JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 6, NUMBER 7 JULY 1965

Weight Factors for the Two-Dimensional Ising Model*

GLEN BAXTER
Department of Mathematics, University of California, San Diego, La Jolla, California
(Received 30 November 1964)

A group $\mathfrak{G}$ of local weights is constructed for the square, honeycomb, and triangular lattices which counts for any closed path in the lattice $1/2\pi$ times the change in the argument of the tangent vector (mod 2) and the number of enclosed units of area (mod 2). These weights are used to evaluate the partition function of the two-dimensional Ising model with nearest-neighbor interaction and with a particular, imaginary external magnetic field. For the square lattice, the method gives a result announced by Lee and Yang.

INTRODUCTION

IN their discussion of the two-dimensional Ising model for the square lattice, Kac and Ward[1] introduced weight factors which proved useful in evaluating the number of closed graphs in the lattice via the evaluation of a large-order determinant. They were able to rederive Onsager's expression[2] for the partition function in the case of nearest-neighbor interaction and zero external magnetic field. Later, Feynman[3] conjectured an identity, valid for a large class of planar graphs, giving the number of closed subgraphs as an exponent of a sum over weighted paths in the graph. Recently, Sherman[4,5] gave a rigorous proof of this conjecture.

An important feature of both methods is that the weight assigned to any path depends only on the "local" behavior of the path. That is, a path is given as a sequence of vertices with connecting straight-line edges and a weight is assigned to each vertex depending on the angle formed by the two corresponding edges. The "past" or the "future" of the path play no role in the assignment at the given vertex. The weight of the whole path is simply the product of the weights assigned to the vertices in the path. Thus, "global" properties are determined by "local" properties, and an explicit evaluation is quite often possible (by the method of inching along!). In the case of the square lattice in the plane, the weight assigned to a "closed" path is $(-1)^{n(p)}$, where $2\pi n(p)$ is the change in the argument of the tangent vector in one traversal of the path p.

We are interested here in another global property which can be determined by local weights. We lose immediately the generality of Feynman's conjecture and must restrict our considerations to three regular lattices in the plane, i.e., square, honeycomb, and triangular. However, for the three lattices just mentioned, we will derive a system of weights which for any closed path will give us, in addition to $1/2\pi$ times the change in the argument of the tangent vector (mod 2), the number of enclosed "units" of area (mod 2). Using these weights, one can rederive a result of Lee and Yang[6] for the square lattice giving the free energy per spin at $z = -1$ (external magnetic field equal to $ikT\pi/2\mu$), and one can derive similar results for the honeycomb and triangular lattices.

* This research was supported in part by the U. S. Air Force under grant number AF-AFOSR-722-65.
[1] M. Kac and J. C. Ward, Phys. Rev. 88, 1332 (1952).
[2] L. Onsager, Phys. Rev. 65, 117 (1944).
[3] Unpublished.
[4] S. Sherman, J. Math. Phys. 1, 202 (1960).
[5] S. Sherman, J. Math. Phys. 4, 1213 (1963).
[6] T. D. Lee and C. N. Yang, Phys. Rev. 87, 410 (1952).

DISCUSSION OF WEIGHTS

For the moment, we restrict our discussion to the square lattice in the plane, even though it is clear that the argument will apply with equal validity to the other lattices. Let the lattice points be (m, n) where m and n are integers, and let the edges be the unit long segments connecting any two nearest points in a horizontal or vertical direction. To each edge we can assign a direction in two ways. By a *path* we mean a well-defined sequence of directed edges, where any two successive edges touch (terminal to initial) and where no two successive edges are the same. By a *closed path* we mean a path such that the last edge touches the first edge (terminal to initial). The local behavior of a closed path at a point is illustrated in Fig. 1. Kac and Ward assigned weights $\alpha = \exp(\frac{1}{4}i\pi)$, 1, $\bar{\alpha} = \exp(-\frac{1}{4}i\pi)$, respectively, to the configurations (i), (ii), (iii) of Fig. 1. The weight assigned a closed path is 1 or -1. Clearly, the weights are elements from a cyclic group of order 8 generated by α.

If one is interested in assigning local weights, say $\hat{\alpha}$, β, α, to the configurations in (i), (ii), and (iii), respectively, in Fig. 1, in such a way that a closed path has a weight from which one can determine "twist" of the tangent vector (mod 2) and enclosed number of squares (mod 2), then four weights are required for describing closed paths. Moreover, the square of all of these four weights must be the same, since traversing any closed path twice around yields a weight which is the square of that assigned to the original closed path, while traversing any closed path twice around doubles both the twist of the tangent vector and the number of enclosed squares. Thus, we are dealing with the four-group $\mathfrak{V} \equiv \{a, b\}$ with $a^2 = b^2 = I$, $ab = ba$. We will require the following interpretation for closed paths which have the weight indicated.

I: even twist of tangent vector and even number of enclosed squares.
a: odd twist of tangent vector and even number of enclosed squares. (1)
b: even twist of tangent vector and odd number of enclosed squares.
ab: odd twist of tangent vector and odd number of enclosed squares.

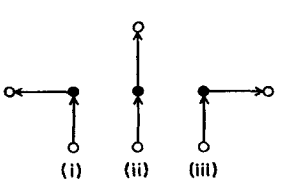


FIG. 1. The three types of local behavior of a path at a vertex in the square lattice.

In general, the weights α, β, $\hat{\alpha}$ will be elements of a noncommutative group $\mathfrak{G}$. Care must be exercised in assigning weights to a path. We will abide by the following rule.

Rule 1. If a path (directed!) has edges $e_0, \cdots, e_n$ and successive vertices $v_1, \cdots, v_n$, and if α_k is the weight assigned to the path configuration at v_k, then the weight assigned to the path itself will be $w = \alpha_1 \cdots \alpha_n$.

A closed path may be described in many ways as a sequence of edges, each description being determined by the selection of a particular edge as the "start." In any case, the weight assigned must be invariant under the obvious cyclic permutation of the α_k's. This means that the four-group $\mathfrak{V} \equiv \{a, b\}$ must lie in the center of the weight group $\mathfrak{G}$. We state this as a second rule.

Rule 2. The four-group $\mathfrak{V}$ must be a subgroup of the center of the group $\mathfrak{G}$ generated by the local weights.

In the next few sections, we will construct the weight groups, according to the stated rules, for the square, honeycomb and triangular lattices, so that weights assigned to closed paths have the interpretation (1).

THE SQUARE LATTICE

The square lattice was described in the previous section. We let the local weights, if any exist, be α for a right turn, β for a move straight ahead, and $\hat{\alpha}$ for a left turn. These weights are to be independent of position in the lattice and of the orientation.

Let us first derive some necessary conditions for α, $\hat{\alpha}$, β. In Fig. 2, we have shown schematically two closed paths, one of which contains one more enclosed square than does the other. Both paths have the same twist of the tangent vector. Thus, the weight assigned to path 2 is b times that assigned to path 1. Since the weights are to form a group, we can cancel out the weight assigned to that part of the closed path lying between $\mathbf{u}_1$ and $\mathbf{u}_0$, and we deduce that $b\alpha^2 = \beta\alpha^2\beta$. Actually, there are several relations of this type which can be deduced in a

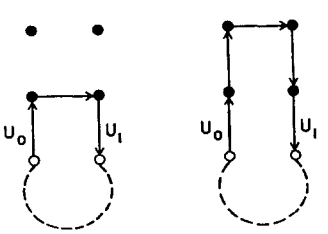


FIG. 2. Two closed paths which differ only by inclusion or noninclusion of sides from a particular square.

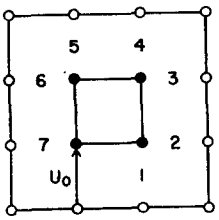


FIG. 3. The numbers indicate seven possible positions for the directed edge of the closed path as the path leaves the central square.

similar manner. In Fig. 3, we have indicated seven possible positions for the vector $\mathbf{u}_1$, each of which leads to a condition. Adding the obvious condition $\alpha^4 = \hat{\alpha}^4 = ab$ to these, we can list eight basic relations:

(1) $b\alpha^2 = \beta\alpha^2\beta$,
(2) $b\alpha\beta = \beta\alpha^2\hat{\alpha}$,
(3) $b\alpha\hat{\alpha}\alpha = \beta\alpha\beta$,
(4) $b\alpha\hat{\alpha}\beta = \beta\alpha\hat{\alpha}$,
(5) $b\alpha\hat{\alpha}^2\alpha = \beta^2$,
(6) $b\alpha\hat{\alpha}^2\beta = \beta\hat{\alpha}$,
(7) $b\alpha\hat{\alpha}^3\alpha = \hat{\alpha}$,
(8) $\alpha^4 = ab$.

It is amusing to note that the weight group $\mathfrak{G}$ (if it exists) must be noncommutative. For, if β commuted with α and $\hat{\alpha}$, (4) would imply that $b = I$, a contradiction. A bit of guess work is needed here to achieve a simplification, the guess being that $\beta^2 = I$. One can then derive from (1)–(8)

(i) $\beta^2 = I$,
(ii) $\hat{\alpha}^2\alpha^2 = b$,
(iii) $\alpha\hat{\alpha} = \hat{\alpha}\alpha$,
(iv) $\beta\alpha\beta\hat{\alpha} = I$,
(v) $\alpha^4 = \hat{\alpha}^4 = ab$.

For example, (ii) follows from (5), (iii) subsequently follows from (7), and in turn (iv) follows from (3) and (ii). It is not difficult to see that the converse is also true, i.e., (i)–(v) imply (1)–(8).

Relations (i)–(v) define a group $\mathfrak{G}$, of order 64, with center $\mathfrak{Z} \equiv \{a, b\}$. This group will now be shown to be the desired weight group. To accomplish this end, we must first check the weights assigned to two more pairs of related paths, as pictured in Fig. 4. It is left to the reader to verify that for the first pair the assigned weights differ by a factor of ab, while for the second pair the assigned weights differ by a factor of a. Now, any closed path can be thought of as being formed step by step (in a non-unique manner) through the addition of some or all of the edges of a single square, starting of course with a closed path surrounding a single square. The considerations above have shown us that the addition of edges from a single square (with, in general, the removal of others) alters the weight in exactly the appropriate manner so that the interpretation (1) remains valid for the new path if it was valid for the old one. The fact that $\mathfrak{G}$ is the desired weight group follows by induction.

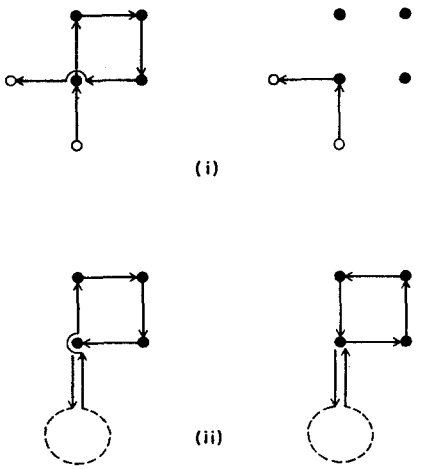


FIG. 4. Two pairs of related closed paths differing only in their motion around the central square.

THE HONEYCOMB LATTICE

The method of the last section can be applied with equal success to find the weight group $\mathfrak{G}$ associated with the honeycomb lattice (see Fig. 5). At any vertex there are two possible motions, a right turn and a left turn, to which we assign the weights α and β, respectively. In Fig. 5, we have shown five possible positions for a vector $\mathbf{u}_1$ directed away from the central hexagon. For each position of $\mathbf{u}_1$, two different paths connecting $\mathbf{u}_0$ and $\mathbf{u}_1$ can be drawn using only sides from the central hexagon and such that the two paths use different sides. In each case the weights assigned to the two different paths must disagree by a factor of b. Adding the obvious condition $\alpha^6 = \beta^6 = ab$, we get six basic relations:

(1) $b\alpha^2 = \beta\alpha^4\beta$,
(2) $b\alpha\beta\alpha = \beta\alpha^3\beta$,
(3) $b\alpha\beta^2\alpha = \beta\alpha^2\beta$,
(4) $b\alpha\beta^3\alpha = \beta\alpha\beta$,
(5) $b\alpha\beta^4\alpha = \beta^2$,
(6) $\alpha^6 = ab$.

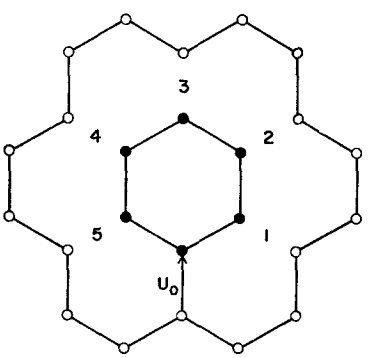


FIG. 5. The numbers indicate five possible positions for the directed edge of the closed path as the path leaves the central hexagon.

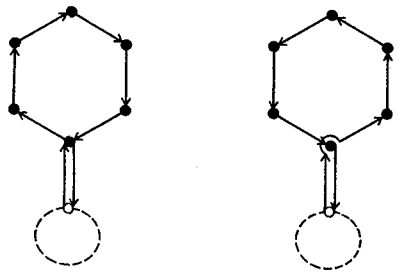


FIG. 6. A pair of related closed paths differing only in the direction of their motion around the central hexagon.

A simplification is achieved by adding the condition $\alpha^3 = \beta^{-3}$. This added condition implies, for example, that (1) and (4) are equivalent and (2) and (5) are equivalent. Our simplified conditions become

(i) $\alpha^3 = \beta^{-3}$
(ii) $\alpha^6 = \beta^6 = ab$,
(iii) $\alpha\beta\alpha\beta = b$.

Note that (iii) follows immediately from (1) or (2). Relations (1)–(6) are implied by (i)–(iii). For example, to show (3) we deduce from (iii) that

$$\alpha\beta^3 = b\beta^2\alpha^2\beta^2 \quad \text{and} \quad \alpha^4 = (b\alpha^2)^2 = \alpha^6(\beta\alpha\beta)^2.$$

Thus,

$$b\alpha^6\beta^2\alpha^2\beta^2 = \alpha^6\beta\alpha\beta^2\alpha\beta,$$

and (3) follows.

Relations (i)–(iii) define a group $\mathfrak{G}$, of order 96, with center $\{\alpha^3, b\} \supset \mathfrak{V} \equiv \{a, b\}$. This is the desired weight group, as can be proved by the same method outlined in the last section. It is left to the reader to verify that the appropriate relationship exists between the weights assigned to the two paths pictured in Fig. 6.

THE TRIANGULAR LATTICE

At any vertex of a triangular lattice there are five possible motions. We assign weight β to a step straight ahead, weight α (or $\hat{\alpha}$) to a turn to the right (or left) through 60°, and weight γ (or $\hat{\gamma}$) to a turn to the right (or left) through 120°. As is indicated in Fig. 7, we would expect to start the analysis with 12 basic relations. We shall not do this, for one is very fortunate here to be in one

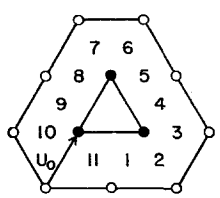


FIG. 7. The numbers indicate eleven possible positions for the directed edge of a closed path as the path leaves the central triangle.

of those rare and lovely situations in mathematics where for no apparent reason a gross simplification is possible. The weight group $\mathfrak{G}$ can be taken to be the commutative group, of order 24, generated by $\alpha^6 = a$ and b. In terms of α and b, the local weights are

$$\alpha = \alpha, \qquad \hat{\alpha} = \alpha^{-1}, \qquad \beta = b.$$
$$\gamma = b\alpha^2, \qquad \hat{\gamma} = b\alpha^{-2},$$

Using the above, it is easy to show that all pairs of paths starting with $\mathbf{u}_0$ in Fig. 7 and differing only in the inclusion of some or all of the sides of the middle triangle have weights which differ by the proper factor. The proof that $\mathfrak{G}$ is the weight group goes as before.

MATRIX REPRESENTATION

In what follows, it will be necessary to have a matrix representation for a certain quotient group of the weight group $\mathfrak{G}$. If $\mathfrak{K} = \{ab\}$ denotes the normal subgroup of $\mathfrak{G}$ of order two which is generated by ab, the group to be represented is $\mathfrak{H} = \mathfrak{G}/\mathfrak{K}$. We will continue to use α, β, etc. to represent (hopefully without confusion) the cosets to which they belong in $\mathfrak{H} = \mathfrak{G}/\mathfrak{K}$. The coset containing a and b will be denoted by $-I$.

A. The Square Lattice

The relations which define $\mathfrak{H} = \mathfrak{G}/\mathfrak{K}$ for the square lattice are

$$\alpha^4 = \beta^2 = \beta\alpha\beta\hat{\alpha} = I,$$
$$\alpha^2\hat{\alpha}^2 = -I, \qquad \alpha\hat{\alpha} = \hat{\alpha}\alpha.$$

A matrix representation in 2×2 matrices of $\mathfrak{H}$ is given by

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \qquad \hat{\alpha} = \begin{pmatrix} -i & 0 \\ 0 & 1 \end{pmatrix}, \qquad \beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

B. The Triangular Lattice

The relations which define $\mathfrak{H} = \mathfrak{G}/\mathfrak{K}$ for the triangular lattice are simply

$$\alpha^6 = \beta = -I.$$

Thus, a representation is possible in terms of twelfth roots of unity. We have

$$\beta = -1, \alpha = \hat{\alpha}^{-}$$
$$= \exp(\pi i/6), \gamma = \hat{\gamma}^{-} = \exp(4\pi i/3),$$

The bar is written as a superscript because of typographical difficulties.

C. The Honeycomb Lattice

The relations defining $\mathfrak{H} = \mathfrak{G}/\mathfrak{K}$ in this case are

$$\alpha^3 = \beta^3 = \alpha\beta\alpha\beta = -I.$$

Unfortunately, we have been unable to find a representation for $\mathfrak{H}$ in terms of matrices of usefully small order. There is one representation in terms of 6×6 matrices which we will mention in passing. We set

$$\alpha = \begin{bmatrix} 0 & A & 0 \\ 0 & 0 & B \\ AB & 0 & 0 \end{bmatrix}, \qquad \beta = \begin{bmatrix} 0 & 0 & A \\ B & 0 & 0 \\ 0 & BA & 0 \end{bmatrix},$$

where A and B are the 2×2 quaternion matrices

$$A = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

and where 0 is a 2×2 zero matrix.

LOW-TEMPERATURE EXPANSION

Let $\mathfrak{L}$ be a lattice of one of the three types under consideration having N vertices labeled $i = 1, 2, \cdots, N$, and let any vertex have an even number of edges touching it. Edges in the lattice have d different orientations or directions. For the square lattice $d = 2$, while for the honeycomb and triangular lattices $d = 3$. In the Ising model a spin coordinate $\sigma_i = \pm 1$ is attached to the ith vertex, and between any two nearest-neighbor vertices i, j in the s direction, there is postulated an interaction energy constant $J_s = K_{ij}$. The partition function is then

$$Z = \sum_{\text{spins}} \exp\left\{A \sum_{\text{edges}} K_{ij}\sigma_i\sigma_j + B \sum_{i=1}^{N} \sigma_i\right\},$$

where A and B are suitable constants depending on the absolute temperature T, a magnetic moment μ, and an external magnetic field $\mathfrak{H}$. Low temperature expansion formulae are known for Z (see, for example, Newell and Montroll[7]). In particular, it is known that

$$Z(z) = C^N \sum_{m, n_1, \cdots, n_d} g_N(m; n_1, \cdots, n_d) x_1^{n_1} \cdots x_d^{n_d} z^m, \tag{2}$$

where C is a suitable constant, and where

$$x_s = \exp(-2J_s/kT), \qquad z = \exp(-2\mu H/kT).$$

The coefficient $g_N(m; n_1, \cdots, n_d)$ is the number of closed graphs in the dual lattice $\mathfrak{L}^*$ (the lattice formed from $\mathfrak{L}$ by drawing an edge bisecting each edge of $\mathfrak{L}$ and connecting these new edges at points in the center of each unit cell of $\mathfrak{L}$) having n_s edges in the direction $s^\perp$ (the direction perpendicular to s) and having m enclosed units of area. The sum in (2) is taken over all closed graphs in $\mathfrak{L}^*$.

We are especially interested in the case in which the magnetic field is so chosen that $z = -1$, i.e., $\mathfrak{H} = ikT\pi/2\mu$. In that case, we can write

$$Z(-1) = C^N \sum_{\substack{n_1, \cdots, n_d \\ m=0,1}} h_N(m; n_1, \cdots, n_d) \times x_1^{n_1} \cdots x_d^{n_d}(-1)^m,$$

where $h_N(m; n_1, \cdots n_d)$ is the number of closed graphs in the dual lattice having n_s edges in the direction $s^\perp$ and having $m (m = 0, 1) \pmod 2$ enclosed units of area. Following the lead of Feynman, we now propose a method for evaluating $Z(-1)$. Let p be any closed path in the dual lattice $\mathfrak{L}^*$ having n_s edges in the direction $s^\perp$. Let $W(p)$ be the weight assigned to p according to Rule 1 by means of the weight group $\mathfrak{G}^*$ of $\mathfrak{L}^*$. If $W(p)$ is either I or ab, we take $\bar{W}(p) = 1$; otherwise, we take $\bar{W}(p) = -1$. We then form the sum

$$S = \sum_p \frac{\bar{W}(p) x_1^{n_1} \cdots x_d^{n_d}}{(n_1 + \cdots + n_d)}, \tag{3}$$

taken over all possible closed (directed) paths p. Then we claim that

$$Z(-1) = C^N \exp(-S/2). \tag{4}$$

One can give a rigorous proof of (4) along the lines of argument in Sherman[4,5] where the evaluation of $Z(1)$ is considered. The only formal difference in the evaluation procedure for $Z(1)$ and $Z(-1)$ is in the choice of weighting $\bar{W}(p)$ for closed paths p. For $Z(1)$, one takes $\bar{W}(p) = 1$ if $W(p)$ is I or b, and $\bar{W}(p) = -1$ otherwise.

A small simplification is achieved if one assumes that the weight $W(p)$ associated with a closed path p is formed as an ordered product of the elements α, β, etc. considered as representatives of the cosets in $\mathfrak{G}^*/\mathfrak{K}$. In this way, $W(p)$ is automatically identified with 1 or -1, i.e., with $\bar{W}(p)$.

CALCULATION OF $Z(-1)$

An explicit calculation of $Z(-1)$ will now be made for the square and honeycomb lattices with N vertices. Theoretically at least, one can also make the corresponding calculation for the triangular lattice. However, there are practical difficulties in tarrying through the computation. In our calculations, boundaries will be ignored. Heuristically, this is equivalent to examining the dominant part (as

[7] G. F. Newell and E. W. Montroll, Rev. Mod. Phys. 25, 353 (1953).

$N \to \infty$) of $Z(-1)$ for lattices embedded on the torus. In the case of torus embedding, the number N of vertices (and hence squares) must be even in order that our evaluation of "enclosed" area (mod 2) of any closed path will make sense. On the torus, there is no real distinction between the inside and outside of a nonintersecting closed path, so the number of area units in the two regions determined by the closed path must have the same parity. Perhaps this explains the point raised by Sherman[4] in connection with the Lee–Yang formula (see below).

In the evaluations, we assume that the weights from $\mathfrak{G}^*$ commute with the indeterminants x_i and with the complex constants $\exp(2\pi i\theta)$ and $\exp(2\pi i\varphi)$.

A. The Square Lattice

To evaluate S in (3), and hence $Z(-1)$, in this case, we employ weighted random walks. The dual lattice $\mathfrak{L}^*$ is again the square lattice. If we wanted simply the number of random walks of length k in $\mathfrak{L}^*$ (ignoring boundaries), we would introduce a "characteristic" function $\Phi = \sum_{\sigma_1,\sigma_2=\pm 1} \exp(\sigma_1 2\pi i\theta + \sigma_2 2\pi i\varphi)$, which describes one step of the random walk, and we would compute $\int_0^1 \int_0^1 \Phi^k \, d\theta \, d\varphi$. Although the computation of S follows the same lines, there is a complication in that each vertex of a path corresponds to four "states," one for each of the directions of the path when entering the vertex. Thus, we introduce the characteristic matrix $\mathfrak{M} = \mathfrak{M}(\theta, \varphi)$ to describe one step of our weighted paths.

$$\mathfrak{M} = \begin{bmatrix} \beta u x_1 & 0 & \hat{\alpha} v x_2 & \alpha \bar{v} x_2 \\ 0 & \beta \bar{u} x_1 & \alpha v x_2 & \hat{\alpha} \bar{v} x_2 \\ \alpha u x_1 & \hat{\alpha} \bar{u} x_1 & \beta v x_2 & 0 \\ \hat{\alpha} u x_1 & \alpha \bar{u} x_1 & 0 & \beta \bar{v} x_2 \end{bmatrix}, \qquad \begin{aligned} u &= \exp(2\pi i\theta), \\ v &= \exp(2\pi i\varphi). \end{aligned}$$

It is noted that m_{ij} gives the appropriate weight (includes indeterminant x_j) to the motion from direction i through the vertex to direction j, according to the labeling of directions in Fig. 8. We interpret $\alpha, \beta, \hat{\alpha}$ to be cosets in $\mathfrak{G}^*/\mathfrak{K}$. It follows that (ignoring boundaries)

$$S = N \sum_{k=1}^{\infty} \operatorname{Tr} \left\{ \int_0^1 \int_0^1 \frac{\mathfrak{M}^k(\theta, \varphi)}{k} \, d\theta \, d\varphi \right\}$$

$$= -N \int_0^1 \int_0^1 \operatorname{Tr} \{\ln (I - \mathfrak{M})\} \, d\theta \, d\varphi.$$

The computation of $\operatorname{Tr}(\mathfrak{M}^k)$ or $\operatorname{Tr}\{\ln(I - \mathfrak{M})\}$ is complicated by the fact that $\mathfrak{M}$ contains noncommutative elements from $\mathfrak{G}^*/\mathfrak{K}$. One avoids this difficulty by introducing the matrix representations

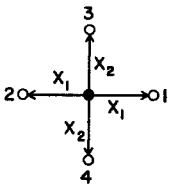


FIG. 8. Labeling of directions, with assignment of indeterminant, for square lattice.

of $\mathfrak{G}^*/\mathfrak{K}$ presented in Sec. 5. Let $\tilde{\mathfrak{M}}(\theta, \varphi) = \tilde{\mathfrak{M}}$ be the 8×8 matrix formed from $\mathfrak{M}$ by replacing $\alpha, \beta, \hat{\alpha}$ (and 0) by their 2×2 matrices

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \qquad \beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{\alpha} = \begin{pmatrix} -i & 0 \\ 0 & 1 \end{pmatrix}.$$

Then,

$$\int_0^1 \int_0^1 \operatorname{Tr}(\mathfrak{M}^k) \, d\theta \, d\varphi = \frac{1}{2} \int_0^1 \int_0^1 \operatorname{Tr}(\tilde{\mathfrak{M}}^k) \, d\theta \, d\varphi,$$

and we have

$$S = -\frac{N}{2} \int_0^1 \int_0^1 \operatorname{Tr} \{\ln (I - \tilde{\mathfrak{M}})\} \, d\theta \, d\varphi$$

$$= -\frac{N}{2} \int_0^1 \int_0^1 \ln |I - \tilde{\mathfrak{M}}| \, d\theta \, d\varphi.$$

An incredibly tedious computation shows that

$$|I - \tilde{\mathfrak{M}}| = (1 - x_1^2)^2 (1 - x_2^2)^2 + 4x_1^2 (1 - x_2^2)^2 \sin^2 2\pi\theta + 4x_2^2 (1 - x_1^2)^2 \sin^2 2\pi\varphi.$$

Thus, to within a constant,

$$\frac{1}{N} \ln Z(-1) \to \frac{1}{4} \int_0^1 \int_0^1 \ln |I - \tilde{\mathfrak{M}}| \, d\theta \, d\varphi,$$

which is the Lee–Yang formula.[6]

B. The Honeycomb Lattice

In this case, the dual lattice $\mathfrak{L}^*$ is the triangular lattice. At each point of the triangular lattice there are six outward directions, which for our purposes are labeled (with indeterminants) in Fig. 9. To keep track of the position in the lattice, we introduce the complex weights $u = \exp(2\pi i\theta)$ for a step in direction 1, $v = \exp(2\pi i\varphi)$ for a step in direction 3, and $w = \exp[2\pi i(\theta + \varphi)]$ for a step in direction 5. Steps in the opposite directions are assigned the conjugates. The characteristic matrix $\mathfrak{M} \equiv \mathfrak{M}(\theta, \varphi)$

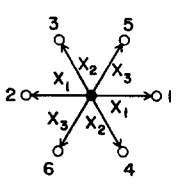


FIG. 9. Labeling of directions, with assignment of indeterminant, for triangular and honeycomb lattices.

in this case is then

$$\mathfrak{M} = \begin{bmatrix} -x_1u & 0 & \gamma x_2v & \alpha x_2\bar{v} & \bar{\alpha}x_3w & \gamma x_3\bar{w} \\ 0 & -x_1\bar{u} & \alpha x_2v & \bar{\gamma}x_2\bar{v} & \gamma x_3w & \bar{\alpha}x_3\bar{w} \\ \gamma x_1u & \bar{\alpha}x_1\bar{u} & -x_2v & 0 & \alpha x_3w & \bar{\gamma}x_3\bar{w} \\ \bar{\alpha}x_1u & \gamma x_1\bar{u} & 0 & -x_2\bar{v} & \bar{\gamma}x_3w & \alpha x_3\bar{w} \\ \alpha x_1u & \bar{\gamma}x_1\bar{u} & \bar{\alpha}x_2v & \gamma x_2\bar{v} & -x_3w & 0 \\ \bar{\gamma}x_1u & \alpha x_1\bar{u} & \gamma x_2v & \bar{\alpha}x_2\bar{v} & 0 & -x_3\bar{w} \end{bmatrix},$$

where $\alpha = \exp(\pi i/6)$ and $\gamma = \exp(4\pi i/3)$. Ignoring boundaries, we have that

$$S = -N\int_0^1\int_0^1 \mathrm{Tr}\ln(I - \mathfrak{M})\,d\theta\,d\varphi = -N\int_0^1\int_0^1 \ln|I - \mathfrak{M}|\,d\theta\,d\varphi.$$

By means of an even more tedious calculation than that for the square lattice, one can show that

$$\begin{aligned}|I - \mathfrak{M}| &= (1 + x_1^2)(1 + x_2^2)(1 + x_3^2) - 8x_1x_2x_3 \\ &\quad + 2x_1(1 - x_2^2)(1 - x_3^2)\cos 2\pi\theta \\ &\quad + 2x_2(1 - x_1^2)(1 - x_3^2)\cos 2\pi\varphi \\ &\quad + 2x_3(1 - x_1^2)(1 - x_2^2)\cos 2\pi(\theta + \varphi). \qquad (5)\end{aligned}$$

Thus, to within a constant,

$$\frac{1}{N}\ln Z(-1) \to \frac{1}{2}\int_0^1\int_0^1 \ln|I - \mathfrak{M}|\,d\theta\,d\varphi,$$

where $|I - \mathfrak{M}|$ is given in (5).

C. The Triangular Lattice

In this case the dual lattice $\mathfrak{L}^*$ is the honeycomb lattice. At each point of the honeycomb lattice, there are *six* possible outward directions, three of them being excluded for the particular point in question. Thus, we can once again take the labeling of Fig. 9. We can also use again the complex weights u, v, and w for keeping track of the position in the lattice. The characteristic matrix $\mathfrak{M} \equiv \mathfrak{M}(\theta, \varphi)$ in this case is

$$\mathfrak{M} = \begin{bmatrix} 0 & 0 & 0 & \alpha x_2\bar{v} & \beta x_3w & 0 \\ 0 & 0 & \alpha x_2v & 0 & 0 & \beta x_3\bar{w} \\ 0 & \beta x_1\bar{u} & 0 & 0 & \alpha x_3w & 0 \\ \beta x_1u & 0 & 0 & 0 & 0 & \alpha x_3\bar{w} \\ \alpha x_1u & 0 & \beta x_2v & 0 & 0 & 0 \\ 0 & \alpha x_1\bar{u} & 0 & \beta x_2\bar{v} & 0 & 0 \end{bmatrix},$$

where α and β satisfy the group relations $\alpha^3 = \beta^3 = \alpha\beta\alpha\beta = -I$. It follows that (ignoring boundaries)

$$S = -N\int_0^1\int_0^1 \mathrm{Tr}\{\ln(I - \mathfrak{M})\}\,d\theta\,d\varphi.$$

The complication that $\mathfrak{M}^k$ contains elements from the noncommutative group $\mathfrak{G}^*/\mathfrak{K}$ comes back to haunt us here. If one attempts to avoid a direct computation of $\mathfrak{M}^k$ by the method of substituting a matrix representation for α and β in $\mathfrak{M}$, one finds the resulting composite matrix of relatively high order. If in particular we form $\tilde{\mathfrak{M}}$ by replacing α, β and 0 by their 6×6 matrix representations given in Sec. 5, we are led to the problem of evaluating the determinant $|I - \tilde{\mathfrak{M}}|$ of order 36×36. (Egad!)

In any case, the final result is

$$\frac{1}{N}\ln Z(-1) \to \frac{1}{12}\int_0^1\int_0^1 \ln|I - \tilde{\mathfrak{M}}|\,d\theta\,d\varphi,$$

so the only problem which remains is to evaluate $|I - \tilde{\mathfrak{M}}|$.

ACKNOWLEDGMENT

The author is grateful to Dr. Colin Thompson for reading through the original manuscript and for offering useful suggestions.

ADDENDUM

Since the appearance of the original version of this paper, Freeman Dyson and the author have succeeded in finishing the evaluation of $Z(-1)$ for the triangular lattice. One begins with the observation that

$$\alpha = \tfrac{1}{2}\begin{pmatrix} 1-i & -1-i \\ 1-i & 1+i \end{pmatrix}, \qquad \beta = \tfrac{1}{2}\begin{pmatrix} 1+i & -1-i \\ 1-i & 1-i \end{pmatrix}$$

is a representation of the honeycomb lattice group $H = G/K$: $\alpha^3 = \beta^3 = \alpha\beta\alpha\beta = -1$. Substituting these matrix representations for α and β into the characteristic matrix $\mathfrak{M}$ yields a matrix $\tilde{\mathfrak{M}}$ and an evaluation of $|I - \tilde{\mathfrak{M}}|$ gives

$$\begin{aligned}|I - \tilde{\mathfrak{M}}| &= 1 + 2x_1^2x_2^2(1 - x_3^2)^2\cos 4\pi(\theta - \varphi) \\ &\quad + 2x_1^2x_3^2(1 - x_2^2)^2\cos 4\pi(2\theta + \varphi) \\ &\quad + 2x_2^2x_3^2(1 - x_1^2)^2\cos 4\pi(\theta + 2\varphi) \\ &\quad - 4x_1^2x_2^2x_3^2 + x_1^4x_2^4 + x_1^4x_3^4 + x_2^4x_3^4. \qquad (6)\end{aligned}$$

To within a constant,

$$\frac{1}{N}\ln Z(-1) \to \frac{1}{4}\int_0^1\int_0^1 \ln|I - \tilde{\mathfrak{M}}|\,d\theta\,d\varphi,$$

where $|I - \tilde{\mathfrak{M}}|$ is given in (6).

Conservation Laws for Free Fields*

T. W. B. KIBBLE
Department of Physics, Imperial College, London, England
(Received 21 July 1964)

The recently discovered conservation laws for the "zilch" of the electromagnetic field (of which a simple derivation is presented) are examined in the context of a general discussion of bilinear conserved quantities in free-field theories. It is shown that there is always an infinite set of these quantities, and a method of finding them all is presented and illustrated by applications to simple theories. The existence of these conserved quantities is shown to be a consequence of the fact that the momentum-space density is constant in time.

1. INTRODUCTION

TEN new conservation laws for the electromagnetic field in vacuum have recently been discovered by Lipkin,[1] who has termed the corresponding conserved quantities the "zilch" of the electromagnetic field. The present paper was motivated by a desire to understand these laws. This investigation has led to a more general study of conservation laws for noninteracting fields.

We begin by presenting an alternative proof of the conservation of zilch. We show that, by using the dual tensor $^*F_{\mu\nu}$, one can write the zilch tensor in a form in which it is obviously conserved in virtue of the wave equation. In Sec. 3 we show that for any noninteracting field there is an infinite set of conserved quantities corresponding to densities which are local bilinear functions of the field variables. In any particular case, the entire set can easily be found, as we show by considering some simple examples.

These conserved quantities take on a particularly simple form in momentum space. Their conservation is a consequence of the invariance in time of the momentum-space density. Thus it is only in very special circumstances that we can expect any of these conservation laws to carry over to the case of interacting fields.

2. CONSERVATION OF ZILCH

The electromagnetic field is described by the antisymmetric tensor $F_{\mu\nu}$. It is often convenient to introduce also the dual tensor $^*F_{\mu\nu}$ defined by[2]

$$^*F^{\mu\nu} \equiv \tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}, \qquad F_{\mu\nu} \equiv -\tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}{}^*F^{\rho\sigma}. \tag{1}$$

Using standard identities it is easy to verify the important identity

$$^*F^{\nu\lambda}F_{\lambda\mu} \equiv \tfrac{1}{4}\delta^\nu_\mu {}^*F^{\kappa\lambda}F_{\lambda\kappa} \equiv \delta^\nu_\mu \mathbf{E}\cdot\mathbf{H}. \tag{2}$$

In terms of F and $^*F$, Maxwell's equations may be written simply

$$F^{\mu\nu}{}_{,\nu} = 0, \qquad {}^*F^{\mu\nu}{}_{,\nu} = 0. \tag{3}$$

We shall define

$$Z^\mu_{\nu\rho} \equiv {}^*F^{\mu\lambda}F_{\lambda\nu,\rho} - F^{\mu\lambda}{}^*F_{\lambda\nu,\rho}. \tag{4}$$

Later we shll show that this is equivalent to Lipkin's definition. Using the identity (2) differentiated with respect to x_ρ, we can write (4) in the alternative form

$$Z^\mu_{\nu\rho} \equiv {}^*F^{\mu\lambda}F_{\lambda\nu,\rho} + {}^*F_{\nu\lambda}F^{\lambda\mu}{}_{,\rho} - \tfrac{1}{2}\delta^\mu_\nu {}^*F^{\kappa\lambda}F_{\lambda\kappa,\rho}. \tag{5}$$

It follows immediately that Z is symmetric in its first two indices,

$$Z^{\mu\nu}{}_\rho \equiv Z^{\nu\mu}{}_\rho, \tag{6}$$

and also traceless in the same pair

$$Z^\mu{}_{\mu\rho} \equiv 0. \tag{7}$$

These are identities. From Maxwell's equations (3) we see that also

$$Z^{\mu\rho}{}_\rho = 0, \tag{8}$$

so that all the contractions of Z vanish. Making explicit use of the symmetry (6) we can rewrite (4) in the form

$$Z^\mu{}_{\nu\rho} \equiv {}^*F^{\mu\lambda}\overleftrightarrow{\partial}_\rho F_{\lambda\nu} + {}^*F_{\rho\lambda}\overleftrightarrow{\partial}_\rho F^{\lambda\mu}, \tag{9}$$

where $A\overleftrightarrow{\partial}_\rho B \equiv \tfrac{1}{2}(AB_{,\rho} - A_{,\rho}B)$. Then it follows from the wave equations

$$\Box F_{\mu\nu} = 0, \qquad \Box {}^*F_{\mu\nu} = 0 \tag{10}$$

that

$$Z^{\mu\nu\rho}{}_{,\rho} = 0, \tag{11}$$

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research OAR through the European Office, Aerospace Research, U. S. Air Force.

[1] D. M. Lipkin, J. Math. Phys. 5, 696 (1964).

[2] The notation is as follows: Greek indices run from 0 to 3, and we use a metric with signature $(1 - 1 - 1 - 1)$ and scalar product $k\cdot x = k^\circ x^\circ - \mathbf{k}\cdot\mathbf{x}$. The completely antisymmetric tensor is $\epsilon^{\lambda\mu\nu\rho}$ with $\epsilon^{0123} = 1$. Derivatives with respect to x^μ are denoted by a comma.

and hence that the ten quantities

$$Z^{\mu\nu} \equiv \int d^3\mathbf{x}\, Z^{\mu\nu 0} \tag{12}$$

are constants of the motion. These are the ten components of zilch. [Actually, in view of (7), there are really only nine components.]

Next we note that

$$\begin{aligned} Z^{\mu}_{\nu\rho} - Z^{\mu}_{\rho\nu} &= {}^*F^{\mu\lambda}(F_{\lambda\nu,\rho} - F_{\lambda\rho,\nu}) \\ &\quad - F^{\mu\lambda}({}^*F_{\lambda\nu,\rho} - {}^*F_{\lambda\rho,\nu}) \\ &= -{}^*F^{\mu\lambda}F_{\nu\rho,\lambda} + F^{\mu\lambda}{}^*F_{\nu\rho,\lambda} \\ &= ({}^*F^{\lambda\mu}F_{\nu\rho} - F^{\lambda\mu}{}^*F_{\nu\rho})_{,\lambda}, \end{aligned} \tag{13}$$

using Maxwell's equations first in their cyclic form and then in the form (3). The right-hand side of this equation can be expressed in terms of the Maxwell stress–energy tensor

$$T^{\mu}{}_{\nu} \equiv F^{\mu\lambda}F_{\lambda\nu} - \tfrac{1}{4}\delta^{\mu}_{\nu}F^{\kappa\lambda}F_{\lambda\kappa}.$$

It is easy to verify that

$$\begin{aligned} {}^*F^{\lambda\mu}F^{\nu\rho} - F^{\lambda\mu}{}^*F^{\nu\rho} &= \tfrac{1}{2}(\epsilon^{\kappa\mu\nu\rho}T^{\lambda}{}_{\kappa} - \epsilon^{\kappa\lambda\nu\rho}T^{\mu}{}_{\kappa} \\ &\quad - \epsilon^{\kappa\rho\lambda\mu}T^{\nu}{}_{\kappa} + \epsilon^{\kappa\nu\lambda\mu}T^{\rho}{}_{\kappa}). \end{aligned}$$

It follows, using $T^{\lambda}_{\kappa,\lambda} = 0$, that

$$\begin{aligned} Z^{\mu\nu\rho} - Z^{\mu\rho\nu} &= \tfrac{1}{2}(-\epsilon^{\kappa\lambda\nu\rho}T^{\mu}{}_{\kappa,\lambda} \\ &\quad + \epsilon^{\kappa\lambda\rho\mu}T^{\nu}{}_{\kappa,\lambda} - \epsilon^{\kappa\lambda\nu\mu}T^{\rho}{}_{\kappa,\lambda}). \end{aligned} \tag{14}$$

Taking the divergence with respect to the index ρ and using (11) we find that

$$Z^{\mu\rho\nu}{}_{,\rho} = 0.$$

Hence Z is divergence-free in all its indices. However, the corresponding conserved quantities are not independent of zilch, because for $\rho = 0$ the right side of (14) can always be written as a spatial divergence (using conservation of $T^{\rho}{}_{\kappa}$).

It is easy to establish the relationship with Lipkin's definition of zilch, using (14). In fact,

$$Z^{\mu\nu\rho} = \tfrac{1}{2}(Z^{\rho\mu\nu} + Z^{\rho\nu\mu}) + \tfrac{1}{2}(\epsilon^{\kappa\lambda\rho\mu}T^{\nu}{}_{\kappa} + \epsilon^{\kappa\lambda\rho\nu}T^{\mu}{}_{\kappa})_{,\lambda},$$

and the right-hand side of this equation may be identified as Lipkin's expression [Eq. (4) of Ref. 1].

It is evident that the conservation of zilch follows directly from the fact that Z can be written in the form (9), and that the fields satisfy the wave equation (10). We shall see in the next section that there is a large class of conservation laws of this type for any free fields.

3. GENERAL FREE FIELD

Let us consider a free field ϕ with n real components ϕ_r which satisfies the Klein–Gordon equation for mass m,

$$(\Box + m^2)\phi = 0, \tag{15}$$

and possibly some additional subsidiary conditions. We write the plane-wave decomposition of ϕ in the form

$$\phi(x) = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,[e^{-ik\cdot x}\phi(\mathbf{k}) + e^{ik\cdot x}\phi^*(\mathbf{k})], \tag{16}$$

where $k^0 \equiv (\mathbf{k}^2 + m^2)^{\frac{1}{2}}$. The Fourier component $\phi(\mathbf{k})$ has the form

$$\phi(\mathbf{k}) = \sum_{\alpha} a_{\alpha}(\mathbf{k})u_{\alpha}(\mathbf{k}), \tag{17}$$

where the $u_{\alpha}(\mathbf{k})$ are a complete set of c-number solutions of the subsidiary equations.

Now let us consider the conditions under which a bilinear expression of the form

$$J_{\rho} = \tilde{\phi}P_{\rho}(i\overleftarrow{\partial}, i\vec{\partial})\phi \tag{18}$$

satisfies the continuity equation

$$J^{\rho}{}_{,\rho} = 0. \tag{19}$$

Here $P_{\rho}(k', k)$ is an $n \times n$ matrix polynomial in the components of the 4-vectors k' and k. The expression (18) may always be symmetrized so that P_{ρ} satisfies the condition

$$\tilde{P}_{\rho}(k, k') = \pm P_{\rho}(k', k), \tag{20}$$

with the upper sign for Bose fields and the lower one for Fermi fields. Clearly, a sufficient condition that (19) should follow from (15) is that P should satisfy

$$(k_{\rho} + k'_{\rho})P^{\rho}(k', k) = 0 \tag{21}$$

for all 4-vectors k and k' such that

$$k^2 = k'^2 = m^2. \tag{22}$$

In particular, it is sufficient if P_{ρ} is proportional to $(k_{\rho} - k'_{\rho})$; that is, if J_{ρ} involves the two-sided derivative $\overleftrightarrow{\partial}_{\rho}$.

To find a necessary condition, we may substitute (16) and (17) into

$$Q = \int d^3\mathbf{x}\, J^0. \tag{23}$$

It is easy to see by inspection that Q will be a constant of the motion if and only if the terms in aa and a^*a^* are identically zero. The condition for this is

$$\tilde{u}_{\alpha}(-\mathbf{k})P^0(k^0, -\mathbf{k}; k^0, \mathbf{k})u_{\beta}(\mathbf{k}) = 0. \tag{24}$$

Since for this case $k'_{\rho} + k_{\rho}$ is in the time direction, this amounts to requiring that (21) should hold for

all the components satisfying the subsidiary conditions. This condition is clearly also sufficient to guarantee that J^ρ satisfies (19). When it is satisfied, Q reduces to

$$Q = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,\phi^\dagger(\mathbf{k})\,\frac{P_0(-k,\,k)}{k^0}\,\phi(\mathbf{k}), \qquad (25)$$

provided that we subtract out the vacuum expectation value in the usual way.

We turn next to the problem of finding all the conserved quantities of this form. Evidently, many different matrix polynomials P_ρ can lead to the same conserved quantity Q. It will be convenient to introduce the variables

$$P = \tfrac{1}{2}(k - k'), \qquad q = \tfrac{1}{2}(k + k'),$$

which in view of (22) satisfy

$$p^2 + q^2 = m^2, \qquad p\cdot q = 0.$$

The conditions (20) and (21) then take the form

$$\tilde{P}_\rho(-p,\,q) = \pm P_\rho(p,\,q), \qquad (26)$$

$$q^\rho P_\rho(p,\,q) = 0. \qquad (27)$$

Since Q depends only on $P_\rho(p,\,0)$, two polynomials which become equal for $q = 0$ lead to the same conservation law. Hence it is unnecessary to consider any terms in q, and we may restrict our considerations to polynomials of p alone. To illustrate the method, we consider some simple examples.

4. EXAMPLES

A. The Scalar Field

For a real one-component field, P_ρ must be an ordinary polynomial in p, and by (26) must be even. Hence we have the sequence of possibilities

$$P_\rho = p_\rho p_\mu, \qquad p_\rho p_\mu p_\nu, \cdots . \qquad (28)$$

All these evidently satisfy (27). The corresponding currents are

$$J_{\rho\mu} = -\phi\,\overleftrightarrow{\partial}_\rho\,\overleftrightarrow{\partial}_\mu\,\phi,$$

$$J_{\rho\mu\nu\sigma} = \phi\,\overleftrightarrow{\partial}_\rho\,\overleftrightarrow{\partial}_\mu\,\overleftrightarrow{\partial}_\nu\,\overleftrightarrow{\partial}_\sigma\,\phi, \qquad (29)$$

etc., and the conserved quantities in momentum space are

$$Q^\mu = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,k^\mu\phi^*(\mathbf{k})\phi(\mathbf{k}),$$

$$Q^{\mu\nu\sigma} = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,k^\mu k^\nu k^\sigma\phi^*(\mathbf{k})\phi(\mathbf{k}), \qquad (30)$$

etc. The first of these is of course the total momentum, and it is easy to verify that $J_{\rho\mu}$ differs from the usual stress–energy tensor

$$T_{\rho\mu} = \phi_{,\rho}\phi_{,\mu} - \tfrac{1}{2}g_{\rho\mu}(\phi^{,\lambda}\phi_{,\lambda} - m^2\phi\phi)$$

by an explicit divergence. In fact,

$$T_{\rho\mu} - J_{\rho\mu} = \tfrac{1}{4}(\phi\phi)_{,\rho\mu} - \tfrac{1}{4}g_{\rho\mu}(\phi\phi)^{,\lambda}{}_{,\lambda},$$

which is always a spatial divergence for $\rho = 0$.

If the field is charged, it has two components which are connected by the charge matrix

$$q = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}.$$

In this case, there are sequences of symmetric matrix polynomials similar to (28), but also an antisymmetric sequence

$$P = qp_\rho, \qquad qp_\rho p_\mu p_\nu, \cdots .$$

The first of these yields the usual charge current $J_\rho = \phi_2\overleftrightarrow{\partial}_\rho\phi_1$, and total charge

$$Q = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,\phi^\dagger(\mathbf{k})q\phi(\mathbf{k}). \qquad (31)$$

B. The Spinor Field

Next we consider a real (Majorana) spinor field satisfying the Dirac equation[3]

$$i\gamma^\mu\psi_{,\mu} = m\psi.$$

The matrix P_ρ in this case must be a linear combination of the sixteen matrices β, $\beta\gamma_\mu$, $\beta\sigma_{\mu\nu}$, $i\beta\gamma_\mu\gamma_5$, $\beta\gamma_5$. However, Q involves only the matrix elements $u^\dagger(\mathbf{p})P_\rho u(\mathbf{p})$. Thus, because of the identities

$$m\bar{u}(\mathbf{p})\gamma_\mu u(\mathbf{p}) = p_\mu\bar{u}(\mathbf{p})u(\mathbf{p}),$$

$$m\bar{u}(\mathbf{p})i\gamma^\mu\gamma^5 u(\mathbf{p}) = \tfrac{1}{2}\epsilon^{\mu\nu\kappa\lambda}p_\nu\bar{u}(\mathbf{p})\sigma_{\kappa\lambda}u(\mathbf{p}),$$

$$m\bar{u}(\mathbf{p})\gamma_5 u(\mathbf{p}) = 0,$$

it is sufficient to consider matrices P_ρ proportional to either β or $\beta\sigma_{\mu\nu}$. Moreover, since

$$p^\mu\bar{u}(\mathbf{p})\sigma_{\mu\nu}u(\mathbf{p}) = 0,$$

it is unnecessary to consider any contractions between p^μ and the spin matrices.

It follows from the symmetry condition (26) that β must be multiplied by an even function of p and $\beta\sigma_{\mu\nu}$ by an odd function. Hence we obtain the two sequences

$$P_\rho = p_\rho p_\mu\beta, \qquad p_\rho p_\mu p_\nu p_\kappa\beta, \cdots ,$$

$$P_\rho = p_\rho\beta\sigma_{\mu\nu}, \qquad p_\rho p_\kappa p_\lambda\beta\sigma_{\mu\nu}, \cdots .$$

[3] We use a Majorana representation of the Dirac matrices in which all four Dirac matrices γ_μ are pure imaginary, and β, $\beta\gamma_\mu$, $\beta\sigma_{\mu\nu}$, $i\beta\gamma_\mu\gamma_5$, $\beta\gamma_5$ are Hermitian. We write $\sigma_{\mu\nu} = \frac{1}{2}i[\gamma_\mu,\,\gamma_\nu]$ and $\bar{\psi} = \psi^\dagger\beta$.

The first two of these yield the currents

$$J_{\rho\mu} = -\bar{\psi}\,\vec{\partial}_\rho\,\vec{\partial}_\mu\,\psi, \qquad J_{\rho\mu\nu} = \bar{\psi}\,\vec{\partial}_\rho\,\sigma_{\mu\nu}\,\psi, \tag{32}$$

and hence the conserved quantities

$$Q^\mu = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,k^\mu\bar{\psi}(\mathbf{k})\psi(\mathbf{k}),$$

$$Q^{\mu\nu} = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,\bar{\psi}(\mathbf{k})\sigma^{\mu\nu}\psi(\mathbf{k}). \tag{33}$$

As before, Q^μ is the total energy–momentum of the field, and $J_{\rho\mu}$ differs from $T_{\rho\mu}$ by an explicit divergence.

C. The Vector Field

For the spin-one field the role of the c-number solutions $u(k)$ is played by the polarization vectors $\varepsilon(\mathbf{k})$. If we denote the matrix elements of P_ρ by $P_\rho{}^{\kappa\lambda}$, then the condition for conservation is

$$q^\rho\varepsilon'_\kappa P_\rho{}^{\kappa\lambda}(p)\varepsilon_\lambda = 0. \tag{34}$$

From the polarization vectors we can form three irreducible tensors,

$$\varepsilon'\cdot\varepsilon,\ \varepsilon'_\kappa\varepsilon_\lambda - \varepsilon'_\lambda\varepsilon_\kappa, \quad\text{and}\quad \varepsilon'_\kappa\varepsilon_\lambda + \varepsilon'_\lambda\varepsilon_\kappa - \tfrac{1}{2}g_{\kappa\lambda}\varepsilon'\cdot\varepsilon.$$

It is unnecessary to consider terms in which any of these are contracted with p, because such terms yield vanishing contributions to Q. Moreover, none of these tensors is itself automatically orthogonal to q. Thus the only way of satisfying (34) is to make P_ρ proportional to p_ρ. Hence we obtain three sets of possible polynomials

$$\begin{gathered} p_\rho p_\mu\varepsilon'\cdot\varepsilon, \qquad p_\rho p_\mu p_\nu p_\sigma\varepsilon'\cdot\varepsilon, \cdots; \\ p_\rho(\varepsilon'_\kappa\varepsilon_\lambda - \varepsilon'_\lambda\varepsilon_\kappa), \cdots; \\ p_\rho p_\mu(\varepsilon'_\kappa\varepsilon_\lambda + \varepsilon'_\lambda\varepsilon_\kappa - \tfrac{1}{2}g_{\kappa\lambda}\varepsilon'\cdot\varepsilon), \cdots. \end{gathered} \tag{35}$$

It is a simple matter to write down the corresponding conserved quantities, but we shall not do so explicitly.

D. The Electromagnetic Field

It is of course possible to regard the electromagnetic field as a special case of the vector field, and the polynomials listed above do indeed yield conserved quantities for this special case. However, most of them are not gauge invariant. The corresponding densities are local functions of the vector potential A_μ but not of the field variables $F_{\mu\nu}$. To find gauge-invariant conserved quantities we could proceed by looking for linear combinations of these polynomials which vanish when ε is replaced by k or ε' by k'. However, it is just as simple to start again, using the $F_{\mu\nu}$ instead of the A_μ as field variables. We have to find polynomials P_ρ such that

$$\tfrac{1}{4}q^\rho f'_{\kappa\lambda}P_\rho{}^{\kappa\lambda,\mu\nu}(p)f_{\mu\nu} = 0 \tag{36}$$

for all $f_{\mu\nu}$ satisfying

$$f_{\mu\nu}k^\nu = 0, \qquad {}^*f_{\mu\nu}k^\nu = 0. \tag{37}$$

We proceed exactly as before. From $f'_{\kappa\lambda}$ and $f_{\mu\nu}$ we can form six irreducible tensors, four symmetric in f' and f, and two antisymmetric. They must be multiplied by even and odd functions of p, respectively. As before, we need not consider terms in which any of the tensors is contracted with p, because these yield vanishing contributions to Q. Two of the six tensors (those corresponding to energy–momentum and zilch) already satisfy (36) by virtue of (37). The remainder must be multiplied by p_ρ.

Three of the six possible types of polynomial are in fact trivial, and yield vanishing conserved quantities Q. These are the ones whose first members are

$$p_\rho p_\mu\tfrac{1}{4}f'^{\kappa\lambda}f_{\lambda\kappa}; \qquad p_\rho p_\mu\tfrac{1}{4}{}^*f'^{\kappa\lambda}f_{\lambda\kappa};$$

$$p_\rho\tfrac{1}{2}(f'_{\mu\lambda}f^\lambda{}_\nu - f'_{\nu\lambda}f^\lambda{}_\mu).$$

[To see this, note that in Q we have $f_{\mu\nu} = i(p_\mu\varepsilon_\nu - p_\nu\varepsilon_\mu)$, $f'_{\mu\nu} = i(p_\mu\varepsilon'_\nu - p_\nu\varepsilon'_\mu)$, with $p^2 = p\cdot\varepsilon = p\cdot\varepsilon' = 0$.] Hence we are left with three families of polynomials, two symmetric ones and one antisymmetric. Their first members are

$$\tfrac{1}{2}(f'_{\rho\lambda}f^\lambda{}_\mu + f'_{\mu\lambda}f^\lambda{}_\rho) - \tfrac{1}{4}g_{\mu\rho}f'_{\kappa\lambda}f^{\lambda\kappa},$$

which yields energy–momentum conservation,

$$p_\rho[\tfrac{1}{2}({}^*f'_{\mu\lambda}f^\lambda{}_\nu + {}^*f'_{\nu\lambda}f^\lambda{}_\mu) - \tfrac{1}{4}g_{\mu\nu}{}^*f'_{\kappa\lambda}f^{\lambda\kappa}],$$

which yields conservation of zilch and, finally,

$$\begin{aligned} p_\rho p_\sigma\tfrac{1}{4}[&(f'_{\kappa\mu}f_{\nu\lambda} + f'_{\kappa\nu}f_{\mu\lambda} + f'_{\lambda\mu}f_{\nu\kappa} + f'_{\lambda\nu}f_{\mu\kappa}) \\ &- \tfrac{1}{2}g_{\mu\nu}(f'_{\kappa\tau}f^\tau{}_\lambda + f'_{\lambda\tau}f^\tau{}_\kappa) \\ &- \tfrac{1}{2}g_{\kappa\lambda}(f'_{\tau\mu}f_\nu{}^\tau + f'_{\tau\nu}f_\mu{}^\tau) + \tfrac{1}{4}g_{\kappa\lambda}g_{\mu\nu}f'_{\alpha\beta}f^{\beta\alpha}]. \end{aligned}$$

These three sets correspond to the three sets (35) obtained for a massive vector field (though of course the correspondence is not one-to-one, since some members of the latter sets cannot be made gauge-invariant).

We note finally that in momentum space the conserved zilch tensor takes the form

$$\begin{aligned} Z^{\mu\nu} = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,&(k^\mu\epsilon^{\nu\sigma\kappa\lambda} + k^\nu\epsilon^{\mu\sigma\kappa\lambda})k_\sigma \\ &\times A^*_\kappa(\mathbf{k})A_\lambda(\mathbf{k}), \end{aligned}$$

or, equivalently, in the Coulomb gauge,

$$Z^{\mu\nu} = \frac{-1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{2k^0}\,\frac{2k^\mu k^\nu}{k^0}\,\mathbf{k}\cdot(\mathbf{A}^*\times\mathbf{A}).$$

5. CONCLUSION

We have shown that for any free field one can find an infinite set of conserved quantities with densities which are local bilinear functions of the fields. Physically, these conservation laws are a direct consequence of the fact that the momentum-space density $\phi(\mathbf{k})\phi^\dagger(\mathbf{k})$ is constant in time. Very few of them can be expected to carry over to the case of interacting fields, apart from the ones associated with known symmetry properties, though of course a conservation law valid for interacting fields must reduce to one of those listed when the interaction is turned off. Most of these laws are probably of little practical importance, though for some purposes it may be useful to have an exhaustive list.

ACKNOWLEDGMENTS

The author is indebted to Dr. J. Charap and Dr. G. Guralnik for valuable discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Tensorial Description of Neutrinos

R. PENNEY
Ford Motor Company, Scientific Laboratory, Dearborn, Michigan
(Received 5 October 1964)

It is shown that the equation $T_\nu{}^\omega(T^{\rho\lambda} + T_d{}^{\rho\lambda})_{|\rho} + T^{\mu\omega}(T_{\mu\nu} - T_{d\mu\nu})^{|\lambda} = 0$, where $T_{\mu\nu}$ is a null antisymmetric tensor and T_d its dual, is necessary and sufficient that there be a two-component Majorana neutrino field. That is, the neutrino field may be described solely in terms of tensorial quantities and operations, without the need for spinors.

I. INTRODUCTION

THE statement is often made in the literature of physics that spinorial quantities, for example the Dirac ψ, are necessary for the description of the properties of the fields of physics. The implication is that it is not possible to use only tensorial quantities to describe nature, and such is simply not the case. One can, in fact, with a little effort, avoid the use of spinors in describing the electron, for example.

Whittaker[1] showed in 1936 that the Dirac equation could be written in vectorial form. Nonetheless, he did not really complete the process. Our aim in the present work is to write down a single equation for a single tensorial quantity which will replace the usual equations for the neutrino. That is, we will show that a tensor equation for an antisymmetric null second-rank tensor is necessary and sufficient to describe the two-component neutrino.

Our analysis might be presumed to be rather academic in scope and aim, but such is not necessarily the case. Quantum mechanics and general relativity are not unified yet, and any avenue of approach to such a unification must necessarily accomplish the aim of describing fermions in a way compatible with the formalism of tensor analysis.

Another point of interest which our analysis is applicable to is the following. Classically at least, the energy and momentum of a field is mathematically described on the basis of a Lagrangian formalism. That is, conventionally one *defines* the energy–momentum tensor as that which is conserved due to translational invariance of the Lagrangian. If, for example, one could describe the neutrino tensorially, one might then find a vastly different Lagrangian formalism, which in turn would lead to a different coupling of neutrinos and gravitation.

Before proceeding, we remark that we will use a very common notation, with Minkowski coordinates ($x_4 = ict$), such that the spatial components of vectors are real, and the time component imaginary. In general, the notation is that of Roman.[2]

II. MAJORANA NEUTRINOS

We consider a two-component description of the

[1] E. T. Whittaker, Proc. Roy. Soc. (London) **A158**, 38 (1936).

[2] P. Roman, *Theory of Elementary Particles* (Interscience Publishers, Inc., New York, 1961), 2nd ed.

5. CONCLUSION

We have shown that for any free field one can find an infinite set of conserved quantities with densities which are local bilinear functions of the fields. Physically, these conservation laws are a direct consequence of the fact that the momentum-space density $\phi(\mathbf{k})\phi^\dagger(\mathbf{k})$ is constant in time. Very few of them can be expected to carry over to the case of interacting fields, apart from the ones associated with known symmetry properties, though of course a conservation law valid for interacting fields must reduce to one of those listed when the interaction is turned off. Most of these laws are probably of little practical importance, though for some purposes it may be useful to have an exhaustive list.

ACKNOWLEDGMENTS

The author is indebted to Dr. J. Charap and Dr. G. Guralnik for valuable discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Tensorial Description of Neutrinos

R. PENNEY
Ford Motor Company, Scientific Laboratory, Dearborn, Michigan
(Received 5 October 1964)

It is shown that the equation $T_\nu{}^\omega(T^{\rho\lambda} + T_d{}^{\rho\lambda})_{|\rho} + T^{\mu\omega}(T_{\mu\nu} - T_{d\mu\nu})^{|\lambda} = 0$, where $T_{\mu\nu}$ is a null antisymmetric tensor and T_d its dual, is necessary and sufficient that there be a two-component Majorana neutrino field. That is, the neutrino field may be described solely in terms of tensorial quantities and operations, without the need for spinors.

I. INTRODUCTION

THE statement is often made in the literature of physics that spinorial quantities, for example the Dirac ψ, are necessary for the description of the properties of the fields of physics. The implication is that it is not possible to use only tensorial quantities to describe nature, and such is simply not the case. One can, in fact, with a little effort, avoid the use of spinors in describing the electron, for example.

Whittaker[1] showed in 1936 that the Dirac equation could be written in vectorial form. Nonetheless, he did not really complete the process. Our aim in the present work is to write down a single equation for a single tensorial quantity which will replace the usual equations for the neutrino. That is, we will show that a tensor equation for an antisymmetric null second-rank tensor is necessary and sufficient to describe the two-component neutrino.

Our analysis might be presumed to be rather academic in scope and aim, but such is not necessarily the case. Quantum mechanics and general relativity are not unified yet, and any avenue of approach to such a unification must necessarily accomplish the aim of describing fermions in a way compatible with the formalism of tensor analysis.

Another point of interest which our analysis is applicable to is the following. Classically at least, the energy and momentum of a field is mathematically described on the basis of a Lagrangian formalism. That is, conventionally one *defines* the energy–momentum tensor as that which is conserved due to translational invariance of the Lagrangian. If, for example, one could describe the neutrino tensorially, one might then find a vastly different Lagrangian formalism, which in turn would lead to a different coupling of neutrinos and gravitation.

Before proceeding, we remark that we will use a very common notation, with Minkowski coordinates ($x_4 = ict$), such that the spatial components of vectors are real, and the time component imaginary. In general, the notation is that of Roman.[2]

II. MAJORANA NEUTRINOS

We consider a two-component description of the

[1] E. T. Whittaker, Proc. Roy. Soc. (London) **A158**, 38 (1936).

[2] P. Roman, *Theory of Elementary Particles* (Interscience Publishers, Inc., New York, 1961), 2nd ed.

neutrino[3] given by

$$\gamma_\mu \partial_\mu \psi = 0, \tag{1}$$

$$\psi = \psi^c, \tag{2}$$

where

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\delta_{\mu\nu}, \tag{3}$$

$$\psi^c \equiv C\bar{\psi}^t = C\gamma_4^t \psi^*, \tag{4}$$

$$\gamma_\mu^t = -C^{-1}\gamma_\mu C. \tag{5}$$

The symbol t denotes ordinary transpose, * denotes complex conjugate. C is the Schwinger matrix.

We use the Majorana description solely for convenience. The Weyl or Lee–Yang description would do as well for our analysis. With our description, the only nonzero bilinear covariants available are

$$V_\mu = i\bar{\psi}\gamma_\mu \psi, \tag{6}$$

$$T_{\mu\nu} = \tfrac{1}{2} i\bar{\psi}(\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu)\psi, \tag{7}$$

$$^*T_{\mu\nu} = \tfrac{1}{2} i\bar{\psi}\gamma_5(\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu)\psi, \tag{8}$$

where $^*T_{\mu\nu}$ is the dual of $T_{\mu\nu}$, i.e.,

$$^*T_{\mu\nu} = -\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta} T_{\alpha\beta}. \tag{9}$$

Using results reported elsewhere,[4] we know that

$$V_\mu V_\mu = 0, \quad T_{\mu\nu} T_{\mu\nu} = 0, \quad T_{\mu\nu} {}^*T_{\mu\nu} = 0, \tag{10}$$

$$T_{\mu\nu} V_\nu = 0, \quad {}^*T_{\mu\nu} V_\nu = 0, \tag{11}$$

$$T_{\mu\nu} T_{\rho\nu} = V_\mu V_\rho = -{}^*T_{\mu\nu} {}^*T_{\rho\nu}, \tag{12}$$

where the summation convention on repeated indices is used.

It follows that $T_{\mu\nu}$ has but *four* independent (real) components. These four components correspond to the *two* independent complex components of ψ.

For further development, it is convenient to use a Majorana representation[5] of the γ_μ. In this representation,

$$\psi^c = \psi^*, \tag{13}$$

so the four components of ψ are *real*. We readily calculate the components of our bilinears. For example,

$$V_1 = \psi_2^2 + \psi_4^2 - \psi_1^2 - \psi_3^2, \tag{14}$$

$$V_2 = -2(\psi_1 \psi_4 + \psi_2 \psi_3), \tag{15}$$

and so on.

Now let us see what differential equations the bilinears obey as a result of ψ obeying the Dirac equation. We easily see that

$$\partial_\mu T_{\mu\nu} = P_\nu, \tag{16}$$

$$\partial_\mu {}^*T_{\mu\nu} = Q_\nu, \tag{17}$$

where we have defined

$$P_\nu \equiv i\bar{\psi}(\partial_\nu \psi) - i(\partial_\nu \bar{\psi})\psi, \tag{18}$$

$$Q_\nu \equiv i\bar{\psi}\gamma_5(\partial_\nu \psi) - i(\partial_\nu \bar{\psi})\gamma_5 \psi. \tag{19}$$

Next, using the results of Whittaker[1] as a guide, we may easily show that

$$(\partial_\lambda V_\mu) T_{\mu\nu} = V_\nu P_\lambda, \tag{20}$$

$$(\partial_\lambda V_\mu^*) T_{\mu\nu} = -V_\nu Q_\lambda, \tag{21}$$

after a trivial calculation which is best done by calculating a few components of each side of the equations.

III. THE CONVERSE PROCESS

We know that if ψ obeys the Dirac equation, our bilinears obey certain differential equations. We may then investigate the problem of whether the opposite is true.

The answer is that we may write down a set of differential equations for $T_{\mu\nu}$ which allow us to retrieve our ψ and its Dirac equation. To see this, write the Dirac equation in our representation:

$$A \equiv -\partial_1 \psi_4 + \partial_2 \psi_1 - \partial_3 \psi_3 - i\partial_4 \psi_4 = 0, \tag{22a}$$

$$B \equiv \partial_1 \psi_3 + \partial_2 \psi_2 - \partial_3 \psi_4 - i\partial_4 \psi_3 = 0, \tag{22b}$$

$$C \equiv \partial_1 \psi_2 - \partial_2 \psi_3 - \partial_3 \psi_1 + i\partial_4 \psi_2 = 0, \tag{22c}$$

$$D \equiv -\partial_1 \psi_1 - \partial_2 \psi_4 - \partial_3 \psi_2 + i\partial_4 \psi_1 = 0. \tag{22d}$$

Now, consider the vector relation expressed by

$$\partial_\mu \{T_{\mu\nu} + i\, {}^*T_{\mu\nu}\} = P_\nu + iQ_\nu, \tag{23}$$

where the i-factors are to ensure the appropriate reality conditions. By writing out our four equations explicitly, we have

$$(\psi_1 + \psi_3)A - (\psi_4 + \psi_2)D + (\psi_4 - \psi_2)B + (\psi_3 - \psi_1)C = 0, \tag{24a}$$

$$(\psi_4 - \psi_2)A + (\psi_1 - \psi_3)D + (\psi_3 + \psi_1)B + (\psi_2 + \psi_4)C = 0, \tag{24b}$$

$$(\psi_2 + \psi_4)A - (\psi_1 + \psi_3)D + (\psi_1 - \psi_3)B + (\psi_2 - \psi_4)C = 0, \tag{24c}$$

$$(\psi_3 - \psi_1)A + (\psi_2 - \psi_4)D - (\psi_2 + \psi_4)B - (\psi_1 + \psi_3)C = 0. \tag{24d}$$

Unless the determinant of coefficients vanishes, we will have that A, B, C, D vanish. However the determinant is, explicitly,

$$-(\psi_1^2 + \psi_3^2 - \psi_2^2 - \psi_4^2)^2 = -V_1^2, \tag{25}$$

and therefore vanishes only if V_1, the first component of V_μ should vanish. In general, it is not possible for V_1 to vanish. That is, we may by accident find that V_1 vanishes in some particular Lorentz frame,

[3] Reference 2, p. 306.
[4] R. Penney, J. Math. Phys. 5, 1657 (1964).
[5] Reference 2, p. 126.

but by a simple transformation we can always obtain a frame in which the determinant does not vanish.

We must conclude that we have a vector equation which implies ψ obeys the Dirac equation. Thus, if it is possible to obtain ψ from knowledge of $T_{\mu\nu}$, we have necessary and sufficient conditions.

Whittaker[1] has shown explicitly that a null self-dual tensor is equivalent to a two-component spinor. We will not repeat the analysis here, but it is clear from his results that, given a null antisymmetric tensor $T_{\mu\nu}$, we have always a corresponding spinor (φ_1, φ_2). The real and imaginary parts of φ_1 and φ_2 correspond to our four *real* components $\psi_1\psi_2\psi_3\psi_4$. It is easy, but unnecessary, to detail the prescription for ψ_i in terms of φ_1 and φ_2 and thus in terms of $T_{\mu\nu}$. We are already in possession of such a correspondence implicitly.

IV. TENSORIAL DESCRIPTION OF NEUTRINOS

We are now able to write down a single tensor equation for a single tensor quantity which is equivalent to a two-component spinor description of the neutrino.

For this purpose we will, at the risk of some confusion, use real coordinates, as appropriate to the tensor analysis of general relativity. First, we recapitulate what we have found.

The analysis in the previous section shows that, given a null antisymmetric tensor $T_{\mu\nu}$,

$$T_{\mu\nu}T_{\mu\nu} = 0, \qquad T_{\mu\nu}{}^*T_{\mu\nu} = 0, \tag{10}$$

we may derive the vectors V_μ, P_μ, Q_μ by the prescriptions (due to Whittaker[1]):

$$T_{\mu\nu}T_{\rho\nu} = V_\mu V_\rho, \tag{12}$$

$$(\partial_\lambda V_\mu)T_{\mu\nu} = V_\nu P_\lambda, \tag{20}$$

$$(\partial_\lambda V_\mu{}^*)T_{\mu\nu} = -V_\nu Q_\lambda, \tag{21}$$

and may also find the spinor $(\varphi_1\varphi_2)$ or the Majorana wavefunction ψ by Whittaker's prescription.[1]

If, then, we subject this tensor to the differential equation

$$\partial_\mu\{T_{\mu\nu} + i\,{}^*T_{\mu\nu}\} = P_\nu + iQ_\nu, \tag{23}$$

we indeed have neutrino physics. However, this prescription does not fulfill our aim.

Let us therefore define the combinations

$$S_{\mu\nu} \equiv T_{\mu\nu} + i\,{}^*T_{\mu\nu}, \tag{26}$$

$$A_{\mu\nu} \equiv T_{\mu\nu} - i\,{}^*T_{\mu\nu}, \tag{27}$$

$$R_\lambda \equiv P_\lambda + iQ_\lambda. \tag{28}$$

Then, if our differential equation is obeyed we have

$$(\partial_\lambda V_\mu)A_{\mu\nu} = V_\nu\partial_\rho S_{\rho\lambda} \tag{29}$$

by eliminating R_λ. We then use

$$A_{\mu\nu}V_\nu = 0 \tag{30}$$

to write

$$-V_\mu\partial_\lambda A_{\mu\nu} = V_\nu\partial_\rho S_{\rho\lambda}, \tag{31}$$

and multiply by V_δ here to obtain

$$V_\delta V_\nu\partial_\rho S_{\rho\lambda} + V_\delta V_\mu\partial_\lambda A_{\mu\nu} = 0. \tag{32}$$

We then use the fact that $V_\alpha V_\beta$ can be expressed in terms of $T_{\mu\nu}$, to write

$$T_{\delta\omega}T_{\nu\omega}\partial_\rho S_{\rho\lambda} + T_{\delta\omega}T_{\mu\omega}\partial_\lambda A_{\mu\nu} = 0, \tag{33}$$

which will be so certainly if

$$T_{\nu\omega}\partial_\rho S_{\rho\lambda} + T_{\mu\omega}\partial_\lambda A_{\mu\nu} = 0. \tag{34}$$

In terms of $T_{\mu\nu}$ alone, this last equation is

$$T_{\nu\omega}\partial_\rho(T_{\rho\lambda} + i\,{}^*T_{\rho\lambda}) + T_{\mu\omega}\partial_\lambda(T_{\mu\nu} - i\,{}^*T_{\mu\nu}) = 0. \tag{35}$$

We write our equation for $T_{\mu\nu}$ in the form appropriate to general relativity, by using real coordinates and the covariant–contravariant notation. With a bar (|) denoting covariant differentiation, we have

$$T^\omega_\nu(T^{\rho\lambda} + T^{\rho\lambda}_{\mathrm{d}})_{|\rho} + T^{\mu\omega}(T_{\mu\nu} - T_{\mathrm{d}\mu\nu})^{|\lambda} = 0, \tag{36}$$

where $T^{\rho\lambda}_{\mathrm{d}}$ is the dual tensor to $T^{\rho\lambda}$.

V. CONCLUSIONS

We have shown that a single tensorial equation involving a single tensorial quantity may be written down which is necessary and sufficient to ensure that there exists a ψ function which obeys the Dirac equation.

The tensor equation which is equivalent to the Majorana theory of the neutrino is *not* very simple, but it does exist. We could, in fact, describe the neutrino *solely* in terms of a null antisymmetric tensor if we so desired, at least classically. Such a procedure would be decidedly inelegant in general, but could be useful for certain purposes.

At least our results should serve to remove some of the mystery surrounding the use of spinors in physics. It may be much *simpler* to describe the neutrino with spinor calculus, but it is not necessary.

A similar program may be carried out for the general four-component electron equation, of course, but there seems no compelling reason to do so at the present time. Whittaker's[1] results show that, in principle, one does not need spinors to describe the electron.

Geometrization of a Massive Scalar Field

R. PENNEY
Ford Motor Company, Scientific Laboratory, Dearborn, Michigan
(Received 16 November 1964)

A set of necessary and sufficient conditions involving only the metric tensor and the Einstein tensor $G_{\mu\nu}$ are given which thereby guarantee that $G_{\mu\nu}$ may represent the stress-energy tensor of a massive real scalar field. One of the conditions is nonlocal, and is a demand that the energy tensor vanish at spatial infinity.

INTRODUCTION

IN 1925, Rainich[1] succeeded in showing that one could geometrize the Maxwell source-free electromagnetic field. That is, Rainich showed that there was a set of algebraic and differential conditions which one could impose upon the Einstein tensor $G_{\mu\nu}$ which would ensure that $G_{\mu\nu}$ would represent the Maxwell stress-energy tensor.

Wheeler[2] and his collaborators have used the Rainich results in attempting to understand the diverse connections between geometry and the classical field theories of physics.

In the present analysis, we wish to show that there exists a set of algebraic and differential conditions which are necessary and sufficient that $G_{\mu\nu}$ should represent the stress-energy tensor of a massive, real, scalar field, i.e., a "meson." Previously, we have displayed the conditions for a massless[3] real scalar field.

The problem of the massive scalar field has been treated by Peres,[4] who found certain necessary conditions which must be true for a tensor to represent the stress-energy tensor of a "meson." However, Peres did not find all of the necessary conditions, and did not show sufficiency.

In the next section, we will examine the necessary conditions that a tensor must fulfill in order that it may represent the energy tensor of a meson. We will then prove the sufficiency of the conditions.

We adopt a notation as follows. The real coordinates are $x^0x^1x^2x^3$ where x^0 is ct. The metric is such that, locally, the signature is $(1, -1, -1, -1)$. The Einstein tensor $G_{\mu\nu}$ is given in terms of the Ricci tensor by

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu}. \tag{1}$$

The trace of a mixed tensor is denoted by the same symbol as the tensor itself, without indices. For example,

$$T \equiv g^{\mu\nu}T_{\mu\nu} = T^{\mu}_{\mu}. \tag{2}$$

Finally, covariant differentiation is indicated by a slash between indices; with this notation, the contracted Bianchi identities are

$$G_{\mu\nu}{}^{|\nu} \equiv 0. \tag{3}$$

II. NECESSARY CONDITIONS

We consider the usual stress-energy tensor for a massive scalar field, viz.

$$T_{\mu\nu} = \varphi_\mu\varphi_\nu - \tfrac{1}{2}\varphi_\rho\varphi^\rho g_{\mu\nu} + \tfrac{1}{2}k^2\varphi^2 g_{\mu\nu}, \tag{4}$$

where

$$\varphi_\mu \equiv \varphi_{|\mu} = \partial\varphi/\partial x^\mu. \tag{5}$$

We now define the tensor

$$S_{\mu\nu} = T_{\mu\nu} - \tfrac{1}{4}Tg_{\mu\nu} = \varphi_\mu\varphi_\nu - \tfrac{1}{4}\varphi_\rho\varphi^\rho g_{\mu\nu}, \tag{6}$$

and also the two tensors

$$U_{\mu\nu} \equiv S_{\mu\nu} + \tfrac{1}{4}(\tfrac{4}{3}S_{ab}S^{ab})^{\frac{1}{2}}g_{\mu\nu} = \varphi_\mu\varphi_\nu, \tag{7}$$

$$W_{\mu\nu} \equiv S_{\mu\nu} - \tfrac{1}{4}(\tfrac{4}{3}S_{ab}S^{ab})^{\frac{1}{2}}g_{\mu\nu} = \varphi_\mu\varphi_\nu - \tfrac{1}{2}\varphi_\rho\varphi^\rho g_{\mu\nu}. \tag{8}$$

Following Peres,[4] we could note that the necessary and sufficient conditions that $U_{\mu\nu}$ be of the form $\varphi_\mu\varphi_\nu$ where φ_μ is a 4-vector are

$$U_{mn}U_{rs} = U_{mr}U_{ns}, \tag{9}$$

but we prefer in the sequel to use the tensor $W_{\mu\nu}$. The reason for our choice is that we already know the conditions which must be imposed upon $W_{\mu\nu}$ to guarantee its form, viz.,

$$W^{\mu\alpha}W_{\nu\alpha} = \tfrac{1}{4}W^2g^\mu_\nu, \tag{10a}$$

$$W < 0, \tag{10b}$$

$$W_{00} > 0, \tag{10c}$$

$$(W^\alpha_\rho - \tfrac{1}{2}Wg^\alpha_\rho)\{W_{\alpha\beta|\gamma} - W_{\alpha\gamma|\beta}\} = 0, \tag{10d}$$

as previously shown.[3]

[1] G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).
[2] J. A. Wheeler, *Geometrodynamics*, (Academic Press, Inc., New York, 1962).
[3] R. Penney, Phys. Letters **11**, 228 (1964).
[4] A. Peres, Bull. Res. Council Israel, **9F**, 129 (1960).

The conditions on $W_{\mu\nu}$ are necessary and sufficient that $W_{\mu\nu}$ be of the desired form, with φ_μ the gradient of a scalar.

The relation between $T_{\mu\nu}$ and $W_{\mu\nu}$ may be also written as

$$T_{\mu\nu} - \tfrac{1}{4}Tg_{\mu\nu} = W_{\mu\nu} - \tfrac{1}{4}Wg_{\mu\nu}, \tag{11}$$

so that, in fact

$$T_{\mu\nu} = W_{\mu\nu} + \lambda^2 g_{\mu\nu}, \tag{12}$$

where λ^2 is a positive-definite scalar function.

We need now write conditions on $T_{\mu\nu}$ which will ensure that λ^2 is no longer arbitrary, but is in fact of the form $\frac{1}{2}k^2\varphi^2$, where φ is the scalar function which determines φ_μ.

First, suppose that λ^2 is indeed of the form desired. Then, we find that

$$\lambda_{|\mu}\lambda_{|\nu} = \tfrac{1}{2}k^2(W_{\mu\nu} - \tfrac{1}{2}Wg_{\mu\nu}). \tag{13}$$

Conversely, if this latter condition is obeyed, we will have

$$\lambda_{|\mu} = \pm(k/\sqrt{2})\varphi_\mu = \pm(k/\sqrt{2})\varphi_{|\mu}, \tag{14}$$

where

$$\lambda = \pm(k/\sqrt{2})\varphi + c_1, \tag{15}$$

where c_1 is a fixed constant.

Now, physically speaking, $T_{\mu\nu}$ must be a bounded tensor in the sense that, for large spatial distances, $T_{\mu\nu}$ should vanish. Otherwise, the energy and momentum of the field described by $T_{\mu\nu}$ would not be bounded.

On the other hand, for large spatial distances, the metric tensor must approach Euclidean values, unless one imagines distant sources of gravitation. We are thereby led to demand that

$$\lim T_{\mu\nu} = 0, \qquad x^i \to \infty. \tag{16}$$

The boundary condition on $T_{\mu\nu}$, which seems *necessary* on physical grounds, is sufficient to eliminate the undetermined constant c_1, as is readily appreciated.

III. SUFFICIENCY OF THE CONDITIONS

Let $T_{\mu\nu}$ be a tensor which is restricted by the following conditions:

$$T_{\mu\nu} = T_{\nu\mu}, \tag{17a}$$

$$T_{\mu\nu}{}^{|\nu} = 0, \tag{17b}$$

$$T_{00} > 0, \tag{17c}$$

$$T < 0, \tag{17d}$$

$$T^{\mu\nu}T_{\mu\nu} \geq T^2, \tag{17e}$$

and the conditions given by Eqs. (10a), (10d), (13), and (16), where $W_{\mu\nu}$ is given by Eq. (8), $S_{\mu\nu}$ by (6),

$$\lambda^2 \equiv \tfrac{1}{4}(T - W), \tag{18}$$

and k^2 is a fixed contant.

Using the tensor conditions, it follows immediately that $W_{\mu\nu}$ is of the form

$$W_{\mu\nu} = \varphi_\mu\varphi_\nu - \tfrac{1}{2}\varphi_\rho\varphi^\rho g_{\mu\nu}, \tag{19a}$$

$$\varphi_\mu = \varphi_{|\mu}, \tag{19b}$$

as we have previously shown. Thus, $T_{\mu\nu}$ is determined to be

$$T_{\mu\nu} = \varphi_\mu\varphi_\nu - \tfrac{1}{2}\varphi_\rho\varphi^\rho g_{\mu\nu} + \lambda^2 g_{\mu\nu}. \tag{20}$$

Next, applying the conditions on λ, we have

$$\lambda_{|\mu}\lambda_{|\nu} = \tfrac{1}{2}k^2\varphi_{|\mu}\varphi_{|\nu}, \tag{21}$$

which gives unambiguously that

$$\lambda = \pm(k/\sqrt{2})\varphi + c_1. \tag{22}$$

Therefore, we have the result

$$T_{\mu\nu} = \varphi_\mu\varphi_\nu - \tfrac{1}{2}\varphi_\rho\varphi^\rho g_{\mu\nu} + (\tfrac{1}{2}k^2\varphi^2 \pm \sqrt{2}c_1 k\varphi + c_1^2)g_{\mu\nu}, \tag{23}$$

which will vanish for large spatial distances only if c_1 vanishes.

Before closing this section, we must emphasize that all of our conditions are conditions on the tensor $T_{\mu\nu}$. λ^2 and $W_{\mu\nu}$ may be expressed entirely in terms of $T_{\mu\nu}$ and are introduced only to avoid clumsy expressions. In particular, we have

$$\lambda^2 \equiv \tfrac{1}{4}T + [\tfrac{1}{12}(T^{\mu\nu}T_{\mu\nu} - \tfrac{1}{4}T^2)]^{\frac{1}{2}}. \tag{24}$$

Thus, we have accomplished our aim. By replacing $T_{\mu\nu}$ in our conditions by $G_{\mu\nu}$, we have geometrized the massive real scalar field, in the sense of the Rainich geometrization of the Maxwell field.

A geometry in which $G_{\mu\nu}$ is restricted by our conditions is interpreted as implying the presence of a "meson" field. Such an interpretation follows from the usual field equations of general relativity,

$$G_{\mu\nu} = \kappa T_{\mu\nu}. \tag{25}$$

IV. THE LIMITING CASE OF ZERO MASS

For the massless meson, we know that

$$T^{\mu\nu}T_{\mu\nu} = T^2, \tag{26}$$

whence we see that $W_{\mu\nu}$ reduces to $T_{\mu\nu}$, or, that is, λ^2 vanishes. The constant k^2 must also vanish, and

our conditions reduce to

$$T_{\mu\nu} = T_{\nu\mu}, \tag{17a}$$

$$T_{\mu\nu}{}^{|\nu} = 0, \tag{17b}$$

$$T_{00} > 0, \tag{17c}$$

$$T < 0, \tag{17d}$$

$$T^{\mu\alpha}T_{\nu\alpha} = \tfrac{1}{4}T^2 g^{\mu}_{\nu}, \tag{27}$$

$$(T^{\alpha}_{\rho} - \tfrac{1}{2}Tg^{\alpha}_{\rho})\{T_{\alpha\beta|\,\gamma} - T_{\alpha\gamma|\,\beta}\} = 0. \tag{28}$$

and Eq. (16).

Aside from the boundary condition, these restrictions are the conditions previously[3] found for the massless meson. The symmetry, and the vanishing divergence of $T_{\mu\nu}$ are trivial conditions since the Einstein tensor obeys such identities.

V. CONCLUSIONS

We have found necessary and sufficient conditions which must be imposed upon a Riemannian geometry in order that we may consistently interpret the geometry in terms of a massive "meson" field.

Analogously to the development of the Maxwell field in terms of geometry,[2] the present analysis permits a geometrical interpretation of a classical field of physics.

Further analysis of the geometrodynamical consequences of our conditions may be expected to lead to deeper understanding of geometrodynamics[2] itself.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Bosons and Fermions

R. PENNEY
Scientific Laboratory, Ford Motor Company, Dearborn, Michigan
(Received 21 December 1964)

It is proven that one cannot construct boson creation and annihilation operators from a finite number of fermion operators. The proof follows from the isomorphism of the fermion algebra and the algebra of Dirac matrices.

I. INTRODUCTION

IN the present analysis, we wish to address ourselves to the problem of "making bosons from fermions." Before proceeding further, we must clarify this concept.

As usual, we consider a *fermion* field to be described by a set of annihilation and creation operators in the Fock scheme. The anticommutation rules for these operators are the usual ones. We wish to investigate the possibility of combining such operators to produce a set of *boson* creation and annihilation operators.

The connection between boson and fermion operators has been studied by Case,[1] who showed that one could not, for example, produce a theory of gravitons using quadrilinear combinations of the operators for a two-component neutrino field. Our investigation will be more restrictive than Case's since we will consider only a finite number of fermion operators, but more general in that we allow more general combinations of the fermion operators.

We intend to prove that one cannot form a boson creation operator from a finite number of fermion operators. Our result may help to explain, for example, why the creation operators for Cooper pairs[2] in the BCS theory of superconductivity retain their Fermi–Dirac statistics.

II. TWO-FERMION PROOF

We consider the possibility of constructing a combination of two fermion creation and annihilation operators to make a boson creation operator. Let us suppose, therefore, that we have two operators A_1, A_2 and their Hermitian conjugates obeying the rules

$$[A_i, A_j]_+ = 0, \quad [A_i, A_j^*]_+ = \delta_{ij}, \quad [A_i^*, A_j^*]_+ = 0. \tag{1}$$

[1] K. M. Case, Phys. Rev. **106**, 1316 (1957).

[2] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1180 (1957).

our conditions reduce to

$$T_{\mu\nu} = T_{\nu\mu}, \tag{17a}$$

$$T_{\mu\nu}{}^{|\nu} = 0, \tag{17b}$$

$$T_{00} > 0, \tag{17c}$$

$$T < 0, \tag{17d}$$

$$T^{\mu\alpha}T_{\nu\alpha} = \tfrac{1}{4}T^2 g^{\mu}_{\nu}, \tag{27}$$

$$(T^{\alpha}_{\rho} - \tfrac{1}{2}Tg^{\alpha}_{\rho})\{T_{\alpha\beta|\,\gamma} - T_{\alpha\gamma|\,\beta}\} = 0. \tag{28}$$

and Eq. (16).

Aside from the boundary condition, these restrictions are the conditions previously[3] found for the massless meson. The symmetry, and the vanishing divergence of $T_{\mu\nu}$ are trivial conditions since the Einstein tensor obeys such identities.

V. CONCLUSIONS

We have found necessary and sufficient conditions which must be imposed upon a Riemannian geometry in order that we may consistently interpret the geometry in terms of a massive "meson" field.

Analogously to the development of the Maxwell field in terms of geometry,[2] the present analysis permits a geometrical interpretation of a classical field of physics.

Further analysis of the geometrodynamical consequences of our conditions may be expected to lead to deeper understanding of geometrodynamics[2] itself.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Bosons and Fermions

R. PENNEY
Scientific Laboratory, Ford Motor Company, Dearborn, Michigan
(Received 21 December 1964)

It is proven that one cannot construct boson creation and annihilation operators from a finite number of fermion operators. The proof follows from the isomorphism of the fermion algebra and the algebra of Dirac matrices.

I. INTRODUCTION

IN the present analysis, we wish to address ourselves to the problem of "making bosons from fermions." Before proceeding further, we must clarify this concept.

As usual, we consider a *fermion* field to be described by a set of annihilation and creation operators in the Fock scheme. The anticommutation rules for these operators are the usual ones. We wish to investigate the possibility of combining such operators to produce a set of *boson* creation and annihilation operators.

The connection between boson and fermion operators has been studied by Case,[1] who showed that one could not, for example, produce a theory of gravitons using quadrilinear combinations of the operators for a two-component neutrino field. Our investigation will be more restrictive than Case's since we will consider only a finite number of fermion operators, but more general in that we allow more general combinations of the fermion operators.

We intend to prove that one cannot form a boson creation operator from a finite number of fermion operators. Our result may help to explain, for example, why the creation operators for Cooper pairs[2] in the BCS theory of superconductivity retain their Fermi–Dirac statistics.

II. TWO-FERMION PROOF

We consider the possibility of constructing a combination of two fermion creation and annihilation operators to make a boson creation operator. Let us suppose, therefore, that we have two operators A_1, A_2 and their Hermitian conjugates obeying the rules

$$[A_i, A_j]_+ = 0, \quad [A_i, A_j^*]_+ = \delta_{ij}, \quad [A_i^*, A_j^*]_+ = 0. \tag{1}$$

[1] K. M. Case, Phys. Rev. **106**, 1316 (1957).

[2] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1180 (1957).

We then define the combinations

$$B_1 \equiv A_1 + A_1^*, \qquad B_3 = A_2 + A_2^*, \tag{2}$$

$$B_2 \equiv i(A_1 - A_1^*), \qquad B_4 = i(A_2 - A_2^*), \tag{3}$$

and we thereby summarize the properties of A_1, A_2 as

$$B_\mu B_\nu + B_\nu B_\mu = 2\delta_{\mu\nu}, \tag{4}$$

$$B_\mu^* = B_\mu. \tag{5}$$

The set of operators B_μ is thus seen to be isomorphic to the Dirac γ-matrices. As usual, therefore, we may form the Clifford algebra[3] of the B_μ with the members

$$1, \quad B_\mu, \quad iB_\mu B_\nu, \quad iB_\mu B_\nu B_\rho, \quad B_1B_2B_3B_4, \tag{6}$$

and generically denote the 16 members by Γ_μ, where Γ_1 is the unit operator.

We may now use all of the well-known properties of the Γ_μ to solve our problem. We ask whether a function of the Γ_μ, say $f(\Gamma_\mu)$ exists with the property:

$$[f, f^*]_- = 1, \tag{7}$$

which is the minimal property of a boson operator which we must demand.

First we note that, most generally,

$$f = \sum_{\mu=1}^{16} a_\mu \Gamma_\mu \tag{8}$$

where the a_μ are c-numbers. The properties of the Γ_μ ensure that no powers of Γ_μ occur. We separate the unit element so that

$$f = a_1 + \sum_{\mu=2}^{16} a_\mu \Gamma_\mu, \tag{9}$$

$$f^* = a_1^* + \sum_{\mu=2}^{16} a_\mu^* \Gamma_\mu, \tag{10}$$

and demand that

$$\sum_{\mu=2}^{16} \sum_{\nu=2}^{16} a_\mu a_\nu^* [\Gamma_\mu \Gamma_\nu]_- = 1. \tag{11}$$

The terms in a_1, a_1^* obviously commute with all others, so our remaining sums do not include Γ_1. Since Γ_1 does not occur in the series, we may use the well-known fact that each Γ_μ commutes with eight other Γ_ν and anticommutes with eight others. Thus every term in our series has the property

$$[\Gamma_\mu \Gamma_\nu]_- = 0, \quad \text{or} \tag{12}$$

$$= 2\Gamma_\mu \Gamma_\nu. \tag{13}$$

[3] P. Roman, *Theory of Elementary Particles* (North-Holland Publishing Company, Amsterdam, 1961), 2nd ed., p. 114.

Our demand therefore reduces to

$$\sum_{\mu=2}^{16}{}' \sum_{\nu=2}^{16}{}' a_\mu a_\nu^* \Gamma_\mu \Gamma_\nu = \frac{1}{2}, \tag{14}$$

where the primes denote the omission of the vanishing terms.

Now we recall that $\Gamma_\mu \Gamma_\nu = 1$ if and only if $\mu = \nu$. But all terms for which $\mu = \nu$ are excluded in the primed sums because they gave zero for the commutators. Next, we realize that $\Gamma_\mu \Gamma_\nu$ is proportional to some $\Gamma_\rho \neq 1$, so our sum reduces to the form

$$\sum_{\mu=2}^{16} Q_\mu \Gamma_\mu = \frac{1}{2}. \tag{15}$$

Renaming $Q_1 = -\frac{1}{2}$ here, we see that we are demanding

$$\sum_{\mu=1}^{16} Q_\mu \Gamma_\mu = 0, \tag{16}$$

which, due to the linear independence of the Γ_μ, would demand that Q_1 vanish. Thus our demand is absurd, and we have proved that we cannot construct a boson from two fermions.

III. N-FERMION PROOF

It is easy to see that method of proof can be extended to any finite number of fermions. We simply form

$$B_i = A_i + A_i^*, \tag{17}$$

$$B_{i+1} = i(A_i - A_i^*) \tag{18}$$

for each fermion operator, and see that

$$B_i B_j + B_j B_i = 2\delta_{ij}, \tag{19}$$

$$B_i^* = B_i. \tag{20}$$

We thereby have the properties of the fermion operators contained in a Clifford algebra of $(2N)^2$ numbers. The basic properties we have used in the proof for $N = 2$ are the same for any N, and the proof follows trivially. Note that we are in general allowing for products of $2N$ fermion operators in our combinations.

IV. APPARENT CONTRADICTION OF THE THEOREM

It is important to realize that, in proving our theorem, we have assumed nothing about the states upon which our operators act. The fact that one is able to construct boson creation operators from fermion operators, as illustrated by Case,[1] is due to *further assumptions concerning the states* utilized in a particular theory. For example, if one supposes

that the boson operators act only upon states of the form of a "Dirac filled sea," in which all negative k states are filled for large k, and all positive k states are empty for large enough k, then one *is* able to construct boson operators.

The well-known analysis[4] of the neutrino theory of light uses the fact that neutrinos may be considered to occupy all negative energy states, and assumes that all positive energy states for high energy are empty. By this assumption, Born and Nath[5] were first able to construct creation and annihilation operators for photons from those for neutrinos.

Thus, our theorem illustrates the important point that the commutation rules for operators may seem to differ depending upon the assumptions concerning the states upon which the operators act. Lieb and Mattis,[6] for example, have found that certain density operators for a one-dimensional electron gas "model" have bosonlike commutation rules, due to the existence of a filled Dirac sea, as first realized by Born and Nath.[5]

Actually, the commutation rules of operators should be independent of any assumptions concerning the states upon which the operators act. Thus, we are faced with an apparent dilemma which must be resolved. To understand the problem, we may consider a trivial example.

Let us suppose we have a single creation operator of the fermion type, with its annihilation operator. The only irreducible representation of the concomitant algebra, as is well known, is of the form of 2×2 matrices,

$$a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad a^* = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{21}$$

which therefore operate upon states of the form

$$\psi = \varphi_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \varphi_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{22}$$

Now it is possible for us to assume that our operators only act upon those states for which

$$a\psi = 0, \tag{23}$$

in which case, our operators, *acting upon such states* obey

$$aa^* - a^*a = 1, \tag{24}$$

as is easily checked. The point is that our choice of states upon which the operators act is a projection of the whole space, and the "altered" commutation rules are true only in the sense that the operators themselves are altered. In the present example, our restriction of states allowed only part of the operators to operate, and the apparent "boson" rules we obtained were meaningless.

In a similar manner, the apparent change of commutation rules for the operators used by Case,[1] or Lieb and Mattis,[6] obtains because of assumptions concerning the Hilbert space wherein the operators perform. To illustrate this point quite clearly, we consider certain operators, first used by Born and and Nath.[5]

Let a_k, a_k^* be a denumerable set of annihilation and creation operators for fermions obeying

$$a_k a_{k'}^* + a_{k'}^* a_k = \delta_{kk'}, \tag{25}$$

$$a_k a_{k'} + a_{k'} a_k = 0, \tag{26}$$

$$a_k^* a_{k'}^* + a_{k'}^* a_k^* = 0, \tag{27}$$

and define

$$f = \sum_{k=-R}^{R} a_k^* a_{k+1}, \tag{28}$$

$$f^* = \sum_{k=-R}^{R} a_{k+1}^* a_k, \tag{29}$$

where R is a large number.

Using the commutation rules, we easily calculate

$$[f, f^*]_- = \sum_{k=-R}^{R} a_k^* a_k - \sum_{k=-R}^{R} a_{k+1}^* a_{k+1}, \tag{30}$$

which reduces to

$$[f, f^*]_- = a_{-R}^* a_{-R} - a_{R+1}^* a_{R+1}. \tag{31}$$

Now, the right-hand side of Eq. (31) has the possible values $0, \pm 1$, depending upon the assumptions concerning the underlying Hilbert space. As examples we may consider three possible subspaces. One subspace contains a finite number of occupied states, in which event we may always take R large enough to obtain 0. Another (unrealistic) subspace has a "filled sea" of positive energy states, with negative energy states empty, and we obtain -1. The last subspace is the usual "filled sea" of negative energy states, which gives $+1$ for our commutator.

Thus we really have no contradiction of the theorem proved in the present analysis. Nonetheless, one can make boson operators from fermion operators, provided one operates only within a projected region of Hilbert space, and that no operations involved remove one from that particular region of Hilbert space.

[4] M. H. L. Pryce, Proc. Roy. Soc. (London) **A165**, 247 (1938).
[5] M. Born and N. S. N. Nath, Proc. Indian Acad. Sci. **A3**, 318 (1936).
[6] E. Lieb and D. Mattis, J. Math. Phys. **6**, 304 (1965).

As long as one considers only a finite number of fermions, one may not construct bosons. If, however, one allows a "filled sea" of fermions, it is possible to obtain bosonlike operators. Our theorem is *not* true in the limit of $N = \infty$.

V. CONCLUSIONS

We have shown that one cannot construct boson creation and annihilation operators from a *finite* number of fermion operators. Incidentally, we have seen that the commutation rules for fermion creation operators are summarized in a Clifford algebra, a result which has apparently not been noticed before Using the isomorphism of the fermion operators with the Clifford algebra, one can deduce the irreducible representations very quickly. For one fermion operator, the algebra is the Pauli algebra, of course, and that fact is commonly used.

We have also seen that the existence of a filled Dirac sea with an infinite number of fermions allows one to construct boson operators. As long as one is careful to stay within the Hilbert subspace containing the filled sea, the commutation rules for the boson operators remain valid. Thus, the boson operators constructed by Born and Nath[5] and recently rediscovered by Lieb and Mattis[6] are not in contradiction with our theorem.

ACKNOWLEDGMENTS

The author is indebted to A. W. Overhauser, A. D. Brailsford, and D. R. Hamann for several enlightening suggestions and conversations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Exact Eigenstates of the Pairing-Force Hamiltonian. II*

R. W. RICHARDSON
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 14 December 1964)

The restrictions on a previously reported class of exact eigenstates of the pairing-force Hamiltonians are removed and it is indicated that all the eigenstates of this Hamiltonian can be included in this class. Explicit expressions are given for the expectation values of one- and two-body operators in the exact, seniority-zero eigenstates of this Hamiltonian. In particular, a simple expression for the occupation probabilities of the levels of the single-particle potential is given. This expression may be easily evaluated for realistic nuclear systems.

I. INTRODUCTION

IN a previous paper,[1] the exact eigenstates of the pairing-force Hamiltonian for finite systems were studied. This study was motivated by the wide use of this Hamiltonian as a model Hamiltonian in nuclear physics.[2] Some of the results of this study were subsequently applied to pairing models of some even isotopes of lead.[3] This application indicated that there is a considerable improvement in the accuracy of the model's description of the excitation spectra of nuclei when exact eigenvalues of the Hamiltonian are used instead of the currently fashionable approximations to these eigenvalues.[2] Similar improvements in the description of other nuclear properties are to be expected from the use of the exact eigenstates of this Hamiltonian. The study of these eigenstates is continued in this paper.

The principal result of I was the demonstration of the existence of a new "restricted class" of eigenstates of the pairing-force Hamiltonian which can be written in a particularly simple form. That is, the wavefunction of an N-pair state in this class was shown to be that of a state of N independent pairs in which each pair interacts through an effective pairing interaction. This result is given below in Eqs. (1.1)–(1.12). The states of this class are restricted by the set of subsidiary requirements that the N single-pair functions which make up an N-pair wavefunction must be distinct. In this paper, we will discuss these restrictions and indicate how they

* This work was supported by the AEC Computing and Applied Mathematics Center, Courant Institute of Mathematical Sciences, New York University, under contract AT(30-1)-1480 with the U. S. Atomic Energy Commission.
[1] R. W. Richardson and N. Sherman, Nucl. Phys. **52**, 221 (1964) (to be referred to as I).
[2] A. M. Lane, *Nuclear Theory* (W. A. Benjamin, Inc., New York, 1964), Part I, and the references cited therein.
[3] R. W. Richardson and N. Sherman, Nucl. Phys. **52**, 253 (1964).

As long as one considers only a finite number of fermions, one may not construct bosons. If, however, one allows a "filled sea" of fermions, it is possible to obtain bosonlike operators. Our theorem is *not* true in the limit of $N = \infty$.

V. CONCLUSIONS

We have shown that one cannot construct boson creation and annihilation operators from a *finite* number of fermion operators. Incidentally, we have seen that the commutation rules for fermion creation operators are summarized in a Clifford algebra, a result which has apparently not been noticed before Using the isomorphism of the fermion operators with the Clifford algebra, one can deduce the irreducible representations very quickly. For one fermion operator, the algebra is the Pauli algebra, of course, and that fact is commonly used.

We have also seen that the existence of a filled Dirac sea with an infinite number of fermions allows one to construct boson operators. As long as one is careful to stay within the Hilbert subspace containing the filled sea, the commutation rules for the boson operators remain valid. Thus, the boson operators constructed by Born and Nath[5] and recently rediscovered by Lieb and Mattis[6] are not in contradiction with our theorem.

ACKNOWLEDGMENTS

The author is indebted to A. W. Overhauser, A. D. Brailsford, and D. R. Hamann for several enlightening suggestions and conversations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Exact Eigenstates of the Pairing-Force Hamiltonian. II*

R. W. RICHARDSON
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 14 December 1964)

The restrictions on a previously reported class of exact eigenstates of the pairing-force Hamiltonians are removed and it is indicated that all the eigenstates of this Hamiltonian can be included in this class. Explicit expressions are given for the expectation values of one- and two-body operators in the exact, seniority-zero eigenstates of this Hamiltonian. In particular, a simple expression for the occupation probabilities of the levels of the single-particle potential is given. This expression may be easily evaluated for realistic nuclear systems.

I. INTRODUCTION

IN a previous paper,[1] the exact eigenstates of the pairing-force Hamiltonian for finite systems were studied. This study was motivated by the wide use of this Hamiltonian as a model Hamiltonian in nuclear physics.[2] Some of the results of this study were subsequently applied to pairing models of some even isotopes of lead.[3] This application indicated that there is a considerable improvement in the accuracy of the model's description of the excitation spectra of nuclei when exact eigenvalues of the Hamiltonian are used instead of the currently fashionable approximations to these eigenvalues.[2] Similar improvements in the description of other nuclear properties are to be expected from the use of the exact eigenstates of this Hamiltonian. The study of these eigenstates is continued in this paper.

The principal result of I was the demonstration of the existence of a new "restricted class" of eigenstates of the pairing-force Hamiltonian which can be written in a particularly simple form. That is, the wavefunction of an N-pair state in this class was shown to be that of a state of N independent pairs in which each pair interacts through an effective pairing interaction. This result is given below in Eqs. (1.1)–(1.12). The states of this class are restricted by the set of subsidiary requirements that the N single-pair functions which make up an N-pair wavefunction must be distinct. In this paper, we will discuss these restrictions and indicate how they

* This work was supported by the AEC Computing and Applied Mathematics Center, Courant Institute of Mathematical Sciences, New York University, under contract AT(30-1)-1480 with the U. S. Atomic Energy Commission.
[1] R. W. Richardson and N. Sherman, Nucl. Phys. **52**, 221 (1964) (to be referred to as I).
[2] A. M. Lane, *Nuclear Theory* (W. A. Benjamin, Inc., New York, 1964), Part I, and the references cited therein.
[3] R. W. Richardson and N. Sherman, Nucl. Phys. **52**, 253 (1964).

may be removed. The members of this class are therefore shown to be completely general. We will also consider the calculation of the expectation values of one- and two-body operators in these eigenstates.

In studying the restrictions on the eigenstates given by Eq. (1.12) below, we find that they can only be violated for a finite set of values of the interaction strength. They are satisfied for all other values. Thus, the results of I are valid everywhere except on this discrete set of values of the interaction strength. This result is derived in Sec. V by showing that if the restrictions are violated, then the interaction strength must be a root of an algebraic equation. If the interaction strength is equal to one of these excluded values, then the wavefunction and energy of the state may be obtained by treating them as continuous functions of the interaction strength. Therefore, this continuity of the wavefunction and energy is sufficient to remove the restrictions and show that the eigenstates have the simple product form given below for all values of the interaction strength.

The remainder of this paper is devoted to the calculation of the expectation values of one- and two-body operators in the seniority-zero eigenstates of the pairing-force Hamiltonian. The normalization of the states is treated in Sec. II and we obtain a simple determinental expression for their normalization coefficients. A simple expression for the occupation probability of the levels of the single-particle potential is derived in Sec. III. The evaluation of this expression is shown to require the solution of an $N \times N$ system of linear algebraic equations whose coefficients depend upon the pair energies E_{p_i}. In Sec. IV, we evaluate the expectation value of $b_f^\dagger b_{f'}$ [see Eq. (1.3) for notation]. While the expression that we derive for this expectation value is similar to that given for the occupation probability, it is not as readily evaluated. However, it does suggest many approximations that may be used in particular calculations. We do not consider off-diagonal matrix elements or matrix elements which involve states of higher seniority. These matrix elements will be the subject of subsequent work.

The practical aspects of our results should be emphasized. For a realistic nuclear system, Eqs. (1.8) may be easily solved for the pair energies. Once the pair energies have been obtained, the occupation probabilities of the single-particle levels may be easily calculated, using the results of Sec. III, and other expectation values may be calculated using the results of Sec. IV.

For reference, we list here the results of I that are used in the subsequent discussion. The pairing-force Hamiltonian is written as

$$H = \sum_f 2\epsilon_f \hat{N}_f - g \sum_{ff'} b_f^\dagger b_{f'}, \tag{1.1}$$

where the quantum numbers $(f\sigma)$, $\sigma = \pm$ denoting states that are conjugate with respect to time reversal, label the states of an external potential and ϵ_f is the energy of the two states $(f\pm)$. The sums in (1.1) are all over a finite set of values of f and, in what follows, all such sums will be understood to be over this same set of values. The operators $\hat{N}_f$ and b_f are given by

$$\hat{N}_f = \tfrac{1}{2}(a_{f-}^\dagger a_{f-} + a_{f+}^\dagger a_{f+}) \tag{1.2}$$

and

$$b_f = a_{f-} a_{f+}, \tag{1.3}$$

where $a_{f\sigma}^\dagger$ and $a_{f\sigma}$ are fermion creation and annihilation operators. The seniority-zero eigenstates of (1.1), which have N pairs of particles, can be written as

$$|\psi\rangle = (N!)^{-\frac{1}{2}} \sum_{f_1 \cdots f_N} \psi(f_1 \cdots f_N) b_{f_1}^\dagger \cdots b_{f_n}^\dagger |0\rangle, \tag{1.4}$$

where $|0\rangle$ is the vacuum state. It is shown in I that the wavefunction $\psi(f_1 \cdots f_N)$ for a restricted class of eigenstates may be written as

$$\psi(f_1 \cdots f_N) = \theta(f_1 \cdots f_N)\phi(f_1 \cdots f_N), \tag{1.5}$$

where

$$\theta(f_1 \cdots f_N) = \prod_{i<j} (1 - \delta_{f_i f_j}) \tag{1.6}$$

and

$$\phi(f_1 \cdots f_N) = \sum_P P\left[\prod_{i=1}^N (2\epsilon_{f_i} - E_{p_i})^{-1}\right]. \tag{1.7}$$

In (1.7), $\sum_P P$ is the sum of the $N!$ permutations P of the indices $p_1 \cdots p_N$ which label the pair-energies E_{p_i}. These pair-energies are roots of the coupled system of equations

$$F(E_{p_i}) = 1/g_i, \qquad i = 1 \cdots N, \tag{1.8}$$

where

$$F(E) = \sum_f \frac{1}{(2\epsilon_f - E)} \tag{1.9}$$

and

$$\frac{1}{g_i} = \frac{1}{g} + 2 \sum_{j=1}^{N}{}' \frac{1}{(E_{p_j} - E_{p_i})}. \tag{1.10}$$

The prime on the sum in (1.10) indicates that it is over those values of j that are not equal to i.

In what follows, a prime on such a sum will always be taken to mean this. The energy of this state is given by

$$E = \sum_{i=1}^{N} E_{p_i}.$$

The indices $p_1 \cdots p_N$ may be defined by

$$\lim_{g \to 0} E_{p_i} = 2\epsilon_{p_i}, \qquad i = 1 \cdots N \tag{1.11}$$

and are to be thought of as labeling the state under consideration. The restrictions on these eigenstates are contained in the additional requirements that the E_{p_i} must be distinct, i.e.,

$$E_{p_i} \neq E_{p_j}, \quad \text{for} \quad i \neq j. \tag{1.12}$$

We defer the discussion and removal of the restrictions (1.12) on the solutions of (1.8) until Sec. V while we consider the normalization of the states in Sec. II and the expectation values of $\hat{N}_f$ and $b_f^\dagger b_{f'}$ in Secs. III and IV. These expectation values are necessary to calculate the expectation value of any one- or two-body operator.

II. NORMALIZATION OF THE WAVEFUNCTIONS

As our first example of the techniques that may be used to calculate with the wavefunctions (1.5), we calculate their normalization coefficients. Our result will be a simple expression in terms of the pair energies that may be easily calculated in realistic situations.

In order to simplify the notation, we will write ϵ_i for ϵ_{f_i} and E_i for E_{p_i}. There will be no confusion over whether i stands for f_i or p_i since this will be clear from the context. In this notation, we have the wavefunction ϕ (a function of the N variables $1 \cdots N$, i.e., $f_1 \cdots f_N$ which is implicitly labeled by the N indices $1 \cdots N$, i.e., $p_1 \cdots p_N$) given by

$$\phi(1 \cdots N) = (N!)^{-\frac{1}{2}} C S_N \prod_{i=1}^{N} (2\epsilon_i - E_i)^{-1}, \tag{2.1}$$

where we have introduced the normalization coefficient $(N!)^{-\frac{1}{2}}C$ and have denoted by S_N the symmetrizer on the N indices of the pair energies, i.e., S_N is the sum of the $N!$ permutations of the N indices which label the E_i.

The normalization coefficient is to be calculated from the normalization condition[1]

$$\langle \psi \mid \psi \rangle = \sum_{1 \cdots N}{}' \theta(1 \cdots N)\phi^2(1 \cdots N) = 1. \tag{2.2}$$

We will show that this implies that C^{-2} is an $N \times N$ determinant. The diagonal element of this determinant in the (i, i) position will be shown to be

$$c_i - 2 \sum_{j=1}^{N}{}' E_{ji}^2,$$

where

$$c_i = \sum_f (2\epsilon_f - E_i)^{-2} \tag{2.3}$$

and

$$E_{ij} = (E_i - E_j)^{-1}. \tag{2.4}$$

The off-diagonal element in the (i, j) position will be shown to be $2E_{ij}^2$.

Our derivation of the above expression for the normalization is in two parts. In the first part, we derive an operator-product form for C^{-2} and, in the second part, we derive a recursion relation between the operator product for the N-pair system and the operator product for an $(N-1)$-pair system. The above results are then proven by starting with the easily calculated normalization of a two-pair system and then using the recursion relation to work up to the desired value of N. Throughout this proof, we will assume that the pair energies are distinct. In Sec. V, we will show that this assumption is satisfied for all values of the interaction strength except a finite set of values. The special cases when the interaction equals one of the values for which our assumption does not hold may be treated by using an appropriate limiting procedure similar to the one used in that section. We now turn to the derivation.

An operator-product expression for C^{-2} may be derived from (2.2) by making use of the recursion relation

$$\theta(1 \cdots N) = \theta(1 \cdots N-1)\left[1 - \sum_{i=1}^{N-1} \delta_{iN}\right], \tag{2.5}$$

between the θ-functions of N and $N-1$ variables, to perform the sum. To make use of this relation, we write C^{-2}, from (2.1) and (2.2), as

$$\begin{aligned} C^{-2} &= (N!)^{-1} \sum_{1 \cdots N}{}' \theta(1 \cdots N)\left[S_N \prod_{i=1}^{N} (2\epsilon_i - E_i)^{-1}\right]^2 \\ &= (N!)^{-1} S_N S_N' \sum_{1 \cdots N}{}' \theta(1 \cdots N) \\ &\qquad \times \prod_{i=1}^{N} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1} \\ &= S_N' \sum_{1 \cdots N}{}' \theta(1 \cdots N) \\ &\qquad \times \prod_{i=1}^{N} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1}, \end{aligned} \tag{2.6}$$

where S_N' symmetrizes the primed indices i' and

sets $i' = i$. Throughout this section, we will assume that the prime on i' is only a mark to distinguish it from i for the purposes of symmetrization. Thus, the factor in the product with $i = 1$ is $(2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1}$. We rewrite (2.6) as

$$C^{-2} = I_N(1 \cdots N), \tag{2.7}$$

where $I_N(1 \cdots N)$ is given by (2.6) and its arguments are the indices on the E_i's. These arguments will be suppressed when it is not important to spell out in detail what they are. We now substitute (2.5) into (2.6) and perform the sum on the variable N

$$\begin{aligned} I_N &= S'_N \sum_{1\cdots N-1} \theta(1 \cdots N-1) \\ &\times \left\{\left[\prod_{i=1}^{N-1} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1}\right]\right. \\ &\times \sum_N (2\epsilon_N - E_N)^{-1}(2\epsilon_N - E_{N'})^{-1} \\ &- \sum_{j=1}^{N-1}\left[\prod_{\substack{i=1\\(i\neq j)}}^{N-1} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1}\right](2\epsilon_j - E_j)^{-1} \\ &\left.\times (2\epsilon_j - E_{j'})^{-1}(2\epsilon_j - E_N)^{-1}(2\epsilon_j - E_{N'})^{-1}\right\}. \end{aligned} \tag{2.8}$$

We next define the "overlap integrals" c_{ij} by

$$c_{ij} = \sum_f (2\epsilon_f - E_i)^{-1}(2\epsilon_f - E_j)^{-1}. \tag{2.9}$$

Note that, by (2.3), $c_{ii} = c_i$. Thus, the last factor in the first term of (2.8) is $c_{NN'}$. Since we have assumed that the pair energies are distinct, we may use a partial fraction expansion to rewrite the last factors in the second term of (2.8) as

$$\begin{aligned} &(2\epsilon_j - E_j)^{-1}(2\epsilon_j - E_{j'})^{-1}(2\epsilon_j - E_N)^{-1}(2\epsilon_j - E_{N'})^{-1} \\ &= E_{jN}E_{j'N'}[(2\epsilon_j - E_j)^{-1} - (2\epsilon_j - E_N)^{-1}] \\ &\quad\times [(2\epsilon_j - E_{j'})^{-1} - (2\epsilon_j - E_{N'})^{-1}] \\ &= \hat{E}_{jN}\hat{E}_{j'N'}(2\epsilon_j - E_j)^{-1}(2\epsilon_j - E_{j'})^{-1}, \end{aligned} \tag{2.10}$$

where we have defined the operators $\hat{E}_{ij}$ by

$$\hat{E}_{ij} = (1 + P_{ij})E_{ij}, \tag{2.11}$$

where P_{ij} is the transposition $i \leftrightarrow j$, operating on the indices of the E_i's. Substituting (2.9) and (2.10) into (2.8), we have

$$\begin{aligned} I_N = S'_N&\left[c_{NN'} - \sum_{j=1}^{N-1} \hat{E}_{jN}\hat{E}_{j'N'}\right] \\ &\times \sum_{1\cdots N-1} \theta(1 \cdots N-1) \\ &\times \prod_{i=1}^{N-1} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1} \end{aligned} \tag{2.12}$$

apart from the reduction of N by one, the sum in (2.12) is the same as the sum in (2.6). We may therefore perform the sums on the variables $1 \cdots N$ by repeated application of (2.12) with the result that

$$\begin{aligned} I_N = S'_N&\left[c_{NN'} - \sum_{j=1}^{N-1} \hat{E}_{jN}\hat{E}_{j'N'}\right] \cdots \\ &\times [c_{22'} - \hat{E}_{12}\hat{E}_{1'2'}]c_{11'}. \end{aligned} \tag{2.13}$$

We may write (2.13) in a little more compact notation if we introduce the operators

$$A_L = c_{LL'} - \sum_{j=1}^{L-1} \hat{E}_{jL}\hat{E}_{j'L'}. \tag{2.14}$$

Note that A_L operates on the indices $1 \cdots L$ and $1' \cdots L'$. In terms of the operators A_L, Eq. (2.13) becomes

$$I_N = S'_N \prod_{k=1}^{N} A_k, \tag{2.15}$$

where the order of the operators A_k in the product is such that k decreases from left to right. In what follows, all such operator products will be assumed to be written in this order. Equations (2.7) and (2.15) yield the desired operator expression for the normalization coefficient.

It should be noted at this point that c_{ij}, for $i \neq j$, can be given explicitly in terms of the E_i's. For, if we first perform a partial fraction expansion of the summand in (2.9) and then sum on f using (1.8), we obtain

$$\begin{aligned} c_{ij} &= E_{ij}(g_i^{-1} - g_j^{-1}) \\ &= -4E_{ij}^2 + 2\sum_{k=1}^{N}{}'' E_{ki}E_{kj}, \qquad i \neq j, \end{aligned} \tag{2.16}$$

where the double prime on the sum indicates that it is over those values of k not equal to i or j.

It is instructive to consider a simple example of Eq. (2.15). We will also need the results of this example to initiate the recursion relation that we are going to derive later in this section. We therefore consider Eq. (2.15) for $N = 2$, in which case we have

$$\begin{aligned} I_2 &= S'_2(c_{22'} - \hat{E}_{12}\hat{E}_{1'2'})c_{11'} \\ &= S'_2[c_{22'}c_{11'} - 2E_{12}E_{1'2'}(c_{11'} + c_{22'})], \end{aligned} \tag{2.17}$$

where we have used

$$\begin{aligned} &S'_2\hat{E}_{12}\hat{E}_{1'2'} \\ &= S'_2(1 + P_{12})(1 + P_{1'2'})E_{12}E_{1'2'} \\ &= S'_2(1 + P_{1'2'})(1 + P_{12}P_{1'2'})E_{12}E_{1'2'} \\ &= 2S'_2E_{12}E_{1'2'}(1 + P_{12}P_{1'2'}) \end{aligned} \tag{2.18}$$

in the second term. Equation (2.17) may be rewritten as a determinant, i.e.,

$$I_2 = S_2' \begin{vmatrix} c_{11'} - 2E_{12}E_{1'2'} & 2E_{12}E_{1'2'} \\ 2E_{12}E_{1'2'} & c_{22'} - 2E_{12}E_{1'2'} \end{vmatrix}. \quad (2.19)$$

When we substitute $c_{12} = -4E_{12}^2$, which we obtain from (2.16), into (2.19) we find that it is only the identity permutation in S_2' that makes a nonzero contribution to I_2. Thus, we have

$$I_2 = \begin{vmatrix} c_1 - 2E_{12}^2 & 2E_{12}^2 \\ 2E_{12}^2 & c_2 - 2E_{12}^2 \end{vmatrix}. \quad (2.20)$$

In the general case, we will see that I_N can be written as the natural generalization of (2.19) and that when it is written in this form only the identity permutation in S_N' contributes to I_N. Therefore, I_N will be given by the natural generalization of (2.20). This, of course, does not mean that it is only the identity permutation in S_N' that contributes to (2.6). For, we clearly needed all the permutations in S_2' in order to write (2.18). However, it does mean that we can collect terms and arrange them in such a fashion that we can neglect all the permutations in S_N' except the identity.

We will now derive a recursion relation that will prove that the above determinental form of I_N is true in general. The derivation is based upon the following observation: if we hypothesize that I_N has a determinental form similar to (2.19), i.e., that

$$I_N = S_N' \begin{vmatrix} c_{11'} - 2\sum_i{}' E_{i1}\bar{E}_{i'1'} & 2E_{12}E_{1'2'} \cdots & 2E_{1N}E_{1'N'} \\ 2E_{21}E_{2'1'} & \cdots & \\ \cdots & & \\ 2E_{N1}E_{N'1'} & \cdots & c_{NN'} - 2\sum_i{}' E_{iN}E_{i'N'} \end{vmatrix}, \quad (2.21)$$

and consider the coefficient of $c_{NN'}$ in this expression, then we find that it has the same form as that hypothesized for $I_{N-1}(1 \cdots N-1)$ in which, however, $c_{ii'}$ has been replaced by $c_{ii'} - 2E_{iN}E_{i'N'}$. It is to be emphasized that we speak here only of the explicit functional dependence of the I_N's on the E_{ij}'s and c_i's. Therefore, the $I_{N-1}(1 \cdots N-1)$ mentioned above will not be related to the normalization coefficient of an $(N-1)$-pair wavefunction by (2.7) because the pair energies $E_1 \cdots E_{N-1}$ that appear in it are solutions of the N-pair equations. We now show that the coefficient of $c_{NN'}$ in I_N is given by I_{N-1}.

Let us denote by $[X]_{ii'}$ the coefficient of $c_{ii'}$ in the quantity X. Following the outline given above, we seek the coefficient of $c_{NN'}$ in I_N, i.e., $[I_N]_{NN'}$. In order to obtain this, we first write

$$S_N' = S_{N-1}'(1' \cdots N-1') \sum_{i=1}^{N} P_{i'N'},$$

where $S_{N-1}'(1' \cdots N-1')$ is the symmetrizer operating on the indices $1' \cdots N-1'$. We next write

$$S_N' A_N = S_N' \left[c_{NN'} - 2\sum_{i=1}^{N-1} \hat{E}_{iN}E_{i'N'} \right],$$

where we have used the rearrangement theorem to write

$$S_N' \hat{E}_{i'j'} = 2S_N' E_{i'j'}.$$

We then have, from (2.15),

$$[I_N]_{NN'} = S_{N-1}'(1' \cdots N-1') \times \left[\sum_{i=1}^{N} P_{i'N'} \left(c_{NN'} - 2\sum_{i=1}^{N-1} \hat{E}_{iN}E_{i'N'} \right) \prod_{k=1}^{N-1} A_k \right]_{NN'}$$

$$= S_{N-1}'(1' \cdots N-1') \left\{ \prod_{k=1}^{N-1} A_k - 2 \times \sum_{i,j=1}^{N-1} P_{iN}P_{i'N'}E_{iN}E_{i'N'} \left[\prod_{k=1}^{N-1} A_k \right]_{ij'} \right\}, \quad (2.22)$$

where we have neglected the identity operators in $\sum_i P_{i'N'}$ and $\hat{E}_{iN}$ in the second term because these operators act on $\prod_{k=1}^{N-1} A_k$, which does not contain $c_{NN'}$, and therefore they can not contribute to $[I_N]_{NN'}$. The evaluation of (2.22) proceeds with the evaluation of $[\prod_{k=1}^{N-1} A_k]_{ij'}$ which we do in two parts. We will first evaluate the "diagonal elements" with $i = j$ and then turn to the "nondiagonal elements" with $i \neq j$.

For the diagonal elements

$$\left[\prod_{k=1}^{N-1} A_k \right]_{ii'},$$

we may write

$$\left[\prod_{k=1}^{N-1} A_k \right]_{ii'} = \prod_{k=i+1}^{N-1} [A_k(i) - E_{ik}E_{i'k'}] \left[\prod_{k=1}^{i} A_k \right]_{ii'}, \quad (2.23)$$

where we have defined

$$A_L(i, j, \cdots) = c_{LL'} - \sum_{\substack{m=1 \\ (m \neq i,j,\cdots)}}^{L-1} \hat{E}_{mL}\hat{E}_{m'L'}. \quad (2.24)$$

To this definition of $A_L(i, j, \cdots)$, we append the convention that any one of the arguments $i, j, \cdots$ that is greater than $L-1$ is to be ignored. With this definition, $A_k(i) - E_{ik}E_{i'k'}$ which appears in (2.23) is the same as A_k with $\hat{E}_{ik}\hat{E}_{i'k'}$ replaced by $E_{ik}E_{i'k'}$. We now prove by induction that

$$\left[\prod_{k=1}^{i} A_k\right]_{ii'} = \prod_{k=1}^{i-1}(A_k - E_{ik}E_{i'k'}). \tag{2.25}$$

For $i = 1$, we have $[A_1]_{11'} = 1$ and, for $i = 2$, we have

$$\begin{aligned}[A_2A_1]_{22'} &= [(c_{22'} - \hat{E}_{12}\hat{E}_{1'2'})A_1]_{22'} \\ &= A_1 - P_{12}P_{1'2'}E_{12}E_{1'2'}[A_1]_{11'} \\ &= A_1 - E_{21}E_{2'1'}.\end{aligned}$$

In the general case, we have

$$\begin{aligned}&\left[\prod_{k=1}^{i} A_k\right]_{ii'} \\ &= \left[\left(c_{ii'} - \sum_{j=1}^{i-1}\hat{E}_{ji}\hat{E}_{j'i'}\right)\prod_{k=1}^{i-1} A_k\right]_{ii'} \\ &= \prod_{k=1}^{i-1} A_k - \sum_{j=1}^{i-1} E_{ij}E_{i'j'}P_{ij}P_{i'j'}\left[\prod_{k=1}^{i-1} A_k\right]_{jj'}.\end{aligned} \tag{2.26}$$

Using (2.23) and the induction hypothesis (2.25), we may write

$$\left[\prod_{k=1}^{i-1} A_k\right]_{jj'} = \prod_{\substack{k=1\\(k\neq j)}}^{i-1}[A_k(j) - E_{jk}E_{j'k'}] \tag{2.27}$$

in the second term of (2.26). Substituting (2.27) into (2.26) and carrying out the indicated transpositions yields

$$\begin{aligned}\left[\prod_{k=1}^{i} A_k\right]_{ii'} &= \prod_{k=1}^{i-1} A_k - \sum_{j=1}^{i-1} E_{ij}E_{i'j'} \\ &\quad\times \prod_{\substack{k=1\\(k\neq j)}}^{i-1}[A_k(j) - E_{ik}E_{i'k'}].\end{aligned} \tag{2.28}$$

We now perform the sum on j in (2.28) by a method that we will use many times in what follows. The sum is done by first noting that the operators in the second term of (2.28) do not operate on the indices ij or $i'j'$. We may therefore commute $E_{ij}E_{i'j'}$ past the first $(i-1-j)$ factors in the product and place it in the position that A_j would occupy if it were present. We next observe that

$$\begin{aligned}&\hat{E}_{ik}\hat{E}_{j'k'}E_{ij}E_{i'j'} \\ &\quad= E_{jk}(E_{ij} - E_{ik})E_{j'k'}(E_{i'j'} - E_{i'k'}) \\ &\quad= E_{ij}E_{ik}E_{i'j'}E_{i'k'}.\end{aligned} \tag{2.29}$$

Thus, if we have commuted $E_{ij}E_{i'j'}$ past the first $(i-1-j)$ factors, then we may replace $E_{ik}E_{i'k'}$ in these factors by $\hat{E}_{ik}\hat{E}_{j'k'}$. Or, equivalently, we may replace $A_k(j) - E_{ik}E_{i'k'}$ by A_k. Equation (2.28) then becomes

$$\begin{aligned}\left[\prod_{k=1}^{i} A_k\right]_{ii'} &= \prod_{k=1}^{i-1} A_k - \sum_{j=1}^{i-1}\left(\prod_{k=j+1}^{i-1} A_k\right)E_{ij}E_{i'j'} \\ &\quad\times \prod_{l=1}^{j-1}(A_l - E_{il}E_{i'l'}).\end{aligned} \tag{2.30}$$

The sum on j may now be performed. For consider the term in (2.30) with $j = 1$. It is

$$-\left(\prod_{k=2}^{i-1} A_k\right)E_{i1}E_{i'1'}$$

and it may be combined with the first term of (2.30) to give

$$\left(\prod_{k=2}^{i-1} A_k\right)(A_1 - E_{i1}E_{i'1'}). \tag{2.31}$$

The term with $j = 2$ in (2.30) is

$$-\left(\prod_{k=3}^{i-1} A_k\right)E_{i2}E_{i'2'}(A_1 - E_{i1}E_{i'1'})$$

and it may be combined with (2.31) to give

$$\left(\prod_{k=3}^{i-1} A_k\right)(A_2 - E_{i2}E_{i'2'})(A_1 - E_{i1}E_{i'1'}).$$

In this way, the sum on j in (2.28) may be done and the result is (2.25) which completes the induction. Substitution of (2.25) into (2.23) yields the desired expression

$$\left[\prod_{k=1}^{N-1} A_k\right]_{ii'} = \prod_{\substack{k=1\\(k\neq i)}}^{N-1}[A_k(i) - E_{ik}E_{i'k'}] \tag{2.32}$$

for the diagonal elements.

For the nondiagonal elements, we have

$$\begin{aligned}\left[\prod_{k=1}^{N-1} A_k\right]_{ii'} &= \prod_{l=i+1}^{N-1}[A_l(ij) - \hat{E}_{jl}E_{i'l'} \\ &\quad - E_{il}\hat{E}_{i'l'}]\left[\prod_{k=1}^{i} A_k\right]_{ii'},\end{aligned} \tag{2.33}$$

for $i > j$, and

$$\begin{aligned}\left[\prod_{k=1}^{N-1} A_k\right]_{ii'} &= \prod_{l=j+1}^{N-1}[A_l(ij) - E_{il}\hat{E}_{i'l'} \\ &\quad - \hat{E}_{jl}E_{i'l'}]\left[\prod_{k=1}^{i} A_k\right]_{ii'},\end{aligned} \tag{2.34}$$

for $i < j$. We prove by induction that

$$\begin{aligned}\left[\prod_{k=1}^{i} A_k\right]_{ij'} &= -E_{ij}E_{i'i'}\sum_{\substack{k=1\\(k\neq j)}}^{i-1}[A_k(j) \\ &\quad - 2E_{ik}E_{j'k'}], \quad \text{for} \quad i > j, \\ &= \left[\prod_{k=1}^{i} A_k\right]_{ij'}, \quad \text{for} \quad i < j.\end{aligned} \tag{2.35}$$

For example, when $ij = 12$ or 21, we have

$$[A_2A_1]_{21'} = [(c_{22'} - \hat{E}_{12}\hat{E}_{1'2'})A_1]_{21'}$$
$$= -P_{12}E_{12}E_{1'2'}[A_1]_{11'} = -E_{21}E_{1'2'}$$

and in the same way $[A_2A_1]_{12'} = -E_{12}E_{2'1'}$. For the general case with $i > j$, we write

$$\left[\prod_{k=1}^{i} A_k\right]_{ij'} = \left[\left(c_{ii'} - \sum_{l=1}^{i-1} \hat{E}_{li}\hat{E}_{l'i'}\right)\prod_{k=1}^{i-1} A_k\right]_{ij'}$$
$$= -P_{ij}E_{ji}E_{j'i'}\left[\prod_{k=1}^{i-1} A_k\right]_{ij'}$$
$$- \sum_{\substack{l=1\\(l\neq j)}}^{i-1} P_{li}E_{li}\hat{E}_{l'i'}\left[\prod_{k=1}^{i-1} A_k\right]_{lj'}. \quad (2.36)$$

(The case $i < j$ should be considered in parallel with the case $i > j$. However, we leave this to the reader since the steps in both cases are essentially the same.) We use (2.32) for the first term of (2.36) and (2.32), (2.34), and the induction hypothesis (2.35) for the second term. We then carry out the indicated transpositions to obtain

$$\left[\prod_{k=1}^{i} A_k\right]_{ij'} = -E_{ij}E_{j'i'}\prod_{\substack{k=1\\(k\neq j)}}^{i} [A_k(j) - E_{ik}E_{j'k'}]$$
$$+ \sum_{l=1}^{i-1} E_{il}\hat{E}_{l'i'}\prod_{m=j+1}^{i-1} [A_m(lj) - E_{im}\hat{E}_{l'k'}$$
$$- \hat{E}_{im}E_{j'm'}]E_{ij}E_{j'l'}\prod_{k=1}^{j-1} [A_k(l) - 2E_{ik}E_{j'k'}]$$
$$+ \sum_{l=j+1}^{i-1} E_{il}\hat{E}_{l'i'}\prod_{m=l+1}^{i-1} [A_m(lj) - E_{im}\hat{E}_{l'm'}$$
$$- \hat{E}_{jm}E_{j'm'}]E_{ij}E_{j'l'}\prod_{k=1}^{l-1} [A_k(j) - 2E_{ik}E_{j'k'}]. \quad (2.37)$$

We note that the operators $\hat{E}_{l'm'}$ and $\hat{E}_{jm}$ in this expression operate on $E_{j'l'}$ and E_{ij}, respectively, and that

$$\hat{E}_{l'm'}E_{j'l'} = E_{l'm'}(E_{j'l'} - E_{j'm'}) = E_{j'l'}E_{j'm'}$$

and

$$\hat{E}_{jm}E_{ij} = E_{jm}(E_{ij} - E_{im}) = E_{ij}E_{im}.$$

We may therefore replace the operators $\hat{E}_{l'm'}$ and $\hat{E}_{jm}$ in (2.37) by $E_{j'm'}$ and E_{im}. With this replacement made, we note that the operators $\hat{E}_{l'i'}$ operate on $E_{j'l'}$ and may be replaced $E_{j'i'}$. Equation (2.37) can therefore be written as

$$\left[\prod_{k=1}^{i} A_k\right]_{ij'} = -E_{ij}E_{j'i'}\left\{\prod_{k=1}^{i-1} [A_k(j) - E_{ik}E_{j'k'}]\right.$$
$$\left. - \sum_{\substack{l=1\\(l\neq j)}}^{i-1} E_{il}E_{j'l'}\prod_{\substack{k=1\\(k\neq l,j)}}^{i-1} [A_k(lj) - 2E_{ik}E_{j'k'}]\right\}. \quad (2.38)$$

The sum on l in (2.38) may be performed by the method that was used to sum (2.28) and the result is (2.35) which completes the induction. Equations (2.33), (2.34), and (2.35) imply that

$$\left[\prod_{k=1}^{N-1} A_k\right]_{ij'}$$
$$= -E_{ij}E_{j'i'}\prod_{\substack{k=1\\(k\neq j)}}^{N-1} [A_k(ij) - 2E_{ik}E_{j'k'}] \quad (2.39)$$

which is the desired expression for the nondiagonal elements.

We may now use (2.32) and (2.39) to evaluate (2.22) which we write as

$$[I_N]_{NN'} = S'_{N-1}(1' \cdots N-1')$$
$$\times \left\{\prod_{k=1}^{N-1} A_k - 2\sum_{i=1}^{N-1}\left(P_{iN}P_{i'N'}E_{iN}E_{i'N'}\left[\prod_{k=1}^{N-1} A_k\right]_{ii'}\right.\right.$$
$$\left.\left. + \sum_{\substack{l=1\\(j\neq i)}}^{N-1} P_{iN}P_{j'N'}E_{iN}E_{i'N'}\left[\prod_{k=1}^{N-1} A_k\right]_{ij'}\right)\right\}. \quad (2.40)$$

Substituting (2.32) and (2.39) into a typical term in the sum on i in (2.40) and carrying out the indicated transpositions yields

$$2P_{iN}P_{i'N'}E_{iN}E_{i'N'}\left[\prod_{k=1}^{N-1} A_k\right]_{ii'}$$
$$+ 2\sum_{\substack{j=1\\(j\neq i)}}^{N-1} P_{iN}P_{j'N'}E_{iN}E_{i'N'}\left[\prod_{k=1}^{N-1} A_k\right]_{ij'}$$
$$= 2E_{Ni}E_{N'i'}\prod_{\substack{k=1\\(k\neq i)}}^{N-1} [A_k(i) - E_{Nk}E_{N'k'}]$$
$$- 2\sum_{\substack{j=1\\(j\neq i)}}^{N-1} E_{Ni}E_{i'j'}E_{Nj}E_{N'i'}$$
$$\times \prod_{\substack{k=1\\(k\neq i,j)}}^{N-1} [A_k(ij) - 2E_{Nk}E_{N'k'}]. \quad (2.41)$$

Since this is to be symmetrized by $S'_{N-1}(1' \cdots N-1')$, and i' and j' do not occur in the product in the second term, we may replace $2E_{i'j'}E_{N'i'}$ by $(1 + P_{i'j'})E_{i'j'}E_{N'i'} = E_{N'i'}E_{N'j'}$. Thus, (2.41) may be written as

$$E_{Ni}E_{N'i'}\prod_{\substack{k=1\\(k\neq i)}}^{N-1} [A_k(i) - E_{Nk}E_{N'k'}]$$
$$+ E_{Ni}E_{N'i'}\left\{\prod_{\substack{k=1\\(k\neq i)}}^{N-1} [A_k(i) - E_{Nk}E_{N'k'}]\right.$$
$$\left. - \sum_{\substack{j=1\\(j\neq i)}}^{N-1} E_{Nj}E_{N'j'}\prod_{\substack{k=1\\(k\neq i,j)}}^{N-1} [A_k(ij) - 2E_{Nk}E_{N'k'}]\right\}.$$

The terms in the curly brackets can be summed by the methods that were used in Eq. (2.28), giving

$$E_{Ni}E_{N'i'}\prod_{k=1}^{N-1}[A_k(i)-E_{Nk}E_{N'k'}] + E_{Ni}E_{N'i'}\prod_{k=1}^{N-1}[A_k(i)-2E_{Nk}E_{N'k'}]$$

for (2.41). Substituting this into (2.40), we have

$$[I_N]_{NN'} = S'_{N-1}(1'\cdots N-1') \times\Big\{\prod_{k=1}^{N-1}A_k - \sum_{i=1}^{N-1}E_{Ni}E_{N'i'}\prod_{\substack{k=1\\(k\neq i)}}^{N-1}[A_k(i)-E_{Nk}E_{N'k'}] - \sum_{i=1}^{N-1}E_{Ni}E_{N'i'}\prod_{\substack{k=1\\(k\neq i)}}^{N-1}[A_k(i)-2E_{Nk}E_{N'k'}]\Big\}.$$

The two sums on i may be done by using the methods of Eq. (2.28), with the result that

$$[I_N]_{NN'} = S'_{N-1}(1'\cdots N-1') \times\prod_{k=1}^{N-1}[A_k-2E_{Nk}E_{N'k'}], \qquad (2.42)$$

which is the desired expression for the coefficient of $c_{NN'}$ in I_N.

If we compare (2.42) with (2.14) and (2.15), we see that our hypothesis, that the coefficient of $c_{NN'}$ in I_N is just $I_{N-1}(1\cdots N-1)$ in which $c_{ii'}$ has been replaced by $c_{ii'}-2E_{iN}E_{i'N'}$, is proven. Eq. (2.42) may be used recursively, starting with the initial value of I_2 given by (2.19), to prove the general form of I_N given by (2.21) if at each recursion we take into account the symmetry of I_N.

We finally prove that it is only the identity permutation in the S'_N of (2.21) that contributes to I_N. For, consider a permutation of S'_N which interchanges $1'$ and $2'$. After applying this permutation and equating the primed indices with the unprimed ones we get, for the top two rows of the resulting determinant,

$$\begin{vmatrix} c_{12}+2E_{12}^2-2\sum_{j=3}^{N}E_{j1}E_{j2} & -2E_{12}^2 & \cdots & 2E_{1N}E_{2N} \\ -2E_{12}^2 & c_{21}+2E_{12}^2-2\sum_{j=3}^{N}E_{j2}E_{j1} & \cdots & 2E_{2N}E_{1N}. \end{vmatrix}$$

After substituting (2.16) for c_{12} and c_{21}, we see that the elements of these two rows are equal and therefore the corresponding contribution to I_N is zero. The same results are obtained for all the permutations of S'_N that are not equal to the identity. We therefore have the determinental form

$$I_N = \begin{vmatrix} c_1-2\sum' E_{j1}^2 & 2E_{12}^2 & \cdots & 2E_{1N}^2 \\ 2E_{12}^2 & c_2-2\sum' E_{j2}^2 & \cdots & 2E_{2N}^2 \\ \cdots & & & \\ 2E_{N1}^2 & \cdots & & c_N-2\sum' E_{jN}^2 \end{vmatrix} \qquad (2.43)$$

for I_N. Then, by (2.7), the normalization coefficient is given by the square root of the reciprocal of (2.43).

IV. THE OCCUPATION PROBABILITY

In this section, we will derive a simple expression for the occupation probability,[1]

$$N_1 = \langle\psi|\hat{N}_1|\psi\rangle = N\sum_{2\cdots N}\theta(1\cdots N)\phi^2(1\cdots N), \qquad (3.1)$$

for the levels of the single-particle well. We will use the notation of Sec. II and our methods will be very similar to those used in that section. The result that we will derive is that N_1 is C^2 times a sum of N terms. The mth term in this sum is a determinant which can be obtained from I_N, (2.43), by replacing the mth column in (2.43) by $(2\epsilon_1-E_m)^{-2}$. When we combine this with the expression for C^2, derived in Sec. II, we obtain

$$N_1 = \sum_{i=1}^{N}\alpha_i(2\epsilon_1-E_i)^{-2}, \qquad (3.2)$$

where the α_i's satisfy

$$\begin{aligned} &(c_1-2\textstyle\sum' E_{j1}^2)\alpha_1+2E_{12}^2\alpha_2+\cdots+2E_{1N}^2\alpha_N=1,\\ &2E_{21}^2\alpha_1+(c_2-2\textstyle\sum' E_{j2}^2)\alpha_2+\cdots+2E_{2N}^2\alpha_N=1,\\ &\cdots\\ &2E_{N1}^2\alpha_1+\cdots(c_N-2\textstyle\sum' E_{jN}^2)\alpha_N=1. \end{aligned} \qquad (3.3)$$

Our proof is along lines similar to those used in the preceding section for the normalization coefficient.

We first derive an operator-product expression for N_1 that is similar to (2.15). Then, after considering the special case with $N = 2$, we derive a recursion relation for these operator products that can be used to prove the above result.

We derive the operator-product form of N_1 from (3.1) after substitution of (2.1), i.e., from

$$N_1 = N(N!)^{-1}C^2 S_N S'_N \sum_{2\cdots N} \theta(1 \cdots N) \times \prod_{i=1}^{N} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1}. \quad (3.4)$$

Since f_1 is not summed in (3.4), the symmetrizer S_N should be written as

$$S_N = \sum_{l=1}^{N} P_{1l} S_{N-1}(2 \cdots N),$$

where $S_{N-1}(2 \cdots N)$ is the symmetrizer operating on the indices $2 \cdots N$ and may be replaced by $(N-1)!$. Making this replacement and performing the sums on $2 \cdots N$ in the same way that (2.6) was summed, we have

$$N_1 = C^2 \sum_{l=1}^{N} P_{1l} S'_N \prod_{k=2}^{N} A_k (2\epsilon_1 - E_1)^{-1} \times (2\epsilon_1 - E_{1'})^{-1}, \quad (3.5)$$

where the operators A_k have been defined in Eq. (2.14). The remainder of this section is devoted to a discussion of this operator product.

As a simple example of Eq. (3.5), let us consider the case for which $N = 2$. We then have

$$\begin{aligned} N_1 &= C^2(1 + P_{12}) S'_2 (c_{22'} - \hat{E}_{12}\hat{E}_{1'2'}) \\ &\quad \times (2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1} \\ &= C^2(1 + P_{12})(1 + P_{1'2'})(c_{22'} - 4E_{12}E_{1'2'}) \\ &\quad \times (2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1} \\ &= C^2(1 + P_{12})(1 + P_{1'2'}) \begin{vmatrix} 1 & 2E_{12}E_{1'2'} \\ 1 & c_{22'} - 2E_{12}E_{1'2'} \end{vmatrix} \\ &\quad \times (2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1}. \end{aligned} \quad (3.6)$$

Carrying out the indicated transpositions and using $c_{12} = c_{21} = -4E_{12}^2$, we have

$$N_1 = C^2 \left[\begin{vmatrix} 1 & 2E_{12}^2 \\ 1 & c_2 - 2E_{12}^2 \end{vmatrix} (2\epsilon_1 - E_1)^{-2} + \begin{vmatrix} c_1 - 2E_{12}^2 & 1 \\ 2E_{12}^2 & 1 \end{vmatrix} (2\epsilon_1 - E_2)^{-2} \right] \quad (3.7)$$

which is in accord with (3.2) when we substitute (2.20) for C^{-2}. We will show that the natural generalization of (3.7) holds for arbitrary N.

We now turn to the derivation of a recursion relation for operator products such as (3.5) that is similar to the one, (2.42), we derived in our calculation of the normalization coefficient. We start by writing the operator product that appears in N_1 as

$$\sum_{l=1}^{N} P_{1l} S'_N \prod_{k=2}^{N} A_k (2\epsilon_1 - A_1)^{-1}(2\epsilon_1 - E_{1'})^{-1} = \sum_{i,j=1}^{N} I_{ij'}(N)(2\epsilon_1 - E_i)^{-1}(2\epsilon_1 - E_{j'})^{-1}. \quad (3.8)$$

The hypothesis we wish to prove is that $I_{ij'}(N)$ is given by a determinant that is a natural generalization of (3.6), e.g.,

$$I_{11'}(N) = S'_{N-1}(2' \cdots N') \begin{vmatrix} 1 & 2E_{12}E_{1'2'} & \cdots & 2E_{1N}E_{1'N'} \\ 1 & c_{22'} - 2\sum{}' E_{i2}E_{i'2'} & \cdots & 2E_{2N}E_{2'N'} \\ \cdots & & & \\ 1 & 2E_{N2}E_{N'2'} & \cdots & c_{NN'} - 2\sum{}' E_{iN}E_{i'N'} \end{vmatrix}, \quad (3.9)$$

The other $I_{ij'}(N)$ can be obtained from $I_{11'}(N)$ by symmetry considerations. Our recursion relation is based upon the observation that if we expand the determinant in (3.9) by the minors of the first row then the first term is $I_{N-1}(2 \cdots N)$ in which $c_{ii'}$ has been replaced by $c_{ii'} - 2E_{1i}E_{1'i'}$ and the remaining terms can be written in terms of $I_{ij'}(N-1)$. To make this more explicit, we introduce the notation

$$I_{ij'}(N; 1 \cdots N) = I_{ij'}(N), \quad (3.10)$$

i.e., $I_{ij'}(N; 1 \cdots N)$ is the coefficient $I_{ij'}(N)$ in the expansion (3.8) in which there are N pair energies E_i which are labeled by the indices $1 \cdots N$. Our hypothesis is that $I_{11'}(N; 1 \cdots N)$ is given by (3.9). Expanding (3.9) by the minors of the first row, we see that our hypothesis implies that

$$I_{11'}(N; 1 \cdots N) = \bar{I}_{N-1}(2 \cdots N) - 2 \sum_{i,j=2}^{N} P_{i'j'} E_{1i}E_{1'i'} \bar{I}_{ij'}(N-1; 2 \cdots N), \quad (3.11)$$

where the bar means that $c_{ii'}$ has been replaced by $c_{ii'} - 2E_{1i}E_{1'i'}$. Conversely, if we can prove (3.11) then, starting with (3.6), we can prove our

hypothesis (3.9). We now turn to the proof of (3.11).

The first step in the proof of (3.11) is an expansion of the product

$$\prod_{k=2}^{N} A_k(2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1}$$

appearing in (3.8). Let us define $T_{ij'}(N)$ by

$$\prod_{k=2}^{N} A_k(2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1}$$

$$= \sum_{i,j=1}^{N} T_{ij'}(N)(2\epsilon_1 - E_i)^{-1}(2\epsilon_1 - E_{j'})^{-1}. \quad (3.12)$$

From this definition, we have

$$\sum_{i,j=1}^{N} T_{ij'}(N)(2\epsilon_1 - E_i)^{-1}(2\epsilon_1 - E_{j'})^{-1}$$

$$= A_N \sum_{i,j=1}^{N-1} T_{ij'}(N-1)(2\epsilon_1 - E_i)^{-1}(2\epsilon_1 - E_{j'})^{-1}$$

from which we can derive the recursion relations

$$T_{ii'}(N) = [A_N(i) - E_{iN}E_{i'N'}]T_{ii'}(N-1), \quad (3.13a)$$

$$T_{ij'}(N) = [A_N(ij) - E_{iN}\hat{E}_{i'N'} - \hat{E}_{jN}E_{i'N'}]T_{ij'}(N-1), \quad (3.13b)$$

$$T_{iN'}(N) = -E_{iN}E_{N'i'}P_{i'N'}T_{ii'}(N-1) - \sum_{\substack{k=1 \\ (k\neq i)}}^{N-1} \hat{E}_{kN}E_{N'k'}P_{k'N'}T_{ik'}(N-1), \quad (3.13c)$$

$$T_{Nj'}(N) = -E_{Nj}E_{j'N'}P_{iN}T_{jj'}(N-1) - \sum_{\substack{k=1 \\ (k\neq j)}}^{N-1} E_{Nk}\hat{E}_{k'N'}P_{kN}T_{kj'}(N-1), \quad (3.13d)$$

$$T_{NN'}(N) = -\sum_{k=1}^{N-1} E_{kN}E_{k'N'}P_{kN}P_{k'N'}T_{kk'}(N-1), \quad (3.13e)$$

for $i \neq j$ and $i, j < N$. These recursion relations are to be solved subject to the initial condition

$$T_{11'}(1) = 1. \quad (3.14)$$

It is easily verified that the solution of Eqs. (3.13) that satisfies (3.14) is

$$T_{11'}(N) = \prod_{k=2}^{N} [A_k(1) - E_{1k}E_{1'k'}], \quad (3.15)$$

$$T_{ii'}(N) = -E_{1i}E_{1'i'} \times \prod_{\substack{k=2 \\ (k\neq i)}}^{N} [A_k(1i) - 2E_{ik}E_{i'k'}], \quad (3.16)$$

$$T_{1j'}(N) = -E_{1j}E_{j'1'} \times \prod_{\substack{k=2 \\ (k\neq j)}}^{N} [A_k(1j) - 2E_{1k}E_{j'k'}], \quad (3.17)$$

$$T_{i1'}(N) = -E_{i1}E_{1'i'} \times \prod_{\substack{k=2 \\ (k\neq i)}}^{N} [A_k(1i) - 2E_{ik}E_{1'N'}], \quad (3.18)$$

and

$$T_{ij'}(N) = 2E_{i1}E_{j'i'}E_{ij}E_{j'i'} \times \prod_{\substack{k=2 \\ (k\neq i,j)}}^{N} [A_k(1ij) - 3E_{iN}E_{j'N'}], \quad (3.19)$$

for $N > 1$, $i \neq j$, and $1 < i, j \leq N$. This completes the expansion of the operator product

$$\prod_{k=2}^{N} A_k(2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1}.$$

We next derive an expression for $I_{11'}(N)$. From (3.8) and (3.12), we have

$$\sum_{ij'} I_{ij'}(N)(2\epsilon_1 - E_i)^{-1}(2\epsilon_1 - E_{j'})^{-1} = \sum_{l=1}^{N} P_{1l}S'_N \sum_{ij'} T_{ij'}(N)(2\epsilon_1 - E_i)^{-1}(2\epsilon_1 - E_j)^{-1}.$$

In this expression, we write

$$S'_N = S'_{N-1}(2' \cdots N') \sum_m P_{1'm'}$$

and pick out the coefficient of $(2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_{1'})^{-1}$ to get

$$I_{11'}(N) = S'_N(2' \cdots N') \times \left\{ \sum_{i,j=1}^{N} P_{1i}P_{1'j'}T_{ij'}(N) \right\}. \quad (3.20)$$

We now use (3.14)–(3.19) for $T_{ij'}(N)$, carry out the indicated transpositions, and partially perform the sums on i and j [using the same techniques that were used to sum (2.28)] to obtain

$$I_{11'}(N) = S'_{N-1}(2' \cdots N') \left\{ \prod_{k=2}^{N} [A_k(1) - 2E_{1k}E_{1'k'}] - 2\sum_{i=2}^{N} E_{1i}E_{1'i'} \prod_{\substack{k=2 \\ (k\neq i)}}^{N} [A_k(1i) - 3E_{1k}E_{1'k'}] \right\} \quad (3.21)$$

which is the desired expression for $I_{11'}(N)$.

As our next step, we derive an equivalent expression for the part of (3.21) that is inside the curly brackets. It will be equivalent to (3.21) in the sense that the two expressions will only differ by terms which vanish on symmetrization by S'_{N-1}. In order

to do this, we write (3.21) as

$$I_{11'}(N) = S'_{N-1}(2' \cdots N')J_{11'}(N; 1 \cdots N), \quad (3.22)$$

where

$$J_{11'}(N; 1 \cdots N) = \prod_{k=2}^{N} \bar{A}_k(1) - 2 \sum_{i=2}^{N} E_{1i}E_{1'i'} \prod_{\substack{k=2 \\ (k\neq i)}}^{N} [\bar{A}_k(1i) - E_{1k}E_{1'k'}]. \quad (3.23)$$

The bar notation has been defined in Eq. (3.11). It is clear that the first term of (3.23) contributes the first term, $\bar{I}_{N-1}(2 \cdots N)$, of (3.11). We therefore concentrate on the second term of (3.23) and we will show that

$$J_{11'}(N; 1 \cdots N) = \prod_{k=2}^{N} \bar{A}_k(1) - 2 \sum_{i=2}^{N} E_{1i}E_{1'i'}\bar{J}_{ii'}(N-1; 2 \cdots N) \quad (3.24)$$

plus terms that vanish after symmetrization by S'_{N-1} and therefore do not contribute to $I_{11'}$.

We begin our proof of Eq. (3.24) by writing (3.23) as

$$J_{11'}(N; 1 \cdots N) = \prod_{k=2}^{N} \bar{A}_k(1) - 2 \sum_{i=2}^{N} P_{2i}P_{2'i'}E_{12}E_{1'2'}K_N(12; 3 \cdots N), \quad (3.25)$$

where

$$K_N(12; 3 \cdots N) = \prod_{k=3}^{N} [\bar{A}_k(12) - E_{1k}E_{1'k'}] \quad (3.26)$$

and writing (3.24) as

$$J_{11'}(N; 1 \cdots N) = \prod_{k=2}^{N} \bar{A}_k(1) - 2 \sum_{i=2}^{N} P_{2i}P_{2'i'} \times E_{12}E_{1'2'}\bar{J}_{22'}(N-1; 2 \cdots N). \quad (3.27)$$

We will then prove the equivalence of (3.25) and (3.27). This equivalence is proven by first deriving expressions for K_N and $\bar{J}_{22'}$ and then showing that the difference between (3.25) and (3.27) vanishes when it is symmetrized by S'_{N-1}.

An expression for K_N may be derived by writing Eq. (3.26) as

$$K_N(12; 3 \cdots N) = [\bar{A}_N(12) - E_{1N}E_{1'N'}]K_{N-1}(12; 3 \cdots N-1)$$

and iterating this equation. In this way we obtain

$$K_N(12; 3 \cdots N) = \prod_{k=3}^{N} \bar{A}_k(12) - \sum_{l=3}^{N} E_{1l}E_{1'l'} \prod_{\substack{k=3 \\ (k\neq l)}}^{N} \bar{A}_k(12l) + 2 \sum_{l<m} E_{1l}E_{1'l'}E_{1m}E_{1'm'} \times \prod_{(k\neq l,m)} \bar{A}_k(12lm) - \cdots + (-)^N(N-2)! \prod_{k=3}^{N} E_{1k}E_{1'k'}. \quad (3.28)$$

We may obtain a similar expression for $\bar{J}_{22'}$ by using (3.24) to write it as

$$\begin{aligned}\bar{J}_{22'}(N-1; 2 \cdots N) &= \prod_{k=3}^{N} [\bar{A}_k(12) - 2E_{2k}E_{2'k'}] \\ &\quad - 2 \sum_{i=3}^{N} \prod_{l=i+1}^{N} [\bar{A}_l(12) - 2E_{2l}E_{2'l'}]E_{2i}E_{2'i'} \\ &\quad \times \prod_{k=3}^{i-1} [\bar{A}_k(12) - 3E_{2k}E_{2'k'}] \\ &= [\bar{A}_N(12) - 2E_{2N}E_{2'N'}]\bar{J}_{22'}(N-2; 2 \cdots N-1) \\ &\quad - 2E_{2N}E_{2'N'} \prod_{k=3}^{N-1} [\bar{A}_k(12) - 3E_{2k}E_{2'k'}].\end{aligned}$$

Solving this for $\bar{J}_{22'}(N-1; 2 \cdots N)$, we have

$$\begin{aligned}&\bar{J}_{22'}(N-1; 2 \cdots N) \\ &= \prod_{k=3}^{N} \bar{A}_k(12) - 4 \sum_{l=3}^{N} E_{2l}E_{2'l'} \prod_{\substack{k=3 \\ (k\neq l)}}^{N} \bar{A}_k(12l) \\ &\quad + 18 \sum_{l<m} E_{2l}E_{2'l'}E_{2m}E_{2'm'} \prod_{(k\neq l,m)} \bar{A}_k(12lm) \\ &\quad - \cdots + (-)^N(N-1)(N-1)! \prod_{k=3}^{N} E_{2k}E_{2'k'}.\end{aligned} \quad (3.29)$$

Equations (3.28) and (3.29) are the expressions for K_N and $\bar{J}_{22'}$ that may be used to prove the equivalence of (3.23) and (3.24).

We now use the results of the preceding paragraph to prove that the difference between (3.23) and (3.24) gives a zero contribution to $I_{11'}(N)$, i.e., that

$$S'_{N-1}(2' \cdots N') \sum_{k=2}^{N} P_{2k}P_{2'k'}E_{12}E_{1'2'} \times \{K_N(12; 3 \cdots N) - \bar{J}_{22'}(N-1; 2 \cdots N)\} = 0. \quad (3.30)$$

From (3.28) and (3.29), we have

$$K_N(12; 3 \cdots N) - \bar{J}_{22'}(N-1; 2 \cdots N) = \sum_{l=3}^{N} (4E_{2l}E_{2'l'} - E_{1l}E_{1'l'}) \prod_{\substack{k=3 \\ (k\neq l)}}^{N} \bar{A}_k(12l)$$

$$- 2 \sum_{l<m} (9E_{2l}E_{2'l'}E_{2m}E_{2'm'} - E_{1l}E_{1'l'}E_{1m}E_{1'm'}) \times \prod_{(k\neq l,m)} \bar{A}_k(12lm) + \cdots - (-)^N(N-2)! \times \left[(N-1)^2 \prod_{k=3}^{N} E_{2k}E_{2'k'} - \prod_{k=3}^{N} E_{1k}E_{1'k'}\right]. \quad (3.31)$$

Consider the contribution to (3.30) of the last term in (3.31). It is proportional to

$$S'_{N-1}(2' \cdots N') \sum_{i=2}^{N} P_{2i}P_{2'i'}E_{12}E_{1'2'} \times \left[(N-1)^2 \prod_{k=3}^{N} E_{2k}E_{2'k'} - \prod_{k=3}^{N} E_{1k}E_{1'k'}\right].$$

which, on taking advantage of the symmetries of the expression, can be written as

$$(N-1)^2(N-2)! \times \left[\left(\sum_{i=2}^{N} P_{2i} \prod_{k=3}^{N} E_{2k}E_{12}\right)^2 - \prod_{k=2}^{N} E_{1k}^2\right]. \quad (3.32)$$

We have dropped the primes since the argument is valid with or without them. We show that (3.32) vanishes by defining

$$X_N = \sum_{i=2}^{N} P_{2i} \prod_{k=3}^{N} E_{2k}E_{12}$$

and proving by induction that

$$X_N = \prod_{k=2}^{N} E_{1k}. \quad (3.33)$$

For example, for $N = 3$, we have

$$X_3 = (1 + P_{23})E_{23}E_{12} = E_{23}(E_{12} - E_{13}) = E_{12}E_{13}. \quad (3.34)$$

For arbitrary N, we have, using $E_{2N} - E_{1N} = E_{1N}E_{2N}/E_{12}$,

$$X_N = \sum P_{2i}\left(E_{1N} + \frac{E_{1N}E_{2N}}{E_{12}}\right) \times \prod_{k=3}^{N-1} E_{2k}E_{12} + P_{2N} \prod_{k=3}^{N} E_{2k}E_{12}$$

$$= E_{1N} \sum_{i=2}^{N-1} P_{2i} \prod_{k=3}^{N-1} E_{2k}E_{12} + E_{1N} \sum_{i=2}^{N-1} P_{2i} \prod_{k=3}^{N} E_{2k} - \prod_{k=1}^{N} E_{Nk}$$

$$= E_{1N}X_{N-1} - E_{1N}P_{1N}X_{N-1} - \prod_{k=1}^{N-1} E_{Nk}. \quad (3.35)$$

Substituting the induction hypothesis (3.33) for X_{N-1} in (3.35) proves that (3.33) is true for X_N and therefore the induction is complete. Substitution of (3.33) into (3.32) indicates that the contribution of the last term of (3.31) to (3.30) is zero. Similar arguments can be made to show that the other terms of (3.31) give a zero contribution to (3.30). Thus the equivalence of (3.23) and (3.24) has been demonstrated.

Equations (3.22) and (3.24) may now be used to prove our recursion relation (3.11). For, on substituting (3.24) into (3.22), we have

$$I_{11'}(N; 1 \cdots N) = S'_{N-1}(2' \cdots N')\left\{\prod_{k=2}^{N} \bar{A}_k(1) - 2 \sum_{i=2}^{N} E_{1i}E_{1'i'}\bar{J}_{ii'}(N-1; 2 \cdots N)\right\}. \quad (3.36)$$

If we write

$$S'_{N-1}(2' \cdots N') = \sum_{j=2}^{N} P_{i'j'}S'_{N-2}(2' \cdots i-1', i+1' \cdots N')$$

and again use (3.22) in the second term of (3.36), then we can write it as

$$I_{11'}(N; 1 \cdots N) = S'_{N-1}(2' \cdots N') \prod_{k=2}^{N} \bar{A}_k(1) - 2 \sum_{i,j=2}^{N} P_{i'j'}E_{1i}E_{1'i'}\bar{I}_{ii'}(N-1; 2 \cdots N). \quad (3.37)$$

As has already been noted, the first term in (3.37) is $\bar{I}_{N-1}(2 \cdots N)$. With this identification, we see that Eq. (3.11) has been proven and this equation can then be used to verify (3.9).

There remain two things to prove before we can complete the derivation of Eq. (3.2). We must first prove that it is only the identity permutation in the S'_{N-1} of (3.9) that contributes to $I_{11'}(N)$. Then, we must prove that $I_{ij'}$ is diagonal in i and j. These are both proven by the same methods that were used to prove that it was only the identity permutation in S'_N which contributed to the normalization and we refer the reader to the end of Sec. II for the proof. For future reference, it should be noted that it is only at this stage of the proof that the specific properties of $c_{ij'}$ are used.

We conclude this section with a few parenthetical remarks on the form of N_1. If we differentiate Eqs. (1.8) with respect to g, we get the system of equations

$$\left[c_i - 2 \sum_j E_{ij}^2\right] \frac{dE_i}{dg} + 2 \sum_j E_{ij}^2 \frac{dE_j}{dg} = \frac{-1}{g^2} \quad (3.38)$$

for the derivatives dE_i/dg of the pair energies. Comparing (3.38) with (3.3), we see that

$$\alpha_i = -g^2 dE_i/dg. \quad (3.39)$$

Substituting this into (3.2), we get

$$N_f = -g^2 \frac{d}{dg}\left(\sum_{i=1}^{N} \frac{1}{2\epsilon_f - E_i}\right). \tag{3.40}$$

In this form it is easy to prove that

$$\sum_f N_f = N$$

since

$$\sum_f \sum_i \frac{1}{(2\epsilon_f - E_i)} = \frac{N}{g}.$$

It should also be noted that (3.40) is exactly the same expression one would get if the Pauli principle were neglected and one considered a system of bosons in a one-body potential described by the Hamiltonian (1.1) with no Pauli principle.

V. THE EXPECTATION VALUE OF $b_f^\dagger b_{f'}$

Let us define the expectation value

$$B(11') = \langle\psi|\ b_1^\dagger b_{1'}\ |\psi\rangle. \tag{4.1}$$

It is essential for the evaluation of the expectation values of one- and two-body operators. For, within the framework of the seniority-zero states of the pairing-force Hamiltonian, we have

$$\langle\psi|\ a_{1\sigma}^\dagger a_{1'\sigma'}\ |\psi\rangle = \delta_{\sigma\sigma'} B(11') \tag{4.2}$$

and, for $1 \neq 1'$,

$$\langle\psi|\ a_{1\sigma_1}^\dagger a_{2\sigma_2}^\dagger a_{2'\sigma_2} a_{1'\sigma_1}\ |\psi\rangle$$
$$= (\delta_{\sigma_1\sigma_1'}\delta_{\sigma_2\sigma_2'} - \delta_{\sigma_1\sigma_2'}\delta_{\sigma_2\sigma_1'})\delta_{12}\delta_{1'2'} B(11'). \tag{4.3}$$

Since $B(11) = N_1$, it is suggested that, in some kind of approximation, an expression for $B(11')$, for $1 \neq 1'$, can be obtained by replacing $(2\epsilon_1 - E_i)^{-2}$ in (4.2) by $(2\epsilon_1 - E_i)^{-1}(2\epsilon_{1'} - E_i)^{-1}$, i.e.,

$$B(11') \cong \sum_{i=1}^{N} \alpha_i (2\epsilon_1 - E_i)^{-1}(2\epsilon_{1'} - E_i)^{-1}. \tag{4.4}$$

In this section, we show that

$$B(11') = \sum_{i=1}^{N} \alpha_i(11')(2\epsilon_1 - E_i)^{-1}(2\epsilon_{1'} - E_i)^{-1}, \tag{4.5}$$

where, in an approximation to be described below, $\alpha_i(11') \cong \alpha_i$. The derivation of (4.5) closely follows the one given for N_1 in Sec. III. We first derive an operator-product expression for $B(11')$ and then show that it is formally the same as the one that we derived for N_1, (3.5), with a modified definition of the c_{ij}'s. We therefore can take over all the results of Sec. III that do not depend upon the particular form of the c_{ij}'s. In particular, we obtain a result that is similar to (3.9). However, due to the modification of the c_{ij}'s, all permutations in S'_{N-1} will contribute to B. If we ignore this modification, we obtain the approximation (4.4) and, if we retain it, we obtain the exact expression (4.5). We conclude this section by calculating the expectation value of the Hamiltonian, using the approximation (4.4). Here, we are able to give an explicit expression for the errors made by the approximation. Throughout this section, we assume that $1 \neq 1'$.

We derive an operator expression for $B(11')$ by starting with

$$B(11') = N \sum_{2\cdots N} \psi(12\cdots N)\psi(1'2\cdots N)$$

which on substitution of (1.5) and (2.1), becomes

$$B(11') = C^2\Big(\sum_{j=1}^{N} P_{1j}\Big) S'_N \sum_{2\cdots N} \theta(12\cdots N)$$
$$\times \prod_{k=2}^{N} (1 - \delta_{1'k})(2\epsilon_1 - E_1)^{-1}(2\epsilon_{1'} - E_{1'})^{-1}$$
$$\times \prod_{i=2}^{N} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1}, \tag{4.6}$$

where we have replaced S_N by

$$(N-1)! \sum_{j=1}^{N} P_{1j},$$

as we did in Eq. (3.4), and we have used

$$\theta(12\cdots N)\theta(1'2\cdots N)$$
$$= \theta(12\cdots N) \prod_{k=2}^{N} (1 - \delta_{1'k}). \tag{4.7}$$

We next use the recursion relation for the θ-functions (2.5) to write (4.7) as

$$\theta(12\cdots N) \prod_{k=2}^{N} (1 - \delta_{1'k}) = \theta(12\cdots N-1)$$
$$\times \prod_{k=2}^{N-1} (1 - \delta_{1'k})\Big[1 - \sum_{l=1}^{N-1} \delta_{lN}\Big](1 - \delta_{1'N})$$
$$= \theta(12\cdots N-1) \prod_{k=2}^{N-1} (1 - \delta_{1'k})$$
$$\times \Big[1 - \delta_{1N} - \delta_{1'N} - \sum_{l=2}^{N-1} \delta_{lN}\Big], \tag{4.8}$$

where, in multiplying the last two factors, we have neglected the term $\delta_{1N}\delta_{1'N}$ because we have assumed $1 \neq 1'$ and we have neglected the terms

$$\delta_{1'N} \sum_{l=2}^{N-1} \delta_{lN}$$

because they do not contribute to (4.6). Substituting

(4.8) into (4.6) and performing the sum on N, as we did in Eq. (2.8), we obtain

$$B(11') = C^2 \sum_{i=1}^{N} P_{1i} S_N' \Big[c_{NN'} - (2\epsilon_1 - E_N)^{-1} \times (2\epsilon_{1'} - E_{N'})^{-1} - (2\epsilon_{1'} - E_N)^{-1}(2\epsilon_{1'} - E_{N'})^{-1} - \sum_{l=2}^{N-1} \hat{E}_{1N}\hat{E}_{1'N'} \Big] \sum_{2\cdots N-1} \theta(1 \cdots N-1) \times \prod_{k=2}^{N-1} (1 - \delta_{1'k})(2\epsilon_1 - E_1)^{-1}(2\epsilon_{1'} - E_{1'})^{-1} \times \prod_{i=2}^{N-1} (2\epsilon_i - E_i)^{-1}(2\epsilon_i - E_{i'})^{-1}. \quad (4.9)$$

Equation (4.9) may be iterated to obtain an operator-product expression for $B(11')$. Before doing so we note that the sum on l in (4.9) does not include the term with $l = 1$. In order to make the resulting operator product as much like the one derived for N_1, (3.5), as is possible, we include the term with $l = 1$ by adding and subtracting $\hat{E}_{1N}\hat{E}_{1'N'}$ to the operator in (4.9). We next note that the operator $\hat{E}_{1N}\hat{E}_{1'N'}$ which we have added to (4.9) operates on $(2\epsilon_1 - E_1)^{-1}(2\epsilon_{1'} - E_{1'})^{-1}$ and that

$$\hat{E}_{1N}\hat{E}_{1'N'}(2\epsilon_1 - E_1)^{-1}(2\epsilon_{1'} - E_{1'})^{-1} = (2\epsilon_1 - E_1)^{-1}(2\epsilon_1 - E_N)^{-1} \times (2\epsilon_{1'} - E_{1'})^{-1}(2\epsilon_{1'} - E_{N'})^{-1}.$$

Thus, this operator may be replaced by $(2\epsilon_1 - E_N)^{-1}(2\epsilon_{1'} - E_{N'})^{-1}$. We may therefore write the operator in the square brackets of (4.9) as $[A_N + U_{NN'}(11')]$, where we define $U_{NN'}$ by

$$U_{NN'}(11') = (2\epsilon_1 - E_N)^{-1}(2\epsilon_{1'} - E_{N'})^{-1} - (2\epsilon_1 - E_N)^{-1}(2\epsilon_1 - E_{N'})^{-1} - (2\epsilon_{1'} - E_N)^{-1}(2\epsilon_{1'} - E_{N'})^{-1} \quad (4.10)$$

and A_N has been defined by (2.14). Making these modifications in (4.9) and then iterating it, we obtain the expression

$$B(11') = C^2 \sum_{i=1}^{N} P_{1i} S_N' \prod_{k=2}^{N} [A_k + U_{kk'}(11')] \times (2\epsilon_1 - E_1)^{-1}(2\epsilon_{1'} - E_{1'})^{-1} \quad (4.11)$$

which is the desired operator form of B.

Equation (4.11) is identical to Eq. (3.5) except for the replacement of one ϵ_1 by $\epsilon_{1'}$ and $C_{kk'}$ by $C_{kk'}(11')$, where

$$C_{kk'}(11') = C_{kk'} + U_{kk'}(11'). \quad (4.12)$$

In order to avoid possible confusion at this point, we emphasize that in (4.12) 1 and 1′ represent the variables f_1 and $f_{1'}$, and k and k' represent the indices p_k and $p_{k'}$. We may now take over the results of Sec. IV that do not depend upon the explicit form of $C_{kk'}$. That is, by analogy with (3.8), we may write $B(11')$ as

$$B(11') = C^2 \sum_{ij} I_{ij'}(11') \times (2\epsilon_1 - E_i)^{-1}(2\epsilon_{1'} - E_{j'})^{-1}, \quad (4.13)$$

where, from (3.9), $I_{ij'}(11')$ is a symmetrizer times a determinant, e.g.,

$$I_{11'}(11') = S_{N-1}'(2' \cdots N') \times \begin{vmatrix} 1 & 2E_{12}E_{1'2'} \cdots 2E_{1N}E_{1'N'} \\ 1 & c_{22'}(11') - 2\sum{}' E_{i2}E_{j'2'} \cdots 2E_{2N}E_{2'N'} \\ \cdots & \\ 1 & 2E_{N2}E_{N'2'} \cdots c_{NN'}(11') - 2\sum{}' E_{iN}E_{j'N'} \end{vmatrix}. \quad (4.14)$$

There are two sources of difficulty that appear when one tries to use (4.14). The first source is that $I_{ij'}(11')$ is no longer diagonal in i and j and all N^2 terms in (4.13) contribute to B. Secondly, and more seriously, all $(N-1)!$ permutations in S_{N-1}' contribute to (4.19). An approximation which removes these difficulties is immediately suggested and is to ignore the $U_{kk'}$'s in (4.14). This is the approximation that leads to (4.4) and it can be thought of as a zeroth-order approximation in the U's. Possibly a better approximation which also solves these difficulties is to ignore the U_{ij}'s for $i \neq j$. This would lead to an expression similar to (4.5) in which $\alpha_i(11')$ is obtained by writing α_i as a ratio of two determinants and then replacing c_i in the numerator by $c_i(11') = c_i + U_{ii}(11')$. Still other approximations might be developed by retaining only those U_{ij}'s that make a significant contribution to the quantity being calculated. However, these would depend upon the particular system being considered. In this paper, we only briefly consider the first of the above approximations.

As a test of the approximation (4.4), we use it to calculate the expectation value of the Hamiltonian. While this test will not indicate the errors of this approximation which have a fluctuating sign, it will indicate any tendency of (4.4) to over or under estimate B. Also, this test has the advantage that the error can be written out explicitly. We start by using (1.1), (3.1), and (4.1) to write the expectation value of the Hamiltonian as

$$\langle\psi| H |\psi\rangle = \sum_1 2\epsilon_1 N_1 - g \sum_{11'} B(11'). \quad (4.15)$$

We may use (1.8), (2.3), and (3.2) to write the first term of (4.15) as

$$\sum_1 2\epsilon_1 N_1 = \sum_i \alpha_i \sum_1 \frac{2\epsilon_1}{(2\epsilon_1 - E_i)^2} = \sum_i \alpha_i [g_i^{-1} + E_i c_i]. \tag{4.16}$$

The second term of (4.15) can be written as

$$-g \sum_{11'} B(11') = -g \sum_i \alpha_i g_i^{-2}, \tag{4.17}$$

where we have used (1.6) and (4.4). Combining these results, we have

$$\langle \psi | H | \psi \rangle = \sum_i \alpha_i [E_i c_i + g(g^{-1} - g_i^{-1})/g_i]. \tag{4.18}$$

If we multiply the ith equation of (3.3) by E_i and then sum on i, we get

$$\sum_i \alpha_i E_i c_i = \sum_i E_i + 2 \sum_{i,j}{}' \alpha_i E_{ij}$$

for the first term of (4.18). Substituting this into (4.18), using the explicit form (1.10) of g_i^{-1}, and noting that $\sum_i E_i$ is the exact energy of the state, we obtain the expression

$$\frac{1}{g} \sum_{i=1}^{N} \alpha_i \left[1 - \left(\frac{g}{g_i} \right)^2 \right] \tag{4.19}$$

for the error in (4.18). This may be easily evaluated when the E_i's are determined.

V. REMOVAL OF THE RESTRICTIONS ON THE EQUATIONS[4]

We now consider the nature of the restrictions (1.12) on the solutions of Eqs. (1.8). In order to simplify the notation, we will write E_i for E_{ν_i}. Therefore, Eqs. (1.8) and (1.10) become

$$F(E_i) = 1/g_i, \qquad i = 1 \cdots N, \tag{5.1}$$

and

$$\frac{1}{g_i} = \frac{1}{g} + 2 \sum_{j=1}^{N}{}' \frac{1}{(E_j - E_i)}. \tag{5.2}$$

We discuss the restrictions by deriving conditions that must be satisfied if the restrictions are to be violated. That is, we will first assume that K of the pair energies E_i are equal and therefore violate the restrictions. We then derive certain conditions that must be fulfilled if this assumption is to be consistent with Eqs. (5.1). These conditions are: (1) The value of the K equal pair energies must be equal to one of the values of $2\epsilon_f$ appearing in the definition (1.9) of $F(E)$. We will call this value $2\epsilon_0$. (2) The pair degeneracy Ω_0 of the level whose energy is ϵ_0 must be $K - 1$. (The pair degeneracy of a level is defined to be one-half the total degeneracy, e.g., in the spherical-shell model of the nucleus, the pair degeneracy of a level j is $j + \frac{1}{2}$.) (3) If one assumes that K of the pair energies are almost, but not quite, equal and retains terms that are first order in the differences $2\epsilon_0 - E_i$, then these differences are proportional to the K Kth roots of one. (4) The interaction strength g must be a root of a Kth-degree algebraic equation.

This last condition is the most important. For, it implies that there are no more than K different values of the interaction strength at which K of the pair energies can be equal. Thus, if we take all the values of K, $K \leq N$, that are consistent with condition (2), then we obtain a finite set of values of g at which the restrictions can be violated for a finite system. This set contains values of g that correspond to excited, seniority-zero states as well as the ground state and the interpretation of any particular value of g in this set must follow from a detailed investigation of Eqs. (5.1). However, the values of g at which the restrictions can be violated for a given state of the system are contained in this set and are therefore finite in number. The energy and wavefunction of a state at one of these values of g may be obtained by a limiting procedure similar to that described in this section. In this way, the energy and wavefunction of a state may be obtained for all values of g and the restrictions removed.

The derivation of the above conditions might proceed by first assuming that the K pair energies, $E_1 \cdots E_k$, are almost equal and that they have power series expansions in some small parameter Δ which, at the end of the calculation, is to be set equal to zero. These expansions would then be substituted into (5.1) and the coefficients of like powers of Δ equated. However, if one carries out this program, one obtains an indeterminate system of equations, with more unknowns than equations, for the expansion coefficients. We circumvent this difficulty by first introducing a new set of unknowns defined by

$$\mathcal{E}_l = \sum_{i=1}^{K} (E_i)^l. \tag{5.3}$$

where we have assumed that the energy scale has been adjusted so that $\epsilon_0 = 0$. We then obtained set of equations for these new unknowns by taking the ith equation of (5.1) and multiplying it by $(E_i)^n$, $n = 0 \cdots K$, and then summing on i from 1 to K. One can then show that each of the variables

[4] R. W. Richardson, Bull. Am. Phys. Soc. 9, 75 (1964).

(5.3) is of order Δ^K and that it is only necessary to consider the lowest-order term in each of the equations in order to derive the four conditions given above.

The derivation proceeds by first considering the case $K = N$ when all the pair energies are equal. We then generalize the results to include the cases $K < N$. Throughout the derivation we assume that $g \neq 0$. We now turn to the derivation of these four conditions for $K = N$.

The first point that we prove is that if $E_1 \cdots E_K$ are equal, then their value must be equal to $2\epsilon_f$ for some values of f. This is done by showing that $E_1 \cdots E_K$ equal but not equal to some $2\epsilon_f$ is inconsistent with Eqs. (5.1). For, if we adjust the energy scale so that the value $E_1 \cdots E_K$ have when they are equal is zero and expand in powers of Δ then the E_i are of order Δ and, if $\epsilon_f \neq 0$ for all f, $F(E_i)$ is of order one. Since we have assumed that $g \neq 0$, the g_i^{-1} are of order $1/\Delta$ and, to this order, are given by

$$g_i^{-1} = 2 \sum_{j=1}^{K}{}' \frac{1}{(E_j - E_i)}, \qquad i = 1 \cdots K.$$

Equations. (2.1), to order $1/\Delta$, are then given by

$$2 \sum_{j=1}^{K}{}' \frac{1}{(E_j - E_i)} = 0, \qquad i = 1 \cdots K. \tag{5.4}$$

Multiplying these equations by E_i and summing on i yields $K(K-1) = 0$. However, since $K \geq 2$, this is inconsistent and therefore $\epsilon_f = 0$ for some value of f. In what follows, we assume that this is so and we denote the pair degeneracy of this level by Ω_0.

We now prove our second point which is $\Omega_0 = K - 1$. This is proven by repeating the above argument with the one modification that, to order $1/\Delta$,

$$F(E_i) = \Omega_0/(-E_i).$$

Thus, Eqs. (5.4) become

$$2 \sum_{j=1}^{K}{}' \frac{1}{(E_j - E_i)} = \Omega_0/(-E_i), \quad i = 1 \cdots K. \tag{5.5}$$

Multiplying (5.5) by E_i and summing on i, we get $K(K-1) = K\Omega_0$, which proves our point.

In what follows, we assume that the first two conditions are satisfied, i.e., $\epsilon_0 = 0$, and $\Omega_0 = K - 1$. We can therefore write

$$F(E) = (K-1)/(-E) + \tilde{F}(E), \tag{5.6}$$

where $\tilde{F}(E)$ has a power series expansion which is valid in some finite domain about $E = 0$. We write this power series expansion as

$$\tilde{F}(E) = \sum_{l=0}^{\infty} F_l E^l. \tag{5.7}$$

Our third and fourth points are that, to lowest order in Δ, the E_i are proportional to the K Kth roots of one and that g must satisfy a Kth-degree equation. These are proven by considering the new set of equations which is obtained from (5.1) by multiplying the ith equation of (5.1) by $(E_i)^n$, $n = 0 \cdots K$, and summing on i, i.e., the set of equations

$$\sum_{i=1}^{K} E_i^n g_i^{-1} = \sum_{i=1}^{K} E_i^n F(E_i), \qquad n = 0 \cdots K, \tag{5.8}$$

to lowest order in Δ.

We first write Eqs. (5.8) in terms of the $\mathcal{E}_l$ defined by (5.3). The left-hand side of (5.8), for $n = 0$, is

$$\sum_i g_i^{-1} = \sum_i g^{-1} + 2 \sum_{i,j}{}' \frac{1}{(E_i - E_j)} = \frac{K}{g} \tag{5.9}$$

and, for $n = 1$, it is

$$\sum_i E_i g_i^{-1} = \mathcal{E}_1 g^{-1} - K(K-1). \tag{5.10}$$

For $n > 1$, we write the left-hand side of (5.8) as

$$\begin{aligned} \sum_i E_i^n g_i^{-1} &= \sum_i E_i^n g^{-1} + 2 \sum_{i,j}{}' \frac{E_i^n}{E_j - E_i} \\ &= \mathcal{E}_n g^{-1} - \sum_{i,j}{}' \frac{E_j^n - E_i^n}{E_j - E_i} \\ &= \mathcal{E}_n g^{-1} - \sum_{i,j}{}' \sum_{l=0}^{n-1} E_i^{n-1-l} E_j^l, \end{aligned} \tag{5.11}$$

where we have used

$$x^n - y^n = (x - y) \sum_{l=0}^{n-1} x^{n-1-l} y^l.$$

Using (5.3), we can write

$$\sum_{i,j}{}' E_i^{n-1-l} E_j^l = \mathcal{E}_{n-1-l}\mathcal{E}_l - \mathcal{E}_{n-1}$$

and, noting that $\mathcal{E}_0 = K$, we can then write (5.11) as

$$\sum_i E_i^n g_i^{-1} = \mathcal{E}_n g^{-1} - (2K - n)\mathcal{E}_{n-1} - \sum_{l=1}^{n-2} \mathcal{E}_{n-1-l}\mathcal{E}_l. \tag{5.12}$$

We may use (5.3), (5.6), and (5.7) to write the right-hand side of Eq. (5.8) as

$$\sum_i E_i^n F(E_i) = -(K-1)\mathcal{E}_{n-1} + \sum_l F_l \mathcal{E}_{l+n}. \tag{5.13}$$

Thus, Eqs. (5.8) may be rewritten as

$$K/g = -(K-1)\varepsilon_{-1} + \sum_l F_l\varepsilon_l, \tag{5.14a}$$

$$\varepsilon_1/g = \sum_l F_l\varepsilon_{l+1}, \tag{5.14b}$$

$$\varepsilon_n/g - (K-n+1)\varepsilon_{n-1} - \sum_{m=1}^{n-2}\varepsilon_{n-1-m}\varepsilon_m = \sum_l F_l\varepsilon_{l+n}, \quad n > 1. \tag{5.14c}$$

This completes the change of variables from the E_i's to the ε_l's. Equations. (5.14) are valid as long as all the E_i's are within the domain of convergence of the power series expansion (5.7).

We next consider the order of the ε_l's with an eye towards linearizing (5.14c). We show that ε_l, for $l > 0$, is at least of order Δ^K. This is best treated by the discussion of a simple example. We therefore consider the case for which $K = 3$. In this case, Eq. (5.14c), for $n = 2$ and 3, becomes

$$\varepsilon_2/g - 2\varepsilon_1 = F_0\varepsilon_2 + F_1\varepsilon_3 + \cdots, \tag{5.15}$$

$$\varepsilon_3/g - \varepsilon_2 - \varepsilon_1^2 = F_0\varepsilon_3 + \cdots. \tag{5.16}$$

From (5.3), we know that if E_i is of order Δ, then ε_l is at least of order Δ^l. This plus (5.15) implies that ε_1 must be at least of order Δ^2 which when substituted into (5.16) implies that ε_2 is at least of order Δ^3. Returning to (5.15), we see that if ε_2 is at least of order Δ^3, then so is ε_1. This line of reasoning terminates at this point because the coefficient of ε_{n-1} in (5.14c), for $n = 4$, vanishes. This result may be easily generalized to arbitrary K where $\varepsilon_1 \cdots \varepsilon_K$ are all at least of order K. We therefore assume that $\varepsilon_1 \cdots \varepsilon_K$ are all of order Δ^K. This allows us to neglect the terms in (5.14a) with $l > 0$, the terms in (5.14b) with $l > K-1$, the terms with $l > K-n$, and the nonlinear terms in (5.14c). We may also ignore Eqs. (5.14c) when $n > K$. Equations (5.14) therefore become

$$KG = -(K-1)\varepsilon_{-1}, \tag{5.17a}$$

$$G\varepsilon_1 = \sum_{l=1}^{K-1} F_l\varepsilon_{l+1}, \tag{5.17b}$$

$$G\varepsilon_n - (K-n+1)\varepsilon_{n-1} = \sum_{l=1}^{K-n} F_l\varepsilon_{l+n}, \quad n = 2 \cdots K, \tag{5.17c}$$

where we have defined

$$G = g^{-1} - F_0 \tag{5.18}$$

and have used $\varepsilon_0 = K$.

We already have sufficient information to prove our third point that the E_i's are proprotional to the K Kth roots of one. For, if we form the polynomial

$$P(x) = \prod_{i=1}^{K}(x - E_i),$$

then, by definition, the pair energies are the roots of the equation $P(x) = 0$. If we write $P(x)$ as

$$P(x) = \sum_{n=0}^{K} a_n x^{K-n}, \tag{5.19}$$

then $a_0 = 1$ and a_n, for $n > 0$, may be formed from combinations of the ε_l's, e.g., $a_1 = \varepsilon_1$ and $a_2 = \frac{1}{2}(\varepsilon_1^2 - \varepsilon_2)$. Therefore a_n, for $n > 0$, is of order Δ^K. Since x is of order Δ, the orders of the terms in (5.19) are Δ^K, for $n = 0$ and K, and Δ^{2K-n}, for $n = 1 \cdots K-1$. Therefore, to lowest order in Δ, $P(x) = x^K + a_K$ and the pair energies are roots of the equation

$$x^K + a_K = 0$$

which proves our assertion.

Returning to Eqs. (5.17), we note that we have $K+1$ equations in the $K+2$ unknowns G, ε_{-1}, $\varepsilon_1 \cdots \varepsilon_K$. We therefore need to derive one more equation before they can be solved. We can, in fact, derive two more equations and this will provide a check on the consistency of our arguments. These equations have their origin in our third point above. For, if we write

$$E_i = b_i\Delta + c_i\Delta^2 + \cdots,$$

then we know from above that $(b_i)^K = 1$ when the expansion parameter Δ is appropriately defined. We therefore have immediately that $\varepsilon_K = K\Delta^K$, to lowest order in Δ. We can also derive a relation between ε_{-1} and ε_{K-1}. For, from its definition

$$\varepsilon_{-1} = \sum_i (b_i\Delta + c_i\Delta^2 + \cdots)^{-1} = -\sum_i \frac{c_i}{b_i^2} + \cdots$$

since $\sum b_i^{-1} = 0$. Also, from its definition,

$$\varepsilon_{K-1} = \sum_i (b_i\Delta + c_i\Delta^2 + \cdots)^{K-1} = (K-1)\left(\sum_i \frac{c_i}{b_i^2}\right)\Delta^K + \cdots.$$

Comparing these two results, we have

$$\varepsilon_{K-1} = -(K-1)\varepsilon_{-1}\Delta^K + \cdots.$$

We now write

$$\varepsilon_n = KA_n\Delta^K + \cdots, \quad n = 1 \cdots K,$$

and combine the above results with Eqs. (5.17) to obtain the K equations

$$G = A_{K-1}, \tag{5.20a}$$

$$GA_1 - \sum_{l=1}^{K-2} F_l A_{l+1} = F_{K-1}, \qquad (5.20b)$$

$$-(K-n+1)A_{n-1} + GA_n - \sum_{l=1}^{K-n-1} F_l A_{l+n} = F_{K-n}, \quad n = 2 \cdots K-1, \qquad (5.20c)$$

for the K unknowns G, $A_1 \cdots A_{K-1}$. Equation (5.17c), with $n = K$, reproduces Eq. (5.20a) and serves as a useful check on the consistency of the above scheme.

Our fourth point, that g satisfies a Kth degree equation, is proven by noting that (5.20b, c) may be solved for A_{K-1} and that the result is a polynomial in G of degree $K-2$ divided by a polynomial in G of degree $K-1$. When this expression for A_{K-1} is substituted into Eq. (5.20a), we have a Kth-degree equation in G which then determines the interaction strength through (5.18). For example, for $K = 2$ and 3, we have the equations

$$G^2 - F_1 = 0$$

and

$$G^3 - 3F_1G + 2F_2 = 0$$

for G.

The above results may be easily generalized to include the case for which $K < N$. In this case we denote the K equal pair energies by $E_1 \cdots E_K$ and adjust the energy scale so that their value is zero. The remaining pair energies $E_{K+1} \cdots E_N$ are distinct and not equal to zero. We split Eqs. (5.1) into two sets: those for which $i = 1 \cdots K$ and those for which $i = K+1 \cdots N$. For the first set we write

$$\frac{1}{g_i} = \frac{1}{g} + 2 \sum_{j=1}^{K}{}' \frac{1}{(E_j - E_i)} + 2 \sum_{j=K+1}^{N} \frac{1}{(E_j - E_i)}, \quad i = 1 \cdots K, \qquad (5.21)$$

and for the second set

$$\frac{1}{g_i} = \frac{1}{g} + \frac{2K}{-E_i} + 2 \sum_{j=K+1}^{N}{}' \frac{1}{(E_j - E_i)}, \quad i = K+1 \cdots N. \qquad (5.22)$$

The corrections to (5.22) are of order Δ^K since they are of the form

$$\sum_{l=1} \varepsilon_l E_i^{-l-1}.$$

Therefore, the corrections to E_i, $i = K+1 \cdots N$, are of order Δ^K and can be ignored in (5.21). The second sum in (5.21) can therefore be transposed to the left-hand side of Eqs. (5.1) and included in a modified definition of $F(E)$. The quantities $F_0 \cdots F_{K-1}$ will now depend upon $E_{K+1} \cdots E_N$. These remaining pair energies are determined by Eqs. (2.1) with $1/g_i$ given by (5.22). Thus, the only difference between this case and the previous case with $K = N$ is a modification of the definitions of the quantities $F_0 \cdots F_{K-1}$.

The highly unlikely case of two or more groups of pair energies being equal but with members of different groups being unequal may also be treated by a simple generalization of the above.

Thus, we have shown that the restrictions on the solution of Eqs. (5.1) can only be violated for a discrete set of values of the interaction strength. However, these violations are of no consequence since the wavefunction and energy can be obtained for these values of the interaction strength by requiring them to be continuous functions of g. Also, since we have chosen to label the states of the Hamiltonian in the same way that the states of the noninteracting system are labeled, the states of the interacting and noninteracting systems may be put into one-to-one correspondence. Therefore, since these states are eigenvectors of finite matrices of the same rank, we may conclude that all the states of the interacting system may be written in the form that we have given.

We conclude this section with some comments on the interpretation of the values of g at which the restrictions are violated. These values have previously been called the singular values[3] of g and we will now indicate in what sense they are singular. It can be shown that the pair energies possess well behaved derivatives of any order with respect to g at any point where they satisfy Eqs. (5.1). This can be shown by successive differentiation of Eqs. (5.1), see Eq. (3.38) above. Thus, the only singularities that the pair energies have on the real-g axis are located at the singular values of g. The work of this section then suggests that the singularities are branch points and that the pair energies behave as $(g_{\text{sing}} - g)^{1/K}$ in the neighborhood of g_{sing}, the singular value of g. Of course, one should not conclude from this that the singular values of g give some information about the domain of convergence of the perturbation series for the energy of the state since these singularities cancel each other in the energy which is the sum of the pair energies. However, a knowledge of the behavior of the pair energies in the neighborhoods of the singular values of g does aid in the solution of Eqs. (5.1).

VI. ACKNOWLEDGMENT

The author wishes to acknowledge many helpful discussions with Professor J. K. Percus about this work.

Quantum Electrodynamics Without Indefinite Metric*

ASHER PERES
Department of Physics, Technion—Israel Institute of Technology, Haifa, Israel
(Received 8 May 1964; final manuscript received 25 January 1965)

The electromagnetic potentials $A_\alpha(x)$ are quantized in such a way that their space components are Hermitian and their time component anti-Hermitian. On the other hand, the metric in Hilbert space and the Hamiltonian are positive definite. The above formalism, which leads to the usual commutator $[A_\alpha(x), A_\beta(y)] = ig_{\alpha\beta}D_0(x - y)$, is shown to be Lorentz-covariant over the manifold of physical states. The latter contain neither longitudinal nor scalar photons (rather than an equal number of them, as in the usual theory). This definition is also Lorentz covariant. The S-matrix is not unitary, but satisfies $PS^*PSP = P$, where P is the projection operator over physical states. In other words, the S-matrix is unitary over the subspace of physical states, this being sufficient for the interpretation of the theory.

1. INTRODUCTION

IT is often stated that the quantization of the electromagnetic potentials leads to serious difficulties, and in particular that that covariant commutator[1]

$$[A_\alpha(x), A_\beta(y)] = ig_{\alpha\beta}D_0(x - y) \tag{1}$$

implies the introduction of states with negative norm,[2,3] i.e., of a pseudo-Hilbert space. Unfortunately, very little is known of the *mathematical* properties of pseudo-Hilbert spaces, except that most theorems which can be proved for Hilbert spaces would not be valid in them (e.g., the Schwarz inequality fails). From the *physical* point of view, the indefinite metric leads to monstrosities such as negative probabilities, or probabilities larger than one (there is no Schwarz inequality), etc.

Traditionally, these nonsensical results have been escaped only at the expense of manifest covariance, e.g., by taking the radiation gauge[4]

$$a_0(x) = 0, \tag{2}$$

$$[a_m(x), a_n(y)] = -i[\delta_{mn} - \partial_m\partial_n(\partial_0)^{-2}]D_0(x - y), \tag{3}$$

or one of its generalizations.[5-7] Alternative quantization schemes have also been proposed, in which manifest gauge independence is achieved at the expense of locality,[8-13] or use is made of various limiting processes.[14,15] Though mathematically and physically sound, such alternatives are rather inconvenient because they lead to a considerable increase of the computational labor, to finally obtain the same results as with Eq. (1).[15a]

In this paper, we propose a new approach, in which we maintain altogether Eq. (1) and the positive definiteness of the metric in Hilbert space and of the Hamiltonian. Physical states contain neither longitudinal nor scalar photons (rather than symmetric combinations of them,[16] as in the usual theory). All these desirable results are obtained in Sec. 2 by making $A_0(x)$ anti-Hermitian, while the $A_k(x)$ remain Hermitian. It then follows that the S-matrix

$$S = T\left\{\exp\left[-i\int j^\alpha(x)A_\alpha(x)d^4x\right]\right\}, \tag{4}$$

is not unitary. It is however shown, in Sec. 3, that it is unitary over the subspace of physical states, this being clearly sufficient for the interpretation

* The research reported in this document has been supported in part by the Aerospace Research Laboratories, OAR, under Grant AF-EOAR-64-20, through the European Office of Aerospace Research, U. S. Air Force.

[1] Greek indices run from 0 to 3, Latin indices from 1 to 3. The signature of the metric $g_{\alpha\beta}$ is $(+---)$. Creation operators are denoted as c^+, annihilation operators as c^-. We use natural units $c = \hbar = 1$.

[2] S. N. Gupta, Proc. Phys. Soc. (London) **A63**, 681 (1950).

[3] K. Bleuler, Helv. Phys. Acta **23**, 567 (1950).

[4] A. A. Sokolov, *Vvedenie v Kvantovuyu Elektrodinamiku* (Gosudarstvennoe Izdatelstvo, Moscow, 1958), p. 56.

[5] J. G. Valatin, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **26**, No. 13 (1951).

[6] L. E. Evans and T. Fulton, Nucl. Phys. **21**, 492 (1960).

[7] A. Peres, Nuovo Cimento **34**, 346 (1964).

[8] K. F. Novobátsky, Z. Physik **111**, 292 (1938).

[9] F. J. Belinfante and J. S. Lomont, Phys. Rev. **84**, 541 (1951).

[10] F. J. Belinfante, Phys. Rev. **84**, 546, 648 (1951).

[11] C. L. Hammer and R. H. Good, Jr., Phys. Rev. **111**, 342 (1958); Ann. Phys. (N. Y.) **12**, 463 (1961).

[12] I. Goldberg, Phys. Rev. **112**, 1361 (1958).

[13] S. Mandelstam, Ann. Phys. (N. Y.) **19**, 1 (1962).

[14] R. Utiyama, T. Imamura, S. Sunakawa, and T. Dodo, Progr. Theoret. Phys. (Kyoto) **6**, 587 (1951).

[15] T. T. Wu, Phys. Rev. **129**, 1420 (1963).

[15a] *Footnote added in proof:* See also S. Sato, Progr. Theoret. Phys. (Kyoto) **31**, 256 (1964); A. Katz, Nuovo Cimento **37**, 342 (1965); S. Weinberg, Phys. Rev. (to be published).

[16] These symmetric combinations are popularly known as "an equal number of longitudinal and scalar photons." Actually, they have the form

$$(c_0^+ - c_3^+)^n\Psi_0 = [(c_0^+)^n - n(c_0^+)^{n-1}c_3^+ + \cdots]\Psi_0.$$

of the theory. Finally, the Lorentz covariance of this formalism is discussed in Appendix A.

From a "practical" point of view, this paper contains of course no new result, but only a better justification of Eq. (1).

2. QUANTIZATION OF THE FREE ELECTROMAGNETIC FIELD

It was shown long ago by Peierls[17] that the commutator

$$[F_{\alpha\beta}(x), F_{\gamma\delta}(x)] = -i(g_{\alpha\gamma}\partial_\beta\partial_\delta - g_{\beta\gamma}\partial_\alpha\partial_\delta + g_{\beta\delta}\partial_\alpha\partial_\gamma - g_{\alpha\delta}\partial_\beta\partial_\gamma)D_0(x-y) \quad (5)$$

can be obtained directly from a variational principle, without any reference to the potentials. If we now introduce the radiation gauge potentials $a_k(x)$ by means of

$$a_k = \int F_{0k}\, dx^0, \quad (6)$$

$$F_{mn} = \partial_m a_n - \partial_n a_m, \quad (7)$$

$$\partial^k a_k = 0, \quad (8)$$

then we readily obtain, from (5), the radiation gauge commutator (3). Note that Eqs. (6)–(8) are compatible, in virtue of the Maxwell equations,

$$\partial^\alpha F_{\alpha\beta} = 0, \quad (9)$$

$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0. \quad (10)$$

As is well known, a Fourier analysis of (3) and (8) leads to

$$a_m(x) = \sum_j \int e_m^{(j)}(\mathbf{k})[c_j^+(\mathbf{k})e^{ikx} + c_j^-(\mathbf{k})e^{-ikx}](4\pi^{3/2}\,|\mathbf{k}|^{\frac{1}{2}})^{-1}d^3\mathbf{k}, \quad (11)$$

where

$$kx \equiv |\mathbf{k}|\,x^0 - \mathbf{k}\cdot\mathbf{x}, \quad (12)$$

and the $c_j^+(\mathbf{k})$ are creation and annihilation operators, satisfying

$$[c_j^+(\mathbf{k})]^* = c_j^-(\mathbf{k}), \quad (13)$$

and

$$[c_i^-(\mathbf{p}), c_j^+(\mathbf{q})] = \delta_{ij}\delta(\mathbf{p}-\mathbf{q}), \quad (14)$$

$$[c_i^-(\mathbf{p}), c_j^-(\mathbf{q})] = [c_i^+(\mathbf{p}), c_j^+(\mathbf{q})] = 0. \quad (15)$$

The unit vectors $\mathbf{e}^{(j)}$ are orthogonal to $\mathbf{k}$ and to each other. (Note that the index j is an *enumerator*, taking the values 1 and 2 only. It is *not* a vectorial index.) For later convenience, we shall also define

$$\mathbf{e}^{(3)} = \mathbf{k}/|\mathbf{k}|, \quad (16)$$

and

$$e_\alpha^{(0)} = (1, 0, 0, 0), \quad (17)$$

so that

$$\sum_\gamma e_\alpha^{(\gamma)} e_\beta^{(\gamma)} = \delta_{\alpha\beta}, \quad (18)$$

where the sum over γ runs from 0 to 3.

It is easily seen that Eqs. (11), (14), and (15) are equivalent to the commutator (3). Unfortunately, the latter has an ugly noncovariant aspect (though actually, it is of course Lorentz covariant)[18] and, in order to obtain the *manifestly* covariant commutator (1), it is customary to introduce unphysical (longitudinal and scalar) photons, with creation and annihilation operators $c_3^\pm(\mathbf{k})$ and $c_0^\pm(\mathbf{k})$, respectively. However, if we write, as usual,

$$A_\alpha(x) = \sum_\gamma \int e_\alpha^{(\gamma)}(\mathbf{k})[c_\gamma^+(\mathbf{k})e^{ikx} + c_\gamma^-(\mathbf{k})e^{-ikx}](4\pi^{3/2}\,|\mathbf{k}|^{\frac{1}{2}})^{-1}d^3\mathbf{k}, \quad (19)$$

then we do not obtain (1), but rather $[A_\alpha(x), A_\beta(y)] = -i\delta_{\alpha\beta}D_0(x-y)$. Thus, in order to maintain (19), one usually takes, instead of (14),

$$[c_0^-(\mathbf{p}), c_0^+(\mathbf{q})] = -\delta(\mathbf{p}-\mathbf{q}). \quad (20)$$

Now, if we multiply (20) by $f^*(\mathbf{p})f(\mathbf{q})$, where f is an arbitrary function, then integrate over $\mathbf{p}$ and $\mathbf{q}$ and take the vacuum expectation value of the result, we obtain

$$\left|\int c_0^+(\mathbf{q})f(\mathbf{q})d^3\mathbf{q}\Psi_0\right|^2 = -\int |f(\mathbf{q})|^2 d^3\mathbf{q}, \quad (21)$$

i.e., we are led to a pseudo-Hilbert space with an indefinite metric, as is well known.[2,3] [It seems that this result could be avoided by reinterpreting $c_0^-(\mathbf{k})$ as a creation operator and $c_0^+(\mathbf{k})$ as an annihilation operator, so that $c_0^-(\mathbf{k})\Psi_0 \neq 0$ and $c_0^+(\mathbf{k})\Psi_0 = 0$. This interpretation, however, leads to other difficulties, as shown in Appendix B.]

The alternative which we propose here is to allow $A_0(x)$ to be anti-Hermitian, by taking instead of (17),

$$e_\alpha^{(0)} = (i, 0, 0, 0). \quad (22)$$

We now have

$$\sum_\gamma e_\alpha^{(\gamma)} e_\beta^{(\gamma)} = -g_{\alpha\beta}, \quad (23)$$

so that we maintain both Eq. (19) and well-behaved commutation relations like (14), and yet obtain the manifestly covariant commutator (1). The drawback of this method is, of course, that $A_0(x)$ is anti-

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

[18] B. Zumino, J. Math. Phys. 1, 1 (1960).

Hermitian. We shall presently see that this does not lead to any serious consequences.

First, let us call *physical states* those which do not contain any longitudinal or scalar photon, i.e., which satisfy

$$c_3^-(\mathbf{k})\Psi = c_0^-(\mathbf{k})\Psi = 0. \tag{24}$$

Only matrix elements between physical states are physically meaningful, since the other states have been introduced artificially, just for mathematical convenience. Now, it is easily seen from (11), (13), (19), and (24) that

$$(\Phi, A_\alpha(x)\Psi) = (\Phi, a_\alpha(x)\Psi), \tag{25}$$

if both Φ and Ψ are physical. It follows that when we compute $(\Phi, F_{\alpha\beta}\Psi)$, we can take

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha, \tag{26}$$

rather than the *correct* definition [see Eqs. (6)–(7)]

$$F_{\alpha\beta} = \partial_\alpha a_\beta - \partial_\beta a_\alpha. \tag{27}$$

In other words, the $A_\alpha(x)$ practically behave as if they were the true electromagnetic potentials. (This property is, of course, shared by our formalism and by the usual one.)

We still have to prove that this method is Lorentz covariant, and in particular that the definition (24) of physical states is invariant. Since this is a rather technical question, we shall postpone it to Appendix A.

3. THE *S*-MATRIX

Quantization of the free field is well known to be relatively trivial with respect to such questions as Lorentz invariance, elimination of the unphysical photons, etc. The real problems arise only when coupling to a quantized current is considered, and we must now examine how the non-Hermitian $A_\alpha(x)$ can interact with a Hermitian current density $j^\alpha(x)$. For instance, it is clear that the operator $A_\alpha(x)$ cannot satisfy Maxwell-like equations. This is not surprising, since $A_\alpha(x)$ is *not* related to the "correct" electromagnetic potential $a_\alpha(x)$ by means of an operator gauge transformation.[7]

In this paper, we shall therefore relinquish the old-fashioned approach to quantum field theory, according to which the field operators must satisfy equations of motion similar to those of classical fields.[19] As is well known, this approach leads to serious mathematical difficulties[20,21] because the fields are not well-behaved operators, but rather operator-valued distributions, and products of distributions are mathematically ill-defined. (Anyhow, the field equations are never used in practical calculations.)

The standpoint which we take here is midway between that of the classic Lagrangian field theories and the more recent *S*-matrix approach.[22,23] Namely, we consider only the initial and final states of the physical system, which are generated by the free fields and related by the *S*-matrix, the latter being also constructed in terms of the free fields. It can be shown that, for a given interaction, the condition of causality alone is sufficient to determine the *S*-matrix completely (except for renormalization).[24] It is given by Eq. (4), above.

The essential difficulty caused by a non-Hermitian $A_\alpha(x)$ is that the *S*-matrix (4) is not unitary (because the current density j^α is Hermitian). Let us however recall the physical motivation for the unitarity of the *S*-matrix.

The expression

$$|(\Psi_b, S\Psi_a)|^2 = (S\Psi_a, \Psi_b)(\Psi_b, S\Psi_a), \tag{28}$$

is the probability of a transition from state Ψ_a to state Ψ_b. It can also be written as $(S\Psi_a, P_bS\Psi_a)$, where P_b is the projection operator over state Ψ_b. If we now take a complete set of orthogonal states Ψ_b, then we must have

$$\sum_b (S\psi_a, P_bS\psi_a) = (S\psi_a, S\psi_a) = 1, \tag{29}$$

because $\sum_b P_b = 1$, and (29) is the total probability of obtaining *any* state. From (29), we obtain $(\Psi_a, S^*S\Psi_a) = 1$, and, since this is valid for any normalized state Ψ_a, we finally have $S^*S = 1$.

The above argument clearly applies only when all the states of Hilbert space have an equal status. In our case, however, we should consider only the physical states, satisfying (24), since the other ones have been introduced only for mathematical convenience. We thus have

$$\sum_b P_b = P, \tag{30}$$

[19] It is therefore somewhat abusive to call the present theory "Quantum Electrodynamics."

[20] I. E. Segal, J. Math. Phys. **5**, 269 (1964).

[21] J. G. Taylor, Nuovo Cimento, Suppl. **1**, 856 (1964).

[22] S. Weinberg, Phys. Rev. **133**, B 1318 (1964); **134**, B 882 (1964).

[23] A. Peres, Phys. Rev. **137**, B 696 (1965).

[24] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), Sec. 18. These authors actually construct the *S*-matrix from the requirements of Lorentz invariance, unitarity and causality. However, a careful examination of their argument shows that the condition of causality alone is sufficient to determine completely the *S*-matrix (except for renormalization). Lorentz invariance and unitarity can then be verified *post facto*. In particular, unitarity requires the interaction Lagrangian to be Hermitian.

where P is the projection operator over physical states, and, instead of (29),

$$(S\Psi_a, PS\Psi_a) = (\Psi_a, S^*PS\Psi_a) = 1, \quad (31)$$

valid for any *physical* Ψ_a. It can easily be shown that (31) is equivalent to

$$PS^*PSP = P, \quad (32)$$

i.e., the S-matrix is unitary over the subspace of physical states. Our task thus is to prove that (32) indeed holds.

Although a direct proof is difficult, it can be seen as follows that this is true. First, we note that all our matrix elements (28) between physical states are the same as in the usual theory (since we use the same propagator) and therefore[18] the same as with the radiation gauge (3). In the latter case, there are no unphysical photons at all and the S-matrix is clearly unitary (in the usual theory with the indefinite metric, S is pseudounitary). It follows then that (32) must indeed hold in the present theory.

4. CONCLUDING REMARKS

From the "practical" point of view, this paper contains no new result.

Our sole aim was to show that the familiar commutator (1) can be derived without introducing the indefinite metric, i.e., without violating the principles of quantum theory. Incidentally, we have also obtained a definition of physical states—no unphysical photon at all—which seems more satisfactory than the traditional one, according to which both kinds of unphysical photons may appear in certain symmetric combinations.[16]

The price we have to pay for these desirable results may seem high: $A_\alpha(x)$ is not Hermitian, and the S matrix is not unitary. However, this does not lead to any difficulty if we consider only matrix leements between physical states, which are the only ones appearing in nature. Unphysical photons—those which violate Hermiticity and unitarity—appear only in the *intermediate* states, where they enable us to use the simple Feynman propagator, rather than the cumbersome radiation-gauge propagator.

Whether or not the present quantization method is preferable to the other ones[2-15a] is of course a matter of taste. Its spirit is closest to the Gupta–Bleuler theory (as far as practical calculations are concerned, it is identical to it). It seems, however, that the use of anti-Hermitian operators is intellectually more honest than the introduction of an indefinite metric. I hope some readers will concur with this view.

ACKNOWLEDGMENT

I am indebted to Professor W. H. Furry for a critical discussion.

APPENDIX A

We still have to prove that the above formalism is Lorentz covariant. It is evident that the $A_\alpha(x)$ field, as we have defined it, is *not* a vector field.[19] Namely, under a homogeneous Lorentz transformation with coefficients $L_\alpha{}^{\beta'}$, which is represented in our Hilbert space of states by a unitary operator $U(L)$, we do *not* have

$$A_{\alpha'}(Lx) = L_{\alpha'}{}^{\beta}U(L)A_\beta(x)U^*(L). \quad (33)$$

Indeed, the unitary operator U cannot alter the Hermitian or anti-Hermitian nature of $A_\beta(x)$, but the matrix $L_{\alpha'}{}^{\beta}$ would mix up Hermitian and anti-Hermitian components.[25]

A formal proof that the theory is nevertheless Lorentz covariant must therefore be given by displaying explicitly the ten generators P_α and $M_{\alpha\beta}$ of the Lorentz group.[26] This is most easily done by taking the latter as those for the purely transverse (i.e., physically correct) electromagnetic potentials.[27] For instance, we have

$$H \equiv P_0 = \sum_i \int |\mathbf{k}|\, c_i^+(\mathbf{k})c_i^-(\mathbf{k})d^3\mathbf{k}, \quad (34)$$

the sum being taken over the physical photons only.

It then follows from Eq. (11) that the transformation properties of $a_\alpha(x)$ are correctly reproduced: the Lorentz transformation includes, in addition to the usual vector transformation, a gauge transformation which restores $a_0(x) \equiv 0$.[28] [It follows that even $a_\alpha(x)$ is not a genuine vector field.][29]

On the other hand, the unphysical part of the potentials

[25] The reader may have in mind the counterexample of the Dirac matrices γ_α which satisfy $\gamma_\alpha\gamma_\beta + \gamma_\beta\gamma_\alpha = 2g_{\alpha\beta}$. Then γ_0 is a Hermitian matrix, while the γ_k are anti-Hermitian matrices. The whole system of matrices is nevertheless Lorentz invariant. However, in the transformation law $\gamma_{\alpha'} = L_{\alpha'}{}^{\beta}S\gamma_\beta S^{-1}$, the matrix S is *not* unitary (the Lorentz group has no finite-dimensional unitary representation).

[26] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949); **34**, 592 (1962).

[27] There are many other possible choices. Ours is the simplest (and also the closest to physical reality). Whether or not these generators can be obtained, as in classical physics, from an action principle, is, in our opinion, an irrelevant question.

[28] See Ref. 18, Appendix B.

[29] One should not be surprised by this result, since even the "polarization vectors" $e_\alpha{}^{(j)}$ are not genuine vectors (their time component is always zero). Their transformation law has been discussed in detail in a lucid paper by S. Weinberg, Phys. Rev. **135**, B 1049 (1964), see especially Appendix A. I am very much indebted to Dr. S. Weinberg for making his results available to me prior to publication.

$$A_\alpha(x) - a_\alpha(x) = \int [i(c_0^+ e^{ikx} + c_0^- e^{-ikx}) + (\mathbf{k}/|\mathbf{k}|)(c_3^+ e^{ikx} + c_3^- e^{-ikx})] d^3\mathbf{k}/4\pi^{3/2}\,|\mathbf{k}|^{\frac{1}{2}}, \quad (35)$$

is *not affected at all* by Lorentz transformations, even by mere translations. For instance, we have

$$i[A_0(x), H] \equiv 0 \neq \partial_0 A_0(x). \quad (36)$$

In other words, the unphysical part of the potential does not transform at all as a local field. At first sight, this result might look somewhat strange. However, there actually is no compelling reason to expect unphysical quantities to have well-behaved transformation laws, because, as long as we restrict ourselves to physical states, all the observables are independent from them [cf. Eq. (25)]. In this connection, we note that Eq. (24), which defines physical states, is Lorentz invariant.

We must still show that the S-matrix (4), which is constructed by coupling a vector current to something which is not a vector, is nevertheless Lorentz invariant. The proof is quite easy: it trivially follows from the fact that, in the actual construction of the S-matrix, the electromagnetic polentials $A_\alpha(x)$ appear only through their commutator (1), or the corresponding propagator, which are *manifestly* covariant.

We may therefore summarize our approach by saying that, in order to avoid the indefinite metric, we have been compelled to give $A_\alpha(x)$ extremely complicated properties. However, the commutator $[A_\alpha(x), A_\beta(y)]$ is a very simple tensor, and since this is the only quantity which actually enters in the calculations, we can claim that the present theory is "manifestly covariant for all practical purposes."

APPENDIX B

It is sometimes proposed to replace the time component of (19) by

$$A_0(x) = \int [c_0^+(\mathbf{k}) e^{-ikx} + c_0^-(\mathbf{k}) e^{+ikx}] d^3\mathbf{k}/4\pi^{3/2}\,|\mathbf{k}|^{\frac{1}{2}}, \quad (37)$$

thereby avoiding the indefinite metric, while maintaining both Eq. (1) and the Hermiticity of $A_0(x)$. Unfortunately, Eq. (37) implies, as is shown below, that the scalar photons have negative energies,[30] a property which leads to well-known troubles, in particular with respect to the definition of the vacuum.

To prove the above assertion, let us recall that if a creation operator a_m^+ creates a particle in a state $u_m(x)$, then one particle states can be mapped by functions of x according to

$$a_m^+ \Psi_0 \rightarrow u_m(x). \quad (38)$$

Let us now consider a complete orthonormal set of such $u_m(x)$, and let us go over to a new orthonormal basis

$$v_\alpha(x) = \sum_m c_{\alpha m} u_m(x), \quad (39)$$

where $c_{\alpha m}$ is a unitary matrix. In virtue of the superposition principle, the mapping (38) must be linear, so that the operator a_α^+, which creates a particle in state $v_\alpha(x)$, is given by

$$a_\alpha^+ = \sum_m c_{\alpha m} a_m^+. \quad (40)$$

Likewise we have, for the annihilation operator

$$a_\alpha^- = \sum_m \bar{c}_{\alpha m} a_m^-. \quad (41)$$

It then follows from the unitarity of $c_{\alpha m}$ that

$$\phi^-(x) = \sum_m a_m^- u_m(x) = \sum_\alpha a_\alpha^- v_\alpha(x), \quad (42)$$

and

$$\phi^+(x) = \sum_m a_m^+ \bar{u}_m(x) = \sum_\alpha a_\alpha^+ \bar{v}_\alpha(x), \quad (43)$$

are independent of the choice of the orthonormal basis, and therefore are acceptable field operators. On the other hand, this property does not hold for $\sum a_m^+ u_m(x)$ or $\sum a_m^- \bar{u}_m(x)$. We are therefore compelled to interpret, in (37), e^{ikx} as the wavefunction of the scalar photon, and e^{-ikx} as the conjugate of its wavefunction. It then follows from (12) that the scalar photons of (37) have a negative energy, which is of course better than a negative probability, but still is a rather unwelcome feature.

[30] This fact is well known [cf., e.g., S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson and Company, Evanston, Illinois, 1961) p. 224], but we shall now prove it without using, as usual, the explicit form of the Hamiltonian, in order to avoid a possible controversy on the choice of the latter.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Quantum Mechanical Time

BAYARD RANKIN
Department of Mathematics, Case Institute of Technology, Cleveland, Ohio
(Received 24 April 1964)

Some suggestions on the probabilistic structure of quantization are first obtained by studying local properties of random variables such as continuity and differentiability with respect to a topology in the sample space. This analysis is extended to analytic functions on the complex plane and conditions are formulated under which probability densities are expressed as sums of absolute squares of complex numbers. Physical restrictions are introduced through the Schrödinger equation for a single bound-particle system. As a result, observations on a physical system become identified with the random selection of points in a topological measure space and the physical observables such as energy, position, and momentum, as well as time, are identified with measurable functions on appropriate spaces. Time, considered as a function, appears to be multiple-valued with spacings between multiple values and a detailed functional structure that characterizes and is characterized by the physical system under observation and the physical observables being measured. Its detailed functional structure is related to the physically measurable probabilities of quantum theory and it is seen to serve in the capacity of a conditioning random variable in the computation of quantum mechanical expectations.

INTRODUCTION

IN the Schrödinger representation for a physical system, time enters as a real parameter, while energy, position, and momentum are represented by operators. From the Schrödinger point of view, it is also the case that statistical assertions enter in the postulates concerning certain Hermitian operators, while time is not treated statistically. We wish to display an alternate quantum mechanical representation, entirely consistent with the Schrödinger picture, in which energy, position, momentum, and time are treated alike and each is represented by a measurable function on an appropriate topological measure space. In this presentation, observations on a physical system are identified with the random selection of points in a topological space. The points selected determine the values of the physical observables, including time, and the functional structures of the energy, position, momentum, and time determine the forms of the physically observable distributions. We are thus justified in using the logic of measure-theoretic probability theory in constructing expectations and conditional expectations. It results that conditional expectations, given the time function, can be identified with the time-dependent quantum mechanical expectations.

While the treatment of the time and the space coordinates as similar entities is in the spirit of relativity theory, all the physical restrictions introduced into the theory are dictated by the nonrelativistic Schrödinger equation. The proof of our basic theorem is limited to the case of a one-dimensional bound-particle system in which potential energy is independent of time, though there appears to be no reason why all our arguments would not generalize.

From the rudiments of measure-theoretic probability, we recall that for pairs of finite, single-valued, real, measurable functions Y and X defined on a normalized measure space S, there exists, under suitable circumstances, a conditional probability function $p(y_n \mid x)$ or a conditional probability density function $p(y \mid x)$ that relates to the conditional expectation of Y, given X, through one of the representations

$$E[Y \mid X = x] = \sum_n y_n p(y_n \mid x),$$

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} yp(y \mid x)\, dy.$$

We can combine these in the Stieltjes representation with conditional distribution function $F(y \mid x)$ of Y, given X,

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} y\, dF(y \mid x).$$

This analysis can be extended to other settings in which S is a measurable subset of the complex space (or, more generally, a vector space) and in which the conditioning variable X is multiple-valued and complex. We will assume in this introduction that the details of such extensions are known.

While the form of the function $F(y \mid x)$ is most frequently implied from the measure-theoretic characterization of the functions Y and X, namely their

joint distribution, it can also be deduced from the analytic or topological properties of Y and X, such as their differentiable properties, if such are defined. As we will see in later sections of this paper, the latter circumstance is more natural to our purposes, in view of the fact that analytical information is given to us through the wavefunction.

In the calculus of quantum theory also, distribution functions enter, though these are not generally related to well-defined measurable functions on measure spaces. That is, for each physical observable with corresponding Hermitian operator L, with eigenspectrum $\Lambda(L) = \{\lambda\}$, and with corresponding resolution of the identity E_L, we write, for state vector ϕ_t,

$$F_{L,t}(\lambda) = (E_L(\lambda)\phi_t, \phi_t),$$

where $F_{L,t}$ is a distribution function that is related to quantum mechanical expectation through the identity

$$(L\phi_t, \phi_t) = \int_{-\infty}^{\infty} \lambda \, dF_{L,t}(\lambda).$$

It will be of particular interest to us to identify a distribution function of the type $F_{L,t}$ with a conditional distribution function F of one measurable function given another.

For each physical observable corresponding to Hermitian operator L with pure or discrete eigenspectrum, no degeneracy being present among the eigenvalues, let us represent the most general normalized solution ϕ_t, $-\infty < t < \infty$, of the Schrödinger equation as a set of ordered pairs,

$$(\omega_t(\lambda), \lambda), \ \omega_t(\lambda) \quad \text{complex}, \qquad \lambda \in \Lambda(L), \qquad \text{(I)}$$

in which the $\omega_t(\lambda)$ are restricted as follows:

$$F_{L,t}(\lambda) = \begin{cases} \sum_{u \le \lambda} |\omega_t(u)|^2, & \text{if } \Lambda \text{ is discrete;} \\ \int_{u \le \lambda} |\omega_t(u)|^2 \, du, & \text{if } \Lambda \text{ is continuous.} \end{cases}$$

As examples, if L is the energy operator, Λ is discrete and $\lambda = E_n$:

$$\omega_t(E_n) = a_n e^{-iE_n t/\hbar}.$$

If L is the position operator, $\Lambda = \{q: -\infty < q < \infty\}$, and for $\lambda = q$

$$\omega_t(q) = \phi_t(q).$$

If L is the momentum operator, $\Lambda = \{p: -\infty < p < \infty\}$, and for $\lambda = p$

$$\omega_t(p) = \psi_t(p),$$

where

$$\psi_t(p) = \frac{1}{(2\pi\hbar)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \phi_t(q) e^{-ipq/\hbar} \, dq.$$

With the aid of the representation (I) of the general solution to the Schrödinger equation, we now formulate two theorems. The contents of Secs. 3, 4, and 5 will provide a proof for the theorems, though these sections will progress informally by exploring the probabilistic meaning of the theorems first and by imposing physical restrictions last. Theorem 2 is the primary conclusion of this paper. From it we can derive a number of physical conclusions concerning the quantum mechanical structure of time.

Theorem 1. Let the set of ordered pairs (I) represent the most general solution to the Schrödinger equation for a one-dimensional bound-particle system with time-independent potential energy. Let L be the Hermitian operator corresponding to energy, position, or momentum. There is a line in the complex plane $t + iv_0$ (v_0 a constant, $-\infty < t < \infty$), a neighborhood N of this line, and two functions $\Omega_z = \{\Omega_z(\lambda), \lambda \in \Lambda(L)\}$, $\Omega'_z = \{\Omega'_z(\lambda), \lambda \in \Lambda(L)\}$ for each complex number $z \in N$, such that:

(a) If $\Lambda(L)$ is discrete, Ω_z and Ω'_z are vectors in Hilbert space and

$$\Omega'_z(\lambda) = d\Omega_z(\lambda)/dz, \quad \text{for each} \quad \lambda \in \Lambda(L).$$

(b) If $\Lambda(L)$ is continuous, Ω_z and Ω'_z are functions in L_2-space and for some orthonormal set $\{\phi_n\}$ and complex coefficients $\{\alpha_n(z)\}$:

$$\Omega_z(\lambda) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{n=1}^{N} \alpha_n(z)\phi_n(\lambda),$$

$$\Omega'_z(\lambda) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{n=1}^{N} \frac{d\alpha_n(z)}{dz} \phi_n(\lambda).$$

(c) Moreover, for all λ in Case (a) and for almost all λ in Case (b):

$$[\Omega'_z(\lambda)]_{z=t+iv_0} = \omega_t(\lambda).$$

We now define for each operator L specified in Theorem 1, a space $S = S(L)$ and a function T defined on S in terms of some neighborhood N and some function Ω_z that satisfy Theorem 1. We write

$$S = \bigcup_{\lambda \in \Lambda} S_\lambda, \qquad S_\lambda = \{\Omega_z(\lambda), \ z \in N\},$$

$$T(\Omega) = z \quad \text{if and only if} \quad \Omega = \Omega_z(\lambda).$$

Theorem 2. Let the set of ordered pairs (I) and the operator L be as defined in Theorem 1. Let $\mathfrak{L}$

be a measurable function defined on the space $S = S(L)$, such that $\mathcal{L}(\Omega) = \lambda$ for $\Omega \in S_\lambda$. Let T be defined above. The distribution function $F_{L,t}$ and the conditional distribution function F of $\mathcal{L}$, given T, are the same:

$$F_{L,t}(\lambda) = F(\lambda \mid t).$$

The proof, found in Secs. 3, 4, and 5, follows from the fact that whether Λ is discrete and $p(\lambda \mid t)$ is the conditional probability function of $\mathcal{L}$, given T, or whether Λ is continuous and $p(\lambda \mid t)$ is the conditional probability density function of $\mathcal{L}$, given T:

$$p(\lambda \mid t) = |\Omega_z'(\lambda)|^2_{z=t+i\nu_0} = |\omega_t(\lambda)|^2,$$

for all λ, if Λ is discrete, and for almost all λ, if Λ is continuous.

This analytic characterization of $p(\lambda \mid t)$ in terms of the derivative of the inverse of T, derives from a more general expression for conditional probability or probability density in which T belongs to a general measurable set B in N

$$p(\lambda \mid T \in B) = \int_B |\Omega_z'(\lambda)|^2 \, dt \, dv,$$

for all λ, if Λ is discrete, and for almost all λ, if Λ is continuous. It thus becomes clear that Theorem 2 depends strongly on the topological representation of conditional probability considered as a neighborhood property of the conditioning function T. The theorem also depends strongly on the fact that for almost all λ in Λ the function Ω_z, that is the inverse of T, is analytic over a region, in the sense defined by its L_2-representation.

Though fairly extensive conclusions concerning the physical meaning of T will be developed in Sec. 6, the following facts may be stated here. The function T plays the role of time. It is studied simultaneously for its statistical (or global) properties and its topological (or local) properties. The term, "micro–macro" function, will be introduced for such functions and T is referred to as "micro–macro time," or "quantum mechanical time."

The functional structure of micro–macro time depends on the operator L and the solutions to the Schrödinger equation. From Theorem 2 it is seen that the probabilities of quantum theory, as they relate to an operator L, are deducible from the local properties (or fine structure, so to speak) of micro–macro time. In general, for each eigenvalue λ, $\Omega_z(\lambda)$ is periodic as a function of z, and thus, from the definition of T, we see that T is multiple-valued with multiple-valued structure depending on the subspace S_λ over which it takes its values.

For example, if L is the energy operator, the spacing between multiple values of T are equal to h/E_n, depending on the subset $S_{(E_n)}$ over which T takes its values, where h is Planck's constant. This property is in conformity with the uncertainty relation between energy and time, as will be clarified in Sec. 6. The multiple-valued structure of T is also shown to be in conformity with the correspondence principle of quantum theory.

We conclude that the functional structure of micro–macro time is consistent with the probabilities of quantum theory and, in fact, can be used as a logical tool for calculating and interpreting those probabilities. Consistent with this point of view, however, the concept of time as a micro–macro function and the classical concept of time as a real parameter merge in the quantum mechanical description of physical systems in the macro-world.

A remark concerning notation may be helpful. Though in stating Theorems 1 and 2 we have used the general notation $\Omega_z(\lambda)$ in both the continuous and discrete cases, it will be useful to vary the notation in the ensuing sections. In the context of measuring energy we will write

$$\Omega_z(\lambda) = \alpha_n(z), \qquad \lambda = E_n.$$

In the contexts of measuring position and momentum, respectively, we will write

$$\Omega_z(\lambda) = \Phi_z(q), \quad \lambda = q, \quad \text{and} \quad \Omega_z(\lambda) = \Psi_z(p), \quad \lambda = p.$$

It is also worth noticing that in the concise language of the theorems it has been convenient to express results in terms of conditional distribution functions. In the body of the paper, it will be better to concentrate on conditional expectation. The relationship between these two entities has been summarized above.

1. SUGGESTIONS FROM PROBABILITY THEORY

The following facts have been enigmatical to investigators in the theories of probability and quantum phenomena. First, the fact that the probabilities of probability theory are measures in a measure space while in quantum theory they are absolute squares of complex numbers. Second, the fact that the familiar structure of measure-theoretic probability theory is in default of any obvious quantization properties corresponding to the quantization of physical measurables. Third, the fact that the law of large numbers in probability theory, in contradistinction to the uncertainty principle of quantum theory, omits any reference to the limitation

of accuracy attainable in measuring two variables simultaneously.

Certain suggestions that one can obtain from probability theory indicate, however, that these apparent enigmas are favored by the different perspectives that are natural to the two theories, and that a new perspective taken in one of the theories or both would minimize the enigmas.

For example, if in probability theory we limit ourselves to a measure space composed of the unit interval and to random variables defined on this space that are continuous and differentiable, and if we express probability density functions and conditional probabilities in some unconventional forms that involve derivatives of the random variables, then there appear sums and sums of products involving these derivatives in a manner suggestive of inner products in Hilbert space. Hilbert-space theory is, of course, characteristic of quantum theory, though not generally of elementary probability theory. An analysis of elementary probability theory that would bring out its Hilbert-space characteristics is, thus, one interesting way to obtain suggestions of the desired type.

If $X = X(\alpha)$ is a random variable defined on the unit interval, $0 \leq \alpha \leq 1$, that is continuous and differentiable, and if the derivative of X is zero for at most countably many points α, then X has at most a countable number of inverses, each of which maps a subinterval of the range of X continuously onto a subinterval of the domain of X (see Fig. 1). We write these inverses as

$$\alpha_1 = \alpha_1(x), \quad \alpha_2 = \alpha_2(x), \cdots, \alpha_n = \alpha_n(x), \cdots. \tag{1}$$

A simple analysis shows that the probability density function of X, that is the function $p = p(x)$ that

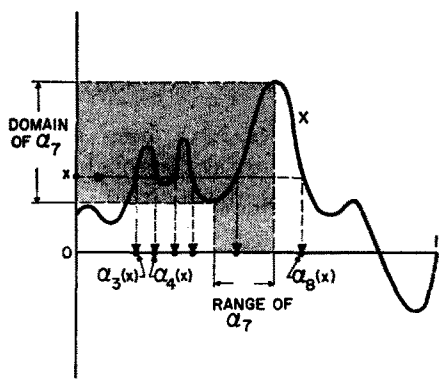


FIG. 1. Graph of a random variable X defined on the unit interval, showing how a point in the range of X is mapped into points between 0 and 1 through the inverse functions $\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots$. For convenience of drawing, inverse functions are given subscripts with values increasing as their range values increase from left to right.

satisfies

$$P[X < u] = \int_{-\infty}^{u} p(x)\,dx, \qquad -\infty < u < \infty \tag{2}$$

can be analyzed for almost all values of x in the range of X as

$$p(x) = |\alpha_1'(x)| + |\alpha_2'(x)| + \cdots + |\alpha_n'(x)| + \cdots, \tag{3}$$

where we define $\alpha_n'(x)$ to be $d\alpha(x)/dx$ if x is interior to the domain of definition of α_n, and 0 otherwise (see Fig. 2).

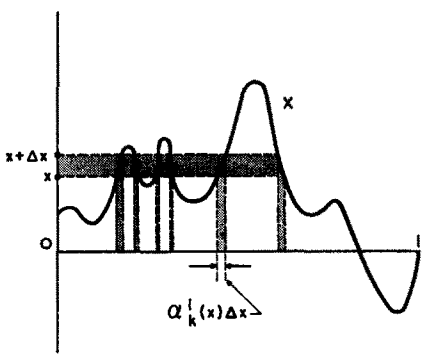


FIG. 2. Graph of a random variable X defined on the unit interval, showing the probability for $x < X < x + \Delta x$ to be approximated by a sum of terms of the type $|\alpha_k'(x)|\,\Delta x$. A point selected at random between 0 and 1 will determine a value of X between x and $x + \Delta x$, if it falls in one of the shaded intervals.

If the vector $\boldsymbol{\alpha}'(x)$ is defined as

$$\boldsymbol{\alpha}'(x) = (|\alpha_1'(x)|, |\alpha_2'(x)|, \cdots, |\alpha_n'(x)|, \cdots), \tag{4}$$

the density $p(x)$ can be expressed as an inner product of $\boldsymbol{\alpha}'(x)$ with the countable-dimensional unit vector $\mathbf{1}$:

$$p(x) = (\boldsymbol{\alpha}'(x), \mathbf{1}). \tag{5}$$

More generally, the joint density between X and an arbitrary random variable Y on the unit interval can be defined as an inner product. For this purpose, let $\mathbf{1}(x, y)$ be the vector with nth-component equal to one if $Y[\alpha_n(x)] = y$ and 0 otherwise. For each x, one has $\mathbf{1}(x, y) = 0$ except for at most a countable number of values of y. Let $y_k(x)$, $k = 1, 2, 3, \cdots$ be those values for which $\mathbf{1}(x, y_k(x)) \neq 0$. The joint density, that is the function $h(x, y)$ that satisfies

$$P[X < v, Y < u] = \int_{x<v} \sum_{y_k(x)<u} h(x, y_k(x))\,dx, \tag{6}$$

can be analyzed for almost all values of x interior to the range of X and for each y as

$$h(x, y) = (\boldsymbol{\alpha}'(x), \mathbf{1}(x, y)). \tag{7}$$

As a consequence of (5) and (7), the conditional

probability of the event $Y = y$ given $X = x$ is expressed as

$$g(y \mid x) = (\alpha'(x), \mathbf{1}(x, y))/(\alpha'(x), \mathbf{1}). \quad (8)$$

When the vector

$$\mathbf{y}(x) = (Y[\alpha_1(x)], Y[\alpha_2(x)], \cdots, Y[\alpha_n(x)], \cdots) \quad (9)$$

is introduced, expression (8) and the usual definition of conditional expectation of Y given $X = x$ yield

$$E[Y \mid x] = (\alpha'(x), \mathbf{y}(x))/(\alpha'(x), \mathbf{1}). \quad (10)$$

This analysis is entirely within the realm of elementary probability theory, but introduces the Hilbert-space geometry by virtue of the rather special character of the random variable X. Thus, it is suggested that a formalism similar to that of quantum theory can be obtained by putting stricter conditions on random variables than are expressible in terms of distribution functions, for example, conditions of continuity, differentiability, and the like. An exact replica of quantum analysis cannot arise without the introduction of a complex space in place of the unit interval and without some restrictions imposed upon the random variables in terms of a mechanics.

Even within the elementary probability framework, quantization characteristics appear. This is particularly evident in the countably many values $y_k(x)$ for which $\mathbf{1}(x, y_k(x)) \neq 0$, or, expressed differently, in the discrete character of the conditional distribution (8) (see Fig. 3).

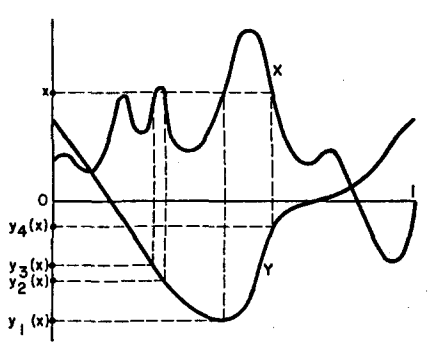


FIG. 3. Graph showing quantization of a continuously distributed random variable Y under the condition that for random variable X, $X = x$. In the diagram the only allowed values of Y in the presence of the condition $X = x$ are $y_1(x)$, $y_2(x)$, $y_3(x)$, $y_4(x)$.

2. MICRO-MACRO FUNCTIONS

In the introductory remarks it was suggested that random variables could be used to describe quantum effects, provided that more analytical detail were known about the random variables than was contained in their distribution functions or joint distribution functions. Specifically, it was seen that certain properties of functions on a topological space, such as their continuity and differential properties, become useful for our immediate purposes, though such properties are quite foreign to the main stream of probabilistic analysis and statistics. More generally speaking, we will find use for *local* properties of random variables (that is, those defined in terms of a topology on the space of definition), as well as *global* properties (that is, those statistical properties defined in terms of the joint distributions). On the other hand, in the present calculations, we wish to avoid the criticism, frequently stemming from the trend toward generalization in modern mathematics, that the concept of "random variable" is being misused or unnecessarily specialized by the introduction of local properties. Thus, we introduce a new term. We will call a function f defined on a space S a *micro–macro function*, if the space S possesses an associated class of measurable subsets, a measure, and a topology, and if the function is measurable with respect to the measurable subsets. It is understood in a qualitative sense that the term "micro–macro function" is used in the context of probabilistic calculations. It should be remarked that this general definition does not exclude spaces S that are complex (as will be apparent in the next section) and does not exclude micro–macro functions that are complex, unbounded, multiple valued, and with unnormalized distribution functions. Thus, the term "micro–macro function" calls upon more specific requirements on the space S but is in some ways more general than the term "random variable." The term "micro–macro function" can in many contexts be used as synonymous with "random variable."

3. THE COMPLEX SAMPLE SPACE

For the analysis of quantum theory we will need the following sample space (possibly unbounded):

$S : \{w : w = x + iy, (x, y) \in S^{(2)}$, where $S^{(2)}$ is a measurable subset of the Euclidean plane$\}$.

The measure m associated with subsets A of S will be defined in terms of the Lebesgue measure μ of two-dimensional real sets $A^{(2)}$ of the form

$$A^{(2)} : \{(x, y) : x + iy \in A\}.$$

That is, we define $m(A) = \mu(A^{(2)})$ for all measurable subsets $A^{(2)}$ of $S^{(2)}$.

A single-valued complex function $T = T(w)$ of the complex variable w is called *measurable* if for

each two-dimensional Borel set $A^{(2)}$ the following set is measurable: $\{(x, y) : T(w) \in A\}$. We will frequently be concerned with complex measurable functions T defined on S that satisfy the following properties.

T_1: For each complex value $z = u + iv$ in the range of T there are in S at most countably many solutions $w_1, w_2, \cdots, w_n, \cdots$ to the equation: $T(w) = z$.

T_2: For almost all complex values z_0 in the range of T and in some neighborhood of each solution w_n of the equation $T(w) = z_0$, the function $T = T(w)$ has a single-valued inverse $\alpha_n = \alpha_n(z)$. The functions α_n are continuous and possess continuous first partial derivatives.

T_3: For almost all complex z in the range of T, the series

$$\sum_n \left| \frac{\partial \alpha_n}{\partial u} \times \frac{\partial \alpha_n}{\partial v} \right|$$

is convergent. The sum is assumed to extend over all n for which $T(w_n) = z$.

Thus, in the sequel, we will assume that a measurable function designated with the symbol "T" has properties T_1, T_2, T_3. The meaning of the property T_3 will become clear when we can interpret the micro–macro function T as a random variable and can identify the sum appearing in assumption T_3 with the probability density of T.

For a micro–macro function $T = U + iV$ with real and imaginary parts $U = U(x, y)$, $V = V(x, y)$, we may define a measure m of the set $\{w : T(w) \in A\}$ in terms of the two-dimensional Lebesgue measure μ:

$$m\{w : T(w) \in A\} = \mu\{(x, y) : (U, V) \in A^{(2)}\}, \qquad (11)$$

and, in particular, if $A(u, v; \delta u, \delta v)$ is the region defined by the double inequality: $A(u, v; \delta u, \delta v)$: $(u \leq U < u + \delta u, v \leq V < v + \delta v)$, we have, for almost all complex $z = u + iv$,

$$m\{w : T(w) \in A(u, v; \delta u, \delta v)\} = \sum_n \left| \frac{\partial \alpha_n}{\partial u} \times \frac{\partial \alpha_n}{\partial v} \right| \delta u \, \delta v + o(\delta u \, \delta v), \qquad (12)$$

with convergence of the sum assured almost everywhere by property T_3. Because the sum in (12), which is a function of $z = u + iv$, behaves like a density, we define the *probability density* of T as follows:

$$p(z) = \sum_n \left| \frac{\partial \alpha_n}{\partial u} \times \frac{\partial \alpha_n}{\partial v} \right| \qquad (13)$$

for all z for which the sum has meaning and exists (see Fig. 4).

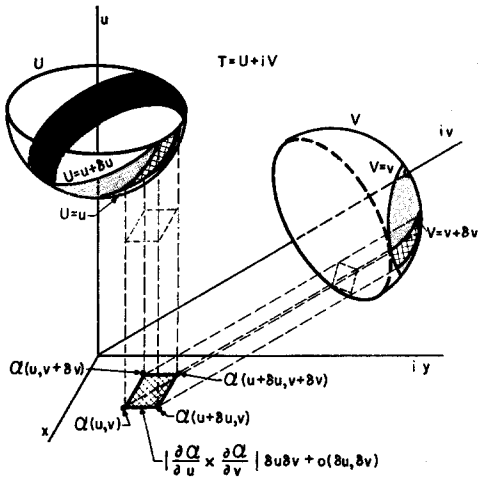


FIG. 4. Schematic graph showing that a summand in the probability density for $T = U + iV$ is $|\alpha'|^2$, where α is one inverse of T. A point chosen at random in a finite region of the x, y plane will determine the simultaneous conditions $u < U < u + \delta u$, $v < V < v + \delta v$, if it falls in the shaded rectangle. If T is analytic, $|\partial\alpha/\partial u \times \partial\alpha/\partial v| = |\alpha'|^2$.

If the micro–macro function T is analytic, we deduce from known theorems that, for almost all z, each inverse $\alpha_n(z)$ is analytic. The Cauchy–Riemann equations hold for each $\alpha_n(z) = X_n(u, v) + iY_n(u, v)$:

$$\frac{\partial X_n}{\partial u} = \frac{\partial Y_n}{\partial v}, \qquad \frac{\partial X_n}{\partial v} = -\frac{\partial Y_n}{\partial u}. \qquad (14)$$

From the definition of the vector product we have

$$\frac{\partial \alpha_n}{\partial u} \times \frac{\partial \alpha_n}{\partial v} = \mathbf{n}\left(\frac{\partial X_n}{\partial v} \frac{\partial Y_n}{\partial u} - \frac{\partial X_n}{\partial u} \frac{\partial Y_n}{\partial v} \right), \qquad (15)$$

where $\mathbf{n}$ is the normal vector. Thus

$$\left| \frac{\partial \alpha_n}{\partial u} \times \frac{\partial \alpha_n}{\partial v} \right| = \left| \left(\frac{\partial X_n}{\partial v} \right)^2 + \left(\frac{\partial Y_n}{\partial v} \right)^2 \right| = \left| \frac{\partial \alpha_n}{\partial v} \right|^2. \qquad (16)$$

Again, from the analytic property of α_n, we conclude

$$p(z) = \sum_n |\alpha_n'|^2, \qquad \alpha_n' = \frac{d\alpha(z)}{dz}, \qquad (17)$$

where the sum is assured to converge for almost all z by property T_3. Equation (17) is the complex analog of (3) with squares of absolute values arising in a natural way from the complex character of the sample space. Due to the desirable form of (17), *we also assume throughout the remainder of this paper the following:*

T_4: The function T is analytic over S.

4. THE TRAJECTORY SAMPLE SPACE

For greater convenience, we will make a reformulation expressed as a further assumption having to do with the inverses of the micro–macro function T (assumptions are not independent).

T_5: The complex sample space S can be partitioned into at most a countable number of parts $S_1, S_2, \cdots$, with $\cup S_n = S$, such that for each positive integer n, T assumes no value twice in the interior of S_n.

Assumption T_5 insures that the function T restricted to the interior of S_n, $n=1, 2, \cdots, n, \cdots$ is schlicht. We may identify α_n with the inverse of T restricted to the interior of S_n. *In the sequel, we will assume that this identification is made.* We will write $\alpha_n(z)=0$ if z is not in the domain of definition of α_n, that is, if z is not in the range of T restricted to S_n.

If, as in Sec. 1, we introduce the vector $\alpha'(z)$ with components $\alpha_n'(z)$, and for any micro–macro function L defined over S we introduce the vector $\lambda = \lambda(z)$ with components $L(\alpha_n) = L[\alpha_n(z)]$, the conditional expectation of L given T, provided it exists, can be analyzed:

$$E[L \mid T(w)] = (\lambda \otimes \alpha', \alpha')/||\alpha'||^2, \quad T(w) = z, \tag{18}$$

where

$$\lambda \otimes \alpha' = (L(\alpha_1)\cdot\alpha_1', L(\alpha_2)\cdot\alpha_2', \cdots, L(\alpha_n)\cdot\alpha_n', \cdots).$$

Moreover, if the micro–macro function L is measurable with respect to the smallest σ-field containing the sets $S_1, S_2, \cdots, S_n, \cdots$, it is clear that L assumes constant values λ_n over the interior of corresponding sets S_n and $\lambda(z)$ is independent of z [see Fig. 5(a)].

Thus, for micro–macro function L measurable with respect to the smallest σ-field that contains the sets $S_1, S_2, \cdots, S_n, \cdots$, we can write

$$E[L \mid T(w)] = \sum_n \frac{\lambda_n |\alpha_n'(z)|^2}{||\alpha'(z)||^2}, \qquad f(w) = z, \tag{19}$$

provided L is real and the series converges. It is significant to remark on the similarity between the expression (19) and the familiar expression in quantum theory for the expectation of a physical measurable whose associated operator L possesses a discrete spectrum. In the latter case, the complex numbers α_n' are the coefficients in an expansion of the wave-function ϕ in terms of an orthonormal set of functions ϕ_n in L_2-space, and the numbers λ_n are eigen-

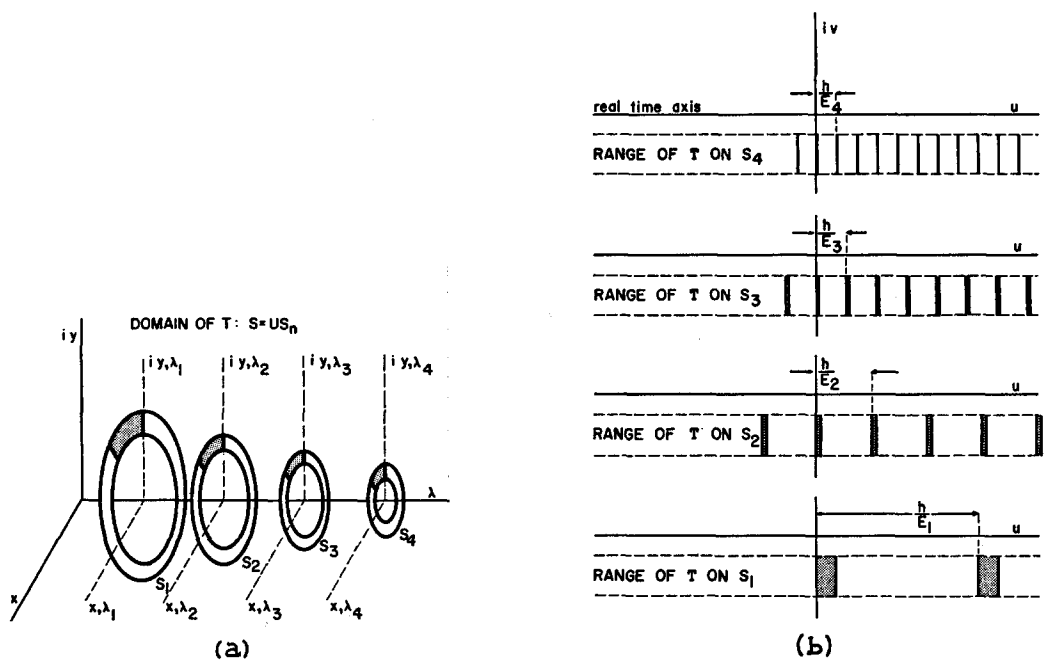


FIG. 5(a). A graph suggesting one way to associate discrete values λ_n of a micro–macro function L with subregions S_n of the complex plane. The logical union $S = \cup S_n$ of the annular subregions S_n is consistent with the domain of T when measurement is made on the energy of the harmonic oscillator. For comparison with later generalization (Fig. 6), the subregions S_n are plotted so as to correspond to values λ_n on the λ-axis. Shaded areas are interpreted in Fig. 5(b). See Section 5. (b) The range of the micro–macro function T as defined over the annular regions S_n of Fig. 5(a). Multiple shaded regions in the range of T are mapped by T from the shaded areas of the corresponding regions S_n. A point chosen at random in S of Fig. 5(a) will be in S_n with probability proportional to the area of S_n. Given that the point is in S_n, T will have a value in each of the multiple shaded regions of its range over S_n with probability proportional to the shaded area in S_n. The decrease in the spacing between multiple values of T as n increases corresponds roughly to the behavior of quantum mechanical time for the harmonic oscillator as the energy level increases.

values of the operator L corresponding to the eigenfunctions ϕ_n : $L\phi_n = \lambda_n\phi_n$. That is, we compare (19) with the quantum mechanical statement, $EL = (L\phi, \phi)/||\phi||^2$, in which L is an operator.

At least two significant aspects of this analysis, as it relates to quantum theory, require further clarification: (1) What extensions to the geometrical reasoning offered up to this point will accommodate the statistical properties of operators with continuous spectra? (2) What interpretive role can be given to the functions T, L, etc.? In the remainder of this section, we attend to question (1).

The following correspondences are significant:

I: There is a one-to-one relationship between points z in the range of T and vectors $\alpha(z) = (\alpha_1(z), \alpha_2(z), \cdots, \alpha_n(z), \cdots)$ composed of components such that for each z, $\alpha_n(z) \in S_n$. The vectors $\alpha(z)$ may or may not belong to Hilbert space.

II: For each point z in the range of T, there is a vector $\alpha'(z)$ in Hilbert space.

It is natural, for the sake of calculational simplicity, *to make the following further assumption on* T;

T_6: For the function T and for almost all z in the range of T, the vector $\alpha(z)$ belongs to Hilbert space. (We allow that the dimensionality of Hilbert space may be finite or countably infinite.)

Correspondences I and II and Assumption T_6, together with relations (17), provide us with the consequence that there exists a point-to-point mapping from the range of T to a subset of Hilbert space which preserves measure in the sense that

$$z \leftrightarrow \alpha(z); \qquad p(z) = ||\alpha'(z)||^2. \tag{20}$$

In order to enrich the geometry sufficiently to accommodate the statistical analysis of operators in Hilbert space with continuous spectra, we must supplement the correspondence (20) with a correspondence between points $\alpha_n = \alpha_n(z)$ in S_n and vectors in L_2-space. In this way we are led to associate each subset S_n with a member $\phi_n = \phi_n(q)$, $-\infty < q < \infty$, of an orthonormal set in L_2-space and to associate the root w_n in S_n of the equation $T(w) = z$ with $\alpha_n(z)\cdot\phi_n$. Writing

$$\Phi_z(q) = \sum_n \alpha_n(z)\cdot\phi_n(q), \quad \Phi'_z(q) = \sum_n \alpha'_n(z)\cdot\phi_n(q), \tag{21}$$

where we understand the sums defined in terms of limits in the mean, we supplement (20) with:

$$\alpha(z) \leftrightarrow \Phi_z; \qquad ||\alpha'(z)||^2 = \int_{-\infty}^{\infty} |\Phi'_z(q)|^2\, dq. \tag{22}$$

For the sake of enriching the geometry, we may now interpret the Hilbert-space point $\alpha(z)$ as a set of ordered pairs, $(\Phi_z(q), q)$, or a trajectory $\{\Phi_z(q), -\infty < q < \infty\}$. Let us consider the space Σ_q of the union of all such trajectories for z in the range of T. The trajectory space Σ_q can be written

$$\Sigma_q = \{(\bar{w}, q) : \bar{w} = \Phi_z(q), -\infty < q < \infty, \quad z \text{ in the range of } T\}.$$

If each pair (z, q) defines a point $(\bar{w}, q)$ on at most one trajectory; that is, if trajectories do not cross, then we may also establish the one-to-one correspondences

$$(z, q) \leftrightarrow (\alpha(z), q) \leftrightarrow (\bar{w}, q). \tag{23}$$

A function $h = h(w)$ on S uniquely defines a function $h = h(w, q)$ on $S \times R$ through the definition: $h(w, q) = h(w)$. In particular (see Fig. 6), if $h = T$, we can further define a function on Σ_q through

$$T(\bar{w}, q) = T(w, q) = T(w), \quad \text{if and only if} \quad \bar{w} = \Phi_{T(w)}(q). \tag{24}$$

In this statement, we take advantage of the many–

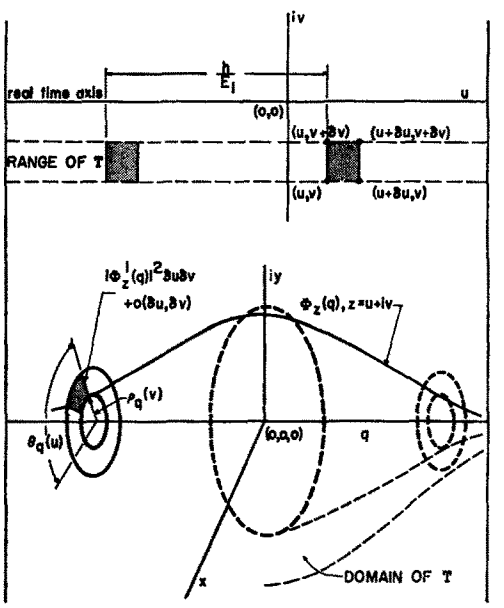


FIG. 6. Illustration of the range and domain (trajectory space) of T when measurement is made on continuous spectra. A point in the shaded region of the domain is mapped through T into the multiple shaded regions of the range. A point in any shaded portion of the range is mapped through an inverse $\Phi_z(q)$ of T into the shaded portion of the domain. Details in the graph having to do with the spacing h/E_1 between multiple values of T, the polar coordinates $\theta_q(u)$, $\rho_q(v)$ and the shape of the trajectory Φ_z are consistent with measurement of position of the harmonic oscillator in its ground state. For these details see Sec. 5 and 7.

one correspondence $(w, q) \rightarrow (\bar{w}, q)$ determined by the equality $\bar{w} = \Phi_{T(w)}(q)$ and the fact that T is constant over the set of pairs (w, q) that map into a fixed point $(\bar{w}, q)$. For certain purposes it is convenient to consider T defined on Σ_q and for others to consider T defined on S or $S \times R$.

We may consider more general functions $Q = Q(\bar{w}, q)$ defined on Σ_q. By assuming that Q is a micro–macro function and by taking advantage of the topology on Σ_q, we can inquire into the existence of the probability density of Q. It occurs, however, that joint probability densities and conditional probability densities are especially important for our study of the probabilistic structure of quantum mechanics.

We proceed by considering the important case of micro–macro functions on Σ_q of the form

$$Q(\bar{w}, q) = Q(q). \tag{25}$$

Such functions can be transferred to $S \times R$ space through the definition $Q(w, q) = Q(\bar{w}, q)$ for every pair (w, q) that maps into $(\bar{w}, q)$. When such a function $Q(w, q)$ is considered as a function of its first argument w, it is seen to be measurable with respect to the smallest σ-field with respect to which T is measurable.

Considering, as a special case of (25) the function $Q(\bar{w}, q) = q$, we seek the joint density between Q and T. For fixed q, $\Phi_z(q)$ is the inverse of $T(\bar{w}, q)$ in the sense that $T[\Phi_z(q), q] = z$. From expansions (21), $\Phi_z(q)$ behaves like an analytic function of z. Thus, by the techniques of the previous section, the conditional probability density of T given Q is seen to be

$$p(z \mid q) = |\Phi_z'(q)|^2 \Big/ \int |\Phi_z'(q)|^2 \, dz, \tag{26}$$

assuming that the normalization integral exists. If the normalization integral exists, we identify it with the probability density of Q. Thus, we write the joint density of T and Q, without qualification as

$$p(z, q) = |\Phi_z'(q)|^2. \tag{27}$$

From this it follows immediately that, for any micro–macro function Q satisfying (25), we have its conditional expectation given T as follows:

$$E[Q \mid T(w)] = \int_{-\infty}^{\infty} \frac{Q(\bar{w}, q) \, |\Phi_z'(q)|^2 \, dq}{||\alpha'(z)||^2}. \tag{28}$$

This integral has, in fact, the familiar form of the expectation of a physical measurable whose corresponding operator has a continuous spectrum.

5. THE RESTRICTION IMPOSED BY THE SCHRÖDINGER EQUATION

While it is clear from the above constructions that a micro–macro function T satisfying assumptions $T_1 - T_6$ determines the function $\Phi_z'(q)$, up to its L_2-representation, it is not necessary that a given L_2-function $\Phi_z'(q)$ admits the above construction in terms of a micro–macro function T that satisfies assumptions $T_1 - T_6$. We will now indicate how a slight broadening of the concept of T makes possible the desired construction of a given $\Phi_z'(q)$.

The reason for concern over this point derives directly from the obvious fact that $\Phi_z'(q)$ appears to play the role of the quantum mechanical wave-function in equations such as (28), and, as such, should be restricted by the Schrödinger equation if a physically meaningful result is to be derived. That is to say, before a verification of the consistency of the above probabilistic construction with the apparatus of modern quantum theory can be borne out, it must be verified that $\Phi_z'(q)$ can be made to satisfy the Schrödinger equation, at least in some sense. Such a verification must entail relating the range of the complex function T with the real axis of time and relating the micro–macro function T, itself, with the usual independent time parameter t.

As an example, let $\phi_t(q)$ satisfy the time-dependent Schrödinger equation

$$H\phi_t(q) = i\hbar \partial \phi_t(q)/\partial t, \tag{29}$$

in which H is the Hamiltonian operator for a bound-particle system with one degree of freedom and with time-independent potential energy. It is well known that the most general solution of (29) is

$$\phi_t(q) = \sum_n a_n e^{-iE_n t/\hbar} \phi_n(q), \tag{30}$$

where the real constants E_n are the eigenvalues and the complex functions $\phi_n(q)$ are the eigenfunctions for the operator H, assuming no degeneracy:

$$H\phi_n(q) = E_n \phi_n(q). \tag{31}$$

Let a function $\Phi_z(q)$ of the complex variable $z = u + iv$ and the real variable q be defined such that for some pure imaginary number iv_0:

$$\Phi_{t+iv_0}'(q) = \phi_t(q). \tag{32}$$

A comparison of Eqs. (21) and (30) suggest an identification of the coefficients $a_n e^{-E_n s i/\hbar}$ and $\alpha_n'(z)$ over a region of the complex plane that includes the line $\{t + iv_0, -\infty < t < \infty\}$. If one attempts such a comparison, however, while retaining $\alpha_n'(z)$ as the derivative of an inverse of a micro–macro

function T, certain difficulties are encountered having to do with the periodic behavior of the exponential function. Thus, the following assumption of T naturally suggests itself as a generalization of the requirement that T be single-valued and the assumption will hold throughout the remainder of our work.

T_7: The function T is single- or multiple-valued in such a way that each of the inverses $\alpha_n(z)$ of the restriction of T to the corresponding set S_n is periodic in u (with finite or infinite period).

The wording of the assumptions T_1–T_6 and the remarks following definition (25) admit obvious interpretation which makes them consistent with assumption T_7. The term "inverse" will be retained for the $\alpha_n(z)$ by virtue of the option to identify $\alpha_n(z)$ over a given period with the inverse of a corresponding branch of T. Mappings I, (20), (23) become many–one.

Thus, with periodicity allowed in the inverses of T, we can, in fact, make the identification

$$\alpha_n'(z) = a_n e^{-iE_n z/\hbar}, \qquad n = 1, 2, \cdots, \tag{33}$$

over certain regions of the complex plane, from which we obtain as formal general solutions to the inverses of T

$$\alpha_n(z) = (ia_n\hbar/E_n)e^{-iE_n z/\hbar} + k_n, \quad n = 1, 2, \cdots, \tag{34}$$

where k_n is a function independent of z. These solutions are analytic, as desired by assumption T_4, over any region of definition in the complex plane.

We must next attend to the requirement that each function $\alpha_n(z)$ can be identified with the inverse of a function T restricted to a set S_n and that the union of such sets S_n constitute the domain of T. At the same time, all the assumptions T_1–T_7 must be satisfied. For this purpose, we choose to set $k_n = 0$, $n = 1, 2, \cdots$, and write

$$\begin{gathered}\alpha_n(z) = \rho_n(v)e^{i\theta_n(u)}; \quad \rho_n(v) = (|a_n\hbar|/E_n)e^{E_n v/\hbar} \\ \theta_n(u) = -[(E_n/\hbar)u - \arg(ia_n)]\end{gathered} \tag{35}$$

defined for all $z = u + iv$, $-\infty < u < \infty$, $v_{1,n} < v < v_{2,n}$, the bounds for v to be defined. It is noticed that $\alpha_n(z)$ is periodic in u with period $(h/E_n)j$, $h = 2\pi\hbar$, $j = 1, 2, \cdots$. We also defined for each positive integer n, the annular region

$$S_n = \{\alpha_n(z) : -\infty < u < \infty, v_{1,n} < v < v_{2,n} < 0\}, \tag{36}$$

where $v_{1,n}$ and $v_{2,n}$ are so chosen for each positive integer n so that the annular regions S_n are disjoint[1] and assumption T_6 is satisfied [see Fig. 5(a)]. In regard to assumption T_6, it is sufficient to select the $v_{2,n}$ such that

$$||\alpha(z)||^2 = \sum_n |\alpha_n(z)|^2 < \sum_n \left|\frac{a_n\hbar}{E_n}\right|^2 e^{2E_n v_{2,n}/\hbar} < \infty, \tag{37}$$

which is always possible. In particular, if the eigenvalues E_n constitute a strictly increasing sequence as a function of n that diverges sufficiently rapidly, then all the annular regions S_n can be defined by bounds $v_{1,n}$ and $v_{2,n}$ that contain a fixed point v_0: $v_{1,n} < v_0 < v_{2,n}$. We will tacitly assume in the sequel that this is the case, though it is not essential to do so. The most common examples of quantum mechanical systems—for example, the harmonic oscillator for which $E_n = (n - \frac{1}{2})\hbar\omega$—suggest this simplifying but unessential assumption.

With the inverses (35) and the subspaces (36) defined, we are free to constitute the desired micro–macro function T as that function whose domain of definition is $S = \cup S_n$ and whose inverse over S_n is $\alpha_n(z)$, $n = 1, 2, \cdots$. Thus, we are led to the function

$$T(w) = \frac{i\hbar}{E_n}\log\left[\frac{E_n w}{i\hbar a_n}\right]; \quad w \in S_n, \quad n = 1, 2, \cdots, \tag{38}$$

where the logarithmic function is the usual complex multiple-valued function. For w in S_n, the range of $T(w)$ is the set of all complex $z = u + iv$, with $-\infty < u < \infty$, $v_{1,n} < v < v_{2,n}$. Assuming that there is a fixed point v_0 such that $v_{1,n} < v_0 < v_{2,n}$, then $u + iv_0$ is common to the range of $T(w)$ for w in S_n, $n = 1, 2, \cdots$. The multiple-valued property of T is displayed by writing, for $z = u + iv_0$, in the range of T

$$\begin{gathered}T[\alpha_n(z)] = z + \frac{h}{E_n}j = \left(u + \frac{h}{E_n}j\right) + iv_0; \\ j = 0, \pm 1, \pm 2, \cdots, h = 2\pi\hbar.\end{gathered} \tag{39}$$

[see Fig. 5(b)].

The accomplishment of the construction for T in the reasonably general case of a bound-particle system with time-independent potential energy, completes the desired demonstration that the function $\Phi_z'(q)$ of definition (21) can, indeed, be interpreted

[1] The disjointness condition is essential here only because we have restricted the domain of T to be a subset of the complex plane. As in the analysis for the continuous spectrum, we could extend the domain of T to be three-dimensional (one dimension corresponding to eigenvalues of the energy). With such extended geometry, the disjointness condition can be dropped. The geometric extension is suggested in Fig. 5(a).

as a quantum mechanical wavefunction, at least in certain cases. We have at the same time completed the proofs of Theorems 1 and 2 as stated in the introduction.

6. PHYSICAL INTERPRETATION OF MICRO-MACRO TIME

The function T, though it is a micro–macro function and has been used in such probabilistic calculations as conditional expectation, is best described as a quantum mechanical representation of time. It is important to observe the close relationship between the structure of micro–macro time and the wavefunction of the physical system under observation. It is also important to observe that T, like the functions L and Q which play the role of quantum mechanical operators, is defined on a space of points. Micro–macro time T is, thus, more like a quantum mechanical operator than an independent real parameter. Its relation to the familiar independent time parameter t is given by (39) with $u = t$. The value of micro–macro time T is determined, up to a multiplicity, by the random selection of a point w in S and, consequently, T behaves like a random variable. It is also true that the function T is complex. Certain clarifications of these diverse characteristics of T are necessary if the time interpretation of T is to be justified.

One is reminded at this point of remarks made by von Neumann[2] in regard to the way that time enters in ordinary quantum mechanics:

"Then how can our assumption of instantaneous measurements be justified? First of all we must admit that this objection points at an essential weakness which is, in fact, the chief weakness of quantum mechanics: its non-relativistic character, which distinguishes the time t from the three space coordinates x, y, z, and presupposes an objective simultaneity concept. In fact, while all other quantities (expecially those x, y, z closely connected with t by the Lorentz transformation) are represented by operators, there corresponds to the time an ordinary number-parameter t, just as in classical mechanics. . . . It may be connected with this non-relativistic character of quantum mechanics that we can ignore the natural law of minimum duration of the measurements. This might be a clarification, but not a happy one!"

It is not out of keeping with von Neumann's remarks to expect that a function such as T, rather than a parameter, could, indeed, represent time. As is well known, von Neumann's question concerning instantaneous measurements is in part answered by relativistic quantum theory, in which time is more in the nature of an operator than a parameter. We conclude that the micro–macro representation of time is in some ways closer to the spirit of relativitic quantum theory than to nonrelativistic. It is important, in this regard, to emphasize, however, that physically realistic restrictions on micro–macro time have been imposed by the Schrödinger equation and not by the relativistic Klein–Gordon equation.

Because we have found that T can be analytic and at the same time serve in the construction of the wavefunction according to the analysis (21), the derivatives $\alpha_n'(z)$ appearing in the expansion of the wavefunction can be taken along any path through the point z in the range of T. This, together with the fact that there may (as in the case of the harmonic oscillator) be a fixed line $u + iv_0$, $-\infty < u < \infty$, common to the range of $T(w)$ for $w \in S_n$, $n = 1, 2, \cdots$, suggest identifying the line $u + iv_0$, $-\infty < u < \infty$, with the real time axis and justify computing derivatives of the type $\alpha_n'(z)$ for points $z = u + iv_0$ as derivatives with respect to the real time parameter. Thus, while a geometric flavor has been introduced into the interpretation of the wavefunction and its expansion coefficients by virtue of the analyticity of T, the familiar properties of the wavefunction, in particular reference to its usual dependence on time, have been preserved.

We are permitted, henceforth, to explore the structure of T for physical interpretation, having been assured that its properties as a micro–macro function do not conflict with our preconceptions of time as the independent parameter on which the wavefunction depends.

Observe, first, that the sets S_n, $n = 1, 2, \cdots$, are associated with distinct eigenstates and that, without preconditioning the value of micro–macro time, the same contingency that determines the eigenstate of the system determines the value of micro–macro time, up to a multiplicity. Such a contingency is representable in our construction by the selection of a point in S. In fact, our geometry which associates area in S with probability, permits the language of random selection of a point in S. We are, thus, led to the physical interpretation that an observation on a physical system is to be identified with a point selected at random in S. (The case in which the space S is extended to the trajectory space Σ_q is discussed in the next section.) In the

[2] J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955).

absence of knowledge about the physical system, we impose no restriction on the subset in S from which a point is to be chosen. (We are phrasing in the language of measure-theoretic probability the usual postulate of quantum theory to the effect that, prior to measurement, the wavefunction is a superposition of eigenstates. In our case, the separate eigenstates correspond to subsets $S_1, S_2, \cdots, S_n, \cdots$ whose logical union comprises S.) Once a particular point is chosen at random from S (that is, an observation is made), one of the subsets $S_1, S_2, \cdots, S_n, \cdots$ is occupied, namely, that one from which the point was chosen. (We express in this way the familiar postulate of quantum theory to the effect that a measurement disturbs the system by causing the wavefunction to change to an eigenstate in an unpredictable way.)

We emphasize here that an observation on a physical system is not assumed necessarily to be conditioned on time having a preassigned instantaneous value. An unconditioned observation on a physical system is interpreted, in the language of micro–macro functions, as determining a value (possibly unmeasurable with precision) for time, up to a multiplicity, as well as values for physical measurables such as energy. That is, we understand that an observation is made on a space–time system. It is in this sense that the relativistic character of the present theory becomes apparent.

It may next be remarked from analysis (39) that micro–macro time is multiple valued with spacing h/E_n between multiple values, depending upon the energy state of the physical system, where h is Planck's constant. This spacing is significant for small energy levels that are characteristic of the micro-world and infinitesimal for high energy levels of the macro-world. Thus, the characteristic that T assumes random values, determinable, up to a multiplicity, by the selection of a point in S, becomes insignificant for observations in the macro-world. In this sense, micro–macro time satisfies the correspondence principle, its random character disappearing in the large [see Fig. 5(b)].

Let us clarify the meaning of the multiple-valuedness and the spacings h/E_n of micro–macro time. These spacings h/E_n must represent physically the degree to which time, as it relates to an observation on the system, is completely random. If we identify an observation on the physical system with a randomly selected point in S, and characterize the result of this observation with an exact energy E_n, then h/E_n must pertain to the uncertainty with which the time of the observation can be known. We find a physical explanation for the spacings from the uncertainty relation between time and energy measurements. We refer to the restriction between the uncertainty in parametric time Δt and the uncertainty in energy ΔE:

$$\Delta t \, \Delta E \gtrsim h, \tag{40}$$

where h is Planck's constant. This uncertainty relation can be inferred from the postulates of ordinary quantum theory (and thus from the present model also), though, unlike other uncertainty relations, it is known to have a rigorous derivation only in relativistic quantum mechanics. From (40) we read that the interval of time over which a measurement extends must be greater than h/E if any knowledge whatsoever is to be obtained about the energy E—that is, if it is not to be the case that $\Delta E > E$. In another sense, we can conclude from (40) that, if a physical system is in a pure energy state E, then time within an interval smaller than h/E is completely random with respect to our measuring devices, provided that the same measuring devices are to be at all sensitive to the energy of the system.

Having given the spacings between multiple values of micro–macro time some physical meaning based on known quantum analysis, we give the following heuristic interpretation: The real component of micro–macro time, as a function of randomly chosen points in S, has the structure of a ladder with rungs placed at the multiple values of T and with spacing h/E_n between rungs depending on the energy of the system under observation. The ladder is translated with reference to each observation, so as to make the exact position of a rung completely random within an interval of length h/E_n. To speak less exactly, the ladder is vibrating erratically relative to the space within which our observations are made, so as to make no position of a given rung within an interval of length h/E_n a reference point for a quantum mechanical measurement. For energies large in comparison to Planck's constant h, the rungs of the ladder are extremely close together and permit approximation by a real continuum, the continuum of the familiar time domain.

While the random nature of micro–macro time helps us avoid the dilemma of instantaneous measurement, the unmeasurable fineness of the spacings between multiple values may signal to the logical positivist a meaningless structure. On the contrary, it is from the detailed structure of micro–macro time that we compute the probabilities and expectations of quantum phenomena. Prior to any measurement on a physical system, we may meaningfully speak

of conditioning observation on an exact value of T, because, prior to any measurement, the origin of time is arbitrary. By imposing the restriction $T = z$, $z = t + iv_0$ on future observations and conceiving of probability density of T as a differential concept involving the derivatives of the inverses of T, we deduce the probability for a future measurement to yield an eigenstate. We see from (19) that the chance that a point arbitrarily chosen from S should be chosen from S_n, if $T = z$, is $|\alpha_n'|^2$. We see this by setting L in (19) to be the characteristic function of the set S_n. (In the language of ordinary quantum theory, an observation at time t disturbs the system by causing the wavefunction to change in an unpredictable way to an eigenstate ϕ_n according to the probability $|a_n|^2$, where a_n is the coefficient in the expansion of the undisturbed wavefunction.) It is important to observe from the standpoint of micro–macro time, that the probabilities $|\alpha_n'|^2$ derive from a differential concept that takes explicit advantage of the local properties of T in the neighborhood of a solution w_n to $T(w) = z$. The probabilities derive from minute variations in T that may in a physical sense be unobservable except through the physical measurement of probabilities, themselves.

More generally, the conditional expectation (19), which is identified with quantum mechanical expectation, relates directly to the detailed structure of micro–macro time T. Intuitively, one conceives of the conditioning of micro–macro time to a fixed value as a means of isolating a differentially small subset in S, namely, the subset over which micro–macro time assumes values in an infinitely small neighborhood of the fixed value. Within the restrictions imposed by this subset in S, one then observes an average of values assumed by a quantum mechanical measurable, such as the micro–macro function L of (19). The reasoning for quantum mechanical expectation is seen to be identical to that for conditional expectation in the measure-theoretic probabilistic sense, provided only that time is viewed as a micro–macro function on a space S.

7. OBSERVATIONS IN TRAJECTORY SPACE AND SIMULTANEOUS MEASUREMENTS

In the last section, we confined our attention to observations on a physical system that could be identified with the random selection of a point in S. Because S is partitioned into subsets $S_1, S_2, \cdots, S_n, \cdots$ that we associate (through the physical restrictions of the Schrödinger equation) with energy states, S is a natural space over which to define micro–macro functions L that are measurable with respect to the smallest σ-field including $S_1, S_2, \cdots, S_n, \cdots$. The same space is, thus, the natural space with which to associate operators that commute with the energy operator. It is, of course, also a space over which we may define micro–macro time. We saw that the deduction of conditional expectations of functions like L in Eq. (19), though they impose a condition on micro–macro time, do not entail a meaningless condition on T. Where probabilistic computations are specifically in reference to energy eigenstates, we interpret that an *a posteriori* observation results in a state that is stationary in time and, thus, prior to observation, the origin of time must be arbitrary.

In contrast to observations relative to S, observations on a physical system that are expressible in terms of the selection of points in trajectory space Σ_q are strikingly different. In order to clarify this fact, we must first examine micro–macro time as defined on Σ_q [see Fig. (6)].

The interpretation of definition (24), in the light of assumption T_7, reveals that the structure of micro–macro time is very different as defined on Σ_q and as defined on S. When T is defined on S we discover the multiple values with spacing h/E_n. When T is defined on Σ_q we discover the possibility of a variety of behaviors depending upon the number of eigenstates. In general, T, as defined on Σ_q, is multiple valued and the spacing between multiple values is determined, in physical language, by the superposition of eigenfunctions in the composition of the wavefunction. The same multiple-valuedness of T as defined on Σ_q becomes reflected in the periodicity in z of such functions as $p(z, q)$ of (27). Here we see in formal language the electrical oscillations that are characteristic of mixed states.

It is important to emphasize once again that the expectations conditioned on time are a differential concept having to do with the detailed structure of micro–macro time in a *neighborhood* of a specified point. This is true, whether micro–macro time is defined on S or on Σ_q. We discover in this way that it is not meaningless to restrict selection of points of Σ_q by imposing a condition on micro–macro time as defined on Σ_q. Questions having to do with instantaneous measurement do not enter. Consequently, while the reasoning of conditional expectation, given T, remains the same as in the last section, application of that reasoning to formulas such as (28) result in truly time-dependent functions for expectations and distributions, as is characteristic of mixed eigenstates.

In quantum theory, conjugate observables are generally expressed in terms of noncommuting operators. In the notation of micro–macro functions, we consider a transformation on L_2-space that supplements the correspondence (22) with

$$\Phi_z \leftrightarrow \Psi_z \tag{40}$$

according to

$$\frac{1}{(2\pi\hbar)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \Phi_z(q)\, e^{-ipq/\hbar}\, dq = \Psi_z(p). \tag{41}$$

This Fourier transform takes the eigenfunctions $\phi_n(q)$ of the orthonormal set $\{\phi_n(q)\}$ into $\psi_n(p)$ of a new orthonormal set $\{\psi_n(p)\}$ and gives us a new trajectory space $\Sigma_p = \{(\bar{w}, p) : \bar{w} = \Psi_z(p), -\infty < p < \infty,$ z in the range of $T\}$. On Σ_p we define a new class of micro–macro functions P in exact analogy to the way we define Q on Σ_q through definition (25). The functions Q and P play the role of conjugate observables. These functions can represent the position operator Q and the momentum operator P of ordinary quantum mechanics. Of course, the familiar uncertainty relations could be derived for them. We omit the details here which are the same as in ordinary quantum mechanics.

Transformation (41) notably fails to give us a point-to-point correspondence between the spaces Σ_q and Σ_p and, instead, gives a correspondence between trajectories. It is noticed that T can be defined on Σ_q, through definition (24), or on Σ_p, through a similar definition. That is, a condition on T characterizes an ensemble of trajectories in Σ_q and an ensemble in Σ_p. We identify simultaneous measurements on position and momentum of a particle with the two selections, made at random, of a point in Σ_q and a point in Σ_p. We condition these selections (as described in the last section) upon the corresponding differentially small subspaces in Σ_q and Σ_p over which micro–macro time assumes values in an infinitely small neighborhood of a fixed time value. The finer questions of the meaning of simultaneity, like the questions relating to instantaneous measurement, must be referred once again to the detailed structure of micro–macro time, though they must inevitably express themselves in terms of the known uncertainty relations between conjugate observables.

Dynamically speaking, we can identify successive L_2-space points Φ_{t+iv_0} or Ψ_{t+iv_0}, considered as functions of the real time parameter t, in much the same way that we speak of the dynamics of the wavefunction. The chief distinction in concept, as it derives from the present thinking, is concentrated in the relation between the points in L_2-space and the probability densities. Through the concepts of micro–macro functions, probability densities are derived from the local behavior of trajectories as they depend on time. Probability densities associated with Q or P derive from the idea of differentiating the inverse of micro–macro time. In this way, the operation of taking the absolute square of the wavefunction has a geometric motivation, justified by the properties of analytic functions and consistent with the identification, as in (16), between absolute values of crossproducts and absolute squares.

8. SUMMARY

We have presented the outlines of a probabilistic construction from which we are motivated to identify observations on physical systems with the random selection of points in topological spaces. The close analogies between certain mathematical entities in the construction and the entities of quantum mechanics lead us to make the entire construction consistent with quantum mechanics. Details are carried out by restricting our work to the one-dimensional case in which the Hamiltonian generates a complete class of orthonormal functions in L_2-space. We are, thus, led to a theory in which quantum mechanical observables including energy, position, momentum, and time, are represented as functions on topological measure spaces. (We call these functions micro–macro functions, as a modification of the random variable concept.) Each observable, including time, has a probability distribution. The conditional distributions and expectations of energy, position, and momentum, conditioned on the event that micro–macro time has a specific value, are identical with the quantum mechanical distributions and expectations (within the generality of the case studied). It appears that the most natural spaces to serve as domains of definition for the energy and position micro–macro functions are distinct, the energy and position being representative of quantum mechanical operators with discrete and continuous spectra, respectively. The space on which the position function is defined is a geometrical enrichment of the other. The selection of points from these topological measure spaces determine values for energy and position, as well as time. The spaces most natural for the domains of definition of conjugate observables, such as position and momentum, are also distinct and are related through the Fourier transform in such a way as to require the separate selection of a point from each space in order to determine "simulta-

neous" values for position and momentum. The conditional distributions of position, given micro–macro time, and momentum, given micro–macro time, possess the dispersion parameters (namely statistical variance) that satisfy the usual uncertainty relations. Details of these latter facts were not carried out.

Micro–macro time was studied at some length, because of its unifying importance to the theory. The way in which it enters in the theory appears to be consistent with the spirit of relativity. The multiple-valued structure of micro–macro time and the spacings between multiple values were explained (at least tentatively) in terms of the correspondence principle of quantum theory and the conjugate relation between energy and time that is characteristic of relativistic quantum theory. We observe that time, considered as a function on a topological space in which points correspond to observations, has strikingly different structure when the space is consistent with observations on energy and when the space is consistent with observations on position.

We have omitted from our present considerations notably the questions having to do with the degree of generality to which our probabilistic structure can be extended and the extent to which the structure satisfies or conflicts with the requirements of a hidden-variable theory of quantum mechanics.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Algebraic Difficulties of Preserving Dynamical Relations When Forming Quantum-Mechanical Operators*

RICHARD ARENS AND DONALD BABBITT
University of California, Los Angeles, California
(Received 12 October 1964)

Proofs are presented showing impossibility of assigning differential operators (quantum observables) to classical mechanical observables in such a way as to preserve the usual bracket formalism. Difficulty is shown to arise even if we limit ourselves to preserving brackets between the Hamiltonian and a rather limited set of observables. Some other algebraic difficulties inherent in the operator assignment problem are also discussed.

1. INTRODUCTION

IT is known that the set of dynamical variables $\mathfrak{C}$ for a (classical) mechanical system has in a natural way a (real) Lie-algebra structure as well as a squaring operation. Lie multiplication of two dynamical variables is just their Poisson bracket and the square of a dynamical variable is just the square of the function representing the dynamical variable.[1] One of the principal objectives in quantizing a mechanical system is to find a natural "operator assignment" map A from a subspace of $\mathfrak{C}$ into the real vector space $\mathfrak{D}_{sa}$ of formally self-adjoint differential operators on the configuration space of the system in question.[2] Since $\mathfrak{D}_{sa}$ has a natural Lie structure and squaring operation defined on it, it is assumed that A preserves a reasonable amount of the algebraic structure of the domain of A.

In practical situations an operator assignment map is assumed to exist, but to our knowledge none has ever been given explicitly. The usual procedure is to define A explicitly for a limited subset of $\mathfrak{C}$ which usually includes linear and angular momenta, total energy and linear configuration observables. This suffices for most applications but hardly reveals a complete operator assignment procedure. It is our purpose to give several examples which indicate the domain of A must be, at most, a modest subset of $\mathfrak{C}$ if one hopes to have the quantum observables (range of A) retain a reasonable amount of the algebraic structure of the corresponding classical observables.

We stress the algebraic aspect of the operator assignment problem because we feel that the problem of preserving brackets, etc., is mainly an algebraic one and should be separated from typical Hilbert-space problems of finding self-adjoint extensions, etc.

* This paper was prepared with the assistance of NSF grants GP-2045 and GP-1988.

[1] $\mathfrak{C}$ also has the structure of a commutative algebra, but this is supposed not to pass over to quantum mechanics, and is thus, except for the squaring, ignored.

[2] We assume that the configuration space has a natural measure on it, and "self-adjoint" is to be defined in terms of it. We ask the coefficients of differential operators to be infinitely differentiable. These operators thus determine symmetric operators in the Hilbert space of square summable functions on configuration space, but we do not use this fact.

neous" values for position and momentum. The conditional distributions of position, given micro–macro time, and momentum, given micro–macro time, possess the dispersion parameters (namely statistical variance) that satisfy the usual uncertainty relations. Details of these latter facts were not carried out.

Micro–macro time was studied at some length, because of its unifying importance to the theory. The way in which it enters in the theory appears to be consistent with the spirit of relativity. The multiple-valued structure of micro–macro time and the spacings between multiple values were explained (at least tentatively) in terms of the correspondence principle of quantum theory and the conjugate relation between energy and time that is characteristic of relativistic quantum theory. We observe that time, considered as a function on a topological space in which points correspond to observations, has strikingly different structure when the space is consistent with observations on energy and when the space is consistent with observations on position.

We have omitted from our present considerations notably the questions having to do with the degree of generality to which our probabilistic structure can be extended and the extent to which the structure satisfies or conflicts with the requirements of a hidden-variable theory of quantum mechanics.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Algebraic Difficulties of Preserving Dynamical Relations When Forming Quantum-Mechanical Operators*

RICHARD ARENS AND DONALD BABBITT
University of California, Los Angeles, California
(Received 12 October 1964)

Proofs are presented showing impossibility of assigning differential operators (quantum observables) to classical mechanical observables in such a way as to preserve the usual bracket formalism. Difficulty is shown to arise even if we limit ourselves to preserving brackets between the Hamiltonian and a rather limited set of observables. Some other algebraic difficulties inherent in the operator assignment problem are also discussed.

1. INTRODUCTION

IT is known that the set of dynamical variables $\mathfrak{C}$ for a (classical) mechanical system has in a natural way a (real) Lie-algebra structure as well as a squaring operation. Lie multiplication of two dynamical variables is just their Poisson bracket and the square of a dynamical variable is just the square of the function representing the dynamical variable.[1] One of the principal objectives in quantizing a mechanical system is to find a natural "operator assignment" map A from a subspace of $\mathfrak{C}$ into the real vector space $\mathfrak{D}_{sa}$ of formally self-adjoint differential operators on the configuration space of the system in question.[2] Since $\mathfrak{D}_{sa}$ has a natural Lie structure and squaring operation defined on it, it is assumed that A preserves a reasonable amount of the algebraic structure of the domain of A.

In practical situations an operator assignment map is assumed to exist, but to our knowledge none has ever been given explicitly. The usual procedure is to define A explicitly for a limited subset of $\mathfrak{C}$ which usually includes linear and angular momenta, total energy and linear configuration observables. This suffices for most applications but hardly reveals a complete operator assignment procedure. It is our purpose to give several examples which indicate the domain of A must be, at most, a modest subset of $\mathfrak{C}$ if one hopes to have the quantum observables (range of A) retain a reasonable amount of the algebraic structure of the corresponding classical observables.

We stress the algebraic aspect of the operator assignment problem because we feel that the problem of preserving brackets, etc., is mainly an algebraic one and should be separated from typical Hilbert-space problems of finding self-adjoint extensions, etc.

* This paper was prepared with the assistance of NSF grants GP-2045 and GP-1988.

[1] $\mathfrak{C}$ also has the structure of a commutative algebra, but this is supposed not to pass over to quantum mechanics, and is thus, except for the squaring, ignored.

[2] We assume that the configuration space has a natural measure on it, and "self-adjoint" is to be defined in terms of it. We ask the coefficients of differential operators to be infinitely differentiable. These operators thus determine symmetric operators in the Hilbert space of square summable functions on configuration space, but we do not use this fact.

Van Hove[3] proved a result similar to our Theorem 5.1. A result similar to our Theorem 3.1 was proved by Groenewold.[4] Groenewold also stated a theorem similar to our Theorem 4.5 but his proof is unconvincing.[5]

2. WAVE-MECHANICAL OPERATOR ASSIGNMENTS

Let n be an integer (usually 3, but often 1) and let H be a function defined on $\mathbf{R}^{2n}$, real Cartesian space of dimension $2n$. The pair $(\mathbf{R}^{2n}, H)$ is a *Hamiltonian system* with Hamiltonian function H, and is the usual point of departure from classical mechanics to wave mechanics.

As usual, we denote the Cartesian coordinates in $\mathbf{R}^{2n}$ by $(x^1, \cdots, x^n, p_1, \cdots, p_n)$, and call $\mathbf{R}^{2n}$ *phase space. Configuration space* here is $\mathbf{R}^n$ and $(x^1, \cdots, x^n)$ are used to denote the coordinates there. Functions defined on $\mathbf{R}^{2n}$ are (classical) *dynamical variables,* and for two such variables, f, g one defines

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial x^i} - \frac{\partial f}{\partial x^i} \frac{\partial g}{\partial p_i} \right). \tag{2.1}$$

A (*wave-mechanical*) *operator assignment* A is a rule which assigns differential operators on $\mathbf{R}^n$ to some of the dynamical variables.[6] If A assigns an operator to the dynamical variable f, we denote this operator by $A(f)$ or when possible by Af. The class of dynamical variables f for which Af is defined shall be denoted by "dom A." We require that (λ, μ being constant scalars) (2.2) for f and g in dom A one has $\lambda f + \mu g$ in dom A, and $A(\lambda f + \mu g) = \lambda Af + \mu Ag$.

For differential operators F, G on $\mathbf{R}^n$ one can define their commutator

$$[F, G] = FG - GF. \tag{2.3}$$

It is a fact that the customary wave-mechanical operator assignment (denoted for the moment by A) "preserves the bracket" in the sense that $A\{f, g\}$ is proportional $[Af, Ag]$, at least to a certain extent. (It is this extent which we will explore.)

To put the matter quantitatively, we choose a positive number h, and define

$$D_i = (h/i)\partial/\partial x^i, \qquad j = 1, \cdots, n.$$

We modify (2.1) by defining a bracket

$$[f, g] = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial p_i} D_i g - \frac{\partial g}{\partial p_i} D_i f \right). \tag{2.4}$$

We shall now say that A *preserves* the bracket $[f, g]$ if $A[f, g] = [Af, Ag]$.

Our notation (2.4) has the advantage that neither i nor h appears explicitly. Nonetheless, this is the identical substance of the familiar problem.

Let us give an example. Our Hamiltonian H shall be the Hamiltonian for a free particle

$$\frac{1}{2} \sum_{i=1}^{n} (p_i)^2. \tag{2.5}$$

We will define a specific operator assignment to be called A_0. The domain dom A_0 of our operator assignment A_0 shall be all dynamical variables f which are polynomials in $p_1, \cdots, p_n$ with coefficients which are indefinitely differentiable functions of $x^1, \cdots, x^n$. Because of (2.2), it will suffice to assign operators to dynamical observables of the form $ap_1^{m_1} p_2^{m_2} \cdots p_n^{m_n}$. Using a notation to be explained presently, we will define

$$A_0(ap_1^{m_1} \cdots p_n^{m_n}) = (\tfrac{1}{2}\theta_{D_1} + R_{D_1})^{m_1} \cdots (\tfrac{1}{2}\theta_{D_n} + R_{D_n})^{m_n} a. \tag{2.5a}$$

The expression appearing in front of a is to be multiplied out according to commutative algebra. Then the terms are applied to the *multiplication operator* a appearing on the right, as in the following example:

$$c\theta_{D_1}^{k_1} \cdots \theta_{D_n}^{k_n} R_{D_1}^{q_1} \cdots R_{D_n}^{q_n} a = c(D_1^{k_1} \cdots D_n^{k_n} a) D_1^{q_1} \cdots D_n^{q_n}.$$

The rationale here is that θ_{D_i} prefixed to a linear differential operator F means the new operator obtained by applying D_i to each coefficient of F, assuming that F is written in the usual way, with coefficients on the left. Further, $R_{D_i}F$ means FD_i. (Thus θ and R commute.) Later, we use also L_{D_i}, where $(L_{D_i}F)(\psi) = D_i(F\psi)$. It is important to note that

$$L_{D_i} = \theta_{D_i} + R_{D_i} \tag{2.5b}$$

so that the L, θ, R symbols commute.

The term of highest order in (2.5a) is a $D_1^{m_1} \cdots D_n^{m_n}$.

[3] L. Van Hove, Acad. Roy. Belg., Classe Sci. Mém. Coll. in 8° 26, No. 6, 1–102 (1951). Van Hove's results involve the terms of Hilbert space theory, and therefore do not make the algebraic nature of the problem as evident as we hope Theorem VI does. A more recent account of these problems can be found in T. F. Jordan and E. C. G. Sudarshan, Lie group dynamical formalism and the relation between quantum mechanics and classical mechanics, Rev. Mod. Phys. **33**, 515 (1961).

[4] H. J. Groenewold, Physica **12**, 405 (1946).

[5] Groenewold says (Ref. 4, p. 449) that if A is a commutative algebra and B is a linear image ring under a map which preserves squares, then B is commutative. Professor P. C. Curtis, Jr. has kindly shown us that this is not true in general; and also that it is true for finite-dimensional semisimple rings.

[6] The relation A between dynamical variables and operators, may be regarded as a dictionary which assigns familiar names to quantum observables.

We list a few specimens: $A_0(a) = a$,

$$A_0(ap_i) = aD_i + \tfrac{1}{2}(D_i a), \tag{2.5c}$$

$$A_0(ap_1^2) = (\tfrac{1}{2}\theta_1 + R_1)^2 a = (\tfrac{1}{4}\theta_1^2 + \theta_1 R_1 + R_1^2)a$$
$$= \tfrac{1}{4}(D_1^2 a) + (D_1 a)D_1 + aD_1^2, \tag{2.5d}$$

where we abbreviate θ_{D_1} by θ_1, etc., and

$$A_0(ap_1p_2) = (\tfrac{1}{2}\theta_1 + R_1)(\tfrac{1}{2}\theta_2 + R_2)a$$
$$= (\tfrac{1}{4}\theta_1\theta_2 + \tfrac{1}{2}\theta_1 R_2 + \tfrac{1}{2}\theta_2 R_1 + R_1R_2)a$$
$$= \tfrac{1}{4}(D_1D_2a) + \tfrac{1}{2}(D_1a)D_2$$
$$+ \tfrac{1}{2}(D_2a)D_1 + aD_1D_2. \tag{2.5e}$$

The extent to which this A_0 preserves brackets is indicated by the following.

Theorem I. For the operator assignment A_0 given by (2.5a) and the Hamiltonian (2.5),

$$A_0([H, f]) = [A_0H, A_0f]. \tag{2.6}$$

For f and g whose degrees as polynomials in $p_1, \cdots, p_n$ have a sum at most two,

$$A_0([f, g]) = [A_0f, A_0g]. \tag{2.7}$$

But (2.7) does not always hold, since

$$A_0[a, bp_1^3] = [a, A_0(bp_1^3)] + \tfrac{1}{4}b(D_1^3a). \tag{2.8}$$

Proof. The assertions concerning (2.7) and (2.8) are established by working out individually the seven different cases that may occur, and we ask the reader to accept them as valid. However, (2.6) involves some technique, and a proof is now given. It suffices to take

$$f = ap_1^{m_1} \cdots p_n^{m_n}.$$

Considering (2.5) and (2.4) we obtain

$$[H, f] = (\sum_i a_i p_i)p_1^{m_1} \cdots p_n^{m_n},$$

where a_i stands for D_ia. Writing θ_i for θ_{D_i}, etc., as we did before, (2.51) yields

$$A_0[H, f] = (\tfrac{1}{2}\theta_1 + R_1)^{m_1} \cdots$$
$$\times (\tfrac{1}{2}\theta_n + R_n)^{m_n} \sum_i (\tfrac{1}{2}\theta_i + R_i)a_i, \tag{2.8a}$$

because (2.5a) says, briefly, that a factor $\frac{1}{2}\theta_i + R_i$ should be included in $A_0(f)$ for each occurrence of p_i in f.

For the Hamiltonian itself we have

$$A_0(H) = \tfrac{1}{2} \sum D_i^2.$$

Hence

$$A_0(f)A_0(H) = (\tfrac{1}{2}\theta_1 + R_1)^{m_1} \cdots$$
$$\times (\tfrac{1}{2}\theta_n + R_n)^{m_n}\tfrac{1}{2} \sum_i R_i^2 a.$$

This is true because $R_ia = aD_i$, etc., as explained before. For $A_0(H)A_0(f)$ we need to use L_i which is an abbreviation for L_{D_i}, since it means "put D_i in front of the operator" i.e., "apply the operator and then apply D_i." Thus

$$A_0(H)A_0(f) = (\tfrac{1}{2}\theta_1 + R_1)^{m_1} \cdots$$
$$\times (\tfrac{1}{2}\theta_n + R_n)^{m_n}\tfrac{1}{2} \sum_i L_i^2 a,$$

whence

$$[A_0(H), A_0(f)] = (\tfrac{1}{2}\theta_1 + R_1)^{m_1} \cdots$$
$$\times (\tfrac{1}{2}\theta_n + R_n)^{m_n}\tfrac{1}{2} \sum (L_i^2 - R_i^2)a. \tag{2.8b}$$

Using (2.5b), we see that $L_i^2 - R_i^2 = (L_i + R_i)\theta_i = (\theta_i + 2R_i)\theta_i$. Hence

$$\tfrac{1}{2} \sum (L_i^2 - R_i^2) = \sum (\tfrac{1}{2}\theta_i + R_i)a_i.$$

Inserting this into (2.8b), and comparing the latter with (2.8a), establishes (2.6). Thus A_0 preserves all brackets involving H, some brackets not involving H, but not all brackets.

3. NO ASSIGNMENT PRESERVES ALL BRACKETS

Theorem II. Let A be any wave-mechanical operator assignment for the free particle with Hamiltonian (2.5). Suppose dom A contains all polynomials in the p's and the x's. Suppose that for every[7] a,

$$A([\sum p_i^2, a]) = [\sum D_i^2, a] \tag{3.1}$$

and

$$A([\sum p_i^2, ap_1]) = [\sum D_i^2, A(ap_1)]. \tag{3.2}$$

Nevertheless, there are a and b for which

$$A([[\sum p_i^2, ap_1], bp_1])$$
$$\neq [A[\sum p_i^2, ap_1], A(bp_1)]. \tag{3.3}$$

Proof. The assignment A_0 of 2.5a has properties (3.1) and (3.2). Let $B = A - A_0$. Then

$$B([\sum p_i^2, a]) = 0, \tag{3.1a}$$

and

$$B([\sum p_i^2, ap_1]) = [\sum D_i^2, B(ap_1)]. \tag{3.2a}$$

Suppose a depends only on x^1. Then (3.1a) says that $B(2p_1a_1) = 0$ where $a_1 = D_1a$. Every polynomial depending only on x^1 is of the form a_1 so that $B(ap_1) \equiv 0$ whenever a depends only on x^1. Thereupon (3.2a) tells us that $B(2p_1a_1p_1) = 0$ and we conclude that $B(ap_1^2) \equiv 0$ whenever a depends only on x^1. Therefore, if $=$ were true in (3.3), it would

[7] As implied in the hypotheses, the a's and b's in (3.1)–(3.3) are polynomials in the x's. An analog of Theorem II in which the a's and b's are analytically different is treated later in Sec. 5.

also be true that (as we shall suppose)

$$A_0([2p_1ap_1,\ bp_1]) = [A_0(2p_1ap_1),\ A_0(bp_1)], \tag{3.3a}$$

wherever a and b depend only on x^1. Whereas A was quite unknown, A_0 is explicitly known. As the reader may verify, the zeroth-order term for the operator on the left of (3.3a) is $a_1b_{111}+a_{11}b_{11}-\frac{1}{2}a_{1111}b$. On the right, it comes out $a_1b_{111} + \frac{3}{2}a_{11}b_{11} - \frac{1}{2}a_{1111}b$. (Here $a_{11} = D_1a_1$, etc.) Thus (3.3a) is absurd, and Theorem II is proved.

Remark. The term $[\sum p_i^2,\ ap_1]$ in (3.3) shows that no operator assignment preserves brackets even if we restrict its domain to the linear dynamical observables [compare (2.7)], plus those obtained from these by bracketing with H.

We now consider a Hamiltonian of the form

$$H = \tfrac{1}{2}\sum p_i^2 + v, \tag{3.4}$$

where v depends only on $x^1, \cdots, x^n$. Addition of (2.6) and (2.8) shows that now, even brackets $[H, f]$ may not be preserved by A_0. Our next theorem shows that the same is true for any operator assignment. For simplicity, we consider only the one-dimensional case, and write D for D_1, x for x^1, etc. But D_1a we denote by a_1.

Theorem III. Let A be a wave-mechanical operator assignment for the one-dimensional system with Hamiltonian $H = \frac{1}{2}p^2 + v$. Suppose dom A contains all polynomials in the x's and p's. Suppose further that $A(a) = a$ for each polynomial a depending only on x, and for each such a,

$$A([\tfrac{1}{2}p^2 + v,\ ap^k]) = [\tfrac{1}{2}D^2 + v,\ A(ap^k)] \tag{3.5a}$$

for $k = 0, 1, 2, 3$. Then the third derivative v''' of the potential is zero.

Proof. First we define $R_0, \cdots, R_3$ by

$$A_0[v,\ ap^k] = [v,\ A_0(ap^k)] + R_k.$$

We can also write the equation

$$A_0[\tfrac{1}{2}p^2,\ ap^k] = [\tfrac{1}{2}D^2,\ A_0(ap^k)].$$

From (2.8) we obtain $R_0 = R_1 = R_2 = 0, R_3 = \frac{1}{4}av_{111}$. From (3.5a) we obtain

$$A[\tfrac{1}{2}p^2,\ ap^k] - A(kav_1p^{k-1}) = [\tfrac{1}{2}D^2,\ A(ap^k)] + [v,\ A(ap^k)].$$

In this equation we let $A = A_0 + B$ and use the two preceding equations about A_0. The result is

$$B[\tfrac{1}{2}p^2,\ ap^k] - kB(av_1p^{k-1}) = [\tfrac{1}{2}D^2 + v,\ B(ap^k)] - R_k,$$

so that

$$B(a_1p^{k+1}) = kB(av_1p^{k-1}) + [\tfrac{1}{2}D^2 + v,\ B(ap^k)] - R_k. \tag{3.5b}$$

Now $B(a) = A(a) - A_0(a) = 0$. Thus (3.5b) written down for $k = 0$ gives $B(a_1p) = 0$. This implies $B(ap) = 0$ for all a. Then (3.5b) for $k = 1$ gives $B(a_1p^2) = 0$. This argument is repeated until $k = 3$, where we get $B(a_1p^4) = -R_3$. Now we select $a = 1$. Then $a_1 = 0$ so that $0 = \frac{1}{4}v_{111}$. Thus $v''' = 0$.

This shows that only for the harmonic oscillator and the falling body do you have preservation of all brackets containing H.

4. PRESERVING SQUARING

Theorems I and II show that for a free particle (2.5), the operator assignment A_0 preserves brackets as well as any other, and in particular those involving H. But brackets are not the only things one likes to see preserved: there is also the algebraic operation of squaring. The impossibility of preserving both, even though dom A is held down to polynomials in x and p of degree not exceeding 4, is shown by the following.

Theorem IV. Let A be a wave-mechanical operator assignment for the system (2.5) ($n = 1$). Suppose x^2, x^4, $[H, x^2]$, $[H, x^2]^2$, and $[H, [H, x^4]]$ are in dom A. Suppose

$$A([\tfrac{1}{2}p^2,\ [\tfrac{1}{2}p^2,\ x^4]]) = [\tfrac{1}{2}D^2,\ [\tfrac{1}{2}D^2,\ x^4]]. \tag{4.1}$$

Then

$$A(3[\tfrac{1}{2}p^2,\ x^2]^2) \neq 3[\tfrac{1}{2}D^2,\ x^2]^2. \tag{4.2}$$

Proof. The 3 in (4.2) has been put in so that the dynamical variables in (4.1) and (4.2) upon which A shall act are identical, namely, $12x\,x^2p^2$. But the operators on the right emerge as $12x_1^2x^2D^2 + 24x_1^3xD + 6x_1^4$ and $12x_1^2x^2D^2 + 24x_1^3xD + 4x_1^4$ respectively, thus proving Theorem IV.

Suppose we forget about preserving brackets, but ask that A have the familiar form

$$A(ap) = aD + \tfrac{1}{2}(Da). \tag{4.3}$$

Then squaring cannot be preserved, either.

Theorem V. Suppose A (for $n=1$) has the property (4.3). Suppose x^3p, $(x^2 \pm x^4)p$, x^6p^2, and $(x^2 \pm x^4)^2p^2$ are in dom A. Then the following three relations cannot all hold:

$$A(x^6p^2) = (A(x^3p))^2, \tag{4.4a}$$

$$A((x^2 + x^4)^2p^2) = (A((x^2 + x^4)p))^2, \tag{4.4b}$$

$$A((x^2 - x^4)^2p^2) = (A((x^2 - x^4)p))^2. \tag{4.4c}$$

Proof. Subtracting (4.4c) from (4.4b), and using (4.4a) leads to the relation

$$2(A(x^3)p)^2 = A(x^2p)A(x^4p) + A(x^4p)A(x^2p).$$

But this is incompatible with (4.3).

5. AN ANALYTIC VARIANT OF AN EARLIER RESULT

It might still be hoped that we could find an operator assignment A with the property that $A[f, g] = [Af, Ag]$ whenever the dynamical variables f, g, and $\{f, g\}$ all generate one-parameter groups of contact transformations. The dynamical variables treated in Sec. 3 do not generate one-parameter groups of contact transformations in general. The following theorem will show no such operator assignment exists even if we restrict ourselves to dynamical variables which are at most quadratic in the p's. Moreover, as in the previous results, the difficulty will be of an algebraic nature and will have nothing to do with Hilbert-space theory.

Theorem VI. There does not exist an operator assignment A for the system $(\mathbf{R}^{2n}, \frac{1}{2}(\sum_{i=1}^{n} p_i^2))$ with the following properties:

(a) dom A contains all dynamical variables which generate one-parameter groups of contact transformations and are at most degree 2 in the p's;

(b) if $f(x^1, \cdots, x^n)$ is an indefinitely differentiable function on $\mathbf{R}^n$ and $\{\frac{1}{2}(\sum_{i=1}^{n} p_i^2), f(x^1, \cdots, x^n)\}$ generates a one-parameter group of contact transformations, then

$$A\left[\frac{1}{2}\left(\sum_{i=1}^{n} p_i^2\right), f(x^1, \cdots, x^n)\right] = \left[\frac{1}{2}\left(\sum_{i=1}^{n} D_i^2\right), f(x, \cdots, x^n)\right];$$

(c) if $f(x^1, \cdots, x^n)$ is as in (b) and

$$\left\{\frac{1}{2}\left(\sum_{i=1}^{n} p_i^2\right), \left\{\frac{1}{2}\left(\sum_{i=1}^{n} p_i^2\right), f(x^1, \cdots, x^n)\right\}\right\}$$

generate a one-parameter group of contact transformations, then

$$A\left[\frac{1}{2}\left(\sum_{i=1}^{n} p_i^2\right), \left[\frac{1}{2}\left(\sum_{i=1}^{n} p_i^2\right), f(x^1, \cdots, x^n)\right]\right] = \left[\frac{1}{2}\left(\sum_{i=1}^{n} D_i^2\right), \left[\frac{1}{2}\left(\sum_{i=1}^{n} D_i^2\right), f(x^1, \cdots, x^n)\right]\right];$$

(d) if $f = \sum_{i,k} a_{ik}(x^1, \cdots, x^n)p_i p_k$, $g = \sum_i b_i(x^1, \cdots, x^n)p_i$ and $\{f, g\}$ generate one-parameter groups of contact transformations, then

$$A[f, g] = [Af, Ag].$$

Proof. We first prove the theorem for the case $n = 1$. Let $0 < c < d$ and define

$$f(x) = \exp\frac{1}{x-d} - \frac{1}{x-c}, \quad \text{if } c < x < d$$
$$= 0, \quad \text{otherwise.}$$

f is indefinitely differentiable as is

$$F(x) = \int_x^d f(t)\,dt \Big/ \int_c^d f(t)\,dt$$

which has value 1 for $x \leq c$ and 0 for $x \geq d$. Let $\tilde{a}(x) = F(x^2)$ and $a(x) = \tilde{a}(x) - \tilde{a}(x - 2d^{\frac{1}{2}})$. Then it is easy to verify that $a(x)$ has the following properties:

(i) $a(x)$ is indefinitely differentiable and has compact support;

(ii) if $\bar{x}$ is such that $a(\bar{x}) = 0$, then $da(\bar{x})/dx = 0$ and $d^2a(\bar{x})/dx^2 = 0$;

(iii) if $\bar{x}$ is such that $da(\bar{x})/dx = 0$, then $d^2a(\bar{x})/dx^2 = 0$;

(iv) $\int_{-\infty}^{\infty} a(x)\,dx = 0$;

(v) $(da/dx)d^2(a^2)/dx^2 \not\equiv 0$.

Properties (i)–(v) guarantee that

$$a(x)p^2 = \left\{\frac{1}{2}p^2, \left\{\frac{1}{2}p^2, \int_{-\infty}^{x}\int_{-\infty}^{t} a(u)\,du\,dt\right\}\right\}, \tag{5.1a}$$

$$a^2(x)p = \left\{\frac{1}{2}p^2, \int_{-\infty}^{x} a^2(t)\,dt\right\}, \tag{5.1b}$$

$$3a^2(x)\frac{da}{dx}\cdot p^2 = \{a(x)p^2, a^2(x)p\} = \left\{\frac{1}{2}p^2, \left\{\frac{1}{2}p^2, \int_{-\infty}^{x} a^3(t)\,dt\right\}\right\} \tag{5.1c}$$

all generate one-parameter groups of contact transformations. This follows on one hand from the fact that $C(x)p^2$ generates a one-parameter group of contact transformations if $C(x)$ is indefinitely differentiable with compact support and has the property that $C(\bar{x}) = 0$ implies $dC(\bar{x})/dx = 0$ [which (i)–(iv) guarantee for (5.1a) and (5.1c)] and on the other hand from the fact that $C(x)p$ generates a one-parameter group of contact transformations if C is indefinitely differentiable with compact support [which (i) guarantees for (5.1b)].

Now note that (5.1a, b, c) and conditions (a), (b), (c) on the operator assignment A determine $A(a(x)p^2)$, $A(a^2(x)p)$, and $A[a(x)p^2, a^2(x)p]$, and thus a straightforward calculation gives

$$[A(a(x)p^2), A(a^2(x)p)] = A[a(x)p^2, a^2(x)p] - \tfrac{1}{4}a_1(a^2)_{11}$$

and then (d) cannot hold by property (v) for $a(x)$.

The proof for arbitrary n follows directly by applying the above arguments to

$$f = \sum_{i=1}^{n} a^2(x^i)p_i, \qquad g = \sum_{i=1}^{n} a(x^i)p_i^2$$

and $\{g, f\}$. Then the theorem is proved.

Functional Expansion in Peratization Theory*

H. M. FRIED
Physics Department, Brown University, Providence, Rhode Island
(Received 1 February 1965)

A functional peratization expansion is defined in the charged vector meson theory of weak interactions. The simplest approximation is carried through and shown to yield an allowed lepton–lepton scattering amplitude, properly damped on the light cone. A basic deficiency in this procedure is the absence of any criteria for the estimation of the character and size of successive corrections.

I. INTRODUCTION

HIGHER-ORDER corrections to the leptonic weak interactions, mediated by charged W-mesons of spin 1, using the technique of the so-called peratization approximation to the ladder approximation Bethe–Salpeter (BS) equation,[1] have recently been discussed by Feinberg and Pais.[2,3] Their main result concerning the effective modification of the one-boson-exchange scattering amplitude at zero momentum transfer has been derived in an alternate manner by Pwu and Wu,[4] and also by Bég,[5] again in the approximation of considering ladder graphs only. An essential part of these calculations is the use of regularized propagators; all iterations are performed in terms of these quantities, and at the end of the calculation the regulator masses are allowed to become infinite and finite results are obtained. However, it was noticed by Bardakci, Bolsterli, and Suura[6] that, if the ladder approximation iterations are performed in configuration space without prior regularization, the resulting amplitude has an essential singularity on the light cone which cannot be regularized away; that is, regularization will permit a Fourier transform to be defined, but in the limit of large regulator masses, the essential singularity will again appear.

The purpose of this paper is to investigate, in one special approximate way, the effect of all the contributing ladder and crossed graphs on the leptonic scattering amplitudes. It has been emphasized by Pwu and Wu[7] that there is no reason to consider the crossed graphs as less significant than the ladder graphs; and it has long been known[8] that, for example, in a neutral vector meson theory all graphs of each order must be included in order to demonstrate gauge invariance. From the work of Zumino[9,10] it is clear that although configuration space essential singularities do occur, they are canceled upon going to the momentum-space mass shell of the amputated scattering amplitudes in gauge-invariant theories; and conversely, such singularities—but without the delicate mass-shell cancellations—may be anticipated in certain approximations to non-gauge-invariant theories, and specifically in the charged vector meson theory considered here. The possibility of finding essential singularities upon iterating the simplest crossed graph has been mentioned in a previous note[11]; in the language of the BS equation, what is of importance here is the light-cone behavior of the entire irreducible amplitude, rather than the behavior of a finite-order approximation thereof.

The aim here is to attempt a treatment of the crossed graphs which isolates possible damping mechanisms corresponding to the summation of successive terms from an infinite number of such graphs. For this discussion, a damped amplitude is one which does not contain an essential singularity on the light cone, and a damping mechanism is understood to mean a method of rearranging and summing selected approximations to all the permitted Feynman graphs which results in a damped amplitude. Such damping is the main goal of this paper; but it should be emphasized immediately that a discussion so limited is unsatisfactory. Aside from the simplifying approximations of neglecting self-energy, vertex-type, and closed fermion loop structure, which one cannot even begin to justify, the basic shortcoming is that there is no obvious way of guaranteeing that corrections to the damped amplitude obtained here will (a) also be damped, and will (b)

* Supported, in part, by the U. S. Atomic Energy Commission.

[1] E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).
[2] G. Feinberg and A. Pais, Phys. Rev. **131**, 2724 (1963).
[3] G. Feinberg and A. Pais, Phys. Rev. **133**, B477 (1964).
[4] Y. Pwu and T. T. Wu, Phys. Rev. **133**, B778 (1964).
[5] M. A. B. Bég, Ann Phys. (N. Y.) **27**, 183 (1964).
[6] K. Bardakci, M. Bolsterli, and H. Suura, Phys. Rev. **133**, B1273 (1964).
[7] Y. Pwu and T. T. Wu, Phys. Rev. **133**, B1299 (1964).
[8] R. P. Feynman, Phys. Rev. **76**, 769 (1949).
[9] B. Zumino, J. Math. Phys. **1**, 1 (1960).
[10] B. Zumino, Nuovo Cimento **17**, 547 (1960).
[11] H. M. Fried, Phys. Rev. **133**, B1562 (1964).

possess Fourier transforms which, for certain values of momentum transfer, for example, are of higher order. Even if (a) is satisfied, it is easy to construct examples which violate (b). The existence of sets of damped amplitudes is in some sense the configuration-space analog of momentum-space power counting arguments,[12] where one sums the leading divergent singularities of each order, the next-to-leading singularities, and so on.

In spite of these inadequacies, this discussion may be of some value in setting the stage for a more useful study. It is possible to state the problem, at various levels of approximation, in terms of corresponding graphical approximations to the complete BS equation. However, it is considerably more convenient to employ a variant of Schwinger's functional techniques in order to enumerate, in succinct fashion, all the Feynman graphs permitted by the selection rules for each specific amplitude. Further, the functional method has the advantage of permitting a simple comparison with gauge-invariant theories, and, indeed, of suggesting approximation schemes based upon gauge techniques. The method employed works directly with the functional expression for the scattering amplitude, with a peratization[13] expansion given in terms of approximations to the basic fermion Green's function defined in the presence of an external, gradient source. Although there is no gauge invariance in this problem, the gauge properties of these Green's functions are pertinent here; and one has the strong feeling that further progress in this subject is dependent upon a better application of these gauge techniques.

The arrangement of these few remarks is as follows. Section II contains the basic notation. The peratization expansion is defined in Sec. III, where the contribution of the simplest approximation to the allowed lepton–lepton scattering amplitude is computed. Corrections may easily be defined, although their calculation may be more difficult, using the same methods; but these are not considered in this paper. Some miscellaneous comments concerning these approximations, and some of the details, have been put into a short Appendix.

II. FUNCTIONAL NOTATION, PAISOTOPICS,[14] AND THE BS EQUATION

The basic problem under consideration is the scattering of a pair of leptons, say A and B, interacting via the exchange of charged W-mesons.[15] To avoid clumsy symmetrization of the configuration space amplitudes, A and B will always be considered as distinct; that is, A will correspond to an electron and/or its neutrino, and B to a muon and/or its neutrino. The coordinates of A will always be designated by the subscript 1, while those of B shall carry the subscript 2. All lepton masses are zero and the W mass is M. Using a convenient p-spin notation, the interaction Lagrangian may be written in the form

$$L' = (ig/\sqrt{2})\bar{\psi}_e\gamma_\mu(1+\gamma_5)\mathfrak{d}\cdot\mathbf{W}_\mu\psi_e + (e\leftrightarrow\mu), \quad (1)$$

where

$$\psi_e = \begin{pmatrix} e \\ \nu_e \end{pmatrix}, \ \psi_\mu = \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \qquad W_\mu^{(1)} = \frac{1}{\sqrt{2}}[W_\mu + W_\mu^\dagger],$$

$$W_\mu^{(2)} = (i/\sqrt{2})[W_\mu - W_\mu^\dagger], \qquad W_\mu^{(3)} = 0,$$

and where g is the coupling constant of Ref. 2; the σ_i are the Pauli matrices. Under the approximation of neglecting all leptonic closed loops, vertex-type, and self-energy structure, the unamputated scattering amplitude is defined functionally (Appendix A) by

$$M(x;y) = M(x_1y_1, x_2y_2) = e^{D_F[\delta/\delta J,\,\delta/\delta J']}G_A(x_1y_1 \mid \mathbf{J}_\alpha) \times G_B(x_2y_2 \mid \mathbf{J}'_\beta)|_{J=J'=0}, \quad (2)$$

where $G_{A,B}$ is the Green's function of lepton A or B in the external c-number field $J_\mu^{(i)}(x)$, $i = 1$ and 2,

$$[\gamma_\mu\partial_\mu^x - (ig/\sqrt{2})\gamma_\mu(1+\gamma_5)\mathfrak{d}\cdot\mathbf{J}_\mu(x)] \times G(xy \mid \mathbf{J}_\alpha) = \delta(x-y). \quad (3)$$

The functional differentiation operator D_F is defined by

$$D_F\left[\frac{\delta}{\delta J}, \frac{\delta}{\delta J'}\right] = -i\int \sum_{i=1}^{2} \Delta_F^{\mu\nu}(z_1 - z_2)\frac{\delta}{\delta J_\mu^{(i)}(z_1)}\cdot\frac{\delta}{\delta J_\nu'^{(i)}(z_2)}, \quad (4)$$

[12] Such as those introduced by T. D. Lee, Phys. Rev. **128**, 899 (1962).

[13] The definition of peratization really needs some clarification. It is not proper to use this word to describe the contributions obtained from the gradient part of the boson propagator only, since many prior gauge-invariance arguments have discussed just these terms; yet the latter hopefully provide the dominant effect when the special gauge cancellations are prevented by the selection rules. Peratization is here used to specify that initial step of an approximation procedure, suggested by the dominance of the boson propagator's gradient terms in those situations where one knows beforehand that the latter do not cancel. Physically, the approximation means that external momenta are neglected compared to virtual momenta.

[14] We single out the isotopic spin formalism useful in describing the purely leptonic weak interactions by assigning to it the name "Paisotopic spin," or p-spin.

[15] The methods used here stem from the introduction of charged vector meson fields in the basic interaction. An alternate approach, in which these bosons appear as resonant objects obtained from a Fermi interaction, has been outlined by J. G. Taylor, Nuovo Cimento Suppl. **1**, 857 (1964).

where $\Delta_F^{\mu\nu}(x) = [\delta_{\mu\nu} - M^{-2}\partial_\mu\partial_\nu]\Delta_F(x)$ is the boson propagator of mass M. It is not difficult to see that the expansion of the exponential factor of (2) generates all the ladder and crossed graphs consistent with the selection rules. The connected, amputated, configuration-space scattering amplitude is denoted by $T(x; y)$, and is related to $M(x; y)$ by

$$M(x; y) = S_F^A(x_1 - y_1)S_F^B(x_2 - y_2) + \int S_F^A(x_1 - u_1)S_F^B(x_2 - u_2)T(u; v) \times S_F^A(v_1 - y_1)S_F^B(v_2 - y_2),$$

where the $S_F^{\alpha,B}$ represent free-lepton propagators of zero mass. The mass-shell Fourier transform of T, between the appropriate lepton spinors, is the corresponding S-matrix scattering amplitude.

Corresponding to the interaction (1), under the restriction of charge conservation, the scattering amplitude T may be separated into three paisotopic amplitudes,

$$T(x; y) = P_a T_a(x; y) + P_f T_f(x; y) + P_c T_c(x; y)$$

as noted in Ref. 11 and repeated here for completeness. The "allowed" amplitude (e.g., for the process $e^- + \nu_\mu \to \nu_e + \mu^-$) is denoted by T_a, and is represented in lowest order by the one-boson-exchange graph; T_f denotes the "forbidden" amplitude (e.g., $e^- + \nu_\mu \to e^- + \nu_\mu$) represented in lowest order by a two-rung ladder graph; and T_c denotes the "forbidden" amplitude (e.g., $e^- + \mu^- \to e^- + \mu^-$) represented in lowest order by the simplest crossed graph. The operators $P_{a,f,c}$ are given by

$$P_a = \frac{1}{2}\sum_{i=1}^{2}\sigma_i^A\sigma_i^B,$$

$$P_f = \tfrac{1}{2}[1 - \sigma_3^A\sigma_3^B],$$

$$P_c = \tfrac{1}{2}[1 + \sigma_3^A\sigma_3^B],$$

and are linearly related to the three p-spin projection operators $P_{1,2,3}$:

$$P_1 = \tfrac{1}{2}[P_f + P_a], \quad P_2 = \tfrac{1}{2}[P_f - P_a], \quad P_3 = P_c.$$

The relation of (2) to the conventional BS equation can be seen by amputating on the $x_{1,2}$ coordinates; this yields, with the aid of (3),

$$\gamma_\mu^A\partial_\mu^{x_1}\cdot\gamma_\nu^B\partial_\nu^{x_2}M(x; y) \equiv M(\bar{x}_1y_1, \bar{x}_2y_2) = \delta(x_1 - y_1)\delta(x_2 - y_2) - \frac{g^2}{2}e^{D_F}\sum_{i,i'=1}^{2}[\gamma_\mu(1 + \gamma_5)\sigma_i]_A[\gamma_\nu(1 + \gamma_5)\sigma_{i'}]_B \times J_\mu^{(i)}(x_1)J_\nu'^{(i')}(x_2)G_A(x_1y_1|\mathbf{J}_\alpha)G_B(x_2y_2|\mathbf{J}_\beta')|_0. \quad (5)$$

The last term of (5) splits naturally into two parts:

$$ig^2[\gamma_\mu(1 + \gamma_5)]_A[\gamma_\nu(1 + \gamma_5)]_B\Delta_F^{\mu\nu}(x_1 - x_2)P_aM(x; y) + \frac{g^2}{2}\sum_{i,i'=1}^{2}[\gamma_\mu(1 + \gamma_5)\sigma_i]_A[\gamma_\nu(1 + \gamma_5)\sigma_{i'}]_Be^{D_F} \times \int \Delta_F^{\mu\sigma}(x_1 - z_2)\Delta_F^{\nu\lambda}(x_2 - z_1)\frac{\delta}{\delta J_\lambda^{(i')}(z_1)}G_A(x_1y_1 \mid \mathbf{J}_\alpha) \times \frac{\delta}{\delta J_\sigma'^{(i)}(z_2)}G_B(x_2y_2 \mid \mathbf{J}_\beta')|_0. \quad (6)$$

In the second term of (6) are included these graphs containing at least one pair of crossed internal lines, among which are all the irreducible crossed graphs; it is evident that the use of only the first term of (6), in (5), generates the coupled ladder graph BS equations for T_a and T_f (written for the partially amputated amplitudes) of Ref. 2.

III. THE PERATIZATION APPROXIMATION

A functional peratization method is now applied to the basic amplitude (2). We first write the functional differential operator $D_F[\delta/\delta J, \delta/\delta J']$ in the form $D_F = D_F^\delta + D_F^g$ where D_F^δ is obtained from the $\delta_{\mu\nu}$ part of the boson propagator in (4), and D_F^g is the contribution of the gradient portion. The unamputated amplitude is then written in the form

$$M(x; y) = e^{D_F^\delta}\cdot e^{D_F^g}G_A(x_1y_1 \mid \mathbf{J}_\alpha)G_B(x_2y_2 \mid \mathbf{J}_\beta')|_0, \quad (7)$$

and (7) is expanded in powers of D_F^δ. Thus $M = \sum_{n=0}^{\infty}M^{(n)}$, where the superscript n refers to the number of "$\delta_{\mu\nu}$ boson lines" exchanged. The object here is to define an expansion in powers of that part of the coupling associated with the conventionally renormalizable part of the interaction, and simultaneously, to sum over all terms associated with the nonrenormalizable part of the interaction. By extracting contributions from all graphs in this way, one hopes to construct a set of amplitudes each of which is well damped on the light cone.

We calculate here only $M^{(0)}$ (or rather, its first approximation, defined below),

$$M^{(0)}(x; y) = G_A(x_1y_1|\partial_\alpha\mathbf{\Lambda})G_B(x_2y_2 \mid \mathbf{J}_\beta')|_0, \quad (8)$$

where

$$\Lambda^{(i)}(z_1) = -\frac{i}{M^2}\int \partial_\sigma^{z_2}\Delta_F(z_1 - z_2)\frac{\delta}{\delta J_\sigma'^{(i)}(z_2)}, \quad (9)$$

and it is now necessary to evaluate $G(xy \mid \partial_\alpha\mathbf{\Lambda})$. Because of the everpresent Paisotopic factors, this problem, trivial in a gauge-invariant theory, is quite

impossible here, and a further set of approximations must be developed.

The method adopted consists of a peratization approximation to the lepton Green's functions defined in the presence of an external gradient source, as in (9). One defines a sequence of approximations to $G(xy \mid \partial_\alpha \mathbf{\Lambda})$, say $G = \sum_{m=0}^{\infty} G^{(m)}$, which can be inserted term by term into (8) and the functional operations performed explicitly. The only qualification is that certain functions, for particular values of their arguments, will have to be considered as continuations of the explicitly obtained functions; this situation is familiar from the work of Ref. 2. In the place of (7) and (8) one then writes

$$M = \sum_{n,m=0}^{\infty} M^{(nm)} = \sum_{n,m=0}^{\infty} \frac{(D_F^{\delta})^n}{n!} \times G_A^{(m)}(x_1 y_1 \mid \mathbf{\Lambda}) G_B(x_2 y_2 \mid \mathbf{J}'_\beta)\mid_0, \tag{10}$$

and in particular, we consider here only the simplest term,

$$M^{(00)}(x; y) = G_A^{(0)}(x_1 y_1 \mid \mathbf{\Lambda}) G_B(x_2 y_2 \mid \mathbf{J}'_\beta)\mid_0. \tag{11}$$

With such an expansion procedure, the corresponding amputated functions $T^{(nm)}(x; y)$ are related to the $M^{(nm)}(x; y)$ by

$$M^{(nm)}(x; y) = S_F^A(x_1 - y_1) S_F^B(x_2 - y_2)\delta_{n0}\delta_{m0} + \int S_F^A(x_1 - u_1) S_F^B(x_2 - u_2) T^{(nm)}(u; v) \times S_F^A(v_1 - y_1) S_F^B(v_2 - y_2). \tag{12}$$

The expansion used here may be defined starting from the integral form of (3),

$$G(xy \mid \partial_\alpha \mathbf{\Lambda}) = S_F(x - y) + i\alpha \int S_F(x - z) \times [\boldsymbol{\vartheta}\cdot\boldsymbol{\gamma}_\mu \partial_\mu^z \mathbf{\Lambda}(z)] G(zy \mid \partial_\alpha \mathbf{\Lambda}), \tag{3a}$$

$$\alpha = (g/\sqrt{2})(1 + \gamma_5).$$

Performing an integration by parts on the last term of (3b), the result may be written in the form

$$G_\lambda(xy \mid \partial_\alpha \mathbf{\Lambda}) = S_F(x - y) - i\alpha \int S_F(x - z) \times [\boldsymbol{\vartheta}\cdot\mathbf{\Lambda}(z)]\gamma_\mu \partial_\mu^z G_\lambda(zy \mid \partial_\alpha \mathbf{\Lambda}) + i\alpha\lambda[\boldsymbol{\vartheta}\cdot\mathbf{\Lambda}(z)] G_\lambda(zy \mid \partial_\alpha \mathbf{\Lambda}), \tag{13}$$

where the desired Green's function is then given by $G[\lambda = 1]$. If the solution to (13) is expanded in powers of λ,

$$G_\lambda = \sum_{m=0}^{\infty} \lambda^m G^{(m)},$$

one will be performing, in momentum space, the approximation of neglecting external (q) compared to virtual (k) momenta in the integral equation for the partially amputated function. The lowest-order approximation takes the familiar form $\gamma\cdot k[\gamma\cdot k - \gamma\cdot q]^{-1} \approx 1$, and generates

$$\gamma_\mu \partial_\mu^x G^{(0)}(xy \mid \mathbf{\Lambda}) \equiv G^{(0)}(\bar{x}y \mid \mathbf{\Lambda}) = [1 + i\alpha\boldsymbol{\vartheta}\cdot\mathbf{\Lambda}(x)]^{-1}\delta(x - y). \tag{14}$$

Using the adjoint form of (3) to extract the disconnected part of the amplitude, one has

$$M^{(00)}(x; y) - S_F^A(x_1 - y_1) S_F^B(x_2 - y_2) = i\frac{g}{\sqrt{2}}(1 + \gamma_5)_A \int G_A^{(0)}(x_1 z_1 \mid \mathbf{\Lambda})[\boldsymbol{\vartheta}_A\cdot\boldsymbol{\gamma}_\mu^A \partial_\mu^{z_1} \mathbf{\Lambda}(z_1)] \times G_B(x_2 y_2 \mid \mathbf{J}'_\beta)\mid_0 \cdot S_F^A(z_1 - y_1), \tag{15}$$

or, in terms of the amputated amplitude,

$$T^{(00)}(x; y) = -i(g/\sqrt{2})(1 - \gamma_5)_A G_A^{(0)}(\bar{x}_1 y_1 \mid \mathbf{\Lambda}) \times [\boldsymbol{\vartheta}_A\cdot\boldsymbol{\gamma}_\mu^A \partial_\mu^{y_1} \mathbf{\Lambda}(y_1)] G_B(\bar{x}_2 \bar{y}_2 \mid \mathbf{J}'_\beta)\mid_0. \tag{16}$$

Inserting (14) into (16), one obtains

$$T^{(00)}(x; y) = -i(g/\sqrt{2})[\gamma_\mu(1 + \gamma_5)]_A \delta(x_1 - y_1) \times [1 + ig\sqrt{2}\boldsymbol{\vartheta}_A\cdot\mathbf{\Lambda}(x_1)]^{-1} \times (\boldsymbol{\vartheta}_A\cdot\partial_\mu^{x_1}\mathbf{\Lambda}(x_1)) G_B(\bar{x}_2\bar{y}_2 \mid \mathbf{J}'_\beta)\mid_0. \tag{17}$$

No further approximation is required for the evaluation of (17); this is carried out in Appendix C, where the result is shown to be

$$T_f^{(00)} = T_e^{(00)} = 0,$$

$$T_a^{(00)}(x) = B^g(x) I(f(x^2)). \tag{18}$$

Here, $B^g(x)$ denotes the gradient portion of the one-boson-exchange graph,

$$B^g(x) = -i\left(\frac{g}{M}\right)^2 \times [\gamma_\mu(1 + \gamma_5)]_A [\gamma_\nu(1 + \gamma_5)]_B \cdot \partial_\mu\partial_\nu \Delta_F(x), \tag{19}$$

and

$$I(f) = \frac{1}{2}\int_0^\infty\int_0^\infty d\xi\, d\eta e^{-(\xi+\eta)}[1 + \cosh(2f\xi^{\frac{1}{2}}\eta^{\frac{1}{2}})] \equiv \tfrac{1}{2} + \tfrac{1}{2}J(f), \tag{20}$$

where $f(x^2) = i(2g/M)^2\Delta_F(x)$. The Fourier transform of (18), between lepton spinors, represents the mass-shell scattering amplitude, and is, in this approximation, a function of momentum transfer only. From (20) one sees that only "half" of the gradient part of the one-boson-exchange term is affected.

The integral (20) exists and may be immediately evaluated when $-1 < \mathrm{Re}\, f < 1$, and the definition of $J(f)$, for arbitrary f, shall be taken as the corresponding continuation of the result of this integration. For this, we note that the continuation of $J(f)$ corresponding to large and real values of f vanishes according as[16] $J(f) \sim f^{-2} \ln |f|$, providing an explicit example of light-cone damping.

Application of the heuristic argument in the Appendix of Ref. 2, which may be expected to be valid in the absence of essential singularities, shows that the η-factor appropriate to this approximation is $\eta = \frac{1}{8}$ instead of the $\frac{1}{4}$ of Ref. 2; this is because of the peculiar decomposition of (20), where the damping occurs for $J(f)$ rather than for $I(f)$. The actual η value resulting from any such finite number of approximations may, of course, bear little relation to the factor defined by the infinite sequence of approximations which have not yet been considered.

IV. SUMMARY

While it is satisfying that the functional peratization method employed leads, in simplest approximation, to a damped amplitude, emphasis must again be placed upon the basic deficiency of all calculational methods, here and elsewhere, thus far proposed: one has no *a priori* way of estimating corrections.

ACKNOWLEDGMENTS

The author is indebted for stimulating conversations to far too many persons to list here; but special acknowledgment should be made to D. Yennie and G. Feinberg, for each correcting a different error in work closely related to this material.

APPENDIX A

The derivation of (2) from Schwinger's functional solution to the field equations will be briefly sketched here, using a simplified example of a scalar interaction, $L' = g\bar{\psi}w\psi$; in this way, the decomposition of the vector boson field into transverse and longitudinal parts, which is not relevant to the particular approximate result desired here, may be bypassed. The generating functional, from which all the unrenormalized and unamputated n-point Green's functions are constructed by functional differentiation, is given by

$$Z[\eta, \bar{\eta}, j] = \frac{1}{N} \exp\left(\frac{i}{2} j\Delta_F j\right) \cdot \exp\left(-\frac{i}{2}\frac{\delta}{\delta J}\Delta_F \frac{\delta}{\delta J}\right) \cdot \exp L[J] \cdot \exp(i\bar{\eta}G[J]\eta), \tag{A1}$$

where N is a normalization constant, $J(z) = \int \Delta_F(z - z')j(z')$, and $L = -\mathrm{Tr} \ln [G[J]\cdot G^{-1}[0]]$ represents the closed loop dependence. We are not interested in external W-lines (hence $j \to 0$), and all closed fermion loops are to be neglected ($L \to 0$, $N \to 1$). The scattering of two identical fermions is then given by

$$M(x; y) = \sum_{P=1}^{2} \exp\left(-\frac{i}{2}\frac{\delta}{\delta J}\Delta_F \frac{\delta}{\delta J}\right) \times G(x_1 y_1 \mid J) G(x_2 y_2 \mid J)|_{J\to 0}, \tag{A2}$$

where $\sum_{P=1}^{2}$ denotes a sum on the two permutations of coordinate pairs. If we are dealing with two distinguishable particles, say A and B, the permutation sum may be omitted,

$$M(x; y) = \exp\left(-\frac{i}{2}\frac{\delta}{\delta J}\Delta_F \frac{\delta}{\delta J}\right) \times G_A(x_1 y_1 \mid J) G_B(x_2 y_2 \mid J)|_0, \tag{A3}$$

or

$$M = \left[\exp\left(-\frac{i}{2}\frac{\delta}{\delta J}\Delta_F \frac{\delta}{\delta J}\right) G_A\right] \times \exp\left(-i\frac{\overleftarrow{\delta}}{\delta J}\Delta_F \frac{\vec{\delta}}{\delta J}\right) \times \left[\exp\left(-\frac{i}{2}\frac{\delta}{\delta J}\Delta_F \frac{\delta}{\delta J}\right) G_B\right]\Bigg|_0. \tag{A4}$$

All the lepton self-energy and vertex-type structure is generated by the exponential operators inside the square brackets of (A4), and the neglect of this structure then provides the simple amplitude,

$$M(x; y) = G_A(x_1 y_1 \mid J) \exp\left(-i\frac{\overleftarrow{\delta}}{\delta J}\Delta_F \frac{\vec{\delta}}{\delta J}\right) \times G_B(x_2 y_2 \mid J)|_0. \tag{A5}$$

The replacement of J by J_μ, and Δ_F by $\Delta_F^{\mu\nu}$ in (A5), produces (2). We follow Ref. 2 in using the symbol Δ_F; this function has also been called Δ_c.

APPENDIX B

In a hypothetical neutral vector meson theory of weak interactions, one could write

$$\partial_\mu^z \frac{\delta}{\delta J_\mu(z)} G(xy \mid J_a) = -ig(1 + \gamma_5) \times [\delta(x - z) - \delta(y - z)] G(xy \mid J_a), \tag{B1}$$

[16] Similar damping can be constructed for the allowed amplitude if one merely multiples the g^{2n} peratized n-rung ladder graph contribution of Ref. 2 by the number of graphs permitted in that order, and sums over all odd n. I am indebted to Professor A. Pais for some correspondence on this point. Identical results can be obtained from a modified Bloch–Nordsieck functional model.

in obvious analogy to Zumino's equation in electrodynamics, and this could then be immediately used to extract all the dependence coming from the gradient portion of the boson propagator. The corresponding gauge statement here is more complicated, but does merit a brief discussion. One may first consider the Dirac equation in the presence of an arbitrary, external, c-number source, $\mathbf{J}_\mu$. Solutions to the equation

$$[\gamma_\mu\partial_\mu^x - i(g/\sqrt{2})\gamma_\mu(1+\gamma_5)\boldsymbol{\delta}\cdot\mathbf{J}_\mu(x)] \times \psi(x \mid \mathbf{J}_\alpha) = 0 \qquad \text{(B2)}$$

possess the following gauge property: under the transformation

$$\mathbf{J}_\mu(x) \to \chi_\mu(x) + \boldsymbol{\varepsilon}\partial_\mu\Lambda(x), \qquad \text{(B3)}$$

where $\boldsymbol{\varepsilon}$ is a constant Paisotopic vector, $\Lambda(x)$ a Lorentz and Paisotopic scalar,

$$\psi(x \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda) = \exp[i(g/\sqrt{2})(1+\gamma_5)\boldsymbol{\delta}\cdot\boldsymbol{\varepsilon}\Lambda(x)]\psi(x \mid \mathbf{J}_\alpha). \qquad \text{(B4)}$$

The compatibility of (B1), (B2), (B3) provides the definition of χ_μ; it must be chosen such that

$$\boldsymbol{\delta}\cdot\chi_\mu(z) = \exp[i(g/\sqrt{2})(1+\gamma_5)\boldsymbol{\delta}\cdot\boldsymbol{\varepsilon}\Lambda(x)]\boldsymbol{\delta}\cdot\mathbf{J}_\mu(x) \times \exp[-i(g/\sqrt{2})(1+\gamma_5)\boldsymbol{\delta}\cdot\boldsymbol{\varepsilon}\Lambda(x)], \qquad \text{(B5)}$$

or

$$\chi_\mu = [\mathbf{J}_\mu - \boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}\cdot\mathbf{J}_\mu)/\boldsymbol{\varepsilon}^2] \times \cos(\sqrt{2}g(1+\gamma_5)\Lambda(\boldsymbol{\varepsilon}^2)^{\frac{1}{2}}) + \frac{\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}\cdot\mathbf{J}_\mu)}{\boldsymbol{\varepsilon}^2} + \frac{\mathbf{J}_\mu \times \boldsymbol{\varepsilon}}{(\boldsymbol{\varepsilon}^2)^{\frac{1}{2}}} \times \sin[\sqrt{2}g(1+\gamma_5)\Lambda(\boldsymbol{\varepsilon}^2)^{\frac{1}{2}}]. \qquad \text{(B6)}$$

The gauge property (B4) is reproduced in the corresponding lepton Green's function,

$$G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda) = \exp[i(g/\sqrt{2})(1+\gamma_5)\boldsymbol{\delta}\cdot\boldsymbol{\varepsilon}\Lambda(x)]G(xy \mid \mathbf{J}_\alpha) \times \exp[-i(g/\sqrt{2})(1-\gamma_5)\boldsymbol{\delta}\cdot\boldsymbol{\varepsilon}\Lambda(y)]. \qquad \text{(B7)}$$

Using the relation

$$\sum_{i=1}^{3}\epsilon_i\frac{\delta\chi_\nu^{(i)}(z')}{\delta J_\mu^{(i)}(z)} = \epsilon_i\delta_{\mu\nu}\delta(z-z'),$$

it follows that

$$\boldsymbol{\varepsilon}\cdot[\delta/\delta\mathbf{J}_\mu(z)]G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda) = \boldsymbol{\varepsilon}\cdot[\delta/\delta\chi_\mu(z)]G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda), \qquad \text{(B8)}$$

and with (B8), the 4-divergence of the functional derivative of (B7) then gives

$$\begin{aligned}\boldsymbol{\varepsilon}\cdot[\partial_\mu^z\delta/\delta\mathbf{J}_\mu(z)]G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda) &= -[\partial/\partial\Lambda(z)]G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda)\\ &= -\left[\frac{\delta}{\delta\Lambda(z)} - \int dz'\,\frac{\delta\chi_\mu(z')}{\delta\Lambda(z)}\cdot\frac{\delta}{\delta\chi_\mu(z')}\right]\\ &\quad\times G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda),\end{aligned} \qquad \text{(B9)}$$

where $\partial/\partial\Lambda(z)$ denotes differentiation of only the explicit Λ dependence, not the implicit Λ dependence of χ_α. From (B6) there follows

$$\delta\chi_\mu(z')/\delta\Lambda(z)|_{\Lambda\to 0} = \sqrt{2}g(1+\gamma_5)\delta(z-z')(\mathbf{J}_\mu(z) \times \boldsymbol{\varepsilon}),$$

which, when substituted into (B9), provides the relation

$$\begin{aligned}&\boldsymbol{\varepsilon}\cdot\partial_\mu^z\frac{\delta}{\delta\mathbf{J}_\mu(z)}G(xy \mid \mathbf{J}_\alpha)\\ &= -[\delta/\delta\Lambda(z)]G(xy \mid \chi_\alpha + \boldsymbol{\varepsilon}\partial_\alpha\Lambda)|_{\Lambda\to 0}\\ &\quad - \sqrt{2}g(1+\gamma_5)\boldsymbol{\varepsilon}\cdot(\mathbf{J}_\mu(z) \times \delta/\delta\mathbf{J}_\mu(z))G(xy \mid \mathbf{J}_\alpha).\end{aligned} \qquad \text{(B10)}$$

The first term on the right side of (B10) may be evaluated using (B7), and the result, for arbitrary $\boldsymbol{\varepsilon}$, is

$$\begin{aligned}\partial_\mu^z[\delta/\delta J_\mu^{(i)}(z)]G(xy \mid \mathbf{J}_\alpha) &= -i(g/\sqrt{2})(1+\gamma_5)\\ &\quad\times\{\delta(x-z)\sigma_i G(xy \mid \mathbf{J}_\alpha) - \delta(z-y)G(xy \mid \mathbf{J}_\alpha)\sigma_i\}\\ &\quad + \sqrt{2}g(1+\gamma_5)\sum_{j,k}\epsilon_{ijk}J_\mu^{(k)}(z)\frac{\delta}{\delta J_\mu^{(j)}(z)}G(xy \mid \mathbf{J}_\alpha),\end{aligned} \qquad \text{(B11)}$$

which is the appropriate generalization of (B1).

The derivation of (B11) is most easily obtained from (3), and the adjoint form of (3), but the method used here serves to illustrate the underlying gauge structure. The arithmetical reason for the relative complexity of (B11) is, of course, the noncommutivity of the Pauli matrices; the physical reason is the existence of charge conservation selection rules which remove certain graphs and prevent those which remain from all being added coherently.[17]

APPENDIX C

The evaluation of $T^{(00)}$ proceeds as follows. The damping denominator of (17) is rationalized and put into exponential form with the aid of the rep-

[17] This discussion is essentially a zero-lepton-mass version of general arguments given by C. N. Yang and R. I. Mills, Phys. Rev. 96, 191 (1954).

resentation

$$[1 + i\sqrt{2}g\mathfrak{d}_A\cdot\mathbf{\Lambda}(x_1)]^{-1} = \frac{1}{\pi}\int_0^\infty d\xi\, e^{-\xi} \times \iint_{-\infty}^{+\infty} d\alpha_1\, d\alpha_2 \left[1 - \frac{\mathfrak{d}_A\cdot\boldsymbol{\alpha}}{\sqrt{\xi}}\right] \times \exp\,[-(\alpha_1^2 + \alpha_2^2) + 2ig(2\xi)^{\frac{1}{4}}\boldsymbol{\alpha}\cdot\mathbf{\Lambda}(x_1)], \quad \text{(C1)}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ and $\mathbf{\Lambda}(x_1) = (\Lambda^{(1)}(x_1), \Lambda^{(2)}(x_1))$ is given by (9). The combination

$$\sum_{i=1}^{2} \sigma_i^A \gamma_\mu^A \partial_\mu^{x_1} \Lambda^{(i)}(x_1) G_B(\bar{x}_2 \bar{y}_2 \mid \mathrm{J}'_\beta)$$

is, using

$$[\delta/\delta J_\lambda^{(i)}(z)]G(xy \mid \mathrm{J}_\alpha) = i(g/\sqrt{2})(1 + \gamma_5)G(xz \mid \mathrm{J}_\alpha)\gamma_\lambda\sigma_i G(zy \mid \mathrm{J}_\alpha),$$

replaced by

$$\frac{g}{\sqrt{2}M^2}(1 + \gamma_5)_B \sum_{i=1}^{2} \sigma_i^A \gamma_\mu^A \int \partial_\mu^{x'}\partial_\nu^{z'}\Delta_F(x_1 - z_2) \times G_B(\bar{x}_2 z_2 \mid \mathrm{J}'_\beta)\gamma_\nu^B\sigma_i^B G_B(z_2\bar{y}_2 \mid \mathrm{J}'_\beta), \quad \text{(C2)}$$

and the J'_β dependence of (C2) is translated by the operator

$$\exp\,[2ig(2\xi)^{\frac{1}{4}}\boldsymbol{\alpha}\cdot\mathbf{\Lambda}(x_1)] = \exp\left[\frac{2g(2\xi)^{\frac{1}{4}}}{M^2}\int \partial_\sigma^{z'}\Delta_F(x_1 - z)\boldsymbol{\alpha}\cdot\frac{\delta}{\delta \mathrm{J}'_\sigma(z)}\right].$$

In the limit $\mathrm{J}'_\beta \to 0$, one has then effectively replaced the J'_β dependence according to the relation

$$\mathrm{J}'_\sigma(z) \to \frac{2g(2\xi)^{\frac{1}{4}}}{M^2}\partial_\sigma^{z'}\Delta_F(x_1 - z)\boldsymbol{\alpha} \equiv \boldsymbol{\alpha}\partial_\sigma^{z'}\phi(z), \quad \text{(C3)}$$

where we have suppressed the x_1 dependence of ϕ.

But, as is clear from the work of Appendix B, the lepton Green's functions are trivial in the presence of a source such as that (C3),

$$G(xy \mid \boldsymbol{\alpha}\partial\phi) = \exp\,[i(g/\sqrt{2})(1 + \gamma_5)\mathfrak{d}\cdot\boldsymbol{\alpha}\phi(x)] \times S_F(x - y)\exp\,[-i(g/\sqrt{2})(1 - \gamma_5)\mathfrak{d}\cdot\boldsymbol{\alpha}\phi(y)], \quad \text{(C4)}$$

where (C4) involves, of course, only the difference $\phi(x) - \phi(y)$. Hence the functionally translated product of the two partially amputated Green's functions of (C2) is given as

$$\gamma_\mu^B\partial_\mu^{z'}\{\exp\,[4i(g^2/M^2)\xi^{\frac{1}{4}}\Delta_F(x_1 - x_2)\mathfrak{d}_B\cdot\boldsymbol{\alpha}]S_F^B(x_2 - z_2) \times \exp\,[-4i(g^2/M^2)\xi^{\frac{1}{4}}\Delta_F(x_1 - z_2)\mathfrak{d}_B\cdot\boldsymbol{\alpha}]\gamma_\nu^B\sigma_i^B \times \exp\,[4i(g^2/M^2)\xi^{\frac{1}{4}}\Delta_F(x_1 - z_2)\mathfrak{d}_B\cdot\boldsymbol{\alpha}]S_F^B(z_2 - y_2) \times \exp\,[-4i(g^2/M^2)\xi^{\frac{1}{4}}\Delta_F(x_1 - y_2)\mathfrak{d}_B\cdot\boldsymbol{\alpha}]\}\gamma_\lambda^B\overleftarrow{\partial}_\lambda^{y'}, \quad \text{(C5)}$$

where the exponential $(1 + \gamma_5)$ factors have been replaced by 2. From Zumino's work we again expect the mass-shell contribution of (C5) to be that of

$$-\delta(x_2 - z_2)\delta(z_2 - y_2)\gamma_\nu^B \times \exp\,[-4i(g^2/M^2)\xi^{\frac{1}{4}}\mathfrak{d}_B\cdot\boldsymbol{\alpha}\Delta_F(x_1 - z_2)]\sigma_i^B. \times \exp\,[4i(g^2/M^2)\xi^{\frac{1}{4}}\mathfrak{d}_B\cdot\boldsymbol{\alpha}\Delta_F(x_1 - z_2)], \quad \text{(C6)}$$

and the Paisotopic dependence of (C6) can easily be reduced,

$$e^{\mathfrak{d}\cdot\mathbf{a}}\mathfrak{d}\cdot\mathbf{b}e^{-\mathfrak{d}\cdot\mathbf{a}} = \left[\mathfrak{d}\cdot\mathbf{b} - \frac{(\mathfrak{d}\cdot\mathbf{a})(\mathbf{a}\cdot\mathbf{b})}{a^2}\right] \times \cosh(2a) + \frac{(\mathfrak{d}\cdot\mathbf{a})(\mathbf{a}\cdot\mathbf{b})}{a^2} + i\mathfrak{d}\cdot(\mathbf{a}\times\mathbf{b})\frac{\sinh(2a)}{a}, \qquad a = [\mathbf{a}^2]^{\frac{1}{2}},$$

as used in Appendix B.

Putting all of these factors together, and separating $T^{(00)}$ into its three Paisotopic amplitudes, one finds that the $T_{f,e}^{(00)}$ contributions vanish, while the allowed amplitude receives the contribution

$$T_a^{(00)}(x; y) = i(g^2/M^2)[\gamma_\mu(1 + \gamma_5)]_A[\gamma_\nu(1 + \gamma_5)]_B \times \delta(x_1 - y_1)\delta(x_2 - y_2) \times I(f(x_1 - x_2))\cdot\partial_\mu^{x'}\partial_\nu^{z'}\Delta_F(x_1 - x_2). \quad \text{(C7)}$$

In momentum space, (C7) leads to a function of momentum transfer given by the Fourier transform of (18). A representation for $J(f) = 2I(f) - 1$ is

$$J(f) = \int_0^1 v\,dv\,[1 - v^2]^{-\frac{1}{2}}\int_0^\infty u\,du\,e^{-u}\cosh(uvf), \quad \text{(C8)}$$

which is valid only in the strip $-1 < \mathrm{Re}\, f < 1$. Near the light cone, on the other hand,

$$f \sim -(g/\pi M)^2(r^2 - x_0^2)^{-1}.$$

Since in this limit $f \sim \pm\infty$, $J(f)$ is here defined by the appropriate continuation of the explicit (C8) integration. The latter is

$$J(f) = [1 - f^2]^{-1}\{1 + [f^{-2} - 1]^{-\frac{1}{2}} \times \tan^{-1}([f^{-2} - 1]^{-\frac{1}{2}})\}, \quad \text{(C9)}$$

and the continuation of (C9) into the light-cone region of f leads to the damping quoted in the text.

Off-Diagonal Long-Range Order and Generalized Bose Condensation*

M. D. GIRARDEAU
Institute of Theoretical Science, and Department of Physics, University of Oregon, Eugene, Oregon
(30 November 1964)

It is shown for boson systems with periodic boundary conditions that the existence of generalized Bose condensation, of which the simple type typified by the ideal Bose gas is a special case, is equivalent to the existence of off-diagonal long-range order (ODLRO) in the single-particle density matrix ρ_1 provided that the two noncomuting limits involved in the criterion for ODLRO are taken in the proper order: first size of system $\to \infty$, then interparticle separation $\to \infty$. It is shown by means of an example that certain assumptions concerning the behavior of the single-particle momentum distribution function involved in proving this equivalence are actually satisfied in thermal equilibrium for some dynamical models. The analysis is generalized to box-enclosure boundary conditions by an extension of an argument due to Schafroth, according to which box-enclosure conditions are equivalent to homogeneous Neumann conditions except for a thermodynamically negligible surface effect, provided that one second-quantizes with respect to Hartree–Fock orbitals rather than free-particle orbitals. A general criterion for generalized Bose condensation in terms of eigenvalues of ρ_1 is proposed. On the basis of the behavior of soluble models it is conjectured that for a boson system in thermal equilibrium subject to arbitrary boundary conditions, the existence of such Bose condensation is equivalent to the existence of ODLRO in ρ_1. The two-particle density matrix ρ_2 is discussed briefly. By means of a simplified model it is shown that for a Bose system generalized condensation implies large eigenvalues of ρ_2 and ODLRO of ρ_2, just as for ρ_1. It is pointed out that the question of the existence or nonexistence of generalized Bose condensation of fermion pairs ought to be investigated.

1. INTRODUCTION

THE concept of off-diagonal long-range order (ODLRO) was invented by Penrose[1] in 1951 in an attempt to explain the anomalous hydrodynamics of helium II, extended by Penrose and Onsager[2] in 1956, and recently given a general formulation and generalization by Yang[3] as a criterion for both superfluidity and superconductivity. We wish here to discuss one aspect of this concept, in particular the relationship between ODLRO and Bose–Einstein condensation. We shall show that the criterion for ODLRO is compatible with Bose–Einstein condensation of a more general form than the simple type typified by the ideal Bose gas.

2. DEFINITION OF ODLRO

The coordinate representative of the single-particle density operator ρ_1 of a system of N identical bosons is defined in the thermal-equilibrium case as

$$\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle = \mathrm{Tr}\,[\psi(\mathbf{x})\rho\psi^\dagger(\mathbf{x}')], \tag{1}$$

where ρ is the normalized canonical density operator, ψ and $\psi^\dagger$ are the usual Bose field operators, and the trace runs over a complete set of N-particle states. An equivalent definition in the Schrödinger representation is

$$\langle \mathbf{x}\,|\rho_1|\,\mathbf{x}'\rangle = \frac{N}{Z}\sum_\alpha e^{-\beta E_\alpha} \times \int \psi_\alpha(\mathbf{x}\mathbf{x}_2\cdots\mathbf{x}_N)\psi_\alpha^*(\mathbf{x}'\mathbf{x}_2\cdots\mathbf{x}_N)\,dx_2^3\cdots d^3x_N, \tag{2}$$

where ψ_α and E_α are the energy eigenfunctions and eigenvalues, and Z is the partition function $\sum_\alpha e^{-\beta E_\alpha}$. We define ODLRO to be present in ρ_1 if and only if

$$\lim_{|\mathbf{x}-\mathbf{x}'|\to\infty} \lim \mathrm{therm}\,\langle \mathbf{x}\,|\rho_1|\,\mathbf{x}'\rangle \neq 0 \tag{3}$$

where "lim therm" means "thermodynamic limit" and is defined for any function of N, the number of particles, and Ω, the volume of the system, by

$$\lim \mathrm{therm}\, f(N,\Omega) \equiv \lim_{\substack{N\to\infty \\ \Omega\to\infty \\ (N/\Omega)\to\rho,\,0<\rho<\infty}} f(N,\Omega) \tag{4}$$

if this limit exists. The precise meaning of (3) is as follows: If the prescribed double limit of $\langle \mathbf{x}|\ \rho_1\ |\mathbf{x}'\rangle$ exists and is equal to zero, then ODLRO is *not* present; if the limit exists and is not zero *or* if it does not exist,[4] then ODLRO *is* present.

The operation lim therm is not explicit in the definitions given by Penrose and Onsager[1,2] and Yang.[3] Nevertheless, it is implicit in their applications of the definition, as is shown by the fact that

* Supported in part by the Public Health Service (GM 09153-03).

[1] O. Penrose, Phil. Mag. **42**, 1373 (1951).

[2] O. Penrose and L. Onsager, Phys. Rev. **104**, 576 (1956), in particular Sec. 4.

[3] C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962) and J. Math. Phys. **4**, 418 (1963). Also J. S. Bell, Phys. Letters **2**, 116 (1962).

[4] This case is relevant to Bose–Einstein condensation into a single-particle state of nonzero and macroscopic momentum. We assume, however, that the thermodynamic limit of $\langle \mathbf{x}\,|\rho_1|\,\mathbf{x}'\rangle$ does exist (before letting $|\mathbf{x}-\mathbf{x}'|\to\infty$).

for a system with periodic boundary conditions, $\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle$ is necessarily triply periodic in both $\mathbf{x}$ and $\mathbf{x}'$ with periods equal to the edges of the periodicity cube Ω; inclusion of the operation lim therm is necessary to remove the boundary-condition effects. It is furthermore clear that the order of the limits in (3) is crucial. If they are interchanged, the double limit never exists for periodic boundary conditions, and is always zero for box-enclosure boundary conditions. If the limits are taken in the order (3) it is easy to show[5] that an ideal Bose gas possesses ODLRO below its condensation temperature T_c but no ODLRO above T_c.

3. EQUIVALENCE OF ODLRO AND GENERALIZED BOSE CONDENSATION

Penrose and Onsager[1,2] and Yang[3] showed that if ρ_1 possesses an eigenvalue of order N, as in the case of the ideal Bose gas below T_c, then ODLRO is necessarily present. However, we shall show that the converse is not true: ODLRO is possible even when ρ_1 has no eigenvalue of order N, although a more general type of Bose condensation must still be present. To this end, consider the case of periodic boundary conditions. Then ρ_1 is diagonal in momentum representation, so that

$$\langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle = \Omega^{-1} \sum_{\mathbf{k}} n_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}, \tag{5a}$$

$$n_{\mathbf{k}} = \langle \mathbf{k} |\rho_1| \mathbf{k}\rangle = \mathrm{Tr}(a_{\mathbf{k}} \rho_1 a_{\mathbf{k}}^{\dagger}). \tag{5b}$$

Suppose the dependence of $n_{\mathbf{k}}$ on N and Ω to be such that[6]

$$\lim_{k_0 \to 0} \lim \mathrm{therm}\ \Omega^{-1} \sum_{k<k_0} n_{\mathbf{k}} = \rho_c > 0, \tag{6}$$

$$\lim \mathrm{therm}\ n_{\mathbf{k}} = \mathfrak{N}(\mathbf{k}) < \infty, \qquad \mathbf{k} \neq 0.$$

Breaking the sum (5) into two parts, one then has[7]

$$\lim \mathrm{therm}\ \langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle = \lim \mathrm{therm}\ \Omega^{-1} \sum_{k<k_0} n_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} + (2\pi)^{-3} \int_{k>k_0} \mathfrak{N}(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\, d^3k. \tag{7}$$

Since the left side of (7) is independent of k_0, and k_0 is independent of N and Ω, one finds[8] on letting $k_0 \to 0$

$$\lim \mathrm{therm}\ \langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle = \rho_c + \langle \mathbf{x} |\rho_1'| \mathbf{x}'\rangle,$$

$$\langle \mathbf{x} |\rho_1'| \mathbf{x}'\rangle \equiv (2\pi)^{-3}\, \mathrm{P} \int \mathfrak{N}(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\, d^3k, \tag{8}$$

where P denotes the principal value of the integral, evaluated by excluding a sphere of radius k_0 about the origin and letting $k_0 \to 0$.[9] If the limit function $\mathfrak{N}(k)$ is sufficiently well behaved in the neighborhood of $\mathbf{k} = 0$, e.g., if it is spherically symmetric and of order k^{-p} with $p < 3$, then $\langle \mathbf{x}| \rho' |\mathbf{x}'\rangle$ will vanish as $|\mathbf{x} - \mathbf{x}'| \to \infty$, and one will have

$$\lim_{|\mathbf{x}-\mathbf{x}'|\to\infty} \lim \mathrm{therm}\ \langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle = \rho_c > 0. \tag{9}$$

Thus the criterion (3) for ODLRO will be satisfied. Although the case of simple Bose condensation [$n_0 = O(N)$ and $n_{\mathbf{k}} = 0(1)$ for $\mathbf{k} \neq 0$] is the simplest way of satisfying (5), it is not the only way. The top Eq. (5) is just the criterion for generalized Bose condensation[10]; it states, crudely speaking, that a nonzero fraction of the total number of particles in the system have momenta less than any macroscopic momentum. We conclude, therefore, that the existence of generalized Bose condensation (which includes simple Bose condensation as a special case) implies the existence of ODLRO in ρ_1 provided that

---

[5] See, e.g., C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962), Sec. 13.

[6] The precise meaning of the second line of (6) is as follows: Let $\mathbf{k}$ be any fixed nonzero momentum vector. There are several allowed (by periodic boundary conditions) momentum vectors within a distance $2\pi\Omega_N^{-1/3}$ from $\mathbf{k}$, for each fixed value of N, where $\Omega_N = N/\rho$ with ρ the mean particle-number density. Pick one such vector $\mathbf{q}_N(\mathbf{k})$. Then we require that the limit

$$\lim_{N\to\infty} n_{\mathbf{q}_N(\mathbf{k})} = \mathfrak{N}(\mathbf{k}) < \infty$$

exist and that it be independent of the particular choice of the sequence $\{\mathbf{q}_N(\mathbf{k})\}$, all such sequences having $\mathbf{k}$ as limit. In general the function $n_{\mathbf{q}_N}$ will depend on N not only explicitly via $\mathbf{q}_N$, but will also have an implicit dependence on N, and also on $\Omega = N/\rho$ because of the boundary conditions.

[7] We assume that the approach of $n_{\mathbf{k}}$ to its limit $\mathfrak{N}(\mathbf{k})$ [second line of (6)] is uniform in the domain $k > k_0$; it then follows that

$$\lim \mathrm{therm}\ \Omega^{-1} \sum_{k>k_0} n_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} = (2\pi)^{-3} \int_{k>k_0} \mathfrak{N}(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\, d^3k$$

provided that the limit function $\mathfrak{N}(\mathbf{k})$ is integrable in this domain, which we also assume.

[8] In order to see that

$$\lim_{k_0\to 0} \lim \mathrm{therm}\ \Omega^{-1} \sum_{k<k_0} n_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} = \rho_c,$$

note that for $k < k_0$, one has by an elementary geometrical argument

$$e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} = 1 + \mathfrak{R}_{\mathbf{k}}(\mathbf{x} - \mathbf{x}'),$$

$$|\mathfrak{R}_{\mathbf{k}}(\mathbf{x} - \mathbf{x}')| \leq |\mathbf{k}\cdot(\mathbf{x} - \mathbf{x}')| \leq k_0\, |\mathbf{x} - \mathbf{x}'|.$$

Thus

$$\lim \mathrm{therm}\ \Omega^{-1} \sum_{k<k_0} n_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} = [1 + \mathfrak{R}(k_0, \mathbf{x} - \mathbf{x}')] \lim \mathrm{therm}\ \Omega^{-1} \sum_{k<k_0} n_{\mathbf{k}},$$

$$|\mathfrak{R}(k_0, \mathbf{x} - \mathbf{x}')| \leq k_0\, |\mathbf{x} - \mathbf{x}'|.$$

On taking the limit $k_0 \to 0$, the contribution from the term proportional to $\mathfrak{R}$ vanishes provided that the limit in the top line of (6) exists, so that the desired result follows immediately.

[9] In most cases the limit function $\mathfrak{N}(\mathbf{k})$ will be such that the principal-value operation may be omitted without changing the value of the integral.

[10] M. Girardeau, Phys. Fluids **5**, 1468 (1962), Eq. (48) ff.

certain assumptions concerning the behavior of $n_{\mathbf{k}}$ in the thermodynamic limit are satisfied.

This relationship is strengthened by the observation that the converse is also true: ODLRO implies generalized Bose condensation, provided that the criterion for ODLRO is taken to be (9), rather than the more general criterion (3). As a matter of fact, the only obvious motivation for the more general criterion (3) is to allow treatment of simple or generalized condensation at $\mathbf{q} \neq 0$, in which case the thermodynamic limit of $\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle$ behaves like $\rho_0 e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')}$ as $|\mathbf{x} - \mathbf{x}'| \to \infty$; however, this case can be reduced to the case of $\mathbf{q} = 0$ [to which (9) applies] by a Galilean transformation. We thus assume the existence of the limit function $\mathfrak{N}(\mathbf{k})$ in the second line of (6) for $k \neq 0$. We furthermore assume the existence of the limit

$$\rho_0 = \lim_{k_0 \to 0} \lim \text{therm } \Omega^{-1} \sum_{k<k_0} n_{\mathbf{k}}, \tag{10}$$

but we do not assume initially that $\rho_0 > 0$, since this is what we wish to prove. Then, retracing the previous derivation, one concludes[7-9] that

$$\lim_{|\mathbf{x}-\mathbf{x}'|\to\infty} \lim \text{therm } \langle \mathbf{x}\, |\rho_1|\, \mathbf{x}'\rangle = \rho_0. \tag{11}$$

But at this stage we may assume $\rho_0 > 0$, since (9) (existence of ODLRO) is satisfied by hypothesis. Then by (10) one concludes that the criterion (6) for generalized Bose condensation is indeed satisfied.

4. A SIMPLE EXAMPLE

In order to see that there exist model many-boson systems for which the assumptions in the previous section concerning the behavior of $n_{\mathbf{k}}$ in the thermodynamic limit are actually realized, consider the Hamiltonian[11]

$$H = \sum_{\mathbf{k}} \tfrac{1}{2} k^2 a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'} a_{\mathbf{k}}^\dagger a_{\mathbf{k}'}^\dagger a_{\mathbf{k}'} a_{\mathbf{k}} + 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'}{}' a_{\mathbf{k}'}^\dagger a_{\mathbf{k}}^\dagger a_{\mathbf{k}'} a_{\mathbf{k}}. \tag{12}$$

This is obtained from the full Hamiltonian[12] of a boson system with pairwise interactions by replacing the interaction in $\mathbf{k}$ space by $4\pi\alpha$, where α is the two-body scattering length at zero energy,[13] and retaining only diagonal terms in the interaction Hamiltonian. The first $\sum_{\mathbf{k}\mathbf{k}'}$ represents the forward-scattering terms, obtained by choosing momentum transfer $\mathbf{q} = 0$ in the general interaction term $a_{\mathbf{k}+\mathbf{q}}^\dagger a_{\mathbf{k}'-\mathbf{q}}^\dagger a_{\mathbf{k}'} a_{\mathbf{k}}$ while $\sum'_{\mathbf{k}\mathbf{k}'}$ represents exchange scattering, arising from terms with $\mathbf{q} = \mathbf{k}' - \mathbf{k}$, with $\mathbf{k} \neq \mathbf{k}'$ since the $\mathbf{k} = \mathbf{k}'$ terms are already included in the forward scattering. Using the Bose commutation relations, one can rewrite (12) as a function only of the occupation-number operators $N_{\mathbf{k}} = a_{\mathbf{k}}^\dagger a_{\mathbf{k}}$. Noting that the first $\sum_{\mathbf{k}\mathbf{k}'}$ can then be replaced by its eigenvalue $N(N-1)$ when acting on eigenstates of the total-particle-number operator $\sum_{\mathbf{k}} N_{\mathbf{k}}$ with eigenvalue N, one finds

$$H = 2\pi(N-1)\rho\alpha + \sum_{\mathbf{k}} \tfrac{1}{2} k^2 N_{\mathbf{k}} + 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'}{}' N_{\mathbf{k}} N_{\mathbf{k}'}, \tag{13}$$

provided that we restrict ourselves to N-particle states; here $\rho = N/\Omega$, the mean particle-number density.

If α is positive then it is trivial to show that the N-particle ground state of H has $N_0 = N$ and $N_{\mathbf{k}} = 0$, $\mathbf{k} \neq 0$, i.e., complete simple Bose condensation; the ground-state energy is then $2\pi(N-1)\rho\alpha$. Let us consider, however, the case $\alpha < 0$. Then, because of the terms involving $N_{\mathbf{k}} N_{\mathbf{k}'}$ in (13), one can obtain a state of lower energy by removing a nonzero fraction of the particles from zero momentum and placing them in the adjacent allowed momentum sites, thus allowing them to contribute a negative interaction energy,[14] since $\alpha < 0$. This will lower the potential energy by an amount of order $N\rho\,|\alpha|$; on the other hand, the kinetic energy will only be raised by an amount of order $\rho\Omega^{\frac{1}{3}}$ since the allowed momenta nearest $\mathbf{k} = 0$ are only a few times $2\pi\Omega^{-\frac{1}{3}}$. More precisely, one sees from (13) that a state in which (N/l) particles are placed into each of l allowed momentum sites nearest the origin has energy

$$2\pi N\rho\alpha + 2\pi N\rho\alpha(1 - l^{-1}) + o(N), \tag{14}$$

where $o(N)$ has the usual meaning [$o(N) < O(N)$, i.e., lim therm $N^{-1}o(N) = 0$]; we have assumed in (14) that $l = o(N)$. Since $\alpha < 0$, lower energy is obtained for larger l, and for large l the energy approaches the value

$$E_0 = -4\pi N\rho\,|\alpha| + o(N). \tag{15}$$

This is in fact the true N-particle ground-state energy of H for fixed ρ, since the kinetic energy

[11] The zero-temperature properties of this model were discussed previously: M. Girardeau, Boeing Scientific Research Laboratories Document D1-82-0119 (June 1961). We repeat the analysis here for completeness.

[12] We use units such that $\hbar = m = 1$.

[13] This is the Fermi pseudopotential approximation, which is sufficient if one wants results correct only to first order in α; we shall limit ourselves here to this approximation.

[14] It is crucial to this argument that the terms with $\mathbf{k} = \mathbf{k}'$ in (13) are excluded, being already included in the c-number term; thus at least two allowed momentum sites must be occupied in order for $\sum'_{\mathbf{k}\mathbf{k}'}$ to contribute.

is positive semidefinite and the $N_{\mathbf{k}}N_{\mathbf{k}'}$ contributions to (15) have actually attained their lower bound [to $O(N)$] given by

$$2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'}{}' N_{\mathbf{k}}N_{\mathbf{k}'} > 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'} N_{\mathbf{k}}N_{\mathbf{k}'} = -2\pi N\rho\,|\alpha|. \quad (16)$$

Thus, although the state in which a large number of allowed momentum sites nearest the origin have *equal* occupations is certainly not the true ground state, its energy only differs from that of the true ground state by a thermodynamically negligible amount $o(N)$.

In order to find the true ground state, we note that the upper limit on l in (14) is set by the increase of kinetic energy with increasing l. Thus, although the amount of "smearing" of the condensation, i.e., the size of the region about $\mathbf{k} = 0$ in which $n_{\mathbf{k}}$ is large, is so small that the kinetic energy is $o(N)$ (thermodynamically negligible) and the ground-state energy to $O(N)$ is independent of the details of the smearing, nevertheless the kinetic energy must be included in determining these details. The energy eigenvalue of an arbitrary N-particle eigenstate of (13) in which the $N_{\mathbf{k}}$ have eigenvalues $n_{\mathbf{k}}$ (non-negative integers) is

$$E(\{n_{\mathbf{k}}\}) = 2\pi(N-1)\rho\alpha + \sum_{\mathbf{k}} \tfrac{1}{2}k^2 n_{\mathbf{k}} + 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'}{}' n_{\mathbf{k}}n_{\mathbf{k}'}. \quad (17)$$

Since we have already seen that the ground state is characterized by the $n_{\mathbf{k}}$ being very large in an infinitesimal neighborhood of $\mathbf{k} = 0$ but zero outside this region, we may treat the $n_{\mathbf{k}}$ as continuous variables in determining the minimizing set $\{n_{\mathbf{k}}\}$ provided that we require $n_{\mathbf{k}} \geq 0$. Then, minimizing (17) with respect to $n_{\mathbf{k}}$ subject to the subsidiary condition

$$\sum_{\mathbf{k}} n_{\mathbf{k}} = N, \quad (18)$$

one obtains the condition

$$\tfrac{1}{2}k^2 - \mu + 4\pi\Omega^{-1}\alpha \sum_{\mathbf{k}'}{}' n_{\mathbf{k}'} = 0, \quad (19)$$

where μ, the chemical potential, is the Lagrange multiplier corresponding to (18). But

$$\sum_{\mathbf{k}'}{}' n_{\mathbf{k}'} = N - n_{\mathbf{k}}, \quad (20)$$

so that

$$\tfrac{1}{2}k^2 - \mu + 4\pi\rho\alpha - 4\pi\Omega^{-1}\alpha n_{\mathbf{k}} = 0, \quad (21)$$

with solution[15]

$$n_{\mathbf{k}} = \Omega(\tfrac{1}{2}k^2 - \mu + 4\pi\rho\alpha)/4\pi\alpha \quad \text{if } \tfrac{1}{2}k^2 - \mu + 4\pi\rho\alpha > 0, \quad (22)$$
$$= 0 \quad \text{if } \tfrac{1}{2}k^2 - \mu + 4\pi\rho\alpha < 0.$$

To determine μ, we use the condition (18), hence, replacing the sum by an integral,

$$\frac{\Omega}{(2\pi)^3}\frac{\Omega}{4\pi\alpha}4\pi\int_0^{k_s} k^2\left(\frac{1}{2}k^2 - \mu + 4\pi\rho\alpha\right)dk = N, \quad (23)$$

where

$$\tfrac{1}{2}k_s^2 - \mu + 4\pi\rho\alpha = 0, \quad (24)$$

or

$$k_s = [2(\mu - 4\pi\rho\alpha)]^{\frac{1}{2}}. \quad (25)$$

Performing the integration (23) and substituting, one finds, recalling that $\alpha < 0$ so that $\alpha = -|\alpha|$,

$$\mu + 4\pi\rho\,|\alpha| = \tfrac{1}{2}[15(2\pi)^3\rho\,|\alpha|/\Omega]^{2/5}. \quad (26)$$

Thus μ differs from $-4\pi\rho\,|\alpha|$ only by a term of order $\Omega^{-2/5}$. Substituting into (22), one finds

$$n_{\mathbf{k}} = \tfrac{1}{2}(k_s^2 - k^2)\Omega/4\pi\,|\alpha|, \quad k < k_s$$
$$= 0, \quad k > k_s, \quad (27)$$

where

$$k_s = [15(2\pi)^3\rho\,|\alpha|/\Omega]^{1/5}. \quad (28)$$

Note that the ratio of k_s to the lattice spacing $2\pi\Omega^{-\frac{1}{3}}$ in $\mathbf{k}$ space becomes infinite in the thermodynamic limit, in spite of the fact that k_0 vanishes in the same limit.

It follows from (27) and (28) that the condition (6) for generalized Bose condensation is satisfied trivially; the condensate density ρ_0 is found to be equal to the total density ρ, while the limiting momentum distribution function $\mathfrak{N}(\mathbf{k})$ vanishes[6] for $\mathbf{k} \neq 0$. Thus the model (12), with $\alpha < 0$, leads to complete generalized Bose condensation at temperature $T = 0$.[16] The other assumptions[7,17] made in Sec. 3 concerning the behavior of $n_{\mathbf{k}}$ in the thermodynamic limit are also trivially satisfied for this model, so that ODLRO of ρ_1 does occur.

[15] Treating $n_{\mathbf{k}}$ as a continuous variable is only permissible for values of $\mathbf{k}$ such that the solution $n_{\mathbf{k}}$ of (21) is positive and very large (infinite in the thermodynamic limit). For values of $\mathbf{k}$ such that the solution of (21) is negative, one must take $n_{\mathbf{k}} = 0$ to obtain the ground state.

[16] For $\alpha > 0$ complete *simple* Bose condensation is present at $T = 0$; the ground-state energy for $\alpha > 0$ is $2\pi N\rho\alpha + o(N)$, whereas for $\alpha < 0$ it is $4\pi N\rho\alpha + o(N) = -4\pi N\rho|\alpha| + o(N)$. The nonanalytic dependence on α at $\alpha = 0$ is a very general feature of many-body problems.

[17] See Eqs. (8), (9), and the associated discussion.

The analysis of the model (12), (13) can be extended to nonzero temperature by Wentzel's method of the thermodynamically equivalent Hamiltonian.[18] To apply this method one decomposes the Hamiltonian (13) into an independent-particle Hamiltonian H_0 and a fluctuation Hamiltonian H_1:

$$H = H_0 + H_1,$$

$$H_0 = 2\pi(N-1)\rho\alpha - 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'}{}' n_{\mathbf{k}} n_{\mathbf{k}'} + \sum_{\mathbf{k}} \omega_{\mathbf{k}} N_{\mathbf{k}},$$

$$H_1 = 2\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\mathbf{k}'}{}' (N_{\mathbf{k}} - n_{\mathbf{k}})(N_{\mathbf{k}'} - n_{\mathbf{k}'}), \tag{29}$$

where the $n_{\mathbf{k}}$ are c-number parameters and

$$\omega_{\mathbf{k}} = \tfrac{1}{2}k^2 + 4\pi\Omega^{-1}\alpha \sum_{\mathbf{k}'}{}' n_{\mathbf{k}'} = \tfrac{1}{2}k^2 + 4\pi\rho\alpha - 4\pi\Omega^{-1}\alpha n_{\mathbf{k}}. \tag{30}$$

Then H_1 will not contribute to the thermodynamic functions in the thermodynamic limit[19] provided that the $n_{\mathbf{k}}$ are taken to be[18]

$$n_{\mathbf{k}} = \langle N_{\mathbf{k}} \rangle_0, \tag{31}$$

where $\langle\ \rangle_0$ denotes the average in the canonical ensemble determined by H_0. Since H_0 is the Hamiltonian of a system of independent bosons, the usual formula

$$n_{\mathbf{k}} = (e^{\beta(\omega_{\mathbf{k}} - \mu)} - 1)^{-1} \tag{32}$$

applies, where the chemical potential μ is determined by (18). In view of (30), Eq. (32) is really a transcendental equation for $n_{\mathbf{k}}$. However, one can obtain explicit results in certain limits, which provide all the information needed to evaluate the behavior in the thermodynamic limit. For sufficiently low temperatures one expects that $n_{\mathbf{k}}$ will become very large as $\mathbf{k} \to 0$, as in the case $T = 0$ [Eqs. (27), (28)]. Then $\beta(\omega_{\mathbf{k}} - \mu)$ must necessarily become very small, so that the exponential may be expanded, and

$$n_{\mathbf{k}} = \kappa T/(\omega_{\mathbf{k}} - \mu), \qquad k \to 0 \tag{33}$$

to leading order for large N and Ω. Inserting (30), one then has

$$n_{\mathbf{k}}(\tfrac{1}{2}k^2 + 4\pi\rho\alpha - \mu - 4\pi\Omega^{-1}\alpha n_{\mathbf{k}}) = \kappa T, \ k \to 0. \tag{34}$$

But since κT is independent of N and Ω, it is clear that $n_{\mathbf{k}}$ can only be of order $\Omega^{3/5}$ [as in (27)] if the expression in parentheses vanishes to leading order. Thus

$$n_{\mathbf{k}} = \tfrac{1}{2}(k_s^2 - k^2)\Omega/4\pi\,|\alpha|, \qquad k < k_s, \tag{35}$$

where k_s is related to μ, α, and Ω by (25). This is the same expression as was obtained for $T = 0$, except that now we shall find that k_s is temperature dependent. At $k = k_s$, $n_{\mathbf{k}}$ does not exactly vanish, since (35) is only correct to order $\Omega^{3/5}$; instead, one sees from (34) and (25) that

$$n_{k_s} = (\Omega\kappa T/4\pi\,|\alpha|)^{\frac{1}{2}}. \tag{36}$$

As k increases, $n_{\mathbf{k}}$ will continue to decrease monotonically. For k which is nonvanishing in the thermodynamic limit, one sees from (32) and (30) that $n_{\mathbf{k}}$ is given by an expression which differs from that for the ideal Bose gas only in the replacement of μ by $\mu + 4\pi\rho\,|\alpha|$. But in view of (25), and the fact that k_s is of order $\Omega^{-1/5}$, one sees that the limit function $\mathfrak{N}(\mathbf{k})$ [Eq. (6)] is just

$$\mathfrak{N}(\mathbf{k}) = (e^{\frac{1}{2}\beta k^2} - 1)^{-1}, \quad \mathbf{k} \neq 0, \quad T < T_c, \tag{37}$$

where T_c is the Bose condensation temperature, which will presently be determined. The limiting behavior (37) is the same as that for the ideal Bose gas.[20] The chemical potential μ is determined, as always, by (18). In view of the formal identity of (35) and the top Eq. (27), the number fN of particles in the condensate is obtained by replacing N by fN in the right side of (23), where f is the condensed fraction. Then (26) is replaced by

$$\mu + 4\pi\rho\,|\alpha| = \tfrac{1}{2}[15(2\pi)^3 f\rho\,|\alpha|/\Omega]^{2/5} \tag{38}$$

and ρ is replaced by $f\rho$ in (26). The number of uncondensed particles is determined by integrating the limiting momentum distribution (37):

$$(1-f)N = \frac{\Omega}{(2\pi)^3} 4\pi \int_0^\infty k^2 (e^{\frac{1}{2}\beta k^2} - 1)^{-1}\, dk. \tag{39}$$

Inserting the known[20] value for the integral, one finds

$$(1-f)\rho = (\kappa T/2\pi)^{\frac{3}{2}}\zeta(\tfrac{3}{2}) = 2.612(\kappa T/2\pi)^{\frac{3}{2}}, \tag{40}$$

where ζ is the Riemann zeta function. The condensation temperature T_c is that temperature at which f and k_0 [Eqs. (25) and (38)] vanish; hence, by (40),

$$T_c = (2\pi/\kappa)(\rho/2.612)^{\frac{2}{3}}. \tag{41}$$

Thus the condensation temperature T_c and the dependence of the condensed fraction f on T are the same as those of the ideal Bose gas; the only effect of the attractive interaction $\alpha < 0$ in the simple

[18] G. Wentzel, Phys. Rev. **120**, 1572 (1960).

[19] More precisely, the contribution of H_1 to intensive quantities will vanish in the thermodynamic limit, and its contribution to extensive quantities will be $o(\Omega)$[lim therm $\Omega^{-1}o(\Omega) = 0$].

[20] See, e.g., F. London, *Superfluids* (John Wiley & Sons, Inc., New York, 1954), Vol. II, pp. 40 ff. London's parameter α is of order N^{-1} for $T < T_c$.

model (12), (13) is to "smear" the condensation, but without changing the gross parameters f and T_c. Above the condensation temperature $n_\mathbf{k}$ is only of order unity for all values of $\mathbf{k}$ including $\mathbf{k} = 0$ and its neighborhood, so that (30) and (32) give

$$\mathfrak{N}(\mathbf{k}) = [e^{\beta(\frac{1}{2}k^2-\mu')} - 1]^{-1}, \quad T > T_c, \quad \text{all } \mathbf{k}, \tag{42}$$

where

$$\mu' = \mu + 4\pi\rho\,|\alpha|. \tag{43}$$

Furthermore, it is clear from the method of determination of μ that for given $T > T_c$, μ' will be equal to the chemical potential of the ideal Bose gas at the same temperature. Thus

$$\mathfrak{N}(\mathbf{k}, T) = \mathfrak{N}_0(\mathbf{k}, T), \qquad T > T_c, \tag{44}$$

where $\mathfrak{N}_0$ is the momentum distribution of the ideal Bose gas in the thermodynamic limit. Although these results might appear surprising in view of the interaction terms in (13), they are readily understandable when one notes that the single-particle energy levels $\omega_\mathbf{k}$ of the thermodynamically equivalent Hamiltonian H_0 [Eq. (29)] differ from those of an ideal Bose gas only by the constant shift $4\pi\rho\alpha = -4\pi\rho\,|\alpha|$, except in an infinitesimal region about $\mathbf{k} = 0$ for $T < T_c$.

One concludes, then, that at all temperatures $T < T_c$, the thermal equilibrium single-particle momentum distribution of the model (12), (13), with $\alpha < 0$,[21] satisfies the sufficient conditions given in Sec. 3 for the existence of ODLRO in ρ_1 due to generalized Bose condensation, the nonzero parameter ρ_c (the condensate density) in (9) being given by

$$\rho_c = f\rho = \rho[1 - (T/T_c)^{\frac{3}{2}}], \tag{45}$$

the same expression as for the ideal Bose gas. The essential difference between this model and the ideal Bose gas is seen by looking at the behavior of $\langle \mathbf{x}|\,\rho_1\,|\mathbf{x}'\rangle$ as a function of $\mathbf{x} - \mathbf{x}'$ for *large but finite* N and Ω. For the special case $T = 0$ one finds by (5) and (27)

$$\begin{aligned}\langle \mathbf{x}\,|\rho_1|\,\mathbf{x}'\rangle &= \tfrac{1}{2}(2\pi)^{-3}(\Omega/|\alpha|)\,|\mathbf{x} - \mathbf{x}'|^{-5} \\ &\times\ [6 - 2k_s^2\,|\mathbf{x} - \mathbf{x}'|^2]\sin(k_s\,|\mathbf{x} - \mathbf{x}'|) \\ &- 6k_s\,|\mathbf{x} - \mathbf{x}'|\cos(k_s\,|\mathbf{x} - \mathbf{x}'|)\end{aligned} \tag{46}$$

to leading order[22] for large N and Ω. If one were to take the limit $|\mathbf{x} - \mathbf{x}'| \to \infty$ in (46) for fixed k_s and Ω, one would reach the incorrect conclusion that ODLRO is not present. However, as shown in Sec. 2, the correct definition of ODLRO involves taking the thermodynamic limit first, then the limit $|\mathbf{x}-\mathbf{x}'| \to \infty$. For fixed $|\mathbf{x} - \mathbf{x}'|$, the expression in curly brackets behaves like $(\frac{2}{15})k_s^5\,|\mathbf{x} - \mathbf{x}'|^5$ as $k_s \to 0(\Omega \to \infty)$, so that by (26) the right side of (46) approaches ρ. Thus ODLRO is present,[23] in agreement with our previous result.

The model we have studied here is physically unrealistic in two respects. In the first place, it retains only the diagonal part of the interparticle interaction; in the second place, the system would collapse to infinitesimal volume if the density ρ were not artificially constrained to remain constant, since the interaction is purely attractive with no repulsive core. The first defect can be removed to the extent of including "pairing-interaction" terms $a_\mathbf{k}^\dagger a_{-\mathbf{k}}^\dagger a_{-\mathbf{k}'} a_{\mathbf{k}'}$, without altering the final conclusion that generalized condensation, and hence ODLRO, are present.[24] The second can be removed by including a hard-core pseudopotential in addition to an attractive tail,[25] although the treatment is not entirely satisfactory.[26] However, the point we wished to emphasize in this section is that ODLRO is implied not only by simple Bose condensation, but also by generalized condensation. The question of the precise nature of the Bose condensation in liquid ^{4}He must, of course, still be regarded as open.

5. EFFECT OF BOUNDARY CONDITIONS

We have restricted ourselves in Secs. 3 and 4 to the case of periodic boundary conditions. However, as emphasized by Yang, the concept of ODLRO is a very general one, and ought also to apply, e.g., to box-enclosure boundary conditions, which are more realistic for most applications to liquid ^{4}He. The operations involved in the criteria (3) and (9) for ODLRO in ρ_1 can still be carried out in the case of box-enclosure boundary conditions (or any other reasonable ones), so that it is natural to inquire whether or not there is any reasonable generalization

---

[21] The same is true if $\alpha > 0$, in which case the condensation is of the usual form ("simple" condensation, i.e., a delta function at $\mathbf{k} = 0$), which is a special case of generalized condensation.

[22] The k-sum (5) has been replaced by an integral in obtaining (46); this is permissible, since the limit $\Omega \to \infty$ is to be taken first, for *fixed* $|\mathbf{x} - \mathbf{x}'|$, before passing to the limit $|\mathbf{x} - \mathbf{x}'| = \infty$.

[23] The final limit $|\mathbf{x} - \mathbf{x}'| \to \infty$ is trivial in this case ($T = 0$) since the thermodynamic limit of (46) is in fact independent of $|\mathbf{x} - \mathbf{x}'|$. For $T > 0$, there is a nonzero term $\langle \mathbf{x}|\rho_1'|\,\mathbf{x}'\rangle$ which vanishes as $|\mathbf{x} - \mathbf{x}'| \to \infty$.

[24] The zero-temperature case is treated in Ref. 10; the analysis can be extended to $T > 0$ by the method of the thermodynamically equivalent Hamiltonian (Ref. 18).

[25] Reference 10, p. 1476.

[26] Although the ground-state energy as a function of density has a minimum so that collapse does not occur, the low-density approximations that must be made in order to obtain explicit results are not accurate at the equilibrium density.

of the concept of generalized Bose condensation which would apply in such cases.

We start from the observation that the single-particle density martix $\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle$ in coordinate representation [(1), (2)], when regarded as an integral kernel, is Hermitian and positive semidefinite, so that it possesses a complete orthonormal set of eigenfunctions[27] $\varphi_i(\mathbf{x})$ and associated real and nonnegative[28] eigenvalues n_i:

$$\int_\Omega \langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle \varphi_i(\mathbf{x}') \, d^3x' = n_i \varphi_i(\mathbf{x}). \tag{47}$$

The generalization of Eq. (5a) is then

$$\langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle = \sum_i n_i \varphi_i(\mathbf{x}) \varphi_i^*(\mathbf{x}'), \tag{48}$$

which is the familiar eigenfunction expansion of a Hermitian kernel. All eigenvalues of ρ_1 [hence of the kernel $\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle$] for an N-particle system are[3] $\leq N$, so that they may be assumed to be ordered in a descending sequence $n_1 \geq n_2 \geq \cdots$. We then define generalized Bose condensation to be present if and only if

$$\lim_{\epsilon \to 0} \lim \text{therm } N^{-1} \sum_{i=1}^{[\epsilon N]} n_i = f > 0, \tag{49}$$

where ϵ is a positive parameter independent of N and Ω, $[\epsilon N]$ is the largest integer $\leq \epsilon N$, and "lim therm" is defined by (4); the limit f will be called the "condensed fraction." In case ρ_1 has one eigenvalue of order N and all other eigenvalues of order unity, (49) will be satisfied, with f equal to the thermodynamic limit of n_1/N.[29] However, the definition (49) is much more inclusive. Consider, e.g., the case of periodic boundary conditions. Then the n_i are the values of the momentum distribution function $n_{\mathbf{k}}$ on the discrete $\mathbf{k}$-lattice, enumerated in order of decreasing magnitude. Suppose, furthermore, that generalized Bose condensation in the sense (6) is present, and that $n_{\mathbf{k}}$ is a decreasing function of k. Then, since there are asymptotically $\frac{4}{3}[\Omega/(2\pi)^3]\pi k_0^3$ allowed $\mathbf{k}$ values in the sphere of radius k_0 about $\mathbf{k} = 0$, one sees on taking $\epsilon = (\frac{4}{3})\pi k_0^3/(2\pi)^3\rho$ that (49) is satisfied with $f = \rho_0/\rho$. More generally, (49) will also be satisfied for Bose condensation at nonzero momentum, condensation into a single-particle state which is not a momentum eigenstate, and hopefully also for "smeared" condensation into a set of "neighboring" (in some physically meaningful sense) single-particle states which are not momentum eigenstates, but are physically meaningful single-particle states for the system at hand.

However, a simple example shows that in discussing the relationship between ODLRO and Bose condensation, complications arise for box-enclosure boundary conditions which are not present for periodic boundary conditions. Consider an ideal Bose gas in the one-dimensional box $0 \leq x \leq L$. Then the lowest single-particle state is

$$\varphi_1(x) = (2/L)^{\frac{1}{2}} \sin(\pi x/L). \tag{50}$$

It follows that at temperature $T = 0$

$$\langle x| \rho_1 |x'\rangle = 2\rho \sin(\pi x/L) \sin(\pi x'/L), \tag{51}$$

where $\rho = N/L$. Thus

$$\lim \text{therm } \langle x| \rho_1 |x'\rangle = 0 \tag{52}$$

for *fixed* x and x'. Then by (3) one concludes that ODLRO is not present, in spite of the fact that at $T = 0$, complete Bose condensation is present. However, a little thought shows what has gone wrong. The essential point is that the purpose of taking the thermodynamic limit is to remove the boundaries to infinity *before* letting $|x - x'| \to \infty$. But if we use the box $0 \leq x \leq L$ as above, and keep x and x' fixed as $L \to \infty$, then only the right end of the box is removed to infinity; x and x' stay within a finite distance of the left end of the box, and (52) is a simple consequence of the normalization of the wavefunction together with the requirement that it vanish at the walls (in particular, at the left end of the box). Thus the thermodynamic limit should be interpreted in such a way that both (in three dimensions, all) walls recede to infinite distance. This will be achieved, e.g., if we take the box $-\frac{1}{2}L \leq x \leq \frac{1}{2}L$. Then instead of (50) we have

$$\varphi_1(x) = (2/L)^{\frac{1}{2}} \cos(\pi x/L), \tag{53}$$

so that

$$\langle x| \rho_1 |x'\rangle = 2\rho \cos(\pi x/L) \cos(\pi x'/L),$$

$$\lim \text{therm } \langle x| \rho_1 |x'\rangle = 2\rho. \tag{54}$$

Then the double limit (3) is 2ρ instead of zero, so that ODLRO is present.

This example suggests that Bose condensation and ODLRO in ρ_1 are closely related for boundary con-

[27] These are the "natural orbitals" of the system; see, e.g., A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963), where the analogous fermion case is discussed. Further references are given there.

[28] ρ_1 may have the eigenvalue zero, which may be degenerate. Thus, e.g., for the ideal Bose gas at zero temperature, $\langle \mathbf{x} |\rho_1| \mathbf{x}'\rangle = N\varphi_1(\mathbf{x})\varphi_1^*(\mathbf{x}')$, where $\varphi_1(\mathbf{x})$ is the lowest single-particle state. This kernel possesses a single nonzero eigenvalue $n_1 = N$ with eigenfunction φ_1, and the infinitely degenerate eigenvalue zero, with eigenfunctions $\varphi_i(\mathbf{x})$, $i > 1$.

[29] Since the eigenvalues n_i with $i \geq 2$ are only of order unity, they possess an upper bound b independent of N and Ω, so that $\sum_{i=1}^{[\epsilon N]} n_i \leq n_1 + \{[\epsilon N] - 1\}b$ and hence $\lim \text{therm } N^{-1} \sum_{i=1}^{[\epsilon N]}. n_i = \lim \text{therm } (n_1/N) + \epsilon b$, so that the total contribution of higher eigenvalues vanishes as $\epsilon \to 0$.

ditions other than the periodic one for which we have already proved the relationship, provided that the thermodynamic limit is interpreted properly. As further evidence, we shall examine a model closely related to that studied in Sec. 4. We begin by noting that the peculiar result (52), although due partially to an incorrect interpretation of the thermodynamic limit, as explained above, is also related to the pathological nature of the completely ideal Bose gas in a box, and in particular to the fact that the single-particle density varies across the box instead of being constant except within a microscopic distance of the walls, as would be the case for a real gas.[30] It follows that for box-enclosure boundary conditions, it is very easy to obtain nonsensical results by a perturbation treatment starting with an ideal Bose gas as unperturbed system. The way around this difficulty was pointed out by Schafroth[31]: one starts not with free-particle states, but with Hartree–Fock states. Then the lowest Hartree–Fock orbital[32] $\varphi_0(\mathbf{x})$ for any reasonable container shape is just the constant $\Omega^{-\frac{1}{2}}$ (where Ω is the volume of the container) except for a region within a microscopic distance from the walls, where $\varphi_0(\mathbf{x})$ drops to its value zero at the walls as required by the boundary condition. Schafroth's analysis was for the special (and exceptional) case of the charged Bose gas, but his final conclusion concerning the nature of the lowest Hartree–Fock orbital appears to be of very general validity. Although a rigorous proof appears difficult because of the nonlinearity of the Hartree–Fock equation, a self-consistent analysis which makes this contention plausible is carried out in the Appendix. It is furthermore shown (not rigorously) that not only the lowest Hartree–Fock orbital but also the higher orbitals differ only in a microscopic boundary layer near the walls from free-particle orbitals $\varphi_{\mathbf{k}}$ appropriate to the homogeneous Neumann boundary condition[33]

$$\hat{\mathbf{n}}\cdot\nabla\varphi_{\mathbf{k}}(\mathbf{x}) = 0 \quad \text{on the walls,} \tag{55}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the wall.

We shall therefore examine the properties of a model Hamiltonian which is diagonal in such a representation. For a cubical box of volume Ω with faces at $x = \pm\frac{1}{2}L$, $y = \pm\frac{1}{2}L$, $z = \pm\frac{1}{2}L$, a complete set of orbitals satisfying (55) is

$$\varphi_{\mathbf{k}}(\mathbf{x}) = f_{k_x}(x)f_{k_y}(y)f_{k_z}(z), \tag{56}$$

where the allowed values of k_x, k_y, and k_z are $n\pi/L$, $n = 0, 1, 2, \cdots$,

$$f_k(x) = (2/L)^{\frac{1}{2}}\cos[k(x + \tfrac{1}{2}L)], \qquad k \neq 0$$
$$f_0(x) = (1/L)^{\frac{1}{2}}, \tag{57}$$

and $L = \Omega^{\frac{1}{3}}$. If we start with a Schrödinger Hamiltonian

$$H_{\text{Sch}} = \sum_{i=1}^{N} -\tfrac{1}{2}\Delta_i^2 + 4\pi\alpha\sum_{i<l}^{N}\delta(\mathbf{x}_i - \mathbf{x}_l), \tag{58}$$

where α is a scattering length and $4\pi\alpha\delta(\mathbf{x}_i - \mathbf{x}_l)$ the corresponding Fermi pseudopotential, then the diagonal part of this Hamiltonian in a representation in terms of annihilation and creation operators $a_{\mathbf{k}}$, $a_{\mathbf{k}}^\dagger$ for particles in the orbitals $\varphi_{\mathbf{k}}$ is

$$\begin{aligned} H = &\sum_{\mathbf{k}} \tfrac{1}{2}k^2 a_{\mathbf{k}}^\dagger a_{\mathbf{k}} \\ &+ \tfrac{1}{2}\sum_{\mathbf{k}\mathbf{k}'}(\mathbf{k}\mathbf{k}'\,|V|\,\mathbf{k}\mathbf{k}')a_{\mathbf{k}}^\dagger a_{\mathbf{k}'}^\dagger a_{\mathbf{k}'}a_{\mathbf{k}} \\ &+ \tfrac{1}{2}\sum_{\mathbf{k}\mathbf{k}'}{}' (\mathbf{k}'\mathbf{k}\,|V|\,\mathbf{k}\mathbf{k}')a_{\mathbf{k}'}^\dagger a_{\mathbf{k}}^\dagger a_{\mathbf{k}'}a_{\mathbf{k}} \end{aligned} \tag{59}$$

in analogy with (12); here

$$\begin{aligned} (\mathbf{k}\mathbf{k}'\,|V|\,\mathbf{k}\mathbf{k}') &= (\mathbf{k}'\mathbf{k}\,|V|\,\mathbf{k}\mathbf{k}') \\ &= 4\pi\alpha\int_\Omega \varphi_{\mathbf{k}}^2(\mathbf{x})\theta_{\mathbf{k}'}^2(\mathbf{x})\,d^3x. \end{aligned} \tag{60}$$

Evaluating the matrix elements with the aid of (56), (57) and using the Bose commutation relations to express H in terms of the occupation-number operators $N_{\mathbf{k}} = a_{\mathbf{k}}^\dagger a_{\mathbf{k}}$, one finds, aside from terms which are negligible in the thermodynamic limit,[34]

$$H = \frac{2\pi\alpha}{\Omega}N_0^2 + \sum_{\mathbf{k}}\tfrac{1}{2}k^2 N_{\mathbf{k}} + \frac{4\pi\alpha}{\Omega}\sum_{\mathbf{k}\mathbf{k}'}{}' N_{\mathbf{k}}N_{\mathbf{k}'}, \tag{61}$$

where the prime on the last summation implies omission of terms with $\mathbf{k} = \mathbf{k}'$; note, however that terms with $\mathbf{k} = 0$ or $\mathbf{k}' = 0$ (but not both) are included.

If $\alpha \geq 0$ then the N-particle ground state has $N_0 = N$ and $N_k = 0$, as in the case of the model

[30] This was emphasized by T. D. Lee, K. Huang, and C. N. Yang, Phys. Rev. **106**, 1135 (1957), Footnote 17.

[31] M. R. Schafroth, Phys. Rev. **100**, 463 (1955), Sec. 4.

[32] For the sake of brevity we adopt the chemical nomenclature "orbital" for "single-particle state."

[33] Note that this insures that there will be no particle flux through the walls, an obvious necessary condition for impenetrable walls.

[34] The omitted terms are

$$\begin{aligned} &(4\pi\alpha/\Omega)\{-(27/16)\sum_{\mathbf{k}\neq 0}N_{\mathbf{k}}(N_{\mathbf{k}}-1) - \tfrac{1}{2}N_0 \\ &+ \sum_{\substack{\mathbf{k}\mathbf{k}' \\ (\mathbf{k}\neq 0,\,\mathbf{k}'\neq 0)}}{}' [D(k_x,k_x') + D(k_y,k_y') + D(k_z,k_z') \\ &+ D(k_x,k_x')D(k_y,k_y') + D(k_x,k_x')D(k_z,k_z') \\ &+ D(k_y,k_y')D(k_z,k_z') + D(k_x,k_x')D(k_y,k_y')D(k_z,k_z')], \end{aligned}$$

where

$$D(k,k') = \delta_{k0} + \delta_{k'0} + \tfrac{1}{2}(\delta_{kk'} + \delta_{k,-k'}).$$

These terms clearly contribute only $O(1)$ to the thermodynamic functions even in the case of simple Bose condensation [$N_0 = O(N)$]; in thermal equilibrium no $N_{\mathbf{k}}$ except N_0 can be $O(N)$.

studied in Sec. 4, which differed only in that periodic boundary conditions were used instead of (55). The corresponding ground-state energy is clearly $2\pi N\rho\alpha$, which differs from the result $2\pi(N-1)\rho\alpha$ of Sec. 4 only by a thermodynamically negligible (nonextensive) term. By (48) and (57) one has[35]

$$\langle \mathbf{x}|\ \rho_1\ |\mathbf{x}'\rangle = N/\Omega = \rho, \tag{62}$$

so that ODLRO is present; again, (62) is identical with what one would calculate for periodic boundary conditions.

The interesting question, however, is whether generalized Bose condensation and ODLRO occur for $\alpha < 0$, as they did in the case of periodic boundary conditions. By essentially the same argument as used in Sec. 4, one sees that then the ground state will have high occupation of a large number of allowed $\mathbf{k}$ states near and including $\mathbf{k} = 0$; minimizing the ground-state energy with respect to the occupation numbers $n_\mathbf{k}$ subject to (18), one obtains

$$4\pi\Omega^{-1}\alpha n_0 - \mu + 8\pi\Omega^{-1}\alpha \sum_{\mathbf{k}\neq 0} n_\mathbf{k} = 0,$$

$$\tfrac{1}{2}k^2 - \mu + 8\pi\rho\alpha - 8\pi\Omega^{-1}\alpha n_\mathbf{k} = 0, \quad \mathbf{k} \neq 0. \tag{63}$$

The solution of the $\mathbf{k} \neq 0$ equation is[36]

$$n_\mathbf{k} = \Omega(\tfrac{1}{2}k^2 - \mu + 8\pi\rho\alpha)/8\pi\alpha \quad \text{if} \quad \tfrac{1}{2}k^2 - \mu + 8\pi\rho\alpha > 0,$$
$$= 0 \quad \text{if} \quad \tfrac{1}{2}k^2 - \mu + 8\pi\rho\alpha < 0. \tag{64}$$

The value, k_s, of k at which $n_\mathbf{k}$ vanishes to leading order in Ω is, in analogy with (25),

$$k_s = [2(\mu - 8\pi\rho\alpha)]^{\frac{1}{2}}, \tag{65}$$

and the analog of (23) is[37]

$$\left(\frac{1}{8}\frac{\Omega}{\pi^3}\frac{\Omega}{8\pi\alpha}\right) 4\pi \int_0^{k_s} k^2(\tfrac{1}{2}k^2 - \mu + 8\pi\rho\alpha)\, dk = N. \tag{66}$$

Thus

$$\mu + 8\pi\rho\,|\alpha| = \tfrac{1}{2}[240\pi^3\rho\,|\alpha|/\Omega]^{2/5}$$

so that by (64) and (65)

$$n_\mathbf{k} = \tfrac{1}{2}(k_s^2 - k^2)\Omega/8\pi|\alpha|, \quad 0 < k < k_s,$$
$$= 0, \quad k > k_s \tag{67}$$

with

$$k_s = [240\pi^3\rho\,|\alpha|/\Omega]^{1/5}. \tag{68}$$

One then sees from (18), (65), and the top Eq. (63) that

$$n_0 = \tfrac{1}{2}k_s^2\Omega/4\pi|\alpha|. \tag{69}$$

Thus, just as in the periodic boundary condition case studied in Sec. 4, n_0 is not $O(N)$, but the sum of $n_\mathbf{k}$ over $k < k_s$ is $O(N)$. The eigenvalues n_1, n_2, $\cdots$ of ρ_1 are just the $n_\mathbf{k}$ ordered in descending magnitude[38]; thus the criterion (49) for generalized Bose condensation is satisfied with $f = 1$, as was the case for periodic boundary conditions.

In order to examine the limiting behavior of $\langle \mathbf{x}|\ \rho_1\ |\mathbf{x}'\rangle$ we use (48) and (67) to write[38,39]

$$\langle \mathbf{x}\ |\rho_1|\ \mathbf{x}'\rangle = \frac{\Omega}{16\pi\,|\alpha|} \sum_{\substack{\mathbf{k} \\ (k<k_s)}} (k_s^2 - k^2)\varphi_\mathbf{k}(\mathbf{x})\varphi_\mathbf{k}^*(\mathbf{x}') \tag{70}$$

to leading order in Ω. In evaluating the thermodynamic limit of (70) some care is required, since the $\varphi_\mathbf{k}$ depend on Ω in a way which is not so trivial as in the case of periodic boundary conditions. For the boundary conditions (55) one has

$$\cos[k(x + \tfrac{1}{2}L)]$$
$$= (-1)^{\frac{1}{2}n}\cos(kx), \quad n = 0, 2, 4, \cdots,$$
$$= (-1)^{\frac{1}{2}(n+1)}\sin(kx), \quad n = 1, 3, 5, \cdots, \tag{71}$$

where $k = n\pi/L$. Since $k_s \to 0$ in the thermodynamic limit, it is clear that contributions to (70) involving sine factors will be negligible compared to those involving only cosine factors in the thermodynamic limit, and that in that limit the cosine factors may be replaced by unity. Furthermore, the factors $(-1)^{\frac{1}{2}n}$ may be dropped since only $[(-1)^{\frac{1}{2}n}]^2 = 1$ occurs in $\varphi_\mathbf{k}(\mathbf{x})\varphi_\mathbf{k}(\mathbf{x}')$. Finally, the set of allowed $\mathbf{k}$ values whose components are all *even* nonnegative integral multiples of π/L, so that by (71) only cosine factors occur, is $\frac{1}{8}$ of the set of all allowed $\mathbf{k}$ values (all components nonnegative integral multiples of π/L), which is in turn $\frac{1}{8}$ of the set of all $\mathbf{k}$ values whose components are positive or negative integral multiples of π/L. Thus one finds with the aid of (57), (71), and (68)

[35] It is clear from (1) that $\langle \mathbf{x}\ |\rho_1|\ \mathbf{x}'\rangle$ reduces at $T = 0$ to the ground-state average of $\psi\dagger(\mathbf{x}')\psi(\mathbf{x})$. Inserting $\psi(\mathbf{x}) = \sum_\mathbf{k} \varphi_\mathbf{k}(\mathbf{x})\, a_\mathbf{k}$, and using the fact that the ground state is diagonal in the $N_\mathbf{k}$ representation with eigenvalues $N_0 = N$ and $N_\mathbf{k} = 0$, $\mathbf{k} \neq 0$, one obtains the desired result.

[36] See Footnote 15.

[37] With the boundary condition (55) and states (57), all components of $\mathbf{k}$ must be nonnegative; thus we integrate only over 1/8 of $\mathbf{k}$ space. Also, the density of states in the allowed region (all components positive) is Ω/π^3 instead of $\Omega/(2\pi)$.[3]

[38] This is proved by the obvious generalization of the argument in Footnote 35.

[39] Since, by (57), φ_k is an even function of the components of $\mathbf{k}$, and furthermore $n_\mathbf{k}$ is a function only of $k = |\mathbf{k}|$, one may extend the integration over all $\mathbf{k}$ with $k < k_s$ and multiply by a factor 1/8.

$$\lim \text{therm} \langle \mathbf{x} | \rho_1 | \mathbf{x}' \rangle$$

$$= \lim_{k_s \to 0} \frac{4\pi}{64} \frac{\Omega}{16\pi |\alpha|} \frac{\Omega}{\pi^3} \frac{8}{\Omega} \int_0^{k_s} (k_s^2 - k^2) k^2 \, dk$$

$$= \lim_{\Omega \to \infty} \frac{2}{15} \frac{\Omega}{32\pi^3 |\alpha|} \frac{240\pi^3 \rho |\alpha|}{\Omega} = \rho. \quad (72)$$

Thus ODLRO in ρ_1 is present, and in fact with the same limiting value, ρ, as was found in Sec. 4 for periodic boundary conditions. It is furthermore clear that the analysis could be extended to nonzero temperature by the same method as was used in Sec. 4, and that one would then find a thermodynamic phase transition at $T = T_c$, with the properties that for $T < T_c$ both generalized Bose condensation [in the sense (49)] and ODLRO in ρ_1 are present, and for $T > T_c$ both are absent.

At this point it is necessary to make the same provisos concerning the unphysical features of the model Hamiltonian as were made at the end of Sec. 4; as there, we remark that the model can be refined so as to remove some but not all[40] of the unphysical features.[41] We further remark, as in Sec. 4, that the existence of generalized Bose condensation and hence ODLRO in ρ_1 for this model raises the question of whether ODLRO in more realistic systems might not be associated with generalized rather than simple Bose condensation.

6. A CONJECTURE

In Secs. 3 and 4 it was shown that for a boson system with periodic boundary conditions, the existence of ODLRO in ρ_1 is equivalent to the existence of generalized Bose condensation. In Sec. 5 it was shown that the relationship also holds for homogeneous Neumann boundary conditions in the case of a simple soluble model, and it was made plausible in the Appendix that the same analysis should hold for box-enclosure boundary conditions. On the basis of these results we shall make the following conjecture:

For a system of bosons in thermal equilibrium and with arbitrary boundary conditions, the existence of ODLRO in ρ_1 in the sense (9) is equivalent to the existence of generalized Bose condensation in the sense (49).

A major difficulty in attempting to prove this conjecture is that the eigenfunctions of ρ_1 are not known for an interacting system with arbitrary boundary conditions. In contrast, for periodic boundary conditions, ρ_1 is necessarily diagonal in momentum representation, so its eigenfunctions are plane waves $\Omega^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{x}}$ for arbitrary translation-invariant interparticle interactions[42] and arbitrary temperatures; this is what enabled us to construct an explicit proof in Sec. 3. It would seem that a proof for general boundary conditions might be based on variational arguments of the type employed, e.g., in Sec. 4 of Penrose and Onsager.[2] However, the author has not succeeded in such an attempt, and wishes here only to express the hope that the conjecture will eventually be either proved or disproved.

7. DIFFICULTIES WITH HARD CORES

The models studied in Secs. 4 and 5 are not directly applicable to systems with hard-core interactions except insofar as one trusts scattering-length or pseudopotential approximations as applied to this problem, since the Hamiltonian does not exist in any rigorous sense in a single-particle representation. This is well known for the special case of plane waves, and can be shown to be true for any complete set of single-particle states.

On the other hand, the conjecture in Sec. 6 is meant to apply also to systems with hard cores (in particular, to liquid ^{4}He). Furthermore, the general relationship between ODLRO and generalized Bose condensation established in Sec. 3 for periodic boundary conditions should be applicable to systems with hard-core interactions, although it does not seem possible to rigorously prove that any such system has a momentum distribution function with the analytical properties assumed in Sec. 3.

8. TWO-PARTICLE DENSITY MATRIX

The two-particle density matrix is defined, in Yang's notation,[3] by

$$\langle \mathbf{x}_1 \mathbf{x}_2 | \rho_2 | \mathbf{x}_1' \mathbf{x}_2' \rangle = \text{Tr} \, [\psi(\mathbf{x}_1)\psi(\mathbf{x}_2)\rho\psi^\dagger(\mathbf{x}_2')\psi^\dagger(\mathbf{x}_1')]. \quad (73)$$

We shall limit ourselves in this section to periodic boundary conditions. Then the Bose field operator $\psi(\mathbf{x})$ can be expanded in terms of plane waves, leading to the expression

[40] An additional unphysical feature, which does not occur for periodic boundary conditions, is that the single-particle density for *large but finite* N and Ω can be shown to be of the form

$$\langle \mathbf{x} | \rho_1 | \mathbf{x} \rangle \sim (1/8)\rho + f(\mathbf{x}),$$

where $f(0) = (7/8)\rho$, whereas $f(\mathbf{x})$ falls to zero for $|\mathbf{x}| \gg k_s^{-1} = O(\Omega^{1/5})$. Thus 7/8 of the particles are concentrated in a region in the middle of the box which is negligible compared to the volume of the system, although very large compared to atomic dimensions. For hard spheres this could not occur.

[41] Removing *all* unphysical features is clearly incompatible with exact solubility of the model.

[42] The interaction potential must be periodically extended in the usual way to ensure compatibility with the periodic boundary conditions.

$$\langle \mathbf{x}_1\mathbf{x}_2 \,|\rho_2|\, \mathbf{x}_1'\mathbf{x}_2'\rangle = \Omega^{-2} \sum_{\mathbf{k}_1\mathbf{k}_2} \sum_{\mathbf{k}_1'\mathbf{k}_2'} e^{i\mathbf{k}_1\cdot\mathbf{x}_1} e^{i\mathbf{k}_2\cdot\mathbf{x}_2} \times e^{-i\mathbf{k}_1'\cdot\mathbf{x}_1'} e^{-i\mathbf{k}_2'\cdot\mathbf{x}_2'} \operatorname{Tr}(a_{\mathbf{k}_1} a_{\mathbf{k}_2} \rho a^\dagger_{\mathbf{k}_2'} a^\dagger_{\mathbf{k}_1'}), \quad (74)$$

analogous to the simpler expression (5) for $\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle$. By momentum conservation,

$$\operatorname{Tr}(a_{\mathbf{k}_1} a_{\mathbf{k}_2} \rho a^\dagger_{\mathbf{k}_2'} a^\dagger_{\mathbf{k}_1'}) = 0 \quad \text{unless} \quad \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_1' + \mathbf{k}_2'; \quad (75)$$

but here this is a much weaker restriction than was the corresponding one for ρ_1, since it does not imply that ρ_2 is diagonal in the free-particle representation. Hence we cannot prove a general relationship between ODLRO of ρ_2 and its eigenvalues, analogous to that proved in Sec. 3 for ρ_1; the best we can do is look at highly simplified special cases, and make the obvious conjecture about the general case.

To this end, consider the case that H (hence ρ) is diagonal in the free-particle representation, as is the case for the model studied in Sec. 4. Then the trace (75) is only nonzero if $\mathbf{k}_1 = \mathbf{k}_1'$, $\mathbf{k}_2 = \mathbf{k}_2'$ or if $\mathbf{k}_1 = \mathbf{k}_2'$, $\mathbf{k}_2 = \mathbf{k}_1'$, and one finds with the aid of the Bose commutation relations

$$\langle \mathbf{x}_1\mathbf{x}_2 \,|\rho_2|\, \mathbf{x}_1'\mathbf{x}_2'\rangle = \Omega^{-2} \sum_{\mathbf{k}_1\mathbf{k}_2} [e^{i\mathbf{k}_1\cdot(\mathbf{x}_1-\mathbf{x}_1')} e^{i\mathbf{k}_2\cdot(\mathbf{x}_2-\mathbf{x}_2')} + e^{i\mathbf{k}_1\cdot(\mathbf{x}_1-\mathbf{x}_2')} e^{i\mathbf{k}_2\cdot(\mathbf{x}_2-\mathbf{x}_1')}] \operatorname{Tr}(\rho N_{\mathbf{k}_1} N_{\mathbf{k}_2}) - \Omega^{-2} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}_1+\mathbf{x}_2-\mathbf{x}_1'-\mathbf{x}_2')} \operatorname{Tr}(\rho N_{\mathbf{k}}^2). \quad (76)$$

Evaluation of the trace in a free-particle basis $\{|\alpha\rangle\}$ gives[43]

$$\operatorname{Tr}(\rho N_{\mathbf{k}_1} N_{\mathbf{k}_2}) = \sum_{\alpha\beta} \langle\alpha\,|\rho N_{\mathbf{k}_1}|\,\beta\rangle\langle\beta\,|N_{\mathbf{k}_2}|\,\alpha\rangle = \sum_{\alpha} \langle\alpha\,|\rho N_{\mathbf{k}_1}|\,\alpha\rangle\langle\alpha\,|N_{\mathbf{k}_2}|\,\alpha\rangle. \quad (77)$$

At $T = 0$, ρ is just the projection operator onto the ground state, which is a simultaneous eigenstate of the $N_{\mathbf{k}}$ and can therefore be chosen to be one of the $|\alpha\rangle$. Hence

$$\operatorname{Tr}(\rho N_{\mathbf{k}_1} N_{\mathbf{k}_2}) \underset{T\to 0}{\longrightarrow} n_{\mathbf{k}_1} n_{\mathbf{k}_2}, \quad (78)$$

where $n_{\mathbf{k}}$ is the eigenvalue of $N_{\mathbf{k}}$ in the ground state. Thus, substituting into (76) and comparing with (5), one finds

$$\langle \mathbf{x}_1\mathbf{x}_2 \,|\rho_2|\, \mathbf{x}_1'\mathbf{x}_2'\rangle \underset{T\to 0}{\longrightarrow} \langle \mathbf{x}_1\,|\rho_1|\,\mathbf{x}_1'\rangle\langle \mathbf{x}_2\,|\rho_1|\,\mathbf{x}_2'\rangle + \langle \mathbf{x}_1\,|\rho_1|\,\mathbf{x}_2'\rangle\langle \mathbf{x}_2\,|\rho_1|\,\mathbf{x}_1'\rangle - \Omega^{-2} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}_1+\mathbf{x}_2-\mathbf{x}_1'-\mathbf{x}_2')} n_{\mathbf{k}}^2. \quad (79)$$

In the case of simple Bose condensation [$n_0 = O(N)$] both the superposition terms (those bilinear in ρ_1) and the sum over $n_{\mathbf{k}}^2$ have nonvanishing thermodynamic limits for all values of $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_1'$, $\mathbf{x}_2'$, so that ODLRO is present in ρ_2, in agreement with Yang's conclusion.[44] In the case of "smeared" condensation typified by the model of Sec. 4, the thermodynamic limit of the sum over $n_{\mathbf{k}}^2$ can be shown to vanish, but that of $\langle \mathbf{x}| \rho_1 |\mathbf{x}'\rangle$ does not (as shown in Sec. 4), so that again ODLRO of ρ_2 is present.

How is this behavior related to the eigenvalues of ρ_2? In the simple case studied here (H diagonal, $T = 0$) the eigenfunctions and eigenvalues of ρ_2 can be found explicitly. Substitution of (78) into (76) yields

$$\langle \mathbf{x}_1\mathbf{x}_2 \,|\rho_2|\, \mathbf{x}_1'\mathbf{x}_2'\rangle = \Omega^{-2} \sum_{\mathbf{k}_1\mathbf{k}_2} n_{\mathbf{k}_1} n_{\mathbf{k}} \times [e^{i\mathbf{k}_1\cdot(\mathbf{x}_1-\mathbf{x}_1')} e^{i\mathbf{k}_2\cdot(\mathbf{x}_2-\mathbf{x}_2')} + e^{i\mathbf{k}_1\cdot(\mathbf{x}_1-\mathbf{x}_2')} e^{i\mathbf{k}_2\cdot(\mathbf{x}_2-\mathbf{x}_1')}] - \Omega^{-2} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}_1+\mathbf{x}_2-\mathbf{x}_1'-\mathbf{x}_2')} n_{\mathbf{k}}^2. \quad (80)$$

Then one finds by direct substitution that the eigenfunctions φ and eigenvalues λ of ρ_2 [in the sense of (47)] are

$$\varphi_{\mathbf{q}_1\mathbf{q}_2}(\mathbf{x}_1\mathbf{x}_2) = 2^{-\frac{1}{2}}\Omega^{-1}(e^{i\mathbf{q}_1\cdot\mathbf{x}_1} e^{i\mathbf{q}_2\cdot\mathbf{x}_2} + e^{i\mathbf{q}_2\cdot\mathbf{x}_1} e^{i\mathbf{q}_1\cdot\mathbf{x}_2}), \quad (81)$$

$$\lambda_{\mathbf{q}_1\mathbf{q}_2} = 2n_{\mathbf{q}_1} n_{\mathbf{q}_2} - \delta_{\mathbf{q}_1\mathbf{q}_2} n_{\mathbf{q}_1}^2.$$

It is therefore clear, at least for this special case, that large eigenvalues of ρ_1 imply large eigenvalues of ρ_2, the largest eigenvalues of ρ_2 being of order of magnitude of the square of the largest eigenvalues of ρ_1. In the case of simple Bose condensation, ρ_2 has a single large eigenvalue (here λ_{00}) of order N^2, as previously pointed out by Yang[44]; in the case of generalized condensation typified by the model of Sec. 4, ρ_2 has no single eigenvalue of order N^2, but there are many large eigenvalues whose sum is $O(N^2)$, in the sense that

$$\sum_{\mathbf{q}_1<k_0} \sum_{\mathbf{q}_2<k_0} \lambda_{\mathbf{q}_1\mathbf{q}_2} = O(N^2), \quad (82)$$

where k_0 is given by (28). We conjecture that this behavior is of very general occurrence; to be precise, we conjecture that generalized Bose condensation is associated not only with ODLRO of ρ_1 and large eigenvalues of ρ_1 in the sense of (49), but also with ODLRO of ρ_2 and large eigenvalues of ρ_2 in the sense

$$\lim_{\epsilon\to 0} \lim \operatorname{therm} N^{-2} \sum_{i=1}^{[\epsilon N^2]} \lambda_i > 0, \quad (83)$$

where the λ_i are the eigenvalues of ρ_2 enumerated in order of decreasing magnitude. It is clear that

[43] This is true even if ρ is not diagonal in the free-particle basis; however, the diagonal property of ρ will be used in further reducing (77).

[44] Section 17 of the first reference in Footnote 3.

(83) is satisfied in the special case (82), but (83) can be applied even when the λ_i are not labeled by free-particle momenta.

The situation is quite different in the Fermi case. The derivation of (79) goes through as in the Bose case, except that the sum of superposition terms is replaced by their difference, and the sum over n_k^2 is absent. But then, since ρ_1 cannot show ODLRO for a Fermi system,[3] one concludes that a model Hamiltonian which is diagonal in the free-fermion representation, but may include arbitrary diagonal interaction terms, cannot lead to ODLRO in ρ_2 at $T = 0$, and hence *a fortiori* also not at higher temperatures. This is entirely consistent with the fact that the canonical example of a Fermi system showing ODLRO in ρ_2, namely the Bardeen, Cooper, Schrieffer (BCS) model of superconductivity,[45] has the property that the interactions included are entirely off-diagonal in the free-particle representation, consisting of pair–pair scattering terms $a_k^\dagger a_{-k}^\dagger a_{-k'} a_{k'}$. For such a model the aforementioned superposition terms in $\langle \mathbf{x}_1\mathbf{x}_2|\rho_2|\mathbf{x}_1'\mathbf{x}_2'\rangle$ are supplemented by an additional term[46] $\chi(\mathbf{x}_1\mathbf{x}_2)\chi^*(\mathbf{x}_1'\mathbf{x}_2')$ which leads to ODLRO in ρ_2 of the type described by Yang,[47] associated with a single large eigenvalue of ρ_2 of order N. This is the analog, for the Fermi system, of simple Bose condensation. It would appear that the following questions deserve to be investigated: Are there any Fermi systems which exhibit a phenomenon analogous to generalized Bose condensation? If so, do they exhibit the phenomena typical of a superconductor, such as the Meissner effect and magnetic flux quantization?

APPENDIX

We shall investigate here the nature of the Hartree–Fock orbitals for a system of bosons in a container with impenetrable walls.

In order to avoid nonessential complications, we shall consider the special case of a cubical box and a repulsive delta-function interparticle interaction. Thus the Hamiltonian is

$$H = \sum_{i=1}^{N} -\tfrac{1}{2}\nabla_i^2 + \sum_{i<l}^{N} v_0\delta(\mathbf{x}_i - \mathbf{x}_l), \qquad \text{(A1)}$$

with $v_0 > 0$. We take box-enclosure boundary conditions:

$$\Phi(\mathbf{x}_1 \cdots \mathbf{x}_N) = 0, \quad x_i = \pm\tfrac{1}{2}L, \quad y_i = \pm\tfrac{1}{2}L, \quad z_i = \pm\tfrac{1}{2}L, \quad 1 \le j \le N, \qquad \text{(A2)}$$

where Φ is the many-boson wavefunction, x_i, y_i, z_i are the rectangular coordinates of $\mathbf{x}_i$, and $\Omega = L^3$. We are interested in the nature of the Hartree–Fock approximations to the exact eigenstates of H. For bosons the lowest Hartree–Fock state is of the form

$$\Phi_0(\mathbf{x}_1 \cdots \mathbf{x}_N) = \prod_{i=1}^{N} \varphi_0(\mathbf{x}_i), \qquad \text{(A3)}$$

where φ_0 (assumed normalized) satisfies the Hartree–Fock equation

$$[-\tfrac{1}{2}\nabla^2 + V(\mathbf{x})]\varphi_0(\mathbf{x}) = \epsilon_0\varphi_0(\mathbf{x}) \qquad \text{(A4)}$$

with a self-consistent potential

$$V(\mathbf{x}) = (N-1)\int_\Omega v_0\delta(\mathbf{x} - \mathbf{x}')\,|\varphi_0(\mathbf{x}')|^2\,d^3x' = (N-1)v_0\,|\varphi_0(\mathbf{x})|^2. \qquad \text{(A5)}$$

Since $\varphi_0(\mathbf{x})$ must vanish at the walls because of (A2), it follows that $V(\mathbf{x})$ also vanishes there; however, we shall find that $V(\mathbf{x})$ is essentially constant except within a microscopic distance of the walls. Since we wish to obtain a complete orthonormal set of orbitals, we shall also consider excited orbitals $\varphi_k(\mathbf{x})$ which are solutions of

$$[-\tfrac{1}{2}\nabla^2 + V(\mathbf{x})]\varphi_k(\mathbf{x}) = \epsilon_k\varphi_k(\mathbf{x}) \qquad \text{(A6)}$$

with the same potential V and satisfying the same boundary conditions. Just as (A4) arises from a variational treatment with a trial ground state of the form (A3), so (A6) arises from a variational treatment (with respect to variations of φ_k) with a single-excitation trial state of the form

$$\Phi_k(\mathbf{x}_1 \cdots \mathbf{x}_N) = N^{-\frac{1}{2}}\sum_{i=1}^{N} \varphi_0(\mathbf{x}_1)\cdots \times \varphi_0(\mathbf{x}_{i-1})\varphi_k(\mathbf{x}_i)\varphi_0(\mathbf{x}_{i+1}) \cdots \varphi_0(\mathbf{x}_N). \qquad \text{(A7)}$$

The physical significance of the eigenvalue parameters ϵ_0 and ϵ_k in (A4) and (A6) is that the energy expectation value of the state Φ_0 is $N\epsilon_0$, whereas the excitation energy of the state Φ_k is $\epsilon_k - \epsilon_0$. The reason for labeling the excited orbitals by a wave vector $\mathbf{k}$ will presently become clear.

Schafroth showed[31] that for the charged Bose gas φ_0 is constant except within a microscopic distance of the walls, where it drops to zero. Since the Coulomb interaction requires a slightly different treatment[31] because of its long range, and furthermore since we wish also to consider the excited orbitals, we shall redo the analysis here. Anticipating the final result, we write

$$\varphi_0(\mathbf{x}) = \Omega^{-\frac{1}{2}}[1 + \chi_0(\mathbf{x})], \qquad \text{(A8)}$$

[45] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
[46] See, e.g., A. J. Coleman, Phys. Rev. Letters **13**, 406 (1964), and other references cited therein.
[47] Section 18 of the first reference in Footnote 3.

where

$$\chi_0(\mathbf{x}) = -1 \quad \text{on the walls} \tag{A9}$$

in order to satisfy the boundary condition (A2); however, we expect $|\chi_0| \ll 1$ in the interior of the box, so that the normalization constant can differ only infinitesimally from $\Omega^{-\frac{1}{2}}$, and this difference can be absorbed into χ_0. Substitution into (A4) gives

$$[-\tfrac{1}{2}\nabla^2 + V(\mathbf{x}) - \epsilon_0]\chi_0(\mathbf{x}) = \epsilon_0 - V(\mathbf{x}), \tag{A10}$$

with

$$V(\mathbf{x}) = \rho \int_\Omega v_0 \delta(\mathbf{x} - \mathbf{x}') \, |1 + \chi_0(\mathbf{x}')|^2 \, d^3x'$$

$$= \rho v_0 \, |1 + \chi_0(\mathbf{x})|^2, \tag{A11}$$

since $(N - 1)/\Omega \to \rho$ in the thermodynamic limit. But if χ_0 is appreciable only in a region of extent λ (independent of N and Ω) near the walls, then it follows from (A8)–(A11) that

$$\epsilon_0 = \rho v_0 + O(\lambda/\Omega^{\frac{1}{3}}), \tag{A12}$$

since $\lambda\Omega^{\frac{1}{3}}$ is the volume within which χ_0 is appreciable and $\lambda\Omega^{\frac{1}{3}}/\Omega = \lambda/\Omega^{\frac{2}{3}}$. Thus in the interior of the box we may linearize (A10) [with (A11)] in the form

$$(-\tfrac{1}{2}\nabla^2 + 2\rho v_0)\chi_0(\mathbf{x}) = 0 \tag{A13}$$

aside from a term which vanishes in the thermodynamic limit; in expanding (A11) we have assumed without loss of generality (since the ground state is nondegenerate) that χ_0 is real.

Since χ_0 will be found to be appreciable only within a microscopic (i.e., independent of Ω) distance of the walls, it is extremely plausible that χ_0 will be a function only of the single variable measuring distance perpendicular to the wall, unless $\mathbf{x}$ is within a microscopic distance of an edge or corner. We therefore assume $\mathbf{x}$ to be near the wall $x = +\frac{1}{2}L$ or $x = -\frac{1}{2}L$ but not near any other, and look for a solution $\chi_0(x)$ independent of y and z (x means the x component of the vector $\mathbf{x}$). Then (A13) reduces to

$$-\tfrac{1}{2}\chi_0''(x) + 2\rho v_0 \chi_0(x) = 0. \tag{A14}$$

By symmetry $\chi_0(x)$ must be an even function of x; the most general real, even solution of (A14) is

$$\chi_0(x) = A \cosh(\alpha x), \qquad \alpha = 2(\rho v_0)^{\frac{1}{2}}, \tag{A15}$$

where A is an arbitrary real constant. Since we shall find that χ_0 is only appreciable near the walls $x = \pm\frac{1}{2}L$, one has

$$\chi_0(x) = \tfrac{1}{2}Ae^{\alpha x} \tag{A16}$$

near the right wall, except for a negligible term of order $Ae^{-\frac{1}{2}\alpha L}$; since χ_0 is symmetric, the left wall need not be considered explicitly. Eliminating A in terms of $\chi_0(\frac{1}{2}L - \lambda)$, one has

$$\chi_0(x) = \chi_0(\tfrac{1}{2}L - \lambda)e^{-\alpha(\frac{1}{2}L - \lambda - x)}, \tag{A17}$$

which shows that χ_0 falls off exponentially as x recedes from the boundary layer (of width λ) toward the interior of the box.

To complete the solution of the full nonlinear equation (A10) [with (A11)], one can choose the thickness λ of the boundary layer such that χ_0 is small enough outside the boundary layer to justify the linearization (A13), joining the analytical solution of the linear equation onto a numerical solution of the nonlinear equation which goes to zero at the walls. It is clear from (A17) that even the nonlinear equation (A10) separates to an excellent approximation except within a microscopic distance of edges (where the problem becomes two dimensional) and corners (where it becomes three dimensional). Thus, within a microscopic distance of the plane $x = \frac{1}{2}L$, χ_0 depends only on x away from edges and corners, so that (A10) [with (A11) and (A12)] becomes one dimensional:

$$-\tfrac{1}{2}\chi_0''(x) + 3\rho v_0\chi_0(x) + 2\rho v_0\chi_0^2(x) + 2\rho v_0\chi_0^3(x) = 0. \tag{A18}$$

To be specific, suppose

$$\chi_0(\tfrac{1}{2}L - \lambda) = -0.1. \tag{A19}$$

Then one may integrate (A18) numerically to the right, starting at $x = \frac{1}{2}L - \lambda$ where χ_0 and its derivative are fitted to (A17), until $\chi_0 = -1$; the value of λ must then be adjusted so that the value of x at which $\chi_0 = -1$ is $x = \frac{1}{2}L$. Defining new independent and dependent variables ξ and ψ_0 by

$$\xi = (\rho v_0)^{\frac{1}{2}}(x + \lambda - \tfrac{1}{2}L), \qquad \psi_0(\xi) = \chi_0(x), \tag{A20}$$

one obtains the following equation for ψ_0:

$$\psi_0'' = 2\psi_0(3 + 2\psi_0 + 2\psi_0^2), \tag{A21}$$

which is to be solved subject to the initial conditions

$$\psi_0(0) = -0.1, \qquad \psi_0'(0) = -0.2, \tag{A22}$$

which follow from (A20), (A19), (A17), and (A15). Multiplying (A21) by ψ_0', one finds

$$(d/d\xi)(\psi_0')^2 = 4\psi_0'\psi_0(3 + 2\psi_0 + 2\psi_0^2), \tag{A23}$$

which can be integrated immediately to give

$$[\psi_0'(\xi)]^2 - [\psi_0'(0)]^2 = F(\psi_0(\xi)) - F(\psi_0(0)),$$
$$F(\psi_0) = 2\psi_0^2(3 + \tfrac{4}{3}\psi_0 + \psi_0^2). \tag{A24}$$

Thus, since $\psi_0' < 0$,

$$\psi_0'(\xi) = -[A + F(\psi_0(\xi))]^{\frac{1}{2}}, \tag{A25}$$

with

$$A = [\psi_0'(0)]^2 - F(\psi_0(0)) = -0.0164. \tag{A26}$$

Then

$$\int_{\psi_0(0)}^{\psi_0(\xi)} [A + F(t)]^{-\frac{1}{2}}\, dt = -\xi. \tag{A27}$$

The value, ξ_w, of ξ at which ψ_0 goes to -1 is then given by

$$\xi_w = -\int_{-0.1}^{-1} [A + F(t)]^{-\frac{1}{2}}\, dt$$
$$= \int_{0.1}^{1} (-0.0164 + 6t^2 - \tfrac{8}{3}t^3 + 2t^4)^{-\frac{1}{2}}\, dt = 1.2. \tag{A28}$$

Then, since $x = \frac{1}{2}L$ at the wall, (A20) gives the thickness λ of the boundary layer, nominally chosen so that $|\chi_0| = 0.1$ at the edge of the boundary layer:

$$\lambda = 1.2(\rho v_0)^{-\frac{1}{2}}. \tag{A29}$$

Within a microscopic distance of edges or corners, the assumption that χ_0 depends only on the distance from one wall (the wall which $\mathbf{x}$ is very near) fails; to obtain the solution of (A10) in such exceptional regions, one would have to solve (A10), (A11) numerically subject to (A9), adjusting $\nabla\chi_0$ on the walls so that χ_0 joins on smoothly to the previously obtained solutions. However, we shall not consider these regions, since their effect is clearly negligible in the thermodynamic limit.

We still have to investigate the solutions of (A6) for $\mathbf{k} \neq 0$. In order to see what the proper generalization of (A8) is, it is necessary first to state our result for the case $\mathbf{k} = 0$ in a form which admits such a generalization. We found that χ_0 in (A8) was appreciable only in a microscopic boundary layer lining the walls of the cubical box within which the system of bosons is confined; throughout the interior of the box, $\varphi_0(\mathbf{x})$ is merely equal to the constant $\Omega^{-\frac{1}{2}}$. The first thought that comes to mind is that this constant $\Omega^{-\frac{1}{2}}$ is the lowest free-particle state for periodic boundary conditions with a periodicity cube of volume Ω; hence one is tempted to look for solutions $\varphi_\mathbf{k}$ which differ from plane waves $\Omega^{-\frac{1}{2}}e^{i\mathbf{k}\cdot\mathbf{x}}$ only in the boundary layer, where they fall to zero. However, no such solutions of (A6) can exist for $\mathbf{k} \neq 0$, since such a solution would have current $\mathbf{k} \neq 0$ in the interior but zero current at the walls (since $\varphi_\mathbf{k}$ must vanish there), thereby violating the equation of continuity. The way around this difficulty is, following Schafroth, to note that the constant $\Omega^{-\frac{1}{2}}$ is also the lowest solution satisfying homogeneous Neumann boundary conditions (55). The free-particle orbitals $\varphi_\mathbf{k}$ [(56), (57)] satisfying this boundary condition are real and therefore have zero current, so that no contradiction arises if one modifies such $\varphi_\mathbf{k}$ in the boundary layer so as to make them vanish on the walls. We therefore make the ansatz

$$\varphi_\mathbf{k}(\mathbf{x}) = \varphi_\mathbf{k}^N(\mathbf{x})[1 + \chi_\mathbf{k}(\mathbf{x})], \tag{A30}$$

where $\varphi_\mathbf{k}^N$ is what is called $\varphi_\mathbf{k}$ in (56) and (57), namely the free-particle orbital satisfying the Neumann condition (55). We then seek solutions of (A6) for which $\chi_\mathbf{k}$ is negligible in the interior of the box, but

$$\chi_\mathbf{k}(\mathbf{x}) = -1 \quad \text{on the walls,} \tag{A31}$$

generalizing (A9). The appropriate generalization of (A10) is found with the aid of (A30), (56), (57), and (A6) to be

$$\{-\tfrac{1}{2}\nabla^2 + \tfrac{1}{2}k^2 + V(\mathbf{x}) + k_x \tan[k_x(x + \tfrac{1}{2}L)]\partial/\partial x + k_y \tan[k_y(y + \tfrac{1}{2}L)]\partial/\partial y + k_z \tan[k_z(z + \tfrac{1}{2}L)]\partial/\partial z - \epsilon_\mathbf{k}\}\chi_\mathbf{k}(\mathbf{x}) = \epsilon_\mathbf{k} - \tfrac{1}{2}k^2 - V(\mathbf{x}), \tag{A32}$$

where $V(\mathbf{x})$ is the same as in (A11) (with χ_0, not $\chi_\mathbf{k}$) and we have assumed that all components of $\mathbf{k}$ are nonzero, since the exceptions form a set of measure zero. Since we are seeking a solution for which $\chi_\mathbf{k}$ vanishes in the interior, where $V(\mathbf{x}) = \rho v_0$ by (A11), it is clear that the appropriate generalization of (A12) is

$$\epsilon_\mathbf{k} = \rho v_0 + \tfrac{1}{2}k^2 + O(\lambda/\Omega^{\frac{1}{2}}). \tag{A33}$$

As in the case $\mathbf{k} = 0$, we assume $\mathbf{x}$ to be near the wall $x = +\frac{1}{2}L$ but not near an edge or corner, and hence seek a solution $g_\mathbf{k}$ depending only on x. Then, introducing the new variable

$$t = \tfrac{1}{2}L - x, \tag{A34}$$

one finds by (A32), (A34), and (A11)

$$-\tfrac{1}{2}g_\mathbf{k}''(t) + 2\rho v_0 g_0(t)[1 + \tfrac{1}{2}g_0(t)]g_\mathbf{k}(t) + k_x \tan(k_x t)g_\mathbf{k}'(t) = -2\rho v_0 g_0(t)[1 + \tfrac{1}{2}g_0(t)], \tag{A35}$$

where $g_0(t) = \chi_0(x)$ and $g_\mathbf{k}(t) = \chi_\mathbf{k}(x)$.

In the interior we can insert the approximation (A17), (A19), which in the present notation reads

$$g_0(t) = g_0(\lambda)e^{-\alpha(t-\lambda)} = -0.1e^{-\alpha(t-\lambda)}, \; t > \lambda, \quad \text{(A36)}$$

and can neglect $\frac{1}{2}g_0(t)$ compared to unity in the terms $1 + \frac{1}{2}g_0(t)$; however, the term $g_0(t)g_k(t)$ in (A35) cannot be neglected due to the singularities of $\tan(k_x t)$ at $k_x t = \frac{1}{2}\pi, \frac{3}{2}\pi, \cdots$. Then (A35) becomes

$$-\tfrac{1}{2}g_k''(t) + k_x \tan(k_x t) g_k'(t) - 0.2\rho v_0 e^{-\alpha(t-\lambda)} g_k(t) = 0.2\rho v_0 e^{-\alpha(t-\lambda)}, \; t > \lambda. \quad \text{(A37)}$$

In view of the singularities of the differential equation (A37) at $k_x t = \frac{1}{2}\pi, \frac{3}{2}\pi, \cdots$, it is convenient to make a change of variables; since the properties of the hypergeometric equation are very well known, it is convenient to choose the new variables so that the transformed differential equation resembles the hypergeometric equation as closely as possible. This is achieved by the substitutions

$$u = \tfrac{1}{2}[1 - \sin(k_x t)], \qquad \psi_k(u) = 1 + g_k(t). \quad \text{(A38)}$$

Then (A37) becomes a homogeneous equation

$$u(1-u)\psi_k''(u) + \tfrac{3}{2}(1-2u)\psi_k'(u) + (0.4\rho v_0/k_x^2)e^{\alpha\lambda}e^{-\alpha k_x^{-1}\sin^{-1}(1-2u)}\psi_k(u) = 0, \quad \text{(A39)}$$

which differs from the hypergeometric equation only through the presence of the exponential factor in the last term. The only singularities are at $u = 0$, 1, and ∞; $u = 0$ corresponds to $k_x t = \pi/2, 5\pi/2, 9\pi/2, \cdots$, $u = 1$ corresponds to $k_x t = 3\pi/2, 7\pi/2, 11\pi/2, \cdots$, whereas $u = \infty$ is not relevant because $0 \leq u \leq 1$ for real t. The indicial equation[48] at $u = 0$ is the same as that for the hypergeometric equation, and the conclusion is therefore the same: of the two linearly independent solutions which must be combined to form the general solution ψ_k, one is analytic at $u = 0$ and the other is equal to $u^{\frac{1}{2}}$ times a function analytic at $u = 0$. It follows that at $k_x t = \pi/2, 5\pi/2, \cdots$, $g_k(t)$ will behave like $[1 - \sin(k_x t)]^{-\frac{1}{2}}$. However, we note from (A30), (57), and (A34) that in obtaining φ_k, g_k is to be multiplied by $\cos(k_x t) = \pm[1 - \sin^2(k_x t)]^{\frac{1}{2}}$. The product therefore behaves like

$$[1 - \sin(k_x t)]^{-\frac{1}{2}}[1 - \sin^2(k_x t)]^{\frac{1}{2}} = [1 + \sin(k_x t)]^{\frac{1}{2}} = 2^{\frac{1}{2}}$$

at $k_x t = \pi/2, 5\pi/2, \cdots$, i.e. these singularities of g_k do not cause any singularities of φ_k. A similar argument applies to the points $k_x t = 3\pi/2, 7\pi/2, \cdots$. In fact, we shall find that the only effect of the singularities of g_k in the interior of the box is to cause a small displacement of the nodes of φ_k away from the positions of those of the unperturbed functions (56), (57), the amount of the displacement decreasing as one recedes from the walls, which are, in fact, responsible for the perturbation [g_k was introduced in order to take into account the effect of the requirement that φ_k must vanish at the walls instead of satisfying (55)].

In order to see this more clearly, it is convenient to make the substitution

$$g_k = h_k \sec(k_x t) - 1. \quad \text{(A40)}$$

Then (A35) reduces to

$$h_k''(t) + \{k_x^2 - 4\rho v_0 g_0(t) \times [1 + \tfrac{1}{2}g_0(t)]\}h_k(t) = 0. \quad \text{(A41)}$$

Although this cannot be solved analytically for general values of k_x, if $k_x \gg \lambda^{-1}$ (λ is the distance in which g_0 varies appreciably) one can use the WKBJ approximation:

$$h_k(t) \approx A \sin\{[k_x^2 - U(t)]^{\frac{1}{2}}t\} + B \cos\{[k_x^2 - U(t)]^{\frac{1}{2}}t\}, \quad \text{(A42)}$$

where

$$U(t) = 4\rho v_0 g_0(t)[1 + \tfrac{1}{2}g_0(t)]. \quad \text{(A43)}$$

Then, since $g_k(0) = -1$ (so φ_k vanishes at the wall), it is necessary to take $B = 0$. Furthermore, normalization and reality require $A = \pm 1$, and since g_k increases from -1 as t recedes from the wall one must choose $A = +1$. But since φ_k is proportional to $(1 + g_k)\cos(k_x t) = h_k$ whereas φ_k^N is proportional to $\cos(k_x t)$ [see (A30)], one concludes that far in the interior the only perturbation of the Neumann free-particle solution due to the wall and the self-consistent field is a phase shift of $\pi/2$.[49]

Although we have not been able to give rigorous derivations because of the nonlinearity of the Hartree–Fock equation, we believe that the foregoing

[48] See, e.g., E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge at the University Press, Cambridge, England, 1927), 4th ed., Chaps X and XIV.

[49] There is also a small secular phase shift beyond the accuracy of the WKBJ approximation, which is negligible over distances small compared to L (the length of the box) but accumulates to another $\frac{1}{2}\pi$ by the time one reaches the center of the box. This can be seen from the fact that in the neighborhood of the center of the box ($t = \frac{1}{2}L$) one must take $A = 0$, $B = 1$ in (A42) in order to satisfy the requirement that φ_k be even or odd under reflection about the center of the box according to whether Lk_x/π is an even or an odd integer.

analysis shows fairly convincingly that the Hartree–Fock orbitals φ_k, the solutions of (A4) and (A6), differ from the free-particle orbitals (56), (57) satisfying the Neumann boundary condition (55) only through a distortion in a microscopic boundary layer near the walls (so as to vanish there) plus a thermodynamically negligible phase shift (microscopic shift in the positions of the nodes). Although the analysis was carried out for the special case of a repulsive delta-function interaction, one expects the same behavior for any finite-range repulsive interaction whose Fourier transform exists, since the "wall potential" $V(\mathbf{x}) - \rho v_0$, where v_0 is the Fourier transform of the interaction at $\mathbf{k} = 0$, will be of the same qualitative form in such cases. It is much less clear how much "attractive tail" the interparticle interaction may have without disturbing these conclusions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Schrödinger Basis for Spinor Representations of the Three-Dimensional Rotation Group

DAVE PANDRES, JR.
Advanced Research Laboratory, Douglas Aircraft Company, Inc., Santa Monica, California
(Received 28 September 1964)

It is shown that the double-valued spherical harmonics provide a basis for the irreducible spinor representations of the three-dimensional rotation group. Pauli's assertion to the contrary is shown to be false. Both infinitesimal and finite rotations are discussed in some detail. It is also shown that there remains a twofold degeneracy in the spherical harmonic Y_{jm} when j and m are specified.

I. INTRODUCTION

LET J_a be a set of Schrödinger-type operators defined by

$$J_1 = i(\cot\theta\cos\phi\partial/\partial\phi + \sin\phi\partial/\partial\theta),$$
$$J_2 = i(\cot\theta\sin\phi\partial/\partial\phi - \cos\phi\partial/\partial\theta),$$
$$J_3 = -i\partial/\partial\phi, \tag{1}$$

where θ and ϕ are, respectively, the usual polar and azimuthal angles locating a point P on the surface of a sphere. The J_a operate upon one-component complex functions Ψ which depend only upon θ and ϕ. Now, while Ψ and J_a bear a formal similarity to the wavefunctions and orbital angular momentum operators of ordinary nonrelativistic quantum mechanics, it is known that we need not interpret them in this way. Instead, we may regard the J_a simply as generators of the infinitesimal rotations associated with the three-dimensional rotation group $O(3)$, and may regard the Ψ simply as elements of a vector space (function space) upon which the J_a are defined.

We wish to consider the functions $Y_{jm}(\theta, \phi)$ which satisfy the equations

$$(J_1^2 + J_2^2 + J_3^2)Y_{jm} = j(j+1)Y_{jm},$$
$$J_3 Y_{jm} = mY_{jm}, \tag{2}$$

and which provide the basis for a $(2j+1)$-dimensional irreducible representation D^j of $O(3)$. It is well known that, for integer values of j, the Y_{jm} are the familiar single-valued spherical harmonics. They provide a basis for the tensor representations of $O(3)$. The spherical harmonics for half-integer values of j are also known.[1] They are double-valued functions of P. It is generally believed, however, that they fail to provide a basis for the genuine spinor representations of $O(3)$. We shall show that this belief is incorrect.

It is convenient to begin by recalling, briefly, why double-valued quantum mechanical wavefunctions have conventionally been regarded as inadmissible. Schrödinger[2] originally required that wavefunctions be single valued because he believed that they must correspond directly to some observable property of a physical system. Pauli[3] early recognized that, since wavefunctions correspond only indirectly to observables, there is no *a priori* reason why they must be single valued. (In this spirit, Bohm[4] has remarked

[1] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), p. 783.
[2] E. Schrödinger, Ann. Physik 79, 361, 489 (1926).
[3] W. Pauli in *Handbuch der Physik*, edited by Geiger Sheek (Verlag Julius Springer, Berlin, 1933), Vol. 24, p. 121.
[4] D. Bohm, *Quantum Theory* (Methuen and Company, Ltd., London, 1934), p. 104.

analysis shows fairly convincingly that the Hartree–Fock orbitals φ_k, the solutions of (A4) and (A6), differ from the free-particle orbitals (56), (57) satisfying the Neumann boundary condition (55) only through a distortion in a microscopic boundary layer near the walls (so as to vanish there) plus a thermodynamically negligible phase shift (microscopic shift in the positions of the nodes). Although the analysis was carried out for the special case of a repulsive delta-function interaction, one expects the same behavior for any finite-range repulsive interaction whose Fourier transform exists, since the "wall potential" $V(\mathbf{x}) - \rho v_0$, where v_0 is the Fourier transform of the interaction at $\mathbf{k} = 0$, will be of the same qualitative form in such cases. It is much less clear how much "attractive tail" the interparticle interaction may have without disturbing these conclusions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Schrödinger Basis for Spinor Representations of the Three-Dimensional Rotation Group

DAVE PANDRES, JR.
Advanced Research Laboratory, Douglas Aircraft Company, Inc., Santa Monica, California
(Received 28 September 1964)

It is shown that the double-valued spherical harmonics provide a basis for the irreducible spinor representations of the three-dimensional rotation group. Pauli's assertion to the contrary is shown to be false. Both infinitesimal and finite rotations are discussed in some detail. It is also shown that there remains a twofold degeneracy in the spherical harmonic Y_{jm} when j and m are specified.

I. INTRODUCTION

LET J_a be a set of Schrödinger-type operators defined by

$$J_1 = i(\cot\theta\cos\phi\partial/\partial\phi + \sin\phi\partial/\partial\theta),$$
$$J_2 = i(\cot\theta\sin\phi\partial/\partial\phi - \cos\phi\partial/\partial\theta),$$
$$J_3 = -i\partial/\partial\phi, \tag{1}$$

where θ and ϕ are, respectively, the usual polar and azimuthal angles locating a point P on the surface of a sphere. The J_a operate upon one-component complex functions Ψ which depend only upon θ and ϕ. Now, while Ψ and J_a bear a formal similarity to the wavefunctions and orbital angular momentum operators of ordinary nonrelativistic quantum mechanics, it is known that we need not interpret them in this way. Instead, we may regard the J_a simply as generators of the infinitesimal rotations associated with the three-dimensional rotation group $O(3)$, and may regard the Ψ simply as elements of a vector space (function space) upon which the J_a are defined.

We wish to consider the functions $Y_{jm}(\theta, \phi)$ which satisfy the equations

$$(J_1^2 + J_2^2 + J_3^2)Y_{jm} = j(j+1)Y_{jm},$$
$$J_3 Y_{jm} = mY_{jm}, \tag{2}$$

and which provide the basis for a $(2j+1)$-dimensional irreducible representation D^j of $O(3)$. It is well known that, for integer values of j, the Y_{jm} are the familiar single-valued spherical harmonics. They provide a basis for the tensor representations of $O(3)$. The spherical harmonics for half-integer values of j are also known.[1] They are double-valued functions of P. It is generally believed, however, that they fail to provide a basis for the genuine spinor representations of $O(3)$. We shall show that this belief is incorrect.

It is convenient to begin by recalling, briefly, why double-valued quantum mechanical wavefunctions have conventionally been regarded as inadmissible. Schrödinger[2] originally required that wavefunctions be single valued because he believed that they must correspond directly to some observable property of a physical system. Pauli[3] early recognized that, since wavefunctions correspond only indirectly to observables, there is no *a priori* reason why they must be single valued. (In this spirit, Bohm[4] has remarked

[1] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), p. 783.
[2] E. Schrödinger, Ann. Physik 79, 361, 489 (1926).
[3] W. Pauli in *Handbuch der Physik*, edited by Geiger Sheek (Verlag Julius Springer, Berlin, 1933), Vol. 24, p. 121.
[4] D. Bohm, *Quantum Theory* (Methuen and Company, Ltd., London, 1934), p. 104.

that, logically, one can only demand that wavefunctions shall satisfy conditions which insure that the average value of every observable is single valued.) Pauli initially argued, however, that multivalued wavefunctions give rise to sources and sinks of probability current, and so are physically inadmissible. It is this argument of Pauli's to which Blatt and Weisskopf[1] refer, although they attribute it to Nordsieck. The argument also appears in other current textbooks,[5] although Pauli[6] later recognized that it is incorrect. He noted that there is a large class of multivalued angular momentum eigenfunctions with well-behaved probability currents, and that the double-valued spherical harmonics are of this class. Pauli then gave another argument in which he concluded that only single-valued wavefunctions are admissible. He considered only single-valued and double-valued functions as possibilities because Schrödinger[7] had shown that only these behave properly under time reversal. (One can also exclude all other functions by noting that it follows from purely abstract arguments[8] that the J_a have only integer and half-integer eigenvalues. Only single-valued and double-valued functions correspond to these eigenvalues.) The quadratically integrable, but not normalized, spherical harmonics for $m = j$ are of the form

$$Y_{jj} = (\sin\theta)^j \epsilon^{ij\phi}. \tag{3}$$

Let $J^{\pm} = J_1 \pm iJ_2$ be the usual raising and lowering operators. We obtain the functions Y_{jm} for m equal to $j - 1, \cdots, -j$ by repeated application of J^- to Y_{jj}. Thus, for $m = j, \cdots, 1 - j$ we write

$$J^- Y_{jm} = N_{jm} Y_{j,m-1}, \tag{4}$$

where N_{jm} is an arbitrary constant. For convenience, we take $N_{jm} = [(j + m)(j - m + 1)]^{\frac{1}{2}}$. Upon applying J^+ to both sides of Eq. (4) and using the identity

$$J^{\pm} J^{\mp} Y_{jm} = [(j \pm m)(j \mp m + 1)] Y_{jm}, \tag{5}$$

we find that

$$J^+ Y_{jm} = [(j - m)(j + m + 1)]^{\frac{1}{2}} Y_{j,m+1} \tag{6}$$

for $m = j - 1, \cdots, -j$.

Now let $\mathcal{S}_j$ be a function space which is spanned by the basis functions Y_{jm} for a given value of j. Pauli rejected the Y_{jm} with half-integer j because they fail to satisfy his requirement that the application of J_a to any element of $\mathcal{S}_j$ shall lead again to an element of $\mathcal{S}_j$. Pauli's reason for making this requirement is contained in his assertion that, otherwise, no unambiguous correlation between Schrödinger operator calculus and matrix calculus would exist. In particular, he asserted that the Y_{jm} with half-integer j fail to provide the basis for an irreducible representation D^j. We shall show that these assertions are false.

We wish to emphasize that we have no quarrel with Pauli's conclusion concerning the inadmissibility of multivalued quantum mechanical wavefunctions, although we take issue with the argument through which he reached that conclusion. Indeed, we have shown[9] previously that the components of linear momentum are not simultaneously observable if multivalued functions are admitted. A simpler argument leading to the same conclusion may be mentioned here: If the components of linear momentum are simultaneously observable, then any admissible wavefunction must be expressible by a superposition of their simultaneous eigenfunctions. Such a superposition is just a Fourier integral, which is always a single-valued function.

On the other hand, we wish to emphasize that the Y_{jm} with half-integer j do, in fact, yield an unambiguous correlation between operator calculus and matrix calculus; and that they do provide the basis for an irreducible representation D^j of $O(3)$. We shall see that if scalar products are defined as in ordinary quantum mechanics, then the representation D^j is unitary for $j = \frac{1}{2}$, nonunitary for $j = \frac{3}{2}, \frac{5}{2}, \cdots$. However, a slight change in the definition of the scalar product results in all representations being unitary. Finally, we shall show that when j and m are fixed (at either integer or half-integer values) there remains a twofold degeneracy in the corresponding spherical harmonic.

II. A CRITIQUE OF PAULI'S ARGUMENT

Let the Dirac ket vector $|jm\rangle$ correspond to Y_{jm} and let $\langle jm|$ be its conjugate imaginary bra vector. An arbitrary bra vector $\langle\Psi|$ may be expressed in the form

$$\langle\Psi| = \sum_{m=-j}^{j} \frac{\langle\Psi \mid jm\rangle}{\langle jm \mid jm\rangle} \langle jm|. \tag{7}$$

Now, suppose that we have a given function Ψ, and that we wish to find the bra vector $\langle\Psi|$ which corresponds to it. If scalar products are defined as

[5] R. M. Sillito, *Quantum Mechanics* (Edinburgh University Press, Edinburgh, Scotland, 1960), p. 51.
[6] W. Pauli, Helv. Phys. Acta **12**, 147 (1939).
[7] E. Schrödinger, Ann. Physik **32**, 49 (1938).
[8] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., p. 141.
[9] D. Pandres, Jr., J. Math. Phys. **3**, 305 (1962).

in ordinary quantum mechanics; i.e.,

$$\langle \Psi \mid \Phi \rangle = \int_0^{2\pi} \int_0^{\pi} \Psi^* \Phi \sin\theta \, d\theta \, d\phi, \tag{8}$$

then $\langle \Psi | jm \rangle$ vanishes if Ψ and Y_{jm} are orthogonal. But, suppose that Ψ is orthogonal to all of the Y_{jm} for the given value of j. It is then clear from Eq. (7) that $\langle \Psi | = 0$. Indeed, Dirac[10] has pointed out explicitly that "a bra vector is considered to be completely defined when its scalar product with every ket vector is given; so that if a bra vector has its scalar product with every ket vector vanishing, the bra vector itself must be considered as vanishing. In symbols, if $\langle P|A \rangle = 0$, all $|A\rangle$, then $\langle P| = 0$." It is precisely this point which is not taken into account in Pauli's argument.

Pauli's assertion that the functions Y_{jm} with half-integer j fail to provide a basis for a representation D^j rests upon the fact that although $J^+ Y_{jj}$ vanishes, the function

$$Q = J^- Y_{j,-j} \tag{9}$$

does not vanish. It is not an element of $\mathcal{S}_j$, and it is certainly inadmissible because it is an eigenfunction with a negative eigenvalue of the nonnegative definite operator $J_1^2 + J_2^2$. We wish to point out, however, that Q is orthogonal to every element of $\mathcal{S}_j$ (because of its ϕ dependence, which is just $\epsilon^{-i(j+1)\phi}$). It is therefore clear that Q is a representative of the vanishing bra vector even though it is not of vanishing functional form. Furthermore, if J^- is applied repeatedly to Q, the result is always a representative of the vanishing bra vector. It seems at first glance that if J^+ is applied to Q, this must lead back into $\mathcal{S}_j$. This would be unsatisfactory since, for example, the matrix representing J^+J^- on the $\mathcal{S}_j$ basis would not equal the product of the matrices representing J^+ and J^- on the same basis. There is no problem, however, because Eq. (5) insures that J^+Q vanishes identically. We see that although the J_a lead us out of $\mathcal{S}_j$, they can never lead us back in again.

It is instructive to examine the representation of finite rotations on the $\mathcal{S}_j$ basis. Every element of $\mathcal{S}_j$ is of the form

$$\Psi = \sum_{m=-j}^{j} C_m Y_{jm}(\theta, \phi), \tag{10}$$

where the C_m are constants. Consider a rotation through the angle B about an axis specified by the unit vector $\mathbf{n}$. The corresponding rotation operator is

$$U = \exp(-iB\mathbf{n}\cdot\mathbf{J}). \tag{11}$$

[10] P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press, New York, 1958), 4th ed., p. 20.

It is obvious that U does not transform Ψ as a scalar. The point is that $U\Psi$ is double valued on encirclement of the $\theta = 0, \pi$ axis, while the function $\bar{\Psi}$ which results from rotating Ψ as a scalar is double valued on encirclement of some new $\bar{\theta} = 0, \pi$ axis into which the $\theta = 0, \pi$ axis has been rotated. By the transformation of Ψ as a scalar, we mean that $\bar{\Psi}$ depends upon $\bar{\theta}, \bar{\phi}$ in the same way as Ψ depends upon θ, ϕ and that $\bar{\theta}, \bar{\phi}$ are referred to rotated axes in the same way as θ, ϕ are referred to the original axes. Thus, we have

$$\bar{\Psi} = \sum_{m=-j}^{j} C_m Y_{jm}(\bar{\theta}, \bar{\phi}). \tag{12}$$

Notice that $\bar{\Psi}$ may be regarded as a function of θ, ϕ, $\mathbf{n}$, and the rotation angle B. When $\bar{\Psi}$ is expanded into an infinite series in powers of B, there is a nonvanishing remainder term. Because of this term, the series converges not to $\bar{\Psi}$ but rather to $U\Psi$. This establishes an interesting functional correspondence between $\bar{\Psi}$ and $U\Psi$, and enables us to obtain a double-valued function by superposition of any two double-valued functions. For example, the function $\Psi + U\Psi$ is only double valued although the function $\Psi + \bar{\Psi}$ is four valued.

Now, the function $U\Psi$ is not an element of $\mathcal{S}_j$. We find instead that

$$U\left(\sum_{m=-j}^{j} C_m Y_{jm}\right) = \sum_{m=-j}^{j} C_m' Y_{jm} + R, \tag{13}$$

where R is a representative of the vanishing bra vector. The constants C_m' are given in terms of the C_m by

$$VC = C', \tag{14}$$

where C is the column matrix

$$\begin{pmatrix} C_j \\ \vdots \\ C_{-j} \end{pmatrix},$$

C' is the column matrix

$$\begin{pmatrix} C_j' \\ \vdots \\ C_{-j}' \end{pmatrix},$$

and V is a unitary matrix given by

$$V = \exp(-iB\mathbf{n}\cdot\boldsymbol{\sigma}). \tag{15}$$

Here the components $\sigma_1, \sigma_2, \sigma_3$ of $\boldsymbol{\sigma}$ are the Hermitian matrices which comprise the well-known spinor representations of $O(3)$. For example, if $j = \frac{1}{2}$, we have the Pauli matrices

$$\sigma_1 = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \tfrac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and, if $j = \frac{3}{2}$,

$$\sigma_1 = \tfrac{1}{2}\begin{pmatrix} 0 & \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 2 & 0 \\ 0 & 2 & 0 & \sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix},$$

$$\sigma_2 = \tfrac{1}{2}\begin{pmatrix} 0 & -\sqrt{3}i & 0 & 0 \\ \sqrt{3}i & 0 & -2i & 0 \\ 0 & 2i & 0 & -\sqrt{3}i \\ 0 & 0 & \sqrt{3}i & 0 \end{pmatrix},$$

$$\sigma_3 = \tfrac{1}{2}\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

The importance of Eq. (13) lies in the fact that it establishes a *functional* interpretation for the genuine spinor representations of $O(3)$. If, in Eq. (4), we had chosen some other value for the constant N_{jm}, the matrices σ_a would have been non-Hermitian corresponding to a nonunitary irreducible representation of $O(3)$.

If one uses the standard formulas of quantum theory to compute matrices M_a representing the J_a, one finds that the element in the rth row and cth column of M_a is

$$(M_a)_{rc} = \langle \psi_r | J_a | \psi_c \rangle / \langle \psi_r \mid \psi_r \rangle^{\frac{1}{2}} \langle \psi_c \mid \psi_c \rangle^{\frac{1}{2}}, \tag{16}$$

where $\psi_r = Y_{j,j+1-r}$. One finds that the M_a obtained in this way do indeed provide $(2j+1)$-dimensional irreducible spinor representations of $O(3)$. The M_a are Hermitian for $j = \frac{1}{2}$, but are not Hermitian for $j = \frac{3}{2}, \frac{5}{2}, \cdots$. This means that the corresponding representations are unitary for $j = \frac{1}{2}$, nonunitary for $j = \frac{3}{2}, \frac{5}{2}, \cdots$. The nonunitary arises because $\langle jm|jm \rangle$ is infinite for $m = -\frac{3}{2}, -\frac{5}{2}, \cdots, -j$. But, it is well known[11] that it is always possible to redefine the scalar product such that the representations to $O(3)$ are unitary. Let us consider how this can be done.

[11] I. M. Gel'fand and Y. Ya Sapiro, "The Representations of the Group of Rotations in Three-Dimensional Space and their Applications," Am. Math. Soc. Trans., Series 2, Vol. 2 (1956).

We let $|\Psi\rangle$ be represented in the usual way by Ψ, but instead of letting $\langle \Psi|$ be represented by Ψ^*, as is customary, we let $\langle \Psi|$ be represented by $(G\Psi^*)$, where G is an operator whose properties are to be determined. (The situation is rather similar to that which exists in field theory where $\Psi^\dagger$ differs from Ψ^*.) Thus, instead of Eq. (8), we write

$$\langle \Psi \mid \Phi \rangle = \int_0^{2\pi} \int_0^{\pi} (G\Psi)^* \Phi \sin\theta \, d\theta \, d\phi. \tag{17}$$

If the space $\mathcal{S}_j$ is to be a Hilbert space, we must have

$$\langle \Psi \mid \Phi \rangle = \langle \Phi \mid \Psi \rangle^*. \tag{18}$$

This just requires that

$$\int_0^{2\pi} \int_0^{\pi} [(G\Psi)^* \Phi - \Psi^* G\Phi] \sin\theta \, d\theta \, d\phi = 0, \tag{19}$$

i.e., that G be Hermitian in the sense of ordinary quantum theory. We now define G by requiring that

$$\langle \psi_r \mid \psi_c \rangle = \delta_{rc}, \tag{20}$$

where δ_{rc} is the usual Kronecker symbol. It follows from Eq. (20) that the element in the rth row and cth column of the matrix representing G is just

$$G_{rc} = (S^{-1}_{rc})^*, \tag{21}$$

where S^{-1} is the inverse to the matrix of overlap integrals; i.e.,

$$S_{rc} = \int_0^{2\pi} \int_0^{\pi} \psi_r^* \psi_c \sin\theta \, d\theta \, d\phi. \tag{22}$$

With this choice of G, the M_a generated by Eq. (16) reduce to the Hermitian σ_a of Eq. (15). We mention in passing that G plays the role of a metric in a Hilbert space whose covariant and contravariant basis vectors are the $\langle \psi_n|$ and $|\psi_n\rangle$, respectively. Notice that if $\Psi = \sum C_n \psi_n$ and $\Phi = \sum K_n \psi_n$ then

$$\langle \Psi \mid \Phi \rangle = \sum C_n^* K_n, \tag{23}$$

as one would expect. Notice also that, for the spaces $\mathcal{S}_j$ considered in this paper, G and S are diagonal matrices which reduce to the identity if the ψ_n are normalized to unity.

III. DEGENERACY OF THE SPHERICAL HARMONICS

We wish to point out that, instead of defining Y_{jj} as in Eq. (3), we could have written

$$Y_{jj} = \alpha_{\frac{1}{2}} Y^{\frac{1}{2}}_{jj} + \alpha_{-\frac{1}{2}} Y^{-\frac{1}{2}}_{jj}, \tag{24}$$

where

$$Y^{\frac{1}{2}}_{jj} = (\sin\theta)^j \epsilon^{ij\phi},$$

$$Y^{-\frac{1}{2}}_{jj} = (\sin\theta)^j \epsilon^{ij\phi} \int^{\cos\theta} (1 - x^2)^{-(j+1)} \, dx,$$

and $\alpha_{\frac{1}{2}}$ and $\alpha_{-\frac{1}{2}}$ are arbitrary constants. Equation (24) gives the general solution of Eq. (2) for $m = j$. It is clear that $Y^{\frac{1}{2}}_{jj}$, and $Y^{-\frac{1}{2}}_{jj}$ are linearly independent for both integer and half-integer j, because they are even and odd functions of cos θ, respectively. The point is that a spherical harmonic is not uniquely determined when j and m are specified. A twofold degeneracy remains because the invariant operator $J_1^2 + J_2^2 + J_3^2$ involves second derivatives with respect to θ. We may remove this degeneracy by classifying our functions by eigenvalues Λ of the operator K_3 defined by

$$K_3 Y^{\Lambda}_{jm} = \Lambda Y^{\Lambda}_{jm}. \tag{25}$$

We have, as yet, no operator connecting $Y^{\frac{1}{2}}_{jm}$ and $Y^{-\frac{1}{2}}_{jm}$. However, we may define such operators $K^{\pm} = K_1 \pm iK_2$ by the relations

$$K^{+} Y^{-\frac{1}{2}}_{jm} = Y^{\frac{1}{2}}_{jm},$$

$$K^{-} Y^{\frac{1}{2}}_{jm} = Y^{-\frac{1}{2}}_{jm}, \tag{26}$$

$$K^{+} Y^{\frac{1}{2}}_{jm} = K^{-} Y^{-\frac{1}{2}}_{jm} = 0.$$

The matrices which represent K_a on the basis $Y^{\frac{1}{2}}_{jm}$, $Y^{-\frac{1}{2}}_{jm}$ are just the Pauli matrices. Notice that the K_a commute with all of the generators J_a of $O(3)$. This implies that the symmetry group $SU(2)$ corresponding to Λ degeneracy is independent of the three-dimensional rotation group.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Valentine Bargmann, Dr. Arthur Bierman, and Dr. Hendricus G. Loos for stimulating discussions and constructive criticisms. Thanks also are due to Dr. H. C. Corben, Dr. Baxter H. Armstrong, Dr. John W. Clark, and Dr. Dominic G. B. Edelen for helpful comments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Asymptotic Reduced-Width Amplitude Distributions

NAZAKAT ULLAH
Tata Institute of Fundamental Research, Bombay, India
(Received 18 September 1964; final manuscript received 1 February 1965)

The method of moments is used to derive the reduced-width amplitude distributions. The explicit dependence of the distribution function on the dimension N of the random orthogonal matrix for large values of N is obtained. It is shown that in the limit $N \to \infty$, the distribution is the same as the one obtained using the explicit assumption of level independence.

I. INTRODUCTION

THE multivariate reduced-width amplitude distribution was derived by Krieger and Porter[1] using the assumptions of level independence and of functional form invariance of the distribution. It was pointed out by Dyson[2] that the assumption of level independence is quite unphysical.

An important consequence of the invariance hypothesis is that the eigenvector components of a random Hamiltonian matrix are distributed independently of its eigenvalues.[3] This fact was used to calculate the correlations of the Hamiltonian matrix elements.[4,5] Using a suggestion due to Rosenzweig,[6] a general technique has been developed to obtain the averages of the components of random orthogonal vectors.[5] A knowledge of these averages enables one to calculate the various moments of the reduced-width amplitudes. These moments are then used to determine the distribution of the reduced-width amplitude. This approach has the advantage that it shows the explicit dependence of the distribution function on the dimension N of the random orthogonal matrix for large values of N. It will be shown that the multivariate reduced-width amplitude distribution in the limit $N \to \infty$ is the same

[1] T. J. Krieger and C. E. Porter, J. Math. Phys. 4, 1272 (1963).
[2] F. J. Dyson, J. Math. Phys. 3, 140 (1962).
[3] C. E. Porter and N. Rosenzweig, Ann. Acad. Sci. Fennicae A6, 44 (1960).
[4] N. Ullah and C. E. Porter, Phys. Letters 6, 301 (1963).
[5] N. Ullah, Nucl. Phys. 58, 65 (1964).
[6] N. Rosenzweig (private communication); Phys. Letters 6, 123 (1963).

and $\alpha_{\frac{1}{2}}$ and $\alpha_{-\frac{1}{2}}$ are arbitrary constants. Equation (24) gives the general solution of Eq. (2) for $m = j$. It is clear that $Y^{\frac{1}{2}}_{jj}$, and $Y^{-\frac{1}{2}}_{jj}$ are linearly independent for both integer and half-integer j, because they are even and odd functions of cos θ, respectively. The point is that a spherical harmonic is not uniquely determined when j and m are specified. A twofold degeneracy remains because the invariant operator $J_1^2 + J_2^2 + J_3^2$ involves second derivatives with respect to θ. We may remove this degeneracy by classifying our functions by eigenvalues Λ of the operator K_3 defined by

$$K_3 Y^{\Lambda}_{jm} = \Lambda Y^{\Lambda}_{jm}. \tag{25}$$

We have, as yet, no operator connecting $Y^{\frac{1}{2}}_{jm}$ and $Y^{-\frac{1}{2}}_{jm}$. However, we may define such operators $K^{\pm} = K_1 \pm iK_2$ by the relations

$$K^{+} Y^{-\frac{1}{2}}_{jm} = Y^{\frac{1}{2}}_{jm},$$

$$K^{-} Y^{\frac{1}{2}}_{jm} = Y^{-\frac{1}{2}}_{jm}, \tag{26}$$

$$K^{+} Y^{\frac{1}{2}}_{jm} = K^{-} Y^{-\frac{1}{2}}_{jm} = 0.$$

The matrices which represent K_a on the basis $Y^{\frac{1}{2}}_{jm}$, $Y^{-\frac{1}{2}}_{jm}$ are just the Pauli matrices. Notice that the K_a commute with all of the generators J_a of $O(3)$. This implies that the symmetry group $SU(2)$ corresponding to Λ degeneracy is independent of the three-dimensional rotation group.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Valentine Bargmann, Dr. Arthur Bierman, and Dr. Hendricus G. Loos for stimulating discussions and constructive criticisms. Thanks also are due to Dr. H. C. Corben, Dr. Baxter H. Armstrong, Dr. John W. Clark, and Dr. Dominic G. B. Edelen for helpful comments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Asymptotic Reduced-Width Amplitude Distributions

NAZAKAT ULLAH
Tata Institute of Fundamental Research, Bombay, India
(Received 18 September 1964; final manuscript received 1 February 1965)

The method of moments is used to derive the reduced-width amplitude distributions. The explicit dependence of the distribution function on the dimension N of the random orthogonal matrix for large values of N is obtained. It is shown that in the limit $N \to \infty$, the distribution is the same as the one obtained using the explicit assumption of level independence.

I. INTRODUCTION

THE multivariate reduced-width amplitude distribution was derived by Krieger and Porter[1] using the assumptions of level independence and of functional form invariance of the distribution. It was pointed out by Dyson[2] that the assumption of level independence is quite unphysical.

An important consequence of the invariance hypothesis is that the eigenvector components of a random Hamiltonian matrix are distributed independently of its eigenvalues.[3] This fact was used to calculate the correlations of the Hamiltonian matrix elements.[4,5] Using a suggestion due to Rosenzweig,[6] a general technique has been developed to obtain the averages of the components of random orthogonal vectors.[5] A knowledge of these averages enables one to calculate the various moments of the reduced-width amplitudes. These moments are then used to determine the distribution of the reduced-width amplitude. This approach has the advantage that it shows the explicit dependence of the distribution function on the dimension N of the random orthogonal matrix for large values of N. It will be shown that the multivariate reduced-width amplitude distribution in the limit $N \to \infty$ is the same

[1] T. J. Krieger and C. E. Porter, J. Math. Phys. 4, 1272 (1963).
[2] F. J. Dyson, J. Math. Phys. 3, 140 (1962).
[3] C. E. Porter and N. Rosenzweig, Ann. Acad. Sci. Fennicae A6, 44 (1960).
[4] N. Ullah and C. E. Porter, Phys. Letters 6, 301 (1963).
[5] N. Ullah, Nucl. Phys. 58, 65 (1964).
[6] N. Rosenzweig (private communication); Phys. Letters 6, 123 (1963).

as the one obtained using the level-independence hypothesis.

The reduced-width amplitude $\gamma_{\lambda c}$ for level λ and channel c is given by

$$\gamma_{\lambda c} = \left(\frac{\hbar^2}{2m_c a_c}\right)^{\frac{1}{2}} \int X_\lambda \phi_c^* \, dS, \tag{1}$$

where m_c is the reduced mass, a_c the channel radius, ϕ_c the channel function and X_λ the wavefunction of the compound system.

Expanding the wavefunction X_λ in terms of a convenient set of basic functions ψ_μ we can write Eq. (1) as

$$\gamma_{\lambda c} = \sum_\mu a_{\mu\lambda} J_{\mu c}, \tag{2}$$

where

$$J_{\mu c} = \left(\frac{\hbar^2}{2m_c a_c}\right)^{\frac{1}{2}} \int \psi_\mu \phi_c^* \, dS,$$

and $a_{\mu\lambda}$ $(1 \leq \mu \leq N)$ are the components which express the wavefunction X_λ in the chosen representation. The components $a_{\mu\lambda}$ behave like the components of randomly oriented unit vectors. The $N \times N$ random orthogonal matrix is formed from N such random unit vectors.

II. SINGLE-CHANNEL DISTRIBUTION

The distribution of the reduced-width amplitude $\gamma_{\lambda c}$ was derived by Porter and Rosenzweig[3] from a knowledge of the eigenvector component distribution and the application of the central limit theorem. Later Rosenzweig[6] had indicated that the standard results of the theory in the limit $N \to \infty$ can be obtained from a knowledge of the distribution of the orthogonal matrix. In particular he had derived the single-channel Porter–Thomas distribution. However, his method does not give the explicit dependence of the distribution on N, for large values of N. Further, one does not know how to get the results for the multilevel case directly from the eigenvector component distribution. The method of moments which we follow here not only gives the explicit dependence on N but can also be used easily for the multilevel case. The moments of the reduced-width amplitude $\gamma_{\lambda c}$ are calculated using the averaging technique described in Ref. 5. The derivation of the distribution makes use of the well-known results of the problem of moments.[7] It is shown that our distribution in the limit $N \to \infty$ gives the results obtained by Porter and Rosenzweig.[3]

The ensemble average of $\gamma_{\lambda c}$ using Eq. (2) can be written as

$$\langle \gamma_{\lambda c}^{2m} \rangle = \frac{\int \delta\left(\sum_{\mu=1}^{N} a_{\mu\lambda}^2 - 1\right)\left(\sum_{\mu=1}^{N} a_{\mu\lambda} J_{\mu c}\right)^{2m} \prod_{\mu=1}^{N} da_{\mu\lambda}}{\int \delta\left(\sum_{\mu=1}^{N} a_{\mu\lambda}^2 - 1\right) \prod_{\mu=1}^{N} da_{\mu\lambda}}. \tag{3}$$

The odd moments of $\gamma_{\lambda c}$ vanish. Putting in the value of the denominator in expression (3) and after some simplification,[5] we get

$$\langle \gamma_{\lambda c}^{2m} \rangle = \frac{\Gamma(\frac{1}{2}N)}{\pi^{\frac{1}{2}N}\Gamma(\frac{1}{2}(N+2m))} \left[\left(\frac{\partial}{\partial\alpha}\right)^{2m} \times \prod_{\mu=1}^{N} \int_{-\infty}^{\infty} \exp\left[-(a_{\mu\lambda}^2 + \alpha J_{\mu c} a_{\mu\lambda})\right] da_{\mu\lambda}\right]_{\alpha=0},$$

which gives

$$\langle \gamma_{\lambda c}^{2m} \rangle = \frac{\Gamma(\frac{1}{2}N)\Gamma(2m+1)}{2^{2m}\Gamma(\frac{1}{2}(N+2m))\Gamma(m+1)} \left(\sum_\mu J_{\mu c}^2\right)^m. \tag{4}$$

It can be easily checked that the set of moments given by expression (4) satisfy Carleman's criteria,[7] and therefore should determine the distribution uniquely.

Using Sterling's formula to expand the gamma function for large values of N, we get

$$\langle \gamma_{\lambda c}^{2m} \rangle = \left[1 + \frac{1}{N}(-m^2 + m)\right] \frac{\Gamma(2m+1)}{\Gamma(m+1)} (\tfrac{1}{2}J_c)^m, \tag{5}$$

where

$$J_c \equiv \frac{1}{N} \sum_{\mu=1}^{N} J_{\mu c}^2.$$

The characteristic function of the distribution $\phi(t)$ is given by

$$\phi(t) = \sum_{m=0}^{\infty} \frac{(it)^{2m}}{\Gamma(2m+1)} \langle \gamma_{\lambda c}^{2m} \rangle.$$

Using expression (5) it can be expressed as

$$\phi(t) = [1 - J_c^2 t^4/4N] \exp(-\tfrac{1}{2}J_c t^2). \tag{6}$$

[7] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Charles Griffin & Company, Ltd., London, 1958), Vol I, Chap. IV; H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946), p. 176.

Using the inversion theorem[7] we find that the frequency function $f_\lambda(\gamma_{\lambda c})$ is given by

$$f_\lambda(\gamma_{\lambda c}) = (2\pi J_c)^{-\frac{1}{2}} \times \left[1 - \frac{1}{4N}\left(\frac{\gamma_{\lambda c}^4}{J_c^2} - \frac{6\gamma_{\lambda c}^2}{J_c} + 3\right)\right] e^{-\frac{1}{2}(\gamma_{\lambda c}^2/J_c)}. \quad (7)$$

If we now take the limit $N \to \infty$, we get the result

$$f_\lambda(\gamma_{\lambda c}) = (2\pi J_c)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}\gamma_{\lambda c}^2/J_c), \quad (8)$$

which is in agreement with the earlier result.[1,3]

We next consider two levels λ, λ' and write the bivariate moment

$$\langle \gamma_{\lambda c}^{2m} \gamma_{\lambda' c}^{2n} \rangle = \frac{\Gamma(N-1)}{2^{N-2}\pi^{N-1}} \times \int (\sum_\mu a_{\mu\lambda} J_{\mu c})^{2m} (\sum_\mu a_{\mu\lambda'} J_{\mu c})^{2n} \times \delta(\sum_\mu a_{\mu\lambda}^2 - 1)\delta(\sum_\mu a_{\mu\lambda'}^2 - 1) \times \delta(\sum_\mu a_{\mu\lambda} a_{\mu\lambda'}) \prod_\mu da_{\mu\lambda}\, da_{\mu\lambda'}, \quad (9)$$

where $2^{N-2}\pi^{N-1}/\Gamma(N-1)$ is the value of the integral[5]

$$\int \delta(\sum_\mu a_{\mu\lambda}^2 - 1)\delta(\sum_\mu a_{\mu\lambda'}^2 - 1) \times \delta(\sum_\mu a_{\mu\lambda} a_{\mu\lambda'}) \prod_\mu da_{\mu\lambda}\, da_{\mu\lambda'}.$$

The derivation of the bivariate characteristic function is a little more involved than the univariate characteristic function and therefore we shall only indicate the essential steps of the derivation.

After some mathematical manipulation and using a simple substitution

$$a_{\mu\lambda} = (1/\sqrt{2})(p_\mu - q_\mu), \qquad a_{\mu\lambda'} = (1/\sqrt{2})(p_\mu + q_\mu),$$

we can write expression (9) as

$$\langle \gamma_{\lambda c}^{2m} \gamma_{\lambda' c}^{2n} \rangle = \frac{\Gamma(N-1)}{2^{N-3}\pi^{N-1}\Gamma(\frac{1}{2}(N+2m-1))\Gamma(\frac{1}{2}(N+2n-1))} \left[\left(\frac{\partial}{\partial\alpha}\right)^{2m}\left(\frac{\partial}{\partial\beta}\right)^{2n} \times \int \exp\left\{-\sum_\mu \left[p_\mu^2 + q_\mu^2 + \frac{1}{\sqrt{2}}(\alpha+\beta)J_{\mu c}p_\mu - \frac{1}{\sqrt{2}}(\alpha-\beta)J_{\mu c}q_\mu\right]\right\}\delta[\sum_\mu (p_\mu^2 - q_\mu^2)] \prod_\mu dp_\mu\, dq_\mu\right]_{\alpha=\beta=0}.$$

Using the Fourier transform of the δ function and carrying out the p and q integrations, we get

$$\langle \gamma_{\lambda c}^{2m} \gamma_{\lambda' c}^{2n} \rangle = \frac{\Gamma(N-1)}{2^{N-2}\Gamma(\frac{1}{2}(N+2m-1))\Gamma(\frac{1}{2}(N+2n-1))} \left\{\left(\frac{\partial}{\partial\alpha}\right)^{2m}\left(\frac{\partial}{\partial\beta}\right)^{2n} \times \int dk\,(1+k^2)^{-N/2} \exp\left[\tfrac{1}{4}(1+k^2)^{-1}(\alpha^2+\beta^2)(\sum_\mu J_{\mu c}^2) + \frac{ik}{2}(1+k^2)^{-1}\alpha\beta(\sum_\mu J_{\mu c}^2)\right]\right\}_{\alpha=\beta=0}.$$

Carrying through the indicated differentation and the integration over k, we get

$$\langle \gamma_{\lambda c}^{2m} \gamma_{\lambda' c}^{2n} \rangle = \frac{\pi^{\frac{1}{2}}\Gamma(N-1)(2m)!(2n)!}{2^{N+2m+2n-2}\Gamma(\frac{1}{2}(N+2m-1))\Gamma(\frac{1}{2}(N+2n-1))\Gamma(\frac{1}{2}(N+2m+2n))} \times (\sum_\mu J_{\mu c}^2)^{m+n} \sum_j (-1)^j \frac{\Gamma(\frac{1}{2}(N+2m+2n-2j-1))}{(m-j)!(n-j)!j!}, \quad (10)$$

where the summation over j goes from 0 to m or n, whichever is smaller. If n is put equal to zero, expression (10) reduces to expression (4), as it should.

Using Stirling's formula to expand the gamma function for large values of N, expression (10) can be written as

$$\langle \gamma_{\lambda c}^{2m} \gamma_{\lambda' c}^{2n} \rangle = (2m)!(2n)!(\tfrac{1}{2}J_c)^{m+n} \times \left[\frac{1}{m!n!} - \frac{1}{N}\left\{\frac{1}{(m-2)!n!} + \frac{1}{m!(n-2)!} + \frac{2}{(m-1)!(n-1)!}\right\}\right]. \quad (11)$$

The bivariate characteristic function $\phi(t, t')$ is then given by

$$\phi(t, t') = [1 - (J_c^2/4N)(t^2 + t'^2)^2] \times \exp[-\tfrac{1}{2}J_c(t^2 + t'^2)]. \quad (12)$$

From expressions (12) and (6) we see that $\phi(t, t')$ cannot be expressed as the product of $\phi(t)$ and $\phi(t')$, except in the limit $N \to \infty$. Therefore in the limit $N \to \infty$, the joint distribution of $\gamma_{\lambda c}$, $\gamma_{\lambda' c}$ can be written as a product of two normal distributions. Thus the joint distribution of the reduced-width amplitudes for the single channel c, in the limit

$N \to \infty$, can be written as

$$P_c(\gamma_{1c}, \gamma_{2c}, \cdots) = (2\pi J_c)^{-\frac{1}{2}l} \times \exp\left[-\frac{1}{2J_c}\left(\sum_{\lambda=1}^{l}\gamma_{\lambda c}^2\right)\right], \quad (13)$$

where l is the number of levels. This is in agreement with the result derived by Krieger and Porter.[1]

III. MULTICHANNEL DISTRIBUTION

The multichannel distribution can be obtained in the same way as single-channel distribution. Let us consider a level λ and m channels $c, c', c'', \cdots$, then the multivariate moment $\langle \gamma_{\lambda c}^m \gamma_{\lambda c'}^n \cdots \rangle$ can be worked out in the same fashion as $\langle \gamma_{\lambda c}^{2m} \rangle$ given by expression (3). In the limit $N \to \infty$, we find that the multivariate characteristic function $\phi(\mathbf{t})$, where $\mathbf{t}$ is m-dimensional vector, can be written as

$$\phi(\mathbf{t}) = \exp\left[-\tfrac{1}{2}(\tilde{\mathbf{t}}\Sigma\mathbf{t})\right], \quad (14)$$

where Σ is $m \times m$ matrix, called the covariance matrix and is defined as[8]

$$\Sigma = \langle \gamma_\lambda \tilde{\gamma}_\lambda \rangle. \quad (15)$$

The frequency function $f_\lambda(\gamma_\lambda)$ is then the m-variate normal distribution[8]

$$f_\lambda(\gamma_\lambda) = (2\pi)^{-m/2} |\Sigma|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\tilde{\gamma}_\lambda \Sigma^{-1} \gamma_\lambda)\right], \quad (16)$$

where $|\Sigma|$ is the determinant of the covariance matrix Σ.

We have shown in Sec. II that in the limit $N \to \infty$, the joint characteristic function of $\gamma_{\lambda c}$, $\gamma_{\lambda' c}$ can be written as the product of two characteristic functions, one belonging to $\gamma_{\lambda c}$ and the other to $\gamma_{\lambda' c}$. This together with (14) and (16) implies that the joint distribution $P_{cc'c''\ldots}$ can be written as

$$P_{cc'c''\ldots} = \prod_\lambda f_\lambda(\gamma_\lambda). \quad (17)$$

IV. CONCLUDING REMARKS

In Secs. II and III we have used the method of moments to derive the reduced-width amplitude distribution for a system invariant under an orthogonal transformation. In a similar way we can derive the results for the unitary and symplectic ensembles defined by Dyson.[2] In the limit $N \to \infty$, these distributions agree with the ones derived using the explicit assumption of level independence.[9]

The various correlation coefficients can be exactly worked out using the expressions (4) and (10). These values agree with the ones obtained earlier.[10] In the limit $N \to \infty$, they give the level–level correlation to be zero and the width channel–channel correlation to be square of the reduced-width-amplitude channel–channel correlation, which is in agreement with the earlier results.[1]

[8] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (John Wiley & Sons, Inc., New York, 1958), Chap. II.

[9] N. Ullah, J. Math. Phys. **4**, 1279 (1963); Phys. Letters **7**, 153 (1963).

[10] N. Ullah, Nucl. Phys., **64**, 349 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Sufficient Conditions for an Attractive Potential to Possess Bound States. II

F. CALOGERO
Istituto di Fisica dell'Università, Roma; Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Rome, Italy
(Received 13 November 1964)

A condition sufficient to secure the existence of a least one bound state for each angular momentum $l \leq L$ is given by the inequality

$$-\int_0^\infty d(qr)V(r)/[(qr)^{2L}q^2 - (qr)^{-2L}V(r)] > 1,$$

where q is an arbitrary constant and $V(r)$ an everywhere-attractive potential.

1. INTRODUCTION

RECENTLY we have given certain simple conditions on an attractive central potential which are sufficient to guarantee the existence of at least one bound state for each angular momentum smaller than a given one.[1] In this paper we derive another condition to the same effect. This condition is best possible, i.e., for each l there exists a potential

[1] F. Calogero, J. Math. Phys. **6**, 161 (1965), hereafter referred to as I.

$N \to \infty$, can be written as

$$P_c(\gamma_{1c}, \gamma_{2c}, \cdots) = (2\pi J_c)^{-\frac{1}{2}l} \times \exp\left[-\frac{1}{2J_c}\left(\sum_{\lambda=1}^{l}\gamma_{\lambda c}^2\right)\right], \quad (13)$$

where l is the number of levels. This is in agreement with the result derived by Krieger and Porter.[1]

III. MULTICHANNEL DISTRIBUTION

The multichannel distribution can be obtained in the same way as single-channel distribution. Let us consider a level λ and m channels $c, c', c'', \cdots$, then the multivariate moment $\langle \gamma_{\lambda c}^m \gamma_{\lambda c'}^n \cdots \rangle$ can be worked out in the same fashion as $\langle \gamma_{\lambda c}^{2m} \rangle$ given by expression (3). In the limit $N \to \infty$, we find that the multivariate characteristic function $\phi(\mathbf{t})$, where $\mathbf{t}$ is m-dimensional vector, can be written as

$$\phi(\mathbf{t}) = \exp\left[-\tfrac{1}{2}(\tilde{\mathbf{t}}\Sigma\mathbf{t})\right], \quad (14)$$

where Σ is $m \times m$ matrix, called the covariance matrix and is defined as[8]

$$\Sigma = \langle \gamma_\lambda \tilde{\gamma}_\lambda \rangle. \quad (15)$$

The frequency function $f_\lambda(\gamma_\lambda)$ is then the m-variate normal distribution[8]

$$f_\lambda(\gamma_\lambda) = (2\pi)^{-m/2} |\Sigma|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\tilde{\gamma}_\lambda \Sigma^{-1} \gamma_\lambda)\right], \quad (16)$$

where $|\Sigma|$ is the determinant of the covariance matrix Σ.

We have shown in Sec. II that in the limit $N \to \infty$, the joint characteristic function of $\gamma_{\lambda c}$, $\gamma_{\lambda' c}$ can be written as the product of two characteristic functions, one belonging to $\gamma_{\lambda c}$ and the other to $\gamma_{\lambda' c}$. This together with (14) and (16) implies that the joint distribution $P_{cc'c''\ldots}$ can be written as

$$P_{cc'c''\ldots} = \prod_\lambda f_\lambda(\gamma_\lambda). \quad (17)$$

IV. CONCLUDING REMARKS

In Secs. II and III we have used the method of moments to derive the reduced-width amplitude distribution for a system invariant under an orthogonal transformation. In a similar way we can derive the results for the unitary and symplectic ensembles defined by Dyson.[2] In the limit $N \to \infty$, these distributions agree with the ones derived using the explicit assumption of level independence.[9]

The various correlation coefficients can be exactly worked out using the expressions (4) and (10). These values agree with the ones obtained earlier.[10] In the limit $N \to \infty$, they give the level–level correlation to be zero and the width channel–channel correlation to be square of the reduced-width-amplitude channel–channel correlation, which is in agreement with the earlier results.[1]

[8] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (John Wiley & Sons, Inc., New York, 1958), Chap. II.

[9] N. Ullah, J. Math. Phys. **4**, 1279 (1963); Phys. Letters **7**, 153 (1963).

[10] N. Ullah, Nucl. Phys., **64**, 349 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Sufficient Conditions for an Attractive Potential to Possess Bound States. II

F. CALOGERO
Istituto di Fisica dell'Università, Roma; Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Rome, Italy
(Received 13 November 1964)

A condition sufficient to secure the existence of a least one bound state for each angular momentum $l \leq L$ is given by the inequality

$$-\int_0^\infty d(qr)V(r)/[(qr)^{2L}q^2 - (qr)^{-2L}V(r)] > 1,$$

where q is an arbitrary constant and $V(r)$ an everywhere-attractive potential.

1. INTRODUCTION

RECENTLY we have given certain simple conditions on an attractive central potential which are sufficient to guarantee the existence of at least one bound state for each angular momentum smaller than a given one.[1] In this paper we derive another condition to the same effect. This condition is best possible, i.e., for each l there exists a potential

[1] F. Calogero, J. Math. Phys. **6**, 161 (1965), hereafter referred to as I.

(in fact, a one-parameter class of potentials) which saturates it. These potentials are given explicitly; or S-waves we find a Hulthén shape. Units are defined so that $\hbar = 2m = 1$, m being the mass of the particle considered.

2. ON THE POLES OF A RICCATI EQUATION

Consider the Riccati equation

$$y'(x) = f(x)[g(x) + y(x)]^2, \tag{2.1}$$

with boundary condition

$$y(0) = 0, \tag{2.2}$$

and with the following limitations on the real functions $f(x)$, $g(x)$:

$$f(x) \geq 0, \tag{2.3a}$$

$$g(0) = 0, \qquad g(x) \geq 0, \tag{2.3b}$$

$$g'(x) \geq 0. \tag{2.3c}$$

We now prove the following

Theorem. A sufficient condition for the function $y(x)$ to have at least one pole within the interval of the real axis between 0 and $X > 0$ is that there exists a value of the positive constant Q such that

$$\int_0^X dx\, g'(x)f(x)/[Q^{-1}g'(x) + Qf(x)] > 1. \tag{2.4}$$

The proof is very simple. Introduce the function $Y(x)$ through

$$y(x) + g(x) = QY(x)/[1 - Y(x)]. \tag{2.5}$$

Note that this implies that $Y(x)$ varies between zero and one while $y(x)$ varies between zero, for $x = 0$, and infinity. We now show that the hypothesis, Eq. (2.4), implies that $Y(X)$ is larger than one, thereby proving the theorem. In fact the differential equation satisfied by $Y(x)$ is

$$Y'(x) = Q^{-1}g'(x)[1 - Y(x)]^2 + Qf(x)Y^2(x), \tag{2.6}$$

with boundary condition

$$Y(0) = 0, \tag{2.7}$$

as implied by Eqs. (2.1), (2.5), and (2.3b). But Eq. (2.6) implies

$$Y'(x) \geq g'(x)f(x)/[Q^{-1}g'(x) + Qf(x)]. \tag{2.8}$$

This equation is obtained simply taking the minimum value of the right-hand side of Eq. (2.6) [at this stage use is made of the conditions Eqs. (2.3a) and (2.3c)]. It follows, through Eqs. (2.7) and (2.4), that

$$Y(X) > 1. \tag{2.9}$$

This completes the proof.

3. APPLICATION TO THE BOUND-STATE PROBLEM

The condition for the occurrence of at least one bound state with angular momentum l corresponds to the requirement that the solution of the Riccati equation

$$a_l'(r) = -(2l + 1)^{-1}[V(r)/q] \times (qr)^{-2l}[(qr)^{2l+1} + a_l(r)]^2, \tag{3.1}$$

with boundary condition

$$a_l(0) = 0, \tag{3.2}$$

has at least one pole on the positive real axis.[1,2] Here q is an arbitrary constant and $a_l(r)$ is connected with the phase function $\delta_l(r)$ by the relation[2]

$$a_l(r) = (2l + 1)!!(2l - 1)!! \lim_{k\to 0} [tg\,\delta_l(r)(q/k)^{2l+1}], \tag{3.3a}$$

where k is the linear momentum. The function $a_l(r)$ is also simply related to the logarithmic derivative of the zero-energy radial wavefunction $u_l(r)$ through

$$a_l(r) = (qr)^{2l+1}(l + \lambda + 1)/(l - \lambda) \tag{3.3b}$$

with

$$\lambda(r) = -ru_l'(r)/u_l(r). \tag{3.3c}$$

Assuming that the potential is attractive

$$V(r) = -|V(r)|, \tag{3.4}$$

we may immediately apply the theorem of the preceding section. Setting $Q = (2L + 1)$ we obtain, as a sufficient condition for the existence of at least one bound state with angular momentum L (and therefore also at least one for all $l \leq L$), the inequality

$$\int_0^\infty \frac{d(qr)\,|V(r)|}{(qr)^{2L}q^2 + (qr)^{-2L}\,|V(r)|} > 1. \tag{3.5}$$

It is easy to derive the potentials which saturate this condition, by enforcing the equality sign in Eq. (2.8). We find

$$V_l(r) = -(1 + \epsilon)q^2(qr)^{4L} \times [\exp\{(qr)^{2L+1}(2L + 1)^{-1}\} - 1]^{-1}, \quad \epsilon > 0. \tag{3.6}$$

These potentials, while possessing at least one bound state for each angular momentum $l \leq L$, have the property that they, when substituted in the integral of Eq. (3.5), yield values which may be made ar-

[2] F. Calogero, Nuovo Cimento **27**, 261 (1963); B. R. Levy and J. B. Keller, J. Math. Phys. **4**, 54 (1963); R. F. Dashen, *ibid.*, p. 388.

bitrarily close to one by the choice of a sufficiently small ϵ. Note that the constant q is arbitrary.

For S-waves the potential takes the Hulthén shape

$$V_0(r) = -(1 + \epsilon)q^2/[e^{qr} - 1], \tag{3.7}$$

while the condition Eq. (3.5) becomes simply

$$\int_0^\infty \frac{d(qr)\ |V(r)|}{q^2 + |V(r)|} > 1. \tag{3.8}$$

For a general potential this condition may be more or less stringent than those given in I. For instance, for an exponential $V(r) = -|V_0| \exp(-r/r_0)$ it yields

$$r_0^2\ |V_0| \geq c/\ln^2(1 + c), \tag{3.9}$$

where $c = |V_0|/q^2$ is an arbitrary constant. The optimal choice $c = 3.92$ yields $r_0^2\ |V_0| > 1.54$. This is a more stringent condition than that obtained in I, and is quite close to the exact minimum value for the existence of one bound state $|V_0|\ r_0^2 = 1.44$. On the other hand for a square-well potential of depth $|V_0|$ and range r_0 we obtain at best the condition $|V_0|\ r_0^2 > 4$, which is less stringent than the one obtained in I. This condition corresponds to $q^2 = |V_0|$. It is remarkable that these optimal choices of q depend on the strength but not on the range of the potential, contrary to what happened in I.

Finally we note that the theorem of Sec. 2 provides also an upper limit for the energy of the lower bound state for each angular momentum, when applied to the relevant Riccati equation, in analogy with what was done in I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Variational Principle for Diffraction of Elastic Waves

J. KORRINGA*
California Research Corporation, La Habra, California
(Received 22 October 1964)

The theory of diffraction of elastic waves is developed for the case of two homogeneous elastic media, welded together at an interface of arbitrary shape. The Green's function of the problem, defined to be the displacement at a point P_1, caused by a periodic force at a point P_0, is expressed in terms of the displacement and tension on the interface. A set of Fredholm integral equations for these interface functions is obtained. From them a variational principle is derived which gives the Green's function, minus the "free" Green's function, as the stationary value of a functional. It is analogous to the variational theorem of Levine and Schwinger for optical diffraction at an aperture in a screen. The variational equations of this functional are the above-mentioned integral equations. Explicit expressions are obtained for the case of isotropic elastic solids and for liquids. A generalization to the case of pulsed waves is indicated.

1. INTRODUCTION

WITH an eye on applications to seismic exploration, the theory of the reflection of waves from an interface between two homogeneous elastic media will be further developed. In cases of interest the problem is often one of diffraction, as the wavelengths produced in an explosion can well be comparable to the lengths characterizing irregularities in the formation. To judge from recent review articles,[1,2] diffraction theory is still primarily a branch of physical optics. Correspondingly, its oldest and best-developed part deals with diffraction from edges and apertures of screens. When diffraction from a nonplanar interface between two media is considered one finds results for a special geometry, e.g., for a sphere, obtained by a direct solution of the boundary-value problem with use of special functions. Such methods are of little help for other, less regular, interfaces. A notable exception is the work of Müller,[3] who applied to refractive interfaces the techniques familiar for the theory of diffraction from an aperture in a screen. These consist of the use of a representation theorem (Green's theorem) to express the field at a point P inside a region $\mathcal{V}$ in terms of certain functions (e.g., electric and magnetic currents in the electromagnetic case) on the surface S enclosing $\mathcal{V}$, and the use of the boundary conditions

* Permanent address: Physics Department, The Ohio State University, Columbus, Ohio.

[1] H. Hönl, A. W. Maue, and K. Westphahl in *Encyclopedia of Physics*, edited by S. Flügge (Springer-Verlag, Berlin, 1961), Vol. XXV/1, pp. 218–573.

[2] C. J. Bouwkamp, Rept. Progr. Phys. 17, 35 (1954).

[3] Cl. Müller, *Grundlagen der mathematischen Theorie electromagnetischer Schwingungen* (Springer-Verlag, Berlin, 1957); Math. Ann. 123, 345 (1951).

bitrarily close to one by the choice of a sufficiently small ϵ. Note that the constant q is arbitrary.

For S-waves the potential takes the Hulthén shape

$$V_0(r) = -(1 + \epsilon)q^2/[e^{qr} - 1], \tag{3.7}$$

while the condition Eq. (3.5) becomes simply

$$\int_0^\infty \frac{d(qr)\ |V(r)|}{q^2 + |V(r)|} > 1. \tag{3.8}$$

For a general potential this condition may be more or less stringent than those given in I. For instance, for an exponential $V(r) = -|V_0| \exp(-r/r_0)$ it yields

$$r_0^2\ |V_0| \geq c/\ln^2(1 + c), \tag{3.9}$$

where $c = |V_0|/q^2$ is an arbitrary constant. The optimal choice $c = 3.92$ yields $r_0^2\ |V_0| > 1.54$. This is a more stringent condition than that obtained in I, and is quite close to the exact minimum value for the existence of one bound state $|V_0|\ r_0^2 = 1.44$. On the other hand for a square-well potential of depth $|V_0|$ and range r_0 we obtain at best the condition $|V_0|\ r_0^2 > 4$, which is less stringent than the one obtained in I. This condition corresponds to $q^2 = |V_0|$. It is remarkable that these optimal choices of q depend on the strength but not on the range of the potential, contrary to what happened in I.

Finally we note that the theorem of Sec. 2 provides also an upper limit for the energy of the lower bound state for each angular momentum, when applied to the relevant Riccati equation, in analogy with what was done in I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Variational Principle for Diffraction of Elastic Waves

J. KORRINGA*
California Research Corporation, La Habra, California
(Received 22 October 1964)

The theory of diffraction of elastic waves is developed for the case of two homogeneous elastic media, welded together at an interface of arbitrary shape. The Green's function of the problem, defined to be the displacement at a point P_1, caused by a periodic force at a point P_0, is expressed in terms of the displacement and tension on the interface. A set of Fredholm integral equations for these interface functions is obtained. From them a variational principle is derived which gives the Green's function, minus the "free" Green's function, as the stationary value of a functional. It is analogous to the variational theorem of Levine and Schwinger for optical diffraction at an aperture in a screen. The variational equations of this functional are the above-mentioned integral equations. Explicit expressions are obtained for the case of isotropic elastic solids and for liquids. A generalization to the case of pulsed waves is indicated.

1. INTRODUCTION

WITH an eye on applications to seismic exploration, the theory of the reflection of waves from an interface between two homogeneous elastic media will be further developed. In cases of interest the problem is often one of diffraction, as the wavelengths produced in an explosion can well be comparable to the lengths characterizing irregularities in the formation. To judge from recent review articles,[1,2] diffraction theory is still primarily a branch of physical optics. Correspondingly, its oldest and best-developed part deals with diffraction from edges and apertures of screens. When diffraction from a nonplanar interface between two media is considered one finds results for a special geometry, e.g., for a sphere, obtained by a direct solution of the boundary-value problem with use of special functions. Such methods are of little help for other, less regular, interfaces. A notable exception is the work of Müller,[3] who applied to refractive interfaces the techniques familiar for the theory of diffraction from an aperture in a screen. These consist of the use of a representation theorem (Green's theorem) to express the field at a point P inside a region $\mathcal{V}$ in terms of certain functions (e.g., electric and magnetic currents in the electromagnetic case) on the surface S enclosing $\mathcal{V}$, and the use of the boundary conditions

* Permanent address: Physics Department, The Ohio State University, Columbus, Ohio.

[1] H. Hönl, A. W. Maue, and K. Westphahl in *Encyclopedia of Physics*, edited by S. Flügge (Springer-Verlag, Berlin, 1961), Vol. XXV/1, pp. 218–573.

[2] C. J. Bouwkamp, Rept. Progr. Phys. 17, 35 (1954).

[3] Cl. Müller, *Grundlagen der mathematischen Theorie electromagnetischer Schwingungen* (Springer-Verlag, Berlin, 1957); Math. Ann. 123, 345 (1951).

on S to obtain integral equations for these surface functions.[4] An advantage of this method is that it leads to convenient approximation methods such as the Kirchhoff approximation, or to more powerful ones such as the variational method of Levine and Schwinger,[5] which can be extended—as will be indicated below—to apply to the integral equations obtained by Müller.

The theory of diffraction of elastic waves, as developed along similar lines independent of the geometry, is of rather recent origin.[6,7] In analogy with optics, only "screens" of different types have been considered. A refractive boundary like that studied by Müller would be more appropriate for geophysical applications, but it has never been used in this connection. In the following this problem is treated in terms of integral equations and subsequently transformed into a variational principle analogous to that of Levine and Schwinger.

2. REPRESENTATION THEOREM

The state of an elastic solid is given by a vector field u_i which is the displacement as a function of position, x, and time, t. With it goes a stress field t_{ij} given by

$$t_{ij} = c_{ij,pq}\partial_q u_p. \tag{1}$$

The elastic coefficients, which can be x dependent, satisfy the relations

$$c_{ij,pq} = c_{ji,pq} = c_{ij,qp} = c_{pq,ij}. \tag{2}$$

∂_q signifies $\partial/\partial x_q$. The tension, τ_i, for a surface element with normal n_j is equal to

$$\tau_i = t_{ij}n_j = \Theta_{ip}u_p, \tag{3}$$

where

$$\Theta_{ip} = c_{ij,pq}n_j\partial_q. \tag{4}$$

The equations of motion are

$$\partial_j t_{ij} - \rho\ddot{u}_i = -f_i \tag{5}$$

or, using (1),

$$\mathfrak{F}_{ij}u_j - \rho\ddot{u}_i = -f_i, \tag{6}$$

where

$$\mathfrak{F}_{ij} = \partial_k c_{ik,jq}\partial_q. \tag{7}$$

f_i is the external (body) force density, and ρ is the mass density.

The solutions of Eq. (6) for any f_i can be obtained, by integration, from the Green's function of the wave equation. For the present, I consider only the Green's function for periodic time dependence, which is a solution of Eq. (6) for a unit force density of frequency ω acting at a point $P_0(x_0)$, in the k direction, i.e.,

$$f_i = \delta_{ik}\delta(x - x_0)e^{i\omega t}. \tag{8}$$

It will be symbolized by $G_{ik}(xx_0)e^{i\omega t}$, thus

$$(\mathfrak{F}_{ij} + \rho\omega^2\delta_{ij})G_{jk}(xx_0) = -\delta_{ik}\delta(x - x_0), \tag{9}$$

and is defined by the additional requirement that it satisfies the radiation condition,[8] according to which the elastic coefficients and the density become constant at large distance and

$$\lim_{|x|\to\infty} |x|\, G_{ik}(xx_0) = 0. \tag{10}$$

Note that G depends on ω, although this is not indicated explicitly.

The representation theorem follows[7] by considering, inside a region $\mathcal{V}$, two displacement fields, $u^{(1)}$ and $u^{(2)}$, and the corresponding stress fields, and by applying Gauss' theorem to the vector field $u_i^{(2)}t_{ji}^{(1)} - u_i^{(1)}t_{ji}^{(2)}$. With Eq. (2) this yields the identity

$$\int_{\mathcal{V}} dx\, (u_i^{(2)}\mathfrak{F}_{ij}u_j^{(1)} - u_i^{(1)}\mathfrak{F}_{ij}u_j^{(2)}) = \int_{\Sigma} d\hat{x}\, (u_i^{(2)}\tau_i^{(1)} - \tau_i^{(2)}u_i^{(1)}). \tag{11}$$

$\hat{x}$ denotes a point on the surface Σ bounding $\mathcal{V}$; τ is the tension on Σ, with respect to the outward-directed normal. By applying Eq. (11) to all space and choosing $u_i^{(1)} = G_{iq}(xx_0)$ and $u_i^{(2)} = G_{ir}(xx_1)$, the right-hand side vanishes on account of Eq. (10). With use of Eq. (9) one therefore finds[9]

$$G_{ij}(x_1x_0) = G_{ji}(x_0x_1). \tag{12}$$

Equation (12) is the "reciprocity theorem,"[10] which will be used frequently in the following.

The representation theorem is obtained from Eq. (11) by selecting for $\mathcal{V}$ a volume in which the elastic coefficients $c_{ij,pq}$ are constant and by taking $u_i^{(1)} = G_{iq}(xx_0)$ and $u_i^{(2)} = G_{iq}^0(xx_1)$, where G^0 is the Green's function for the case that the elastic coefficients have the same constant value in all space. The point $P_1(x_1)$ is in $\mathcal{V}$. This gives, using Eq. (12) for G^0,

$$G_{ij}(x_1x_0) - \epsilon G_{ij}^0(x_1x_0) = \int_{\Sigma} d\hat{x}'\, (G_{ki}^0(\hat{x}'x_1)T_{kj}(\hat{x}'x_0) - T_{ki}^0(\hat{x}'x_1)G_{kj}(\hat{x}'x_0)), \tag{13}$$

[4] Reference 1, p. 353.
[5] H. Levine and J. Schwinger, Phys. Rev. 74, 958 (1948).
[6] W. D. Kupradse, *Randwertaufgaben der Schwingungstheorie und Integralgleichungen* (Deutsche Verlag der Wissenschaft, Berlin, 1956), pp. 124–135.
[7] A. T. de Hoop, "Representation Theorems for the Displacement in an Elastic Solid and their Application to Elastodynamic Diffraction Theory," thesis, Delft (Excelsior, The Hague, 1958).
[8] Reference 1, p. 247.
[9] L. Knopoff and A. F. Gangi, Geophysics 24, 681 (1959).
[10] Lord Rayleigh, *The Theory of Sound* (Dover Publications, Inc., New York, 1945), 1st Amer. ed., Vol. 1, Sec. 107, p. 151.

where $\epsilon = 1$ when $P_0(x_0)$ is in $\mathcal{V}$, $\epsilon = 0$ otherwise. T is the tension on Σ associated with the displacement field G:

$$T_{kj}(\hat{x}'x_0) = [\Theta'_{kp}G_{pj}(x'x_0)]_{x'=\hat{x}'}. \tag{14}$$

The prime on Θ indicates that the differentiation is with respect to the primed variables, and that the normal n'_i at the point $\hat{x}'$ of Σ is to be used. T^0 is similarly related to G^0. In Eq. (13) G^0 and T^0 are considered to be known functions, although their evaluation for all but isotropic elastic solids may pose some difficulties. Thus, Eq. (13) reduces the problem of finding G inside $\mathcal{V}$ to that of finding the displacement-tension field G, T on Σ; but at the same time, by letting x_1 approach Σ, it yields important identities for this surface field.

Before deriving these identities, it is useful to introduce a new notation, which is aimed at giving the integrand of Eq. (13) the form of a matrix product. In the first term one has, from Eq. (12)

$$G^0_{ki}(\hat{x}'x_1) = G^0_{ik}(x_1\hat{x}'). \tag{15}$$

In the second term, I write

$$T^0_{ki}(\hat{x}'x_1) \equiv T^{0\dagger}_{ik}(x_1\hat{x}') = [\Theta'_{kq}G^0_{iq}(x_1x')]_{x'=\hat{x}'}, \tag{16}$$

which serves as definition of $T^{0\dagger}$. The symbol $\dagger$ is thus used to indicate simultaneous interchange of indices and variables. This operation will be called conjugation. According to Eq. (15), G^0 is invariant under conjugation.

This brings the integrand in Eq. (13) into the form

$$G^{0\dagger}_{ik}(x_1\hat{x}')T_{kj}(\hat{x}'x_0) - T^{0\dagger}_{ik}(x_1\hat{x}')G_{kj}(\hat{x}'x_0). \tag{17}$$

This can be further simplified by introducing a two-component "vector"

$$H_{kj}(\hat{x}'x_0) = \begin{bmatrix} G_{kj}(\hat{x}'x_0) \\ T_{kj}(\hat{x}'x_0) \end{bmatrix}, \tag{18}$$

and, as an extension of the meaning of the symbol $\dagger$, a "conjugate vector"

$$H^{0\dagger}_{ik}(x_1\hat{x}') = (G^{0\dagger}_{ik}(x_1\hat{x}'), T^{0\dagger}_{ik}(x_1\hat{x}')). \tag{19}$$

The expression (17) has the form of an alternating product

$$H^{0\dagger}_{ik}(x_1\hat{x}')\sigma H_{kj}(\hat{x}'x_0), \tag{20}$$

where

$$\sigma = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{21}$$

In (20) one can dispense with the indices i, j, k, by adopting the convention that products will be matrix products with respect to these indices as well. Thus, Eq. (13) becomes

$$G(x_1x_0) - \epsilon G^0(x_1x_0) = \int_\Sigma d\hat{x}' H^{0\dagger}(x_1\hat{x}')\sigma H(\hat{x}'x_0). \tag{22}$$

Extending the definition (19) to apply also to H, and because of $H^{0\dagger\dagger} = H^0$, one obtains by conjugation of Eq. (22):

$$G^\dagger(x_1x_0) - \epsilon G^{0\dagger}(x_1x_0) = -\int_\Sigma d\hat{x}' H^\dagger(x_1\hat{x}')\sigma H^0(\hat{x}'x_0). \tag{23}$$

According to the reciprocity relation, Eq. (12) and (15), i.e., $G^\dagger = G$, $G^{0\dagger} = G^0$, the expressions (22) and (23) have the same value.

3. IDENTITIES FOR H

In Eq. (22) one can take for x_1 a variable point x and let it approach Σ : $x \to \hat{x}$. In the limit one obtains $G(\hat{x}x_0) - \epsilon G^0(\hat{x}x_0)$ with $\epsilon = 1$ or 0 as above. One can also apply the operator Θ, with n_i being the normal to Σ at the point $\hat{x}$, and then take the limit $x \to \hat{x}$. The result is $T(\hat{x}x_0) - \epsilon T^0(\hat{x}x_0)$. These relations can be written in a compact form as follows:

$$H(\hat{x}x_0) - \epsilon H^0(\hat{x}x_0) = \lim_{x\to\hat{x}} \Xi \int_\Sigma d\hat{x}' H^{0\dagger}(x\hat{x}')\sigma H(\hat{x}'x_0), \tag{24}$$

where Ξ is a two-component operator defined by

$$\Xi_{ij} = \begin{pmatrix} \delta_{ij} \\ \Theta_{ij} \end{pmatrix}, \tag{25}$$

where Θ differentiates with respect to the unprimed variable x in Eq. (24). In order to take the limit in Eq. (24) one chooses an area S_0 of Σ containing the point $\hat{x}$, separates the integral into $\int_{\Sigma-S_0} + \int_{S_0}$, and takes the limit $x \to \hat{x}$ and $S_0 \to 0$ in that order.[11] It is easily shown[12] that one has, quite generally,

$$\lim_{S_0\to 0} \lim_{x\to\hat{x}} \Xi \int_{S_0} d\hat{x}' H^{0\dagger}(x\hat{x}')\sigma H(\hat{x}'x_0) = \tfrac{1}{2}H(\hat{x}x_0). \tag{26}$$

In the remaining integral, the limit $x \to \hat{x}$ can be taken by substitution after operating with Ξ. This gives

$$\tfrac{1}{2}H(\hat{x}x_0) - \epsilon H^0(\hat{x}x_0) = \lim_{S_0\to 0}\left[\Xi \int_{\Sigma-S_0} d\hat{x}' H^{0\dagger}(x\hat{x}')\sigma H(\hat{x}'x_0)\right]_{x=\hat{x}}. \tag{27}$$

For finite S_0, one can apply Ξ under the integral sign, but the limit $S_0 \to 0$ of the resulting integral does not exist. Nevertheless Eq. (27) will be expressed as a symbolic integration, f, which is defined by

[11] Reference 1, p. 234.

[12] Equation (26) follows by applying Eq. (22) to a small cylinder with base S_0, axis along n_i and height equaling twice the distance from P to S, and by considering the limit that the height is much smaller than the linear dimensions of S_0, while both approach zero.

$$\oint_{\Sigma} d\hat{x}'[\Xi F(x\hat{x}')]_{x=\hat{x}}$$

$$\equiv \lim_{S_0 \to 0} [\Xi \int_{\Sigma - S_0} d\hat{x}' F(x\hat{x}')]_{x=\hat{x}}. \quad (28)$$

This gives

$$\tfrac{1}{2}H(\hat{x}x_0) - \epsilon H^0(\hat{x}x_0) = \oint_{\Sigma} d\hat{x}' L(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0), \quad (29)$$

where

$$L(\hat{x}\hat{x}') = [\Xi H^{0\dagger}(x\hat{x}')]_{x=\hat{x}}, \quad (30)$$

or

$$L(\hat{x}\hat{x}') = \begin{bmatrix} G^0 & T^{0\dagger} \\ T^0 & Q^0 \end{bmatrix}, \quad (31)$$

with

$$Q^0_{ij}(\hat{x}\hat{x}') = [\Theta_{iq}\Theta'_{jp}G^0_{qp}(x\hat{x}')]_{x=\hat{x},x'=\hat{x}'}. \quad (32)$$

It is easily verified that $Q^{0\dagger} = Q^0$ and therefore

$$L^{\dagger}(\hat{x}\hat{x}') = L(\hat{x}\hat{x}'). \quad (33)$$

The conjugate of Eq. (29) can similarly be obtained from Eq. (23).

When P_0 is in $\mathcal{V}(\epsilon = 1)$, Eq. (29) can be combined with Eq. (23) to eliminate H^0. This gives

$$G^{\dagger}(x_1x_0) - G^{0\dagger}(x_1x_0) = -\int_{\Sigma} d\hat{x} \oint_{\Sigma} d\hat{x}' H^{\dagger}(x_1\hat{x})\sigma L(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0). \quad (34)$$

The other term, arising from the first term in Eq. (29), vanishes on account of the identity

$$\int_{\Sigma} d\hat{x}\, H^{\dagger}(x_1\hat{x})\sigma H(\hat{x}x_0) = 0, \quad (35)$$

which is found by taking in Eq. (11), $u^{(1)} = G(xx_0)$ and $u^{(2)} = G(xx_1)$, and using Eq. (12). From the conjugate of Eq. (29) (with x_0 replaced by x_1) and from Eq. (22) one finds similarly, for P_0 in $\mathcal{V}$:

$$G(x_1x_0) - G^0(x_1x_0) = -\int_{\Sigma} d\hat{x} \oint_{\Sigma} d\hat{x}'\, H^{\dagger}(x_1\hat{x}')\sigma L(\hat{x}'\hat{x})\sigma H(\hat{x}x_0). \quad (36)$$

The right-hand sides of Eq. (34) and (36) differ only in the manner in which the symbolic integration is performed, but one sees from the reciprocity theorem that the results are the same. On the other hand, Eq. (34) follows from Eq. (36) by conjugation. This shows that the invariance of L, expressed by Eq. (33), embodies the reciprocity, and suggests that the symbolic integration can be treated in a manner which preserves this symmetry.

4. INTEGRAL EQUATIONS FOR H

The only special assumption made in deriving the relations (29) is that the elastic coefficients are constants inside $\mathcal{V}$. Considered as integral equations for H, they have therefore a multitude of solutions. In order to arrive at a unique solution, the model must be fully specified. This I will do by stipulating that space is filled with two homogeneous solids, welded together on a surface S. S is supposed to divide space into two parts, which will be called region 1 and region 2. Region 1 contains the source, i.e., the point P_0. The elastic constants in this region are indicated simply as c, those in region 2 as c^a. Other quantities, e.g., ρ, G^0, L, are similarly distinguished for the two regions. The normal n_i on S points from region 1 into region 2. Across S, where ρ and c are discontinuous, one must impose the boundary condition that the displacement and the tension, i.e., G and T, are continuous.

The Green's function in all of region 1 can be expressed, with the above mentioned methods, in terms of a surface function H on S. To this end one takes for the surface Σ of the previous section the interface S, closed at infinity (if necessary) to surround region 1, and notes that the latter part does not contribute on account of the radiation condition Eq. (10). This gives an expression like Eq. (22), with $\epsilon = 1$:

$$G(x_1x_0) - G^0(x_1x_0) = \int_S d\hat{x}'\, H^{0\dagger}(x_1\hat{x}')\sigma H(\hat{x}'x_0). \quad (37)$$

For the present model, an expression just like this but with $\epsilon = 0$ and with H^0 replaced by H^{0a} holds for G in region 2. Approaching now with the point x_1 the surface S from the respective sides, Eq. (37) and its Θ derivative lead to an integral relation for H of the type (29), while its counterpart in region 2 gives a similar relation for the surface function on the other side of S. The continuity of G and T implies that

$$\lim_{x \to \hat{x} \text{ in } 1} H(xx_0) = \lim_{x \to \hat{x} \text{ in } 2} H(xx_0). \quad (38)$$

One therefore has

$$\tfrac{1}{2}H(\hat{x}x_0) - H^0(\hat{x}x_0) = \oint_S d\hat{x}'\, L(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0), \quad (39)$$

$$\tfrac{1}{2}H(\hat{x}x_0) = -\oint_S d\hat{x}'\, L^a(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0). \quad (40)$$

The minus sign in Eq. (40) corresponds to the fact that the normal n_i is directed inward for region 2.

Obviously, neither Eq. (39) nor (40) alone suffices to determine H. Together, however, they fully account for the wave equation in regions 1 and 2 and for the boundary conditions. The set (39) and (40) must, therefore, have a unique solution. Mathematically speaking, the ambiguity of the solution

of either (39) or (40) can be traced to the singularity, at the point $\hat{x}' = \hat{x}$, of the kernel L, in particular of the matrix element Q^0 of L which, as seen from Eq. (32), contains second derivatives of the singular function G^0. This necessitates the use of the limiting process defined by Eq. (28).

I will now show that a unique method exists to eliminate these divergencies, while preserving the general character and symmetry of the equations. To this end I cancel the strong singularities of Q^0 at any point of S by subtracting an appropriate linear combination of matrix elements of $Q^{0\alpha}$. This defines a matrix $B(x)$ such that the expression

$$Q^0_{ij}(\hat{x}\hat{x}') - B_{ik}(\hat{x})Q^{0\alpha}_{kj}(\hat{x}\hat{x}') \tag{41}$$

is free of strong singularities. I now take the corresponding linear combinations of Eqs. (39) and (40) and obtain

$$A^u(\hat{x})H(\hat{x}x_0) - H^0(\hat{x}x_0) = \int_S d\hat{x}'\, K^u(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0), \tag{42}$$

where

$$K^u(\hat{x}\hat{x}') = L(\hat{x}\hat{x}') - B(\hat{x})L^a(\hat{x}\hat{x}') \tag{43}$$

and

$$A^u_{ij}(x) = \tfrac{1}{2}[\delta_{ij} + B_{ij}(\hat{x})]. \tag{44}$$

In Eq. (42), the integral is the principal value, i.e.,

$$\int_S d\hat{x}' = \lim_{S_0 \to 0} \int_{S - S_0} d\hat{x}'. \tag{45}$$

The integral equation (42) does not have the symmetry suggested by Eq. (39), because K^u is not invariant under conjugation:

$$K^{u\dagger}(\hat{x}\hat{x}') = L(\hat{x}\hat{x}') - L^a(\hat{x}\hat{x}')\tilde{B}(\hat{x}') \neq K^u(\hat{x}\hat{x}'), \tag{46}$$

and, moreover, the matrix A^u is in general not symmetric, $\tilde{A}^u \neq A^u$ ($\sim$ indicates transposition).

Substitution of H^0 from Eq. (42) in Eq. (23) gives

$$\begin{aligned} G^\dagger(x_1x_0) - G^{0\dagger}(x_1x_0) &= -\int_S d\hat{x} \int_S d\hat{x}'\, H^\dagger(x_1\hat{x})\sigma K^u(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0) \\ &\quad + \int_S d\hat{x}\, H^\dagger(x_1\hat{x})A^u(\hat{x})\sigma H(\hat{x}x_0). \end{aligned} \tag{47}$$

One can similarly derive the conjugate equation

$$\begin{aligned} G(x_1x_0) - G^0(x_1x_0) &= -\int_S d\hat{x} \int_S d\hat{x}'\, H^\dagger(x_1\hat{x})\sigma K^{u\dagger}(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0) \\ &\quad + \int_S d\hat{x}\, H^\dagger(x_1\hat{x})\tilde{A}^u(\hat{x})\sigma H(\hat{x}x_0). \end{aligned} \tag{48}$$

On account of the reciprocity relations, Eqs. (47) and (48) must be the same. Consequently, one has

$$\begin{aligned} \int_S d\hat{x} \int_S d\hat{x}'\, H^\dagger(x_1\hat{x})\sigma(K^u(\hat{x}\hat{x}') - K^{u\dagger}(\hat{x}\hat{x}'))\sigma H(\hat{x}'x_0) \\ = \int_S d\hat{x}\, H^\dagger(x_1\hat{x})(A^u(\hat{x}) - \tilde{A}^u(\hat{x}))\sigma H(\hat{x}x_0). \end{aligned} \tag{49}$$

It was argued earlier that reciprocity, i.e., the validity of Eq. (49), should be a consequence of the integral equations for H. This would be the case if Eq. (49) could be replaced by the stronger relation:

$$\begin{aligned} \int_S d\hat{x}'\, (K^u(\hat{x}\hat{x}') - K^{u\dagger}(\hat{x}\hat{x}'))\sigma H(\hat{x}'x_0) \\ = (A^u(\hat{x}) - \tilde{A}^u(\hat{x}))H(\hat{x}x_0), \end{aligned} \tag{50}$$

from which Eq. (49) follows by multiplication with $H^\dagger(x_1x)\sigma$ (with x_1 arbitrary) and integration. Indeed Eq. (50) is the necessary and sufficient condition that Eq. (42) can be symmetrized, giving

$$A(\hat{x})H(\hat{x}x_0) - H^0(\hat{x}x_0) = \int_S d\hat{x}'\, K(\hat{x}\hat{x}')\sigma H(\hat{x}'x_0), \tag{51}$$

where

$$A_{ij}(\hat{x}) = \tfrac{1}{4}[2\delta_{ij} + B_{ij}(\hat{x}) + B_{ji}(\hat{x})] \tag{52}$$

and

$$K(\hat{x}\hat{x}') = L(\hat{x}\hat{x}') - \tfrac{1}{2}[B(\hat{x})L^a(\hat{x}\hat{x}') + L^a(\hat{x}\hat{x}')\tilde{B}(\hat{x}')]. \tag{53}$$

Expressions for G and $G^\dagger$ similar to (47) and (48), but derived from Eq. (51) are, obviously, identical. I will assume in the following, on physical rather than on mathematical grounds, that Eq. (50) is valid.

The derivation of the integral equation (51) was the purpose of this section. It will not be attempted here to prove that it has a unique solution. Its structure is, however, analogous to the integral equations for the electromagnetic case,[3] apart from the fact that there B is proportional to the unit matrix, so that symmetrization is not necessary. A uniqueness proof for that case can be found in Ref. 3.

5. THE VARIATIONAL PRINCIPLE

The integral equation (51) and its conjugate can be obtained from the following variational principle:

$$\delta X[H(\hat{x}x_0), H^\dagger(x_1\hat{x})] = 0, \tag{55}$$

where

$$X[H, H^\dagger] = \left(\int H^{0\dagger}\sigma H\right)\left(\int H^\dagger\sigma H^0\right) \times \left[\iint H^\dagger\sigma K\sigma H - \int H^\dagger A\sigma H\right]^{-1}. \tag{56}$$

$H(\hat{x}x_0)$ and $H^\dagger(x_1\hat{x})$ are two functions of $\hat{x}$ which are varied independently; arguments are to be inserted in (56) as usual. This result is obtained as follows. The coefficients of $\delta H^\dagger\sigma$ in Eq. (55) gives

$$AH - EH^0 = \int K\sigma H, \tag{57}$$

$$E = \left[\int\int H^\dagger \sigma K \sigma H - \int H^\dagger A \sigma H\right] \Big/ \int H^\dagger \sigma H^0. \tag{58}$$

E is a constant, i.e., depends only on the parameters x_0 and x_1. The solution of Eq. (57) is E times the solution of Eq. (51). Equation (58) is also obtained as a consequence of Eq. (57) by multiplying with $H^\dagger\sigma$ and integrating. Therefore E is an arbitrary normalization constant. It will be clear that the existence of a variational principle of the form (56) hinges on the validity of the symmetrized integral equation (51).

The variational principle given by Eqs. (55) and (56) is a generalization of the variational principle of Levine and Schwinger,[5] which applies to diffraction of scalar waves (e.g., sound waves in a liquid) by an aperture in an infinite plane screen. The importance of this variational principle lies in the fact that the stationary value of X is precisely the quantity in which one is normally interested. This follows from the fact that X is homogeneous of degree zero in H and $H^\dagger$, making it independent of normalization, and from the fact that Eq. (57) with $E = 1$ leads to Eq. (37) which gives, using Eq. (58),

$$X_{\text{stat}}[H, H^\dagger] = G(x_1x_0) - G^0(x_1x_0). \tag{59}$$

This makes Eq. (56) ideally suited for approximate calculations of $G(x_1x_0) - G^0(x_1x_0)$, because it is the quantity least sensitive to errors in H and $H^\dagger$.

6. ISOTROPIC SOLIDS

The elastic coefficients of an isotropic solid are given in terms of the Lamé coefficients λ and μ by

$$c_{ij,pq} = \lambda\delta_{ij}\delta_{pq} + \mu(\delta_{ip}\delta_{jq} + \delta_{iq}\delta_{jp}). \tag{60}$$

The free Green's function G^0 is given by[7]

$$G^0_{ij}(xx') = (4\pi\rho\omega^2)^{-1}[\partial_i\partial'_i R^{-1}(e^{-ik_sR} - e^{-ik_pR}) - k_s^2\delta_{ij}R^{-1}e^{-ik_sR}], \tag{61}$$

where

$$R^2 = R_iR_i, \qquad R_i = x_i - x'_i, \tag{62}$$

$$k_p^2 = \omega^2/c_p^2 = \rho\omega^2/(\lambda + 2\mu), \tag{63}$$

$$k_s^2 = \omega^2/c_s^2 = \rho\omega^2/\mu. \tag{64}$$

The singular part of G^0 is

$$G^0_{ij} \approx (4\pi R^3)^{-1}(\nu_+\delta_{ij}R^2 + \nu_-R_iR_j), \tag{65}$$

where

$$\nu_\pm = \tfrac{1}{2}[\mu^{-1} \pm (\lambda + 2\mu)^{-1}]. \tag{66}$$

From this one finds, using Eq. (32), the following expression for the singular part of Q^0:

$$Q^0_{ij}(\hat{x}\hat{x}') \approx -\mu^2(4\pi\hat{R}^5)^{-1}[2(\nu_- - \nu_+)\delta_{ij}\hat{R}^2 + (\nu_+ - 3\nu_-)(3\hat{R}_i\hat{R}_j + 2\hat{R}^2n_in_j)]. \tag{67}$$

On a frame of reference with z axis parallel to the normal in the point $\hat{x}$, the unwanted singularities reside in the diagonal elements of Q^0, and are given by

$$Q^0_\parallel = \mu(4\pi\hat{R}^3)^{-1}[1 + \tfrac{1}{2}\lambda/(\lambda + 2\mu)], \tag{68}$$

$$Q^0_\perp = \mu(4\pi\hat{R}^3)^{-1}[1 + \lambda/(\lambda + 2\mu)]. \tag{69}$$

Consequently the matrix $B(x)$ of Eq. (41) is equal to

$$B_{ik}(\hat{x}) = \gamma_\parallel\delta_{ik} + (\gamma_\perp - \gamma_\parallel)n_in_k, \tag{70}$$

$$\gamma_\parallel = Q^0_\parallel/Q^{0a}_\parallel, \qquad \gamma_\perp = Q^0_\perp/Q^{0a}_\perp. \tag{71}$$

$\gamma_\parallel$ and $\gamma_\perp$ are constants; the $\hat{x}$-dependence of B comes from its dependence on the normal n, in the point $\hat{x}$.

The equations for an isotropic solid can be considerably simplified by using an isotropic source function, i.e., radially directed forces distributed uniformly over a small sphere with center P_0. Averaging Eq. (22) over this force distribution is equivalent to taking the divergence:

$$\tilde{G}_i(x_1x_0) = \partial^0_i G_{ij}(x_1x_0), \tag{72}$$

with similar relations for $G^0(x_1x_0)$ and $H(\hat{x}'x_0)$. From Eq. (61) one finds for the average free Green's function

$$\tilde{G}^0_i(xx') = \partial'_jG^0_{ij}(xx') = (\lambda + 2\mu)^{-1}\partial_i(e^{-ik_pR}/R). \tag{73}$$

The resulting equations are not invariant under conjugation. This invariance is restored if one also takes the divergence with respect to the first index, in the point P_1. This amounts to calculating only the spherically symmetric or p-part of the wave arriving in P_1, just as the averaging (72) generates only a p-wave. With this the free Green's function becomes a scalar quantity, i.e.,

$$\bar{\bar{G}}^0(x_1x_0) = -\rho\omega^2(\lambda + 2\mu)^{-2}e^{-ik_pR}/R \tag{74}$$

and Eq. (22) becomes

$$\bar{\bar{G}}(x_1x_0) - \bar{\bar{G}}^0(x_1x_0) = \int_\Sigma d\hat{x}'\ \bar{H}^{0\dagger}_i(x_1\hat{x}')\sigma\bar{H}_i(\hat{x}'x_0). \tag{75}$$

$\bar{H}_i$ follows from the average of Eq. (51) (with respect to the second index only). The kernel K remains therefore unchanged, but $H^0_{ij}(\hat{x}x_0)$ is replaced by $\bar{H}^0_i(\hat{x}x_0)$. In Eq. (56), finally, the denominator is unchanged except that the unknown tensor H is now a vector quantity, while the numerator is drastically simplified because $\bar{H}^0$ and $\bar{H}^{0\dagger}$ contains only p-waves. I want to emphasize that the equations thus obtained are not approximations, but give the exact relation between a p-wave emitted at P_0 and the p-part of the refracted wave arriving at P_1.

7. LIQUIDS

Although a displacement field in a liquid is most conveniently described in terms of a scalar potential,

the present formalism can also be directly applied. G^0 is obtained from Eq. (61) by taking the limit $\mu \to 0$, assuming that a small imaginary part causes the exponential in k_t to vanish. This gives

$$G^{00}_{ij}(xx') = -(\rho\omega^2)^{-1}\partial_i\partial'_j\varphi(x - x') , \tag{76}$$

where

$$\varphi(x - x') = (4\pi R)^{-1}e^{-ikR} \tag{77}$$

with

$$k^2 = \rho\omega^2/\lambda. \tag{78}$$

φ is the Green's function of the Laplace wave equation,

$$\partial_i\partial_i\varphi + k^2\varphi = -\delta(x - x'). \tag{79}$$

It is convenient in this case to use a slightly different definition of H given by

$$H_{ij} = \begin{bmatrix} \rho\omega^2 G_{ij} \\ T_{ij} \end{bmatrix}. \tag{80}$$

With (76) and (80), L of Eq. (30) takes the form

$$L^0_{ij} = \begin{bmatrix} -\partial_i\partial'_j & \partial_i n'_j \\ n_i\partial'_j & n_i n'_j \end{bmatrix}\varphi. \tag{81}$$

It follows that T has the direction of the normal, $T_{ij} = n_i n_k T_{kj}$, and that Eq. (39) gives rise to equations relating only the tension and the normal component of G. Indicating the normal component of H by the corresponding lower-case letter,

$$\eta_j(\hat{x}x_0) = n_i H_{ij}(\hat{x}x_0), \tag{82}$$

Eq. (39) becomes

$$\tfrac{1}{2}\eta_j(\hat{x}x_0) - \eta^{00}_j(\hat{x}x_0) = \oint d\hat{x}'\, L^0(\hat{x}\hat{x}')\sigma\eta_j(\hat{x}'x_0), \tag{83}$$

where L^0 is a scalar quantity given by

$$L^0 = n_i L^0_{ij} n_j = \begin{bmatrix} -\partial_n\partial'_n & \partial_n \\ \partial'_n & 1 \end{bmatrix}\varphi. \tag{84}$$

Here ∂_n is the normal derivative, $\partial_n = n_i\partial_i$. The boundary conditions have to be modified, requiring continuity of the tension and of the normal component of the displacement only. Equation (83) can then be combined with an equation similarly obtained from Eq. (40). One sees that the case of liquids is anomalous in that the strong singularities reside in the first element of L^0, i.e., in $n_i G^{00}_{ij} n_j$. The matrix B defining the linear combination is proportional to the unit matrix, and one obtains

$$\tfrac{1}{2}(1 + \gamma)\eta_j(\hat{x}x_0) - \eta^{00}_j(\hat{x}x_0) = \int d\hat{x}'\, K^0(\hat{x}\hat{x}')\sigma\eta_j(\hat{x}'x_0), \tag{85}$$

$$K^0 = L^0 - \begin{bmatrix} -\partial_n\partial'_n & \gamma\partial_n \\ \gamma\partial'_n & \gamma^2 \end{bmatrix}\varphi^a, \tag{86}$$

where

$$\varphi^a = e^{-ik^a R}/R, \tag{87}$$

$$\gamma = \rho^a/\rho. \tag{88}$$

One can finally take the average for an isotropic source, replacing η_j by $\bar{\eta}$, in the same manner as for an isotropic solid. One finds

$$\bar{\eta}^{00}(\hat{x}x_0) = k^2\begin{bmatrix} \partial_n \\ -1 \end{bmatrix}\varphi, \tag{89}$$

and, averaging the equation corresponding to Eq. (22) (with $\epsilon = 1$) and remembering the definition (80) one has

$$\rho\omega^2(\bar{G}(x_1x_0) - \bar{G}^{00}(x_1x_0)) = \int_S d\hat{x}'\, \bar{\eta}^{00\dagger}(x_1\hat{x}')\sigma\bar{\eta}(\hat{x}'x_0), \tag{90}$$

with

$$\rho\omega^2\bar{G}^{00} = -k^4\varphi. \tag{91}$$

Equations (86) and (89) are to be used in the variational equation.

8. PULSED WAVES

The above presentation of diffraction was based on harmonic analysis of the source function f_i. It is sometimes convenient to use a Laplace analysis[7]

$$f(t) = \int_0^\infty ds\, f(s)e^{-st}, \tag{92}$$

or to use a presentation in terms of pulsed waves. The latter case will be discussed in this section.

A time-dependent Green's function is here defined as a solution of Eq. (6) with the source function[13]

$$f_i(xt) = \delta_{ij}\delta(x - x_0)\delta(t - t_0). \tag{93}$$

This gives $G_{ij}(xtx_0t_0)$. A time-dependent free Green's function, $G^0{}_{ij}(xtx_0t_0)$, is similarly defined with respect to a wave equation with constant coefficients.

The quantity of physical interest is the causal Green's function, defined by the additional requirement

$$G(xtx_0t_0) = 0 \qquad (t \leq t_0). \tag{94}$$

An anticausal Green's function will be distinguished by a *, and satisfies

$$G^*(xtx_0t_0) = 0 \qquad (t \geq t_0). \tag{95}$$

From the time independence of the coefficients of

[13] M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Ltd., London, 1959), p. 377.

the wave equation, it follows that G is a function of $t - t_0$. Therefore

$$G^*(xtx_0t_0) = G(x, -t, x_0, -t_0). \tag{96}$$

The reciprocity theorem for causal Green's functions is obtained from Eq. (11) by choosing

$$\begin{aligned} u_i^{(1)} &= G_{iq}(xtx_0t_0), \\ u_i^{(2)} &= G_{ir}^*(xtx_1t_1), \end{aligned} \tag{97}$$

with $t_1 > t_0$ and integrating from $t_0 - \kappa$ to $t_1 + \kappa$, $\kappa > 0$. Upon substitution from Eq. (6) the term involving $\ddot{u}$ vanishes on account of Eqs. (94) and (95). Extending the integration over all space one obtains

$$G_{ij}(x_1t_1x_0t_0) = G_{ji}(x_0, -t_0, x_1, -t_1). \tag{98}$$

The generalization of the representation theorem, Eq. (22), to time-dependent Green's functions follows similarly by taking

$$\begin{aligned} u_i^{(1)} &= G_{iq}(xtx_0t_0), \\ u_i^{(2)} &= G_{iq}^{0*}(xtx_1t_1). \end{aligned} \tag{99}$$

This gives

$$\begin{aligned} &G(x_1t_1x_0t_0) - \epsilon G^0(x_1t_1x_0t_0) \\ &= \int_{t_0-\kappa}^{t_1+\kappa} dt' \int_\Sigma d\hat{x}' H^{0\dagger}(x_1t_1\hat{x}'t')\sigma H(\hat{x}'t'x_0t_0), \end{aligned} \tag{100}$$

where the conjugation $\dagger$ is now defined by

$$H^\dagger(x_1t_1\hat{x}t) = \tilde{H}(\hat{x}, -t, x_1, -t_1), \tag{101}$$

$\sim$ meaning transposition of the binary and ternary indices. Because of the fact that all quantities depend only on the time difference, this is the same as interchanging variables and indices and not interchanging times. The time integration in Eq. (100) can, of course, be just as well extended from $-\infty$ to $+\infty$, because the second factor is zero for $t < t_0$ and the first factor is zero for $t > t_1$.

From here on, the development of the foregoing sections can be repeated, the difference being that G^0 and its derivatives, e.g., L, become known functions of time, and that everywhere the integration over S is accompanied by integration over a time from $-\infty$ to $+\infty$. One still has $L^\dagger = L$ representing the reciprocity. The matrix $B(\hat{x})$ is time independent. In the variational theorem H and $H^\dagger$ must now be varied as functions of x and t.

For isotropic solids one finds, from Ref. 7,

$$\begin{aligned} &G_{ij}^0(xtx't') = (4\pi\rho)^{-1}\{\partial_i\partial_j' R^{-1} \\ &\times [D(t - t' - R/c_s) - D(t - t' - R/c_p)] \\ &+ c_s^{-2}\delta_{ij}R^{-1}\delta(t - t' - R/c_s)\}, \end{aligned} \tag{102}$$

where

$$\begin{aligned} D(x) &= 0 \qquad (x < 0), \\ D(x) &= x \qquad (x > 0). \end{aligned} \tag{103}$$

Because G^0 and therefore also H^0 and K contain only δ-functions of time and their derivatives, the time integration in both factors in the numerator of X, Eq. (56), and one time integration in the double integral of the denominator, can be carried out. This requires a partial integration and is similar to the optical case discussed in Ref. 13.

A generalization of the Green's function to pulsed waves of arbitrary time dependence is considered in Ref. 7. Taking the source function

$$f_i(xt) = \delta_{ij}\delta(x - x_0)g(t - t_0), \tag{104}$$

one can define a Green's function $\tilde{G}^g$ by

$$\tilde{G}_{ij}^g(xtx_0t_0) = \int dt'\, G_{ij}(xtx_0t_0')g(t_0 - t_0'). \tag{105}$$

For the free Green's function one obtains, for an isotropic solid,[7] using Eq. (102):

$$\begin{aligned} &\tilde{G}_{ij}^{0g}(xtx't') = (4\pi\rho)^{-1}\{\partial_i\partial_j' R^{-1} \\ &\times [D^g(t - t' - R/c_s) - D^g(t - t' - R/c_p)] \\ &+ c_s^{-2}\delta_{ij}R^{-1}g(t - t' - R/c_s)\}, \end{aligned} \tag{106}$$

$$D^g(x) = \int_0^\infty d\xi\, g(x - \xi)\xi. \tag{107}$$

With the one-sided averaging in the definition (105) one loses the symmetry of the equations, and thereby the variational theorem. Symmetry can be restored by also averaging the signal arriving in P_1, with the same shape function $g(t - t_1)$. This gives

$$\begin{aligned} X[\bar{H}, \bar{H}^\dagger] &= \left(\int \bar{H}^{0\dagger g}\sigma\bar{H}\right)\left(\int \bar{H}^\dagger\sigma\bar{H}^{0g}\right) \\ &\times \left[\iint \bar{H}^\dagger\sigma K\sigma\bar{H} - \int \bar{H}^\dagger A\sigma\bar{H}\right]^{-1}, \end{aligned} \tag{108}$$

where all integrations include time integration. Note that the kernel is unchanged. The solution, $\bar{H}^\dagger$ and $\bar{H}$, obtained from $\delta X = 0$ will be the time average,

$$\bar{H} = \bar{H}^g \tag{109}$$

and the stationary value of X is

$$X_{\text{Stat.}} = \bar{\bar{G}}^g - \bar{\bar{G}}^{0g}. \tag{110}$$

ACKNOWLEDGMENTS

The author is indebted to the California Research Corporation for the hospitality extended during the period when this work was performed. He also wishes to thank Dr. R. J. S. Brown and other members of the Geophysics Research Section for valuable criticism and suggestions.

Scattering by Singular Logarithmic Potential*

H. H. Aly,† Riazuddin, and A. H. Zimerman‡
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received 4 August 1964)

Potential of the form $(gr^{-4}\ln^2 r - g^{\frac{1}{2}}r^{-3})\theta(r - r_0)$ is considered in connection with the applicability of peratization technique. The advantage of this potential is the fact that while it is dominated by a logarithmic part near the origin, the exact solution of the zero-energy and s-wave Schrödinger equation is obtained in a closed form. We show that the peratization technique gives the correct answers.

1. INTRODUCTION

In a previous paper[1] we have considered a singular logarithmic potential of the form $gr^{-4}\ln^2 r$. In that paper, we have shown that the summation of the leading singular terms in each order of the perturbation series, where we introduced a cutoff parameter, gives a nonsensical answer. Recently, Wu,[2] dealing with a similar problem, pointed out that if all the singular terms are summed up, the final answer can be made meaningful when the cutoff $\Lambda \to \infty$. Unfortunately, since the zero-energy and s-wave Schrödinger equation for both the potential considered by Wu $(-gr^{-4}\ln r)$ and that considered by the present authors $(gr^{-4}\ln^2 r)$ cannot be solved exactly, it is difficult to draw any definite conclusion with regard to the applicability of the technique of peratization.

The technique of peratization has recently been discussed and applied[3,4] to scattering problems with repulsive potential of the form gr^{-n} for $g > 0$ and $n > 3$, that is a potential which has a singularity of the pole type at the origin. (More generally we have a branch-point type of singularity when $n \neq$ integer.) It has also been shown by the present authors that peratization technique also gives the correct answer for a potential which has an essential type of singularity at the origin, i.e., potential of the form $gr^{-4}(e^{2/r}) + \frac{1}{4}r^{-4}$. In this paper we shall discuss the applicability of peratization technique for a potential which has a branch-point singularity of the logarithmic type at the origin. For this purpose we consider a potential of the form

$$V(r) = \left(g\frac{\ln^2 r}{r^4} - g^{\frac{1}{2}}\frac{1}{r^3}\right)\theta(r - r_0), \tag{1}$$

where

$$\theta(r - r_0) = 1 \quad \text{for} \quad r < r_0,$$
$$= 0 \quad \text{for} \quad r > r_0.$$

Our selection of a potential of the form given in Eq. (1) is motivated by the following consideration: It is the logarithmic term which dominates at the origin, since the $g^{\frac{1}{2}}r^{-3}$ term is less singular, and this is what we want. The addition of the $g^{\frac{1}{2}}r^{-3}$ term helps in the sense that the Schrödinger equation for the potential (1) can now be solved exactly in a closed form for zero energy and zero angular momentum. On the other hand, the $g^{\frac{1}{2}}r^{-3}$ term does not fall sufficiently fast at ∞ so as to define phase shift at zero energy. To avoid this difficulty, we have introduced the θ function. This is no handicap because it is the behavior of the potential near the origin that we are interested in studying in connection with the peratization technique. Since we know the exact solution of the Schrödinger equation for the potential (1) for zero energy and s wave as discussed below, we show by comparing it with the one obtained by the peratization technique, that the peratization technique does in fact work for the potential (1) also.

2. EXACT SOLUTION

Let us consider the radial s-wave Schrödinger equation for zero energy for the potential (1):

$$\frac{d^2\psi}{dr^2} - \left(g\frac{\ln^2 r}{r^4} - \frac{g^{\frac{1}{2}}}{r^3}\right)\theta(r - r_0)\psi = 0, \tag{2}$$

with the boundary conditions

$$\psi(r) \underset{r\to 0}{\longrightarrow} 0 \quad \text{and} \quad \psi(r) \underset{r\to\infty}{\longrightarrow} r.$$

The exact solution of (2) is:

$$\psi(r) = r\exp\{g^{\frac{1}{2}}[(\ln r + 1)/r]\} \quad \text{for} \quad r < r_0$$
$$= \alpha r + \beta \quad \text{for} \quad r > r_0, \tag{3}$$

* Work supported by the U. S. Atomic Energy Commission and the U. S. Air Force.
† Present address: Max Planck Institute für Physik und Astrophysik, Munich, West Germany.
‡ Present address: The Instituto de Fisica Teorica, Sao Paulo, Brazil.
[1] H. H. Aly, Riazuddin, and A. H. Zimerman, Phys. Rev. **136**, B1174 (1964).
[2] T. T. Wu, Phys. Rev. **136**, B1176 (1964).
[3] N. N. Khuri and A. Pais, Rev. Mod. Phys. **36**, 590 (1964).
[4] G. Tiktojoulos and S. B. Treiman, Phys. Rev. **134**, B844 (1964).

which is regular at the origin. The continuity conditions for $\psi(r)$ and $d\psi/dr$ at $r = r_0$ gives

$$\alpha = \exp\left\{g^{\frac{1}{2}}\left[\frac{\ln r_0 + 1}{r_0}\right]\right\}\left[1 - g^{\frac{1}{2}}\frac{\ln r_0}{r_0}\right]$$

$$\beta = \exp\{g^{\frac{1}{2}}[(\ln r_0 + 1)/r_0]\}g^{\frac{1}{2}}\ln r_0. \quad (4)$$

From (3) and (4) it follows that the zero-energy scattering amplitude is given by

$$a = g^{\frac{1}{2}}\ln r_0(1 - g^{\frac{1}{2}}\ln r_0/r_0)^{-1}. \quad (5)$$

3. SOLUTION OBTAINED BY PERATIZATION

Let us now apply the peratization procedure to Eq. (2) in order to get a (the zero-energy scattering amplitude). For this purpose we shall follow the method which Khuri and Pais[3] used in connection with the inverse-power potential. Introducing

$$\Psi(r) = (1/r)\psi(r), \quad (6)$$

and regulating the potential by introducing a parameter α, the regulated wavefunction $\Psi(r, \alpha)$ is written as

$$\Psi(r, \alpha) = \Psi_1(r, \alpha) + \Psi_2(r, \alpha), \quad (7)$$

where Ψ_1 satisfies a singular integral equation for $\alpha \to 0$ while Ψ_2 satisfies a regular integral equation for $\alpha \to 0$. These integral equations are

$$\Psi_1(r, \alpha) = -\frac{1}{r}\int_0^\infty y^2\,dy\;V(y, \alpha) \times [\Psi_1(y, \alpha) + \Psi_2(y,\alpha)], \quad (8)$$

$$\Psi_2(r, \alpha) = 1 - \frac{1}{r}\int_r^\infty y\,dy\,(r - y)V(y, \alpha)\Psi_1(y, \alpha) - \frac{1}{r}\int_r^\infty y\,dy\,(r - y)V(y, \alpha)\Psi_2(y, \alpha). \quad (9)$$

For $\alpha \neq 0$ there exists always a solution of Eq. (8) and is given by

$$\Psi_1(r, \alpha) = a(\alpha)/r, \quad (10)$$

where

$$a(\alpha) = -\int_0^\infty y^2\,dy\;V(y, \alpha) \times [\Psi_1(y, \alpha) + \Psi_2(y, \alpha)], \quad (11)$$

which is the zero-energy scattering amplitude.

Substituting (10) in (9), we obtain a Volterra type of equation for Ψ_2, the solution of which we write in the form

$$\Psi_2(r, \alpha) = \Psi_2^{(1)}(r, \alpha) + a(\alpha)\Psi_2^{(2)}(r, \alpha).$$

Assuming that the limit of a (α) as $\alpha \to 0$ exists, Eq. (9) being of the Volterra type is nonsingular when $\alpha \to 0$. Putting $\Psi_2^{(2)}(r) = F(r) - 1/r$, $\Psi_2^{(1)}(r)$ and $F(r)$ satisfy the integral equations

$$\Psi_2^{(1)}(r) = 1 - \frac{1}{r}\int_r^\infty y\,dy\,(r - y)V(y)\Psi_2^{(1)}(y), \quad (12)$$

$$F(r) = \frac{1}{r} - \frac{1}{r}\int_r^\infty y\,dy\,(r - y)V(y)F(y). \quad (13)$$

From (12) it follows that

$$\Psi_2^{(1)} = A\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln r + 1}{r}\right]\right\} + B\exp\left\{\mathfrak{D}g^{\frac{1}{2}}\left[\frac{\ln r + 1}{r}\right]\right\}L(r), \quad \text{for } r < r_0,$$

$$\Psi_2^{(1)} = 1, \quad \text{for } r > r_0, \quad (14)$$

where

$$L(r) = \int_\xi^r \frac{1}{r'^2}\exp\left\{-2g^{\frac{1}{2}}\left[\frac{\ln r' + 1}{r'}\right]\right\}dr',$$

ξ being an arbitrary point. The continuity condition for $\Psi_2^{(1)}(r)$ and $d\Psi_2^{(1)}(r)/dr$ at $r = r_0$ gives

$$A = \exp\{-g^{\frac{1}{2}}[(\ln r_0 + 1)/r_0]\} - BL(r_0),$$

$$B = g^{\frac{1}{2}}\ln r_0\exp\{g^{\frac{1}{2}}[(\ln r_0 + 1)/r_0]\}. \quad (15)$$

From Eq. (13) it follows that:

$$F(r) = A'\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln r + 1}{r}\right]\right\} + B'\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln r + 1}{r}\right]\right\}L(r) \quad \text{for } r < r_0$$

$$= 1/r \quad \text{for } r > r_0. \quad (16)$$

The continuity condition for $F(r)$ and $dF(r)/dr$ at $r = r_0$ gives

$$A' = \frac{1}{r_0}\exp\left\{-g^{\frac{1}{2}}\left[\frac{\ln r_0 + 1}{r_0}\right]\right\} - B'L(r_0),$$

$$B' = \exp\left\{g^{\frac{1}{2}}\left[\frac{\ln r_0 + 1}{r_0}\right]\right\}\left(g^{\frac{1}{2}}\frac{\ln r_0}{r_0} - 1\right). \quad (17)$$

Using the expression

$$a = \lim_{\sigma\to 0}\frac{-\int_\sigma^\infty y^2\,dy\;V(y, \alpha)\Psi_2^{(1)}(y, \alpha)}{1 + \int_\sigma^\infty y^2\,dy\;V(y, \alpha)[\Psi_2^{(2)}(y, \alpha) + 1/y]},$$

we get

$$a = \lim_{\sigma\to 0}\frac{-\int_\sigma^\infty y^2\,dy\;V(y)\left(A\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln y + 1}{y}\right]\right\} + B\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln y + 1}{y}\right]\right\}L(y)\right)}{1 + \int_\sigma^\infty y^2\,dy\;V(y)\left(A'\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln y + 1}{y}\right]\right\} + B'\exp\left\{g^{\frac{1}{2}}\left[\frac{\ln y + 1}{y}\right]\right\}L(y)\right)}. \quad (18)$$

Taking the leading singularities both in the numerator and the denominator we get

$$a = -B/B',$$

which because of (15) and (17) reproduces the exact result (5).

In conclusion we can say that the peratization technique can be applied to a special type of singular potential of logarithmic type. As it was shown to apply also for other special cases like potentials having poles[3,4] or essential singularity[5] at the origin, we may conjecture that the peratization technique can be applied to all singular potentials independently of the nature of the singularity at the origin, provided the potential must be sufficiently singular at the origin and have some good behavior at infinity.

[5] H. H. Aly, Riazuddin, and A. H. Zimerman, Nuovo Cimento 35, 324 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Evaluation of Feynman's Functional Integrals

ROBERT L. ZIMMERMAN
Lawrence Radiation Laboratory, University of California, Livermore, California
(Received 7 December 1964)

Several relationships for Feynman's functional integrals are derived. From these relationships, we construct two different schemes for approximating Feynman's functional integral. The methods of approximation are expected to converge sufficiently rapidly in many cases so that only the lowest orders of the approximation are required to give reliable answers.

I. INTRODUCTION

IN his Princeton dissertation of 1942, Feynman introduced a new formulation of nonrelativistic quantum theory.[1] This formulation is different from that of Schrödinger or Heisenberg. However, because of the great difficulty in evaluating the functional integrals, there are only a few applications of the Feynman formalism.

It is the purpose of this paper to derive some relationships among the functional integrals. From these relations we will construct two approximation schemes. The approximation schemes are not based on an expansion about a coupling constant and, therefore, are expected to be useful in strong interactions. These approximation schemes are expected to converge very rapidly since we have kept the predominant contributions to the functional integrals. The solution becomes exact as the order of the approximation goes to infinity.

We have constructed one of the approximations to give correct answers for harmonic potentials [i.e., $V(X) = a + \lambda x + \frac{1}{2}m_\omega^2 x^2$] in all orders of the approximation scheme. In a future paper, this method will be applied to the solution of the functional integrals that arise for the matrix elements in field theoretic calculations and will be shown to yield the exact answer in all orders of the approximation for noninteracting fields. Also, since we are keeping the predominant contribution to the functional integrals, it is expected that these integrals will converge very rapidly and hence give reliable answers for the first few terms of the approximation for interacting fields.

There is an integral which is closely related to the Feynman integral; it is called the Wiener integral.[2] In fact the Feynman integral is in many cases defined in terms of the Wiener integral.[3] There have appeared in the literature various approximations of the Wiener integral and hence the Feynman integral.

The first such results were given by Cameron.[4] He has constructed two possible numerical integration rules which are based on approximating the functional of interest by a finite series of trigonometric functions and replacing the infinite-dimensional integral by a finite-dimensional one. One of Cameron's methods which he calls his rectangular rule consists of constructing a sequence $\{I_n\}$ where

[1] R. P. Feynman, Rev. Mod. Phys. 20, 367 (1948).

[2] See for example I. M. Koval'chik, Russian Math. Surveys 18, No. 1, 97 (1963) or G. E. Shilov, Russian Math. Surveys 18, No. 2, 97 (1963).

[3] R. H. Cameron, J. Math. and Phys. 39, 126 (1960).

[4] R. H. Cameron, Duke Math. J. 18, 111 (1951).

Taking the leading singularities both in the numerator and the denominator we get

$$a = -B/B',$$

which because of (15) and (17) reproduces the exact result (5).

In conclusion we can say that the peratization technique can be applied to a special type of singular potential of logarithmic type. As it was shown to apply also for other special cases like potentials having poles[3,4] or essential singularity[5] at the origin, we may conjecture that the peratization technique can be applied to all singular potentials independently of the nature of the singularity at the origin, provided the potential must be sufficiently singular at the origin and have some good behavior at infinity.

[5] H. H. Aly, Riazuddin, and A. H. Zimerman, Nuovo Cimento 35, 324 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Evaluation of Feynman's Functional Integrals

ROBERT L. ZIMMERMAN
Lawrence Radiation Laboratory, University of California, Livermore, California
(Received 7 December 1964)

Several relationships for Feynman's functional integrals are derived. From these relationships, we construct two different schemes for approximating Feynman's functional integral. The methods of approximation are expected to converge sufficiently rapidly in many cases so that only the lowest orders of the approximation are required to give reliable answers.

I. INTRODUCTION

IN his Princeton dissertation of 1942, Feynman introduced a new formulation of nonrelativistic quantum theory.[1] This formulation is different from that of Schrödinger or Heisenberg. However, because of the great difficulty in evaluating the functional integrals, there are only a few applications of the Feynman formalism.

It is the purpose of this paper to derive some relationships among the functional integrals. From these relations we will construct two approximation schemes. The approximation schemes are not based on an expansion about a coupling constant and, therefore, are expected to be useful in strong interactions. These approximation schemes are expected to converge very rapidly since we have kept the predominant contributions to the functional integrals. The solution becomes exact as the order of the approximation goes to infinity.

We have constructed one of the approximations to give correct answers for harmonic potentials [i.e., $V(X) = a + \lambda x + \frac{1}{2}m_\omega^2 x^2$] in all orders of the approximation scheme. In a future paper, this method will be applied to the solution of the functional integrals that arise for the matrix elements in field theoretic calculations and will be shown to yield the exact answer in all orders of the approximation for noninteracting fields. Also, since we are keeping the predominant contribution to the functional integrals, it is expected that these integrals will converge very rapidly and hence give reliable answers for the first few terms of the approximation for interacting fields.

There is an integral which is closely related to the Feynman integral; it is called the Wiener integral.[2] In fact the Feynman integral is in many cases defined in terms of the Wiener integral.[3] There have appeared in the literature various approximations of the Wiener integral and hence the Feynman integral.

The first such results were given by Cameron.[4] He has constructed two possible numerical integration rules which are based on approximating the functional of interest by a finite series of trigonometric functions and replacing the infinite-dimensional integral by a finite-dimensional one. One of Cameron's methods which he calls his rectangular rule consists of constructing a sequence $\{I_n\}$ where

[1] R. P. Feynman, Rev. Mod. Phys. 20, 367 (1948).

[2] See for example I. M. Koval'chik, Russian Math. Surveys 18, No. 1, 97 (1963) or G. E. Shilov, Russian Math. Surveys 18, No. 2, 97 (1963).

[3] R. H. Cameron, J. Math. and Phys. 39, 126 (1960).

[4] R. H. Cameron, Duke Math. J. 18, 111 (1951).

I_n is an n-dimensional integral. For reasonably well-behaved functions, $\lim_{n\to\infty} I_n$ will be equal to the Wiener integral. Unfortunately the limit is not approached very rapidly, which makes his rectangular rule rather impractical.

By making a slight modification in his rectangular rule he constructs his other approximation sequence $\{Jn\}$ which he calls his "Simpson rule." This sequence converges much more rapidly. In fact, for third-degree polynomial functionals, Jn is exact for all n. Therefore, the first few terms of the Simpson rule should give very reliable results if the functional can be approximated sufficiently well by a third-degree polynomial.

Vladimirov has also constructed an alternate approximation scheme which corresponds to replacing the Wiener integral by a sum.[5] The parameters in the sum are chosen in such a way to give exact equality for polynomial functionals of degree three, for odd functionals, and also for functionals of the form

$$F(x) = \int_0^1 [x(s)]^2\, ds \int_0^1 \int_0^1 X(s_1)X(s_2)d^2_{s_1,s_2}\sigma(s_1, s_2),$$

where $\sigma(S_1, S_2)$ is an arbitrary function of bounded variation. Vladimirov's formula is more accurate than Cameron's Simpson rule. However, both Cameron's and Vladimirov's methods applied to most problems of physics should be very poor because the functional of interest is of the exponential type and consequently is poorly approximated by a polynomial function of degree 3.

Gel'fand and Chentsov[6] have approximated the Wiener integral by a finite-dimensional Stieltzes integral of sufficiently high multiplicity, and evaluated it by a Monte-Carlo method.

Rosen[7] has approximated the Feynman integral by taking a subset of the paths and consequently gets the Green's function which is correct to lowest orders of the time.

The motivation of this work is based on Cameron's paper.[4] In this paper he derived a relationship for Wiener integrals called the mixed integration formula, which expressed them as multiple integrals, where one of the integrals is again a Wiener integral. In this paper we derive an analogous mixed integration formula for Feynman integrals and using the stationary-phase method[8] construct an approximation for Feynman integrals.

The arrangement of the paper is as follows: In Sec. II, a brief review of Feynman's formalism of quantization is presented. The definition of Feynman's functional integral and its relation to the Green's function of the Schrödinger equation is given. In Sec. III, we discuss an alternate method of the formulation of Feynman functional integral. Instead of partitioning the particle trajectory into polygonal paths and integrating over the endpoints as was done by Feynman,[1] we expand the particle trajectory into a complete orthonormal set and then integrate over all the coefficients of the orthonormal expansion. By varying the coefficients we obtain the set of all possible paths. With the aid of this latter formulation of Feynman integrals in terms of an orthonormal expansion we easily get Cameron's mixed integration formula.[4] The free-particle approximation is the derived in Sec. IV with the aid of Cameron's mixed integration formula and the method of stationary phases. The free-particle approximation is constructed to give the correct answer for potentials of the form $V(x) = a + bX$. Since the major contribution to the functional integrals comes from the classical path it is expected that the first few terms of the free-particle approximation to give reliable answers for other potentials. To show that this is indeed the case we will solve an example with a potential of the form $V(x) = \frac{1}{2}m\omega^2x^2$. In Sec. V, we improve the free-particle approximation so as to give better results. We call this improved procedure the "harmonic-oscillator approximation." The harmonic-oscillator approximation gives the exact results in all orders of the approximation for potentials of the form $V(x) = a + bx + cx^2$. For other potentials it is expected to give reliable results for the first few terms of the approximation. The reason for this is the same as given for the free-particle approximation. Section VI is a summary and discussion of the results.

II. THE FEYNMAN FUNCTIONAL INTEGRAL

In the Feynman approach to nonrelativistic quantum mechanics, the basic idea of a wavefunction is maintained. Instead of solving the Schrödinger equation to obtain the wavefunction of the system, a new postulate is introduced. In this formulation, the Green's function $K(x, t; x_0, t_0)$ connecting the wavefunction $\psi(x, t)$ at time t, with the wavefunction $\psi(x_0, t_0)$ at an earlier time t_0 is given by an average

[5] V. S. Vladimirov, Uspekhi Mat. Nauk **15**, No. 4 (94), 129 (1960).

[6] I. M. Gel'fand and N. N. Chentsov, Zh. Eksper. Teoret. Fiz. **31**, 1106 (1956) [English transl.: Soviet Phys.—JETP **4**, 945 (1957)].

[7] G. Rosen, J. Math. Phys. **4**, 1327 (1963).

[8] R. Abé, Busseiron Kenyo, No. 79, 101 (1954).

over all real continuous paths connecting the space–time points (x, t) and (x_0, t_0) multiplied by a normalization factor in order to insure unitarity of the wavefunction. All continuous paths are assumed to be equally probable in magnitude, but the contributions to $K(x, t; x_0t_0)$ differ in phase by an amount proportional to the action. Therefore, if we know the initial wavefunction $\psi(x_0, 0)$, the wavefunction at a latter time, t, is postulated to be

$$\psi(x, t) = \int_{-\infty}^{+\infty} K(x, t; x', 0)\psi(x', 0)\, dx', \qquad (1)$$

where the Green's function kernel is given symbolically by

$$K(x, t; x_0, 0) = \frac{1}{N}\int \exp\left\{\frac{i}{\hbar}\int_0^t L[\dot{x}(\tau), x(\tau)]\, d\tau\right\}\delta[x(\tau)], \qquad (2)$$

with

$$L[\dot{x}(\tau), x(\tau)] = \tfrac{1}{2}m[\dot{x}(\tau)]^2 - V[x(\tau)].$$

The variable $x(\tau)$ is the continuous path that goes from $x(0) = x_0$ to $x(t) = x$ in a time, t. This path is arbitrary and not restricted by the equations of motion. $1/N$ is a normalizing factor introduced to insure unitarity of the wavefunction. As a consequence of unitarity, $K(x, t; x_0t_0)$ must satisfy the integral equation

$$K(x, t; x_0, t_0) = \int_{-\infty}^{+\infty} K(x', t'; x_0, t_0)K(x, t; x', t')\, dx'. \qquad (3)$$

This condition will determine the normalization constant $1/N$. The integration $\int \delta[x(\tau)]$ is meant to imply that the integration is taken over all real continuous functions $x(\tau)$ for $0 \leq \tau \leq t$ satisfying the boundary conditions $x(0) = x_0$ and $x(t) = x$. The meaning of the symbolic notation is as follows: Partition the interval $[0, t]$ into n points $\tau_0 = 0$, $\tau_1, \tau_2, \cdots, \tau_n = t$ and let

$$\int_0^t L[\dot{x}(\tau), x(\tau)]\, d\tau = \sum_{r=1}^{n}\left\{\frac{1}{2}m\frac{(x_r - x_{r-1})^2}{(\tau_r - \tau_{r-1})^2} - V(x_r)\right\}(\tau_r - \tau_{r-1}). \qquad (4)$$

Then substitute Eq. (4) into Eq. (2) and integrate over each X_r from $-\infty$ to $+\infty$ except for the end points X and X_0 which are kept fixed. Then we take the limit as $n \to \infty$. If we use Eq. (3) to solve for the normalization constant, then Eq. (2) can be written as

$$K(x, t; x_0, 0) = \lim_{n\to\infty}\left(\frac{im}{2\pi\hbar}\right)^{n/2}[\tau_1(\tau_2 - \tau_1)(\cdots)(\tau_n - \tau_{n-1})]^{\frac{1}{2}} \times \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\exp\left\{\frac{i}{\hbar}\sum_{r=1}^{n}\left[\frac{1}{2}m\frac{(x_r - x_{r-1})^2}{(\tau_r - \tau_{r-1})^2} - V(X_r)\right](\tau_r - \tau_{r-1})\right\} dx_1\, dx_2 \cdots dx_{n-1}. \qquad (5)$$

We note immediately that Eq. (5) is not well defined. The integrals are over rapidly oscillating functions and will not, in general, converge. In order to assign the integrals a definite meaning, one must resort to certain devices, such as letting h have a small negative imaginary part to insure convergence, which is allowed to go to zero after the integration is performed.

For our purpose, Eq. (2) can be expressed in a more convenient form if we make the change of variables

$$x(\tau) = x_0 + (x - x_0)\tau/t + y(\tau), \qquad (6)$$

where $x(0) = x_0$ and $x(t) = x$. The function $y(\tau)$ has the boundary conditions

$$y(0) = y(t) = 0.$$

With this change of variables, Eq. (2) becomes

$$K(x, t; x_0, 0) = \exp\frac{im(x - x_0)^2}{2\hbar t}\frac{1}{N}\int \exp\left(\frac{i}{\hbar}\int_0^t\left\{\frac{1}{2}m\dot{y}(\tau)^2 - V\left[x_0 + (x - x_0)\frac{\tau}{t} + y(\tau)\right]\right\}d\tau\right)\delta[y(\tau)], \qquad (7)$$

where the functional integral is taken over all continuous paths $y(\tau)$ in the interval $(0, t)$ such that $y(0) = y(t) = 0$.

III. ALTERNATE FORMULATION OF THE FUNCTIONAL INTEGRAL

The evaluation of Eq. (5) is usually very difficult. Davison,[9] Burton and deBorde,[10] and Davies[11] have considered a different approach. They carried out the calculations of these integrals by representing the set of all continuous paths going from x_0 to x in a time t in terms of a complete set of orthogonal functions. They calculated the action integral in terms of the orthogonal expansion and then integrated over the coefficients of the expansion, thereby taking into account the contribution of all continuous paths.

[9] B. Davison, Proc. Roy. Soc. (London) **A225**, 252 (1954).
[10] W. K. Burton and A. H. de Borde, Nuovo Cimento **2**, 197 (1955).
[11] H. Davies, Proc. Cambridge Phil. Soc. **53**, 199 (1957).

Let $\{\phi_n(z)\}$ be a complete set of orthonormal functions in the interval $0 < z < 1$ such that $\phi_0(z) = 1$. Then, $y(t')$ can be expanded in terms of $\phi_n(t'/t)$

$$\dot{y}(t') = \left(\frac{2\pi\hbar}{mt}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} a_n \phi_n\left(\frac{t'}{t}\right). \tag{8}$$

Although this method is valid for an arbitrary complete set $\{\phi_n(z)\}$ of orthonormal functions such that $\phi_0(z) = 1$, we will consider the case where

$$\phi_n(z) = \sqrt{2}\cos(n\pi z) \quad \text{for} \quad n \neq 0,$$
$$\phi_0(z) = 1. \tag{9}$$

The expansion of $y(\tau)$ is then obtained by integrating Eq. (8) subject to the boundary condition that $y(0) = y(t) = 0$

$$y(\tau) = \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{a_n}{n} \sin\left(\frac{n\pi\tau}{t}\right). \tag{10}$$

By varying the coefficients a_n, we can obtain the set of all possible paths.

In analogy to the paper by Davison, Eq. (7) can be expressed in terms of the orthonormal set given in Eq. (9) as

$$K(x, t; x_0, 0) = \exp\left(\frac{im(x - x_0)^2}{2\hbar t}\right)\left(\frac{m}{2\pi i\hbar t}\right)^{\frac{1}{2}} \tag{11}$$
$$\times \lim_{n\to\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} i^{-\frac{1}{2}n}$$
$$\times \exp\left\{\frac{i}{\hbar}\int_0^t L[\dot{y}_n(\tau), y_n(\tau)]\, d\tau\right\} da_1\, da_2 \cdots da_n, \tag{11}$$

where

$$\dot{y}_n(\tau) = \left(\frac{4\pi\hbar}{mt}\right)^{\frac{1}{2}} \sum_{l=1}^{n} a_l \cos\left(\frac{l\pi\tau}{t}\right) \tag{11a}$$

and

$$y_n(\tau) = \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{l=1}^{n} \frac{a_l}{l} \sin\left(\frac{l\pi\tau}{t}\right). \tag{11b}$$

The exponent in Eq. (11) expressed in terms of the expansion is

$$\frac{i}{\hbar}\int_0^t L[\dot{y}_n(\tau), y_n(\tau)]\, d\tau$$
$$= \frac{i}{\hbar}\int_0^t \{\tfrac{1}{2}m\dot{y}^2 - V[x + (x - x_0)\tau/t + y(\tau)]\}\, d\tau$$
$$= i\pi \sum_{n=1}^{\infty} a_n^2 - \frac{i}{t}\int_0^t V\left[x + (x - x_0)\frac{\tau}{t} + \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{a_n}{n} \sin\frac{n\pi\tau}{t}\right] d\tau. \tag{12}$$

Using Eqs. (11) and (12), $K(x, t; x_0, 0)$ can be expressed as

$$K(x, t; x_0, 0) = K_0(x, t; x_0, 0) I[V(y)], \tag{13}$$

where

$$I[V(y)] = \lim_{n\to\infty} I_n[V(y_n)]$$
$$= \lim_{n\to\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} i^{-\frac{1}{2}n} \exp\left\{i\pi \sum_{l=1}^{n} a_l^2 - \frac{i}{\hbar}\int_0^t V\left[x + (x - x_0)\frac{t}{\tau} + y_n(\tau)\right] d\tau\right\} da_1\, da_2 \cdots da_n$$

and

$$K_0(x, t; x_0, 0) = \left(\frac{m}{2\pi i\hbar t}\right)^{\frac{1}{2}} \exp\left[\frac{im(x - x_0)^2}{2\hbar t}\right].$$

It can easily be shown that $K_0(x, t; x_0, 0)$ is just the free-particle Green's function. Equation (13) is not well defined since the integrals do not in general converge. To make it well defined, we will let a_l in Eq. (13) have a small positive imaginary part in the exponent to insure convergence, and then after we are finished integrating, let it go to zero.

We have now reduced the calculation of the Green's function to solving for $I[V(y)]$. For all but the simplest kind of potentials, this is still very difficult to do. We could approximate $I[V(y)]$ by the nth term of its sequence $I_n[V(y_n)]$. The result is in general very poor since the convergence of the sequence is very slow. However, by modifying $I_n[V(y_n)]$ we can obtain a sequence that approaches its limit faster than before. In fact, it will give the exact answer for all n when the potential is of the form $V(x) = a + \lambda x + \frac{1}{2}m\omega^2 x^2$ where a, λ, $\frac{1}{2}m\omega^2$ are constants. However, before we can do this, we must derive the mixed integration formula

$$I[V(y)] = \frac{1}{(i)^{\nu/2}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left[i\pi \sum_{n=1}^{\nu} \zeta_n^2\right]$$
$$\times I\{V[y - y_\nu + \psi_\nu(\zeta)]\}\, d\zeta_1 \cdots d\zeta_\nu, \tag{14}$$

where

$$\psi_\nu(\zeta) = \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{l=1}^{\nu} \frac{\zeta_l}{l} \sin\left(\frac{l\pi\tau}{t}\right) \tag{14a}$$

and ν is an arbitrary positive integer.

To prove Eq. (14), we let ν be a given positive

integer and rewrite $I_n[V(y_n)]$ in Eq. (13) when $n > \nu$ in the form

$$I_n[V(y_n)] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} d\zeta_1 \cdots d\zeta_\nu\, i^{\frac{1}{2}\nu} \exp\left(i\pi \sum_{l=1}^{\nu} \zeta_l^2\right) \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} i^{-\frac{1}{2}(n-\nu)}$$
$$\times \exp\left\{i\pi \sum_{l=\nu+1}^{n} a_l^2 - \frac{i}{\hbar}\int_0^t V\left[x + (x - x_0)\frac{t}{\tau} + \psi_\nu(\zeta) + y_n - y_\nu\right] d\tau\right\} da_{\nu+1} \cdots da_n. \quad (15)$$

We have just changed the name of the first ν terms of a_l to ζ_l which does not affect the results. Since

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} i^{\frac{1}{2}\nu} \exp\left(i_\tau \sum_{l=1}^{\nu} a_l^2\right) da_1 \cdots da_\nu = 1$$

and

$$V[x + (x - x_0)\tau/t + \psi_\nu(\zeta) + y_n - y_\nu] = V\left\{x + (x - x_0)\frac{\tau}{t} + \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \times \left[\sum_{l=1}^{\nu} \frac{\zeta_l}{l} \sin\left(\frac{\pi l\tau}{t}\right) + \sum_{l=\nu+1}^{n} \frac{a_l}{l} \sin\left(\frac{2\pi\tau}{t}\right)\right]\right\}$$

is independent of $a_1, a_2, \cdots, a_\nu$, Eq. (15) can be written as

$$I_n[V(y_n)] = (-i)^{\nu/2} \int_{-\infty}^{+\infty} \cdots \int d\zeta_1 \cdots d\zeta_\nu \exp\left(i\pi \sum_{l=1}^{\nu} \zeta_l^2\right) \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} i^{-\frac{1}{2}n}$$
$$\times \exp\left\{i\pi \sum_{l=1}^{n} a_l^2 - i/\hbar \int_0^t V[x + (x - x_0)\tau/t + \psi_\nu(\zeta) + y_n - y_\nu]\, d\tau\right\} da_1 \cdots da_n$$
$$= (-i)^{\nu/2} \int_{-\infty}^{+\infty} \cdots \int \exp\left(i\pi \sum_{l=1}^{\nu} \zeta_l^2\right) I_n\{V[y_n + \psi_\nu(\zeta) - y_\nu]\}\, d\zeta_1 \cdots d\zeta_\nu. \quad (16)$$

By taking the limit of Eq. (16) as $n \to \infty$ and interchanging the order of integration with it, we obtain

$$I[V(y)] = (-i)^{\nu/2} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(i\pi \sum_{l=1}^{\nu} \zeta_l^2\right)$$
$$\times I\{V[y + \psi_\nu(\zeta) - y_\nu]\}\, d\zeta_1 \cdots d\zeta_\nu. \quad (14)$$

When $I[V(y)]$ exists, there is no difficulty in interchanging the limit and order of integration.

The mixed integration formula (14) can be used as a starting point for various approximations to $I[V(y)]$ or $K(x, t; x_0, 0)$. In Secs. IV and V, we will construct two such approximation methods starting from Eq. (14).

IV. FREE-PARTICLE APPROXIMATION

In this section, we use the mixed integration formula and the stationary-phase approximation to find a new sequence $I'_\nu[V(y)]$ which in the limit as $\nu \to \infty$ will equal $I[V(y)]$. The convergence of this new sequence will in many cases be very rapid, so that we may get very reliable results in approximation $I[V(y)]$ by the first few terms of the sequence $I'_\nu[V(y)]$. In fact,

$$I'_\nu[V(y)] = I[V(y)] \quad (17)$$

for all positive integers ν if the potential is of the form

$$V(x) = a + \lambda x, \quad (18)$$

where a and λ are constants.

In constructing the sequence $I'_\nu[V(y)]$, we must make use of the stationary phase approximation developed by Abe.[8] So before proceeding further, we will outline the main steps in using the stationary phase approximation on $I[V(y)]$.

Looking at Eq. (11), we notice that since $\hbar$ is small, the exponential function $e^{i/\hbar S}$, where S is the action, oscillates extremely rapidly even for small changes in S. Therefore, the contributions of the neighboring trajectories will, in general, mutually cancel, making only a small contribution to the value of the integral. The largest contribution to the integral will come from that trajectory which causes S to change the most slowly, i.e., $\delta S = 0$. But this last condition determines precisely the classical trajectory. Expanding the action $S[x(t)]$ about the classical trajectory $x^c(t)$, we have

$$S[x(t)] = S[x^c(t)] + \tfrac{1}{2}\delta^2 S[x^c(t)] + \cdots. \quad (19)$$

Keeping only the first two terms of this expansion, we obtain the result

$$I[V(y)] \simeq \left(\frac{t}{D}\right)^{\frac{1}{2}} \exp\left(\frac{i}{\hbar}\int_0^t \{\tfrac{1}{2}m[\dot{y}^c(\tau)^2] - V[x_0 + (x - x_0)\tau/t + y^c(t)]\}\, d\tau\right), \quad (20)$$

where

$$m\,d^2y^c(\tau)/d\tau^2 = -\partial V(y^c)/\partial y^c,$$
$$y^c(0) = y^c(t) = 0 \qquad (21)$$

and D is determined by the differential equation

$$md^2D/dt^2 = -D\partial^2V/\partial y^{c2}$$

with the boundary conditions

$$D(0) = 0, \qquad dD/dt\,|_{t=0} = 1. \qquad (22)$$

It has been shown that the stationary phase approximation will give the exact results for $I[V(y)]$ in Eq. (20) if the potential is either constant, linear, or quadratic functions of y.[12]

In this section, we seek to find a method of obtaining the correct solution to Eq. (13) for all potentials of the form $V(x) = a + \lambda x$. We see that this is possible if we approximate $I\{V[y-y_\nu+\psi_\nu(\zeta)]\}$ in the right-hand side of Eq. (14) by

$$I\{V[y - y_\nu + \psi_\nu(\zeta)]\} \simeq \exp(i/\hbar)S_c\{V[y^c - y^c_\nu + \psi_\nu(\zeta)]\}, \qquad (23)$$

where

$$S_c\{y^c - y^c_\nu + \psi_\nu(\zeta)\} = \hbar\pi \sum_{n=\nu+1}^{\infty} b_n^2 - \int_0^t V\Big[x + (x - x_0)\frac{\tau}{t} + \psi_\nu(\zeta) + \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=\nu+1}^{\infty} \frac{b_n}{n} \sin\frac{n\pi\tau}{t}\Big]\,d\tau \qquad (24)$$

and the b_n is determined from the set of equations

$$0 = 2\hbar\pi b_j - \frac{\partial}{\partial b_j}\int_0^t V\Big(x + (x - x_0)\frac{\tau}{t} + \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{b_n}{n} \sin\frac{n\pi\tau}{t}\Big)\,d\tau. \qquad (25)$$

These b_j are just those coefficients of the expansion

$$X(\tau) = X + (x - x_0)\frac{\tau}{t} + \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{b_n}{n} \sin\frac{n\pi\tau}{t}, \qquad (26)$$

which make $x(\tau)$ the classical path for a particle in a potential $V(x)$ and satisfying the boundary conditions $x(t) = x$, $x(0) = x_0$.

That is instead of finding the b_j by the relations given in Eq. (25). We could have obtained the classical path from the equation

$$md^2x/d\tau^2 = -\partial V(x)/\partial x \qquad (27)$$

satisfying the boundary conditions

$$x(0) = x_0, \qquad x(t) = x'.$$

Then the b_j could be found by expanding $y^c = [x^c - x - (x - x_0)\tau/t]$ into the Fourier series

$$\left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{b_n}{n} \sin\frac{n\pi\tau}{t}. \qquad (28)$$

To obtain the free-particle approximation, we substituted Eq. (23) into the right-hand side of the mixed integration formula of Eq. (14) and obtained the sequence

$$I'_\nu[V(y)] = (-i)^{\nu/2}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} d\zeta_1\cdots d\zeta_\nu \times \exp\left(i\pi\sum_{l=1}^{\nu}\zeta_l^2\right)\exp\left\{\frac{i}{\hbar}\int_0^t \tfrac{1}{2}m(y^c)^2\,d\tau - \frac{i}{\hbar}\int_0^t V\Big[x + (x - x_0)\frac{\tau}{t} + \psi_\nu(\zeta) + y^c - y^c_\nu\Big]\,dt\right\} \qquad (29)$$

where y^c is given by Eqs. (28) with the b_j determined by Eq. (25). Notice that for all $\nu > 0$

$$I'_\nu[V(y)] = I[V(y)] \qquad (30)$$

if the potential is of the form

$$V = a + \lambda x. \qquad (31)$$

This is true since Eq. (23) is exact for potentials of this form and, therefore, the mixed integration formula with Eq. (23) in its right-hand side is exact.

For other potentials, we have

$$I[V(y)] = \lim_{\nu\to\infty} I'_\nu[V(y)]. \qquad (32)$$

This is true since when we take the limit as $\nu \to \infty$ of Eq. (24), then the potential becomes

$$\lim_{\nu\to\infty} V[x + (x - x_0)t/\tau + \psi_\nu(\zeta) + y^c - y^c_\nu] = V[x + (x - x_0)t/\tau + \psi_\infty(\zeta)]. \qquad (33)$$

The potential is independent of the coordinate y, hence the classical path $y^c(t)$ with the boundary condition $y^c(\tau) = y^c(0) = 0$ is $y^c(t) \equiv 0$. Therefore,

$$S_c = -\int_0^t V[x + (x + x_0)\tau/t + \psi_\nu(\zeta)]\,dt \qquad (34)$$

and Eq. (29) becomes in the limit as $\nu \to \infty$

$$\lim_{\nu\to\infty} I'_\nu[V(y)] = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\exp\frac{i}{\hbar} \times \int_0^t V[x + (x - x_0)\tau/t + \psi_\infty(\zeta)]\,d\tau \times \exp\left(i\pi\sum_{l=1}^{\infty}\zeta_l^2\right)\frac{d\zeta_1}{(i)^{1/2}}\frac{d\zeta_2}{(i)^{1/2}}\cdots\frac{d\zeta_\infty}{(i)^{1/2}}. \qquad (35)$$

[12] C. Morette, Phys. Rev. 81, 848 (1951).

Comparing Eq. (35) with Eq. (13), we see that

$$I[V(y)] = \lim_{\nu\to\infty} I'_\nu[V(y)]. \tag{36}$$

The advantage of this new sequence $I'_\nu[V(y)]$ is that it is expected to converge very rapidly since the classical path makes the greatest contribution to the integral.

As a matter of illustration, we will let

$$V(x) = \tfrac{1}{2}m\omega^2x^2, \tag{37}$$

and calculate $I'_\nu[V(y)]$ from Eq. (29). From Eq. (A5) in the appendix, we see that the classical path y^c that goes in Eq. (29) is

$$y^c(\tau) = \frac{2}{\pi}\left(\frac{\omega\tau}{\pi}\right)^2 \sum_{n=\nu+1}^{\infty} \frac{[x_0 - (-1)^n x]}{n^3(1 - \omega^2\tau^2/n^2\pi^2)} \sin\left(\frac{n\pi\tau}{t}\right).$$

Therefore, $(i/\hbar)S_c$ in Eq. (29) becomes

$$\begin{aligned}\frac{i}{\hbar}S_c = {} & \frac{imt}{2\hbar}\left[-\frac{\omega^2}{3}(x^2 + xx_0 + x_0^2)\right] \\ & - \frac{im\omega^4t^3}{\hbar\pi^4}\sum_{n=\nu+1}^{\infty}\frac{[x^2 + x_0^2 - 2xx_0(-1)^n]}{n^2[n^2 - (\omega t/\pi)^2]} \\ & + i\sum_{n=1}^{\nu}\left[\pi\left(-\frac{\omega^2t^2}{n^2\pi^2}\right)\zeta_n^2\right. \\ & \left. - \left(\frac{\omega t}{n\pi}\right)^2\left(\frac{4\pi m}{\hbar t}\right)^{\frac{1}{2}}[x - (-1)^n x_0]\zeta_n\right].\end{aligned} \tag{38}$$

Substituting this into Eq. (29), we get

$$\begin{aligned}I'_\nu[V(y)] = {} & \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} i^{-\frac{1}{2}\nu} \\ & \times \exp\left[i\sum_{n=1}^{\nu}\pi\left(1 - \frac{\omega^2t^2}{n^2\pi^2}\right)\zeta_n^2\right] \\ & \times \exp\left\{-i\left(\frac{\omega t}{n\pi}\right)^2\left(\frac{4\pi m}{\hbar t}\right)^{\frac{1}{2}}\sum_{n=1}^{\nu}[x_0 - (-1)^n x]\zeta_n\right\} \\ & \times \exp\left\{-\frac{im\omega^4t^3}{\hbar\pi^4}\sum_{n=\nu+1}^{\infty}\frac{[x^2 + x_0^2 - 2xx_0(-1)^n]}{n^2[n^2 - (\omega t/\pi)^2]}\right. \\ & \left. - \frac{imt}{6\hbar}\omega^2(x^2 + xx_0 + x_0^2)\right\} d\zeta_1\cdots d\zeta_\nu.\end{aligned} \tag{39}$$

After performing the integration, Eq. (39) becomes

$$\begin{aligned}I'_\nu[V(y)] = {} & \prod_{n=1}^{\nu}\left(1 - \frac{\omega^2t^2}{n^2\pi^2}\right)^{-\frac{1}{2}} \\ & \times \exp\left\{\frac{-im\omega^4t^3}{\hbar\pi^4}\sum_{n=1}^{\infty}\frac{[x^2 + x_0^2 - 2xx_0(-1)^n]}{n^2[n^2 - (\omega t/\pi)^2]}\right\} \\ & \times \exp\left[-i\frac{mt}{6\hbar}\omega^2(x^2 + xx_0 + x_0^2)\right] \\ = {} & \prod_{n=1}^{\nu}\left(1 - \frac{\omega^2t^2}{n^2\pi^2}\right)^{-\frac{1}{2}} \exp\left\{\frac{im\omega}{2\hbar}[(x^2 + x_0^2)\operatorname{ctn}\omega t\right. \\ & \left. - 2xx_0\csc\omega t]\right\}\exp\left[\frac{-im}{2\hbar}\frac{(x - x_0)^2}{t}\right].\end{aligned} \tag{40}$$

Taking the limit $\nu\to\infty$ of Eq. (40), we get

$$\begin{aligned}I[V(y)] = {} & \lim_{\nu\to\infty} I'_\nu[V(y)] \\ = {} & \left(\frac{\omega t}{\sin\omega t}\right)^{\frac{1}{2}}\exp\left\{\frac{-im}{2\hbar t}(x - x_0)^2\right. \\ & \left. + \frac{im\omega}{2\hbar}[(x^2 + x_0^2)\operatorname{ctn}\omega t - 2xx_0\csc\omega t]\right\}.\end{aligned} \tag{41}$$

This is by no means the simplest way to calculate $I[V(y)]$ for the harmonic potential, but we are interested in $I'_\nu[V(y)]$. Notice that we would have gotten very good agreement between $I[V(y)]$ and $I'_\nu[V(y)]$ if we had chosen ν such that

$$\omega t/\nu\pi < 1. \tag{42}$$

This is true since

$$I'_\nu[V(y)] = \prod_{n=\nu}^{\infty}\left(1 - \frac{\omega^2t^2}{n^2\pi^2}\right)I[V(y)], \tag{43}$$

and since $\omega t/\nu\pi < 1$,

$$I'_\nu[V(y)] \simeq I[V(y)] + O(\omega t/2\nu\pi). \tag{44}$$

V. HARMONIC-OSCILLATOR APPROXIMATION

We now wish to generalize the results of the last section. We generalize the sequence $I'_\nu[V(y)]$ given by Eq. (29) so that it will be exact for all ν if the potential is of the form

$$V(x) = a + \lambda x + \tfrac{1}{2}m\omega^2x^2. \tag{45}$$

The sequence $I''_\nu[V(y)]$ that has these properties is

$$I''_\nu[V(y)] = \prod_{n=\nu}^{\infty}\left(1 - \frac{\omega^2t^2}{n^2\pi^2}\right)^{-\frac{1}{2}} I'_\nu[V(y)], \tag{46}$$

where $I'_\nu[V(y)]$ is defined by Eq. (29). For $\omega = 0$ this is identical to Eq. (29). Also in the limit $\nu\to\infty$

$$\begin{aligned}\lim_{\nu\to\infty} I''_\nu[V(y)] = {} & \lim_{\nu\to\infty}\prod_{n=\nu}^{\infty}\left(1 - \frac{\omega^2t^2}{n^2\pi^2}\right)I'_\nu[V(y)] \\ = {} & \lim_{\nu\to\infty} I'_\nu[V(y)] \\ = {} & I[V(y)].\end{aligned} \tag{47}$$

It is also straightforward to show that

$$I''_\nu[V(y)] = I[V(y)] \tag{48}$$

for all ν if $V(x)$ is of the form

$$V(x) = a + \lambda x + \tfrac{1}{2}m\omega^2x^2.$$

Therefore, $I''_\nu[V(y)]$ is the desired generalized sequence. This sequence as defined in Eq. (46) will be called the harmonic-oscillator approximation. For the same reasons stated in the last section, the convergence of $I''_\nu[V(y)]$ is expected to be very rapid.

We have now reduced the problem of solving the Green's function $K(x, t; x_0, 0)$ to a solution of Eq. (46). In order to solve Eq. (46), the classical trajectory must be known. However, in most problems of interest in quantum mechanics, the classical path itself is a very complicated function when written in the form $x(t)$. So the problem of solving the lowest order terms in Eq. (46) will be straightforward, but rather tedious.

VI. CONCLUSION

In Sec. III, we derived the mixed integration formula (14). From this formula various methods for solving the functional integrals can be derived. For example Cameron uses it to get his Simpson rule. The Simpson rule was constructed to give good results for functionals that could be approximated by third-degree polynomial functionals. This is usually not the case for most problems in quantum mechanics where one is working with expontenial functionals and a polynomial functional of third degree is usually a poor approximation. So in Secs. IV and V we have derived with the aid of Eq. (14) two different methods of solving functional integrals.

The free-particle approximation given by Eq. (29) is constructed to give exact results in all orders of the approximation for exponential functionals of the form $\exp[\int_0^t (a + bx)\, d\tau]$, where a and b are independent of x. In Sec. V we improved the free-particle approximation so as to give exact answer in all orders of the approximation for exponential functionals of the form $\exp[\int_0^t (a + bx + cx^2)\, d\tau]$, where a, b, and c are independent of x. We called this new method the harmonic-oscillator approximation and it is given by Eq. (46).

For other types of exponential functionals the harmonic-oscillator and free-particle approximations can be made as accurate as desired by taking higher order terms in the sequence [cf. Eqs. (29) and (46)]. However, for the same reasons as previously stated it is felt that only the first few terms of the sequence will be needed to give reliable answers for most potentials. Although these methods of approximations were only derived in one dimension they can easily be generalized to many dimensions.

Many other approximating schemes can be constructed from the mixed integration formula (14); however, the two that were developed in this paper seem to be the most useful for the problems that arise in nonrelativistic quantum mechanics.

ACKNOWLEDGMENT

The author is grateful to Professor Ernest M. Henley for comments and criticisms.

APPENDIX

We wish to solve the classical trajectory with the potential of the form

$$V(y) = \tfrac{1}{2}m\omega^2[x + (x - x_0)\tau/t + \psi_\nu(\zeta) + y - y_\nu]^2, \quad \text{(A1)}$$

where the notation is defined in Eqs. (11a) and (14a). The classical trajectory could be obtained from Eq. (21); however, for our case it is much simpler to find those coefficients that make S in Eq. (24) an extremum, that is, those coefficients that are determined by Eq. (25).

Since $y(\tau)$ must satisfy the boundary conditions $y(0) = y(t) = 0$, we can expand it in the series

$$y(\tau) = \left(\frac{4\hbar t}{m\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{a_n}{n} \sin\left(\frac{n\pi\tau}{t}\right). \quad \text{(A2)}$$

We can write the action S in terms of the expansion as

$$S = \hbar\pi \sum_{n=1}^{\infty} a_n^2 - \sum_{n=\nu+1}^{\infty} \frac{\pi\hbar\omega^2 t^2 a_n^2}{n^2\pi^2} - \sum_{n=\nu+1}^{\infty} \left(\frac{\omega t}{n\pi}\right)^2 \left(\frac{4\pi m\hbar}{t}\right)^{\frac{1}{2}} [x_0 - (-1)^n x] a_n - \tfrac{1}{3}\omega^2(x^2 + xx_0 - x_0^2) + \cdots. \quad \text{(A3)}$$

We then find the values of a_n that make S an extremum. Substituting these values of a_n in Eq. (A2) will give us the classical trajectory.

The values of a_n that make S an extremum can be found by differentiating S in Eq. (A3) with respect to the coefficients a_n and then setting the expression equal to zero and solving for the a_n's. This is equivalent to Eq. (25). Doing this we get the following values for the a_n's:

$$a_n = 0 \quad \text{for} \quad 1 \leq n < \nu,$$

$$a_n = \left(\frac{\omega t}{n\pi}\right) \frac{(m/\hbar\pi t)^{\frac{1}{2}}[x_0 - (-1)^n x]}{(1 - \omega^2 t^2/n^2\pi^2)} \quad \text{for} \quad \nu < n. \quad \text{(A4)}$$

Therefore,

$$y^c(\tau) = \frac{2}{\pi}\left(\frac{\omega\tau}{\pi}\right)^2 \sum_{n=\nu+1}^{\infty} \frac{[x_0 - (-1)^n x]}{n^3(1 - \omega^2\tau^2/n^2\pi^2)} \sin\left(\frac{n\pi\tau}{t}\right). \quad \text{(A5)}$$

Time-Dependent One-Speed Albedo Problem for a Semi-Infinite Medium

I. KUŠČER* AND P. F. ZWEIFEL
Department of Nuclear Engineering, The University of Michigan, Ann Arbor, Michigan
(Received 19 May 1964)

A Laplace transformation technique is used to determine the neutron distribution in a semi-infinite medium which has been irradiated by a neutron pulse. The result is given in terms of known solutions of Milne's problem and of the steady-state albedo problem, which in turn are expressed by aid of Case's X-function. Simple asymptotic approximations, valid for $t \gg 1$, are deduced from the exact result.

I. INTRODUCTION

IT is well known that time-dependent transport problems with given initial values can be formally converted to steady-state problems by Laplace transformation. In simple cases the transformed equation can be solved rigorously, e.g., by the singular eigenfunction method of Case.[1-3] Then the solution of the time-dependent problem is constructed by inverse Laplace transformation.

The indicated method has been used by Bowden[4] for a problem with slab geometry, the general aspects of which problem have been clarified previously by Lehner and Wing.[5] A slightly different approach has been used by Case[1] for an infinite medium with a pulsed plane source. It seems worthwhile to extend these investigations also to the semi-infinite medium, in which case several explicit results can be deduced.

We restrict our attention to the one-speed equation with isotropic scattering and seek the neutron distribution everywhere in an infinite half-space following irradiation of the surface with a monodirectional pulse of neutrons at $t=0$. The appropriate equation (using units in which $\sigma = v = 1$) is[1]

$$\psi(x, \mu, t; \mu_0) + \frac{\partial \psi}{\partial t} + \mu \frac{\partial \psi}{\partial x} = \frac{c}{2} \int_{-1}^{1} \psi(x, \mu', t; \mu_0)\, d\mu', \qquad (1a)$$

where $x \geq 0$, $\mu_0 > 0$, and the boundary and initial conditions are

$$\psi(0, \mu, t; \mu_0) = \mu_0^{-1} \delta(\mu - \mu_0)\delta(t) \quad \text{for} \quad \mu > 0, \qquad (1b)$$

$$\psi(x, \mu, t; \mu_0) \to 0 \quad \text{for} \quad x \to \infty, \qquad (1c)$$

and

$$\psi(x, \mu, t; \mu_0) = 0 \quad \text{for} \quad t < 0.$$

We shall also be interested in the distribution

$$\psi^*(x, \mu, t) = \int_0^1 \psi(x, \mu, t; \mu_0)\mu_0\, d\mu_0, \qquad (2)$$

produced by a pulsed isotropic incident distribution.

Finally, we shall need the values of the neutron densities and net currents, defined by

$$\rho(x, t; \mu_0) = \int_{-1}^{1} \psi(x, \mu, t; \mu_0)\, d\mu, \qquad (3)$$

$$j(x, t; \mu_0) = \int_{-1}^{1} \psi(x, \mu, t; \mu_0)\mu\, d\mu, \qquad (4)$$

(and similarly for ρ^*, j^*). For convenience the factor 2π has been omitted here, which can be justified by saying that ψ represents the angular density integrated over the azimuth.

Certain general properties of the solution are immediately apparent. First, we notice that the pulse initiates some transient discontinuities in the neutron distribution. Evidently $\psi = 0$ for $x > t$, since the neutrons enter the medium with a speed which is unity in the present notation. Moreover, a term $\delta(x - \mu_0 t)\delta(\mu - \mu_0)e^{-t}$, describing the distribution of the uncollided neutrons, is contained in ψ. However, all such singularities die out exponentially, and ψ becomes a smooth function for $t \gg 1$.

Second, according to a reciprocity theorem,[6] the following relation for the angular density of the reflected neutrons must hold

$$\psi(0, -\mu, t; \mu_0) = \psi(0, -\mu_0, t; \mu), \qquad \mu > 0. \qquad (5)$$

Finally, for an absorbing medium ($c < 1$), we expect that the decay of the neutron distribution is governed mainly by the true absorption rate, i.e.,

* On leave of absence from the University of Ljubljana, Yugoslavia.

[1] K. M. Case, Ann. Phys. (N. Y.) 9, 1 (1960).

[2] K. M. Case, *Recent Developments in Neutron Transport Theory* (Michigan Memorial Phoenix Project, 1961).

[3] K. M. Case and P. F. Zweifel, *Neutron Transport Theory* (to be published).

[4] R. L. Bowden, thesis, Virginia Polytechnic Institute (Report TID 18 884, 1963). See also: R. L. Bowden and C. D. Williams, J. Math. Phys. 5, 1527 (1964).

[5] G. M. Wing, *An Introduction to Transport Theory* (J. Wiley & Sons, Inc., New York, 1962).

[6] L. M. Biberman and B. A. Veklenko, Zh. Eksperim. i Teor. Fiz. 39, 88 (1960) [English transl.: Soviet Phys.—JETP 12, 64 (1961)].

ψ should be roughly proportional to $e^{-(1-c)t}$. After an appropriate substitution is made, Eq. (1a) shows that

$$\psi^{(c)}(x, \mu, t; \mu_0) = ce^{-(1-c)t}\psi^{(1)}(cx, \mu, ct; \mu_0), \quad (6)$$

where the value of c is indicated by a superscript. Hence it is sufficient to study the problem for a nonabsorbing medium ($c = 1$), and therefore the subsequent discussion will be limited to this case only.

Following Lehner and Wing[5] and Bowden,[4] we multiply both sides of (1a)–(1c), where now $c = 1$, by $e^{(1-s)t}dt$, and integrate from 0 to ∞. The integral converges for Re $(s) > 1$, and the transform

$$\psi_s(x, \mu; \mu_0) = \int_0^\infty \psi(x, \mu, t; \mu_0)e^{(1-s)t}\, dt \quad (7)$$

is found to obey the equation

$$s\psi_s(x, \mu; \mu_0) + \mu \frac{\partial \psi_s}{\partial x} = \frac{1}{2}\int_{-1}^{1} \psi_s(x, \mu'; \mu_0)\, d\mu', \quad (8a)$$

with the boundary conditions

$$\psi_s(0, \mu; \mu_0) = \mu_0^{-1}\delta(\mu - \mu_0) \quad \text{for} \quad \mu > 0, \quad (8b)$$

and

$$\psi_s(x, \mu; \mu_0) \to 0 \quad \text{for} \quad x \to \infty. \quad (8c)$$

From $\psi_s(x, \mu; \mu_0)$ the solution of the time-dependent problem will be computed by inverse Laplace transformation,

$$\psi(x, \mu, t; \mu_0) = \lim_{\omega\to\infty} \frac{1}{2\pi i}\int_{\gamma-i\omega}^{\gamma+i\omega} \psi_s(x, \mu; \mu_0)e^{-(1-s)t}\, ds, \quad (9)$$

where $\gamma > 1$. However, before carrying out this inverse transformation it seems advisable to modify it in the usual way by shifting and bending the path of integration as far as possible to the left in the complex s-plane. In order to be able to do this we first must check the analyticity properties of $\psi_s(x, \mu; \mu_0)$ as a function of s. We shall start with explicit expressions for this function.

II. PROPERTIES OF THE TRANSFORM OF THE SOLUTION

According to Eqs. (8a)–(8c), the function $\psi_s(x, \mu; \mu_0)$ coincides with the solution of the steady-state albedo problem, normalized to unit ingoing net current, for a semi-infinite medium with a macroscopic total cross section s and a macroscopic scattering cross section equal to unity. This problem has been solved by Case, and we can copy his results, at least for real s. The only novelty encountered with the present problem lies in the necessity of performing an analytical continuation to complex values of s. We shall postpone this task temporarily, and start with the assumption that s is real and > 1.

Besides $\psi_s(x, \mu; \mu_0)$ we shall need later on the solution $\psi_s(x, \mu)$ of Milne's problem, which is defined by an equation like (8a), and the boundary conditions

$$\psi_s(0, \mu) = 0 \quad \text{for} \quad \mu > 0, \quad (10a)$$

$$\psi_s(x, \mu) < O(e^{sx}) \quad \text{for} \quad x \to \infty. \quad (10b)$$

In both cases we shall be interested also in the neutron densities and net currents, which will be denoted by $\rho_s(x; \mu_0)$, $\rho_s(x)$ and $j_s(x; \mu_0)$, $j_s(x)$, respectively. We normalize the solution of Milne's problem to unit emerging net current: $j_s(0) = -1$.

All these quantities can be expressed in terms of Case's X-function, or equivalently, in terms of Chandrasekhar's H-function,[7] $H(\mu, s) = [(1 - s^{-1})^{\frac{1}{2}}(\nu_0 + \mu)X(-\mu, s)]^{-1}$. In Case's notation the formulas for $\psi_s(x, \mu; \mu_0)$, etc., are[2,3,7,8]

$$\psi_s(x, \mu; \mu_0) = \frac{1}{(\nu_0 + \mu_0)X(-\mu_0, s)} \times \left\{\frac{4sX(-\nu_0, s)}{\nu_0\Lambda_s'(\nu_0)}\phi_{s+}(\mu_0)\phi_{s+}(\mu)e^{-sx/\nu_0} + \int_0^1 \frac{(\nu_0 + \nu)X(-\nu, s)}{\nu\Lambda_s^+(\nu)\Lambda_s^-(\nu)}\phi_{s\nu}(\mu_0)\phi_{s\nu}(\mu)e^{-sx/\nu}\, d\nu\right\}, \quad (11)$$

$$\psi_s(0, -\mu; \mu_0) = [2(s - 1)(\mu + \mu_0)(\nu_0 + \mu_0) \times X(-\mu_0, s)(\nu_0 + \mu)X(-\mu, s)]^{-1}, \quad \mu \geq 0, \quad (12)$$

$$\rho_s(0; \mu_0) = [(1 - s^{-1})^{\frac{1}{2}}\mu_0(\nu_0 + \mu_0)X(-\mu_0, s)]^{-1}, \quad (13)$$

$$j_s(0; \mu_0) = [(\nu_0 + \mu_0)X(-\mu_0, s)]^{-1}, \quad (14)$$

$$\psi_s(x, \mu) = \frac{1}{2(1 - s^{-1})\nu_0^2}\left[\frac{1}{X(\nu_0, s)}\phi_{s+}(\mu)e^{-sx/\nu_0} + \frac{1}{X(-\nu_0, s)}\phi_{s-}(\mu)e^{sx/\nu_0}\right] - \frac{1}{2s}\int_0^1 \frac{X(-\nu, s)}{\Lambda_s^+(\nu)\Lambda_s^-(\nu)}\phi_{s\nu}(\mu)e^{-sx/\nu}\, d\nu, \quad (15)$$

$$\psi_s(0, -\mu) = [2(s - 1)(\nu_0^2 - \mu^2)X(-\mu, s)]^{-1}, \quad \mu \geq 0, \quad (16)$$

$$\rho_s(0) = [(1 - s^{-1})^{\frac{1}{2}}\nu_0]^{-1}. \quad (17)$$

The following functions appear in the above formulas:

$$X(z, s) = \frac{1}{1 - z}\exp\left[\frac{1}{2\pi i}\int_0^1 \ln\frac{\Lambda_s^+(\nu)}{\Lambda_s^-(\nu)}\frac{d\nu}{\nu - z}\right] \quad (18)$$

[7] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London and New York, 1950).
[8] I. Kuščer, Can. J. Phys. **31**, 1187 (1953).

(with the integrand = 0 at $\nu = 0$),

$$\Lambda_s(z) = 1 - (z/s)\tanh^{-1}(1/z) \qquad (19a)$$

(defined in the complex plane, cut along $-1 < z < 1$),

$$\Lambda_s^{\pm}(\nu) = \lambda_s(\nu) \pm \pi i\nu/2s \quad \text{for} \quad -1 < \nu < 1, \qquad (19b)$$

$$\lambda_s(\nu) = 1 - (\nu/s)\tanh^{-1}\nu, \qquad (20)$$

$$\pm\nu_0(s) = \text{roots of } \Lambda_s(\nu_0) = 0,$$

$$\Lambda_s'(\nu_0) = \nu_0 s^{-1}(\nu_0^2 - 1)^{-1} - \nu_0^{-1}, \qquad (21)$$

$$\phi_{s\pm}(\mu) = \frac{\nu_0}{2s}\frac{1}{\nu_0 \mp \mu}, \qquad (22)$$

$$\phi_{s\nu}(\mu) = \frac{\nu}{2s}P\frac{1}{\nu - \mu} + \lambda_s(\nu)\delta(\nu - \mu), \qquad (23)$$

where P indicates that we have to take the Cauchy principal value of any integral over ν or μ of the expression $1/(\nu - \mu)$ following that symbol. The integral in (11), with the two singularities of the integrand merging when $\mu \to \mu_0$, has to be understood in the same sense as with the orthogonality relation

$$\int_{-1}^{1} \phi_{s\nu}(\mu)\phi_{s\nu'}(\mu)\mu\, d\mu = \nu\Lambda_s^+(\nu)\Lambda_s^-(\nu)\delta(\nu - \nu'), \qquad (24)$$

used in full-range developments.[1-4] It can then be seen that the right-hand side of Eq. (11) contains the discrete term $\mu_0^{-1}\delta(\mu - \mu_0)e^{-sx/\mu_0}$, corresponding to the uncollided neutron beam.

The neutron densities and net currents, belonging to (11) and (15), follow immediately if we observe that $\phi_{s\pm}$ and $\phi_{s\nu}$ are normalized to unit density, and that the corresponding net currents are $\pm(1 - s^{-1})\nu_0$ and $(1 - s^{-1})\nu$, respectively.

We may introduce the "extrapolation distance" $q(s)$ and another parameter $Q(s)$ by

$$-X(\nu_0, s)/X(-\nu_0, s) = e^{2q/\nu_0}, \qquad (25)$$

$$-X(\nu_0, s)X(-\nu_0, s) = \frac{\Lambda_s'(\nu_0)}{2\nu_0(1 - s^{-1})} = \frac{1}{Q^2\nu_0^6(1 - s^{-1})^2}, \qquad (26)$$

with the purpose of expressing $\rho_s(x)$ in a shorter form

$$\rho_s(x) = Q\nu_0 \sinh\frac{sx + q}{\nu_0} - \frac{1}{2s}\int_0^1 \frac{X(-\nu, s)}{\Lambda_s^+(\nu)\Lambda_s^-(\nu)} e^{-sx/\nu}\, d\nu. \qquad (27)$$

When $s \to 1$ one should use the well-known approximation

$$\nu_0 \approx [3(s - 1)]^{-\frac{1}{2}}, \qquad (28)$$

which leads to $Q(1) = 3$, whereas $q(1) = 0.71045$.

Let us now turn to complex values of s. By retracing the derivation of Eqs. (11)–(17) one verifies that they remain valid so long as s is such that $\Lambda_s(z)$ has a pair of zeros. The condition for this to happen is, according to Bowden,[4] that s belongs to a certain region S_i of the complex plane, as shown by Fig. 1. This region is the conformal map, produced by the function $s = \nu_0\tanh^{-1}(1/\nu_0)$, of the (say) right-hand half of the complex plane of ν_0, cut as mentioned before. Hence the boundary C of S_i is the (double) conformal map of half of this cut.

The analytic behavior of $\psi_s(x, \mu; \mu_0)$ inside the region S_i of the complex s-plane is linked to the properties of $\nu_0(s)$. This is the inverse of the previously mentioned function $s = \nu_0\tanh^{-1}(1/\nu_0)$, and its values can be read from the quoted figure. We see there that the point $s = 1$ is a branch point of $\nu_0(s)$, as shown also by the approximation (28). Hence, if we want ν_0 to be uniquely determined for $s \in S_i$, a cut has to be drawn in the s-plane, most conveniently to the left of that point. If we chose ν_0 to be the particular root which is positive for $s > 1$, then $\mathrm{Re}\,(\nu_0) \geq 0$ in the whole cut region S_i.

Expression (11) shows that inside S_i the function $\psi_s(x, \mu; \mu_0)$ is regular in s, except for the branch cut $(0 \leq s \leq 1)$ due to $\nu_0(s)$. The reason why this cut is inherited by $\psi_s(x, \mu; \mu_0)$ is that, by definition, only one of the discrete eigenfunctions, $\phi_{s+}(\mu)e^{-sx/\nu_0}$, is involved in the expansion (11). Consequently, when s approaches the branch cut from above or below, two different limits ψ_s^+ and ψ_s^- are obtained, involving the negative and positive imaginary ν_0 in (11), respectively. Since ψ_s^+ and ψ_s^- both are solutions of Eqs. (8a, b), the difference $\psi_s^+ - \psi_s^-$ is a solution of the corresponding homogeneous problem, i.e., of Milne's problem. (For Milne's problem $s = 1$ is no branch point because going around this point merely interchanges the two discrete terms in (15) and leaves the sum unchanged.) Taking account of the value (14) of the net current at the surface we find that

$$\begin{aligned}\psi_s^+(x, \mu; \mu_0) - \psi_s^-(x, \mu; \mu_0) &= -2i\,|\nu_0|\,[(|\nu_0|^2 + \mu_0^2)X(-\mu_0, s)]^{-1}\psi_s(x, \mu) \\ &= -4i(1 - s)\,|\nu_0|\,\psi_s(0, -\mu_0)\psi_s(x, \mu), \quad 0 \leq s \leq 1.\end{aligned} \qquad (29)$$

If s is in the external region S_e (Fig. 1), the situation is different because $\Lambda_s(z)$ then has no zero and

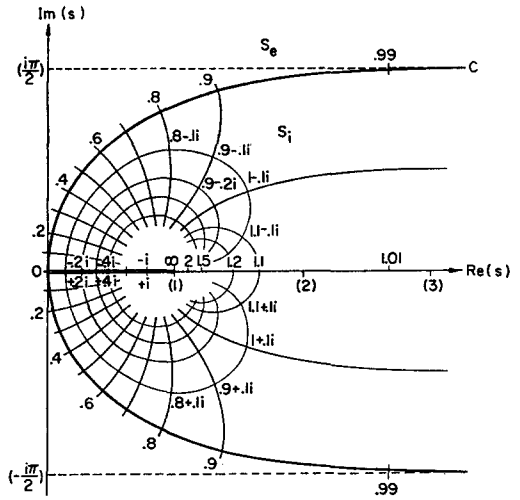


FIG. 1. Values of $\nu_0(s)$ plotted in the complex plane of s. The values of s are given in parentheses.

the corresponding discrete term in the expansion is missing. By the use of $X_0(z, s) = (1 - z)X(z, s)$ instead of $X(z, s)$ the following formulas are obtained for this case:

$$\psi_s(x, \mu; \mu_0) = \frac{1}{X_0(-\mu_0, s)} \times \int_0^1 \frac{X_0(-\nu, s)}{\nu \Lambda_s^+(\nu)\Lambda_s^-(\nu)} \phi_{s\nu}(\mu_0)\phi_{s\nu}(\mu) e^{-sx/\nu}\, d\nu, \qquad (30)$$

$$\psi_s(0, -\mu; \mu_0) = [2(s - 1)(\mu + \mu_0) \times X_0(-\mu_0, s)X_0(-\mu, s)]^{-1}, \qquad \mu \geq 0. \qquad (31)$$

Similarly as for $s \in S_i$, we now conclude from (30) that $\psi_s(x, \mu; \mu_0)$ is regular in s also for $s \in S_e$, Re $(s) \geq 0$. However, we are still uncertain about what happens when s crosses the boundary C separating the two regions.

One way to assure the analyticity of $\psi_s(x, \mu; \mu_0)$ across C would be to extend the existence theorems, worked out by Lehner and Wing for the slab case,[4,5] to the semi-infinite medium. An alternative method, chosen in the following, consists in the comparison of the limits of the explicit expressions (11) and (30), when s approaches the line C from one or the other side. First however, we have to insert a discussion about the X-function, which itself is discontinuous at $s \in C$.

It can be inferred from the definition (18) that the change of $X(z, s)$, as s crosses C, is expressed by

$$[(\nu_0 - z)X(z, s)]_{\substack{\text{inside} \\ \text{limit}}} = [(1 - z)X(z, s)]_{\substack{\text{outside} \\ \text{limit}}}. \qquad (32)$$

Thus, $(\nu_0 - z)X(z, s)$ for $s \in S_i$ and $(1 - z)X(z, s)$ for $s \in S_e$ represent one and the same analytic function, which we may denote by $X_0(z, s)$, if the definition is adapted as follows:

$$X_0(z, s) = \begin{cases} (1 - z)X(z, s) & \text{for } s \in S_e, \\ (\nu_0 - z)X(z, s) & \text{for } s \in S_i, \end{cases} \qquad (33)$$

with $X(z, s)$ given by (18).

It may be mentioned that the analyticity of $X_0(z, s)$ in both variables is obvious from a complex representation, which in a different form has been given by Chandrasekhar,[7] and which also readily ensues from the above definition:

$$\ln X_0(z, s) = \begin{cases} \dfrac{1}{2\pi i}\displaystyle\int_{-i\infty}^{i\infty} \ln \frac{\Lambda_s(z')}{\Lambda_s(\infty)} \frac{dz'}{z' - z} \\ \quad + \ln \dfrac{\Lambda_s(z)}{\Lambda_s(\infty)}, \qquad \text{Re } (z) > 0 \\ \dfrac{1}{2\pi i}\displaystyle\int_{-i\infty}^{i\infty} \ln \frac{\Lambda_s(z')}{\Lambda_s(\infty)} \frac{dz'}{z' - z}, \\ \qquad\qquad \text{Re } (z) < 0, \end{cases} \qquad (34)$$

with $\Lambda_s(\infty) = 1 - s^{-1}$. For fixed s the function $X_0(z, s)$ has no singularity in the z-plane, cut along (0, 1), and it has one simple zero at $z = \nu_0(s)$, Re $(\nu_0) \geq 0$, only if $s \in S_i$. The zero disappears by crossing the cut when s crosses the boundary C. Vice versa, for fixed z there is no singularity in the s-plane, cut along (0, 1), and for Re $(z) > 0$ there is only one zero at $s = z \tanh^{-1} (1/z)$. The zero disappears by crossing the cut when z crosses the imaginary axis.

We return now to the problem of the behavior of $\psi_s(x, \mu; \mu_0)$ at $s \in C$. For the case $x = 0$, it is immediately clear, in view of Eq. (33), that (31) is an analytical continuation of (12). For $x > 0$, an apparent difficulty arises from the discontinuity of the individual terms in (11) and (30), when s crosses C. Moreover, certain terms have poles at those values of s which make $\nu_0(s)$ equal to μ or μ_0. However, a closer inspection proves that all these singularities cancel each other, so that $\psi_s(x, \mu; \mu_0)$ is, in fact, continuous across C, and consequently regular in the whole right-hand half-plane of s, cut along $0 \leq s \leq 1$. This is what we wanted to know.

The tedious term-by-term comparison of (11) with (30) can be avoided by transforming both expressions into a unique complex representation, from which the analyticity in s is evident:

$$\psi_s(x, \mu; \mu_0) = \mu_0^{-1}\delta(\mu - \mu_0)e^{-sx/\mu_0} - \frac{1}{2sX_0(-\mu_0, s)} \frac{1}{2\pi i} \times \int \frac{X_0(-z, s)}{\Lambda_s(z)} \frac{1}{(z - \mu)(z - \mu_0)} e^{-sx/z}\, dz. \qquad (35)$$

The integration over z is carried out along a contour which starts and ends at $z = 0$, with Re $(z/s) \geq 0$ at $z \to 0$, and embraces the branch cut $0 < z < 1$, as well as the pole $z = \nu_0$ of the integrand.

We see, by the way, that the discrete term in (11) is due to the residue of the integrand in (35) at $z = \nu_0$, and that the expressions (12), (31) stem from the residue at $z = \mu$.

III. FINAL FORM OF THE SOLUTION

The above conclusions permit us to deform the integration path in (9) as shown by Fig. 2. Thereby and by the use of relation (29) the integral in (9) is put into a more convenient form,

$$\psi(x, \mu, t; \mu_0)$$

$$= \frac{2}{\pi}\int_0^1 (1 - s)\,|\nu_0(s)|\,\psi_s(0, -\mu_0)\psi_s(x, \mu)e^{-(1-s)t}\,ds + \lim_{\omega\to\infty}\frac{1}{2\pi i}\int_{-i\omega}^{i\omega}\psi_s(x, \mu; \mu_0)e^{-(1-s)t}\,ds. \quad (36)$$

This, with the expressions (16), (15), and (30) substituted, represents the final result. Expressions for $\rho(x, t; \mu_0)$ and $j(x, t; \mu_0)$ follow immediately.

For $x = 0$ a further simplification is possible because the expression (31) can be analytically continued to Re $(s) < 0$, which for $x > 0$ was impossible, because of the factor $e^{-sx/\nu}$ in the integrand in (30). Now the integration path can be bent still further to the left, and we end up with a closed loop encircling the branch cut. This means that the last term in (36) drops out for $x = 0$, so that

$$\psi(0, -\mu, t; \mu_0) = \frac{2}{\pi}\int_0^1 (1 - s)\,|\nu_0(s)| \times \psi_s(0, -\mu_0)\psi_s(0, -\mu)e^{-(1-s)t}\,ds, \quad \mu \geq 0. \quad (37)$$

The validity of the reciprocity relation (5) is clearly demonstrated. The values of $X(-\mu, s)$ involved in $\psi_s(0, -\mu)$ through Eq. (16) can be taken from graphs presented by Bowden.[4]

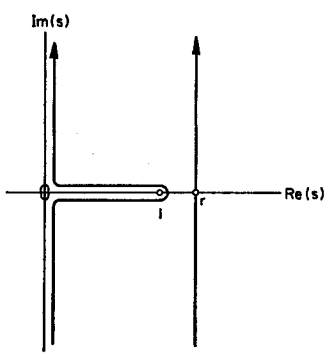


FIG. 2. Integration paths for inverse Laplace transformation of $\psi_s(x, \mu; \mu_0)$.

The neutron density and the net current at the surface of the medium are obtained from (37) by substituting the factor $\psi_s(0, -\mu)$ of the integrand by the expression (17) and by $j_s(0) = -1$, respectively, and by adding the contribution due to the incident neutrons, $\rho_{inc}(0, t; \mu_0) = \mu_0^{-1}\delta(t)$, $j_{inc}(0, t; \mu_0) = \delta(t)$. Especially simple formulas follow for the case of isotropic angular distribution of the incident neutrons:

$$\rho^*(0, t) = \delta(t) + t^{-1}e^{-t/2}I_1(\tfrac{1}{2}t) \quad (38a)$$

$$= \delta(t) + \frac{1}{4}\left[1 - \frac{1}{2}t + \frac{5}{32}t^2 - \cdots\right] \quad (38b)$$

$$= (\pi t^3)^{-\frac{1}{2}}\left[1 - \frac{3}{4}t^{-1} - \frac{15}{32}t^{-2} + O(t^{-3})\right], \quad (38c)$$

$$j^*(0, t) = \tfrac{1}{2}\delta(t) - \pi^{-1}\int_0^1 2(1 - s)\,|\nu_0(s)|\,e^{-(1-s)t}\,ds \quad (39a)$$

$$= \tfrac{1}{2}\delta(t) - \tfrac{1}{3}(1 - \ln 2) + O(t) \quad (39b)$$

$$= -(3\pi t^3)^{-\frac{1}{2}}[1 - (27/20)t^{-1} + O(t^{-2})]. \quad (39c)$$

The formula for $\rho^*(0, t)$ has been reduced to an expression containing the modified Bessel function I_1 by aid of the substitution $s = \frac{1}{2}(1 - \cos\theta)$, which leads to Poisson's integral representation for this function.[9]

The initial values of the reflected angular density could be computed from Eq. (37) by substituting $t = 0$. However, an easier way is to expand the previously mentioned closed-loop integration path into a very large circle, instead of shrinking it onto the branch cut. Observing that $X_0(z, s) = 1 + O(s^{-1})$ for $s \to \infty$, as can easily be shown, we obtain, using (31),

$$\psi(0, -\mu, 0; \mu_0) = [2(\mu + \mu_0)]^{-1}, \quad \mu \geq 0. \quad (40)$$

This angular density is entirely due to neutrons scattered only once, as one can infer directly from the transport equation.

IV. DISCUSSION

The above results closely resemble those obtained by Bowden for the slab problem. The main difference is that in the latter case two discrete terms, involving the factors $e^{\pm sx/\nu_0}$, enter a development analogous to (11). Therefore the function $\psi_s(x, \mu; \mu_0)$ for the slab needs no branch cut, but has instead a finite number of poles at certain "critical" values of s inside the interval $0 < s < 1$. The poles fill

[9] *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II.

up this interval more and more densely as the thickness of the slab is increased.

Thus in the slab case the integral over the branch cut in (36) is replaced by a sum over the residues. Actually, Eq. (36) can be deduced as a limit from Bowden's result [Ref. 4, Eq. (5.12)]. This is done by proving that the factor $(2/\pi)(1-s)\,|\nu_0(s)|$ of the first integrand in (36) is equal to the limit of the product of the pole number density and a normalization factor.

The individual terms of the mentioned sum, just as the integrand of the branch-cut term in (36), can be pictured by standing waves decaying at various rates, slower than e^{-t}. Each wave corresponds in the slab case to a solution of the critical problem, and in the present case to a solution of Milne's problem for a multiplying medium.

The last term in (36), when $\psi_s(x, \mu; \mu_0)$ is developed according to Eq. (30), represents a sum over a continuous family of traveling waves,[10] all decaying like e^{-t}, i.e., with a decay time equal to the mean time between collisions of a neutron. Only ingoing waves, with speeds ν ranging from 0 to 1, are present in the case of a semi-infinite medium, whereas waves propagating in both directions are included in the slab solution. As shown by Eq. (37) those waves do not contribute to the angular density of the neutrons reflected by a semi-infinite medium.

In view of the fast decay rate of the traveling waves we may say that their sum describes the transient effects mentioned in the introduction. Actually this sum contains the uncollided beam term $\delta(x-\mu_0 t)\delta(\mu-\mu_0)e^{-t}$, since the Laplace transform of this term, $\mu_0^{-1}\delta(\mu-\mu_0)e^{-sx/\mu_0}$, is contained in $\psi_s(x, \mu; \mu_0)$.

On the other hand, one expects the branch-cut term in (36) alone to describe the behavior of ψ for large values of t, so that this term represents an asymptotic approximation. Some simplification, consistent with this kind of approximation, can be achieved by using (28) and by substituting $\psi_s(0, -\mu_0) \approx \psi_1(0, -\mu_0)$, and, for small x only, $\psi_s(x, \mu) \approx \psi_1(x, \mu)$. An expression results, which contains the integral $\int_0^1 (1-s)^{\frac{1}{2}} e^{-(1-s)t}\,ds$. For $t \gg 1$ it is permissible to shift the lower limit of this integral to $-\infty$. Then, with (16), the expression simplifies to

$$\psi(x, \mu, t; \mu_0) \approx (\tfrac{4}{3}\pi t^3)^{-\frac{1}{2}}[X(-\mu_0, 1)]^{-1}\psi_1(x, \mu). \quad (41)$$

Approximations for ρ, j, and for $\psi(0, -\mu, t; \mu_0)$ follow in a simple way upon application of the formulas (17), and (16), with (28).

In a similar way, by substituting the asymptotic part of (15) into (36), we arrive at a different asymptotic approximation, valid for $t \gg 1$, $x \gg 1$. Let us write down only the expression for the neutron density, which follows from the asymptotic part of (27):

$$\rho(x, t; \mu_0) \approx 3(\tfrac{4}{3}\pi t^3)^{-\frac{1}{2}}[X(-\mu_0, 1)]^{-1}[x + q(1)] \times \exp\{-\tfrac{3}{4}t^{-1}[x + q(1)]^2\}. \quad (42)$$

Tables of $H(\mu, 1) = \sqrt{3}/X(-\mu, 1)$ and of $\rho_1(x)$, needed for the evaluation of $\psi(0, -\mu, t; \mu_0)$ and $\rho(x, t; \mu_0)$, according to the approximations (41) and (42), are available.[7,11]

Various refined asymptotic approximations could be conceived by making less crude substitutions for the functions involved in the exact expressions. For instance, we observe that the factor $(1-s)^{\frac{1}{2}}e^{-(1-s)t}$ of the first integrand in (36) is zero at $s = 1$. Hence it seems advisable to approximate the remaining (finite) factor $(1-s)^{\frac{1}{2}}\,|\nu_0(s)|\,\psi_s(0, -\mu_0)\psi_s(x, \mu)$ by its value at s slightly below 1, rather than at $s = 1$. We may require that this procedure should be correct if the latter factor were a linear function of s. We find then that for $t \gg 1$ the appropriate value of s is $1 - \frac{3}{2}t^{-1}$. This, with (15) and (16), has to be inserted into

$$\psi(x, \mu, t; \mu_0) \approx (\pi t^3)^{-\frac{1}{2}}(1-s)^{\frac{1}{2}}\,|\nu_0(s)|\,\psi_s(0, -\mu_0)\psi_s(x, \mu), \quad (43)$$

which is valid for $t \gg 1$, $x \ll t^{\frac{1}{2}}$, as one can show. The improvement of (43) over (41) can be judged from the fact that the first two terms in (38c) and (39c) follow from (43), whereas only the first term is obtained from (41).

It should be mentioned that the approximations (41)–(43) can be deduced also without knowing the exact result, solely by considerations based upon the diffusion equation and upon a reciprocity theorem. Such a derivation, though not rigorous, has the advantage of being amenable to generalizations to anisotropic scattering and to energy-dependent problems.

ACKNOWLEDGMENTS

We should like to thank Dr. G. J. Mitsis, Professor G. C. Summerfield, and Mr. M. R. Mendelson for helpful suggestions. One of us (I.K.) is indebted to the Department of Nuclear Engineering, The University of Michigan for hospitality, and to the Conference Board of Associated Research Councils for the award of a Fulbright–Hays grant.

[10] A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chain Reactors* (University of Chicago Press, Chicago, 1958), p. 235.

[11] C. Mark, Phys. Rev. **72**, 558 (1947).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

Asymptotic Expansions of Solutions of Differential Equations*

J. G. FIKIORIS†
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received 29 April 1964)

A generalization of Ford's method, concerning the asymptotic expansions of solutions of differential equations with polynomial coefficients and with three or more regular singular points and one irregular at infinity, is presented. The analysis is subsequently extended to the special case of integral values for the difference of exponents of the differential equation, thus providing the complete asymptotic expansion of the second, logarithmic solution of the equation. Explicit formulas for the evaluation of the constant coefficients of the expansions are given; each coefficient is expressed in terms of a single solution of the adjoint difference equation associated with the original differential equation. Differential equations possessing singularities in excess of the hypergeometric equation (3 regular) or its variations, appear as separated solutions of the wave equation in certain stratified media whose index of refraction is a continuous function of position.

1. INTRODUCTION

THE asymptotic expansion of a function $f(z)$ defined by a Maclaurin series has been the subject of numerous investigations. A comprehensive and original treatment is due to Ford[1-3] to whom the reader is also referred for a fairly complete list of references. Ford's method is based on the behavior of the coefficients a_n of the Maclaurin series for large values of n and can be applied to series with finite or infinite radius of convergence. In the latter case $f(z)$ is an entire function. The applicability of the method is restricted by the nature of the singularities of $f(z)$ in the complex z-plane.

An important application is concerned with the asymptotic expansions of solutions of linear ordinary differential equations with polynomial coefficients. The method is successful in the following two cases: (1) All singularities of the equation are regular. (2) Only one singularity (conveniently placed at infinity) is irregular with rank not exceeding 1, while all others, in the finite z-plane, are regular. In both cases the equation and its solutions possess a finite number of singularities one of which (regular) may be at the origin. The order of the equation is finite but unrestricted. The classical method of Frobenius applied around the origin expresses a solution $R(z)$ in terms of a Maclaurin series whose coefficients a_n obey a recurrence formula with a finite number of terms. Essentially $R(z)$ is defined by the recurrence formula. Based solely on this definition, that is without reference to the fact that $R(z)$ satisfies a certain differential equation, Ford's theory proceeds to obtain its complete asymptotic expansion. In an entirely independent manner it yields results in agreement with the theory of differential equations and, at the same time, provides explicit formulas for the evaluation of the constant coefficients that linearly connect $R(z)$ with the asymptotic series in the expansion. Only in special cases, admitting a straightforward application of an integral transformation, is the theory of differential equations able to determine these coefficients.

As a general outline, the method proceeds through the following steps: Transformation of the recurrence formula into a difference equation with polynomial coefficients; complete solution of the latter; detailed examination of the analytic properties of the solutions and of their behavior for large values of the argument in the right half-plane. Based on this information the asymptotic expansion of $R(z)$ is obtained with the use of certain theorems contained in Ref. 2 by Ford. In the last chapter of his book[2] Ford applies his method to second-order equations and obtains explicit results for the two independent solutions R_1 and R_2 around the origin. The constant coefficients A_{13}, etc., in the asymptotic expansions

$$R_1(z) \sim A_{13}R_3(z) + A_{14}R_4(z), \qquad (1.1)$$

$$R_2(z) \sim A_{23}R_3(z) + A_{24}R_4(z), \qquad (1.2)$$

are determined in terms of the solutions of the difference equations obtained from the recurrence formulas defining R_1 and R_2. Each coefficient depends on all the independent solutions of the cor-

* Supported by NSF Grant 20225 and Contract Nonr-1866(32).
† Present address: Research and Advanced Development Division, Avco Corporation, Wilmington, Massachusetts.
[1] J. G. Fikioris, Tech. Rept. No. 395, Cruft Laboratory, Harvard University (1963). Also Part II of Ph.D. thesis in Applied Physics, Harvard University (1963).
[2] W. B. Ford, *The Asymptotic Developments of Functions Defined by Maclaurin Series* (Chelsea Publishing Company, New York, 1960).
[3] W. B. Ford, *Studies on Divergent Series and Summability* (Chelsea Publishing Company, New York, 1960).

responding difference equation. As long as the difference of exponents $\sigma_{12} = \sigma_1 - \sigma_2$ ($\geqq 0$) at the singular point $z = 0$ differs from 0 or a positive integer, the process is applicable to both R_1 and R_2. If σ_{12} is equal to 0 or a positive integer the method, without alteration, applies to the first solution R_1; the second solution R_2 becomes logarithmic, and this case, to the author's best knowledge, has not been treated up to the present time. Ford refers to it as a subject for further research.

In this paper, Ford's method is slightly modified and subsequently extended to the second, logarithmic solution of the differential equation. In addition to the associated difference equation its adjoint is also introduced; $A_{13}(A_{14})$ can then be expressed in terms of a single solution of the latter, say $N_3(y)[N_4(y)]$, which one proceeds to determine explicitly without bothering about the rest. A corresponding statement can be made about A_{23}, A_{24} when σ_{12} differs from 0 or an integer. Notice also that the order of the difference equations (and, therefore, the number of their independent solutions) is independent of the order of the differential equation and, in general, higher than 2. The main advantage of this more compact formulation, however, is the fact that it can be extended to the special but important case of the second, logarithmic solution R_2 of the differential equation, whenever $\sigma_{12} = 0$ or a positive integer. $A_{23}(A_{24})$ is now determined in terms of $N_3(y)$, $N_3'(y)$ (N_4, N_4') the same particular solution of the adjoint difference equation that expresses $A_{13}(A_{14})$. The final formulas for all A's are especially adapted to desk or machine computations. They possess a so to speak "inherent-checking" property, which, at the same time, reveals the accuracy of the computations. This statement, and certain previous ones, will become clear in the course of the analysis and after the explicit development of the final formulas.

Differential equations with polynomial coefficients possessing singularities in excess of the hypergeometric equation (3 regular) or its variations, appear as separated solutions of the wave equation in certain stratified media whose index of refraction is a continuous function of position. In particular, the radial equation for radially stratified media in spherical or cylindrical coordinates, where both the scalar and vector wave equations are still separable, may be put under the above form for a wide variety of "stratification functions." Outside a finite sphere (or cylinder), the radial solution that satisfies the radiation condition corresponds to integral values of the separation constant. Satisfaction of boundary conditions for finite values of the radius and "near-field" computations require the analytic continuation of this solution in the vicinity of the origin, in the form

$$R_4(z) = A_{41}R_1(z) + A_{42}R_2(z). \tag{1.3}$$

In this case $\sigma_{12} = 0, 1, 2, \cdots$ and $R_2(z)$ becomes logarithmic. Thus, in an indirect way, $R_2(z)$ appears in a physical problem and the necessity of obtaining its complete asymptotic expansion (1.2) arises.

2. SERIES SOLUTIONS OF THE DIFFERENTIAL EQUATION

Convergent Series Solutions Around $x = 0$

Without loss of generality the following specific differential equation is considered

$$\frac{d^2R}{dx^2} + \frac{a-b}{(x+a)(x+b)}\frac{dR}{dx} + \left[1 + \frac{a-b}{x+b} - \frac{\nu(\nu+1)}{x^2}\right]R(x) = 0. \tag{2.1}$$

a, b are constant parameters, in general complex, while ν, the separation constant, is real and positive. The above form is a special case of the radial equation

$$R''(x) - \frac{f'(x)}{f(x)}R'(x) + \left[f(x) - \frac{\nu(\nu+1)}{x^2}\right]R(x) = 0 \tag{2.2}$$

for transverse magnetic waves in spherically stratified media[4,5]; $x = k_0 r$ is the radial distance in electrical units and $f(x)$ the stratification function expressing the dependence of the index of refraction on the radius. For the simple form

$$f(x) = \frac{x+a}{x+b} = 1 + \frac{c}{x+b}\,; \quad c = a - b, \tag{2.3}$$

Eq. (2.2) reduces to (2.1). [The explicit results obtained in this paper have been used to calculate the behavior of a biconical antenna immersed in a radially stratified medium characterized by (2.3). This application is published elsewhere.[6]]

Equation (2.1) has three regular singular points at $x = 0$, $x = -a$, $x = -b$ and an irregular singularity of the first rank at $x = \infty$ (Ref. 7, pp. 417–428; Ref. 8, pp. 58–77). Since it contains only three

[4] C. T. Tai, Appl. Sci. Res. Sec. B 7, 113 (1958).
[5] C. Yeh and Z. A. Kaprelian, University of Southern California Techn. Rept., Elect. Eng. Dept. (November 1960).
[6] J. G. Fikioris, Tech. Rept. No. 390, Cruft Laboratory, Harvard University (1963). Also Part I of Ph.D. thesis in Applied Physics, Harvard University (1963) and IEEE Trans. Ant. and Prop., AP-13, 289 (March 1965).
[7] E. L. Ince, *Ordinary Differential Equations* (Dover Publications, Inc., New York, 1956).
[8] A. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956).

independent parameters (a, b, ν) it is not the most general type of second-order equation with the same singularities. It was chosen because it corresponds to a certain physical problem and because the differences between this and the general case are trivial. Ford has shown that the analysis depends mainly on the nature and number of singularities of the equation,[2] the number of independent parameters it contains being unimportant in this respect. Variations of the above form, e.g., five regular singular points, can be treated in essentially the same way. The particular choice of (2.1) does not restrict in the least the representative character of the case under consideration.

Following the classical method of Frobenius first write (2.1) in the form:

$$\begin{aligned}&[x^2 + (a + b)x + ab]x^2R''(x) + cx^2R'(x)\\&\quad + [x^4 + 2ax^3 + (a^2 - \nu^2 - \nu)x^2\\&\quad - \nu(\nu + 1)(a + b)x - \nu(\nu + 1)ab]R(x)\\&\quad = T_0(x)x^2R''(x) + T_1(x)xR'(x) + T_2(x)R(x) = 0.\end{aligned} \tag{2.4}$$

Substituting the formal series

$$R(x) = \sum_{n=0}^{\infty} a_n x^{n+\sigma}$$

a five-term recurrence formula is obtained for the coefficients a_n:

$$\sum_{m=0}^{4} a_{n-m} f_m(n + \sigma - m) = 0;$$

$$a_{-j} = 0, \qquad j = 1, 2, 3, \cdots, \tag{2.5}$$

where

$$\begin{aligned}f_0(z) &= ab[z(z - 1) - \nu(\nu + 1)]\\&= ab(z + \nu)(z - \nu - 1),\\f_1(z) &= (a + b)z(z - 1) + cz\\&\quad - (a + b)\nu(\nu + 1),\\f_2(z) &= z(z - 1) + a^2 - \nu(\nu + 1),\\f_3(z) &= 2a,\\f_4(z) &= 1.\end{aligned} \tag{2.6}$$

Indicial equation: $f_0(\sigma) = ab(\sigma + \nu)(\sigma - \nu - 1) = 0$ with roots

$$\sigma_1 = \nu + 1, \quad \sigma_2 = -\nu, \quad \sigma_{12} = \sigma_1 - \sigma_2 = 2\nu + 1. \tag{2.7}$$

If $2\nu + 1$ differs from an integer two independent particular solutions of (2.1) are defined as follows:

$$R_1(x) = x^{\nu+1}\left[1 + \sum_{n=1}^{\infty} a_n x^n\right], \quad |x| < \min(|a|, |b|), \tag{2.8}$$

$$R_2(x) = x^{-\nu}\left[1 + \sum_{n=1}^{\infty} b_n x^n\right], \quad |x| < \min(|a|, |b|), \tag{2.9}$$

where the a_n's are obtained from (2.5) with $\sigma = \sigma_1 = \nu + 1$ and $a_0 = 1$, while for the b_n's (2.5) is used again with $\sigma = \sigma_2 = -\nu$ and $b_0 = 1$. Notice that $R_1(0) = 0$, $R_2(0) = \infty$; $x = 0$ is a branch point for both solutions.

If $2\nu + 1$ is equal to a positive integer or zero the first solution $R_1(x)$ is again defined by (2.8), (2.5). The second independent solution $R_2(x)$, however, has a logarithmic singularity at $x = 0$ and is no longer defined by (2.9) (Ref. 9, pp. 200–201). In this case put

$$R_2(x) = \ln x R_1(x) + S(x); \; S(x) = \sum_{n=0}^{\infty} b_n x^{n-\nu}, \tag{2.10}$$

where $R_1(x) = \sum_{m=0}^{\infty} a_m x^{m+\nu+1}$, $a_0 = 1$, is the first solution defined by (2.8), (2.5). Substitute into (2.4) noticing that $R_1(x)$ satisfies (2.4):

$$\begin{aligned}&T_0(x)x^2S'' + T_1(x)xS' + T_2(x)S\\&\qquad = T_0(x)[R_1 - 2xR_1'] - T_1(x)R_1.\end{aligned} \tag{2.11}$$

Use the series expressions for S, R_1 and multiply by x^ν:

$$\sum_{n=0}^{\infty} b_n x^n f(x, n - \nu) = x^{2\nu+1} \sum_{m=0}^{\infty} a_m x^m F(x, m + \nu + 1), \tag{2.12}$$

where

$$f(x, z) = \sum_{i=0}^{4} f_i(z)x^i, \quad f_i(z) \text{ as in } (2.6), \tag{2.13}$$

$$F(x, z) = \sum_{i=0}^{2} F_i(z)x^i, \tag{2.14}$$

$$\begin{aligned}F_0(z) &= -ab(2z - 1),\\F_1(z) &= -(a + b)(2z - 1) - c\\&= -2(a + b)z + 2b,\\F_2(z) &= -2z + 1.\end{aligned} \tag{2.15}$$

Since $f(x, n - \nu)$ and $F(x, m + \nu + 1)$ are polynomials in x, it is obvious, incidentally, that Eq. (2.12) cannot be satisfied unless $2\nu + 1$ is equal to a positive integer. Now put $n = m + 2\nu + 1$

[9] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, England, 1946), American ed.

in its right-hand side:

$$\sum_{n=0}^{\infty} b_n x^n f(x, n-\nu) = \sum_{n=2\nu+1}^{\infty} a_{n-2\nu-1} x^n F(x, n-\nu). \tag{2.16}$$

Since $a_{-1} = a_{-2} = \cdots = 0$, the second summation can be started from $n = 0$. The lowest exponent of x is $n = 0$ with coefficient $b_0 f_0(0 - \nu) = 0$. For $1 \leqq n \leqq 2\nu$,

$$\sum_{m=0}^{4} f_m(n - \nu - m)\, d_{n-m} = 0,$$

$$d_{-j} = 0, \quad j = 1, 2, 3, \cdots, \tag{2.17}$$

$$d_n = b_n/b_0, \quad d_0 = 1, \quad n = 0, 1, 2, \cdots. \tag{2.18}$$

Since $f_0(n - \nu)$ differs from 0 for all these values of n, the recurrence formula can be useds uccessively to yield $d_1, d_2, \cdots, d_{2\nu}$. For $n = 2\nu + 1$, however, $f_0(\nu + 1) = 0$; the corresponding equation can only be satisfied by choosing conveniently the value of b_0, left undetermined so far, i.e.,

$$b_0 = -ab(2\nu + 1)[d_{2\nu} f_1(\nu) + d_{2\nu-1} f_2(\nu - 1) + d_{2\nu-2} f_3(\nu - 2) + d_{2\nu-3} f_4(\nu - 3)]^{-1}. \tag{2.19}$$

$b_1, b_2, \cdots, b_{2\nu}$ may now be evaluated using $b_n = b_0 d_n$.

For $n > 2\nu + 1$, Eq. (2.16) yields

$$\sum_{m=0}^{4} f_m(n - \nu - m) b_{n-m} = a_{n-2\nu-1} F_0(n - \nu) + a_{n-2\nu-2} F_1(n - \nu - 1) + a_{n-2\nu-3} F_2(n - \nu - 2). \tag{2.20}$$

This equation may be used for all integral values of n, including (2.17) as a special case, since $a_0 = 1$, $a_{-1} = a_{-2} = \cdots = 0$. The initial conditions

$$b_{-3} = b_{-2} = b_{-1} = 0, \qquad b_0 \text{ as in (2.19)} \tag{2.21}$$

are sufficient to satisfy all requirements for the coefficients b_n of $S(x)$. Thus, for $n = 0$, Eq. (2.20) reduces to $b_{-4} = 0$ and by repeated applications it yields $b_{-5} = b_{-6} = \cdots = 0$, as required.

The process so far has left $b_{2\nu+1}$ undetermined. Inspection of (2.20) shows, however, that in addition to the set $(b_n; n = 0, 1, 2, \cdots)$ it can also be satisfied by the set $(b_n + k a_{n-2\nu-1})$, where k is a constant, the reason being that the set $(a_{n-2\nu-1})$ satisfies its homogeneous part. In fact for

$$n = 2\nu + 1 + m, \qquad b_n = b_{2\nu+1+m} = B_m, \tag{2.22}$$

Eq. (2.20) becomes

$$\sum_{j=0}^{4} B_{m-j} f_j(m + \nu + 1 - j) = a_m F_0(m + \nu + 1) + a_{m-1} F_1(m + \nu) + a_{m-2} F_2(m + \nu - 1). \tag{2.23}$$

Its homogeneous part is satisfied by the set $a_0, a_1, \cdots$ as (2.5) with $\sigma = \sigma_1 = \nu + 1$ reveals.

$b_{2\nu+1}$ can be chosen arbitrarily. In the process of using (2.20) for the b_n's the simplest choice is

$$b_{2\nu+1} = B_0 = 0. \tag{2.24}$$

The remaining b_n's are evaluated using (2.20), (2.21). A nonvanishing value for $b_{2\nu+1} = B_0$ means simply, that $S(x)$, and consequently $R_2(x)$, contain the additive solution $b_{2\nu+1} R_1(x)$, which can be discarded in the definition of the second independent solution $R_2(x)$. Without ambiguity the following definition is adopted

$$R_2(x) = \ln x R_1(x) + x^{-\nu} \sum_{n=0}^{\infty} b_n x^n; \quad |x| < \min(|a|, |b|);$$

$$b_{2\nu+1} = 0, \qquad b_0 \text{ as in (2.19)}. \tag{2.25}$$

The power series representations that have been obtained for the two independent solutions $R_1(x)$ and $R_2(x)$ are valid only within the circle $|x| < \min(|a|, |b|)$, where they both converge uniformly and absolutely. Their analytic continuation outside this circle, in fact into whole regions of the x-plane where x varies from 0 to ∞, can be obtained with the use of a bilinear transformation of the independent variable in the form

$$t = x/(x + p), \qquad x = pt/(1 - t). \tag{2.26}$$

The constant p is conveniently chosen in each case so as to optimize the convergence of the resulting series in t, or to map the half-plane of interest, wherein x varies from 0 to ∞, within the unit circle $t = 1$, placing $x = -a$, $x = -b$ outside it. $x = 0$, $-a$, $-b$, ∞ are mapped at $t = 0, a/(a-p), b/(b-p), 1$, respectively,[1] which now appear as the only singular points of the resulting differential equation in t, preserving their nature and rank. This is a well-known property of bilinear transformations (Ref. 7, p. 437). Power-series solutions $R_1(t)$, $R_2(t)$ around $t = 0$ ($x = 0$) can be obtained in a similar way and can be easily connected to the previously defined $R_1(x)$, $R_2(x)$ by simple comparison of behavior in the vicinity of the origin.[1] The former series converge within $|t| = 1$, for example, and provide the analytic continuation of (2.8) and (2.25) into certain regions of the x-plane outside $|x| = \min(|a|, |b|)$. Numerical computations have shown that the series in t can be used for values of $|x|$ three to four times larger than the ones possible in connection with (2.8), (2.25). In certain cases they proved capable of carrying the computations into the region of validity of the asymptotic expansions

of $R_1(x)$, $R_2(x)$ and provided a check on the formulas. Since $R_1(t)$, $R_2(t)$ have no connection with the specific aim of this analysis no explicit development or further mention will be made of them.

Asymptotic Solutions $R_3(x)$, $R_4(x)$ Around $x = \infty$

As $x \to \infty$ the coefficients of R' and R in Eq. (2.1) vary as

$$\frac{c}{(x+a)(x+b)} \sim O(x^{-2}),$$

$$1 + \frac{c}{x+b} - \frac{\nu(\nu+1)}{x^2} \sim 1 + O(x^{-1}).$$

Point $x = \infty$ is an irregular singularity of rank 1 (Ref. 8, pp. 58–77). Formal descending series solutions around $x = \infty$ can be obtained by well known procedures in the theory of differential equations[7,8] (based essentially on the method of undetermined coefficients). For example the method explained in Ref. 8 (pp. 58–77) yields[1]

$$R_{3,4}(x) \underset{|x|\to\infty}{\sim} \left(\frac{x+a}{x+b}\right)^{\frac{1}{2}} e^{\omega x} x^{-\rho}\left[1 + \sum_{n=1}^{\infty} h_n x^{-n}\right], \quad (2.27)$$

where $\omega_3 = i$, $\rho_3 = -ic/2$ are used for R_3 and $\omega_4 = -i$, $\rho_4 = ic/2$ for R_4. The expansion coefficients h_n are determined from the recurrence formula

$$2\omega n h_n = \left(n + \frac{c}{2\omega}\right)\left(n + \frac{c}{2\omega} - 1\right)h_{n-1} + \sum_{m=2}^{n+1} q_m h_{n+1-m};$$

$$h_0 = 1, \qquad n = 1, 2, \cdots, \quad (2.28)$$

where

$$q_0 = 1, \quad q_1 = c, \quad q_2 = -\nu(\nu+1) - cb, \quad (2.29)$$

$$q_n = (-1)^n\left[b^{n-1}\left(\frac{n-1}{4b} - \frac{1}{2c} - c\right) + a^{n-1}\left(\frac{1}{2c} - \frac{3(n-1)}{4a}\right)\right], \qquad n = 3, 4, \cdots.$$

The series appearing in the above formal representations (2.27) are normal asymptotic series in the precise sense of Poincaré's definition (Ref. 7, pp. 168–174, 444–445, Ref. 8, pp. 69–72). If $[(x+a)/(x+b)]^{\frac{1}{2}}$ is expanded for $|x| \to \infty$ by the binomial theorem, the alternative expressions

$$R_{3,4} \sim e^{\omega x} x^{-\rho}\left(1 + \left\{\frac{c}{2} + \left[\left(\frac{c}{2\omega}\right)\left(\frac{c}{2\omega} + 1\right) - \nu(\nu+1) - cb\right](2\omega)^{-1}\right\}\frac{1}{x} + \frac{g_2}{x^2} + \frac{g_3}{x^3} + \cdots\right) \quad (2.30)$$

are obtained, which will serve for comparison later.

For real a, b: $\rho_3 = \bar{\rho}_4$, where the bar signifies the complex conjugate, while always $\omega_3 = \bar{\omega}_4$. Since from (2.29) all q_n are real the recurrence formula (2.28) yields as coefficients h_n for R_3 the complex conjugates of the coefficients of R_4. The same is obviously true for the coefficients g_n in (2.30). This means

$$R_3(x) = \bar{R}_4(\bar{x}); \qquad a, b \text{ real}, \quad (2.31)$$

and for real x

$$R_3(x) = \bar{R}_4(x). \quad (2.32)$$

All the above expressions are valid irrespectively of the value of ν. The critical line, or Stokes line, is given by Re $(\omega x) = 0$ (Ref. 8, p. 72). With $\omega = \pm i$ it turns out to be the real axis: Im $x = 0$. Any solution $R(x)$ of Eq. (2.1) can be expressed asymptotically as a linear combination of the above formal solutions

$$R(x) \underset{x\to\infty}{\sim} A_3 R_3(x) + A_4 R_4(x); \quad (2.33)$$

this expression holds uniformly in each of the upper and lower half x-planes separated by the Stokes line Im $x = 0$. The coefficients A_3, A_4 may, or course, change values from one plane to another for the same solution $R(x)$ (Ref. 8, p. 73). A similar combination expresses $R(x)$ along the real axis.

The main purpose of the analysis that follows is to find explicitly expressions like (2.33) for the particular solutions $R_1(x)$, $R_2(x)$ defined in the preceding section, in both halves of the x plane and along the real axis.

3. ASYMPTOTIC EXPANSIONS OF $R_1(z)$, $R_2(z)$ FOR LARGE $|z|$

Solution of the Associated Difference Equation

As outlined in the Introduction, the recurrence formula (2.5) is first transformed into a difference equation. Replace the index n by the continuous variable x and a_n by the general function $u(x)$, such that

$$u(n) = a_n, \qquad n \text{ being an integer}, \quad (3.1)$$

to obtain

$$\sum_{m=0}^{4} f_m(x + \sigma - m)u(x - m) = 0.$$

Advance x to $x + 4$, i.e., write $x + 4$ for x:

$$\sum_{m=0}^{4} f_m(x + \sigma + 4 - m)u(x + 4 - m) = 0. \quad (3.2)$$

At this point Ford's procedure is slightly modified by substituting

$$x + \sigma = y, \quad (3.3)$$

$$u(x) = v(x + \sigma) = v(y). \quad (3.4)$$

The equation becomes

$$p_4(y)v(y + 4) + p_3(y)v(y + 3) + p_2(y)v(y + 2) + p_1(y)v(y + 1) + p_0(y)v(y) = 0, \quad (3.5)$$

where

$$\begin{aligned} p_4(y) &= f_0(y + 4) \\ &= ab[(y + 4)(y + 5) - 2(y + 4) - \nu(\nu + 1)], \\ p_3(y) &= f_1(y + 3) \\ &= (a + b)(y + 3)(y + 4) \\ &\quad - (a + 3b)(y + 3) - (a + b)\nu(\nu + 1), \quad (3.6) \\ p_2(y) &= f_2(y + 2) \\ &= (y + 2)(y + 3) - 2(y + 2) \\ &\quad + a^2 - \nu(\nu + 1), \\ p_1(y) &= f_3(y + 1) = 2a, \\ p_0(y) &= f_4(y) = 1. \end{aligned}$$

These equations express the $p(y)$'s in forms suitable for the solution of Eq. (3.5) by the method of Laplace's transformation (Ref. 2, Chap. VIII, Ref. 10, pp. 478–501, Ref. 11, pp. 57–88). Being a linear difference equation of the fourth order with polynomial coefficients, Eq. (3.5) possesses four independent solutions. For the general theory of these equations and for certain of its results, of which continuous use will be made in what follows, the following references are cited: Ref. 11 (Chaps. I, III), Ref. 10 (Chaps. XII, XV), Ref. 2 (Chap. VIII). The aim is to find that particular analytic solution $\bar{v}(y)$ of (3.5), which satisfies condition (3.1), or, in terms of y and $\bar{v}(y)$, the conditions

$$\bar{v}(n + \nu + 1) = a_n \quad \text{for all integers } n \text{ and for } R_1(z), \quad (3.7)$$

$$\bar{v}(n - \nu) = b_n \quad \text{for all integers } n,\ 2\nu + 1 \neq \text{integer}, \quad \text{and for } R_2(z). \quad (3.8)$$

The method of Laplace's transformation is applied by assuming a solution in the form

$$v(y) = \frac{1}{2\pi i}\int_l t^{y-1}\psi(t)\,dt, \quad (3.9)$$

where l is a line of integration in the complex t-plane, suitably determined later, and $\psi(t)$ is found from a certain differential equation, as will be seen in the following. Substitution in (3.5), integration by parts, and use of Eqs. (3.6) for the $p(y)$'s yield

$$\int_l t^{y-1}[t^2\phi_2(t)\psi''(t) - t\phi_1(t)\psi'(t) + \phi_0(t)\psi(t)]\,dt + [I(\psi, t)]_l = 0,$$

where

$$\phi_2(t) = abt^2(t + a^{-1})(t + b^{-1}), \quad (3.10)$$

$$\phi_1(t) = -2abt^2(t + a^{-1})(t + b^{-1}) + ct^3, \quad (3.11)$$

$$\phi_0(t) = -\nu(\nu + 1)abt^2(t + a^{-1})(t + b^{-1}) + a^2(t + a^{-1})^2, \quad (3.12)$$

$$I(\psi, t) = abt^{y+2}(t + b^{-1}) \times \{\psi(t)[(y + 1)(t + a^{-1}) + 2t] - \psi'(t)t(t + a^{-1})\}. \quad (3.13)$$

The conclusion is that (3.9) provides a solution of (3.5) if $\psi(t)$ is a solution of the differential equation

$$abt^4(t + a^{-1})(t + b^{-1})\psi''(t) - t\phi_1(t)\psi'(t) + \phi_0(t)\psi(t) = 0, \quad (3.14)$$

and the path of integration l is chosen so that $I(\psi, t)$ has the same value at both extremities of the path when it is open, or so that $I(\psi, t)$ returns to the same value if l is closed and t returns to the same point after describing it.

The choice of l is based on the behavior of the solutions of (3.14) at the vicinity of its singularities. The equation has three regular singular points at $t_1 = -1/a$, $t_2 = -1/b$, $t = \infty$ and an irregular at $t = 0$. Following standard methods (Ref. 7, pp. 168–174; Ref. 8, pp. 69–72) two normal solutions around $t = 0$ are obtained

$$\psi_I(t) \sim e^{i/t}t^{-ic/2}(1 + g_1t + g_2t^2 + \cdots), \quad (3.15)$$

$$\psi_{II}(t) \sim e^{-i/t}t^{ic/2}(1 + g_1't + g_2't^2 + \cdots). \quad (3.16)$$

[10] L. M. Milne-Thomson, *The Calculus of Finite Differences* (MacMillan and Company Ltd., London, 1951).

[11] N. -E. Nörlund, *Leçons sur les équations linéaires aux différences finies* (Gauthier-Villars, Paris, 1929).

Both expressions are asymptotic as $t \to 0$. The theory also assures that ψ_I, ψ_{II} are twice differentiable and that the asymptotic expansions for ψ_I', ψ_{II}' (even for ψ_I'', ψ_{II}'') may be obtained by formal differentiation of (3.15) and (3.16), respectively. Since any solution $\psi(t)$ of (3.14) is a linear combination of ψ_I and ψ_{II}, expressions (3.15) and (3.16) in connection with (3.13) show that

$$I[\psi(t), t]|_{t=0} = 0, \quad \text{if} \quad \operatorname{Re} y > -1 + |\operatorname{Im} c|/2, \tag{3.17}$$

and as long as t approaches 0 along the real t axis. This statement serves to fix the path of integration l in (3.9), thus making the integral expression a solution of the difference equation (3.5).

The roots of the indicial equations for the regular singular points $t_1 = -a^{-1}$ and $t_2 = -b^{-1}$ are 0, 2 and 0, 0, respectively. There exist one solution $\psi_1(t)$ of Eq. (3.14) around $t_1 = -a^{-1}$ and another $\psi_2(t)$ around $t_2 = -b^{-1}$, which, in the neighborhood of the corresponding point and up to the nearest singularity can be expressed by the following convergent series:

$$\psi_1(t) = (t + a^{-1})^2[1 + p_1(t + a^{-1}) + p_2(t + a^{-1})^2 + \cdots], \tag{3.18}$$

$$\psi_2(t) = 1 + q_1(t + b^{-1}) + q_2(t + b^{-1})^2 + \cdots. \tag{3.19}$$

Using (3.13) it may be verified that $I(\psi_1, t)|_{t=-a^{-1}} = I(\psi_2, t)|_{t=-b^{-1}} = 0$. In combination with (3.17) these results show that there exist two corresponding solutions of (3.5) in the form

$$v_1(y) = \frac{1}{2\pi i}\int_{l_1} t^{y-1}\psi_1(t)\,dt, \tag{3.20}$$

$$v_2(y) = \frac{1}{2\pi i}\int_{l_2} t^{y-1}\psi_2(t)\,dt, \tag{3.21}$$

where the paths l_1 and l_2 are shown in Fig. 1. The dotted lines represent the branch lines of $\psi_1(t)$ and $\psi_2(t)$. The paths may be deformed as long as they do not cross branch lines and approach $t = 0$ along the real t axis. The so defined solutions $v_1(y)$, $v_2(y)$

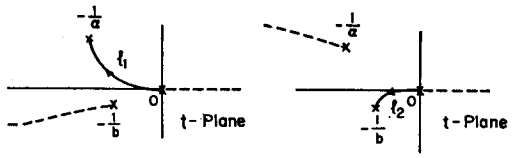


FIG. 1. Integration paths l_1, l_2 for $v_1(y)$, $v_2(y)$.

are analytic for (Ref. 2, Chap. VIII, Ref. 10, pp. 478–501, Ref. 11, Chap. III),

$$\operatorname{Re} y > -1 + |\operatorname{Im} c|/2. \tag{3.22}$$

It is shown later that they are also independent.

The third and fourth independent solutions of (3.5) may be obtained by putting (Ref. 2, Chap. VIII)

$$v(y) = w(y)/\Gamma(y). \tag{3.23}$$

Substitution in (3.5) and multiplication by $\Gamma(y + 4)$ yields a fourth-order linear difference equation for $w(y)$ with polynomial coefficients. The application of Laplace's transformation

$$w(y) = \frac{1}{2\pi i}\int_l t^{y-1}\psi(t)\,dt$$

shows that $\psi(t)$ must satisfy the differential equation

$$(t^2 + 1)t^4\psi^{IV}(t) - \phi_3(t)t^3\psi'''(t) + \phi_2(t)t^2 \times \psi''(t) - \phi_1(t)t\psi'(t) + \phi_0(t)\psi(t) = 0, \tag{3.24}$$

where

$$\begin{aligned}
\phi_3(t) &= (a + b)t^3 - 6t^2 + 2at,\\
\phi_2(t) &= -2abt^4 - (3a + 5b)t^3 + [a^2 + 6 - \nu(\nu + 1)]t^2,\\
\phi_1(t) &= -2abt^4 + [a + 3b - (a + b)\nu(\nu + 1)]t^3,\\
\phi_0(t) &= -ab\nu(\nu + 1)t^4.
\end{aligned} \tag{3.25}$$

In this case also (Ref. 10, pp. 478–501, Ref. 11, Chap. III):

$$\begin{aligned}
I(\psi, t) = {} & \psi(t)\sum_{m=0}^{3}\frac{d^m}{dt^m}[t^{\nu+m}\phi_{m+1}(t)]\\
& - \psi'(t)\sum_{m=0}^{2}\frac{d^m}{dt^m}[t^{\nu+m+1}\phi_{m+2}(t)]\\
& + \psi''(t)\sum_{m=0}^{1}\frac{d^m}{dt^m}[t^{\nu+m+2}\phi_{m+3}(t)]\\
& - \psi'''(t)t^{\nu+3}\phi_4(t).
\end{aligned} \tag{3.26}$$

Equation (3.24) has three regular singular points at $t = 0, t_3 = i, t_4 = -i$, and an irregular singularity at $t = \infty$. The roots of the indicial equation around $t = 0$ are $\rho = 0, 1, 2, 3$. Four independent solutions of (3.24) correspond to these roots and any solution of it can be expressed as a linear combination of them. These solutions have the general form

$$\psi_s(t) = t^{\rho_s}[f_0(t) + f_1(t)\ln t + \cdots + f_m(t)(\ln t)^m],$$

where $s = 1, 2, 3, 4$; $m \leqq 3$, and $f_0(t), \cdots, f_m(t)$ are analytic at $t = 0$. Thus, if $\operatorname{Re} y > 0$, $t^y\psi_s(t)$

vanishes at $t = 0$ for all $\psi_s(t)$. Reference to (3.26) shows that

$$\text{if} \quad \operatorname{Re} y > 0, \quad \text{then} \quad I(\psi, t)\,|_{t=0} = 0 \tag{3.27}$$

for any solution $\psi(t)$ of (3.24). Under the same condition ($\operatorname{Re} y > 0$) $[I(\psi, t)]_l = 0$, where the path starts and ends at $t = 0$.

The roots of the indicial equations around $t_3 = i$ and $t_4 = -i$ are 0, 1, 2, $\beta_3 = -ic/2$, and 0, 1, 2, $\beta_4 = ic/2$, respectively. As before two solutions $\psi_3(t)$ and $\psi_4(t)$ of (3.24) exist, the first defined around $t = i$, the second around $t = -i$, which can be expressed by the convergent series

$$\psi_3(t) = (t - i)^{-ic/2}[1 + h_1(t - i) + h_2(t - i)^2 + \cdots], \quad |t - i| < 1, \tag{3.28}$$

$$\psi_4(t) = (t + i)^{ic/2}[1 + j_1(t + i) + j_2(t + i)^2 + \cdots], \quad |t + i| < 1. \tag{3.29}$$

The remaining two independent solutions of (3.5) can now be defined by

$$v_3(y) = \frac{1}{2\pi i \Gamma(y)} \int_{l_3} t^{y-1}\psi_3(t)\,dt, \tag{3.30}$$

$$v_4(y) = \frac{1}{2\pi i \Gamma(y)} \int_{l_4} t^{y-1}\psi_4(t)\,dt, \tag{3.31}$$

where l_3 and l_4 are shown in Fig. 2; they are analytic for

$$\operatorname{Re} y > 0. \tag{3.32}$$

Four independent solutions of the fourth-order linear difference equation (3.5) have been found: v_1 and v_2 defined and analytic for $\operatorname{Re} y > -1 + |\operatorname{Im} c|/2$; v_3 and v_4 defined and analytic for $\operatorname{Re} y > 0$. For $\operatorname{Re} y < -1 + |\operatorname{Im} c|/2$ and $\operatorname{Re} y < 0$ the analytic continuation of v_1, v_2 and v_3, v_4, respectively, is defined by the difference equation (3.5), i.e.,

$$v_s(y) = -[p_4(y)v_s(y + 4) + p_3(y)v_s(y + 3) + p_2(y)v_s(y + 2) + 2av_s(y + 1)]; \quad s = 1, 2, 3, 4. \tag{3.33}$$

This definition makes all four solutions analytic in the whole y plane.

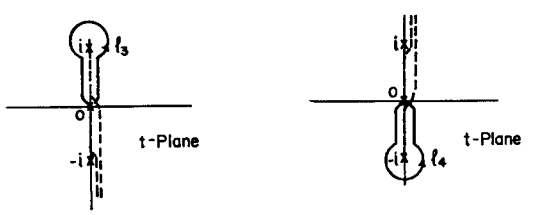


FIG. 2. Integration paths l_3, l_4 for $v_3(y)$, $v_4(y)$.

As a next step these functions are expanded in inverse factorial series and their behavior is investigated as $y \to \infty$ in the right-half-plane, from now on indicated by $y \to |\ \infty$. It is first observed that for $\operatorname{Re} y > y_s$ ($y_s = -1 + |\operatorname{Im} c|/2$ for $s = 1, 2$ and 0 for $s = 3, 4$), v_1, v_2, w_3, w_4 can be expressed in the general form

$$u_s(y) = \frac{1}{2\pi i} \int_{l_s} t^{y-1}(t - t_s)^{\beta_s} f_s(t)\,dt, \tag{3.34}$$

where $f_s(t)$ is analytic at $t = t_s$ but singular at $t = 0$. The singularity at $t = 0$ is regular for $f_3(t)$ and $f_4(t)$, but irregular for $f_1(t)$, $f_2(t)$. In the first case ($s = 3, 4$) it is a familiar fact of the theory of linear difference equations (Ref. 2, Chap. VIII; Ref. 10, pp. 485–487; Ref. 11, pp. 61–64) that, if a sufficiently large positive number ω is selected, $u_s(y)$, except for a constant factor, can be expressed in the following form:

$$u_s(y) = t_s^y \frac{\Gamma(y/\omega)}{\Gamma(y/\omega + \beta_s + 1)} \Omega^{(s)}(y), \tag{3.35}$$

where

$$\Omega^{(s)}(y) = 1 + \sum_{n=1}^{\infty} \times \frac{g_n^{(s)}}{(y + \omega\beta_s + \omega)(y + \omega\beta_s + 2\omega)\cdots(y + \omega\beta_s + n\omega)} \tag{3.36}$$

is an inverse factorial series convergent for $\operatorname{Re} y > y_s$. It may be seen from Fig. 2 that the paths l_3, l_4 can be deformed into the circles $|t - i| = 1$ and $|t + i| = 1$, respectively. As a result of this possibility it is allowed to take in this case $\omega = 1$ (Ref. 10, pp. 485–487; Ref. 11, pp. 61–64). For additional details Ref. 1 is cited. The above expansions, except for a constant factor, are immediately applicable to $v_{3s}(y) = w_{3s}(y)/\Gamma(y)$ with $\omega = 1$, $t_{3s} = \pm i = e^{\pm i\pi/2}$, $\beta_{3s} = \mp ic/2$. The inverse factorial series $\Omega^{(3)}(y)$, $\Omega^{(4)}(y)$ converge uniformly for $\operatorname{Re} y > 0$. Their coefficients $g_n^{(3)}$, $g_n^{(4)}$ can be obtained by standard methods [starting from the integral representations (3.30), (3.31), term-by-term integration, and use of the integral representation of the beta function], or by direct substitution into the difference equation (3.5) and application of the method of undetermined coefficients. They will be found explicitly later using the former method. The analytic continuation for $\operatorname{Re} y < 0$ is again provided by (3.23) making v_3 and v_4 analytic throughout the y-plane. From now on these definitions are adopted for v_3, v_4 instead of the original integral forms (3.30), (3.31). They are constant multiples of the latter.

When $t = 0$ is an irregular singular point of $f_s(t)$ in (3.34), convergent expansions of the form (3.35) and (3.36) can not, in general, be obtained. However, with $\omega = 1$ these expressions provide at least an asymptotic expansion for $u_s(y)$ as $y \to \infty$ in the sector $-\frac{1}{2}\pi + \epsilon < \arg y < \frac{1}{2}\pi - \epsilon$ ($\epsilon > 0$ and arbitrarily small) (Ref. 2, pp. 309–318; Ref. 3, pp. 70–74; Ref. 10, pp. 457–459; Ref. 12). Therefore, except for a constant factor, the following relations are valid

$$v_1(y) = \left(-\frac{1}{a}\right)^y \frac{\Omega^{(1)}(y)}{y(y+1)(y+2)};$$
$$v_2(y) = \left(-\frac{1}{b}\right)^y \frac{\Omega^{(2)}(y)}{y} \tag{3.37}$$

$$\Omega^{(1)}(y) \underset{y\to|\infty}{\sim} 1 + \frac{g_1^{(1)}}{y+3} + \frac{g_2^{(1)}}{(y+3)(y+4)} + \cdots,$$
$$\Omega^{(2)}(y) \underset{y\to|\infty}{\sim} 1 + \frac{g_1^{(2)}}{y+1} + \frac{g_2^{(2)}}{(y+1)(y+2)} + \cdots. \tag{3.38}$$

The above expressions hold in the sector $-\frac{1}{2}\pi + \epsilon < \arg y < \frac{1}{2}\pi - \epsilon$. The coefficients $g_n^{(s)}$, $s = 1, 2$ can be determined by both methods mentioned previously, the only difference being that the process is now at least formal. As is seen in the following it is not necessary to find them explicitly.

The linear independence of the so obtained solutions v_1, v_2, v_3, v_4 can be demonstrated by referring to Nörlund's theorem (Ref. 3, pp. 70–74; Ref. 12; Ref. 1) or by making use of a general theorem given in Ref. 10 (p. 360), based upon the behavior of the solutions $v_s(y)$ as $y \to |\infty$, which is now known.

For v_3, v_4 two convergent and for v_1, v_2 two at least asymptotic expansions have been obtained in terms of inverse factorial series valid in a half-plane limited to the left. It is a fact of the theory of linear difference equations that when $|y|$ is large (without any limitation to the left), the corresponding expansions $\Omega^{(s)}(y)$, though not necessarily convergent any more, are in any case developable asymptotically in series of the same form (Ref. 2, pp. 309–319; Ref. 10, pp. 457–459; Ref. 1). Thus, it is possible to write

$$v_1(y) = \left(-\frac{1}{a}\right)^y \frac{1+\epsilon_1(y)}{y(y+1)(y+2)};$$
$$v_2(y) = \left(-\frac{1}{b}\right)^y \frac{1+\epsilon_2(y)}{y}; \tag{3.39}$$
$$v_3(y) = \frac{i^y \Omega^{(3)}(y)}{\Gamma(y+1-ic/2)};$$
$$v_4(y) = \frac{(-i)^y \Omega^{(4)}(y)}{\Gamma(y+1+ic/2)}, \tag{3.40}$$

where

$$\lim_{|y|\to\infty} \epsilon_s(y) = 0, \qquad s = 1, 2, \tag{3.41}$$

$$\Omega^{(s)}(y) \underset{|y|\to\infty}{\sim} 1 + \frac{g_1^{(s)}}{y+1\mp ic/2} + \frac{g_2^{(s)}}{(y+1\mp ic/2)(y+2\mp ic/2)} + \cdots, \qquad s = 3, 4; \tag{3.42}$$

the upper sign should be used for $s = 3$, the lower for $s = 4$.

Asymptotic Expansions of $R_1(z)$, $R_2(z)$

The results of the preceding sections and Ford's theorems I and VI (Ref. 2, Chaps. I and VI, respectively; Ref. 1) can now be applied to yield the asymptotic expansions of $R_1(z)$, $R_2(z)$. For the latter function, the procedure applies only when $2\nu + 1$ differs from an integer. As stated by Eqs. (3.7), (3.8), the particular solution $\bar{v}(y)$ of (3.5) is sought that satisfies the conditions

$$\bar{v}(n+\sigma) = a_n \quad \text{for all integers } n \qquad (\sigma_1 = \nu + 1, \sigma_2 = -\nu). \tag{3.43}$$

In terms of the four independent solutions of the fourth-order difference equation (3.5), defined in the preceding section,

$$\bar{v}(y) = Av_1(y) + Bv_2(y) + Cv_3(y) + Dv_4(y), \tag{3.44}$$

where A, B, C, D are constants. The connection between (3.5) and the recurrence formula (2.5) for the a_n's shows immediately that the four initial conditions

$$\bar{v}(-3+\sigma) = 0, \quad \bar{v}(-2+\sigma) = 0, \quad \bar{v}(-1+\sigma) = 0, \quad \bar{v}(\sigma) = 1 \qquad (\sigma = \sigma_1 \text{ or } \sigma_2) \tag{3.45}$$

[or, for that matter, any four such relations obtained from (3.43) for different but arbitrary values of n] are enough for the satisfaction of (3.43) for all n. These conditions serve to determine the coefficients A, B, C, D of the linear combination (3.44). Deferring this matter for the moment it is possible to express the series

$$R_{1,2}(z) = z^\sigma \sum_{n=0}^{\infty} a_n z^n = z^\sigma \sum_{n=0}^{\infty} \bar{v}(n+\sigma) z^n (\sigma_1 = \nu + 1, \sigma_2 = -\nu)$$

[12] N. -E. Nörlund, Acta Math. 34, 16 (1911).

as follows:

$$R_{1,2}(z) = A\left(\frac{-z}{a}\right)^{\sigma} \sum_{n=0}^{\infty} G_1(n)\left(\frac{-z}{a}\right)^{n} + B\left(\frac{-z}{b}\right)^{\sigma} \sum_{n=0}^{\infty} G_2(n)\left(\frac{-z}{b}\right)^{n} + C(iz)^{\sigma} \sum_{n=0}^{\infty} G_3(n)(iz)^{n} + D(-iz)^{\sigma} \sum_{n=0}^{\infty} G_4(n)(-iz)^{n}, \quad (3.46)$$

where $G_s(n) = v_s(n + \sigma)/t_s^{n+\sigma}$, $t_1 = -a^{-1}$, $t_2 = -b^{-1}$, $t_3 = i$, $t_4 = -i$. In other words $G_s(n)$, when considered as functions of $w = x + iy$, are given by

$$G_1(w) = \frac{1 + \epsilon_1(w + \sigma)}{(w + \sigma)(w + \sigma + 1)(w + \sigma + 2)}; \qquad \lim_{|w|\to\infty} \epsilon_1(w + \sigma) = 0 \quad (3.47)$$

$$G_2(w) = \frac{1 + \epsilon_2(w + \sigma)}{w + \sigma}; \qquad \lim_{|w|\to\infty} \epsilon_2(w + \sigma) = 0 \quad (3.48)$$

$$G_3(w) = \Omega^{(3)}(w + \sigma)[\Gamma(w + \sigma + 1 - ic/2)]^{-1}$$

$$\Omega^{(3)}(w + \sigma) \underset{|w|\to\infty}{\sim} 1 + \frac{g_1^{(3)}}{w + \sigma + 1 - ic/2} + \frac{g_2^{(3)}}{(w + \sigma + 1 - ic/2)(w + \sigma + 2 - ic/2)} + \cdots \quad (3.49)$$

$$G_4(w) = \Omega^{(4)}(w + \sigma)[\Gamma(w + \sigma + 1 + ic/2)]^{-1}$$

$$\Omega^{(4)}(w + \sigma) \underset{|w|\to\infty}{\sim} 1 + \frac{g_1^{(4)}}{w + \sigma + 1 + ic/2} + \frac{g_2^{(4)}}{(w + \sigma + 1 + ic/2)(w + \sigma + 2 + ic/2)} + \cdots . \quad (3.50)$$

The functions $G_s(w)$ are analytic for all finite w just as the functions $v_s(w + \sigma)$ are. For Re $w > -\sigma$ the expansions for $\Omega^{(3)}$ and $\Omega^{(4)}$ are convergent series. The above expressions show that $G_1(w)$ and $G_2(w)$ satisfy all the conditions of Ford's theorem I (subject to certain remarks and generalizations), while $G_3(w)$ and $G_4(w)$ satisfy all the conditions of Ford's theorem VI (Ref. 2, Chaps I and VI; Ref. 1). Application of the ratio test to the four power series in (3.46) reveals that the first converges for $|z| < |a|$, the second for $|z| < |b|$, while the last two are entire functions. For large z the first two series can be expanded asymptotically using Theorem I, the last two by using Theorem VI. The process is straightforward and details will be omitted. They can be found in Ref. 1, or, for a different example, in Chapter VIII of Ref. 2. With $z = |z|\, e^{i\phi}$, $-\pi < \phi \leq \pi$ the final result is

$$\begin{aligned} R_{1,2}(z) &\sim -z^{\sigma} \sum_{n=1}^{\infty} \frac{\bar{v}(-n + \sigma)}{z^n} + De^{-\pi c/4} e^{-iz} z^{-ic/2}\left[1 + \sum_{n=1}^{\infty} \frac{g_n^{(4)}}{(-iz)^n}\right], && 0 < \phi < \pi \\ &\sim -z^{\sigma} \sum_{n=1}^{\infty} \frac{\bar{v}(-n + \sigma)}{z^n} + Ce^{-\pi c/4} e^{iz} z^{ic/2}\left[1 + \sum_{n=1}^{\infty} \frac{g_n^{(3)}}{(iz)^n}\right], && -\pi < \phi < 0 \\ &\sim -z^{\sigma} \sum_{n=1}^{\infty} \frac{\bar{v}(-n + \sigma)}{z^n} + Ce^{-\pi c/4} e^{iz} z^{ic/2}\left[1 + \sum_{n=1}^{\infty} \frac{g_n^{(3)}}{(iz)^n}\right] \\ &\qquad + De^{-\pi c/4} e^{-iz} z^{-ic/2}\left[1 + \sum_{n=1}^{\infty} \frac{g_n^{(4)}}{(-iz)^n}\right], && \phi = 0. \end{aligned} \quad (3.51)$$

Since $\bar{v}(-n + \sigma) = a_{-n} = 0$ $(n = 1, 2, 3, \cdots)$ the first term in all three expansions drops out. It is understood that $\sigma_1 = \nu + 1$ must be used for $R_1(z)$, $\sigma_2 = -\nu$ for $R_2(z)$, when $2\nu + 1$ differs from an integer. For large $|z|$ the branch lines starting from $z = -a$ and $z = -b$, for both R_1 and R_2, are drawn along $\phi = \pi$. This ray is excluded from the above expansions. Finally the problem is completed by evaluating explicitly the coefficients $g_n^{(3)}$ and $g_n^{(4)}$. Two standard methods have been outlined in the previous section [following Eq. (3.36)] and the first has been carried out in detail elsewhere (Ref. 1; see also Ref. 10, Chap. XV and Ref. 11, Chap. III), yielding four-term recurrence formulas for these coefficients. Omitting details the final results are

$$\sum_{m=0}^{3} T_m(n) g_{n-m}^{(s)} = 0; \qquad g_0^{(s)} = 1; \qquad g_{-j}^{(s)} = 0 (j = 1, 2, \cdots); \qquad s = 3, 4, \quad (3.52)$$

$$T_0(n) = \pm 2in, \quad (3.53)$$

$$T_1(n) = \mp i(n - 1 \mp ic/2)[n \mp \tfrac{1}{2}i(5a + 3b)] + c(1 \mp ia) \pm iv(v + 1), \quad (3.54)$$

$$T_2(n) = (a + b)(n - 2 \mp ic/2)^2 + 2b(1 \mp ia) \times (n - 2 \mp ic/2) - (a + b)\nu(\nu + 1), \quad (3.55)$$

$$T_3(n) = \pm iab(n - 3 - \nu \mp ic/2) \times (n - 2 + \nu \mp ic/2), \quad (3.56)$$

where in the last four expressions the upper sign is used for $s = 3$, the lower for $s = 4$. Obviously, for real a, b

$$g_n^{(4)} = \bar{g}_n^{(3)} \qquad (\text{real } a, b). \quad (3.57)$$

It remains to identify the asymptotic series in (3.51) with the expansions (2.30) for R_3 and R_4 that have already been obtained independently. This can be done without difficulty, for example by evaluating

$$g_1^{(34)} = -T_1(1)/T_0(1) = \pm i\{\tfrac{1}{2}c + [(\mp ic/2)(1 \mp ic/2) - \nu(\nu + 1) - cb](\pm 2i)^{-1}\}.$$

It is then obvious that the corresponding expansions in (3.51) are simply $Ce^{-\pi c/4}R_3(z)$ and $De^{-\pi c/4}R_4(z)$. Summarizing, it may be stated that in an entirely independent manner the required expansions

$$R_{1,2}(z) \underset{|z|\to\infty}{\sim} \begin{cases} A_4R_4(z), & 0 < \phi < \pi, \\ A_3R_3(z), & -\pi < \phi < 0, \\ A_3R_3(z) + A_4R_4(z), & \phi = 0, \end{cases} \quad (3.58)$$

where

$$A_4 = De^{-\pi c/4}, \qquad A_3 = Ce^{-\pi c/4}, \quad (3.59)$$

have been arrived at, in complete agreement with the results of the preceding section. At the same time, explicit relations have been obtained for the precise evaluation of the coefficients of the linear relations. This evaluation is carried out in the next section.

A final remark may be added, based on (3.57). For real a, b, z : $R_3(z) = \bar{R}_4(z)$, in agreement with (2.32), while $R_1(z)$, $R_2(z)$ are real functions. Therefore $A_3 = \bar{A}_4$ [from (3.58) and for $\phi = 0$], or $C = \bar{D}$ for real a, b.

4. DETERMINATION OF THE COEFFICIENTS OF THE LINEAR RELATIONS

The Multipliers of $v_s(y)$

In view of (3.59), only C and D must be determined. Four linear equations are obtained for A, B, C, D if (3.44) is substituted into the four relations (3.45). Their solutions can be written

$$A = \mu_1^{(4)}(\sigma - 3), \qquad B = \mu_2^{(4)}(\sigma - 3),$$
$$C = \mu_3^{(4)}(\sigma - 3), \qquad D = \mu_4^{(4)}(\sigma - 3), \quad (4.1)$$

where $\mu_s^{(4)}(y)$ $(s = 1, 2, 3, 4)$ are the cofactors of the last row of the determinant $P(y)$ divided by $P(y)$ and where

$$P(y) = \begin{vmatrix} v_1(y) & v_2(y) & v_3(y) & v_4(y) \\ v_1(y+1) & v_2(y+1) & v_3(y+1) & v_4(y+1) \\ v_1(y+2) & v_2(y+2) & v_3(y+2) & v_4(y+2) \\ v_1(y+3) & v_2(y+3) & v_3(y+3) & v_4(y+3) \end{vmatrix} \quad (4.2)$$

is Casorati's determinant (Ref. 10, Chap. XII; Ref. 11, Chap. I) for the four particular solutions $v_s(y)$ $(s = 1, 2, 3, 4)$ of (3.5). It satisfies the first-order difference equation (Heymann's theorem)

$$\frac{P(y+1)}{P(y)} = \frac{p_0(y)}{p_4(y)} = \frac{1}{ab(y + 3 - \nu)(y + 4 + \nu)}, \quad (4.3)$$

whose solution is given by (Ref. 10, pp. 327–328)

$$P(y) = \left(\frac{1}{ab}\right)^y \frac{\bar{w}(y)}{\Gamma(y + 3 - \nu)\Gamma(y + 4 + \nu)}; \quad (4.4)$$

$\bar{w}(y)$ is, in general, an arbitrary periodic function of y with period 1. For the particular solutions $v_s(y)$ $(s = 1, 2, 3, 4)$, whose explicit behavior is known as $y \to |\infty$, the corresponding $\bar{w}(y)$ can be determined. Thus letting $y \to |\infty$, substituting (3.37), (3.38), (3.40), and (3.42) in (4.2), and making use of the relation (Ref. 2, Chap. VIII; Ref. 10, pp. 254–255)

$$\frac{\Gamma(y + \tau)}{\Gamma(y + \rho)} \underset{y\to\infty}{\to} y^{\tau-\rho}, \qquad -\pi + \epsilon < \arg y < \pi - \epsilon, \quad (4.5)$$

it can be found[1] that $\bar{w}(y) = 2ic/(ab)^3$, i.e., a constant. Therefore

$$P(y) = \left(\frac{1}{ab}\right)^{y+3} \frac{2ic}{\Gamma(y + 3 - \nu)\Gamma(y + 4 + \nu)}. \quad (4.6)$$

It is now convenient to introduce the so-called multipliers $N_s(y)$ $(s = 1, 2, 3, 4)$ of the solutions $v_s(y)$ of (3.5) (Ref. 10, pp. 372–374):

$$N_s(y) = \mu_s^{(4)}(y + 1)/p_4(y) = \mu_s^{(4)}(y + 1)P(y + 1)/P(y), \quad (4.7)$$

where $p_0(y) = 1$ was used in writing the second expression. From their definition it is always possible to write them in determinantal form. With $p_0(y) = 1$ they are simply the negatives of the cofactors of the first row of $P(y)$ divided by $P(y)$. Using (4.7) in (4.1) it is possible to express A, B, C, D

as follows:

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = [ab(y+3-\nu)(y+4+\nu)\begin{bmatrix} N_1(y) \\ N_2(y) \\ N_3(y) \\ N_4(y) \end{bmatrix}]_{\nu=-4+\sigma}, \tag{4.8}$$

where $\sigma = \nu + 1$ is used for the expansion of $R_1(z)$, $\sigma = -\nu$ for that of $R_2(z)$.

The behavior of $N_s(y)$ as $y \to |\infty$ can also be found as was done for $P(y)$:

$$N_1(y) \underset{y\to|\infty}{\sim} \frac{(-a)^{y+3}}{c}\, y,$$

$$N_2(y) \underset{y\to|\infty}{\sim} -\frac{(-b)^{y+3}}{c}\frac{1}{y}, \tag{4.9}$$

$$N_{3,4}(y) \underset{y\to\infty}{\sim} -\tfrac{1}{2}(\mp i)^y \Gamma(y+3-\nu) y^{\nu-2\mp ic/2}$$

$$\sim -\tfrac{1}{2}(\mp i)^y \Gamma(y+4+\nu) y^{-\nu-3\mp ic/2}. \tag{4.10}$$

The two expansions given for N_3, N_4 are asymptotically equivalent, as may be shown with the use of (4.5).

The Adjoint Difference Equation

It is a well-known fact that the multipliers $N_s(y)$ are independent solutions of the difference equation adjoint to (3.5), i.e., they satisfy (Ref. 10, pp. 372–374; Ref. 11, Chap. I)

$$\sum_{m=0}^{4} p_m(y+4-m)N(y+4-m) = 0.$$

Direct application of Laplace's transformation to this equation yields two solutions proportional to $N_1(y)$ and $N_2(y)$, as may be seen by comparing their asymptotic behavior as $y \to |\infty$ with (4.9). In order to determine $N_3(y)$ and $N_4(y)$ put

$$N(y) = \Gamma(y+4+\nu)M(y), \tag{4.11}$$

to obtain the difference equation

$$Q_4(y)M(y+4) + Q_3(y)M(y+3) + Q_2(y)M(y+2) + Q_1(y)M(y+1) + Q_0(y)M(y) = 0, \tag{4.12}$$

where

$$Q_s(y) = p_{4-s}(y+s)\frac{\Gamma(y+4+\nu+s)}{\Gamma(y+5+\nu)}; \qquad s = 0, 1, 2, 3, 4. \tag{4.13}$$

The equation can be solved by application of Laplace's transformation; the integral representations, defining analytic functions in a certain right half-plane, are then expanded into inverse factorial series, asymptotic for $y \to |\infty$, as was outlined previously for $v_3(y)$, $v_4(y)$; proportionality factors can be determined comparing with (4.10); the continuation of these solutions to the left is provided by the difference equation itself, resulting in $N_3(y)$, $N_4(y)$ having simple poles at

$$y = \nu - 3 - n \quad \text{and} \quad y = -\nu - 4 - n, \qquad n = 0, 1, 2, \cdots. \tag{4.14}$$

Details, omitted here, can be found in Ref. 1. The final results for the asymptotic expansions follow:

$$N_{4,3}(y) = -\frac{(\pm i)^y}{2}\frac{\Gamma(y+4+\nu)\Gamma(y)}{\Gamma(y+\nu+3\mp ic/2)}T(y), \tag{4.15}$$

$$T(y) = 1 + \frac{d_1}{y+\nu+3\mp ic/2} + \frac{d_2}{(y+\nu+3\mp ic/2)(y+4+\nu\mp ic/2)} + \cdots. \tag{4.16}$$

Recurrence formula for the coefficients d_n:

$$d_n = \frac{1}{2n}\sum_{m=1}^{6}\tau_m(x)d_{n-m}; \qquad d_0 = 1; \qquad d_{-j} = 0, \qquad j = 1, 2, 3, \cdots, \tag{4.17}$$

$$x = n + \nu \mp ic/2,$$

$$\tau_1(x) = 11x^2 - [14\nu + 17 \mp i(7a - 3b)]x + 4\nu^2 + 8\nu - ac \mp i(4a\nu + 3c),$$

$$\tau_2(x) = -25x^3 + [39\nu + 102 \mp i(3b - 17a)]x^2 - [18\nu^2 + 81\nu + 89 + ab - 3a^2 \pm i(9b - a(16\nu + 39))]x + 2\nu^3 + 10\nu^2 + (12 - a^2 - ab)\nu - 3ac \mp i[(3a + b)\nu^2 + (11a + b)\nu + 3c],$$

$$\tau_3(x) = x\{30x^3 - [56\nu + 228 \pm i(b - 19a)]x^2 + [3(11\nu^2 + 85\nu + 172 - a^2) \pm 2i(3b - a(12\nu + 50))]x - 6\nu^3 - 65\nu^2 + (2a^2 - 245)\nu + 9a^2 - 330 \pm i[(7a + b)\nu^2 + (55a + b)\nu + 108a - 8b]\},$$

$$\tau_4(x) = x(x-1)(x-\nu-5)[-20x^2 + (24\nu + 122 \mp i10a)x + a^2 - 7\nu^2 - 67\nu - 174 \pm i2a(3\nu + 16)],$$

$$\tau_5(x) = x(x-1)(x-2)(x-\nu-6)(x-\nu-5)(7x-4\nu-25 \pm 2ia),$$

$$\tau_6(x) = -x(x-1)(x-2)(x-3)(x-\nu-7)(x-\nu-6)(x-\nu-5). \tag{4.18}$$

Throughout the above relations the upper sign is used for $N_4(y)$, the lower for $N_3(y)$.

Different but equivalent expansions for $N_4(y)$, $N_3(y)$ can be found by employing Boole's operational method for solving difference equations with rational coefficients, as explained in Ref. 8, pp. 434–461. It is essentially the method of undetermined coefficients, completely analogous to the method of Frobenius for differential equations. The final results are

$$N_{4,3}(y) = -\frac{(\pm i)^y}{2}\frac{\Gamma(y+4+\nu)\Gamma(y+3-\nu)}{\Gamma(y+6\mp ic/2)}T(y), \tag{4.19}$$

$$T(y) = 1 + \frac{d_1}{y+6\mp ic/2} + \frac{d_2}{(y+6\mp ic/2)(y+7\mp ic/2)} + \cdots. \tag{4.20}$$

Recurrence formula for the coefficients d_n:

$$d_n = \frac{1}{2n}\sum_{m=1}^{5} d_{n-m}f_{5-m}(x); \qquad d_0 = 1; \qquad d_{-j} = 0, \qquad j = 1, 2, 3, \cdots, \tag{4.21}$$

$$x = 2 - \nu - n \pm ic/2,$$

$$f_0(x) = x(x-1)(x-2)(x+2\nu+1)(x+2\nu)(x+2\nu-1),$$

$$f_1(x) = 2(x-1)(x-2)(x+2\nu)(x+2\nu-1)(3x+3\nu-3\mp ia),$$

$$f_2(x) = (x-2)(x+2\nu-1)[14x^2 + 4(7\nu-10\mp i2a)x + 12\nu^2 - 42\nu + 30 - a^2 \mp i4a(2\nu-3)],$$

$$f_3(x) = 16x^3 + [48\nu - 90 \mp i(11a-b)]x^2 + 2[20\nu^2 - 94\nu + 85 - a^2 \mp i((11a-b)(\nu-2)+b)]x$$
$$+ 8\nu^3 - 84\nu^2 + (184 - 2a^2)\nu + 4a^2 - 108 \mp i[8a\nu^2 + (7b-47a)\nu + 44a - 8b],$$

$$f_4(x) = 9x^2 + [18\nu - 39 \mp i2(3a-b)]x + 8\nu^2 - 40\nu + 42 - ac \mp i[2(3a-b)\nu + 7b - 15a]. \tag{4.22}$$

The sign convention is the same.

The inverse factorial series in all these expressions are at least asymptotic for large $|y|$. It is also obvious that for real a, b : $d_n^{(3)} = \bar{d}_n^{(4)}$. Then $N_3(y) = \bar{N}_4(\bar{y})$, or for real y: $N_3(y) = \bar{N}_4(y)$. From (4.8) it may be seen that in this case $C = \bar{D}$, for both R_1 and R_2, a fact that was inferred previously, Eq. (3.60), through different considerations.

C and D may now be evaluated using (4.8), with $y = \sigma - 4 = \nu - 3$ for $R_1(z)$ and $y = -\nu - 4$ for $R_2(z)$, when, in the latter case, $2\nu + 1$ differs from an integer. In general, the expansions given will not be useful for direct evaluation of $N(\nu - 3)$ and $N(-\nu - 4)$. However, $N_4(y)$ and $N_3(y)$ may be evaluated at y, $y+1$, $y+2$, $y+3$ with an adequately large y and then the difference equation satisfied by $N(y)$ can be used to obtain $N(y-1)$, $N(y-2)$, etc., up to $N(\nu-3)$ and $N(-\nu-4)$. All these considerations can be expressed in a more compact and general form, which will also prove advantageous in the next section, where the special case of integral values for $2\nu + 1$ is investigated.

Starting from (4.2), the definition of Casorati's determinant, multiply the first, second, third, and fourth columns by A, B, C, D, respectively, add columns, use the sum as the new fourth column and make use of (3.44) to obtain

$$D = P_4(y)/P(y), \tag{4.23}$$

where $P_4(y)$ is Casorati's determinant for the set of solutions $v_1, v_2, v_3, \bar{v}$ of the difference equation (3.5). Thus D is equal to the constant ratio of two Casorati's determinants of this equation, differing only in their last column. Development of $P_4(y)$ with respect to its last column yields

$$D = \frac{\bar{v}(y+3)}{P(y)}\begin{vmatrix} v_1(y) & v_2(y) & v_3(y) \\ v_1(y+1) & v_2(y+1) & v_3(y+1) \\ v_1(y+2) & v_2(y+2) & v_3(y+2) \end{vmatrix} - \frac{\bar{v}(y+2)}{P(y)}\begin{vmatrix} v_1(y) & v_2(y) & v_3(y) \\ v_1(y+1) & v_2(y+1) & v_3(y+1) \\ v_1(y+3) & v_2(y+3) & v_3(y+3) \end{vmatrix}$$
$$+ \frac{\bar{v}(y+1)}{P(y)}\begin{vmatrix} v_1(y) & v_2(y) & v_3(y) \\ v_1(y+2) & v_2(y+2) & v_3(y+2) \\ v_1(y+3) & v_2(y+3) & v_3(y+3) \end{vmatrix} - \frac{\bar{v}(y)}{P(y)}\begin{vmatrix} v_1(y+1) & v_2(y+1) & v_3(y+1) \\ v_1(y+2) & v_2(y+2) & v_3(y+2) \\ v_1(y+3) & v_2(y+3) & v_3(y+3) \end{vmatrix}. \quad (4.24)$$

Call (a), (b), (c), (d) the four terms of this equation. Recall the definition of $N_4(y)$ as the minor of the upper left element of $P(y)$ divided by $P(y)$ and use $P(y+1) = P(y)/p_4(y)$ to obtain

$$(a) = [\bar{v}(y+3)/P(y)]P(y-1)N_4(y-1)$$
$$= \bar{v}(y+3)p_4(y-1)N_4(y-1), \quad (4.25)$$
$$(d) = -\bar{v}(y)N_4(y). \quad (4.26)$$

For (b) substitute the elements of the last row by

$$v_s(y+3) = -[p_3(y-1)v_s(y+2) + p_2(y-1)v_s(y+1) + 2av_s(y) + v_s(y-1)]/p_4(y-1); \quad s = 1,$$

2, 3 and break it up into four determinants with the same upper two rows, as in (b), and with last rows containing each of the four terms in the above expansion, respectively. The second and third of these determinants vanish having two proportional rows. The remaining terms lead to

$$(b) = \bar{v}(y+2)p_3(y-1)N_4(y-1) + \bar{v}(y+2)p_4(y-2)N_4(y-2). \quad (4.27)$$

Treating the elements of the first row in (c) in a similar way yields

$$(c) = -\bar{v}(y+1)N_4(y+1) - 2a\bar{v}(y+1)N_4(y). \quad (4.28)$$

Substitution in (4.24) finally yields

$$D = \bar{v}(y+2)p_4(y-2)N_4(y-2) + [\bar{v}(y+3)p_4(y-1) + \bar{v}(y+2)p_3(y-1)]N_4(y-1) - [\bar{v}(y) + 2a\bar{v}(y+1)]N_4(y) - \bar{v}(y+1)N_4(y+1). \quad (4.29)$$

For $y = n + \sigma$ and the relation $\bar{v}(n+\sigma) = a_n$, for any integer n, the last expression reduces to

$$D = a_{n+2}p_4(n-2+\sigma)N_4(n-2+\sigma) + [a_{n+3}p_4(n-1+\sigma) + a_{n+2}p_3(n-1+\sigma)]N_4(n-1+\sigma) - [a_n + 2a\,a_{n+1}]N_4(n+\sigma) - a_{n+1}N_4(n+1+\sigma). \quad (4.30)$$

For an appropriately large value of n, the four values of $N_4(y)$ appearing in (4.30) can be evaluated from the asymptotic expansions given previously. The computations are checked automatically if more than one value of n, not necessarily consecutive, are used, the result being independent of n. At the same time the accuracy of evaluation is revealed. For $n = -3$: $D = p_4(\sigma-4)N_4(\sigma-4)$ is obtained, as in (4.8). C is given by the same equation (4.30), if $N_3(y)$ is substituted in place of $N_4(y)$.

5. THE LOGARITHMIC SOLUTION: $2\nu + 1$ = INTEGER

Asymptotic Expansion of $R_2(z)$ for Large $|z|$

For the coefficients B_m of the logarithmic solution $R_2(z)$ the recurrence formula (2.23) has been obtained in Sec. 2. The initial conditions are given by (2.21). An equivalent set is

$$B_0 = 0, \quad B_{-1} = b_0 d_{2\nu}, \quad B_{-2} = b_0 d_{2\nu-1}, \quad B_{-3} = b_0 d_{2\nu-2}, \quad (5.1)$$

where b_0, $d_{2\nu}$, $d_{2\nu-1}$, $d_{2\nu-2}$ are definite numbers obtained in Sec. 2. Now write $m+4$ for m in (2.23) and replace a_m by $\bar{v}(m+\nu+1)$. Next introduce in place of $m+\nu+1$ the general variable y and in place of B_m the function $v(y)$ such that

$$v(m+\nu+1) = B_m = b_{2\nu+1+m} \quad \text{for all integral values of } m. \quad (5.2)$$

As a result (2.23) assumes the form

$$p_4(y)v(y+4) + p_3(y)v(y+3) + p_2(y)v(y+2) + p_1(y)v(y+1) + p_0(y)v(y) = q_2(y)\bar{v}(y+4) + q_1(y)\bar{v}(y+3) + q_0(y)\bar{v}(y+2), \quad (5.3)$$

where the right-hand side is identical with (3.5), (3.6), while

$$q_2(y) = F_0(y+4) = -ab[2(y+4)-1] = -dp_4(y)/dy,$$

$$q_1(y) = F_1(y+3) = -2(a+b)(y+3) + 2b = -dp_3(y)/dy, \quad (5.4)$$

$$q_0(y) = F_2(y+2) = -2y - 3 = -dp_2(y)/dy.$$

The last expressions for $q_s(y)$ may be obtained from (3.6) by differentiation. $\bar{v}(y)$, in the right-hand side, is given by (3.44) as the particular solution of (3.5) which corresponds to $R_1(z)$ and which satisfies $\bar{v}(n+\nu+1) = a_n$; it has been determined in Secs. 3 and 4.

The corresponding to this case difference equation turns out to be an inhomogeneous equation, whose homogeneous part is identical with (3.5). The particular solution $v_i(y)$ of this equation is sought, which satisfies the conditions (5.2), in which B_m may be considered as definite, given constants. As before for R_1, just four of these conditions are sufficient. The general solution of (5.3) consists of the general solution of the homogeneous equation plus a particular solution $v_5(y)$ of the inhomogeneous equation (Ref. 10, Chap. XII), i.e.,

$$v_i(y) = Ev_1(y) + Fv_2(y) + Gv_3(y) + Hv_4(y) + v_5(y) = V(y) + v_5(y), \quad (5.5)$$

where $v_s(y)$ ($s = 1, 2, 3, 4$) are the four particular solutions of (3.5) corresponding to $R_1(z)$ and determined in Sec. 3, while E, F, G, H are constants, which will be determined so that four of the initial conditions (5.2) are satisfied. Now write $\bar{v}(y)$ for $v(y)$ in (3.5) and differentiate the equation with respect to y. Making use of (5.4) and the relations $dp_1(y)/dy = dp_0(y)/dy = 0$ it can be seen that a particular solution of (5.3) is

$$v_5(y) = \frac{d\bar{v}(y)}{dy} = A\frac{dv_1(y)}{dy} + B\frac{dv_2(y)}{dy} + C\frac{dv_3(y)}{dy} + D\frac{dv_4(y)}{dy}, \quad (5.6)$$

where A, B, C, D are the definite constants corresponding to $R_1(z)$ with $\sigma = \nu + 1$ and obtained in Sec. 4. A complete solution of (5.3) has thus been found.

The asymptotic expansion of $R_2(z)$ for large $|z|$ is obtained as before, for R_1, by making use of Theorems I and VI. From its definition (2.25) and the preceding expressions (5.2) and (5.5), $R_2(z)$ can be put in the form

$$R_2(z) = z^{-\nu}\sum_{n=0}^{2\nu} b_n z^n + u_1(z) + u_2(z), \quad |z| < \min(|a|, |b|), \quad (5.7)$$

where

$$u_1(z) = (\ln z)z^{\nu+1}\sum_{n=0}^{\infty}\bar{v}(n+\nu+1)z^n + z^{\nu+1}\sum_{n=0}^{\infty}v_5(n+\nu+1)z^n, \quad (5.8)$$

$$u_2(z) = z^{\nu+1}\sum_{n=0}^{\infty}V(n+\nu+1)z^n = z^{\nu+1} \times \sum_{n=0}^{\infty}[Ev_1(n+\nu+1) + Fv_2(n+\nu+1) + Gv_3(n+\nu+1) + Hv_4(n+\nu+1)]z^n. \quad (5.9)$$

For large $|z|$ the asymptotic expansion of $u_2(z)$ is obtained in exactly the same manner as for $R_1(z)$ with $\sigma = \nu + 1$. Thus, as in (3.51)

$$u_2(z) \sim -z^{\nu+1}\sum_{n=1}^{\infty}\frac{V(-n+\nu+1)}{z^n} + He^{-\pi c/4}R_4(z), \quad 0 < \phi < \pi,$$

$$-z^{\nu+1}\sum_{n=1}^{\infty}\frac{V(-n+\nu+1)}{z^n} + Ge^{-\pi c/4}R_3(z), \quad -\pi < \phi < 0,$$

$$-z^{\nu+1}\sum_{n=1}^{\infty}\frac{V(-n+\nu+1)}{z^n} + e^{-\pi c/4}[GR_3(z) + HR_4(z)], \quad \phi = 0, \quad (5.10)$$

where the bracketed series expansions in (3.51) have been identified with $e^{\pm iz}z^{\pm ic/2}R_{4,3}(z)$.

Next consider $u_1(z)$. By making use of (5.6) it can be put in the form

$$u_1(z) = \left[\frac{\partial u_3(z,\sigma)}{\partial\sigma}\right]_{\sigma=\nu+1}, \quad (5.11)$$

where

$$u_3(z,\sigma) = z^\sigma\sum_{n=0}^{\infty}\bar{v}(n+\sigma)z^n, \quad |z| < \min(|a|, |b|). \quad (5.12)$$

In fact the part $u_1(z)$ in R_2, in the form shown by (5.11), could be written down from the start, if, in order to obtain the logarithmic solution R_2, the general method of Frobenius had been followed, as applied in this special case and explained in Ref. 7 (pp. 396–404). The asymptotic expansion of $u_3(z,\sigma)$ for large $|z|$ can be written down in exactly the same manner as for $R_1(z)$ and $u_2(z)$, using σ as a variable parameter

$$u_3(z,\sigma) \underset{|z|\to\infty}{\sim} -z^\sigma \sum_{n=1}^{\infty} \frac{\bar{v}(-n+\sigma)}{z^n} + De^{-\pi c/4}R_4(z), \qquad 0<\phi<\pi,$$

$$\underset{|z|\to\infty}{\sim} -z^\sigma \sum_{n=1}^{\infty} \frac{\bar{v}(-n+\sigma)}{z^n} + Ce^{-\pi c/4}R_3(z), \qquad -\pi<\phi<0,$$

$$\underset{|z|\to\infty}{\sim} -z^\sigma \sum_{n=1}^{\infty} \frac{\bar{v}(-n+\sigma)}{z^n} + e^{-\pi c/4}[CR_3(z)+DR_4(z)], \qquad \phi=0. \tag{5.13}$$

The parameter σ varies around the root $\nu+1$ of the indicial equation about $z=0$ of Eq. (2.4). In (5.12), the function $u_3(z,\sigma)$ is defined for $|z|<\min(|a|,|b|)$, by a uniformly convergent series of analytic functions of σ [$\bar{v}(y)$ was proved to be analytic for all y and z^σ is an analytic function of σ] and can, therefore, be differentiated any number of times with respect to σ (Ref. 11, p. 400). On the other hand, in the asymptotic expansion (5.13) the dependence of $u_3(z,\sigma)$ on σ appears (for z in any sector) only in the series

$$-z^\sigma \sum_{n=1}^{\infty} \frac{\bar{v}(-n+\sigma)}{z^n}.$$

Each term of this series is an analytic function of σ; in particular, for $\sigma=\nu+1$: $\bar{v}(-n+\nu+1)=a_{-n}=0$ $(n=1,2,\cdots)$, i.e., each term of the series vanishes. Then, with σ varying in the vicinity of $\nu+1$ and for sufficiently large $|z|$,

$$-z^\sigma \sum_{n=1}^{\infty} \frac{\bar{v}(-n+\sigma)}{z^n}$$

is a uniformly convergent series of analytic functions of σ. Therefore, differentiation with respect to σ any number of times is again permissible. Based on these remarks the asymptotic expansion of $\partial u_3/\partial\sigma$ may be obtained by differentiation of (5.13)

$$\frac{\partial u_3(z,\sigma)}{\partial\sigma} \underset{|z|\to\infty}{\sim} -(\ln z)\, z^\sigma \sum_{n=1}^{\infty} \frac{\bar{v}(-n+\sigma)}{z^n} - z^\sigma \sum_{n=1}^{\infty} \frac{\partial\bar{v}(-n+\sigma)}{\partial(-n+\sigma)} \frac{1}{z^n}, \qquad -\pi<\phi<\pi. \tag{5.14}$$

Letting $\sigma=\nu+1$, use (5.11), (5.6), and $\bar{v}(-n+\nu+1)=0$ $(n=1,2,\cdots)$ to obtain

$$u_1(z) \underset{|z|\to\infty}{\sim} -z^{\nu+1} \sum_{n=1}^{\infty} \frac{v_5(-n+\nu+1)}{z^n}. \tag{5.15}$$

This result can be combined with the first term of $u_2(z)$ in (5.10); in connection with (5.5) and (5.2) the combination yields

$$-z^{\nu+1} \sum_{n=1}^{\infty} \frac{v_i(-n+\nu+1)}{z^n} = -z^{\nu+1} \sum_{n=1} \frac{b_{-n+2\nu+1}}{z^n} = -z^{\nu+1} \sum_{n=1}^{2\nu+1} \frac{b_{-n+2\nu+1}}{z^n},$$

where $b_{-1}=b_{-2}=\cdots=0$ was used to reduce the series into a finite sum. Change of the summation index, $n=2\nu+1-m$, yields

$$-z^{\nu+1} \sum_{m=2\nu}^{m=0} \frac{b_m}{z^{2\nu+1-m}} = -z^{-\nu} \sum_{m=0}^{2\nu} b_m z^m$$

and this exactly cancels the finite sum in (5.7), which, of course, remains unchanged for large $|z|$. Therefore

$$R_2(z) \underset{|z|\to\infty}{\sim} A_{24}R_4(z), \qquad 0<\phi<\pi,$$

$$A_{23}R_3(z), \qquad -\pi<\phi<0,$$

$$A_{23}R_3(z)+A_{24}R_4(z), \qquad \phi=0, \tag{5.16}$$

where

$$A_{24}=He^{-\pi c/4}, \qquad A_{23}=Ge^{-\pi c/4}. \tag{5.17}$$

The result is in agreement with theory (Ref. 7, pp. 168–174; Ref. 8, pp. 72–73) and analogous to the result for $R_1(z)$.

Determination of the Coefficients *G*, *H*

Following the steps that led from Casorati's determinant (4.2) to expression (4.23) for D, it is possible to obtain a similar expression for H by making use of the equation

$$v_i(y)-\bar{v}'(y)=Ev_1(y)+Fv_2(y)+Gv_3(y)+Hv_4(y). \tag{5.18}$$

It follows that $H=\bar{P}_4(y)/P(y)$, where $\bar{P}_4(y)$ is Casorati's determinant for the set of solutions v_1, v_2, v_3, $v_i-\bar{v}'$ of the difference equation (3.5). Moreover the procedure that led from (4.23) to (4.29) shows that H may be expressed as follows

$$H=[v_i(y+2)-\bar{v}'(y+2)]p_4(y-2)N_4(y-2)+\{[v_i(y+3)-\bar{v}'(y+3)]p_4(y-1)+[v_i(y+2)-\bar{v}'(y+2)]p_3(y-1)\}N_4(y-1)-\{v_i(y)-\bar{v}'(y)+2a[v_i(y+1)-\bar{v}'(y+1)]\}N_4(y)-[v_i(y+1)-\bar{v}'(y+1)]N_4(y+1). \tag{5.19}$$

Differentiation of (4.29) with respect to y followed by addition with (5.19) yields

$$H = v_i(y+2)p_4(y-2)N_4(y-2) + [v_i(y+3)p_4(y-1) + v_i(y+2)p_3(y-1)]N_4(y-1)$$
$$- [v_i(y) + 2av_i(y+1)]N_4(y) - v_i(y+1)N_4(y+1) + \bar{v}(y+2)(d/dy)[p_4(y-2)N_4(y-2)]$$
$$+ \bar{v}(y+3)(d/dy)[p_4(y-1)N_4(y-1)] + \bar{v}(y+2)(d/dy)[p_3(y-1)N_4(y-1)]$$
$$- [\bar{v}(y) + 2a\bar{v}(y+1)][dN_4(y)/dy] - \bar{v}(y+1)[dN_4(y+1)/dy]. \tag{5.20}$$

Finally set $y = n + \nu + 1$ and use $v_i(n+\nu+1) = B_n$, $\bar{v}(n+\nu+1) = a_n$ to obtain

$$H = [B_{n+2}p_4(n+\nu-1) + a_{n+2}p_4'(n+\nu-1)]N_4(n+\nu-1) + [B_{n+3}p_4(n+\nu) + B_{n+2}p_3(n+\nu)$$
$$+ a_{n+3}p_4'(n+\nu) + a_{n+2}p_3'(n+\nu)]N_4(n+\nu) - [B_n + 2aB_{n+1}]N_4(n+\nu+1) - B_{n+1}N_4(n+\nu+2)$$
$$+ a_{n+2}p_4(n+\nu-1)N_4'(n+\nu-1) + [a_{n+3}p_4(n+\nu) + a_{n+2}p_3(n+\nu)]N_4'(n+\nu)$$
$$- [a_n + 2aa_{n+1}]N_4'(n+\nu+1) - a_{n+1}N_4'(n+\nu+2). \tag{5.21}$$

Exactly the same expression gives G, if $N_3(y)$ is substituted in place of $N_4(y)$.

An expression for $N'(y)$, at least asymptotic for $y \to |\infty$, can be found by direct differentiation of (4.15), (4.16), or (4.19), (4.20) (Ref. 10, pp. 434–461 and 457–459). The corresponding results are

$$N_{4_s}'(y) = [\pm\tfrac{1}{2}i\pi + \psi(y+\nu+4) + \psi(y)]N_{4_s}(y) + \tfrac{1}{2}(\pm i)^y \frac{\Gamma(y+\nu+4)\Gamma(y)}{\Gamma(y+\nu+3\mp\frac{1}{2}ic)}$$
$$\times\left[\psi(y+\nu+3\mp\tfrac{1}{2}ic) + \frac{d_1\psi(y+\nu+4\mp\frac{1}{2}ic)}{y+\nu+3\mp\frac{1}{2}ic} + \frac{d_2\psi(y+\nu+5\mp\frac{1}{2}ic)}{(y+\nu+3\mp\frac{1}{2}ic)(y+\nu+4\mp\frac{1}{2}ic)} + \cdots\right] \tag{5.22}$$

$$N_{4_a}'(y) = [\pm\tfrac{1}{2}i\pi + \psi(y+\nu+4) + \psi(y+3-\nu)]N_{4_a}(y) + \tfrac{1}{2}(\pm i)^y \frac{\Gamma(y+\nu+4)\Gamma(y+3-\nu)}{\Gamma(y+6\mp\frac{1}{2}ic)}$$
$$\times\left[\psi(y+6\mp\tfrac{1}{2}ic) + \frac{d_1\psi(y+7\mp\frac{1}{2}ic)}{y+6\mp\frac{1}{2}ic} + \frac{d_2\psi(y+8\mp\frac{1}{2}ic)}{(y+6\mp\frac{1}{2}ic)(y+7\mp\frac{1}{2}ic)} + \cdots\right], \tag{5.23}$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$, $\psi(z+1) = \psi(z) + 1/z$ and $\psi(1) = -\gamma$ = Euler's constant.

It may be seen from the above results that for real a, b : $G = \bar{H}$. $H(G)$ have been expressed in terms of N_4, N_4' (N_3, N_3') only, just as $D(C)$ were found in terms of N_4 (N_3). The remarks following (4.30) and concerning the computation of C, D apply to the numerical evaluation of G, H as well. For large n, $\Gamma(x)$ and $\psi(x)$ are easily evaluated with the use of their well-known asymptotic expansions

$$\Gamma(x) \sim e^{-x}x^{x-\frac{1}{2}}(2\pi)^{\frac{1}{2}}\left[1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \frac{571}{2488320x^4} + \frac{163879}{209018880x^5} + \frac{5246819}{75246796800x^6} - \cdots\right] \tag{5.24}$$

$$\psi(x) \sim \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \frac{1}{240x^8} - \frac{1}{132x^{10}} + \frac{691}{32760x^{12}} - \frac{1}{12x^{14}} + \cdots. \tag{5.25}$$

Even for $x = 2$, the second series yields $\psi(2)$ with an accuracy of six decimals, while the same is true for $\Gamma(x)$ and $x = 3$.

Notice, however, that n cannot be given very large values in (4.30), (5.21). Since a_n's and B_n's are involved in these relations and the accuracy of their evaluation diminishes as n increases, there is a limitation to the values of n that can be used. It was also observed that for large n the summation of the terms in the right-hand sides of (4.30) and (5.21) destroyed the accuracy rapidly by eliminating the first significant decimals of the individual terms. In each particular case, i.e., for given values of a, b, there is an optimum range of values of n for which (4.30) and (5.21), with a given accuracy of computation, yield the most accurate results. Additional remarks regarding the numerical evaluation of C, D, G, H, as well as of the functions R_1, R_2, R_3, R_4, can be found in Ref. 6.

As an indication, a few results are given, obtained with eight-decimal machine computations for the special case $a = 12$, $b = 10$, $c = 2$, $C = \bar{D}$, $G = \bar{H}$,

ν	D	H
1	$4.14617 - i4.06427$	$-252.508 - i265.563$
3	$-133.193 + i117.365$	$898.741 + i1306.918$
5	$(1.24526 - i0.965800) \times 10^4$	$(-0.875955 - i4.30989) \times 10^4$
7	$(-2.32660 + i1.5730) \times 10^6$	$(-2.05026 + i3.83348) \times 10^6$
9	$(7.2577 - i4.2497) \times 10^8$	$(10.0441 - i7.42212) \times 10^8$

ν	x	$R_2(x)$ from convergent series	$R_2(x)$ from (5.16)
1	12	38.0483	38.1064
	14	114.081	114.082
3	12	-327.000	-326.902
	14	-401.884	-401.875
5	12	1.46444×10^4	1.46278×10^4
	14	7.44758×10^3	7.44795×10^3
7	12	-1.95279×10^6	-1.95571×10^6
	14	-2.95473×10^5	-2.95481×10^5
9	12	4.80712×10^8	4.80351×10^8
	14	-3.75249×10^8.	-3.75230×10^8.

$R_3(x) = \bar{R}_4(x)$. Besides D and H, the values of $R_2(x)$ for $x = 12$ and $x = 14$ are given for $\nu = 1, 3, 5, 7, 9$. These values of x fall in the overlapping region between the asymptotic series (5.16) for R_2 and a convergent series obtained with the use of the transformation (2.26). Both values of R_2 are given for comparison. R_1 is not given, since it is included in (2.25), defining R_2.

For $x = 14$, falling roughly in the middle of the overlapping region in this case, the agreement is good up to five significant decimals.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. P. King and Professor T. T. Wu of Harvard University for their advice and encouragement during the course of this work. The numerical computations were done at the Computation Center at MIT, Cambridge, Massachusetts. The generous allotment of computer time by the staff of the Center is hereby thankfully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

On Polynomial Systems in a Banach Ring*

J. G. TAYLOR†

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, and Churchill College, Cambridge, England

(Received 30 June 1964)

We define and discuss equations on Banach rings (algebras) which are of polynomial form. We prove a local uniqueness theorem for the homogeneous case, and an existence and local uniqueness theorem for the nonhomogeneous case. In order to apply these results to the equations of Lagrangian quantum field theory we find it necessary to extend the concept of a ring to that of an n-ring. The resulting theory is applied to a simple model equation arising in quantum field theory.

1. INTRODUCTION

NONLINEAR problems are of great interest in theoretical physics. In particular the equations of quantum field theory and general relativity are highly nonlinear. The typical system of nonlinear equations is between a set of elements forming a linear space over the complex numbers. The nonlinearity in the system of equations allows a multiplication to be introduced in the linear space, so turning the space into a ring.[1] The system of equations involves N elements (where N may be infinite), and is often in the form of the vanishing of a set of N polynomials in the N elements. The degree of these polynomials has a finite upper bound M. We call such a system of equations a polynomial system. Polynomial systems arise, for example, in Lagrangian quantum field theory.[2] Here N is infinite, corresponding to the fact that there is no upper bound on the number of particles existing

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research, OAR, under Grant No. AF EOAR 63-79 with the European Office of Aerospace Research, U. S. Air Force.

† Present address: Department of Physics, Rutgers, The State University, New Brunswick, New Jersey.

[1] We are using the terminology of M. A. Naimark *Normed Rings* (P. Noordhoff Ltd., Groningen, The Netherlands, 1959), English transl.

[2] This is discussed in more detail in papers by this author in Nuovo Cimento Suppl. 1, 857 (1963).

ν	D	H
1	$4.14617 - i4.06427$	$-252.508 - i265.563$
3	$-133.193 + i117.365$	$898.741 + i1306.918$
5	$(1.24526 - i0.965800) \times 10^4$	$(-0.875955 - i4.30989) \times 10^4$
7	$(-2.32660 + i1.5730) \times 10^6$	$(-2.05026 + i3.83348) \times 10^6$
9	$(7.2577 - i4.2497) \times 10^8$	$(10.0441 - i7.42212) \times 10^8$

ν	x	$R_2(x)$ from convergent series	$R_2(x)$ from (5.16)
1	12	38.0483	38.1064
	14	114.081	114.082
3	12	-327.000	-326.902
	14	-401.884	-401.875
5	12	1.46444×10^4	1.46278×10^4
	14	7.44758×10^3	7.44795×10^3
7	12	-1.95279×10^6	-1.95571×10^6
	14	-2.95473×10^5	-2.95481×10^5
9	12	4.80712×10^8	4.80351×10^8
	14	-3.75249×10^8.	-3.75230×10^8.

$R_3(x) = \bar{R}_4(x)$. Besides D and H, the values of $R_2(x)$ for $x = 12$ and $x = 14$ are given for $\nu = 1, 3, 5, 7, 9$. These values of x fall in the overlapping region between the asymptotic series (5.16) for R_2 and a convergent series obtained with the use of the transformation (2.26). Both values of R_2 are given for comparison. R_1 is not given, since it is included in (2.25), defining R_2.

For $x = 14$, falling roughly in the middle of the overlapping region in this case, the agreement is good up to five significant decimals.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. P. King and Professor T. T. Wu of Harvard University for their advice and encouragement during the course of this work. The numerical computations were done at the Computation Center at MIT, Cambridge, Massachusetts. The generous allotment of computer time by the staff of the Center is hereby thankfully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

On Polynomial Systems in a Banach Ring*

J. G. TAYLOR†

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, and Churchill College, Cambridge, England

(Received 30 June 1964)

We define and discuss equations on Banach rings (algebras) which are of polynomial form. We prove a local uniqueness theorem for the homogeneous case, and an existence and local uniqueness theorem for the nonhomogeneous case. In order to apply these results to the equations of Lagrangian quantum field theory we find it necessary to extend the concept of a ring to that of an n-ring. The resulting theory is applied to a simple model equation arising in quantum field theory.

1. INTRODUCTION

NONLINEAR problems are of great interest in theoretical physics. In particular the equations of quantum field theory and general relativity are highly nonlinear. The typical system of nonlinear equations is between a set of elements forming a linear space over the complex numbers. The nonlinearity in the system of equations allows a multiplication to be introduced in the linear space, so turning the space into a ring.[1] The system of equations involves N elements (where N may be infinite), and is often in the form of the vanishing of a set of N polynomials in the N elements. The degree of these polynomials has a finite upper bound M. We call such a system of equations a polynomial system. Polynomial systems arise, for example, in Lagrangian quantum field theory.[2] Here N is infinite, corresponding to the fact that there is no upper bound on the number of particles existing

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research, OAR, under Grant No. AF EOAR 63-79 with the European Office of Aerospace Research, U. S. Air Force.

† Present address: Department of Physics, Rutgers, The State University, New Brunswick, New Jersey.

[1] We are using the terminology of M. A. Naimark *Normed Rings* (P. Noordhoff Ltd., Groningen, The Netherlands, 1959), English transl.

[2] This is discussed in more detail in papers by this author in Nuovo Cimento Suppl. 1, 857 (1963).

at any one time. M is the degree of the interaction in the Lagrangian. The elements of interest are S-matrix elements for different numbers of particles, and the nonlinearity is that obtained by multiplying together two S-matrix elements and integrating over a certain number of the particle momenta. A similar type of equation arises in the bootstrap approach to elementary particles, with both M and N finite.

The homogeneous equations (involving no fixed elements of the ring), have been discussed recently from the viewpoint of representation theory for commutative rings.[3] We wish to discuss here the more general case of not necessarily homogeneous polynomial systems in a not necessarily commutative ring. The equations we are considering are not expected to have an exact solution, so we must set up approximation procedures to solve them. The questions which we need to answer before doing this are: (a) do the equations have a solution? (b) if so, is this solution unique?

In this paper we want to give answers to these questions for a certain class of polynomial systems on a normed ring.[1] We first give a precise definition of polynomial systems in the next section. We then show, in the third section, that for homogeneous polynomial systems solutions arbitrarily close to zero cannot occur. In the corresponding nonhomogeneous case we show in the fourth section that at least one nontrivial solution will occur. These results may be applied directly to certain simplified models of the quantum field equations. We do this for a simple case in the last section, after generalizing the notion of a ring in Sec. 5.

2. POLYNOMIAL SYSTEMS

We recall that a normed ring[1] R, sometimes known as a normed algebra, is a collection of elements $x, y, z, \cdots$ which is a normed linear space over the complex numbers with the norm $|x|$ of x satisfying $|xy| \leq |x|\,|y|$ for any pair of elements x, y of R. Since we can always embed a noncomplete normed ring in a complete normed ring we will assume R is complete, and thus is a Banach ring.

Let us consider N elements $x_1, \cdots, x_N$ in R, with N finite. A system of s polynomial equations of finite degree M will be of the form

$$\sum_{\mathbf{n}} a_{\mathbf{n}}^{(r)} x_1^{n_1} \cdots x_N^{n_N} = 0, \qquad r = 1, \cdots, s, \tag{1}$$

where the summation in (1) is over all values of the N-vector $\mathbf{n} = (n_1, \cdots, n_N)$ with $|n| = \sum_{i=1}^{n} n_i \leq M$. We say that this system is a homogeneous polynomial system if $s = N$, and the N polynomial equations can be written in the form

$$x_r = \sum_{\mathbf{n}} b_{\mathbf{n}}^{(r)} x_1^{n_1} \cdots x_N^{n_N}. \tag{2}$$

In (2) the summation is the same as in (1), though now $b_{\mathbf{n}}^{(r)} = 0$ if $|\mathbf{n}| = 0$ or 1. Evidently every system of N polynomial equations is not a homogeneous polynomial system of the form (2). The results we derive in the next two sections may still be true for the more general system (1).

We now extend (2) to a polynomial system as follows: we interpret the quantities $b_{\mathbf{n}}^{(r)}$ in (2) as given elements of the normed ring, with the condition $b_{\mathbf{n}}^{(r)} = 0$ if $|\mathbf{n}| = 1$. Thus we now allow an inhomogeneous term x_r^0 on the right-hand side of (2). Then we have the

Definition 1. A finite polynomial system on a Banach ring R is a set of N equations in the N elements $x_1, \cdots, x_N$ of R of the form

$$x_r = \sum_{\mathbf{n}} b_{\mathbf{n}}^{(r)} x_1^{n_1} \cdots x_N^{n_N}, \tag{3}$$

where the summation is over vectors $\mathbf{n} = (n_1, n_2, \cdots, n_N)$ with $|\mathbf{n}| \leq M$, $b_{\mathbf{n}}^{(r)} = 0$ if $|\mathbf{n}| = 1$, and the $b_{\mathbf{n}}^{(r)}$ are given elements of R.

Definition 2. A homogeneous finite polynomial system on a Banach ring R is the same set of equations as in Definition 1 except that in addition $b_{\mathbf{n}}^{(r)} = 0$ if $|\mathbf{n}| = 0$.

We may let N become infinite above, provided that the resulting series on the right-hand side of (3) are absolutely convergent for any set $(x_1, x_2, \cdots)$. Thus we have

Definition 3. A polynomial system on a Banach ring R is a set of equations, one for each of a set of elements x_α, where α runs over an arbitrary index set A, of the form

$$x_\alpha = \sum_{\mathbf{n}} b_{\mathbf{n}}^{(\alpha)} \prod x_\beta^{n_\beta}, \tag{4}$$

where the summation is over vectors $\mathbf{n}$ with components n_α satisfying $|\mathbf{n}| = \sum_{\alpha \in A} n_\alpha \leq M$. The elements $b_{\mathbf{n}}^{(\alpha)}$ are in R, and $b_{\mathbf{n}}^{(\alpha)} = 0$ if $|\mathbf{n}| = 1$. The series on the right of (4) must converge absolutely in R for any choice of the set x_β and each α.

This requirement means that for each α only a denumerable infinity of $b_{\mathbf{n}}^{(\alpha)}$ are nonzero. Also the condition $|\mathbf{n}| \leq M$ in the summation means that only a finite number of factors enter in the product $x_\beta^{n_\beta}$.

From Definition 3 we may define a homogeneous polynomial system by adding to that definition the requirement $b_{\mathbf{n}}^{(\alpha)} = 0$ if $|\mathbf{n}| = 0$, for any α.

[3] M. M. Broido, "On Homogeneous Equations," Cambridge University preprint, 1964 (to be published).

The case when A corresponds to the set of positive integers arises in quantum field theory,[2] the homogeneous case arising for bootstrapped particles.[4]

3. HOMOGENEOUS POLYNOMIAL SYSTEMS

We wish to prove that the only solution to a finite homogeneous polynomial system with $x_1, \cdots, x_N$ small enough is the trivial one $x_1 = x_2 = \cdots = x_N = 0$. A similar result should hold for the more general polynomial system of Definition 3, though our method doesn't immediately apply to it.

To prove our result we consider the finite polynomial system as defining a transformation on the direct product $R \times R \cdots \times R = R^N$ of R with itself N times. This transformation F is defined for any element $x = (x_1, \cdots, x_N)$ of R^N by $x' = F(x)$, with

$$x_r' = \sum b_{\mathbf{n}}^{(r)} x_1^{n_1} \cdots x_N^{n_N}. \tag{5}$$

Let C be some positive number with $|b_{\mathbf{n}}^{(r)}| \leq C$ for all $\mathbf{n}$ and r. We introduce the product norm on R^N:

$$|x| = \sup_{1 \leq i \leq N} |x_i|. \tag{6}$$

Then from (5) and (6), for any two $x^{(1)}$, $x^{(2)}$ with $|x^{(1)}| < f$, $|x^{(2)}|, < f$, for some given positive f,

$$|x^{(1)\prime} - x^{(2)\prime}| \leq Ch(f)\, |x^{(1)} - x^{(2)}|, \tag{7}$$

where the function $h(z)$ is defined by

$$h(z) = \sum_{|\mathbf{n}| \leq M} C\, |\mathbf{n}|\, z^{\mathbf{n}-1}. \tag{8}$$

The summation over $\mathbf{n}$ in (8) is only over $|\mathbf{n}| > 1$, so $h(f)$ is a decreasing positive function as f decreases. Thus we may choose an f small enough so that $Ch(f) < 1$. Also $|x'| \leq Cg(f)$ where

$$g(z) = \sum_{\mathbf{n}} Cz^{\mathbf{n}}. \tag{9}$$

Again the summation in (9) is over all $\mathbf{n}$ with $1 < |\mathbf{n}| \leq M$. Hence we may choose f small enough so that $Cg(f) < f$. If we let d denote the maximum value of f to satisfy both $Ch(f) < 1$ and $Cg(f) < f$, then for all x in R^N with $|x| < d$ the transformation F maps this sphere into itself. From (7), F also satisfies the conditions necessary for application of the contraction mapping principle.[5] Then there is a unique solution to the polynomial system (3) in the sphere $|x| < d$ in R^N. Since there is already the trivial solution $x = 0$, then this is the only solution there. This proves the

Theorem 1. The only solution to a finite homogeneous polynomial system with $x_1, \cdots, x_N$ small enough is the trivial one $x_1 = x_2 = \cdots = x_N = 0$.

We would like to extend this local uniqueness theorem to a global uniqueness theorem, as done by Broido.[6] This is not possible, as is shown by the homogeneous polynomial system in two variables

$$x_1 = x_1 x_2, \qquad x_2 = x_2^2.$$

This has the solution $x_2 = e$ (the identity) for any x_1, as well as the trivial solution $x_1 = x_2 = 0$. This example also shows that we may have an infinity of different solutions to the system.

4. EXISTENCE OF A SOLUTION

We now turn to the problem of constructing a solution for a not-necessarily-homogeneous finite polynomial system. If the system is homogeneous, the solution we construct will be the trivial one. For the nonhomogeneous system we may again consider the mapping defined by the system. Now it is of the form $x \to F(x) + x^{(0)}$, where $F(x)$ has the same expression as (5), with $b_{\mathbf{n}}^{(r)} = 0$ if $|\mathbf{n}| \leq 1$, and $x^{(0)}$ is a given element of R^N. We may again use the contraction mapping principle to prove the existence of a unique solution of the system

$$x = x^{(0)} + F(x) \tag{10}$$

in some sphere $|x| < d$, provided d is chosen so that $|x^0 + F(x)| < d$ if $|x| < d$. This will be true if

$$|x^0| + Cg(d) < d. \tag{11}$$

We also require $Ch(d) < 1$, where h is the function given in Eq. (8). It is evident that if C is small enough (11) is satisfied for some $d > |x^0|$. So we have proved that for some range of C small enough there exists a unique solution to (10) in some suitable sphere round the origin in R^N. This proves the

Theorem 2. For a suitable upper bound on the norms of the elements $b_{\mathbf{n}}^{(r)}$ of R the nonhomogeneous finite polynomial system (3) has a unique nontrivial solution in some suitable sphere round the origin in R^N.

We cannot discuss the global problem by the above methods. However, the simple example of a nonhomogeneous polynomial system

$$x_1 = x_1 x_2, \qquad x_2 = \tfrac{1}{2}e + \tfrac{1}{2}x_2^2$$

has the infinite set of solutions $x_2 = e$, x_1 arbitrary. This means that we cannot hope to give a global extension to Theorem 2, nor that we can extend the upper bound on the norm of the set of elements $b_{\mathbf{n}}^{(r)}$ without limit in that theorem.

[4] A. Salam, Nuovo Cimento 25, 224 (1962).

[5] M. A. Krasnoselskii, *Topological Methods in the Theory of Non-Linear Integral Equations* (Pergamon Press, Inc., New York, 1964), p. 141.

[6] M. M. Broido, "On the Complete Unitarity Equations for Pion–Pion Scattering," Cambridge University preprint, 1964 (to be published).

5. A GENERALIZATION OF RING STRUCTURE

In order to apply the results of Theorems 1 and 2 to quantum field theory we have to extend the notion of a ring. We introduce the term n-ring for a set of elements $a, b, \cdots$ which is a linear space over the complex numbers, and which has an operation of multiplication for any n of its elements, denoted by $(a_1 \cdots a_n)$. This operation is

(i) multilinear in all its elements:

$$(a_1 \cdots a_{i-1}\alpha a_i a_{i+1} \cdots a_n) = \alpha(a_1 \cdots a_n)$$

(ii) distributive in all its elements:

$$(a_1 \cdots a_{i-1}a_i + a_i' a_{i+1} \cdots a_n) = (a_1 \cdots a_{i-1}a_i \cdots a_n) + (a_1 \cdots a_{i-1}a_i' a_{i+1} \cdots a_n)$$

(iii) associative, in the sense that we may form the product of the $2n - 1$ elements a_i in sets n at a time, all in the same order, and the result always agrees with

$$(a_1 \cdots a_n)a_{n+1} \cdots a_{2n-1}.$$

Evidently any n-ring R is also a $(2n - 1)$-ring, and so on, but we assume there is no $m < n$ so that R is also an m-ring, and the n-product is that generated by a suitable set of m-products.

The notion of an n-ring for $n > 2$ appears to be new in the mathematical literature. A natural reason for this may be that up to now such a concept has not been useful to consider. However, such rings certainly appear in approximations to the equations of quantum field theory[2] and in bootstrap equations.[4]

We define a normed n-ring as an n-ring R which is a normed linear space for which the norm $|x|$ of x satisfies

$$|x_1 \cdots x_n| \leq |x_1| \cdot |x_2| \cdots |x_n|. \tag{12}$$

Then a complete normed n-ring will be called a Banach n-ring.

We may extend Theorems 1 and 2 of the previous two sections to the case of polynomial systems on a Banach m-ring, where the terms on the right-hand sides of (3) and (5) only involve products defined in terms of the m-products, i.e., $|\mathbf{n}|$ only takes values $m, 2m - 1, 4m - 3, \cdots$, in the summations.

We hope to return elsewhere to an analysis of the algebraic and topological properties of n-rings.

6. A MODEL FROM QUANTUM FIELD THEORY

We consider the application of our two theorems to a simple model arising in quantum field theory. We consider a self-interacting-scalar particle with $\sum_{i=1}^{3} p_i = 0$ corresponds to energy–momentum conservation. We consider these variables as one dimensional, and in a limited region of the real axis, say the interval $(-a, a)$. We require that M is a symmetric function of its variables. We define a product of these symmetric functions of three variables as

$$M_1M_2M_3(p_1p_2p_3) = \sum \int dp_4\, dp_5\, \delta(p_4 + p_5 - p_1) \times M_1(p_1, -p_4, -p_5)M_2(p_3, p_4, -p_3 - p_4) \times M_3(p_2, p_5, -p_2 - p_5), \tag{13}$$

where the summation in (13) is over all permutations of the variables p_1, p_2, p_3. If we introduce the norms on the functions as

$$|M| = \alpha \sup |M|, \tag{14}$$

the supremum being over all values of the variables conserving momentum, then

$$|M_1M_2M_3| \leq |M_1|\,|M_2|\,|M_3|$$

if $\alpha > (2a)^{\frac{1}{2}}$. Hence we may regard the set of symmetric functions of three variables satisfying momentum conservation, furnished with the product of any triple of elements (13) and with a norm (14) with $\alpha > (2a)^{\frac{1}{2}}$, as a Banach 3-ring. The triple product of (13) is not associative in the sense of Sec. 4. However, if we consider the polynomial equation, for some constant k and fixed vertex function $M_{(0)}$,

$$M = M_{(0)} + kM^3, \tag{15}$$

then the proofs of Theorems 1 and 2 go through. This is because the three products of a given element M of the 3-ring define an associative 3-ring. Hence we have the results that if $M_{(0)} \equiv 0$, the homogeneous (bootstrap) equation (15) has no solution except the trivial one, inside a certain sphere with center the origin. If $M_{(0)} \not\equiv 0$, then the inhomogeneous equation (15) has a unique solution inside a certain sphere with center the origin. We will discuss elsewhere[7] the more interesting infinite polynomial systems which represent realistic Lagrangian field equations.

ACKNOWLEDGMENT

I would like to thank M. M. Broido for stimulating discussions on this and related questions.

[7] J. G. Taylor, "On the Existence of Field Theory," Rutgers University Preprints (unpublished).

Effect of n-Spin-Wave Interaction on the Low-Temperature Spontaneous Magnetization*

TOHRU MORITA
The Catholic University of America, Washington, D. C.
AND
TOMOYASU TANAKA
The Catholic University of America, Washington, D. C., and U. S. Naval Ordnance Laboratory, White Oak, Maryland
(Received 11 August 1964)

For a Heisenberg model of a ferromagnet, it is known that the commutator of the Hamiltonian $\mathcal{H}$ and the operator $S_0^- = \sum_j S_j^-$ is proportional to S_0^-. From this fact, a conjecture is made on the low-lying energy levels of two-, three-, $\cdots$, n-spin-wave problems. With the aid of the conjecture and the assumption that there is no low-lying n-spin-wave bound state, it is concluded that the contributions to the low-temperature expansion of the spontaneous magnetization due to two-, three-, $\cdots$, n-spin-wave problems are of $O(T^4)$, $O(T^{13/2})$, $\cdots$, $O(T^{6n/2-1})$, respectively.

THE low-temperature behavior of a Heisenberg ferromagnet is described well by the theory of spin waves introduced by Bloch.[1] The leading correction to the ideal spin-wave approximation was first evaluated by Dyson.[2] Dyson solved the two-spin-wave problem and calculated the leading correction of order T^4 for the spontaneous magnetization. He discusses the correction due to the three-spin-wave problem and showed that it is of order $T^{13/2}$; though he was not certain about $O(T^6)$. The purpose of this paper is to give a general discussion of the order of magnitude of the corrections due to two-, three-, ..., n-spin wave problems. This discussion is based on a simple conjecture on the low-lying energy levels of two-, three-, ..., n-spin-wave problems[3] and a compact virial-expansion formula for an assembly of quasiparticles, obtained quite recently by the present authors.[4]

The Hamiltonian of the Heisenberg ferromagnet is given by

$$\mathcal{H} = -H \sum_f (S_{fz} - S) - \sum_f \sum_g J_{fg} S_g^- S_f^+ - \sum_f \sum_g J_{fg}(S_{fz} S_{gz} - S^2),$$

where two constant terms are added so that the eigenvalue is zero for the ground state $|0\rangle$ where $S_f^+ |0\rangle$ is zero for all lattice sites f; H is the external magnetic field along z direction expressed in a suitable unit, J_{fg} is equal to the exchange integral J when f and g are nearest-neighbor lattice sites and zero otherwise, and $J(0)$ is the value of $J(\mathbf{k})$ for $\mathbf{k} = 0$ where

$$J(\mathbf{k}) = \sum_f J_{fg} \exp i\mathbf{k}\cdot(\mathbf{r}_f - \mathbf{r}_g).$$

An operator $S_{\mathbf{k}}^-$ is introduced by

$$S_{\mathbf{k}}^- = N^{-\frac{1}{2}} \sum_f S_f^- \exp i\mathbf{k}\cdot\mathbf{r}_f,$$

where N is the total number of lattice sites in the system. Then it is easily confirmed that

$$[\mathcal{H}, S_0^-] = HS_0^-.$$

This implies that if $|\ \rangle$ is an eigenstate of $\mathcal{H}$ with eigenvalue E, then $S_0^- |\ \rangle$ is also an eigenstate with eigenvalue $E + H$. In particular,

$$S_0^- |0\rangle, \quad S_0^- S_0^- |0\rangle, \cdots, S_0^{-n} |0\rangle$$

are exact eigenfunctions of $\mathcal{H}$ with eigenvalues $H, 2H, \cdots, nH$, respectively.

The commutation relation between the Hamiltonian $\mathcal{H}$ and $S_{\mathbf{k}}^-$ is given by

$$[\mathcal{H}, S_{\mathbf{k}}^-] = HS_{\mathbf{k}}^- - 2N^{-\frac{1}{2}} \sum_f \sum_i J_{fi} \exp(i\mathbf{k}\cdot\mathbf{r}_i) \times [\exp i\mathbf{k}\cdot(\mathbf{r}_f - \mathbf{r}_i) - 1] S_{fz} S_i^- = HS_{\mathbf{k}}^- + O(\mathbf{k}).$$

If $\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_n$ are very small,

$$S_{\mathbf{k}_1}^- |0\rangle, \quad S_{\mathbf{k}_1}^- S_{\mathbf{k}_2}^- |0\rangle, \cdots, S_{\mathbf{k}_1}^- \cdots S_{\mathbf{k}_n}^- |0\rangle$$

are the eigenfunctions of $\mathcal{H}$ corresponding to the eigenvalues $H, 2H, \cdots, nH$ with errors both of $O(\mathbf{k}_1)$, $O(\mathbf{k}_1, \mathbf{k}_2)$, $\cdots$, $O(\mathbf{k}_1, \cdots, \mathbf{k}_n)$, respectively.[5] Denoting the errors by $\varphi(\mathbf{k}_1, \cdots, \mathbf{k}_n)$ and $\epsilon(\mathbf{k}_1, \cdots, \mathbf{k}_n)$, the eigenfunctions and eigenvalues are expressed as

* This work was supported in part by the Air Force Office of Scientific Research Grant AF-AFOSR-445-63.

[1] F. Bloch, Z. Physik **61**, 206 (1930); **74**, 295 (1932).

[2] F. J. Dyson, Phys. Rev. **102**, 1217, 1230 (1956).

[3] It is assumed that there exist no low-lying bound states of n-spin waves; for two-spin waves, this has been established by Wortis. M. Wortis, Phys. Rev. **132**, 85 (1963).

[4] T. Morita and T. Tanaka, Phys. Rev. **138**, A 1088 (1965).

[5] $O(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \cdots, \mathbf{k}_n) = O(\mathbf{k}_1) + O(\mathbf{k}_2) + \cdots + O(\mathbf{k}_n) + O(\mathbf{k}_1)O(\mathbf{k}_2) + \cdots + O(\mathbf{k}_1)O(\mathbf{k}_2)O(\mathbf{k}_3) + \cdots + O(\mathbf{k}_1)O(\mathbf{k}_2) \cdots O(\mathbf{k}_n)$.

$$\Psi(\mathbf{k}_1) = S^-_{\mathbf{k}_1}|0\rangle + \varphi(\mathbf{k}_1),$$

$$\Psi(\mathbf{k}_1, \mathbf{k}_2) = S^-_{\mathbf{k}_1}S^-_{\mathbf{k}_2}|0\rangle + S^-_{\mathbf{k}_1}\varphi(\mathbf{k}_2) + S^-_{\mathbf{k}_2}\varphi(\mathbf{k}_1) + \varphi(\mathbf{k}_1, \mathbf{k}_2),$$

$$\Psi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = S^-_{\mathbf{k}_1}S^-_{\mathbf{k}_2}S^-_{\mathbf{k}_3}|0\rangle + S^-_{\mathbf{k}_1}S^-_{\mathbf{k}_2}\varphi(\mathbf{k}_3) + S^-_{\mathbf{k}_2}S^-_{\mathbf{k}_3}\varphi(\mathbf{k}_1) + S^-_{\mathbf{k}_3}S^-_{\mathbf{k}_1}\varphi(\mathbf{k}_2) + S^-_{\mathbf{k}_1}\varphi(\mathbf{k}_2, \mathbf{k}_3) + S^-_{\mathbf{k}_2}\varphi(\mathbf{k}_1, \mathbf{k}_3) + S^-_{\mathbf{k}_3}\varphi(\mathbf{k}_1, \mathbf{k}_2) + \varphi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3),$$

$$\Psi(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_n) = S^-_{\mathbf{k}_1}S^-_{\mathbf{k}_2}\cdots S^-_{\mathbf{k}_n}|0\rangle + \sum_{j=1}^{n} S^-_{\mathbf{k}_1}\cdots S^-_{\mathbf{k}_{j-1}}S^-_{\mathbf{k}_{j+1}}\cdots S^-_{\mathbf{k}_n}\varphi(\mathbf{k}_j) + \cdots + \varphi(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_n),$$

$$E(\mathbf{k}_1) = H + \epsilon(\mathbf{k}_1),$$

$$E(\mathbf{k}_1, \mathbf{k}_2) = 2H + \epsilon(\mathbf{k}_1) + \epsilon(\mathbf{k}_2) + \epsilon(\mathbf{k}_1, \mathbf{k}_2),$$

$$E(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 3H + \epsilon(\mathbf{k}_1) + \epsilon(\mathbf{k}_2) + \epsilon(\mathbf{k}_3) + \epsilon(\mathbf{k}_1, \mathbf{k}_2) + \epsilon(\mathbf{k}_2, \mathbf{k}_3) + \epsilon(\mathbf{k}_3, \mathbf{k}_1) + \epsilon(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3),$$

$$E(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_n) = nH + \sum_{i=1}^{n} \epsilon(\mathbf{k}_i) + \sum_{n \geq j > l \geq 1} \epsilon(\mathbf{k}_j, \mathbf{k}_l) + \cdots + \epsilon(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_n).$$

Quantities $\varphi(\mathbf{k}_1, \cdots, \mathbf{k}_n)$ and $\epsilon(\mathbf{k}_1, \cdots, \mathbf{k}_n)$ vanish if one of the arguments is zero, because a state in which one of the arguments is zero represents an exact eigenstate obtained by operating S^-_0 on an eigenstate with one less arguments and hence there is no need for corrections. Note that if $\mathbf{k}_1 = 0$ in $\Psi(\mathbf{k}_1)$ and $E(\mathbf{k}_1)$, then

$$\Psi(0) = S^-_0|0\rangle + \varphi(0), \qquad E(0) = H + \epsilon(0),$$

where $S^-_0|0\rangle$ is an exact eigenfunction which belongs to the eigenvalue H; hence $\varphi(0) = 0$, and $\epsilon(0) = 0$. Now if $\mathbf{k}_2 = 0$ in $\Psi(\mathbf{k}_1, \mathbf{k}_2)$ and $E(\mathbf{k}_1, \mathbf{k}_2)$, then

$$\Psi(\mathbf{k}_1, 0) = S^-_0[S^-_{\mathbf{k}_1}|0\rangle + \varphi(\mathbf{k}_1)] + \varphi(\mathbf{k}_1, 0) = S^-_0\Psi(\mathbf{k}_1) + \varphi(\mathbf{k}_1, 0),$$

$$E(\mathbf{k}_1, 0) = 2H + \epsilon(\mathbf{k}_1) + \epsilon(\mathbf{k}_1, 0) = H + E(\mathbf{k}_1) + \epsilon(\mathbf{k}_1, 0),$$

where $S^-_0\Psi(\mathbf{k}_1)$ is also an exact eigenfunction whose eigenvalue is $H + E(\mathbf{k}_1)$; hence $\varphi(\mathbf{k}_1, 0) = 0$ and $\epsilon(\mathbf{k}_1, 0) = 0$. Then by mathematical induction, the statement is true. Thus orders of magnitude of these corrections are estimated to be

$$\varphi(\mathbf{k}_1, \cdots, \mathbf{k}_n) = O(\mathbf{k}_1)\cdots O(\mathbf{k}_n),$$

$$\epsilon(\mathbf{k}_1, \cdots, \mathbf{k}_n) = O(\mathbf{k}_1)\cdots O(\mathbf{k}_n).$$

The order of magnitude of $\epsilon(\mathbf{k}_1, \cdots, \mathbf{k}_n)$ as a function of the size of the system is evaluated as of $O(N^{-n+1})$ under the assumption that there exist no low-lying bound states. The order of magnitude of ϵ's is estimated as follows: As far as the bound state does not appear, the shift of eigenvalue is considered to be due to the change of effective volume for particles due to the potential. For the two-spin-wave problem, the disturbance of the Hamiltonian appears when two spin waves come to the same or nearest-neighbor lattice sites. This changes the effective number of lattice sites for two-spin waves from N^2 by a number of order N. This results in a shift of $O(1/N)$ in the eigenvalue: $\epsilon(\mathbf{k}, \mathbf{k}') = O(1/N)$. In the three-spin-wave problem, $\epsilon(\mathbf{k}, \mathbf{k}', \mathbf{k}'')$ is due to the change of the effective number of lattice sites, caused by the cases when all three come to the same or nearest neighbors of each other; that is, a number of $O(N)$ in N^3. This results in the correction of $O(1/N^2)$ to the energy eigenvalue: $\epsilon(\mathbf{k}, \mathbf{k}', \mathbf{k}'') = O(1/N^2)$. Similarly, $\epsilon(\mathbf{k}_1, \cdots, \mathbf{k}_n)$ is estimated to be of $O(1/N^{n-1})$.

For cubic lattices, we know that

$$\varphi(\mathbf{k}) = 0, \qquad \epsilon(\mathbf{k}) = 2S[J(0) - J(\mathbf{k})] = O(\mathbf{k}^2),$$

$$\varphi(\mathbf{k}_1, \mathbf{k}_2) = O(\mathbf{k}_1\cdot\mathbf{k}_2) + O(\mathbf{k}_1^2, \mathbf{k}_2^2),$$

$$\epsilon(\mathbf{k}_1, \mathbf{k}_2) = O(\mathbf{k}_1\cdot\mathbf{k}_2) + O(\mathbf{k}_1\cdot\mathbf{k}_2)^2 + O(\mathbf{k}_1^2, \mathbf{k}_2^2) + O(\mathbf{k}_1, \mathbf{k}_2)^6.$$

The above eigenfunctions can be specified by a set $\{n_\mathbf{k}\}$ and are expressed as

$$\Psi\{n_\mathbf{k}\} = \prod_\mathbf{k} S^{-n_\mathbf{k}}_\mathbf{k}|0\rangle + \cdots.$$

The corresponding energies are expressed as

$$E\{n_\mathbf{k}\} = \sum_\mathbf{k} n_\mathbf{k}(H + \epsilon(\mathbf{k})) + \sum_{\mathbf{k}\leq\mathbf{k}'} n_\mathbf{k}n_{\mathbf{k}'}\epsilon(\mathbf{k}, \mathbf{k}') + \cdots.$$

The partition function, Z, of this system is given by

$$Z = \sum_{\{n_\mathbf{k}\}} \exp(-\beta E\{n_\mathbf{k}\}), \qquad \beta = 1/k_B T.$$

This is also the grand partition function for an assembly of Bose quasi-particles which has been analyzed in a separate paper.[4] The virial expansion formula obtained there is given by

$$\ln Z = -\beta E\{\rho(\mathbf{k})\} + S\{\rho(\mathbf{k})\}/k_B$$

with

$$S\{\rho(\mathbf{k})\}/k_B = -\sum_{\mathbf{k}} \{\rho(\mathbf{k}) \ln \rho(\mathbf{k}) - [1 + \rho(\mathbf{k})] \ln [1 + \rho(\mathbf{k})]\},$$

where $\rho(\mathbf{k}) = \langle n_{\mathbf{k}} \rangle$. The expression for $\rho(\mathbf{k})$ is obtained by

$$0 = \partial \ln Z/\partial\rho(\mathbf{k})$$

or

$$\rho(\mathbf{k}) = \{\exp \beta[H + \epsilon(\mathbf{k}) + \Delta\epsilon(\mathbf{k})] - 1\}^{-1}$$

where

$$H + \epsilon(\mathbf{k}) + \Delta\epsilon(\mathbf{k}) = \partial E\{\rho(\mathbf{k})\}/\partial\rho(\mathbf{k}) = H + \epsilon(\mathbf{k}) + \sum_{\mathbf{k}_2} \rho(\mathbf{k}_2)\epsilon(\mathbf{k}, \mathbf{k}_2) + \cdots + \frac{1}{(n-1)!} \sum_{\mathbf{k}_2} \cdots \sum_{\mathbf{k}_n} \rho(\mathbf{k}_2) \cdots \times \rho(\mathbf{k}_n)\epsilon(\mathbf{k}, \mathbf{k}_2, \cdots, \mathbf{k}_n) + \cdots .$$

Noticing that

$$\sum_{\mathbf{k}} \rho(\mathbf{k}) = \sum_{\mathbf{k}} \langle n_{\mathbf{k}} \rangle = \frac{\partial \ln Z}{\partial(-\beta H)} = \frac{\partial \ln \mathrm{Tr}\, e^{-\beta\mathcal{H}}}{\partial(-\beta H)} = -\Sigma_f[\langle S_{fz} \rangle - S],$$

one obtains the total magnetization in terms of $\rho(\mathbf{k})$,[6]

$$M = NS - \sum_{\mathbf{k}} \rho(\mathbf{k}).$$

Using the above estimate of $\epsilon(\mathbf{k}_1, \mathbf{k}_2)$, the leading correction to the spin-wave energy $\Delta\epsilon(\mathbf{k})$ is found to be

$$\sum_{\mathbf{k}_2} \rho(\mathbf{k}_2)\epsilon(\mathbf{k}_1, \mathbf{k}_2) \cong O(\mathbf{k}_1^2) \cdot O(T^{5/2}),$$

since $\rho(\mathbf{k}_2)$ is even in $\mathbf{k}_2$ and so only the part which is even in $\mathbf{k}_2$ in $\epsilon(\mathbf{k}_1, \mathbf{k}_2)$ contributes. That part is $O(\mathbf{k}_1^2) \cdot O(\mathbf{k}_2^2)$ and summation over $\mathbf{k}_2$ gives us $O(\mathbf{k}_1^2) \cdot O(T^{5/2})$. If one expands $\rho(\mathbf{k})$ in powers of this correction to $\epsilon(\mathbf{k})$, one obtains a correction $O(T^4)$ to the spontaneous magnetization. In general, the contribution to $\Delta\epsilon(\mathbf{k})$ of the n-spin-wave problem is estimated as

$$\sum_{\mathbf{k}_2} \cdots \sum_{\mathbf{k}_n} \rho(\mathbf{k}_2) \cdots \rho(\mathbf{k}_n)\epsilon(\mathbf{k}, \mathbf{k}_2, \cdots, \mathbf{k}_n) \sim O(\mathbf{k}^2) \cdot O(T^{5(n-1)/2})$$

by noticing the fact that the part of $\epsilon(\mathbf{k}, \mathbf{k}_2, \cdots, \mathbf{k}_n)$ which is even in $\mathbf{k}_2, \cdots, \mathbf{k}_n$ is $O(\mathbf{k}^2) \cdot O(\mathbf{k}_2^2) \cdots O(\mathbf{k}_n^2)$. This results in the correction $O(T^{3/2}T^{5(n-1)/2}) = O(T^{5n/2-1})$ to the spontaneous magnetization.

As a result, the corrections to the spontaneous magnetization are found to be

$$O(T^4), \quad O(T^{13/2}), \quad O(T^9), \quad O(T^{5n/2-1})$$

for two-, three-, four- and n-spin-wave problems, respectively. This means that the coefficients of T^5 and T^6 would be calculated by solving the two-spin-wave problem by generalizing Dyson's calculation of the coefficient of T^4; that is, by determining the eigenfunction $\rho(\mathbf{k}_1, \mathbf{k}_2)$ correctly up to $O(\mathbf{k}_1, \mathbf{k}_2)^4$ or $O(\mathbf{k}_1, \mathbf{k}_2)^6$ and correspondingly $\epsilon(\mathbf{k}_1, \mathbf{k}_2)$ correctly up to $O(\mathbf{k}_1, \mathbf{k}_2)^6$ or $O(\mathbf{k}_1, \mathbf{k}_2)^8$. The coefficients of $O(T^{9/2})$ and $O(T^{11/2})$ are equal to those given by the ideal spin-wave approximation. Actually the coefficients of T^7 and T^8 would also be calculated by solving the two-spin-wave problem, but these are not significant unless the first two terms of three spin-wave contribution are known.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Kishin Moorjani for his help in preparing the manuscript.

APPENDIX

The expression for the total magnetization $M = \sum_f \langle S_{fz} \rangle$ in terms of $\rho(\mathbf{k})$ is found on a basis of a general property of the Hamiltonian as follows: The Hamiltonian $\mathcal{H}$ consists of two terms, the Zeeman term, $H \sum_f S_{fz}$, and the exchange term plus a constant. These terms commute with each other and one of them, the Zeeman term, appears with an arbitrary factor H, so that an eigenfunction of $\mathcal{H}$ is an eigenfunction of both of these operators. This implies that an eigenfunction of $\mathcal{H}$ is expressed as a linear combination of eigenfunctions of operator $\sum_f S_{fz}$, all belonging to the same eigenvalue. It has been shown that the eigenfunction of $\mathcal{H}$ is specified by $\{n_{\mathbf{k}}\}$ and it has a term $\prod_{\mathbf{k}} S_{\mathbf{k}}^{n_{\mathbf{k}}} |0\rangle$. Now it can be seen that this is an eigenfunction of $\sum_f S_{fz}$ with eigenvalue $NS - \sum_{\mathbf{k}} n_{\mathbf{k}}$, therefore the expectation value as well as the eigenvalue of $\sum_f S_{fz}$ in the eigenstate $\Psi\{n_{\mathbf{k}}\}$ is equal to $NS - \sum_{\mathbf{k}} n_{\mathbf{k}}$. Hence the equilibrium magnetization $M = \sum_f \langle S_{fz} \rangle$ is given by

$$M = NS - \sum_{\mathbf{k}} \rho(\mathbf{k}).$$

The derivation given in the text is based on the fact that the energy corrections $\epsilon(\mathbf{k}_1, \mathbf{k}_2, \cdots)$ are independent of H. Now the eigenfunction of $\mathcal{H}$ is a simultaneous eigenfunction of both the total magnetization $\sum_f S_{fz}$ and the exchange term, hence the eigenvalue is the sum of eigenvalues of the Zeeman term and the exchange term. Therefore the energy corrections $\epsilon(\mathbf{k}_1, \mathbf{k}_2, \cdots)$ which are due to the exchange term are independent of H.

[6] See Appendix for a more detailed derivation.

Proof and Refinements of an Inequality of Feynman

KURT SYMANZIK
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 17 December 1964)

A lower bound given by Feynman for the quantum mechanical free energy of an oscillator is proved, refined, and generalized. The method, application of a classical inequality to path integrals, also gives upper bounds for one-temperature Green's functions.

FOR an oscillator with Hamiltonian

$$H = p^2/2m + V(q), \qquad qp - pq = i\hbar, \tag{1}$$

the inequality for the Helmholtz free energy (with $\beta = (kT)^{-1}$)

$$F \equiv -\beta^{-1} \ln \operatorname{Tr} e^{-\beta H}$$
$$\geq \tfrac{1}{2}\beta^{-1} \ln (2\pi\beta\hbar^2/m) - \beta^{-1} \ln \left(\int e^{-\beta V(x)}\, dx \right) \tag{2}$$

has been given by Feynman[1] on the basis of an intuitive argument relating to the path integral representing $\exp(-\beta F)$. In this note, we prove Eq. (2), show how it may be sharpened, and compare it with corresponding upper bounds for F.

The one-temperature Green's function for a system of N distinguishable nonrelativistic spinless particles of mass m may be represented by the Feynman–Kac integral[2]

$$\langle \mathbf{x}_1 | e^{-\beta H} | \mathbf{x}_2 \rangle = \int_{\substack{\mathbf{x}(0)=\mathbf{x}_1 \\ \mathbf{x}(\beta)=\mathbf{x}_2}} \mathfrak{D}(\mathbf{x}) \times \exp\left[-\frac{m}{2\hbar^2} \int_0^\beta \dot{\mathbf{x}}(\tau)^2\, d\tau \right] \times \exp\left[-\int_0^\beta V(\mathbf{x}(\tau))\, d\tau \right]. \tag{3}$$

Here $\mathbf{x}(\tau)$ is a continuous trajectory in $3N$-space with $\tau \in [0, \beta]$ (dimensionally, $\hbar\tau$) the "time" variable. $V(\mathbf{x})$ is the potential assumed position-dependent only. $\mathfrak{D}(\mathbf{x})$ together with the first exponential represents Wiener measure.

We choose a comparison potential $V_0(\mathbf{x})$ and write, with $\Delta V = V - V_0$, the second exponential in (3) as

$$\exp\left[-\int_0^\beta V_0(\mathbf{x}(\tau))\, d\tau \right] \times \exp\left[-\int_0^\beta \Delta V(\mathbf{x}(\sigma))\, d\sigma \right]. \tag{4}$$

We combine the first term in (4) with the first exponential in (3) and apply to the integral so obtained over the second term in (4) the generalized Hölder inequality[3]

$$\int_{S_1} \exp\left\{ \int_S \ln |K(s, s_1)|\, \mu(ds) \right\} \mu_1(ds_1) \leq \exp\left\{ \int_S \left[\ln \left(\int_{S_1} |K(s, s_1)|\, \mu_1(ds_1) \right] \mu(ds) \right\}, \tag{GH}$$

valid for $\mu(S) = 1$, interpreting s_1 as trajectory parameter and s as σ/β. Thereupon (GH) becomes

$$\langle \mathbf{x}_1 | e^{-\beta H} | \mathbf{x}_2 \rangle \leq \exp\left\{ \beta^{-1} \int_0^\beta d\sigma \ln \left[\int_{\substack{\mathbf{x}(0)=\mathbf{x}_1 \\ \mathbf{x}(\beta)=\mathbf{x}_2}} \mathfrak{D}(\mathbf{x}) \times \exp\left\{ -\frac{m}{2\hbar^2} \int_0^\beta \dot{\mathbf{x}}(\tau)^2\, d\tau - \int_0^\beta V_0(\mathbf{x}(\tau))\, d\tau \right\} \times \exp\{ -\beta \Delta V(\mathbf{x}(\sigma)) \} \right] \right\}$$

or, if we make use of (3), of the completeness of states in coordinate representation (i.e., the definition of the Wiener integral), and abbreviate $H - V + V_0 \equiv H_0$,

$$\langle \mathbf{x}_1 | e^{-\beta H} | \mathbf{x}_2 \rangle \leq \exp\left\{ \beta^{-1} \int_0^\beta d\sigma \ln \left[\int d\mathbf{x} \langle \mathbf{x}_1 | e^{-\sigma H_0} | \mathbf{x} \rangle \times \exp[-\beta \Delta V(\mathbf{x})] \langle \mathbf{x} | e^{-(\beta-\sigma) H_0} | \mathbf{x}_2 \rangle \right] \right\} \tag{5}$$

or, in brief

$$\langle \mathbf{x}_1 | e^{-\beta H} | \mathbf{x}_2 \rangle \leq \exp\left\{ \beta^{-1} \int_0^\beta d\sigma \ln \langle \mathbf{x}_1 | e^{-\sigma H_0} e^{-\beta \Delta V} e^{-(\beta-\sigma) H_0} | \mathbf{x}_2 \rangle \right\}.$$

(5) is an upper bound for the perturbed Green's

[1] R. P. Feynman, *Lectures on Statistical Mechanics* (Rand Corporation, 1959) (unpublished).
[2] R. P. Feynman, thesis, Princeton (1942). M. Kac, Trans. Amer. Math. Soc. 65, 1 (1949).
[3] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience Publishers, Inc., New York, 1958), Vol. I, p. 535.

function in terms of the unperturbed one and the perturbing potential ΔV. This bound may be contrasted with the lower bound

$$\langle \mathbf{x}_1| e^{-\beta H} |\mathbf{x}_2\rangle \geq \langle \mathbf{x}_1| e^{-\beta H_0} |\mathbf{x}_2\rangle \exp\left\{-\langle \mathbf{x}_1| e^{-\beta H_0} |\mathbf{x}_2\rangle^{-1} \times \int_0^\beta d\sigma \int d\mathbf{x} \langle \mathbf{x}_1| e^{-\sigma H_0} |\mathbf{x}\rangle \Delta V(\mathbf{x}) \langle \mathbf{x}| e^{-(\beta-\sigma)H_0} |\mathbf{x}_2\rangle\right\} \quad (6)$$

or, in brief

$$\langle \mathbf{x}_1| e^{-\beta H} |\mathbf{x}_2\rangle \geq \langle \mathbf{x}_1| e^{-\beta H_0} |\mathbf{x}_2\rangle \exp\left\{-\langle \mathbf{x}_1| e^{-\beta H_0} |\mathbf{x}_2\rangle^{-1} \times \int_0^\beta d\sigma \langle \mathbf{x}_1| e^{-\sigma H_0} \Delta V e^{-(\beta-\sigma)H_0} |\mathbf{x}_2\rangle\right\}$$

obtained by applying to (3), (4) the inequality of the arithmetic and geometric means[3,4]

$$\int_S |f(s)|\, \mu(ds) \geq \exp\left\{\int_S \ln |f(s)|\, \mu(ds)\right\}, \quad \text{(AG)}$$

valid for $\mu(S) = 1$, which was exploited by Feynman.[1] That the right-hand side of (6) is smaller than the right-hand side of (5) can be seen directly by applying (AG) on the $\mathbf{x}$-integration.

The trace of (5) gives an upper bound for the partition function. To obtain a simpler although less sharp bound, one may apply (AG) to the σ-integration in (5), obtaining

$$\langle \mathbf{x}_1| e^{-\beta H} |\mathbf{x}_2\rangle \leq \frac{1}{\beta} \int_0^\beta d\sigma \int d\mathbf{x} \langle \mathbf{x}_1| e^{-\sigma H_0} |\mathbf{x}\rangle \times e^{-\beta \Delta V(\mathbf{x})} \langle \mathbf{x}| e^{-(\beta-\sigma)H_0} |\mathbf{x}_2\rangle, \quad (7)$$

and taking the trace gives

$$\operatorname{Tr} e^{-\beta H} \leq \int d\mathbf{x} \langle \mathbf{x}| e^{-\beta H_0} |\mathbf{x}\rangle e^{-\beta \Delta V(\mathbf{x})} = \operatorname{Tr}(e^{-\beta H_0} e^{-\beta \Delta V}). \quad (8)$$

Setting in (8) $V_0 = 0$, $\Delta V = V$ gives for the oscillator (1) the inequality (2). The upper bound obtained from (5) by the same substitution is, of course, lower.

Taking the trace in (6) and, to simplify although hereby worsening the bound, applying (AG) to the $x_1 = x_2$ integration gives

$$\operatorname{Tr} e^{-\beta H} \geq \operatorname{Tr} e^{-\beta H_0} \times \exp\{-[\operatorname{Tr} e^{-\beta H_0}]^{-1} \beta \operatorname{Tr}(e^{-\beta H_0} \Delta V)\}, \quad (9)$$

[4] (GH) may be derived from (AG) by substituting for $f(s)$ the expression $K(s, s_1)[\int_{s_1} |K(s, s_1)\mu(ds_1)]^{-1}$ and integrating over s_1.

the well-known inequality of Bogoliubov.[5] For the same H_0 as in (9) the Peierls inequality[6] always gives a better bound as follows from (AG), this last bound, however, seems not to be easy to compare with the trace of (6). That the right-hand side of (9) is smaller than the right-hand side of (8) follows from (AG).

The bounds (5), (7), and (8) may be optimized by varying parameters in V_0 in the same manner as lower bounds are. For $\beta \to \infty$ all these bounds give the same inequalities for the ground-state energy E_G, with ground state $|G_0\rangle$ to H_0,

$$E_{G_0} + \min_{\mathbf{x}} \Delta V(\mathbf{x}) \leq E_G \leq E_{G_0} + \langle G_0| \Delta V |G_0\rangle.$$

Also path integrals involving a potential term depending on two "times" instead of one as they arise, e.g., in the polaron problem[7] may be treated by an obvious extension [e.g., taking in (GH) the s-integration over a two-dimensional domain] of the method shown here. Also V_0 may be chosen nonlocal in "time" provided (GH) can still be applied and the expressions required in (5), (7), or (8) be calculated. This is the case (up to elementary analysis at least) for V_0 bilinear in x. However, these methods seem not to give in the polaron problem[7] a finite lower bound for the ground-state energy in a simple manner for the Hamiltonian chosen in Ref. 7.

Note added in proof: S. Golden [Phys. Rev. **137**, B 1127 (1965)] has shown that the inequality (8), together with a class of refinements of it, holds for general Hermitian H_0 and ΔV. The above method is easily modified to yield an extension of Golden's result to Green's functions for the present case of classical and for Bose–Einstein statistics. It is hereby seen that symmetrizing the bound for the Green's function of classical statistics always yields a result at least as good as that obtained using Bose–Einstein statistics throughout. This remark also applies to the following other generalization (for both statistics): Write the semigroup formula for $\langle \mathbf{x}_1| e^{-\beta H} |\mathbf{x}_2\rangle$ for an arbitrary partition of the interval $[0, \beta]$ and insert the bounds (5), (6), and (7). The upper and lower bounds so obtained monotonically decrease or increase, respectively, if the partition is refined.

[5] N. N. Bogoliubov, Dokl. Akad. Nauk. SSSR **119**, 244 (1958) [English transl.: Soviet Phys.—Doklady **3**, 292 (1958)].
[6] R. E. Peierls, Phys. Rev. **54**, 918 (1938).
[7] R. P. Feynman, Phys. Rev. **97**, 660 (1955).

Singularities of Analytic Functions Having Integral Representations, with a Remark about the Elastic Unitarity Integral*

R. P. GILBERT, H. C. HOWARD, AND S. AKS
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
(Received 15 October 1964)

In this paper a survey is given of some results which have been obtained recently concerning the singularities of holomorphic functions having integral representations. These results are all essentially extensions or modifications of those developed by Hadamard (for the proof of his multiplication of singularities theorem) to the case of several complex variables. As a concluding remark we consider the connection between the original Hadamard idea and the elastic unitarity integral of the quantum theory of fields.

I. INTRODUCTION

THE classical problem of the connection between the singularities of the holomorphic function $K(z;\zeta)$ and the contour integral $F(z) = \int_{\mathcal{L}} K(z;\zeta)\,d\zeta$ has found application in the quantum theory of fields. When $K(z;\zeta)$ is a rational function of the complex variables $(z;\zeta) \in \mathbf{C}^m \times \mathbf{C}^n$ [i.e., when $K(z;\zeta) = P(z;\zeta)/Q(z;\zeta)$ with $P(z;\zeta)$ and $Q(z;\zeta)$ polynomials] and $\mathcal{L}$ denotes the integration domain $(\mathcal{L}_1, \cdots, \mathcal{L}_n)$ in $\mathbf{C}^n$, Landau[1,2] has shown that $F(z)$ is singular only on the set

$$\left\{z \mid Q(z;\zeta) = \frac{\partial}{\partial\zeta_1} Q(z;\zeta) = \frac{\partial}{\partial\zeta_2} Q(z;\zeta) = \cdots = \frac{\partial}{\partial\zeta_n}(z;\zeta) = 0\right\}.$$

In this paper we generalize this result to a somewhat wider class of holomorphic functions, the holomorphic functions singular on analytic sets. It will be shown that many of the qualitative concepts associated with the (rational) propagators of perturbation theory have a natural transcription to this wider class.

The results are collected in a series of theorems starting with Hadamard's treatment of the holomorphic functions[3] determined by germs of the form $\sum_{n=0}^{\infty} a_n b_n z^n$ and go on to the physically interesting multiple integrals. Some of the results are already within the arsenal of the quantum field theorists in one form or another[4]; but, particularly in the case of the multiple integrals, the problem has not been precisely formulated. This we propose to do. A number of the theorems are tailored to special requirements of the quantum theory of fields, and potential scattering,[5] notably, Theorems 4 and 5 which are valuable tools for studying the unitarity integral. We conclude with a remark on the connection between the original Hadamard idea and the unitarity integral.

II. SINGULARITIES OF ANALYTIC FUNCTIONS WITH INTEGRAL REPRESENTATIONS

Hadamard[6] proved the following result in 1898 (the multiplication of singularities theorem[7]).

Theorem 0: If

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,\ |z| < R,$$

$$g(z) = \sum_{n=0}^{\infty} b_n z^n,\ |z| < R';$$

furthermore if $f(z)$ has singularities at $\alpha_1, \alpha_2, \cdots$, and $g(z)$ has singularities at $\beta_1, \beta_2, \cdots$ then the singularities of $F(z) = \sum a_n b_n z^n$ are to be found at the points $\alpha_m\beta_n$.

To prove this Hadamard considers the following representation for $F(z)$:

$$F(z) = \frac{1}{2\pi i}\int_C f(\zeta)g(z/\zeta)\,d\zeta/\zeta,$$

where C is a simple contour lying in the annulus $|z|/R' < |\zeta| < R$. This gives us a representation for $F(z)$ in a neighborhood of the origin. This rep-

* This research was supported in part by the National science Foundation under Grants GP-3, GP-2067, and GP-3937, and also in part by the U. S. Air Force Office of Scientific Research under Grant AFOSR 400-64.

[1] D. Landau, Nucl. Phys. 13, 181 (1959).

[2] D. Landau, Zh. Eksperim. i Teor. Fiz. 37, 62 (1959) [English transl.: Soviet Phys.—JETP 10, 45 (1960)].

[3] J. Hadamard, Acta Math. 22, 55 (1898).

[4] For example see: G. R. Screaton, *Dispersion Relations* (Oliver & Boyd, Ltd., Edinburgh, 1961), p. 65; G. Källén, Nucl. Phys. 25, 568 (1961).

[5] S. Ø. Aks, R. P. Gilbert, and H. C. Howard, Technical Note BN-376, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Md., J. Math. Phys. (to be published).

[6] J. Hadamard, Acta Math. 22, 55 (1898).

[7] P. Dienes, *The Taylor Series* (Oxford University Press, Oxford, England, 1931).

resentation may be continued analytically along a curve Γ to z_1, provided no point of Γ coincides with a singularity of the integrand on the path of integration. Hadamard notes that this representation may be continued further along Γ by continuously deforming C, such that C at no time crosses a singularity of the integrand. Since the singularities of $f(z)$ are at $\{\alpha_m\}$ and those of $g(z/\zeta)$ are at $\{z/\beta_n\}$ we see that it is always possible to make such a deformation except when a point z/β_n coincides with a point α_m, that is for $z = \alpha_m\beta_n$. For further details see Ref. 8.

Gilbert[9] has modified the idea used by Hadamard in the proof of the above theorem to prove a theorem concerning the location of singularities of a harmonic function of three complex variables. We give below a theorem for holomorphic functions of n complex variables whose proof is an immediate extension of the proof given for the case of harmonic functions.[9,10]

Theorem 1: Let $K(z;\zeta)$ be an analytic function of $(n+1)$ variables $(z;\zeta) \equiv (z_1, \cdots, z_n;\zeta)$. Furthermore, let $K(z;\zeta)$ be regular-analytic in a product domain $(\mathfrak{D}^{(n)} \times \mathfrak{B}^{(1)}) \subset (\mathbf{C}^n \times C^1)$, such that, for a rectifiable contour $\mathfrak{L} \subset \mathfrak{B}^{(1)}$, the function $F(z)$ defined by the integral representation

$$F(z) \equiv \frac{1}{2\pi i}\int_{\mathfrak{L}} K(z;\zeta)\,\frac{d\zeta}{\zeta},$$

$$\mathfrak{L} \equiv \{\zeta \mid \zeta = \zeta(t) \neq 0; 0 \leq t \leq 1\} \qquad (2.1)$$

is regular-analytic in $\mathfrak{D}^{(n)}$. Then if the singularities of $K(z;\zeta)$ are located on the analytic set $\mathfrak{S}^{(n)} \equiv \{S(z;\zeta) = 0\}$, $F(z)$ is regular for all points $(z) \notin \mathfrak{S}^{(n)} \cap \mathfrak{Q}^{(n)}$,[11] where $\mathfrak{Q}^{(n)} \equiv \{\partial S/\partial\zeta(z;\zeta) = 0\}$.

Proof: Since we assume $F(z)$ is regular-analytic in a domain $\mathfrak{D}^{(n)}$, we may choose a point $(z_0) \in \mathfrak{D}^{(n)}$ and a neighborhood of (z_0), $\mathfrak{N}^{(n)}(z_0)$, in which we may define a regular function element, or "germ" of $F(z)$. Starting with this germ we may continue it analytically along a contour $\mathfrak{C}^{(1)}_{(z)}$ starting with (z_0) and terminating with the point (z) providing no point of $\mathfrak{C}^{(1)}_{(z)}$ corresponds to a singularity of the integrand on the path of integration, $\mathfrak{L}$. We shall refer to the union of all such points (z) as the *initial domain of definition* of $F(z)$. Now, since the singularities of $K(z;\zeta)$ lie on $\mathfrak{S}^{(n)}$ we realize that, as we continue $F(z)$ along $\mathfrak{C}^{(1)}_{(z)}$, the singularities of the integrand move in the ζ-plane. Hence, it is clear that the initial domain of definition must consist of all those points (z) which are reached without a singularity in the ζ-plane passing over $\mathfrak{L}$. We note, however, that we may extend the integral representation for $F(z)$ to other points by continuously deforming $\mathfrak{L} \to \mathfrak{L}'$, provided that in so doing we do not cross a singularity in the ζ-plane. All such points (z) reached by continuing $F(z)$ along paths $\mathfrak{C}^{(1)}_{(z)}$ which can be reached by suitably deforming the path of integration we shall refer to as the *domain of association* of $F(z)$. It is clear that the domain of association consists of the set of all points at which there exists a regular germ of $F(z)$, and furthermore, if one considers all such possible paths $\mathfrak{C}^{(1)}_{(z)}$, one constructs in this way all the regular germs associated with $F(z)$.

We now consider when it may no longer be possible to continue $F(z)$ along a path $\mathfrak{C}^{(1)}_{(z_1)}$. This case occurs [assuming we can solve $S(z,\zeta) = 0$ as a function of z, $\zeta = \alpha(z)$, say] when a singularity $\zeta = \alpha(z)$ tends to cross the contour $\mathfrak{L}$ and we may no longer deform $\mathfrak{L}$ to avoid it. We assume we have continued $F(z)$ along $\mathfrak{C}^{(1)}_{(z_1)}$ up to a point (z_1), and that at $(z) = (z_1)$ there exists a singularity $\zeta = \alpha$ on the path of integration. If $S(z_1;\zeta)$ has a simple zero at $\zeta = \alpha$, we may approximate

$$S(z_1;\zeta) \approx (\zeta - \alpha)\,\partial S/\partial\zeta(z_1;\alpha)$$

in a suitably small neighborhood $\mathfrak{N}(\alpha) \equiv \{\zeta \mid |\zeta - \alpha| < \epsilon\}$. In this case, it is clear that for $\zeta \in \mathfrak{N}(\alpha)$, $S(t;\zeta) \neq 0$ except for $\zeta = \alpha$, and we may deform $\mathfrak{L}$ about the point $\zeta = \alpha$ by letting it follow a portion of the circle $\{|\zeta - \alpha| = \frac{1}{2}\epsilon\}$. This completes the proof of our theorem.

Remark 1: Note that any singularities of $K(z,\zeta)$ contained in the boundary of the holomorphy domain of K can be included in a topological sum of analytic sets, so the singularities arising from these points will be accounted for by Theorem 1.

Remark 2: One might give this result a geometric interpretation by saying, that unless one is on the "envelope" $\mathfrak{E}^{(n)} \equiv \mathfrak{S}^{(n)} \cap \mathfrak{Q}^{(n)}$ of $\mathfrak{S}^{(n)}$ it is always possible to avoid a singularity by slightly varying $\mathfrak{L}$.

Remark 3: If $S(z_1;\zeta) = 0$ has a second- (or higher-) order zero at $\zeta = \alpha$, then we know that the inverse of the function $w = S(z_1;\zeta)$, i.e., $\zeta = f_\nu(w;z_1)$ $(\nu = 1, 2)$ has (at least) two branches in a suitably small neighborhood of $w = 0$. In this case it is *possible* that both branches of the singularity manifold may pinch together on opposite sides of $\mathfrak{L}$, and then, according to the original Hadamard idea of singu-

[8] See Ref. 7.
[9] R. P. Gilbert, Pacific J. Math. **10**, 1243 (1960).
[10] R. P. Gilbert, J. Reine Angew. Math. **205**, 75 (1960); Arch. Ratl. Mech. Anal. **6**, 171 (1960); J. Math. Phys. **5**, 983 (1964).
[11] Throughout this paper the expression $A \cap B$, where $A = \{(z,\zeta) \mid F(z,\zeta) = 0\}$ and $B = \{(z,\zeta) \mid G(z,\zeta) = 0\}$ is taken to be $\{z \mid$ there exists ζ, $F(z,\zeta) = G(z,\zeta) = 0\}$.

larities that "pinch," this corresponds to a possible singularity of $F(z)$ at $(z) = (z_1)$.

Remark 4: To show that this theorem is not empty we mention that it applies to kernels, $K(z; \zeta)$, that are meromorphic on $(n + 1)$-dimensional, finitely sheeted, analytically ramified covering spaces[12] of a domain $\mathfrak{D}^{(n+1)} \subset \mathbf{C}^{(n+1)}$. For example, we may consider for kernels rational functions, $K(z_1, \cdots, z_n; \zeta)$, defined on an *algebraic manifold* given by an expression of the form

$$p_0(z_2, \cdots, z_n; \zeta)z_1^N + p_1(z_2, \cdots, z_n; \zeta)z_1^{N-1} + \cdots + p_N(z_2, \cdots, z_n; \zeta) = 0,$$

where the $p_\nu(z_2, \cdots, z_n; \zeta)$ $(\nu = 0, 1, \cdots, N)$ are polynomials in the variables $z_2, z_3, \cdots, z_n, \zeta$.

Theorem 2: Let $K(z; \zeta)$ be an analytic function of $(n + 1)$-variables as in Theorem 1, with a decomposition in $(\mathfrak{D}^{(n)} \times \mathfrak{B}^{(1)})$ into a product of non-constant, regular-analytic functions $K(z; \zeta) \equiv f_1(z; \zeta)f_2(z; \zeta)$. Furthermore, let the function $f_\nu(z; \zeta)$ $(\nu = 1, 2)$ be singular on the analytic set $\mathfrak{S}_\nu^{(n)} \equiv \{\zeta = \alpha_\nu(z)\}$ $(\nu = 1, 2)$, respectively, with $\mathfrak{S}_\nu^{(n)} \neq \phi$, $\nu = 1, 2$. Then the "possible singularities" of the function

$$F(z) = \frac{1}{2\pi i}\int_{\mathfrak{L}} f_1(z; \zeta)f_2(z; \zeta)\frac{d\zeta}{\zeta}, \tag{2.2}$$

generated from the singularity sets $\mathfrak{S}_\nu^{(n)}$ above, as given by either the Hadamard method or the "envelope method" are the same.

Proof: By "possible singularities" we mean those points (z) which we are unable to list as regular points by either Theorem 0 or Theorem 1. [The phrase "the Hadamard method" used in the theorem above is Gilbert's name[13] for a natural generalization of Hadamard's theorem—Theorem 0 above—which says $F(z)$ is regular at all points $(z) \notin \mathfrak{S}_1^{(n)} \cap \mathfrak{S}_2^{(n)}$ and which follows readily from the technique of proof used for Theorem 0]. We apply the "envelope method" of Theorem 1 by constructing as our singularity manifold for $K(z; \zeta)$ the analytic set $\mathfrak{S}^{(n)} \equiv \{[\zeta - \alpha_1(z)][\zeta - \alpha_2(z)] = 0\}$. Following Theorem 1 we compute $\mathfrak{Q}^{(n)} \equiv \{[\zeta - \alpha_1(z)] + [\zeta - \alpha_2(z)] = 0\}$. Since the singularities must lie on $\mathfrak{S}^{(n)} \cap \mathfrak{Q}^{(n)}$, we conclude this means (z) is regular providing $(z) \notin \{\zeta - \alpha_1(z) = 0\} \cap \{\zeta - \alpha_2(z) = 0\} \equiv \mathfrak{S}_1^{(n)} \cap \mathfrak{S}_2^{(n)}$.

The assumption of "explicit representations" of the singularity sets of $f_\nu(z; \zeta)$, $\nu = 1, 2$, [i.e., $\zeta = \alpha_\nu(z)$, $\nu = 1, 2$] rules out singularities due to coincidences involving $\nu = 1$ or $\nu = 2$ alone. To be more specific, let us consider the envelope method where $f_\nu(z; \zeta)$, $\nu = 1, 2$ is singular on $\mathfrak{S}_\nu^{(n)} = \{(z; \zeta) \mid S_\nu(z; \zeta) = 0\}$ and where $S_\nu(z; \zeta)$ is a holomorphic function. Then if we set

$$\mathfrak{S}^{(n)} = \{(z; \zeta) \mid S_1(z; \zeta)S_2(z; \zeta) = 0\}$$

and

$$\mathfrak{Q}^{(n)} = \{(z; \zeta) \mid \partial/\partial\zeta)S_1S_2 = 0\},$$

it is possible that $F(z)$ may be singular on $\mathfrak{S}_1^{(n)} \cap \mathfrak{Q}_1^{(n)}$ since this set is a subset of $\mathfrak{S}^{(n)} \cap \mathfrak{Q}^{(n)}$. If $S_1 = 0$ but $S_2 \not\equiv 0$ and if $\partial S_1/\partial\zeta = 0$ it is evident that z may satisfy both intersection conditions but not be one of the singularities mentioned in Theorem 2. However, in this case it is not possible to represent $\mathfrak{S}_1^{(n)}$ in the form $\{(z; \zeta) \mid \zeta = \alpha(z)\}$ where the first derivative $\partial S_1/\partial\zeta$ vanishes. The requirement that the singularity sets of $f_\nu(z; \zeta)$, $\nu = 1, 2$, be explicitly represented is equivalent to

$$\bigcup_{\nu=1,2} \mathfrak{S}_\nu^{(n)} \cap \mathfrak{Q}_\nu^{(n)} = \phi.$$

A natural generalization of these results may be made by considering multiple integrals, and combinations of the Hadamard method and envelope method used alternately. As a first attempt in this direction we give the following theorem, which is a form of a result given by Gilbert.[14]

Theorem 3: Let $K(z; \zeta_1, \zeta_2)$ be an analytic function of $(n + 2)$ variables, $(z; \zeta_1, \zeta_2) \equiv (z_1, \cdots, z_n; \zeta_1, \zeta_2)$, regular in a product domain $(\mathfrak{D}^{(n)} \times \mathfrak{B}^{(2)}) \subset (\mathbf{C}^n \times \mathbf{C}^2)$, where $\mathfrak{B}^{(2)} \equiv \mathfrak{B}_1^{(1)} \times \mathfrak{B}_2^{(1)}$. Let $\mathfrak{L}_\nu \subset \mathfrak{B}_\nu^{(1)}$, $(\nu = 1, 2)$, be closed, rectifiable curves in the ζ_ν-planes, respectively, and let $K(z; \zeta_1, \zeta_2)$ be singular on the analytic set $\mathfrak{S}^{(n+1)} \equiv \{S(z; \zeta_1, \zeta_2) = 0\}$. Then, the function $F(z)$ defined by the representation

$$F(z) \equiv \left(\frac{1}{2\pi i}\right)^2 \int_{\mathfrak{L}_1} \frac{d\zeta_1}{\zeta_1} \int_{\mathfrak{L}_2} \frac{d\zeta_2}{\zeta_2} K(z; \zeta_1, \zeta_2), \quad \text{for } (z) \in \mathfrak{D}^{(n)}, \tag{2.3}$$

is regular for all $(z) \notin \mathfrak{S}_1^{(n+1)} \cap \mathfrak{Q}_1^{(n+1)} \cap \mathfrak{Q}_2^{(n+1)}$, where

$$\mathfrak{Q}_\nu^{(n+1)} \equiv \left\{\frac{\partial S}{\partial \zeta_\nu}(z; \zeta_1, \zeta_2) = 0\right\}, \quad (\nu = 1, 2).$$

Proof: Our proof parallels those given before. We consider all points (z) which may be reached by continuing $F(z)$ along a contour $\mathfrak{C}_{(z)}^{(1)}$, such that no

[12] H. Behnke and H. Grauert, "Analysis of Non-Compact Complex Spaces" in *Analytic Functions* (Princeton University Press, Princeton, New Jersey, 1960), p. 11.

[13] R. P. Gilbert, J. Math. Phys. 5, 983 (1964); also R. P. Gilbert (Ref. 9).

[14] R. P. Gilbert, J. Reine Angew. Math. 205, 75 (1960); J. Math. Phys. 5, 983 (1964).

point of this contour corresponds to a singularity of the integrand on the domain of integration $\mathcal{L}_1 \times \mathcal{L}_2$. Again we attempt to enlarge this initial domain of definition by continuously deforming $\mathcal{L}_1$, and $\mathcal{L}_2$ providing we do not let either curve pass over a singularity in its respective plane. (As we deform one of the curves we must be careful, that a singularity does not move over the "stationary" curve in the other plane.) We assume we have been able to continue $F(z)$ to the point $(z) = (z_1)$, where the singularity $(\zeta_1, \zeta_2) = (\alpha_1, \alpha_2)$ threatens to meet $\mathcal{L}_1 \times \mathcal{L}_2$. If both $\partial S/\partial \zeta_\nu(z; \alpha_1, \alpha_2)$ $(\nu = 1, 2)$ are not zero then in a suitably small bi-cylindrical neighborhood of (α_1, α_2), $\mathfrak{N}^{(2)}(\alpha_1, \alpha_2) \equiv \{(\zeta_1, \zeta_2) \mid |\zeta_\nu - \alpha_\nu| < \epsilon_\nu; \nu = 1, 2\}$ we may approximate:

$$S(z_1; \zeta_1, \zeta_2) \approx (\zeta_1 - \alpha_1)\frac{\partial S(z_1; \alpha_1, \alpha_2)}{\partial \zeta_1} + (\zeta_2 - \alpha_2)\frac{\partial S(z_1; \alpha_1, \alpha_2)}{\partial \zeta_2}.$$

In this case we may deform $\mathcal{L}_1$, $\mathcal{L}_2$ about the point (α_1, α_2) in such a manner that $S(z_1; \zeta_1, \zeta_2) \neq 0$ for (ζ_1, ζ_2) on the set

$$\left\{|\zeta_1 - \alpha_1| = \frac{\epsilon_1}{2}\right\} \times \left\{|\zeta_2 - \alpha_2| = \frac{\epsilon_1}{4}\frac{|\partial S/\partial \zeta_1|}{|\partial S/\partial \zeta_2|}\right\}.$$

If one of the terms $\partial S/\partial \zeta_\nu$ $(z; \alpha_1, \alpha_2)$, $(\nu = 1, 2)$ vanishes, then the proof follows that of Theorem 1. This concludes our argument.

We consider next several extensions of this last result. Let us assume first that $K(z; \zeta_1, \zeta_2)$ may be decomposed into the form $K(z; \zeta_1, \zeta_2) \equiv f_1(z; \zeta_1, \zeta_2)f_2(z; \zeta_1, \zeta_2)$, where f_1, f_2 are analytic regular in $(\mathfrak{D}^{(n)} \times \mathfrak{B}^{(2)})$, $\mathfrak{B}^{(2)} = \mathfrak{B}_1^{(1)} \times \mathfrak{B}_1^{(1)}$. Further, let $f_\nu(z; \zeta_1, \zeta_2)$ $(\nu = 1, 2)$ be singular on the analytic set $\mathfrak{S}_{\nu,0}^{(n+1)} \equiv \{S_\nu(z; \zeta_1, \zeta_2) = 0\}$; let $\mathfrak{C}_z^{(1)}$ be a curve in the space of the (z)-variables starting at an initial point $(z_0) \in \mathfrak{D}^{(n)}$, and terminating at (z). We wish to consider now the continuation of the germ [defined about an initial point $(z_0) \in \mathfrak{D}^{(n)}$] given by the integral representation

$$F(z) = \left(\frac{1}{2\pi i}\right)^2 \int_{\mathcal{L}_1}\int_{\mathcal{L}_2} f_1(z; \zeta_1, \zeta_2)f_2(z; \zeta_1, \zeta_2)\, d\zeta_1\, d\zeta_2, \tag{2.4}$$

and to this end we introduce the following notation:

$$\mathfrak{S}_{\nu,1}^{(n+1)} \equiv \left\{\frac{\partial}{\partial \zeta_1} S_\nu(z; \zeta_1, \zeta_2) = 0\right\}, (\nu = 1, 2), \tag{2.5a}$$

$$\mathfrak{S}_{\nu,2}^{(n+1)} \equiv \left\{\frac{\partial}{\partial \zeta_2} S_\nu(z; \zeta_1, \zeta_2) = 0\right\}, (\nu = 1, 2), \tag{2.5b}$$

$$\mathfrak{F}^{(n+1)} \equiv \left\{\frac{\partial S_1}{\partial \zeta_1}\frac{\partial S_2}{\partial \zeta_2} - \frac{\partial S_1}{\partial \zeta_2}\frac{\partial S_2}{\partial \zeta_1} = 0\right\}, \tag{2.5c}$$

$$\mathfrak{Q}_0^{(n+1)} \equiv \{S_1(z; \zeta_1, \zeta_2)S_2(z; \zeta_1, \zeta_2) = 0\}, \tag{2.6a}$$

$$\mathfrak{Q}_1^{(n+1)} \equiv \left\{\frac{\partial S_1}{\partial \zeta_1} S_2 + S_1\frac{\partial S_2}{\partial \zeta_1} = 0\right\}, \tag{2.6b}$$

$$\mathfrak{Q}_2^{(n+1)} \quad \left\{\frac{\partial S_1}{\partial \zeta_2} S_2 + S_1\frac{\partial S_2}{\partial \zeta_2} = 0\right\}. \tag{2.6c}$$

Now let us consider how singularities may occur. Certainly, it is possible for singularities to correspond to the "envelope" of either $\mathfrak{S}_{1,0}^{(n+1)}$ or $\mathfrak{S}_{2,0}^{(n+1)}$. Hence, singular points of $F(z)$ may be contained in the set

$$\bigcup_{\nu=1,2}\left\{\bigcap_{\mu=0,1,2}\mathfrak{S}_{\nu,\mu}^{(n+1)}\right\} \equiv \bigcup_{\nu=1,2}\mathfrak{E}_\nu^{(n-1)} = \mathfrak{E}_{1,2}^{(n-1)},$$

where $\mathfrak{E}_\nu^{n-1}$ is the corresponding "envelope" for $\mathfrak{S}_{\nu,0}^{(n+1)}$. On the other hand the Hadamard approach, that is, eliminating ζ_1 or ζ_2 between the equations $S_1 = 0$ and $S_2 = 0$, applied to either variable ζ_1, or ζ_2 first, followed by the envelope method applied to the eliminant yields the following set as containing possible singularities:

$$\mathfrak{H}^{(n-1)} \equiv \mathfrak{S}_{1,0}^{(n+1)} \cap \mathfrak{S}_{2,0}^{(n+1)} \cap \mathfrak{H}_{1,2}^{(n+1)}, \mathfrak{H}_{1,2}^{(n+1)} \equiv \left\{\frac{\partial S_1}{\partial \zeta_1}\frac{\partial S_2}{\partial \zeta_2} - \frac{\partial S_1}{\partial \zeta_2}\frac{\partial S_2}{\partial \zeta_1} = 0\right\}.$$

The set $\mathfrak{H}_{1,2}^{(n+1)}$ occurs upon applying the envelope method to the eliminant of $\mathfrak{S}_{1,0}^{(n+1)} \cap \mathfrak{S}_{2,0}^{(n+1)}$ regardless which variable ζ_1 or ζ_2 has been eliminated (by the Hadamard approach), providing *no* $\partial S_\nu/\partial \zeta_\mu = 0$ $(\nu, \mu = 1, 2)$. The case where all the $\partial S_\nu/\partial \zeta_\mu = 0$ is contained above (in the envelope method approach) and if *not all* the $\partial S_\nu/\partial \zeta_\mu = 0$ we may compute $\mathfrak{H}^{(n+1)}$ by eliminating a *particular* variable first. Finally, by applying Theorem (3) to the combined singularity manifold of the integrand, $\mathfrak{Q}_0^{(n+1)}$, we obtain as the set of possible singularities for $F(z)$

$$\mathfrak{E}^{(n-1)} \equiv \mathfrak{Q}_0^{(n+1)} \cap \mathfrak{Q}_1^{(n+1)} \cap \mathfrak{Q}_2^{(n+1)}.$$

Moreover it is easy to see from the definitions of $\mathfrak{H}^{(n-1)}$ and $\mathfrak{E}^{(n-1)}$ that $\mathfrak{H}^{(n-1)} \subset \mathfrak{E}^{(n-1)}$. We summarize the above discussion with the following theorem, which has already appeared in a more general form as Lemma (3.2) in a Technical Note by Gilbert and Howard[15]

Theorem 4: Let $f_\nu(z; \zeta_1, \zeta_2)$ $(\nu = 1, 2)$ be regular analytic in $(\mathfrak{D}^{(n)} \times \mathfrak{B}^{(2)})$ with the singular sets $\mathfrak{S}_{\nu,0}^{(n+1)}$ as described above. Further, let $\mathcal{L}_\nu$ be rectifiable contours in the ζ_ν-planes, respectively; then the continuation of the analytic function element given in $\mathfrak{D}^{(n)}$ by

[15] R. P. Gilbert and H. C. Howard, Technical Note BN-366 (July 1964), Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland (to appear in the Australian Math. Soc. Journal).

$$F(z) = \left(\frac{1}{2\pi i}\right)^2 \int_{\mathcal{L}_1}\int_{\mathcal{L}_2} f_1(z;\zeta_1,\zeta_2) f_2(z;\zeta_1,\zeta_2)\, d\zeta_1\, d\zeta_2$$

is regular at all points (z) which may be reached by continuation along $\mathcal{C}_z^{(1)}$ provided no point of $\mathcal{C}_z^{(1)}$ meets the set $\mathfrak{E}^{(n-1)}$. A somewhat sharper estimate for the singular set of $F(z)$ is the set of points $\mathfrak{H}^{(n-1)}$.

Finally, we consider integrals of the form

$$F(z) = \left(\frac{1}{2\pi i}\right)^2 \int_{\mathcal{L}_1}\int_{\mathcal{L}_2} K(z;\zeta_1,\zeta_2) \times f_1(z;\zeta_1) f_2(z;\zeta_2)\, d\zeta_1\, d\zeta_2 \qquad (2.7)$$

and apply our previous results. Let $\mathfrak{S}_{0,0}^{(n+1)} \equiv \{S(z;\zeta_1,\zeta_2) = 0\}$, $\mathfrak{S}_{\nu,0}^{(n)} \equiv \{\zeta_\nu = \phi_\nu(z)\}$ $(\nu = 1, 2)$ be singular sets of $K(z;\zeta_1,\zeta_2)$, and $f_\nu(z;\zeta_\nu)$ $(\nu = 1, 2)$, respectively. Let $\mathfrak{S}_{0,\mu}^{(n+1)} \equiv \{\partial/\partial\zeta_\mu S(z;\zeta_1,\zeta_2) = 0\}$, $\mu = 1, 2$. The singularity manifold of the integrand is

$$\mathfrak{Q}_0^{(n+1)} \equiv \{S(z;\zeta_1,\zeta_2)(\zeta_1 - \phi_1(z))(\zeta_2 - \phi_2(z)) = 0\},$$

and we define, as usual,

$$\mathfrak{Q}_1^{(n+1)} = \{S_{\zeta_1}(z;\zeta)(\zeta_1 - \phi_1)(\zeta_2 - \phi_2) - S(z;\zeta)(\zeta_2 - \phi_2) = 0\}$$

and

$$\mathfrak{Q}_2^{(n+1)} \quad \{S_{\zeta_2}(z;\zeta)(\zeta_1 - \phi_1)(\zeta_2 - \phi_2) - S(z;\zeta)(\zeta_1 - \phi_1) = 0\}.$$

Clearly, if we apply the envelope method to the entire integrand we have as "candidates" for singularities the set $\bigcap_{\nu=0,1,2} \mathfrak{Q}_\nu^{(n+1)}$.

If we apply the envelope method to the singularities of the kernel we obtain the set

$$\bigcap_{\mu=0,1,2} \mathfrak{S}_{0,\mu}^{(n+1)} \subset \bigcap_{\nu=0,1,2} \mathfrak{Q}_\nu^{(n+1)}.$$

Applying the Hadamard method to both ζ_ν $(\nu = 1, 2)$ variables we obtain

$$\mathfrak{S}_{0,0}^{(n+1)} \cap \mathfrak{S}_{1,0}^{(n)} \cap \mathfrak{S}_{2,0}^{(n)} \equiv \{S(z;\phi_1(z),\phi_2(z)) = 0\} \subset \bigcap_{\nu=0,1,2} \mathfrak{Q}_\nu^{(n+1)}.$$

Corollary to Theorem 4: The possible singularities of the analytic continuation of the function element given in $\mathfrak{D}^{(n)}$ by (2.7) lie on the set

$$\mathfrak{S}_{0,0}^{(n+1)} \cap [(\bigcap_{\mu=1,2} \mathfrak{S}_{0,\mu}^{(n+1)}) \cup (\bigcap_{\nu=1,2} \mathfrak{S}_{\nu,0}^{(n)})] \subset \bigcap_{\mu=0,1,2} \mathfrak{Q}_\mu^{(n+1)}.$$

We have finally Theorem 5, which appeared earlier in more general form (for n-fold integrals) as Lemma (4.2) in a Technical Note by Gilbert and Howard.[16]

[16] See Ref. 15.

Theorem 5: Let $K(z;\zeta)$ be a holomorphic function of the $(n+2)$-complex variables $(z_1, \cdots, z_n;\zeta_1,\zeta_2)$ in a product space $\mathcal{B}^{(n)} \times \mathfrak{D}^{(2)} \subset \mathbf{C}^{(n+2)}$. Let Γ be topologically a 2-dimensional chain whose boundary $\partial\Gamma$ is fixed. Furthermore, let the singularities of $K(z;\zeta)$ lie on the analytic set,

$$\mathfrak{S}_0^{(n+1)} \equiv \{S(z;\zeta) = 0\}.$$

Then the holomorphic function defined by the integral representation

$$F(z) = \int_\Gamma F(z;\zeta)\, d\zeta_1 \wedge d\zeta_2,$$

is regular at all points not lying on the set $\Lambda_1 \cup \Lambda_2$, where

$$\Lambda_1 \equiv \bigcap_{\mu=0}^{2} \mathfrak{S}_\mu^{(n+1)},$$

$$\Lambda_2 \equiv \bigcup_{(\zeta)\in\partial\Gamma} \mathfrak{S}_0^{(n+1)},$$

and

$$\mathfrak{S}_\mu^{(n+1)} = \left\{\frac{\partial S(z;\zeta)}{\partial\zeta_\mu} = 0\right\} (\mu = 1, 2).$$

Proof: Let us continuously deform $\Gamma \to \tilde{\Gamma}$, such that the boundary remains fixed, i.e., $\partial\Gamma = \partial\tilde{\Gamma}$. If in so doing this we have not passed over a singularity of the integrand, i.e., if the kernel $K(z;\zeta)$ is holomorphic-regular for a fixed (z) and for $(\zeta \in \mathbf{V}$, where $\partial\mathbf{V} \equiv \Gamma - \tilde{\Gamma}$, then by the Cauchy–Poincaré theorem we have

$$\int_\Gamma K(z;\zeta)\, d\zeta_1 \wedge d\zeta_2 = \int_{\tilde{\Gamma}} K(z;\zeta)\, d\zeta_1 \wedge d\zeta_2.$$

From this it is clear that as we continue $F(z)$ along a curve originating at say (z^0) we may treat all singularities that threaten to cross at the interior points of Γ in the same manner as we did in our earlier results. That is it follows from before that the points lying on $\Lambda_1 = \bigcap_{\mu=0}^{2} \mathfrak{S}_\mu^{(n+1)}$ are possible singularities.

Singularities which tend to cross Γ at a boundary point $(\zeta) \in \partial\Gamma$ are however of a different type. We cannot deform the boundary since it is given to be fixed. Hence these points *may* be singular points, and so we conclude the points on $\Lambda_2 \equiv \bigcap_{(\zeta)\in\partial\Gamma} \mathfrak{S}_0^{(n+1)}$ may be singularities of $F(z)$. We remark that since Λ_2 is contained in a set of real-dimension $(2n-1)$ that these points *may* correspond to a natural boundary of $F(z)$.

III. CONCLUDING REMARKS

In closing we note an interesting connection between the original Hadamard idea (Theorem 0) and

the elastic unitarity integral of the quantum theory of fields.[17] The unitarity integral is given by[18]

$$\phi(s;x) - \phi_2(s;x) = \frac{i}{8}\left(\frac{s-4m^2}{s}\right)^{\frac{1}{2}} \int_{-1}^{+1} dx_1 \times \int_{-1}^{+1} dx_2 \frac{\theta(1-x_1^2-x_2^2-x^2+2x_1x_2x)}{(1-x_1^2-x_2^2-x^2+2x_1x_2x)^{\frac{1}{2}}} \times \phi_2(s;x_1)\phi(s;x_2). \tag{3.1}$$

For $-1 \leq x_1, x_2, x \leq +1$, Omnes[19] has given a replacement for the factor used for the kernel in (3.1), namely,

$$\frac{\theta(1-x_1^2-x_2^2-x^2+2x_1x_2x)}{(1-x_1^2-x_2^2-x^2+2x_1x_2x)^{\frac{1}{2}}} = \frac{\pi}{2}\sum_{l=0}^{\infty}(2l+1)P_l(x_1)P_l(x_2)P_l(x), \tag{3.2}$$

where the $P_l(\xi)$ are Legendre polynomials. On introducing Legendre coefficients, or partial waves,

$$\phi_{l,2}(s) = \tfrac{1}{2}\int_{-1}^{+1} dx_1\, P_l(x_1)\phi_2(s;x_1), \tag{3.3}$$

$$\phi_l(s) = \tfrac{1}{2}\int_{-1}^{+1} dx_2\, P_l(x_2)\phi(s;x_2),$$

the unitarity condition (3.1) becomes

$$\phi(s;x) - \phi_2(s;x) = \frac{i\pi}{4}\left(\frac{s-4m^2}{s}\right)^{\frac{1}{2}} \times \sum_{l=0}^{\infty}(2l+1)\phi_{l,2}(s)\phi_l(s)P_l(x). \tag{3.4}$$

The right-hand side is *similar* to the sum $\sum_{n=0}^{\infty} a_n b_n z^n$ contained in Hadamard's multiplication of singularities theorem, and the functions

$$\phi(s;x) = \sum_{l=0}^{\infty}(2l+1)\phi_l(s)P_l(x), \tag{3.5}$$

$$\phi_2(s;x) = \sum_{l=0}^{\infty}(2l+1)\phi_{2,l}(s)P_l(x),$$

are similar to the sums $\sum_{n=0}^{\infty} a_n z^n$, and $\sum_{n=0}^{\infty} b_n z^n$.

We shall show, in a sequel to this paper, on the elastic unitarity integral, that, by applying the extended Hadamard criteria to the function

$$F(s;x) = \sum_{l=0}^{\infty}(2l+1)\phi_{2,l}(s)P_l(x), \tag{3.6}$$

$F(s;x)$ can be singular only whenever the singularities of the kernel and those of $\phi(s;x)$ and $\phi_2(s;x)$ coincide.[20] This is whenever $x = x_1x_2 \pm [(1-x_1^2)(1-x_2^2)]^{\frac{1}{2}}$ and when, for each fixed value of s, the singularities of $\phi(s;x_1)$ and $\phi_2(s;x_2)$ are represented by $x_1 = \alpha(s)$ and $x_2 = \beta(s)$, respectively. One obtains in this case

$$x = \alpha(s)\beta(s) \pm [(1-\alpha^2(s))(1-\beta^2(s)]^{\frac{1}{2}}. \tag{3.7}$$

If we replace α, β by $\alpha = \frac{1}{2}(\zeta_1 + 1/\zeta_1)$, $\beta = \frac{1}{2}(\zeta_2 + 1/\zeta_2)$, respectively, we may represent this as

$$x = \frac{1}{4}\left(\zeta_1 + \frac{1}{\zeta_1}\right)\left(\zeta_2 + \frac{1}{\zeta_2}\right) \pm \frac{1}{4}\left(\zeta_1 - \frac{1}{\zeta_1}\right)\left(\zeta_2 - \frac{1}{\zeta_2}\right), \tag{3.8}$$

which becomes for $|\zeta_k| = 1$, $\theta_k = \arg \zeta_k$, $(k = 1, 2)$,

$$x = \cos\theta_1 \cos\theta_2 \pm \sin\theta_1 \sin\theta_2 = \cos(\theta_1[s] \pm \theta_2[s]). \tag{3.9}$$

This result may also be obtained as a simplified form of an analogous Hadamard theorem for generalized axially symmetric potentials (GASPT)[21] and is also contained in the result of Nehari.[22] It is also interesting to note that other results[23] concerning the location and number of singularities of functions with the representations (3.5) and (3.6) lead immediately to new interpretations here. In fact if one knows information about the partial waves, one can obtain complete information about the scattering amplitude singularities by using the Mandelbrojt–Hadamard analogs.[24–26] For instance, see the note by Gilbert and Shieh[27] to see how this may be done for the case of potential scattering.

[17] A complete discussion of the analytic properties of the elastic unitarity integral using the methods of this paper is contained in the paper by S. Ø. Aks, R. P. Gilbert, and H. C. Howard, Technical Note BN-376 (September 1964), Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland, J. Math. Phys. (to be published).

[18] The elastic unitarity integral is a direct consequence of the unitarity of the S-matrix (operator) at energies below the first production threshold. For details see: S. Ø. Aks, Technical Note BN-363 (July 1964), Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland, J. Math. Phys. (to be published).

[19] R. Omnes, Nuovo Cimento **25**, 806 (1962).

[20] See Ref. 17.

[21] R. P. Gilbert, J. Math. Phys. **5**, 983 (1964).

[22] Z. Nehari, J. Ratl. Mech. Anal. **5**, 987 (1956).

[23] R. P. Gilbert, Arch. Ratl. Mech. Anal. **6**, 171 (1960).

[24] S. Bergman, J. Anal. Math. **11**, 249 (1963).

[25] See: R. P. Gilbert (Ref. 21).

[26] R. P. Gilbert, Pacific J. Math. **13**, 79 (1963).

[27] R. P. Gilbert and S. Y. Shieh, Technical Note BN-401 (May 1965) Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland.

Solution of the Klein–Gordon and Dirac Equations for a Particle with a Plane Electromagnetic Wave and a Parallel Magnetic Field*

PETER J. REDMOND
Defense Research Corporation, Santa Barbara, California
(Received 24 December 1964)

Solutions of the Klein–Gordon and Dirac equations are obtained for a particle moving in an arbitrary plane electromagnetic wave in vacuo plus a uniform static magnetic field parallel to the direction of propagation of the electromagnetic wave.

INTRODUCTION

RECENTLY Roberts and Buchsbaum[1] solved the classical, relativistic equations of motion for a charged particle moving in a particular configuration of electromagnetic fields. The field configuration considered consisted of a plane electromagnetic wave plus a static magnetic field such that the direction of the static field was parallel to the direction of propagation of the wave.

In this paper we show that the corresponding quantum mechanical problem is also soluble. We obtain solutions of the Klein–Gordon and the Dirac equations for this field configuration. The solutions are exact within a framework in which the quantum nature of the electromagnetic field is ignored; and in which quantum mechanical radiative corrections and the classical radiative reaction force are neglected.

The quantum mechanical wavefunctions obtained are very closely related to the classical trajectories. For the Klein–Gordon equation, the relationship is so intimate that the classical solutions for the motion transverse to the magnetic field are also valid quantum mechanical solutions provided the equations are properly interpreted.

The problem we treat is a generalization of one discussed by Volkov[2] many years ago. Volkov solved the Klein–Gordon and Dirac equations with a plane electromagnetic wave. We generalize this to include a static magnetic field. It is interesting to note that the Volkov solutions have recently proven useful in studies of the Compton scattering of an electron by a laser beam.[3]

Although the classical problem has already been solved by Roberts and Buchsbaum we present an alternative derivation of their results for the special case when the electromagnetic wave propagates in the vacuum. It is only for this special case that the quantum mechanical problem can be solved. (This is true even when the static magnetic field is absent.) Because we restrict ourselves to a very special case it is possible to obtain a very much simpler derivation than that given by Roberts and Buchsbaum (who allow propagation in a dielectric medium). The resultant simplicity facilitates comparison of the quantum mechanical and classical solutions.

I. PRELIMINARIES

In this section we establish our notation and conventions and obtain a suitable expression for the electromagnetic field tensor.

We consider a particle with charge $-e$ and mass m. Units are chosen such that $\hbar = c = 1$. We shall need the Dirac matrices satisfying the usual anticommutation relations $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\,g_{\mu\nu}$; $\mu, \nu = 1, 2, 3, 0$. The metric tensor has signature 1, 1, 1, -1.

In considering the motion of a charged particle with a magnetic field in the z direction it is convenient to introduce the quantities $x \pm iy$. In order to have a convenient notation for these combinations we introduce the vectors ϵ and ϵ^* with components $\epsilon_\mu = (1/\sqrt{2})(1, i, 0, 0)$ and $\epsilon^*_\mu = (1/\sqrt{2})(1, -i, 0, 0)$. These vectors satisfy the conditions $\epsilon\cdot\epsilon = \epsilon^*\cdot\epsilon^* = 0$ and $\epsilon\cdot\epsilon^* = 1$. We shall frequently need to consider the projection of a vector in the x–y plane. We shall call such a projection the transverse part of the vector. In general, if a vector has components C_μ then the transverse part of C has components, $C^\perp_\mu$, given by

$$C^\perp_\mu = \epsilon_\mu\epsilon^*\cdot C + \epsilon^*_\mu\epsilon\cdot C. \tag{1.1}$$

The electromagnetic field tensor $F_{\mu\nu}$ is derivable from a potential so that $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. We write the potential as a sum of two terms distinguished from each other by their arguments. Thus

$$A_\mu(x) = A_\mu(\mathbf{x}_\perp) + A_\mu(\theta). \tag{1.2}$$

The first term on the right-hand side of Eq. (1.2) generates the static magnetic field and is a function

* This work was supported by the U. S. Army Research Office (Durham).
[1] C. S. Roberts and S. J. Buchsbaum, Bull. Am. Phys. Soc. 9, 14 (1964).
[2] D. M. Volkov, Z. Physik 94, 250 (1934).
[3] L. S. Brown and T. W. B. Kibble, Phys. Rev. 133, A703 (1964).

only of the transverse components of the position vector. The second term generates the plane wave and for a wave propagating in the z direction this is an arbitrary function of the single variable $t - z$ ($\equiv \theta$). By introducing the vector n with components $n_\mu = (0, 0, -1, -1)$ we may write $\theta = n\cdot x$. Note that $n^2 = n\cdot\epsilon = n\cdot\epsilon^* = 0$.

The electromagnetic field tensor may then be written

$$F_{\mu\nu} = F_{\mu\nu}(H) + F_{\mu\nu}(\theta)$$
$$= iH(\epsilon_\mu\epsilon^*_\nu - \epsilon_\nu\epsilon^*_\mu) + n_\mu\, dA_\nu/d\theta - n_\nu\, dA_\mu/d\theta, \quad (1.3)$$

where H is the magnitude of the static magnetic field.

By insisting that the potential have the form given by Eq. (1.2) we have placed a restriction on the permissible gauge for the potential. It will be convenient for our purposes to restrict the gauge even further and we shall insist that all the vectors appearing in Eq. (1.2) be transverse.

Within these restrictions, $dA_\mu/d\theta$ is determined by $F_{\mu\nu}$. In the following we shall consider only plane waves of finite pulse length which implies that $\lim_{|\theta|\to\infty} dA_\mu/d\theta = 0$. We may then uniquely determine $A_\mu(\theta)$ by the requirement that $A_\mu(-\infty) = 0$. In general then $A_\mu(+\infty)$ will exist but will not vanish. Although $A_\mu(\theta)$ is completely determined by $F_{\mu\nu}$ there is still considerable freedom in the choice of the first term $A_\mu(\mathbf{x}_\perp)$. We shall not find it necessary to further specify the gauge of $A_\mu(\mathbf{x}_\perp)$.

If the reader finds these gauge restrictions unduly onerous they can be easily removed. It is only necessary to allow an additional term $\partial_\mu f$ on the right-hand side of Eq. (1.2), where f is an arbitrary function of space–time, and to multiply the wavefunctions determined in the following by a factor $\exp(-ief)$.

In the following, all our manipulations are formally covariant. The results are therefore valid not only for the "standard" field configuration, which we use as a basis for our discussion, but also for any more general field configuration which may be transformed into the standard configuration by a Lorenz transformation. The general field tensor will still be described by Eq. (1.4) except that the vectors ϵ, ϵ^*, and n are any set of vectors such that the components of ϵ and ϵ^* are complex conjugates of each other, the components of n are real, $\epsilon\cdot\epsilon = \epsilon^*\cdot\epsilon^* = \epsilon\cdot n = \epsilon^*\cdot n = n\cdot n = 0$, and $\epsilon\cdot\epsilon^* = 1$.[4] The field configuration then consists of the following: (a) a plane wave; (b) a static magnetic field with magnitude H and a direction parallel to the direction of propagation of the plane wave; (c) a static electric and a static magnetic field. These additional fields have an arbitrary magnitude except that the extra electric field is equal to the extra magnetic field. The extra fields are perpendicular to each other and to H.

[4] Once the field tensor $F_{\mu\nu}$ is known, the vectors ϵ, ϵ^*, and n may be determined by solving the eigenvalue equation $F_{\mu\nu}S^\nu = \lambda S_\mu$. The vectors ϵ, ϵ^*, and n correspond to the eigenvalues $i\lambda$, $-i\lambda$, and 0, respectively. The vectors ϵ and ϵ^* are then determined except for an arbitrary phase and n is given except for an arbitrary real multiplier.

II. DETERMINATION OF CLASSICAL TRAJECTORIES

With our conventions the classical equations of motion are

$$m\ddot{x}_\mu = -eF_{\mu\nu}\dot{x}^\nu, \quad (2.1)$$

where the dots indicate differentiation with respect to the proper time(s). By using the expression given by Eq. (1.3) for $F_{\mu\nu}$, and noting that $\dot{A}_\mu = n\cdot\dot{x}\; dA_\mu/d\theta$, this may be written as

$$m\ddot{x}_\mu = -ieH(\epsilon_\mu\epsilon^*\cdot\dot{x} - \epsilon^*_\mu\epsilon\cdot\dot{x}) - en_\mu\dot{x}\cdot(dA/d\theta) + e\dot{A}_\mu. \quad (2.2)$$

Each term on the right-hand side of Eq. (2.2) is orthogonal to the vector n. By taking the inner product of this equation with n we obtain

$$n\cdot\ddot{x} = 0. \quad (2.3)$$

It follows then, that by an appropriate choice for the origin of the proper time, we may set

$$\theta = \alpha s \quad (2.4)$$

with α constant. The electron moves in such a way that its "clock" keeps time by the phase of the wave.

If we now introduce the quantity

$$P_\mu(s) = m\dot{x}_\mu + ieH\epsilon_\mu\epsilon^*\cdot x - ieH\epsilon^*_\mu\epsilon\cdot x - eA_\mu, \quad (2.5)$$

we find that P_μ satisfies the equation

$$\dot{P}_\mu(s) = -en_\mu(dA_\nu/d\theta)\dot{x}^\nu. \quad (2.6)$$

We may then write

$$P_\mu(s) = P_\mu(-\infty) + n_\mu I(s) \quad (2.7)$$

with

$$\dot{I}(s) = -e(dA/d\theta)\cdot\dot{x}. \quad (2.8)$$

Since the longitudinal parts of P_μ and the kinetic momentum $m\dot{x}_\mu$ are equal, and since n_μ is in the direction of the energy and momentum vector of the plane wave, Eq. (2.6) has an obvious interpretation in terms of the absorption of photons.

The right-hand side of Eq. (2.8) depends only on the transverse components of x. Inspection of the

transverse part of Eq. (2.2) reveals that this equation is formally equivalent to the problem of determining the transverse motion of a nonrelativistic particle moving in a static field and a time-dependent electric field. In this formal equivalence the proper time plays the role of the nonrelativistic ordinary time. These equations can be solved. Once the transverse motion is determined, the right-hand side of Eq. (2.8) is known and the longitudinal motion can be obtained by an integration.

A more explicit expression for $I(s)$ is easily obtained. According to the analogy, $\dot{I}(s)$ is essentially given by the electric field multiplied by the velocity. As a result $I(s)$ is the change in the nonrelativistic kinetic energy. Guided by these considerations we immediately obtain

$$I(s) = -\frac{1}{\alpha}\left[\frac{m}{2}\dot{\mathbf{x}}_\perp^2(s) - \lim_{s\to-\infty}\frac{m}{2}\dot{\mathbf{x}}_\perp^2(s)\right]. \tag{2.10}$$

An alternate derivation of Eq. (2.10) starts with the observation that $\dot{x}^2 = -1$. By considering the longitudinal and transverse parts of this expression separately and using Eqs. (2.7) and (2.5) for the longitudinal parts we again arrive at Eq. (2.10).

In order to complete the solution it is necessary to determine the transverse motion. By taking the inner product of Eq. (2.5) with ϵ, and noting that $\epsilon\cdot P = \epsilon\cdot P(-\infty)$ we obtain

$$m\epsilon\cdot x - ieH\epsilon\cdot x - e\epsilon\cdot A = \epsilon\cdot P(-\infty). \tag{2.11}$$

This has the solution

$$\epsilon\cdot x = \frac{i}{eB}\epsilon\cdot P(-\infty) + C(-\infty)e^{i\omega_0 s} + \int_{-\infty}^{s} ds'\left(\frac{e}{m}\right)e^{i\omega_0(s-s')}\epsilon\cdot A(\alpha s'), \tag{2.12}$$

where $C(-\infty)$ is a constant and $\omega_0(=eH/m)$ is the cyclotron frequency. This, plus the equation for $\epsilon^*\cdot x$ obtained by complex conjugation, completes the solution.

An alternative form of the solution is of some interest. If we write

$$\epsilon\cdot x = \lambda(s) + C(s)e^{i\omega_0 s}, \tag{2.13}$$

and

$$\epsilon\cdot\dot{x} = i\omega_0 C(s)e^{i\omega_0 s}, \tag{2.14}$$

then the two unknowns $\lambda(s)$ and $C(s)$ can be determined in terms of $\epsilon\cdot x$ and $\epsilon\cdot\dot{x}$. This form is of some interest in that λ and C constant corresponds to the solution without the radiation field. In this case λ locates the guiding center for the orbit and C gives the radius and phase angle for the orbit. The proper time-dependent λ and C may then be interpreted as instantaneous values of these quantities. We find

$$\lambda(s) = (i/eB)P(-\infty)\cdot\epsilon + (ie/m\omega_0)A(\alpha s)\cdot\epsilon \tag{2.15}$$

and

$$C(s) = C(-\infty) - \frac{ie}{m\omega_0}\int_{-\infty}^{s} ds'\, e^{-i\omega_0 s'}\frac{d}{ds'}\epsilon\cdot A(\alpha s'). \tag{2.16}$$

Thus, λ and C have finite limits as $|s| \to \infty$ which represent the final position of the guiding center and the final radius and phase of the circular orbit, respectively.

The relations

$$\lambda(\infty) = \lambda(-\infty) + (ie/m\omega_0)\epsilon\cdot A(\infty) \tag{2.17}$$

$$C(\infty) = C(-\infty) - \frac{ie}{m\omega_0}\int_{-\infty}^{+\infty} ds'\, e^{-i\omega_0 s'}\frac{d}{ds'}\epsilon\cdot A(\alpha s') \tag{2.18}$$

constitute the classical scattering matrix.

It is easy to see from these relations that it is possible to have a resonance. Consider a right circularly polarized plane wave which is sinusoidal with frequency ω for a period of time T. For such a wave the electric field is rotating with the same sense as the electron. While the field is on $\epsilon\cdot A$ will have the form $\epsilon\cdot A = |\epsilon\cdot A|e^{i\omega\alpha s}$ with $|\epsilon\cdot A|$ a complex constant. For those electrons with an initial momentum in the z direction such that the Doppler-shifted frequency is equal to the cyclotron frequency ($\omega\alpha = \omega_0$) we will have

$$C(\infty) = C(-\infty) + (e/m)T\,|\epsilon\cdot A|. \tag{2.19}$$

The radius of the circular orbit increases linearly with the time T that the wave is present. [For T sufficiently large that the second term in Eq. (2.19) dominates.] The energy then goes up quadratically with T. It is interesting to note that the change in the position of the guiding center goes as the zeroeth power of T, depending only on the net impulse.

III. KLEIN-GORDON EQUATION

The Klein–Gordon equation is

$$\{m^2 + [-i\partial + eA(\mathbf{x}_\perp) + eA(\theta)]^2\}\Phi(x) = 0. \tag{3.1}$$

Since the coefficients depend only on $\mathbf{x}_\perp$ and θ, solutions of the form

$$\Phi(x) = \Phi_p(\mathbf{x}_\perp, \theta)e^{ipx} \tag{3.2}$$

exist. Without loss of generality we require that p be longitudinal. We also introduce the operators $\pi_\mu = -i\partial_\mu + eA_\mu(\mathbf{x}_\perp)$. In the absence of the radiation field, these operators represent the kinetic mo-

mentum of the particle. They satisfy the commutation relations

$$[\pi_\mu, \pi_\nu] = -ieF_{\mu\nu}(H). \tag{3.3}$$

In order to facilitate comparison with the classical solutions we replace the variable θ by the proper time, s, defined so that

$$\theta = n\cdot p/m = \alpha s. \tag{3.4}$$

We then have

$$\left[m^2 + \left(p - in\frac{m}{n\cdot p}\frac{\partial}{\partial s} + \pi_\perp + eA(\alpha s)\right)^2\right]\Phi_p = 0. \tag{3.5}$$

This may be rewritten in the form

$$\left[i\frac{\partial}{\partial s} + \frac{p_\perp^2}{2m} - \frac{1}{2m}\left(\pi_\perp + eA(\alpha s)\right)^2\right]\Phi_p = 0, \tag{3.6}$$

with $p_\perp^2 = -(m^2 + p^2)$. In complete analogy with the classical situation, this is the Schrödinger equation describing the nonrelativistic motion of a charged particle in a static magnetic field with an additional time-dependent electric field present.

We will solve Eq. (3.6) by making a suitable transformation. It seems appropriate to indicate the reasoning which led to the selection of such a transformation. The quantum mechanical problem is most easily solved in the Heisenberg representation. In this representation the classical equations of motion are also true quantum mechanically. Because the equations are linear, the classical solutions given by Eqs. (2.15) and (2.16) are also valid quantum mechanically provided $\lambda(s)$, $C(s)$, $\lambda(-\infty)$, and $C(-\infty)$ are treated as noncommuting operators. Since $\lambda(s)$ and $C(s)$ satisfy the same commutation relations as $\lambda(-\infty)$ and $C(-\infty)$ the operators at a time s are related to the operators at a time $-\infty$ by a unitary transformation. By going from the Heisenberg representation to the Schrödinger representation it is found that the same unitary transformation serves to solve the Schrödinger equation. Rather than go through this lengthy procedure we introduce the appropriate unitary transformation as an ansatz.

If $K(s)$ is a transverse vector the following identities hold:

$$e^{iK\cdot\pi}\pi_\mu^\perp e^{-iK\cdot\pi} = \pi_\mu^\perp - eF_{\mu\nu}(H)K^\nu \tag{3.7}$$

and

$$e^{iK\cdot\pi}i\frac{\partial}{\partial s}e^{-iK\cdot\pi} = i\frac{\partial}{\partial s} + \pi\cdot\dot{K} - \frac{e}{2}\dot{K}^\mu F_{\mu\nu}(H)K^\nu. \tag{3.8}$$

The combination $e^{iK\cdot\pi}\Phi_p$ then satisfies the equation

$$\left[i\frac{\partial}{\partial s} + K\cdot\pi - \frac{e}{2}\dot{K}\cdot F(H)\cdot K + \frac{p_\perp^2}{2m} - \frac{1}{2m}\left(\pi_\perp - eF(H)\cdot K + eA(\alpha s)\right)^2\right]e^{iK\cdot\pi}\Phi_p = 0. \tag{3.9}$$

We now determine K by the requirement that the terms linear in π drop out of Eq. (3.9). This leads to the classical equation

$$m\dot{K}_\mu = -eF_{\mu\nu}(H)K^\nu + eA_\mu(\alpha s). \tag{3.10}$$

The condition $K(-\infty) = 0$ then determines K completely. The component $\epsilon\cdot K$ is given by the third term on the right-hand side of Eq. (2.12) and $\epsilon^*\cdot K$ is obtained by complex conjugation.

When the terms linear in π have been eliminated, it is found that the equation separates so that

$$\left[i\frac{\partial}{\partial s} - J(s)\right]e^{iK\cdot\pi}\Phi_p = 0, \tag{3.11}$$

with

$$J(s) = \int_{-\infty}^{s} ds'\left[\frac{e}{2}\dot{K}(s')\cdot F(H)\cdot K(s') + \frac{1}{2m}\left(-eF(H)\cdot K(s') + eA(\alpha s')\right)^2\right], \tag{3.12}$$

and

$$(\pi_\perp^2 - p_\perp^2)e^{iK\cdot\pi}\Phi_p = 0. \tag{3.13}$$

The separation constant has been determined by demanding that p be the incoming longitudinal momentum. This implies that $\lim_{s\to-\infty}\partial\Phi_p/\partial s = 0$.

These equations are now quite tractable. In particular Eq. (3.13) describes a particle in a uniform static magnetic field. This is a problem which has been discussed by many authors. We give a brief sketch of the procedure by which solutions to Eq. (3.13) are obtained and then quote the results.

One proceeds by noting the relation

$$[\epsilon^*\cdot\pi, \epsilon\cdot\pi] = eH. \tag{3.14}$$

This is the commutation relation for annihilation and creation operators. If $e > 0$ then $\epsilon^*\cdot\pi$ is the annihilation operator and the ground state is determined by

$$\epsilon^*\cdot\pi\Psi_0(\mathbf{x}_\perp) = 0. \tag{3.15}$$

This is a first-order partial differential equation which separates if $A(\mathbf{x}_\perp)$ is a linear function of $\mathbf{x}_\perp$. Once the ground state has been determined the excited states are given by

$$\Psi_n(\mathbf{x}_\perp) = (n!)^{-\frac{1}{2}}(eH)^{-n/2}(\epsilon\cdot\pi)^n\Psi_0(\mathbf{x}_\perp), \tag{3.16}$$

Since

$$\pi_\perp^2 = \epsilon^*\cdot\pi\ \epsilon\cdot\pi + \epsilon\cdot\pi\ \epsilon^*\cdot\pi \tag{3.17}$$

the allowed values of $p_\perp^2$ are given by

$$p_\perp^2 = (2n + 1)eH. \tag{3.18}$$

The separation constant which occurs in the course of solving Eq. (3.15) provides an additional quantum label for the state vectors which we do not indicate. The additional quantum number may be chosen in a large number of ways. For example it may be the z component of the angular momentum. The infinite degeneracy associated with this extra quantum label arises because of the invariance of the statement of the physical problem under transverse translations. (Equivalently, the degeneracy arises because the location of the guiding center is a constant of the motion.)

Our discussion follows very closely the work of Johnson and Lippmann.[5] We refer the interested reader to their paper for a more complete discussion of the degeneracy and for explicit forms for $\Psi_n(\mathbf{x}_\perp)$.

A complete set of solutions of the Klein–Gordon equation, Eq. (3.1), is then given by

$$\Phi(p, n, x) = e^{ipx}e^{-iJ}e^{-iK\cdot\pi}\Psi_n(\mathbf{x}_\perp). \tag{3.19}$$

IV. INTERPRETATION OF SOLUTION TO KLEIN-GORDON EQUATION

The solution of the Klein–Gordon equation obtained in the last section is formal in that the combination $e^{-iK\cdot\pi}\Psi_n(\mathbf{x}_\perp)$ must be evaluated. In this section we obtain two expressions for this quantity.

The expression $K\cdot\pi$ is the sum of two terms which do not commute with one another, that is

$$K\cdot\pi = -iK\cdot\partial + eK\cdot A(\mathbf{x}_\perp). \tag{4.1}$$

The following simple trick[6] is useful when it is necessary to evaluate the exponential of a sum of noncommuting operators. Consider the operator $G(\lambda)$ defined by

$$G(\lambda) = \exp\left[-\lambda K\cdot\partial - ie\lambda K\cdot A(\mathbf{x}_\perp)\right]\exp\lambda K\cdot\partial. \tag{4.2}$$

By differentiating this with respect to λ we obtain

$$\frac{\partial G(\lambda)}{\partial\lambda} = \exp\left\{-[\lambda K\cdot\partial + ie\lambda K\cdot A(\mathbf{x}_\perp)]\right\} \times [-ieK\cdot A(\mathbf{x}_\perp)]\exp(\lambda K\cdot\partial). \tag{4.3}$$

Using the relation

$$\exp(-\lambda K\cdot\partial)A(\mathbf{x}_\perp)\exp(+\lambda K\cdot\partial) = A(\mathbf{x}_\perp - \lambda K) \tag{4.4}$$

this can be written in the form

$$\partial G(\lambda)/\partial\lambda = G(\lambda)[-ieK\cdot A(\mathbf{x}_\perp - \lambda K)] \tag{4.5}$$

with the solution

$$G(\lambda) = \exp\left[-ie\int_0^\lambda d\lambda'\ K(s)\cdot A(\mathbf{x}_\perp - \lambda' K)\right]. \tag{4.6}$$

Setting $\lambda = 1$ we then obtain

$$e^{-ik\cdot\pi}\Psi_n(\mathbf{x}_\perp) = G(1)\exp(-K\cdot\partial)\Psi_n(\mathbf{x}_\perp) = \exp\left[-ie \times \int_0^1 d\lambda\ K(s)\cdot A(\mathbf{x}_\perp - \lambda K)\right] \times \Psi_n(\mathbf{x}_\perp - K(s)). \tag{4.7}$$

It is interesting to verify that this expression has the proper transformation properties when a gauge transformation is made. If $A_\mu(\mathbf{x}_\perp) \to A_\mu(\mathbf{x}_\perp) + \partial_\mu f(\mathbf{x}_\perp)$ then the wavefunction $\Phi(x) \to e^{-ief}\Phi(x)$. Now $\Psi_n(\mathbf{x}_\perp)$ has the required transformation property but $\Psi_n(\mathbf{x}_\perp - K)$ does not. However, the argument of the exponential in Eq. (4.6) contains the term

$$\begin{aligned} -ie\int_0^1 d\lambda\ K\cdot\partial f(\mathbf{x}_\perp - \lambda K) &= +ie\int_0^1 d\lambda\,\frac{d}{d\lambda}f(\mathbf{x}_\perp - \lambda K) \\ &= ie(f(\mathbf{x}_\perp - K) - f(\mathbf{x}_\perp)), \end{aligned} \tag{4.8}$$

which is just what is needed to restore the correct transformation properties.

It is also possible to express $e^{-iK\cdot\pi}\Psi_n$ as an expansion in the functions Ψ_n. In order to do this we note that

$$K\cdot\pi = K\cdot\epsilon\ \epsilon^*\cdot\pi + K\cdot\epsilon^*\ \epsilon\cdot\pi \tag{4.9}$$

is the sum of an annihilation and a creation operator. With the aid of the identity

$$e^{\lambda(A+B)} = e^{\lambda A}e^{\lambda B}e^{-\frac{1}{2}\lambda^2[A,B]}, \tag{4.10}$$

valid when the commutator of A and B commutes with A and B, we may write

$$e^{-iK\cdot\pi} = e^{-iK\cdot\epsilon^*\ \epsilon^*\cdot\pi}e^{-iK\cdot\epsilon\ \epsilon\cdot\pi}e^{-\frac{1}{2}eHK^2}. \tag{4.11}$$

By expanding each of the two exponentials and using

[5] M. H. Johnson and B. A. Lippmann, Phys. Rev. 76, 828 (1949). A bibliography of the earlier work on this problem is contained in this paper. Our notation and conventions have been chosen so as to be consistent with this paper except that we set $\hbar = c = 1$.

[6] This trick of introducing a parameter and then examining the equation obtained by differentiating the parameter provides a simple means of proving the various identities involving exponential operators that we have used.

the matrix elements of the annihilation and creation operators, one obtains[7]

$$e^{-iK\cdot\pi}\Psi_n(\mathbf{x}_\perp) = e^{-\frac{1}{4}eHK^2}\sum \frac{(-iK\cdot\epsilon^*)^j}{j!}\frac{(-iK\cdot\epsilon)^k}{k!}\frac{(eH)^{\frac{1}{2}(j+k)}}{} \times\left(\frac{n!n'!}{(n-k)!(n'-j)!}\right)^{\frac{1}{2}}\Psi_{n'}(\mathbf{x}_\perp), \quad (4.12)$$

where the summation runs over integral values of j, k, and n' subject to the restrictions that $n - k = n' - j$ and that the arguments of the factorials are all nonnegative. It is interesting to observe that if the system is initially in the ground state the probability, P_n, that it will be in a state n is given by the Poisson distribution

$$P_n = (\tfrac{1}{2}eHK^2)^n(1/n!)e^{-\frac{1}{4}eHK^2}. \quad (4.13)$$

V. DIRAC EQUATION

The Dirac equation is, with our conventions,

$$[m - i\gamma(-i\partial + eA(x))]\Phi = 0. \quad (5.1)$$

Any solution of the Dirac equation also satisfies the equation obtained by multiplying on the left with $[m + i\gamma(-i\partial + eA)]$ so that

$$[m^2 + (-i\partial + eA(x))^2 - \tfrac{1}{2}ie\gamma_\mu\gamma_\nu F^{\mu\nu}]\Phi = 0. \quad (5.2)$$

Our procedure shall be to first find a complete set of solutions to Eq. (5.2). We then find the linear combinations of these solutions which asymptotically satisfy the Dirac equation (5.1). It is easily seen that these linear combinations then satisfy the Dirac equation for all times.

As in the case of the Klein–Gordon equation we may take solutions having the form

$$\Phi(x) = e^{ipz}\Phi_p(\mathbf{x}_\perp, s). \quad (5.3)$$

We make the same transformations which were successful for the Klein–Gordon equation to obtain

$$\left[2mi\frac{\partial}{\partial s} - p_\perp^2 + \pi_\perp^2 + \frac{eH}{2}(\gamma\epsilon\gamma\epsilon^* - \gamma\epsilon^*\gamma\epsilon) \cdot - \frac{ie}{\alpha}\gamma n\gamma\dot{A}(\alpha s)\right]e^{iJ}e^{iK\cdot\pi}\Phi_p = 0. \quad (5.4)$$

Now the combinations $\sigma_{\mu\nu} = (1/2i)(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$ are the well known relativistic generalizations of the Pauli spin matricies. The expression $S = \frac{1}{2}(\gamma\epsilon\gamma\epsilon^* - \gamma\epsilon^*\gamma\epsilon)$ is simply the component of the spin matrix in the z direction. As $\gamma n\gamma\dot{A} = -\gamma\dot{A}\gamma n$ the $\sigma_{\mu\nu}$ also appear in this term. Since the term $\gamma n\gamma\dot{A}$ vanishes asymptotically we may choose solutions of Eq. (5.4) which are asymptotically eigenfunctions of S with eigenvalues ± 1.

The equation separates with the separation constant determined just as in the case of the Klein–Gordon equation. We have

$$[p_\perp^2 - (\pi_\perp^2 \pm eH)]e^{iJ}e^{iK\cdot\pi}\Phi_p^\pm = 0 \quad (5.5)$$

and

$$[2mi\,\partial/\partial s + eH(S \mp 1) - (ie/\alpha)\gamma n\gamma\dot{A}(\alpha s)]\, e^{iJ}e^{iK\cdot\pi}\Phi_p^\pm = 0. \quad (5.6)$$

The eigenfunctions and eigenvalues of $\pi_\perp^2$ have already been discussed. The allowed values of $p_\perp^2$ are then given by[8]

$$p_\perp^2 = 2neH. \quad (5.7)$$

A complete set of solutions to the Dirac equation is then given by

$$\Phi(p, n, x) = e^{ipz - iJ - iK\cdot\pi} \times \{\Psi_{n-1}(\mathbf{x}_\perp)f_+(s) + \Psi_n(\mathbf{x}_\perp)f_-(s)\}, \quad (5.8)$$

where $f_\pm$ are spinors determined by Eq. (5.6) and the requirement that the right-hand side of Eq. (5.8) satisfies the Dirac equation asymptotically.

The equation for $f_+(s)$ is

$$2mi\dot{f}_+ = [eH\gamma\epsilon^*\gamma\epsilon + (ie/\alpha)\gamma n\gamma\dot{A}]f_+. \quad (5.9)$$

The form of the solution can be determined by solving the equation by iteration starting with f_+ on the right-hand side approximated by a constant spinor u_+ satisfying $S\,u_+ = u_+$. Now it is readily seen that $S\,u_+ = u_+$ implies $\gamma\epsilon u_+ = 0$. As a result we find that f_+ has the form

$$f_+(s) = [1 + \gamma n\gamma\epsilon^* R(s)]u_+ \quad (5.10)$$

with $R(s)$ a scalar function of s. Substituting this into Eq. (5.9) we obtain, after some manipulation of the γ matrices,

$$[2mi\dot{R} - 2eHR - (ie/\alpha)\epsilon\cdot\dot{A}]\gamma n\gamma\epsilon^* u_+ = 0. \quad (5.11)$$

This can be satisfied nontrivially only if

$$R(s) = \frac{ie}{2m\alpha}e^{i\omega s}\int_{-\infty}^{s} ds'\, e^{-i\omega s'}\frac{d}{ds'}\,\epsilon\cdot A(\alpha s'). \quad (5.12)$$

The same procedure applied to f_- leads to the solution

$$f_- = (1 + R^*\gamma n\gamma\epsilon)u_-. \quad (5.13)$$

[7] This structure occurs in any problem which is equivalent to a quantum harmonic oscillator driven by a classical force. The result is derived by J. Schwinger, Phys. Rev. 91, 728 (1953) in connection with electrodynamics with a classical current. He shows that polynomials involved are related to Laguerre polynomials.

[8] M. H. Johnson and B. A. Lippmann, Phys. Rev. 77, 702 (1950).

The functions R and R^* also occur in the classical solution where they describe the radius and phase of the circular orbit [see Eq. (2.16)].

In order to complete the solution it is only necessary to determine the correct linear combination which asymptotically satisfies the Dirac equation. Suitable choices for the appropriate linear combinations of u_+ and u_- can be obtained by using Eqs. (25), (30), or (32) of Ref. 8. If this is done the results will not be in an explicitly covariant form. Use of the results of Ref. 8 corresponds to a choice of representation of the Dirac matrices given by

$$\gamma_0 = i\beta, \quad \beta = \begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix}, \quad \boldsymbol{\gamma} = -i\begin{pmatrix} 0, & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}, & 0 \end{pmatrix}.$$

An explicity covariant solution is also possible. The asymptotic form of the Dirac equation is

$$(m - i\gamma p - i\gamma\pi_\perp)\Phi^A(\mathbf{x}_\perp) = 0. \tag{5.14}$$

In order to treat this equation we require a representation of the Dirac matrices. A convenient representation is obtained by considering the spinors satisfying

$$Su(\lambda, \sigma) = \sigma u(\lambda, \sigma), \tag{5.15}$$

and

$$\gamma pSu(\lambda, \sigma) = i\lambda(-p^2)^{\frac{1}{2}}u(\lambda, \sigma), \tag{5.16}$$

where λ and σ take on the values ± 1. In this representation γp is also diagonal. The combination γpS is more convenient since it commutes with both γp and $\gamma\pi_\perp$. These spinors may be used to form the basis for a representation of the Dirac matrices. The representation is not unique since the relative phases of the different $u(\lambda, \sigma)$ are arbitrary. By choosing an explicit representation the relative phases are specified.

It is convenient to introduce the quantities $\chi(n, \lambda, \sigma, \mathbf{x}_\perp)$ defined by

$$\chi(n, \lambda, +1, \mathbf{x}_\perp) = \Psi_{n-1}(\mathbf{x}_\perp)u(\lambda, +1) \tag{5.17}$$

and

$$\chi(n, \lambda, -1, \mathbf{x}_\perp) = \Psi_n(\mathbf{x}_\perp)u(\lambda, -1). \tag{5.18}$$

It is also simpler to introduce the vector q defined uniquely by the conditions

$$q\cdot\epsilon = q\cdot\epsilon^* = q\cdot p = 0. \tag{5.19}$$

and

$$q\cdot n = p\cdot n = m\alpha. \tag{5.20}$$

It is readily found that $q^2 = -p^2$ and

$$n_\mu = (m\alpha/p^2)(p_\mu - q_\mu). \tag{5.21}$$

A completely satisfactory representation of the Dirac matrices is then given by:

$$\gamma\pi\chi(n, \lambda, \sigma, \mathbf{x}_\perp) = (2neH)^{\frac{1}{2}}\chi(n, \lambda, -\sigma, \mathbf{x}_\perp) \tag{5.22}$$

$$\gamma p\chi(n, \lambda, \sigma, \mathbf{x}_\perp) = i(-p^2)^{\frac{1}{2}}(\lambda\sigma)\chi(n, \lambda, \sigma, \mathbf{x}_\perp) \tag{5.23}$$

$$\gamma\epsilon\chi(n, \lambda, +1, \mathbf{x}_\perp) = \gamma\epsilon^*\chi(n, \lambda, -1, \mathbf{x}_\perp) = 0 \tag{5.24}$$

$$\gamma\epsilon\chi(n, \lambda, -1, \mathbf{x}_\perp) = \sqrt{2}\chi(n + 1, \lambda, +1, \mathbf{x}_\perp) \tag{2.25}$$

$$\gamma\epsilon^*\chi(n, \lambda, +1, \mathbf{x}_\perp) = \sqrt{2}\chi(n - 1, \lambda, -1, \mathbf{x}_\perp) \tag{5.26}$$

and

$$\gamma q\chi(n, \lambda, \sigma, \mathbf{x}_\perp) = \sigma(-p^2)^{\frac{1}{2}}\chi(n, -\lambda, \sigma, \mathbf{x}_\perp). \tag{5.27}$$

In this representation the Dirac matrices are self-adjoint if the adjoint (†) is related to the Hermitian conjugate (*) by

$$\gamma_\mu^\dagger = -p^{-2}\gamma p\gamma_\mu^*\gamma p = \gamma_\mu. \tag{5.28}$$

The asymptotic solutions to the Dirac equation then have the form

$$\Phi^A(p, n, \lambda, \mathbf{x}_\perp) = a\chi(n, \lambda, +1, \mathbf{x}_\perp) + b\chi(n, \lambda, -1, \mathbf{x}_\perp) \tag{5.29}$$

and the Dirac equation reduces to

$$(m + \lambda\sqrt{-p^2})a - i(2neH)^{\frac{1}{2}}b = 0 \tag{5.30}$$

and

$$-i(2neH)^{\frac{1}{2}}a + (m - \lambda\sqrt{-p^2})b = 0. \tag{5.31}$$

The equations are consistent only if $p_\perp^2 \equiv -(m^2 + p^2) = 2neH$ which is the result we had previously obtained in Eq. (5.7). For correct normalization we must choose a and b so that $|a|^2 + |b|^2 = 1$.

Finally, the solutions of the complete problem are

$$\Phi(p, n, \lambda, x) = e^{ipz - iJ - iK\cdot r}[(1 + R\gamma n\gamma\epsilon^*)a\chi(n, \lambda, +1, \mathbf{x}_\perp) + (1 + R^*\gamma n\gamma\epsilon)b\chi(n, \lambda, -1, \mathbf{x}_\perp)]. \tag{5.32}$$

Solution of the Integral Equation for $V\theta$ Scattering in the Lee Model*

CHARLES M. SOMMERFIELD
Department of Physics, Yale University, New Haven, Connecticut
(Received 7 December 1964)

A deductive procedure is presented for obtaining the solution of the integral equation for $V\theta$ scattering in the Lee model.

A SOLUTION of the Källén–Pauli integral equation[1] for $V\theta$ scattering in the Lee model[2] was presented recently by Kenschaft and Amado.[3,4] In this note we describe a deductive procedure for obtaining this solution and for demonstrating its uniqueness.

The equation may be written in the form

$$\psi(\omega) = -\frac{1}{\omega} - \frac{1}{\pi}\int_{\mu} d\omega' \frac{\omega'}{\omega_0 - \omega' + i0} \times \frac{\operatorname{Im} K_+(\omega')}{K_+(\omega_0 - \omega')} \frac{\psi(\omega')}{\omega' + \omega - \omega_0 - i0}, \quad (1)$$

where $\mu > 0$ and where $K_+(\omega)$[5] is the boundary value $K(\omega + i0)$ of a function of a complex variable $K(z)$ given by

$$K(z) = 1 + \frac{z}{\pi}\int_{\mu}^{\infty} dx \frac{U(x)}{x(x - z)}.$$

The real function $U(x)$ is continuous on the interval of integration and approaches zero at both ends thereof so that $U(x) = O[(x^2 - \mu^2)^{\frac{1}{2}}]$ for $x \to \mu$ and $\int_{\mu}^{\infty} dx[U(x)/x] < \infty$. It is also such that $K(z)$ has no zeros in the complex plane. The domain of "physical" interest for the real parameter ω_0 is $\mu \leq \omega_0 < \infty$. We will first solve the equation for the simpler case $0 < \omega_0 < \mu$, however, and then indicate how the method is applied to the "physical" situation.

We see that for $0 < \omega_0 < \mu$ Eq. (1) is a nonsingular Fredholm equation with square-integrable kernel for the function $\psi(\omega)$ on the interval $\omega_0 - \mu < \omega < \infty$. The solution in this interval is thus unique. We use Eq. (1) to extend the range of definition of $\psi(\omega)$ into the rest of the complex ω plane. It is clear that $\psi(\omega)$ is then an analytic function of ω except for a pole at $\omega = 0$ with residue -1 and a cut extending along the real axis from $-\infty$ to $\omega_0 - \mu$. The discontinuity across the cut is given by

$$(2i)^{-1}[\psi(\omega + i0) - \psi(\omega - i0)] = \begin{cases} \dfrac{\omega_0 - \omega}{\omega} \dfrac{\operatorname{Im} K_+(\omega_0 - \omega)}{K_+(\omega)} \psi(\omega_0 - \omega), & \omega \leq \omega_0 - \mu \\ 0, & \omega > \omega_0 - \mu. \end{cases} \quad (2)$$

Referring to Fig. 1, we see that Eq. (1) may be

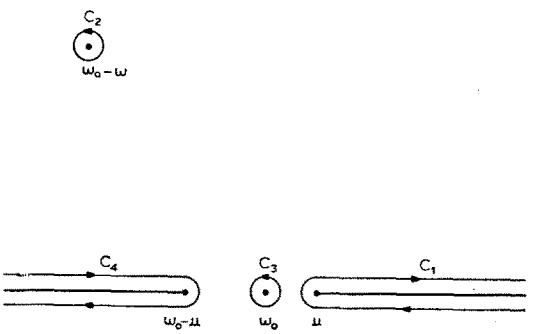


FIG. 1. Integration contours for $\omega_0 < \mu$.

written in terms of a contour integral:

$$\psi(\omega) = -\frac{1}{\omega} - \frac{1}{2\pi i}\int_{C_1} d\omega' \frac{\omega'}{\omega_0 - \omega'} \times \frac{K(\omega')}{K(\omega_0 - \omega')} \frac{\psi(\omega')}{\omega' + \omega - \omega_0}.$$

It follows from Eq. (1) that $\psi(\omega) = O(|\omega|^{-1})$ as $\omega \to \infty$ along the positive real axis, and indeed that $\psi(\omega) = O(|\omega|^{-1})$ for $\omega \to \infty$ in any direction in the complex plane. The contour C_1 may then be deformed so as to yield integrals along C_2, C_3, and C_4 together with a vanishing contribution from a

* Research supported in part by a grant from the National Science Foundation.

[1] G. Källén and W. Pauli, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **30**, No. 7 (1955).

[2] T. D. Lee, Phys. Rev. **95**, 1329 (1954).

[3] R. P. Kenschaft and R. D. Amado, J. Math. Phys. **5**, 1340 (1964).

[4] The solution to the $V\theta$ problem has also been obtained using different techniques by M. Vaughn (private communication) and A. Pagnamenta (University of Maryland preprint).

[5] In the notation of Ref. 3, $K_+(\omega) = h(\omega)/\omega$.

contour at infinity. We obtain by this procedure

$$\psi(\omega) + \frac{\omega_0 - \omega}{\omega}\frac{K(\omega_0 - \omega)}{K(\omega)}\psi(\omega_0 - \omega)$$
$$= -\frac{1}{\omega} + \frac{\omega_0 K(\omega_0)\psi(\omega_0)}{\omega}$$
$$+ \frac{1}{\pi}\int_{-\infty}^{\omega_0-\mu} d\omega' \frac{\omega'}{\omega_0 - \omega'}\frac{K(\omega')}{\omega' + \omega - \omega_0}\frac{1}{2i}$$
$$\times\left[\frac{\psi(\omega' + i0)}{K(\omega_0 - \omega' - i0)} - \frac{\psi(\omega' - i0)}{K(\omega_0 - \omega' + i0)}\right], \quad (3)$$

where, for convenience in what follows, we have moved the contribution arising from C_2 to the left-hand side.

Consider for a moment the function

$$L(\omega) = \psi(\omega) + \frac{\omega_0 - \omega}{\omega}\frac{K(\omega_0 - \omega)}{K(\omega)}\psi(\omega_0 - \omega). \quad (4)$$

According to Eq. (2), $L(\omega)$ has no cut for $\omega \leq \mu$, but can be discontinuous for $\omega > \mu$. Except for a pole at $\omega = 0$ it has no other singularity and is $O(|\omega|^{-1})$ at ∞. The function $\omega K(\omega)L(\omega)$ then has a possible cut for $\omega > \mu$, no pole, and goes to a constant at ∞. But according to the construction of Eq. (4),

$$\omega K(\omega)L(\omega) = (\omega_0 - \omega)K(\omega_0 - \omega)L(\omega_0 - \omega),$$

so that $\omega K(\omega)L(\omega)$ cannot have a cut at all, is thus an entire function, and is indeed just a constant, as dictated by its behavior for large $|\omega|$. Evaluating the constant in terms of $L(\omega_0)$ we obtain

$$L(\omega) = [\psi(\omega_0)\omega_0 K(\omega_0) - 1]/\omega K(\omega). \quad (5)$$

In arriving at Eq. (5) we used the statements $\lim_{\omega\to\omega_0}(\omega - \omega_0)\psi(\omega - \omega_0) = -1$ and $K(0) = 1$.

We now observe that Eq. (3) provides an expression for the discontinuity of the function

$$Q(\omega) = -\psi(\omega_0 - \omega)/K(\omega)$$

across the real axis in terms of the known discontinuity of $L(\omega)$ as follows:

$$\frac{\psi(\omega_0)\omega_0 K(\omega_0) - 1}{\omega}\operatorname{Im}\left[\frac{1}{K_+(\omega)}\right]$$
$$= -\frac{\omega_0 - \omega}{\omega}K(\omega_0 - \omega)\frac{1}{2i}$$
$$\times [Q(\omega + i0) - Q(\omega - i0)], \qquad \omega \geq \mu.$$

According to its definition $Q(\omega)$ has no cut for $\omega < \mu$, has a pole at $\omega = \omega_0$ with residue $-1/K(\omega_0)$ and is $O(|\omega|^{-1})$ at ∞. Thus we may use the Cauchy integral theorem to express $Q(\omega)$ in terms of its pole and the discontinuity across its cut and so discover that

$$\psi(\omega) = -\frac{K(\omega_0 - \omega)}{K(\omega_0)}\frac{1}{\omega}$$
$$+ [\psi(\omega_0)\omega_0 K(\omega_0) - 1]K(\omega_0 - \omega)A(\omega), \quad (6)$$

where

$$A(\omega) = \frac{1}{\pi}\int_\mu^\infty d\omega' \frac{1}{\omega' - \omega_0 + \omega}\frac{1}{\omega_0 - \omega'}$$
$$\times\frac{1}{K(\omega_0 - \omega')}\operatorname{Im}\frac{1}{K_+(\omega')}. \quad (7)$$

The value of $\psi(\omega_0)$ may be found by solving the algebraic equation obtained from Eq. (6) by the substitution $\omega = \omega_0$:

$$\psi(\omega_0) = -\frac{1}{\omega_0 K(\omega_0)} - \frac{2A(\omega_0)}{1 - \omega_0 K(\omega_0)A(\omega_0)}.$$

We obtain, consequently,

$$\psi(\omega) = -K(\omega_0 - \omega)$$
$$\times\left[\frac{1}{\omega K(\omega_0)} + \frac{2A(\omega)}{1 - \omega_0 K(\omega_0)A(\omega_0)}\right]. \quad (8)$$

If $1 - \omega_0 K(\omega_0)A(\omega_0) = 0$, Eq. (1) has no solution

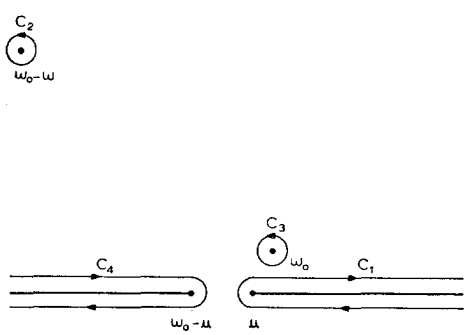


FIG. 2. Integration contours for $\mu \leq \omega_0 < 2\mu$.

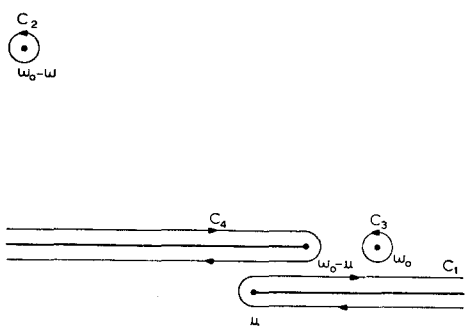


FIG. 3. Integration contours for $2\mu \leq \omega_0$.

by the associated homogeneous equation does. This would correspond to a bound state in the $V\theta$ channel.

We note that the same procedure we have illustrated will work for $2\mu > \omega_0 \geq \mu$. Now however, ω_0 has moved above the right-hand cut as in Fig. 2, as demanded by the denominator $\omega_0 - \omega' + i0$ in Eq. (1). And for $\omega_0 > 2\mu$, $\psi(\omega)$ is discontinuous within the interval of integration of Eq. (1). Then, according to the denominator $\omega' + \omega - \omega_0 - i0$ we are to interpret $\psi(\omega')$ as $\psi(\omega' - i0)$ so that the integration contours are as in Fig. 3. In both cases everything goes through as before except that in Eqs. (7) and (8), ω_0, as it appears in all places except in the denominator $\omega' - \omega_0 + \omega$, is to be interpreted as $\omega_0 + i0$.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. Lowell S. Brown for several helpful suggestions and to express my appreciation for the hospitality extended to me by Dr. L. Van Hove at the Theoretical Division of CERN where part of this work was done.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

On the Approximation of Quantum Field Theories

ARTHUR JAFFE
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 7 December 1964)

We assume that the Wightman distributions are tempered and that they are boundary values of functions analytic in the forward tube. Under these conditions a sequence of such distributions may converge in Borchers' topology. The necessary and sufficient conditions for such convergence are spelled out in terms of the corresponding analytic functions. The two cases with and without the assumption of the spectral condition are separately treated. A discussion of other topologies and some examples of the use of this technique are given.

I. INTRODUCTION

LET $\phi(x)$ be a quantum field,[1] so that smeared with a test function f in $\mathcal{S}$, it yields an operator $\phi(f)$. We can then ask when the field $\phi(f)$ and the theory it describes can be approximated, either by a sequence of fields or by some other method. In order to discuss the approximation it is necessary to introduce a topology into the space in question. We first must decide on a suitable approximation concept, and toward this end we consider a few examples.

(1) Let $\phi_\infty(f)$ and $\phi_j(f)$ be fields all of which satisfy the Borchers–Zimmermann growth condition[2] with radius of convergence greater than R_0. Then for all real test functions $f \in \mathcal{D}$, the closure of each of the operators $\phi_n(f)$ is self-adjoint. Thus, we might say that ϕ_j approximates ϕ_∞ (ultra) weakly, (ultra) strongly or uniformly if, for all real $f \in \mathcal{D}$, the corresponding unitaries $U_j(f) = e^{i\phi_j(f)}e^{-i\phi_\infty(f)}$ converge to the identity in the respective operator topology as $j \to \infty$. Furthermore, we know from the Borchers–Zimmermann condition that for t_1, $\cdots$, t_n in some sufficiently small polycircle, then for all j the vacuum expectation functionals

$$(\psi_0, e^{i\phi_j(t_1f_1+\cdots+t_nf_n)}\psi_0)$$

are holomorphic functions in $t_1, \cdots, t_n$. Since weak convergence of

$$U_j\Big(\sum_{i=1}^{n} t_if_i\Big)$$

to the identity implies the convergence of the vacuum expectation functions, we know from the analyticity that the coefficients $\langle\phi_j(f_1) \cdots \phi_j(f_n)\rangle$ tend to $\langle\phi_\infty(f_1) \cdots \phi_\infty(f_m)\rangle$ as j tends to ∞. However, the converse statement is not true; convergence of the Wightman functions one by one in $\mathcal{D}'$ does not imply weak convergence of the U_n, nor even convergence of $(\psi_0, e^{i\phi_n(f)}\psi_0)$. In fact, if each $(\psi_0, e^{i\phi_n(f)t}\psi_0)$ is analytic for all n with a radius of convergence in t

[1] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964), Chap. III.
[2] H. J. Borchers and W. Zimmermann, Nuovo Cimento **31**, 1047 (1964).

by the associated homogeneous equation does. This would correspond to a bound state in the $V\theta$ channel.

We note that the same procedure we have illustrated will work for $2\mu > \omega_0 \geq \mu$. Now however, ω_0 has moved above the right-hand cut as in Fig. 2, as demanded by the denominator $\omega_0 - \omega' + i0$ in Eq. (1). And for $\omega_0 > 2\mu$, $\psi(\omega)$ is discontinuous within the interval of integration of Eq. (1). Then, according to the denominator $\omega' + \omega - \omega_0 - i0$ we are to interpret $\psi(\omega')$ as $\psi(\omega' - i0)$ so that the integration contours are as in Fig. 3. In both cases everything goes through as before except that in Eqs. (7) and (8), ω_0, as it appears in all places except in the denominator $\omega' - \omega_0 + \omega$, is to be interpreted as $\omega_0 + i0$.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. Lowell S. Brown for several helpful suggestions and to express my appreciation for the hospitality extended to me by Dr. L. Van Hove at the Theoretical Division of CERN where part of this work was done.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 7 JULY 1965

On the Approximation of Quantum Field Theories

ARTHUR JAFFE
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 7 December 1964)

We assume that the Wightman distributions are tempered and that they are boundary values of functions analytic in the forward tube. Under these conditions a sequence of such distributions may converge in Borchers' topology. The necessary and sufficient conditions for such convergence are spelled out in terms of the corresponding analytic functions. The two cases with and without the assumption of the spectral condition are separately treated. A discussion of other topologies and some examples of the use of this technique are given.

I. INTRODUCTION

LET $\phi(x)$ be a quantum field,[1] so that smeared with a test function f in $\mathcal{S}$, it yields an operator $\phi(f)$. We can then ask when the field $\phi(f)$ and the theory it describes can be approximated, either by a sequence of fields or by some other method. In order to discuss the approximation it is necessary to introduce a topology into the space in question. We first must decide on a suitable approximation concept, and toward this end we consider a few examples.

(1) Let $\phi_\infty(f)$ and $\phi_j(f)$ be fields all of which satisfy the Borchers–Zimmermann growth condition[2] with radius of convergence greater than R_0. Then for all real test functions $f \in \mathcal{D}$, the closure of each of the operators $\phi_n(f)$ is self-adjoint. Thus, we might say that ϕ_j approximates ϕ_∞ (ultra) weakly, (ultra) strongly or uniformly if, for all real $f \in \mathcal{D}$, the corresponding unitaries $U_j(f) = e^{i\phi_j(f)}e^{-i\phi_\infty(f)}$ converge to the identity in the respective operator topology as $j \to \infty$. Furthermore, we know from the Borchers–Zimmermann condition that for t_1, $\cdots$, t_n in some sufficiently small polycircle, then for all j the vacuum expectation functionals

$$(\psi_0, e^{i\phi_j(t_1f_1+\cdots+t_nf_n)}\psi_0)$$

are holomorphic functions in $t_1, \cdots, t_n$. Since weak convergence of

$$U_j\Big(\sum_{i=1}^{n} t_if_i\Big)$$

to the identity implies the convergence of the vacuum expectation functions, we know from the analyticity that the coefficients $\langle\phi_j(f_1) \cdots \phi_j(f_n)\rangle$ tend to $\langle\phi_\infty(f_1) \cdots \phi_\infty(f_m)\rangle$ as j tends to ∞. However, the converse statement is not true; convergence of the Wightman functions one by one in $\mathcal{D}'$ does not imply weak convergence of the U_n, nor even convergence of $(\psi_0, e^{i\phi_n(f)}\psi_0)$. In fact, if each $(\psi_0, e^{i\phi_n(f)t}\psi_0)$ is analytic for all n with a radius of convergence in t

[1] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964), Chap. III.
[2] H. J. Borchers and W. Zimmermann, Nuovo Cimento **31**, 1047 (1964).

greater than R_0, then a necessary and sufficient condition for $(\psi_0, e^{i\phi_n(f)^t}\psi_0) \to (\psi_0, e^{i\phi_\infty(f)^t}\psi_0)$ is that the Wightman functions $(\psi_0, \phi_n(f)^m\psi_0)$ converge to $(\psi_0, \phi_\infty(f)^m\psi_0)$ uniformly in m as $n \to \infty$. The relation of convergence of the Wightman functions to convergence in any of the above operator topologies is not clear.

(2) The ϕ_n may be fields but not self-adjoint. In this case the unitaries U_j could be replaced by a discussion of convergence of the Stone matrices[3] corresponding to ϕ_j, in one of the above topologies. However, these operators are even farther removed from the vacuum expectation values or other computable matrix elements.

(3) A weaker notion of convergence has been discussed by Borchers.[4] He introduces a topology directly on the space Σ of sequences of test functions, so that Σ is a locally convex, separable, complete space. The bounded sets of Σ have only a finite number of nonzero components, so that an element T of the dual Σ' when applied to $f \in \Sigma$ will be a finite sum of functionals in $\mathcal{S}'$ applied to test functions in $\mathcal{S}$. Hence, a sequence $T_n \subset \Sigma'$ will converge as $n \to \infty$ to T_∞ in the weak topology of Σ' if and only if each element $T_{k,n}$ of $\mathcal{S}'_{kl}$ converges as $n \to \infty$ to an element $T_{k,\infty}$ of $\mathcal{S}'_{kl}$. Here

$$T_n = \bigoplus_{k=0}^{\infty} T_{k,n}.$$

As in $\mathcal{S}'$, this implies convergence in the strong topology of Σ'.

This notion of convergence is useful, since Σ' is complete. Then a Wightman functional $W \in \Sigma'$ is defined as one which satisfies

Lorentz invariance:	$W((a, \Lambda)g) = W(g)$;	
Spectrum:	$W(g) = 0,$	$\forall g \in M_{SP};$ (W)
Locality:	$W(g) = 0,$	$\forall g \in I_e;$
Positivity:	$W(g^+ \times g) \geq 0,$	$\forall g.$

These conditions and uniqueness of the vacuum allow the construction of a Hilbert space and a field operator with W as Wightman functional.[1,4] Since Σ' is complete, if $\{W_n\}$ is a sequence of Wightman functionals, and $\{W_n\}$ converges in Σ' to W_∞ as $n \to \infty$, then W_∞ also satisfies the conditions to be a Wightman functional, and hence has a corresponding field.[4] Actually, if only any one of the above conditions (W) is satisfied for a convergent sequence of functionals $\{T_n\} \subset \Sigma'$, then the limit also satisfies this condition, positivity for example. Thus, we can speak of approximations of a field theory by distributions which do not have all the properties necessary to give rise to fields.

As mentioned above, the necessary and sufficient condition for convergence in Σ' is convergence of each component distribution $T_{k,n} \in \mathcal{S}'_{kl}$ of the sequence $T_k \in \Sigma'$, where $T_{k,n}$ is a distribution in k vector variables of l components and each converges in $\mathcal{S}'_{kl}$. The purpose of this note is to translate convergence of these functionals into convergence of the holomorphic functions, the Wightman functions, which give the distributions as boundary values. That is, we spell out the induced topology on the analytic functions. A more general case will be considered than the one analyzed in Ref. 4; we only require temperedness of the distributions, and the fact that they are boundary values of analytic functions. Then the spectral condition is formulated and we give two conditions on the convergence of the analytic functions depending on whether or not the distributions obey the spectral condition. In Sec. 3, a number of examples of applications are given for the theorems, some of which prove other properties of the limiting functionals than the four mentioned above.

We note that every Wightman functional $W \in \Sigma'$ can be expressed as

$$W = \bigoplus_{k=0}^{\infty} W_k,$$

where W_k is the distribution boundary value of a function analytic in the tube $\mathfrak{T}^{kl} = R^{kl} - i\Gamma$. The necessary and sufficient condition for convergence of a sequence of such W_k (and hence a sequence of such W functionals) will be given in Theorem 1. Hence, it is possible to put a topology on the space of analytic functions having Wightman functionals as boundary values by using this criterion of convergence. This topology will give this space of analytic functions a structure which is topologically equivalent to the structure on $W \cap \Sigma'$ analyzed by Borchers in Ref. 4. Therefore, Borchers' theorem A6 shows this space of analytic functionals is complete.

II. EXCHANGE OF LIMITS

Notation: We consider distributions in ml variables consisting of m l-vectors. The scalar product xp will be taken to mean

$$\sum_{i=1}^{m} \sum_{\mu=0}^{l-1} x^{(i)\mu} p^{(i)}{}_{\mu},$$

where μ indexes the components of a given vector

[3] M. H. Stone, J. Indian Math. Soc. 15, 155 (1951).
[4] H. J. Borchers, Nuovo Cimento 24, 214 (1962), Appendix.

and i the different vectors. In Theorem 1 we restrict the support of T_n (denoted from now on by "supp T_n") to lie in Γ. For each set Γ we define $\tilde{\Gamma}$ to be the dual of Γ and those points such that

$$\tilde{\Gamma} = \{x; xp \geq 0 \quad \text{for all} \quad p \in \Gamma\}.$$

Note that for all Γ, $\tilde{\Gamma}$ is a convex cone. In ordinary applications, Γ would be the direct product of n forward light cones, $\Gamma = V_+^{\otimes n}$. In this case $\tilde{\Gamma} = \Gamma$. We also denote the Euclidean length of a vector by

$$||x||^2 = \sum_{i=1}^{m} \sum_{\mu=0}^{l-1} |x^{(i)\mu}|^2.$$

In general we follow the notation of Ref. 1, Chap. 2.

Definition 1. Spectral Condition: We say that a sequence of tempered distributions $\{T_n\}$ obey the spectral condition if there is a compact C and a Γ such that the interior of $\tilde{\Gamma}$ is nonempty, and (supp T_n) $\subset (\Gamma \cup C)$ for all n. Note that C allows us to extend the support of our distributions outside Γ but only in some fixed compact. In one dimension, for instance, we could take $\Gamma = \{p; p \geq 0\}$, $C = \{p; -a \leq p \leq a\}$ so the support in p can run down to $-a$.

We can now state the following *Limit Theorem with Spectral Condition:*

Theorem 1. Let F_n be a sequence of tempered distributions such that the Fourier transforms $\mathfrak{F}F_n = T_n$ have support in $\Gamma \cup C$. Then if $\tilde{\Gamma}$ is nonempty, F_n is the boundary value of a function $\mathbf{F}_n$ holomorphic in $\tilde{\mathfrak{T}} = R^{ml} - i\tilde{\Gamma}$. Suppose that $\{F_n\}$ is a weakly bounded set in $\mathcal{S}'$, that is, for each $f \in \mathcal{S}$, $F_n(f)$ is bounded as n varies. Then the holomorphic functions are uniformly bounded in the sense that for each compact κ of $\tilde{\Gamma}$ there is a polynomial P_κ and an integer r such that for all ξ for all t in the range $0 < t < 1$, and for $\eta \in \kappa$,

$$|\mathbf{F}_n(\xi - it\eta)| < P_\kappa(\xi)/t^r \tag{1}$$

independently of n. Moreover, if for all $f \in \mathcal{S}$, $F_n(f) \to F_\infty(f)$ as $n \to \infty$, then the analytic functions $\mathbf{F}_n(\xi - i\eta)$ converge to $\mathbf{F}_\infty(\xi - i\eta)$ uniformly on all compacts in $\tilde{\mathfrak{T}}$.

Conversely, let $\mathbf{F}_n(\xi - i\eta)$ be a sequence of functions holomorphic in the tube $\tilde{\mathfrak{T}} = R^{ml} - i\tilde{\Gamma}$. Suppose that for each compact κ of $\tilde{\Gamma}$ there is a polynomial P_κ and an integer r such that for all ξ for all t in the range $0 < t < 1$ and for $\eta \in \kappa$

$$|\mathbf{F}_n(\xi - it\eta)| < P_\kappa(\xi)/t^r$$

independently of n. Then as η approaches zero in the interior of $\tilde{\Gamma}$, $\mathbf{F}_n(\xi - i\eta)$ has a tempered distribution boundary value F_n, and as n varies, F_n varies over a strongly bounded set of $\mathcal{S}'$. In addition, if the analytic functions converge pointwise, $\mathbf{F}_n(\xi - i\eta) \to \mathbf{F}_\infty(\xi - i\eta)$ for all $\xi - i\eta$ in $\tilde{\mathfrak{T}}$, then the boundary values F_n converge to F_∞ in the strong topology of $\mathcal{S}'$, that is the convergence is uniform on bounded sets of $\mathcal{S}$.

Proof. Suppose that $\{F_n\}$ is a weakly bounded set in $\mathcal{S}'$. Thus, by the theorem of Mackey[5] it is bounded in the strong topology of $\mathcal{S}'$, that is bounded on bounded sets of $\mathcal{S}$. Let the seminorms which define $\mathcal{S}$ be given an increasing order, $||f||_i \leq ||f||_{i+1}$ and let $\mathcal{S}_i$ be the space of functions with finite i-norm. Then $\mathcal{S}_i \supset \mathcal{S}_{i+1}$ and every bounded set of $\mathcal{S}'$ is contained in some $\mathcal{S}_i'$ and is bounded in norm in this space.[6] Thus we know that the F_n are equicontinuous, that is, there exists a given seminorm $||\ ||_i$ and a constant C such that for all f in $\mathcal{S}$ and independently of n

$$|F_n(f)| \leq C\,||f||_i. \tag{2}$$

On the other hand, it is shown in Ref. 1 that the Fourier transform of F_n, $T_n = \mathfrak{F}F_n$, has the property that it can be written

$$e^{-p\eta}T_n(p) = a(p; \eta)S_n(p), \tag{3}$$

where

$$a(p; \eta) = e^{-p\eta}\left\{\sum_{i=1}^{M} e^{-p\eta_i}\right\}^{-1}$$

with $\eta_i \in \tilde{\Gamma}$ and η in the convex hull of the η_i. In this form $a(p; \eta) \in \mathcal{S}_p$, $S_n \in \mathcal{S}_p'$.[7]

Thus, $S_n = \{\sum_{i=1}^{M} e^{-p\eta_i}\}T_n$. Since supp $T_n \subset \Gamma \cup C$, it is possible to choose $h(p)$ so that $h(p) = 1$ if $p \in \Gamma \cup C$ and

$$h(p)\left\{\sum_{i=1}^{M} e^{-p\eta_i}\right\} \in \mathcal{S}.$$

Hence

$$S_n = h(p)\left\{\sum_{i=1}^{M} e^{-p\eta_i}\right\}\mathfrak{F}F_n \tag{4}$$

and as F_n varies over a strongly bounded set of $\mathcal{S}'$, so does S_n. This follows since both Fourier transformation and multiplication by an element of $\mathcal{S}$ are continuous linear transformations from $\mathcal{S}'$ to $\mathcal{S}'$. Hence, by estimate (2) for bounded sets we

[5] N. Bourbaki, *Espaces Vectorials Topologiques* (Hermann & Cie., Paris, 1955), Chap. IV, p. 70, Sec. 2, Theorem 3. See also L. Schwartz, *Théorie des Distributions* (Hermann & Cie., Paris, 1959), Vol. I, Chap. III, Sec. 2 and Vol. II, p. 91.

[6] I. M. Guelfand and G. E. Silov, *Les Distributions* (Dunod Cie., Paris, 1964), Vol. II, page 45, Theorem 2. See also Ref. 5.

[7] Streater and Wightman, Ref. 1, p. 55–56.

know there exist a constant C and integer i such that for all $f \in \mathcal{S}$ and independently of n

$$|S_n(f)| \leq C\,||f||_i. \tag{5}$$

By taking $f(p, \xi, \eta) = a(p; \eta)e^{i(p,\xi)}$ we have

$$S_n(f) = \mathbf{F}_n(\xi - i\eta) \tag{6}$$

and so

$$|\mathbf{F}_n(\xi - i\eta)| \leq C\,||a(p; \eta)e^{ip\xi}||_i. \tag{7}$$

Furthermore, as η varies over a compact κ in the cone $\bar{\Gamma}$, and $0 < t < 1$, we have from (7) that there is a polynomial P_κ and integer r such that[8]

$$|\mathbf{F}_n(\xi - it\eta)| < P_\kappa(\xi)/t^r. \tag{8}$$

This proves that the $\mathbf{F}_n$ satisfy a uniform bound. If, furthermore, F_n converges weakly to F_∞ in $\mathcal{S}'$, then since $f = ae^{ip\xi} \in \mathcal{S}$, we know that $S_n(f)$ converges to $S_\infty(f)$ as S_n is a continuous function of a convergent series F_n. Hence,

$$|\mathbf{F}_n(\xi - it\eta) - \mathbf{F}_\infty(\xi - it\eta)| = |(S_n - S_\infty)(a(p; t\eta)e^{ip\xi})| \to 0. \tag{9}$$

Thus, $\mathbf{F}_n(\xi - i\eta)$ converges pointwise in the tube $\tilde{\mathfrak{T}} = R^{\eta l} - i\bar{\Gamma}$. However, (8) gives a uniform bound on all compacts in $\tilde{\mathfrak{T}}$ and thus by the Vitale convergence theorem,[9] $\mathbf{F}_n(\xi - i\eta) \to \mathbf{F}_\infty(\xi - i\eta)$ uniformly on all compacts in the tube.

Conversely, suppose that the $\mathbf{F}_n(\xi - i\eta)$ are holomorphic in $\tilde{\mathfrak{T}}$ and satisfy bound (1) for η in a compact of $\bar{\Gamma}$ as described above. Define the sequence of functions

$$\mathbf{G}_n^{(m)}(\xi - it\eta) = \int_1^t dt_1 \int_1^{t_1} dt_2 \cdots \int_1^{t_{m-1}} dt_m \mathbf{F}_n(\xi - it_m\eta). \tag{10}$$

For $\eta \in \bar{\Gamma}$, $0 < t \leq 1$, $\mathbf{F}_n(\xi - it\eta)$ is holomorphic, so $\mathbf{G}_n^{(m)}$ is also holomorphic in the tube. Furthermore, for $\eta \in \kappa$, estimate (1) yields the bounds

$$\begin{aligned}
|\mathbf{G}_n^{(r-1)}(\xi - it\eta)| &\leq \left|\int_1^t dt_1\right| \int_1^{t_1} dt_2 \cdots \\
&\quad \times \left|\int_1^{t_{m-1}} dt_m\right| \mathbf{F}_n(\xi - it_m\eta)\ ||\cdots|| \\
&\leq |P_\kappa(\xi)|/t \\
|\mathbf{G}_n^{(r)}(\xi - it\eta)| &\leq |P_\kappa(\xi)|\ |\log t| \\
|\mathbf{G}_n^{(r+1)}(\xi - it\eta)| &\leq |P_\kappa(\xi)| \\
|\mathbf{G}_n^{(r+2)}(\xi - it\eta)| &< |P_\kappa(\xi)|.
\end{aligned} \tag{11}$$

[8] Streater and Wightman, Ref. 1, p. 62.

[9] L. Bieberbach, *Lehrbuch der Funktionen Theorie* (B. G. Teubner, Leipzig, 1921), Vol. I, p. 165. This proves the theorem in one variable. A similar proof holds for n variables.

Also

$$|\mathbf{G}_n^{(r+2)}(\xi - it\eta) - G_n^{(r+2)}(\xi)| = \left|\int_0^t dt_1 \mathbf{G}_n^{(r+1)}(\xi - it_1\eta)\right| \leq t\,|P_\kappa(\xi)|. \tag{12}$$

Thus, for any $g(\xi) \in \mathcal{S}$, and $0 < t < 1$, it follows that for $\eta \in \kappa$, and for any polynomial $Q(\xi)$ we have

$$\begin{aligned}
&\left|\int \mathbf{G}_n^{(r+2)}(\xi - it\eta)g(\xi)\,d\xi\right| \\
&\leq \sup_\xi \left\{\frac{|\mathbf{G}_n^{(r+2)}(\xi - it\eta)|}{1 + |P_\kappa(\xi)|}\,\frac{1}{1 + |Q(\xi)|}\right\} \\
&\quad \times \int d\xi' \{1 + |P_\kappa(\xi')|\}\{1 + |Q(\xi')|\}\ |g(\xi')|.
\end{aligned} \tag{13}$$

The integral converges for $g \in \mathcal{S}$ and is independent of n. Furthermore, since a polynomially bounded function is a tempered distribution, there is a seminorm $||\ ||_i$ so that the integral is majorized by $C\,||g||_i$. On the other hand, estimate (11) shows that the sup term is less than one. Hence

$$\left|\int \mathbf{G}_n^{(r+2)}(\xi - it\eta)g(\xi)\,d\xi\right| \leq C\,||g||_i. \tag{14}$$

Furthermore, estimate (12) shows that as $t \to 0$, so does

$$\int \left[\mathbf{G}_n^{(r+2)}(\xi - it\eta) - G_n^{(r+2)}(\xi)\right]g(\xi)\,d\xi \to 0 \tag{15}$$

so that

$$\left|\int G_n^{(r+2)}(\xi)g(\xi)\,d\xi\right| \leq C\,||g||_i. \tag{16}$$

Let us choose

$$g(\xi) = \sum_{p=0}^{l-1} \left(i\eta^\mu \frac{\partial}{\partial \xi^\mu}\right)^{r+2} f(\xi)$$

for any $f \in \mathcal{S}_\xi$. Then, for $0 < t < 1$,

$$\int \mathbf{G}_n^{(r+2)}(\xi - it\eta)g(\xi)\,d\xi = \int \mathbf{F}_n(\xi - it\eta)f(\xi)\,d\xi.$$

We know that the left-hand side has a continuous boundary value as $t \to 0$, as does the right-hand side. Thus

$$G_n^{(r+2)}(g) = \lim_{t\to 0} \int \mathbf{G}_n^{(r+2)}(\xi - it\eta)g(\xi)\,d\xi = \lim_{t\to 0} \int \mathbf{F}_n(\xi - it\eta)f(\xi)\,d\xi = F_n(f). \tag{17}$$

Hence, for all f in $\mathcal{S}$, we know from (16) that there exists a constant C and norm $||\ ||$, such that

$$|F_n(f)| = |G_n^{(r+2)}(g)| \leq C\,||g||_i \leq \tilde{C}\,||f||, \tag{18}$$

independently of n. This proves that F_n is a strongly bounded set in $\mathcal{S}'$.

Let us suppose that $\mathbf{F}_n(\xi - i\eta)$ converges pointwise in the tube $\tilde{\mathfrak{T}}$ to $\mathbf{F}_\infty(\xi - i\eta)$. Since (1) gives a uniform bound for $F_n(\xi - i\eta)$ as $\xi - i\eta$ varies over any compact in the tube, we have from the Vitale convergence theorem[9] that $\mathbf{F}_n(\xi - i\eta)$ converges to $\mathbf{F}_\infty(\xi - i\eta)$ uniformly on compacts in the tube. Hence, $\mathbf{G}_n^{(r+2)}(\xi - it\eta)$ converges to $\mathbf{G}_\infty^{(r+2)}(\xi - it\eta)$ uniformly on compacts in $\tilde{\mathfrak{T}}$, since the G's are primatives of a series which converges uniformly on compacts. We consider

$$\left| \int \{\mathbf{G}_n^{(r+2)}(\xi - it\eta) - \mathbf{G}_\infty^{(r+2)}(\xi - it\eta)\} g(\xi)\, d\xi \right| \tag{19}$$

which is majorized by estimate (13) with $\mathbf{G}_n^{(r+2)}$ replaced by $\mathbf{G}_n^{r+2} - \mathbf{G}_\infty^{(r+2)}$. Again the integral term is bounded by $\tilde{C}\, ||f||_\nu$. By estimate (11),

$$\left| \frac{\mathbf{G}_n^{(r+2)}(\xi - it\eta) - \mathbf{G}_\infty^{(r+2)}(\xi - it\eta)}{1 + |P_\kappa(\xi)|} \right| < 2.$$

Hence, given any $\epsilon > 0$, we can choose a polynomial Q and M so that if $||\xi||^2 > M$, then $\{1 + |Q(\xi)|\}^{-1} < \frac{1}{2}\epsilon$. However, in the compact $||\xi||^2 \leq M$, $t\eta \in \kappa$, we know that $|\mathbf{G}_n^{r+2}(\xi - it\eta) - \mathbf{G}_\infty^{r+2}(\xi - it\eta)|$ converges to zero uniformly. Thus, there is an n_0 such that for all $n > n_0$ the sup term is less than ϵ. Furthermore, for the case $t = 0$ we have

$$\begin{aligned} &|G_n^{(r+2)}(g) - G_\infty^{(r+2)}(g)| \\ &\leq \left| G_n^{(r+2)}(g) - \int \mathbf{G}_n^{(r+2)}(\xi - it\eta) g(\xi)\, d\xi \right| \\ &\quad + \left| \int \{\mathbf{G}_n^{(r+2)}(\xi - it\eta) - \mathbf{G}_\infty^{(r+2)}(\xi - it\eta)\} g(\xi)\, d\xi \right| \\ &\quad + \left| \int \mathbf{G}_\infty^{(r+2)}(\xi - it\eta) g(\xi)\, d\xi - G_\infty^{(r+2)}(g) \right|. \end{aligned} \tag{20}$$

For the first and third terms we have from estimate (12) that for η fixed in $\tilde{\Gamma}$

$$\begin{aligned} &\left| G_n^{(r+2)}(g) - \int \mathbf{G}_n^{(r+2)}(\xi - it\eta) g(\xi)\, d\xi \right| \\ &< t \int |P_\kappa(\xi)|\, g(\xi)\, d\xi < Ct\, ||g||_\iota < \tilde{C}t\, ||f||_\lambda \end{aligned} \tag{21}$$

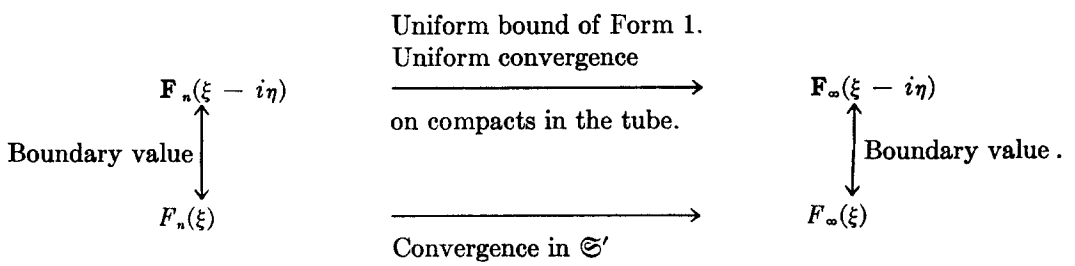
independently of n.

Hence, given $\epsilon > 0$, we can choose one value of t and one λ such that (20) is majorized by $\epsilon\, ||f||_\lambda$ for all n. Now having fixed these two terms we use the majorization of (19) where we saw that given $\epsilon > 0$, there exists n_0 such that $n > n_0$ (19) is majorized by $\epsilon\, ||f||_\nu$. Hence for a given $k > \max(\nu, \lambda)$ we have from these estimates and from (17) that given $\epsilon > 0$ then for all $f \in \mathcal{S}$, there is an n_0 such that for $n > n_0$

$$\begin{aligned} &|F_n(f) - F_\infty(f)| \\ &\quad = |G_n^{(r+2)}(g) - G_\infty^{(r+2)}(g)| < 3\epsilon\, ||f||_k. \end{aligned} \tag{22}$$

This just says that F_n converges to F_∞ uniformly on bounded sets of $\mathcal{S}$, which was to be proved. It completes the proof.

Theorem 1 characterizes the connection between limits of distributions and those of the corresponding analytic functions in the case when the distributions obey the spectral condition. It gives the necessary and sufficient conditions under which the order of these two limiting processes can be interchanged. That is, it tells when the following diagram commutes:



As usual in the case of interchange of limits, uniformity is essential. The uniformity of convergence in $\mathcal{S}'$ is assured from simple convergence since $\mathcal{S}$ is a Montel space. However, this does not imply uniformity for the analytic functions without the spectral condition or some other assumption. The following example shows the importance of the spectral condition for the above proof. Consider the case of one variable and let $T_n(p) = \delta(p + n) + \delta(p - n)$, and $T_\infty(p) = 0$. Then clearly for all $f \in \mathcal{S}$, $T_n(f) \to T_\infty(f) = 0$, and in fact the limit is uniform on bounded sets of $\mathcal{S}$. Furthermore, $\mathbf{F}_n(\xi - i\eta) = 2\cos n(\xi - i\eta)$ is an entire function of $\xi - i\eta$ for each n, and for all η,

$$\lim_{\eta \to 0} \int \mathbf{F}_n(\xi - i\eta) f(\xi)\, d\xi = F_n(f).$$

However, there is no uniform bound of type (1), and for fixed η,

$$\lim_{n\to\infty} \int \mathbf{F}_n(\xi - i\eta) f(\xi)\, d\xi = 0$$

is true for all $f \in \mathcal{S}$, only if $\eta = 0$. Thus the limiting operations can not be interchanged.

Moreover, in the case with spectral condition, such as $T_n(p) = \delta(p-n)$, $T_\infty(p) = 0$, $\Gamma = \{p;\, p \geq 1\}$, $\tilde{\Gamma} = \{\eta;\, \eta \geq 0\}$, the theorem only insures a uniform bound and convergence on compacts inside $R - i\tilde{\Gamma}$, that is the lower half-plane. We see that while for each n, $\mathbf{F}_n(\xi - i\eta)$ is an entire function of $\xi - i\eta$, as $n \to \infty$ the $\mathbf{F}_n$ diverge for $\eta < 0$.

In the more general case of any tempered distributions which are boundary values of holomorphic functions (without the requirement that supp $T_n \subset \Gamma \cup C$ where $\tilde{\Gamma}$ has a nonempty interior) we can make the following statements.

Theorem 2. Let $T_n(p)$ be a set of distributions. Let Γ_1 be the set of all η such that $e^{-p\eta}T_n(p)$ is in $\mathcal{S}'_p$ for all n. Let $\Gamma_2 \subset \Gamma_1$ be the set of all η in Γ_1 such that as n varies for η fixed, $e^{-p\eta}T_n$ is a bounded set of distributions in $\mathcal{S}'_p$. Let $\Gamma_3 \subset \Gamma_2$ be the set of all η in Γ_2 such that for fixed η, $e^{-p\eta}T_n$ is a convergent set of distributions in $\mathcal{S}'_p$ as n tends to infinity. Then Γ_1, Γ_2, and Γ_3 are convex. If the origin belongs to Γ_i, then so does $t\Gamma_i$ for $0 \leq t \leq 1$.

Proof. Suppose that η_1, η_2 are in Γ_i. Then for $\eta = \alpha\eta_1 + (1-\alpha)\eta_2$ we have

$$e^{-p\eta}T_n = a(p;\eta)\{e^{-p\eta_1}T_n + e^{-p\eta_2}T_n\},$$

where $a(p;\eta) = e^{-p\eta}\{e^{-p\eta_1} + e^{-p\eta_2}\} \in \mathcal{S}_p$. However, since multiplication by a function of $\mathcal{S}$ is a continuous linear map from $\mathcal{S}'$ to $\mathcal{S}'$, it sends bounded sequences into bounded sequences and convergent sequences into convergent ones. Hence $\eta \in \Gamma_i$. The statement about t is just convexity.

We can now state the following *Limit Theorem without Spectral condition.*

Theorem 3. Let $T_n(p)$ be a sequence of tempered distributions and suppose that Γ_1, Γ_2, and Γ_3 are defined as in Theorem 2. If Γ_1 has a nonempty interior, then each Fourier transform $F_n = \mathfrak{F}T_n$ is the boundary value of a function holomorphic in the interior of $\mathfrak{T}_1 = R^{ml} - i\Gamma_1$. If the origin is in Γ_2, and Γ_2 has a nonempty interior, then the holomorphic functions $\mathbf{F}_n(\xi - i\eta)$ are uniformly bounded on compacts in $\mathfrak{T}_2 = R^{ml} - i\Gamma_2$ in the sense that for each compact κ of Γ_2 there is a polynomial P_κ and an integer r such that for all ξ and all t in the range $0 < t < 1$

$$|\mathbf{F}_n(\xi - it\eta)| < P_\kappa(\xi)/t^r,$$

independently of n. Furthermore, if the origin is in Γ_3 and Γ_3 has a nonempty interior, then the analytic functions $\mathbf{F}_n(\xi - i\eta)$ converge to $\mathbf{F}_\infty(\xi - i\eta)$ uniformly on all compacts in $\mathfrak{T}_3 = R^{ml} - i\Gamma_3$.

Conversely, let $\mathbf{F}_n(\xi - i\eta)$ be a sequence of functions holomorphic in the tube $\mathfrak{T}_1$. Then if each $\mathbf{F}_n(\xi - i\eta)$ is bounded for all ξ, all η in a compact κ of Γ_1 and all t in the range $0 < t < 1$ by

$$|\mathbf{F}_n(\xi - it\eta)| < P'_\kappa(\xi)/t^r,$$

then $\mathbf{F}_n(\xi - i\eta)$ has a distribution boundary value as $\eta \to 0$ in Γ_1. If Γ_2 has a nonempty interior and for each compact κ of Γ_2 there is a uniform bound such that for all ξ, all $\eta \in \kappa$ and all t in the range $0 < t < 1$,

$$|\mathbf{F}_n(\xi - it\eta)| < P_\kappa(\xi)/t^r$$

independently of n, then as n varies the boundary values $\{F_n\}$ vary over a bounded set of $\mathcal{S}'$. Furthermore, $\int \mathbf{F}_n(\xi - i\eta) f(\xi)\, d\xi$ define a bounded set of functionals in $\mathcal{S}'$ as n varies and as η varies over any compact in Γ_2. If in addition the analytic functions $\mathbf{F}_n(\xi - i\eta)$ converge pointwise in the interior of $\mathfrak{T}_3 \subset \mathfrak{T}_2$, that is for $\eta \in \Gamma_3 \subset \Gamma_2$, then the boundary values $\{F_n\}$ converge to F_∞ in the strong topology of $\mathcal{S}'$. Furthermore, the functionals $\mathbf{F}_n(\xi - i\eta)$ converge to $\mathbf{F}_\infty(\xi - i\eta)$ uniformly as n tends to ∞ and η varies over a compact of Γ_3.

Remark. This theorem says that the above diagram pertaining to Theorem 1 will commute if the analytic functions are considered in $\mathfrak{T}_3 = R^{ml} - i\Gamma_3$ and the boundary values taken as $\eta \to 0$ in Γ_3.

Proof. The proof is just an extension of the proof of Theorem 1. In that proof it could be shown that if 0 were contained in Γ_1, Γ_2, or Γ_3, then that Γ_i was contained in Γ. However here the various Γ_i's must be considered separately. The only modification of the first half of the proof is that we deduced in (4) that S_n was a strongly bounded set. Now, this is assured if each $\eta_i \in \Gamma_2$. Thus if η is a vector in the convex hull of a set of vectors in Γ_2,

$$e^{-p\eta}T_n = a(p;\eta)\left\{\sum_{i=1}^{M} e^{-p\eta_i}T_n\right\}.$$

Each $\{e^{-p\eta_i}T_n\}$ will now be a bounded set by assumption, and multiplication by $a \in \mathcal{S}$ maps this into a bounded set. Furthermore, as η varies over a compact in Γ_2, and stays in the convex hull of the η_i, $a(p;\eta)$ varies over a bounded set of $\mathcal{S}$, and so

$\{e^{-p\eta}T_n\}$ is a bounded set of $\mathcal{S}'$ for all n and all η in a compact of Γ_2. We note that by Theorem 2, Γ_2 contains $t\Gamma_2$ for $0 \leq t \leq 1$, so that Γ_2 extends as a cone to the origin, and the t estimate can be made. This same argument is used for the convergence with η in the interior of Γ_3.

The converse of the theorem is a straightforward extension of the results of Theorem 1.

III. EXAMPLES OF APPLICATIONS

(1) Let ϕ_n be fields, Ω_0 the vacuum. Suppose that the analytic Wightman functions corresponding to

$$(\Omega_0, \phi_n^*(y_1) \cdots \phi_n^*(y_r)\phi_n(y_{r+1}) \cdots \phi_n(y_{2r})\Omega_0)$$
$$+ (\Omega_0, \phi_\infty^*(y_1) \cdots \phi_\infty^*(y_r)\phi_\infty(y_{r+1})\phi_\infty(y_{2r})\Omega_0)$$
$$- (\Omega_0, \phi_n^*(y_1) \cdots \phi_n^*(y_r)\phi_\infty(y_{r+1}) \cdots \phi_\infty(y_{2r})\Omega_0)$$
$$- (\Omega_0, \phi_\infty^*(y_1) \cdots \phi_\infty^*(y_r)\phi_n(y_{r+1}) \cdots \phi_n(y_{2r})\Omega_0)$$

with the difference variables placed in the forward tube converge to zero as required by the theorem. Then the vectors

$$\phi_n(f_1) \cdots \phi_n(f_r)\Omega_0 \xrightarrow{s} \phi_\infty(f_1) \cdots \phi_\infty(f_r)\Omega_0$$

converge in the strong topology.

(2) Let ψ be any Wick polynomial of the free field ϕ defined on the domain D_0, the polynomial algebra of ϕ applied to the vacuum. Then the closure $(\psi|_{D0})^{**}$ has in its domain all Wick powers applied to D_0.[10]

(3) It has been shown possible to prove positive definiteness of

$$:\exp ig\phi(x):$$

on the states of definite charge where $\phi(x)$ is a zero-mass free scalar field in two-dimensional space–time.[11]

(4) The author has used this method to prove positive definiteness for the Thirring model field on a large (but not complete) set of states.

IV. QUESTIONS

(1) What is the relation of this convergence to convergence of Stone projections or other bounded functions of fields? That is, what is the induced topology on these operators in the case when the approximating distributions are actually a Wightman functional?

(2) Can any of the forms of convergence in examples 1 and 2 of section I be related to convergence of Wightman functionals?

ACKNOWLEDGMENTS

The writer is grateful for discussions with H. Epstein, O. Lanford, D. Prill, Professor A. Wightman, and M. Zerner.

The author wishes to thank Professor L. Motchane for hospitality at the Institut des Hautes Etudes Scientifiques, and Professor M. Lévy for hospitality at the Institut des Etudes Scientifiques de Cargèse. He is grateful to the National Science Foundation for a Fellowship.

[10] A. M. Jaffe, Ann. Phys. (N. Y.) **32**, 127 (1965), Lemma 4. Other applications can be found in this reference.

[11] A. S. Wightman, *Cargèse, Corsica Lectures* (1964) (to be published).

Announcement: ZIP Code Information

THE U. S. Post Office Department has announced that the use of ZIP Codes will be mandatory on all U. S. addresses for subscriptions and other mailings by 1 January 1967. Accordingly, we solicit your cooperation in providing your ZIP Code to the American Institute of Physics.

In September 1965, the AIP will send renewal invoices to all members and renewal order forms to all nonmember subscribers. Instructions will be given requesting that your ZIP Code be furnished at the time of sending your remittance. Your compliance with this request will enable the AIP to avoid an expensive, time-consuming task of looking up the ZIP Code to your address.

Please do not send the AIP a request for a change of address merely to add your ZIP Code. Your only response to this notice should be your supplying the ZIP Code information requested when you send in your renewal.