begins with the density operator D_n, which is the solution of an initial value problem for an equation analogous to the classical Liouville equation. The solution of this problem is equivalent to the solution of the quantum-mechanical n-body problem and is therefore not practical when n is large.

In order to circumvent this difficulty one may introduce reduced density operators F_s ($s = 1, 2, \cdots$) which, in an appropriate representation, are defined as integrals of the density operator. These operators satisfy an infinite system of equations (in the limit $n \to \infty$) called the hierarchy equations, and this system is, in turn, equivalent to a single functional differential equation for a generating functional L which generates the operators F_s.

In Sec. 4 of this paper we obtain a formal solution of the functional differential equation, and from that solution we derive expansions for the operators F_s as power series in the number density. These expansions have the advantage that, for small densities, F_s may be approximated by a few terms of the series. Then to obtain an explicit expression for F_s, only certain k-body problems, where k is small, need to be solved. However the expansions probably suffer from the defect that the remainder terms grow rapidly with time.

The work in Sec. 4 parallels our earlier work[1] on the corresponding problem in classical mechanics. In fact, by introducing the appropriate representation of the operators, the two treatments become formally almost identical. We call this representation the Wigner representation because in it the density operator is represented by the well-known Wigner function. The Wigner representation is discussed in Sec. 3.

In Sec. 2 we review the basic notions of quantum statistical mechanics which we shall require. This is necessary because these notions are presented from a point of view which is slightly different from customary formulations. The unifying idea in this point of view is the representation of Hilbert space in terms of a complete set of commuting observables.

In Sec. 5 we apply the expansion of F_1, which was derived in Sec. 4, as the starting point for a derivation of the quantum-mechanical Boltzmann equation. By careful use of operator representations we are able to avoid most of the mathematical difficulties which appear in earlier derivations. However, as we point out in Sec. 5, one basic difficulty remains in that, at one point, we are forced to argue by analogy with the case of classical mechanics. The significance of this difficulty and the validity of the equation so derived is discussed in the concluding remarks.

2. BASIC NOTATIONS OF QUANTUM STATISTICAL MECHANICS

Let $\mathfrak{K}$ denote a Hilbert space of elements Φ with inner product (Φ, Φ_1) and norm $||\Phi|| = (\Phi, \Phi)^{1/2}$. The states of a quantum-mechanical system correspond to elements Φ of $\mathfrak{K}$ with $||\Phi|| = 1$. For many purposes it is convenient to represent the elements Φ by functions $\psi(q) = \psi(q^{(1)}, \cdots, q^{(k)})$ such that

$$\int |\psi(q)|^2 \, dq = \int |\psi(q)|^2 \, dq^{(1)} \cdots dq^{(k)} < \infty. \quad (2.1)$$

This is the well-known Schrödinger, or coordinate, representation. For our purposes it is convenient to indicate the correspondence between elements Φ and functions $\psi(q)$ by writing

$$R_q(\Phi) = \psi(q). \quad (2.2)$$

* The research in this paper was supported by Contract No. AF49(638)341. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.
[1] R. M. Lewis, J. Math. Phys. 2, 222 (1961).

If $q = \{q^{(1)}, \cdots, q^{(k)}\}$ denotes[2] the complete set of coordinate operators then

$$R_q(q^{(j)}\Phi) = q^{(j)}\psi(q); \qquad j = 1, \cdots, k; \tag{2.3}$$

and

$$||\Phi||^2 = \int |\psi(q)|^2 \, dq. \tag{2.4}$$

More generally, if $Y = \{Y^{(1)}, \cdots, Y^{(k)}\}$ is any complete set of commuting observables we shall denote the representation of the Hilbert space in terms of these operators by

$$R_Y(\Phi) = c(y) = c(y^{(1)}, \cdots, y^{(k)}), \tag{2.5}$$

$$R_Y(Y^{(j)}\Phi) = y^{(j)}c(y); \qquad j = 1, \cdots, k. \tag{2.6}$$

Here $c(y)$ is a complex-valued function. In general $||\Phi||^2$ will not be given by a simple integral of the type (2.4) but may be given by a sum (discrete spectrum), a combination of an integral and a sum, etc. In order to admit all possibilities we shall write[3-6]

$$||\Phi||^2 = \int |c(y)|^2 \, dm(y). \tag{2.7}$$

With the aid of the representation R_Y, functions of the operators $Y^{(1)}, \cdots, Y^{(k)}$ are conveniently defined by

$$R_Y[f(y)\Phi] = f(y)c(y) = f(y)R_Y[\Phi]. \tag{2.8}$$

In this paper we shall be concerned with a system consisting of n identical particles. Since an individual particle can also be considered as a quantum-mechanical system it is essential to examine the relationships between the Hilbert spaces of the component systems and the Hilbert space of the composite system.

Given n identical quantum-mechanical systems with Hilbert spaces $\mathfrak{H}_\nu$; $\nu = 1, \cdots, n$; let $Y_\nu = \{Y_\nu^{(1)}, \cdots, Y_\nu^{(k)}\}$ be a complete set of commuting observables in $\mathfrak{H}_\nu$, with representation

$$R_{Y_\nu}(\Phi_\nu) = c_\nu(y_\nu) = c_\nu(y_\nu^{(1)}, \cdots, y_\nu^{(k)}); \tag{2.9}$$

$$||\Phi_\nu||^2 = \int |c_\nu(y_\nu)|^2 \, dm(y_\nu). \tag{2.10}$$

Although the subscript ν will be used to distinguish the systems, we choose the operators, and hence the representations, to be the same in the n systems. Let $\mathfrak{K}$ be the space of complex-valued functions

$$c(y_{(n)}) = c(y_1, \cdots, y_n) = c(y_1^{(1)}, \cdots, y_n^{(k)}) \tag{2.11}$$

for which

$$\int |c(y_{(n)})|^2 \, dm(y_1) \cdots dm(y_n) < \infty. \tag{2.12}$$

Then $\mathfrak{K}$ may be viewed as representing a Hilbert space $\mathfrak{H}$ whose elements Φ correspond to the functions $c(y_{(n)})$, and for which

$$||\Phi||^2 = \int |c(y_{(n)})|^2 \, dm(y_1) \cdots dm(y_n). \tag{2.13}$$

We denote this correspondence by

$$R_{Y_{(n)}}(\Phi) = c(y_{(n)}). \tag{2.14}$$

The Hilbert space $\mathfrak{H}$ is then said to be the *product*[7] of the spaces $\mathfrak{H}_1, \cdots, \mathfrak{H}_n$.

Each observable A_ν acting on $\mathfrak{H}_\nu$, has a representation as a functional operator [e.g., a differential or integral operator] acting on the functions $c_\nu(y_\nu)$. We now define the operator A_ν acting on $\mathfrak{H}$ by requiring that the corresponding functional operator acting on the functions $c(y_{(n)})$ be the same functional operator, the variables y_j $(j \neq \nu)$ being held fixed.

The *composite* of the n identical systems introduced above is the quantum-mechanical system whose Hilbert space is the product space $\mathfrak{H}$ and whose observables include the operators A_ν just defined, as well as all functions of these. Thus, in constructing the composite system we have identified every physical observable of the component systems with one of the composite system. It is easily seen that the correspondence (2.14) is the representation of $\mathfrak{H}$ in terms of the complete set of commuting observables

$$Y_{(n)} = \{Y_1, \cdots, Y_n\} = \{Y_1^{(1)}, \cdots, Y_n^{(k)}\}. \tag{2.15}$$

The composite system defined above is called a *Maxwell-Boltzmann system* to distinguish it from *Fermi-Dirac* and *Bose-Einstein* systems to be discussed shortly.

Let P be any permutation of the integers $\{1, \cdots, n\}$. Then P assigns to each integer j in this set another integer Pj in the set. Let $Py_{(n)}$ be defined by

$$Py_{(n)} = P\{y_1, \cdots, y_n\} = \{y_{P1}, \cdots, y_{Pn}\}. \tag{2.16}$$

[2] Following customary notation we sometimes use the same symbol to denote an operator and the independent variable (eigenvalue) in its representation.

[3] In a rigorous treatment of quantum mechanics, R_Y is called the simultaneous spectral representation of the complete set of commuting operators $Y^{(1)}, \cdots, Y^{(k)}$, and $m(y)$ is a measure function. For an excellent treatment of these notions see references 4–6.

[4] K. O. Friedrichs, Communs. Pure and Appl. Math. **1**, 361 (1948).

[5] K. O. Friedrichs, Communs. Pure and Appl. Math. **4**, 161 (1951).

[6] K. O. Friedrichs, "Spectral Theory of Operators in Hilbert Space," N. Y. U. Lecture Notes (1960).

[7] P. A. M. Dirac, *Principles of Quantum Mechanics* (Oxford University Press, New York, 1958), 4th ed., p. 82.

The *permutation operator* P on $\mathfrak{H}$ may now be defined by means of the representation $R_{\gamma(n)}$:

$$R_{\gamma(n)}(P\Phi) = c(Py_{(n)}). \tag{2.17}$$

By means of these permutation operators, the *symmetry operator* $\mathfrak{S}_n$ and the *antisymmetry operator* $\mathfrak{A}_n$ are defined by

$$\mathfrak{S}_n = \frac{1}{n!}\sum P; \qquad \mathfrak{A}_n = \frac{1}{n!}\sum (\pm)P. \tag{2.18}$$

Here the summation is over the $n!$ permutations P, and the positive or negative sign is taken for even or odd permutations, respectively. It is easy to show that the permutation operators are unitary and that the operators $\mathfrak{S}_n$ and $\mathfrak{A}_n$ are Hermitian,

$$\mathfrak{S}_n^* = \mathfrak{S}_n, \qquad \mathfrak{A}_n^* = \mathfrak{A}_n, \tag{2.19}$$

and satisfy the condition

$$\mathfrak{S}_n^2 = \mathfrak{S}_n, \qquad \mathfrak{A}_n^2 = \mathfrak{A}_n. \tag{2.20}$$

The last two equations imply that $\mathfrak{S}_n$ and $\mathfrak{A}_n$ are orthogonal projectors.

Let $B = B(A_\nu^{(i)})$ be a function of all the operators $A_\nu^{(i)}$, $\nu = 1, \cdots, n$; $i = 1, \cdots, j_\nu$; where $A_\nu^{(i)}$ is an observable in $\mathfrak{H}_\nu$. Then B is an observable in $\mathfrak{H}$, and is said to be a *symmetric observable* if $B(A_P^{(i)}) = B(A_\nu^{(i)})$ for all permutations P of $\{1, \cdots, n\}$. It can easily be shown that an operator B is symmetric if and only if it commutes with all the permutation operators. From this it follows that $\mathfrak{S}_n$ and $\mathfrak{A}_n$ are symmetric and that any symmetric operator B commutes with $\mathfrak{S}_n$ and $\mathfrak{A}_n$:

$$[B, \mathfrak{S}_n] = 0; \qquad [B, \mathfrak{A}_n] = 0. \tag{2.21}$$

Since $\mathfrak{S}_n$ (or $\mathfrak{A}_n$) is an orthogonal projector, we may introduce the subspace $\mathfrak{H}_\mathfrak{S}$ (or $\mathfrak{H}_\mathfrak{A}$) of $\mathfrak{H}$ onto which $\mathfrak{S}_n$ (or $\mathfrak{A}_n$) projects. This is the space of all elements Φ of $\mathfrak{H}$ for which $\mathfrak{S}_n\Phi = \Phi$ (or $\mathfrak{A}_n\Phi = \Phi$), i.e., the eigenspace of $\mathfrak{S}_n$ (or $\mathfrak{A}_n$) corresponding to the eigenvalue 1. The elements Φ of $\mathfrak{H}_\mathfrak{S}$ (or $\mathfrak{H}_\mathfrak{A}$) are called *symmetric vectors* (or *antisymmetric vectors*) and have the property that for every permutation operator P,

$$P\Phi = \Phi \quad (\text{or} \quad P\Phi = \pm\Phi). \tag{2.22}$$

Since every symmetric observable B commutes with $\mathfrak{S}_n$, it follows that if Φ is an element of $\mathfrak{H}_\mathfrak{S}$ then $B\Phi = B\mathfrak{S}_n\Phi = \mathfrak{S}_nB\Phi$ is also an element of $\mathfrak{H}_\mathfrak{S}$. Hence B is an observable on $\mathfrak{H}_\mathfrak{S}$. The same assertion is true for $\mathfrak{H}_\mathfrak{A}$. Thus we see that there is a natural correspondence between observables in the component systems and symmetric functions of these observables in $\mathfrak{H}_\mathfrak{S}$ and $\mathfrak{H}_\mathfrak{A}$. A quantum mechanical system whose Hilbert space is $\mathfrak{H}_\mathfrak{S}$ (or $\mathfrak{H}_\mathfrak{A}$) and whose observables correspond in the above way to the observables of the component systems is called a *Bose-Einstein system* (or *Fermi-Dirac system*).

In the remainder of this paper we shall treat simultaneously the three types of systems mentioned. This is very conveniently done by introducing the operator Q_n, where $Q_n = \mathfrak{S}_n$ for Bose-Einstein systems, $Q_n = \mathfrak{A}_n$ for Fermi-Dirac systems, and $Q_n = 1$ for Maxwell-Boltzmann systems. Thus, e.g., $\mathfrak{H}_Q = \mathfrak{H}$ for the case $Q_n = 1$.

Let $A = A(1, \cdots, j)$ be an operator whose functional representer in the representation $R_{\gamma(n)}$ is an integral operator acting on the variables $(1, \cdots, j)$, i.e.,

$$\begin{aligned} R_{\gamma(n)}[A(1, \cdots, j)\Phi] &= \int a(y_1, \cdots, y_j, y_1', \cdots, y_j') \\ &\quad \times c(y_1', \cdots, y_j', y_{j+1}, \cdots, y_n) \\ &\quad \times dm(y_1') \cdots dm(y_j'). \end{aligned} \tag{2.23}$$

In this case, we say that "the operator A is represented by the kernal a," and indicate the correspondence by writing[8]

$$\begin{aligned} R_{\gamma(n)}[A(1, \cdots, j)] &= a(y_1, \cdots, y_j, y_1', \cdots, y_j') \\ &= a(y_{(j)}, y_{(j)}'). \end{aligned} \tag{2.24}$$

In terms of the notation introduced in (2.23), $A(2, \cdots, j+1)$ is the operator such that

$$\begin{aligned} R_{\gamma(n)}[A(2, \cdots, j+1)\Phi] &= \int a(y_2, \cdots, y_{j+1}, y_2', \cdots, y_{j+1}') \\ &\quad \times c(y_1, y_2', \cdots, y_{j+1}', y_{j+2}, \cdots, y_n) \\ &\quad \times dm(y_2') \cdots dm(y_{j+1}'), \end{aligned} \tag{2.25}$$

etc. It is easy to see that if $A(1, \cdots, j)$ and $B(j+1, \cdots, j+m)$ operate on different sets of variables then $[A, B] = AB - BA = 0$, and

$$R_{\gamma(n)}[AB] = R_{\gamma(n)}[A]R_{\gamma(n)}[B]. \tag{2.26}$$

Let $\{\Phi_\nu\}$ be a complete orthonormal set of elements of $\mathfrak{H}$ and let B be a linear operator on $\mathfrak{H}$. The *trace* of the operator B is defined by[9]

---

[8] It is important to note the distinction between (2.14) and (2.24). The former is a correspondence between elements Φ and functions c; the latter is a correspondence between operators A and kernels a, which is induced by the former.

[9] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955), pp. 178–195.

$$\mathrm{Tr}\,(B) = \sum_\nu (\Phi_\nu, B\Phi_\nu), \tag{2.27}$$

provided the series converges absolutely. It can be shown that the value of Tr (B) is independent of the choice of the set $\{\Phi_\nu\}$, and if

$$R_{Y(n)}[B] = b(y_{(n)}, y'_{(n)}) \tag{2.28}$$

then

$$\mathrm{Tr}\,(B) = \int b(y_{(n)}, y_{(n)})\, dm(y^{(1)}) \cdots dm(y^{(n)}). \tag{2.29}$$

It is convenient also to define the *partial trace* $T_{s+1,\cdots,j}$ of an operator with respect to some of the variables. Thus, e.g., for $s < j$, $T_{s+1,\cdots,j}A(1, \cdots, j)$ is the operator such that

$$\begin{aligned} R_{Y(n)}\{T_{s+1,\cdots,j}A(1, \cdots, j)\} &= \int a(y_1, \cdots, y_s, z_{s+1}, \cdots, z_j, \\ &\quad \times y'_1, \cdots, y'_s, z_{s+1}, \cdots, z_j) \\ &\quad \times dm(z_{s+1}) \cdots dm(z_j). \end{aligned} \tag{2.30}$$

Of course

$$T_{1,\cdots,n}A(1, \cdots, n) = \mathrm{Tr}\, A(1, \cdots, n). \tag{2.31}$$

In quantum statistical mechanics, the mean value of an arbitrary observable B is given by

$$\langle B\rangle = \mathrm{Tr}\,(D_n B), \tag{2.32}$$

where D_n is a nonnegative, Hermitian, bounded, completely continuous linear operator, called the *density operator*. It is normalized by the condition

$$\mathrm{Tr}\,(D_n) = 1. \tag{2.33}$$

D_n varies with time according to the equation

$$D_n(t) = \exp(-itH_n/\hbar)\, D_n(0) \exp(itH_n/\hbar), \tag{2.34}$$

where H_n is the Hamiltonian operator of the system. It follows from (2.34) that D_n satisfies the differential equation

$$i\hbar\, \partial D_n/\partial t = [H_n, D_n]. \tag{2.35}$$

If, now, we have a quantum-mechanical system with Hilbert space $\mathfrak{H}_Q$, we must interpret the trace in (2.32) and (2.33) as a trace, $\mathrm{Tr}_{(Q)}$, over $\mathfrak{H}_Q$, i.e., the orthonormal set $\{\Phi_\nu\}$ in (2.27) must consist of elements of $\mathfrak{H}_Q$. It is necessary to relate the trace $\mathrm{Tr}_{(Q)}$ to the trace Tr over $\mathfrak{H}$, and this can be done as follows:

Let $\{\Phi_\nu\}$ be a complete orthonormal set in $\mathfrak{H}_Q$, and $\{\Psi_\mu\}$ a complete orthonormal set in the orthocomplement $\mathfrak{H}_Q^\perp$ of $\mathfrak{H}_Q$. Then the combined set is a complete orthonormal set in $\mathfrak{H}$, and

$$\mathrm{Tr}\,(B) = \sum_\nu (\Phi_\nu, B\Phi_\nu) + \sum_\mu (\Psi_\mu, B\Psi_\mu). \tag{2.36}$$

We assume that B is an observable in $\mathfrak{H}_Q$. (Thus in the cases $Q_n = \mathcal{S}_n$ and $Q_n = \mathcal{A}_n$, B is symmetric.) It follows from (2.21) that

$$[B, Q_n] = 0. \tag{2.37}$$

Now $Q_n\Phi_\nu = \Phi_\nu$, $Q_n\Psi_\mu = 0$, and $Q_n^2 = Q_n$. Thus, from (2.27) and (2.19),

$$\begin{aligned} \mathrm{Tr}\,(Q_nB) &= \mathrm{Tr}\,(Q_n^2B) = \mathrm{Tr}\,(Q_nBQ_n) \\ &= \sum_\nu (\Phi_\nu, Q_nBQ_n\Phi_\nu) = \sum_\nu (Q_n\Phi_\nu, BQ_n\Phi_\nu) \\ &= \sum_\nu (\Phi_\nu, B\Phi_\nu) = \mathrm{Tr}_{(Q)}\,(B). \end{aligned} \tag{2.38}$$

Thus, in all three cases,

$$\langle B\rangle = \mathrm{Tr}_{(Q)}\,(D_nB) = \mathrm{Tr}\,(Q_n\, D_nB) = \mathrm{Tr}\,(D_nQ_nB), \tag{2.39}$$

and

$$\mathrm{Tr}\,(D_nQ_n) = \mathrm{Tr}\,(Q_n\, D_n) = 1. \tag{2.40}$$

For a system of n identical components, H_n and D_n are symmetric, hence commute with Q_n. Therefore,

$$Q_n\, D_n(t) = \exp(-itH_n/\hbar)Q_n\, D_n(0) \exp(itH_n/\hbar) \tag{2.41}$$

and

$$i\hbar\, \frac{\partial}{\partial t}(Q_n\, D_n) = [H_n, Q_n\, D_n]. \tag{2.42}$$

The object of quantum statistical mechanics is to calculate the mean values $\langle B\rangle$. To this end, one must study the solutions of Eq. (2.42). Although this can be done in abstract operator notation, or in terms of any convenient representation, we shall work instead with a Wigner representation which we shall introduce in the next section. This representation has the advantage of producing formulas which resemble closely the corresponding formulas of classical statistical mechanics.

The quantum-mechanical system we shall consider consists of n identical monatomic spinless particles contained in a finite volume V. The particles obey Bose-Einstein, Fermi-Dirac, or Maxwell-Boltzmann statistics, according as $Q_n = \mathcal{S}_n$, $Q_n = \mathcal{A}_n$, or $Q_n = 1$. The Hamiltonian of the system is given by

$$\tilde{H}_n = T_n + U_n + \sum_{i=1}^{n} u_V(q_i), \tag{2.43}$$

$$T_n = \sum_{i=1}^{n} \frac{p_i^2}{2m} = \sum_{i=1}^{n}\sum_{\alpha=1}^{3} \frac{(p_i^\alpha)^2}{2m}, \tag{2.44}$$

$$U_n = \sum_{1\le i<j\le n} \phi(|q_i - q_j|). \tag{2.45}$$

Here m denotes the mass of a particle, $q_i = \{q_i^1, q_i^2, q_i^3\}$ and $p_i = \{p_i^1, p_i^2, p_i^3\}$ denotes the Cartesian coordinates and the conjugate moments of the ith particle, ϕ is the interparticle potential, and $v_V(q_i)$ is the potential due to the containing boundary. Thus, $u_V(q)$ is constant inside V and rapidly approaches infinity at the boundary. At an appropriate point we shall let n and V tend to infinity in such a way that $v = V/n$ approaches a finite limit (the specific volume). In this process, the effect of the terms $u_V(q_i)$ vanishes[10]; hence, we shall omit these terms from further equations. The resulting Hamiltonian is

$$H_n = T_n + U_n. \tag{2.46}$$

In our analysis we shall make use of Hamiltonians

$$H_s = T_s + U_s \text{ and } H_{s,0} = T_s; \qquad s = 1, 2, \cdots \tag{2.47}$$

corresponding to systems of s interacting and noninteracting particles.

In what follows we shall make frequent use of a representation in terms of the complete set of momentum operators

$$p_{(n)} = \{p_1, \cdots, p_n\} = \{p_1^1, \cdots, p_n^3\}. \tag{2.48}$$

Since this representation will be used so frequently we will denote it simply by R, rather than $R_{p(n)}$. The momentum representation is characterized by

$$R(\Phi) = \phi(p_{(n)}), \tag{2.49}$$

$$R(p_i^\alpha \Phi) = p_i^\alpha \phi(p_n), \tag{2.50}$$

$$||\Phi||^2 = \int |\phi(p_{(n)})|^2 \, dp_{(n)}. \tag{2.51}$$

Here $dp_{(n)} = dp_1 \cdots dp_n = dp_1^1 \cdots dp_n^3$, and the integration is over all values of $p_{(n)}$.

3. THE WIGNER REPRESENTATION

Let $A(1, \cdots, s)$ be an operator such that

$$R[A(1, \cdots, s)] = \alpha(p_{(s)}, p'_{(s)}). \tag{3.1}$$

Here, as in Sec. 2, R denotes the momentum representation. Let

$$a(q_{(s)}, p_{(s)}) = (\pi\hbar)^{-3s} \int \exp\left\{\frac{-2i}{\hbar} q_{(s)} \cdot z_{(s)}\right\} \times \alpha(p_{(s)} - z_{(s)}, p_{(s)} + z_{(s)}) \, dz_{(s)}. \tag{3.2}$$

It follows, by inversion of the Fourier transform, that

$$\alpha(p_{(s)}, p'_{(s)}) = \int \exp\left\{\frac{i}{\hbar} q_{(s)} \cdot (p'_{(s)} - p_{(s)})\right\} \times a[q_{(s)}, \tfrac{1}{2}(p_{(s)} + p'_{(s)})] \, dq_{(s)}. \tag{3.3}$$

These equations define a correspondence between functions $a(q_{(s)}, p_{(s)})$ and operators $A(1, \cdots, s)$. It is not apparent that a is the representer of the operator $A(1, \cdots, s)$ in terms of any complete set of commuting observables. Nevertheless it is convenient to call a "the representer of A in a *Wigner representation*" and to indicate the correspondence by writing

$$W[A(1, \cdots, s)] = a(q_{(s)}, p_{(s)}). \tag{3.4}$$

With the aid of (2.26) it is easy to see that

$$W[A(1, \cdots, j)B(j+1, \cdots, j+m)] = W[A(1, \cdots, j)]W[B(j+1, \cdots, j+m)]. \tag{3.5}$$

For convenience, we shall sometimes denote the pair q, p by x. Thus, e.g.,

$$q_{(s)}, p_{(s)} = x_{(s)};$$

$$q_1, \cdots, q_s, p_1, \cdots, p_s = x_1, \cdots, x_s; \tag{3.6}$$

etc.

In terms of the above definitions, the well-known *Wigner function*, $w_n(t, x_{(n)})$ is simply the Wigner representer of the operator $Q_n D_n(t)$:

$$w_n(t, x_{(n)}) = w_n(t, q_{(n)}, p_{(n)}) = W[Q_n D_n(t)]. \tag{3.7}$$

The properties of the Wigner function are discussed extensively by Irving and Zwanzig.[11] Here we shall outline briefly those which we shall require.

Since Q_n and D_n are Hermitian and commute, it follows that their product is Hermitian and hence that w is real, although not necessarily nonnegative. Given an arbitrary function $g(q_{(n)}, p_{(n)})$, it is well known that because of ambiguities arising from the ordering of noncommuting operators there may exist more than one operator corresponding to g. However, among these operators one may specify a unique operator, called the *Weyl assignment*. If G is the operator corresponding, under this assignment, to $g(q_{(n)}, p_{(n)})$ then

$$\langle G \rangle = T_r(Q_n D_n G) = \int w_n(t, q_{(n)}, p_{(n)}) g(q_{(n)}, p_{(n)}) \, dq_{(n)} \, dp_{(n)}. \tag{3.8}$$

The Weyl assignment and the proof of (3.8) are given in reference 11.

[10] See N. N. Bogolyubov and K. P. Gurov, J. Exptl. Theoret. Phys. (U.S.S.R.) 17, 614 (1947).

[11] J. H. Irving and R. W. Zwanzig, J. Chem. Phys. 19, 1173 (1951).

In classical statistical mechanics, the mean value of a phase function $g(q_{(n)}, p_{(n)})$ is given by a formula identical to (3.8) where w_n is the phase space probability density function, and satisfies the Liouville equation (see reference 1). The great virtue of the Wigner function formalism is this similarity to the classical case. Indeed the Wigner function satisfies a quantum-mechanical Liouville equation, similar to the classical equation (but differing in the crucial fact that it is not of first order).

This equation is simply the Wigner representation of (2.42) and is given by

$$\frac{\partial w_n}{\partial t} + \{w_n; T_n\} + \sum_{1 \le j < k \le n} \theta_{jk} \cdot w_n = 0. \tag{3.9}$$

Here $\{w_n; T_n\}$ denotes the Poisson bracket, and the operator $\theta_{jk}\cdot$ is given by

$$\theta_{jk}\cdot w_n = -\frac{2}{\hbar}\sin\left(\frac{\hbar}{2}\nabla_{p(n)}\cdot\nabla_{q(n)}\right) \times \phi(|q_j - q_k|)w_n(t, r_{(n)}, p_{(n)})\,|_{r(n)=q(n)}. \tag{3.10}$$

We note that

$$\begin{aligned}\theta_{jk}\cdot w_n &= -\nabla_{q(n)}\phi(|q_j - q_k|)\cdot\nabla_{p(n)}w_n(t, q_{(n)}, p_{(n)}) \\ &\quad + O(\hbar^2) \\ &= -\{\phi(|q_j - q_k|); w_n\} + O(\hbar^2).\end{aligned} \tag{3.11}$$

The classical Liouville equation is given by (3.9) with $\theta_{jk}\cdot w_n$ replaced by $-\{\phi(|q_j - q_k|); w_n\}$; thus it is clearly the limiting form of the quantum-mechanical equation for $\hbar \to 0$. Various forms of the operator $\theta_{jk}\cdot$ are given in reference 11.

With the aid of the Wigner representation, we can introduce a formal solution of (3.9) derived from (2.41). This solution is given in terms of a solution operator $S_t^{(k)}(1, \cdots, k)$ defined as follows: If $a(q_{(s)}, p_{(s)}) = W[A(1, \cdots, s)]$, and $k \le s$ then

$$\begin{aligned}S_t^{(k)}(1, \cdots, k)a(x_{(s)}) &= S_t^{(k)}(1, \cdots, k)a(q_{(s)}, p_{(s)}) \\ &= W\left\{\exp\left[\frac{it}{\hbar}H_k(1, \cdots, k)\right]A(1, \cdots, s) \right. \\ &\quad \left. \times \exp\left[\frac{-it}{\hbar}H_k(1, \cdots, k)\right]\right\}.\end{aligned} \tag{3.12}$$

Here H_k is defined by (2.47). The solution operator $S_t^{(k,0)}(1, \cdots, k)$ is defined by (3.12) with H_k replaced by $H_{k,0}$. In an analogous way we can define the operator $S_t^{(k)}(2, \cdots, k+1)$, etc. Where no confusion can arise, we omit the explicit indication of the variables. Thus, e.g.,

$$S_t^{(s)}a(q_{(s)}, p_{(s)}) = S_t^{(s)}(1, \cdots, s)a(q_{(s)}, p_{(s)}). \tag{3.13}$$

It follows immediately from the definitions that $S_0^{(k)}a = a$, and for all t_1, t_2

$$S_{t_1}^{(k)}S_{t_2}^{(k)} = S_{t_1+t_2}^{(k)}. \tag{3.14}$$

From (2.41) it is clear that the formal solution to the initial value problem for the quantum-mechanical Liouville equation is given by

$$w_n(t, q_{(n)}, p_{(n)}) = S_{-t}^{(n)}w_n(0, q_{(n)}, p_{(n)}). \tag{3.15}$$

For use in Sec. 5, it will be convenient to define an operator $\mathfrak{Q}_k$ which, in the Wigner representation, corresponds to multiplication of an operator by the operator Q_k. Thus if $a(q_{(s)}, p_{(s)}) = W[A(1, \cdots, s)]$, let

$$\mathfrak{Q}_k a(q_{(s)}, p_{(s)}) = W[Q_k A(1, \cdots, s)]. \tag{3.16}$$

In the case of classical statistical mechanics solution operators may also be defined.[1] One first introduces the solution operator for the system of Hamilton's equations corresponding to the Hamiltonian H_k (or $H_{k,0}$). If we denote the solution of these equations by $x_{(k)}(t) = \{q_{(k)}(t), p_{(k)}(t)\}$, then this solution operator is defined by

$$S_t^{(k)}x_{(k)}(0) = x_{(k)}(t). \tag{3.17}$$

This operator maps each point in $6k$-dimensional phase space into its image after time t along the classical trajectory. The classical solution operator for functions is then defined by

$$S_t^{(k)}(1, \cdots, k)a(x_{(s)}) = a[S_t^{(k)}x_{(k)}, x_{k+1}, \cdots, x_s], \tag{3.18}$$

$$S_t^{(k)}(2, \cdots, k+1)a(x_{(s)}) = a[x_1, S_t^{(k)}\{x_2, \cdots, x_{k+1}\}, x_{k+2}, \cdots, x_s], \tag{3.19}$$

etc. If, now, we interpret $S_t^{(n)}$ as the classical solution operator, the solution of the initial value problem for the classical Liouville equation is given by (3.15).

It is important to observe that in the case of free particles ($\phi \equiv 0$; $H_n = H_{n,0}$) the classical and quantum-mechanical solution operators coincide. This is seen most easily from the fact that in this case the two Liouville equations are identical.

In both the classical and the quantum-mechanical cases, the determination of the operator $S_t^{(n)}$ is equivalent to the solution of the corresponding mechanical problem and hence is, in general, not practical for $n > 2$. In an attempt to circumvent this difficulty, we introduce s-particle Wigner functions defined by

$$\begin{aligned}f_{n,s}(t, x_{(s)}) &= f_{n,s}(t, x_1, \cdots, x_s) \\ &= V^s \int w_n(t, x_1, \cdots, x_n)\, dx_{s+1} \cdots dx_n; \\ &\qquad s = 1, 2, \cdots, n.\end{aligned} \tag{3.20}$$

By integrating Eq. (3.9) with respect to $x_{s+1}, \cdots, x_n$, one obtains the system of equations for the functions $f_{n,s}$

$$\frac{\partial f_{n,s}}{\partial t} + \{f_{n,s}; T_s\} + \sum_{1 \le j < k \le s} \theta_{jk} \cdot f_{n,s} = -\frac{1 - s/n}{v} \times \sum_{i=1}^{s} \int \theta_{i,s+1} \cdot f_{n,s+1}(t, x_1, \cdots, x_{s+1})\, dx_{s+1};$$

$$s = 1, 2, \cdots, n-1. \qquad (3.21)$$

Here T_s is defined by (2.44) and $v = V/n$. We observe that the left side of (3.21) is identical to (3.9) for $s = n$. In the derivation of (3.21) one makes use of the symmetry of w_n, which is a consequence of the symmetry of the operator $Q_n D_n$. The details of the derivation are given in reference 11.

In the next section we shall obtain a formal solution of the initial value problem for the system of equations (3.21) in the limit $V \to \infty$, $n \to \infty$, $V/n = v$. In the following section we shall use that solution to derive a quantum-mechanical analog of the Boltzmann equation.

4. THE FUNCTIONAL DIFFERENTIAL EQUATION

The work in this section parallels the development of the classical case given in Secs. 2 and 3 of reference 1. In fact, if we simply replace the classical density function D_n of reference 1 by the Wigner function w_n, and the operation $\{\phi(|q_i - q_k|);\}$ by $-\theta_{ik}\cdot$, the analysis becomes identical. Hence we shall only outline the development here, and refer to reference 1 for the details. A generalization of the classical case to systems of n species of particles is presented by Stell.[12]

We begin by introducing the functional

$$L_n[t, u] = \int w_n(t, x_1, \cdots, x_n) \times \prod_{i=1}^{n} [1 + v u(x_i)]\, dx_1 \cdots dx_n, \qquad (4.1)$$

which is defined on the domain of functions $u(x)$ for which the integral converges. By functional differentiation we obtain

$$\frac{\delta^s L_n}{\delta u(x_1) \cdots \delta u(x_s)}\bigg|_{u=0} = \frac{n!}{n^s (n-s)!} f_{n,s}(t, x_1, \cdots, x_s);$$

$$s = 0, 1, 2, \cdots, n. \qquad (4.2)$$

With the aid of (4.2), L_n may now be expressed as a (finite) series expansion around $u = 0$.

$$L_n[t, u] = 1 + \sum_{s=1}^{n} \frac{1}{s!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{s-1}{n}\right) \times \int f_{n,s}(t, x_1, \cdots, x_s) u(x_1) \cdots u(x_s)\, dx_1 \cdots dx_s. \qquad (4.3)$$

A differential equation for L_n may be obtained by multiplying (3.9) by

$$\prod_{i=1}^{n} [1 + v u(x_i)]$$

and integrating with respect to $x_1, \cdots, x_n$. We then let $n \to \infty$, $V \to \infty$ in the resulting equation. In this process we always assume that $V/n = v$ approaches a finite limit. Thus we obtain

$$\frac{\partial L}{\partial t} = \int \left[u(x_1) + \frac{1}{v}\right]\left[T(p_1); \frac{\delta L}{\delta u(x_1)}\right] dx_1 - \frac{1}{2} \int \left[u(x_1) + \frac{1}{v}\right]\left[u(x_2) + \frac{1}{v}\right] \theta_{12} \times \frac{\delta^2 L}{\delta u(x_1)\, \delta u(x_2)}\, dx_1\, dx_2. \qquad (4.4)$$

Here

$$T(p) = \frac{p^2}{2m} \quad \text{and} \quad L[t, u] = \lim_{\substack{n \to \infty \\ V \to \infty}} L_n[t, u]. \qquad (4.5)$$

If we set

$$f_s(t, x_1, \cdots, x_s) = \lim_{\substack{n \to \infty \\ V \to \infty}} f_{n,s}(t, x_1, \cdots, x_s)$$

and pass to the limit in (3.21) and (4.3), we obtain

$$\frac{\partial f_s}{\partial t} + \{f_s; T_s\} + \sum_{1 \le j < k \le s} \theta_{jk} \cdot f_s = -\frac{1}{v} \sum_{i=1}^{s} \times \int \theta_{i,s+1} \cdot f_{s+1}(t, x_1, \cdots, x_{s+1})\, dx_{s+1};$$

$$s = 1, 2, \cdots \qquad (4.6)$$

and

$$L[t, u] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \times \int f_s(t, x_1, \cdots, x_s) u(x_1) \cdots u(x_s)\, dx_1 \cdots dx_s. \qquad (4.7)$$

It follows from (4.7) that

$$f_s(t, x_1, \cdots, x_s) = \frac{\delta^s L}{\delta u(x_1) \cdots \delta u(x_s)}\bigg|_{u=0}. \qquad (4.8)$$

[12] G. Stell, Doctoral dissertation, New York University, (1961).

The equations of the infinite system (4.6) are called the quantum-mechanical *hierarchy equations*. The functional L is called the *generating functional* because it generates the functions f_s by means of (4.8).

In order to solve the initial value problem for (4.4), we begin by examining the case of zero density, $1/v = 0$. We shall use superscripts 0 to denote this case. It is convenient now to denote the independent function by w instead of u. Thus from (4.7), (4.4), and (4.8),

$$L^0[t, w] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \times \int f_s^0(t, x_1, \cdots, x_s) w(x_1) \cdots w(x_s)\, dx_1 \cdots dx_s, \tag{4.9}$$

$$\frac{\partial L^0}{\partial t} - \int w(x_1) \left[T(p_1); \frac{\delta L^0}{\delta w(x_1)} \right] dx_1 + \frac{1}{2} \int w(x_1) w(x_2) \theta_{12} \cdot \frac{\delta^2 L^0}{\delta w(x_1)\, \delta w(x_2)}\, dx_1\, dx_2 = 0, \tag{4.10}$$

$$f_s^0(t, x_1, \cdots, x_s) = \frac{\delta^s L^0}{\delta w(x_1) \cdots \delta w(x_s)} \bigg|_{w=0}; \tag{4.11}$$

and (4.6) reduces to

$$\frac{\partial f_s^0}{\partial t} + \{f_s^0; T_s\} + \sum_{1 \le i < k \le s} \theta_{ik} \cdot f_s^0 = 0; \qquad s = 1, 2, \cdots. \tag{4.12}$$

Comparing (4.12) with (3.9) we observe that f_s^0 satisfies the quantum-mechanical Liouville equation for s particles. It follows from (3.15) that the solution of (4.12) is given by

$$f_s^0(t, x_1, \cdots, x_s) = S_{-t}^{(s)} f_s^0(0, x_1, \cdots, x_s); \qquad s = 1, 2, \cdots, \tag{4.13}$$

where $S_t^{(s)}$ is the quantum-mechanical solution operator for a system of s particles. On inserting (4.13) in (4.9) we obtain the solution of (4.10) subject to the initial conditions

$$L^0[0, w] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \times \int f_s^0(0, x_1, \cdots, x_s) w(x_1) \cdots w(x_s)\, dx_1 \cdots dx_s. \tag{4.14}$$

The $f_s^0(0, x_1, \cdots, x_s)$ are the given initial data.

In order to solve the general equation (4.4) we compare it with (4.10). This comparison suggests that we try a solution of the form

$$L[t, u] = L^0[t, w]; \qquad w(x) = u(x) + 1/v. \tag{4.15}$$

Inserting this expression in (4.4) we see at once that the equation is satisfied by virtue of the fact that L^0 satisfies (4.10).

But (4.4) must be solved subject to the initial conditions

$$L[0, u] = \mathfrak{L}[u] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \times \int f_s(0, x_1, \cdots, x_s) u(x_1) \cdots u(x_s)\, dx_1 \cdots dx_s. \tag{4.16}$$

The functions $f_s(0, x_1, \cdots, x_s)$ are the given initial data. In terms of (4.15) the initial conditions become

$$L^0[0, w] = \mathfrak{L}[u]; \qquad u = w - 1/v. \tag{4.17}$$

Now let $L^0[t, w]$ be the solution of the initial value problem for (4.10) with initial conditions presented in (4.17). Then $L[t, u]$, defined by (4.15), is the solution of the initial value problem for (4.4) with initial conditions presented in (4.16).

One may now obtain expansions of the functions $f_s(t, x_1, \cdots, x_s)$ in powers of the density $1/v$ as follows: From (4.8) and (4.15),

$$f_s(t, x_1, \cdots, x_s) = \frac{\delta^s L[t, u]}{\delta u(x_1) \cdots \delta u(x_s)} \bigg|_{u=0} = \frac{\delta^s L^0[t, w]}{\delta w(x_1) \cdots \delta w(x_s)} \bigg|_{w=1/v}. \tag{4.18}$$

If we apply $\delta^s / \delta w(x_1) \cdots \delta w(x_s)$ to (4.9) and set $w = 1/v$, we obtain $f_s(t, x_1, \cdots, x_s)$ as a power series in $1/v$ with coefficients involving the functions $f_s^0(t, x_1, \cdots, x_s) = S_{-t}^{(s)} f_s^0(0, x_1, \cdots, x_s)$. Now from (4.11) and (4.17)

$$f_s^0(0, x_1, \cdots, x_s) = \frac{\delta^s L^0[0, w]}{\delta w(x_1) \cdots \delta w(x_s)} \bigg|_{w=0} = \frac{\delta^s \mathfrak{L}[u]}{\delta u(x_1) \cdots \delta u(x_s)} \bigg|_{u=-1/v}. \tag{4.19}$$

If we apply $\delta^s / \delta u(x_1) \cdots \delta u(x_s)$ to (4.16) and set $u = -1/v$, we obtain $f_s^0(0, x_1, \cdots, x_s)$ as a power series in $-1/v$ with coefficients involving the functions $f_s(0, x_1, \cdots, x_s)$, which are the given initial data. If we now insert the series expression for $f_s^0(0, x_1, \cdots, x_s)$ into the series expression for $f_s(t, x_1, \cdots, x_s)$ we obtain[13]

[13] Except for slight differences of notation, (4.20) is identical to Eq. (38) of reference 1. The latter was obtained by a process somewhat more involved than the one indicated here, in order to avoid introducing divergent integrals at an intermediate step.

$$f_s(t, x_1, \cdots, x_s) = \sum_{k=0}^{\infty} \left(\frac{1}{v}\right)^k \times \int \left[\sum_{j=0}^{k} \frac{(-1)^{k-j}}{j!\,(k-j)!} S_{-t}^{(j+s)}(1, \cdots, j+s) \times S_{-t}^{(k-j,0)}(j+s+1, \cdots, k+s) \times f_{k+s}(0, x_1, \cdots, x_{k+s}) \right] \times dx_{s+1}, \cdots, dx_{s+k}; \quad s = 1, 2, \cdots. \tag{4.20}$$

The series expansion (4.20) for f_s has the advantage that for small densities ($1/v \ll 1$) the function f_s is approximated by terminating the series after a few terms. Now the functions f_s of main interest are those for which s is small, and for these functions, the calculation of the leading terms of the expansion requires a knowledge only of solution operators $S_t^{(k)}$ where k is small, and initial data $f_j(0, x_1, \cdots, x_j)$ where j is small. The expansion (4.20) probably suffers, however, from the same defect as the corresponding expansion in the classical case, namely that the remainder terms grow rapidly with time.

In order to obtain the operator analog of (4.20) we must first introduce s-particle density operators

$$F_{n,s}(t, 1, \cdots, s) = V^n T_{s+1,\cdots,n}[Q_n D_n(t)], \tag{4.21}$$

and

$$F_s(t, 1, \cdots, s) = \lim_{\substack{n\to\infty \\ V\to\infty}} F_{n,s}(t, 1, \cdots, s). \tag{4.22}$$

Then it is easy to show that

$$f_s(t, x_1, \cdots, x_s) = W[F_s(t, 1, \cdots, s)] \tag{4.23}$$

and the analog of (4.20) is

$$F_s(t, 1, \cdots, s) = \sum_{k=0}^{\infty} \left(\frac{1}{v}\right)^k T_{s+1,\cdots,s+k} \times \left[\sum_{j=0}^{k} \frac{(-1)^{k-j}}{j!\,(k-j)!} \times \exp\left\{-\frac{it}{\hbar}[H_{j+s}(1, \cdots, j+s) + H_{k-j,0}(j+s+1, \cdots, k+s)]\right\} \times F_{k+s}(0, 1, \cdots, k+s) \times \exp\left\{\frac{it}{\hbar}[H_{j+s}(1, \cdots, j+s) + H_{k-j,0}(j+s+1, \cdots, k+s)]\right\}\right]; \quad s = 1, 2, \cdots. \tag{4.24}$$

In the case of classical mechanics, s-particle density functions are defined as in (3.20), where w_n is replaced by the classical density function. They satisfy a system of hierarchy equations, and the solution of the initial value problem for these equations is given by (4.20) where the solution operators are those appropriate to classical mechanics. These matters are discussed in detail in reference 1. In that paper we applied the leading terms of the expansion (4.20) for $s = 1$ to obtain a simple derivation of the Boltzmann equation. In the next section, we shall attempt to derive the quantum-mechanical analog of the Boltzmann equation in a similar manner.

5. DERIVATION OF THE QUANTUM-MECHANICAL BOLTZMANN EQUATION

In 1933, Uehling and Uhlenbeck[14] suggested that certain modifications of the collision integral of the classical Boltzmann equation would be appropriate to the case of quantum mechanics. Since that time various authors[10,15–19] have attempted to verify the validity of the resulting equation by deriving it from the basic formulas of quantum statistical mechanics. In our opinion, none of these derivations, including the one to be presented here, is entirely satisfactory. Some of the inconsistencies of the earlier derivations can be removed by proper use of operator representations and application of recent developments in time-dependent scattering theory. The purpose of the present derivation is to remove these inconsistencies and to isolate what we consider to be the real difficulty.

At one point in our derivation we, like the earlier authors, are forced to argue by analogy with the classical case. It is at this point that we feel the basic difficulty remains. After the derivation is completed we shall add some concluding remarks about this difficulty and the possibility of removing it.

The starting point of our derivation is the solution (4.20) of the quantum-mechanical hierarchy equations. Following the suggestion of Kirkwood[20] and others, we shall make use of a time average of f_1. The form of the molecular chaos assumption appropriate to each of our three cases ($Q_n = S_n$,

[14] E. A. Uehling and G. E. Uhlenbeck, Phys. Rev. **43**, 552 (1933).
[15] H. Mori and S. Ono, Progr. Theoret Phys. (Kyoto) **8**, 327 (1952).
[16] S. Ono, Progr. Theoret Phys. (Kyoto) **12**, 113 (1954).
[17] J. Ross and J. Kirkwood, J. Chem. Phys. **22**, 1094 (1954).
[18] A. W. Säenz, Phys. Rev. **105**, 545 (1957).
[19] H. Mori and J. Ross, Phys. Rev. **109**, 1877 (1958).
[20] J. Kirkwood, J. Chem. Phys. **14**, 180 (1946).

$Q_n = \mathfrak{a}_n$, $Q_n = 1$) may be stated in terms of density operators as follows:

$$F_2(t, 1, 2) = Q_2 F_1(t, 1) F_1(t, 2). \tag{5.1}$$

The justification of this assumption is discussed in reference 10. We shall assume that it holds for all t. It follows from (3.5), (3.16), and (4.23) that

$$f_2(t, x_1, x_2) = \mathfrak{Q}_2 f_1(t, x_1) f_1(t, x_2). \tag{5.2}$$

Thus (5.2) is the Wigner representation of the molecular chaos assumption.

If we set $s = 1$ in (4.20), we obtain

$$f_1(t, x_1) = S^{(1)}_{-t} f_1(0, x_1) + (1/v) \int [S^{(2)}_{-t} f_2(0, x_1, x_2) - S^{(1)}_{-t}(1) S^{(1)}_{-t}(2) f_2(0, x_1, x_2)]\, dx_2 + O(1/v^2). \tag{5.3}$$

Since initial values may be specified at an arbitrary time (say at $t - \tau/2$) we obtain, by replacing t by τ,

$$f_1(t + \tau/2, x_1) = S^{(1)}_{-\tau} f_1(t - \tau/2, x_1) + \frac{1}{v} \int [S^{(2)}_{-\tau} f_2(t - \tau/2, x_1, x_2) - S^{(1)}_{-\tau}(1) S^{(1)}_{-\tau}(2) f_2(t - \tau/2, x_1, x_2)]\, dx_2 + O(1/v^2). \tag{5.4}$$

In a similar manner we obtain

$$f_1(t - \tau/2, x_1) = S^{(1)}_{\tau/2} f_1(t, x_1) + O(1/v);$$
$$S^{(1)}_s f_1(t + s, x_1) = f_1(t, x_1) + O(1/v). \tag{5.5}$$

If we apply the operator $S^{(1)}_{\tau/2}(1)$ to (5.4) the result is

$$S^{(1)}_{\tau/2} f_1(t + \tau/2, x_1) - S^{(1)}_{-\tau/2} f_1(t - \tau/2, x_1) = Z, \tag{5.6}$$

where

$$Z = \frac{1}{v} \int [S^{(1)}_{\tau/2}(1) S^{(2)}_{-\tau} f_2(t - \tau/2, x_1, x_2) - S^{(1)}_{-\tau/2}(1) S^{(1)}_{-\tau}(2) f_2(t - \tau/2, x_1, x_2)]\, dx_2 + O(1/v^2). \tag{5.7}$$

We now introduce the transformation of integration variables, $x_2 = S^{(1)}_{\tau/2} x_2'$ in (5.7). Since $S^{(1)}_t x = S^{(1)}_t(q, p) = (q + (t/m)p, p)$ we see that $dx_2' = dx_2$. If we then change x_2' to x_2 in the resulting integral and note that $S^{(2,0)}_t(1, 2) = S^{(1)}_t(1) S^{(1)}_t(2)$, we obtain

$$Z = \frac{1}{v} \int [S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} f_2(t - \tau/2, x_1, x_2) - S^{(1)}_{-\tau/2}(1) S^{(1)}_{-\tau/2}(2) f_2(t - \tau/2, x_1, x_2)]\, dx_2 + O(1/v^2). \tag{5.8}$$

From the molecular chaos assumption (5.2) and, from (5.5),

$$f_2(t - \tau/2, x_1, x_2) = \mathfrak{Q}_2 f_1(t - \tau/2, x_1) f_1(t - \tau/2, x_2)$$
$$= \mathfrak{Q}_2 S^{(1)}_{\tau/2}(1) f_1(t, x_1) S^{(1)}_{\tau/2}(2) f_1(t, x_2) + O(1/v)$$
$$= \mathfrak{Q}_2 S^{(2,0)}_{\tau/2} f_1(t, x_1) f_1(t, x_2) + O(1/v). \tag{5.9}$$

Thus, since $\mathfrak{Q}_2$ commutes with $S^{(2)}_t$ and $S^{(2,0)}_t$,

$$Z = \frac{1}{v} \int [S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2} \mathfrak{Q}_2 f_1(t, x_1) f_1(t, x_2) - \mathfrak{Q}_2 f_1(t, x_1) f_1(t, x_2)]\, dx_2 + O(1/v^2). \tag{5.10}$$

We now introduce a time average f of the one-particle Wigner function f_1 defined by

$$f(t, x_1) = \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} S^{(1)}_s f_1(t + s, x_1)\, ds, \tag{5.11}$$

and we define the total time derivative of a function $h(t, x_1)$,

$$D_t h(t, x_1) = (d/dz) S^{(1)}_z h(t + z, x_1)\,|_{z=0}. \tag{5.12}$$

Then

$$D_t h(t, x_1) = (d/dt) h(t + z, q_1 + (z/m)p_1, p_1)\,|_{z=0} = [\partial/\partial t + (1/m)p_1 \cdot \nabla_{q_1}] h. \tag{5.13}$$

From (5.11), (5.12), and (5.6),

$$D_t f(t, x_1) = \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} \frac{d}{ds} S^{(1)}_s f_1(t + s, x_1)\, ds$$
$$= 1/\tau [S^{(1)}_{\tau/2} f_1(t + \tau/2, x_1) - S^{(1)}_{-\tau/2} f_1(t - \tau/2, x_1)]$$
$$= (1/\tau) Z, \tag{5.14}$$

and from (5.11) and (5.5)

$$f_1(t, x_1) = f(t, x_1) + O(1/v). \tag{5.15}$$

If we insert (5.15) into (5.10), and the latter in (5.14) we obtain

$$D_t f(t, x_1) = \frac{1}{\tau v} \int [S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2} \mathfrak{Q}_2 f(t, x_1) f(t, x_2) - \mathfrak{Q}_2 f(t, x_1) f(t, x_2)]\, dx_2 + O(1/\tau v^2). \tag{5.16}$$

We recognize the left side of (5.16) as that of the Boltzmann equation. Our object now will be to transform the right side into the Uehling-Uhlenbeck modification of the collision integral.

So far the analysis in this section is valid for the case of classical mechanics as well as for the case of quantum mechanics, provided we interpret the solution operators according to their classical definition, and take $\mathfrak{Q}_2 = 1$. Let us, for the moment, assume that we are dealing with the classical case and examine the solution operators appearing in (5.16). The following discussion depends on the fact that the classical solution operator is defined

first for points in phase space (3.17) and then for functions (3.18). With the aid of Fig. 1 we examine the effect of the operator $S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2}$ on points $\{x_1, x_2\} = \{q_1, p_1, q_2, p_2\}$ for fixed values of p_1, p_2, and q_1, and various choices of q_2.

In Fig. 1, A, B, C together form a cylindrical region with spherical end surfaces. The radius of the cross section of the cylinder is r_1, the interaction distance of the interparticle potential; i.e., we assume that

$$\phi(r) = 0 \quad \text{for} \quad r \geq r_1. \tag{5.17}$$

For points q_2 in C, the operator $S^{(2,0)}_{\tau/2}$ maps q_2 to q_2', the position of $S^{(2,0)}_{\tau/2} q_2$ in a coordinate system in which q_1 is fixed. Under $S^{(2,0)}_{\tau/2}$, q_2 moves relative to q_1, a distance $(\tau/2m)\,|p_2 - p_1|$ in the direction $p_2 - p_1$, along a trajectory which is a straight line because $S^{(2,0)}_{\tau/2}$ corresponds to a Hamiltonian $H_{2,0}$ with no interaction. Then $S^{(2)}_{-\tau}$ maps q_2' to q_2'' along a trajectory which is first straight, then curved in the interaction region (the sphere of radius r_1 about q_1) and then straight, with direction $-(\hat{p}_2 - \hat{p}_1)$. Finally $S^{(2,0)}_{\tau/2}$ maps q_2'' back along the straight trajectory with direction $\hat{p}_2 - \hat{p}_1$ to q_2'''. The momenta p_1, p_2 are unaffected by the operators $S^{(2,0)}_{\tau/2}$ but are transformed into $\hat{p}_1$, $\hat{p}_2$ by $S^{(2)}_{-\tau}$. Now if τ is increased, the points q_2' and q_2'' move outward along the straight trajectories but q_2''' does not change its position. Thus, for points q_2 in C,

$$S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2} \{x_1, x_2\} = \lim_{\tau \to \infty} S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2} \{x_1, x_2\}. \tag{5.18}$$

It is easy to see that points outside the region A, B, C simply map into themselves under the operator $S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2}$. (Of course, the length of the cylindrical region increases linearly with τ.) It follows that the integrand in (5.16) is zero for points q_2 outside the region A, B, C, hence, certainly for points in the region

$$|q \cdot \rho| > (\tau/2m)\,|p_2 - p_1| + r_1, \tag{5.19}$$

where

$$q = q_2 - q_1; \qquad \rho = (p_2 - p_1)/|p_2 - p_1|. \tag{5.20}$$

Therefore, we can replace $f(t, x_1) f(t, x_2)$ in (5.16) by a function $g(t, q_1, q_2, p_1, p_2)$ where

$$g = \begin{Bmatrix} f\,f \text{ for } |q \cdot \rho| \leq (\tau/2m)\,|p_2 - p_1| + r_1 \\ 0 \text{ for } |q \cdot \rho| > (\tau/2m)\,|p_2 - p_1| + r_1 \end{Bmatrix}. \tag{5.21}$$

We now assume that f is *spatially homogeneous*, i.e., independent of q_1 for all t:

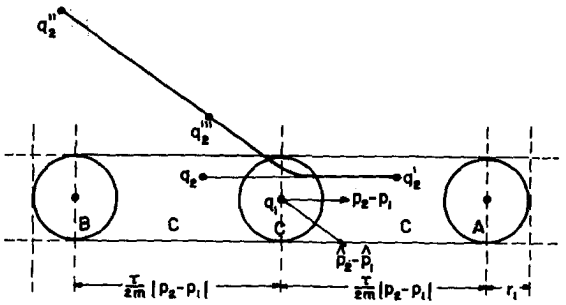


FIG. 1. Configuration space for classical binary collision.

$$f(t, q_1, p_1) = f(t, p_1). \tag{5.22}$$

It follows that g depends on q_1, q_2 only through the value of $(q \cdot \rho)$:

$$g(t, q_1, q_2, p_1, p_2) = g(t, q \cdot \rho, p_1, p_2) = g\left(t, q \cdot \frac{(p_2 - p_1)}{|p_2 - p_1|}, p_1, p_2\right) \tag{5.23}$$

and from (5.21) we may easily derive the identity

$$\frac{1}{\tau v} \int_{-\infty}^{\infty} g(t, q \cdot \rho, p_1, p_2)\, d(q \cdot \rho) = (|p_2 - p_1|/vm) f(t, p_1) f(t, p_2) + O(1/\tau v). \tag{5.24}$$

After replacing ff by g in (5.16), we see from (5.18) that we may replace the solution operators by their limiting values (for $\tau \to \infty$) except for q_2 in the spheres A and B (where incomplete collisions occur). The fact that the cylinder increases in length with τ (and in fact becomes infinite for $\tau \to \infty$) is rendered harmless by the fact that g vanishes for $|q \cdot \rho| > (\tau/2m)\,|p_2 - p_1| + r_1$. Thus we obtain

$$D_t f(t, p_1) = \frac{1}{\tau v} \int [\lim_{\tau \to \infty} S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2} \mathcal{Q}_2 g - \mathcal{Q}_2 g] \times dq_2\, dp_2 + O(1/\tau v). \tag{5.25}$$

The term $O(1/\tau v)$ represents the error due to incomplete collisions.

Although we have derived (5.25) and (5.24) by an argument which is valid only for the case of classical mechanics, we shall nevertheless proceed to show that the Uehling-Uhlenbeck form of the quantum-mechanical Boltzmann equation follows from (5.24) and (5.25) provided the solution operators and the operator $\mathcal{Q}_2$ in the latter are given their quantum-mechanical meaning.

Let G be the operator whose Wigner representer is g, i.e.,

$$W(G) = g(t, q_1, q_2, p_1, p_2). \tag{5.26}$$

Then from (3.16)

$$W(Q_2G) = \mathfrak{Q}_2 g. \tag{5.27}$$

Let

$$a(t, q_1, q_2, p_1, p_2) = \lim_{\tau\to\infty} S^{(2,0)}_{\tau/2} S^{(2)}_{-\tau} S^{(2,0)}_{\tau/2} \mathfrak{Q}_2 g - \mathfrak{Q}_2 g = W(A). \tag{5.28}$$

Then from our definition of the solution operators (3.12),

$$A = \lim_{\tau\to\infty} \exp\left(\frac{i\tau}{2\hbar} H_{2,0}\right) \exp\left(-\frac{i\tau}{\hbar} H_2\right) \times \exp\left(\frac{i\tau}{2\hbar} H_{2,0}\right) Q_2G \exp\left(-\frac{i\tau}{2\hbar} H_{2,0}\right) \times \exp\left(\frac{i\tau}{\hbar} H_2\right) \exp\left(-\frac{i\tau}{2\hbar} H_{2,0}\right) - Q_2G. \tag{5.29}$$

Here H_2 and $H_{2,0}$ are the Hamiltonians corresponding to $S^{(2)}_t$ and $S^{(2,0)}_t$, i.e.,

$$H_2 = T_2 + U_2 = (1/2m)(p_1^2 + p_2^2) + \phi(|q_1 - q_2|); \tag{5.30}$$

$$H_{2,0} = T_2 = (1/2m)(p_1^2 + p_2^2). \tag{5.31}$$

In order to apply the results of scattering theory to (5.29) we introduce center-of-mass and relative coordinates:

$$\left.\begin{aligned} q_c &= \tfrac{1}{2}(q_1 + q_2), \quad q = q_2 - q_1, \\ q_1 &= q_c - \tfrac{1}{2}q, \quad q_2 = q_c + \tfrac{1}{2}q \\ p_c &= p_1 + p_2, \quad p = \tfrac{1}{2}(p_2 - p_1), \\ p_1 &= \tfrac{1}{2}p_c - p, \quad p_2 = \tfrac{1}{2}p_c + p \\ M &= 2m, \quad \mu = \frac{m}{2}, \quad H_c = \frac{1}{2M}p_c^2, \\ H &= E + \phi(|q|), \quad E = \frac{1}{2\mu}p^2 \end{aligned}\right\}. \tag{5.32}$$

It follows that

$$H_2 = H + H_c; \quad H_{2,0} = E + H_c. \tag{5.33}$$

From (5.29),

$$A = \lim_{\tau\to\infty} N(\tau) Q_2 G N^*(\tau) - Q_2 G, \tag{5.34}$$

where

$$N(\tau) = \exp\left(\frac{i\tau}{2\hbar} H_{2,0}\right) \exp\left(-\frac{i\tau}{2\hbar} H_2\right) \times \exp\left(-\frac{i\tau}{2\hbar} H_2\right) \exp\left(\frac{i\tau}{2\hbar} H_{2,0}\right) = M(\tau) M^*(-\tau). \tag{5.35}$$

Here

$$M(\tau) = \exp\left(\frac{i\tau}{2\hbar} H_{2,0}\right) \exp\left(-\frac{i\tau}{2\hbar} H_2\right). \tag{5.36}$$

Since $[E, H_c] = 0$ and $[H, H_c] = 0$, it follows that

$$M(\tau) = \exp\left[\frac{i\tau}{2\hbar}(E + H_c)\right] \exp\left[-\frac{i\tau}{2\hbar}(H_c + H)\right] = \exp\left[\frac{i\tau}{2\hbar} E\right] \exp\left[-\frac{i\tau}{2\hbar} H\right]. \tag{5.37}$$

From the theory of time-dependent quantum-mechanical scattering,[4,6,21] under appropriate conditions[22] on the 'perturbing operator' $\phi(|q|)$, we have the following result:

$$\lim_{\tau\to\infty} M(\tau) = SU; \quad \lim_{\tau\to\infty} M(-\tau) = U. \tag{5.38}$$

Here S is the quantum-mechanical *scattering operator* and U is a unitary operator which transforms the unperturbed Hamiltonian E into the perturbed Hamiltonian H:

$$H = U^*EU; \quad UU^* = U^*U = 1. \tag{5.39}$$

The scattering operator is also unitary, i.e.,

$$SS^* = S^*S = 1. \tag{5.40}$$

From (5.38)

$$\lim_{\tau\to\infty} M^*(\tau) = U^*S^*; \quad \lim_{\tau\to\infty} M^*(-\tau) = U^*. \tag{5.41}$$

Thus from (5.35)

$$\lim_{\tau\to\infty} N(\tau) = SUU^* = S, \tag{5.42}$$

and from (5.34)

$$A = SQ_2GS^* - Q_2G. \tag{5.43}$$

It is somewhat more convenient to work with the operator $J = S - 1$ instead of the scattering operator S. Thus

$$S = 1 + J, \tag{5.44}$$

and from (5.43),

$$A = JQ_2G + Q_2GJ^* + JQ_2GJ^*. \tag{5.45}$$

From its definition (5.21) we see that g is a symmetric function, i.e., $g(t, x_1, x_2) = g(t, x_2, x_1)$. From this it follows that $G = G(1, 2)$ is a symmetric operator, and from (5.42) and (5.35) we see that S and J are symmetric operators, since H_2 and $H_{2,0}$

[21] H. E. Moses, Nuovo cimento 1, 103 (1955).

[22] The conditions given in references 4 and 6 are such as to exclude point eigenvalues of H. In reference 21 the theory is extended to include point eigenvalues.

are. It follows from (2.27) that G and J commute with Q_2. Since $Q_2 = Q_2^2$ and $Q_2^* = Q_2$, (5.45) becomes

$$\begin{aligned} A &= JQ_2G + G(JQ_2)^* + JQ_2(JQ_2)^* \\ &= TG + GT^* + TGT^*, \end{aligned} \tag{5.46}$$

where

$$T = JQ_2. \tag{5.47}$$

From (5.40) and (5.44)

$$J + J^* = -JJ^* = -J^*J \tag{5.48}$$

hence from (5.47)

$$T + T^* = -TT^* = -T^*T \tag{5.49}$$

The operator J is most conveniently discussed by means of the representation in terms of a complete set of commuting observables that includes the unperturbed Hamiltonian E; hence, we shall introduce that representation. To do this, let

$$r = |p| = \tfrac{1}{2}|p_2 - p_1|;$$
$$\rho = r^{-1}p = (p_2 - p_1)/|p_2 - p_1|. \tag{5.50}$$

Then ρ is a unit vector, and

$$E = r^2/(2\mu) = |p_2 - p_1|^2/(8\mu). \tag{5.51}$$

From (5.32) one may determine that the Jacobian of the transformation from p_1, p_2 to p, p_c is 1, i.e.,

$$dp_1\, dp_2 = dp\, dp_c \tag{5.52}$$

and from (5.50) we see that

$$dp = r^2\, dr\, dA(\rho). \tag{5.53}$$

Here $dA(\rho)$ is the area element on the unit sphere $|\rho| = 1$. From (5.53) we obtain

$$dp = \lambda^2\, dE\, dA(\rho), \tag{5.54}$$

where

$$\lambda = (2\mu^3 E)^{1/4} = (\mu r)^{1/2}. \tag{5.55}$$

In Sec. 2 we introduced the momentum representation R. Let

$$\begin{aligned} &\phi_0(E, \rho, p_c, p_3, \cdots, p_n) \\ &\quad = \lambda\phi(p_1, p_2, p_3, \cdots, p_n) = \lambda R(\Phi). \end{aligned} \tag{5.56}$$

Then from (2.8), (5.32), and (5.51)

$$\begin{aligned} \lambda R(E\Phi) &= \lambda(p^2/2\mu)\phi(p_1, \cdots, p_n) \\ &= E\phi_0(E, \rho, p_c, p_3, \cdots, p_n), \end{aligned} \tag{5.57}$$

and from (2.51), (5.52), (5.54), and (5.56)

$$\begin{aligned} ||\Phi||^2 &= \int |\phi|^2 \lambda^2\, dE\, dA(\rho)\, dp_c\, dp_3 \cdots dp_n \\ &= \int |\phi_0|^2\, dE\, dA(\rho)\, dp_c\, dp_3 \cdots dp_n. \end{aligned} \tag{5.58}$$

(The integrations indicated are, with respect to E over the real semi-axis $0 \leq E$, with respect to ρ over the unit sphere, and with respect to the vectors $p_c, p_3, \cdots, p_n$ over all space.) It follows from (5.56)–(5.58) that the correspondence (5.56) between elements Φ of $\mathfrak{H}$ and functions $\phi_0(E, \rho, p_c, p_3, \cdots, p_n)$ is the representation in terms of the unperturbed Hamiltonian operator E and the operators ρ, p_c, $p_3, \cdots, p_n$. For brevity we denote this representation by R_0. Thus

$$R_0(\Phi) = \phi_0(E, \rho, p_c, p_3, \cdots, p_n) = \lambda R(\Phi). \tag{5.59}$$

Now, in terms of the representation R_0, the operator J is given by[4,6,21]

$$\begin{aligned} R_0(J\Phi) = \int j(E, \rho, \rho') \\ \times \phi_0(E, \rho', p_c, p_3, \cdots, p_n)\, dA(\rho'), \end{aligned} \tag{5.60}$$

and is related to the quantum-mechanical *differential scattering cross section*, $\eta(E, \rho, \rho')$, by

$$\eta(E, \rho, \rho') = (2\pi\hbar)^2\, |r^{-1}j(E, \rho, \rho')|^2. \tag{5.61}$$

The operator J can be put in the form of an integral operator,

$$\begin{aligned} R_0(J\Phi) = \int j(E, \rho, \rho')\, \delta(E - E')\, \delta(p_c - p_c') \\ \times \delta(p_3 - p_3') \cdots \delta(p_n - p_n') \\ \times \phi_0(E', \rho', p_c', p_3', \cdots, p_n') \\ \times dE'\, dA(\rho')\, dp_c'\, dp_3' \cdots dp_n', \end{aligned} \tag{5.62}$$

with a singular kernel. In earlier treatments, representations of J of the form (5.62) were used, and this led to inconsistencies involving the square of the δ function. To avoid these difficulties we shall make no use of (5.62) and shall work directly from (5.60).

In the two cases $Q_2 = \mathfrak{S}_2$, and $Q_2 = \mathfrak{A}_2$, we have, from (2.42),

$$Q_2 = \tfrac{1}{2}(1 \pm P_{1,2}), \tag{5.63}$$

where $P_{1,2}$ is the permutation that interchanges 1 and 2. From (2.17)

$$\begin{aligned} R(Q_2\Phi) = \tfrac{1}{2}[\phi_0(p_1, p_2, p_3, \cdots, p_n) \\ \pm \phi(p_2, p_1, p_3, \cdots, p_n)] \end{aligned} \tag{5.64}$$

and from (5.56) and (5.50)

$$\begin{aligned} R_0(Q_2\Phi) = \tfrac{1}{2}[\phi_0(E, \rho, p_c, p_3, \cdots, p_n) \\ \pm \phi_0(E, -\rho, p_c, p_3, \cdots, p_n)]. \end{aligned} \tag{5.65}$$

Thus, from (5.47) and (5.60)

$$R_0(T\Phi) = R_0(JQ_2\Phi)$$
$$= \tfrac{1}{2}\int j(E, \rho, \rho')[\phi_0(E, \rho', p_c, p_3, \cdots, p_n) \pm \phi_0(E, -\rho', p_c, p_3, \cdots, p_n)]\, dA(\rho')$$
$$= \int b(E, \rho, \rho')\phi_0(E, \rho', p_c, p_3, \cdots, p_n)\, dA(\rho'), \tag{5.66}$$

where

$$b(E, \rho, \rho') = \tfrac{1}{2}[j(E, \rho, \rho') \pm j(E, \rho, -\rho')]. \tag{5.67}$$

Corresponding to these two cases, we define the symmetrized, and the antisymmetrized cross sections

$$\sigma(E, \rho, \rho') = (2\pi\hbar)^2\, |r^{-1}b(E, \rho, \rho')|^2$$
$$= (2\pi\hbar)^2\, |(1/2r)[j(E, \rho, \rho') \pm j(E, \rho, -\rho')]|^2. \tag{5.68}$$

Of course, in the case of Maxwell-Boltzmann statistics, $Q_2 = 1$, $T = J$, $b = j$, and $\sigma = \eta$.

From (5.66)

$$R_0(T^*\Phi) = \int b^*(E, \rho', \rho) \times \phi_0(E, \rho', p_c, p_3, \cdots, p_n)\, dA(\rho'), \tag{5.69}$$

and from (5.49) we obtain

$$b(E, \rho, \rho') + b^*(E, \rho', \rho) = -\int b(E, \rho, \rho'')b^*(E, \rho', \rho'')\, dA(\rho''). \tag{5.70}$$

For $\rho = \rho'$ this becomes

$$b(E, \rho, \rho) + b^*(E, \rho, \rho) = -\int |b(E, \rho, \rho')|^2\, dA(\rho'). \tag{5.71}$$

Our object now is to express the operator A, as given by (5.46) first in R_0 representation and then in R representation. From the latter it will be easy to transform to Wigner representation and then to evaluate the right side of (5.25) using (5.28). In order to do this, let

$$R(G) = \gamma(p_1, p_2, p_1', p_2') \tag{5.72}$$

and

$$R_0(G) = \gamma_0(E, \rho, p_c, E', \rho', p_c'), \tag{5.73}$$

i.e.,

$$R_0(G\Phi) = \int \gamma_0(E, \rho, p_c, E', \rho', p_c') \times \phi_0(E', \rho', p_c', p_3, \cdots, p_n)\, dE'\, dA(\rho')\, dp_c'. \tag{5.74}$$

Then, from (5.59), (5.56), (5.52), (5.54), and (5.72),

$$\lambda R(G\Phi) = R_0(G\Phi) = \int \gamma_0(E, \rho, p_c, E', \rho', p_c')\lambda' \times \phi(p_1', p_2', p_3, \cdots, p_n)(\lambda')^{-2}\, dp_1'\, dp_2'$$
$$= \lambda \int \gamma(p_1, p_2, p_1', p_2')\phi(p_1', p_2', p_3, \cdots, p_n)\, dp_1'\, dp_2'. \tag{5.75}$$

From this it follows that

$$(\lambda\lambda')^{-1}\gamma_0(E, \rho, p_c, E', \rho', p_c') = \gamma(p_1, p_2, p_1', p_2') = \gamma(\tfrac{1}{2}p_c - r\rho, \tfrac{1}{2}p_c + r\rho, \tfrac{1}{2}p_c' - r'\rho', \tfrac{1}{2}p_c' + r'\rho'). \tag{5.76}$$

Although γ and γ_0 were defined as the R and R_0 representers of $G(1, 2)$, it is clear that (5.76) is an identity which holds for the R and R_0 representers of an arbitrary operator $B(1, 2)$.

Now from (5.66) and (5.73)

$$R_0(TG\Phi) = \int b(E, \rho, \rho'')\gamma_0(E, \rho'', p_c, E', \rho', p_c') \times \phi_0(E', \rho', p_c', p_3, \cdots, p_n) \times dE'\, dA(\rho')\, dp_c'\, dA(\rho''). \tag{5.77}$$

Similarly, from (5.69) and (5.73)

$$R_0(GT^*\Phi) = \int b^*(E', \rho', \rho'')\gamma_0(E, \rho, p_c, E', \rho'', p_c') \times \phi_0(E', \rho', p_c', p_3, \cdots, p_n) \times dE'\, dA(\rho')\, dp_c'\, dA(\rho''), \tag{5.78}$$

and

$$R_0(TGT^*\Phi) = \int b(E, \rho, \rho''')b^*(E', \rho', \rho'') \times \gamma_0(E, \rho''', p_c, E', \rho'', p_c') \times \phi_0(E', \rho', p_c', p_3, \cdots, p_n) \times dE'\, dA(\rho')\, dp_c'\, dA(\rho'')\, dA(\rho'''). \tag{5.79}$$

Let $R_0(A) = \alpha_0(E, \rho, p_c, E', \rho', p_c')$. Then, from (5.46),

$$\alpha_0(E, \rho, p_c, E', \rho', p_c') = R_0(A)$$
$$= \int b(E, \rho, \rho'')\gamma_0(E, \rho'', p_c, E', \rho', p_c')\, dA(\rho'') + \int b^*(E', \rho', \rho'')\gamma_0(E, \rho, p_c, E', \rho'', p_c')\, dA(\rho'') + \int b(E, \rho, \rho''')b^*(E', \rho', \rho'') \times \gamma_0(E, \rho''', p_c, E', \rho'', p_c')\, dA(\rho'')\, dA(\rho'''). \tag{5.80}$$

If we set $R(A) = \alpha(p_1, p_2, p_1', p_2')$, then (5.76) is valid for α and α_0 as well as for γ and γ_0. Thus, from (5.76) and (5.80),

$$\alpha(p_1, p_2, p_1', p_2') = R(A) = \int b(E, \rho, \rho'')\gamma$$
$$\times (\tfrac{1}{2}p_c - r\rho'', \tfrac{1}{2}p_c + r\rho'', p_1', p_2')\, dA(\rho'')$$
$$+ \int b^*(E', \rho', \rho'')\gamma$$
$$\times (p_1, p_2, \tfrac{1}{2}p_c' - r'\rho'', \tfrac{1}{2}p_c' + r'\rho'')\, dA(\rho'')$$
$$+ \int b(E, \rho, \rho''')b^*(E', \rho', \rho'')$$
$$\times \gamma(\tfrac{1}{2}p_c - r\rho''', \tfrac{1}{2}p_c + r\rho''', \tfrac{1}{2}p_c' - r'\rho'', \tfrac{1}{2}p_c' + r'\rho'')$$
$$\times dA(\rho'')\, dA(\rho'''). \tag{5.81}$$

By definition $\gamma = R(G)$ and $g = W(G)$. Hence from (3.3)

$$\gamma(p_1, p_2, p_1', p_2')$$
$$= \int \exp\left\{\frac{i}{\hbar}[s_1\cdot(p_1' - p_1) + s_2\cdot(p_2' - p_2)]\right\}$$
$$\times g[t, s_1, s_2, \tfrac{1}{2}(p_1 + p_1'), \tfrac{1}{2}(p_2 + p_2')]\, ds_1\, ds_2. \tag{5.82}$$

We introduce new variables of integration, $s = s_2 - s_1$, $s_c = \frac{1}{2}(s_1 + s_2)$. Then $ds_1 ds_2 = ds ds_c$, and

$$\gamma(p_1, p_2, p_1', p_2')$$
$$= \int \exp\left\{\frac{i}{\hbar}[s_c\cdot(p_1' + p_2' - p_1 - p_2)\right.$$
$$\left. + \tfrac{1}{2}s\cdot(p_1 - p_1' + p_2' - p_2)]\right\}$$
$$\times g[\tfrac{1}{2}(p_1 + p_1'), \tfrac{1}{2}(p_2 + p_2')]\, ds\, ds_c. \tag{5.83}$$

Here we do not explicitly indicate the dependence of g on t, s, and s_c. But from (5.23) g is independent of s_c; therefore,

$$\gamma(p_1, p_2, p_1', p_2') = (2\pi\hbar)^3\, \delta(p_1' + p_2' - p_1 - p_2)$$
$$\times \int \exp\left[\frac{i}{2\hbar}s\cdot(p_1 - p_1' + p_2' - p_2)\right]$$
$$\times g[\tfrac{1}{2}(p_1 + p_1'), \tfrac{1}{2}(p_2 + p_2')]\, ds. \tag{5.84}$$

If we insert (5.84) in (5.81) we obtain

$$\alpha(p_1, p_2, p_1', p_2') = (2\pi\hbar)^3\, \delta(p_1' + p_2' - p_1 - p_2)$$
$$\times \left\{\int b(E, \rho, \rho'') \exp\left[\frac{i}{\hbar}s\cdot(r'\rho' - r\rho'')\right]\right.$$
$$\times g[\tfrac{1}{2}(2p_1' + r'\rho' - r\rho''),$$
$$\tfrac{1}{2}(2p_2' - r'\rho' + r\rho'')]\, ds\, dA(\rho'')$$
$$+ \int b^*(E', \rho', \rho'') \exp\left[\frac{i}{\hbar}s\cdot(r'\rho'' - r\rho)\right]$$
$$\times g[\tfrac{1}{2}(2p_1 + r\rho - r'\rho''),$$
$$\tfrac{1}{2}(2p_2 - r\rho + r'\rho'')]\, ds\, dA(\rho'')$$
$$+ \int b(E, \rho, \rho''')b^*(E', \rho', \rho'')$$
$$\times \exp\left[\frac{i}{\hbar}s\cdot(r'\rho'' - r\rho''')\right]$$
$$\times g[\tfrac{1}{2}(p_c - r\rho''' - r'\rho''),$$
$$\left.\tfrac{1}{2}(p_c + r\rho''' + r'\rho'')]\, ds\, dA(\rho'')\, dA(\rho''')\right\}. \tag{5.85}$$

By definition $\alpha = R(A)$ and $a = W(A)$. Therefore, from (3.2)

$$a(t, q_1, q_2, p_1, p_2) = (\pi\hbar)^{-6}$$
$$\times \int \exp\left[\frac{-2i}{\hbar}(q_1\cdot z_1 + q_2\cdot z_2)\right]$$
$$\times \alpha(p_1 - z_1, p_2 - z_2, p_1 + z_1, p_2 + z_2)\, dz_1\, dz_2, \tag{5.86}$$

and from (5.25) and (5.28)

$$D_t f(t, p_1) = \frac{1}{\tau v}(\pi\hbar)^{-6} \int \exp\left[\frac{-2i}{\hbar}(q_1\cdot z_1 + q_2\cdot z_2)\right]$$
$$\times \alpha(p_1 - z_1, p_2 - z_2, p_1 + z_1, p_2 + z_2)$$
$$\times dz_1\, dz_2\, dq_2\, dp_2 + O(1/\tau v). \tag{5.87}$$

The integrations with respect to q_2 and z_2 may be carried out in (5.87). The result is

$$D_t f(t, p_1) = \frac{1}{\tau v}(\pi\hbar)^{-3} \int \exp\left(\frac{-2i}{\hbar}q_1\cdot z\right)$$
$$\times \alpha(p_1 - z, p_2, p_1 + z, p_2)\, dz\, dp_2 + O(1/\tau v). \tag{5.88}$$

We now insert (5.85) in (5.88). In so doing we replace p_1 by $p_1 - z$, p_1' by $p_1 + z$, and p_2' by p_2. Then $\delta(p_1' + p_2' - p_1 - p_2)$ becomes $\delta(2z) = 2^{-3}\delta(z)$. The result is

$$D_t f(t, p_1) = \frac{1}{\tau v}\left\{\int b(E, \rho, \rho') \exp\left[\frac{ir}{\hbar}s\cdot(\rho - \rho')\right]\right.$$
$$\times g\left[t, s\cdot\frac{(\rho + \rho')}{|\rho + \rho'|}, p_1 + \frac{r}{2}(\rho - \rho'),\right.$$
$$\left.\left. p_2 - \frac{r}{2}(\rho - \rho')\right]\right.$$
$$\times ds\, dA(\rho')\, dp_2 + \int b^*(E, \rho, \rho')$$

$$\times \exp\left[\frac{-ir}{\hbar}\, s\cdot(\rho-\rho')\right]$$
$$\times g\Big[t, s\cdot\frac{(\rho+\rho')}{|\rho+\rho'|}, p_1+\frac{r}{2}(\rho-\rho'),$$
$$p_2-\frac{r}{2}(\rho-\rho')\Big]\, ds\, dA(\rho')\, dp_2$$
$$+\int b(E,\rho,\rho'')b^*(E,\rho,\rho')\exp\left[\frac{ir}{\hbar}\, s\cdot(\rho'-\rho'')\right]$$
$$\times g\Big[t, s\cdot\frac{(\rho'+\rho'')}{|\rho'+\rho''|}, \frac{p_c}{2}-\frac{r}{2}(\rho''+\rho'),$$
$$\frac{p_c}{2}+\frac{r}{2}(\rho''+\rho')\Big]\, ds\, dA(\rho')\, dA(\rho'')\, dp_2\Big\}$$
$$+O(1/\tau v). \tag{5.89}$$

In (5.89) we have indicated explicitly the dependence of g on t and s. This dependence is obtained from (5.23).

From (5.32), $p_2 = p_1 + 2p$. Therefore, for fixed p_1,

$$dp_2 = 2^3\, dp = 2^3 r^2\, dr\, dA(\rho). \tag{5.90}$$

If we use (5.90) to replace dp_2 in (5.89), we may then use the result (A1), (A36) of the Appendix to simplify (5.89). This yields

$$D_t f(t,p_1) = \frac{(2\pi\hbar)^2}{\tau v}$$
$$\times\Big\{\int r^{-2}b(E,\rho,\rho)g[t, s\cdot\rho, p_1, p_2]\, d(s\cdot\rho)\, dp_2$$
$$+\int r^{-2}b^*(E,\rho,\rho)g[t, s\cdot\rho, p_1, p_2]\, d(s\cdot\rho)\, dp_2$$
$$+\int r^{-2}b(E,\rho,\rho')b^*(E,\rho,\rho')g[t, s\cdot\rho', \hat{p}_1, \hat{p}_2]$$
$$\times d(s\cdot\rho')\, dA(\rho')\, dp_2\Big\} + O(1/\tau v). \tag{5.91}$$

Here

$$\hat{p}_1 = p_c/2 - r\rho' = p_1 + r(\rho-\rho'),$$
$$\hat{p}_2 = p_c/2 + r\rho' = p_2 - r(\rho-\rho'). \tag{5.92}$$

Let $\hat{p} = \frac{1}{2}(\hat{p}_2 - \hat{p}_1) = r\rho'$. From (5.92) we easily determine that $\hat{p}_1 + \hat{p}_2 = p_1 + p_2$ and $\hat{p}_1^2 + \hat{p}_2^2 = p_1^2 + p_2^2$. Thus $\hat{p}_1$ and $\hat{p}_2$ are the final momenta after a collision of two particles whose initial momenta are p_1 and p_2. The final relative momentum is $\hat{p}$ and the final relative direction is ρ'. The collision preserves total momentum and the kinetic energy of both the total and relative motions.

We now use (5.71) to simplify (5.91). Thus we obtain

$$D_t f(t,p_1) = \frac{(2\pi\hbar)^2}{\tau v}\int r^{-2}\,|b(E,\rho,\rho')|^2$$
$$\times\Big\{\int g(t, s\cdot\rho', \hat{p}_1, \hat{p}_2)\, d(s\cdot\rho')$$
$$-\int g(t, s\cdot\rho, p_1, p_2)\, d(s\cdot\rho)\Big\}$$
$$\times dA(\rho')\, dp_2 + O(1/\tau v). \tag{5.93}$$

Finally, from (5.24) and (5.68),

$$D_t f(t,p_1)$$
$$= \frac{1}{v}\int \sigma(E,\rho,\rho')[f(t,\hat{p}_1)f(t,\hat{p}_2) - f(t,p_1)f(t,p_2)]$$
$$\times |(p_2-p_1)/m|\, dA(\rho')\, dp_2 + O(1/\tau v). \tag{5.94}$$

For sufficiently large τ, we may neglect the remainder term in (5.94) and the result is the Uehling-Uhlenbeck form of the quantum-mechanical Boltzmann equation, for the spatially homogeneous case. The Uehling-Uhlenbeck equation for the inhomogeneous case is given by

$$D_t f(t,q_1,p_1)$$
$$= \frac{1}{v}\int \sigma(E,\rho,\rho')[f(t,q_1,\hat{p}_1)f(t,q_1,\hat{p}_2)$$
$$- f(t,q_1,p_1)f(t,q_1,p_2)]$$
$$\times |(p_2-p_1)/m|\, dA(\rho')\, dp_2. \tag{5.95}$$

The derivation of (5.95) would require further work.

VI. CONCLUSION

It is quite clear that the weak point in the present derivation is the argument leading to (5.25) and (5.24) which proceeds by analogy with the classical case. In fact, in our opinion, this point is so weak that it leaves the validity of the result (5.94) in considerable doubt. In recent years, some work has been done[23] to clarify the asymptotic approach (as $t\to\infty$) of the operator $\exp[(it/\hbar)E]\exp[(-it/\hbar)H]$ to the operator SU, as in (5.38). It may be that this type of study will resolve the difficulty that we have pointed out.

APPENDIX

The purpose of this section is to apply the formula for the inversion of Fourier transforms in order to

[23] K. O. Friedrichs, Nachr. Akad. Wiss. Göttingen, Math-physik. Kl., IIa, 7, 43 (1952).

simplify integrals of the form

$$I = \int \exp\left[\frac{ir}{\hbar}\, s\cdot(\rho - \omega)\right] \times F\left[s\cdot\frac{\rho+\omega}{|\rho+\omega|}\,,\ \rho, \omega\right] dA(\rho)\, dA(\omega)\, ds. \tag{A1}$$

Here ρ and ω are unit vectors and dA is the area element on the unit sphere. We shall make use of the Fourier inverse relation in the form

$$\int \exp\left[\pm\frac{ir}{\hbar}\, y(x - x')\right] f(x')\, dx'\, dy = (2\pi\hbar/r) f(x). \tag{A2}$$

The area element dA can be replaced by an area element in the plane by introducing a Cartesian coordinate system such that ρ has components $\rho = (\rho_1, \rho_2, \rho_3)$; $\rho_1^2 + \rho_2^2 + \rho_3^2 = 1$. Then

$$dA(\rho) = d\rho_1\, d\rho_2/|\rho_3|. \tag{A3}$$

To transform the integral I, we introduce a vector

$$\tau = \rho - \omega \tag{A4}$$

and a unit vector

$$\sigma = (\rho + \omega)/|\rho + \omega|. \tag{A5}$$

Then

$$\sigma\cdot\tau = 0. \tag{A6}$$

Let

$$j_4 = (\sigma_1^2 + \sigma_2^2)^{-1/2}(\sigma_2, -\sigma_1, 0), \tag{A7}$$

$$j_5 = \sigma \times j_4 = (\sigma_1^2 + \sigma_2^2)^{-1/2}(\sigma_1\sigma_3, \sigma_2\sigma_3, -\sigma_1^2 - \sigma_2^2), \tag{A8}$$

$$j_6 = \sigma. \tag{A9}$$

j_4, j_5, and j_6 form an orthonormal basis. Since $\sigma\cdot\tau = 0$,

$$\tau = \tau_4 j_4 + \tau_5 j_5, \tag{A10}$$

where

$$\tau_4 = \tau\cdot j_4 = B^{-1/2}[\sigma_2(\rho_1 - \omega_1) - \sigma_1(\rho_2 - \omega_2)], \tag{A11}$$

$$\tau_5 = \tau\cdot j_5 = B^{-1/2}[\sigma_1\sigma_3(\rho_1 - \omega_1) + \sigma_2\sigma_3(\rho_2 - \omega_2) - B(\rho_3 - \omega_3)], \tag{A12}$$

$$B = \sigma_1^2 + \sigma_2^2. \tag{A13}$$

Let

$$A = (\rho + \omega)^2 = 2(1 + \rho\cdot\omega) = 2(1 + \rho_1\omega_1 + \rho_2\omega_2 + \rho_3\omega_3). \tag{A14}$$

Then

$$\sigma_1 = A^{-1/2}(\rho_1 + \omega_1), \quad \sigma_2 = A^{-1/2}(\rho_2 + \omega_2), \tag{A15}$$

$$B = A^{-1}[\rho_1^2 + 2\rho_1\omega_1 + \omega_1^2 + \rho_2^2 + 2\rho_2\omega_2 + \omega_2^2]. \tag{A16}$$

We proceed to calculate the Jacobian of the transformation from ρ and ω to τ and σ, *evaluating the results only at* $\tau = 0$, *i.e., at* $\rho = \omega = \sigma$. By using relations of the form

$$\partial\rho_3/\partial\rho_1 = -\rho_1/\rho_3 \tag{A17}$$

we successively obtain, at $\tau = 0$,

$$A = 4, \tag{A18}$$

$$\frac{\partial A}{\partial\rho_1} = \frac{\partial A}{\partial\rho_2} = \frac{\partial A}{\partial\omega_1} = \frac{\partial A}{\partial\omega_2} = 0, \tag{A19}$$

$$\frac{\partial\sigma_1}{\partial\rho_1} = \frac{\partial\sigma_1}{\partial\omega_1} = \frac{\partial\sigma_2}{\partial\rho_2} = \frac{\partial\sigma_2}{\partial\omega_2} = \frac{1}{2}, \tag{A20}$$

$$\frac{\partial\sigma_1}{\partial\rho_2} = \frac{\partial\sigma_1}{\partial\omega_2} = \frac{\partial\sigma_2}{\partial\rho_1} = \frac{\partial\sigma_2}{\partial\omega_1} = 0, \tag{A21}$$

$$\frac{\partial B}{\partial\rho_1} = \frac{\partial B}{\partial\omega_1} = \sigma_1, \qquad \frac{\partial B}{\partial\rho_2} = \frac{\partial B}{\partial\omega_2} = \sigma_2, \tag{A22}$$

$$\left.\begin{aligned} \frac{\partial\tau_4}{\partial\rho_1} &= B^{-1/2}\sigma_2, & \frac{\partial\tau_4}{\partial\rho_2} &= -B^{-1/2}\sigma_1, \\ \frac{\partial\tau_4}{\partial\omega_1} &= -B^{-1/2}\sigma_2, & \frac{\partial\tau_4}{\partial\omega_2} &= B^{-1/2}\sigma_1, \end{aligned}\right\} \tag{A23}$$

$$\frac{\partial\tau_5}{\partial\rho_1} = C\sigma_1, \quad \frac{\partial\tau_5}{\partial\rho_2} = C\sigma_2,$$

$$\frac{\partial\tau_5}{\partial\omega_1} = -C\sigma_1, \quad \frac{\partial\tau_5}{\partial\omega_2} = -C\sigma_2, \tag{A24}$$

where

$$C = B^{-1/2}\sigma_3 + B^{1/2}\sigma_3^{-1}. \tag{A25}$$

Let

$$J_1 = \frac{d\sigma_1\, d\sigma_2\, d\tau_4\, d\tau_5}{d\rho_1\, d\rho_2\, d\omega_1\, d\omega_2} = \begin{Vmatrix} \frac{\partial\sigma_1}{\partial\rho_1} & \frac{\partial\sigma_1}{\partial\rho_2} & \frac{\partial\sigma_1}{\partial\omega_1} & \frac{\partial\sigma_1}{\partial\omega_2} \\ \frac{\partial\sigma_2}{\partial\rho_1} & \frac{\partial\sigma_2}{\partial\rho_2} & \frac{\partial\sigma_2}{\partial\omega_1} & \frac{\partial\sigma_2}{\partial\omega_2} \\ \frac{\partial\tau_4}{\partial\rho_1} & \frac{\partial\tau_4}{\partial\rho_2} & \frac{\partial\tau_4}{\partial\omega_1} & \frac{\partial\tau_4}{\partial\omega_2} \\ \frac{\partial\tau_5}{\partial\rho_1} & \frac{\partial\tau_5}{\partial\rho_2} & \frac{\partial\tau_5}{\partial\omega_1} & \frac{\partial\tau_5}{\partial\omega_2} \end{Vmatrix}. \tag{A26}$$

If we insert (A20)–(A24) in (A26) we obtain easily, at $\tau = 0$,

$$J_1 = 1/|\sigma_3|. \tag{A27}$$

Let J be defined by

$$dA(\rho)\, dA(\omega) = J\, d\tau_4\, d\tau_5\, dA(\sigma). \quad \text{(A28)}$$

Then from (A3),

$$\frac{d\rho_1\, d\rho_2\, d\omega_1\, d\omega_2}{|\rho_3|\, |\omega_3|} = J\, d\tau_4\, d\tau_5 \frac{d\sigma_1\, d\sigma_2}{|\sigma_3|}, \quad \text{(A29)}$$

and, at $\tau = 0$,

$$J = \frac{|\sigma_3|}{|\rho_3|\, |\omega_3|} J_1^{-1} = 1. \quad \text{(A30)}$$

Now define F to be zero in (A1) for $|\rho| \neq 1$ or $|\omega| \neq 1$. The integral may be written as a sum of four terms depending on the domain of integration,

$$I_{++}(\rho_3 > 0, \omega_3 > 0), \quad I_{--}(\rho_3 < 0, \omega_3 < 0),$$
$$I_{+-}(\rho_3 > 0, \omega_3 < 0), \quad I_{-+}(\rho_3 < 0, \omega_3 > 0) \quad \text{(A31)}$$

and by (A28)

$$I = \int \exp\left(\frac{ir}{\hbar} s\cdot\tau\right) F[s\cdot\sigma, \rho, \omega] J\, d\tau_4\, d\tau_5\, dA(\sigma)\, ds. \quad \text{(A32)}$$

In the transformed terms, I_{++} corresponds to $\sigma_3 > 0$, and I_{--} to $\sigma_3 < 0$, and the τ domains of both of these integrals include the point $\tau = 0$. The τ domains for I_{+-} and I_{-+} exclude the point $\tau = 0$. (A32) may be written

$$I = \int \exp\left(\frac{ir}{\hbar}\left[s_4\tau_4 + s_5\tau_5\right]\right) \times F[s_6, \rho, \omega] J\, ds_6\, ds_4\, ds_5\, d\tau_4\, d\tau_5\, dA(\sigma), \quad \text{(A33)}$$

and from (A2)

$$I = \left(\frac{2\pi\hbar}{r}\right)^2 \int F[s_6, \rho, \omega] J\, ds_6\, |_{\tau=0}\, dA(\sigma). \quad \text{(A34)}$$

But at $\tau = 0$, $\rho = \omega = \sigma$ and $J = 1$; hence,

$$I = (2\pi\hbar/r)^2 \int F[s_6, \sigma, \sigma]\, ds_6\, dA(\sigma). \quad \text{(A35)}$$

Since the point $\tau = 0$ is outside the domain of integration, $I_{+-} = I_{-+} = 0$. The contributions I_{++} and I_{--} may be added to produce an integral over the entire unit sphere $|\sigma| = 1$. By changing σ to ρ, we obtain the result

$$I = (2\pi\hbar/r)^2 \int F[(s\cdot\rho), \rho, \rho]\, d(s\cdot\rho)\, dA(\rho). \quad \text{(A36)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Theory of Transport Coefficients. III. Quantum Statistical Systems

S. FUJITA
Université Libre de Bruxelles, Brussels, Belgium
(Received November 10, 1961)

The electrical conductivity of an electron-phonon system is calculated from Kubo's formula using a perturbation method on the assumptions that the coupling is weak, the system is infinitely large, and that the electrons do not interact between them and obey the Fermi-Dirac statistics and phonons are in thermodynamic equilibrium. The result is identical with that one would obtain from the usually assumed Boltzmann-Bloch equation. The calculation is a logical extension of the previous treatment I of the present series, where the same problem is treated on the assumption that the electrons obey the Boltzmann statistics. The relation between the present calculation and the derivation of master equation is critically discussed. A brief sketch is given for the calculation of the viscosity coefficient of a dilate quantum statistical gas.

1. INTRODUCTION

IN the two previous parts I and II[1] of the present series, transport coefficients: electrical conductivity and viscosity coefficients of quantum systems obeying the Boltzmann statistics were calculated, starting from Kubo's formulas and using the perturbation diagrams.[2] In this paper the same problem for quantum statistical systems is treated.

We have formulated in I the general program in terms of second quantized quantities. Therefore,

Supported by the Air Research and Development Command, U. S. Air Force through its European Office.

[1] S. Fujita and R. Abe, J. Math. Phys. **3**, 350 (1962); S. Fujita, *ibid.* **3**, 359 (1962). These will be quoted in the text as I and II, respectively.

[2] The references concerning Kubo's formulas and diagram techniques can be found in I and II.

Let J be defined by

$$dA(\rho)\, dA(\omega) = J\, d\tau_4\, d\tau_5\, dA(\sigma). \quad \text{(A28)}$$

Then from (A3),

$$\frac{d\rho_1\, d\rho_2\, d\omega_1\, d\omega_2}{|\rho_3|\, |\omega_3|} = J\, d\tau_4\, d\tau_5 \frac{d\sigma_1\, d\sigma_2}{|\sigma_3|}, \quad \text{(A29)}$$

and, at $\tau = 0$,

$$J = \frac{|\sigma_3|}{|\rho_3|\, |\omega_3|} J_1^{-1} = 1. \quad \text{(A30)}$$

Now define F to be zero in (A1) for $|\rho| \neq 1$ or $|\omega| \neq 1$. The integral may be written as a sum of four terms depending on the domain of integration,

$$I_{++}(\rho_3 > 0, \omega_3 > 0), \quad I_{--}(\rho_3 < 0, \omega_3 < 0),$$
$$I_{+-}(\rho_3 > 0, \omega_3 < 0), \quad I_{-+}(\rho_3 < 0, \omega_3 > 0) \quad \text{(A31)}$$

and by (A28)

$$I = \int \exp\left(\frac{ir}{\hbar} s\cdot\tau\right) F[s\cdot\sigma, \rho, \omega] J\, d\tau_4\, d\tau_5\, dA(\sigma)\, ds. \quad \text{(A32)}$$

In the transformed terms, I_{++} corresponds to $\sigma_3 > 0$, and I_{--} to $\sigma_3 < 0$, and the τ domains of both of these integrals include the point $\tau = 0$. The τ domains for I_{+-} and I_{-+} exclude the point $\tau = 0$. (A32) may be written

$$I = \int \exp\left(\frac{ir}{\hbar}\left[s_4\tau_4 + s_5\tau_5\right]\right) \times F[s_6, \rho, \omega] J\, ds_6\, ds_4\, ds_5\, d\tau_4\, d\tau_5\, dA(\sigma), \quad \text{(A33)}$$

and from (A2)

$$I = \left(\frac{2\pi\hbar}{r}\right)^2 \int F[s_6, \rho, \omega] J\, ds_6\, |_{\tau=0}\, dA(\sigma). \quad \text{(A34)}$$

But at $\tau = 0$, $\rho = \omega = \sigma$ and $J = 1$; hence,

$$I = (2\pi\hbar/r)^2 \int F[s_6, \sigma, \sigma]\, ds_6\, dA(\sigma). \quad \text{(A35)}$$

Since the point $\tau = 0$ is outside the domain of integration, $I_{+-} = I_{-+} = 0$. The contributions I_{++} and I_{--} may be added to produce an integral over the entire unit sphere $|\sigma| = 1$. By changing σ to ρ, we obtain the result

$$I = (2\pi\hbar/r)^2 \int F[(s\cdot\rho), \rho, \rho]\, d(s\cdot\rho)\, dA(\rho). \quad \text{(A36)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Theory of Transport Coefficients. III. Quantum Statistical Systems

S. FUJITA
Université Libre de Bruxelles, Brussels, Belgium
(Received November 10, 1961)

The electrical conductivity of an electron-phonon system is calculated from Kubo's formula using a perturbation method on the assumptions that the coupling is weak, the system is infinitely large, and that the electrons do not interact between them and obey the Fermi-Dirac statistics and phonons are in thermodynamic equilibrium. The result is identical with that one would obtain from the usually assumed Boltzmann-Bloch equation. The calculation is a logical extension of the previous treatment I of the present series, where the same problem is treated on the assumption that the electrons obey the Boltzmann statistics. The relation between the present calculation and the derivation of master equation is critically discussed. A brief sketch is given for the calculation of the viscosity coefficient of a dilate quantum statistical gas.

1. INTRODUCTION

IN the two previous parts I and II[1] of the present series, transport coefficients: electrical conductivity and viscosity coefficients of quantum systems obeying the Boltzmann statistics were calculated, starting from Kubo's formulas and using the perturbation diagrams.[2] In this paper the same problem for quantum statistical systems is treated.

We have formulated in I the general program in terms of second quantized quantities. Therefore,

Supported by the Air Research and Development Command, U. S. Air Force through its European Office.

[1] S. Fujita and R. Abe, J. Math. Phys. **3**, 350 (1962); S. Fujita, *ibid.* **3**, 359 (1962). These will be quoted in the text as I and II, respectively.

[2] The references concerning Kubo's formulas and diagram techniques can be found in I and II.

the necessary quantum statistical effect was included there and the Boltzmann statistics were only introduced for simplifying the diagram analysis. In Sec. 2 of the present paper we calculate the electrical conductivity of an electron-phonon system, assuming that electrons obey the Fermi-Dirac statistics. The diagram analysis is greatly facilitated by the use of a new diagram technique and a theorem about left-multidentate diagrams. As before, we obtain the result identical with that which we would obtain from the usually assumed Boltzmann-Bloch equation.

Although this identity has been conjectured and seemingly accepted,[3] its explicit demonstration has, to the knowledge of the author, never been worked out for a quantum statistical system in which there exists no universal relaxation time. However, the main interest of the present paper is rather in the techniques connected with quantum statistical diagrams, i.e., the use of O diagrams and a theorem about left-multidentate diagrams. In fact these techniques are used, e.g., in the following paper (IV) of the present series to obtain a new formula for the magnetoresistance of an electron-phonon system to which situation the Boltzmann equation in the usual sense does not exist.

In Sec. 3 we give several remarks relative to the diagrams in general and to the relation between the present calculation and the derivation of master equation. A brief sketch for the calculation of the viscosity coefficient of a dilute quantum statistical gas is also given there.

2. ELECTRICAL CONDUCTIVITY

Let us consider the same system as in I, an electron-phonon system characterized by the Hamiltonian[4]

$$H = \sum_{\mathbf{p}} p^2 a_{\mathbf{p}}^\dagger a_{\mathbf{p}} + \sum_{\mathbf{q}} \omega_{\mathbf{q}} b_{\mathbf{q}}^\dagger b_{\mathbf{q}} + g \sum_{\mathbf{p},\mathbf{q}} V^{-1/2} \gamma_{\mathbf{q}}^{1/2} (a_{\mathbf{p}+\mathbf{q}}^\dagger a_{\mathbf{p}} b_{\mathbf{q}} + a_{\mathbf{p}}^\dagger a_{\mathbf{p}+\mathbf{q}} b_{\mathbf{q}}^\dagger), \quad (2.1)$$

where the same notations are used: $(a_{\mathbf{p}}, a_{\mathbf{p}}^\dagger)$ and $(b_{\mathbf{q}}, b_{\mathbf{q}}^\dagger)$ are annihilation and creation operators for electron and phonon, V is the volume, and g is the coupling constant. We want to calculate the static electrical conductivity σ from Kubo's formula

$$\sigma = \beta V^{-1} \operatorname{Re} \left\{ \int_0^\infty du \langle J_x J_x[u] \rangle \right\}, \quad (2.2)$$

where $\beta \equiv (kT)^{-1}$ is the reciprocal temperature, k the Boltzmann constant, and $J_x[u]$ is the Heisenberg operator representing the x component of the electric current; the symbol $\langle\ \rangle$ means the grand canonical average taken with $e^{\alpha N - \beta H}$, e^α being fugacity and N the number of electrons.

The integrand in (2.2) can be approximated as follows in the weak coupling limit [see (2.5)][5]:

$$\langle J_x J_x[u] \rangle \to 4e^2 \sum_{\mathbf{p},\mathbf{p}'} p_x p_x' \langle e^{iuH} a_{\mathbf{p}}^\dagger a_{\mathbf{p}'} e^{-iuH} a_{\mathbf{p}}^\dagger a_{\mathbf{p}} \rangle_0 \equiv 4e^2 \sum_{\mathbf{p},\mathbf{p}'} p_x p_x' U(\mathbf{p}', \mathbf{p}; u), \quad (2.3)$$

where the symbol $\langle\ \rangle_0$ means the average with $e^{\alpha N - \beta H_0}$ and e the electronic charge. The function $U(\mathbf{p}', \mathbf{p}; u)$ is expanded in a perturbation series, which is represented by linked Feynman diagrams.

The linked diagrams were analyzed in I by assuming the four following conditions:

(1) The system is infinitely large:

$$N \to \infty, \qquad V \to \infty \ (N/V = \text{finite}). \quad (2.4)$$

(2) The coupling is weak:

$$g \to 0, \qquad t \to \infty \ (g^2 t = \text{finite}). \quad (2.5)$$

(3) The electrons obey the Boltzmann statistics:

$$f \to 0, \qquad 1 - f \to 1, \quad (2.6)$$

where f and $1 - f$ are the statistical factors associated with the electron line, f being the Fermi distribution function.

The first two conditions will be assumed unchanged in the present analysis. The third was introduced only for simplifying the diagram analysis. We now replace it by 3′.

(3′) Electrons obey the Fermi-Dirac statistics. This is, of course, more satisfactory for practical purposes. The analysis becomes somewhat more complicated.

(4) Phonons are in thermodynamic equilibrium.

Let us now enumerate various contributions to $U(\mathbf{p}, \mathbf{p}'; u)$. In the zeroth order one has

$$\delta^{(3)}(\mathbf{p} - \mathbf{p}') f_{\mathbf{p}} (1 - f_{\mathbf{p}}), \quad (2.7)$$

where $\delta^{(3)}(\mathbf{p})$ denotes three-dimensional Dirac's delta function.

[3] See, e. g., papers by Chester-Thellung, Verboven, cited in I. Some incomplete investigations into the same problem as that treated here can be found in the following papers: M. I. Klinger, Soviet Phys.—Solid State 1, 782(1960); R. Zigenlaub, *ibid.* 1, 964 (1960); Lang, *ibid.* 2, 2077 (1961).

[4] Throughout in the text the units are chosen such that $2M = \hbar = 1$, where M is the mass of an electron.

[5] The function $U(p', p; u)$ defined in (2.3) is different from U in (I, 2.2b) by a factor $e^{\alpha - \beta p^2}$. Accordingly the following change is to be made in the prescription in I for obtaining mathematical expressions from diagrams: Treat as if the point at $t = 0$ were below the boundary, and abolish all the rules exceptionally prescribed for the point at $t = 0$.

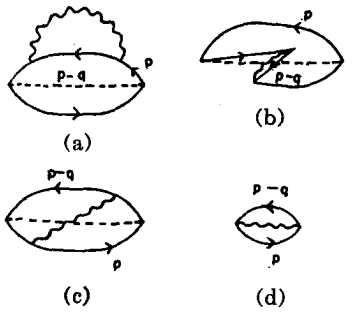


FIG. 1. Three O diagrams of the second order (a), (b), and (c). Their basis diagram (d).

In the second order there exist, in all, 48 non-vanishing distinct diagrams which can be obtained by following the prescription given in I. We may construct them in two steps as follows:

(i) In *one plane* draw diagrams which are constructed by the same procedure as before except for the boundary line being suppressed. Such a diagram is called an *O diagram*.

With the convention that *an undirected phonon line* corresponds to *two* lines directed in the mutually opposite senses, there are six such diagrams of which three are drawn in Fig. 1(a, b, c). The diagram shown in (d) can be obtained from any of the three in (a), (b), and (c) by suppressing two particle lines with a common momentum running in opposite directions without changing the nature of interaction vertices. This type of diagram is called *basis diagram*. It may be useful for generating O diagrams. The meaning of basis diagrams will be discussed in the next section.

(ii) From an O diagram construct the number 2^n (n being the order in g) of diagrams by distributing vertices representing interactions H_I' above and below the boundary line without changing their time ordering. The diagrams so obtained are said to be mutually *similar*.

This process is illustrated in Fig. 2, where four diagrams are drawn corresponding to the O diagram in Fig. 1(a). According to the prescription given in I those contributions corresponding to similar diagrams have a common time integral and may be summed up into a simple expression. For the diagrams in Fig. 2 the sum yields a contribution

$$-u\,\delta^{(3)}(p'-p)\Gamma_+(\mathbf{p})f_{\mathbf{p}}(1-f_{\mathbf{p}}), \tag{2.8}$$

where

$$\Gamma_-(\mathbf{p}) \equiv \pi g^2 \frac{1}{(2\pi)^3}\int d^2q\, W_-(\mathbf{p}-\mathbf{q},\mathbf{p}) \tag{2.9}$$

$$\begin{aligned} W_-(\mathbf{p}-\mathbf{q},\mathbf{p}) \equiv \gamma_q\{&[(1-f_{\mathbf{p}-\mathbf{q}})(1+n_q) \\ &+ f_{\mathbf{p}-\mathbf{q}}n_q]\,\delta_-(\epsilon_1) + [(1-f_{\mathbf{p}-\mathbf{q}})n_q \\ &+ f_{\mathbf{p}-\mathbf{q}}(1+n_q)]\,\delta_-(\epsilon_2)\} \end{aligned} \tag{2.10}$$

$$\delta_-(\epsilon) = \delta(\epsilon) \pm (i/\pi)P(1/\epsilon) \tag{2.11}$$

$$\begin{aligned} \epsilon_1 &= p^2 - (\mathbf{p}-\mathbf{q})^2 - \omega_q, \\ \epsilon_2 &= p^2 - (\mathbf{p}-\mathbf{q})^2 + \omega_q \end{aligned} \tag{2.12}$$

with

$$n_q = n_{-q} \tag{2.13}$$

being the Planck distribution function.

We shall say that the sum of those contributions corresponding to similar diagrams is the contribution corresponding to the generating O diagram.

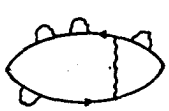


FIG. 3. A general O diagram contributing a finite amount in the limits (2.4) and (2.5).

It can be shown that the contribution corresponding to the O diagram in Fig. 1(b) vanishes. In general, it is found[6] that *an O diagram*, such as the above, *representing a vanishing contribution can be easily recognized by its special structure*, i.e., a structure containing one or more vertices which have leaving and entering particle lines all on the right. Such a structure will be called *left-multidentate structure*.[6]

The above fact as well as the use of O diagrams greatly simplifies the analysis of diagrams. In fact O diagrams contributing finite amounts in the limits of (2.4) and (2.5) have easily recognizable structures. Such a general structure is given in Fig. 3. It is noted that this diagram is identical in structure, except for the boundary line being suppressed, with the structure shown in Fig. 5(b) of I, which was important for the Boltzmann statistical electrons. It is further seen that all of the contributing O diagrams have a common factor $f_{\mathbf{p}}(1-f_{\mathbf{p}})$. The

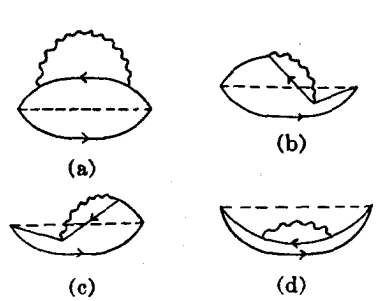


FIG. 2. The set of similar diagrams generated from an O diagram in Fig. 1(a).

[6] The same is true for the case of Bose statistics.

total contribution written as

$$U(\mathbf{p}, \mathbf{p}'; u) \rightarrow Y(\mathbf{p}', \mathbf{p}; u) f_{\mathbf{p}}(1 - f_{\mathbf{p}}) \tag{2.14}$$

can be obtained by the same procedure as in I. The u integral of Y defined by

$$\psi(\mathbf{p}', \mathbf{p}) \equiv \int_0^\infty du Y(\mathbf{p}', \mathbf{p}; u) \tag{2.15}$$

can be shown to satisfy the following integral equation

$$\psi(\mathbf{p}', \mathbf{p}) = \frac{\delta^{(3)}(\mathbf{p}' - \mathbf{p})}{\Gamma(\mathbf{p})} + \frac{g^2}{(2\pi)^2} \times \int \frac{W(\mathbf{p} - \mathbf{q}, \mathbf{p})}{\Gamma(\mathbf{p})} \psi(p', p - q)\, d^3q, \tag{2.16}$$

where Γ and W are defined as

$$\Gamma(\mathbf{p}) \equiv \tfrac{1}{2}[\Gamma_+(\mathbf{p}) + \Gamma_-(\mathbf{p})], \tag{2.17}$$

$$W(\mathbf{p}', \mathbf{p}) \equiv \tfrac{1}{2}[W_+(\mathbf{p}', \mathbf{p}) + W_-(\mathbf{p}', \mathbf{p})]. \tag{2.18}$$

If we define a function $\varphi_x(\mathbf{p})$ by

$$\varphi_x(\mathbf{p}) \equiv \frac{V}{(2\pi)^3} \int p_x' \psi(\mathbf{p}', \mathbf{p})\, d^3p', \tag{2.19}$$

we obtain from (2.16) an equation

$$p_x = \frac{g^2}{(2\pi)^2} \int W(\mathbf{p} - \mathbf{q}, \mathbf{p})[\varphi_x(\mathbf{p}) - \varphi_x(\mathbf{p} - \mathbf{q})]\, d^3q. \tag{2.20}$$

In view of (2.2), (2.3), (2.14), (2.15), and (2.19) the electrical conductivity is given by

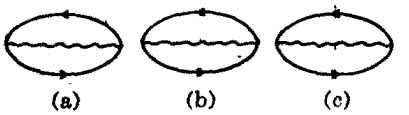
$$\sigma = \frac{4\beta e^2}{(2\pi)^2} \int d^3p p_x \varphi_x(\mathbf{p}) f_{\mathbf{p}}(1 - f_{\mathbf{p}}). \tag{2.21}$$

It is easy to verify that the conductivity calculated above is exactly equal to the one that we would obtain from the usual Boltzmann equation. The latter equation for the present case is written as[7]

$$\begin{aligned} e\mathcal{E} \frac{\partial F}{\partial p_x} = \frac{g^2}{(2\pi)^2} \int d^3q \gamma_{\mathbf{q}} \\ \times \{[n_{\mathbf{q}} F_{\mathbf{p}-\mathbf{q}}(1 - F_{\mathbf{p}}) \\ - (1 + n_{\mathbf{q}}) F_{\mathbf{p}}(1 - F_{\mathbf{p}-\mathbf{q}})]\, \delta(\epsilon_1) \\ + [(1 + n_{\mathbf{q}}) F_{\mathbf{p}-\mathbf{q}}(1 - F_{\mathbf{p}}) \\ - n_{\mathbf{q}} F_{\mathbf{p}}(1 - F_{\mathbf{p}-\mathbf{q}})]\, \delta(\epsilon_2)\}. \end{aligned} \tag{2.22}$$

Here $F_{\mathbf{p}}$ is the distribution function for the electron and $n_{\mathbf{q}}$ that for the phonon, which is assumed to have the same form as in the equilibrium state, and $\mathcal{E}$ is the magnitude of the stationary electric field applied to the system along the x axis.

FIG. 4. Succession of rings relevant to the derivation of master equation in the limits (2.4) and (2.5).



3. REMARKS

A. Master Equation Approach

In the preceding section we have seen that the electrical conductivity calculated from Kubo's formula gives in the weak coupling limit a result identical to that one would obtain from the usually assumed Boltzmann equation. This implies the possibility of deriving this equation from the Von Neumann equation using the two limits (2.4) and (2.5). This derivation is usually done in two steps: (i) derivation of the master equation and (ii) the reduction to the equation for one-body distribution function. This was done, for example, by Van Hove.[8]

When one uses the diagram technique[3,8] for calculating the transport coefficient and for deriving the master equation referring to the same system, a remarkable correspondence appears between the diagrams for the two cases. Loosely speaking, those diagrams which appear in the derivation of the master equation are basis diagrams introduced in Sec. 2. It is well known[8,9] that those diagrams relevant in the limits (2.4) and (2.5) for the master-equation derivation are characterized by successions of rings as shown in Fig. 4. This structure can be recognized as the basis diagram associated with that given in Fig. 3.

Some remarks are in order for the detail of this correspondence.

In the derivation of master equations,[8-10] diagrams are usually drawn relative to the final state specifying the density matrix, and each region sectioned by points of interaction represents a many-particle state. In our theory of transport coefficients complete contractions with respect to the grand canonical ensemble are represented by diagrams. The representation is exact, but a region in the above sense does not represent any definite state. Both diagram

[7] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, New York, 1954), 2nd ed., p. 258.

[8] L. Van Hove, *La théorie des gaz neutres et ionisés*, edited by C. De Witt and J. F. Detoeuf (Hermann et Cie., Paris, 1960), p. 151.

[9] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, to be published).

[10] K. Nishikawa, J. Phys. Soc. Japan, 15, 78 (1960).

representations fail in the presence of Bose-Einstein degeneracy.[11]

In the theory of transport coefficients only linked diagrams are important. According to the analysis due to Hugenholtz,[12] this means that every important diagram is of order 1 with respect to N, the total number of particles. In the theory of the master equation diagonal fragments[9] play an important role. Such a fragment however may contain several unlinked diagrams and therefore contribute an amount of high order N^p, $p \geq 1$. This gives a difficulty to the meaning of the master equation valid up to an arbitrarily high order. This difficulty never arises in the theory of transport coefficients or in that of the one-body distribution function.

The Boltzmann equation describes the evolution of the one-body distribution function. Although this equation can be considered in many cases as a satisfactory starting point for the calculation of transport coefficients, it is not always so. The electrical conduction in the presence of magnetic field can not be discussed on the basis of the Boltzmann equation.[13] This interesting case will be treated in the subsequent paper IV of the present series.

B. Operator Diagrams and c-Number Diagrams.

In Sec. 2, O diagrams are introduced which represent sets of similar diagrams. The relation between an O diagram and its corresponding similar diagrams is the same as that between an operator[14] and its corresponding c-number diagrams, which was discussed earlier by the present author.[15] The operator diagrams are, in general, more convenient in discussing the time development of quantum statistical quantities.

C. The Viscosity Coefficient of a Dilute Quantum Statistical Gas

We shall only make brief remarks pertinent to the problem.

The calculation is made on the assumption that the gas is very dilute:

$$c \equiv N/V \to 0, \qquad t \to 0(ct = \text{finite}) \tag{3.1}$$

and the potential function $v(r)$ of r (distance) may be strong but of short range:

$$rv(r) \to 0 \quad \text{as} \quad r \to \infty. \tag{3.2}$$

The limit (3.1) replaces in the present case the weak coupling limit (2.5). Since the potential is strong and has the property (3.2), it is advantageous and necessary to use the binary-collision expansion,[11] which may be represented by diagrams similar to the perturbation ones used in II. Except for this, the calculation essentially proceeds in the same line as in Sec. 2, i.e., using the new techniques, O-diagram representation and a theorem about left-multidentate diagrams.

The resulting expression for the viscosity is identical to that one would obtain from the Uehling-Uhlenbeck equation,[16] whose collision term contains the statistical effects on the final states.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor I. Prigogine for the kind hospitality extended to him. The author wishes to thank Dr. F. Takano, Tokyo University of Education, for pointing out the misstatement in a preprint version on the theorem about left-multidentate structures.

[11] See the Appendix of S. Fujita, Physica **27**, 930 (1961).
[12] N. M. Hugenholtz, Physica **23**, 481 (1957).
[13] See for an instructive review M. Dresden, Revs. Modern Phys. **33**, 265 (1961).
[14] I. Prigogine, *Superfluidité et equation du transport quantique*, edited by P. Résibois (Institute of Nuclear Science, Brussels, 1960).
[15] S. Fujita, Physica **27**, 940 (1961).
[16] E. A. Uehling and G. E. Uhlenbeck, Phys. Rev. **43**, 552 (1933).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Numerical Estimation of the Partition Function in Quantum Statistics*

LLOYD D. FOSDICK
Digital Computer Laboratory, University of Illinois, Urbana, Illinois
(Received July 20, 1962)

A method for estimating the partition function of a quantum mechanical system is described. The method is based on a technique for evaluating the Wiener integrals in terms of which the partition function may be expressed. This technique involves first, an approximation of the Wiener integral by an n-dimensional integral and, second, a Monte Carlo estimation of the value of the n-dimensional integral. Application of the method to a harmonic oscillator and a pair of interacting particles in a box in two dimensions is described.

1. INTRODUCTION

THE purpose of this paper is to describe a method for the numerical estimation of the partition function Z of a quantum mechanical system, given by

$$Z = \sum_n e^{-\beta E_n} \tag{1.1}$$

where $\beta = 1/kT$, E_n is the energy of state n, and the summation extends over all states with proper symmetry. This method is based on the fact that Z can be expressed in terms of certain Wiener integrals and that values for these integrals can be estimated by a "sampling" or Monte Carlo procedure. It will be seen that this method is related in spirit to the Monte Carlo method which has been used to treat certain problems in classical[1-3] statistics. Like the classical method, the present method can provide a detailed picture of the important terms in the partition function and the nature of these terms is such that their physical significance remains clear.

The discussion will begin with a description of the method used here to estimate the Wiener integral, followed by a simple example. Next, we will consider the connection of the Wiener integral to the partition function and the sampling method for estimating the particular Wiener integrals arising in this connection. Finally, an illustration of the method is given by using it to estimate the partition function for a harmonic oscillator and for a pair of particles in a two-dimensional box. In the latter example we will treat both the case of no interaction and an interaction represented by the potential function

$$V(r) = \begin{cases} \infty & r < a \\ -v & a \leq r \leq b, \\ 0 & b < r \end{cases} \tag{1.2}$$

where r is the distance between the particles and a, b, v are adjustable parameters. In all of the examples, except the last, exact results can be obtained by standard techniques, affording a direct test of the accuracy of the method. The last example is included mainly to demonstrate the fact that the presence of the interaction does not introduce any essential complication for the method. The numerical computations described below were performed on the Illiac, an automatic computer located at the University of Illinois.

2. ESTIMATION OF THE WIENER INTEGRAL

Let $\{x(\tau), \tau \in t\}$ be a stochastic process with a continuous "time" parameter $0 \leq \tau \leq t$, normalized by the condition $x(0) = 0$. When the increments $x(\tau) - x(s)$ are independent and have a Gaussian distribution with expectation values

$$\text{Ex}\,\{x(\tau) - x(s)\} = 0 \tag{2.1}$$

$$\text{Ex}\,\{[x(\tau) - x(s)]^2\} = \tfrac{1}{2}\,|\tau - s|, \tag{2.2}$$

the stochastic process is called a Wiener process.[4] If F denotes a functional, $F[x(\cdot)]$, of the sample functions of the Wiener process, then its expectation value, Ex $\{F\}$, is called the Wiener integral of F. A helpful intuitive notion of the Wiener integral may be obtained from the following description: Let S denote the ensemble of functions $x(\tau)$ generated by the Wiener process, and let the functional $F[x(\cdot)]$ be evaluated for each member $x(\tau)$ of the ensemble, denoting a particular evaluation by $F[x_\omega(\cdot)]$, then the Wiener integral Ex $\{F\}$ is given by averaging $F[x_\omega(\cdot)]$ over the ensemble S.

* This work was supported in part by the Office of Naval Research.

[1] L. D. Fosdick, Phys. Rev. 116, 565 (1959).

[2] J. R. Ehrman, L. D. Fosdick, and D. C. Handscomb, J. Math. Phys. 1, 547 (1960).

[3] W. W. Wood and F. R. Parker, J. Chem. Phys. 27, 720 (1957).

[4] J. L. Doob, *Stochastic Processes* (John Wiley & Sons, Inc., New York, 1953).

Most of our attention here will not be directed at the Wiener integral, but rather at a closely related integral called the conditional Wiener integral, denoted by Ex $\{F \mid C\}$. The notion of the conditional Wiener integral is made clear from the following description: Let S' denote a subset of the ensemble S, all of whose members $x_\omega(\tau)$ satisfy some condition C, and let $F[x(\cdot)]$ be evaluated for each $x_\omega(\tau)$ in S', then the conditional Wiener integral Ex $\{F \mid C\}$ is given by averaging $F[x_\omega(\cdot)]$ over the subset S'. In this work the condition which will be of interest is represented by $x(t) = a$; that is, we require the terminal value of the function $x(\tau)$ to have a fixed value. The associated conditional Wiener integral will be denoted by Ex $\{F \mid x(t) = a\}$.

The method employed here for estimating Wiener integrals consists of two stages. In the first stage a theorem due to Cameron[5] is used to obtain an approximation to the Wiener integral in terms of an n-dimensional Riemann integral which approaches the Wiener integral as $n \to \infty$. In the second stage the value of the n-dimensional Riemann integral is estimated by a Monte Carlo sampling procedure.

Cameron's theorem, which is the basis for the "rectangle approximation" is as follows, where the time interval has been normalized to $0 \leq \tau \leq 1$, and C is the space of all continuous functions on the interval $0 \leq \tau \leq 1$ which satisfy $x(0) = 0$.

Theorem 1. Let $F[x(\cdot)]$ be continuous in the Hilbert topology on the space C and let

$$F\,|[x(\cdot)]| \leq H\left(\int_0^1 x^2(\tau)\,d\tau\right) \tag{2.3}$$

on C where $H(u)$ is monotonically increasing and the Wiener integral Ex $\{H\}$ satisfies

$$\text{Ex}\,\{H\} < \infty. \tag{2.4}$$

Let $\alpha_1(\tau), \alpha_2(\tau), \cdots$ be a complete orthonormal set on the interval $0 \leq \tau \leq 1$ such that $\alpha_j(\tau) \in C$ $(j = 1, 2, \cdots)$ and let $\beta_1(\tau), \beta_2(\tau), \cdots$ be obtained by applying the Schmidt orthogonalization process to the sequence of integrals

$$\int_\tau^1 \alpha_1(s)\,ds, \qquad \int_\tau^1 \alpha_2(s)\,ds, \cdots.$$

Let the relation between these integrals and the β_k's be given by

$$\int_\tau^1 \alpha_k(s)\,ds = \sum_{j=1}^{k} \gamma_{jk}\beta_j(\tau). \tag{2.5}$$

Then the Wiener integral of $F[x(\cdot)]$ is given by the formula

$$\text{Ex}\,\{F\} = \lim_{n\to\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F\left[\sum_{k=1}^{n}\sum_{j=1}^{k} \gamma_{j,k}\xi_j\alpha_k(\cdot)\right] \times e_n(\xi)\,d\xi_1 \cdots d\xi_n, \tag{2.6}$$

where

$$e_n(\xi) = \pi^{-n/2} \exp\left(-\xi_1^2 - \xi_2^2 - \cdots - \xi_n^2\right). \tag{2.7}$$

In the present work we have taken, rather arbitrarily, the orthonormal set

$$\alpha_j(\tau) = \sqrt{2}\sin\left(j - \tfrac{1}{2}\right)\pi\tau. \tag{2.8}$$

With this choice we have

$$\int_\tau^1 \alpha_j(s)\,ds = \frac{\sqrt{2}\cos\left(j - \tfrac{1}{2}\right)\pi\tau}{\left(j - \tfrac{1}{2}\right)\pi}, \tag{2.9}$$

so that

$$\gamma_{jk} = \delta_{jk}[(j - \tfrac{1}{2})\pi]^{-1}, \tag{2.10}$$

where δ_{jk} is the Kronecker delta, and Cameron's theorem gives

$$\text{Ex}\,\{F\} = \lim_{n\to\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F[x_n(\cdot)]e_n(\xi)\,d\xi_1 \cdots d\xi_n, \tag{2.11}$$

where

$$x_n(\tau) = \sum_{j=1}^{n} \xi_j \frac{\sqrt{2}\sin\left(j - \tfrac{1}{2}\right)\pi\tau}{\left(j - \tfrac{1}{2}\right)\pi}. \tag{2.12}$$

We define

$$\text{Ex}\,\{F\}_n = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F[x_n(\cdot)]e_n(\xi)\,d\xi_1 \cdots d\xi_n \tag{2.13}$$

and use this as an approximation for the Wiener integral Ex $\{F\}$; this is called the rectangle approximation by Cameron. This represents the first stage in our method for obtaining an estimate of Ex $\{F\}$ and it can be described rather simply as follows: The ensemble $\{x(\tau)\}$ over which one should average $F[x_\omega(\cdot)]$ to obtain Ex $\{F\}$ is replaced by the ensemble $\{x_n(\tau)\}$ having members which are n-term sine series with coefficients $\xi_j/(j - 1/2)\pi$, where ξ_j has a Gaussian distribution with expectation zero and variance one-half.

The second stage of approximation consists in evaluating the n-dimensional integral in Eq. (2.13) by Monte Carlo sampling. This sampling scheme is described as follows. A set k of n numbers is randomly selected from a Gaussian distribution; we represent this set by $\xi_{1,k}, \xi_{2,k}, \cdots, \xi_{n,k}$. With this set, the functional $F[x_{n,k}(\cdot)]$ is evaluated where

$$x_{n,k} = \sum_{j=1}^{n} \frac{\xi_{j,k}\sqrt{2}\sin\left(j - \tfrac{1}{2}\right)\pi\tau}{\left(j - \tfrac{1}{2}\right)\pi}. \tag{2.14}$$

Letting the integer subscript, k, denote the ordinal

[5] R. H. Cameron, Duke Math. J. 18, 111 (1951).

number of the quantities, $F[x_{n,1}(\cdot)]$, $F[x_{n,2}(\cdot)]$, $\cdots$ obtained by repeating this random process, we have

$$\text{Ex } \{F\}_{n,R} = \frac{1}{R}\sum_{k=1}^{R} F[x_{n,k}(\cdot)] \tag{2.15}$$

and, by the law of large numbers

$$\text{Ex } \{F\}_{n,R} \to \text{Ex } \{F\}_n \quad \text{as} \quad R \to \infty,$$

in the probabilistic sense, i.e.,

$$\Pr\{|\text{Ex } \{F\}_{n,R} - \text{Ex } \{F\}_n| > \epsilon\} \to 0$$

for every $\epsilon > 0$, as $R \to \infty$. The use of Ex $\{F\}_{n,R}$ in place of Ex $\{F\}_n$ represents the second stage of approximation. It is to be noted that the Gaussian weight factor $e_n(\xi)$ which appears in the integrand of Eq. (2.13) is implicit in the approximation given by Eq. (2.15) because the sets of random numbers $\xi_{i,k}$ are selected from the indicated Gaussian distribution.

(a) Example

In order to illustrate the application of these ideas let us now consider a specific example, namely, the estimation of the Wiener integral of

$$F[x(\cdot)] = \exp\left[-\int_0^1\int_0^1 \tau\tau' x(\tau)x(\tau')\,d\tau\,d\tau'\right]. \tag{2.16}$$

This example is useful because the Wiener integral of this functional can be evaluated exactly, moreover the approximation Ex $\{F\}_n$ can be evaluated exactly for arbitrary n, and finally the functional is related in form to the functional to be treated later in the consideration of the partition function.

The exact value of this Wiener integral is easily obtained with the help of the following[6]:

Theorem 2. Let $\beta(\tau)$ be a real function of class L_1 on $0 \le \tau \le 1$, let

$$A = \left[\int_0^1\left[\int_\tau^1 \beta(\xi)\,d\xi\right]^2 d\tau\right]^{1/2} \tag{2.17}$$

and let $F(u)$ be a (complex) measurable function on $-\infty < u < \infty$. Then a necessary and sufficient condition that the Wiener integral of

$$F\left[\int_0^1 \beta(\tau)x(\tau)\,d\tau\right]$$

exists is that $e^{-u^2}F(Au)$ be of class L_1 on $-\infty < u < \infty$. Moreover, we will then have

$$\text{Ex } \{F\} = \pi^{-1/2}\int_{-\infty}^{+\infty} e^{-u^2}F(Au)\,du. \tag{2.18}$$

Applying this result to the present problem we have

$$\beta(\tau) = \tau, \tag{2.19}$$

and

$$A = \left[\int_0^1\left[\int_\tau^1 \xi\,d\xi\right]^2 d\tau\right]^{1/2}, \tag{2.20}$$

$$A = \sqrt{\tfrac{2}{15}}. \tag{2.21}$$

Also,

$$u = \int_0^1 \tau x(\tau)\,d\tau, \tag{2.22}$$

$$F(u) = e^{-u^2}, \tag{2.23}$$

$$F(Au) = e^{-(2/15)u^2}, \tag{2.24}$$

and finally,

$$\text{Ex } \{F\} = \pi^{-1/2}\int_{-\infty}^{+\infty} e^{-(17/15)u^2}\,du, \tag{2.25}$$

which yields

$$\text{Ex } \{F\} = \sqrt{\tfrac{15}{17}} \tag{2.26}$$

$$= 0.939336436\cdots.$$

Let us now consider the first stage of the approximation, namely, Ex $\{F\}_n$ given by Eq. (2.13). The functional $F[x_{n,k}(\cdot)]$ can be written (suppressing the subscript k)

$$F[x_n(\cdot)] = \exp\left[-\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j c_{ij}\right], \tag{2.27}$$

where

$$c_{ij} = (8/\pi^2)(2j-1)^{-1}(2i-1)^{-1} \times \int_0^1\int_0^1 \tau\tau'[\sin(j-\tfrac{1}{2})\pi\tau] \times [\sin(i-\tfrac{1}{2})\pi\tau']\,d\tau\,d\tau', \tag{2.28}$$

giving

$$c_{ij} = (-1)^{i+j}\left(\frac{128}{\pi^6}\right)\left(\frac{1}{2i-1}\right)^3\left(\frac{1}{2j-1}\right)^3. \tag{2.29}$$

It is easily verified that Ex $\{F\}_n$ can now be expressed as

$$\text{Ex } \{F\}_n = \pi^{-n/2} \times \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \exp\left[-\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j\,d_{ij}\right] \times d\xi_1\cdots d\xi_n, \tag{2.30}$$

where

$$d_{ij} \equiv c_{ij} + \delta_{ij} \tag{2.31}$$

(δ_{ij} = Kronecker delta). The quadratures on the right side of Eq. (2.30) are executed with a standard formula yielding

$$\text{Ex } \{F\}_n = |d|^{-1/2}, \tag{2.32}$$

where $|d|$ is the determinant with elements d_{ij} and, after a little calculation, one can obtain the result

[6] R. H. Cameron and W. T. Martin, Ann. Math. **45**, 386 (1944).

$$|d| = [1 + \gamma + \gamma/3^6 + \gamma/5^6 + \cdots + \gamma/(2n-1)^6], \tag{2.33}$$

where

$$\gamma = 128/\pi^6. \tag{2.34}$$

It is easy to verify that $\text{Ex}\,\{F\}_n \to \sqrt{\frac{15}{17}}$ as $n \to \infty$: The mth Bernoulli number is given by

$$B_m = \frac{2(2m)!}{\pi^{2m}(2^{2m}-1)}\left(1 + \frac{1}{3^{2m}} + \frac{1}{5^{2m}} + \cdots\right), \tag{2.35}$$

and setting $m = 3$ and using the fact that $B_3 = 1/42$ we have

$$\lim_{n\to\infty} |d| = 1 + \frac{2^7}{\pi^6}\left(\frac{\pi^6(2^6-1)}{84\cdot(6!)}\right) = \frac{17}{15}. \tag{2.36}$$

It is clear from Eq. (2.33) that $\text{Ex}\,\{F\}_n$ converges rapidly. To explicitly indicate this, the values (to nine significant figures) for $n = 1, 2, 3, 4, \infty$, are listed.

n	$\text{Ex}\,\{F\}_n$
1	0.939416290
2	0.939340593
3	0.939337062
4	0.939336593
∞	0.939336436

Thus, it is seen that considerable accuracy can be attained in this example with a small value for n.

Let us now suppose that the Monte Carlo sampling procedure is applied to obtain an estimate of $\text{Ex}\,\{F\}_n$, denoted by $\text{Ex}\,\{F\}_{n,R}$, which is defined by Eq. (2.15). If it is assumed that R is large enough to permit use of the central limit theorem, then the probability distribution for the error

$$\Delta_{n,R} = \text{Ex}\,\{F\}_{n,R} - \text{Ex}\,\{F\}_n \tag{2.37}$$

is

$$\Pr\{-\alpha < \Delta_{n,R} < \alpha\} \cong \frac{1}{\sigma\sqrt{2\pi}}\int_{-\alpha}^{+\alpha} \exp\left[-\left(\frac{\Delta_{n,R}}{\sqrt{2}\,\sigma}\right)^2\right] d\Delta_{n,R}, \tag{2.38}$$

where

$$\sigma = \sigma'/R^{1/2}, \tag{2.39}$$

and σ' is the standard deviation of the random variable $F[x_n(\cdot)]$, Eq. (2.27). The expectation value of $\{F[x_n(\cdot)]^2\}$ is readily obtained when it is noted that the computation is the same as for the evaluation of the right side of Eq. (2.30) with $2c_{ij}$ replacing c_{ij}. After a short computation, one finds for the case $n = 2$

$$\sigma' = 0.0785. \tag{2.40}$$

With this result and Eqs. (2.38) and (2.39) we can estimate the statistical error in the Monte Carlo sampling. One finds, for example, that with probability one-half

$$|\Delta_{2,500}| < 0.002. \tag{2.41}$$

The Monte Carlo sampling procedure was programmed for the Illiac. In this program, the process of selecting the random numbers $\xi_{i,k}$ is carried out by a subroutine which generates pseudo-random numbers $\xi^*_{i,k}$ for which the distribution function is approximately Gaussian. This generation of $\xi^*_{i,k}$ is done by taking a set of twenty pseudo-random numbers (each containing nine binary digits) which are approximately uniformly distributed on the interval $(-1, 1)$ and summing them and finally multiplying the result by a normalization factor. It is clear from this that the distribution function for $\xi^*_{i,k}$ will certainly deviate from the Gaussian distribution at the "tails"; however, experiments with this generator gave satisfactory agreement with the Gaussian distribution out to three standard deviations, and this was judged to be adequate for this computation.

In Fig. 1, results from fifty independent computations of $\Delta_{2,100}$ and $\Delta_{2,500}$ are shown. The ordinate is the value obtained for the probability distribution on the basis of the fifty evaluations of $\Delta_{2,R}$ and the abscissa is the normalized value of $\Delta_{2,R}$, i.e., $\Delta_{2,R}/\sigma$: For $R = 100$, $\sigma = 0.00785$ and for $R = 500$, $\sigma = 0.00350$. The distribution of the Δ's should be approximately Gaussian and, for comparison, the Gaussian curve is shown in this figure.

3. PARTITION FUNCTION CALCULATION

There are a number of areas in physics where problems may be formulated in terms of Wiener integrals or the related Feynman path integrals.[7] Here we are concerned with the formulation of the partition function for a quantum mechanical system in terms of certain Wiener integrals. The fundamental relation required for the present considerations has been obtained by Kac.[8] It relates the

[7] S. G. Brush, Revs. Modern Phys. **33**, 79 (1961).

[8] M. Kac, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley,* edited by J. Neyman (University of California Press, Berkeley, California, 1951); also *Lectures in Applied Mathematics, Volume I, Proceedings of the Summer Seminar, Boulder, Colorado, 1957* (Interscience Publishers, Inc., New York, 1958).

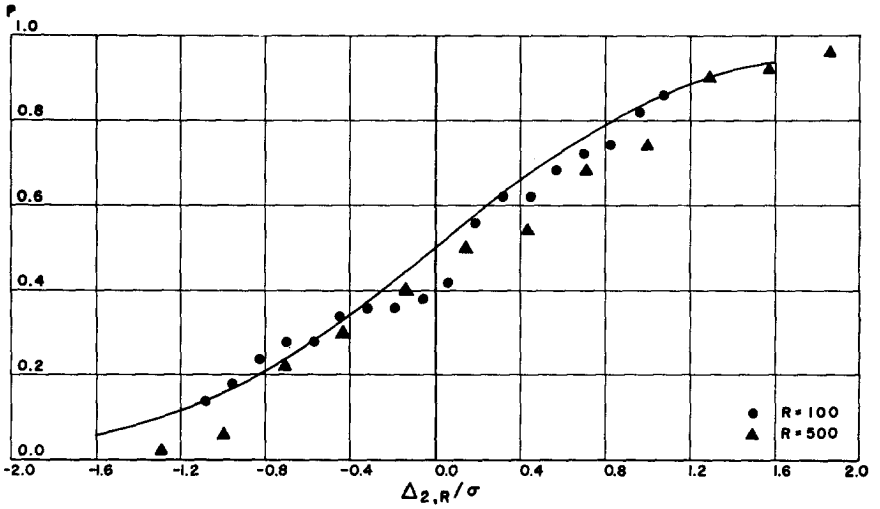


FIG. 1. Distribution of the error $\Delta_{2,R}$ in the estimate of the Wiener integral of the functional defined by Eq. (2.16). Solid curve is the Gaussian distribution function.

normalized eigenvectors $\psi_1(x)$, $\psi_2(x)$, $\cdots$ and eigenvalues E_1, E_2, $\cdots$ of the differential equation

$$-\tfrac{1}{4}\, d^2\psi/dx^2 + V(x)\psi = E\psi, \qquad (3.1)$$

to a conditional Wiener integral:

$$\frac{\exp\left[-(a-X)^2/t\right]}{(\pi t)^{1/2}} \mathrm{Ex}\left\{\exp\left[-\int_0^t V(x(\tau)+X)\, d\tau\right] \times \Big| x(t) = a - X\right\} = \sum_{i=1}^{\infty} \exp(-E_i t)\psi_i(a)\psi_i(X). \qquad (3.2)$$

It is clear that Eq. (3.1) is the one-dimensional Schrödinger equation for a particle of mass m, with a change of variables:

$$-(\hbar^2/2m)(d^2\psi/dy^2) + V(y)\psi = E\psi, \qquad (3.3)$$

with

$$x = (1/\hbar)(m/2)^{1/2} y, \qquad (3.4)$$

we obtain

$$-\tfrac{1}{4}\, d^2\psi/dx^2 + V((2\hbar^2/m)^{1/2}x) = E\psi. \qquad (3.5)$$

Thus, if we replace $V(x(\tau) + X)$ in Eq. (3.2) by $V((2\hbar^2/m)^{1/2}(x(\tau) + X))$, we may interpret the E_i's as the energy eigenvalues and the ψ_i's as the corresponding normalized eigenvectors of Eq. (3.3). It has been pointed out by Kac and others that if one sets $a = X$ and $t = \beta$ $(=1/kT)$ in Eq. (3.2) and then integrates over the variable X, the right side of Eq. (3.2) becomes

$$Z = \sum_{i=1}^{\infty} e^{-\beta E_i}, \qquad (3.6)$$

and thus one obtains the following expression for the partition function for the one particle problem:

$$Z = \left(\frac{1}{\pi\beta}\right)^{1/2} \int_{-\infty}^{+\infty} \mathrm{Ex}\left\{\exp\left[-\int_0^\beta V\, d\tau\right] \times \Big| x(\beta) = 0\right\} dX, \qquad (3.7)$$

where the argument of V is

$$(2\hbar^2/m)^{1/2}[x(\tau) + X].$$

This result has been used to obtain a series expansion for Z in powers of $\hbar$ obtained earlier in another way by Wigner and Kirkwood.[7] It will now be shown how the ideas developed in the preceding section can be used to estimate values of Z. In the following example our attention is restricted to a one-dimensional, single-particle problem, namely, the one-dimensional harmonic oscillator.

(a) One-Dimensional Harmonic Oscillator

The essential difference between the present calculation and the example of the preceding section is that now an estimate of the value of a *conditional* Wiener integral is required. In particular, the stochastic process represented by $x(\tau)$ is now required to have the value zero after a "time" interval equal to β has elapsed, as well as having the value zero at $\tau = 0$. This condition on $x(\tau)$ leads to an approximation formula of the type given in Eq. (2.13) with basis functions $\alpha_i(\tau)$ which satisfy the conditions $\alpha_i(0) = 0$ and $\alpha_i(\beta) = 0$. This new approximation formula results from the following theorem which is proved in the Appendix.

Theorem 3. Let C_a be the space of all functions $x(\tau)$ which are continuous on the interval $0 \leq \tau \leq 1$ and which satisfy the conditions $x(0) = 0$ and $x(1) = a$. Let $F[x(\cdot)]$ be continuous in the Hilbert

topology on the space C_a and let

$$|F[x(\cdot)]| \leq H\left(\int_0^1 x^2(\tau)\, d\tau\right) \tag{3.8}$$

on C_a where $H(u)$ is a monotonically increasing function of u and the conditional Wiener integral $\mathrm{Ex}\,\{H \mid x(1) = a\}$ satisfies

$$\mathrm{Ex}\,\{H \mid x(1) = a\} < \infty. \tag{3.9}$$

Let $\alpha_1(\tau), \alpha_2(\tau), \cdots$ be a complete orthonormal set on the space C_a.

Let

$$\int_\tau^1 \alpha_i(s)\, ds = \delta_i + \omega_i(\tau), \tag{3.10}$$

where δ_i is a constant and $\omega_i(\tau)$ contains no additive constant and let $1, \beta_1(\tau), \beta_2(\tau), \cdots$ be obtained by applying the Schmidt orthogonalization process to the sequence $1, \omega_1(\tau), \omega_2(\tau), \cdots$. Let the relation between the ω_i's and the β_i's be given by

$$\omega_k(\tau) = \sum_{i=1}^{k} \gamma_{ik}\beta_i(\tau), \tag{3.11}$$

and define

$$\beta_0(\tau) = 1. \tag{3.12}$$

Then

$$\mathrm{Ex}\,\{F \mid x(1) = a\}$$

$$= \lim_{n\to\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F[x_n(\cdot)]e_n(\xi)\, d\xi_1 \cdots d\xi_n, \tag{3.13}$$

where

$$x_n(\tau) = \sum_{k=1}^{n} \left(a\,\delta_k + \sum_{i=1}^{k} \gamma_{ik}\xi_i\right)\alpha_k(\tau), \tag{3.14}$$

and

$$e_n(\xi) = \pi^{-n/2} e^{-\xi_1^2 - \xi_2^2 - \cdots - \xi_n^2}. \tag{3.15}$$

We define

$$\mathrm{Ex}\,\{F \mid x(1) = a\}_n$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F[x_n(\cdot)]e_n(\xi)\, d\xi_1 \cdots d\xi_n, \tag{3.16}$$

where $x_n(\tau)$ is given Eq. (3.14) and use

$$\mathrm{Ex}\,\{F \mid x(1) = a\}_n$$

as an approximation for the conditional Wiener integral $\mathrm{Ex}\,\{F \mid x(1) = a\}$.

For this calculation the conditional Wiener integral appearing in Eq. (3.7) must be approximated and to simplify later formulas we normalize the time interval so that it has unit length and introduce the thermal wavelength λ, given by

$$\lambda = h(\beta/m)^{1/2}, \tag{3.17}$$

with

$$\lambda\!\!\!^{-} = \lambda/2\pi. \tag{3.18}$$

It is to be noted that λ is the de Broglie wavelength of a particle having kinetic energy equal to $\frac{1}{2}kT$. With this change in variables Eq. (3.7) becomes

$$Z = \frac{1}{\pi^{1/2}} \times \int_{-\infty}^{+\infty} \mathrm{Ex}\left\{\exp\left[-\beta \int_0^1 V(\sqrt{2}\,\lambda\!\!\!^{-}(x(\tau) + X))\, d\tau\right] \times \Big| x(1) = 0\right\} dX. \tag{3.19}$$

For the approximation of the conditional Wiener integral appearing here the orthonormal set

$$\alpha_j(\tau) = \sqrt{2}\sin j\pi\tau \tag{3.20}$$

is used, and it follows from Theorem 3 that

$$x_n(\tau) = \sum_{j=1}^{n} \xi_j \sqrt{2}\,\frac{\sin j\pi\tau}{j\pi}. \tag{3.21}$$

The potential energy function is

$$V(q) = \tfrac{1}{2}m\omega^2 q^2, \tag{3.22}$$

and it follows that

$$V(\sqrt{2}\,\lambda\!\!\!^{-}(x_n(\tau) + X)) = \lambda\!\!\!^{-2} m\omega^2 \left(X + \sum_{j=1}^{n} \frac{\xi_j \sqrt{2}\sin j\pi\tau}{j\pi}\right)^2. \tag{3.23}$$

Integrating this expression with respect to τ, and using the approximation given by Eq. (3.16) in (3.19) to form an approximation Z_n for Z, we obtain

$$Z_n = \pi^{-(n+1)/2} \int_{-\infty}^{+\infty} \cdots \times \int_{-\infty}^{+\infty} \exp\left[-\sum_{i=0}^{n}\sum_{j=0}^{n} d_{ij}\xi_i\xi_j\right] d\xi_0 \cdots d\xi_n, \tag{3.24}$$

where we define

$$\xi_0 = X, \qquad \alpha = (\hbar\omega\beta)^2$$

$$d_{ij} = \begin{cases} \alpha & (i = 0,\ j = 0) \\ 1 + \alpha/j^2\pi^2 & (i = j \neq 0) \\ 2^{3/2}\alpha/j^2\pi^2 & (i = 0,\ j \text{ odd}) \\ 2^{3/2}\alpha/i^2\pi^2 & (i \text{ odd},\ j = 0) \\ 0 & (\text{otherwise}). \end{cases} \tag{3.25}$$

Executing the $n+1$ quadratures one obtains

$$Z_n = |d|^{-1/2}, \tag{3.26}$$

where d is the determinant with elements d_{ij}. It is easy to show that this determinant can be written

$$|d| = \left(\prod_{j=0}^{n} d_{jj}\right)\left(1 - \sum_{\text{odd } j}^{n} \frac{d_{0j}^2}{d_{00}\, d_{jj}}\right). \tag{3.27}$$

It is now readily verified that we obtain the well-known result

$$Z = [2 \sinh (\hbar\omega\beta/2)]^{-1}, \tag{3.28}$$

when $n \to \infty$ in Eq. (3.26). The first factor on the right side of Eq. (3.27) converges to the limit

$$\lim_{n\to\infty}\left(\prod_{j=0}^{n} d_{jj}\right) = \lim_{n\to\infty}\left(\alpha \prod_{j=1}^{n}\left(1 + \frac{\alpha}{j^2\pi^2}\right)\right) = \alpha^{1/2} \sinh (\alpha)^{1/2}, \tag{3.29}$$

and the second factor on the right side of Eq. (3.27) converges to the limit

$$\lim_{n\to\infty}\left(1 - \sum_{\text{odd } j}^{n} \frac{d_{0j}^2}{d_{00}\, d_{jj}}\right) = \lim_{n\to\infty}\left(1 - \sum_{\text{odd } j}^{n} \frac{8\alpha}{j^2\pi^2(j^2\pi^2 + \alpha)}\right) = (2/\alpha^{1/2}) \tanh \tfrac{1}{2}\alpha^{1/2} \tag{3.30}$$

and the result (3.28) follows directly.

These results make it possible in this example to compute exactly the error committed by making the approximation represented in Eq. (3.24). Table I shows the least value of n for which the relative error

$$\rho = (Z_n - Z)/Z \tag{3.31}$$

satisfies the indicated conditions for different $\alpha^{1/2} = (\hbar\omega/kT)$. This table shows quite clearly that for $kT \approx 10\hbar\omega$, considerable accuracy is possible when $x(\tau)$ is represented by a sine series containing only a few terms. However, at very low temperatures, $kT \approx \hbar\omega$, the number of terms required for similar accuracy is considerably increased and, in particular, 51 terms are required for an accuracy of 0.1%. It is also to be noted that the relative error is inversely proportional to n; thus, one can predict that for $\alpha^{1/2} = 2.56$ about 3300 terms would be required for an accuracy of 0.01%. We also note, as expected from the above equations, that the number of terms required for a given accuracy is proportional to α, and thus inversely proportional to T^2.

Let us now estimate the statistical error which would result if Monte Carlo sampling is used to execute the quadratures in Eq. (3.24). The quadratic form Q, which appears in the exponent of the integrand of this expression, is

$$Q = \left[\alpha\xi_0^2 + \sum_{1}^{n} \xi_j^2\left(\frac{1}{j^2\pi^2}\right) + 2\xi_0 \sum_{\text{odd } j}^{n} \frac{\xi_j 2^{3/2}}{j^2\pi^2}\right] + \sum_{1}^{n} \xi_j^2. \tag{3.32}$$

Let us set up the Monte Carlo scheme following the same steps as in the example of the last section, ignoring the fact that we could diagonalize Q and execute the quadratures directly. The expression for Q is separated into two terms

$$Q = C + D, \tag{3.33}$$

where

$$C = \alpha\xi_0^2 + \sum_{1}^{n} \xi_j^2, \tag{3.34}$$

and D represents the remainder. The sampling procedure would now consist in picking random variables $\alpha^{1/2}\xi_0, \xi_1, \xi_2, \cdots, \xi_n$ from a normal distribution, evaluating $\exp(-D)$ for each such sample, and averaging the results as in Eq. (2.15). It is interesting to note that this procedure can get into serious difficulty, because for a certain range in α the variance is infinite. This is seen from a calculation of the expectation value of $\exp(-2D)$:

$$\text{Ex}\{\exp(-2D)\} = \pi^{-(n+1)/2} \times \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp(-C - 2D), \times d\xi_0 \cdots d\xi_n. \tag{3.35}$$

After a short calculation, similar to the one above in which Z_n was computed, it is found that in the limit $n \to \infty$ the determinant $|d|'$ of the quadratic form $C + 2D$ is

$$|d|' = (1/\alpha) \sinh (2\alpha)^{1/2}[-(\alpha/2)^{1/2} + 2 \tanh (\alpha/2)^{1/2}]. \tag{3.36}$$

As α increases from zero this determinant vanishes

TABLE I. Tabulation of the minimum value of n required for a specified relative error in the partition function.

$\alpha^{1/2}$	$\lvert\rho\rvert < 10^{-2}$	$\lvert\rho\rvert < 10^{-3}$	$\lvert\rho\rvert < 10^{-4}$
0.01	1	1	1
0.04	1	1	1
0.16	1	1	13
0.64	2	21	208
1.00	5	51	> 500[a]
2.56	33	332	> 500[b]

[a] $|\rho| = 1.015 \times 10^{-4}$ when $n = 500$.
[b] $|\rho| = 6.637 \times 10^{-4}$ when $n = 500$.

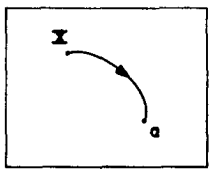


FIG. 2. Path in Γ space.

and becomes negative when α is about 7.3, from which it follows that the variance will be infinite for α above 7.3. The location of this infinity has no real significance, but is simply a consequence of the rather arbitrary way the variables in Eqs. (3.33) and (3.34) were split up to obtain a weight factor for the sampling. However, this indicates that one must be cautious in picking a weight factor for the Monte Carlo sampling. As it did not seem to be worthwhile to perform any numerical experiments with this example this concludes our consideration of the harmonic oscillator and we turn our attention to the two particle problem.

(b) Pair of Identical Interacting Particles

In this example we consider a system composed of two identical particles with a potential energy V, given by

$$V = V_{\text{box}} + V_{\text{int}} \tag{3.37}$$

where V_{box} is the usual "box" potential, equal to zero when both particles are inside a box having edge length L, and equal to infinity when either particle is outside of this box; and where V_{int} represents a spherically symmetric interaction between the particles given by

$$V_{\text{int}} = \begin{cases} \infty & r < a \\ -v & a \leq r \leq b \\ 0 & b < r, \end{cases} \tag{3.38}$$

where r is the distance between the particles and a, b, v are adjustable parameters. The discontinuous nature of this potential function does not cause any practical difficulties, for it may be regarded as the limiting case of a continuous function with very rapid changes taking place at the points $r = a$, $r = b$, and the sides of the box. For simplicity we restrict our attention to a two-dimensional system with the origin located at the center of the box.

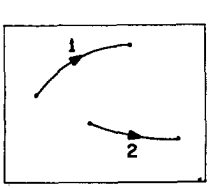


FIG. 3. Paths in μ space.

The Schrödinger equation for this system contains four coordinates, two for each particle. Equation (3.2) may be generalized in the obvious way for this four-dimensional problem[9]: $x(\tau)$ is now a vector with components, $x_1(\tau)$, $y_1(\tau)$, $x_2(\tau)$, $y_2(\tau)$ and X is a vector with components X_1, Y_1, X_2, Y_2, where the subscripts identify the particles. In this generalization it is to be understood that the components of $x(\tau)$ are generated by four, independent Wiener processes. A particular Wiener process is represented by a directed "path" in a four-dimensional space, called Γ space, as indicated symbolically in Fig. 2, where X denotes the initial point and a the terminal point, in accord with the notation in Eq. (3.2). The conditional Wiener integral is now viewed as an average taken over an ensemble of these paths with fixed end points. The path in Γ space can also be viewed as a pair of paths in the two-dimensional particle space, called μ space, as indicated in Fig. 3, where numbers distinguish the particles.

There are two classes of paths in Γ space which we distinguish, "closed" and "open." These are represented in Fig. 4 with their counterparts in μ

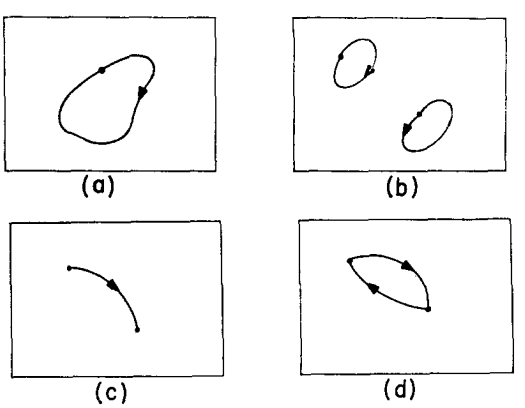


FIG. 4. (a) Closed path in Γ space; (b) Closed path in μ space; (c) Open path in Γ space; (d) Open path in μ space.

space. It can be seen from these figures that a closed path is characterized by having identical initial and terminal points. An open path is characterized by having different initial and terminal points, which are related in that one is a permutation of the coordinates of the other. For a closed path Eq. (3.2) becomes

$$\frac{1}{(\pi t)^{1/2}} \operatorname{Ex}\left\{\exp\left[-\int_0^t V(x(\tau) + X)\, d\tau\right] \middle| x(t) = 0\right\} = \sum_{i=1}^{\infty} \exp(-E_i t)\psi_i^2(X), \tag{3.39}$$

[9] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. 1, 48 (1960).

and, for an open path Eq. (3.2) becomes

$$\frac{\exp\left[-(a-X)^2/t\right]}{(\pi t)^{1/2}} \times \mathrm{Ex}\left\{\exp\left[-\int_0^t V(x(\tau)+X)\,d\tau\right] \,\middle|\, x(t)=a-X\right\} = \sum_{i=1}^{\infty} \exp\left(-E_i t\right)(-1)^{\sigma_i}\psi_i^2(X)\ , \tag{3.40}$$

where $\sigma_i = 0$ when ψ_i is symmetric with respect to interchange of the particles and $\sigma_i = 1$ when ψ_i is antisymmetric with respect to interchange of the particles. Addition of Eq. (3.40) to (3.39) yields an equation with only the symmetric eigenfunctions represented and subtraction of Eq. (3.40) from (3.39) yields an equation with only the antisymmetric eigenfunctions. It follows from this that the partition function for Bose-Einstein statistics is represented by

$$Z_B = \tfrac{1}{2}(Z_{\text{closed}} + Z_{\text{open}}), \tag{3.41}$$

and for Fermi-Dirac statistics it is represented by

$$Z_F = \tfrac{1}{2}(Z_{\text{closed}} - Z_{\text{open}}), \tag{3.42}$$

where

$$Z_{\text{closed}} = \frac{1}{(\pi\beta)^2}\iiiint_{-\infty}^{+\infty} \mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V(\sqrt{2}\,\lambda(x(\tau)+X))\,d\tau\right] \,\middle|\, x(1)=\Delta\right\} dX_1\,dY_1\,dX_2\,dY_2, \tag{3.43}$$

$$Z_{\text{open}} = \frac{1}{(\pi\beta)^2}\iiiint_{-\infty}^{+\infty} g(\Delta)\,\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V(\sqrt{2}\,\lambda(x(\tau)+X))\,d\tau\right] \,\middle|\, x(1)=\Delta\right\} dX_1\,dY_1\,dX_2\,dY_2, \tag{3.44}$$

with the definitions

$$\Delta = a - X, \tag{3.45}$$

(note that a is a function of X) and

$$g(\Delta) = \exp\left(-\Delta^2/\beta\right). \tag{3.46}$$

It should be recognized that Δ^2 is simply

$$\Delta^2 = 2\{[X_1 - X_2]^2 + [Y_1 - Y_2]^2\}; \tag{3.47}$$

i.e., it is twice the square of the initial distance (or terminal distance) between the particles. With these results one can apply the method described in earlier sections to obtain estimates of Z_{closed} and Z_{open} from which the partition function is obtained using Eq. (3.41) or Eq. (3.42).

It is convenient to make a change in variables in Eqs. (3.43) and (3.44). Let

$$X' = \sqrt{2}\,\lambda X/L, \tag{3.48}$$

where L is the edge length of the box which encloses the system, and making use of the fact that $V = \infty$ outside of the box, Eqs. (3.43) and (3.44) become

$$Z_{\text{closed}} = Z_0\iiiint_{-1/2}^{+1/2} \mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V(\sqrt{2}\,\lambda x(\tau) + X'L)\,d\tau\right] \,\middle|\, x(1)=0\right\} dX_1'\,dY_1'\,dX_2'\,dY_2', \tag{3.49}$$

$$Z_{\text{open}} = Z_0\iiiint_{-1/2}^{+1/2} g\left(\frac{L}{\sqrt{2}}\,\lambda\,\Delta'\right)\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V(\sqrt{2}\,\lambda x(\tau) + X'L)\,d\tau\right] \,\middle|\, x(1)=(L/\sqrt{2}\,\lambda)\,\Delta'\right\} \times dX_1'\,dY_1'\,dX_2'\,dY_2', \tag{3.50}$$

where

$$Z_0 = 4\pi^2(L/\lambda)^4, \tag{3.51}$$

and $(\Delta')^2$ is twice the square of the initial distance between the particles in terms of the new coordinates, X'. It will be recognized that $Z_0/2$ is the partition function for the perfect Boltzmann gas with two particles when L/λ is large, i.e., in the classical limit. It is readily verified that $Z_{\text{open}} \to 0$ as $L/\lambda \to \infty$ and, consequently, that Z_F and Z_B approach the classical partition function as $L/\lambda \to \infty$.

Let us now consider the application of our method to the evaluation of the integrals appearing on the right side of Eq. (3.49). For the conditional Wiener integral which appears on the right side of this equation we will use the approximation represented by Eq. (3.16) with $a = 0$ and the understanding that $x_n(\tau)$ has four components represented by

$$x_n^{(k)}(\tau) = \sum_{j=1}^{n} \xi_j^{(k)}\frac{\sin j\pi\tau}{j\pi} \qquad (k = 1, 2, 3, 4),$$

where $\xi_j^{(k)}$ represents a set of $4n$ independent random variables, each having a Gaussian distribution, and accordingly the weight factor in Eq. (3.16) becomes

$$e_n(\xi) = \exp\left(-\sum_{j=1}^{n}\sum_{k=1}^{4}[\xi_j^{(k)}]^2\right), \tag{3.53}$$

TABLE II. $Z_{n,R}$ for $L/\lambda = 2$; $Z_{\text{exact}} = 0.655$.

R	n = 10	20	30
100	0.690	0.690	0.710
200	0.720	0.735	0.710
300	0.757	0.733	0.723
400	0.772	0.725	0.715
500	0.774	0.736	0.720

and the integration extends over the $4n$ variables $\{\xi_i^{(k)}\}$. A similar approximation may be obtained for Eq. (3.46); however, we will only concern ourselves here with the estimation of Z_{closed}.

The Monte Carlo estimation of the $4n$-fold integrals is performed by the obvious extension of the first example [see Eq. (2.15)] except that we also perform the integration over X_1', Y_1', X_2', Y_2' by Monte Carlo sampling. This simply amounts to randomly picking values for X_1', Y_1', X_2', Y_2', from a distribution which is uniform on the interval $(-1/2, 1/2)$.

In the examples already considered it has been possible to evaluate the functional analytically for any $x_n(\tau)$ in the ensemble. However, in general, it will be necessary to evaluate $\int_0^1 V\, d\tau$ numerically and the trapezoidal rule may be used for this. It does not seem necessary to use a more sophisticated integration formula, inasmuch as the dominant errors can be expected to come from the approximations already made.

An analysis of the error that is to be expected from a numerical calculation of the partition function when the present method is used proves quite formidable and, excepting some relatively trivial cases, little progress in this direction has been made. We have examined the error "experimentally" by performing the computation of the partition function according to this method on the Illiac.

The program which was written to perform these computations uses the potential function given by Eqs. (3.37) and (3.38). When $b = 0$ the results can be checked against the partition function for the ideal gas. This program computes only the integrals appearing in the expression for Z_{closed}, Eq. (3.49), hence our results for $b = 0$ apply to the perfect Boltzmann gas.

In Tables II, III, IV estimates of $Z_{n,R}$ for different λ/L are shown, where $Z_{n,R}$ is the approximation

$$Z_{n,R} \cong \frac{Z_{\text{closed}}}{Z_0} = \iiiint_{-1/2}^{+1/2} \text{Ex}\left\{\exp\left[-\beta \int_0^1 V(\sqrt{2}\,\lambda x(\tau) + X'L)\, d\tau\right] \,\middle|\, x(1) = 0\right\} \times dX_1'\, dY_1'\, dX_2'\, dY_2', \qquad (3.54)$$

and R is the number of samples in the estimate and n represents the number of terms in the sine series in the approximation formula Eq. (3.52). In all cases there is no interaction, that is $b = 0$, and the exact value, $Z(\text{exact}) = Z_{\text{closed}}/Z_0$, is given for comparison. The figures in a single column of these tables are cumulative; thus, the estimate for $R = 200$ is obtained from the estimate for $R = 100$ by adding 100 new samples to the ensemble, etc. The figures in a single line of these tables are correlated in the following respect. A single random number sequence is used to pick the initial point for each path and this sequence is the same for each n. Thus, the paths of the ensemble used to obtain the estimates on any line of these tables have the same set of starting points for each n. The figure in the lower right corner of each table should be the most reliable estimate, and comparing these estimates with Z (exact) one finds, as expected, that the error decreases with λ/L, the error for the case $\lambda/L = \frac{1}{8}$ being about 2%. The figures in the last line of each table should be the best estimate for the corresponding value of n, and it is observed, as expected, that the accuracy of the best estimate increases with n,

TABLE III. $Z_{n,R}$ for $L/\lambda = 4$; $Z_{\text{exact}} = 0.812$.

R	n = 10	20	30
100	0.820	0.830	0.860
200	0.850	0.855	0.850
300	0.880	0.860	0.853
400	0.892	0.852	0.842
500	0.888	0.850	0.844

TABLE IV. $Z_{n,R}$ for $L/\lambda = 8$; $Z_{\text{exact}} = 0.904$.

R	n = 10	20	30
100	0.950	0.930	0.890
200	0.940	0.945	0.920
300	0.950	0.950	0.933
400	0.952	0.948	0.932
500	0.944	0.942	0.922

but very slowly. Finally we note that the convergence to the exact value is from above. An indication of the dependence of the results on the choice of the random number sequence is shown in Table V which displays the same results as for Table IV except that a different random number sequence has been used. The results in Table V show a smaller variation with n than those in Table IV. A comparison of these results indicates that the variation in estimates due to different random number sequences is of the same order as the variation due to different n; compare, for example, $Z_{10\,500}$ in Table V with $Z_{30\,500}$ in Table IV.

Two sets of computations were performed in which interactions between the particles were present. In the first set of calculations Z_{closed}/Z_0 was estimated for $L/\lambda = 8$, $a = 0$ (i.e. no hard core) and a range of values for b/λ and βv; this range included all combinations of $b = \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$ and $\beta v = -0.1, -0.2, -0.4, -1, -2, -4$. The sample size was $R = 400$, and $n = 20$. The results had the expected qualitative behavior. In particular, for $b/\lambda = \frac{1}{16}$ and $b/\lambda = \frac{1}{8}$ the estimate $Z_{20\,400}$ was independent of βv and identical to the estimate for the case $b = 0$, while for $b/\lambda = \frac{1}{4}$ and $\frac{1}{2}$ there was a dependence on βv in which $Z_{20\,400}$ decreased with decreasing βv and with increasing b. Thus, the particles did not interact (in this small set of samples) while the diameter of the "well" was less than one-half of a thermal wavelength.

In the second set of computations estimates of Z_{closed}/Z_0 were obtained for $a/L = \frac{1}{128}$, $b/L = \frac{1}{32}$, and a range of temperature given by starting values $L/\lambda = 32$, $\beta v = \frac{1}{4}$ and decreasing temperature in steps of 25%. The classical value of the partition function for this case can be calculated exactly:

$$Z_{\text{classical}} = Z_0 \iiiint_{-1/2}^{+1/2} e^{-\beta V(X_1', Y_1', X_2', Y_2')} \times dX_1'\, dY_1'\, dX_2'\, dY_2', \qquad (3.55)$$

and performing the quadratures on the right with $a/L = $, $b/L_{\frac{1}{128}} = \frac{1}{32}$ one obtains

TABLE V. $Z_{n,R}$ for $L/\lambda = 8$ and different random number sequence than that used in results for Table IV.

R	n = 10	20	30
100	0.920	0.910	0.940
200	0.925	0.920	0.920
300	0.927	0.913	0.910
400	0.915	0.905	0.905
500	0.922	0.914	0.914

TABLE VI. Estimated value of the quantum mechanical partition function compared with the classical partition function; the sample size R is 900 except for results marked by * where it is 4500. In all cases $n = 20$.

βv	$Z_{n,R} \cong Z_{\text{closed}}/Z_0$	$Z_{\text{classical}}/Z_0$
0.25	0.99	1.00
0.31	0.99	1.00
0.39	0.99	1.00
0.49	0.99	1.00
0.61	0.98	1.00
0.76	0.98	1.00
0.95	0.98	1.00
1.19	0.97	1.01
1.49	0.97	1.01
1.86	0.98	1.02
2.33	0.98	1.03
2.91	0.99	1.05
3.64	1.01	1.10
4.55	1.05	1.26
5.68	1.09	1.82
7.10	1.76	4.41
8.88	2.10*	21.1
11.10	5.83*	185.3

$$Z_{\text{classical}} = Z_0[0.997 + 2.80 \times 10^{-3} e^{\beta v}]. \qquad (3.56)$$

In Table VI the ratios $Z_{\text{classical}}/Z_0$ and Z_{closed}/Z_0 estimated by this method are shown. In these computations $n = 20$.

It is to be kept in mind that the thermal wavelength increases with decreasing temperature so that as the temperature decreases more paths tend to go out of the box, decreasing the estimate of the partition function; on the other hand the weight given to paths which interact via the attractive well increases with decreasing temperature tending to increase the estimate of the partition function. The apparent decrease in Z_{closed}/Z_0 with increasing βv at the start of the table is probably due to the predominance of the former of these effects while the latter increase is certainly due to the predominance of the latter effect. One might suspect that as $T \to 0$, Z_{closed}/Z_0 would once again decrease because every path tends to go out of the box; however, the weight given to the interacting paths is increasing exponentially to ∞ and therefore the limiting behavior is not clear from these simple intuitive observations. The larger value of $Z_{\text{classical}}/Z_0$ compared with Z_{closed}/Z_0 can be explained qualitatively, as due to the paths which wander out of the box, and which interact via the hard core. The statistical error in the estimate, or more precisely, the variance of the estimate increases with decreasing temperature. This error is indicated in Fig. 5 where the estimate of Z_{closed}/Z_0 for $\beta v = 1.49, 8.88, 11.10$ is plotted as a function of sample size R; the estimates for increasing sample size are cumulative.

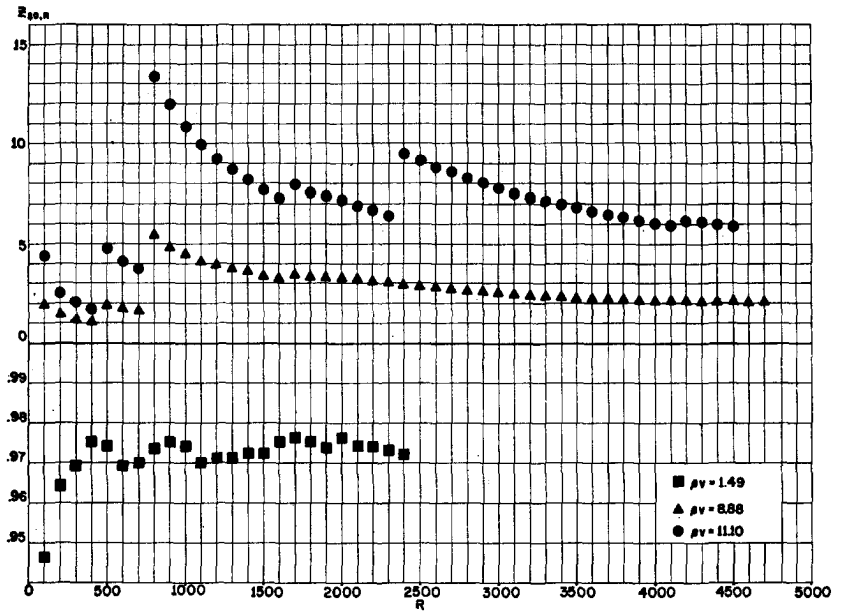


FIG. 5. Estimate of Z_{closed}/Z_0 as a function of sample size R, for three different values of βv.

4. CONCLUSION

In this work we have presented a method for numerically estimating the partition function of a quantum mechanical system by a Monte Carlo sampling scheme, in which the evaluation of the partition function is reduced to a simple repetitive process, well suited for automatic computation. Moreover, this process which consists simply in generating samples for averaging may be terminated, and then resumed at a later time to obtain more samples and an improved average. Finally, subject to the limitation in accuracy imposed by the expansion selected for $x(\tau)$, a result of arbitrary accuracy can be obtained by indefinitely continuing the sampling procedure.

The accuracy indicated by our numerical experiments suggests that the method can be useful in practical problems. For example, with a computer of the IBM 7090 variety one might compute Z_{open} for liquid helium, which is a fundamental parameter in Brush's[10] approximation of the partition function. However, with this method it is possible (at least in principle) to calculate the whole partition function instead of just part of it and it would be worthwhile to attempt to do this, even for small systems. There are three important points that need study in this connection. Before taking up these points it should be noted that in a practical problem it might be better to calculate the various derivatives of the partition function directly rather than by numerical differentiation which would tend to amplify the errors in the partition function.

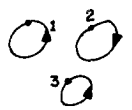
The first point needing investigation is the choice of the series approximation for $x(\tau)$. It is almost certain that a better choice than the sine series can be found. The present results appear to indicate that one can get good accuracy with a short sine series when $kT \approx 10 \times$ (ground state energy) but the number of terms required becomes quite large when $kT \approx$ (ground state energy). These results certainly depend on the potential function, and these estimates are to be taken only as a rough guide. Since it is known that $x(\tau)$ for the Wiener process is not a differentiable function it can be expected that basis functions with similar properties would be better than sine functions. In particular the functions obtained by integrating the Walsh functions might be a better choice. Another way to achieve an improvement would be to pick basis functions which reflect the fact that when the endpoints are fixed the largest deviations in the path tend to occur at the center of the time interval. Finally, the Simpson's rule formula of Cameron[5] might be tried; however, it appears to be complicated to use and may not be well suited for automatic computation; nevertheless, the application of the Simpson's rule formula needs further study before rejecting it as a useful tool.

The second point needing investigation is concerned with speeding up the sampling procedure. It would be possible to generate a "library" of paths and store them on magnetic tape. If this were

[10] S. G. Brush, Proc. Roy. Soc. (London) **A242**, 544 (1957); **A247**, 225 (1958).

TABLE VII. Terms for the 3-particle partition function.

Term value Z_0

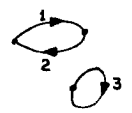


$$Z_0 = (\pi\beta)^{-9/2} \iiint_{-\infty}^{+\infty} \mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V\,d\tau\right] \,\middle|\, x(1) = 0\right\} dX_1\,dX_2\,dX_3$$

Term value Z_1

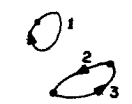
$$Z_1 = (\pi\beta)^{-9/2} \iiint_{-\infty}^{+\infty} g(\Delta_{1,3})\,\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V\,d\tau\right] \,\middle|\, x(1) = \Delta_{1,3}\right\} dX_1\,dX_2\,dX_3$$

Term value Z_2



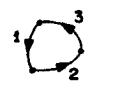
$$Z_2 = (\pi\beta)^{-9/2} \iiint_{-\infty}^{+\infty} g(\Delta_{1,2})\,\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V\,d\tau\right] \,\middle|\, x(1) = \Delta_{1,2}\right\} dX_1\,dX_2\,dX_3$$

Term value Z_3



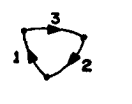
$$Z_3 = (\pi\beta)^{-9/2} \iiint_{-\infty}^{+\infty} g(\Delta_{2,3})\,\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V\,d\tau\right] \,\middle|\, x(1) = \Delta_{2,3}\right\} dX_1\,dX_2\,dX_3$$

Term value Z_4



$$Z_4 = (\pi\beta)^{-9/2} \iiint_{-\infty}^{+\infty} g(\Delta_{1,2,3})\,\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V\,d\tau\right] \,\middle|\, x(1) = \Delta_{1,2,3}\right\} dX_1\,dX_2\,dX_3$$

Term value Z_5



$$Z_5 = (\pi\beta)^{-9/2} \iiint_{-\infty}^{+\infty} g(\Delta_{1,3,2})\,\mathrm{Ex}\left\{\exp\left[-\beta\int_0^1 V\,d\tau\right] \,\middle|\, X(1) = \Delta_{1,3,2}\right\} dX_1\,dX_2\,dX_3$$

done the time required for generation of random numbers and evaluation of the series expression for the path could be eliminated, and with careful programming the tape-reading time could be made negligible.

The third point requiring study concerns a technique for picking the important contributions to the partition function in the N-particle problem. Since we have considered in detail only a two-particle problem it will be worthwhile to outline the application of the present method to a three-particle system.

In a three-particle problem there are six types of path in Γ space corresponding to the six permutation operations on the three particles. These paths in μ space are indicated in Table VII along with the term value for each; the term value being the contribution to the partition function made by the indicated class of paths. These expressions are written with the following notational conventions: X_i represents the initial coordinates, in a three-dimensional space, of particle i; Δ_{ij} represents a vector with components $X_i - X_j$, $X_j - X_i$, and zero; Δ_{ijk} represents a vector with components $X_i - X_k$, $X_j - X_i$, $X_k - X_j$; $g(\Delta)$ is given by Eq. (3.46) with Δ^2 being the sum of the squares of the components of Δ.

It is clear from symmetry that $Z_1 = Z_2 = Z_3$ and $Z_4 = Z_5$. Also, the terms Z_0, Z_4, Z_5 represent even permutations and Z_1, Z_2, Z_3 represent odd permutations. From Eq. (3.2) it follows that the partition function for Bose-Einstein statistics is given by the following linear combination of the term values,

$$Z_B = \tfrac{1}{6}Z_0 + \tfrac{3}{6}Z_1 + \tfrac{2}{6}Z_4 \qquad (4.1)$$

where the denominator 6 comes from normalization of the symmetrized eigenfunction. Similarly, the partition function for Fermi-Dirac statistics is given by

$$Z_F = \tfrac{1}{6}Z_0 - \tfrac{3}{6}Z_1 + \tfrac{2}{6}Z_4. \qquad (4.2)$$

It is convenient to think of the magnitude of the coefficients as representing a weight factor or a priori probability for the associated term value. Following these ideas one can construct the partition function for a system of N identical particles as a sum of terms of the above type multiplied by certain coefficients whose magnitude is the weight of the associated term. This suggests the possibility that an N-particle computation might include random sampling of terms according to these *a priori* weights.

Referring to the formulas for the terms in Table VII, one can readily see that all of the terms except

Z_0 tend to zero as the system becomes classical (i.e. as $\lambda/L \to 0$) because the factor $g(\Delta)$ becomes small except when Δ itself is very small. It is also noted that the factor $g(\Delta)$ favors terms representing a small number of permutations. These observations suggest the use of Δ itself as a random variable in the Monte Carlo sampling to automatically select the important contributions to the partition function. In a similar way one might anticipate the contribution that the Wiener integral itself will give for a particular Δ. With importance sampling along these lines it might be possible to greatly reduce the labor in N-particle computations.

ACKNOWLEDGMENTS

The author wishes to thank R. Haag and D. Woodward for helpful discussions.

APPENDIX

In this Appendix a proof of Theorem 3 of the text is presented. This proof requires a preliminary result obtained below.

Let $u_0, u_1, u_2, \cdots u_n$ be a set of $n+1$ Gaussian random variables with joint probability density

$$p = (2\pi)^{-(n+1)/2} |C|^{-1/2} e^{-U'C^{-1}U/2}, \tag{A1}$$

where C is the covariance matrix with elements

$$c_{ij} = \text{Ex}\{u_i u_j\}, \tag{A2}$$

and U is a vector with components $u_0, u_1, \cdots, u_n$.

Let $\{x(\tau), \tau \in (0,1)\}$ be a Wiener process, and let $\beta_0(\tau), \beta_1(\tau), \cdots \beta_n(\tau)$ be an orthonormal set of functions on the interval $(0,1)$. Let the u_i's be given by the Stieltjes integrals:

$$u_i = \int_0^1 \beta_i(\tau)\, dx(\tau) \qquad (i = 0, 1, \cdots, n). \tag{A3}$$

It follows from the definition of the Wiener process that the u's will be Gaussian random variables, since they are defined as sums of Gaussian random variables. Under these requirements, the conditional Wiener integral $\text{Ex}\{F(u_0, u_1, u_2, \cdots, u_n) \mid u_0 = a\}$ is given by

$$\begin{aligned}&\text{Ex}\{F(u_0, u_1, \cdots, u_n) \mid u_0 = a\}\\ &= (2\pi\sigma_0^2)^{1/2} \exp\left(\frac{a^2}{2\sigma_0^2}\right) \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F(a, u_1, \cdots, u_n)\\ &\quad \times p(a, u_1 \cdots u_n)\, du_1 \cdots du_n,\end{aligned} \tag{A4}$$

where σ_0^2 is the variance of u_0.

We now compute the elements of the covariance matrix.

$$\begin{aligned}c_{ij} &= \text{Ex}\{u_i u_j\}\\ &= \text{Ex}\left\{\int_0^1 \int_0^1 \beta_i(\tau)\beta_j(\tau')\, dx(\tau)\, dx(\tau')\right\},\end{aligned} \tag{A5}$$

and since $x(\tau)$ represents a Wiener process, it follows that

$$c_{ij} = \tfrac{1}{2} \int_0^1 \beta_i(\tau)\beta_j(\tau)\, d\tau, \tag{A6}$$

and making use of the fact that the β_i's are orthonormal

$$c_{ij} = \begin{cases} \frac{1}{2} & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{A7}$$

and we have, finally, the result

$$\begin{aligned}&\text{Ex}\{F(u_0, u_1, \cdots, u_n) \mid u_0 = a\}\\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F(a, \xi_1, \cdots, \xi_n) e_n(\xi)\, d\xi_1 \cdots d\xi_n,\end{aligned} \tag{A8}$$

where

$$e_n(\xi) = \pi^{-n/2} e^{-\xi_1^2 - \xi_2^2 - \cdots - \xi_n^2}. \tag{A9}$$

With this result we now proceed to the proof of Theorem 3. This proof is the same as Cameron's for Theorem 1 with necessary changes imposed by the use of a conditional Wiener integral.

For all $x \in C_a$, $\lim_{n\to\infty} x_n(\tau) = x(\tau)$ where

$$\begin{aligned}x_n(\tau) &= \sum_{k=1}^{n} \alpha_k(\tau) \int_0^1 \alpha_k(s) x(s)\, ds\\ &= \sum_{k=1}^{n} \alpha_k(\tau) \int_0^1 \left[\int_s^1 \alpha_k(u)\, du\right] dx(s)\\ &= \sum_{k=1}^{n} \alpha_k(\tau)\left[\int_0^1 \delta_k\, dx(s) + \int_0^1 \omega_k(s)\, dx(s)\right]\end{aligned} \tag{A10}$$

and, finally,

$$x_n(\tau) = \sum_{k=1}^{n} \alpha_k(\tau)\left\{\sum_{j=0}^{k} \gamma_{jk} \int_0^1 \beta_j(s)\, dx(s)\right\}, \tag{A11}$$

with $\gamma_{0,k} = \delta_k$.

Since $F[x(\cdot)]$ is continuous in the Hilbert topology, we have, for all $x \in C_a$, $\lim_{n\to\infty} F[x_n(\cdot)] = F[x(\cdot)]$. It follows from Bessel's inequality and monotonicity of $H(u)$ that

$$|F[x_n(\cdot)]| \leq H\left(\int_0^1 [x_n(\tau)]^2\, d\tau\right) \leq H\left(\int_0^1 [x(t)]^2\, dt\right), \tag{A12}$$

thus $F[x_n(\cdot)]$ is dominated by a Wiener integrable functional for all n. It follows that

$$\text{Ex}\{F \mid x(1) = a\} = \lim_{n\to\infty} \text{Ex}\{F[x_n(\cdot)] \mid x(1) = a\}. \tag{A13}$$

Noting that $F[x_n(\cdot)]$ is a function of the functionals $\int_0^1 \beta_0(\tau)\, dx(\tau), \cdots, \int_0^1 \beta_n(\tau)\, dx(\tau)$ and using the result established above (A8) we have the result given in Eq. (3.13) of the text.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Dimers on Rectangular Lattices*

TAI TSUN WU†
Brookhaven National Laboratory, Upton, New York
(Received August 17, 1962)

A simplified calculation of a known result of the dimer problem on a rectangular lattice is presented.

1. INTRODUCTION

THE combinatorial problem of filling a planar rectangular lattice completely by dimers has recently been solved by Kasteleyn[1] and by Temperley and Fisher,[2,3] All these authors make use of the Pfaffian. It is the purpose here to point out that the computation can be shortened by using a simple property of the permutation factor of the Pfaffian. No new result is obtained.

2. REVIEW OF THE PROBLEM

In order to explain the notations to be used, this dimer problem is first briefly reviewed. The terminology follows closely that of Kasteleyn.[1] For details the reader is referred to the papers of Kasteleyn[1] and Fisher.[3]

Consider a planar rectangular lattice with lattice points (j, k), $1 \leqq j \leqq m$ and $1 \leqq k \leqq n$, where m is even. A dimer can be used to cover symmetrically two consecutive lattice points, either $[(j, k), (j + 1, k)]$ (horizontal) or $[(j, k), (j, k + 1)]$ (vertical), and the straight line joining the two lattice points covered by a dimer is called a bond, again horizontal in the former case and vertical in the latter case. The problem is to find the combinatorial factor $g(N, N')$, defined as the number of different ways of covering the lattice points by N horizontal dimers and N' vertical dimers such that each lattice point is covered exactly once. Clearly

$$N + N' = \tfrac{1}{2}mn. \tag{1}$$

In particular, it is clear that $g(\frac{1}{2}mn, 0) = 1$. The bonds that appear in this particularly simple way of covering the lattice points are called C_0-bonds. In order to determine g, it is convenient to use the generating function

$$Z_{mn}(z, z') = \sum_{N,N'} g(N, N')z^N z'^{N'}, \tag{2}$$

where (1) is satisfied for each term on the right-hand side.

A covering of the lattice points by dimers can be expressed as a permutation P of (j, k):

$$[(j_1, k_1), (j_2, k_2)], [(j_3, k_3), (j_4, k_4)], \cdots$$
$$[(j_{mn-1}, k_{mn-1}), (j_{mn}, k_{mn})], \tag{3}$$

such that (i) either $j_{2\mu} = j_{2\mu-1} + 1$, $k_{2\mu} = k_{2\mu-1}$ or $j_{2\mu} = j_{2\mu-1}$, $k_{2\mu} = k_{2\mu-1} + 1$, and (ii) $j_{2\mu-1} \leqq j_{2\mu+1}$ and, if $j_{2\mu-1} = j_{2\mu+1}$, $k_{2\mu-1} < k_{2\mu+1}$. If the covering is the simple one corresponding to $N = \frac{1}{2}mn$ and $N' = 0$, then P is the identity; if (i) and (ii) are satisfied, then P is said to be permissible. Thus each covering corresponds to a permissible permutation. Let the permutation factor δ_P be 1 or -1 according as P is even or odd.

Some of the bonds that appear in a covering may be C_0-bonds, others are then called non-C_0-bonds. These non-C_0-bonds together with all C_0-bonds form, besides possibly some isolated C_0-bonds, a number s of (connected, closed) polygons of alternating C_0-bonds and non-C_0-bonds. An example may be found in Fig. 1 of Kasteleyn,[1] who also shows that

$$\delta_P = (-1)^s \tag{4}$$

and that (p. 1214 of reference 1) the number of strips in each polygon is odd.

3. A PROPERTY OF THE PERMUTATION FACTOR

Since all C_0-bonds are horizontal, all vertical bonds in each polygon must be non-C_0-bonds. Since the boundary of each strip contains exactly two vertical bonds, it follows from the last remark of Sec. 2 that the number of vertical non-C_0-bonds in each polygon must be twice an odd number. Thus it follows from (4) that

$$\delta_P = (-1)^{N'/2} = (-i)^{N'}. \tag{5}$$

Let (j, k) and (j', k') be two lattice points such

* Work performed under the auspices of the U. S. Atomic Energy Commission.
† Alfred P. Sloan Foundation Fellow. On leave from Harvard University, Cambridge, Massachusetts.
[1] P. W. Kasteleyn, Physica 27, 1209 (1961).
[2] H. N. V. Temperley and M. E. Fisher, Phil. Mag. 6, 1061 (1961).
[3] M. E. Fisher, Phys. Rev. 124, 1664 (1961).

that $j \leqq j'$ and, if $j = j'$, $k < k'$. Let $D(j, k; j', k')$ be defined by

$$\begin{cases} D(j, k;\ j+1, k) = z, \\ D(j, k;\ j, k+1) = Kz', \\ D(j, k;\ j', k') \quad = 0 \quad \text{otherwise}. \end{cases} \tag{6}$$

Then the Pfaffian of D is by definition

$$\text{Pf } D = \sum_P \delta_P z^N (Kz')^{N'}, \tag{7}$$

where the sum is over all permissible permutations, and $N(N')$ is the number of horizontal (vertical) bonds in the covering corresponding to P. Equation (7) should be compared with the alternative form of (2):

$$Z_{mn}(z, z') = \sum_P z^N z'^{N'}. \tag{8}$$

Accordingly, by (5), if

$$K = i, \tag{9}$$

then

$$Z_{mn}(z, z') = \text{Pf } D. \tag{10}$$

Thus the generating function is expressed as a Pfaffian.

4. EVALUATION OF THE PFAFFIAN

Let D be extended to be an antisymmetrical matrix by

$$\begin{cases} D(j, k; j, k) = 0, \\ D(j', k'; j, k) = -D(j, k; j', k'), \end{cases} \tag{11}$$

then

$$\text{Pf } D = (\det D)^{1/2}. \tag{12}$$

Let E_l be the $l \times l$ identity matrix and Q_l be an $l \times l$ matrix defined by

$$(Q_l)_{j,j+1} = -(Q_l)_{j,j-1} = 1, \tag{13}$$

other elements being zero. Then

$$D = zQ_m \times E_n + iz'E_m \times Q_n. \tag{14}$$

By a unitary transformation, Q_l can be diagonalized with the diagonal elements $2i \cos [j\pi/(l+1)]$. Therefore, by (10) and (12),

$$[Z_{mn}(z, z')]^2 = \prod_{j=1}^{m} \prod_{k=1}^{n} 2i\{z \cos [j\pi/(m+1)] + iz' \cos [k\pi/(n+1)]\}. \tag{15}$$

This is the desired answer obtained by Kasteleyn[1] and Temperley and Fisher.[2,3]

5. DISCUSSION

It must be emphasized that the present calculation follows almost step by step the beautiful work of Kasteleyn[1] and Temperley and Fisher.[2,3] The only new point here is the connection between δ_P and N', as expressed by (5). However, this simple observation does make the calculation shorter and more transparent. In Kasteleyn's assignment of ± 1 to the vertical bonds and in Fisher's reversal numbering of lattice points, loosely half of the translational symmetry is lost. By using the simple relation (5), the full translational symmetry can be maintained, and this leads to a sizable saving of labor in evaluating the Pfaffian.

ACKNOWLEDGMENTS

For very helpful discussions, I am greatly indebted to Professor T. D. Lee and Professor E. W. Montroll.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Approximate Solutions of the Modified Bloch Equations for Low Magnetic Fields

RICHARD J. RUNGE
California Research Corporation, La Habra, California
(Received June 5, 1962)

A set of modified Bloch equations proposed by Torrey, Olds, and Codrington for the description of the time dependence of the magnetic moment expectation value in low magnetic fields, is solved using two approximate methods. The equations are solved approximately for the steady state with a constant field and linearly polarized radio-frequency field applied and an error analysis of the approximate solution is given. A low-harmonic solution is developed and alternative numerical integration techniques are given. Results are exhibited for the low-field electron paramagnetic resonance in anthracene negative ion solution and asphaltene. The methods developed are applicable to a wide range of frequencies, applied field strengths, and ratios of transverse to longitudinal relaxation times.

I. INTRODUCTION

WITHIN recent years, efforts have been made[1-3] to improve the formulation of the problem of magnetic resonance and associated phenomena in its macroscopic aspect by modifying the equations given originally by Bloch[4] that govern the time dependence of the magnetic moment expectation value **M** of a single magnetogyric species in an applied magnetic field **H**. These modifications consist principally in introducing, in a phenomenological way, damping and driving terms of a more general sort than those given by Bloch in his original treatment. It is hoped, in these formulations, that a wider range of applied fields and frequencies can be properly included in these modified models.

In the low-field problem it is clear that the Bloch equations cannot apply since they are formulated for essentially the opposite case; hence, in order to better understand low-field phenomena, some modified set of equations must be considered. We chose to investigate a set of modified Bloch equations (MBE) proposed by Torrey, Olds, and Codrington[1] (TOC) since these appeared to be as general as any we could find in the sense that no restrictions on magnetic field amplitudes, frequency, or ratios of relaxation times were obviously imposed by the semiquantitative arguments used to support the case that such a system of equations might hold.

Specifically, we set out to obtain the best approximate steady-state solutions we could find to these equations for the case of a constant applied field in one direction with a linearly polarized radio-frequency field applied perpendicular to the constant field, for all possible field amplitude ratios, frequencies, and values of the longitudinal and transverse relaxation times.

In Sec. II some general properties of the solutions of the TOC-MBE are discussed, and their bearing on the methods of solution will become clear. In Sec. III, an error analysis applicable to a certain class of approximate steady-state solutions is developed and a set of functions, called saturation-frequency functions, are introduced that enable us to apply sufficient conditions in estimating the validity of particular approximate methods of solution. Section IV treats the solution of the equations in the general case via a method, previously applied by TOC,[1] Purcell,[3] and Garstens and Kaplan[5] to the case of equal relaxation times, which consists of keeping only the lowest harmonic terms of a Fourier series expansion of the solution. In Sec. V, this approximate method is developed in detail for the case of equal relaxation times in such a manner that saturation (by the rf field) can be included. Section VI proceeds to re-examine the problem from a numerical methods point of view and develops the solution of the MBE through forward numerical integration from an arbitrary starting point or from a starting point near the steady-state solution, using a Runge-Kutta solution to the MBE. The numerical approach is needed when the sufficient conditions for validity of the low-harmonic approximate solutions fail to hold; this is an augmentation to the analytical procedure. It is planned to present experimental results on low-field EPR in a subsequent paper.

[1] H. C. Torrey, J. D. Olds, and R. S. Codrington, Tech. Report, Nonr-45403, U. S. Office of Naval Research.
[2] A. Abragam, *Principles of Nuclear Magnetism* (Oxford University Press, New York, 1961).
[3] E. M. Purcell, Suppl. Nuovo cimento **6**, No. 3, 961 (1957).
[4] F. Bloch, Phys. Rev. **70**, 7 (1946).
[5] M. A. Garstens and J. I. Kaplan, Phys. Rev. **99**, 459 (1955).

II. MODIFIED BLOCH EQUATIONS

Letting $\mathbf{M}$ be the magnetic-moment expectation value, $\mathbf{H}$ the applied field, γ the magnetogyric ratio, χ_0 the static susceptibility, and T_1 and T_2 the longitudinal and transverse relaxation times, respectively, the TOC-MBE can be expressed[1] as

$$(d/dt)\mathbf{M} + \mathbf{M}/T_2 + (1/T_1 - 1/T_2)(\mathbf{M}\cdot\mathbf{H}/H^2)\mathbf{H} - \gamma\mathbf{M}\times\mathbf{H} = \mathbf{H}\chi_0/T_1. \quad (1)$$

Inspecting (1) we see that when $\mathbf{M}$ and $\mathbf{H}$ are perpendicular, the term in $\mathbf{M}\cdot\mathbf{H}$ vanishes, and $\mathbf{M}$ relaxes at a rate $1/T_2$; however, if $\mathbf{M} = f_0\mathbf{H}$, $\mathbf{M}$ relaxes at a rate $1/T_1$. When $T_1 \neq T_2$, the term in $\mathbf{M}\cdot\mathbf{H}$ complicates Eq. (1) considerably from an algebraic point of view.

We observe from (1) that the homogeneous equations associated with (1) are linear in $\mathbf{M}$, but not in $\mathbf{H}$. Thus, if $\mathbf{M}_a$ solves Eq. (1) with $\mathbf{H} = \mathbf{H}_a$ and $\mathbf{M}_b$ solves Eq. (1) with $\mathbf{H} = \mathbf{H}_b$, then $\mathbf{M}_a + \mathbf{M}_b$ can at best be only an approximate solution of (1) when $\mathbf{H} = \mathbf{H}_a + \mathbf{H}_b$. It is for this reason, in weak field problems, that one cannot represent a linearly polarized rf field as the sum $\mathbf{H}_+ + \mathbf{H}_-$ of two oppositely rotating, circularly polarized fields (a well-known device in strong field problems involving the original Bloch equations) and expect the solution $\mathbf{M}$ to be the sum $\mathbf{M}_+ + \mathbf{M}_-$ of the two circularly polarized solutions, even approximately. Thus, we are forced to solve (1) directly using a linearly polarized rf field if we want the solution for such an rf field.

We take $\mathbf{H}$ to be of the form

$$\mathbf{H} = (2H_1 \cos\omega t, 0, H_0), \quad (2)$$

where $\omega = 2\pi f$ is the circular frequency of the rf field, and H_0 is constant. Placing (2) in (1), the equations become, when written out,

$$(d/dt)M_x + M_x/T_2 + (1/T_1 - 1/T_2) \times (M_xH_x + M_yH_0)H_x/H^2 - \gamma H_0 M_y = \chi_0 H_x/T_1, \quad (3a)$$

$$(d/dt)M_y + M_y/T_2 + \gamma(M_xH_0 - M_zH_x) = 0, \quad (3b)$$

$$(d/dt)M_z + M_z/T_2 + (1/T_1 - 1/T_2) \times (M_xH_x + M_zH_0)H_0/H^2 + \gamma H_x M_y = \chi_0 H_0/T_1, \quad (3c)$$

where

$$H_x = 2H_1 \cos\omega t, \qquad H^2 = 4H_1^2 \cos^2\omega t + H_0^2 . \quad (4)$$

The weak field case is characterized by having H_1 comparable to or greater than H_0. Our special emphasis will be on the weak field case; however, we shall develop steady-state solutions to Eqs. (3) for *all* values of H_1, H_0, and thus obtain, among other things, Bloch's original solutions.

If $\mathbf{G}$ is a solution of the homogeneous version of (1), then we have, for such a transient solution,

$$(d/dt)\mathbf{G} + \mathbf{G}/T_2 + (1/T_1 - 1/T_2)(\mathbf{G}\cdot\mathbf{H})\mathbf{H}/H^2 - \gamma\mathbf{G}\times\mathbf{H} = 0, \quad (5)$$

and hence G satisfies

$$\frac{d}{dt}G = -\left(\frac{\sin^2\theta(t)}{T_2} + \frac{\cos^2\theta(t)}{T_1}\right)G, \quad (6)$$

where $\theta(t)$ is the angle between $\mathbf{G}$ and $\mathbf{H}$. Thus, G is of the form

$$G = G_0 \exp\left[-\int_{t_0}^{t}\left(\frac{\sin^2\theta}{T_2} + \frac{\cos^2\theta}{T_1}\right)dt'\right], \quad (7)$$

where the trigonometric quantity is always positive and lies in the range

$$1/T_1 \leq \sin^2\theta/T_2 + \cos^2\theta/T_1 \leq 1/T_2, \quad (8)$$

since $T_1 \geq T_2$. Thus, we arrive at the result that any transient solution of (1) will decay in magnitude to zero with increasing time at a rate no slower than e^{-t/T_1} decays. If T^* is the time a transient solution of (1) decays by e^{-2} (about one order of magnitude), then

$$T^* \leq 2T_1. \quad (9)$$

Let τ be the period of the rf field, then $\tau = 2\pi/\omega$. A quantity useful in our numerical discussion is K^*, where

$$K^* = 2T_1/\tau. \quad (10)$$

The positive integer closest to K^* is a measure of the minimum number of rf cycles which must elapse before we can be certain that transient solution of (1) has decayed to $1/e^2$.

Now suppose $\mathbf{M}$ is a solution of (1) and let $\mathbf{A}(t)$, the aperiodicity, be

$$\mathbf{A}(t) = \mathbf{M}(t + \tau) - \mathbf{M}(t). \quad (11)$$

Then, if $\mathbf{H}(t)$ is periodic of period τ [this includes $\mathbf{H}$ of (2)], we have from (1) that $\mathbf{A}$ is a solution of the homogeneous equation

$$(d/dt)\mathbf{A} + \mathbf{A}/T_2 + (1/T_1 - 1/T_2)(\mathbf{A}\cdot\mathbf{H})\mathbf{H}/H^2 - \gamma\mathbf{A}\times\mathbf{H} = 0, \quad (12)$$

which is (5). Hence, from (7) *et seq.*, we have the

result that **A** decays to zero at a rate no slower than $1/T_1$, i.e., in K^* rf cycles the aperiodicity in **M** must be down by $1/e^2$. Thus, in the case of a τ-periodic **H**, the solution approaches a steady-state, periodic function of period τ with transients disappearing at the above rate or faster. One can show, for **H** given by (2), that M is also bounded above by $\chi_0(H_0^2 + 4H_1^2)^{1/2}$ as the steady state is approached; hence, in the linearly polarized harmonic case, **M** describes an orbit in the steady state and is periodic of period τ and interior to a sphere of radius $\chi_0(H_0^2 + 4H_1^2)^{1/2}$.

Considering such a periodic $\mathbf{M}(t)$ which solves Eqs. (3), we can, by changing the variable in (3) from t to $t + \tau/2$, demonstrate that the components of $\mathbf{M}(t)$ satisfy

$$M_x(t + \tau/2) = -M_x(t), \tag{13a}$$

$$M_y(t + \tau/2) = -M_y(t), \tag{13b}$$

$$M_z(t + \tau/2) = M_z(t). \tag{13c}$$

Thus, in view of these remarks, including Eqs. (13), we are justified in assuming that the periodic steady-state solutions of (3) can be written as

$$M_x = A_1 \cos \omega t + B_1 \sin \omega t + (\text{higher odd harmonics}), \tag{14a}$$

$$M_y = C_1 \cos \omega t + D_1 \sin \omega t + (\text{higher odd harmonics}), \tag{14b}$$

$$M_z = M_0 + K_1 \cos 2\omega t + J_1 \sin 2\omega t + (\text{higher even harmonics}). \tag{14c}$$

III. ERROR BOUNDS FOR APPROXIMATE SOLUTIONS OF THE MODIFIED BLOCH EQUATIONS

In general, let $\mathbf{M}_e$ be an exact solution of (1) while $\mathbf{M}_a$ is an approximate solution to (1) given by some method of constructing such which is under investigation. Since $\mathbf{M}_a$ is not an exact solution of (1), the result of placing $\mathbf{M}_a$ in (1) will yield

$$\frac{d}{dt}\mathbf{M}_a + \frac{\mathbf{M}_a}{T_2} + \left(\frac{1}{T_1} - \frac{1}{T_2}\right)\left(\frac{\mathbf{M}_a \cdot \mathbf{H}}{H^2}\right)\mathbf{H} - \gamma \mathbf{M}_a \times \mathbf{H} - \chi_0 \mathbf{H}/T_1 = \boldsymbol{\delta}(t), \tag{15}$$

where $\boldsymbol{\delta}(t)$ is a discrepancy vector. We define a difference vector $\boldsymbol{\varepsilon}(t)$ between $\mathbf{M}_e$ and $\mathbf{M}_a$ as

$$\boldsymbol{\varepsilon}(t) = \mathbf{M}_a - \mathbf{M}_e. \tag{16}$$

Hence, combining (15), (16), and (1), we note that $\boldsymbol{\varepsilon}(t)$ satisfies

$$\frac{d}{dt}\boldsymbol{\varepsilon} + \frac{\boldsymbol{\varepsilon}}{T_2} + \left(\frac{1}{T_1} - \frac{1}{T_2}\right)(\boldsymbol{\varepsilon} \cdot \mathbf{H})\frac{\mathbf{H}}{H^2} - \gamma \boldsymbol{\varepsilon} \times \mathbf{H} = \boldsymbol{\delta}(t). \tag{17}$$

Plainly, if $\mathbf{M}_a = \mathbf{M}_e$, $\boldsymbol{\delta} = 0$. The scalar multiplication of $\boldsymbol{\varepsilon}$ through (17) yields the equation

$$\frac{d}{dt}\epsilon = -\epsilon\left(\frac{\sin^2\theta}{T_2} + \frac{\cos^2\theta}{T_1}\right) + \delta \cos \psi, \tag{18}$$

where $\theta(t)$ is the angle between $\boldsymbol{\varepsilon}$ and **H** while $\psi(t)$ is the angle between $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$.

Now consider the class of approximate solutions to (1) which have a discrepancy vector $\boldsymbol{\delta}$ which is bounded for all t by $\delta_{\max}$, i.e.,

$$0 \leq \delta(t) \leq \delta_{\max}. \tag{19}$$

From (18) we have the result that $d\epsilon/dt < 0$ whenever

$$\delta \cos \psi < \epsilon(\sin^2 \theta/T_2 + \cos^2 \theta/T_1). \tag{20}$$

Since $\delta \cos \psi \leq \delta_{\max}$, we can then say that if

$$\epsilon(\sin^2 \theta/T_2 + \cos^2 \theta/T_1) \geq \delta_{\max} \tag{21}$$

holds, then (20) holds and hence $(d/dt)\epsilon < 0$. For $T_1 > T_2$, a sufficient condition for (21) to hold is

$$\epsilon/T_1 \geq \delta_{\max} \tag{22}$$

since $\sin^2\theta/T_2 + \cos^2 \theta/T_1 \geq 1/T_1$.

Thus, from (22) if ϵ is in excess of $T_1\,\delta_{\max}$, $(d/dt)\epsilon$ will be negative, thus the error cannot grow indefinitely when δ is bounded above. Now, if $\mathbf{M}_e$ and $\mathbf{M}_a$ are continuous, then $\boldsymbol{\varepsilon}$ is also continuous. If the approximate method which provides us with $\mathbf{M}_a$ is adequate enough to reduce ϵ below $T_1\,\delta_{\max}$ at some time $t = t_0$, then for all $t \geq t_0$, ϵ will *remain* below $T_1\,\delta_{\max}$. This follows from the continuity of ϵ and the fact that $(d/dt)\epsilon$ is negative if ϵ exceeds $T_1\,\delta_{\max}$, i.e., ϵ cannot pass through $T_1\,\delta_{\max}$ from below with positive slope and, hence, being continuous, cannot escape the upper bound $T_1\,\delta_{\max}$ once below it.

We shall now develop some criteria applicable to methods for solving (1) approximately in the steady state [**H** being given by (2)] which will enable us to assert sufficient conditions, given the boundedness of $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$, for the validity of those approximate solutions in a mean square sense.

One would like to have an approximate solution $\mathbf{M}_a$ close enough to $\mathbf{M}_e$ such that $|\epsilon/M_a|$ was small compared to unity at all times. This is exceedingly stringent, however, and we shall adopt a compromise criterion, namely that η, given by

$$\eta = [\langle \epsilon^2 \rangle_{av} / \langle M_a^2 \rangle_{av}]^{1/2} \tag{23}$$

be small compared to one.

In the steady state, $\mathbf{M}_a$ will be required to be periodic of period τ in order that the method yielding it be acceptable; hence, $\boldsymbol{\epsilon}$ will be periodic, and the time average can be taken over one period of duration τ. Thus, in the steady state, the value of η is independent of the end points of the averaging interval, so long as this interval is of length τ, and we can state our minimum criterion as

$$\eta \ll 1. \tag{24}$$

Sufficient conditions for (24) to hold in the case of bounded δ and ϵ are as follows. Let ϵ fall below $T_1\ \delta_{max}$ at $t = t_0$, then for all $t \geq t_0$

$$\epsilon \leq T_1\ \delta_{max}. \tag{25}$$

Hence,

$$\langle \epsilon^2 \rangle_{av} \leq T_1^2\ \delta_{max}^2, \tag{26}$$

and thus η is bounded above by

$$\eta \leq T_1\ \delta_{max} / (\langle M_a^2 \rangle_{av})^{1/2}. \tag{27}$$

Consequently, (24) will hold provided

$$\delta_{max} \Big/ \left[\frac{\langle M_a^2 \rangle_{av}}{T_1^2}\right]^{1/2} \ll 1. \tag{28}$$

Now, if in the steady state, δ_{max} and $\langle \delta^2 \rangle_{av}$ are comparable (a circumstance which will prevail in the method to be developed later), then the strong inequality (28) can be replaced by the strong inequality,

$$[\langle \delta^2 \rangle_{av}]^{1/2} \Big/ \left[\frac{\langle M_a^2 \rangle_{av}}{T_1^2}\right]^{1/2} \ll 1. \tag{29}$$

Thus, (29) is sufficient to ensure (24). Inequality (29) suggests a definition, namely, if we call S the saturation-frequency (SF) function where

$$S^2 = \langle \delta^2 \rangle_{av} \Big/ \left(\frac{\langle M_a^2 \rangle_{av}}{T_1^2}\right), \tag{30}$$

then (29) is

$$S \ll 1, \tag{31}$$

and the smallness of the saturation-frequency function S, compared to unity is a sufficient criterion for (24) to hold and thus guarantees that $\mathbf{M}_a$ is an adequate mean square approximation to $\mathbf{M}$, in the steady state. The use of this terminology will become clear below where S will be directly related to the saturation $\gamma^2 H_1^2 T_1 T_2$ and the rf frequency ω.

One can proceed further and define component SF functions S_x, S_y, S_z as

$$S_{x,y,z}^2 = \langle \delta_{x,y,z}^2 \rangle_{av} \Big/ \left(\frac{\langle M_{ax,y,z}^2 \rangle_{av}}{T_1^2}\right), \tag{32}$$

where δ_x, δ_y, δ_z are the $\boldsymbol{\delta}$ components and M_{ax}, M_{ay}, M_{az} are the components of $\mathbf{M}_a$. Then the inequalities

$$S_{x,y,z} \ll 1 \tag{33}$$

are sufficient to guarantee component-wise that

$$\eta_{x,y,z} \ll 1, \tag{34}$$

where

$$\eta_{x,y,z} = \left(\frac{\langle \epsilon_{x,y,z}^2 \rangle_{av}}{\langle M_{ax,y,z}^2 \rangle_{av}}\right)^{1/2}. \tag{35}$$

The reason for introducing component SF functions came about from the experience in certain cases that S could be small, because of the large size of $\langle M_a^2 \rangle_{av}$ while a component (usually S_y) was comparable to unity, due to the fact that the mean square value of an $\mathbf{M}_a$ component was small compared to $\langle M_a^2 \rangle_{av}$ while the corresponding $\boldsymbol{\epsilon}$ component was not proportionally diminished. Consequently, we sharpen our criterion of acceptability of a solution to include (34) and (24). Note that (34) implies (24), hence, our sufficient conditions for (34) to hold are now (33) and its implicate (31).

Our standard procedure is thus to obtain $\mathbf{M}_a$, following the prescription of the approximate method, then use Eq. (15) to compute $\boldsymbol{\delta}$. Equations (32) and (30) are then employed to evaluate the SF functions S, S_x, S_y, and S_z, and, if these are all small compared to unity, the result ($\mathbf{M}_a$) is accepted with confidence.

IV. LOW HARMONIC SOLUTIONS. THE GENERAL CASE

Consider the problem of solving Eqs. (3) approximately in the steady state. Our method will be to employ the lowest harmonic terms in the Fourier series expansions (14) of $\mathbf{M}$, that is, we shall discard all odd harmonics in M_x, M_y from the third on and all the even harmonics in M_z from the fourth on. Torrey,[1] Purcell,[3] and Garstens[5] followed the same method for the case $T_1 = T_2$. However, they discarded in addition the second-harmonic terms in M_z. Attempting to keep all the harmonics would lead to an infinite system of equations in an infinite number of Fourier coefficients, hence some truncation procedure is indicated.

The truncated series of our method are therefore

$$M_x = A_1 \cos \omega t + B_1 \sin \omega t, \tag{36a}$$

$$M_y = C_1 \cos\omega t + D_1 \sin\omega t, \tag{36b}$$

$$M_z = M_0 + K_1 \cos 2\omega t + J_1 \sin 2\omega t. \tag{36c}$$

A quantity of considerable interest to us is the complex Bloch susceptibility χ'' defined through

$$2\omega H_1^2\chi''(\omega) = \frac{1}{\tau}\int_0^\tau \mathbf{H}\cdot\frac{d}{dt}\mathbf{M}\,dt \tag{37}$$

for the steady state. The quantity $2\omega H_1^2\chi''\mu$ is the average, over one rf period, of the rate of energy absorption of the magnetogyric species per unit volume. For the linearly polarized $\mathbf{H}$ given by (2), one obtains from (36) and (37) that χ'' is (whether higher harmonics in $\mathbf{M}$ are kept or not)

$$\chi'' = B_1/2H_1. \tag{38}$$

Consequently, the Fourier coefficient B_1 is of particular interest.

If we now place the low-harmonic expressions (36) in Eqs. (3), we obtain a set of two equations from (3a) and (3b) of type

$$(\cdots)\cos\omega t + (\cdots)\sin\omega t = \delta_x(t), \tag{39a}$$

$$(\cdots)\cos\omega t + (\cdots)\sin\omega t = \delta_y(t), \tag{39b}$$

and a third, from (3c), of type

$$(\cdots) + (\cdots)\cos 2\omega t + (\cdots)\sin 2\omega t = \delta_z(t), \tag{39c}$$

the components of the vector $\boldsymbol{\delta}$ appearing as the terms involving harmonics higher than the second which have been shifted to the right side of the equations. At this point, no restrictions have been placed on the Fourier coefficients of Eqs. (36). Setting the quantities in parentheses in Eqs. (39) equal to zero imposes such a restriction; however, the result is not an exact solution of Eqs. (3) but an approximate one, since $\boldsymbol{\delta}$, which is now the discrepancy vector defined in (15), does not vanish. We set the bracketed quantities in Eqs. (39) equal to zero, then, and obtain a system of seven equations in the seven Fourier coefficients A_1, B_1, C_1, D_1, K_1, J_1, M_0, namely, after abbreviating

$$\theta_0 = (1/T_1 - 1/T_2)/(H_0^2 + 2H_1^2), \tag{40}$$

$$\beta_0 = H_1^2/(H_0^2 + 2H_1^2), \tag{41}$$

$$\omega_0 = \gamma H_0, \tag{42}$$

$$\omega_1 = \gamma H_1, \tag{43}$$

we obtain the set

$$-A_1\omega T_2(1-\beta_0) + B_1(1-\beta_0+\theta_0T_2H_1^2) - \omega_0T_2(1-\beta_0)D_1 + \theta_0T_2H_0H_1J_1 = 0, \tag{44a}$$

$$A_1(1+\beta_0+3\theta_0T_2H_1^2) + \omega T_2B_1(1+\beta_0) - \omega_0T_2(1+\beta_0)C_1 + \theta_0T_2H_0H_1K_1 + 2\theta_0T_2H_0H_1M_0 = 2H_1T_2\chi_0(1+\beta_0)/T_1, \tag{44b}$$

$$\omega_0T_2B_1 - \omega T_2C_1 + D_1 - \omega_1T_2J_1 = 0, \tag{44c}$$

$$\omega_0T_2A_1 + C_1 + \omega T_2\,D_1 - \omega_1T_2K_1 - 2\omega_1T_2M_0 = 0, \tag{44d}$$

$$\theta_0T_2H_0H_1B_1 + \omega_1T_2\,D_1 - 2\omega T_2K_1 + (1+\theta_0T_2H_0^2)J_1 = 0, \tag{44e}$$

$$\theta_0T_2H_0H_1A_1 + \omega_1T_2(1+2\beta_0)C_1 + K_1(1+\theta_0T_2H_0^2) + 2\omega T_2J_1 + 2\beta_0M_0 = 2\beta_0\chi_0H_0T_2/T_1, \tag{44f}$$

$$\theta_0T_2H_0H_1A_1 + \omega_1T_2(1+\beta_0)C_1 + \beta_0K_1 + 2\omega T_2\beta_0J_1 + M_0(1+\theta_0T_2H_0^2) = \chi_0H_0T_2/T_1. \tag{44g}$$

The components of the discrepancy vector turn out to be the finite quantities in the third and fourth harmonics,

$$\delta_x = -H_1\cos 3\omega t[(1/T_2 - 1/T_1) \times (H_1A_1 + H_0K_1) - H_1a_{11}] - H_1\sin 3\omega t[(1/T_2 - 1/T_1) \times (H_1B_1 + H_0J_1) - H_1b_{11}], \tag{45a}$$

$$\delta_y = -\omega_1(K_1\cos 3\omega t + J_1\sin 3\omega t), \tag{45b}$$

$$\delta_z = H_1^2(a_{13}\cos 4\omega t + b_{13}\sin 4\omega t), \tag{45c}$$

where we have

$$a_{11} = \omega B_1 + A_1/T_2 - \omega_0C_1 - 2H_1\chi_0/T_1, \tag{46a}$$

$$b_{11} = B_1/T_2 - \omega A_1 - \omega_0\,D_1, \tag{46b}$$

$$a_{13} = 2\omega J_1 + K_1/T_2 + \omega_1C_1, \tag{47a}$$

$$b_{13} = J_1/T_2 - 2\omega K_1 + \omega_1\,D_1. \tag{47b}$$

It is then clear from Eqs. (45) that, for each $\boldsymbol{\delta}$ component as well as $\boldsymbol{\delta}$ itself, we have

$$\delta_{\max}^2 \approx 2\langle\delta^2\rangle_{av}. \tag{48}$$

Hence, saturation-frequency functions, defined through (30) and (32), are appropriate. Thus, S, S_x, S_y, and S_z can be calculated from (30), and (32) using the expressions

$$\langle M_x^2\rangle_{av} = \tfrac{1}{2}(A_1^2 + B_1^2), \tag{49a}$$

$$\langle M_y^2\rangle_{av} = \tfrac{1}{2}(C_1^2 + D_1^2), \tag{49b}$$

$$\langle M_z^2\rangle_{av} = M_0^2 + \tfrac{1}{2}(K_1^2 + J_1^2), \tag{49c}$$

obtained from Eqs. (36) and the expressions

$$\langle \delta_x^2\rangle_{av} = \tfrac{1}{2}H_1^2\{[(1/T_2 - 1/T_1)(H_1A_1 + H_0K_1) - H_1a_{11}]^2 + [(1/T_2 - 1/T_1) \times (H_1B_1 + H_0J_1) - H_1b_{11}]^2\}, \tag{50a}$$

$$\langle \delta_y^2\rangle_{av} = \tfrac{1}{2}\omega_1^2(K_1^2 + J_1^2), \tag{50b}$$

$$\langle \delta_z^2\rangle_{av} = \tfrac{1}{2}H_1^4(a_{13}^2 + b_{13}^2), \tag{50c}$$

obtained from Eqs. (45). From (50b) we note that the rms value of δ_y is proportional to the amplitude of the time dependent part of M_z of (36c).

Attempts to solve Eqs. (44) algebraically for the Fourier coefficients were begun and then abandoned in favor of a computer approach. These equations are solved for a given ω and a given set of H_1, H_0, γ, χ_0, T_1, and T_2 by a program written for an IBM type 7090 computer. The program also computes the SF functions of (30) and (32) via Eqs. (44), (46), (47), (49), and (50) and χ''/χ_0 from (38). The complete job is accomplished in about 1/2 sec for a given frequency, the output is then our solution and consists in a listing of A_1, B_1, C_1, D_1, M_0, K_1, J_1; S, S_x, S_y, S_z, and χ''/χ_0. Having this information, the SF functions are then inspected to see if sufficient conditions hold which enable us to have confidence in the results.

In order to test the low-harmonic approximation, a known exact solution of Eqs. (3) is employed. This is the linearly polarized Debye case, $H_0 = 0$, $H_x = 2H_1 \cos \omega t$. In this case, the solutions of Eqs. (3) are as follows, M_y and M_z decay to zero at a rate $1/T_2$ as t increases while M_x satisfies (3a) which becomes

$$\frac{d}{dt}M_x + \frac{M_x}{T_1} = \frac{2\chi_0 H_1}{T_1}\cos \omega t. \tag{51}$$

Hence, in the steady state, $M_y = M_z = 0$ while

$$M_x = \frac{2\chi_0 H_1}{1 + \omega^2 T_1^2}(\cos \omega t + T_1\omega \sin \omega t). \tag{52}$$

Consequently the only nonzero Fourier coefficients are A_1 and B_1, where

$$A_1 = 2H_1\chi' = 2\chi_0 H_1/(1 + \omega^2 T_1^2), \tag{53}$$

$$B_1 = 2H_1\chi'' = 2\chi_0 H_1 T_1\omega/(1 + \omega^2 T_1^2), \tag{54}$$

and we note that (54) yields the familiar Debye absorption law

$$\chi''/\chi_0 = \omega T_1/(1 + \omega^2 T_1^2). \tag{55}$$

Plainly, in this case, no higher harmonics are needed for an exact solution; hence, $\delta = 0$, thus the SF functions must all vanish.

When the computer program for solving Eqs. (44) is employed using $H_0 = 0$ as input, the above statements are confirmed in every detail to within the limits of accuracy of the IBM 7090 (i.e., the S, S_x values are of the order of 10^{-10} or so, rather than being exact zeros).

An additional test of the low-harmonic method was made in connection with a Bloch case problem. These are the strong-field cases for which the inequalities

$$H_1/H_0 \ll 1, \tag{56}$$

$$|(\omega - \omega_0)/\omega_0| \ll 1, \tag{57}$$

hold, i.e., strong-field problems involving a pronounced Larmor resonance absorption at $\omega_0 = \gamma H_0$. The original Bloch theory using a circularly polarized field superposition and the original unmodified Bloch equations, yields for such cases the familiar result[4]

$$\frac{\chi''}{\chi_0} = \frac{\omega_0 T_2}{2(1 + T_2^2(\omega_0 - \omega)^2 + \gamma^2 H_1^2 T_1 T_2)}, \tag{58}$$

for the complex part of the susceptibility. The solutions of Eqs. (44) using the low-harmonic method will then provide us with, among other things, a set of χ''/χ_0 values for a linearly polarized rf field to be compared with (58). A hypothetical electron resonance problem was computed in which $T_1 = 10^{-6}$ sec, $T_2 = 10^{-7}$ sec, $\gamma = |\gamma_e| = 1.76 \times 10^7$ G^{-1} sec^{-1}, $H_0 = 6.2832$ G, and $H_1 = 10^{-3}$ G. In this case, $\omega_0 T_2 = 11.058$ while $\gamma^2 H_1^2 T_1 T_2 \approx 10^{-3}$; hence, the case is a highly unsaturated one and the inequality (56), holds strongly. The Bloch equation (58) is insensitive to T_1. The resonance frequency here is $f_0 = \omega_0/2\pi = 1.76$ Mc and several frequencies above resonance were used to compute χ''/χ_0 from Eq. (58) and then from our Eqs. (44). Results are shown in Table I for this problem.

The results of our calculation showed that S_x exceeded all the other SF functions (S was of the order 10^{-8}) and hence only S_x is listed above. It is more than adequately small for acceptance of our solution as an accurate one over the entire frequency range. The agreement between the two theories is best at resonance, where the discrepancy is only about 0.2%. The steady deterioration of χ''/χ_0, obtained from the familiar Bloch equation, is clear in column 2 out on the higher frequency wings of the curve. This is not surprising inasmuch as at high frequencies (57) begins to weaken, i.e., $|(\omega - \omega_0)/\omega_0| = 0.28$ when $f = 2.26$ Mc/sec.

V. LOW HARMONIC SOLUTIONS FOR EQUAL RELAXATION TIMES

The case $T_1 = T_2$ provides a considerable simplification to the problem of solving the MBE. This case can occur in exchange narrowed paramagnetic resonance. It has been considered by Torrey,[1] Purcell,[3] and Garstens[5] using minor modifications of the low-harmonic method used here. None of these authors, however, have given a theoretical error analysis of their results. In this case, Eq. (1) is

$$\frac{d}{dt}\mathbf{M} + \frac{\mathbf{M}}{T_2} - \gamma\mathbf{M}\times\mathbf{H} = \frac{\chi_0\mathbf{H}}{T_2}, \tag{59}$$

and Eqs. (3), for the linearly polarized rf field are correspondingly simplified. Assuming that the steady-state solutions can be approximated by the low-frequency expressions (36), the general equations (44), in this case (note $\theta_0 = 0$) reduce to

$$T_2 b_{11} = -\omega T_2 A_1 + B_1 - \omega_0 T_2 D_1 = 0, \tag{60a}$$

$$T_2 a_{11} = A_1 + \omega T_2 B_1 - \omega_0 T_2 C_1 - 2\chi_0 H_1 = 0, \tag{60b}$$

$$\omega_0 T_2 B_1 - \omega T_2 C_1 + D_1 - \omega_1 T_2 J_1 = 0, \tag{60c}$$

$$A_1\omega_0 T_2 + C_1(1 + 2\omega_1^2 T_2^2) + \omega T_2 D_1 - K_1\omega_1 T_2 = 2\chi_0 H_0\omega_1 T_2, \tag{60d}$$

$$T_2 b_{13} = \omega_1 T_2 D_1 - 2\omega T_2 K_1 + J_1 = 0, \tag{60e}$$

$$T_2 a_{13} = \omega_1 T_2 C_1 + K_1 + 2\omega T_2 J_1 = 0, \tag{60f}$$

$$M_0 = \chi_0 H_0 - \omega_1 T_2 C_1. \tag{60g}$$

Hence, in this case we have from (60a), (60b)

$$a_{11} = b_{11} = 0 \tag{61}$$

and from Eqs. (60e) and (60f)

$$a_{13} = b_{13} = 0. \tag{62}$$

Consequently, from Eqs. (45), we obtain zero discrepancies in x and z,

$$\delta_x = \delta_z = 0. \tag{63}$$

The discrepancy δ_y, however, does not vanish and has a mean square value

$$\langle\delta_y^2\rangle_{av} = (\omega_1^2/2)(K_1^2 + J_1^2). \tag{64}$$

Thus for $T_1 = T_2$, $S_x = S_z = 0$, and $S < S_y$ since $\langle M^2\rangle_{av} > \langle M_y^2\rangle_{av}$ while $\langle\delta^2\rangle_{av} = \langle\delta_y^2\rangle_{av}$. We need only to concern ourselves with the SF component function S_y in this case.

From Eqs. (60e) and (60f), we obtain

$$K_1 = \frac{\omega_1 T_2(2\omega T_2 D_1 - C_1)}{1 + 4\omega^2 T_2^2}, \tag{65a}$$

TABLE I. Comparative values of χ''/χ_0, Bloch case.

f(frequency)	χ''/χ_0 Eq. (58)	χ''/χ_0 Eqs. (44), (38)	S_x
1.76 Mc/sec	5.529	5.540	0.457×10^{-5}
1.86	3.964	4.201	0.559×10^{-5}
1.96	2.144	2.399	0.667×10^{-5}
2.06	1.214	1.433	0.781×10^{-5}
2.16	0.756	0.939	0.901×10^{-5}
2.26	0.509	0.664	1.027×10^{-5}

$$J_1 = -\frac{\omega_1 T_2(2\omega T_2 C_1 + D_1)}{1 + 4\omega^2 T_2^2}. \tag{65b}$$

Applying Eqs. (65) to (64) we obtain

$$\langle\delta_y^2\rangle_{av} = \frac{\omega_1^2(\omega_1 T_2)^2}{1 + 4\omega^2 T_2^2}\left(\frac{D_1^2 + C_1^2}{2}\right), \tag{66}$$

Hence, using (36b), we obtain the result from (66) and (32) that

$$S_y(\omega) = \frac{(\omega_1 T_2)^2}{(1 + 4\omega^2 T_2^2)^{1/2}}. \tag{67}$$

At zero frequency, we set

$$S_0 = S_y(0) = (\omega_1 T_2)^2 = (\gamma H_1 T_2)^2. \tag{68}$$

Thus, we see that S_0 is the familiar saturation term $\gamma^2 H_1^2 T_1 T_2$ for the case $T_1 = T_2$, thus $S_y(\omega)$, which is expressible as

$$S_y(\omega) = \frac{S_0}{(1 + 4\omega^2 T_2^2)^{1/2}} \leq S_0, \tag{69}$$

depends on the saturation S_0 and the frequency ω; hence, the name saturation-frequency function. We observe from (69) that even in saturated cases (S_0 comparable to or greater than unity), S_y will be small compared to unity if the frequency ω is sufficiently large. Hence, the low-harmonic method will provide accurate solutions in saturated cases, provided the frequency is high enough so that $S_y \ll 1$. In *all* unsaturated cases, $S_0 \ll 1$, hence S_y will be small compared to unity regardless of the frequency. Consequently in all unsaturated cases, the low-harmonic method will yield accurate solutions of Eqs. (3) with $T_1 = T_2$. This last remark holds even if $H_1 \gg H_0$. In the Bloch cases, for which inequalities (56) and (57) hold, we have at $\omega = \omega_0$, if $\omega_0 T_2 \gg 1$,

$$S_y(\omega_0) \approx \frac{S_0}{2\omega_0 T_2} = \frac{\omega_1 T_2}{2}\left(\frac{H_1}{H_0}\right). \tag{70}$$

Therefore, in saturated Bloch cases, the low-harmonic solutions are valid for frequencies near and

beyond resonance provided $\gamma H_1 T_2$ is small compared to $2H_0/H_1$.

Equations (60a) through (60d) can now be solved for the remaining Fourier coefficients. We obtain for C_1 and D_1 the results

$$C_1 = (1/\omega_0 T_2)(A_1 + \omega T_2 B_1 - 2\chi_0 H_1), \quad (71a)$$

$$D_1 = (1/\omega_0 T_2)(B_1 - \omega T_2 A_1). \quad (71b)$$

The solutions for A_1 and B_1 are obtained for the case

$$S_v(\omega) \ll 1, \quad (72)$$

as

$$\frac{\chi'}{\chi_0} = \frac{A_1}{2H_1\chi_0} = \frac{[2(1+S_0)\omega^2T_2^2 + (1+2S_0+\omega_0^2T_2^2)(1+T_2^2(\omega_0^2-\omega^2))]}{[1+T_2^2(\omega_0^2-\omega^2)]^2 + 4\omega^2T_2^2 + 2S_0(1+T_2^2(\omega_0^2+\omega^2))} \quad (73a)$$

and

$$B_1 = \frac{2\chi_0 H_1[2\omega T_2(1+2S_0+\omega_0^2T_2^2) - \omega T_2\{1+2S_0+(\omega_0^2-\omega^2)T_2^2\}]}{[1+T_2^2(\omega_0^2-\omega^2)]^2 + 4\omega^2T_2^2 + 2S_0[1+T_2^2(\omega_0^2+\omega^2)]}. \quad (73b)$$

From Eqs. (38), (68), and (73b), we obtain the value of χ''/χ_0, subject to small S_v, as

$$\chi''/\chi_0 = \frac{\omega T_2(1+\omega^2T_2^2+\omega_0^2T_2^2+2\omega_1^2T_2^2)}{[1+T_2^2(\omega^2-\omega_0^2)]^2 + 4\omega_0^2T_2^2 + 2\omega_1^2T_2^2[1+T_2^2(\omega_0^2+\omega^2)]} \quad (74)$$

in which the saturation term, $S_0 = \omega_1^2T_2^2$, appears directly. Formula (74) is equivalent to one given earlier by Garstens[5] and Kaplan for H_1 twice the value used here. Garstens ignored the second harmonic M_z terms in his analysis; however, this does not effect the solution for A_1 and B_1, provided (72) holds.

In the unsaturated cases, the term $2\omega_1^2T_2^2$ in (74) has a negligible effect and (74) reduces to

$$\left(\frac{\chi''}{\chi_0}\right)_{u.s.} = \frac{\omega T_2(1+\omega^2T_2^2+\omega_0^2T_2^2)}{[1+T_2^2(\omega^2-\omega_0^2)]^2 + 4\omega_0^2T_2^2}. \quad (75)$$

Equation (75) is equivalent algebraically to results given earlier by Torrey[1] and then by Purcell,[3] the latter using twice our H_1. It should be emphasized that (74) and (75) are all-field results subject only to the restriction of small S_v, i.e., (72). If we set $\omega_0 = 0$ in (74), terms cancel and (74) reduces to the Debye result, (55). In addition, if inequalities (56) through (57) hold and terms such as $\omega + \omega_0$ are replaced by $2\omega_0$ in (74), then (74) reduces to the Bloch theory Eq. (58) for $T_1 = T_2$ in the saturation term, provided $\omega_0 T_2 \gg 1$.

A discussion of the case $T_1 = T_2$ with its particularly simple SF function, S_v of (69) reveals something about the circumstances under which low-harmonic solutions might be expected to fail. In cases of high saturation, where the numerator S_0 in (69) is not small compared to unity and for frequencies ω low enough so that the denominator in (69) does not outweigh this effect, S_v would not be small compared to unity and our confidence in the accuracy of the low-harmonic solutions would be shaken.

Such cases, if they arise in general for $T_1 \neq T_2$ or $T_1 = T_2$, can often be resolved by means of the numerical methods of the next section.

VI. NUMERICAL METHODS FOR SOLVING THE MODIFIED BLOCH EQUATIONS

In cases where one or more of the SF functions are not small compared to unity, numerical methods can be successfully used to obtain the steady-state solutions of Eqs. (3). The exact solution of Eqs. (3) can be written as $\mathbf{G}(t) + \mathbf{M}(t)$, where $\mathbf{G}(t)$ is a transient satisfying the homogeneous version of Eqs. (3) while $\mathbf{M}(t)$ is the steady-state solution we are after.

One can solve Eqs. (3) by taking an arbitrary starting value $\mathbf{G}(0) + \mathbf{M}(0) = \mathbf{M}_0$ and then forward integrating the equations numerically out to some time t. If the starting value, $\mathbf{M}_0$, is equal to the true steady-state value of $\mathbf{M}(t')$ for some t' in $0 \leq t' \leq \tau$, then $\mathbf{G}(0) = 0$ and the numerical integration gives the steady-state solution at once. If this fortuitous circumstance does not occur, then $\mathbf{G}(0) \neq 0$, and the numerical integration must be carried out far enough so that the transient $\mathbf{G}(t)$ will decay in accordance with Eq. (7) to an insignificantly small value. Thus, in order to be sure of reducing such a perturbing transient by about one order of magnitude, the numerical integration of Eqs. (3) must be carried out through K^{**} rf cycles where K^{**} is the integer closest to K^* of (10) and above K^*.

Once the transients have been reduced by a sufficient amount, the numerical solution becomes

noticeably periodic and the degree of periodicity, measured by the smallness of

$$\langle|\mathbf{M}(t+\tau)-\mathbf{M}(t)|^2\rangle_{av}/\langle M^2\rangle_{av} \tag{76}$$

can be used to gauge the stopping point of the forward integration. A simpler rule, however, would be to require that $|\mathbf{M}(t'+\tau)-\mathbf{M}(t')|$ be a small fraction of rms M for a few selected values t' in $t \geq t' \geq t-\tau$.

If K^{**} is a large integer, then a casual approach is not recommended in choosing the starting value $\mathbf{M}_0$. In cases of this type, we successfully used the low-harmonic solutions of Secs. IV and V in some instances, to provide the starting value $\mathbf{M}_0$. In such cases, the numerical solution is assisted by the low-harmonic theory, the idea being that even when the low-harmonic results are associated with SF functions comparable to or large compared to unity, they may still be close enough to the correct results since the SF functions are essentially upper bounds to the error. Should this in fact be the case, the numerical solutions will quickly bear it out. In using the low-harmonic theory to assist the numerical integration, we conveniently set $t = 0$ in Eqs. (36) and take as starting values the $\mathbf{M}_0$ components,

$$M_{0x} = A_1, \tag{77a}$$

$$M_{0y} = C_1, \tag{77b}$$

$$M_{0z} = M_0 + K_1, \tag{77c}$$

where A_1, C_1, M_0, and K_1 are obtained from the solution of Eqs. (44).

When K^{**} is a relatively small integer, it makes little difference what starting value of $\mathbf{M}_0$ is chosen, provided $G(0)$ is not excessively large. Taking any $\mathbf{M}_0$ inside the steady-state sphere of radius $\chi_0(H_0^2+4H_1^2)^{1/2}$ guarantees that G_0 cannot exceed the diameter of this sphere. We customarily took $\mathbf{M}_0 = 0$ in small K^{**} cases (i.e., $K^{**} \leq 10$), starting in effect from quiescence.

The numerical scheme chosen to perform the forward integration of Eqs. (3) was the Runge-Kutta scheme suitable to three simultaneous first-order equations,[6,7] i.e., Eqs. (3). This method also goes by the name[7] of Kutta's Simpson's 1/6th rule. Using this rule, $M_x(t+\Delta t)$, for example, is given as

$$M_x(t+\Delta t) = M_x + \Delta M_x, \tag{78}$$

where

$$\Delta M_x = \tfrac{1}{6}(\alpha_1 + 2\alpha_2 + 2\alpha_3 + \alpha_4), \tag{79}$$

and the differential quantities α_r are coefficients evaluated by the prescription of the method and are listed in standard references.[6,7] In using the Runge-Kutta scheme, we took time steps Δt normally in the range

$$\tau/50 \leq \Delta t \leq \tau/100, \tag{80}$$

and obtained very stable results even when the forward integration was carried out as far as 4000 time steps or about 40 to 80 rf cycles. The computation was done via a numerical program written for the IBM type 7090 computer. Each calculation was run at normal time step Δt and then rerun out to the same total time at twice the time step. The error in the Runge-Kutta scheme is given[6] as

$$\epsilon_{RK} = |\mathbf{M}^{(2)}(t) - \mathbf{M}^{(1)}(t)|/15, \tag{81}$$

where $\mathbf{M}^{(1)}(t)$ is the result at t for normal time step Δt and $\mathbf{M}^{(2)}(t)$ is the result at t for the same calculation at double time step, $2\Delta t$.

At the termination of each pair of Δt and $2\Delta t$ calculations, the error is obtained from (81) applied to M_x for one or more values of t' in the range $t-\tau \leq t' \leq t$, where t is the total forward integration time. In addition, information is provided by the numerical program which enables us to calculate ϵ_{RKy} and ϵ_{RKz} at t. Values of $\mathbf{M}$ at t and at $t-\tau$ are listed for purposes of checking the periodicity attained over the last cycle of the numerical integration for both Δt and $2\Delta t$. The numerical program also computes χ''/χ_0 directly from the integral (37) which in the linearly polarized case takes the form, using (37), (1), and (2),

$$\frac{\chi''}{\chi_0} = \frac{\tau(H_0^2+2H_1^2)}{4\pi T_1 H_1^2} - \frac{1}{(4\pi T_1\chi_0 H_1^2)} \times \int_0^\tau (2H_1 M_x \cos\omega t + H_0 M_z)\, dt. \tag{82}$$

The integral in (82) is evaluated by summing over elementary rectangles of width Δt in our numerical program. This is sufficiently accurate for our purposes, especially since Δt is usually no greater than 1/50 of τ. Normal IBM type 7090 times for these numerical solutions of Eqs. (3) were about 1/15 sec per time step Δt for a given ω. Hence, for $\Delta t = \tau/100$, 40 rf cycles could be forward integrated accurately in about 4.5 min.

VII. RESULTS

In Fig. 1, χ''/χ_0 is exhibited for a case of equal relaxation $T_1 = T_2 = 5.07 \times 10^{-8}$ sec, describing

[6] L. Collatz, *The Numerical Treatment of Differential Equations* (Springer-Verlag, Berlin, Germany, 1960), 3rd ed.-translated, Chap. II, Sec. 2.

[7] H. Levy and E. A. Baggot, *Numerical Solutions of Differential Equations* (Dover Publications, New York, 1950), 1st U. S. edition, Chap. III.

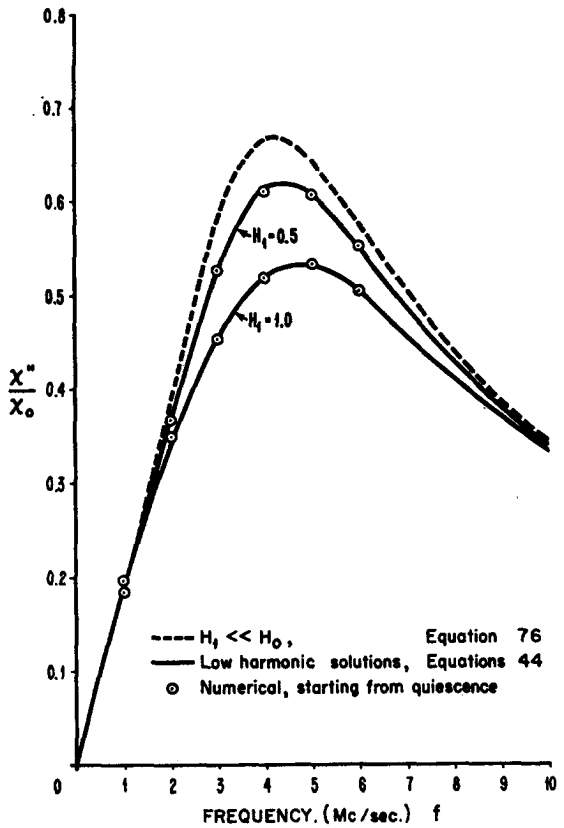


FIG. 1. Anthracene negative ion electron resonance absorption in low fields. For all curves, $H_0 = 1$ G, $\gamma = 1.76 \times 10^7$ G^{-1} sec^{-1}, $T_2 = T_1 = 5.07 \times 10^{-8}$ sec.

the electron resonance absorption of an exchange narrowed solution of anthracene negative ion (ANI) with γ taken as the electron value 1.76×10^7 G^{-1} sec^{-1}. The figure shows as solid curves χ''/χ_0 obtained from the solution of the low-harmonic Eqs. (44) for H_1 assuming the values 0.5 and 1.0 G. In each curve H_0 is 1.0 G. The dashed curve is a plot of Eq. (75) for the limiting case $H_1 \ll H_0$. The frequency range is from 0 to 10 Mc/sec.

The small circles in the figure are the values obtained using the Runge-Kutta numerical integration of Sec. VI, taking $\mathbf{M}_0 = 0$. In this case K^* of (10) for the shortest period ($\tau = 10^{-7}$ sec) is $K^* = 1.014$; hence, in one rf cycle or less the aperiodicity is diminished by an order of magnitude. These numerical calculations were carried forward through 10 rf cycles with $\Delta t = \tau/50$, and the solutions were periodic to one part in 10^6 at termination. Disagreement between $M_x^{(1)}$ and $M_x^{(2)}$ occurred only in the fourth significant figure if at all for the Δt and $2\Delta t$ calculations.

For $H_1 = 1.0$ G, we have $S_y(0) = 0.796$ and S_y does not approach 0.1 until frequencies just beyond 10 Mc/sec are obtained. Thus, this case is saturated over the entire frequency range, particularly at low frequencies. Nevertheless, the low-harmonic solution is quite adequate to compute χ''/χ_0. The very accurate numerical calculation bears this out, as evidenced in Fig. 1. The largest discrepancy between the two was about 3% at 2 Mc/sec.

For $H_1 = 0.5$ G, we have $S_y(0) = 0.199$, with S_y diminishing to 0.06 at 5 Mc/sec. This shows a lower level of saturation than the preceding and again the result of the numerical solution for χ''/χ_0 is in excellent agreement with the low-harmonic value. The maximum discrepancy is 0.3% at 2 Mc/sec. It should be noted here that we use the computed solutions (44) rather than Eq. (74) to obtain the solid curves in Fig. 1.

For H_1 below 0.25 G, we have $S_y(0) \leq 0.05$ and the effects of saturation are negligible. In these cases, Eq. (74) [and ultimately (75)] suffices to described the susceptibility in ANI.

To illustrate the situation for $T_1 \neq T_2$, we considered a highly saturated problem involving asphaltene, using the electron γ, $T_2 = 10^{-8}$ sec and taking $T_1 = 5 \times 10^{-6}$ sec. This T_1 was obtained from EPR saturation measurements made by Thompson[8] of the California Research Corporation. The fields were $H_1 = 1.0$ G and $H_0 = 0.5$ G; hence, the saturation term is $\omega_1^2 T_1 T_2 = 15.5$. Frequencies as high as 20 Mc/sec were considered and for $\tau = 0.5 \times 10^{-7}$ sec, we have for K^{**} the value 200. Thus, in the numerical solutions, the starting value $\mathbf{M}_0$ was obtained from the low-harmonic theory [Eqs. (77)]. When Eqs. (44) were solved, it was noted that the SF functions were enormous, i.e., S_y attained a maximum value of 516 at 1 Mc/sec, while the smallest, S_x, never fell below 58.

The numerical calculations were carried out through 20 rf cycles and then, using the 20-cycle/sec solutions as input, were extended another 20 cycles, making a total of 40 rf cycles at $\Delta t = \tau/100$. No change in χ''/χ_0 out to the sixth figure resulted from this additional computing. Periodicity to one part in 10^6 was obtained, and the error in M_x was reflected in the fourth or fifth significant figure. The results for three selected frequencies are shown in Table II.

At 10 Mc/sec, the discrepancy is about 5% in χ''/χ_0 values and decreases at higher frequencies. Thus, it would appear that assisting the numerical

[8] D. D. Thompson (private communication). The validity of extending such T_1 values to low fields is a separate problem which will not be discussed here, even though it may have a profound effect on the usefulness of the MBE.

solution with the low-harmonic solution is worthwhile in this instance, despite the large SF function values (a few of which are shown in Table II).

Closer investigation revealed that the low-harmonic solution was indeed a poor approximation for **M** at the lowest frequency, 1 Mc/sec. Equation (36a) asserts that in the low-harmonic approximation, M_x is a pure sinusoid of frequency. The numerical solution, upon reaching the steady state, introduces any necessary higher harmonics in **M** not included in the low-harmonic solutions. The numerical program plots the values of M_x for each time step Δt over the last rf cycle of integration. This result is shown in Fig. 2 in this case for a frequency of 1 Mc/sec, in which it is observed that M_x is not a pure sinusoid but contains a significant amount of higher harmonics. The presence of these higher harmonics, however, has apparently not seriously affected the value of B_1, as evidenced in Table II by the small percent discrepancies in χ''/χ_0. The numerical calculations of M_x over the last cycle for frequencies of 10 and 20 Mc/sec are somewhat more sinusoidal appearing than M_x of Fig. 2, this alone being no guarantee of a good solution, however. A further point to report on is that very little difference exists between M_x for the 20th cycle and M_x for the 40th cycle, a fact also reflected in χ''/χ_0 not changing noticeably in going from 20 to 40 rf cycles. This suggests that the error transients damp out considerably faster than at the minimum rate $1/T_1$, i.e., they appear to have dissipated themselves before the 20th rf cycle was reached.

TABLE II. χ''/χ_0 values for asphaltene. $H_1 = 1$, $H_0 = \frac{1}{2}$ G.

Frequency	χ''/χ_0 [Low-harmonic theory, Eq. (44)]	χ''/χ_0 (Numerical 40 rf cycles)	S	S_x	S_y	S_z
1 Mc/sec	3.0465×10^{-2}	2.9213×10^{-2}	82	58	516	185
10 Mc/sec	3.3933×10^{-3}	3.5712×10^{-3}	148	110	182	302
20 Mc/sec	1.7285×10^{-3}	1.7889×10^{-3}	145	119	89	212

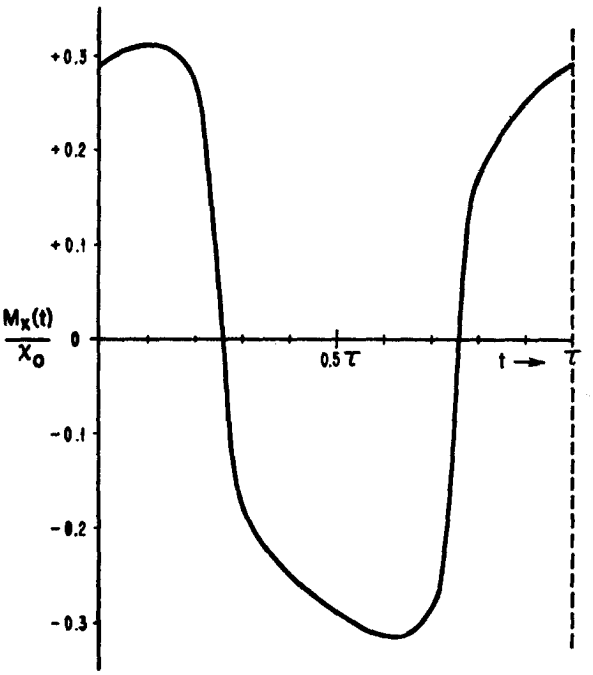


FIG. 2. Asphaltene, M_x/χ_0 (numerical), 40th and 20th rf cycles. The frequency is 1 Mc/sec, $H_1 = 1$G, $H_0 = \frac{1}{2}$ G, $T_1 = 5 \times 10^{-6}$ sec, $T_2 = 10^{-8}$ sec.

ACKNOWLEDGMENTS

The author would like to thank D. D. Thompson, J. Korringa, and H. C. Torrey for many interesting and profitable discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Selection Rules for Integrals of Bloch Functions

J. ZAK
Department of Physics, Israel Institute of Technology, Haifa, Israel
(Received June 11, 1962)

A simple selection rule for integrals of three Bloch functions is derived by using the irreducible representations of the complete space group. In the final formula only the characters of the small representations appear.

INTRODUCTION

We shall deal in this note with selection rules for integrals of the form

$$\int \psi_k \psi_{k'} \psi^*_{k''} \, dV, \tag{1}$$

where ψ_k, $\psi_{k'}$, $\psi_{k''}$ are Bloch functions corresponding to the vectors $\mathbf{k}$, $\mathbf{k}'$, and $\mathbf{k}''$. Two papers dealing with this question have already been published.[1,2] In the first, the representations of the factor group G_k/T_k are used. The selection rule requires, therefore, the knowledge of all irreducible representations of the factor group. In the second paper only the small representations are used, but the selection rules are still rather complicated.

We can look at a function ψ_k in (1) in two ways: first, as a function of the basis for a small representation of the group of $\mathbf{k}$; secondly, as a function of the basis for an irreducible representation of the whole space group. Methods for obtaining the small representations[3,4] and the structure of the representations of the whole space group are well known.[3,5]

We will show in this note that by using the representations of the whole space group, a simple selection rule for integrals of the form (1) can be derived.

DERIVATION OF THE SELECTION RULE

Let us look at the functions ψ_k, $\psi_{k'}$, $\psi_{k''}$ in (1) as belonging to bases for irreducible representations Γ_k, $\Gamma_{k'}$, and $\Gamma_{k''}$ of the whole space group, and let χ_k, $\chi_{k'}$, and $\chi_{k''}$ be the characters of these representations. The integral (1) equals zero if the direct product,

$$\Gamma_k \times \Gamma_{k'} \times \Gamma^*_{k''}, \tag{2}$$

does not contain the identity representation. The number of times the identity representation appears in the direct product (2) is given by

$$\frac{1}{gh} \sum_G \chi_k \chi_{k'} \chi^*_{k''}, \tag{3}$$

where gh is the order of the space group (h—the order of the invariant subgroup of translations H; g—the order of the factor group of G with H as a subgroup); the summation is over the space group.

The character χ_k can be written by means of the characters ξ_k of the small representations γ_k of the group of $\mathbf{k}$ as follows:

Let $\psi_{1k}, \psi_{2k}, \cdots, \psi_{rk}$ be the basis of the small representation γ_k of the group of the vector $\mathbf{k}$:

$$(\alpha_i \mid \mathbf{a}_i)\psi_{mk} = \sum_q \{A(\alpha_i \mid \mathbf{a}_i)\}_{mq} \psi_{qk} \cdots, \tag{4}$$

where $(\alpha_i \mid \mathbf{a}_i)$ are the elements of the group of $\mathbf{k}$ and $A(\alpha_i \mid \mathbf{a}_i)$ the matrices of the representation γ_k. Assign the group of $\mathbf{k}$ by K, then the whole space group G can be written:

$$G = K + (\beta_2 \mid \mathbf{b}_2)K + \cdots + (\beta_f \mid \mathbf{b}_f)K. \tag{5}$$

The functions

$$\psi_{1k}, \cdots, \psi_{rk}, \quad (\beta_2 \mid \mathbf{b}_2)\psi_{1k}, \cdots, \quad (\beta_2 \mid \mathbf{b}_2)\psi_{rk}, \cdots, \quad (\beta_f \mid \mathbf{b}_f)\psi_{rk}, \cdots, \tag{6}$$

form a basis for an irreducible representation of the space group G.[3] This representation is known as one formed by the star of $\mathbf{k}$. Moreover, every set of functions in (6), for example, $(\beta_i \mid \mathbf{b}_i)\psi_{1k}, \cdots, (\beta_i \mid \mathbf{b}_i)\psi_{rk}$ form a basis for a small representation, in this case for the group of the vector $\beta_i\mathbf{k} \doteq \mathbf{k}_i$ ($\doteq$ means equal or equivalent).[2] The elements of the group of $\mathbf{k}_i$ are

$$(\beta_i \mid \mathbf{b}_i)(\alpha_j \mid \mathbf{a}_j)(\beta_i \mid \mathbf{b}_i)^{-1}, \tag{7}$$

where $(\alpha_j \mid \mathbf{a}_j)$ are all the elements of the group of $\mathbf{k}$. It is easy to show that the character

[1] R. J. Elliott and R. Loudon, J. Phys. Chem. Solids **15**, 146 (1960).
[2] M. Lax and J. J. Hopfield, Phys. Rev. **124**, 115 (1961).
[3] G. F. Koster, *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic Press Inc., New York, 1957), Vol. 5, p. 173.
[4] J. Zak, J. Math. Phys. **1**, 165 (1960).
[5] J. Zak, thesis for degree of Doctor of Science, Haifa, Israel (unpublished).

$\xi_{k_i}[(\beta_i \mid \mathbf{b}_i)(\alpha_j \mid \mathbf{a}_j)(\beta_i \mid \mathbf{b}_i)^{-1}]$ of the small representation of the group of $\mathbf{k}_i$ equals the character $\xi_k(\alpha_j \mid \mathbf{a}_j)$ of the group of $\mathbf{k}$[3]:

$$\xi_{k_i}[(\beta_i \mid \mathbf{b}_i)(\alpha_j \mid \mathbf{a}_j)(\beta_i \mid \mathbf{b}_i)^{-1}] = \xi_k(\alpha_j \mid \mathbf{a}_j). \tag{8}$$

By using the form of the basis (6) one can find the character χ_k of the irreducible representation of the whole space group G. Let $(\delta \mid \mathbf{d})$ be any element of G. To the character of this element $\chi_k(\delta \mid \mathbf{d})$ only those functions of (6) will contribute, the $\mathbf{k}$'s of which satisfy the condition $\delta\,\mathbf{k} \doteq \mathbf{k}$. Hence,

$$\chi_k(\delta \mid \mathbf{d}) = \sum_i \xi_{k_i}(\delta \mid \mathbf{d}). \tag{9}$$

The sum is over all the $\mathbf{k}_i$'s in the star of $\mathbf{k}$, satisfying the condition $\delta\,\mathbf{k}_i \doteq \mathbf{k}_i$. $\xi_{k_i}(\delta \mid \mathbf{d})$ is the character corresponding to the element $(\delta \mid \mathbf{d})$ in the small representation of the group of $\mathbf{k}_i$ in the basis

$$(\beta_i \mid \mathbf{b}_i)\psi_{1k}, \cdots, \qquad (\beta_i \mid \mathbf{b}_i)\psi_{rk}, \qquad (\beta_i\mathbf{k} \doteq k_i).$$

We have to calculate the sum (3). By using (9) we have

$$\frac{1}{gh}\sum_G \chi_k\chi_{k'}\chi^*_{k''} = \frac{1}{gh}\sum_G \sum_{ij1} \xi_{k_i}(\delta \mid \mathbf{d})\xi_{k_j'}(\delta \mid \mathbf{d})\xi^*_{k_1''}(\delta \mid \mathbf{d}).$$

Let us first take the sum over all pure translations. Then, since $\xi_{k_i}(\delta \mid \mathbf{d} + \mathbf{t}) = e^{i\mathbf{k}_i\cdot\mathbf{t}}\xi_{k_i}(\delta \mid \mathbf{d})$ we are left with a summation over representative elements $(\rho_m \mid \mathbf{r}_m)$ from every coset of the factor group G/H:

$$\frac{1}{g}\sum_{(\rho_m \mid \mathbf{r}_m)} \sum_{ij1} \xi_{k_i}(\rho_m \mid \mathbf{r}_m)\xi_{k_j'}(\rho_m \mid \mathbf{r}_m)\xi^*_{k_1''}(\rho_m \mid \mathbf{r}_m). \tag{10}$$

The sum here is over all the $\mathbf{k}_i$, $\mathbf{k}_j'$, and $\mathbf{k}_1''$ in the stars of $\mathbf{k}$, $\mathbf{k}'$, and $\mathbf{k}''$ satisfying the condition

$$\mathbf{k}_i + \mathbf{k}_j' - \mathbf{k}_1'' \doteq 0. \tag{11}$$

and over the elements $(\rho_m \mid \mathbf{r}_m)$ for which the equations

$$\rho_m\mathbf{k}_i \doteq \mathbf{k}_i, \qquad \rho_m\mathbf{k}_j' \doteq \mathbf{k}_j', \qquad \rho_m\mathbf{k}_1'' \doteq \mathbf{k}_1'' \tag{12}$$

hold.

The sum (10) becomes *very simple* if the summation is carried out as follows: Let us take a triad $\mathbf{k}_i$, $\mathbf{k}_j'$, and $\mathbf{k}_1''$ satisfying the condition (11) and sum up in (10) over the elements $(\rho_m \mid \mathbf{r}_m)$ which satisfy the condition (12). Assign these elements by

$$(\rho_s \mid \mathbf{r}_s), \qquad s = 1, 2, \cdots, p. \tag{13}$$

Then take an element $(\rho_m \mid \mathbf{r}_m)$ which does not belong to (13) and form a new triad of $\mathbf{k}$'s:

$$\rho_m\mathbf{k}_i, \qquad \rho_m\mathbf{k}_j', \qquad \rho_m\mathbf{k}_1''. \tag{14}$$

Now we have to sum up in (10) over the elements

$$(\rho_m \mid \mathbf{r}_m)(\rho_s \mid \mathbf{r}_s)(\rho_m \mid \mathbf{r}_m)^{-1}, \qquad s = 1, 2, \cdots, p. \tag{15}$$

because if ρ_s satisfies the condition (12) for the triad (11), then $\rho_m\rho_s\rho_m^{-1}$ will satisfy the condition (12) for the triad (14) [since the condition (12) is fulfilled no matter which element of the given coset of the factor group G/H is taken in the summation (10)]. According to (8) the sum over the elements (15) will be equal to the sum over the elements (13). Continuing the summation in (10) in such a way we will obtain g/p equal sums where p is the number of elements in (13). The final form of (10) will therefore be:

$$\frac{1}{p}\sum_s \xi_{k_i}(\rho_s \mid \mathbf{r}_s)\xi_{k_j'}(\rho_s \mid \mathbf{r}_s)\xi^*_{k_1''}(\rho_s \mid \mathbf{r}_s). \tag{16}$$

The last formula is very simple. In order to establish the selection rule for an integral of the form (1) we take a triad $\mathbf{k}_i$, $\mathbf{k}_j'$, and $\mathbf{k}_1''$ [$\mathbf{k}_i$ from the star of $\mathbf{k}$, $\mathbf{k}_j'$ from the star of $\mathbf{k}'$, and $\mathbf{k}_1''$ from the star of $\mathbf{k}''$] satisfying (11) [if no such triad exists, then the integral (1) is zero] and sum up over those elements $(\rho_m \mid \mathbf{r}_m)$ of the factor group G/H (one element from every coset), which satisfy the condition (12).

An Orthogonality Property of Hydrogenlike Radial Functions*

S. PASTERNACK AND R. M. STERNHEIMER
Brookhaven National Laboratory, Upton, New York
(Received June 11, 1962)

An integral involving r^{-s} times the product of two hydrogenlike radial functions of the same n and different l is shown to be zero for a number of values of s.

THERE is an interesting orthogonality property of the hydrogenlike radial functions $R_{nl}(r)$ which, to the best of our knowledge, has not been recorded before.[1] Namely, for $l > l'$,

$$K_{nll',s} \equiv \int_0^\infty \frac{1}{r^s} R_{nl}(r)R_{nl'}(r)r^2\,dr = 0$$

$$\text{if} \quad s = 2, 3, \cdots, l - l' + 1.$$

The proof is quite simple when one makes use of the well-known generating function for the Laguerre polynomials.[2] Proceeding in the same manner as when one obtains the normalizing factor N_{nl} for the radial functions,[2] the value of the integral is expressible as $N_{nl}N_{nl'}(n + l)!(n + l')!$ times the coefficient of $u^{n-l-1}v^{n-l'-1}$ in

$$\int_0^\infty \frac{e^{-\rho u/(1-u)}}{(1-u)^{2l+2}} \frac{e^{-\rho v/(1-v)}}{(1-v)^{2l'+2}} e^{-\rho}\rho^{-s+l+l'+2}\,d\rho$$

$$= (l + l' - s + 2)!\,(1 - u)^{-s-l+l'+1} \times (1 - v)^{-s+l-l'+1}(1 - uv)^{+s-l-l'-3}.$$

The power of v that we seek exceeds the power of u by $(l - l')$. It is clear on inspection that we will not have an infinite series of powers of v in excess of those of u if $s \leq l - l' + 1$; and indeed, if the power of $(1 - v)$ is thus nonnegative, the maximum power of v in excess of that of u is $l - l' + 1 - s$. If this is less than $l - l'$ (i.e., if $s \geq 2$), then the desired coefficient must be zero. Thus the theorem is proved.

The procedure outlined above leads directly to the following expression for the nonzero values of the integral:

$$K_{nll',s} = (-1)^{l+l'}(2Z/n)^s \times \left[\frac{(n - l - 1)!\,(n - l' - 1)!}{(n + l)!\,(n + l')!}\right]^{1/2} \times \frac{(l + l' - s + 2)!}{2n} \sum_\tau (-1)^\tau \binom{l' - l - s + 1}{n - l - 1 - \tau} \times \binom{l - l' - s + 1}{n - l' - 1 - \tau}\binom{-l - l' + s - 3}{\tau},$$

where $\binom{\alpha}{\beta}$ is the usual binomial coefficient. For $l = l'$ (i.e., the mean value of r^{-s}), some more convenient expressions are available.[3]

These integrals occurred in connection with a problem concerning the hyperfine structure due to a nuclear quadrupole moment.[4]

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] For $s = 2$, the property has been proved by G. Feinberg, Phys. Rev. **112**, 1637 (1958); see p. 1641.

[2] See, for example, L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1935), p. 451.

[3] S. Pasternack, Proc. Natl. Acad. Sci. (U. S.) **23**, 91 (1937). In Eq. (4) of that paper, the factor $(2l-q)!/(2l+q+1)!$ should be inverted.

[4] R. M. Sternheimer, Phys. Rev. **105**, 158 (1957).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Green's Distributions and the Cauchy Problem for the Multi-Mass Klein-Gordon Operator

JOHN JUDSON BOWMAN
Department of Physics, University of Michigan, Ann Arbor, Michigan
AND
JOSEPH DAVID HARRIS
Departments of Physics and Biochemistry, Dartmouth University, Hanover, New Hampshire
(Received April 9, 1962)

Explicit forms of the Green's functions (which are to be regarded as distributions in the sense of Schwartz) for the multi-mass Klein-Gordon operator in n-dimensional spaces are presented. The homogeneous Green's functions $G_N(x)$ and $G_N{}^1(x)$, defined in the usual way by independent paths of integration in the k_0 plane, are investigated in the neighborhood of the light cone. The parameter N indicates the total number of masses involved. The singularities on the light cone reflect the well-known difference between even- and odd-dimensional wave propagation. It is found that $G_N(x;$ odd $n)$ contains a finite jump on the light cone as well as a linear combination of derivatives up to order $\frac{1}{2}(n - 2N - 1)$ of $\delta(x^2)$; the singular part of $G_N{}^1(x;$ odd $n)$ consists of a logarithmic singularity $\ln(|x^2|)$ along with a polynomial in $(x^2)^{-1}$ of degree $\frac{1}{2}(n - 2N - 1)$. For even-dimensional spaces, the singular part of both Green's functions consists of a polynomial in $(x^2)^{-1/2}$ of degree $n - 2N + 1$ vanishing *outside* the light cone for G_N and vanishing *inside* the light cone for $G_N{}^1$. In all cases no singularities or finite jumps occur when the order $2N$ of the operator is greater than the number $n + 1$ of space-time dimensions. The general solution of the Cauchy problem is given both for the data carrying surface $t = 0$ and for arbitrary spacelike data surfaces.

1. INTRODUCTION

MULTI-MASS equations have enjoyed a long history in field theory and often arise from attempts to eliminate the typical divergences occurring within the theory.[1] For example, the regularized propagators of Pauli and Villars,[2] obtained with the help of an introduction of discrete auxiliary masses, may easily be seen to satisfy equations of the form

$$F(\Box)\varphi(x) = \rho(x), \tag{1}$$

where

$$F(\Box) = \prod_i (\Box + \mu_i^2). \tag{2}$$

In the neighborhood of the light cone, no singularities or finite jump discontinuities appear in the propagators provided the total number of masses is at least three.

Equations of the type given in (1) have been investigated in detail by Pais and Uhlenbeck.[3] As these authors have shown, if $F(\Box)$ is a polynomial in $\Box$ with arbitrary constant coefficients, then (1) is an equation of hyperbolic type and a well-defined initial value of Cauchy problem exists for the homogeneous equation. The case of polynomials in $\Box$ has also been treated by Rzewuski,[4] who gives an invariant form of the general solution of the homogeneous equation, but leaves the corresponding Green's functions implicit in their Fourier representations.

The present paper is devoted to the Green's functions and the Cauchy problem for the multi-mass Klein-Gordon operator in multi-dimensional spaces. To fix the notation, we shall seek a solution of the Cauchy problem for the equation

$$(\Box + \mu_1^2)^{\lambda_1} \cdots (\Box + \mu_L^2)^{\lambda_L}\varphi(x) = 0, \tag{3}$$

where the d'Alembert operator is given by

$$\Box = \partial_0^2 - \partial_1^2 - \cdots - \partial_n^2. \tag{4}$$

The sum of the nonnegative integers $\lambda_1 \cdots \lambda_L$ will be denoted

$$N \equiv \lambda_1 + \cdots + \lambda_L \tag{5}$$

where $N = 1, 2, \cdots$ gives the total number of rest masses involved. The rest masses $\mu_1, \cdots, \mu_L$, will be taken to be distinct, real and positive (although many of the results can be extended to complex μ_ν); if one of them must be zero, the limit $\mu \to 0$ may be taken after the calculations are made. Clearly, the wave equation is of order $2N$; hence, the Cauchy data on the spacelike plane $t = 0$ consist of the time

[1] Multi-mass equations also arise naturally when one considers particles of higher spin; see, e. g., J. D. Harris, Phys. Rev. **112**, 2124 (1958).
[2] W. Pauli and F. Villars, Revs. Modern Phys. **21**, 434 (1949).
[3] A. Pais and G. E. Uhlenbeck, Phys. Rev. **79**, 145 (1950).
[4] J. Rzewuski, Acta Phys. Polon. **12**, 100 (1953).

derivatives $\partial_0^M \varphi(\mathbf{x}, 0)$ with $M = 0, 1, \cdots, 2N - 1$.

In a previous paper[5] we obtained the explicit form of the homogeneous Green's function $\Delta(x; m; \mu)$ associated with the multi-dimensional, iterated Klein-Gordon operator $(\Box + \mu^2)^m$. The Fourier representation of the Green's function was expressed, after some angular integrations, as a one-dimensional, infinite integral of the Sonine type. It was shown that although this integral is classically divergent when the order of the operator is less than the number of space dimensions, it can be treated rigorously under these conditions using the concepts of distribution analysis. The Green's function is then to be regarded as a (tempered) distribution in the sense of Schwartz.[6] A new distribution introduced for the purpose of giving the improper Sonine integral its generalized meaning was used to investigate the singularities of the Green's function on the light cone.

These results may be easily extended to include the Green's functions associated with the multi-mass operator $(\Box + \mu_1^2)^{\lambda_1} \cdots (\Box + \mu_L^2)^{\lambda_L}$ either by a partial fraction decomposition of the integrand of the Fourier representations,[7] or by the method given in Sec. 2. Explicit expressions for the complete set of Green's functions for the multi-mass operator will be presented; this includes an expression for $\Delta^1(x; m; \mu)$ which has not been previously given in (I). (The two distributions Δ and Δ^1 are linearly independent.)

The behavior of the Green's functions in the neighborhood of the light cone will be explicitly investigated in Sec. 3. Because of the well-known difference between wave propagation in spaces with an even, and spaces with an odd, number of dimensions, the two cases must be treated separately. Some very interesting results are obtained. For *odd-dimensional spaces*, the Green's function $G_N(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$, intimately related to $\Delta(x; m; \mu)$, contains a finite linear combination of derivatives of the Dirac delta function $\delta(x^2)$ as well as a finite jump discontinuity on the light cone. The highest derivative appearing is of order $\frac{1}{2}(n-2N-1)$. The singular part of $G_N^1(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$, closely connected with $\Delta^1(x; m; \mu)$, consists of a polynomial in $(1/x^2)$ of degree $\frac{1}{2}(n - 2N - 1)$ along with a logarithmic singularity ln $(|x^2|)$. On the other hand, for even-dimensional spaces, the singular part of both Green's functions consists essentially of a polynomial in $1/(x^2)^{1/2}$ of degree $(n - 2N + 1)$; but the polynomial vanishes *outside* the light cone for G_N, and vanishes *inside* the light cone for G_N^1. In all cases, no singularities or finite jumps appear when the order of the operator is greater than the number of space-time dimensions (i.e., $2N > n + 1$). When this is true, the first $N - (n + 3)/2$ derivatives with respect to x^2 when n is odd, or first $N - 1 - n/2$ derivatives when n is even, of both $G_N(x)$ and $G_N^1(x)$, are also continuous on the light cone.

The remainder of the paper is concerned with obtaining the general solution of Cauchy's problem for the homogeneous equation (3). In this enterprise, the set of Green's functions $G_\eta(x; r_1 \cdots r_L; \mu_1 \cdots \mu_1)$ with $0 \leq r_p \leq \lambda_p$, $\eta = r_1 + \cdots + r_L$, and $1 \leq \eta \leq N$ is particularly useful. The initial conditions satisfied by these functions are given in Sec. 4 and the general solution $\varphi(x)$ is obtained in Sec. 5. Finally, an invariant form of the general solution is presented in Sec. 6 and agrees with that given by Rzewuski.[4] There are two Appendices.

2. THE GREEN'S FUNCTIONS

All of the Green's functions for the multi-mass operator $(\Box + \mu_1^2)^{\lambda_1} \cdots (\Box + \mu_L^2)^{\lambda_L}$ may be obtained using the general Fourier representation

$$K(x) = (2\pi)^{-n-1} \times \int dk e^{-ikx} (\mu_1^2 - k^2)^{-\lambda_1} \cdots (\mu_L^2 - k^2)^{-\lambda_L}. \quad (6)$$

As is well known, such an expression is not completely defined until the path of integration around the poles of the integrand has been specified. In the k_0 plane, open paths of integration that coincide with the real k_0 axis at $\pm\infty$ give rise to inhomogeneous Green's functions. Closed paths which encircle one or more of the poles lead to homogeneous Green's functions.

There are only $2L$ independent ways of encircling the $2L$ poles of the integrand. However, by reducing the multiplicity of the poles, $2N - 2L$ more independent solutions of the homogeneous equation may be obtained, giving a total of $2N$ independent homogeneous Green's functions. The general inhomogeneous Green's function may be written in the form of a sum of the particular solution of the equation

$$(\Box + \mu_L^2)^{\lambda_L} \cdots (\Box + \mu_L^2)^{\lambda_L} K(x) = \delta(x) \quad (7)$$

[5] J. J. Bowman and J. D. Harris, J. Math. Phys. 3, 396 (1962), hereafter called (I).

[6] L. Schwartz, *Théorie des distributions I, II* (Hermann et Cie, Paris, 1950–51).

[7] Green's functions for multi-mass operators like $[\Box^m - (-\mu^2)^m]^l$ may be calculated directly without recourse to a partial fraction expansion. Such operators have been investigated by J. J. Bowman and J. D. Harris, J. Math. Phys. 3, 1291 (1962).

and of the $2N$ independent solutions of the homogeneous equation with arbitrary coefficients. The multi-mass operator thus has a total of $2N + 1$ independent Green's functions.

The homogeneous Green's function $G_N(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$ is defined by choosing the path of integration in the k_0 plane to consist of a closed curve C encircling *all* of the poles in a clockwise fashion. The notation $G_N(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$ will often be abbreviated to $G_N(x)$ or simply G_N. The inhomogeneous Green's function $\bar{G}_N(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$, defined by taking the principal part of the k_0 integration over the singularities, is related to the homogeneous Green's function $G_N(x)$ by the formula

$$G_N(x) = 2\epsilon(x)\bar{G}_N(x), \tag{8}$$

where $\epsilon(x) = \pm 1$ for $t \gtrless 0$. A second homogeneous Green's function $G_N^1(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$ is characterized by a closed path of integration C^1 which encircles all of the poles on the positive real k_0 axis in a clockwise fashion and all of the poles on the negative real k_0 axis in a counterclockwise direction. Clearly the paths C^1 and C are not equivalent, hence G_N^1 and G_N are linearly independent distributions. We will give explicit expressions for $G_N(x)$ and $G_N^1(x)$.

In what follows, the notation $G_h(x)$ will be used to denote any homogeneous Green's function defined by a path of integration $\mathcal{C}$ which consists of a linear combination of the paths C and C^1. Using the theory of residues and the properties of Dirac δ functions, one easily obtains the identity

$$\Gamma(\lambda) \int_{\mathcal{C}} dk(\mu^2 - k^2)^{-\lambda} = 2\pi i \int_{-\infty}^{\infty} dk f(k)\, \delta^{(\lambda-1)}(k^2 - \mu^2), \tag{9}$$

where $f(k)$ is determined by the path $\mathcal{C}$; in particular,

$$f(k) = \begin{cases} \epsilon(k) & \text{for } \mathcal{C} = C \\ 1 & \text{for } \mathcal{C} = C^1. \end{cases} \tag{10}$$

Making use of (9), and taking into account the contribution of all the poles in (6), one may write $G_h(x)$ in the form

$$G_h(x) = \sum_{p=1}^{L} \frac{i(2\pi)^{-n}}{\Gamma(\lambda_p)} \int_{-\infty}^{\infty} dk e^{-ikx} f(k) \times \delta^{(\lambda_p-1)}(k^2 - \mu_p^2) \prod_{j=1}^{L}{}' (\mu_j^2 - k^2)^{-\lambda_j}, \tag{11}$$

where the prime on the product means that the factor for $j = p$ is to be omitted. Each summand can be expressed in terms of the operator

$$\frac{1}{\Gamma(\lambda_p)} \left(-\frac{d}{d\mu_p^2}\right)^{\lambda_p - 1}$$

acting on

$$i(2\pi)^{-n} \int_{-\infty}^{\infty} dk e^{-ikx} f(k)\, \delta(k^2 - \mu_p^2) \prod_{j=1}^{L}{}' (\mu_j^2 - k^2)^{-\lambda_j} = \Delta_h(x; \mu_p)\Phi(-\mu_p^2), \tag{12}$$

with Δ_h representing the corresponding Green's function for the Klein-Gordon operator $\Box + \mu^2$, and

$$\Phi(-\mu_p^2) \equiv \prod_{j=1}^{L}{}' (\mu_j^2 - \mu_p^2)^{-\lambda_j}. \tag{13}$$

Finally we obtain

$$G_h(x) = \sum_{p=1}^{L} \sum_{m=1}^{\lambda_p} \frac{\Phi^{(\lambda_p - m)}(-\mu_p^2)}{(\lambda_p - m)!} \Delta_h(x; m; \mu_p) \tag{14}$$

by applying the Leibniz rule for differentiating a produce of two functions and using the result

$$\frac{1}{\Gamma(m)} \left(-\frac{d}{d\mu^2}\right)^{m-1} \Delta_h(x; \mu) = \Delta_h(x; m; \mu), \tag{15}$$

where $\Delta_h(x; m; \mu)$ is the homogeneous Green's function for the iterated Klein-Gordon operator $(\Box + \mu^2)^m$. We note that (15) is easily obtained from the Fourier representation

$$\Delta_h(x; m; \mu) = (2\pi)^{-n-1} \int_{\mathcal{C}} dk e^{-ikx} (\mu^2 - k^2)^{-m}, \tag{16}$$

although a rigorous justification of the identity follows only from considerations of distribution theory [cf. (I)].

The result expressed in (14) may alternatively be obtained using the general partial fraction expansion (Appendix A)

$$(\mu_1^2 - k^2)^{-\lambda_1} \cdots (\mu_L^2 - k^2)^{-\lambda_L} = \sum_{p=1}^{L} \sum_{m=1}^{\lambda_p} \frac{\Phi^{(\lambda_p - m)}(-\mu_p^2)}{(\lambda_p - m)!} (\mu_p^2 - k^2)^{-m} \tag{17}$$

for the integrand of (6). Equation (14) follows immediately.

Equation (14) yields an expansion of $G_N(x)$ in terms of the N independent Δ solutions of the homogeneous equation (3); similarly, an expansion of $G_N^1(x)$ in terms of the N independent Δ^1 solutions is obtained. Clearly the distributions $G_\eta(x; r_1 \cdots r_L; \mu_1 \cdots \mu_L)$ and $G_\eta^1(x; r_1 \cdots r_1; \mu_1 \cdots \mu_L)$ with $0 \leq r_p \leq \lambda_p$, $\eta = r_1 + \cdots = r_L$, and $1 \leq \eta \leq N$ are also solutions of the homogeneous equation. There is a large degeneracy with respect to the parameter η and not all of these solutions are

linearly independent. Out of the combined set $\{G_\nu, G_\nu^1\}$ one may choose $2N$ independent homogeneous solutions; in particular, the functions $\Delta(x; m; \mu_\rho)$ and $\Delta^1(x; m; \mu_\rho)$ form a complete set.

Explicit forms of the distributions Δ and Δ^1 may be calculated (see Appendix B) using the method set forth in (I); the results are

$$\Delta(x; m; \mu) = \epsilon(x) \operatorname{Re} \mathbf{K}(x; m; \mu), \tag{18}$$

$$\Delta^1(x; m; \mu) = i \operatorname{Im} \mathbf{K}(x; m; \mu), \tag{19}$$

where

$$\mathbf{K}(x; m; \mu) = \frac{(2\pi)^{(1-n)/2}}{\Gamma(m)2^m} \left(\frac{\sqrt{x^2}}{\mu}\right)^{m-(n+1)/2} \times H^{(2)}_{m-(n+1)/2}(\mu\sqrt{x^2}). \tag{20}$$

Here $H_\nu^{(2)}$ is the Hankel function of the second kind.[8] We stress the fact that $\Delta(x; m; \mu)$ and $\Delta^1(x; m; \mu)$ are in general to be considered as distributions. The $\Delta(x; m; \mu)$ vanish outside the light cone, whereas the $\Delta^1(x; m; \mu)$ do not.

The general homogeneous Green's function for the multi-mass operator $(\Box + \mu_1^2)^{\lambda_1} \cdots (\Box + \mu_L^2)^{\lambda_L}$ is now obviously

$$G_{\text{hom}}(x) = \sum_{\rho=1}^{L} \sum_{m=1}^{\lambda_\rho} [a_{\rho m} \Delta(x; m; \mu_\rho) + b_{\rho m} \Delta^1(x; m; \mu_\rho)], \tag{21}$$

where $a_{\rho m}$ and $b_{\rho m}$ are some constants which may involve the μ_ρ. The general inhomogeneous Green's function for that operator is therefore

$$G_{\text{inhom}}(x) = \tilde{G}_N(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L) + G_{\text{hom}}(x), \tag{22}$$

where $G_{\text{hom}}(x)$ is a linear combination of the N functions $\Delta(x; m; \mu_\rho)$ vanishing outside, and the N functions $\Delta^1(x; m; \mu_\rho)$ not vanishing outside the light cone.

3. SINGULARITIES OF THE GREEN'S FUNCTIONS

Using the results of the preceding section, we may write the distributions $G_N(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$ and $G_N^1(x; \lambda_1 \cdots \lambda_L; \mu_1 \cdots \mu_L)$ in the following manner:

$$\{G_N, G_N^1\} = \sum_{\rho=1}^{L} \frac{1}{\Gamma(\lambda_\rho)} \left(\frac{-d}{d\mu_\rho^2}\right)^{\lambda_\rho - 1} \Phi(-\mu_\rho^2) \times \{\Delta(x; \mu_\rho), \Delta^1(x; \mu_\rho)\}, \tag{23}$$

where the Green's functions for the Klein-Gordon operator have the explicit forms

$$\Delta(x; \mu) = \frac{\epsilon(x)\theta(x^2)}{2} \left(\frac{\mu}{2\pi}\right)^{(n-1)/2} (\sqrt{x^2})^{(1-n)/2} \times J_{(1-n)/2}(\mu\sqrt{x^2}), \tag{24}$$

$$\Delta^1(x; \mu) = -i \frac{\theta(x^2)}{2} \left(\frac{\mu}{2\pi}\right)^{(n-1)/2} (\sqrt{x^2})^{(1-n)/2} \times Y_{(1-n)/2}(\mu\sqrt{x^2}) + i \frac{\theta(-x^2)}{2} \frac{2}{\pi} \left(\frac{\mu}{2\pi}\right)^{(n-1)/2} (\sqrt{-x^2})^{(1-n)/2} \times K_{(1-n)/2}(\mu\sqrt{-x^2}). \tag{25}$$

In the above, J_ν is the Bessel function of the first kind, Y_ν the Neumann function, K_ν the modified Bessel function of the third kind, and $\theta(y) = 1$ for $y > 0$, $\theta(y) = 0$ for $y < 0$.

The cases of even- and odd-dimensional spaces must be considered separately in order to investigate the nature of the singularities near the light cone. The singular part of $\Delta(x; \mu)$ has already been determined in (I); for small x^2 we find

$$\Delta(x; \mu) \simeq \frac{\epsilon(x)}{2} \left(\frac{-\mu^2}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(n-1)/2} \frac{(-\mu^2/4)^{-m}\theta^{(m)}(x^2)}{\Gamma(n/2 - m + \frac{1}{2})}\right\} \quad (n \text{ odd}) \tag{26}$$

$$\Delta(x; \mu) \simeq \frac{\epsilon(x)\theta(x^2)}{2} \left(\frac{-\mu^2}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(1/2)n-1} \frac{(-\mu^2/4)^{-m-1/2}(x^2)^{-m-1/2}}{\Gamma(n/2 - m)\Gamma(\frac{1}{2} - m)}\right\} \quad (n \text{ even}) \tag{27}$$

where the terms that vanish for $x^2 \to 0$ have been omitted. Since $\theta'(x^2) = \delta(x^2)$, one finds that $\Delta(x; \mu$: odd $n)$ contains δ-function singularities as well as a finite jump discontinuity on the light cone; the singular part of $\Delta(x; \mu$: even $n)$ consists of a polynomial in $(x^2)^{-1/2}$ vanishing outside the light cone, and no finite discontinuity is present.

The singular part of $\Delta^1(x; \mu)$ may easily be obtained using well-known formulas[8] of the Bessel functions. In the neighborhood of the light cone, one finds, for odd dimensional spaces

$$\Delta^1(x; \mu) \simeq -\frac{i}{\pi} \left(\frac{-\mu^2}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(n-1)/2} \frac{(-\mu^2/4)^{-m}}{\Gamma(n/2 - m + \frac{1}{2})} \left(\frac{d}{dx^2}\right)^m \ln(|x^2|)^{1/2}\right\} + \frac{i}{\pi} \left(\frac{-\mu^2}{4\pi}\right)^{(n-1)/2} \frac{h(n; \mu)}{\Gamma(n/2 + \frac{1}{2})} \quad (n \text{ odd}) \tag{28}$$

where

$$h(n; \mu) \equiv \tfrac{1}{2}[1 + \tfrac{1}{2} + \tfrac{1}{3} + \cdots + 2/(n-1)] - \ln(\gamma\mu/2), \tag{29}$$

with $\ln \gamma$ $(=0.5772 \cdots)$ representing the Euler-

[8] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, New York, 1944), 2nd ed.

Mascheroni constant. For even-dimensional spaces we have

$$\Delta^1(x;\mu) \simeq \frac{\theta(-x^2)}{2}\left(\frac{-\mu^2}{4\pi}\right)^{(1/2)(n-1)} \times \left\{\sum_{m=0}^{(1/2)n-1} \frac{(-\mu^2/4)^{-m-1/2}(x^2)^{-m-1/2}}{\Gamma(n/2-m)\Gamma(\frac{1}{2}-m)}\right\} - \tfrac{1}{2}\left(\frac{-\mu^2}{4\pi}\right)^{(1/2)(n-1)} \frac{1}{\Gamma(n/2+\frac{1}{2})} \quad (n \text{ even}), \tag{30}$$

where the factor of i is contained implicitly. In both (28) and (30), the terms that vanish for $x^2 \to 0$ have been omitted. Again, the remarkable difference between even- and odd-dimensional spaces is reflected in the nature of the singularities on the light cone: the singular part of $\Delta^1(x;\mu$: even $n)$ consists of a polynomial in $(x^2)^{-1/2}$ vanishing *inside* the light cone, whereas $\Delta^1(x;\mu$: odd $n)$ contains logarithmic and multiple pole singularities.

When one substitutes (24) and (25) into (23) to get explicit expressions for $G_N(x)$ and $G_N^1(x)$, one finds that sums of the form

$$S(l) = \sum_{p=1}^{L} \frac{1}{\Gamma(\lambda_p)}\left(\frac{-d}{d\mu_p^2}\right)^{\lambda_p-1}(-\mu_p^2)^l\Phi(-\mu_p^2) \tag{31}$$

must be evaluated. Such sums are easily determined explicitly (when l is an integer) using the algebra of partial fractions (see Appendix A); in particular we have

$$S(l) = \begin{cases} 0 & \text{for } l = 0, 1, \cdots, N-2; \\ 1 & \text{for } l = N-1. \end{cases} \tag{32}$$

Carrying out the indicated procedure, one easily finds the following explicit results:

For both even and odd values of n,

$$G_N(x) = \frac{\epsilon(x)}{2\pi^{(n-1)/2}} \sum_{m=0}^{\infty} \frac{S(N+m-1)}{4^{N+m-1}\Gamma(N+m)} \times \mathfrak{H}_{N+(1-n)/2+m}(x^2), \tag{33}$$

where [cf. I, Eqs. (14) to (22)]

$$\mathfrak{H}_\beta(y) = \theta(y)y^{\beta-1}/\Gamma(\beta) \qquad (\beta \neq 0, -1, -2, \cdots), \tag{34a}$$

$$\mathfrak{H}_{-\beta}(y) = \delta^{(\beta)}(y) \qquad (\beta = 0, 1, 2, \cdots). \tag{34b}$$

For even n,

$$G_N^1(x) = \tfrac{1}{2}\left(\frac{1}{\pi}\right)^{(n-1)/2} \times \left\{\theta(-x^2)\sum_{m=0}^{\infty} \frac{S(m)(x^2)^{m+(1-n)/2}}{4^m\Gamma(m+1)\Gamma[(3-n)/2+m]} - \sum_{m=0}^{\infty} \frac{S[m+(n-1)/2](x^2)^m}{4^{m+(n-1)/2}\Gamma(m+1)\Gamma[(n+1)/2+m]}\right\}. \tag{35}$$

For odd n,

$$G_N^1(x) = \frac{i}{\pi}\left(\frac{1}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{\infty} \frac{(x^2)^m}{4^m\Gamma(m+1)\Gamma((n+1)/2+m)} \times \left[S((n-1)/2+m)\left[\frac{1}{2}\left(1+\cdots+\frac{1}{m}\right) + \frac{1}{2}\left(1+\cdots+\frac{1}{(n-1)/2+m}\right) - \ln\frac{\gamma\,|x^2|^{1/2}}{2}\right] - \sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(-\frac{d}{d\mu_p^2}\right)^{\lambda_p-1}(-\mu_p^2)^{(n-1)/2+m}\Phi(-\mu_p^2)\ln\mu_p\right] - \sum_{m=1}^{(n-1)/2}\frac{S((n-1)/2-m)}{4^{-m}\Gamma((n+1)/2-m)}\left(\frac{d}{dx^2}\right)^m \ln|x^2|^{1/2}\right\}. \tag{36}$$

where the sum $1 + \cdots + 1/m$ is understood to be zero when $m = 0$.

From Eq. (32) and the immediately preceding Eqs. (33) to (36), it easily follows that on the light cone $x^2 = 0$,

(i) There are no singularities in $G_N(x)$, and $G_N^1(x)$ contains no multiple pole singularities, when $2N > n$, that is, when the order of the wave equation is greater than the number of space dimensions. However, for odd n and $2N = n + 1$, a finite jump discontinuity occurs in $G_N(x)$ and a logarithmic singularity occurs in $G_N^1(x)$.

(ii) Both $G_N(x)$ and $G_N^1(x)$ are continuous together with their first $N - (n+3)/2$ derivatives if n is odd, or with their first $N - 1 - n/2$ derivatives if n is even, provided that $2N > n + 1$, that is provided the order of the wave equation is greater than the number of space-time dimensions.

We observe that the inequality $2N > n + 1$ determines the minimum number of masses required for the regularization of the n-dimensional propagators according to the Pauli-Villars[2] procedure.

When $2N \leq n$, one may write, by means of (A6), expressions which exhibit the singularities of $G_N(x)$ and $G_N^1(x)$ explicitly. For odd-dimensional spaces we find

$$G_N^{(\text{sing})}(x) = \frac{\epsilon(x)}{2}\left(\frac{1}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(n+1)/2-N} \frac{S(n/2-m-\frac{1}{2})}{4^{-m}\Gamma(n/2-m+\frac{1}{2})}\,\theta^{(m)}(x^2)\right\}, \tag{37}$$

$$G_N^{1(\text{sing})}(x) = -\frac{i}{\pi}\left(\frac{1}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(n+1)/2-N} \frac{S(n/2-m-\frac{1}{2})}{4^{-m}\Gamma(n/2-m+\frac{1}{2})} \times \left(\frac{d}{dx^2}\right)^m \ln(|x^2|)^{1/2}\right\}. \tag{38}$$

It is clear that $G_N(x;$ odd $n)$ contains a finite jump $\theta(x^2)$ on the light cone in addition to a linear combination of derivatives up to order $\frac{1}{2}(n-2N-1)$ of $\delta(x^2)$, whereas the singular part of $G_N^1(x;$ odd $n)$ consists of a logarithmic singularity ln $(|x^2|)$ along with a polynomial in $(x^2)^{-1}$ of degree $\frac{1}{2}(n-2N-1)$.

For spaces with an even number of dimensions, one obtains

$$G_N^{(\text{sing})}(x) = \frac{\epsilon(x)\theta(x^2)}{2}\left(\frac{1}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(n-N)/2} \frac{S(n/2-m-1)(x^2)^{-m-1/2}}{4^{-m-1/2}\Gamma(n/2-m)\Gamma(\frac{1}{2}-m)}\right\}, \quad (39)$$

$$G_N^{1(\text{sing})}(x) = \frac{\theta(-x^2)}{2}\left(\frac{1}{4\pi}\right)^{(n-1)/2} \times \left\{\sum_{m=0}^{(n-N)/2} \frac{S(n/2-m-1)(x^2)^{-m-1/2}}{4^{-m-1/2}\Gamma(n/2-m)\Gamma(\frac{1}{2}-m)}\right\}. \quad (40)$$

Here the singular part of the two Green's functions consists of a polynomial in $(x^2)^{-1/2}$ of degree $n-2N+1$, with $G_N^{(\text{sing})}(x;$ even $n)$ vanishing *outside*, and $G_N^{1(\text{sing})}(x;$ even $n)$ vanishing *inside*, the light cone.

In the last four equations, the sum $S(l)$ is given by

$$S(N+q-1) = \sum_{\zeta_1+\cdots+\zeta_L=q} \frac{(\lambda_1)_{\zeta_1}\cdots(\lambda_L)_{\zeta_L}}{\zeta_1!\cdots\zeta_L!} \times (-\mu_1^2)^{\zeta_1}\cdots(-\mu_L^2)^{\zeta_L}, \quad (41)$$

where $q = 0, 1, \cdots$ and $(\lambda_p)_{\zeta_p} = \lambda_p(\lambda_p+1)\cdots(\lambda_p+\zeta_p-1)$.

Finally, it is clear that $G_{\text{hom}}(x)$ in (21) is regular on the light cone, provided

$$\sum_{p=1}^{L}\sum_{m=1}^{\lambda_p} \frac{C_{pm}}{\Gamma(m)}\left(\frac{-d}{d\mu_p^2}\right)^{m-1}(-\mu_p^2)^l = 0 \qquad (l = 0, 1, \cdots, \nu) \quad (42)$$

where $C_{pm} = a_{pm} + b_{pm}$ and

$$\nu = \begin{Bmatrix}\frac{1}{2}(n-1)\\ \frac{1}{2}n-1\end{Bmatrix} \quad \text{for} \quad \begin{Bmatrix}\text{odd} & n\\ \text{even} & n\end{Bmatrix}. \quad (43)$$

In the case of a multi-mass operator with *distinct* rest masses ($\lambda_p \equiv 1$), the singularities and finite jumps cancel if

$$\sum_{p=1}^{L} C_p(\mu_p^2)^l = 0 \qquad (l = 0, 1, \cdots, \nu). \quad (44)$$

These equations represent the multi-dimensional analog of the well-known regularization conditions of Pauli and Villars,[2] guaranteeing the absence of singularities on the light cone.

If one considers (42) with the condition $\nu = N-2$ instead of the condition (43), the homogeneous system (42) contains $N-1$ equations in the N parameters C_{pm}. From (31) and (32), after applying Leibniz's rule, it follows that

$$C_{pm} = \text{const} \times \frac{1}{\Gamma(\lambda_p - m + 1)}\left(-\frac{d}{d\mu_p^2}\right)^{\lambda_p - m}\Phi(-\mu_p^2) \quad (45)$$

is a solution of this system. Since the solution of such a system is determined within a multiplicative constant, it is the only solution, apart from the trivial one

$$C_{pm} = 0. \quad (46)$$

Substituting these solutions in (21) and remembering (15) and (23), it is easily seen that the regularization condition (42) with $\nu = N-2$ is satisfied only for the homogeneous functions

$$G_{\text{hom}} = AG_N + BG_N^1 \quad (47)$$

and

$$G_{\text{hom}} = \sum_{p=1}^{L}\sum_{m=1}^{\lambda_p} d_{pm}[\Delta(x; m; \mu_p) - \Delta^1(x; m; \mu_p)], \quad (48)$$

where A, B, and the d_{pm} are arbitrary constants. Since, as one can now see, Eq. (42) with $\nu = N-2$ is equivalent to the vanishing on the light cone of the first $N-(n+3)/2$ (if n is odd) or $N-n/2-1$ (if n is even) derivatives of G_{hom} with respect to x^2, it follows that no Green's function vanishing outside the light cone can be more regular than G_N, and that G_N is determined uniquely within a multiplicative constant by these regularity and causality conditions.

4. INITIAL CONDITIONS OF THE GREEN'S FUNCTIONS

It will become evident that the set of homogeneous Green's functions $G_\eta(x; r_1\cdots r_L; \mu_1\cdots\mu_L)$ with $0 \le r_p \le \lambda_p$, $\eta = r_1 + \cdots + r_L$, and $1 \le \eta \le N$ is particularly useful for constructing the solution of the Cauchy problem. For this reason, we shall briefly investigate the behavior of these functions and their time derivatives at $t = 0$. The initial behavior of the G_η^1 functions will not be considered, aside from the obvious statement that for $m = 0, 1, 2, \cdots$ the equations

$$\partial_0^{2m+1}G_\eta^1(\mathbf{x}, 0) = 0, \quad (49)$$

$$\partial_0^{2m}G_\eta(\mathbf{x}, 0) = 0 \quad (50)$$

follow immediately from (B1) and (B2). Because of (50), we need only calculate the time derivatives of odd order for $G_\eta(x)$.

There are of course many ways of performing the calculation, the shortest of which seems to be the following. We use a result given in (I),

$$\partial_0^{2m+1}\,\Delta(\mathbf{x}, 0; \mu) = (\nabla^2 - \mu^2)^m\,\delta(\mathbf{x}), \tag{51}$$

which, when substituted into (23), gives

$$\partial_0^{2m+1}G_N(\mathbf{x}, 0) = \sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(\frac{d}{db_p}\right)^{\lambda_p-1}(b_p)^m\Phi(b_p)\,\delta(\mathbf{x}), \tag{52}$$

where

$$b_p = \nabla^2 - \mu_p^2 \quad\text{and}\quad \Phi(b_p) = \prod_j{}'\,(b_p - b_j)^{-\lambda_j}.$$

The sum is evaluated explicitly in Appendix A; using those results, we obtain the initial conditions

$$\partial_0^m G_N(\mathbf{x}, 0) = \begin{cases} 0 & \text{for } m = 0, 1, \cdots, 2N-2; \\ \delta(\mathbf{x}) & \text{for } m = 2N-1; \\ 0 & \text{for all even } m; \end{cases} \tag{53}$$

along with

$$\partial_0^{2N+2m-1}G_N(\mathbf{x}, 0; \lambda_1, \cdots, \lambda_L; \mu_1, \cdots, \mu_L) = \sum_{\zeta_1+\cdots+\zeta_L=m}\frac{(\lambda_1)_{\zeta_1}\cdots(\lambda_L)_{\zeta_L}}{\zeta_1!\cdots\zeta_L!} \times (\nabla^2 - \mu_1^2)^{\zeta_1}\cdots(\nabla^2 - \mu_L^2)^{\zeta_L}\,\delta(\mathbf{x}), \tag{54}$$

where $(\lambda_p)_{\zeta_p} = \lambda_p(\lambda_p + 1)\cdots(\lambda_p + \zeta_p - 1)$ is the Pochhammer symbol. Of course similar initial conditions are obtained for the $G_\eta(x; r_1\cdots r_L; \mu_1\cdots\mu_L)$.

5. GENERAL SOLUTION OF THE CAUCHY PROBLEM

The solution of

$$(\Box + \mu_1^2)^{\lambda_1}\cdots(\Box + \mu_L^2)^{\lambda_L}\varphi(x) = 0$$

taking specified values for $\varphi(x)$, $\partial_0\varphi(x)$, $\cdots$, $\partial_0^{2N-1}\varphi(x)$ on the spacelike plane $t = 0$, may be written in the form

$$\varphi(x) = \int_{t'=0} d\mathbf{x}' \sum_{m=0}^{N-1}\sum_{\eta=m+1}^{N} A_{m\eta}\,\partial_0^{2\eta-2m-2}G_\eta(x - x') \times (\overleftarrow{\partial}_0' - \overrightarrow{\partial}_0')\,\partial_0'^{2m}\varphi(x'), \tag{55}$$

where the arrows indicate the direction in which the differentiation is to be carried out. For each $m = 0, 1, \cdots, N-1$, the $N - m$ constants $A_{m\eta}$ are to be determined from the $N - m$ equations

$$\sum_{\eta=m+1}^{N} A_{m\eta}\,\partial_0^{2\eta+2j-1}G_\eta(\mathbf{x}, 0) = \delta_{0j}\,\delta(\mathbf{x}), \tag{56}$$

with $j = 0, 1, \cdots, N-m-1$.

We first show that (55), subject to the conditions (56), does indeed represent the solution of Cauchy's problem. Consider

$$[\partial_0^{2M}\varphi(x)]_{t=0} = \int_{t'=t=0} d\mathbf{x}' \sum_{m=0}^{M}\sum_{\eta=m+1}^{N} A_{m\eta}\,\partial_0^{2\eta+2(M-m)-1} \times G_\eta(x - x')\,\partial_0'^{2m}\varphi(x'), \tag{57}$$

where the upper limit to the sum over M is determined by (53). Clearly the integral in (55) corresponding to $\overleftarrow{\partial}_0'$ makes no contribution because of (50). Hence, summing over η by virtue of (56), one obtains the desired result

$$\partial_0^{2M}\varphi(\mathbf{x}, 0) = \int d\mathbf{x}' \sum_{m=0}^{M}\delta_{mM}\,\delta(\mathbf{x} - \mathbf{x}')\,\partial_0'^{2m}\varphi(\mathbf{x}', 0). \tag{58}$$

An exactly analogous proof may be used for $\partial_0^{2M+1}\varphi(\mathbf{x}, 0)$; thus (55) is the required solution.

The problem now is to find the $A_{m\eta}$ explicitly. Evidently the existence of a solution to (56) depends on what G_η functions are employed; one must choose an independent set. In the present case, we can guarantee that a solution exists by using *all* of the $G_\eta(x; r_1\cdots r_L; \mu_1\cdots\mu_L)$. Because of (54), the conditions (56) may be reduced to algebraic equations for the $A_{m\eta}$; namely,

$$\sum_{\eta=m+1}^{N} A_{m\eta}\sum_{\zeta_1+\cdots+\zeta_L=j}\frac{(r_1)_{\zeta_1}\cdots(r_L)_{\zeta_L}}{\zeta_1!\cdots\zeta_L!} \times (b_1)^{\zeta_1}\cdots(b_L)^{\zeta_L} = \delta_{0j}, \tag{59}$$

where the sum over $\eta = r_1 + \cdots + r_L$ is understood as a sum over all the r_p consistent with $0 \le r_p \le \lambda_p$ and $m + 1 \le r_1 + \cdots + r_L \le N$. Since we expect $A_{m\eta} = A_{m\eta}(r_1\cdots r_L; \lambda_1\cdots\lambda_L)$ to be independent of the rest masses, we shall seek a solution of

$$\sum_{\eta=m+1}^{N} A_{m\eta}(r_1)_{\zeta_1}\cdots(r)_{\zeta_L} = \delta_{0j} \tag{60}$$

with $j = \zeta_1 + \cdots + \zeta_L$. Such a solution will, of course, be a solution of (59).

We assert that

$$A_{m\eta} = (-)^{\eta-m-1}\binom{\eta-1}{m}\binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L} \tag{61}$$

satisfies (60) in addition to satisfying the requirements $0 \le r_p \le \lambda_p$. To prove this, consider the function

$$F(x, b) = -\frac{(-)^m}{m!}\sum_{r_1,\cdots,r_L}(-)^\eta\binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L}b^{-\eta}x^{\eta-1} = -\frac{(-)^m}{m!}\frac{1}{x}\frac{(b-x)^N}{b^N}. \tag{62}$$

Differentiating with respect to x, one finds

$$(\partial^m/\partial x^m)F(1, b) = -1 + \sum_{\eta=m+1}^{N} (-)^{\eta-m-1} \times \binom{\eta-1}{m}\binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L} b^{-\eta} \quad (63)$$

where the -1 comes from the $r_1 + \cdots + r_L = 0$ term. Differentiating with respect to b, one may write

$$\frac{\partial^i}{\partial b^i}\frac{\partial^m}{\partial x^m} F(1, 1) + \delta_{0i} = \sum_{\zeta_1+\cdots+\zeta_L=i} \frac{(-)^i i!}{\zeta_1!\cdots\zeta_L!} \sum_{\eta=m+1}^{N} (-)^{\eta-m-1} \times \binom{\eta-1}{m}\binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L}(r_1)_{\zeta_1}\cdots(r_L)_{\zeta_L}, \quad (64)$$

where we have used the multinomial differentiation rule

$$\frac{d^N}{dx^N} u_1 u_2 \cdots u_t = \sum_{r_1+\cdots+r_t=N} \frac{N!}{r_1!\cdots r_t!} u_1^{(r_1)}\cdots u_t^{(r_t)} \quad (65)$$

which follows by induction from Leibniz's rule.

A direct calculation from the second equality in (62) gives

$$\frac{\partial^i}{\partial b^i}\frac{\partial^m}{\partial x^m} F(1, 1) = 0 \quad (66)$$

for $m = 0, 1, \cdots, N-1$ and $j = 0, 1, \cdots, N-m-1$; therefore,

$$\sum_{\zeta_1+\cdots+\zeta_L=i} \frac{1}{\zeta_1!\cdots\zeta_L!} \sum_{\eta=m+1}^{N} (-)^{\eta-m-1} \times \binom{\eta-1}{m}\binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L}(r_1)_{\zeta_1}\cdots(r_L)_{\zeta_L} = \delta_{0i}. \quad (67)$$

However, because of the symmetry of the summand, we clearly must have

$$\sum_{\eta=m+1}^{N} (-)^{\eta-m-1}\binom{\eta-1}{m}\binom{\lambda_1}{r_1}\cdots \times \binom{\lambda_L}{r_L}(r_1)_{\zeta_1}\cdots(r_L)_{\zeta_L} = \delta_{0i}, \quad (68)$$

so our assertion is proved.[9]

Returning to the original solution (55) of the Cauchy problem, using (61) and summing over m, we find

$$\varphi(x) = \int_{t'=0} d\mathbf{x}' \sum_{\eta=1}^{N} \binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L} G_\eta \times (x - x'; r_1, \cdots, r_L; \mu_1, \cdots, \mu_L) \times (\vec{\partial}_0' - \overleftarrow{\partial}_0')(\vec{\partial}_0'^2 - \overleftarrow{\partial}_0'^2)^{\eta-1}\varphi(x'). \quad (69)$$

All possible G_η solutions are here involved, although only the N functions $\Delta(x; m; \mu_p)$ are ultimately present. Upon introducing the differential operator

$$X \equiv (\vec{\partial}_0 - \overleftarrow{\partial}_0) \sum_{\eta=1}^{N} \binom{\lambda_1}{r_1}\cdots \times \binom{\lambda_L}{r_L}(\vec{\partial}_0^2 - \overleftarrow{\partial}_0^2)^{\eta-1} \prod_{p=1}^{L} (\overleftarrow{\Box} + \mu_p^2)^{\lambda_p - r_p} \quad (70)$$

we can write the Cauchy solution in the neat form

$$\varphi(x) = \int_{t'=0} d\mathbf{x}' G_N(x - x') X' \varphi(x'). \quad (71)$$

We note that X may be written

$$(\vec{\partial}_0^2 - \overleftarrow{\partial}_0^2)X = (\vec{\partial}_0 - \overleftarrow{\partial}_0)\Big[\prod_{p=1}^{L} (\overleftarrow{\Box} + \mu_p^2 + \vec{\partial}_0^2 - \overleftarrow{\partial}_0^2)^{\lambda_p} - \prod_{p=1}^{L} (\overleftarrow{\Box} + \mu_p^2)^{\lambda_p}\Big]. \quad (72)$$

6. INVARIANT FORM OF THE SOLUTION

All of the Green's functions we have discussed are invariant within the proper Lorentz group, so that an invariant form of the field $\varphi(x)$ is easily obtained. As is well known, the derivative of $\Delta(x; \mu)$ normal to an arbitrary spacelike surface $\sigma(x)$ with normal $n_\beta(x)$ is given by[10]

$$n_\beta(x)\, \partial_\beta\, \Delta(x; \mu) = \delta_\sigma(x) \quad (x^2 < 0) \quad (73)$$

where $\delta_\sigma(x)$ is the invariant surface δ function with the properties

$$\delta_\sigma(x) = 0 \quad (x \not\subset \sigma), \quad (74)$$

$$\int d\sigma(x)\ \delta_\sigma(x) = 1.$$

The corresponding behavior of $G_N(x)$ on spacelike surfaces may be obtained as follows.

Because Δ satisfies the Klein-Gordon equation, it follows that

$$\Box^m\, \Delta(x; \mu) = (-\mu^2)^m\, \Delta(x; \mu); \quad (75)$$

consequently, for $x^2 < 0$ we have

$$n_\beta(x)\, \partial_\beta\, \Box^m\, G_N(x) = S(m)\, \delta_0(x), \quad (76)$$

[9] A rigorous proof showing that (61) satisfies Eq. (59) directly is not hard to construct, but seems longer than this demonstration.

[10] See, e. g., J. Rzewuski, *Field Theory* (Hafner Publishing Company, New York, 1958).

where $S(m)$ is the sum given by (31). Utilizing the explicit expression for the sum, one immediately finds, for spacelike x,

$$n_\beta(x)\,\partial_\beta\,\Box^m\,G_N(x) = \begin{cases} 0 & \text{for } m = 0, 1, \cdots, N-2; \\ \delta_\sigma(x) & \text{for } m = N-1, \end{cases} \tag{77}$$

and

$$n_\beta(x)\,\partial_\beta\,\Box^{N+m-1}G_N(x) = \sum_{\zeta_1+\cdots\zeta_L=m} \frac{(\lambda_1)_{\zeta_1}\cdots(\lambda_L)_{\zeta_L}}{\zeta_1!\cdots\zeta_L!}(-\mu_1^2)^{\zeta_1}\cdots \times(-\mu_L^2)^{\zeta_L}\,\delta_\sigma(x). \tag{78}$$

Furthermore,

$$\Box^m\,G_N(x) = 0(x^2 < 0) \tag{79}$$

follows directly from the fact that G_N is an odd invariant function vanishing outside the light cone.

One may now write an invariant analog of (55) with constants $A_{m\eta}$ that satisfy the same algebraic equations. The result gives the general solution in the invariant form[11]

$$\varphi(x) = \int d\sigma_\beta(x')G_N(x-x')X'_\beta\varphi(x'), \tag{80}$$

where

$$X_\beta = (\vec\partial_\beta - \overleftarrow\partial_\beta)\sum_{\eta=1}^{N}\binom{\lambda_1}{r_1}\cdots\binom{\lambda_L}{r_L}\vec\Box - \overleftarrow\Box)^{\eta-1} \times\prod_{p=1}^{L}(\overleftarrow\Box+\mu_p^2)^{\lambda_p-r_p}. \tag{81}$$

Equation (72) is replaced by

$$(\vec\Box - \overleftarrow\Box)X_\beta = (\vec\partial_\beta - \overleftarrow\partial_\beta) \times\left[\prod_{p=1}^{L}(\vec\Box+\mu_p^2)^{\lambda_p} - \prod_{p=1}^{L}(\overleftarrow\Box+\mu_p^2)^{\lambda_p}\right]. \tag{82}$$

Using (82), the integral in (80) is easily seen to be independent of the data carrying surface $\sigma(x)$; we have

$$\delta\varphi(x)/\delta\sigma = \partial'_\beta G_N(x-x')X'_\beta\varphi(x') = G_N(x-x')[\prod(\vec\Box'+\mu_p^2)^{\lambda_p} - \prod(\overleftarrow\Box'+\mu_p^2)^{\lambda_p}]\varphi(x') = 0, \tag{83}$$

since both $G_N(x)$ and $\varphi(x)$ are solutions of the homogeneous equation.

It is well to remark that these results can be extended to the case of complex μ_p. In general, for a polynomial $F(z)$ with real or complex coefficients, the homogeneous equation

$$F(\Box)\varphi(x) = 0 \tag{84}$$

has a complete solution in the form

$$\varphi(x) = \int d\sigma_\beta(x')\,\Delta_N(x-x')X'_\beta\varphi(x'), \tag{85}$$

where

$$\Delta_N(x) = (2\pi)^{-n-1}\int_C dk e^{-ikx}/F(-k^2) \tag{86}$$

and

$$(\vec\Box - \overleftarrow\Box)X_\beta = (\vec\partial_\beta - \overleftarrow\partial_\beta)[F(\vec\Box) - F(\overleftarrow\Box)]. \tag{87}$$

APPENDIX A. PARTIAL FRACTIONS

Partial fraction decompositions for quotients of polynomials have long been known (although the explicit formulas are difficult to find in textbooks[12]). We shall list here the pertinent results for quotients of the form $z^q\prod(z-b_j)^{-\lambda_j}$ assuming the b_j to be distinct and nonzero, and setting $N \equiv \lambda_1 + \cdots + \lambda_L$.

When $q = 0, 1, \cdots N-1$ the following expansion holds:

$$\frac{z^q}{\prod_{j=1}^{L}(z-b_j)^{\lambda_j}} = \sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(\frac{d}{db_p}\right)^{\lambda_p-1} \times\left[\frac{(b_p)^q\prod_{j=1}^{L}{}'(b_p-b_j)^{-\lambda_j}}{z-b_p}\right] = \sum_{p=1}^{L}\sum_{m=1}^{\lambda_p}\frac{(z-b_p)^{-m}}{(\lambda_p-m)!} \times\left[\left(\frac{d}{db_p}\right)^{\lambda_p-m}(b_p)^q\prod_{j=1}^{L}{}'(b_p-b_j)^{-\lambda_j}\right], \tag{A1}$$

where the second equality is obtained using Leibniz's rule. The factor for $j = p$ is to be left out of the primed products. The important result

$$\sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(\frac{d}{db_p}\right)^{\lambda_p-1}\left[(b_p)^l\prod_{j=1}^{L}{}'(b_p-b_j)^{-\lambda_j}\right] = 0 \qquad (l = 1, 2, \cdots, N-2) \tag{A2}$$

is obtained immediately upon setting $z = 0$ and $q = l+1$.

The partial fraction theorem also gives

$$\frac{z^{N+q}}{\prod_{j=1}^{L}(z-b_j)^{\lambda_j}} = E_q(z) + \sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(\frac{d}{db_p}\right)^{\lambda_p-1} \times\left[\frac{(b_p)^{N+q}\prod_{j=1}^{L}{}'(b_p-b_j)^{-\lambda_j}}{z-b_p}\right], \tag{A3}$$

[11] Compare with reference 4.

[12] A complete treatment of partial fractions is given by J. A. Serret, *Cours d'algèbre supréieure* (Gauthier-Villars, Paris, 1885), Tome I.

where $q = 0, 1, \cdots$, and $E_q(z)$ is a polynomial of degree q. An explicit form for $E_q(z)$ is easily obtained as follows: Since

$$E_q(z) = \left[\frac{z^{N+q}}{(z-b_1)^{\lambda_1}\cdots(z-b_L)^{\lambda_L}}\right]_{\text{entire part}}, \quad \text{(A4)}$$

expand the denominator using the binomial theorem to find

$$E_q(z) = [z^q(1-b_1/z)^{-\lambda_1}\cdots(1-b_L/z)^{-\lambda_L}]_{\text{entire part}}$$
$$= \sum_{\zeta_1+\cdots+\zeta_L\le q}\binom{\lambda_1-1+\zeta_1}{\lambda_1-1}\cdots\binom{\lambda_L-1+\zeta_L}{\lambda_L-1}$$
$$\times (b_1)^{\zeta_1}\cdots(b_L)^{\zeta_L}z^{q-\zeta_1-\cdots-\zeta_L}. \quad \text{(A5)}$$

In particular, then

$$E_q(0) = \sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(\frac{d}{db_p}\right)^{\lambda_p-1}\left[\frac{(b_p)^{N+q-1}}{\prod_{j=1}{}'^{L}(b_p-b_j)^{\lambda_j}}\right]$$
$$= \sum_{\zeta_1+\cdots+\zeta_L=q}\frac{(\lambda_1)_{\zeta_1}\cdots(\lambda_L)_{\zeta_L}}{\zeta_1!\cdots\zeta_L!}(b_1)^{\zeta_1}\cdots(b_L)^{\zeta_L}$$
$$= \frac{1}{q!}\frac{d^q}{d\zeta^q}[(1-b_1\zeta)^{-\lambda_1}\cdots(1-b_L\zeta)^{-\lambda_L}]_{\zeta=0}, \quad \text{(A6)}$$

where the third equality is obtained using the multinomial differentiation rule (65). The important formula

$$\sum_{p=1}^{L}\frac{1}{\Gamma(\lambda_p)}\left(\frac{d}{db_p}\right)^{\lambda_p-1}\left[(b_p)^{N-1}\prod_{j=1}{}'^{L}(b_p-b_j)^{-\lambda_j}\right] = 1 \quad \text{(A7)}$$

is obtained because $E_0(0) = 1$.

APPENDIX B. CALCULATION OF Δ AND Δ^1

We first calculate the Green's functions for the Klein-Gordon operator following the procedure of (I). Upon performing the k_0 integration over the paths C and C^1, one finds the well-known integral representations

$$\Delta(x;\mu) = \frac{\epsilon(x)}{(2\pi)^n}\int_{-\infty}^{\infty}d\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{x}}\left(\frac{\sin\omega\,|t|}{\omega}\right) \quad \text{(B1)}$$

$$\Delta^1(x;\mu) = \frac{i}{(2\pi)^n}\int_{-\infty}^{\infty}d\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{x}}\left(\frac{\cos\omega t}{\omega}\right), \quad \text{(B2)}$$

where $\omega = +(\mathbf{k}^2+\mu^2)^{1/2}$. Since

$$(1/\omega)(\sin\omega t + i\cos\omega t) = (\pi t/2\omega)^{1/2}H^{(2)}_{1/2}(\omega t), \quad \text{(B3)}$$

we introduce (for $t > 0$) the function

$$\mathbf{K}(x;\mu) \equiv (2\pi)^{-n}\int_{-\infty}^{\infty}d\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{x}}\left(\frac{\pi t}{2\omega}\right)^{1/2}H^{(2)}_{1/2}(\omega t), \quad \text{(B4)}$$

and note that

$$\Delta(x;\mu) = \epsilon(x)\,\mathrm{Re}\,\mathbf{K}(x;\mu), \quad \text{(B5)}$$

$$\Delta^1(x;\mu) = i\,\mathrm{Im}\,\mathbf{K}(x;\mu). \quad \text{(B6)}$$

Integrating (B4) over the angles, one obtains

$$\mathbf{K}(x;\mu) = \frac{1}{2}\left(\frac{1}{2\pi}\right)^{(n-1)/2}$$
$$\times\int_0^{\infty}\frac{J_{(n-1)/2}(kR)}{R^{(n-1)/2}}\frac{H^{(2)}_{1/2}(\omega t)}{t^{-1/2}}\frac{k^{n/2}\,dk}{\omega^{1/2}}, \quad \text{(B7)}$$

where the integral over k is an integral of the Sonine-Gegenbauer type [Bateman,[13] Sec. 7.14.2 (48)]; thus, finally

$$\mathbf{K}(x;\mu) = \frac{1}{2}\left(\frac{\mu}{2\pi}\right)^{(n-1)/2}(\sqrt{x^2})^{(1-n)/2}H^{(2)}_{(1-n)/2}(\mu\sqrt{x^2}). \quad \text{(B8)}$$

In (B8), $(x^2)^{1/2}$ is defined as $-i(-x^2)^{1/2}$ when $x^2 < 0$ and Bateman's Sec. 7.2.2 (16) has also been used.

Just as in (I), the Sonine-Gergenbauer integral must be defined as a distribution for general orders of the Bessel functions in the integrand. Equation (20) now follows by an application of (15) and the well-known formula

$$(-d/z\,dz)^m\{z^{-\nu}H^{(2)}_{\nu}(z)\} = z^{-\nu-m}H^{(2)}_{\nu+m}(z). \quad \text{(B9)}$$

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. T. L. McCoy for informative correspondence and Professor H. Uberall and R. A. Leacock for reading the manuscript. This work was supported in part by U. S. Public Health Service Training Grant 2G-174 to Dartmouth Medical School.

[13] H. Bateman, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vols. I, II, III.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Green's Distributions Associated with the Operator $[\Box^m - (-\mu^2)^m]^l$

JOHN JUDSON BOWMAN* AND JOSEPH DAVID HARRIS
Dartmouth College and Dartmouth Medical School, Hanover, New Hampshire
(Received May 18, 1962)

The results of a previous paper on homogeneous Green's functions for a multi-dimensional iterated Klein-Gordon operator $(\Box + \mu^2)^l$ are extended to include homogeneous Green's functions associated with the operator $[\Box^m - (-\mu^2)^m]^l$ in multi-dimensional spaces. The Fourier representation of the Green's functions may be expressed, after some angular integrations, as a one-dimensional infinite integral which does not in general converge. Using the concepts of distribution analysis, it is shown how this improper integral can be evaluated directly to get explicit expressions for the Green's functions. The Green's functions themselves must then be interpreted as distributions in the sense of Schwartz. Several distributions instrumental in this treatment are introduced and their properties studied. Explicit expressions for the singularities of the Green's functions on the light cone are presented. The well-known difference between even- and odd-dimensional spaces is reflected in the nature of these singularities. The singularities appearing for odd-dimensional spaces consist of a finite linear combination of derivatives of the Dirac delta function $\delta(s^2)$, where s is the space-time distance. The highest derivative appearing is of order $\frac{1}{2}(n-2m-l)$ with n giving the number of space dimensions. The singular part for even-dimensional spaces consists of a polynomial in $1/s$ of degree $n - 2ml + 1$. No singularities appear when the order of the operator is greater than the number of space dimensions. Finally, a complete set of homogeneous Δ-function solutions is given along with their initial conditions at zero time. All of these functions would be needed in obtaining the general solution to the Cauchy problem for the operator considered.

1. INTRODUCTION

In a previous paper[1] we obtained the explicit form of the homogeneous Green's function associated with a multi-dimensional iterated Klein-Gordon operator $(\Box + \mu^2)^l$. Starting from its Fourier representation, the Green's function was expressed as a one-dimensional, infinite integral of the Sonine type. It was shown that although this integral is classically divergent when the order of the operator is less than the number of space dimensions, it can be treated rigorously under these conditions using the concepts of distribution analysis. The Green's function is then to be regarded as a (tempered) distribution in the sense of Schwartz.[2] A new distribution introduced for the purpose of giving the improper Sonine integral its generalized meaning was used to investigate the singularities of the Green's function on the light cone.

In the present paper we extend these results to include Green's functions for equations of a more general nature. It is interesting to consider equations of the type

$$[\Box^m - (-\mu^2)^m]^l \varphi = \rho, \tag{1}$$

where m and l are positive integers, and $\Box$ is the multi-dimensional d'Alembert operator

$$\Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \cdots - \frac{\partial^2}{\partial x_n^2}. \tag{2}$$

Clearly the Fourier representation of the Green's function can be given in the form

$$K(x) = (2\pi)^{-n-1} \int dk e^{-ikx} [(-k^2)^m - (-\mu^2)^m]^{-l}. \tag{3}$$

As is well-known, the path of integration in the k_0 plane determines which Green's functions (homogeneous or inhomogeneous) are being represented.

The integrand in (3) contains $2n$ poles of order l for $(-k^2)^m = (-\mu^2)^m$. The location of these poles in the k_0 plane is given by

$$k_0 = [k^4 + \mu^4 + 2k^2\mu^2 \cos(2p\pi/m)]^{1/4} e^{i(q\pi + \vartheta_p/2)}, \tag{4}$$

$$\tan \vartheta_p = \sin(2p\pi/m)/[(k^2/\mu^2) + \cos(2p\pi/m)], \tag{5}$$

where $p = 0, 1, \cdots, m-1$ and $q = 0, 1$. The singularities in the k_0 plane are thus confined within an annular ring of outside radius $(k^2 + \mu^2)^{1/2}$ and inside radius $(k^2 - \mu^2)^{1/2}$. The retarded and advanced Green's functions may be defined in the usual way. The retarded solution is obtained by taking the path of the k_0 integration above the real axis such that all the singularities in the k_0 plane lie below this path. The advanced solution corresponds to a path below the real axis with all the singularities situated above the path. The homogeneous Green's function $\Delta_{n,l,m}(x)$ is then given by

$$\Delta_{n,l,m}(x) = \Delta^{\text{ret}}_{n,l,m}(x) - \Delta^{\text{adv}}_{n,l,m}(x), \tag{6}$$

---

* Present address: University of Michigan, Ann Arbor, Michigan.

[1] J. J. Bowman and J. D. Harris, J. Math. Phys. 3, 396 (1962), hereafter referred to as (I).

[2] L. Schwartz, *Théorie des distributions, I, II* (Hermann & Cie., Paris, 1950–51).

and the inhomogeneous Green's function $\bar{\Delta}_{n,l,m}$ is related to $\Delta_{n,l,m}$ by the formula

$$\Delta_{n,l,m}(x) = 2\epsilon(t)\,\bar{\Delta}_{n,l,m}(x) \tag{7}$$

with $\epsilon(t) = \pm 1$ for $t \gtrless 0$.

A direct calculation from (3) shows that $\Delta_{n,l,m}$ may be written in the form [cf. (I)]

$$\Delta_{n,l,m}(\mathbf{x}, t) = \frac{\epsilon(t)}{(2\pi)^{n/2}} \times \int_0^\infty dk\, \frac{J_{n/2-1}(kr)}{r^{n/2-1}}\, k^{n/2} G_{l,m}(k, |t|), \tag{8}$$

where $J_\nu(z)$ is the Bessel function of the first kind and $G_{l,m}(k, t)$ is given by

$$G_{l,m}(k, t) = \frac{1}{2\pi i} \times \int_{c-i\infty}^{c+i\infty} dz e^{zt}[(z^2 + k^2)^m - (-\mu^2)^m]^{-l}. \tag{9}$$

Here $c > 0$ is a real number such that all the singularities of the integrand in (9) are on the left of the straight line from $c - i\infty$ to $c + i\infty$.

Clearly $G_{l,m}$ is a solution of the differential equation

$$[(d^2 + k^2)^m - (-\mu^2)^m]^l G(t) = \delta(t), \tag{10}$$

where $d \equiv d/dt$.

In general, convergence difficulties are encountered when one attempts to evaluate the integral in (8) to get explicit expressions for the Green's functions and it soon becomes clear that the Green's function itself must be given some generalized interpretation. Such an interpretation is of course available in Schwartz's theory of distributions.[2]

Following the program of (I) we introduce (cf. Sec. 2) several distributions instrumental in the treatment of Green's functions. An integral formula is developed which enables the explicit evaluation of integrals in the form of (8). These results are then applied (cf. Sec. 3) in obtaining directly the homogeneous Green's function $\Delta_{n,l,m}$ for Eq. (1).

Explicit expressions for the singularities of the Green's function $\Delta_{n,l,m}$ on the light cone are given in Sec. 3. Here special attention must be paid to the cases of even n and odd n due to the well-known difference between even- and odd-dimensional spaces in the solutions of hyperbolic differential equations. This remarkable difference is clearly reflected in the nature of the singularities on the light cone.

For spaces with an odd number of dimensions, the singular part of the Green's function consists of a finite sum of derivatives of the Dirac delta function $\delta(s^2)$ where s is the space-time distance given by $s^2 = t^2 - r^2$. The highest derivative appearing is of order $\frac{1}{2}(n - 2ml - 1)$. For even-dimensional spaces, however, the singular part is a polynomial in $1/s$ of degree $n - 2ml + 1$. In both cases, no singularities appear on the light cone when the order of the differential equation is greater than the number of space dimensions.

A complete set of Δ-function solutions is required in the general solution of the Cauchy initial value problem for Eq. (1). Because the integrand in (3) contains $2m$ poles of order l, there are ml independent Δ-function solutions to Eq. (1). An explicit expression for the complete set of these solutions is presented in Sec. 3 along with the initial conditions satisfied then at time $t = 0$.

2. SPECIAL DISTRIBUTIONS

Before introducing specific distributions it may be well to summarize briefly the basic definitions[2] involved in distribution analysis and recapitulate some results presented in (I).

Distributions are defined as linear operators which map a space of functions into a space of numbers. Different spaces of functions generate different distribution spaces. Following Schwartz we denote various spaces as follows:

(i) The spaces ($\mathfrak{D}$) consists of all (C^∞) functions on R^n with compact support.

(ii) The space of distributions ($\mathfrak{D}'$) is the dual of ($\mathfrak{D}$).

(iii) The space ($\mathfrak{S}$) is the space of all (C^∞) functions on R^n that "decay rapidly at infinity."

(iv) The space of tempered distributions ($\mathfrak{S}'$) is the dual of ($\mathfrak{S}$).

Every distribution $\mathfrak{T}$ can be represented by a continuous linear functional $\mathfrak{T}\{\phi\}$ on ($\mathfrak{D}$). In the space of distributions, every distribution has derivatives of all orders which are themselves distributions, such derivatives being defined by[3]

$$\mathfrak{T}^{(n)}\{\phi\} \equiv (-)^n \mathfrak{T}\{\phi^{(n)}\}. \tag{11}$$

With every locally summable point function $\mathfrak{T}(x)$ we may associate the distribution $\mathfrak{T}\{\phi\}$ by the formula

$$\mathfrak{T}\{\phi\} = \int_{-\infty}^{\infty} \mathfrak{T}(x)\phi(x)\,dx; \tag{12}$$

[3] From now on we consider only distribution defined for the real axis. In Eq. (9) we put $\phi^{(n)}(x) = d^n\phi/dx^n$.

thus, "ordinary" functions appear as distributions, and the spaces above satisfy the relation $(\mathcal{D}) \subset (\mathcal{S}) \subset (\mathcal{S}') \subset (\mathcal{D}')$. For nonsummable functions it is often convenient to define "generalized functions" in the manner of Eq. (12), where now, the symbolic or generalized function $\mathfrak{T}(x)$ has no meaning independent of the defining formula (12).

In the space of tempered distributions $(\mathcal{S}')$ the Fourier transformation is defined, and the Fourier transformation of each tempered distribution is again a tempered distribution.[2] All Green's functions obtained from (3) are to be regarded as tempered distributions. The Laplace transformation is defined for distributions whose support is bounded on the left.[4] The support of a distribution $\mathfrak{T}$ is bounded on the left if $\mathfrak{T}\{\phi\} = 0$ for all functions $\phi \in (\mathcal{D})$ whose support is contained in some half-line $(-\infty, c)$. The space of all distributions with support bounded on the left is denoted by $(\mathcal{D}_+')$.

In (I) extensive use was made of the distribution $\mathfrak{H}_\beta$ defined for the entire finite β plane by the following equations:

$$\mathfrak{H}_\beta\{\phi\} = \frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta-1}\phi(x)\, dx \quad (\mathrm{Re}\,\beta > 0), \tag{13}$$

$$\mathfrak{H}_\beta\{\phi\} \equiv (-)^n \mathfrak{H}_{\beta+n}\{\phi^{(n)}\} \quad (\mathrm{Re}\,\beta + n > 0), \tag{14}$$

where n is a positive integer or zero. Note that $\mathfrak{H}_\beta$ is a member of both $(\mathcal{D}_+')$ and $(\mathcal{S}')$.

When $\beta = 0$, the Dirac distribution,

$$\mathfrak{H}_0\{\phi\} = \delta\{\phi\} = \phi(0) \tag{15}$$

is obtained. Derivatives of $\mathfrak{H}_\beta$ are given by

$$\mathfrak{H}_\beta^{(n)} = \mathfrak{H}_{\beta-n}, \tag{16}$$

and this formula may be extended to define *fractional* derivatives. In terms of the Dirac distribution we have

$$\mathfrak{H}_{-n} = \delta^{(n)}, \tag{17}$$

and the familiar property

$$\delta^{(n)}\{\phi\} = (-)^n \phi^{(n)}(0). \tag{18}$$

The product $x^n \mathfrak{H}_\beta$ with $n = 0, 1, 2, \cdots$ is defined by

$$x^n \mathfrak{H}_\beta = (\beta)_n \mathfrak{H}_{\beta+n}, \tag{19}$$

where

$$(\beta)_n = \beta(\beta+1)\cdots(\beta+n-1), \quad (\beta)_0 = 1. \tag{20}$$

[4] J. Lavoine, *Calcul symbolique* (Centre National de la Recherche Scientifique, Paris, 1959).

Now it is evident that a host of new distributions can be obtained by taking various linear combinations of the $\mathfrak{H}$ distributions. At the same time a wide variety of "special functions" occurring in classical analysis can be given extended interpretations when they are regarded as generalized functions. In (I), for example, we introduced the distribution

$$\Omega_\beta(\alpha; \lambda) = \sum_{q=0}^\infty \frac{\lambda^q \mathfrak{H}_{\beta+q+1}}{\Gamma(\alpha + q + 1)}, \tag{21}$$

which may also be written in the form

$$\Omega_\beta(\alpha; \lambda) = {}_1F_2(1; \alpha+1, \beta+1; \lambda x)\mathfrak{H}_{\beta+1}/\Gamma(\alpha+1), \tag{22}$$

where F is a hypergeometric series. This distribution is defined for all complex values of the parameters α, β, and λ by requiring, in accordance with (14) that

$$\Omega_\beta\{\phi\} \equiv (-)^n \Omega_{\beta+n}\{\phi^{(n)}\}, \tag{23}$$

where the nonnegative integer n is such that $\mathrm{Re}\,\beta + n + 1 > 0$.

Other properties of Ω are to be found in our previous paper; here we need to introduce a more general distribution $\mathfrak{Z}$ of which Ω is a special case:

$$\mathfrak{Z}_{\beta,\tau,\nu}(\alpha; \lambda) = \sum_{q=0}^\infty \frac{\lambda^q (\tau)_q \mathfrak{H}_{\beta+\nu\tau+\nu q}}{q!\,\Gamma(\alpha + \nu\tau + \nu q)}. \tag{24}$$

To insure that (14) is applicable $\mathrm{Re}\,\nu$ is taken to be positive, although negative values for which the series terminates may be admitted. By means of the formula

$$(\alpha)_{nm} = m^{nm} \prod_{n=0}^{m-1} \left(\frac{\alpha+n}{m}\right)_n, \tag{25}$$

where m is a positive integer, we obtain

$$\mathfrak{Z}_{\beta,\tau,m}(\alpha; \lambda) = {}_1F_{2m}\left(\tau; \tau + \frac{\alpha}{m}, \cdots, \tau + \frac{\alpha+m-1}{m}; \tau + \frac{\beta}{m}, \cdots, \tau + \frac{\beta+m-1}{m}; \frac{\lambda x^m}{m^{2m}}\right) \times \frac{\mathfrak{H}_{\beta+m\tau}}{\Gamma(\alpha+m\tau)}, \tag{26}$$

giving the connection with a hypergeometric function.

The differential equations satisfied by a distribution often determine its real personality; in the

case of $\mathfrak{Z}$ these equations are indeed interesting. Derivatives are given by

$$\mathfrak{Z}_\beta^{(n)} = \mathfrak{Z}_{\beta-n}. \tag{27}$$

Introducing the operator

$${}_{\alpha\beta}\mathfrak{d} = x(d^2/dx^2) + (1 + \alpha - \beta)(d/dx), \tag{28}$$

we obtain

$${}_{\alpha\beta}\mathfrak{d}^n\mathfrak{Z}_\beta(\alpha) = \mathfrak{Z}_{\beta-n}(\alpha - n) \tag{29}$$

as an easy consequence of (24) and

$${}_{\alpha\beta}\mathfrak{d}^n\mathfrak{H}_{\beta+\nu}/\Gamma(\alpha + \nu) = \mathfrak{H}_{\beta+\nu-n}/\Gamma(\alpha + \nu - n). \tag{30}$$

Furthermore, we have

$$({}_{\alpha\beta}\mathfrak{d}^m - \lambda)^l\mathfrak{Z}_{\beta,\tau,m}(\alpha;\lambda) = \mathfrak{Z}_{\beta,\tau-l,m}(\alpha;\lambda), \tag{31}$$

which leads to the equations

$$({}_{\alpha\beta}\mathfrak{d}^m - \lambda)^l\mathfrak{Z}_{\beta,l,m}(\alpha;\lambda) = \mathfrak{H}_\beta/\Gamma(\alpha), \tag{32}$$

and

$${}_{\alpha\beta}\mathfrak{d}^n({}_{\alpha\beta}\mathfrak{d}^m - \lambda)^l\mathfrak{Z}_{\beta+n,l,m}(\alpha + n;\lambda) = \mathfrak{H}_\beta/\Gamma(\alpha). \tag{33}$$

A homogeneous equation for $\mathfrak{Z}_{\beta,l,m}(\alpha;\lambda)$ may easily be obtained from (32) using

$$[x(d/dx) + 1 - \beta]\mathfrak{H}_\beta = 0. \tag{34}$$

Due to the poles of the gamma functions in (32) and (33) these equations are homogeneous whenever α is zero or a negative integer. That there are just ml independent $\mathfrak{Z}$ solutions to the homogeneous equation

$$({}_{\alpha\beta}\mathfrak{d}^m - \lambda^m)^l\psi = 0 \tag{35}$$

may be seen in the following manner:

Consider the solutions of

$$({}_{\alpha\beta}\mathfrak{d}^m - \lambda^m)\psi = 0. \tag{36}$$

It can be shown that all $\mathfrak{Z}$ solutions of this equation are of the form

$$\psi = \mathfrak{Z}_{\beta-\alpha-k,1,n}(-k;\lambda^n) \tag{37}$$

where n is a positive integer such that $m = na$ for some $a = 1, 2, 3, \cdots$, and k is a nonnegative integer. Not all of these solutions are independent. From (24) one may easily obtain the expansion

$$\mathfrak{Z}_{\beta,1,n}(\alpha;\lambda^n) = \sum_{q=0}^{m/n-1} \lambda^{m-nq-n}\mathfrak{Z}_{\beta-nq,1,m}(\alpha - nq;\lambda^m) \tag{38}$$

provided m/n is a positive integer. Further, the relation

$$\mathfrak{Z}_{\beta-mp-k,1,m}(-mp - k;\lambda^m) = \lambda^{mp}\mathfrak{Z}_{\beta-k,1,m}(-k;\lambda^m) \tag{39}$$

can be derived when p and k are positive integers. Equations (38) and (39) indicate that there are only m independent $\mathfrak{Z}$ solutions of (36); they may be taken in the form

$$\psi = \mathfrak{Z}_{\beta-\alpha-k,1,m}(-k;\lambda^m) \tag{40}$$

with $k = 0, 1, \cdots, m - 1$. Therefore, by virtue of the formula

$$(d^p/d\lambda^p)\mathfrak{Z}_{\beta,\tau,\nu}(\alpha;\lambda) = (\tau)_p\mathfrak{Z}_{\beta,\tau+p,\nu}(\alpha;\lambda), \tag{41}$$

we see that only ml independent $\mathfrak{Z}$ solutions of (35) can be constructed. Clearly these solutions may be given in the form

$$\psi = \mathfrak{Z}_{\beta-\alpha-k,p,m}(-k;\lambda^m) \tag{42}$$

where $k = 0, 1, \cdots, m - 1$ and $p = 1, 2, \cdots, l$.

It should be mentioned that our interest in the operator ${}_{\alpha\beta}\mathfrak{d}$ is motivated by the fact that it may be associated with the spherically symmetric d'Alembert operator

$$\Box = \frac{d^2}{ds^2} + \frac{n}{s}\frac{d}{ds} \qquad (s^2 = t^2 - r^2) \tag{43}$$

for n space dimensions. Transformed to a new independent variable $x = s^2/4$ the d'Alembert operator becomes

$$\Box = x\frac{d^2}{dx^2} + \frac{n+1}{2}\frac{d}{dx}, \tag{44}$$

which is a special case of (28). As we shall see, the $\mathfrak{Z}$ distributions are closely related to the Green's functions for Eq. (1).

The distribution obtained from Eq. (24) by omitting the gamma function in the denominator, that is, the distribution

$$\mathfrak{Z}_{\beta,\tau,\nu}(;\lambda) = \sum_{q=0}^{\infty} \frac{\lambda^q(\tau)_q}{q!}\mathfrak{H}_{\beta+\nu\tau+\nu q}, \tag{45}$$

no longer satisfies (29) and (31), but satisfies Eq. (27) and

$$(d^m - \lambda)^l\mathfrak{Z}_{\beta,\tau,m}(;\lambda) = \mathfrak{Z}_{\beta,\tau-l,m}(;\lambda) \tag{46}$$

where $d \equiv d/dx$. Equation (32) is similarly replaced by

$$(d^m - \lambda)^l\mathfrak{Z}_{\beta,l,m}(;\lambda) = \mathfrak{H}_\beta, \tag{47}$$

which is always inhomogeneous.

The connection with a hypergeometric series is given by

$$\mathfrak{Z}_{\beta,\tau,m}(;\lambda) = {}_1F_m\left(\tau; \tau + \frac{\beta}{m}, \cdots, \tau + \frac{\beta + m - 1}{m}; \frac{\lambda x^m}{m^m}\right)\mathfrak{H}_{\beta+m\tau}, \quad (48)$$

an equation similar to (26). Using the Laplace transform[4] of $\mathfrak{H}_\beta$,

$$\mathfrak{L}\{\mathfrak{H}_\beta\} = \mathfrak{H}_\beta\{e^{-sx}\} = z^{-\beta}, \quad (49)$$

the Laplace transform of $\mathfrak{Z}_{\beta,\tau,\nu}(;\lambda)$ is easily obtained; it is

$$\mathfrak{L}\{\mathfrak{Z}_{\beta,\tau,\nu}(;\lambda)\} = z^{-\beta}(z^\nu - \lambda)^{-\tau}. \quad (50)$$

Further generalizations are available in the distributions

$$\mathfrak{L}_{\beta,\tau,\nu,\mu}(\alpha;\lambda,\omega) = \sum_{q=0}^{\infty} \frac{\omega^q(\tau)_q}{q!}\mathfrak{Z}_{\beta,\nu\tau+\nu q,\mu}(\alpha;\lambda) \quad (51)$$

and

$$\mathfrak{L}_{\beta,\tau,\nu,\mu}(;\lambda,\omega) = \sum_{q=0}^{\infty} \frac{\omega^q(\tau)_q}{q!}\mathfrak{Z}_{\beta,\nu\tau+\nu q,\mu}(;\lambda), \quad (52)$$

where Re ν and Re μ are taken as positive. The following differential equations are satisfied

$$[({}_\alpha\delta^n - \lambda)^m - \omega]^l\mathfrak{L}_{\beta,l,m,n}(\alpha;\lambda,\omega) = \mathfrak{H}_\beta/\Gamma(\alpha), \quad (53)$$

$$[(d^n - \lambda)^m - \omega]^l\mathfrak{L}_{\beta,l,m,n}(;\lambda,\omega) = \mathfrak{H}_\beta. \quad (54)$$

A glance at the last equation indicates that $\mathfrak{L}_{\beta,l,m,n}(;\lambda,\omega)$ will be of use in obtaining $G_{l,m}(k,t)$, the solution of Eq. (10). The Laplace transform is given by

$$\mathfrak{L}\{\mathfrak{L}_{\beta,\tau,\nu,\mu}(;\lambda,\omega)\} = z^{-\beta}[(z^\mu - \lambda)^\nu - \omega]^{-\tau}. \quad (55)$$

At this point all the necessary distributions for calculating the Green's function $\Delta_{n,l,m}$ associated with Eq. (1) have been developed.

As we have seen, a direct calculation of the Green's function from its Fourier representation leads in general to an improper integral [cf. (8)] which must be interpreted in some generalized sense. In (I) a similar calculation led to the improper Sonine integral which was treated in terms of generalized functions with the result

$$\int_0^\infty du\Omega_\mu(0; x - a: u)\Omega_\beta(0; \lambda - u: a) = \Omega_{\beta-\mu-1}(0; \lambda: x), \quad (56)$$

where a is a positive real number. In this formula the complex values of β and μ are unrestricted from the standpoint of distribution theory even though the classical condition for convergence requires Re $\beta >$ Re $\mu > -1$. It would be nice if this result could be extended to permit an explicit evaluation of the integral occurring in Eq. (8).

As a matter of fact, various extensions are possible. From (21) and (24) we find

$$\Gamma(\tau)\mathfrak{Z}_{\beta,\tau,1}(0;\lambda) = \Omega_{\beta+\tau-1}(0;\lambda), \quad (57)$$

which, when applied to Eq. (51) with $\mu = 1$, indicates that $\mathfrak{L}_{\beta,\tau,\nu,1}(0;\lambda,\omega)$ is a series of Ω distributions. Making use of (56) inside the summation gives

$$\int_0^\infty du\Omega_\mu(0; x - a: u)\mathfrak{L}_{\beta,\tau,\nu,1}(0;\lambda - u,\omega: a) = \mathfrak{L}_{\beta-\mu-1,\tau,\nu,1}(0;\lambda,\omega: x). \quad (58)$$

In a similar manner using

$$\Gamma(\tau)\mathfrak{Z}_{0,\tau,2}(;\lambda: 2x^{1/2}) = (\pi)^{1/2}\Omega_{\tau-1/2}(0;\lambda: x), \quad (59)$$

we obtain

$$\int_0^\infty du\Omega_\mu(0; x - a: u)\mathfrak{L}_{0,\tau,\nu,2}(;\lambda - u,\omega: 2a^{1/2}) = (\pi)^{1/2}\mathfrak{L}_{-\mu-1/2,\tau,\nu,1}(0;\lambda,\omega: x). \quad (60)$$

Equation (60) becomes, when $\lambda = 0$,

$$\int_0^\infty du\Omega_\mu(0; x - a: u)\mathfrak{L}_{0,\tau,\nu,2}(;-u,\omega: 2a^{1/2}) = (\pi)^{1/2}\mathfrak{Z}_{-\mu-1/2,\tau,\nu}(0;\omega: x). \quad (61)$$

This is the result we are after; using it the Green's function $\Delta_{n,l,m}$ may be calculated explicitly.

3. DISCUSSION OF THE GREEN'S FUNCTION

Consider now the homogeneous Green's function $\Delta_{n,l,m}(x)$ associated with the equation

$$[\Box^m - (-\mu^2)^m]^l\varphi = 0. \quad (62)$$

In order to apply (8) we must of course know the solution $G_{l,m}(k,t)$ of

$$\left[\left(\frac{d^2}{dt^2} + k^2\right)^m - (-\mu^2)^m\right]^l G(t) = \delta(t) \quad (63)$$

corresponding to (9). Evidently it follows from (54) and (55) that

$$G_{l,m}(k,t) = \mathfrak{L}_{0,l,m,2}[; -k^2, (-\mu^2)^m: t]. \quad (64)$$

The Bessel function in the integrand of (9) can

be expressed in terms of the Ω function by means of the formula[1]

$$\Omega_\beta(0; -\lambda : x) = h(x)(x/\lambda)^{\beta/2} J_\beta(2\lambda^{1/2}x^{1/2}) \tag{65}$$

where $h(x) = 1$ for $x > 0$ and $h(x) = 0$ for $x < 0$. Upon changing the variable of integration to $u = k^2$ the Green's function may be written in the form

$$\Delta_{n\ l,m}(\mathbf{x}, t) = \frac{\epsilon(t)}{2^n \pi^{n/2}} \int_0^\infty du \Omega_{n/2-1}(0; -r^2/4 : u) \times \mathfrak{L}_{0,l,m,2}[; -u, (-\mu^2)^m : t], \tag{66}$$

and in view of (61) we have immediately

$$\Delta_{n,l,m}(x) = \frac{\epsilon(t)}{2^n \pi^{(n-1)/2}} \mathfrak{Z}_{(1-n)/2,l,m}[0; (-\mu^2)^m : s^2/4]. \tag{67}$$

In terms of familiar hypergeometric functions, Eq. (67) becomes

$$2^{2ml-1}\pi^{(n-1)/2}\Gamma(ml)\Gamma\left(\frac{1-n}{2} + ml\right) \Delta_{n,l,m}(x) = h(s^2)s^{2ml-n-1}\ {}_1F_{2m}\left(l; l, l + \frac{1}{m}, \cdots, l + \frac{m-1}{m}, l + \frac{1-n}{2m}, \cdots, l + \frac{1-n}{2m} + \frac{m-1}{m}; \left(\frac{i\mu s}{2m}\right)^{2m}\right) \tag{68}$$

by virtue of Eq. (26). However the series expansion of (67) is often more useful. Recalling (24) one easily obtains

$$\Delta_{n,l,m}(x) = \frac{\epsilon(t)}{2^n \pi^{(n-1)/2}} \sum_{q=0}^\infty \frac{(-\mu^2)^{mq}(l)_q}{q!\ \Gamma(ml + mq)} \times \mathfrak{H}_{(1-n)/2+ml+mq}(s^2/4) = \frac{\epsilon(t)h(s^2)s^{2ml-n-1}}{2^{2ml-1}\pi^{(n-1)/2}} \times \sum_{q=0}^\infty \frac{(-\mu^2/4)^{mq}(l)_q s^{2mq}}{q!\ \Gamma(ml + mq)\Gamma((1-n)/2 + ml + mq)}, \tag{69}$$

from which it is obvious that there are no singularities when $2ml > n$ (remembering that l, m, n, are integers). *That is, singularities do not appear on the light cone when the order of the differential equation is greater than the number of space dimensions.*

When $2ml \leq n$ the Green's function may be written as the sum of a singular part and a regular part. The singular part strikingly exemplifies the well-known difference between even- and odd-dimensional spaces. The regular part represents a finite jump discontinuity across the light cone and is of similar nature in the two cases.

The singular part is readily split off using the series form (69) of the Green's function. For odd-dimensional spaces the singularities appearing are

$$\Delta^{(\text{sing})}_{n,l,m} = \frac{\epsilon(t)(-\mu^2)^{ma-ml}}{2^{2am-1}\pi^{am+c}\Gamma(l)} \times \sum_{q=0}^{a-l} \frac{(-\mu^2/4)^{-mq}\Gamma(a-q)}{(a-l-q)!\ \Gamma(ma-mq)}\ \delta^{(mq+c)}(s^2), \tag{70}$$

where $n = 2am + 2c + 1$ with a, c integers such that $a \geq l$ and $0 \leq c \leq m - 1$. Thus for odd n the singular part consists of a finite linear combination of derivatives of the Dirac delta function $\delta(s^2)$, the highest derivative appearing being of order $\frac{1}{2}(n - 2ml - 1)$.

Spaces with an even number of dimensions have the following singularities in the Green's function:

$$\Delta^{(\text{sing})}_{n,l,m} = \frac{\epsilon(t)(-\mu^2)^{ma-ml}}{2^{2am-1}\pi^{am+c-1/2}\Gamma(l)} \times \sum_{q=0}^{a-l} \frac{(-\mu^2/4)^{-mq}\Gamma(a-q)h(s^2)(1/s)^{2mq+2c+1}}{(a-l-q)!\ \Gamma(ma-mq)\Gamma(\frac{1}{2}-c-mq)}, \tag{71}$$

where $n = 2am + 2c$. Here we have a polynomial in $1/s$ of degree $n - 2ml + 1$.

Various recurrence relations are satisfied by the Green's function $\Delta_{n,l,m}(x)$ and may be obtained using Eq. (67) along with the properties of the $\mathfrak{Z}$ distribution. We have

$$\Delta_{n+2p,l,m} = (1/\pi^p)(d/ds^2)^p\ \Delta_{n,l,m} \tag{72}$$

as a consequence of (27), and recalling (41) we obtain

$$\Delta_{n,l+p,m} = ((-)^{mp}\Gamma(l)/\Gamma(l+p))(d/d\mu^{2m})^p\ \Delta_{n,l,m}. \tag{73}$$

Furthermore, the "method of descent" used by Hadamard[5] finds expression in [cf. Appendix in (I)]

$$\Delta_{n-p,l,m} = \int_{-\infty}^{\infty} \Delta_{n,l,m}\, dx_1 \cdots dx_p. \tag{74}$$

As a result of these recurrence formulas all the Green's functions $\Delta_{n,l,m}$ are derivable from $\Delta_{0,1,m}$.

So far we have only l independent Δ-function solutions to Eq. (1); they are $\Delta_{n,p,m}(x)$ with $p = 1, 2, \cdots, l$. However the integrand of the Fourier integral representing the Green's function contains $2m$ poles of order l in the k_0 plane corresponding to the $2m$ roots of $(-k^2)^m = (-\mu^2)^m$. One expects then ml independent Δ-function solutions. In fact

[5] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Yale University Press, New Haven, Connecticut, 1923).

we can write them down immediately in view of (42); they are

$$\Delta^b_{n,p,m}(x) = \frac{\epsilon(t)}{2^n \pi^{(n-1)/2}} \mathfrak{Z}_{(1-n)/2-b,p,m}[-b; (-\mu^2)^m : s^2/4] \quad (75)$$

where $p = 1, 2, \cdots, l$ and $b = 0, 1, \cdots, (m-1)$. Series expansions and recurrence formulas for $\Delta^b_{n,p,m}$ similar to those for $\Delta_{n,l,m}$ may easily be derived.

The initial conditions satisfied by the Green's functions $\Delta^b_{n,p,m}$ at $t = 0$ may be obtained as follows. Consider the Δ-function solutions of the equation $\Box^p \varphi = 0$; they can be given in the form[1]

$$D_{n,p}(x) = \epsilon(t) \mathfrak{P}_{(1-n)/2+p}(s^2/4)/2^n \pi^{(n-1)/2} \Gamma(p), \quad (76)$$

where the notation D means $\lim_{\mu \to 0} \Delta$. The series expansion of (75) may thus be written[6]

$$\Delta^b_{n,p,m} = \sum_{q=0}^{\infty} \frac{(-\mu^2)^{mq}(p)_q}{q!} D_{n,mp+mq-b}. \quad (77)$$

Now the initial conditions satisfied by $D_{n,p}$ are[1]

$$\partial_0^j D_{n,p}(\mathbf{x}, 0) = 0 \quad \text{for} \quad j = 0, 1, \cdots, (2p-2)$$
$$= \delta(\mathbf{x}) \quad \text{for} \quad j = 2p-1$$
$$= 0 \quad \text{for all even} \quad j, \quad (78)$$

where $\partial_0 = \partial/\partial t$. Therefore the initial conditions satisfied by $\Delta^b_{n,p,m}$ are

$$\partial_0^j \Delta^b_{n,p,m}(\mathbf{x}, 0)$$
$$= 0 \quad \text{for} \quad j = 0, 1, \cdots, (2mp - 2b - 2)$$
$$= \delta(\mathbf{x}) \quad \text{for} \quad j = 2mp - 2b - 1$$
$$= 0 \quad \text{for all even} \quad j. \quad (79)$$

These initial conditions indicate that all ml solutions (75) are needed in the general solution of the Cauchy initial value problem[7] for Eq. (1).

ACKNOWLEDGMENT

This work was supported in part by U. S. Public Health Service Training Grant 2G-174 to Dartmouth Medical School.

[6] Equation (77) along with $\Box^j D_{n,p} = D_{n,p-j}$ may be used to show that $\Delta^b_{n,p,m}$ is indeed a solution of (62).

[7] See (I) where the complete solution of the Cauchy problem for the iterated Klein-Gordon operator is calculated in detail. The general solution for an arbitrary polynomial in $\Box$ appears in J. J. Bowman and J. D. Harris, J. Math. Phys. 3, 1281 (1962), preceding article.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Input Admittance of Infinitely Long Dipole Antennas Driven from Coaxial Lines

TAI TSUN WU*
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received April 30, 1962)

For an infinitely long dipole antenna driven from a coaxial line, the reflection coefficient and hence the apparent terminal admittance are determined approximately when the radii of the coaxial line are small compared with the wavelength. This result is useful because the differences of admittances for antennas with identical geometries near the driving point are less sensitive to the driving condition and hence can be found approximately by many existing theories.

I. INTRODUCTION

BECAUSE of the geometrical complexity of a linear antenna driven from either a coaxial line or a two-wire transmission line, a number of antenna problems have been studied on the basis of a delta-function generator, sometimes called a slice generator. Such an idealization is adequate for many purposes, such as the determination of the field pattern or the current distribution not close to the generator, but causes difficulty in connection with the input admittance because of a singularity in the current distribution at the driving point.[1] Various methods have been suggested to deal with this singularity, such as subtracting out a logarithmic term. It is the purpose of this paper to consider an alternative procedure, which is more natural and hence more satisfactory so far as the author is concerned. Only the case of the antenna driven from a coaxial line is to be studied.

When a delta-function generator is postulated and when the transverse dimensions of the antenna are small, the abovementioned singularity in the current distribution is a local effect in the sense that it is independent of the logitudinal dimensions such as the length of the antenna. In other words, differences of current distributions on antennas with the same transverse geometry are bounded in the vicinity of the driving point. Accordingly, the approximate determination of the differences of input admittances is comparatively less troublesome. For the purpose of obtaining finite input admittances that can be directly compared with experimental results, it is sufficient to study just one canonical problem once the differences of admittances are known. This paper is concerned exclusively with the study of such a canonical case.

The canonical case chosen is that of an infinite straight antenna driven from a coaxial line. This case is by far the simplest choice, and the geometry is shown in Fig. 1. Cylindrical coordinates (r, θ, z) are to be used.

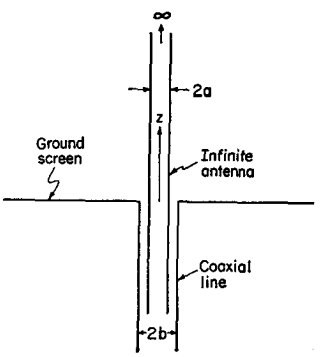


FIG. 1. Geometry of the problem.

This case has been studied by Papadopoulos.[2] However, the approximation in his work is such that it can give at best a qualitative understanding of the situation. The present paper seeks to provide a procedure to obtain quantitatively useful results in the sense that they are to be comparable in accuracy with available experimental measurements. For this purpose all terms of the order kb, which were neglected by Papadopoulos, must be retained since they can be estimated to be numerically quite significant in many cases. In the present solution the input admittance, which is a function of the two dimensionless constants ka and b/a, is found to have a surprisingly simple form as given by (29).

2. FORMULATION OF THE PROBLEM IN TERMS OF AN INTEGRAL EQUATION

Let the incident current on the inner conductor of the coaxial line be

$$I^{\text{inc}}(z) = e^{ikz}. \tag{1}$$

Since rotational symmetry obtains, let $E_r(r, z)$,

* Alfred P. Sloan Foundation Fellow. Work also supported in part by Grant G-9721 from the National Science Foundation.

[1] T. T. Wu and R. W. P. King, J. Appl. Phys. 30, 76 (1959).

[2] V. M. Papadopoulos, Quart. Appl. Math. 17, 423 (1960).

$E_r(r, z)$, and $H_\theta(r, z)$ be the nonvanishing components of the electromagnetic field due to (1). In terms of $E_r(r, \theta)$, the magnetic field H_θ is given by

$$H_\theta(r, z) = -2ik\zeta_0^{-1} \int_a^b r' \, dr' E_r(r', 0) G_0(r, r', z) \quad (2a)$$

for $z \geq 0$ and $r \geq a$, and

$$H_\theta(r, z) = (\pi r)^{-1} \cos kz + 2ik\zeta_0^{-1} \times \int_a^b r' \, dr' E_r(r', 0) G_e(r, r', z) \quad (2b)$$

for $z \leq 0$ and $a \leq r \leq b$, provided that the antenna, the coaxial line, and the ground plane are all assumed to be perfectly conducting. In (2), ζ_0 is the characteristic impedance of free space, while the Fourier transforms of the Green's functions G_0 and G_e are given by

$$\tilde{G}_0(r, r', \zeta) = \int_{-\infty}^{\infty} dz G_0(r, r', z) e^{-i\zeta z}$$
$$= \tfrac{1}{2} i\pi [H_0^{(1)}(\xi a)]^{-1} H_1^{(1)}(\xi r_>) \times [H_0^{(1)}(\xi a) J_1(\xi r_<) - J_0(\xi a) H_1^{(1)}(\xi_{4<})], \quad (3a)$$

and

$$\tilde{G}_e(r, r', \zeta) = \int_{-\infty}^{\infty} dz G_e(r, r', z) e^{-i\zeta z}$$
$$= \tfrac{1}{2} i\pi [J_0(\xi a) H_0^{(1)}(\xi b) - H_0^{(1)}(\xi a) J_0(\xi b)]^{-1} \times [H_0^{(1)}(\xi b) J_1(\xi r_>) - J_0(\xi b) H_1^{(1)}(\xi r_>)] \times [H_0^{(1)}(\xi a) J_1(\xi r_<) - J_0(\xi a) H_1^{(1)}(\xi r)_<], \quad (3b)$$

where $r_>$ and $r_<$ are, respectively, the larger one and the smaller one of r and r', and

$$\xi = (k^2 - \zeta^2)^{1/2} \quad (4)$$

is defined by $\xi = k$ for $\zeta = 0$ and by the branch cuts shown in Fig. 2. The continuity of H_θ for $z = 0$ and $a \leq r \leq b$ gives the following equation for E_r:

$$\int_a^b r' \, dr' E_r(r', 0) [G_0(r, r', 0) + G_e(r, r'0)] = i\zeta_0 (2\pi r k)^{-1}. \quad (5)$$

This is the desired integral equation.

It follows from (3b) that, as $\xi \to 0$,

$$\tilde{G}_e(r, r', \zeta) \sim -(\xi^2 r r')^{-1} [\ln (b/a)]^{-1}. \quad (6)$$

Accordingly, as $z \to -\infty$,

$$G_e(r, r', z) \sim \tfrac{1}{2} i (r r' k)^{-1} [\ln (b/a)]^{-1} e^{-ikz}, \quad (7)$$

and hence

$$rH_\theta(r, z) \sim \pi^{-1} \cos kz - \zeta_0^{-1} e^{-ikz} [\ln (b/a)]^{-1} \times \int_a^b dr' E_r(r', 0), \quad (8)$$

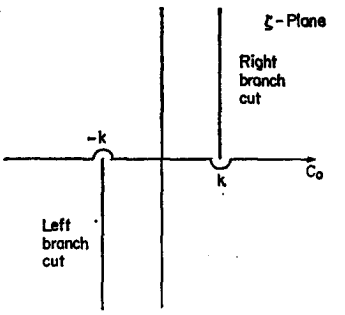


FIG. 2. The ζ plane and the contour C_0.

provided that the coaxial line does not support any higher mode. The reflection coefficient is thus

$$\Gamma = 1 - 2\pi\zeta_0^{-1} [\ln (b/a)]^{-1} \int_a^b dr' E_r(r', 0), \quad (9)$$

and the apparent terminal admittance for this infinitely long antenna, defined by

$$Y_{a\infty} = 2\pi [\zeta_0 \ln (b/a)]^{-1} (1 + \Gamma)/(1 - \Gamma),$$

is given by

$$Y_{a\infty} = -2\pi [\zeta_0 \ln (b/a)]^{-1} + 2\left[\int_a^b dr' E_r(r', 0)\right]^{-1}. \quad (10)$$

3. APPROXIMATIONS FOR SMALL b

From here on the assumption is made that

$$kb \ll 1. \quad (11)$$

In order to make use of (11), note first that

$$\int_a^b \tilde{G}_e(r, a, \zeta) \, dr = -(\xi^2 a)^{-1}, \quad (12)$$

and similarly

$$\int_a^b \tilde{G}_0(r, a, \zeta) \, dr = -(\xi^2 a)^{-1} [1 - H_0^{(1)}(\xi b)/H_0^{(1)}(\xi a)]. \quad (13)$$

On the basis of (12) and (13), define

$$G(r, r'; k) = G_0(r, r', 0) + G_e(r, r', 0), \quad (14)$$

$$G_d(r, r'; k) = G(r, r'; k) - \tau/(r r'), \quad (15)$$

and

$$G_d(r, r') = G_d(r, r'; 0), \quad (16)$$

where

$$\tau = -[2\pi \ln (b/a)]^{-1} \times \int_{C_0} \xi^{-2} \, d\zeta [2 - H_0^{(1)}(\xi b)/H_0^{(1)}(\xi a)], \quad (17)$$

with the contour of integration C_0 shown in Fig. 2.

When (11) holds, $G_d(r, r'; k)$ differs only slightly from $G_d(r, r')$. Thus (5) becomes approximately

$$r\int_a^b r'\,dr'E_r(r', 0)G_d(r, r') = i\zeta_0(2\pi k)^{-1} - \tau\int_a^b dr'E_r(r', 0). \quad (18)$$

If $f(r)$ is defined as the solution of the integral equation

$$r\int_a^b r'\,dr'f(r')G_d(r, r') = 1, \quad (19)$$

and

$$A = a\int_a^b drf(r), \quad (20)$$

then

$$E_r(r, 0) = i\zeta_0(2\pi k)^{-1}f(r)[1 + A\tau/a]^{-1}, \quad (21)$$

and

$$Y_{a\infty} = 2\pi\zeta_0^{-1}\{-2ik(aA^{-1} + \tau) - [\ln (b/a)]^{-1}\}. \quad (22)$$

Note that A is a function of b/a only.

It only remains to write down the kernel G_d of the integral equation explicitly. This is provided by

$$G_d(r, r') = \pi^{-1}\int_0^\infty d\zeta\Big\{[K_0(\zeta a)I_1(\zeta r_<) + I_0(\zeta a)K_1(\zeta r_<)] \times \left[\frac{K_1(\zeta r_>)}{K_0(\zeta a)} - \frac{K_0(\zeta b)I_1(\zeta r_>) + I_0(\zeta b)K_1(\zeta r_>)}{I_0(\zeta a)K_0(\zeta b) - K_0(\zeta a)I_0(\zeta b)}\right] - (\zeta^2 rr')^{-1}[2 - K_0(\zeta b)/K_0(\zeta a)]/[\ln (b/a)]\Big\}. \quad (23)$$

4. APPROXIMATE EVALUATION OF $Y_{a\infty}$

Equation (17) may be further simplified under the assumption (11). First of all, it can be written alternatively as

$$\tau = [\pi \ln (b/a)]^{-1}\int_{-ik}^\infty \eta^{-1}\,d\eta(k^2 + \eta^2)^{-1/2} \times [2 - K_0(\eta b)/K_0(\eta a)], \quad (24)$$

where the path of integration is taken in the fourth quadrant. When η is small, the last factor in the integrand is given approximately by

$$2 - K_0(\eta b)/K_0(\eta a) \sim 1 + \eta aK_1(\eta a)\ln (b/a)/K_0(\eta a). \quad (25)$$

Thus, by (24), τ may be approximately split into two parts as follows:

$$2ik\tau + [\ln (b/a)]^{-1} = -B_1(ka) + kaiB_2(b/a), \quad (26)$$

where

$$B_1(ka) = 2ika\pi^{-1}\Big\{\int_{-ik}^\infty d\eta(k^2 + \eta^2)^{-1/2} \times [1 - K_1(\eta a)/K_0(\eta a)] + \ln (ka) - i\tfrac{1}{2}\pi\Big\}, \quad (27)$$

and

$$B_2(b/a) = -(2/\pi) \times \lim_{D\to\infty}\Big\{\ln (2D) + [a\ln (b/a)]^{-1}\int_0^{D/a}\eta^{-2}\,d\eta \times \left[1 - \frac{K_0(\eta b)}{K_0(\eta a)} - \eta a\left(\ln\frac{b}{a}\right)\frac{K_1(\eta a)}{K_0(\eta a)}\right]\Big\}. \quad (28)$$

Finally, the substitution of (26) into (22) gives

$$Y_{a\infty} = 2\pi\zeta_0^{-1}\{B_1(ka) - kai[2A^{-1}(b/a) + B_2(b/a)]\}. \quad (29)$$

This is the required answer.

Note that the B_1 term does not depend on b and hence may be interpreted as the intrinsic admittance of the antenna, while the terms B_2 and A^{-1} are independent of k and may be interpreted together as a capacitive end correction. As seen from the first term on the right-hand side of (28), this splitting is arbitrary up to a constant in the capacitance.

5. CONCLUSIONS

When the radius of the outer conductor of the coaxial line is small compared with the wavelength, as expressed by (11), the apparent terminal admittance of an infinitely long dipole antenna driven from a coaxial line has been found quite accurately. It is, however, not trivial to obtain numerical results, since the solution of an integral equation (19) is needed. It is perhaps worth reemphasizing that the admittance has a rather simple form (29) as a function of ka and b/a.

Many approximate theories may be expected to give differences of admittances more accurately than the admittances themselves because of the complexity of the current distribution near the driving point. Accordingly, when a coaxial line is used, the present result supplements these theories. More precisely, it is here proposed that the apparent admittance Y_a of a dipole antenna with length $h \gg b$ may be found approximately as follows when driven from a coaxial line. Express Y_a in the form

$$Y_a = (Y_a - Y_{a\infty}) + Y_{a\infty}. \quad (30)$$

The term $Y_{a\infty}$ is given by the present theory, while the difference $Y_a - Y_{a\infty}$ is found on the basis of a

delta-function generator, as mentioned in the introduction. This procedure, in particular, takes account properly of the effects of the size of the coaxial line. In order to use this procedure, the approximate theory used to evaluate the difference $Y_a - Y_{a\infty}$ must be valid for the infinitely long dipole. In other words, a theory of the long dipole antenna is needed. Furthermore, for comparison with experimental results, this theory of the long dipole antenna must also be reasonably accurate even when the antenna is not excessively long. A theory with precisely this purpose in mind has been given previously.[3]

This point of view is also useful in some cases more complicated than that of the dipole antenna. For example, it is applicable to the case of the thin circular loop antenna provided that the radius of the loop is large compared with b. If (11) is replaced by $|kb| \ll 1$, this entire calculation is also valid for dissipative media, where k is complex.

Procedures very similar to the present one can easily be found for various other problems, such as a discontinuity in radius.

ACKNOWLEDGMENT

For very helpful discussions, I am indebted to Professor R. W. P. King.

[3] T. T. Wu, J. Math. Phys. 2, 550 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Theory of the Thin Circular Loop Antenna

TAI TSUN WU*
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received April 30, 1962)

The current distribution on a thin circular loop transmitting antenna driven by a delta-function generator is determined approximately by Fourier series expansion. A difficulty encountered in previous analysis is shown to be due to an inadequate approximation.

I. INTRODUCTION

As early as 1897, Pocklington[1] studied the excitation of a thin loop antenna by a plane wave. Using methods very similar to that of Pocklington, Hallén[2] and later Storer[3] considered the case of the driven antenna. All these authors used Fourier series expansion, as appropriate for the geometry under consideration, and the latter authors found a difficulty in this approach. Their difficulty takes the form of the appearance of either a singularity or a very large term in the Fourier series expansion when the index n is close to a certain large number determined by the geometry. Hallén then concluded that the series is divergent and attributed this divergence to the approximation of a "one-dimensional" equation, whereas Storer avoided the contribution from this large term by first replacing the Fourier series by the corresponding integral and then evaluating the integral in the sense of the Cauchy principle value. This procedure of Storer seems at best to be of doubtful validity. More recently, a similar difficulty has been found to appear also in the case of dipole antennas.[4,5] In this case, however, it is clear from the derivation that the trouble has nothing to do with the originally posed problem, but instead, is a consequence of the approximations used. It is the purpose of this paper to point out that this is also the case with the thin loop antenna, and a procedure is proposed that avoids this difficulty by invoking approximations that are valid over larger ranges of the parameters. It may be added that the difficulty under consideration does not have anything to do with the so-called "gap problem."

* Alfred P. Sloan Foundation Fellow. Work also supported in part by National Science Foundation Grant 9721.

[1] H. C. Pocklington, Proc. Cambridge Phil. Soc. 9, 324 (1897).

[2] E. Hallén, Nova Acta Regiae Soc. Sci. Upsaliensis 2, No. 4 (1938).

[3] J. E. Storer, Trans. A. I. E. E. 75, Part I, 606 (1956).

[4] T. T. Wu, J. Math. Phys. 2, 550 (1961).

[5] T. T. Wu, J. Research Natl. Bur. Standards (to be published).

delta-function generator, as mentioned in the introduction. This procedure, in particular, takes account properly of the effects of the size of the coaxial line. In order to use this procedure, the approximate theory used to evaluate the difference $Y_a - Y_{a\infty}$ must be valid for the infinitely long dipole. In other words, a theory of the long dipole antenna is needed. Furthermore, for comparison with experimental results, this theory of the long dipole antenna must also be reasonably accurate even when the antenna is not excessively long. A theory with precisely this purpose in mind has been given previously.[3]

This point of view is also useful in some cases more complicated than that of the dipole antenna. For example, it is applicable to the case of the thin circular loop antenna provided that the radius of the loop is large compared with b. If (11) is replaced by $|kb| \ll 1$, this entire calculation is also valid for dissipative media, where k is complex.

Procedures very similar to the present one can easily be found for various other problems, such as a discontinuity in radius.

ACKNOWLEDGMENT

For very helpful discussions, I am indebted to Professor R. W. P. King.

[3] T. T. Wu, J. Math. Phys. 2, 550 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 6 NOVEMBER-DECEMBER 1962

Theory of the Thin Circular Loop Antenna

TAI TSUN WU*
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received April 30, 1962)

The current distribution on a thin circular loop transmitting antenna driven by a delta-function generator is determined approximately by Fourier series expansion. A difficulty encountered in previous analysis is shown to be due to an inadequate approximation.

I. INTRODUCTION

As early as 1897, Pocklington[1] studied the excitation of a thin loop antenna by a plane wave. Using methods very similar to that of Pocklington, Hallén[2] and later Storer[3] considered the case of the driven antenna. All these authors used Fourier series expansion, as appropriate for the geometry under consideration, and the latter authors found a difficulty in this approach. Their difficulty takes the form of the appearance of either a singularity or a very large term in the Fourier series expansion when the index n is close to a certain large number determined by the geometry. Hallén then concluded that the series is divergent and attributed this divergence to the approximation of a "one-dimensional" equation, whereas Storer avoided the contribution from this large term by first replacing the Fourier series by the corresponding integral and then evaluating the integral in the sense of the Cauchy principle value. This procedure of Storer seems at best to be of doubtful validity. More recently, a similar difficulty has been found to appear also in the case of dipole antennas.[4,5] In this case, however, it is clear from the derivation that the trouble has nothing to do with the originally posed problem, but instead, is a consequence of the approximations used. It is the purpose of this paper to point out that this is also the case with the thin loop antenna, and a procedure is proposed that avoids this difficulty by invoking approximations that are valid over larger ranges of the parameters. It may be added that the difficulty under consideration does not have anything to do with the so-called "gap problem."

* Alfred P. Sloan Foundation Fellow. Work also supported in part by National Science Foundation Grant 9721.

[1] H. C. Pocklington, Proc. Cambridge Phil. Soc. 9, 324 (1897).

[2] E. Hallén, Nova Acta Regiae Soc. Sci. Upsaliensis 2, No. 4 (1938).

[3] J. E. Storer, Trans. A. I. E. E. 75, Part I, 606 (1956).

[4] T. T. Wu, J. Math. Phys. 2, 550 (1961).

[5] T. T. Wu, J. Research Natl. Bur. Standards (to be published).

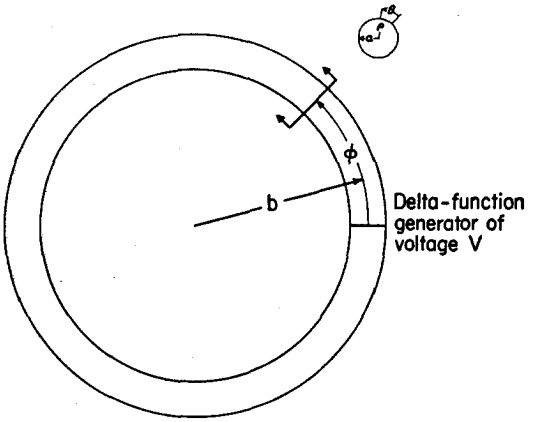


FIG. 1. Geometry of the problem.

2. FORMULATION OF THE PROBLEM

In Fig. 1 is shown the geometry and the coordinate system for a loop antenna driven by a delta-function generator. If the loop is assumed to be perfectly conducting, then there are two components for the surface current density, namely, $J_\phi(\phi, \theta)$ and $J_\theta(\phi, \theta)$. By setting the tangential components of the electric field to zero in the surface of the loop, the following integral equations are obtained for J_ϕ and J_θ:

$$(\partial/\partial\phi)\int d\phi'\, d\theta'(4\pi R)^{-1}e^{ikR}\{a(\partial/\partial\phi')J_\phi(\phi', \theta') + (\partial/\partial\theta')[(b + a\cos\theta')J_\theta(\phi', \theta')]\} + k^2a(b + a\cos\theta) \times \int d\phi'\, d\theta'(4\pi R)^{-1}e^{ikR}(b + a\cos\theta') \times [J_\phi(\phi', \theta')\cos(\phi - \phi') - J_\theta(\phi', \theta')\sin(\phi - \phi')\sin\theta'] = ik\zeta_0^{-1}V\,\delta(\phi) \tag{1a}$$

and

$$(\partial/\partial\theta)\int d\phi'\, d\theta'(4\pi R)^{-1}e^{ikR}\{a(\partial/\partial\phi')J_\phi(\phi', \theta') + (\partial/\partial\theta')[(b + a\cos\theta')J_\theta(\phi', \theta')]\} + k^2a^2\int d\phi'\, d\theta'(4\pi R)^{-1}e^{ikR}(b + a\cos\theta') \times \{J_\phi(\phi', \theta')\sin(\phi - \phi')\sin\theta + J_\theta(\phi', \theta') \times [\cos(\phi - \phi')\sin\theta\sin\theta' + \cos\theta\cos\theta']\} = 0. \tag{1b}$$

Here, all integrals are from $-\pi$ to π, R is the Euclidean distance between the points (ϕ, a, θ) and (ϕ', a, θ'), ζ_0 is the characteristic impedance of free space, and V is the voltage supplied by the delta-function generator.

As a first approximation to these rather complicated integral equations, the so-called one-dimensional equation is obtained under the following circumstances

$$a \ll b, \quad \text{and} \quad ka \ll 1 \tag{2a}$$

so that

$$J_\phi(\phi, \theta) \sim (2\pi a)^{-1}I(\phi), \tag{2b}$$

and

$$J_\theta(\phi, \theta) \sim 0. \tag{2c}$$

Equation (1b) is omitted altogether and (1a) is approximated by

$$(\partial/\partial\phi)\int d\phi' K(\phi - \phi')(\partial/\partial\phi')I(\phi') + k^2b^2\int d\phi' K(\phi - \phi')\cos(\phi - \phi')I(\phi') = 4\pi ik\zeta_0^{-1}V\,\delta(\phi), \tag{3}$$

where

$$K(\phi) = (2\pi)^{-1}\int_{-\pi}^{\pi} d\theta[(2b\sin\tfrac{1}{2}\phi)^2 + (2a\sin\tfrac{1}{2}\theta)^2]^{-1/2} \times \exp\{ik[(2b\sin\tfrac{1}{2}\phi)^2 + (2a\sin\tfrac{1}{2}\theta)^2]^{1/2}\}. \tag{4}$$

Equation (3) is to be studied by Fourier series expansion.

The integral equation used previously differs from the present one in that $K(\phi)$ is replaced by

$$\bar{K}(\phi) = [(2b\sin\tfrac{1}{2}\phi)^2 + a^2]^{-1/2} \times \exp\{ik[(2b\sin\tfrac{1}{2}\phi)^2 + a^2]^{1/2}\}. \tag{4'}$$

This difference has the following consequence. When Fourier series expansion is used, as given by (5) and (6) below, it is easily verified that the series for $I(\phi)$ converges for $\phi \neq 0$. If $\bar{K}$ had been used, the corresponding Fourier coefficients κ_n decreases exponentially for large n, and consequently the series for $I(\phi)$ diverges everywhere. In the work of Hallén and Storer, this violent divergence is avoided by an approximation on the coefficients κ_n which does not decrease exponentially for large n. Equation (4) is quite similar to the expression of the kernel in the case of the dipole antenna.[6]

[6] T. T. Wu and R. W. P. King, J. Appl. Phys. 30, 76 (1959).

3. FOURIER SERIES EXPANSION

Let

$$I(\phi) = \sum_{n=-\infty}^{\infty} I_n e^{in\phi}, \tag{5}$$

and

$$K(\phi) = \sum_{n=-\infty}^{\infty} \kappa_n e^{in\phi}, \tag{6}$$

then

$$I_n = ikV\pi^{-1}\zeta_0^{-1}[\tfrac{1}{2}k^2b^2(\kappa_{n+1} + \kappa_{n-1}) - n^2\kappa_n]^{-1}. \tag{7}$$

Thus, the major task here is to evaluate the coefficients κ_n.

If

$$M_n(A) = \int_{-\pi}^{\pi} d\phi e^{-in\phi}[(2b \sin \tfrac{1}{2}\phi)^2 + A^2]^{-1/2} \times \exp\{ik[(2b \sin \tfrac{1}{2}\phi)^2 + A^2]^{1/2}\} \tag{8}$$

for $n \geq 0$, then

$$\kappa_n = \pi^{-2} \int_0^{2a} dA(4a^2 - A^2)^{-1/2} M_{|n|}(A). \tag{9}$$

Thus, it is sufficient to consider the case $0 < A \ll b$. If

$$A \ll b/n, \tag{10}$$

then the analysis of Oseen[7] and Storer[3] applies with the result

$$bM_n(A) = 2 \ln (8b/A) - 4 \sum_{m=0}^{n-1} (2m + 1)^{-1} - \pi \int_0^{2kb} dx[\Omega_{2n}(x) - iJ_{2n}(x)] \tag{11}$$

approximately, where Ω is the Lommel-Weber function

$$\Omega_m(x) = \pi^{-1} \int_0^{\pi} \sin (x \sin \theta - m\theta)\, d\theta. \tag{12}$$

In particular, (11) implies that

$$bM_0(A) = 2 \ln (8b/A) - \pi \int_0^{2kb} dx[\Omega_0(x) - iJ_0(x)]. \tag{13}$$

On the other hand, the integral

$$N_n(A) = \int_{-\infty}^{\infty} d\phi e^{-in\phi}[b^2\phi^2 + A^2]^{-1/2} \times \exp\{ik[b^2\phi^2 + A^2]^{-1/2}\} \tag{14}$$

may be readily computed to be given by

$$bN_n(A) = \pi i H_0^{(1)}[(k^2 - n^2/b^2)^{1/2}A]. \tag{15}$$

When (10) holds, (15) is approximated by

$$bN_n(A) = -2\gamma - 2 \ln [(-k^2 + n^2/b^2)^{1/2}A/2], \tag{16}$$

where γ is Euler's constant, numerically about 0.57722. At least for n not too large, the difference $M_n(A)$ and $N_n(A)$ is approximately independent of A so that it may be computed for small values of A. Thus it follows from (15) and a comparison of (11) with (16) that

$$bM_n(A) = 2K_0[(-k^2 + n^2/b^2)^{1/2}A] + \ln (1 - k^2b^2/n^2) + 2C_n - \pi \int_0^{2kb} dx[\Omega_{2n}(x) - iJ_{2n}(x)], \tag{17}$$

where

$$C_n = \ln (4n) + \gamma - 2 \sum_{m=0}^{n-1} (2m + 1)^{-1}. \tag{18}$$

Since A is small, (17) may be slightly simplified to

$$bM_n(A) = 2K_0(nA/b) + 2C_n - \pi \int_0^{2kb} dx[\Omega_{2n}(x) - iJ_{2n}(x)]. \tag{19}$$

It is worth noting that both C_n and the integral on the right-hand side of (19) are of the order of magnitude of n^{-2} as $n \to \infty$. Thus (19) does not hold if $n \gg b/A$. Fortunately, this does not cause any trouble. Finally the substitution of (13) and (19) into (9) gives

$$\kappa_0 = (\pi b)^{-1}\left\{\ln (8b/a) - \tfrac{1}{2}\pi \int_0^{2kb} dx[\Omega_0(x) - iJ_0(x)]\right\} \tag{20a}$$

and

$$\kappa_n = \kappa_{-n} = (\pi b)^{-1}\left\{K_0(na/b)I_0(na/b) + C_n - \tfrac{1}{2}\pi \int_0^{2kb} dx[\Omega_{2n}(x) - iJ_{2n}(x)]\right\} \tag{20b}$$

for $n \geq 1$. This is the desired answer. The current distribution may be found from (5), (7), and (20).

The difficulty of Hallén and Storer is simply not encountered, since for large n,

$$I_n^{-1} = -(ikV)^{-1}\pi\zeta_0 n^2 \kappa_n \tag{21}$$

approximately, and this is never very small.

4. THE INPUT ADMITTANCE

Similar to the known case of the dipole antenna,[6] $\lim_{\phi \to 0} I(\phi)$ does not exist. Let Y_∞ be the admittance

[7] C. W. Oseen, Arkiv Mat. Astron. Fys. 9, No. 12 (1913).

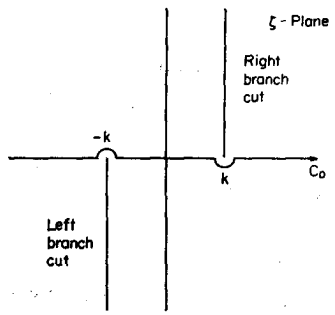


FIG. 2. The ζ plane and the contour C_0.

of an infinite dipole antenna of radius a, then the difference between $Y = I(0)/V$ and Y_∞ is finite. More specifically

$$Y - Y_\infty = \lim_{\phi \to 0} \Bigg(V^{-1} \sum_{n=-\infty}^{\infty} I_n e^{in\phi} + (2/\pi) k \zeta_0^{-1} \times \int_{C_0} d\zeta e^{i\zeta\phi b} (\zeta^2 - k^2)^{-1} \times \{ J_0[a(k^2 - \zeta^2)^{1/2}] H_0^{(1)}[a(k^2 - \zeta^2)^{1/2}] \}^{-1} \Bigg), \quad (22)$$

where the contour of integration C_0 is shown in Fig. 2. Equation (22) may be simplified to

$$Y - Y_\infty = \lim_{N \to \infty} ik\pi^{-1} \zeta_0^{-1} \times \Bigg(\sum_{n=-N}^{N} [\tfrac{1}{2} k^2 b^2 (\kappa_{n+1} + \kappa_{n-1}) - n^2 \kappa_n]^{-1} - 2i \int_{C_0(N)} d\zeta (\zeta^2 - k^2)^{-1} \times \{ J_0[a(k^2 - \zeta^2)^{1/2}] H_0^{(1)}[a(k^2 - \zeta^2)^{1/2}] \}^{-1} \Bigg), \quad (23)$$

where $C_0(N)$ is the part of C_0 with $|\zeta| < N + \frac{1}{2}$.

5. DISCUSSIONS

Within the framework of one-dimensional approximation, the current distribution on a thin circular loop transmitting antenna driven by a delta-function generator has been determined completely. Within the same approximation, the receiving antenna can be treated in a very similar manner without any new difficulty.

One-dimensional approximations, however, are never entirely satisfactory. Although difficult, it is highly desirable to study the complete Eqs. (1a) and (1b) directly, expanding the various quantities in the small parameters a/b and ka. This remains to be done.

ACKNOWLEDGMENT

For very helpful discussions, I am indebted to Professor R. W. P. King.