Diagrammatic methods and renormalization in the Iwamoto–Yamada cluster expansion

Lothar Schäfer

Institut für Theoretische Physik, Abteilung Kernphysik, Universität Heidelberg, Heidelberg, Germany
(Received 14 February 1973)

We develop diagrammatic methods which give the explicit expression for the nth order term of the linked Iwamoto–Yamada cluster expansion in terms of the cluster integrals. The method applies independently of the number A of particles of the system. We carry through partial summations which for finite A result in a representation of the expansion by skeleton diagrams, whereas in the limit $A \to \infty$ the expansion is reduced to two analytical equations. In addition, for finite A the renormalization establishes a connection between the expectation value of an arbitrary operator and the probability to find n hole states, $1 \leq n \leq A$, occupied. Simple considerations raise some doubts concerning the convergence of the unrenormalized expansion.

1. INTRODUCTION

The last years have seen a considerable revival of the interest in the application of cluster expansion methods to nuclear systems. The cluster expansion expresses the expectation value $\langle\Theta\rangle = \langle\Psi|\Theta|\Psi\rangle/\langle\Psi|\Psi\rangle$ of an operator Θ by a series of terms the nth of which, roughly speaking, gives the contribution to $\langle\Theta\rangle$ of n-particle correlations contained in the wavefunction Ψ. Several different cluster expansions have been established. Most applications employ the expansion of Iwamoto and Yamada[1] (I.Y.) or its factorized version,[2] which is intimately related[3–8] to the hole-line expansion of Brueckner theory. It has been found that variational calculations based on that method compete successfully with Brueckner-type calculations. For these reasons we concentrate in the following on the I.Y. expansion.

Up to now a discussion of the general structure of the I.Y. expansion has not been given since in contrast to perturbation theory no general diagrammatic representation is available. Instead methods have been developed[2,9,10] which allow the calculation of every term of the expansion without giving the explicit form of the nth order term. Three exceptions must be mentioned. Feenberg and Wu[11,12] have established a graphical representation of the I.Y. expansion which, however, is limited to *infinite* systems. Providencia[13] has given a diagrammatic representation of a cluster expansion which in contrast to the I.Y. expansion uses nondiagonal cluster integrals. Recently, Gaudin *et al.*[14] have given a simple formulation of the I.Y. expansion which is restricted to the original Jastrow choice of the wavefunction. In all these approaches the possibility of performing partial summations was not considered (see, however, Ref. 15).

Here we develop a diagrammatic representation of the I.Y. expansion which is applicable to *finite and infinite* systems and to *any form* of the wavefunction. We use rather straightforward diagrammatic methods which are very similar to those of perturbation theory. Our derivation can be divided into the following steps. (i) In Sec. 2 we establish the formal representation of the generalized normalization integral $I(\beta) = \langle\Psi|\exp(\beta\Theta)|\Psi\rangle$ in terms of the cluster integrals ("unlinked" cluster expansion). The form of the wavefunction influences only the cluster integrals but not the formal representation of $I(\beta)$ in terms of these quantities. All the following steps depend only on the formal structure of the unlinked expansion. (ii) In full analogy with the unlinked perturbation expansion, the unlinked cluster expansion for any $m > 0$ contains terms which behave like A^m with increasing particle number A. Thus this expansion is useful only for small A. We derive an exponential form which constitutes a linked expansion with uniform A dependence (Sec. 3). (iii) Using the relation $\langle\Theta\rangle = (d/d\beta)\times \ln I(\beta)|_{\beta=0}$ we arrive at the linked I.Y. expansion which can be expressed in terms of "unlabeled" diagrams (Sec. 4).

For an infinite system the expansion can be resummed (Sec. 5) and the resulting "renormalized" expansion is expressed in terms of two analytic equations.[16] Similar manipulations can be carried through for a finite system, leading either to an interesting identity (Sec. 6A) or to the representation of the expectation value by "skeleton diagrams" (Sec. 6B). In Sec. 7 we give some remarks on convergence questions, and Sec. 8 contains a comparison with other approaches. In the Appendix we present diagrammatic methods for the evaluation of the cluster integrals starting either from the product ansatz of Ψ or from the particle–hole representation.

We have tried to state clearly all essential steps of the derivation. We have skipped, however, the proofs of some statements. These proofs are rather elementary, though in some cases a mathematically rigorous presentation will become lengthy.

2. THE UNLINKED I.Y. EXPANSION

Here we establish the unlinked cluster expansion following Ref. 2. For definiteness we use the product form of Ψ which is a generalization of the Jastrow form[17] consisting of a product of state-dependent n-particle correlation functions $f_{i_1\cdots i_n}(1,\ldots,n)$ multiplied by a set of orthonormal single-particle functions $\varphi_i(i)$:

$$\Psi(1,\ldots,A) = \mathfrak{A}\Big\{\prod_n\Big[\prod_{i_1<\cdots<i_n} f_{i_1\cdots i_n}(i_1,\ldots,i_n)\Big]\prod_i \varphi_i(i)\Big\}. \tag{2.1}$$

If the range is not explicitly given, summations and products over the indices $i,j,\ldots,n$ always range from 1 to A, and the argument (i) denotes the space, spin, and isospin coordinates of the ith particle. The symbol $\mathfrak{A}$ denotes the antisymmetrization operator.

In addition we define functions $\Psi_{K_1\cdots K_m}$, $K_1 < \cdots < K_m \leq A$, $m \leq A$, by the relation

$$\Psi_{K_1\cdots K_m}(K_1,\ldots,K_m) = \mathfrak{A}\Big\{\prod_{n=1}^{m}\Big[\prod_{i_1<\cdots<i_n}^{<K_1,\ldots,K_m>} f_{i_1\cdots i_n}(i_1,\ldots,i_n)\Big]\prod_i^{<K_1,\ldots,K_m>}\varphi_i(i)\Big\}. \tag{2.2}$$

The symbol $\prod_{i_1<\cdots<i_n}^{<K_1,\ldots,K_m>}$ ($\sum_{i_1\cdots i_n}^{<K_1,\ldots,K_m>}$) denotes the product (sum) over all possible terms in which the

$i_1, \ldots, i_n$ are taken out of the set $\langle K_1, \ldots, K_m \rangle$. We introduce the quantities

$$J_i \equiv J_i(\beta) = \langle \varphi_i \mid \exp(\beta \Theta^{(1)}) \mid \varphi_i \rangle, \tag{2.3}$$

where $\Theta^{(1)}$ denotes the one-body part of the operator Θ, and we define the (reduced) subnormalization integrals by

$$I^R_{K_1 \cdots K_m} \equiv I^R_{K_1 \cdots K_m}(\beta) = \langle \Psi_{K_1 \cdots K_m} \mid e^{\beta\Theta} \mid \Psi_{K_1 \cdots K_m} \rangle \times \prod_i^{<K_1, \ldots, K_m>} J_i^{-1}, \quad K_1 < \cdots < K_m \leq A. \tag{2.4}$$

The reduced normalization integral, which we want to calculate, is given by

$$I^R(\beta) = I^R_{1 \cdots A}(\beta) = \langle \Psi \mid e^{\beta\Theta} \mid \Psi \rangle \times \prod_i J_i^{-1}. \tag{2.5}$$

The (reduced) cluster integrals $x_{K_1 \cdots K_n}$, $1 \leq n \leq A$, all $K_i \leq A$, are defined by the following requirements:

(i) $x_{K_1 \cdots K_n}$ is symmetric in its indices.

(ii) $x_{K_1 \cdots K_n} = 0$ if any two (or more) indices are equal.

(iii) $$I^R_K = 1 + x_K, \quad K \leq A, \tag{2.6}$$

$$I^R_{K_1 K_2} = 1 + x_{K_1} + x_{K_2} + x_{K_1} x_{K_2} + x_{K_1 K_2}, \quad K_1 < K_2 \leq A; \tag{2.7}$$

in general

$$I^R_{K_1 \cdots K_n} = \sum_{l_1 \cdots l_n = 0}^{\infty} S^{K_1 \cdots K_n}\{l_1, \ldots, l_n\}, \quad K_1 < \cdots < K_n \leq A \tag{2.8}$$

$$S^{K_1 \cdots K_n}\{l_1, \ldots, l_n\} = 1 \quad \text{if all } l_p = 0; \tag{2.9}$$

otherwise

$$S^{K_1 \cdots K_n}\{\mathbf{l}\} = S^{K_1 \cdots K_n}\{l_1, \ldots, l_n\} = \prod_{m=1}^{n} (m!)^{-l_m} (l_m!)^{-1} \times \sum_{i_1 \cdots i_r}^{<K_1, \cdots, K_n>} \{x_{i_1} \times \cdots \times x_{i_{l_1}}\} \times \{x_{i_{l_1+1} i_{l_1+2}} \times \cdots \times x_{i_{l_1+2 l_2-1} i_{l_1+2 l_2}}\} \times \cdots, \quad \times \; r = \sum_m m l_m. \tag{2.10}$$

Each term on the rhs of Eq. (2.10) consists of products of l_m cluster integrals with m indices, $1 \leq m \leq n$, and we sum over all sets of indices $i_1 \cdots i_r$ restricted by

$$i_K \neq i_j \quad \text{if } K \neq j. \tag{2.11}$$

By virtue of the symmetry (i) of the cluster integrals and of the fact that the sum is not changed if we permute the factors within the curly brackets, the definition given here is identical to that of Ref. 2. Combined with Eq. (2.5), Eqs. (2.8)–(2.10) represent the unlinked cluster expansion of $I^R(\beta)$. The formal structure of the expansion is completely independent of the choice of Ψ. All that is required is the definition of a set of subnormalization integrals.

The form of the unlinked cluster expansion is not as arbitrary as the derivation, given above, might suggest. We illustrate this by giving a short account of the physical ideas[1] which have led to the creation of that expansion. The guiding principle is the assumption that the correlations are of short range, i.e., that $|f_{i_1 \cdots i_n}(1, \ldots, n)|^2$ tends to 1 rapidly if the distance between any pair of coordinates tends to infinity. (This statement applies rigorously[5] only in infinite systems.) Expanding $I^R(\beta)$ in powers of the quantity $(|f|^2 - 1)$ and carrying through the integrations, we find that the general term of that expansion separates into a product of integrals. This separation automatically leads to the unlinked cluster expansion,[14] and the cluster integrals are intimately related to the irreducible integrals which emerge from that procedure. The cluster integral with p indices describes the contribution to $I^R(\beta)$ which arises from the correlations of p particles. If the correlations are of short range, we expect that the probability of finding p correlated particles will decrease rapidly as p increases, and we thus hope that the contribution of $x_{i_1 \cdots i_p}$ to $I^R(\beta)$ will decrease rapidly with increasing p. More precise statements on the structure of the cluster integrals are given in the Appendix.

3. DERIVATION OF THE LINKED EXPANSION

It is well known that the contribution of $x_{i_1 \cdots i_n}$, if summed over all indices, asymptotically increases proportional to the number A of particles. (This statement is based on the short range nature of the correlation functions. See the Appendix for a proof.) Thus a term of the unlinked expansion containing a product of m cluster integrals increases with increasing m like A^m. As a consequence for big A the unlinked expansion is not directly useful and we must resum it into a linked expansion.

We first have to get rid of condition (2.11) which restricts the summation on the rhs of Eq. (2.10). We formally define a set of Fermion operators a_i, a_i^+ and multiply each term of the sum by $\langle 0 | a_{i_1} \cdots a_{i_r} a^+_{i_r} \cdots a^+_{i_1} | 0 \rangle$, where $|0\rangle$ denotes the vacuum state. This yields the expression

$$I^R(\beta) = \sum_{l_1 l_2 \cdots = 0}^{\infty} S\{\mathbf{l}\}, \tag{3.1}$$

$$S\{\mathbf{l}\} = \prod_{m=1}^{\infty} (m!)^{-l_m} (l_m!)^{-1} \times \sum_{i_1 \cdots i_r} \langle 0 | a_{i_1} \cdots a_{i_r} a^+_{i_r} \cdots a^+_{i_1} | 0 \rangle \{{}^{l_1}x \times \cdots \times {}^{l_1}x\} \times \{x_{i_{l_1+1} i_{l_1+2}} \times \cdots \times x_{i_{l_1+2 l_2-1} i_{l_1+2 l_2}}\} \times \cdots. \tag{3.2}$$

$S\{\mathbf{0}\}$ and r are defined as in Eqs. (2.9), (2.10), respectively, and we have omitted the upper indices $1, \ldots, A$ on $S\{\mathbf{l}\}$. We have extended the definition of $S\{\mathbf{l}\}$ to include also terms with $r > A$ which vanish identically.

According to Wick's theorem[18] the expectation value $\langle 0 \mid a_{i_1} \cdots a_{i_r} a^+_{i_r} \cdots a^+_{i_1} \mid 0 \rangle$ is given by the sum of all fully contracted terms. All nonvanishing contractions are of the form $\underbrace{a_i a^+_K} = \delta_{iK}$. In terms of diagrams we represent a cluster integral $x_{i_1 \ldots i_n}(\beta)$ by a vertical beam carrying n points which from bottom to top are labelled by the indices i_1 to i_n. The set $\{\mathbf{l}\}$ is represented by a collection of l_m m-point beams, $m = 1, 2, \cdots$, drawn in such a way that all m-point beams stand left of all n-point beams if $m < n$. The points are labelled by indices i_K in such a way that within each beam K increases from bottom to top and going from one beam to the next K increases from left to right (natural labelling, see Fig. 1). The contraction $\underbrace{a_i a^+_K}$ is represented by a directed line starting at point K and ending at point i. Each fully contracted term gives rise to a diagram

which has the property that in each point there starts exactly one line and there ends exactly one line. The contraction lines thus form closed loops and according to point (ii) of the definition of $x_{i_1 \cdots i_n}$ no loop may touch a given beam more than once. In the diagrammatic representation of $[I^R(\beta) - 1]$ all diagrams with these properties occur exactly once. The rules for calculating the contribution of a diagram are obvious from our construction, except for the sign which is given by $(-1)^{\sum_{\tau=1}^{t}(r_\tau - 1)} = (-1)^{r-t}$. Here $r(t)$ denotes the number of points (closed loops) of the diagram, and r_τ denotes the number of points which are connected by the τth loop. This sign rule can easily be deduced from the ordering of the operators a^+, a in the matrix element $\langle 0 | a_{i_1} \cdots a_{i_r} a^+_{i_r} \cdots a^+_{i_1} | 0 \rangle$. Examples of diagrams are given in Fig. 1.

We introduce the notion of a connected part of a diagram.

Definition 3.1: A connected part is any part of a diagram with the following properties: (i) It cannot be separated into two parts without cutting contraction lines or a beam. (ii) We cannot include more beams or contraction lines without invalidating property (i).

We introduce an arbitrary fixed ordering of the sets $\{1\}$ which is indicated by an index (s):

$$\{1\} \to \{1\}^{(s)} \equiv \{ l_1^{(s)}, l_2^{(s)}, \cdots \}, \quad s = 1, 2, \cdots .$$

The sum of all diagrams of class $\{1\}^{(s)}$ which consist of exactly one connected part (connected diagrams) is denoted by $S_c^{(s)}$. We define quantities $A^{(s)}$ by

$$A^{(s)} = [\prod_m (m!)^{l_m^{(s)}} l_m^{(s)}!] \times S_c^{(s)}. \tag{3.3}$$

We consider all those diagrams of class $\{1\}^{(s_0)}, l_m^{(s_0)} = \sum_{s=1}^{\infty} \nu_s l_m^{(s)}$, which consist of $\nu_s \geq 0$ connected parts of type $\{1\}^{(s)}$, $s = 1, 2, \cdots$. The contribution of all those diagrams in which the $l_m^{(s_0)}$ m-point beams, $m = 1, 2, \cdots$, are distributed to these connected parts according to a fixed scheme is given by

$$[\prod_m (m!)^{-l_m^{(s_0)}} (l_m^{(s_0)}!)^{-1}] \prod_s (A^{(s)})^{\nu_s}, \tag{3.4}$$

as is obvious from the rules given above. There are $l_m^{(s_0)}! \times \Pi_s (l_m^{(s)}!)^{-\nu_s}$ possibilities of distributing the $l_m^{(s_0)}$ m-point beams to the different connected parts, and by taking all these possibilities into account each diagram is counted $\Pi_{s=1}^{\infty} \nu_s!$ times. Thus the contribution of all diagrams which consist of ν_s connected parts of type s is given by

$$[\prod_m (m!)^{-l_m^{(s_0)}} (l_m^{(s_0)}!)^{-1}] \times [\prod_n l_n^{(s_0)}! \prod_s (l_n^{(s)}!)^{-\nu_s}] \times \prod_{s'} (\nu_{s'}!)^{-1} [A^{(s')}]^{\nu_{s'}} = \prod_s (\nu_s!)^{-1} [S_c^{(s)}]^{\nu_s}. \tag{3.5}$$

Since $[I^R(\beta) - 1]$ is given by the sum of all different diagrams, we find

$$I^R(\beta) = \exp[G(\beta)], \tag{3.6}$$

$$G(\beta) = \sum_{s=1}^{\infty} S_c^{(s)}. \tag{3.7}$$

These equations constitute the linked cluster expansion (L.C.E.).

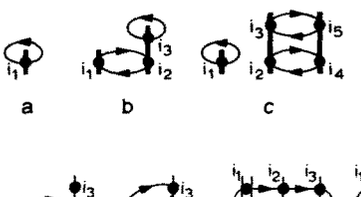


FIG. 1. Some diagrams contributing to $I^R(\beta)$ representing the following terms: a) $\sum_{i_1} x_{i_1}(\beta)$; b) $-1/2 \sum_{i_1 i_2 i_3} x_{i_1}(\beta) \times x_{i_2 i_3}(\beta) \delta_{i_1 i_2}$; c) $1/8 \sum_{i_1 \cdots i_5} x_{i_1}(\beta) x_{i_2 i_3}(\beta) x_{i_4 i_5}(\beta) \delta_{i_2 i_4} \delta_{i_3 i_5}$.

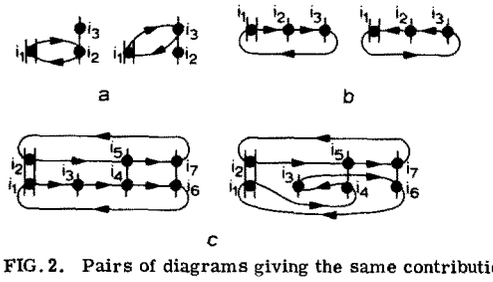


FIG. 2. Pairs of diagrams giving the same contribution, respectively.

4. THE EXPANSION OF THE EXPECTATION VALUE; EQUIVALENT DIAGRAMS

Equations (2.5) and (3.6) show that the expectation value $\langle \Theta \rangle$ is given by

$$\langle \Theta \rangle = \sum_i \frac{d}{d\beta} J_i(\beta) \Big|_{\beta=0} + \frac{d}{d\beta} G(\beta) \Big|_{\beta=0}. \tag{4.1}$$

The differentiation of a diagram of class $\{1\}$ yields $\sum_m l_m$ diagrams which differ in the position of the beam which has been selected for the differentiation. We represent the differentiated cluster integral $x'_{i_1 \cdots i_n} \equiv d/d\beta \, x_{i_1 \cdots i_n}(\beta) \mid_0$ by a double beam, and we alter the ordering of the beams in such a way that the double beam stands left of all others. The thin simple beams now represent factors $x_{i_1 \cdots i_n} \equiv x_{i_1 \cdots i_n}(0)$. We define the class $\{n;1\}$ to contain all different connected diagrams which consist of one n-point double beam (standing left of all other beams) and l_m m-point simple beams, $m = 1, 2, \cdots$, ordered as before. A given diagram of class $\{n;1\}$ is created exactly $(l_n + 1)$ times by differentiation of diagrams of class $\{1'\}$, $l'_m = l_m$ if $m \neq n$, $l'_n = l_n + 1$. Thus the following theorem holds.

Theorem 4.1: The quantity $(d/d\beta) G(\beta) \mid_0$ is given by the sum of the contributions of all classes $\{n;1\}$. The weight factor of the diagrams of class $\{n;1\}$ is equal to $(n!)^{-1} \Pi_{m=1}^{\infty} (l_m!)^{-1} (m!)^{-l_m}$.

The diagrammatic representation of $\langle \Theta \rangle$ established by Eq. (4.1) and Theorem 4.1 contains a lot of diagrams which give the same contribution. This is due to the fact that (A) permutations of points within a beam (Fig. 2a) or (B) permutations of simple beams (Fig. 2b) do not alter the contribution of the diagram. In addition (C) all diagrams in which the same sets of points are connected by closed loops give the same contribution (Fig. 2c). We evaluate the consequences of the symmetries A and B. Symmetry C eliminates the direction of the contraction lines. This direction, however, will prove useful in partial summations. Besides this in many diagrams the symmetries B and C coincide, as can be seen, e.g., from Fig. 2b.

In the following we are concerned with diagrams of an arbitrary fixed class $\{n;1\}$. We define sets C_A and C_B of operators which express the symmetries A and B in a more precise form.

Definition 4.1: The set C_A contains all mappings P_A of $\{n;1\}$ onto $\{n;1\}$ which are defined by the following steps:

(i) Permutation of the labels of the points within each (double or simple) beam separately.

(ii) Shifting of the points with the contraction lines fixed to them along the beams in such a way that the natural labelling (see Fig. 2) is restored.

Definition 4.2: The set C_B contains all mappings P_B of $\{n;1\}$ onto $\{n;1\}$ which are defined in the following way:

(i) Permutation of the labels of all points of the respective beams among different simple m-point beams. The relative ordering of the labels within each beam has to be preserved.

(ii) Shifting of the beams with fixed contraction lines in such a way that the natural labelling is restored.

The set $C_A[C_B]$ contains $n!\,\Pi_m (m!)^{l_m}[\Pi_m l_m!]$ elements. The mappings P_A and P_B are special permutations of the labels of the points, except for the shifting of points or beams. Thus the following statements are obvious.

Lemma 4.1:

(i) Let α denote either A or B. If $P_\alpha, P'_\alpha \in C_\alpha$, then $P_\alpha \times P'_\alpha \in C_\alpha$. There exists an operation $P_\alpha^{-1} \in C_\alpha$ with $P_\alpha \times P_\alpha^{-1} = I$. Here I denotes the identity operation.

(ii) $C_A \cap C_B = I$.

(iii) If $P_B \in C_B$, $P_A \in C_A$, then $P_B^{-1} \times P_A \times P_B = \tilde{P}_A \in C_A$.

As can be seen from Lemma 4.1 the operations $P_A \times P_B$ induce an equivalence relation among the diagrams of class $\{n;1\}$.

Definition 4.3: Two elements $x, x' \in \{n;1\}$ are equivalent ($x \sim x'$) if there exist operations $P_A \in C_A, P_B \in C_B$ such that $P_A \times P_B x = x'$. We denote by $\{x\}$ the class of diagrams which are equivalent to x.

Definition 4.4: The symmetry number $S(x)$ of $x \in \{n;1\}$ is equal to the number of operations $P_A \times P_B$; $P_A \in C_A$, $P_B \in C_B$ with $P_A \times P_B x = x$.

Lemma 4.2: If $x \sim x'$, then $S(x) = S(x')$. For fixed x the operations $P = P_A \times P_B$, $P_A \in C_A$, $P_B \in C_B$, can be grouped together into groups of $S(x)$ elements such that $Px = P'x$ if and only if P and P' belong to the same group.

Theorem 4.2: The equivalence relation (Definition 4.3) divides the class $\{n;1\}$ into subclasses $\{x\}$ of diagrams giving the same contribution. The number of elements of $\{x\}$ is equal to $n!\,\Pi_{m=1}^{\infty}(m!)^{l_m} l_m! \times [S(x)]^{-1}$.

The symmetry number can be calculated by $S(x) = S_A(x) \times S_B(x)$. Here $S_A(x)$ is equal to the number of operations $P_A \in C_A$ with $P_A x = x$, and $S_B(x)$ is given by the number of operations $P_B \in C_B$ with the property that there exists a $P_A \in C_A$ with $P_A \times P_B x = x$. Our results can be summarized by the following rules.

Rule 4.1: Construction of diagrams: A diagram consists out of one double beam and any number of simple beams. The simple beams are ordered at the rhs of the double beam in such a way that the number of points of the simple beams does not decrease as we go from left to right. Each diagram is completely connected by directed contraction lines. In each point of a beam there starts exactly one line and there ends exactly one line. No closed loop of contraction lines may touch a given beam more than once.

Rule 4.2: Evaluation of diagrams:

(i) Label the points of the diagram in such a way that different points carry the same label if and only if they belong to the same loop.

(ii) An n-point double (simple) beam carrying indices $i_1, \ldots, i_n$ represents a factor $x'_{i_1 \ldots i_n}$ ($x_{i_1 \ldots i_n}$).

(iii) Sum over all indices and multiply by $S^{-1} \times (-1)^{t+r}$. Here t (r) denotes the number of loops (points) of the diagram and S denotes the symmetry number.

The quantity $(d/d\beta)\,G(\beta)\,|_0$ is given by the sum of the contributions of all inequivalent connected diagrams. In Fig. 3 we have given all diagrams containing up to four points. We have omitted the directions of the lines where they are not necessary, and we represent a loop which contains only one point by the point itself. Using our method, we have checked the contributions to $\langle \mathcal{K} \rangle$ evaluated in Ref. 19.

5. RENORMALIZED EXPANSION FOR AN INFINITE SYSTEM

In the limit $A \to \infty$ it is possible to resum the L.C.E. in such a way that the result is expressed by two analytic equations.[16] This is due to the fact that in that limit most of the diagrams vanish, as is shown by the following considerations.

Definition 5.1: A connected diagram is called simply connected if we cannot cut any closed loop by two cuts without dividing the diagram into disconnected pieces. Otherwise the diagram is called multiply connected.

Lemma 5.1: If the asymptotic order of magnitude of $x'_{i_1 \ldots i_n}$ for any n is equal to A^α, then the asymptotic order of any simply (multiply) connected diagram is equal to A^α ($A^{\alpha - K}$, $K \geq 1$).

Lemma 5.1 expresses a well-known fact, the proof of which rests upon the properties of the cluster integrals (see the Appendix). It implies that in the limit $A \to \infty$ we can neglect all multiply connected diagrams (provided that the L.C.E. converges uniformly in A). Besides this in an infinite system the $x_i(\beta)$ vanish identically by virtue of momentum conservation.

The structure of a diagram which contributes in the limit $A \to \infty$ is shown in Fig. 4a. The shaded boxes denote insertions, i.e., simply connected diagrams which instead of a double beam contain one external point with fixed index (Fig. 4b). We explicitly admit the trivial insertion which contains no beam.

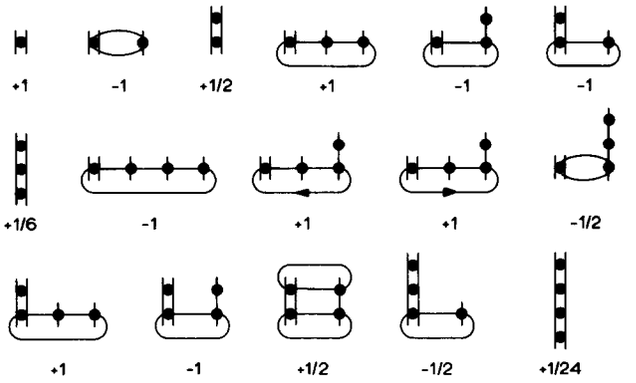


FIG. 3. All diagrams up to the order of four points which contribute to $\langle \mathcal{H} \rangle$. Below each diagram there is given the corresponding sign and symmetry factor.

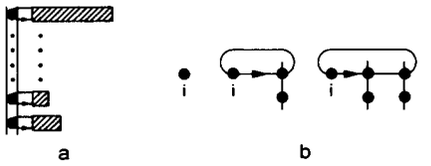


FIG. 4. a) Structure of a diagram contributing in the limit $A \to \infty$. b) Examples of insertions. These diagrams contribute the following terms: $+1; -\sum_j x_{ij}; +\sum_{j_1 j_2} x_{ij_1} x_{ij_2}$.

Definition 5.2: The weight factor g_i is given by the sum of all inequivalent simply connected insertions in which the external point carries the label i. The diagrams are to be evaluated according to rule 4.2, and the external point has to be counted in the sign rule.

With due regard to Definition 4.4 of the symmetry number, it is straightforward to prove the following theorem.

Theorem 5.1:

$$\langle \Theta \rangle = \sum_i \frac{d}{d\beta} J_i \Big|_{\beta=0} + \sum_{n=2}^{\infty} \frac{1}{n!} \sum_{i_1 \cdots i_n} x'_{i_1 \cdots i_n} \prod_{K=1}^{n} g_{i_K}. \tag{5.1}$$

The weight factor g_i obeys the equation

$$g_i = 1 - \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{j_1 \cdots j_n} x_{i j_1 \cdots j_n} g_i \prod_{K=1}^{n} g_{j_K}. \tag{5.2}$$

Equation (5.2) can be proved by induction with respect to the number of cluster integrals. In that proof the directions of the lines are important as they establish a one-to-one relation between the terms found by iteration of Eq. (5.2) and the insertion diagrams: If the line that starts at the external point ends at the beam $x_{i j_1 \cdots j_n}$, then the diagram contributes to $x_{i j_1 \cdots j_n} \times g_i \prod_{K=1}^{n} g_{j_K}$. Theorem 5.1 establishes the renormalized cluster expansion. It holds independently of the form of Ψ or of Θ.

A comparison of Eqs. (5.1) and (5.2) shows that g_i can be interpreted as the expectation value of an operator which fulfils the relations [see also Ref. 19, Eqs. (44)–(46)]

$$\frac{d}{d\beta} J_j \Big|_{\beta=0} = \delta_{ij}, \tag{5.3}$$

$$x'_{i_1 \cdots i_n} = - x_{i_1 \cdots i_n} \Big\{ \sum_{K=1}^{n} \delta_{i i_K} \Big\}. \tag{5.4}$$

If we use a special form of Ψ such an operator can easily be identified [see Eq. (6.1) and the Appendix].

Theorem 5.2: If Ψ is represented in the particle-hole form $\Psi = \exp(S)\Phi$ (see the Appendix for a precise definition), the weight factor g_i becomes equal to the occupation probability $p_i = \langle \Psi \mid a_i^+ a_i \mid \Psi \rangle / \langle \Psi \mid \Psi \rangle$ of the ith single-particle level.

The renormalization by occupation probabilities is a well-known concept in perturbation theory. In the framework of a cluster expansion, it has first been put forward by Providencia and Shakin.[15] Equation (5.2) is implicit in the approach of Feenberg and Wu [see Ref. 12, Eq. (14)] who used it as an ansatz for establishing the I.Y. expansion in the limit $A \to \infty$. These authors, however, did not realize the important role the quantity g_i plays in connection with partial summations. From an analysis of the first five orders of the factorized I.Y. expansion and from an inspection of the results of Refs. 12, 15, the equations (5.1) and (5.2) were recently proposed by Ristig and Clark,[19,20] independently of the present author. These authors also give an interesting formulation of the renormalized expansion which exhibits a relation to expansion methods used in statistical mechanics.

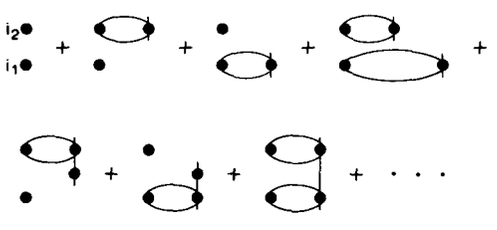


FIG. 5. All diagrams up to the order of two internal points which contribute to $p_{i_1 i_2}$.

6. RENORMALIZED EXPANSIONS FOR A FINITE SYSTEM

We discuss two different possibilities of generalizing the results of Sec. 5 to finite systems. In subsection 6A we sum up all diagrams which contain a fixed n-point double beam. This procedure is useful only if the resulting weight factor has a physical meaning, and we therefore restrict our discussion to the particle–hole form of Ψ. In subsection 6B we give a representation of $\langle \Theta \rangle$ by skeleton diagrams with renormalized lines which is applicable to any form of Ψ.

A. Summation of all diagrams containing $x'_{i_1 \cdots i_n}$

We first evaluate the expectation values $p_{i_1 \cdots i_n}$ of the operators $P_{i_1 \cdots i_n} = a_{i_1}^+ \cdots a_{i_n}^+ a_{i_n} \cdots a_{i_1}$ with respect to $\Psi = \exp(S)\Phi$. Denoting the corresponding differentiated cluster integrals by $x^{i_1 \cdots i_n}_{j_1 \cdots j_m}$, we find from Appendix A2

$$x^{i}_{j_1 \cdots j_m} = - x_{j_1 \cdots j_m} \Big\{ \sum_{K=1}^{m} \delta_{i j_K} \Big\}, \tag{6.1}$$

$$x^{i_1 \cdots i_n}_{j_1 \cdots j_m} = \delta_{nm} \sum_{P} \prod_{K=1}^{n} \delta_{i_{P(K)} j_K}, \qquad n \geq 2. \tag{6.2}$$

In Eq. (6.2) the sum ranges over all permutations P of the numbers 1 to n. Due to the simple form of Eqs. (6.1) and (6.2) we can represent $p_{i_1 \cdots i_n}$ by diagrams in which the double beams are replaced by external points.

Lemma 6.1: The expectation value $p_{i_1 \cdots i_n}$ is given by the sum of the contributions of all inequivalent diagrams containing n external points i_1 to i_n ordered in such a way that i_l stands below i_K if $l < K$. Both in calculating the symmetry number and in defining the equivalence of diagrams we are not allowed to permute the labels of the external points. The diagrams need not be connected. They must be connected, however, if we connect the external points by a beam.

In Fig. 5 we give the first diagrams contributing to $p_{i_1 i_2}$. We should note that in going from the representation employing double beams to the representation using external points the symmetry number is altered. This effect can be handled by splitting the symmetry operations $P_A \times P_B$ into one factor which does not affect the points of the double beam and an other factor which affects only these points. Defining $p^c_{i_1 \cdots i_n}$, $n \leq 2$, by the sum of the contributions of all *connected* diagrams with the external points i_1 to i_n, we find from Lemma 6.1 (see also Fig. 5)

$$p_{i_1 i_2} = p_{i_1} p_{i_2} + p^c_{i_1 i_2}, \tag{6.3}$$

$$p_{i_1 i_2 i_3} = p_{i_1} p_{i_2} p_{i_3} + p_{i_1} p^c_{i_2 i_3} + p_{i_2} p^c_{i_1 i_3} + p_{i_3} p^c_{i_1 i_2} + p^c_{i_1 i_2 i_3}, \tag{6.4}$$

and so on. We will come back to these results in Sec. 7.

Turning to the evaluation of the expectation value of a general operator Θ, we first discuss the contribution $\sigma^{(1)}$ of the set $\{\alpha\}$ of all diagrams containing a one point double beam. Substituting the double beam by the external point (i) we establish a one-to-one mapping $f^{(i)}$ of $\{\alpha\}$ onto the set $\{P_i\}$ of diagrams contributing to p_i. If the contribution of $y \in \{\alpha\}$ is equal to

$y = \sum_j x'_j A_i$, then the contribution of $f^{(i)}y$ is given by $f^{(i)}y = A_i$. Since the range of $f^{(i)}$ exhausts the set $\{P_i\}$, we find

$$\sigma^{(1)} = \sum_j x'_j p_j . \tag{6.5}$$

The contribution to $\langle\Theta\rangle$ of the set of all diagrams containing a n-point double beam ($n \geq 2$) can be evaluated in the same fashion, using however the diagrams containing the double beam $p^{i_1 \cdots i_n}_{j_1 \cdots j_n}$ and evaluating the diagrams by virtue of Eq. (6.2). We find the following result.

Theorem 6.1: The expectation value of any operator Θ with respect to $\Psi = \exp(S)\Phi$ can be represented by

$$\langle\Theta\rangle = \sum_i \frac{d}{d\beta} J_i \Big|_{\beta=0} + \sum_{n=1}^{A} \frac{1}{n!} \sum_{i_1 \cdots i_n} x'_{i_1 \cdots i_n} p_{i_1 \cdots i_n} . \tag{6.6}$$

Applying Theorem 6.1 to P_i and using Eq. (6.1), we find

$$p_i = 1 - \sum_{n=0}^{A-1} \frac{1}{n!} \sum_{j_1 \cdots j_n} x_{i j_1 \cdots j_n} p_{i j_1 \cdots j_n} \tag{6.7}$$

Introducing into Eq. (6.6) the quantities

$$X_{i_1 \cdots i_n} = x'_{i_1 \cdots i_n} + x_{i_1 \cdots i_n} \sum_{K=1}^{n} \frac{d}{d\beta} J_{i_K} \Big|_{\beta=0} \tag{6.8}$$

and using Eq. (6.7), we find a representation of $\langle\Theta\rangle$ where also the term $(d/d\beta)J_i|_{\beta=0}$ is renormalized:

$$\langle\Theta\rangle = \sum_i \frac{d}{d\beta} J_i \Big|_{\beta=0} \times p_i + \sum_{n=1}^{A} \frac{1}{n!} \sum_{i_1 \cdots i_n} X_{i_1 \cdots i_n} p_{i_1 \cdots i_n} . \tag{6.9}$$

Equations (6.6) and (6.7) are generalizations of the results expressed in Theorem 5.1 and Eqs. (5.1) and (5.2).

Equation (6.9) resembles a result given by Providencia[13] in the framework of his nondiagonal cluster expansion. Equations (6.6), (6.7), and (6.9) hold identically, irrespective of convergence problems. This is proved by noting that these equations can be derived directly from the unlinked cluster expansions of the expectation values involved.

B. Summation of insertions

Any part of a diagram which does not contain the double beam and which can be separated from the rest of the diagram by two cuts affecting one closed loop is called an insertion. We close the cut loop via an external point which is drawn left of all beams. We explicitly admit the trivial insertion which consists only of the external point with the attached contraction line and which gives the contribution (+1). The sum of all insertions containing the external point (i) is denoted by g_i. In the special case $A \to \infty$ this definition coincides with Definition 5.2. A diagram which does not contain any nontrivial insertion is called a *skeleton diagram*. A *skeleton insertion* is defined to be an insertion which can not be separated into two nontrivial insertions by twice cutting one closed loop.

Theorem 6.2: The expectation value $\langle\Theta\rangle$ is given by the sum of the contributions of all inequivalent skeleton diagrams where each line with summation index i carries a factor g_i. The weight factor g_i is given by the sum of all inequivalent skeleton insertions where each line j is renormalized by a factor g_j except for the line *starting* at the external point. The symmetry number and the equivalence of diagrams are defined in exactly the same way as in the unrenormalized theory.

The proof of Theorem 6.2 is simplest if one first derives the corresponding results for the labeled diagrams introduced in Sec. 3. The results for unlabeled diagrams then follow trivially since the symmetry operations defined in Sec. 4 are not affected by introducing renormalization factors.

In Fig. 6 we give the first terms of $\langle\Theta\rangle$ and of g_i as determined from Theorem 6.2. Theorem 6.2 is another generalization of Theorem 5.1. It holds for any form of Ψ. Using $\Psi = \exp(S)\Phi$, we find from Lemma 6.1 that the factor g_i coincides with the occupation probability p_i.

7. SOME SIMPLE REMARKS ON THE CONVERGENCE PROBLEM

Since for $A = \infty$ the cluster expansion can be expressed analytically, it does not seem hopeless to discuss convergence questions, at least for infinite systems. Here we will not tackle that problem but rather point out some fairly obvious facts which may shed some light on certain questions involved in a rigorous treatment.

We restrict ourselves to finite A and we define a function $H(\beta, \eta)$ by [see Eq. (3.1)]

$$H(\beta, \eta) = \sum_{n=1}^{A} \eta^n \sum_{\{1\},\, \Sigma m^l m = n} S\{1\} . \tag{7.1}$$

We have the relation $I^R(\beta) = 1 + H(\beta, 1)$, and $H(\beta, \eta)$ can be constructed from $I^R(\beta) - 1$ by the substitution

$$x_{i_1 \cdots i_n}(\beta) \to \eta^n x_{i_1 \cdots i_n}(\beta). \tag{7.2}$$

For our purpose we can replace $x_{i_1 \cdots i_n}(\beta)$ by $x_{i_1 \cdots i_n} + \beta x'_{i_1 \cdots i_n}$, and as a consequence $H(\beta, \eta)$ is a polynomial of finite order in β and η. Since $H(0, 0)$ vanishes there exists a neighborhood in $\mathbf{C} \times \mathbf{C}$ of $(\beta, \eta) = (0, 0)$ in which $\ln[1 + H(\beta, \eta)]$ can be expanded into an absolutely convergent series $\sum_{K=1}^{\infty} \eta^K a_K(\beta)$. It is easily proved that this series represents the L.C.E. in which the substitution (7.2) has been carried through. By differentiation with respect to β we find that the radius of convergence of the L.C.E. of the expectation value in the complex η plane is determined by that solution of the equation

$$F(\eta) = H(0, \eta) + 1 = 0 \tag{7.3}$$

which is nearest to $\eta = 0$. The radius of convergence depends only on Ψ and not on the operator involved. This result, however, may be a special feature of the finite system.

Using the form $\Psi = \exp(S)\Phi$, we can introduce the parameter η into the wavefunction itself by the substitution $S^{(n)} \to \eta^{n/2} S^{(n)}$, $\eta \geq 0$ [see Appendix A2], and the L.C.E. for $\eta \geq 0$ becomes identical to the expansion of

$$\langle\Psi(\eta)|\Theta|\Psi(\eta)\rangle / [\langle\Phi|\Phi\rangle + \langle\Psi(\eta) - \Phi|\Psi(\eta) - \Phi\rangle] \tag{7.4}$$

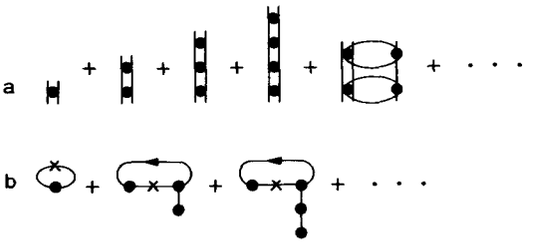


FIG. 6. a) First diagrams which in finite systems contribute to $\langle\Theta\rangle$ according to the renormalized expansion. b) First diagrams of the expansion of the renormalization factor. All but crossed lines are renormalized. Renormalized lines containing only one point are represented by the point itself. The directions of the lines are only given where they are important.

in powers of η. (This result shows that the methods used in Refs. 9, 10 give the same expansion as our method.) The coefficient of η^K in $F(\eta)$ is equal to $\langle \chi^k \mid \chi^k \rangle \geq 0$, where χ^k denotes the (unlinked) k-particle k-hole amplitude of Ψ. Thus $F(\eta)$ has the following properties:

(i) $F(0) = 1$;

(ii) $F(1) = \langle \Psi \mid \Psi \rangle$;

(iii) $| F(\eta) - 1 | \leq F(|\eta|) - 1$;

(iv) $(d/d\eta) F(\eta) \geq 0$ if $\eta \geq 0$.

Using these relations, we can establish the following lemma.

Lemma 7.1: The L.C.E. converges if $\langle \Psi \mid \Psi \rangle < 2$. The expansion diverges if $\langle \Psi \mid \Psi \rangle > 2^A$. Better estimates cannot be obtained without restricting Ψ. (Note that Ψ is normalized according to $\langle \Psi \mid \Phi \rangle = 1$).

In view of Eq. (7.3) the first part of that lemma is proved by noting that $\langle \Psi \mid \Psi \rangle < 2$ implies $| F(\eta) - 1 | \leq F(|\eta|) - 1 \leq F(1) - 1 < 1$, if $|\eta| \leq 1$. The second statement is obvious if we represent $F(\eta)$ by

$$F(\eta) = \prod_{m=1}^{M} \left(\frac{\eta}{|a_m|} - \frac{a_m}{|a_m|} \right) \left(\frac{\eta}{|a_m|} - \frac{a_m^*}{|a_m|} \right) \prod_{l=1}^{A-2M} \left(\frac{\eta}{b_l} + 1 \right). \tag{7.5}$$

Here the quantities a_m, a_m^*, $m = 1, \ldots, M$ denote the pairs of complex zeros of $F(\eta)$ and the $(-b_l)$, $l = 1, \ldots, A - 2M$ denote the zeros on the negative real axes. In deriving Eq. (7.5) we have used properties (i) and (iv). The estimates cannot be improved without restricting Ψ as is shown by the examples $F(\eta) = 1 + \eta^K$ and $F(\eta) = (\eta/(1 + \epsilon) + 1)^A, \epsilon > 0$, respectively. Note that the results of Ref. 21 are consistent with the estimate $\langle \Psi \mid \Psi \rangle < 2$.

Equation (7.5) shows that the L.C.E. will converge for $\langle \Psi \mid \Psi \rangle \sim 2^A$ only if the zeros of $F(\eta)$ are concentrated near the unit circle. We see no reason why in physical applications $F(\eta)$ should show that behavior, and we believe that the upper bound of the radius of convergence established in Lemma 7.1 is much too big. In view of the first part of Lemma 7.1 we feel rather pessimistic concerning the convergence of the unrenormalized expansion in physical problems. The convergence properties of the renormalized expansion may be much better. This hope is based on the observation that the renormalization partially resums the expansion of the normalization denominator in Eq. (7.4). Besides this the identity (6.6) also seems to favor the renormalized theory.

In the limit $A \to \infty$ a necessary condition for the convergence of the cluster expansion is given by the equation $\lim_{A\to\infty} p_{i_1 \ldots i_n} = \prod_{k=1}^{n} p_{i_k}$ which is found by a comparison of Eqs. (5.1) and (6.6). Equations (6.3) and (6.4) show that this condition is equivalent to the vanishing of $p^c_{i_1 \ldots i_n}$, which in the diagrammatic representation employing double beams contains only multiply connected diagrams. This condition seems to be directly related to the short range nature of the correlations which is assumed in the cluster expansion. The B.C.S. wavefunction, for instance, violates it because of the long-range correlations between certain pairs of particles.

8. COMPARISON WITH OTHER APPROACHES

We have shown that the I.Y. expansion has a very simple structure. This fact, however, to some extent is compensated by the relatively complicated structure of the cluster integrals. The n-point cluster integral contains parts which describe a product of several clusters of less than n particles which are bound together by an exchange of single-particle functions. We can isolate these "reducible" parts by defining nondiagonal irreducible clusters[13,19] (see Appendix A1), and the terms of the L.C.E. developed here can be absorbed as diagonal elements into an L.C.E. in terms of products of these irreducible clusters. This idea is systematically carried through in Ref. 13, and it is implicitly used in Ref. 14. Compared to the I.Y. expansion, it has the advantage that the cluster integrals are simplified without complicating too much the L.C.E. in terms of these quantities. (However, there are indications that the renormalization is more complicated.)

On the other hand this treatment mixes the effect of the *normalization* of the wavefunction with the effect of the *antisymmetrization*. Indeed, the reducible parts of $x_{i_1 \ldots i_n}(\beta)$ are present already in the unlinked expansion and vanish if we omit the antisymmetrization operator in Eq. (2.1). The diagrams containing several cluster integrals, which are mixed with these reducible parts, arise by transforming the unlinked into a linked expansion. From the discussion of Sec. 7 it is clear that we equivalently can interpret these diagrams to be created by dividing out the normalization denominator $\langle \Psi \mid \Psi \rangle^{-1}$, and thus these terms keep track of the normalization. They occur independently of the symmetry of the wavefunction. Thus we believe that there is no reason to treat these two types of terms on the same footing.

Especially we feel that it is a good procedure to renormalize the expansion in the way given here as we thus resum diagrams which are created by the same physical effect. In connection with variational methods the renormalization most probably will resolve the "Emery difficulty,"[22] which is caused by the fact that in the lowest order of the I.Y. expansion there is no mechanism which prevents the trial wavefunction from building up an arbitrarily strong correlation within the interaction region of each pair of particles, and thus gaining an indefinite amount of binding energy. Clearly the renormalization would supply us with such a mechanism as it takes into account the norm of Ψ, which also increases if we build up such a correlation. In a further publication we will come back to these problems, which are related[4,15] to the self-consistence of single-particle energies used in Brueckner theory.

ACKNOWLEDGMENT

The author is indebted to Professor H. A. Weidenmüller for many discussions and for a careful reading of the manuscript.

APPENDIX: DIAGRAMMATICAL REPRESENTATION OF THE CLUSTER INTEGRALS

1. Product form of Ψ

Irrespective of the nature of the symmetric operator Θ the reduced subnormalization integral [Eq. (2.4)] symbolically is written in the form

$$I^R_{K_1 \ldots K_m} = \sum_{P\{K_1, \ldots, K_m\}} \operatorname{sgn} P \int \prod_{i}^{<K_1, \ldots, K_m>} \times [di\, \varphi_{\bar{i}}^*(i)\, e^{\beta\Theta^{(1)}(i)} \varphi_i(i) J_i^{-1}(\beta)] \times \prod_{n=1}^{m} \prod_{i_1 < \cdots < i_n}^{<K_1, \ldots, K_m>} f^*_{\bar{i}_1 \ldots \bar{i}_n}(i_1, \ldots, i_n)$$

$$\times e^{\beta\Theta^{(n)}(i_1,\ldots,i_n)[1-\delta_{1n}]} f_{i_1\ldots i_n}(i_1,\ldots,i_n). \quad \text{(A1)}$$

Here $\Theta^{(m)}$ is the m-body part of Θ and the symbol $\bar{i}$ denotes the image $P(i)$ of the index i under the permutation P of the set $\{K_1,\ldots,K_m\}$. The correlation functions by definition fulfil the symmetry relations

$$f_{i_{\bar{1}}\ldots i_{\bar{n}}}(i_{\bar{1}},\ldots,i_{\bar{n}}) = f_{i_1\ldots i_n}(i_1,\ldots,i_n) \text{ for all } P\{1,\ldots n\}.$$

In the diagrammatic representation a simple (double) k-body link [Fig. 7a (b)] connecting the points i_1 to i_K represents a factor $f^*_{\bar{1}\ldots\bar{i}_K}(i_1,\ldots,i_K) f_{i_1\ldots i_K}(i_1,\ldots,i_K) - 1$

$$[f^*_{\bar{1}\ldots\bar{i}_K}(i_1,\ldots,i_K) \times \Theta^{(K)}(i_1,\ldots,i_K) \times f_{i_1\ldots i_K}(i_1,\ldots,i_K)].$$

A [crossed] directed line connecting the point i to the point $\bar{i}$ represents the integration over the coordinates of the ith particle with the weight factor $\varphi^*_{\bar{i}}(i)\varphi_i(i)$ $[\varphi^*_{\bar{i}}(i)\Theta^{(1)}(i)\varphi_i(i)]$. The lines thus fix the permutation considered. The cluster integral $x_{i_1\ldots i_n}$ is given by the sum of the contributions of all different *connected* diagrams which are constructed according to the following rule.

Rule A1: (i) Draw n points labelled from left to right by i_1 to i_n. (ii) Draw a structure of directed lines such that in each point there starts and ends exactly one line, respectively. (iii) Draw any number of simple r-body links ($1 \le r \le n$). An r-body link connecting a definite set of points may occur only once. (iv) Each point must be connected with at least one link.

The rules for evaluating a diagram are obvious. The sign is given by $(-1)^{n-l}$, where l is the number of closed loops of integration lines. In Fig. 7c–e we present some examples which give the following contributions:

(c) $\int d_1(|f_i(1)|^2 - 1)|\varphi_i(1)|^2,$

(d) $-\int d_1 d_2 [f^*_{i_1}(2) f_{i_2}(2) - 1][f^*_{i_2 i_1}(1,2) f_{i_1 i_2}(1,2) - 1] \times \varphi^*_{i_2}(1)\varphi_{i_1}(1)\varphi^*_{i_1}(2)\varphi_{i_2}(2),$

(e) $-\int d_1 d_2 [f^*_{i_1 i_3}(1,2) f_{i_1 i_2}(1,2) - 1] \times |\varphi_{i_1}(1)|^2 \varphi^*_{i_3}(2)\varphi_{i_2}(2) \int d_3 d_4 [f^*_{i_2 i_4}(3,4) \times f_{i_3 i_4}(3,4) - 1]|\varphi_{i_4}(4)|^2 \varphi^*_{i_2}(3)\varphi_{i_3}(3).$

Evaluating the expectation value of $\Theta^{(m)}$, $m \ge 2$, we determine $x'_{i_1\ldots i_n}$, $n \ge m$, by the rules given above except that we add the statement: (iiia) Each diagram contains exactly one double m-point link. If we are concerned with the expectation value of a one-body operator, then $x'_{i_1\ldots i_n}$ is given by $x'_{i_1\ldots i_n} = X_{i_1\ldots i_n} - x_{i_1\ldots i_n} \sum_{K=1}^{n} \langle i_K | \Theta^{(1)} | i_K \rangle$, where $X_{i_1\ldots i_n}$ is constructed according to Rule A1 altered in the following respects. (iia) Exactly one line is crossed. (iva) If $n \ge 2$, the starting point of the crossed line may be excepted from statement (iv). Depending on the exact choice of Ψ these rules can be considerably simplified.

We now use that representation to exhibit the asymptotic behavior of $x_{i_1\ldots i_n}$, a proof of which to our knowledge has not been given in the literature. In the limit $A \to \infty$ the functions φ_i are taken to be plane waves and by virtue of momentum conservation all $f_i(i)$ are identical to one. All one-body links are zero and the directed line (i)

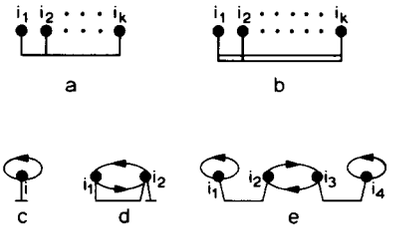


FIG. 7. a, b) Basic ingredients of the cluster integrals for the product form of Ψ. See the text for a definition. c) to e) Examples of diagrams contributing to $x_i, x_{i_1 i_2}, x_{i_1 i_2 i_3 i_4}$, respectively.

carries definite momentum $\mathbf{k}_i$. We use the term "irreducible" cluster to denote any part of a diagram which is completely connected by links and which is not a part of a bigger structure with that property. Each diagram contributing to $x_{i_1\ldots i_n}$ contains m, $1 \le m \le n/2$ irreducible clusters. A typical diagram containing two irreducible clusters is given in Fig. 7e. By virtue of the short range character of the correlation functions and of the normalization $\langle \varphi_i | \varphi_i \rangle = 1$, the integration over the (r) coordinates of an irreducible cluster yields an expression of the asymptotic order $\Omega^{1-r} = \rho^{r-1} A^{1-r}$ where Ω and ρ denote volume and density of the system, respectively. Furthermore, conservation of momentum yields a δ function $\delta[\sum \mathbf{k}^{\text{in}} - \sum \mathbf{k}^{\text{out}}]$ where $\mathbf{k}^{\text{in}}$ ($\mathbf{k}^{\text{out}}$) denotes the momentum of an ingoing (outgoing) line. It is easily proved that exactly $(m-1)$ of these δ functions are independent. As a result any term which contributes to $x_{i_1\ldots i_n}$ and which consists of m irreducible clusters asymptotically behaves like $\rho^{n-m} A^{m-n}$ and in addition contains $(m-1)$ independent δ functions connecting the indices i_1 to i_n. Similar results can be deduced for the differentiated cluster integrals. The asymptotic behavior of $x_{i_1\ldots i_n}$ as derived here, underlies Lemma 5.2.

The irreducible clusters constitute the basic quantities of the L.C.E. constructed in Ref. 13.

2. Particle-hole form of Ψ

This form of Ψ is defined[13,15] by $\Psi = \exp(S)\Phi$ where, in obvious notation, $\Phi = a^+_A \cdots a^+_1 |0\rangle$ is the unperturbed ground state and S is given by $S = \sum_{n=1}^{A} S^{(n)}$,

$$S^{(n)} = \frac{1}{(n!)^2} \sum_{\substack{i_1\ldots i_n \le A \\ b_1\ldots b_n > A}} S^{b_1\ldots b_n}_{i_1\ldots i_n} a^+_{b_n} \cdots a^+_{b_1} a_{i_1} \cdots a_{i_n}, \quad \text{(A2)}$$

$$S^{b_{\bar{1}}\ldots b_{\bar{n}}}_{i_{\bar{1}}\ldots i_{\bar{n}}} = \text{sgn}P\, S^{b_{\bar{1}}\ldots b_{\bar{n}}}_{i_1\ldots i_n} = \text{sgn}P\, S^{b_1\ldots b_n}_{i_{\bar{1}}\ldots i_{\bar{n}}} = S^{b_1\ldots b_n}_{i_1\ldots i_n} \quad \text{(A3)}$$

for all permutations $P\{1,\ldots,n\}$.

The wavefunctions $\Psi_{K_1\ldots K_m}$ are defined in obvious fashion. We introduce pairs of points connected by a (not directed) vertical line to represent the states i_K. The diagrams contributing to $x_{i_1\ldots i_n}$ are constructed and evaluated according to the following rules (see also Ref. 13).

Rule A2: (i) Draw n pairs of points which from left to right are labelled by the indices i_1 to i_n. (ii) Each lower (upper) point is touched by exactly one hanging (standing) link. (iii) In each lower (upper) point there starts (ends) exactly one directed contraction line.

Rule A3: (i) Label the lower (upper) point of the pair i_K by a_K (b_K). (ii) A hanging (standing) link connecting the points i_{K_1} to i_{K_r} contributes a factor $S^{a_{K_1}\cdots a_{K_r}}_{i_{K_1}\cdots i_{K_r}}$ ($S^{b_{K_1}\cdots b_{K_r}}_{i_{K_1}\cdots i_{K_r}}$*). (iii) A directed contraction line connecting the points a_K and b_l contributes a factor $\delta_{a_K b_l}$. (iv) Sum all indices a_K, b_K independently over all states $\alpha > A$ and multiply by the factor $(-1)^{n-l}\sigma^{-1}$. Here l is

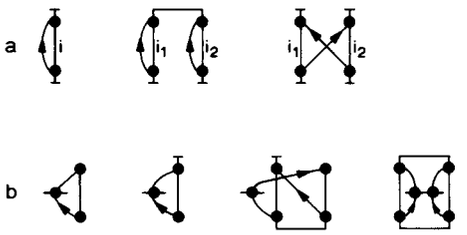


FIG. 8. a) Diagrams contributing to the (not differentiated) cluster integrals for the particle-hole form of Ψ. b) Diagrams contributing to the differentiated cluster integrals.

the number of closed loops consisting out of contraction lines and nondirected lines connecting a pair i_K, and σ is the symmetry number.

The equivalence of diagrams and the symmetry number are defined below. In Fig. 8a we give some examples which give the following contributions:

$$\sum_{a>A} |S_i^a|^2 ; \sum_{a_1 a_2 > A} S_{i_1}^{a_1} S_{i_2}^{a_2} S_{i_1 i_2}^{a_1 a_2 *} ; - \sum_{a_1 a_2 > A} S_{i_1}^{a_1} S_{i_1}^{a_2 *} S_{i_2}^{a_1 *} .$$

A diagram is called connected if we cannot separate it into two parts without cutting a line or a link (or a beam representing the operator). The cluster integral $x_{i_1 \cdots i_n}$ is given by the sum of all inequivalent connected diagrams.

In order to represent the expectation value of the (symmetric) m-body operator

$$\Theta = \left(\frac{1}{m!}\right)^2 \sum_{\alpha_1 \cdots \alpha_m, \beta_1 \cdots \beta_m} \Theta_{\alpha_1 \cdots \alpha_m, \beta_1 \cdots \beta_m} a^+_{\alpha_1} \cdots a^+_{\alpha_m} a_{\beta_m} \cdots a_{\beta_1}, \quad \text{(A4)}$$

we again introduce the quantity $X_{i_1 \cdots i_n} = x'_{i_1 \cdots i_n} + x_{i_1 \cdots i_n} \times \sum_{K=1}^n \Theta_{i_K, i_K}$, where the last term contributes only if $m = 1$. $X_{i_1 \cdots i_n}$ is given by the sum of all inequivalent connected diagrams constructed according to the following version of Rule A2. We add: (ia) Draw a horizontal beam with m points. In each point of that beam there ends one contraction line starting at a lower point i_K and there starts one contraction line ending at an upper point i_l. We alter (ii) into (ii'): Each lower (upper) point is touched by at most one hanging (standing) link. It must be touched by a link if it is not connected to the horizontal beam by one contraction line. Examples are given in Fig. 8b.

From X_{i_1} the trivial diagram (Fig. 8b, the first diagram) has to be excluded. The contribution of a diagram is calculated according to Rule A3, supplemented by the following statements. (ia) The line ending (starting) at the jth point of the horizontal beam is labelled by β_j (α_j). The beam represents $\Theta_{\alpha_1 \cdots \alpha_m, \beta_1 \cdots \beta_m}$. (iiia) Use the corresponding δ factors to eliminate the α_j and β_j. If a point α_j (β_j) is contracted with a point i_K not attached to a link, the δ factor is $\delta_{i_K \alpha_j}$ ($\delta_{i_K \beta_j}$); otherwise it is $\delta_{b_K \alpha_j}$ ($\delta_{a_K \beta_j}$). According to these rules the contributions of the last three diagrams of Fig. 8b are given by

$$\sum_{b>A} S_i^{b *} \Theta_{b,i}, \quad - \sum_{a_1 a_2 > A} S_{i_1}^{a_2 *} S_{i_1 i_2}^{a_1 a_2} \Theta_{i_2, a_1},$$

$$\tfrac{1}{4} \sum_{a_1 a_2 b_1 b_2 > A} S_{i_1 i_2}^{b_1 b_2 *} S_{i_1 i_2}^{a_1 a_2} \Theta_{b_1 b_2, a_1 a_2}.$$

By virtue of the symmetries of Θ two diagrams which differ only in the ordering in which the contraction lines are fixed to the m-point beam are taken to be identical. Taking into account the symmetries of $S^{(n)}$, we can characterize the symmetry operations P, which leave the contribution of a diagram unchanged in the following way: (i) Permute the labels a_i, b_i among themselves in such a way that we interchange only labels which are attached to the same link. (ii) Shift the points along the links with the *contraction lines* fixed to them in such a way that the original labelling is restored. Two diagrams X and X' are equivalent if there exists a P with $PX = X'$. (Remember that the ordering in which the lines are attached to the beam is irrelevant.) The symmetry number $\sigma(X)$ is equal to the number of operations P with $PX = X$.

According to these rules it is easy to prove that the cluster integrals $x'_{i_1 \cdots i_n}$ corresponding to the operators $P_{i_1 \cdots i_m} = a^+_{i_1} \cdots a^+_{i_m} a_{i_m} \cdots a_{i_1}$ are given by Eqs. (6.1) and (6.2).

[1] F. Iwamoto and M. Yamada, Prog. Theor. Phys. **17**, 543 (1957).
[2] J. W. Clark and P. Westhaus, J. Math. Phys. **9**, 131 (1968); J. Math. Phys. **9**, 149 (1968).
[3] K. Schäfer and G. Schütte, Nucl. Phys. A **183**, 1 (1972).
[4] L. Schäfer, Nucl. Phys. A **194**, 497 (1972).
[5] L. Schäfer and H. A. Weidenmüller, preprint.
[6] C. W. Wong, Phys. Rev. C **3**, 1058 (1971).
[7] J. W. Clark and M. L. Ristig, Phys. Rev. C **5**, 1233 (1972).
[8] J. W. Clark and M. L. Ristig, preprint.
[9] J. W. Clark, D. Chakkalakal, and P. Westhaus, Prog. Theor. Phys. **34**, 726 (1965).
[10] S. Fantoni and S. Rosati, Nuovo Cimento A **10**, 145 (1972).
[11] F. Y. Wu and E. Feenberg, Phys. Rev. **128**, 943 (1962).
[12] F. Y. Wu, J. Math. Phys. **4**, 1438 (1963).
[13] J. da Providencia, Nucl. Phys. **44**, 572 (1963); Nucl. Phys. **46**, 401 (1963).
[14] M. Gaudin, J. Gillespie, and G. Ripka, Nucl. Phys. A **176**, 237 (1971).
[15] J. da Providencia and C. M. Shakin, Phys. Rev. C **4**, 1560 (1971); Phys. Rev. C **5**, 53 (1972).
[16] L. Schäfer, Phys. Lett. B **41**, 419 (1972).
[17] R. Jastrow, Phys. Rev. **98**, 1479 (1955).
[18] N. H. March, W. A. Young, and S. Sampanthar, *The Many-Body Problem in Quantum Mechanics* (Cambridge U.P., Cambridge, 1967), Appendix 4A1.
[19] M. L. Ristig and J. W. Clark, Nucl. Phys. A **199**, 351 (1973).
[20] E. Krotschek, J. Nitsch, M. L. Ristig, and J. W. Clark, Nuovo Cimento Lett. **6**, 143 (1973).
[21] H. R. Petry and D. Schütte, Nucl. Phys. A **195**, 623 (1972).
[22] V. J. Emery, Nucl. Phys. **6**, 585 (1958).

Schrödinger equation with a random potential: A functional approach

Thomas P. Eggarter

Laboratoire de Physique des Solides, Université Paris-Sud, Centre d'Orsay, 91405 Orsay, France*
(Received 9 April 1973)

We present a new functional integral approach to the average density of states of a quantum particle in a random potential. As an illustration of the method, the formulas are explicitly evaluated for some one-dimensional problems; results previously obtained by other methods are recaptured. The formalism for the study of wavefunction localization is also derived.

I. INTRODUCTION

In recent years, and probably because of the technological importance of amorphous solids, a great deal of work has been done on the electronic structure of disordered materials. The natural starting point for an understanding of their properties is a knowledge of the one-electron spectrum, and the nature of the associated wave functions. In crystalline materials, the Bloch-Floquet theorem immediately reduces the study of the whole crystal to the study of one elementary cell, thus reducing the difficulty enormously. Unfortunately, no analogous theorem exists for the disordered case, and one is confronted in principle with the formidable task of diagonalizing a random matrix of order 10^{23}. Several methods have been developed to deal with this problem, but the mathematical difficulties involved are of such magnitude that our understanding of the problem is still far from being satisfactory.

A few idealized problems in disordered materials are at present known in which the energy spectrum (but generally no transport properties) can be computed exactly. Dyson[1] found the frequency spectrum of a linear chain of oscillators with random masses and spring constants (equivalent to the density of states of a one-dimensional tight-binding alloy with nearest neighbor hopping, and with random on-site energies and hopping elements), Frisch and Lloyd[2] derived the electron spectrum in a random set of δ-function potentials on a line, and Halperin[3] solved the same problem for a random one-dimensional potential which may have an arbitrary probability distribution at any given point, but must have the property that any two points are statistically independent. In the same paper, Halperin also sets up the formalism for transport properties, but does not do any actual calculations. The formulas are hard to evaluate, even with modern electronic computers.

In three dimensions, the only exact solutions are the density of states found by Lloyd[4] for the tight-binding alloy with a Lorentzian distribution for the on-site energy, and a generalization thereof studied by Eggarter, Cohen, and Economou,[5] in which they have incorporated short-range order in the Lloyd model. In both cases the possibility of solving exactly is due to a particular property of the Lorentzian distribution (only one pole on each side of the real axis), and the elegant result obtained (average density of states = pure crystal density of states convoluted with probability density for on-site energy) is only valid in this special case and provides no help for understanding other three-dimensional situations.

Among the approximate methods, the most widely used is certainly the "coherent potential approximation" (CPA),[6,7] which in several cases seems to describe the overall band structure reasonably well. But it fails near the band edges (the most interesting region in semiconductors); it has also been shown recently[8] that in the strong scattering limit the density of states of a binary alloy has considerable structure in the center of each subband, a fact which is not even qualitatively described by CPA.

The determination of transport properties involves even greater difficulties than the energy spectrum, since in addition information on the wave functions is required. Mott, and Cohen, Fritzsche, and Ovshinsky[9] have conjectured that a disordered material exhibits energy bands inside which there are "mobility edges" separating extended from localized states. Considerable effort has been devoted to prove this conjecture, but no clean proof has so far been given. Anderson's original work,[10] and several extensions thereoff,[11,12] are based on the idea that the convergence of a certain renormalized perturbation series is equivalent to localization of the eigenstates of the Hamiltonian under consideration. The energy enters as a parameter in the series, so that the determination of mobility edges reduces to finding the energies at which the series changes from convergent to divergent. This idea can not be rigorously true; in the case of a one-dimensional tight-binding nearest-neighbor-hopping perfect crystal the eigenstates are extended (Bloch states), while the renormalized perturbation series reduces to a finite sum (just two terms) for this problem. It has also been attempted to decide the question of localization by studying some appropriate series or other expression for the ensemble averaged Green's function $\langle G(E)\rangle$. These efforts cannot be fruitful; it was shown recently[5] that the process of averaging over an ensemble of potential configurations destroys the information on the localized or delocalized character of the wave functions contained in $G(E)$. Percolation theory has also been used to study localization. For a classical particle in a random potential the Mott–CFO conjecture is easy to understand;[13] the mobility edge coincides with the percolation threshold. But the percolation treatment neglects interference phenomena, which may be important. Consider as an example the one-dimensional problem with a potential $V(x)$ bounded from above. Let $V_0 = \sup\{V(x); -\infty < x < \infty\}$. Borland[14] has shown that all states are localized in one dimension, so that no mobility edges exist. But the percolation threshold exists and is at V_0. The localization here is due to the wave nature of the particle, which percolation theory does not describe. The same considerations apply to a three-dimensional separable potential $V(\mathbf{r}) = V_1(x) + V_2(y) + V_3(z)$.

It is the authors belief that new mathematical methods will eventually have to evolve in order to provide a better understanding of the above questions. If this belief is correct, the search for new approaches to the problem constitutes a worthwhile effort.

In this paper, we present one possible alternative approach, based on a conceptually very simple idea. For an arbitrary wave function $\psi(\mathbf{r})$, the potential $V(\mathbf{r})$ which has ψ as an eigenstate of energy E can be written down

immediately by solving the Schrödinger equation for $V(\mathbf{r})$. If, moreover, a probability density is given in the space of all possible potentials, this can be converted to a probability density in ψ space. Integration over all of ψ space with fixed E then gives the density of states, while integration over some selected subset in ψ space (integration over all "localized" states, for example) could given information on the nature of the wave functions. The mathematical difficulties in carrying out this program will of course be considerable, and have not yet been fully explored.

In Sec. II, we set up our formalism: an equation for the average density of states as an integral over ψ space is derived. The way in which the functional integrals can be evaluated as the limit $N \to \infty$ of N-dimensional integrals is discussed. In Sec. III, we consider as specific examples two one-dimensional problems, and show that our formalism permits to recapture the exact solution of the Dyson problem as given by Economou and Papatriantafillhou,[15] and the Frisch and Lloyd solution[2] of the δ-function problem. In Sec. IV, we discuss the question of localization in the present context without being able, unfortunately, to work out any specific examples.

II. FORMALISM

We consider the Schrödinger equation

$$-\nabla\psi + V(\mathbf{r})\psi = E\psi \tag{1}$$

and suppose for simplicity that (1) has to be solved in a finite volume Ω with some appropriate boundary condition on ψ at $\partial\Omega$. For a given $V(\mathbf{r})$, the density of states is usually expressed in terms of the Green's function

$$G(z, V(\mathbf{r})) \equiv (z - H)^{-1} \tag{2}$$

by the well-known equation

$$\rho(E, V(\mathbf{r}) = -(1/\pi)\ \mathrm{Im}\ \mathrm{Tr}G(E^+, V(\mathbf{r})). \tag{3}$$

We will be interested in the case that $V(\mathbf{r})$ is a random function, and our aim will be, among other things, to compute the average density of states

$$\rho(E) = \langle \rho(E, V(\mathbf{r}))\rangle. \tag{4}$$

The fact that $V(\mathbf{r})$ is a random function can be stated mathematically by saying that a probability measure $d\mu$ has been given on the function space X of all possible potentials $V(\mathbf{r})$, so that for any subset $U \subset X$

$$\mathrm{Prob}[V(\mathbf{r}) \in U] = \int_U d\mu. \tag{5}$$

In order to make our formulas more transparent, we will also use the notation[16]

$$d\mu = f[V(\mathbf{r})]\mathfrak{D}[V(\mathbf{r})], \tag{6}$$

suggesting that a probability density $f[V(\mathbf{r})]$ is associated with each point of X. In this notation $\mathfrak{D}[V(\mathbf{r})]$ stands for a "volume element" in the function space X.

The average density of states can then be written as

$$\rho(E) = -\frac{1}{\pi}\int \mathrm{Im}\ \mathrm{Tr}G(E^+, V(\mathbf{r}))f[V(\mathbf{r})]\mathfrak{D}[V(\mathbf{r})]. \tag{7}$$

One difficulty in the explicit evaluation of (7) is that the functional $G(z, V(\mathbf{r}))$ cannot be written down explicitly. Perturbation expansions for G, moreover, are hard to handle because G has closely spaced poles on the energy region of interest.

The alternative approach we want to propose is based on the observation that, although the computation of the eigenvalues and wavefunctions for a given $V(\mathbf{r})$ is difficult, the inverse problem is trivial. If a wavefunction ψ and the corresponding energy E are given, the potential which gives rise to ψ as an eigenfunction of energy E is, according to (1),

$$V(\mathbf{r}) = \Delta\psi/\psi + E. \tag{8}$$

Using this fact we can derive an equation for $\rho(E)$ as an integral over the space Y of all possible wave functions $\psi(\mathbf{r})$.[17] Equation (8) defines a map $M: Y \times R \to X$; it assigns a potential $V(\mathbf{r})$ to each pair (ψ, E). This map will not be one-to-one since for a given $V(\mathbf{r})$ there exist infinitely many eigenfunctions. However, for a sufficiently small neighborhood A of any point $p = (\psi, E) \in Y \times R$ the correspondence between A and $M(A)$ given by (8) is one-to-one.[18] We define a measure $d\tilde{\mu}$ on $Y \times R$ by the condition that for any such sufficiently small set A

$$\int_A d\tilde{\mu} = \int_{M(A)} d\mu. \tag{9}$$

This measure can then be extended to all of $Y \times R$ in the usual way. In our more explicit notation we write

$$d\tilde{\mu} = f[V\{\psi(\mathbf{r}), E\}]|J|\mathfrak{D}[\psi]dE, \tag{10}$$

where $J = J(\psi, E)$ is the appropriate Jacobian of the transformation (8).

We will now prove that

$$\rho(E) = \frac{1}{2}\int_Y f[V\{\psi, E\}]|J|\mathfrak{D}[\psi]. \tag{11}$$

For this purpose, let us define as $V_k(E_0) \subset X$ the set of all those $V(\mathbf{r})$ which have exactly k eigenvalues below E_0. Obviously

$$V_k(E_0) \cap V_{k'}(E_0) = \phi(= \text{empty set}) \quad \text{if } k \neq k', \tag{12}$$

$$\bigcup_{k=0}^{\infty} V_k(E_0) = X. \tag{13}$$

We define the integrated density of states $\mathfrak{N}(E_0, V(\mathbf{r}))$ by

$$\mathfrak{N}(E_0, V(\mathbf{r})) = k \quad \text{iff} V(\mathbf{r}) \in V_k(E_0); \tag{14}$$

this quantity is related to the average density of states in the usual way

$$\langle \mathfrak{N}(E_0, V(\mathbf{r}))\rangle = \int_{-\infty}^{E_0} \rho(E)dE. \tag{15}$$

Consider next the integral

$$F(E_0) = \frac{1}{2}\int_{Y\times(-\infty,E_0)} d\tilde{\mu} = \frac{1}{2}\int_{-\infty}^{E_0} dE \int_Y f[V(\psi, E)]|J|\mathfrak{D}[\psi]. \tag{16}$$

As the point (ψ, E) sweeps out the region $Y \times (-\infty, E_0)$, each $V(\mathbf{r}) \in V_k(E_0)$ will appear exactly $2k$ times as an argument in f, since it appears each time p coincides with one of the eigenstates (ψ_i, E_i) of $V(\mathbf{r})$, and there is a trivial denencracy between $+\psi_i$ and $-\psi_i$ which accounts for the factor $\frac{1}{2}$. Therefore, using Eq. (9), we can write

$$F(E_0) = \frac{1}{2}\sum_{k=0}^{\infty} 2k \int_{V_k(E_0)} d\mu = \sum_{k=0}^{\infty} \int_{V_k(E_0)}$$

$$\mathfrak{N}(E_0, V(\mathbf{r}))d\mu = \int_X \mathfrak{N}(E_0, V(\mathbf{r}))d\mu = \langle \mathfrak{N}(E_0, V(\mathbf{r}))\rangle. \tag{17}$$

Combining Eqs. (15), (16), and (17) we have established the validity of (11).

Equation (11) has a clear geometrical meaning: the integrand $f[V(\psi,E)]|J|$ represents the probability density in ψ space that a given wave function will occur as an eigenfunction of energy E. The average density of states involves an integration over $\mathfrak{D}[\psi]$ because we are only interested in having an eigenvalue E, regardless of what the eigenfunction is. In this sense Eq. (11) is much more transparent than the usual expression (7) for $\rho(E)$. A further advantage is that no divergences, or complex quantities, which are present in (7), occur in (11). The main problem with (11) is the evaluation of the Jacobian. We will later present explicit calculations of this quantity for one-dimensional problems. As usual, we can express our functional integrals as the limit $N \to \infty$ of N-dimensional integrals. For this, we divide space into an infinitesimal cubic lattice (spacing $= \eta$), we label our lattice points by an index $i = 1, 2, \ldots, N$, and put

$$\psi(\mathbf{r}_i) \equiv a_i.$$

The finite difference expression for the Laplace operator is then

$$\Delta\psi(\mathbf{r}_i) \simeq \eta^{-1}\left(\sum_j' a_j - Z a_i\right), \tag{18}$$

where Z is the coordination number (dependent on the number of dimensions), and $\sum_j'$ indicates a sum over all those j's which are neighbors of i. The Schrödinger equation (1) now becomes

$$-W\sum_j' a_j + \epsilon_i a_i = E a_i. \tag{19}$$

Here we have made the identifications

$$W = \eta^{-2} \tag{20}$$

$$\epsilon_i = V(\mathbf{r}_i) + Z\eta^{-2}. \tag{21}$$

Equation (19) is precisely the eigenvalue equation for a tight-binding, nearest-neighbor-hopping Hamiltonian, widely used in the alloy problem. Once we know how to handle (19), the continuum case will be recaptured by letting $\eta \to 0$ in the final results. Equation (8) which gives the potential in terms of the amplitudes $\{a_i\}$ becomes now

$$\epsilon_i = E + W\sum_j' a_j/a_i \;\; = \epsilon_i\{a_j\}. \tag{22}$$

Our integration in ψ space involves only wavefunctions satisfying some normalization condition $\sum_1^N a_1^2 = u_0^2$, it is therefore natural to introduce polar coordinates

$$\begin{aligned} a_1 &= u\cos\phi_1, \\ a_2 &= u\sin\phi_1\cos\phi_2, \\ &\vdots \\ a_{N-1} &= u\sin\phi_1\sin\phi_2\ldots\cos\phi_{N-1}, \\ a_N &= u\sin\phi_1\sin\phi_2\ldots\sin\phi_{N-1}, \end{aligned} \tag{23}$$

and to take $\{\phi_1, \phi_2, \ldots, \phi_{N-1}, E\}$ as independent variables. The normalization constant u_0 disappears from the problem, as can be seen from (22). If $f(\{\epsilon_i\})$ is the probability density in ϵ space,[19] our Eq. (11) for the density of states becomes

$$\rho(E) = \frac{1}{2}\int d\phi_1\ldots d\phi_{N-1} f(\{\epsilon_i\})\left|\frac{\partial(\epsilon_1,\ldots,\epsilon_N)}{\partial(\phi_1,\ldots,\phi_{N-1},E)}\right|. \tag{24}$$

It is possible to transform this into an integral over all wavefunctions, normalized or not. For this we write

$$\begin{aligned} \rho(E) &= \frac{1}{2}\int d\phi_1\ldots d\phi_{N-1}\int_0^\infty du\,\delta(u-u_0)f(\{\epsilon_i\}) \\ &\quad\times\left|\frac{\partial(\epsilon_1,\ldots,\epsilon_N,u)}{\partial(\phi_1\ldots,\phi_{N-1},u,E)}\right| \\ &= \frac{1}{2}\int_{-\infty}^\infty\cdots\int_{-\infty}^\infty da_1\cdots da_N\delta(u-u_0)f(\{\epsilon_i\}) \\ &\quad\times\left|\frac{\partial(\epsilon_1,\ldots,\epsilon_N,u)}{\partial(\phi_1,\ldots,\phi_{N-1},u,E)}\right|\left|\frac{\partial(\phi_1,\ldots\phi_{N-1},u,E)}{\partial(a_1,\ldots,a_N,E)}\right| \\ &= \frac{1}{2}\int_{-\infty}^\infty da_1\cdots da_N\delta(u-u_0)f(\{\epsilon_i\})\left|\frac{\partial(\epsilon_1,\ldots,\epsilon_N,u)}{\partial(a_1,\ldots,a_N,E)}\right|. \end{aligned} \tag{25}$$

Next we notice that $\rho(E)$ is actually independent of u_0. We therefore pick an arbitrary function $\chi(u_0)$ with the only condition that $\int_0^\infty \chi(u)du = 1$; we multiply both sides of (25) by $\chi(u_0)du_0$ and integrate to get

$$\rho(E) = \frac{1}{2}\int_{-\infty}^\infty\cdots\int_{-\infty}^\infty da_1\cdots da_N\chi\left[\left(\sum a_i^2\right)^{1/2}\right]f(\{\epsilon_i\})\left|\frac{\partial(\epsilon_1,\epsilon_2,\ldots,\epsilon_N,u)}{\partial(a_1,\ldots,a_N,E)}\right|. \tag{26}$$

The function χ may eventually be so chosen as to facilitate the evaluation of the integral (26).

All this applies equally well to any number of dimensions. We will next consider some specific one-dimensional problems and evaluate the above formulas.

III. THE ONE-DIMENSIONAL CASE

We begin by studying the discrete eigenvalue problem defined by Eq. (19), without making reference to any underlying continuum. This is the Dyson problem,[1] which appeared first in the study of the spectrum of eigenfrequencies of a linear chain of oscillators with random masses. We will interpret it here as the eigenvalue problem for an electron in a tight-binding band. We lable our points or sites from left to right by an index $i = -N, -N+1, \ldots, 0, 1, \ldots N$, and, since some boundary conditions have to be imposed, we assume amplitude zero to the left of $-N$ and to the right of N. [This amounts to saying that there are impenetrable walls at $-(N+1)$ and at $N+1$.] Equation (19) can then be written more explicitly as

$$\begin{aligned} \epsilon_{-N} &= E + W(a_{-N+1}/a_{-N}), \\ \epsilon_i &= E + W(a_{i-1}/a_i + a_{i+1}/a_i), \\ &\quad i = -N+1,\ldots,N-1, \\ \epsilon_N &= E + W(a_{N-1}/a_N). \end{aligned} \tag{27}$$

It is clear that the most natural choice of independent variables now is not the set of polar angles of Eq. (23), but instead the set $\{\xi_i, E\}$ where the ξ_i's are the ratios

$$\xi_i \equiv a_{i+1}/a_i, \quad i = -N, -N+1, \ldots, N-1 \tag{28}$$

in terms of which

$$\begin{aligned} \epsilon_{-N} &= E + W\xi_{-N}, \\ \epsilon_i &= E + W(\xi_i + 1/\xi_{i-1}), \quad i = -N+1,\ldots,N-1, \\ \epsilon_N &= E + W/\xi_{N-1}. \end{aligned} \tag{29}$$

The Jacobian can now be written down

$$J = \left|\frac{\partial(\epsilon_{-N}, \epsilon_{-N+1}, \ldots \epsilon_N)}{\partial(\xi_{-N}, \ldots, \xi_{N-1}, E)}\right| = \begin{vmatrix} W & 0 & 0 & \cdots & \cdots & 0 & 1 \\ -\frac{W}{\xi_{-N}^2} & W & 0 & \cdots & \cdots & 0 & 1 \\ 0 & -\frac{W}{\xi_{-N+1}^2} & W & 0 & \cdots & 0 & 1 \\ & & & 0 & -\frac{W}{\xi_{N-2}^2} & W & 1 \\ & & & & 0 & -\frac{W}{\xi_{N-1}^2} & 1 \end{vmatrix} \equiv J(-N,N). \tag{30}$$

We have written the last term of (30) in such a way as to show explicitly that all sites from $-N$ to N are included. Expanding the determinant in the last row and using the same notation we obtain the recurrence relation

$$J(-N,N) = W^{N-(-N)} + (W/\xi_{N-1}^2)J(-N,N-1), \tag{31}$$

from which, by iteration,

$$J(-N,N) = W^{2N}\left(1 + \sum_{s=-N}^{N-1} (\xi_{N-1}\xi_{N-2}\cdots\xi_s)^{-2}\right). \tag{32}$$

We will write this as

$$J(-N,N) = W^{2N}\sum_{s=-N}^{N} (\xi_{N-1}\xi_{N-2}\cdots\xi_s)^{-2} \tag{33}$$

with the explicit convention that the term $s = N$ (undefined so far) is unity

Our Eq. (11) written in terms of the ξ_i's is

$$\rho(E) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} d\xi_{-N}\cdots d\xi_{N-1} f[\{\epsilon_i\}]W^{2N}\sum_{s=-N}^{N} (\xi_{N-1}\xi_{N-2}\cdots\xi_s)^{-2} \tag{34}$$

The factor $\frac{1}{2}$ in front of the integral is absent here because each set $\{\epsilon_i\}$ corresponds to a pair of wavefunctions $+\psi$ and $-\psi$. To further simplify matters, we will consider the case in which all ϵ_i's are statistically independent so that the probability density $f[\{\epsilon_i\}]$ factorizes in the form

$$f[\{\epsilon_i\}] = \prod_{j=-N}^{N} g(\epsilon_j) = g(E + W\xi_{-N})g(E + W\xi_{-N+1} + W/\xi_{-N}) \cdots g(E + W\xi_{N-1} + W/\xi_{N-2})g(E + W/\xi_{N-1}), \tag{35}$$

with $g(\epsilon)$ = probability density for a single site.

Inserting (35) into (34) and making the change of variable $u_i = W\xi_i$ (all i) in the integrals, we obtain

$$\rho(E) = \sum_{s=-N}^{N}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} du_{-N}\cdots du_{N-1} g(E + u_{-N}) \times g(E + u_{-N+1} + W^2/u_N)\cdots g(E + W^2/u_{N-1}) \times \frac{W^{2(N-s)}}{(u_{N-1}u_{N-2}\cdots u_s)^2}. \tag{36}$$

It is now convenient to define two integral operators K_1 and K_2 which act on an arbitrary function $h(x)$ in the following way;

$$K_1 h|_x = \int_{-\infty}^{\infty} h(y)g(E + x + W^2/y)dy, \tag{37}$$

$$K_2 h|_x = \int_{-\infty} \frac{W^2}{x^2} h(y)g(E + y + W^2/x)dy. \tag{38}$$

Each term in the summation Eq. (36) can be interpreted as follows: integration over $u_{-N}, u_{-N+1}, \ldots, u_{s-1}$ amounts to applying K_1 $N + s$ times in succession to the function $\vartheta_1(x) = g(E + x)$. Integration over $u_{N-1}, u_{N-2}, \ldots, u_{s+1}$ amounts to apply K_2 $N - s - 1$ times to the function $\vartheta_2(x) = g(E + W^2/x)$. Finally, the integration over u_s gives the density of states as

$$\rho(E) = \sum_{s=-N}^{N}\int_{-\infty}^{\infty} du_s (K_1^{N+s}\vartheta_1)|_{u_s}(K_2^{N-s-1}\vartheta_2)|_{u_s}. \tag{39}$$

It can be shown[20] that, provided g is reasonably well behaved, the limit

$$\lim_{A\to\infty} K_1^A \vartheta_1|_x = \phi(x) \tag{40}$$

exists and is given by the solution of the integral equation

$$\phi(x) = \int_{-\infty}^{\infty} \phi(y)g(E + x + W^2/y)dy, \tag{41}$$

supplemented by the normalization condition $\int_{-\infty}^{\infty}\phi(x)dx = \int_{-\infty}^{\infty}\vartheta_1(x)dx$ which in our case is simply

$$\int_{-\infty}^{\infty}\phi(x)dx = 1. \tag{42}$$

The convergence in (40), moreover, is uniform in x. Similarly, there exists the limit

$$\lim_{A\to\infty} K_2^A \vartheta_2|_x = f(x) \tag{43}$$

with f given by the solution of

$$f(x) = \frac{W^2}{x^2}\int_{-\infty}^{\infty} f(y)g(E + y + W^2/x)dy \tag{44}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{45}$$

It is also easy to verify by the change of variable $x \to W^2x^{-1}$ and $y \to W^2y^{-1}$ in (41) that

$$\phi(x) = (W^2/x^2)f(W^2/x). \tag{46}$$

Using these results, it is immediate that all except the end terms $s \simeq -N$ or $s \simeq N$ in (39) are approximately equal to $\int_{-\infty}^{\infty}\phi(u_s)f(u_s)du_s$. In the limit of an infinitely long chain $N \to \infty$, the density of states per site is therefore also given by this expression

$$\rho(E)_{\text{per site}} = \int_{-\infty}^{\infty} dx f(x)f(W^2/x)(W^2/x) = \int_{-\infty}^{\infty} dx f(W^2/x)f(x). \tag{47}$$

This is precisely the result obtained in Ref. 15 by the use of a Green's function approach.

As our second example, we will now work out the problem of a particle on a line with a potential

$$V(x) = c \sum_i \delta(x - x_i), \tag{48}$$

where the $\{x_i\}$ are points chosen independently and at random with an average density ρ. This is the Frisch and Lloyd problem.[2] The ϵ_i's of Eq. (21) have here a probability distribution

$$\epsilon_i = \begin{cases} \frac{c}{\eta} + \frac{2}{\eta^2} & \text{with probability } \eta\rho \\ \frac{2}{\eta} & \text{with probability } 1 - \eta\rho \end{cases} \tag{49}$$

while $W = \eta^{-2}$. This problem can be handled by combining our former example with a suitable limiting procedure to let $\eta \to 0$. We assume an infinite system, so that (41) applies, and we introduce the function

$$p(x) = \eta\rho\delta(x - c/\eta) + (1 - \rho\eta)\delta(x) \tag{50}$$

and write (41) as

$$\phi(x) = \int \phi(y) p\left(E + x + \frac{1}{\eta^4 y} - \frac{Z}{\eta^2}\right) dy. \tag{51}$$

It is convenient to define next

$$\chi_\eta(x) = (1/\eta)\phi(x/\eta + 1/\eta^2). \tag{52}$$

Expression (51) as an integral equation for χ_η and using (50), one obtains immediately

$$\chi_\eta(x) = (1 + \eta x_1)^2 \rho\eta \chi_\eta(x_1) + (1 - \eta\rho)(1 + \eta x_0)^2 \chi_\eta(x_0) \tag{53}$$

with

$$x_0 \equiv \frac{x + \eta E}{1 - \eta(x + \eta E)} \tag{54}$$

and

$$x_1 \equiv \frac{x - c + \eta E}{1 - \eta(x - c + \eta E)}. \tag{55}$$

In the limit $\eta \to 0$ Eq. 53 reduces to a trivial identity. However, the first derivative with respect to η followed by $\eta \to 0$ leads to an equation for $\chi(x) \equiv \lim_{\eta\to 0}\chi_\eta(x)$:

$$\frac{\partial}{\partial x}(E^2 + x^2)\chi(x) = \rho[\chi(x) - \chi(x - c)]. \tag{56}$$

This is the same equation derived Frisch and Lloyd in their approach based on stochastic processes.[20] Once (56) has been solved with the appropriate normalization condition

$$\int_{-\infty}^{\infty} \chi(x) dx = 1, \tag{57}$$

the density of states can be obtained by a similar limiting procedure in (47). We notice first that

$$\rho(E)_{\text{unit length}} = (1/\eta)\rho(E)_{\text{per site}}. \tag{58}$$

We also write (47) as

$$\rho(E)_{\text{per site}} = \int_{-\infty}^{\infty} \phi(x)\phi(W^2/x) dx, \tag{59}$$

which is correct because of (46). Next, using the definition of χ_η, Eq. (52), we have

$$\rho(E)_{\text{unit length}} = \frac{1}{\eta}\int_{-\infty}^{\infty} \phi(x)\phi\left(\frac{W^2}{x}\right) dx = \int_{-\infty}^{\infty} \chi_\eta(u)\frac{1}{\eta}\phi\left(\frac{1}{\eta^2(1 + \eta u)}\right) du$$

$$= \int_{-\infty}^{\infty} \chi_\eta(u)\chi_\eta\left[\frac{1}{\eta}\left(\frac{1}{1 + \eta u} - \frac{1}{\eta}\right)\right] du, \tag{60}$$

which in the limit $\eta \to 0$ goes over into

$$\rho(E)_{\text{unit length}} = \int_{-\infty}^{\infty} \chi(u)\chi(-u) du. \tag{61}$$

Frisch and Lloyd derived instead

$$N(E)_{\text{unit length}} = \int_{-\infty}^{\infty} \rho(E')_{\text{unit length}} dE' = \lim_{u\to\infty} \rho^{-1} u\chi(u), \tag{62}$$

but the two expressions are equivalent. The fact that $N(E)$ can be written as a functional linear in χ, Eq. (62), while $\rho(E)$ is given by a quadratic functional of the same χ, Eq. (61), is already discussed in Halperin's work.[3] A direct proof of the equivalence has also been given recently.[21]

IV. THE QUESTION OF WAVEFUNCTION LOCALIZATION

Suppose we have agreed on some definition of what is meant by a "localized" wavefunction, and let $L \subset Y$ be the set of all those ψ's which satisfy this definition. It is clear that if in Eq. (11) we restrict the domain of integration to L instead of all ψ space, we obtain a partial density of states $\rho_{\text{loc}}(E)$, associated with those wavefunctions that belong to L. Introducing the characteristic function of L

$$\chi_{\text{loc}}(\psi) = \begin{cases} 1 & \text{if } \psi \text{ is localized} \\ 0 & \text{otherwise} \end{cases}, \tag{63}$$

we can write

$$\rho_{\text{loc}}(E) = \int_Y \chi_{\text{loc}}(\psi) f[\{\epsilon_i\}] \mathfrak{D}[\Psi] |J| \tag{64}$$

which has the same structure as (11), the only difference being that the factor $\chi_{\text{loc}}(\psi)$ "erases" all extended s states. The Mott-CFO conjecture would be verified if we could prove that in the limit of an infinite system $\rho_{\text{loc}}(E)/\rho(E) \to 1$ in certain parts of the energy spectrum, and $\rho_{\text{loc}}(E)/\rho(E) \to 0$ in other parts. We see that the information enabling one to decide on the question of localization is, at least in principle, contained in our formalism.

To make the above rigorous, we have to define $\chi_{\text{loc}}(\psi)$ precisely. For an infinite system matters are very simple: we call a state "localized" if it can be normalized to unity, and extended or delocalized otherwise; this seems to be the only reasonable definition. But it is of no particular interest; infinite systems do not exist in nature. On the other hand, and from a physical point of view, it is clear that in a finite but macroscopic sample of material there are states we want to call localized (like, for example, donor or acceptor states in a slightly doped semiconductor) and states we want to call extended (for example, band states in a very pure crystalline sample). We will try to define χ_{loc} on the basis of these physical ideas. To avoid unnecessary complications, we will only talk about the tight-binding band problem; the case of the continuum can then be obtained by inserting the "mesh spacing" η of Sec. II in the appropriate places and taking the limit $\eta \to 0$.

Let $\alpha(N)$ be any function with the properties

$$\lim_{N\to\infty} \alpha(N) = 0 \tag{65}$$

and for every $\epsilon > 0$,

$$\lim_{N\to\infty} \alpha(N) N^\epsilon = \infty. \tag{66}$$

The typical example would be $\alpha(N) = (\log N)^{-1}$.

We propose the definition

$$\chi_{\text{loc}}(\psi) = \exp\left(-\alpha(N)\sum_i^N |a_i|\right), \tag{67}$$

where a_i is the amplitude of ψ (which we assume normalized to unity) at site i. It is clear that this definition has all the desired properties. For a macroscopic system $\chi_{\text{loc}}(\psi) \simeq 0$ if ψ is a Bloch state, a modulated Bloch state, or even a quasi-one-dimensional percolation channel, and $\chi_{\text{loc}}(\psi) \simeq 1$ if ψ is of the type usually considered as localized, with most of the amplitude concentrated in a small region of space. The above approximate equalities become increasingly accurate if the size of the sample is increased.

Strictly speaking, our $\chi_{\text{loc}}(\psi)$ is not the characteristic function of any decent subset of Y, but of what is called a "fuzzy set."[22] This is, however, irrelevant for our purposes; all that matters is that the right-hand side of (64) is a well-defined quantity for any N, and we simply take it as the definition of $\rho_{\text{loc}}(E)$.

An exact evaluation of (64) with the definition (67) does not seem an impossible task in one dimension, since the structures of (67) and (11) are very similar. But since no mobility edges are expected in one dimension anyway, we have made no serious effort to carry out this calculation.

We conclude this work with a few additional remarks. One may want to know what fraction of ψ space corresponds to localized states.

We can calculate this fraction x as

$$x = \frac{\int_Y \chi_{\text{loc}}(\psi)\mathfrak{D}[\psi]}{\int_Y \mathfrak{D}[\psi]} = \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} da_1\cdots da_N e^{-\alpha(N)\sum_1^N |a_i|}\,\delta\left[1-\sum_1^N a_i^2\right]}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} da_N\,\delta\left[1-\sum_1^N a_i^2\right]}, \tag{68}$$

which resembles an average of the quantity $\exp\left[-\alpha(N)\sum_1^N|a_i|\right]$ over a microcanonical ensemble (because of the δ function in the integral). It is known from statistical mechanics that for a fixed α (N-independent) $\lim_{N\to\infty}\left\langle\exp\left(-\alpha\sum_1^N|a_i|\right)\right\rangle$ can equally well be computed using the corresponding canonical ensemble with $\left\langle\sum_1^N a_i^2\right\rangle = 1$. For our case, and because of the very weak dependence of $\alpha(N)$ on N the same proof still holds, and we have

$$x = \left[\left(\frac{N}{2\pi}\right)^{1/2}\int_{-\infty}^{\infty} e^{-\alpha(N)|a|}e^{-Na^2/2}da\right]^N \simeq \exp(-(2/\pi)^{1/2}\alpha(N)N^{1/2}). \tag{69}$$

Thus, a negligibly small fraction of ψ space corresponds to localized states. But these states must receive an enormous weight in (11) because of the Jacobian. For a localized state, one can considerably change all ϵ_i's outside the region of localization without appreciable change in the wave function. If ψ has vanishing amplitude at some site, it is not affected by the ϵ at that site. From this, we conclude that $|J|^{-1} = |\partial(a_1,\ldots,a_N,E)/\partial(\epsilon_1,\epsilon_2,\ldots,\epsilon_N,u)|$ must be extremely small for a localized state. Thus, the Jacobian tends to favor localization, while the available volume in ψ space favors delocalization of the states as said before. It would be instructive to see which of these tendencies dominates; and we can do so very easily in one dimension.

Let us take any state defined by amplitudes $\{a_i\}$ and suppose these numbers rearranged into a decreasing sequence. If the state is localized, it should survive multiplication by $\exp\left[-\alpha(N)\sum_j|a_j|\right]$, and for this to happen the asymptotic behavior for large N must be

$$|a_j| < 1/j, \quad j \text{ large}. \tag{70}$$

Thus

$$|J|_{\text{loc}} \simeq \left|\frac{1}{a_1 a_2\ldots a_N}\right| > e^{N\log N}. \tag{71}$$

For a typical extended state, on the other hand,

$$|J|_{\text{ext}} \simeq e^{(N/2)\log N}; \tag{72}$$

therefore,

$$\frac{x|J|_{\text{loc}}}{1|J|_{\text{ext}}} \simeq \exp\left(\frac{N}{2}\log N - (2/\pi)^{1/2}\alpha(N)N^{1/2}\right) \underset{N\to\infty}{\to} \infty \tag{73}$$

indicating a dominance of the localizing tendency represented by J. Of course, a rigorous theory should consider $f[\{\epsilon_i\}]$ also, and refer to the three-dimensional case in order to be of interest. We have not succeeded so far in elaborating such a theory on the basis of the present formalism.

*Laboratoire associé au C.N.R.S.

[1]F. J. Dyson, Phys. Rev. **92**, 1331 (1953).
[2]H. L. Frisch and S. P. Lloyd, Phys. Rev. **120**, 1175 (1960).
[3]B. I. Halperin, Phys. Rev. **139**, A104 (1965).
[4]P. Lloyd, J. Phys. C **2**, 1717 (1969).
[5]T. P. Eggarter, M. H. Cohen, and E. N. Economou, Solid St. Comm. (to be published).
[6]P. Soven, Phys. Rev. **156**, 809 (1967).
[7]B. Velicky, S. Kirkpatrick, and H. Ehrenreich, Phys. Rev. **175**, 747 (1968).
[8]S. Kirkpatrick and T. P. Eggarter, Phys. Rev. B **6**, 3598 (1972).
[9]N. F. Mott, Adv. Phys. **16**, 49 (1967); M. H. Cohen, H. Fritzsche and S. R. Ovshinsky, Phys. Rev. Lett. **22**, 1065 (1969).
[10]P. A. Anderson, Phys. Rev. **109**, 1492 (1958).
[11]D. J. Thouless, J. Phys. C 3, 1559 (1970).
[12]E. N. Economou and M. H. Cohen, Phys. Rev. Lett. **25**, 1445 (1970).
[13]T. P. Eggarter, Phys. Rev. A **5**, 2496 (1972).
[14]R. Borland, Proc. R. Soc. A **274**, 529 (1963).
[15]E. N. Economou and C. Papatriantafillhou, Phys. Letters (to be published).
[16]This notation is used for example in R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[17]We assume that all our wavefunctions are real and satisfy some normalization condition $\int|\psi|^2\,dr = u_0^2$. This exhausts all possible wavefunctions since the Hamiltonian in Eq. (1) is real and we are dealing with a finite volume.
[18]It is sufficient to choose A so that it contains no two points $p = (\psi, E)$ and $p' = (\psi', E')$ with $(\psi|\psi') = 0$.
[19]Actually f depends also on η, but as long as we do not take the limit $\eta \to 0$ there is no need to show this dependence explicitly.
[20]J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[21]T. P. Eggarter, Phys. Rev. B **7**, 1727 (1973).
[22]Hofti Zadeh, Inf. Control **8**, 338 (1965); O. Borghi, Rev. Univ. Nac. Tucuman A **20**, 133 (1970).

On the utility of separable expansions for the t matrix

J. S. Levinger*

Institut de Physique Nucléaire, Division de Physique Théorique,† 91406, Orsay, France
(Received 5 April 1973)

Osborn used the property of compactness to show that there is an infinite mean square deviation between the t matrix for a local potential and any separable t matrix of finite rank. I present an alternate proof, using only the property of square integrability. The divergence of the mean square deviation arises from very large momenta. I argue that a separable t matrix can give a good approximation to the trinucleon energy, which is insensitive to values of the t matrix at very large momenta.

In a recent paper, Osborn[1] shows that a separable expansion of finite rank[2] does not converge to the t matrix for a local potential. In this note I summarize his argument, present a simplified proof of his conclusion, and then argue that the *approximation* of a separable t matrix is *likely* useful for certain calculations, such as the energy of the trinucleon.

Osborn's argument uses the separation of operators into those which are compact and those which are noncompact.[3] He also utilizes the lemma that a noncompact operator is not square-integrable. He first states that a local operator is noncompact. The t matrix for a local potential obeys the Lippmann–Schwinger equation

$$t(z) = v - vg_0(z)t(z). \tag{1}$$

Osborn shows that the second term on the right is square-integrable (provided z is not at a pole) and therefore compact. The difference of the noncompact operator v, and the compact second operator gives a noncompact $t(z)$. But a separable t matrix, $t^N(z)$ is square-integrable and therefore compact. Then the difference

$$\Delta t(z) \equiv t(z) - t^N(z) \tag{2}$$

is a noncompact operator, and therefore not square-integrable. That is, the mean-square deviation of $t^N(z)$ from $t(z)$ is infinite.

Sloan and Gray[4] have very recently examined Osborn's argument, and shown that the noncompactness of a local potential arises from large values of momenta. Their conclusion agrees with that of this note: that a separable approximation can still be useful for calculating quantities which are insensitive to the value of the t matrix at high momenta. Sloan's argument uses the property of compactness; while in this paper we confine ourselves to the property of square-integrability.

We simplify Osborn's proof, and obtain some insight as to the source of the failure of convergence, as follows. First, we show that a local potential v is not square-integrable. Since (from Osborn's paper) the term $vg_0(z)t(z)$ is square-integrable, the t matrix $t(z)$ is not square integrable. But $t^N(z)$ is square-integrable; then $\Delta t(z)$ defined by Eq. (2) is not square-integrable (in agreement with Osborn's result) giving our Eq. (7), below.

By definition, a local potential has a matrix element in $(\mathbf{r}, \mathbf{r}')$ space

$$\langle \mathbf{r}|v|\mathbf{r}'\rangle = v(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'). \tag{3}$$

Here $\delta(\mathbf{r} - \mathbf{r}')$ is a three-dimensional Dirac delta function. For determination of square-integrability, we evaluate the trace as a six-dimensional integral

$$\mathrm{Tr}(vv^\dagger) = \int \langle \mathbf{r}|v|\mathbf{r}'\rangle^2 d^3r d^3r' = \infty. \tag{4}$$

The infinity results from the squared delta function in the integrand.

It is of interest to evaluate $\mathrm{Tr}(vv^\dagger)$ for a local potential in the momentum representation, since we are concerned with applications to the three-nucleon system, which are generally made using this representation. From (3), the potential has matrix elements in momentum-space,

$$\langle \mathbf{p}|v|\mathbf{p}'\rangle = f(\mathbf{q}). \tag{5}$$

Here $\mathbf{q} = \mathbf{p} - \mathbf{p}'$, and $f(\mathbf{q})$ is the Fourier transform of $v(\mathbf{r})$. Note that a local potential is independent of the vector $\mathbf{Q} = \mathbf{p} + \mathbf{p}'$. We examine the integral, changing variables from $(\mathbf{p}, \mathbf{p}')$ to $(\mathbf{q}, \mathbf{Q})$:

$$\mathrm{Tr}(vv^\dagger) = \int \langle \mathbf{p}|v|\mathbf{p}'\rangle^2 d^3p d^3p' \equiv \int [f(\mathbf{q})]^2 d^3q \int d^3Q = \infty. \tag{6}$$

The infinity clearly arises from very large values of the magnitude Q; that is, from very large magnitudes of momenta p and p'. [This source could be anticipated, since the integrand in (6) is bounded; so the divergence which must occur to give agreement with (4) must come from the infinite limits for integration in momentum space.]

We now use the simple lemma that the sum (or difference) of an operator which is not square-integrable and an operator which is square-integrable gives us an operator which is not square-integrable. We apply this lemma twice: first where the operators are v and $vg_0(z)t(z)$, respectively; and then where the operators are $t(z)$ and $t^N(z)$, respectively. We thus obtain one of Osborn's results,

$$\mathrm{Tr}(\Delta t \Delta t^\dagger) = \int \langle \mathbf{p}|\Delta t(z)|\mathbf{p}'\rangle^2 d^3p\, d^3p' = \infty. \tag{7}$$

The divergence shown in (7) is very disturbing at first, since it appears that a separable t matrix is a very poor approximation. This seems paradoxical, since results for the trinucleon energy using a separable approximation to the t-matrix agree quite well with those using a t matrix for a local potential.[5] How can we reconcile this contradiction: t^N disagrees with t even for large N, as shown by (7); but it is useful to approximate t by t^N, even for the extreme case $N = 1$ (the unitary pole approximation t^u).

It is crucial to remind ourselves that we are arguing the utility of a separable approximation t^N for some specified purpose, such as calculation of the trinucleon energy E_T. Equations (6) and (7) show us that the approximation of separability breaks down at very large magnitudes of momenta, p and p'. But the trinucleon energy E_T is insensitive[6] to the values of the t matrix at very large values of p, p', or z, so the failure of t^N to be a good approximation for these conditions is

irrelevant to the utility of the approximation t^N for this specified purpose.

We can understand Lavine's result[6] using Fuda's t-matrix perturbation theory,[7] for the first-order correction to the triton energy:

$$\Delta E^{(1)} = \sum_{\alpha=1}^{3} [\langle\phi| - \langle\phi_\alpha|]\Delta t_\alpha(E_0)[|\phi\rangle - |\phi_\alpha\rangle]. \quad (8)$$

Here $|\phi\rangle$ is the triton state vector found solving the Faddeev equations using approximation t^N for the two-body t matrix. α is the "channel" for interaction between nucleons β and γ (different from each other, and different from α). The error $\Delta t_\alpha(E_0)$ uses (2) for channel α, at the energy E_0 found using t^N. $|\phi_\alpha\rangle$ is the component of the triton state vector in the α channel. When we evaluate (8) in the momentum representation, the factor $[|\phi\rangle - |\phi_\alpha\rangle]$ and its transpose each die off rapidly at high momenta (for usual choices of t^N). This convergence factor avoids the infinite result of (7), and in fact gives a rather small numerical result, of order $\frac{1}{5}$ MeV for central forces with soft cores.[5]

One could also argue that the divergence (7) is unlikely to occur if we replace $t(z)$ for a local potential by the (unknown) $t^R(z)$ for the real nucleon–nucleon t matrix that occurs in nature. Even a small nonlocality in the potential would smear out the delta function (3), and give some dependence on Q in (6).

But even if the real potential were exactly local, and the divergence (7) held, the approximation $t^N(z)$ is useful for calculation of E_T for central forces. The divergence (7) is useful as a "danger signal" to remind us that t^N is only an approximation, and that we must expect that there are limits to its utility. If we are calculating a quantity which is sensitive to $\langle \mathbf{p}|t(z)|\mathbf{p}'\rangle$ at very large magnitudes p and p', we should anticipate that the separable approximation t^N would be inaccurate. Since these matrix elements enter in calculation of electron–trinucleon elastic scattering for very large momentum transfers, we would expect failure of the separable approximation here. (It is likely that three other effects are also of significance in this problem: relativistic effects; effects of poorly-known three-body forces; and nonadditive effects in electron scattering.)

Returning for a moment to the problem of E_T, the main difficulty with use of the unitary pole separable approximation is the not very good agreement between t^u and t found[8] for noncentral forces.

ACKNOWLEDGMENT

I am grateful to T. A. Osborn and I. H. Sloan for sending preprints of their papers, and for their letters concerning these problems. I wish to thank P. Benoist-Gueutal, C. Marty, and P. Sauer for helpful oral discussions.

*On leave from Rensselaer Polytechnic Institute, Troy, New York 12181.
†Laboratoire Associé au C.N.R.S.

[1]T. A. Osborn, J. Math. Phys. **14,** 373 (1973).
[2]See for example E. Harms, Phys. Rev. C **1,** 1667 (1970); T. A. Osborn, Nucl. Phys. A **138,** 305 (1969).
[3]See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Sec. 7.3.
[4]I. H. Sloan and J. D. Gray, Phys. Lett. B **44,** 354 (1973).
[5]E. Harms and J. S. Levinger, Phys. Lett. B **30,** 449 (1969).
[6]J. P. Lavine, S. K. Mukhopady, and G. J. Stephenson, Phys. Rev. C **7,** 968 (1973).
[7]M. G. Fuda, Phys. Rev. **166,** 1064 (1968).
[8]B. Siebert, J. S. Levinger, and E. Harms, Nucl. Phys. A **197,** 33 (1972).

The relativistically invariant expansion of a scalar function on imaginary Lobachevski space

E. G. Kalnins

Centre de Recherches Mathématiques, Université de Montréal, Case Postale 6128, Montréal 101, Québec, Canada

(Received 15 September 1972; final revised manuscript received 7 March 1973)

Using the previous analysis of Gel'fand and Graev a new relativistically invariant expansion of a scalar function on three-dimensional imaginary Lobachevski space $L_3(I)$ is given. The coordinate system used corresponds to the horospherical reduction $SO(3,1) \supset E_2 \supset SO(2)$ and covers all of $L_3(I)$.

INTRODUCTION AND SUMMARY

Explicit relativistically invariant expansions of functions defined on the three transitivity surfaces of the proper Lorentz group in Minkowksi space have been studied to varying degrees in recent years.[1-3] Of these surfaces explicit expansions on the upper sheet H_2 of the double sheeted hyperboloid[4] $[x,x] = 1$ and on the cone[5] $[\xi,\xi] = 0$ have been well developed.[6] (Note: x is a 4-vector in Minkowski space with $[x,x] = x_0^2 - \mathbf{x}^2$ the usual scalar product.) The explicit expansions on H_2 and on the cone are based on the expansion formulas due to Gel'fand *et al.*[7] The invariant expansion of a scalar function $f(x)$ $(x \in H_2)$ is obtained by observing that H_2 corresponds to a realization of three-dimensional real Lobachevski space $L_3(R)$. An invertable horospherical integral transform then associates a function $h(\xi)$ on the cone with each $f(x)$. The invariant expansion of $f(x)$ then reduces to the invariant expansion of $h(\xi)$. The latter expansion is achieved by the decomposition of $h(\xi)$ into homogeneous components.

An analogous geometry and irreducible decomposition of a function $f(x)$ on the single sheeted hyperboloid H_1, with equation $[x,x] = -1$, has also been given in Ref. 7. The geometry of H_1 corresponds to a realization of imaginary Lobachevski space $L_3(I)$ and identifies diametrically opposed points [so that $f(x) = f(-x)$]. The irreducible decomposition on H_1 differs from that on H_2 in that it contains a discrete spectrum as well as the usual continuous spectrum.

Previously there has been (to the author's knowledge) one paper by Kuznetzov and Smorodinski[8] which has considered an explicit complete set of functions on H_1 realized as $L_3(I)$. This analysis uses the results of Ref. 7 only insofar as they consider a parametrization of $x \in H_1$ for which the discrete spectrum term is not necessary. [More specifically, they choose a coordinate system which only parametrizes points at a real distance from $x = (0,0,0,1)$.] Verdiev,[9] on the other hand, has given his attention to finding an explicit set of complete functions with spin on H_1. There are some shortcomings in Verdiev's work in that the continuous spectrum expansion functions have not been normalized and the method used to obtain the normalized discrete spectrum expansion functions needs some explanation. Zmuidzinas[2] has given a complete account of the expansion of a scalar function defined on H_1 using the eigenfunction expansion methods of Titchmarsh.[10] This analysis has been done in the canonical group reduction $SO(3,1) \supset SO(3) \supset SO(2)$ or S system. Limic *et al.*[3] have treated the general problem of the expansion of square integrable functions defined on the transitivity surfaces of $SO(p,q)$ in the canonical group reduction and hence include the results of Zmuidzinas as a special case.

In this paper we examine the expansion of a square integrable function defined on $L_3(I)$ in the noncanonical group reduction $SO(3,1) \supset E(2) \supset SO(2)$ or horospherical system. This expansion is new and serves to illustrate how the analysis of Gel'fand and Graev should be treated to yield the correct expansion formulas. There is only one other group reduction which parametrizes all of $L_3(I)$ (apart from the group reduction $SO(3,1) \supset E(2) \supset T_1 \otimes T_2$, which differs little from the horospherical system). This is the S system. We do not however give this expansion here as it differs little from the results of Zmuidzinas and Limic *et al.*

The study of the horospherical system group reduction of $SO(3,1)$ has received attention previously in application to particle physics[11,12] and is also of intrinsic group theoretical interest.

The content of this paper is arranged as follows. In Sec. 1 we collect the pertinent facts concerning the Gel'fand–Graev analysis on $L_3(I)$. In Sec. 2 we give the horospherical system expansion.

1. THE HARMONIC ANALYSIS OF A SCALAR FUNCTION ON $L_3(I)$

The central problem here is the decomposition of the representation

$$[T_g f](x) = f(xg), \quad x \in L_3(I) \tag{1.1}$$

into components which transform according to unitary irreducible representations (UIRs) of the proper Lorentz group $SO(3,1)$. The Gel'fand–Graev transform on $L_3(I)$ invertably maps $f(x)$ into a pair of functions $h(\xi)$ and $\phi(\xi,b)$. The function $h(\xi)$ gives the representation

$$[Q_g h](\xi) = h(\xi g) \tag{1.2}$$

and the functions $\phi(\xi,b)$ define the representation

$$[R_g \phi](l) = \beta^{-1}(l,g)\phi(lg), \tag{1.3}$$

where $\phi(l) = \phi(\xi,b)$ and $\beta(l,g)$ is the zeroth coordinate of ξg. This pair of functions are obtained by integration of $f(x)$ over the two distinct manifolds of horospheres on $L_3(I)$. [We assume that the reader is familiar with the rudiments of the geometry of $L_3(I)$ as found for instance in Ref. 7.] Accordingly, we have

(i) Horospheres of the first kind.

$$h(\xi) = \int f(x)\delta(|[x,\xi]| - 1)dx \tag{1.4}$$

with dx the invariant measure on $L_3(I)$

$$dx = \frac{dx_1 dx_2 dx_3}{|x_0|}. \tag{1.5}$$

Here a typical horosphere of the first kind has the equation

$$|[x,\xi]| = 1. \tag{1.6}$$

(ii) Horospheres of the second kind.

In this case $\phi(\xi,b)$ is obtained by integration of $f(x)$ over the isotropic line $x = b + t\xi$ according to

$$\phi(\xi,b) = \int_{-\infty}^{\infty} f(b + t\xi)dt, \tag{1.7}$$

where

$$[b,b] = -1, \quad [b,\xi] = [\xi,\xi] = 0, \quad b_0 = 0.$$

The choice of integration over an isotropic line is more convenient than over the horosphere itself. We note that each horosphere of the second kind given by $[x,\xi] = 0$ consists of all mutually parallel isotropic lines passing through the point ξ on the cone.

$f(x)$ is given in terms of $h(\xi)$ and $\phi(\xi,b)$ by the formula

$$f(x) = \frac{1}{(4\pi)^2}\int h(\xi)\delta^{(2)}(|[x,\xi]| - 1)d\xi + \frac{1}{(2\pi)^2}\int_0^{\pi}\cot^2\theta\, d\theta \int_\Gamma \phi(\xi,\theta)dw, \tag{1.8}$$

where

$$d\xi = \frac{d\xi_1 d\xi_2 d\xi_3}{|\xi_0|},$$

with $\phi(\xi,\theta)$ the value of $\phi(\xi,b)$ for the isotropic line $y = b + t\xi$ lying in the $[x,y] = \cos\theta$ plane (i.e., $[x,b] = \cos\theta$). Γ is a contour on the cone intersecting each generator once and the measure dw is defined by

$$dw = |\xi_0|^{-1}(\xi_1 d\xi_2 d\xi_3 - \xi_2 d\xi_1 d\xi_3 + \xi_3 d\xi_1 d\xi_2). \tag{1.9}$$

In order to achieve the decomposition of $f(x)$ into irreducible parts it is necessary to expand the "Fourier components" $h(\xi)$ and $\phi(\xi,\theta)$ into homogeneous components. For $h(\xi)$ this is done exactly as for the case of $L_3(R)$, i.e.,

$$h(\xi) = \frac{1}{2\pi i}\int_{\delta-i\infty}^{\delta+i\infty} F(\xi;\sigma)d\sigma, \tag{1.10}$$

$$F(\xi;\sigma) = \int_0^{\infty} h(t\xi)t^{-\sigma-1}dt. \tag{1.11}$$

The expansion of $\phi(\xi,\theta)$ into irreducible (homogeneous) components is achieved by Fourier analyzing $\phi(\xi,\theta)$ with respect to the angle $\theta(0 \le \theta < \pi)$ which specifies each isotropic line in a given horosphere of the second kind. The appropriate decomposition is

$$\phi(\xi,\theta) = \frac{1}{\pi}\sum_{n=-\infty}^{\infty}\tilde{F}(\xi;x;2n)e^{-2in\theta}. \tag{1.12}$$

The "Fourier coefficients" satisfy the homogeneity condition

$$\tilde{F}(\xi;b;2n) = \tilde{F}(\xi;x;2n)e^{-2in\theta}. \tag{1.13}$$

The invariant decomposition of $f(x)$ is then

$$f(x) = \frac{i}{4(2\pi)^3}\int_{\delta-i\infty}^{\delta+i\infty}\sigma(\sigma+1)\int_\Gamma F(\xi;\sigma)|[x,\xi]|^{-\sigma-2}d\xi d\sigma + \frac{1}{\pi^2}\sum_{n=1}^{\infty} n\int_\Gamma \tilde{F}(\xi;b;2n)e^{2in\theta}\delta([x,\xi])d\xi \tag{1.14}$$

and the inversion formulas are

$$F(\xi;\sigma) = \int f(x)|[x,\xi]|^{\sigma}dx, \tag{1.15}$$

$$\tilde{F}(\xi;b;2n) = \int f(x)e^{-2in\theta}\delta([x,\xi])dx. \tag{1.16}$$

Group theoretically the "Fourier coefficients" in (1.14) transform according to the irreducible representations (IRs) of $SO(3,1)$ as follows:

(i) $F(\xi;\sigma)$ transform according to the IRs

$$c = \sigma + 1 = \delta + 1 + i\rho, \quad (-\infty < \rho < \infty), \quad k_0 = 0, \tag{1.17}$$

where $[c,k_0]$ labels each IR of $SO(3,1)$. (This is the notation due to Naimark[13] that we are using here.) We obtain the unitary case (i.e., the principal series) when $\delta = -1$.

(ii) $\tilde{F}(\xi;b;2n)$ transform according to the UIRs $SO(3,1)$

$$c = 0, \quad k_0 = 2n, \quad n = 1,2,3,\cdots. \tag{1.18}$$

2. THE HOROSPHERICAL OR *Ho* SYSTEM EXPANSION ON $L_3(I)$

The *Ho* system 4-vector x on the single sheet hyperboloid H_1 is given by

$$x = (\tfrac{1}{2}[-e^{-a} + (1+r^2)e^{a}], re^{a}\cos\phi, re^{a}\sin\phi, \tfrac{1}{2}[-e^{-a} + (r^2-1)e^{a}], \tag{2.1}$$

$$-\infty < a < \infty, \quad 0 \le r < \infty, \quad 0 \le \phi < 2\pi.$$

This parametrization covers the $x_0 - x_3 \ge 0$ half of the $[x,x] = -1$ hyperboloid and so covers all of $L_3(I)$.

For the *Ho* system expansion the contour Γ is taken to be

$$\xi_0 - \xi_3 = 2 \tag{2.2}$$

and ξ is parametrized according to

$$\xi = (1 + u^2 + v^2, 2u, 2v, -1 + u^2 + v^2), \quad -\infty < u,v < \infty. \tag{2.3}$$

$F(\xi;\sigma)$ is expanded in a double Fourier series according to

$$F(\xi;\sigma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_{\lambda\mu}(\sigma)e^{i\lambda u}e^{i\mu v}d\lambda d\mu \tag{2.4}$$

and the measure on the cone is

$$d\xi = 4\, du\, dv. \tag{2.5}$$

Taking $x = (\mathrm{sha}, 0, 0, -\mathrm{cha})$ the continuous spectrum part of expansion (1.14) then reduces to the calculation of the integral

$$I^{\sigma}_{\lambda\mu}(a) = \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv\, |e^{a}(u^2+v^2) - e^{-a}|^{-\sigma-2}e^{i\lambda u}e^{i\mu v}. \tag{2.6}$$

This integral can be calculated by using the identity

$$|t|^{\beta} = t_+^{\beta} + t_-^{\beta}, \tag{2.7}$$

as well as the known Fourier transforms in two dimensions of the functions $(b^2 - u^2 - v^2)_{\pm}^{-\sigma-2}$ which are

given by

$$\text{F.T.}\left[\frac{(b^2-u^2-v^2)_-^{-\sigma-2}}{\Gamma(-\sigma-1)}\right] = -i(2b)^{-\sigma-1}\left[\frac{K_{-\sigma-1}(b(Q-i0))}{(Q-i0)^{-(\sigma+1)}}-\frac{K_{-\sigma-1}(b(Q+i0))}{(Q+i0)^{-(\sigma+1)}}\right], \tag{2.8}$$

$$\text{F.T.}\left[\frac{(b^2-u^2-v^2)_+^{-\sigma-2}}{\Gamma(-\sigma-1)}\right] = -i(2b)^{-\sigma-1} \times\left[e^{i\pi(\sigma+1)}\frac{K_{-\sigma-1}(b(Q-i0))}{(Q-i0)^{-(\sigma+1)}}-e^{-i\pi(\sigma+1)}\frac{K_{-\sigma-1}(b(Q+i0))}{(Q+i0)^{-(\sigma+1)}}\right], \tag{2.9}$$

where $Q=-\lambda^2-\mu^2$.

These formulas are special cases of the general formulas for the Fourier transforms in n dimensions of the generalized functions $(b^2+P)_\pm^\lambda$ ($\lambda \neq$ integer) as given by Gel'fand and Shilov.[14] (Note: P is a general quadratic form in the n Cartesian coordinate variables expressed in canonical or diagonal form.) $I^\sigma_{\lambda\mu}(a)$ is then found to be

$$I^\sigma_{\lambda\mu}(a) = i\pi\left(\frac{\chi}{2}\right)^{\sigma+1}\Gamma(-\sigma-1)e^{-a}[J_{\sigma+1}(e^{-a}\chi)+J_{-\sigma-1}(e^{-a})], \tag{2.10}$$

where $\chi=(\lambda^2+\mu^2)^{1/2}$.

For the discrete part of expansion (1.14), $\tilde{F}(\xi;2n)$ is expanded according to

$$\tilde{F}(\xi;2n) = \sum_{m=-\infty}^{\infty}\int_0^\infty \chi d\chi a_m(\chi;2n)J_{2n-m}(\chi p)e^{im\psi}, \tag{2.11}$$

where $u=p\cos\psi$, $v=p\sin\psi$

The evaluation of the discrete part of (1.14) then requires the calculation of

$$I_\chi(2n) = \int J_{2n-m}(\chi p)e^{im\psi}\delta([x,\xi])d\xi. \tag{2.12}$$

This integral is readily calculated, using the identity

$$\delta(a^2-x^2) = \frac{1}{2a}[\delta(a+x)+\delta(a-x)], \tag{2.13}$$

to be

$$I_\chi(2n) = 4\pi e^{-a}J_{2n}(\chi e^{-a})J_m(\chi r)e^{im\phi}. \tag{2.14}$$

The *Ho* system expansion on $L_3(I)$ is then

$$f(x) = \sum_{m=-\infty}^{\infty}\int_0^\infty \chi d\chi e^{-a}\left(\frac{-1}{8(2\pi)^2}\int_{\delta-i\infty}^{\delta+i\infty} \times\sigma(\sigma+1)a_m(\chi,\sigma)\Gamma(-\sigma-1) \times[J_{\sigma+1}(e^{-a}\chi)+J_{-\sigma-1}(e^{-a}\chi)]d\sigma + \frac{4}{\pi}\sum_{n=1}^{\infty} n a_m(\chi,2n)J_{2n}(\chi e^{-a})\right)J_m(\chi r)e^{im\phi}. \tag{2.15}$$

For the continuous part of (1.14) we have changed the expansion of $F(\xi;\sigma)$ to polar coordinates and used the identity

$$e^{i\chi r\cos(\phi-\theta)} = \sum_{m=-\infty}^{\infty} i^m J_m(\chi r)e^{im(\phi-\theta)}, \tag{2.16}$$

where $\tan\theta=\lambda/\mu$.

$a_m(\chi;\sigma)$ is then given by

$$a_m(\chi;\sigma) = i^m\int_0^{2\pi} a_{\lambda\mu}(\sigma)e^{-im\theta}d\theta. \tag{2.17}$$

The inversion formulas of (2.15) are

$$a_m(\chi;\sigma) = \frac{i}{4\pi}\Gamma(\sigma+1)\int f(x)e^{-a}[J_{\sigma+1}(e^{-a}\chi) + J_{-\sigma-1}(e^{-a}\chi)]J_m(\chi r)e^{-im\phi}dx, \tag{2.18}$$

$$a_m(\chi;2n) = \tfrac{1}{2}\int f(x)e^{-a}J_{2n}(\chi e^{-a})J_m(\chi r)e^{-im\phi}dx, \tag{2.19}$$

where

$$dx = e^{2a}da\,r\,dr\,d\phi. \tag{2.20}$$

For the principal series $\sigma=-1+i\rho$ the continuous part of (3.15) is an expansion in terms of the functions

$$\Psi^{Ho}_{\rho\chi m}(a,r,\phi) = e^{-a}\tilde{J}_{i\rho}(\chi e^{-a})J_m(\chi r)e^{im\phi} \tag{2.21}$$

which satisfy the orthogonality relations

$$\int \Psi^{Ho}_{\rho\chi m}(a,r,\phi)\overline{\Psi^{Ho}_{\bar\rho\bar\chi\bar m}(a,r,\phi)}dx = \frac{2\,sh\,\pi\rho}{\chi\rho}\delta(\rho-\bar\rho)\delta(\chi-\bar\chi)\delta_{m\bar m}, \tag{2.22}$$

where we have put

$$\tilde{J}_{i\rho}(x) = J_{i\rho}(x)+J_{-i\rho}(x). \tag{2.23}$$

We observe that the a dependant part of (2.22) reproduces the completeness relation for the Titchmarsh integral transform,[15] i.e.,

$$\int_0^\infty \tilde{J}_{i\rho}(x)\tilde{J}_{i\bar\rho}(x)x^{-1}dx = \frac{2\,sh\,\pi\rho}{\rho}\delta(\rho-\bar\rho). \tag{2.24}$$

We also note that for the discrete spectrum expansion functions we have the orthogonality relation

$$\int_0^\infty J_{2n}(x)J_{2m}(x)x^{-1}dx = \frac{1}{4n}\delta_{nm}. \tag{2.25}$$

This is just a special case of the formula

$$\int_0^\infty J_\mu(x)J_\nu(x)x^{-1}dx = \frac{2\sin\frac{1}{2}\pi(\nu-\mu)}{\pi(\nu^2-\mu^2)}. \tag{2.26}$$

CONCLUDING REMARKS

We have given here an expansion of a function $f(x)\in L_3(I)$ in a coordinate system which is an alternative to the canonical or S system, viz., the *Ho* system. The parametrization of x we used is obtained from the corresponding coordinate system vector on H_2 via the analytic continuation $a\to a+i\pi/2$. This example illustrates not only the application of the analysis of Gel'fand and Graev in obtaining explicit expansion formulas but also that the group reduction parametrizations of $x\in H_2$ when continued in the manner above do not always cover

all of H_1. The S system is the only one that covers all of H_1. It should be mentioned here that the expansion functions used for $\tilde{F}(\xi, 2n)$ in (2.11) are the natural ones in the sense that they are the basis functions for the UIR $\{o, 2n\}$ of $SO(3,1)$ when realized in a Ho system basis in the space of square-integrable functions in the plane.

In the future we intend to study all possible coordinate systems on H_1 which cover at least all of $L_3(I)$ and for which the angular part of the Laplacian Δ_L admits a separation of variables.

[1]For perhaps the most complete set of references see P. Winternitz "Two variable expansions based on the Lorentz and conformal groups," talk presented at Symposium on the de Sitter and conformal groups, Boulder, Colorado, 1970 and references contained therein.
[2]J. S. Zmuidzinas, J. Math. Phys. **7**, 764 (1966).
[3]N. Limic, J. Niederle, and R. Raczka, J. Math. Phys. **8**, 1079 (1967).
[4]N. Ya. Vilenkin and Ya. A. Smorodinski, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Sov. Phys.-JETP **19**, 1209 (1964)].
[5]M. A. Liberman and A. A. Makarov, Sov. J. Nucl. Phys. **9**, 766 (1969).
[6]E. G. Kalnins, Ph.D. thesis, University of Western Ontario (1972).
[7]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[8]G. I. Kuznetsov and Ya. A. Smorodinski, Sov. J. Nucl. Phys. **3**, 275 (1966).
[9]I. A. Verdiev, Sov. J. Nucl. Phys. **10**, 726 (1970).
[10]E. C. Titchmarsh, *Eigenfunction Expansions* (Clarendon Press, Oxford, 1962), Part I.
[11]S. J. Chang and L. O'Raifeartaigh, J. Math. Phys. **10**, 21 (1969).
[12]P. Winternitz, Ya. A. Smorodinski, and M. B. Sheftel, Sov. J. Nucl. Phys. **7**, 785 (1968).
[13]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).
[14]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.
[15]See in this connection Eq. 5, p. 150 of *Formulas and Theorems for the Special Functions of Mathematical Physics* by W. Magnus, F. Oberhettinger, and R. P. Soni (Springer-Verlag, New York, 1966).

Exact motion in noncentral electric fields

André Hautot*

University of Liège, Institute Physics, Sart Tilman par 4000 Liège I, Belgium

We study the problem of the motion of a charged particle in noncentral potentials of the type $f(\theta)/r^2 + V(r)$. Newton's and Schrödinger's mechanics are considered. Exact solutions exist if $V(r) = -H/r$ or Kr^2 (i.e., Coulomb or harmonic oscillator potentials) while $f(\theta)$ may have at least three different expressions as a function of θ if the problem is three-dimensional and seven expressions if it is two-dimensional. The classical trajectories are computed and the energy levels in the corresponding quantum problem are given. Analogies between the two treatments are discussed.

For a presumed complete bibliography about the problem of finding exact solutions to the equations of motion in the presence of unusual types of potentials see Refs. 1-10. Central magnetic fields were treated in a previous paper.[10] We now turn to noncentral electric potentials of the type

$$J = (\mu/\epsilon)[f(\theta)/r^2 + V(r)], \tag{1}$$

where μ is the mass and ϵ the charge of the particle, (r, θ, φ) its spherical coordinates. V represents its velocity, γ its acceleration. i equals $\sqrt{-1}$. We shall perform the calculations both for Newton's and for Schrödinger's mechanics.

I. NEWTON'S MECHANICS

Newton's equation can be written in the form

$$\gamma = -\,\mathbf{grad}\,[f(\theta)/r^2 + V(r)]$$

or in more detail

$$\gamma = -\,(1/r)V'(r)\mathbf{r} + 2f(\theta)\,\mathbf{r}/r^4 - (1/r^4)f'(\theta) \times [xz(x^2+y^2)^{-1/2}, yz(x^2+y^2)^{-1/2}, -(x^2+y^2)^{1/2}]. \tag{2}$$

A. The radial integration

The conservation of energy implies

$$v^2 + 2f(\theta)/r^2 + 2\,V(r) = a \quad (= \text{const}). \tag{3}$$

By scalar multiplication of (2) by $\mathbf{r}$ we obtain:

$$\mathbf{r}\cdot\gamma = -\,rV'(r) + 2f(\theta)/r^2.$$

From this equation we deduce

$$d(\mathbf{r}\cdot\mathbf{v})/dt = \mathbf{r}\cdot\gamma + v^2 = a - rV'(r) - 2V(r).$$

Remembering that $2\mathbf{r}\cdot\mathbf{v} = dr^2/dt$, we have after a classical integration

$$\mathbf{r}\cdot\mathbf{v} = (ar^2 - 2r^2V(r) - b)^{1/2}. \tag{4}$$

Finally,

$$\int dt = \int r(ar^2 - 2r^2V - b)^{-1/2}\,dr = F(r). \tag{5}$$

It is very remarkable that the radial motion is independent of the noncentral term in the potential (1). Equation (5) is exactly integrable by means of circular functions in the two classical cases:

$V = V_1 = -H/r$ (Coulomb potential),

$V = V_2 = Kr^2$ (harmonic oscillator).

In what follows we shall always restrict ourselves to these two possibilities and concentrate on the noncentral term in (1). It must be pointed out that $a > 0$ if $V = Kr^2$, but that $a < 0$ if $V = -H/r$ (for bound states).

B. Angular integrations

Let us define $\mathbf{P} = \mathbf{r} \times \mathbf{v}$; its modulus squared P^2 can be written as

$$P^2 = r^2v^2 - (\mathbf{r}\cdot\mathbf{v})^2.$$

Using (3) and (4) one finds

$$P^2 = b - 2f(\theta).$$

Although the modulus of the angular momentum is not a constant of the motion, the expression $P^2 + 2f(\theta)$ is conserved. On the other hand, in spherical coordinates, P^2 equals $r^4(\dot\theta^2 + \sin^2\theta\dot\varphi^2)$ where the point denotes time differentiation, and

$$r^4(\dot\theta^2 + \sin^2\theta\dot\varphi^2) = b - 2f(\theta). \tag{6}$$

By vector multiplication of (2) by $\mathbf{r}$ one finds

$$d\mathbf{P}/dt = -\,[f'(\theta)/r^2][-y(x^2+y^2)^{-1/2}, x(x^2+y^2)^{-1/2}, 0].$$

One concludes that $dP_z/dt = 0$, and after integration

$$P_z = r^2\sin^2\theta\varphi = d \quad (d = \text{const}). \tag{7}$$

Equations (6) and (7) allow us to find the two last integrations needed for the complete solution of this problem:

$$I = \int \frac{\sin\theta\,d\theta}{\sqrt{[b - 2f(\theta)]\sin^2\theta - d^2}} = \int \frac{1}{r}(ar^2 - 2r^2V - b)^{-1/2}\,dr, \tag{8}$$

$$J = d\int \frac{d\theta}{\sin\theta\sqrt{[b - 2f(\theta)]\sin^2\theta - d^2}} = \int d\varphi. \tag{9}$$

The problem is now solved. An exact solution exists provided the integrals present in (8) and (9) are elementary. The expression "exact solution" has been defined in our previous paper[10] as solutions expressible in terms of circular or at most elliptic functions.

Remark 1: Equation (7) indicates that P_z is a constant of the motion. If we can choose the Oxz plane so that it contains the initial vectors $\mathbf{r}(t=0)$ and $\mathbf{v}(t=0)$ $\mathbf{P}$ is directed along the y axis, so that $P_z = d = 0$. The trajectory is entirely contained in the Oxz plane: r and θ are the polar coordinates in the Oxz plane. Equation (9) vanishes and (8) determines the polar equation of the trajectory.

Remark 2: In connection with Remark 1 it must be pointed out that classical mechanics makes no distinction between the motion in the three-dimensional potential (1) (where $r^2 = x^2 + y^2 + z^2$) and the motion in the two-dimensional potential (1) (where $r^2 = x^2 + y^2$) provided $d = 0$ in the first case and $v_z = 0$ in the second case. In both cases the trajectory is located in a plane. It can happen (see examples below) that quadratures (8) and (9) may be exactly performed when $d = 0$ but not $d \neq 0$: then the exact solution for the three-dimensional problem only exists with suitable initial conditions (see Remark 1) without equivalent in the quantum formalism. With respect to the Schrödinger equation, an exact solution is to be expected only for the two-dimensional problem (i.e., in cylindrical coordinates).

1. Elementary integrations

In Remark 2 it was shown that if a Newtonian problem involving a potential of the type (1) is soluble in two dimensions it is also soluble in three dimensions provided suitable initial conditions are imposed. In fact, it suffices to choose adequately the orientation of the axis of reference. In view of future convenience in the comparison between the classical and quantum treatments of a same problem we must however distinguish potentials which lead to elementary quadratures for arbitrary d from those which need $d = 0$. It is not difficult to see that the following functions $f(\theta)$ satisfy the required condition.

Elementary integrations for arbitrary d-values:

$$\text{(a)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \cos^2\theta + \beta \cos\theta + \gamma) \sin^{-2}\theta, \tag{10}$$

$$\text{(b)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \cos^4\theta + \beta \cos^2\theta + \gamma) \sin^{-2}\theta \cos^{-2}\theta, \tag{11}$$

$$\text{(c)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \cot^2\theta + \beta \cot\theta + \gamma). \tag{12}$$

Elementary integrations when $d = 0$ only:

$$\text{(d)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \sin^2\theta + \beta \sin\theta + \gamma) \cos^{-2}\theta, \tag{13}$$

$$\text{(e)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \tan^2\theta/2 + \beta \tan\theta/2 + \gamma), \tag{14}$$

$$\text{(f)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \cot^2\theta/2 + \beta \cot\theta/2 + \gamma), \tag{15}$$

$$\text{(g)}\quad f(\theta) = (\hbar^2/2\mu^2)(\alpha \tan^2\theta + \beta \tan\theta + \gamma). \tag{16}$$

The parameters $\alpha, \beta,$ and γ have arbitrary constant values (apart from conditions specified below). The factor $\hbar^2/2\mu^2$ is introduced in view of future convenience. We shall solve in detail the problem involving the first of these potentials. For the others we shall restrict ourselves to the equation of the trajectory. We shall also investigate the condition under which the motion is stable, i.e.,

$$0 < r_{\min} < r < r_{\max}.$$

(a) Except for the trivial case $\alpha = \beta = \gamma = 0$, only one special case of (10) seems to have been investigated in the literature:[1] $\beta = 0$, $\alpha = -\gamma$ so that the total potential (1) remains central. The study of the motion in the electrical potential

$$J = (\mu/\epsilon)\,[(\hbar^2/2\mu^2) \times (\alpha \cos^2\theta + \beta \cos\theta + \gamma)/(r^2 \sin^2\theta) + V(r)], \tag{17}$$

where $V(r) = -H/r$ or Kr^2, seems to be new. We only integrate the case $d = 0$ for the sake of brevity. With the new variable $u = \cos\theta$ Eq. (8) becomes

$$-\int [-(b + \alpha\hbar^2/\mu^2)u^2 - (\beta\hbar^2/\mu^2)u + (b - \hbar^2\gamma/\mu^2)]^{-1/2} \times du = \int (ar^2 - 2r^2 V - b)^{-1/2}\,(1/r)\,dr.$$

We shall see in the final discussion that these quadratures lead to circular functions only if $\alpha, \beta,$ and γ obey the conditions

$$\alpha + \beta + \gamma \geq 0, \qquad \alpha - \beta + \gamma \geq 0. \tag{18}$$

If not, the solutions are logarithmic so that the motion is not stable. We shall therefore impose (18). Finally, for $V = V_1 = -H/r$ one obtains the polar equation of the trajectory:

$$(b + \alpha\hbar^2/\mu^2)^{-1/2} \arcsin \times \frac{2(b + \alpha\hbar^2/\mu^2)\cos\theta + \beta\hbar^2/\mu^2}{[\beta^2\hbar^4/\mu^4 + 4(b + \alpha\hbar^2/\mu^2)(b - \gamma\hbar^2/\mu^2)]^{1/2}} = b^{-1/2} \arcsin \frac{b - Hr}{r(H^2 + ab)^{1/2}} + C^t. \tag{19}$$

The trajectory is a rosette contained between two circles so that

$$b(H + \sqrt{H^2 + ab})^{-1} < r < b(H - \sqrt{H^2 + ab})^{-1}.$$

It is not difficult to see that the trajectory is closed if

$$b^{-1/2}(b + \alpha\hbar^2/\mu^2)^{1/2} = m/n, \text{ a rational number.}$$

When $V = V_2 = Kr^2$, the trajectory is found to be

$$(b + \alpha\hbar^2/\mu^2)^{-1/2} \arcsin \times \frac{2(b + \alpha\hbar^2/\mu^2)\cos\theta + \beta\hbar^2/\mu^2}{[\beta^2\hbar^4/\mu^4 + 4(b + \alpha\hbar^2/\mu^2)(b - \gamma\hbar^2/\mu^2)]^{1/2}} = (1/2)\,b^{-1/2} \arcsin \frac{2b - ar^2}{r^2(a^2 - 8bK)^{1/2}} + C^t$$

which again presents the aspect of a rosette so that

$$(2b)^{1/2}\,(a + \sqrt{a^2 - 8bK})^{-1/2} < r < (2b)^{1/2} \times (a - \sqrt{a^2 - 8bK})^{-1/2}.$$

It is closed if the above condition is fulfilled.

(b) We now study the motion in the electrical potential

$$J = (\mu/\epsilon)[(\hbar^2/2\mu^2) (\alpha \cos^4\theta + \beta \cos^2\theta + \gamma)/(r^2 \sin^2\theta \cos^2\theta) + V(r)]. \tag{20}$$

Equation (8) can immediately be integrated provided $\cos^2\theta$ is taken as new variable. The equation of the trajectory is (take $V = V_1$ for example)

$$(1/2)(b + \alpha\hbar^2/\mu^2)^{-1/2} \arcsin \times \frac{2(b + \alpha\hbar^2/\mu^2)\cos^2\theta - (b - \beta\hbar^2/\mu^2)}{[(b - \beta\hbar^2/\mu^2)^2 - 4(b + \alpha\hbar^2/\mu^2)\hbar^2\gamma/\mu^2]^{1/2}} = b^{-1/2} \arcsin \frac{b - Hr}{r(H^2 + ab)^{1/2}} + C^t.$$

The following conditions are needed for a stable motion:

$$\alpha + \beta + \gamma \geq 0, \qquad \gamma \geq 0. \tag{21}$$

(c) We study the motion in the electric potential

$$J = (\mu/\epsilon)\,[(\hbar^2/2\mu^2)(\alpha\,\cot^2\theta + \beta\,\cot\theta + \gamma) + V(r)]. \tag{22}$$

Equation (8) can be integrated provided the new (complex!) variable $e^{2i\theta}$ is taken. The equation of the trajectory is ($V = V_1$):

$$(i/2)[\hbar^2(\gamma - \alpha + i\beta)/\mu^2 - b]^{-1/2}\,\arcsin \times \frac{[\hbar^2(\beta + 2i\alpha)/\mu^2]\,\cot\theta + [\hbar^2(2\gamma + i\beta)/\mu^2] - 2b}{(\cot\theta - i)[(\hbar^4\beta^2/\mu^4) + 4(\hbar^2\alpha/\mu^2)(b - \hbar^2\gamma/\mu^2)]^{1/2}} + \text{c.c.} = b^{-1/2}\,\arcsin\frac{b - Hr}{r(H^2 + ab)^{1/2}} + \text{const.} \tag{23}$$

No restrictions on α, β, and γ have to be imposed.

In the cases (a), (b), and (c) both integrals (8) and (9) lead to elementary functions for arbitrary d-values. It is only for the sake of brevity that we put $d = 0$ in the equation of the trajectory. In the following problems (d), (e), (f), and (g) the condition $d = 0$ is needed.

(d) We study the motion in the electric potential

$$J = (\mu/\epsilon)[(\hbar^2/2\mu^2)(\alpha\,\sin^2\theta + \beta\,\sin\theta + \gamma)/(r^2\cos^2\theta) + V(r)]. \tag{24}$$

Here it is necessary to choose the orientation of the axis of reference so that $d = 0$ in (8) (see Remarks 1 and 2) if we want elementary quadratures. The equation of the trajectory is obtained by simply replacing θ by $\pi/2 - \theta$ in (19). Indeed potential (13) deduces from (10) in that way. However, (8) indicates that the sign of the first member of the equation of the trajectory must be inverted. The condition of stability is again (18).

(e) We study the motion in the electric potential

$$J = (\mu/\epsilon)\,[(\hbar^2/2\mu^2)(\alpha\,\tan^2(\theta/2] + \beta\,\tan(\theta/2) + \gamma) + V(r)]. \tag{25}$$

$$\left.\begin{aligned} f(\theta) &= (\alpha\,\cos^2\theta + \beta\,\cos\theta + \gamma)\,\cos^{-2}\theta \quad (\text{put } u = \cos\theta),\\ f(\theta) &= (\alpha\,\sin^2\theta + \beta\,\sin\theta + \gamma)\,\sin^{-2}\theta \quad (\text{put } z = \sin\theta),\\ f(\theta) &= (\alpha\,\cos\theta + \beta\,\sin\theta + \gamma)\,\cos^{-1}\theta\\ f(\theta) &= (\alpha\,\cos\theta + \beta\,\sin\theta + \gamma)\,\sin^{-1}\theta\\ f(\theta) &= (\alpha\,\cos\theta + \beta\,\sin\theta + \gamma\,\sin^2\theta + \delta\,\sin\theta\,\cos\theta + \epsilon) \end{aligned}\right\} \quad [\text{put } y = \tan(\theta/2)].$$

The list does not terminate here but we think it is of little interest to write it *in extenso*.

3. Other possibilities of exact motion

The conclusions of the preceding sections are valid with arbitrary initial conditions (see, however, Remarks 1 and 2). In this section we deal with exact motions allowed by suitable initial conditions. It was recently shown by Armenti and Havas[12] that an exact motion is sometimes possible outside the plane of symmetry $\theta = \pi/2$ when a monopole-prolate quadrupole potential acts on the particle. However, very special initial conditions are needed to this end. The authors noted that the conclusions are also valid when one considers noncentral potentials if in addition to an attractive radial force, there is a θ-component of the noncentral force directed away from the plane of symmetry. It must be pointed out that since the existence of such a movement depends on the initial conditions no equivalent can exist in the quantum formalism.

Equation (8) can be integrated with the new variable $\tan(\theta/2)$.

The trajectory is described by an equation of a rather unusual type:

$$i[\hbar^2(\gamma - \alpha + i\beta)/\mu^2 - b]^{-1/2}\,\arcsin \times \frac{[\hbar^2(\beta + 2i\alpha)/\mu^2]\,\tan(\theta/2) + [\hbar^2(2\gamma + i\beta)/\mu^2] - 2b}{[\tan(\theta/2) - i][(\hbar^4\beta^2/\mu^4) + 4(\hbar^2\alpha/\mu^2)(b - \hbar^2\gamma/\mu^2)]^{1/2}} + \text{c.c.} = -\,b^{-1/2}\,\arcsin\frac{b - Hr}{r(H^2 + ab)^{1/2}} + C^t. \tag{26}$$

Complex quantities are mixed to give a final real result; the problem (c) led to the same remark; there are no restrictions on the values of α, β, and γ.

(f) We study the motion in the electric potential

$$J = (\mu/\epsilon)\,[(\hbar^2/2\mu^2)(\alpha\,\cot^2(\theta/2) + \beta\,\cot(\theta/2) + \gamma) + V(r)]. \tag{27}$$

Since it follows from (25) by the substitution $\theta \to \pi - \theta$ Equation (8) indicates that the equation of the trajectory is obtained by carrying the same substitution in (26) after having inverted the sign of the first member. No restrictions about α, β, and γ.

(g) We study the motion in the electric potential

$$J = (\mu/\epsilon)[(\hbar^2/2\mu^2)(\alpha\,\tan^2\theta + \beta\,\tan\theta + \gamma) + V(r)] \tag{28}$$

The equation of the trajectory obviously follows from (23) by the substitution $\theta \to \pi/2 - \theta$, after inverting the sign of the first member. No restrictions about α, β, and γ.

2. Elliptic integrations

Let us make the equation of the trajectory (8) rational by a suitable change of variables. If the irrationality is of the third or of the fourth degree the trajectory may be written with the aid of elliptic functions. The potentials (1) for which $f(\theta)$ has the following values lead to elliptic integrals:

(a) We prove that the following exact motion is possible: $\dot\theta = 0, \dot\varphi = \omega = \text{const} \to \varphi = \omega t + \varphi_0$. The trajectory is thus a circle located in a plane at the distance $d = r\cos\theta$ from the plane of symmetry.

From the equations of motion[11]:

$$V'(r) - 2f(\theta)/r^3 = \omega^2 r\,\sin^2\theta,$$
$$f'(\theta)/r^3 = \omega^2 r\,\sin\theta\,\cos\theta.$$

We deduce

$$\omega^2 = f'(\theta)/(r^4\,\sin\theta\,\cos\theta),$$

and

$$r^3 V'(r) = 2f(\theta) + f'(\theta)\tan\theta.$$

Such a motion is therefore possible provided $f'\cos\theta > 0$ and $r^3 V' - 2f > 0$. The first equation determines the angular velocity while the second connects r and θ, i.e., it gives the distance $d = r\cos\theta$ between the plane of the trajectory and the plane of symmetry. As an example, let us investigate the case $V(r) = Ar^{n-2}$. Simple algebraic calculations show that

$$d = \{[2f(\theta) + f'(\theta)\tan\theta]/A(n-2)\}^{1/n}\cos\theta,$$

$$\omega = \pm\,[f'(\theta)/\sin\theta\cos\theta]^{1/2}\{A(n-2)/[2f(\theta)+f'\tan\theta]\}^{2/n}.$$

(b) Another special motion is deduced from: $\dot\theta = 0$, $\dot\varphi = 0$. The trajectory is located on a straight line passing through the origin. Of course such a motion also requires special initial conditions. Furthermore, if we restrict ourselves to potentials of the type (1), we must ensure that $\mathbf{r}\wedge\gamma = 0$ which leads to $(-y, x, 0)f' = 0$. If $f' = 0$, $f = \text{const}$ is fulfilled the problem may be solved exactly like a one-dimensional problem on account of the fact that the potential remains central. An exact straight line motion in a noncentral potential is also possible along the Oz axis because in this case $x = y = 0$.

II. SCHRÖDINGER'S MECHANICS

We have already seen (see Remark 2) that the quantum problem involving potentials like (1) is soluble in three dimensions if $f(\theta)$ is given by (10), (11) or (12) and that it is soluble in two dimensions in all cases (10) to (16).

A. The three-dimensional problem

We use spherical coordinates r, θ, φ. Schrödinger's equation takes the form:

$$\frac{\partial^2\psi}{\partial r^2} + (2/r)\frac{\partial\psi}{\partial r} + (1/r^2)\frac{\partial^2\psi}{\partial\theta^2} + (\cot\theta/r^2)\frac{\partial\psi}{\partial\theta} - (m^2/r^2\sin^2\theta)\psi + \frac{2\mu}{\hbar^2} \times [E - \mu f(\theta)/r^2 - \mu V(r)]\psi = 0.$$

The variables can be separated in the usual way: $\psi = \exp(im\varphi)\,\Theta(\theta)R(r)$ (m is the usual magnetic quantum number; we assume $m > 0$; calculations are analogous when $m < 0$). One has

$$r^2R'' + 2rR' + (2\mu/\hbar^2)r^2(E - \mu V(r))R + sR = 0, \qquad (29)$$

$$\Theta'' + \cot\theta\,\Theta' - (m^2/\sin^2\theta)\Theta - (2\mu^2/\hbar^2)f(\theta)\Theta - s\Theta = 0. \qquad (30)$$

As in the Newtonian formalism, the radial motion does not depend on the term $f(\theta)/r^2$ present in the potential. We next investigate the two cases mentioned in Sec. I.A.

Case 1: $V = V_1 = -H/r$.

Equation (29) reduces to the radial equation of a hydrogen-like system. The energy levels are

$$E = -(\mu^3H^2/2\hbar^2)[n + 1/2 + (1/4 - s)^{1/2}]^{-2}, \qquad (31)$$

where the parameter s may only take special values to be determined from (30).

Case 2: $V = V_2 = Kr^2$.

The radial equation (29) is the same as in the theory of the three-dimensional harmonic oscillator; the energy levels are given by

$$E = \hbar\sqrt{2K}\,[2n + 1 + (1/4 - s)^{1/2}], \qquad (32)$$

where again s is quantized. Of course when $f(\theta) = 0$, $s = -l(l+1)$ and we obtain the classical formulas for the energy levels of the hydrogen atom or of the harmonic oscillator.

It only remains to solve the θ-equation (which does not depend on the choice $V = V_1$ or $V = V_2$) to discover the allowed s-values.

The θ-equation is exactly soluble by means of known transcendental functions only when $f(\theta)$ is given by (10), (11) or (12).

(a) The quantum motion in the electric potential (17).

The θ-equation (30) becomes

$$\Theta'' + \cot\theta\,\Theta' - (m^2/\sin^2\theta)\Theta - (\alpha\cos^2\theta + \beta\cos\theta + \gamma)\sin^{-2}\theta\,\Theta - s\Theta = 0. \qquad (33)$$

We make the following substitutions:

$$v = \cos^2(\theta/2), \qquad \Theta = v^\rho(1-v)^\sigma T,$$

where

$$\rho = (1/2)(m^2 + \alpha - \beta + \gamma)^{1/2},$$

$$\sigma = (1/2)(m^2 + \alpha + \beta + \gamma)^{1/2},$$

$$v(1-v)T'' + [(2\rho + 1) - (2\rho + 2\sigma + 2)v]T' - [2\rho\sigma + s + \sigma + \beta/2 + 2\rho^2 + \rho - \alpha]T = 0.$$

We recognize the hypergeometric equation. The polynomial condition gives the allowed s-values. One finds ($k = 0, 1, 2, \cdots$)

$$T = F(-k, k + (m^2 + \alpha - \beta + \gamma)^{1/2} + (m^2 + \alpha + \beta + \gamma)^{1/2} + 1; 1 + (m^2 + \alpha - \beta + \gamma)^{1/2}; v),$$

$$/\,1/4 - s = -\alpha + (k + \rho + \sigma + 1/2)^2.$$

This relation must be introduced into (31) and (32) to obtain the energy levels when $V = V_1$ or $V = V_2$, respectively. In what follows we only consider $V = V_1$. (The other case is analogous.) The energy levels are by (31)

$$E = -(\mu^3H^2/2\hbar^2)\{n + 1/2 + \sqrt{-\alpha + [k + (1/2)(m^2 + \alpha - \beta + \gamma)^{1/2} + (1/2)(m^2 + \alpha + \beta + \gamma)^{1/2} + 1/2]^2}\}^{-2},$$

when $n,k = 0,1,2,\cdots$.

(b) The quantum motion in the electric potential (20).

The θ-equation (30) becomes

$$\Theta'' + \cot\theta\,\Theta' - (m^2/\sin^2\theta)\Theta - (\alpha\cos^4\theta + \beta\cos^2\theta + \gamma) \times \sin^{-2}\theta\cos^{-2}\theta\,\Theta - s\Theta = 0. \quad (34)$$

We make the following substitutions:

$$w = \cos^2\theta, \qquad \Theta = w^\rho(1-w)^\sigma T,$$

where

$$\rho = 1/4 + (1/4)(1+4\gamma)^{1/2},$$
$$\sigma = (1/2)(m^2+\alpha+\beta+\gamma)^{1/2},$$

$$w(1-w)T'' + [(2\rho+1/2) - (2\rho+2\sigma+3/2)w]T' - (1/4)(s+8\rho\sigma+m^2+\beta+2\gamma+2\sigma+4\rho)T = 0.$$

The solution is again the hypergeometric function

$$T = F(-k, k+1+(1/2)(1+4\gamma)^{1/2} + (m^2+\alpha+\beta+\gamma)^{1/2}; 1+(1/2)(1+4\gamma)^{1/2}; w),$$

$$1/4 - s = -\alpha + [2k+1+(1/2)(1+4\gamma)^{1/2} + (m^2+\alpha+\beta+\gamma)^{1/2}]^2,$$

with the energy levels:

$$E = -(\mu^3H^2/2\hbar^2)\{n+1/2+\sqrt{-\alpha+[2k+1+(1/2)(1+4\gamma)^{1/2}+(m^2+\alpha+\beta+\gamma)^{1/2}]^2}\}^{-2}, \quad \text{where } n,k = 0,1,2,\cdots.$$

(c) The quantum motion in the electric potential (22)

The θ-equation (30) becomes

$$\Theta'' + \cot\theta\,\Theta' - (m^2/\sin^2\theta)\Theta - (\alpha\cot^2\theta + \beta\cot\theta + \gamma)\Theta - s\Theta = 0.$$

We make the following substitutions:

$$z = e^{2i\theta}, \qquad \Theta = z^\sigma(1-z)^\tau T,$$

where

$$\sigma = (1/4) + (1/2)(1/4 - \gamma - s + i\beta + \alpha)^{1/2},$$
$$\tau = (m^2+\alpha)^{1/2}.$$

The T-equation is again hypergeometric. One finds ($k = 0,1,2,\cdots$)

$$s = (1/4) - \gamma + \alpha - \frac{(2k+1+2\sqrt{m^2+\alpha})^4 - 4\beta^2}{4(2k+1+2\sqrt{m^2+\alpha})^2},$$

$$T = F[-k, k+1+(1/4-\gamma-s+i\beta+\alpha)^{1/2} + 2(m^2+\alpha)^{1/2}; 1+(1/4-\gamma-s+i\beta+\alpha)^{1/2}; z].$$

The energy levels are given by (31) (we only consider the case $V = V_1$):

$$E = -(\mu^3H^2/2\hbar^2)\left[n+1/2 + \sqrt{\gamma-\alpha+\frac{(2k+1+2\sqrt{m^2+\alpha})^4-4\beta^2}{4(2k+1+2\sqrt{m^2+\alpha})^2}}\right]^{-2},$$

where $n,k = 0,1,2,\cdots$.

B. The two-dimensional problem

We use cylindrical coordinate r,θ,z. Schrödinger's equation is then written as

$$r^2\frac{\partial^2\psi}{\partial r^2} + r\frac{\partial\psi}{\partial r} + \frac{\partial^2\psi}{\partial\theta^2} + r^2\frac{\partial^2\psi}{\partial z^2} + (2\mu/\hbar^2)[Er^2 - \mu r^2V(r) - \mu f(\theta)]\psi = 0.$$

Variables can be separated in the usual way: $\psi = \exp(ip_z z/\hbar)R(r)\Theta(\theta)$. ($p_z$ is the z component of the momentum; it is a constant of the motion.) One has

$$r^2R'' + rR' + (2\mu/\hbar^2)r^2[E - (p_z^2/2\mu) - \mu V(r)]R - s^2R = 0, \quad (35)$$

$$\Theta'' - (2\mu^2/\hbar^2)f(\theta)\Theta + s^2\Theta = 0. \quad (36)$$

Once again the radial motion is independent of $f(\theta)$. $V = V_1$ or V_2 and the radial equation is analogous to that encountered in the problem of the two-dimensional hydrogen atom[12] or that of the two-dimensional oscillator. Energy levels are given by

Case 1: $V = V_1 = -H/r$,

$$E = (p_z^2/2\mu) - 2(\mu^3H^2/\hbar^2)(2n+2s+1)^{-2}, \quad (37)$$

Case 2: $V = V_2 = Kr^2$,

$$E = (p_z^2/2\mu) + \hbar\sqrt{2K}(2n+s+1), \quad (38)$$

$n = 0,1,2,\cdots$ while s may only take quantized values. These are found by solving (36).

The θ-equation is soluble in terms of known transcendental functions in seven cases: when $f(\theta)$ is given by (10)$\cdots$ or (16). Three of them have been treated in three dimensions. Therefore we shall omit the corresponding two-dimensional treatments.

(d) The quantum motion in the electric potential (24) (two-dimensional). The θ-equation (36) becomes

$$\Theta'' - (\alpha\sin^2\theta + \beta\sin\theta + \gamma)\cos^{-2}\theta\,\Theta + s^2\Theta = 0.$$

We make the following substitutions:

$$y = (1-\sin\theta)/2, \qquad \Theta = y^\rho(1-y)^\sigma T,$$

where

$$\rho = 1/4 + 1/4(1+4\alpha+4\beta+4\gamma)^{1/2},$$
$$\sigma = 1/4 + 1/4(1+4\alpha-4\beta+4\gamma)^{1/2},$$

$$y(1-y)T'' + [(2\rho+1/2) - (2\rho+2\sigma+1)y]T' - (1/2)(-2s^2+\rho+\sigma+4\rho\sigma+\gamma-\alpha)T = 0.$$

The solution is hypergeometric:

$$T = F(-k, k+1+(1/2)(1+4\alpha+4\beta+4\gamma)^{1/2} + (1/2)(1+4\alpha-4\beta+4\gamma)^{1/2}; 1+(1/2)(1+4\alpha+4\beta+4\gamma)^{1/2}; y),$$

$$s^2 = -\alpha + [k+1/2+(1/4)(1+4\alpha+4\beta+4\gamma)^{1/2} + (1/4)(1+4\alpha-4\beta+4\gamma)^{1/2}]^2.$$

The energy levels follow from

$$E = (p_z^2/2\mu) - 2(\mu^3 H^2/\hbar^2) \times \{2n + 1 + 2\sqrt{-\alpha + [k + 1/2 + (1/4)(1 + 4\alpha + 4\beta + 4\gamma)^{1/2} + (1/4)(1 + 4\alpha - 4\beta + 4\gamma)^{1/2}]^2}\}^{-2}$$

where $n, k = 0, 1, 2, \cdots$.

(e) The quantum motion in the electric potential (25) (two-dimensional). The θ-equation (36) becomes

$$\Theta'' - [\alpha \tan^2(\theta/2) + \beta \tan(\theta/2) + \gamma]\Theta + s^2\Theta = 0.$$

We make the following substitutions:

$$z = -e^{i\theta}, \quad \Theta = z^\rho (1 - z)^\sigma T,$$

where

$$\rho = (s^2 + \alpha - i\beta - \gamma)^{1/2},$$
$$\sigma = 1/2 + (1/2)(1 + 16\alpha)^{1/2},$$
$$z(1 - z)T'' + [(2\rho + 1) - (2\rho + 2\sigma + 1)z]T' - (2\rho\sigma + \sigma + 4\alpha - 2i\beta)T = 0.$$

The solution is hypergeometric:

$$T = F(-k, k + 2\rho + 1 + (1 + 16\alpha)^{1/2}; 2\rho + 1; z),$$

where

$$\rho^2 = -i\beta + (1/4)\{[k + (1/2) + (1/2)(1 + 16\alpha)^{1/2}]^4 - 4\beta^2\} \times [k + (1/2) + (1/2)(1 + 16\alpha)]^{-2}.$$

One finds

$$s^2 = -\alpha + \gamma + \frac{[k + (1/2) + (1/2)(1 + 16\alpha)^{1/2}]^4 - 4\beta^2}{4[k + (1/2) + (1/2)(1 + 16\alpha)^{1/2}]^2}.$$

The energy levels follow from

$$E = (p_z^2/2\mu) - 2(\mu^3 H^2/\hbar^2)\left\{2n + 1 + 2\sqrt{\gamma - \alpha + \frac{[k + (1/2) + (1/2)(1 + 16\alpha)^{1/2}]^4 - 4\beta^2}{4[k + (1/2) + (1/2)(1 + 16\alpha)^{1/2}]^2}}\right\}^{-2}, \quad \text{where } n, k = 0, 1, 2, \cdots.$$

(f) The quantum motion in the electric potential (27) (two-dimensional). The θ-equation (36) becomes

$$\Theta'' - (\alpha \cot^2(\theta/2) + \beta \cot(\theta/2) + \gamma)\Theta + s^2\Theta = 0.$$

It deduces from the θ-equation of Sec. II. B e by the substitution $\theta \to \pi - \theta$. Therefore, the wavefunction is obtained through the same procedure while the energy levels are given by the same formula.

(g) The quantum motion in the electric potential (28) (two-dimensional). The θ-equation (36) becomes

$$\Theta'' - (\alpha \tan^2\theta + \beta \tan\theta + \gamma)\Theta + s^2\Theta = 0.$$

We make the following substitutions:

$$z = 1 + e^{2i\theta}, \quad \Theta = z^\rho (1 - z)^\sigma T,$$

where

$$\rho = (1/2) + (1/2)(1 + 4\alpha)^{1/2},$$
$$\sigma = (1/2)(s^2 + \alpha - i\beta - \gamma)^{1/2},$$
$$z(1 - z)T'' + [-(2\rho + 2\sigma + 1)z]T' - [2\rho\sigma + \rho + \alpha - (i\beta/2)]T = 0.$$

The solution is hypergeometric:

$$T = F(-k, k + 1 + (1 + 4\alpha)^{1/2} + (s^2 + \alpha - i\beta - \gamma)^{1/2}; 1 + (1 + 4\alpha)^{1/2}; z)$$

$$s^2 = \gamma - \alpha + \frac{[(1 + 4\alpha)^{1/2} + 1 + 2k]^4 - 4\beta^2}{4[(1 + 4\alpha)^{1/2} + 1 + 2k]^2},$$

with energy levels

$$E = (p_z^2/2\mu) - 2(\mu^3 H^2/\hbar^2)\left\{2n + 1 + 2\sqrt{\gamma - \alpha + \frac{[(1 + 4\alpha)^{1/2} + 1 + 2k]^4 - 4\beta^2}{4[(1 + 4\alpha)^{1/2} + 1 + 2k]^2}}\right\}^{-2}, \quad \text{where } n, k = 0, 1, 2, \cdots.$$

C. Conditionally soluble quantum motions

In this section we shall investigate the solubility of Schrödinger's equation when more complicated potentials are considered, precisely those which lead to elliptic functions in the classical theory. Since these potentials are numerous we shall restrict ourselves to a special case in view of illustrating what we have called in a previous paper[10] the "conditional solubility".

We choose the special case

$$f(\theta) = (\hbar^2/2\mu^2)\,\alpha \cos^{-1}\theta$$

so that we investigate the quantum motion in the electric potential

$$J = (\mu/\epsilon)\,[\alpha(\hbar/2\mu^2)/(r^2 \cos\theta) + V(r)].$$

Considering the three-dimensional problem so that one has in spherical coordinates, the θ-equation (30) can be written as follows in spherical coordinates:

$$\Theta'' + \cot\theta\,\Theta' - m^2 \sin^{-2}\theta\,\Theta - \alpha \cos^{-1}\theta\,\Theta - s\Theta = 0.$$

We make the following substitutions:

$$u = \cos\Theta, \quad \Theta = (1-u^2)^{-m/2}\,T;$$

one finds

$$u(1-u)(-1-u)\,T'' + (2-2m)u^2\,T' + [(m^2-m+s)u+\alpha]\,T = 0.$$

This equation is of the general type

$$u(1-u)(\alpha-u)f'' + (au^2+bu+c)f' + (d+eu)f = 0.$$

We have studied it previously.[6] Polynomials solutions exist which ensure the integrability of $|\Theta|^2$ provided the four following conditions are fulfilled:[10]

$$\begin{cases} a+b+c = -j'(1-\alpha) \\ a\alpha^2+b\alpha+c = -j''\alpha(\alpha-1) \qquad (\text{where } j' \text{ and } j'' \text{ are integers} \geqslant 0) \\ e = -n(n+a-1) \\ + \text{ a "continuant" condition (see Ref. 10).} \end{cases}$$

Then one has $T = z^{j'+1}(1-z)^{j''+1}P^{(\nu)}$, where $P^{(\nu)}$ denotes a polynominal of degree ν.

The first two conditions are satisfied if $j' = j'' = m-1$.

The third implies $m^2 - m + s = -(\nu+2m)(m+\nu+1)$ which gives the allowed s-values.

The continuant condition is expressed by the vanishing of a continuant of order $\nu+1$. We have previously seen[10] that each value of ν must be analyzed separately leading to a quantization of the parameter α entering into the definition of the potential. For example, when $\nu = 2$ the determinant is of order three:

$$\begin{vmatrix} d & c & 0 \\ e & d+b & 2(c-1) \\ 0 & a+e & d+2b \end{vmatrix} = 0.$$

In this case, the problem is soluble only if $\alpha = \pm 2\sqrt{4m+2}$. Performing the same operation for each value of ν we arrive at the list of the allowed values for α. Inversely α being fixed (among the allowed values of course) ν, m and s are also fixed so that arbitrary angular momentum states are automatically forbidden. We retrieve the conclusions of our preceding paper.[10]

If $\nu = 2$, one finds $s = -(m+2)(m+3)$ and the energy levels are

$$E = -(\mu^3 H^2/2\hbar^2)[n+m+3]^{-2}.$$

Remark 3: This formula is analogous to the one giving the hydrogen spectrum except for the fact that the ground state and the first excited state are missing. Such a truncated hydrogen-like spectrum is found for every value of ν.

III. DISCUSSION AND CONCLUSION

In this paper we examined the various potentials of the type (1) which allow a complete integration of the equations of motion in both classical and quantum nonrelativistic mechanics. Our first conclusion is that the three-dimensional problem is completely soluble if the particle experiences the potentials (17), (20), or (22) while the two-dimensional problem is soluble when potentials (17), (20), (22), (24), (25), (27), or (28) are considered. To our knowledge these potentials were not treated before in the literature. It is interesting to compare the classical and the quantum treatments. As in our previous paper[10] it is possible to exhibit analogies from two different points of view:

(a) Firstly, there are some purely formal analogies: the resolution of Newton's equation and that of Schrödinger's equation offer many common points in spite of their well distinct origins. Variables separate in both equations for the same potentials. The changes of variables needed for the complete calculation are often identical. The analogy is sometimes very suggestive, e.g., the classical handling of potentials (22) or (25) leads to a trajectory whose equation contains complex quantities but so mixed that the overall result is real. The quantum equation leads to the same result: it is impossible to avoid the use of complex numbers in the calculation of the energy levels though the final expression is of course real. Beside the potentials mentioned above, there is a large class of potentials which allow a complete integration of the classical equation of motion in terms of elliptic functions. The corresponding problem in quantum mechanics leads to a conditional solubility analog to that previously encountered in a paper dealing with the motion in various magnetic fields.[10]

(b) Physical analogies also exist. Firstly we note that in classical mechanics potentials (17), (20), and (22) allow an exact solution whatever the choice of the axis of reference while potentials (24), (25), (27), (28) are soluble only for a special choice of these axes. In quantum mechanics the potentials of the first category are soluble in three dimensions (also in two) while those of the second category are soluble in two dimensions only. A Schrödinger's equation soluble by means of elementary transcendental functions (with energy levels) corresponds in the classical theory to a bounded trajectory expressible by means of circular functions: the parameters entering into the definition of the potential are submitted to analogous conditions in both mechanics in order to warrant a stable motion. Before terminating let us illustrate this last point with an example. Consider the motion in the potential (17); the quantum solution is given in Sec. II. A(a). The formula which determines the energy levels is only valid under the conditions

$$\begin{aligned} &\alpha - \beta + \gamma \geqslant 0, \\ &\alpha + \beta + \gamma \geqslant 0. \end{aligned} \tag{39}$$

It is not difficult to show that these conditions are sufficient to ensure the stability of the classical trajectory. Let us return to Sec. I. B(a). It is easily seen that the θ-integration leads to circular functions only if $b + \alpha\hbar^2/\mu^2 > 0$. Otherwise, the quadrature leads to logarithms and the trajectory spirals to $r = 0$ or to infinity. Since $b - 2f(\theta) = P^2 > 0$ it is sufficient that

$$b + \alpha\hbar^2/\mu^2 \geqslant b - 2f(\theta), \quad \text{where } f(\theta) \text{ is given by (10).}$$

After some reductions one obtains the condition

$$\alpha + \beta \cos\theta + \gamma \geq 0$$

to be compared with (39).

In conclusion, the connection between the classical and the quantum treatments of the problem here investigated can be stated as follows: when the classical trajectory is stable ($0 < r_{\min} < r < r_{\max}$) and when it is expressible by means of circular functions, the corresponding quantum problem is exactly soluble in terms of known functions and the energy spectrum contains a discrete part. If the trajectory is stable but can only be expressed by means of elliptic functions, the connection with the classical motion disappears. However, in that case the example treated in Sec. II. C indicates that the quantum energy spectrum may be discrete if the parameters entering in the definition of the potential are limited to discrete values. Finally, it must be pointed out that exact motions in classical mechanics sometimes occur when very special initial conditions are imposed. In this case no equivalence seems to exist in quantum mechanics.

*Presently Professor at the National University of Zaïre, Kinshasa.

[1]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles* (Cambridge U.P., Cambridge, 1960).
[2]André Hautot, Physica (Utr.) **58**, 37 (1972).
[3]N. D. Sengupta, *Satyendranath Bose, 70th birthday Commemoration Volume*, p. 55, Calcutta (1965).
[4]L. Lam, Phys. Lett. A **31**, 406 (1970).
[5]André Hautot, Phys. Lett. A **35**, 129 (1971).
[6]André Hautot, J. Math. Phys. **13**, 710 (1972).
[7]Harish-Chandra, Phys. Rev. **74**, 883 (1948).
[8]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[9]N. D. Sengupta, Ind. J. Pure Appl. Phys. 8, 765 (1970).
[10]André Hautot, J. Math. Phys. 14 (1973) (to be published).
[11]A. Armenti and P. Havas, *Relativity and Gravitation* (Gordan and Breach, New York, 1971).
[12]B. Zaslow and M. E. Zandler, Am. J. Phys. **35**, 1118 (1967).

Transition from time-dependent to time-independent multichannel quantum scattering theory

Colston Chandler*

Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131

A. G. Gibson†

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131
(Received 14 December 1972)

Rigorous proofs are given of several theorems establishing the connection between time-dependent and time-independent multichannel scattering theory. The method of proof involves a two-Hilbert-space formulation of time-dependent multichannel theory and the theory of spectral integrals. In particular, the time-independent theory in the form proposed by Alt, Grassberger, and Sandhas is derived.

1. INTRODUCTION

Several authors have recently reconsidered the problem of deriving the standard time-independent nonrelativistic quantum scattering theory from the time-dependent theory.[1-7] The principal technical flaw in the early work[8,9] on the subject was the absence of a rigorous justification for a certain interchange of the order of two integrations. The general theorem involved is by no means trivial and its proof occupies a sizeable number of pages in the most recent works.[1-3]

The object of this paper is to extend previous analyses[1-5] which were confined to single channel scattering to the general multichannel case. At the same time the limited results of previous multichannel papers[6,7] are extended. The conclusion is that the time-independent theory is specified by the multichannel scattering operator proposed by Alt, Grassberger, and Sandhas.[10,11]

It is also hoped that in addition to providing a rigorous justification for a formalism already in widespread use in theoretical physics, the analysis contained herein will also contribute to the discussion, current among certain mathematicians,[1,12] of what form a time-independent theory should have.

On the technical side, a question concerning the natural domains of definition of the time-independent formulae is resolved. The proofs in this paper are also somewhat cleaner than those given earlier for the single channel case.[1-4] The key to this improvement is a theorem, given here as Lemma 5, that is a new contribution to the (presently meager) literature on spectral integrals.

The paper is organized as follows:

In Sec. 2 the essential features of a two-Hilbert-space formulation of the multichannel time-dependent theory are presented in abstract form. This formulation, the validity of which is established in an appendix, is preferred over the more traditional one because of the greater structural clarity it lends to the theory. Also discussed in Sec. 2 is a curious feature of the multichannel problem that restricts the operator topology that can be used.

In Sec. 3A certain general mathematical results, including the previously mentioned Lemma 5, are collected. The theory of spectral integrals[1,2,13-15] is then exploited to obtain time-independent representations of the multichannel wave operators (Sec. 3B) and the multichannel scattering operator (Sec. 3C). The theorems and proofs of this section are presented without comment, the formulas obtained being largely self-explanatory.

The paper is concluded, in Sec. 4, with further discussion of certain aspects of the analysis.

2. TIME-DEPENDENT THEORY

In this paper time-dependent multichannel quantum scattering theory is defined by the following assumptions, collectively called assumption (W).

Assumption (W):

(W1) A spectral family E_λ defines a self-adjoint linear operator $H \equiv \int \lambda dE_\lambda$ with domain $\mathfrak{D}(H)$ dense in a separable Hilbert space $\mathfrak{K}$.

(W2) A spectral family E'_λ defines a self-adjoint linear operator $H' \equiv \int \lambda dE'_\lambda$ with domain $\mathfrak{D}(H')$ dense in a separable Hilbert space $\mathfrak{K}'$. The operator H' has absolutely continuous spectrum.

(W3) A bounded linear operator $J: \mathfrak{K}' \to \mathfrak{K}$ is defined. The operator J maps $\mathfrak{D}(H')$ into $\mathfrak{D}(H)$, and the adjoint operator J^* maps $\mathfrak{D}(H)$ into $\mathfrak{D}(H')$.

(W4) Multichannel wave operators $\Omega_\pm: \mathfrak{K}' \to \mathfrak{K}$,

$$\Omega_\pm \equiv \mathrm{s}-\lim_{t\to\pm\infty} e^{iHt}Je^{-iH't}, \tag{2.1}$$

are defined on $\mathfrak{K}'$. The adjoint wave operators $\Omega_\pm^*: \mathfrak{K} \to \mathfrak{K}'$ have on $\mathfrak{K}$ the representation

$$\Omega_\pm^* = \mathrm{w}-\lim_{t\to\pm\infty} e^{iH't}J^*e^{-iHt}. \tag{2.2}$$

The wave operators satisfy the equations

$$\Omega_\pm^*\Omega_\pm = I \quad \text{and} \quad \Omega_\pm\Omega_\pm^* = E_\pm, \tag{2.3}$$

where I is the identity on $\mathfrak{K}'$, and where the operators $E_\pm$ are the orthogonal projections of $\mathfrak{K}$ onto the ranges of $\Omega_\pm$. The operators $\Omega_\pm$ in addition map $\mathfrak{D}(H')$ into $\mathfrak{D}(H)$, and on $\mathfrak{D}(H')$ they satisfy the intertwining relation

$$H\Omega_\pm = \Omega_\pm H'. \tag{2.4}$$

(W5) The multichannel scattering operator $S: \mathfrak{K}' \to \mathfrak{K}'$ is defined by

$$S \equiv \Omega_+^*\Omega_-. \tag{2.5}$$

It is unitary if and only if $E_+ = E_-$.

The validity of Assumption (W) is established in an appendix on the basis of certain properties that are known[16,17] to be satisfied by systems of distinguishable spinless particles interacting via square integrable

pair potentials. As it is doubtless of wider validity, it is presented here as the defining assumption.

A curious, and important, feature of Assumption (W) is that the weak convergence in Eq. (2. 2) cannot be replaced by strong convergence. More precisely, it is proved in the Appendix that with the J defined there, strong convergence in Eq. (2. 2) implies that no bound subsystems of particles are possible. That is, if Eq. (2. 2) holds in the strong topology, the theory must necessarily be a single channel theory.

A consequence of the preceding remark is that the elegant theory of Belopol'skii and Birman,[5] which is couched in language similar to that of this paper, is a single channel theory. They prove that strong convergence in Eq. (2. 2) is a necessary consequence of a certain trace-class condition (A). Since condition (A) is fundamental to their paper, it appears that their theory is restricted in an essential way to the single channel problem.

Formulas derived in this paper for the multichannel scattering operator S are therefore valid only in the weak topology, in marked contrast to single channel formulas. This restriction causes no inconvenience in the present work, but it may complicate studies of the unitarity of the operator S. Unitarity arguments, such as those advanced by Lovelace[18] and by Sandhas and coworkers,[10,19] within the time-independent framework are certainly made more complicated and perhaps invalidated.

Efforts usefully to circumvent this inconvenience have so far failed, the only method the authors have found for obtaining strong topology formulas being the following. In multichannel scattering theory the space $\mathcal{H}'$ is actually a direct sum space (cf. Appendix), $\mathcal{H}' = \oplus_\alpha \mathcal{H}_\alpha$, where the separable Hilbert spaces $\mathcal{H}_\alpha$ are the spaces of asymptotic states of the various channels α. Define for all $\Phi = \oplus_\alpha \phi_\alpha$ in $\mathcal{H}'$ the injection operators

$$J_\pm \Phi = \sum_\alpha E_\pm^{(\alpha)} \phi_\alpha , \tag{2.6}$$

where $E_\pm^{(\alpha)}$ denotes the orthogonal projection of $\mathcal{H}$ onto the image under $\Omega_\pm$ of the channel subspace $\mathcal{H}_\alpha$. The operator J_+ is then to be used instead of J in formulas for Ω_+, and J_- in formulas for Ω_-. It is not difficult to prove that if J is replaced by $J_\pm$ in Assumption (W), then the assumption remains true even if the weak convergence in Eq. (2. 2) is replaced by strong convergence. Repetition of the analysis of this paper then leads to formulas that are valid in the strong topology. These formulas are, however, rather useless since advance knowledge of the ranges of the operators $\Omega_\pm$ is presumed. The formulas analogous to those of Theorem 5 of this paper are in addition hopelessly more complicated.

3. TIME-INDEPENDENT THEORY

A. Lemmata

Lemma 1: Let H_1 be a self-adjoint linear operator with domain $\mathfrak{D}(H_1)$ dense in a separable Hilbert space $\mathcal{H}_1$, and let H_2 be a self-adjoint linear operator with domain $\mathfrak{D}(H_2)$ dense in a separable Hilbert space $\mathcal{H}_2$. Let B: $\mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator with the property that B^*, the adjoint of B, maps $\mathfrak{D}(H_2)$ into $\mathfrak{D}(H_1)$. Then, for all complex numbers z with nonzero imaginary part, the operators $H_1 B^*(z - H_2)^{-1}$ and $(z - H_2)^{-1} B H_1$ are bounded.

Proof: The lemma has been established elsewhere (Lemma 1 of Ref. 1) for the case $\mathcal{H}_1 = \mathcal{H}_2$ and B the identity. Extension of that proof to the present case is straightforward, requiring only a modification of notation. QED

Lemma 2: Assume the following.

(i) Spectral families $E_\lambda^{(1)}$ and $E_\lambda^{(2)}$ are defined on respective separable Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$.

(ii) There is a family of bounded linear operators B_t: $\mathcal{H}_1 \to \mathcal{H}_2$ that are labeled by a parameter t that varies over a (finite or infinite) interval Δ of the real line.

(iii) There exists on $\Delta \times \Lambda$, where Λ is a (finite or infinite) interval on the real line, a complex-valued function $u(t, \lambda)$. For all $\lambda \in \Lambda$ and all $\phi \in \mathcal{H}_1$, the vector $u(t, \lambda) B_t \phi$ is Bochner integrable on Δ.

(iv) There is a real-valued Lebesgue integrable function $v(t)$ defined on Δ with the property that $|u(t, \lambda) \leq v(t)$ for all $t \in \Delta$, $\lambda \in \Lambda$.

Then the following statements are true.

(1) Suppose that the spectral integral $\int_\Lambda u(t, \lambda) dE_\lambda^{(1)} \phi$ exists for all $t \in \Delta$ and all $\phi \in \mathcal{H}_1$. Then, the existence for some $\psi \in \mathcal{H}_1$ of one of the integrals

$$\int_\Delta dt\, B_t \Big(\int_\Lambda u(t, \lambda)\, dE_\lambda^{(1)} \psi \Big) \quad \text{or} \quad \int_\Lambda \Big(\int_\Delta dt\, B_t u(t, \lambda) \Big) dE_\lambda^{(1)} \psi$$

implies the existence of the other and their equality.

(2) Suppose that the spectral integral $\int_\Lambda u(t, \lambda) dE_\lambda^{(2)} \phi$ exists for all $t \in \Delta$ and all $\phi \in \mathcal{H}_2$. Then the existence for some $\psi \in \mathcal{H}_1$ of one of the integrals

$$\int_\Delta dt\, \Big(\int_\Lambda u(t, \lambda) dE_\lambda^{(2)} \Big) B_t \psi \quad \text{or} \quad \int_\Lambda dE_\lambda^{(2)} \Big(\int_\Delta dt\, u(t, \lambda) B_t \psi \Big)$$

implies the existence of the other and their equality.

Proof: Suppose first that $\mathcal{H}_1 = \mathcal{H}_2$ and that $E_\lambda^{(1)} = E_\lambda^{(2)}$. Then the lemma is an adaptation of Theorems 3 and 3′ of Ref. 1 to the case of bounded (instead of unbounded) operators B_t. Extension of these previous proofs to the case $\mathcal{H}_1 \neq \mathcal{H}_2$, $E_\lambda^{(1)} \neq E_\lambda^{(2)}$ requires only a modification of notation. QED

Lemma 3: Assume the following.

(i) A spectral family $E_\lambda^{(1)}$ defines a self-adjoint linear operator $H_1 \equiv \int \lambda dE_\lambda^{(1)}$ with domain $\mathfrak{D}(H_1)$ dense in a separable Hilbert space $\mathcal{H}_1$.

(ii) Families of (possibly unbounded) linear operators $U(\lambda)$ and $V(\lambda)$ are defined with domains containing $\mathfrak{D}(H_1)$ and ranges in a separable Hilbert space $\mathcal{H}_2$. The parameter λ varies over a (finite or infinite) interval Λ of the real line. For all $\lambda \in \Lambda$ the operator $U(\lambda) - V(\lambda)$ is zero on $\mathfrak{D}(H_1)$.

Then, the existence for some $\psi \in \mathcal{H}_1$ of one of the integrals

$$\int_\Lambda U(\lambda) dE_\lambda^{(1)} \psi \quad \text{or} \quad \int_\Lambda V(\lambda) dE_\lambda^{(1)} \psi$$

implies the existence of the other one and their equality.

Proof: By assumption for any finite interval $[a, b]$ the equation $[U(\lambda) - V(\lambda)][E_b^{(1)} - E_a^{(1)}]\psi = 0$ holds for all λ in that interval. The truth of the lemma is now obvious upon expressing the integrals as strong limits of Riemann–Stieltjes sums. QED

Lemma 4: Let E_λ be a spectral family on a separable Hilbert space $\mathcal{H}$. Let $u(\lambda)$ and $v(\lambda)$ be two bounded opera-

tor-valued functions on a (finite or infinite) interval Λ of the real line. Assume that on any finite interval $K \subset \Lambda$ that

(i) $\sup_{\lambda \in K} \|u(\lambda)\| = M(K) < \infty$, and

(ii) there exists $L(K) < \infty$ and $\alpha(K) > \frac{1}{2}$ such that for all $\lambda, \mu \in K$ the condition $\|v(\lambda) - v(\mu)\| \leq L(K)|\lambda - \mu|^{\alpha(K)}$ holds.

Then the following statements are true.

(1) The existence for some $\psi \in \mathcal{H}$ of one of the integrals

$$\int_\Lambda u(\lambda)\left(\int_\Lambda v(\mu) dE_\mu\right) dE_\lambda \psi \quad \text{or} \quad \int_\Lambda u(\lambda) v(\lambda) dE_\lambda \psi$$

implies the existence of the other and their equality.

(2) The existence for some $\psi \in \mathcal{H}$ of one of the integrals

$$\int_\Lambda dE_\lambda \left(\int_\Lambda dE_\mu v(\mu)\right) u(\lambda)\psi \quad \text{or} \quad \int_\Lambda dE_\lambda v(\lambda) u(\lambda)\psi$$

implies the existence of the other and their equality.

Proof: Statement (1) is a slight generalization of Lemma 5 of Ref. 1. The proof of statement (2) does not differ in any important way from that of statement (1). QED

Lemma 5: Assume the following.

(i) A spectral family $E_\lambda^{(1)}$ defines a self-adjoint linear operator $H_1 = \int \lambda dE_\lambda^{(1)}$ with domain $\mathfrak{D}(H_1)$ dense in a separable Hilbert space $\mathcal{H}_1$.

(ii) A family of (possibly unbounded) linear operators U_λ is defined on $\mathfrak{D}(U_\lambda) \subset \mathcal{H}_1$ and has range in a separable Hilbert space $\mathcal{H}_2$. The labeling parameter λ varies over a (finite or infinite) interval Λ of the real line $\mathbf{R}$. At each point $\lambda \in \Lambda$ the domain $\mathfrak{D}(U_\lambda)$ contains $\mathfrak{D}(H_1)$, and for each finite subset $K \subset \Lambda$ there exist nonnegative constants $\alpha = \alpha(K)$ and $\beta = \beta(K)$ such that for all $\lambda \in K$ the inequality

$$\|U_\lambda \phi\|_2 \leq \alpha \|H_1 \phi\|_1 + \beta \|\phi\|_1 \tag{3.1}$$

holds for each $\phi \in \mathfrak{D}(H_1)$. In Eq. (3.1) the subscripts 1 and 2 indicate that the norms are to be taken in the spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively.

(iii) There is a complex-valued function $w(\eta, \lambda)$ defined on $\mathbf{R} \times \Lambda$ such that the integral

$$w(H_1, \lambda)\phi \equiv \int_{\mathbf{R}} w(\eta, \lambda) dE_\eta^{(1)} \phi \tag{3.2}$$

exists for all $\lambda \in \Lambda$ and all $\phi \in \mathfrak{D}(H_1)$. For every finite subset $K \subset \Lambda$ there are nonnegative constants $L = L(K)$ and $\gamma = \gamma(K)$, with $\gamma > \frac{1}{2}$, such that the inequality

$$|w(\eta, \lambda) - w(\eta', \lambda)| \leq L|\eta - \eta'|^\gamma \tag{3.3}$$

holds for all $\eta, \eta', \lambda \in K$.

Then, the existence for some $\psi \in \mathcal{H}_1$ of one of the spectral integrals

$$\int_\Lambda U_\lambda w(H_1, \lambda) dE_\lambda^{(1)} \psi \quad \text{or} \quad \int_\Lambda U_\lambda w(\lambda, \lambda) dE_\lambda^{(1)} \psi \tag{3.4}$$

implies the existence of the other and their equality. Further, if $\psi \in \mathcal{H}_2$ belongs to $\mathfrak{D}(U_\lambda^*)$ for all $\lambda \in \Lambda$, the existence of one of the spectral integrals

$$\int_\Lambda dE_\lambda^{(1)} w^*(H_1, \lambda) U_\lambda^* \psi \quad \text{or} \quad \int_\Lambda dE_\lambda^{(1)} w^*(\lambda, \lambda) U_\lambda^* \psi \tag{3.5}$$

implies the existence of the other and their equality.

Proof: It is sufficient to prove the theorem for finite Λ. For if Λ is infinite the integrals in Eq. (3.4) or Eq. (3.5) are understood to be strong limits of integrals over finite subsets $K \subset \Lambda$ as $K \to \Lambda$. If the theorem is true for all finite intervals K, and if one of the integrals over K has a limit as $K \to \Lambda$, then the other integral has a limit and the two limits are equal.

Assume therefore that Λ is finite with closure $[a, b]$. Let $\pi = \{a = \lambda_0 < \lambda_1 < \cdots < \lambda_n = b\}$ be a partition of $[a, b]$ and define

$$|\pi| = \sup_{1 \leq i \leq n} |\lambda_i - \lambda_{i-1}|. \tag{3.6}$$

For each i, $1 \leq i \leq n$, choose $\lambda_i' \in \Lambda_i = (\lambda_{i-1}, \lambda_i]$, and let $E^{(1)}(\Lambda_i) \equiv E^{(1)}\lambda_i - E^{(1)}\lambda_{i-1}$.

The theorem will follow from the definition of spectral integrals[1,2,9-11] if one can show that, for all $\psi \in \mathcal{H}_1$, $\|\Gamma_\pi \psi\|_2 \to 0$ as $|\pi| \to 0$, where

$$\Gamma_\pi \psi = \sum_{i=1}^n U_{\lambda_i'} B_i \psi, \tag{3.7}$$

$$B_i \psi \equiv [w(H_1, \lambda_i') - w(\lambda_i', \lambda_i')] E^{(1)}(\Lambda_i)\psi. \tag{3.8}$$

From Eq. (3.2) and Eq. (3.8) one concludes that $E^{(1)}(\Lambda_i)B_i = B_i$, and hence that B_i maps $\mathcal{H}_1$ into $\mathfrak{D}(H_1)$. It therefore follows from Eq. (3.1) with $K = \Lambda$ that

$$\|\Gamma_\pi \psi\|_2 \leq \sum_i \|U_{\lambda_i'} B_i \psi\|_2, \tag{3.9}$$

$$\leq \sum_i \{\alpha \|H_1 B_i \psi\|_1 + \beta \|B_i \psi\|_1\}. \tag{3.10}$$

Equation (3.3) now implies the easily derived bounds

$$\|H_1 B_i \psi\|_1 \leq L|\lambda_i - \lambda_{i-1}|^\gamma \|H_1 E^{(1)}(\Lambda_i)\psi\|_1, \tag{3.11a}$$

$$\|B_i \psi\|_1 \leq L|\lambda_i - \lambda_{i-1}|^\gamma \|E^{(1)}(\Lambda_i)\psi\|_1. \tag{3.11b}$$

These two bounds and the Schwarz inequality imply the further inequalities

$$\sum_i \|H_1 B_i \psi\|_1 \leq L|\pi|^{\gamma - 1/2}(b-a)^{1/2} \|H_1 E^{(1)}((a,b])\psi\|_1, \tag{3.12a}$$

$$\sum_i \|B_i \psi\|_1 \leq L|\pi|^{\gamma - 1/2}(b-a)^{1/2} \|E^{(1)}((a,b])\psi\|_1, \tag{3.12b}$$

where $E^{(1)}((a,b]) = E_b^{(1)} - E_a^{(1)}$. It is now apparent, upon substitution of inequalities (3.12) into Eq. (3.10) that $\|\Gamma_\pi \psi\|_2 \to 0$ as $|\pi| \to 0$. This proves the theorem with regard to Eq. (3.4). The remainder of the theorem follows from the preceding argument with Γ_π in Eq. (3.7) replaced by Γ_π^*, since the integrals in Eq. (3.5) are just the adjoints of those in Eq. (3.4). QED

B. Wave operators

In this subsection time-independent formulas for the multichannel wave operators are given.

There is first a two-Hilbert-space version of a result that has been proved for individual channels by other authors.[1-3,6,7]

Theorem 1: If Assumption (W) is true, the formula

$$\begin{aligned}\Omega_\pm &= \mathrm{s}\text{-}\lim_{\epsilon \to 0^+} \mp i\epsilon \int (\lambda \mp i\epsilon - H)^{-1} J dE_\lambda' \\ &= \mathrm{s}\text{-}\lim_{\epsilon \to 0^+} \pm i\epsilon \int dE_\lambda J (\lambda \pm i\epsilon - H')^{-1}\end{aligned} \tag{3.13}$$

is valid on $\mathcal{H}'$, and the formula

$$\Omega_\pm^* = \mathrm{w}-\lim_{\epsilon\to 0^+} \pm i\epsilon\int dE_\lambda' J^*(\lambda \pm i\epsilon - H)^{-1}$$
$$= \mathrm{w}-\lim_{\epsilon\to 0^+} \mp i\epsilon\int(\lambda \mp i\epsilon - H')^{-1}J^* dE_\lambda \qquad (3.14)$$

is valid on $\mathfrak{K}$.

Note: In the above and following formulas the $\pm$ and $\mp$ signs are to be read with either all upper or all lower signs, and all integrals are definite integrals over $(-\infty, \infty)$ unless otherwise indicated.

Proof: The proof is given for the first of Eq. (3.13), that of the other equations being essentially the same. Replace the asymptotic limit in Eq. (2.1) by the Abel limit

$$\Omega_\pm = \mathrm{s}-\lim_{\epsilon\to 0^+} \epsilon\int_0^\infty dt e^{-\epsilon t} e^{\pm iHt} J e^{\mp iH't}. \qquad (3.15)$$

Use Stone's theorem to replace the unitary operator $e^{\mp iH't}$ in Eq. (3.15) by its spectral representation. Then use Lemma 2 to reverse the order of spectral integration and t-integration. The resulting formulas are

$$\Omega_\pm = \mathrm{s}-\lim_{\epsilon\to 0^+} \int(\epsilon\int_0^\infty dt e^{(\mp i\lambda-\epsilon)t} e^{\pm iHt} J)\, dE_\lambda'. \qquad (3.16)$$

Evaluation of the Bochner integrals in Eq. (3.16) yields the first of Eq. (3.13) QED

A related problem of interest is the multichannel generalization of the single channel operator $\Omega_\pm - I$. The difficulty is that in multichannel theory the ranges of the operators $\Omega_\pm$ and I lie in different spaces. A simple way around this problem is to replace I by J. This leads to the following theorem.

Theorem 2: Let Assumption (W) be true, and let the multichannel potential operator V be defined by $V \equiv HJ - JH'$. Then the formula

$$\Omega_\pm - J = \mathrm{s}-\lim_{\epsilon\to 0^+} \int(\lambda \mp i\epsilon - H)^{-1} V dE_\lambda'$$
$$= \mathrm{s}-\lim_{\epsilon\to 0^+} -\int dE_\lambda\, V(\lambda \mp i\epsilon - H')^{-1} \qquad (3.17)$$

is valid on $\mathfrak{K}'$, and the formula

$$\Omega_\pm^* - J^* = \mathrm{w}-\lim_{\epsilon\to 0^+} \int dE_\lambda' V^*(\lambda \pm i\epsilon - H)^{-1}$$
$$= \mathrm{w}-\lim_{\epsilon\to 0^+} -\int(\lambda \pm i\epsilon - H')^{-1} V^* dE_\lambda \qquad (3.18)$$

is valid on $\mathfrak{K}$.

Proof: A proof is given of the first of Eq. (3.17), that of the other equations being essentially the same. Write $\Omega_\pm - J$ as an Abel limit,

$$\Omega_\pm - J = \mathrm{s}-\lim_{\epsilon\to 0^+} \epsilon\int_0^\infty dt e^{-\epsilon t}(e^{\pm iHt} J - J e^{\pm iH't}) e^{\mp iH't}. \qquad (3.19)$$

Use Stone's theorem to replace the unitary operator $e^{\mp iH't}$ in Eq. (3.19) by its spectral representation. Then use Lemma 2 to reverse the order of spectral integration and t-integration. The resulting formulas are

$$\Omega_\pm - J = \mathrm{s}-\lim_{\epsilon\to 0^+} \int\left(\epsilon\int_0^\infty dt e^{(\mp i\lambda-\epsilon)t}(e^{\pm iHt} J - J e^{\pm iH't})\right) dE_\lambda'. \qquad (3.20)$$

Evaluating the Bochner integral and applying the identity

$$(z - H)^{-1}J - J(z - H')^{-1} = (z - H)^{-1}V(z - H')^{-1} \qquad (3.21)$$

yields

$$\Omega_\pm - J = \mathrm{s}-\lim_{\epsilon\to 0^+} \mp i\epsilon\int(\lambda \mp i\epsilon - H)^{-1} V(\lambda \mp i\epsilon - H')^{-1} dE_\lambda'. \qquad (3.22)$$

The operator $(\lambda \mp i\epsilon - H)^{-1}V$ is, by Lemma 1, bounded. Lemma 5 can therefore be applied to replace the factor $(\lambda \mp i\epsilon - H')^{-1}$ by the numerical factor $(\mp i\epsilon)^{-1}$. This proves the first of Eq. (3.17). QED

The multichannel operators $\Omega_\pm - J$ are not the only multichannel generalizations of the single channel operators $\Omega_\pm - I$. Another possibility is the multichannel operators $J^*\Omega_\pm - I$. Time-independent representations of these operators can be derived in a manner analogous to Theorem 2. Since they add nothing of interest, they will not be reproduced here.

C. Scattering operator

In this subsection time-independent formulas for the multichannel scattering operator $S = \Omega_+^*\Omega_-$ are given.

There is first a two-Hilbert-space version of a result that was discovered for individual channels by Hunziker.[20]

Theorem 3: If Assumption (W) is true, the scattering operator S has on $\mathfrak{K}'$ the representation

$$S = \mathrm{w}-\lim_{\epsilon\to 0^+} (i\epsilon/2)\int dE_\lambda'\left(\int J^*[\tfrac{1}{2}(\lambda + \mu + i\epsilon) - H]^{-1} J dE_\mu'\right). \qquad (3.23)$$

The integral in Eq. (3.23) is a repeated spectral integral which may be evaluated in either order of integration.

Proof: According to Eq. (2.1) and Eq. (2.2) the operator S has the representation

$$S = \mathrm{w}-\lim_{t\to\infty} e^{iH't} J^* e^{-2iHt} J e^{iH't}. \qquad (3.24)$$

Replace the asymptotic limit in Eq. (3.24) by the Abel limit

$$S = \mathrm{w}-\lim_{\epsilon\to 0^+} \epsilon\int_0^\infty dt e^{-\epsilon t} e^{iH't} J^* e^{-2iHt} J e^{iH't}. \qquad (3.25)$$

Use Stone's theorem to replace the left-hand factor $e^{iH't}$ in Eq. (3.25) with its spectral representation. Then use Lemma 2 to reverse the order of spectral integration and t-integration. The resulting formula is

$$S = \mathrm{w}-\lim_{\epsilon\to 0^+} \int dE_\lambda'\left(\epsilon\int_0^\infty dt e^{(i\lambda-\epsilon)t} J^* e^{-2iHt} J e^{iH't}\right). \qquad (3.26)$$

Use Stone's theorem again to replace the factor $e^{iH't}$ in Eq. (3.26) with its spectral representation. Appeal to Lemma 2 to reverse the order of spectral integration and t-integration. The resulting formula is

$$S = \mathrm{w}-\lim_{\epsilon\to 0^+} \int dE_\lambda'\left(\int\left\{\epsilon\int_0^\infty dt e^{i(\lambda+\mu+i\epsilon)t} J^* e^{-2iHt} J\right\} dE_\mu'\right). \qquad (3.27)$$

Evaluation of the Bochner integral in Eq. (3.27) yields Eq. (3.23) with the indicated order of integration. To obtain the reverse order of integration apply Stone's theorem first to the right-hand factor in Eq. (3.25). The proof then proceeds in the same way as before. QED

A second theorem is a two-Hilbert-space version of a result that was stated for individual channels by Ekstein[9] and later modified by Sandhas *et al.*[10,11] A proof (using other methods) for three-body systems is found in the book of Faddeev.[7]

Theorem 4: Let Assumption (W) be true, and let $R \equiv S - I$, where I is the identity on $\mathfrak{K}'$. The operator R has on $\mathfrak{K}'$ the representation

$$R = \mathrm{w}-\lim_{\epsilon\to 0^+} (-2\pi i)\int dE'_\lambda \int \delta_\epsilon(\lambda-\mu)T([\lambda+\mu+i\epsilon]/2)dE'_\mu, \tag{3.28}$$

where

$$\delta_\epsilon(x) \equiv (\epsilon/\pi)(x^2+\epsilon^2)^{-1} \tag{3.29}$$

for all real x and

$$T(z) \equiv (z-H')J^*(z-H)^{-1}J(z-H') - (z-H') \tag{3.30}$$

for all complex z with nonzero imaginary part. The integral in Eq. (3.28) is a repeated spectral integral that may be evaluated in either order of integration.

Proof: Using Eq. (3.25) write $R = S - I$ as

$$R = \mathrm{w}-\lim_{\epsilon\to 0^+} R_\epsilon, \tag{3.31}$$

$$R_\epsilon = \epsilon\int_0^\infty dt e^{-\epsilon t}e^{iH't}(J^*e^{-2iHt}J - e^{-2iH't})e^{iH't}. \tag{3.32}$$

Proceed as in the proof of Theorem 3 to derive

$$R_\epsilon = \int dE'_\lambda X(\epsilon,\lambda), \tag{3.33}$$

$$X(\epsilon,\lambda) = \epsilon\int_0^\infty dt e^{(i\lambda-\epsilon)t}(J^*e^{-2iHt}J - e^{-2iH't})e^{iH't}, \tag{3.34}$$

$$= (i\epsilon/2)\int C([\lambda+\mu+i\epsilon]/2)dE'_\mu. \tag{3.35}$$

The operator C in Eq. (3.35) is defined for all complex z with nonzero imaginary part by

$$C(z) \equiv J^*(z-H)^{-1}J - (z-H')^{-1}. \tag{3.36}$$

$$= (z-H')^{-1}T(z)(z-H')^{-1}. \tag{3.37}$$

Consider next the operator

$$Y(\epsilon,\lambda) \equiv -2\epsilon\int_0^\infty dt e^{(i\lambda-\epsilon)t}(\lambda-H') \times (J^*e^{-2iHt}J - e^{-2iH't})(\lambda-i\epsilon-H')^{-1}e^{iH't}. \tag{3.38}$$

The operator $(J^*e^{2iHt}J - e^{2iH't})$ satisfies the conditions required of the operator B of Lemma 1. For each real λ and each $\epsilon > 0$ the integrand in Eq. (3.38) is therefore continuous in t and bounded in norm by a constant times the Lebesgue integrable function $e^{-\epsilon t}$. The Bochner integral $Y(\epsilon,\lambda)$ therefore exists. Proceed as with Eq. (3.34) to obtain

$$Y(\epsilon,\lambda) = -i\epsilon\int(\lambda-H')C([\lambda+\mu+i\epsilon]/2) \times (\lambda-i\epsilon-H')^{-1}dE'_\mu. \tag{3.39}$$

The factor $(\lambda - H')$ in the integrand of Eq. (3.39) is a closed operator. It can therefore be factored to the left of the integral (p. 413 of Ref. 1). The factor $(\lambda - i\epsilon - H')^{-1}$ is a bounded operator (for fixed $\epsilon > 0$) which commutes with E'_μ and can be factored to the right of the integral. It follows that the operator $X(\epsilon,\lambda)(\lambda - i\epsilon - H')^{-1}$ maps $\mathfrak{H}'$ into $\mathfrak{D}(H')$ and that

$$Y(\epsilon,\lambda) = -2(\lambda-H')X(\epsilon,\lambda)(\lambda-i\epsilon-H')^{-1}. \tag{3.40}$$

Apply Lemma 5 to obtain

$$0 = \int dE'_\lambda Y(\epsilon,\lambda). \tag{3.41}$$

Add Eq. (3.41) to Eq. (3.33) to obtain

$$R_\epsilon = \int dE'_\lambda \tilde{X}(\epsilon,\lambda), \tag{3.42}$$

$$\tilde{X}(\epsilon,\lambda) = X(\epsilon,\lambda) + Y(\epsilon,\lambda). \tag{3.43}$$

The operator $H'C(z)$ is, by Lemma 1, bounded. One may therefore apply Lemma 5 to Eq. (3.39) to replace the factor $(\lambda - i\epsilon - H')^{-1}$ by $(\lambda - i\epsilon - \mu)^{-1}$. Substitute the resulting representation for $Y(\epsilon,\lambda)$, together with the representation Eq. (3.35) for $X(\epsilon,\lambda)$, into Eq. (3.43) to obtain

$$\tilde{X}(\epsilon,\lambda) = (i\epsilon/2)\int[1 - 2(\lambda-\mu-i\epsilon)^{-1}(\lambda-H')] \times C([\lambda+\mu+i\epsilon]/2)\,dE'_\mu. \tag{3.44}$$

Substitution of the representation Eq. (3.37) for the operator C and straightforward manipulation of the integrand in Eq. (3.44) yields

$$\tilde{X}(\epsilon,\lambda) = -2\pi i\int\delta_\epsilon(\lambda-\mu)T([\lambda+\mu+i\epsilon]/2) \times \{\tfrac{1}{2}(\lambda-\mu+i\epsilon)[\tfrac{1}{2}(\lambda+\mu+i\epsilon) - H']^{-1}\}dE'_\mu. \tag{3.45}$$

By Lemma 1 the operator $T(z)(z-H')^{-1}$ is bounded for all z with nonzero imaginary part. The operator $T([\lambda+\mu+i\epsilon]/2)$ therefore satisfies Eq. (3.1) and Lemma 5 can be applied to replace the operator in braces in Eq. (3.45) by unity. This proves Eq. (3.28) in the indicated order of integration.

To prove the result with the reverse order of integration proceed as in the proof of Theorem 3 to derive from Eq. (3.32) the formula

$$R_\epsilon = \int U(\epsilon,\mu)dE'_\mu, \tag{3.46}$$

$$U(\epsilon,\mu) = \epsilon\int_0^\infty dt e^{(i\mu-\epsilon)t}e^{iH't}(J^*e^{-2iHt}J - e^{-2iH't}). \tag{3.47}$$

Apply Lemma 5 to prove that the operator $U(\epsilon,\mu)$ can be replaced in Eq. (3.46) by

$$\tilde{U}(\epsilon,\mu) = U(\epsilon,\mu) - 2(\mu-i\epsilon-H')^{-1}U(\epsilon,\mu)(\mu-H'). \tag{3.48}$$

Apply the lemmata as before to prove that the operator $\tilde{U}(\epsilon,\lambda)$ has on $\mathfrak{D}(H')$ the representation

$$\tilde{U}(\epsilon,\mu) = -2\pi i\int dE'_\lambda\delta_\epsilon(\lambda-\mu)T([\lambda+\mu+i\epsilon]/2). \tag{3.49}$$

This representation is now substituted for $U(\epsilon,\mu)$ in Eq. (3.46) to obtain Eq. (3.28) in the reverse order of integration. The use of a representation for $\tilde{U}$ that is valid only on $\mathfrak{D}(H')$ is justified by Lemma 3. QED

An alternative version of the previous theorem is a two-Hilbert-space version of a result established in a slightly different form in single channel theory by previous authors.[1,2]

Theorem 5: Let Assumption (W) be true, and let $R \equiv S - I$, where I is the identity on $\mathfrak{H}'$. The operator R has on $\mathfrak{H}'$ the representations

$$R = \mathrm{w}-\lim_{\epsilon_1\to 0^+}\mathrm{s}-\lim_{\epsilon_2\to 0^+}(-2\pi i)\int\left(\int dE'_\mu\delta_{\epsilon_1}(\lambda-\mu)\right) \times T(\lambda+i\epsilon_2)dE'_\lambda, \tag{3.50}$$

$$= \mathrm{s}-\lim_{\epsilon_1\to 0^+}\mathrm{w}-\lim_{\epsilon_2\to 0^+}(-2\pi i)\int dE'_\lambda T(\lambda+i\epsilon_2) \times\left(\int dE'_\mu\delta_{\epsilon_1}(\lambda-\mu)\right). \tag{3.51}$$

The operator T and the function δ_ϵ are the same as those defined in Theorem 4.

Proof: A proof is first given for Eq. (3.50). By Assumption (W) the operator $R = \Omega_+^*\Omega_- - I = (\Omega_+^* - \Omega_-^*)\Omega_-$ has the representations

$$R = \mathrm{w}-\lim_{t\to\infty}(e^{iH't}J^*\Omega_-e^{-iH't} - e^{-iH't}J^*\Omega_-e^{iH't}), \tag{3.52}$$

$$= \mathrm{w}-\lim_{\epsilon_1\to 0^+} \epsilon_1 \int_0^\infty dt e^{-\epsilon_1 t}(e^{iH't}J^*\Omega_- e^{-iH't} - e^{-iH't}J^*\Omega_- e^{iH't}). \tag{3.53}$$

The operator $J^*\Omega_-$ in Eq. (3.53) has the representation

$$J^*\Omega_- = \mathrm{s}-\lim_{\epsilon_2\to 0^+} F(\epsilon_2), \tag{3.54}$$

$$F(\epsilon_2) = \epsilon_2 \int_0^\infty ds e^{-\epsilon_2 s} J^* e^{-iHs} J e^{iH's}. \tag{3.55}$$

The operator $F(\epsilon_2)$ is clearly bounded for each $\epsilon_2 > 0$. It therefore follows from the Lebesgue dominated convergence theorem for Bochner integrals that

$$R = \mathrm{w}-\lim_{\epsilon_1\to 0^+} \mathrm{s}-\lim_{\epsilon_2\to 0^+} G_-(\epsilon_1, \epsilon_2), \tag{3.56}$$

where the operators $G_\pm(\epsilon_1, \epsilon_2)$ are defined by

$$G_\pm(\epsilon_1, \epsilon_2) = \epsilon_1 \int_0^\infty dt e^{-\epsilon_1 t}(e^{iH't}F(\epsilon_2)e^{-iH't} \pm e^{-iH't}F(\epsilon_2)e^{iH't}). \tag{3.57}$$

It is clear from Eq. (3.57) and the foregoing analysis that

$$\mathrm{s}-\lim_{\epsilon_2\to 0^+} (\epsilon_2/\epsilon_1)G_+(\epsilon_1, \epsilon_2) = 0 \tag{3.58}$$

for all $\epsilon_1 > 0$. It follows that

$$R = \mathrm{w}-\lim_{\epsilon_1\to 0^+} \mathrm{s}-\lim_{\epsilon_2\to 0^+} G(\epsilon_1, \epsilon_2), \tag{3.59}$$

$$G(\epsilon_1, \epsilon_2) = G_-(\epsilon_1, \epsilon_2) + (\epsilon_2/\epsilon_1)[G_+(\epsilon_1, \epsilon_2) - 2I]. \tag{3.60}$$

To obtain integral representations of the operators $G_\pm(\epsilon_1, \epsilon_2)$ use Stone's theorem to replace the right-hand factors $e^{\pm iH't}$ in Eq. (3.57) with their spectral representation. Apply Lemma 2 to reverse the order of spectral integration and t-integration. Perform the remaining Bochner integration to obtain

$$G_\pm(\epsilon_1, \epsilon_2) = -i\epsilon_1 \int [(\lambda - i\epsilon_1 - H')^{-1} \mp (\lambda + i\epsilon_1 - H')^{-1}]F(\epsilon_2)\, dE'_\lambda. \tag{3.61}$$

Substitution of Eq. (3.61) into Eq. (3.60) and manipulating the integrand yields

$$G(\epsilon_1, \epsilon_2) = -2(\epsilon_2/\epsilon_1)I + \int \{(\lambda + i\epsilon_2 - H')[(\lambda + i\epsilon_1 - H')^{-1} - (\lambda - i\epsilon_1 - H')^{-1}]\} F(\epsilon_2) dE'_\lambda. \tag{3.62}$$

By theorem 1,

$$F(\epsilon_2) = i\epsilon_2 \int J^*(\mu + i\epsilon_2 - H)^{-1} J dE'_\mu. \tag{3.63}$$

Substitute Eq. (3.63) into Eq. (3.62). By Lemma 1 the operator in the braces in Eq. (3.62) is a bounded operator. Lemma 4 may therefore be applied to obtain

$$G(\epsilon_1, \epsilon_2) = -2(\epsilon_2/\epsilon_1)I + i\epsilon_2 \int (\lambda + i\epsilon_2 - H')[(\lambda + i\epsilon_1 - H')^{-1} - (\lambda - i\epsilon_1 - H')^{-1}] J^*(\lambda + i\epsilon_2 - H)^{-1} J dE'_\lambda. \tag{3.64}$$

The operator $J^*(\lambda + i\epsilon_2 - H)^{-1}J$ maps, by Assumption (W3), the space $\mathfrak{K}'$ into $\mathfrak{D}(H')$. It is therefore permissible to commute the two left-hand factors in Eq. (3.64) to obtain

$$G(\epsilon_1, \epsilon_2) = \int [(\lambda + i\epsilon_1 - H')^{-1} - (\lambda - i\epsilon_1 - H')^{-1}] \times T(\lambda + i\epsilon_2)[i\epsilon_2(\lambda + i\epsilon_2 - H')^{-1}] dE'_\lambda + Z, \tag{3.65}$$

$$Z \equiv -2(\epsilon_2/\epsilon_1)I + i\epsilon_2 \int [(\lambda + i\epsilon_1 - H')^{-1} - (\lambda - i\epsilon_1 - H')^{-1}] dE'_\lambda. \tag{3.66}$$

The operator Z in Eq. (3.66) is, by Lemma 5, identically zero. The factor $i\epsilon_2(\lambda + i\epsilon_2 - H')^{-1}$ in Eq. (3.65) can, by Lemma 5, be replaced by unity. The final result, Eq. (3.50), is then obtained by replacing the operator $[(\lambda + i\epsilon_1 - H')^{-1} - (\lambda - i\epsilon_1 - H')^{-1}]$ by its spectral representation.

To prove Eq. (3.51) start with $R = \Omega_+^*(\Omega_- - \Omega_+)$ and proceed analogously. QED

It is amusing to compare Eq. (3.50) with the analogous formula that one obtains with the injection operators $J_\pm$ of Eq. (2.6). Define, for all complex numbers z with nonzero imaginary parts, the operators

$$T_\pm(z) \equiv (z - H')J_\pm^*(z - H)^{-1}J_-(z - H') - (z - H'). \tag{3.67}$$

Then the operator $R \equiv S - I$ has on $\mathfrak{K}'$ the representation

$$R = \mathrm{s}-\lim_{\epsilon_1\to 0^+} \mathrm{s}-\lim_{\epsilon_2\to 0^+} (-2\pi i) \int [R^{(1)}(\lambda, \epsilon_1, \epsilon_2) + R^{(2)}(\lambda, \epsilon_1, \epsilon_2)] dE'_\lambda, \tag{3.68}$$

where

$$R^{(1)}(\lambda, \epsilon_1, \epsilon_2) = \int dE'_\mu \delta_{\epsilon_1}(\lambda - \mu) T_+(\lambda + i\epsilon_2), \tag{3.69}$$

$$R^{(2)}(\lambda, \epsilon_1, \epsilon_2) = \int dE'_\mu [(\epsilon_1 - \epsilon_2)/2\pi](\lambda - \mu + i\epsilon_1)^{-1} \times (\lambda - \mu + i\epsilon_2)^{-1}[T_-(\lambda + i\epsilon_2) - T_+(\lambda + i\epsilon_2)]. \tag{3.70}$$

The term $R^{(1)}$ is what one expects, and the term $R^{(2)}$ is a remainder present because $J_+ \neq J_-$. The presence of this complicated remainder is of course a serious disadvantage of the formula Eq. (3.68), making it relatively useless.

4. DISCUSSION

It is to be emphasized that the analysis in Sec. 3 does not lead to any new conclusions. The operator $T(z)$ defined in Eq. (3.30) was originally introduced by Ekstein[9] and later modified by Sandhas and coworkers[10,11] into the form presented here. The importance of the present work is that it provides apparently the first rigorous proof that the multichannel operator $T(z)$ is the one to study.

A curious feature of the analysis is that if the weak convergence of the adjoint wave operators $\Omega_\pm^*$ in Eq. (2.2) is replaced by strong convergence, then the theory is restricted to single channel problems. Further comment on this point is found at the end of Sec. 2.

Another important feature of the analysis is that partial integration of Bochner integrals is avoided.[21] This gap in technique is filled by the use of Lemma 5, which is apparently new.[22] One advantage of proceeding this way is that the proofs of theorems are generally shortened. For example, the proof of Theorem 2 requires one paragraph, while previous proofs[1,2] were considerably longer. Another benefit of avoiding partial integrations is that the time-independent formulas are seen to be valid on all of $\mathfrak{K}'$, instead of being restricted[1,2,6] to $\mathfrak{D}(H')$. The approach of this paper has, therefore, certain advantages over previous approaches to the problem.

A third point concerns the interpretation of the function δ_ϵ defined in Eq. (3.29). It is this function that as $\epsilon \to 0^+$

is supposed to become the delta function that enforces energy conservation. Although this interpretation is probably correct, some caution in making it is needed. Lack of proper treatment of this point is known, for example, to lead to erroneous results in Coulomb scattering.[23,24] As an example of the pitfalls in the short-range case one may use the techniques of Sec. 2 to derive the formula

$$S = I + \mathrm{s}-\lim_{\epsilon\to 0^+}(-2\pi i)\int\left(\int dE'_\lambda(\lambda-\mu)\delta_\epsilon(\lambda-\mu)\Omega_+^* J\right)dE'_\mu . \tag{4.1}$$

The integral is a repeated spectral integral that can be evaluated in either order of integration. With a literal interpretation of δ_ϵ as a delta function, the quantity $(\lambda-\mu)\delta_\epsilon(\lambda-\mu)$ would be zero in the limit $\epsilon\to 0^+$. The conclusion would be that $S = I$, a clear absurdity. Fortunately, closer analysis reveals that this literal interpretation of δ_ϵ is not warranted in this example. When proper notice is taken of the order in which the operations in Eq. (4.1) are performed, it becomes clear that the factor $(\lambda-\mu)J$ is equivalent to, and can be replaced by, the operator $V = HJ - JH'$. Thus, Eq. (4.1) is equivalent to the representation

$$S = I + \mathrm{s}-\lim_{\epsilon\to 0^+}(-2\pi i)\int\left(\int dE'_\lambda\delta_\epsilon(\lambda-\mu)\Omega_+^* V\right)dE'_\mu . \tag{4.2}$$

The remaining factor δ_ϵ in Eq. (4.2) probably does enforce energy conservation, although no rigorous proof of that fact has been published. In the absence of such a proof, a degree of skepticism about the interpretation of δ_ϵ should be maintained.

In summary, the connection between time-dependent and time-independent multichannel quantum scattering theory has been rigorously established using techniques that effect certain mathematical economies over previous treatments of the single channel problem. Whether these techniques have further application in quantum scattering theory is a question for further research.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge discussions with Alan H. Paxton and Alan D. Sheall during the early stages of this work.

APPENDIX

In this appendix the validity of Assumption (W) is established for systems obeying Assumptions (A. I)–(A. III) below. This discussion is appended both for completeness and because previous formulations of a similar sort[9,25-27] have omitted proofs of certain details important for this paper. Specifically, proofs of the boundedness and domain preserving characteristics of J and J^* and of the existence of the limit in Eq. (2.1) have not been given. These are provided in Propositions 1 and 2. In addition it is proved in Proposition 3 that the weak limit in Eq. (2.2) cannot be replaced by a strong limit, a result apparently new.[28]

Consider a system of particles undergoing a scattering process. Asymptotically the particles arrange themselves into clusters, each of which is in a specific quantum mechanical bound state. A specification of both a clustering A and the corresponding bound states b_A is called a channel. The temporal evolution of the system from one channel to another is assumed to be consistent with the following principles.

(A. I) The temporal evolution of the complete system is governed by a one-parameter group e^{-iHt}, where the total Hamiltonian H is a self-adjoint operator with (dense) domain $\mathfrak{D}_H$ in a separable Hilbert space $\mathfrak{K}$.

(A. II) The temporal evolution of each channel α is governed by a one-parameter group $e^{-iH_\alpha t}$, where the channel Hamiltonians H_α are self-adjoint operators with (dense) domains $\mathfrak{D}_\alpha$ in separable Hilbert spaces $\mathfrak{K}_\alpha \subset \mathfrak{K}$ and with absolutely continuous spectra. The Hilbert space $\mathfrak{K}_0$ corresponding to a clustering with only one particle per cluster is the entire space $\mathfrak{K}$. The operators $\mathfrak{K}_\alpha$ for all channels α with a fixed clustering A have a common self-adjoint extension H_A with domain $\mathfrak{D}_A = \mathfrak{D}_H \subset \mathfrak{K}$.

(A. III) The channel wave operators

$$\Omega_\pm^{(\alpha)} = \mathrm{s}-\lim_{t\to\pm\infty} e^{iHt}e^{-iH_\alpha t}P_\alpha , \tag{A1}$$

where the P_α are the orthogonal projections of $\mathfrak{K}$ onto $\mathfrak{K}_\alpha$, exist on $\mathfrak{K}$ for all channels α. The orthogonality relation

$$P_\alpha P_\beta = \delta_{\alpha\beta}P_\alpha , \tag{A2}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta, is true if the channels α and β have the same clustering. The orthogonal projections $E_\pm^{(\alpha)}$ of $\mathfrak{K}$ onto the ranges of $\Omega_\pm^{(\alpha)}$ satisfy

$$E_\pm^{(\alpha)}E_\pm^{(\beta)} = \delta_{\alpha\beta}E_\pm^{(\alpha)} \tag{A3}$$

for all channels α and β.

Assumptions (A. I)–(A. III) are known to be true for systems of spinless particles that interact via square integrable pair potentials.[16,17]

To place the formalism in a two-Hilbert-space setting the direct sum Hilbert space $\mathfrak{K}' = \oplus_\alpha\mathfrak{K}_\alpha$ is formed, and the multichannel Hamiltonian H' is defined for all $\Phi = \oplus_\alpha\phi_\alpha$ in $\mathfrak{D}_{H'} \subset \mathfrak{K}'$ by the equation

$$H'\Phi = \oplus_\alpha H_\alpha\phi_\alpha . \tag{A4}$$

Communication between $\mathfrak{K}'$ and $\mathfrak{K}$ is provided by the injection operator J: $\mathfrak{K}' \to \mathfrak{K}$ defined for all $\Phi = \oplus_\alpha\phi_\alpha$ in $\mathfrak{K}'$ by the equation

$$J\Phi = \sum_\alpha \phi_\alpha . \tag{A5}$$

With these definitions Assumptions (W1) and (W2) follow from Assumptions (A. I) and (A. II). Assumption (W3) is verified by the following proposition.

Proposition 1: The operator J is bounded and maps $\mathfrak{D}_{H'}$ into $\mathfrak{D}_H$. The adjoint operator J^* is bounded and maps $\mathfrak{D}_H$ into $\mathfrak{D}_{H'}$.

Proof: For each clustering A let J_A: $\mathfrak{K}' \to \mathfrak{K}$ be defined for all $\Phi = \oplus_\alpha\phi_\alpha$ in $\mathfrak{K}'$ by the equation

$$J_A\Phi = \sum_\alpha{}^{(A)}\phi_\alpha , \tag{A6}$$

where $\sum_\alpha^{(A)}$ denotes summation over all channels with the clustering A. By Eq. (A2) the operators J_A are bounded: $\|J_A\Phi\|_{\mathfrak{K}} \le \|\Phi\|_{\mathfrak{K}'}$. Since

$$J\Phi = \sum_A J_A\Phi , \tag{A7}$$

where now the sum is over all clusterings A and hence is a finite sum, the boundedness of J and J^* is proved.

By Assumption (A. II), $H_A\phi_\alpha = H_\alpha\phi_\alpha$ for $\phi_\alpha \in \mathfrak{D}_\alpha$ with H_A being self-adjoint and hence closed. It follows that

$$J_A H'\Phi = \sum_\alpha {}^{(A)}H_\alpha \phi_\alpha = \sum_\alpha {}^{(A)}H_A\phi_\alpha = H_A J_A \Phi \tag{A8}$$

for $\Phi \in \mathfrak{D}_{H'}$. Thus J_A maps $\mathfrak{D}_{H'}$ into $\mathfrak{D}_A = \mathfrak{D}_H$. Equation (A7) now implies that J maps $\mathfrak{D}_{H'}$ into $\mathfrak{D}_H$. In a similar way the adjoint equation

$$H'J^*\psi = \sum_A J_A^* H_A \psi \tag{A9}$$

is proved for $\psi \in \mathfrak{D}_H$, thus proving that J^* maps $\mathfrak{D}_H$ into $\mathfrak{D}_{H'}$. QED

Assumption (W4) is established by noting that multichannel operators $\Omega_\pm$: $\mathcal{K}' \to \mathcal{K}$ are standardly defined by[9,16,17,29]

$$\Omega_\pm \Phi \equiv \sum_\alpha \Omega_\pm^{(\alpha)} \phi_\alpha . \tag{A10}$$

Equations (2.3) and (2.4) then follow from the corresponding single channel properties.[16,17] Equation (2.1) is verified in the following proposition.

Proposition 2: The operators $\Omega_\pm$: $\mathcal{K}' \to \mathcal{K}$ defined by Eqs. (2.1) and (A10) are the same. They exist if and only if the channel wave operators $\Omega_\pm^{(\alpha)}$ exist for all channels α.

Proof: Define $\Omega(t)$: $\mathcal{K}' \to \mathcal{K}$ for all $\Phi = \oplus_\alpha \phi_\alpha$ in $\mathcal{K}'$ by

$$\Omega(t)\Phi \equiv e^{iHt} J e^{-iH't}\Phi = \sum_\alpha e^{iHt} e^{-iH_\alpha t}\phi_\alpha . \tag{A11}$$

Let $\Phi \in \mathcal{K}'$ and $\epsilon > 0$ be arbitrary but fixed. Choose a vector $\Psi \in \mathcal{K}'$ with only a finite number of nonzero components ψ_α such that

$$\|\phi - \Psi\|_{\mathcal{K}'} < (\epsilon/2)(1 + \|J\|)^{-1}, \tag{A12}$$

where $\|J\|$ is the norm of J. Such a Ψ exists since vectors of that form are dense in $\mathcal{K}'$.

If the $\Omega_\pm^{(\alpha)}$ exist and if $\Omega_\pm$ are defined by Eq. (A10), then Proposition 1 and Eq. (A12) imply

$$\|[\Omega_\pm - \Omega(t)]\Phi\|_{\mathcal{K}} \le (\epsilon/2) + \sum_\alpha \|[\Omega_\pm^{(\alpha)} - e^{iHt}e^{-iH_\alpha t}]\phi_\alpha\|_{\mathcal{K}}. \tag{A13}$$

Since there are only a finite number of nonzero terms in the sum on the right side of Eq. (A13) there exists a T such that for $\pm t > T$ the sum is less than $(\epsilon/2)$. Existence of the limit of Eq. (2.1) and its identity with the operator defined in Eq. (A10) is established provided the channel wave operators exist. The converse part of the proposition follows immediately upon restriction of Eq. (2.1) to the subspaces $\mathcal{K}_\alpha$. QED

Equation (2.2) follows from Eq. (2.1) and the definition of an adjoint operator. Assumption (W4) is thus established.

Assumption (W5) follows from Eqs. (2.3) and (2.5).

This completes the verification of Assumption (W).

The weak convergence in Eq. (2.2) cannot be replaced by strong convergence. One might think, on the basis of experience with the singlechannel theory that strong convergence should hold on the subspace of $\mathcal{K}$ of absolute continuity of H. This is not generally the case, as the following proposition shows.

Proposition 3: Let P be the orthogonal projection of $\mathcal{K}$ onto the subspace of absolute continuity of H, and suppose that the equation

$$\Omega_\pm^* = \mathrm{s} - \lim_{t \to \pm\infty} e^{iH't} J^* e^{-iHt} P \tag{A14}$$

is true. Then the only nonzero channel wave operator is that for the free particle channel (the channel in which each cluster contains only one particle).

Proof: Denote the free particle channel by $\alpha = 0$. The Hilbert space $\mathcal{K}_0$ corresponding to the free particle channel is the entire space $\mathcal{K}$. Hence, if both Eqs. (2.1) and (A14) are true, then both of the strong limits

$$\Omega_\pm^{(0)} = \mathrm{s} - \lim_{t \to \pm\infty} e^{iHt} e^{-iH_0 t} \tag{A15}$$

and

$$\Omega_\pm^{(0)*} = \mathrm{s} - \lim_{t \to \pm\infty} e^{iH_0 t} e^{-iHt} P \tag{A16}$$

exist on $\mathcal{K}$. Thus $E_\pm^{(0)} = P$ (Ref. 30, Theorem X 3.5). However, $E_\pm^{(\alpha)}\mathcal{K} \subset P\mathcal{K}$ for all channels α (Ref. 30, Theorem X 3.2). It now follows from Eq. (A3) that, except for the free particle channel, all $E_\pm^{(\alpha)}$ and hence all $\Omega_\pm^{(\alpha)}$ are the zero operator. QED

Note added in proof: After seeing a preliminary version of this paper, E. Prugovečki [J. Math. Phys. **14**, 957 (1973)] independently proved some two-Hilbert-space formulas similar to those in this paper. F. Coester has pointed out to the authors that his two-Hilbert-space formulation in Refs. 25–27 also satisfies Assumption (W) of this paper.

*Supported by NASA during summer 1971.
†Supported in part by NSF Grant GP-12365.

[1] W. O. Amrein, V. Georgescu, and J. M. Jauch, Helv. Phys. Acta **44**, 407 (1971).
[2] D. B. Pearson, Nuovo Cimento A **2**, 853 (1971).
[3] E. Prugovečki, Nuovo Cimento B **4**, 124 (1971).
[4] C. Chandler and A. G. Gibson, Helv. Phys. Acta **45**, 734 (1972).
[5] A. L. Belopol'skii and M. S. Birman, Math. USSR Izv. 2, 1117 (1968).
[6] K. Hepp, Helv. Phys. Acta **42**, 425 (1969).
[7] L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering Theory* (Israel Program for Scientific Translations, Jerusalem, 1965), Chap. 11.
[8] B. A. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).
[9] H. Ekstein, Phys. Rev. **101**, 880 (1956).
[10] E. O. Alt, P. Grassberger, and W. Sandhas, Nucl. Phys. B **2**, 167 (1967).
[11] P. Grassberger and W. Sandhas, Nucl. Phys. B **2**, 181 (1967).
[12] T. Kato and S. T. Kuroda, Rocky Mtn. J. Math. **1**, 127 (1971).
[13] J. L. Dalecki and S. G. Krein, Am. Math. Soc. Transl. **47**, 1 (1965).
[14] M. S. Birman and M. Z. Solomyak, Topics in Math. Phys. **1**, 25 (1967).
[15] E. Prugovečki, SIAM J. Math. Anal. 3, 183 (1972).
[16] W. Hunziker, in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Britten (Gordon and Breach, New York, 1968), Vol. X-A.
[17] J. D. Dollard, Rocky Mtn. J. Math. **1**, 5 (1971).
[18] C. Lovelace, in *Strong Interactions and High Energy Physics*, edited by R. G. Moorhouse (Oliver and Boyd, London, 1964).
[19] W. Sandhas, "The Three-Body Problem," Universität Mainz preprint, 1972.
[20] W. Hunziker, *Lectures on Scattering Theory* (ETH, Zurich, 1968).
[21] The authors are indebted to Bill Celmaster, Harvard University, for correspondence concerning the subtleties of partial integrations.
[22] A result much less general than Lemma 5 but in the same vein is to be found on page 873 of Pearson's paper, Ref. 2, above.
[23] J. Nuttall and R. W. Stagat, Phys. Rev. A **3**, 1355 (1971).
[24] A. C. Chen and J. C. Y. Chen, Phys. Rev. A **4**, 2226 (1971).
[25] F. Coester, Helv. Phys. Acta **38**, 7 (1965).
[26] F. Coester, in *Quanta*, edited by P. Freund (University of Chicago Press, Chicago, 1970).
[27] F. Coester and L. Schlessinger, Ann. Phys. (N.Y.) **78**, 90 (1973).
[28] H. Ekstein and J. M. Jauch (private communications).
[29] J. M. Jauch, Helv. Phys. Acta **31**, 661 (1958).
[30] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), pp. 529–532.

On the structure of the canonical tensor operators in the unitary groups. III. Further developments of the boson polynomials and their implications*

J. D. Louck

Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87544

L. C. Biedenharn

Department of Physics, Duke University, Durham, North Carolina 27706
(Received 11 April 1973)

Three structurally distinct and explicit expressions are developed for the boson polynomials. The relationship of these polynomials to the representations of the general linear group and the Gel'fand–Graev generalized beta functions is noted. A by-product of these results is a new, closed expression for the irreducible representations of the symmetric group. Some similarities, as well as dissimilarities, between the boson polynomial forms and the canonical tensor operator forms are presented and discussed, the origin of these properties being traced to the similarities and distinctions between Wigner coefficients and Racah coefficients. One of the boson polynomial expressions is used to prove an important new relation in the Racah–Wigner calculus: the identity of the set of extended projective coefficients to a subset of Racah coefficients. This relationship becomes one-to-one for $SU(2)$ and establishes a pattern calculus for the Racah coefficients of angular momentum theory.

I. INTRODUCTION

In two previous papers[1,2] (hereafter referred to as I and II), we have utilized three principal tools in our investigations of the structural properties of the canonical tensor operators in the unitary groups: (a) the boson factorization lemma[3]; (b) the pattern calculus[4]; and (c) the algebraic relations which define the abstract $U(n)$ Racah–Wigner calculus[5–10] (coupling laws of Wigner operators, projective operators, etc.). Each of these tools together with their rich structural interrelations was demonstrated in I and II to be very effective for the calculation of certain explicit results (a class of reduced Wigner coefficients and Racah coefficients).

These three basic tools are developed still further in the present paper. Such a step appears to be necessary because, despite the fact that we have, for example, given a complete description in I of how to calculate *all* $U(3)$ reduced Wigner coefficients, the task of implementing this general procedure is extremely difficult. In broadening our knowledge of the basic tools, we not only uncover interesting details about these objects as mathematical structures, but also increase our capacity to calculate explicit matrix elements, which is our ultimate goal.

The factorization lemma has been a powerful tool for calculating explicit matrix elements,[1,4,8,11] but its usefulness has at the same time been somewhat limited by the lack of a fully explicit expression for the *general* "boson polynomial" occurring in it. In Sec. II we supply briefly the past developments, and then give three alternative forms for these boson polynomials (each of these forms corresponds to a different structural development), the derivation in each case proceeding directly from the factorization lemma itself. The significance[7] of these polynomials as irreducible representations (irreps) of the general linear group is emphasized, and the connection with the more recent work of Gel'fand and Graev[12] is noted. An interesting by-product of these general considerations is a remarkable formula for the real, orthogonal irreps of the symmetric group.

One form of the general boson polynomial is given in Sec. II as a sum over monomials in the "*fundamental bosons*" $a_i^j, i,j = 1, 2, \cdots, n$. It is also known[7,8] that a general $U(n)$ Wigner operator transforming as irrep $[m_{1n} m_{2n} \cdots m_{nn}]$ $(m_{nn} \geq 0)$ is a sum of monomials in the *fundamental Wigner operators*[13]

$$\left\langle \begin{matrix} & \tau & \\ [1 & & \dot{0}] \\ & i & \end{matrix} \right\rangle \tau, i, = 1, 2, \cdots, n.$$

By analogy to the explicit boson form of Sec. II, we are led in Sec. III to consider a similar sum of monomials in the fundamental Wigner operators. The relation of these irreducible tensor operators to the canonical Wigner operators is discussed, and, in particular, a fully explicit expression is given for the $\langle p \quad \dot{0} \rangle$ canonical Wigner operators in terms of the fundamental ones, the analogy in this case with the corresponding boson expression being quite striking. Aside from their possible calculational value, such relations provide us with considerable insight into the structure of the Racah–Wigner calculus and the significance of operator patterns. What we have in mind here may be described briefly in the following manner: If the fundamental Wigner operators all commuted, then we would build an arbitrary Wigner operator as a sum of monomials in the fundamental Wigner operators (over the field of complex numbers instead of over group invariants for scalars) in exactly the same way that we build an arbitrary boson polynomial from the fundamental bosons. If this were actually the case, then Racah coefficients would be identical with Wigner coefficients and operator patterns would be Gel'fand patterns, i.e., have a subgroup significance. *More precisely then it is the noncommutivity of the fundamental Wigner operators which accounts for the distinction between operator patterns and Gel'fand patterns and between Racah coefficients and Wigner coefficients.*

In Sec. IV and Appendix B a structural relation is established between two objects which enter the Racah–Wigner calculus in quite distinct ways—these objects are the Racah invariants and the extended projective operators.[8] This relation has particular significance for $U(2)$ because it now places the calculation of a class of $U(2)$ Racah coefficients within the framework of the (extended)

pattern calculus rules. Furthermore, the detailed form of this relation suggests that it may be possible to generalize the pattern calculus beyond its present stage. The proof of the relation for $U(2)$ is quite straightforward, but given in some detail because of its significance for angular momentum theory. The proof of the general relation for $U(n)$ is difficult, drawing as it does on (a) the factorization lemma in a form using the results of Sec. II; (b) the abstract coupling law for extended projective operators; and (c) the complicated relation between Racah coefficients derived in Ref. 9 (as a consequence of the associativity law for Wigner operator multiplication). Despite this complexity, the proof is quite interesting because it places the Racah coefficient identity of Ref. 9 in new perspective as an essential and important relation in the Racah–Wigner calculus.

II. THE BOSON POLYNOMIALS

A. Preliminary remarks

The explicit determination of all finite dimensional unitary irreducible representations of the family of unitary groups, $U(n)$, is an essential and important task. Although an implicit construction was already known from the fundamental work of Weyl,[14] the first major step in the *explicit* construction was the determination of the matrices of the generators of all $U(n)$ by Gel'fand and Zetlin[15] in 1950; the details of this determination, as well as the construction of the representations themselves, was given only later, by Gel'fand and Graev,[12] in 1965. An alternative construction of the matrices of the generators of $U(n)$—using boson operator techniques—was given by Baird and Biedenharn[16] in 1963; the explicit representations in this latter construction are designated as "boson polynomials," and have been discussed subsequently by many authors.[17]

The finite dimensional irreps of the generators of $U(n)$ also appear in a wider context: *they are at the same time irreducible representations of the generators of the general linear group, $GL(n, C)$.* This fact is basic to Gel'fand's approach; the special functions associated with these irreps turn out to be generalizations of Euler's beta function.

From the boson operator point-of-view these irreps play a fundamental role in the problem of constructing tensor operators; this results from the fact that the boson polynomials occur in the factorization lemma,[3,4] and their explicit form is required to evaluate the Wigner operators which occur in that lemma.

The present section has as one of its main purposes the explicit determination of the boson polynomials:

$$B\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(A). \tag{2.1}$$

Here A denotes the square array

$$A=\begin{pmatrix} a_1^1 & a_1^2 \cdots a_1^n \\ a_2^1 & a_2^2 \cdots a_2^n \\ \vdots & \\ a_n^1 & a_n^2 \cdots a_n^n \end{pmatrix}, \tag{2.2}$$

with each a_i^j being interpreted as an independent boson operator. The labels (m) and (m') denote Gel'fand patterns associated with the irrep $[m]$.[13]

Under the mapping $A \to U$, where U is a unitary matrix, these polynomials become the elements of a finite dimensional unitary matrix irrep $[m]$ of $U(n)$. Under the mapping $A \to Z$, where Z is an arbitrary nonsingular complex matrix, these same polynomials become the elements of a matrix irrep $[m]$ of $GL(n, C)$.[7]

One of the purposes of the present section is to discuss in some detail the relationship between the Gel'fand–Graev construction of these $GL(n, C)$ irreps and the boson polynomial approach.

Let us suppose that one has been given the task of determining all unitary irreps of $U(n)$: how might one proceed? The most direct procedure would be to proceed along the path laid out by Weyl: here one knows how to construct explicitly the unique vector in the carrier space of an irrep $[m]$ having maximal weight. Using lowering operators one then proceeds step-by-step to generate a basis of the carrier space, and then, using the matrices of the generators, to construct the finite transformations of the group. Such a procedure is quite practical for $U(2)$, useful even for $U(3)$, but loses its utility for arbitrary n. It is important to note that if one uses a boson operator realization in the process, then the construction of the basis vectors themselves is already a determination of certain of the boson polynomials[18,19] and hence of the finite transformations. The most complete results using this direct "lowering operator" approach have been given by Nagel and Moshinsky,[20] Louck,[21] and Holman.[11]

The disadvantage of this direct approach is that it becomes exceedingly cumbersome as n increases, and correspondingly difficult to implement. Moreover the very cumbersomeness of the technique itself obscures insight into the structural properties of the results being obtained.

The method employed by Gel'fand and Graev is very different. First they note that the irreps actually belong to $GL(n, C)$, as mentioned above. This has the important consequence that one is allowed to use transformations generated by (nonsingular) *upper* (or *lower*) *triangular matrices*; such matrices are (in general) not unitary and belong only to $GL(n, C)$.

Let us illustrate this method for $U(2)$. The representation consists in associating with each matrix

$$g=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$$

the operator $T(g)$ acting in the space $H_{m_1 m_2}$ of polynomials $P(z)$ whose degree is $(m_1 - m_2)$ or less. Then one has

$$T(g)P(z) = P\left(\frac{\alpha z+\gamma}{\beta z+\delta}\right)(\beta z+\delta)^{m_1-m_2}(\det g)^{m_2}. \tag{2.3}$$

Upon introducing an explicit basis, f_m, for $H_{m_1 m_2}$,

$$f_m(z) \equiv z^{m-m_2}, m_2 \leq m \leq m_1, \tag{2.4}$$

one may proceed to determine the explicit matrix elements of the representation. That is,

$$T(g)f_m = \sum_{m'}(m', m\,|g)f_{m'}, \quad \text{(Gel'fand–Graev notation)} \tag{2.5a}$$

$$= \sum_{m'} D^{j}_{m',m}(g)f_{m'}, j=\frac{m_1-m_2}{2} \quad \text{(Condon–Shortley–Wigner notation).} \tag{2.5b}$$

So far, this procedure is quite familiar; the $(m', m\,|g)$, or $D^j_{m',m}(g)$, are simply related to the Jacobi polynomials.

The novelty of the Gel'fand–Graev procedure is that they factorize the matrix g into three simpler $GL(2,C)$ matrices:

$$g \equiv \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \alpha \neq 0,$$

$$\equiv z\delta\zeta, \tag{2.6}$$

where

$$z = \begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}, \delta = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \zeta = \begin{pmatrix} 1 & \zeta \\ 0 & 1 \end{pmatrix}. \tag{2.7}$$

It follows that the representation $T(g)$ similarly factors:

$$T(g) = T(z)T(\delta)T(\zeta). \tag{2.8}$$

Hence $T(g)$ is completely determined by the special transformations $T(z)$, $T(\delta)$, and $T(\zeta)$, which are easily found.

In order to make the method practical for $GL(n,C)$, the calculation is embedded recursively into $GL(n+1,C)$. That is, one partitions the $(n+1)\times(n+1)$ matrix g_{n+1} in the form:

$$g_{n+1} = \left(\begin{array}{c|c} n\times n & n\times 1 \\ \hline 1\times n & 1\times 1 \end{array}\right), \tag{2.9}$$

and considers at each step the special matrices,

$$Z^-: z = \left(\begin{array}{c|c} 1_{n\times n} & 0 \\ \hline z_n & 1 \end{array}\right), \tag{2.10}$$

$$Z^+: \zeta = \left(\begin{array}{c|c} 1_{n\times n} & \zeta_n \\ \hline 0 & 1 \end{array}\right), \tag{2.11}$$

and embeds G_n in G_{n+1} by the rule

$$g_n \to \left(\begin{array}{c|c} g_n & 0 \\ \hline 0 & 1 \end{array}\right). \tag{2.12}$$

One then puts $g_{n+1} \in G_{n+1}$ into the form[22]

$$g_{n+1} = \lambda z g_n \zeta, \tag{2.13}$$

where λ = scalar, $z \in Z^-$, $g_n \in G_n$, $\zeta \in Z^+$.

The transformation $T(g_{n+1})$ similarly factors:

$$T(g_{n+1}) = T(\lambda 1)T(z)T(g_n)T(\zeta). \tag{2.14}$$

In Ref. 12, the explicit (but recursive) form for these special transformations $T(z)$ and $T(\zeta)$ are given. (We discuss these results further in Sec. IIF below.)

The method by which the boson polynomials are determined by boson operator techniques—as will be discussed in detail below—stands in marked contrast to the method of Gel'fand–Graev. The essential element in the boson approach is to employ *a coupling law for the boson polynomials so that general polynomials are constructed from more elementary constituents*. We may express the coupling law in the symbolic form:

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(A) = B\begin{pmatrix} \cdot \\ [m'] \\ \cdot \end{pmatrix}(A) \overset{\langle W\rangle}{\underset{\langle W\rangle}{\quad}} B\begin{pmatrix} \cdot \\ [m''] \\ \cdot \end{pmatrix}(A). \tag{2.15}$$

In this symbolic expression the $\langle W\rangle$ represents a Wigner coupling, the dots represent Gel'fand patterns which are summed over, and the A represents the $n\times n$ boson matrix given in Eq. (2.2).

The significant feature in this coupling approach is that the "group element" (represented by A above) is quite arbitrary, but fixed; it is the irrep labels themselves which change. By contrast, the Gel'fand–Graev approach generalizes the admissible group elements, but fixes the irrep labels. Both procedures employ an embedding of $U(n-1)$ in $U(n)$ to achieve a recursive structure. In the boson approach this step utilizes the factorization lemma (see Sec. IIC below), and the fact that special polynomials in $U(n)$ are, in fact, $U(n-1)$ structures in their own right. That is, for $m_{nn} = 0$,

$$B\begin{pmatrix} (m') \\ [\max]_{n-1} \\ [m]_n \\ [\max]_{n-1} \\ (m) \end{pmatrix}(A_n) = B\begin{pmatrix} (m') \\ [m_{1n}\cdots m_{n-1n}] \\ (m) \end{pmatrix}(A_{n-1}). \tag{2.16}$$

(Note that singular manifolds of lower dimension[22] do not enter in the boson approach.)

It is one of the main purposes of the present paper to detail the construction of these boson polynomials, using boson operator coupling techniques. Such a procedure offers many advantages:

(a) The coefficients occurring in the polynomials are always identifiable in terms of Wigner coefficients—hence the structural properties of the polynomials are always clearly in evidence;

(b) the polynomial nature of the functions is also clearly in evidence (this contrasts with the Gel'fand–Graev approach where reciprocal powers of $\det A_{n-1}$ occur); and

(c) a variety of distinct expressions for the polynomials may be derived depending on how one chooses to effect the couplings.

We illustrate this latter point by constructing three different forms for the boson polynomials: a recursive $U_n : U_{n-1}$ form [Eq. (2.28)]; a completely explicit form built up from the totally symmetric monomials (Eq. (2.29)]; and a form similar in structure to the Cartan product form for $U(n)$ irreps [Eq. (2.47)].

B. The totally symmetric boson polynomials B ([p 0̇])

Let us begin the development of the general boson polynomial by considering first a special class: the totally symmetric polynomials

$$B\begin{pmatrix} (m') \\ [p\ \dot{0}] \\ (m) \end{pmatrix}$$

corresponding to the irrep labels [p 0̇]. This class serves as the "elementary constituents" for constructing one form of the general polynomial (see Sec. IID), but is also of interest in itself, in that it possesses an elegantly simple defining form, Eq. (2.19), and interesting discrete symmetry properties.

The boson polynomials corresponding to the irrep labels [p 0̇] are easily derived from the general product law[7]:

$$B\begin{pmatrix}(\mu')\\ [m]\\ (\mu)\end{pmatrix}(\tilde{X}AY) = \sum_{(m)(m')} B\begin{pmatrix}(\mu)\\ [m]\\ (m)\end{pmatrix}(X)B\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(A)B\begin{pmatrix}(\mu')\\ [m]\\ (m')\end{pmatrix}(Y). \tag{2.17}$$

Here the tilde denotes matrix transposition; X and Y are arbitrary complex matrices.

We now specialize the labels in Eq. (2.17): take $[m]$ to be $[p\ \dot{0}]$ and choose $(\mu) = (\mu') = (\max)$. Then the left-hand side of Eq. (2.17) takes the form:

$$B\begin{pmatrix}(\max)\\ [p\ \dot{0}]\\ (\max)\end{pmatrix}(\tilde{X}AY) = \left(\sum_{i,j=1}^{n} x_i^1 a_i^j y_j^1\right)^p. \tag{2.18}$$

Expanding this expression by using the multinomial theorem, one easily identifies the terms occurring in the right-hand side of Eq. (2.17) (particularized to the special case under consideration). The result is expressed most elegantly in the following form:

$$B\begin{pmatrix}(m')\\ [p\ \dot{0}]\\ (m)\end{pmatrix}(A) = \left[\prod_{i=1}^{n}(W_i)!(W_i')!\right]^{1/2} \times \sum_{\boxed{\alpha}} \prod_{i,j=1}^{n} (a_i^j)^{\alpha_i^j}/(\alpha_i^j)!, \tag{2.19}$$

where $[W]$ and $[W']$ denote the *weights* of the lower and upper Gel'fand patterns, respectively, and $\boxed{\alpha}$ denotes the following square matrix of nonnegative integers with constraints on the sums of the entries in the rows and columns:

$$\boxed{\alpha} = \begin{array}{c}\left|\begin{matrix}\alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^n\\ \alpha_2^1 & \alpha_2^2 & \cdots & \alpha_2^n\\ \vdots & & & \\ \alpha_n^1 & \alpha_n^2 & \cdots & \alpha_n^n\end{matrix}\right| \begin{matrix}W_1\\ W_2\\ \\ W_n\end{matrix}\\ W_1' \quad W_2' \quad \cdots \quad W_n'\end{array}. \tag{2.20}$$

The symbols W_i (W_j') written to the right of row i (below column j) designate that the entries in row i (column j) are constrained to add to W_i (W_j'). The sum over $\boxed{\alpha}$ in Eq. (2.19) is to be taken over all nonnegative integers α_i^j (for $i,j = 1, 2, \ldots, n$) which satisfy these constraints. [The letter W is used to accord with the fact that W_i and W_j' are the *weights* of the patterns (m) and (m'), respectively.]

It is clear that the summation over $\boxed{\alpha}$ in Eq. (2.19) involves redundant summation parameters, but there are compensations. For example, one sees that Eq. (2.19) makes quite transparent some of the discrete symmetry properties of these totally symmetric polynomials.

These symmetry properties can be described succinctly in the following manner. Let $\mathcal{P}$ denote a permutation $1 \to i_1, 2 \to i_2, \ldots, n \to i_n$, where $(i_1 i_2 \cdots i_n)$ is a rearrangement of $(12 \cdots n)$. Define the action of $\mathcal{P}$ on the matrix A from the right by the rule

$$A\mathcal{P} = A' = [A^{i_1} A^{i_2} \cdots A^{i_n}], \tag{2.21}$$

where A^i denotes the ith column of A. Similarly, define the action of $\mathcal{P}$ on the matrix A from the left by the rule

$$\mathcal{P}A = A'' = \begin{bmatrix} A_{i_1} \\ A_{i_2} \\ \vdots \\ A_{i_n}\end{bmatrix}, \tag{2.22}$$

where A_i denotes the ith row of A.

The permutation $\mathcal{P}$ induces an action on the weights given by the rule: $[\mathcal{P}W] = [W_{i_1} W_{i_2} \cdots W_{i_n}]$.

Let us note now that for the irrep $[p\ \dot{0}]$ the Gel'fand pattern (m) is uniquely determined by the weight W. Accordingly, we may just as well denote these symmetric polynomials by the notation:

$$D\begin{pmatrix}[W']\\ [p\ \dot{0}]\\ [W]\end{pmatrix}(A) = B\begin{pmatrix}[m']\\ [p\ \dot{0}]\\ (m)\end{pmatrix}(A). \tag{2.23}$$

The symmetry properties which are now obvious from Eq. (2.19) may be expressed in the following two results:

$$D\begin{pmatrix}[W']\\ [p\ \dot{0}]\\ [W]\end{pmatrix}(A\mathcal{P}) = D\begin{pmatrix}[\mathcal{P}W']\\ [p\ \dot{0}]\\ [W]\end{pmatrix}(A), \tag{2.24a}$$

$$D\begin{pmatrix}[W']\\ [p\ \dot{0}]\\ [W]\end{pmatrix}(\mathcal{P}A) = D\begin{pmatrix}[W']\\ [p\ \dot{0}]\\ [\mathcal{P}W]\end{pmatrix}(A), \tag{2.24b}$$

for each $\mathcal{P} \in S_n$. To these symmetries we may also add the general transpositional symmetry which now assumes the form:

$$D\begin{pmatrix}[W']\\ [p\ \dot{0}]\\ [W]\end{pmatrix}(\tilde{A}) = D\begin{pmatrix}[W]\\ [p\ \dot{0}]\\ [W']\end{pmatrix}(A). \tag{2.24c}$$

We would like to emphasize once again the general validity of these results, Eqs. (2.19)–(2.24); *in particular A may be replaced by an arbitrary complex matrix Z*. Accordingly, these results apply to irreps $[p\ \dot{0}]$ of both $U(n)$ and $GL(n, C)$.

C. A recursive definition of the general polynomial

Using the explicit form of the symmetric polynomials derived in Sec. IIB, above, it is quite easy to obtain a recursive definition for the general polynomial, using the coupling law approach. In particular, the following two special couplings lead almost immediately to the desired form:

Coupling form (a):

$$\sum_{(\mu)(b)} \left\langle \begin{pmatrix}[m]\\ (m)\end{pmatrix} \middle| \begin{pmatrix}(\gamma)\\ [b\ \dot{0}]\\ (b)\end{pmatrix} \middle| \begin{pmatrix}[m']_{n-1} & 0\\ & (\mu)\end{pmatrix} \right\rangle \times B\begin{pmatrix}(0)\\ [b\ \dot{0}]\\ (b)\end{pmatrix}(A)\, B\begin{pmatrix}(m')\\ [m']_{n-1} \quad 0\\ (\mu)\end{pmatrix}(A)$$

$$= \left\langle \begin{pmatrix} [m] \\ (m') \end{pmatrix} \middle| \begin{pmatrix} (\gamma) \\ [b\ \dot{0}] \\ (0) \end{pmatrix} \middle| \begin{pmatrix} [m']_{n-1}\ 0 \\ (m') \end{pmatrix} \right\rangle \times B \begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix} (A); \tag{2.25a}$$

Coupling form (b):

$$\sum_{(\mu')(c)} \left\langle \begin{pmatrix} [m']_{n-1}\ 0 \\ [\max]_{n-1}\ 0 \\ (m') \end{pmatrix} \middle| \begin{pmatrix} (\gamma') \\ [\max]_{n-1} \\ [c\ \dot{0}]_n \\ [\max]_{n-1} \\ (c) \end{pmatrix} \middle| \begin{pmatrix} [\mu]_{n-1}\ 0 \\ [\max]_{n-1} \\ (\mu') \end{pmatrix} \right\rangle \times B \begin{pmatrix} (c) \\ [\max]_{n-1} \\ [c\ \dot{0}]_n \\ (0) \end{pmatrix} (A)\, B \begin{pmatrix} (\mu') \\ [\max]_{n-1} \\ [\mu]_{n-1}\ 0 \\ [\max]_{n-1} \\ (\mu) \end{pmatrix} (A)$$

$$= \left\langle \begin{pmatrix} [m']_{n-1}\ 0 \\ (\mu) \end{pmatrix} \middle| \begin{pmatrix} (\gamma') \\ [\max]_{n-1} \\ [c\ \dot{0}]_n \\ (0) \end{pmatrix} \middle| \begin{pmatrix} [\mu]_{n-1}\ 0 \\ [\max]_{n-1} \\ (\mu) \end{pmatrix} \right\rangle \times B \begin{pmatrix} (m') \\ [m']_{n-1}\ 0 \\ (\mu) \end{pmatrix} (A). \tag{2.25b}$$

The existence of these Wigner coefficient couplings used above follows from the significance of the boson polynomials, $B(\cdot)$, as the carrier space of the groups $U(n)*U(n)$. Several features in these two equations, (2.25a, b), should be noted. Observe that it is the irrep labels $[m]$ and the Gel'fand patterns (m) and (m') which are *specified* (fixed) in Eq. (2.25a). All other labels appearing in this equation are either dummy summation patterns or are uniquely determined from $[m]$, (m), and (m').[23]

{Thus, b is determined to be

$$b = \sum_{i=1}^{n} m_{in} - \sum_{i=1}^{n-1} m'_{i,n-1},$$

and the operator pattern (γ) is uniquely determined by the Δ pattern

$$\Delta(\gamma) = [m] - [[m']_{n-1}\ 0].$$

Also the Gel'fand pattern (b) is uniquely determined in terms of (μ) and (m) by the weight

$$W(b) = W(m) - W(\mu),$$

so that the summation may, in fact, be considered to be only over the patterns (μ). Similar conventions apply to Eq. (2.25b) where in addition, the notation $[\max]_{n-1}$ has been introduced to designate that the labels appearing in row $n-1$ of a particular pattern are chosen to be maximal.}

The idea now is to substitute Eq. (2.25b) into Eq. (2.25a) after evaluating the Wigner coefficients appearing in the right-hand side of these equations. Observing that the couplings in the left-hand sides of these equations are in each case that of "kinematically independent" boson polynomials, it follows immediately that the two Wigner coefficients in Eqs. (2.25a, b) have the following values, respectively:

$$\left[\mathfrak{N}([b\ \dot{0}])\,\mathfrak{N}([m']_{n-1}\ 0)/\mathfrak{N}([m])\right]^{1/2}, \tag{2.26a}$$

$$\left[\mathfrak{N}([c\ \dot{0}])\,\mathfrak{N}([\mu]_{n-1}\ 0)/\mathfrak{N}([m']_{n-1}\ 0)\right]^{1/2}. \tag{2.26b}$$

Here $\mathfrak{N}([m])$ refers to the measure of the highest weight tableau, and is just the norm of the boson polynomial specified by the irrep labels $[m]$ [Cf. Ref. 16]. We also note that the $U(n)$ Wigner coefficient appearing in the left-hand side of Eq. (2.25b) is equal to the $U(n-1)$ Wigner coefficient obtained by deleting row n throughout the $U(n)$ Wigner coefficient (and letting the two equal rows $n-1$ of the Wigner operator merge). This identity of a special $U(n)$ Wigner coefficient to a $U(n-1)$ Wigner coefficient is an easy consequence of the factorization lemma and the fact that

$$B \begin{pmatrix} (m') \\ [\max]_{n-1} \\ [m]_n \\ [\max]_{n-1} \\ (m) \end{pmatrix} (A_n) = B \begin{pmatrix} (m') \\ [m_{1n} \cdots m_{n-1n}] \\ (m) \end{pmatrix} (A_{n-1}) \tag{2.27}$$

for $m_{nn} = 0$, where A_n now designates the $n \times n$ array (a_i^j), $i,j = 1, 2, \ldots, n$, and A_{n-1} designates the $n-1 \times n-1$ array (a_i^j) $i,j = 1, 2 \ldots, n-1$.

Putting together all of these results, we obtain the following explicit recursive definition of the general boson polynomial[11]:

$$[\mathfrak{N}([m])]^{-1/2} B \begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix} (A_n) = \sum_{[\mu]_{n-1}} \sum_{\substack{(b)(\mu) \\ (b')(\mu')}} \left\langle \begin{pmatrix} [m] \\ (m) \end{pmatrix} \middle| \begin{pmatrix} (\gamma) \\ [b\ \dot{0}] \\ (b) \end{pmatrix} \middle| \begin{pmatrix} [m']_{n-1}\ 0 \\ (\mu) \end{pmatrix} \right\rangle \times \left\langle \begin{pmatrix} [m']_{n-1} \\ (m') \end{pmatrix} \middle| \begin{pmatrix} (\gamma') \\ [b'\ \dot{0}] \\ (b') \end{pmatrix} \middle| \begin{pmatrix} [\mu]_{n-1} \\ (\mu') \end{pmatrix} \right\rangle \times [b!\,(b')!\,\mathfrak{N}([\mu]_{n-1})]^{-1/2} \times B \begin{pmatrix} (0) \\ [b\ \dot{0}] \\ (b) \end{pmatrix} (A_n)\, B \begin{pmatrix} (b') \\ [\max] \\ [b'\ \dot{0}] \\ (0) \end{pmatrix} (A_n)\, B \begin{pmatrix} (\mu') \\ [\mu]_{n-1} \\ (\mu) \end{pmatrix} (A_{n-1}). \tag{2.28}$$

D. The general boson polynomial in an explicity totally symmetric expansion

The object of this section is to combine the results of Secs. IIB and IIC into an explicit formulation of the general boson polynomial as a sum over $\boxed{\alpha}$. We shall demonstrate the result:

$$B\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(A) = \mathfrak{N}^{1/2}([m]) \sum_{\boxed{\alpha}} C\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(\alpha) \times \prod_{i,j=1}^{n} (a_i^j)^{\alpha_i^j}/[(\alpha_i^j)!]^{1/2}. \quad (2.29)$$

Clearly a form of this type must be valid, and it could be obtained by the direct iteration of Eq. (2.28). However, the simplest way to obtain the coefficients

$$C\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(\alpha) \quad (2.30)$$

is by appeal to the factorization lemma. Directly from Eq. (2.29), we obtain

$$C\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(\alpha) = \left\langle \begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix} \middle| \prod_{i=1}^{n} [(W_i)!]^{-1/2} B\begin{pmatrix}(\alpha_i)\\[W_i\ \dot{0}]\\ i\end{pmatrix}(A) \middle| \begin{pmatrix}(0)\\[\dot{0}]\\(0)\end{pmatrix} \right\rangle, \quad (2.31)$$

where i denotes the Gel'fand pattern having weight $[0 \dots 0 W_i 0 \dots 0]$ (W_i in position i), and (α_i) denotes the Gel'fand pattern (inverted)

$$\begin{pmatrix}[W_i\ \dot{0}]\\(\alpha_i)\end{pmatrix} = \begin{pmatrix} \alpha_i^1 + \alpha_i^2 + \cdots + \alpha_i^n & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ & \alpha_i^1 + \alpha_i^2 & & 0 \\ & & \alpha_i^1 & \end{pmatrix}. \quad (2.32)$$

In Eq. (2.31) we now use the factorization lemma in the form

$$B\begin{pmatrix}(\alpha_i)\\[W_i\ \dot{0}]\\ i\end{pmatrix}(A) = \mathfrak{N}^{1/2} \left(\sum_{\Gamma} \left\langle \begin{matrix}(\Gamma)\\[W_i\ \dot{0}]\\ i\end{matrix} \right\rangle_{l} \left\langle \begin{matrix}(\Gamma)\\[W_i\ \dot{0}]\\(\alpha_i)\end{matrix} \right\rangle_{u} \right) \mathfrak{N}^{-1/2}, \quad (2.33)$$

arranging the factors in the order $\langle W_n\ \dot{0}\rangle \cdots \langle W_2\ \dot{0}\rangle \langle W_1\ \dot{0}\rangle$. With this ordering only one term from each of the sums appears in the matrix element between the initial and final states appearing in Eq. (2.31), and hence we find:

$$C\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(\alpha) = [\mathfrak{N}([m])/\prod_{i=1}^{n}(W_i)!]^{1/2} \times \left\langle \begin{pmatrix}(m)\\(m')\end{pmatrix} \middle| \left\langle \begin{matrix}(\Gamma_n)\\[W_n\ \dot{0}]\\(\alpha_n)\end{matrix} \right\rangle \cdots \left\langle \begin{matrix}(\Gamma_2)\\[W_2\ \dot{0}]\\(\alpha_2)\end{matrix} \right\rangle \left\langle \begin{matrix}(\Gamma_1)\\[W_1\ \dot{0}]\\(\alpha_1)\end{matrix} \right\rangle \middle| (0) \right\rangle \times \left\langle \begin{pmatrix}(m)\\(m')\end{pmatrix} \middle| \left\langle \begin{matrix}(\Gamma_n)\\[W_n\ \dot{0}]\\ n\end{matrix} \right\rangle \cdots \left\langle \begin{matrix}(\Gamma_2)\\[W_2\ \dot{0}]\\ 2\end{matrix} \right\rangle \left\langle \begin{matrix}(\Gamma_1)\\[W_1\ \dot{0}]\\ 1\end{matrix} \right\rangle \middle| (0) \right\rangle, \quad (2.34)$$

where (Γ_k) is the operator pattern which is uniquely determined by the Δ pattern

$$[\Delta(\Gamma_k)] = [m_{1k} m_{2k} \cdots m_{kk}\ \dot{0}] - [m_{1k-1} m_{2k-1} \cdots m_{k-1k-1}\ \dot{0}]. \quad (2.35)$$

We now assert that

$$\left\langle \begin{pmatrix}[m]\\(m)\end{pmatrix} \middle| \left\langle \begin{matrix}(\Gamma_n)\\[W_n\ \dot{0}]\\ n\end{matrix} \right\rangle \cdots \left\langle \begin{matrix}(\Gamma_1)\\[W_1\ \dot{0}]\\ 1\end{matrix} \right\rangle \middle| (0) \right\rangle = \left[\prod_{i=1}^{n} (W_i)! \Big/ \mathfrak{N}([m]) \right]^{1/2}. \quad (2.36)$$

The proof is given in the paragraph following Eq. (2.41).

The final form of the coefficient appearing in Eq. (2.29) is thus given by

$$C\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(\alpha) = \left\langle \begin{pmatrix}[m]\\(m')\end{pmatrix} \middle| \left\langle \begin{matrix}(\Gamma_n)\\[W_n\ \dot{0}]\\(\alpha_n)\end{matrix} \right\rangle \cdots \left\langle \begin{matrix}(\Gamma_2)\\[W_2\ \dot{0}]\\(\alpha_2)\end{matrix} \right\rangle \left\langle \begin{matrix}(\Gamma_1)\\[W_1\ \dot{0}]\\(\alpha_1)\end{matrix} \right\rangle \middle| (0) \right\rangle. \quad (2.37)$$

Since only the (known) matrix elements of totally symmetric Wigner operators[1,24,25] *enter into this determination* [Eq. 2.37], *we assert that the coefficient* $C(\alpha)$ *of* (2.30) *is thereby known explicitly.*

Let us make the following definition pertaining to these coefficients

$$C\begin{pmatrix}(m')\\[m]\\(m)\end{pmatrix}(\alpha) \equiv 0, \quad (2.38)$$

unless

(a) $[m]$ is a set of irrep labels of $U(n)$ for $m_{nn} \geq 0$, (m) and (m') are Gel'fand patterns;

(b) $\boxed{\alpha}$ is a square array of nonnegative integers (α_i^j), $i,j = 1, 2, \dots n$;

(c) the following relations hold between the Gel'fand patterns and the array (α):

$$\sum_{j=1}^{n} \alpha_i^j = W_i = \sum_{j=1}^{i} m_{ji} - \sum_{j=1}^{i-1} m_{j,i-1}, \; i = 1, 2, \dots, n, \quad (2.39a)$$

$$\sum_{i=1}^{n} \alpha_i^j = W_j' = \sum_{i=1}^{j} m'_{ij} - \sum_{i=1}^{j-1} m'_{i,j-1}, j = 1, 2, \dots, n. \quad (2.39b)$$

Thus, if $[m]$, (m), and (m') are specified, then the coefficient is zero unless $\boxed{\alpha}$ satisfies the constraints indicated in Eq. (2.20). On the other hand, if $\boxed{\alpha}$ is specified, then the coefficient is zero unless $[m]$ is a set of integers satisfying $m_{1n} \geq m_{2n} \geq \dots \geq m_{nn} \geq 0$,

$$\sum_{i=1}^{n} m_{in} = \sum_{i,j=1}^{n} \alpha_i^j,$$

and (m) and (m') are Gel'fand patterns having weights given by

$$W_i = \sum_{j=1}^{n} \alpha_i^j (i = 1, 2, \dots, n), \text{ and } W_i' = \sum_{j=1}^{n} \alpha_j^i,$$

respectively.

Using now the fact[3,21] that the set of boson polynomials, Eq. (2.1), corresponding to all partitions $[m]$ of an integer k also span the space of polynomials homogeneous of degree k in the $a_i^\dagger (i,j = 1,2\ldots,n)$, we now deduce directly from Eq. (2.29) the following orthogonality relations:

$$\sum_{\boxed{\alpha}} C\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(\alpha)\, C\begin{pmatrix}(\mu')\\ [\mu]\\ (\mu)\end{pmatrix}(\alpha) = \delta_{[m][\mu]}\delta_{(m)(\mu)}\delta_{(m')(\mu')}, \tag{2.40a}$$

$$\sum_{[m]}\sum_{(m)(m')} C\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(\alpha)\, C\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(\beta) = \delta_{(\alpha)(\beta)}. \tag{2.40b}$$

We also note that these coefficients satisfy the following symmetry relation in consequence of the transpositional symmetry of the boson polynomials[7]:

$$C\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(\tilde{\alpha}) = C\begin{pmatrix}(m)\\ [m]\\ (m')\end{pmatrix}(\alpha). \tag{2.41}$$

The proof of Eq. (2.36) may now be given by the following argument: one can verify[26] directly that the coefficient given by Eq. (2.37) satisfies Eq. (2.40a). Using this result and Eq. (2.34), we now see that relation (2.36) must be correct, except possibly for a sign. [That the sign is also correct may be ascertained by direct expansion of the left-hand side (the intermediate states are uniquely determined, so that no summation is involved) using then the phase conventions of our earlier work to verify that the phase of each matrix element in the product is + 1.]

Several special cases of Eq. (2.37) are worthy of particular notice.

(1) For $n = 2$, the coefficients may be expressed in the standard notation of angular momentum theory:

$$C\begin{pmatrix} j_1+j_2+m & & \\ j_1+j_2+j & j_1+j_2-j & \\ & 2j_1 & \end{pmatrix}\begin{pmatrix} j_1+m_1 & j_1-m_1 \\ j_2+m_2 & j_2-m_2\end{pmatrix} = C\begin{matrix} j_1\ j_2\ j \\ m_1\ m_2\ m\end{matrix}. \tag{2.42}$$

The transpositional symmetry, Eq. (2.41), of these coefficients may then be used to obtain all the known symmetries[27] of the $SU(2)$ Wigner coefficients entirely within the framework of $SU(2)$ as pointed by the Bincer.[28] Let us also remark that the symmetries of Eqs. (2.24a–c) and the symmetry of the representation matrices under complex conjugation already suffice to deduce the full seventy-two symmetries of the $3j$ symbols directly from Wigner's original formulation.[29]

(2) For $n = 3$, we consider the special coefficient having $m_{13} = m_{23} = m_{33} = k = j_1 + j_2 + j$. This coefficient may also be expressed in the jm notation of angular momentum theory:

$$C\begin{pmatrix} k & & & \\ k & k & & \\ k & k & k & \\ k & k & & \\ k & & & \end{pmatrix}\begin{pmatrix} j_1+m_1 & j_1-m_1 & -j_1+j_2+j \\ j_2+m_2 & j_2-m_2 & j_1-j_2+j \\ j-m & j+m & j_1+j_2-j\end{pmatrix}$$

$$= (-1)^{j_1-j_2+m}\left[\frac{2(k!)}{(2j+1)(k+2)!}\right]^{1/2} C\begin{matrix} j_1\ j_2\ j \\ m_1\ m_2\ m\end{matrix}. \tag{2.43}$$

The derivation of this result follows straightforwardly from Eq. (2.37) upon using the subgroup reduction law and the expressions for the $U(3):U(2)$ reduced matrix elements given in Ref. 24. Since the left-hand side of Eq. (2.29) becomes $(a_{123}^{123})^k$ for the special case under consideration [the measure factor is given by $\mathfrak{M}(kkk) = k!(k+1)!(k+2)!/2$], we see that Eq. (2.29) becomes just a boson interpretation of Regge's result[27] from which the seventy-two symmetries of the $3j$-symbol are obvious.

Before closing this subsection, we would like to note the connection of one of our results, Eq. (2.29), with combinatorial mathematics: Suppose we denote by $M(W,W')$ the number of square matrices $\boxed{\alpha}$ having specified row sum W and specified column sum W' [See Eq. (2.20)], and by $K(m,W)$ the number of Gel'fand patterns having irrep labels $m = [m]$ and weight W. Then it is a direct consequence of the fact that the coefficients (2.30) are the elements of an *orthogonal matrix*, hence, a square matrix, that the following identity must hold:

$$M(W,W') = \sum_m K(m,W)K(m,W'), \tag{2.44}$$

where the sum is over all partitions m of the positive integer

$$N = \sum_{i=1}^n W_i = \sum_{i=1}^n W_i'.$$

(The fact that the coefficients (2.30) are the elements of an orthogonal matrix is, of course, itself a consequence of the group property $U(n^2) \supset U(n) \times U(n)$ *and* the reduction of the totally symmetric irrep $[N\ 0]$ of $U(n^2)$ into its $U(n) \times U(n)$ irreducible constituents as discussed in Refs. 3 and 21. Note further that if we sum Eq. (2.44) over all weights W and W' which add to N, then we obtain precisely the relations between dimension formulas noted in Refs. 3 and 21.)

It is a remarkable result that Eq. (2.44) has a purely combinatorial proof.[30] The relation is an immediate consequence of Knuth's algorithm[31] which establishes (by a direct construction) the existence of a one-to-one correspondence between the set of matrices $\boxed{\alpha}$ having row sum W and column sum W' and the set of ordered pairs of column-strict plane partitions[32] of the same "shape" and "type" W and W', respectively.

A column-strict plane partition of shape m is just a Young tableau which has m_{1n} boxes in the first row, m_{2n} boxes in the second row, $\ldots, m_{nn}$ boxes in the nth row, which has been "filled in" in the usual way[16] with the integers $1,2,\ldots,n$. *Accordingly, the set of column-strict plane partitions of shape m is in one-to-one correspondence with the set of Gel'fand patterns having irrep labels* $m = [m]$, *i.e., with the set of triangular patterns*

$$\left\{\begin{pmatrix}[m]\\ (m)\end{pmatrix}: \begin{array}{l}(m) \text{ is a lexical pattern, i.e.,}\\ \text{satisfies the betweenness conditions}\end{array}\right\}.$$

The "type" of a column-strict plane partition simply refers to the fact that W_1 1's, W_2 2's, $\ldots, W_n$ n's appear in the Young tableau of shape m, that is, W is just the weight of the Gel'fand pattern. *Accordingly, there is also a one-to-one correspondence between the set of column-strict plane partitions of shape m and type W*

and the set of Gel'fand patterns having irrep labels m and weight W.

Knuth's algorithm may now be restated in terms of the language of Gel'fand patterns: *There exists a one-to-one correspondence between the set of matrices* $\boxed{\alpha}$ *having row sum W and column sum W′ and the set of double Gel'fand patterns*

$$\left\{ \begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix} : \begin{array}{l} [m] \text{ is a partition of } N = \sum_{i=1}^{n} W_i = \sum_{i=1}^{n} W'_i, \\ (m) \text{ has weight } W, \\ (m') \text{ has weight } W'. \end{array} \right\}. \tag{2.45}$$

This result, of course, implies immediately that the matrix C whose rows are enumerated by the matrices $\boxed{\alpha}$ of row sum W and column sum W' and whose columns are enumerated by the double Gel'fand patterns of type (2.45) is a square matrix. Indeed, the Knuth algorithm may now be used to give an alternative proof of the reduction relation[3,21] between the totally symmetric irreps of $U(n^2)$ and the irreps of $U(n) \times U(n)$. (These relations, of course, still remain valid for the general linear group.)

E. The general boson polynomial in a 'Cartan' decomposition

While the forms, Eqs. (2.28) and (2.29), are reasonably explicit (to the extent that one is able to write out the coefficients), neither of these forms reduces directly to

$$B\begin{pmatrix} (\max) \\ [m] \\ (\max) \end{pmatrix}(A) = \prod_{k=1}^{n} (a_{12\cdots k}^{12\cdots k})^{m_{kn}-m_{k+1,n}}, \tag{2.46}$$

but rather to an expanded version of the right-hand side. One recognizes that Eq. (2.46) expresses the highest weight vector in $[m]$ as a product of highest weight vectors of its elementary constituents: $[\dot{k}_i\ \dot{0}_{n-i}]$. Such an expression is familiar from the work of Cartan, and accordingly we seek to determine an explicit form for the general boson polynomial in the "Cartan" form.

A form which does reduce immediately to Eq. (2.46) is given by

$$\begin{aligned} B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(A) = \sum_{\substack{\gamma_1\gamma_2\cdots\gamma_n \\ \gamma'_1\gamma'_2\cdots\gamma'_n}} & C\begin{pmatrix} [m] \\ (m) \end{pmatrix}(\gamma_1\gamma_2\cdots\gamma_n) \\ & \times C\begin{pmatrix} [m] \\ (m') \end{pmatrix}(\gamma'_1\gamma'_2\cdots\gamma'_n)\, B\begin{pmatrix} (\gamma'_1) \\ [k_1\ \dot{0}] \\ (\gamma_1) \end{pmatrix}(A)\, B\begin{pmatrix} (\gamma'_2) \\ [k_2k_2\ \dot{0}] \\ (\gamma_2) \end{pmatrix}(A) \\ & \times\cdots\times B\begin{pmatrix} (\gamma'_n) \\ [k_n\cdots k_n] \\ (\gamma_n) \end{pmatrix}(A), \end{aligned} \tag{2.47a}$$

where

$$C\begin{pmatrix} [m] \\ (m) \end{pmatrix}(\gamma_1\gamma_2\cdots\gamma_n) = \left\langle \begin{pmatrix} [m] \\ (m) \end{pmatrix} \middle| \begin{pmatrix} (\max) \\ [k_1\ \dot{0}] \\ (\gamma_1) \end{pmatrix} \begin{pmatrix} (\max) \\ [k_2k_2\ \dot{0}] \\ (\gamma_2) \end{pmatrix} \cdots \begin{pmatrix} (\max) \\ [k_n\cdots k_n] \\ (\gamma_n) \end{pmatrix} \middle| \begin{pmatrix} [\dot{0}] \\ (0) \end{pmatrix} \right\rangle \tag{2.47b}$$

$$k_i = m_{in} - m_{i+1,n},\ i = 1, 2, \cdots, n. \tag{2.47c}$$

[This form is established directly from the coupling law, Eq. (2.15). For symmetry of form, a sum over γ_n and γ'_n is indicated in Eq. (2.47a), although these two Gel'fand patterns are, in fact, both maximal.]

The following two results are useful for giving Eq. (2.47a) a more explicit form:

$$B\begin{pmatrix} (\max) \\ [kk\cdots k\ \dot{0}_{n-j}] \\ (\max) \end{pmatrix}(A) = (a_{12}^{12\cdots j})^k, \tag{2.48}$$

$$B\begin{pmatrix} (\tilde{m}') \\ [\dot{k}\ 0] \\ (\tilde{m}) \end{pmatrix}(A) = (-1)^{\phi(m)-\phi(m')}\, B\begin{pmatrix} (m') \\ [k\ \dot{0}] \\ (m) \end{pmatrix}(C), \tag{2.49}$$

where $\phi(m)$ denotes the sum of all the entries in the Gel'fand pattern

$$\begin{pmatrix} [k\ \dot{0}] \\ (m) \end{pmatrix};$$

C denotes the cofactor matrix of A, i.e., the element in row i and column j of C is given by $c_i^j = (-1)^{i-j} \times a_{12\cdots\hat{j}\cdots n}^{12\cdots\hat{i}\cdots n}$, and, $\tilde{m}_{ij} = m_{1n} - m_{j-i+1,j}$.

(The notation $12\cdots\hat{i}\cdots n$ designates that i is missing from the string of integers.)

For $n = 3$, Eqs. (2.47) become fully explicit, i.e., all quantities entering into these equations are completely known. For arbitrary n all quantities entering into Eqs. (2.47) are, of course, known in principle (in the sense that definite calculational procedures exist for determining the quantities).

F. Relationship to the Gel'fand-Graev results

As we have discussed in Sec. IIA, Gel'fand and Graev determined $GL(n, C)$ irreps by the finding the special functions appropriate to their factorization of the group element: $g_{n+1} = \lambda z g_n \zeta$. It is not difficult now to determine the boson polynomials appropriate to these specialized group elements used by Gel'fand and Graev.

The first particular result follows most easily from Eq. (2.29):

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}\begin{pmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & z_n \end{pmatrix} = \delta_{(m)(m')} \prod_{i=1}^{n} (z_i)^{W_i}. \tag{2.50}$$

The second particular result follows most easily from Eq. (2.28): Define the matrix $\mathfrak{z}$ by

$$\mathfrak{z} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & \\ \dot{0} & 0 & \cdots & 1 & 0 \\ z_1 & z_2 & \cdots & z_{n-1} & 1 \end{pmatrix}. \tag{2.51}$$

Then

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(\mathfrak{z}) = B\begin{pmatrix} (m) \\ [m] \\ (m') \end{pmatrix}(\tilde{\mathfrak{z}}) = \left[\frac{\mathfrak{M}([m]_n)}{b!\,\mathfrak{M}([m]_{n-1})} \right]^{1/2}$$

$$\times \left\langle \begin{pmatrix} [m]_n \\ [m]_{n-1} \end{pmatrix} \middle| \begin{bmatrix} (\gamma) \\ [b\ \ \dot{0}]_n \\ (0) \end{bmatrix} \middle| \begin{pmatrix} [m']_{n-1} & 0 \\ [m]_{n-1} & \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [m']_{n-1} \\ (m') \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma') \\ [b'\ \ \dot{0}]_{n-1} \\ (b') \end{matrix} \right\rangle \middle| \begin{pmatrix} [m]_{n-1} \\ (m) \end{pmatrix} \right\rangle$$

$$\times \prod_{i=1}^{n-1} (z_i)^{W_i' - W_i} \Big/ [(W_i' - W_i)!]^{1/2}. \tag{2.52}$$

Specializing Eq. (2.52) still further by setting $z_1 = z_2 = \cdots = z_{n-2} = 0, z_{n-1} = t$, and noting that $\mathcal{Z}$ becomes $I + te_{n,n-1}$, where e_{ij} denotes a matrix unit, we obtain the following result:

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(I + te_{n,n-1}) = B\begin{pmatrix} (m) \\ [m] \\ (m') \end{pmatrix}(I + te_{n-1,n})$$

$$= \delta_{(m)_{n-2}(m')_{n-2}} \left[\frac{\mathfrak{M}([m]_n}{b!(b')!\mathfrak{M}([m]_{n-1})} \right]^{1/2}$$

$$\times \left\langle \begin{pmatrix} [m]_n \\ [m]_{n-1} \end{pmatrix} \middle| \begin{bmatrix} (\gamma) \\ [\dot{b}\ \ \dot{0}]_n \\ (0) \end{bmatrix} \middle| \begin{pmatrix} [m']_{n-1} & 0 \\ [m]_{n-1} & \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [m']_{n-1} \\ [m]_{n-2} \end{pmatrix} \middle| \begin{bmatrix} (\gamma') \\ [b'\ \ \dot{0}]_{n-1} \\ (0) \end{bmatrix} \middle| \begin{pmatrix} [m]_{n-1} \\ [m]_{n-2} \end{pmatrix} \right\rangle t^{b'}. \tag{2.53}$$

The functions defined by Eq. (2.52) are the generalized beta functions introduced by Gel'fand and Graev.[12] Note that these important coefficients, which they studied, appear in the present work automatically, in terms of the matrix elements of a totally symmetric Wigner operator in $U(n-1)$ and a very simple $U(n): U(n-1)$ projective operator. [We would like to emphasize that the coefficient appearing in Eq. (2.53) was actually calculated explicitly by Gel'fand and Graev. Hence, the explicit formula for the matrix elements of the totally symmetric $U(n): U(n-1)$ projective operators having minimal lower operator pattern is already contained in their result.]

The results given above suffice to determine the relation of the Gel'fand–Graev results to the present boson operator approach. For the convenience of the reader who wishes to explore this relationship in more detail let us note that these authors[12] introduced a new basis, $|m\rangle'$, for the state vector $|m\rangle$, employed in our work. The explicit relationship between these two bases is

$$|(m)\rangle' = \left[\prod_{s=2}^{n} \frac{\prod_{i<j=1}^{s-1} (p_{i,s-1} - p_{j,s-1})}{\prod_{i\leq j=1}^{s-1} (p_{is} - p_{j,s-1} - 1)! \prod_{i<j=1}^{s} (p_{i,s-1} - p_{js})!} \right]^{1/2} |(m)\rangle. \tag{2.54}$$

This change of basis is advantageous in that it removes all square root factors from the matrices of the generators. We will discuss the significance of this change of basis for Wigner operators in a future paper.

G. Representations of the symmetric group as special cases of the boson polynomials

The representation theory of the unitary group, $U(n)$, necessarily contains as a special case the representation theory of the symmetric group; more importantly, there is, in fact, a deep interrelationship between these two groups, a subject which has been developed in detail in the monograph of Robinson.[33] Our purpose in the present section is rather more modest: we wish only to demonstrate how our general results for $U(n)$ may be specialized to yield the Young–Yamanouchi real-orthogonal representations of the symmetric group S_n. Our result appears to be of mathematical interest,[34] since it is both completely explicit and in a form hitherto unobtained.

Let us begin by considering the Cayley $n \times n$ permutation representation of S_n. For this one lets $\mathcal{P}$ denote a permutation by the rule:

$$\mathcal{P} = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}. \tag{2.55}$$

Then the correspondence

$$\mathcal{P} \to [e_{i_1} e_{i_2} \cdots e_{i_n}] \equiv I_{\mathcal{P}}, \tag{2.56}$$

—where e_i denotes a unit column vector with 1 in row i and zeroes elsewhere—is a representation of S_n by $n \times n$ matrices.

Since the general boson polynomial admits of an arbitrary interpretation of the argument A by an $n \times n$ indeterminate, it is a well-defined operation to replace A by $I_{\mathcal{P}}$, in Eq. (2.29). One obtains

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(I_{\mathcal{P}})$$

$$= [\mathfrak{M}([m])]^{1/2} \delta_{W_1' W_{i_1}} \delta_{W_2' W_{i_2}} \cdots \delta_{W_n' W_{i_n}} C\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(\alpha_{\mathcal{P}}), \tag{2.57}$$

where $\alpha_{\mathcal{P}}$ denotes the $n \times n$ numerical array

$$(\alpha_{\mathcal{P}}) = [W_{i_1} e_{i_1}, W_{i_2} e_{i_2}, \cdots, W_{i_n} e_{i_n}]. \tag{2.58}$$

Let us next specialize to representations having labels $[m]$ which are partitions of n, and at the same time *restrict the two Gel'fand patterns* (m) *and* (m') *such that the weights* $[W] = [W'] = [1]$. It follows at once from Eq. (2.57), that these special "boson polynomials" take the form:

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(I_{\mathcal{P}}) = [\mathfrak{M}([m])]^{1/2} C\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(I_{\mathcal{P}}). \tag{2.59}$$

It is useful to give a special notation to these objects; let us define

$$D^{[m]}_{(m),(m')}(\mathcal{P}) = B\begin{pmatrix}(m')\\ [m]\\ (m)\end{pmatrix}(I_{\mathcal{P}}), \tag{2.60}$$

and regard the Gel'fand patterns (m) and (m') as enumerating rows and columns, respectively, of the matrix $D^{[m]}(\mathcal{P})$. This matrix is indeed a unitary matrix representation of the group S_n, since for each $\mathcal{P} \in S_n$ we have the correspondence $\mathcal{P} \to D^{[m]}(\mathcal{P})$, which is a representation because of (a) the multiplication properties proven earlier for the $B(\cdots)$, and (b) the multiplication properties of the $\{I_{\mathcal{P}}\}$.

The general transpositional symmetry of the boson polynomials now implies that

$$D^{[m]}_{(m)(m')}(\mathcal{P}^{-1}) = D^{[m]}_{(m')(m)}(\mathcal{P}), \tag{2.61}$$

that is, these matrix representations are *orthogonal*, and hence *real* (since unitary by construction). That these representations are irreducible (for S_n) also follows from the construction, alternatively from the general results of Robinson[33] (see also Moshinsky[35]).

Consider now the specific form taken by the matrix elements of these irreps. From Eq. (2.37) we obtain

$$D^{[m]}_{(m),(m')}(\mathcal{P}) = [n!/\dim[m]]^{1/2} \times \left\langle \begin{pmatrix}[m]\\ (m)\end{pmatrix} \middle| \begin{pmatrix}(\gamma_n)\\ [1\ \dot{0}]\\ i_n\end{pmatrix} \cdots \begin{pmatrix}(\gamma_2)\\ [1\ \dot{0}]\\ i_2\end{pmatrix} \begin{pmatrix}(\gamma_1)\\ [1\ \dot{0}]\\ i_1\end{pmatrix} \middle| \begin{matrix}[\dot{0}]\\ (0)\end{matrix} \right\rangle, \tag{2.62}$$

where

(a) the operator pattern (γ_k) is uniquely determined by the associated Δ pattern

$$[\Delta(\gamma_k)] = [m'_{1k}m'_{2k}\cdots m'_{kk}0\cdots 0] - [m'_{1k-1}m'_{2k-1}\cdots m'_{k-1k-1}0\cdots 0]; \tag{2.63}$$

(b) the notation $([1_i\dot{0}])$ denotes the unique Gel'fand pattern having the weight $[0\ldots010\ldots0]$, with the 1 appearing in the ith place, and $\dim[m]$ denotes the dimension of irrep $[m]$ of S_n.

Note, in regard to item (a) above, that since

$$\sum_{i=1}^{n}\Delta_{in}(\gamma_k) = \sum_{i=1}^{k}m'_{ik} - \sum_{i=1}^{k-1}m'_{i,k-1} = 1$$

[that is, the Gel'fand pattern (m') has weight $[W] = [\dot{1}]]$, each pattern $\Delta(\gamma_k)$ has the form $[0\ldots010\ldots0]$.

The result obtained in Eq. (2.62) for the specific matrix elements of the S_n irrep $D^{[m]}(\mathcal{P})$ has some quite remarkable features:

(a) *It is a completely explicit general result.* Given the group element $\mathcal{P}$, the irrep label $[m]$, and the two patterns (m), (m'), one finds the associated number $D^{[m]}_{(m)(m')}(\mathcal{P})$ directly from the rules of the pattern calculus for the fundamental Wigner operators $\langle 1\ \dot{0}\rangle$.

The importance of this result lies in the fact that all three of the known irreps for the symmetric group [Young's normal form[36] (characterized by integral matrix elements), Young's semi-normal form[37] (characterized by rational matrix elements), and the Young–Yamanouchi real-orthogonal irreps[33,37-40]] are defined through recursive procedures.

(b) It is equivalent to the Young–Yamanouchi real-orthogonal irreps. (For $n \leq 4$ we have verified that the equivalence is, in fact, an identity; the explicit matrix elements involve phase choices, hence, the identity of the two results cannot be asserted *a priori*.)

(c) The Gel'fand patterns (γ_k) which enter into the construction of Eq. (2.62) are closely related to the Yamanouchi symbol, Y, which is itself associated with the Gel'fand pattern (m'). In order to make this point clear it is simplest to consider an example.

Take the Gel'fand pattern (m') belonging to the S_6 irrep [321] to be

$$(m') = \begin{pmatrix}3&2&1&0&0&0\\ &3&2&0&0&0\\ &&2&2&0&0\\ &&&2&1&0\\ &&&&1&1\\ &&&&&1\end{pmatrix}.$$

One may associate[5] to the pattern (m'), in a unique, 1 — 1 fashion, the lexical Young tableau:

1	3	5
2	4	
6		

.

Replacing every number, i, in this tableau by the value of $7 - i$, we obtain a new (inversely lexical) tableau:

6	4	2
5	3	
1		

.

The virtue of this new tableau is that the numbers $1,\ldots,6$ now give an *order* in which the boxes may be removed to yield the subgroup reduction: $S_6 \supset S_5 \supset \ldots \supset S_1 = E$. The Yamanouchi symbol, Y, is a sequence of numbers denoting the *row* from which these successive removals occur. In the example above one finds: $Y = (3,1,2,1,2,1)$. Thus to each Gel'fand pattern (m) belonging to weight $[\dot{1}]$, one may associate a Yamanouchi symbol; this was first pointed out by Moshinsky.[35]

We have already noted that for each $\Delta(\gamma_k)$ we have the form $[0\ldots010\ldots0]$; accordingly we may denote each $\Delta(\gamma_k)$ by an integer $j = 1\ldots n$, where j denotes the position of the 1 associated with $\Delta(\gamma_k)$. It is not difficult to see that this mapping: $(\gamma_k) \to \Delta(\gamma_k) \to j_k$ associates to the sequence of patterns $(\gamma_n)(\gamma_{n-1})\ldots(\gamma_1)$ a sequence of integers: $j_n j_{n-1}\ldots j_1$. *This sequence is precisely the Yamanouchi symbol* of (m'). Note that this occurrence of the Yamanouchi symbol is distinct from the correspondence[35]: $(m) \to Y$—which

is an equivalence of labels—since the $\{(\gamma_k)\} \to Y$ mapping expresses now a *structural property*. It is quite intriguing to find that the Yamanouchi symbol enters so directly into the structure of the matrix element; clearly the general result, Eq. (2.29), implies a corresponding generalization of these ideas.

(d) As Dr. de Vries has pointed out to us, the fact that the two (formally equivalent) labels (m) and (m') enter so unsymmetrically into the final formula, Eq. (2.62), is quite surprising and may indicate that other, more symmetric, formulas exist.

III. GENERAL STRUCTURAL FORMS OF THE CANONICAL WIGNER OPERATORS

In this section, we develop expressions for the canonical Wigner operators in terms of sums of monomials in the fundamental Wigner operators. Our principal reason for presenting such expressions is to point out the structural similarities, as well as dissimilarities, existing between the relation of "boson polynomials" to "fundamental bosons" and the relation of "canonical Wigner operators" to "fundamental Wigner operators."

It is quite remarkable that the totally symmetric Wigner operators may be expressed as a sum of monomials in the fundamental Wigner operators by using essentially the form (2.19) which expresses a totally symmetric boson polynomial in terms of the fundamental bosons $\{a_i^j\}$. This striking analogy in structures between two quite dissimilar sets of objects may be traced to two properties of the fundamental Wigner operators:

(a) Any two fundamental Wigner operators having a common operator pattern commute, i.e.,

$$\left[\left\langle \begin{matrix} \tau \\ [1\ \dot{0}] \\ i \end{matrix} \right\rangle, \left\langle \begin{matrix} \tau \\ [1\ \dot{0}] \\ j \end{matrix} \right\rangle\right] = 0, i,j = 1, 2, \ldots n. \tag{3.1}$$

(The nontrivial proof of this property is given in Appendix A).

(b) Under the following unitary mapping

$$\left\langle \begin{matrix} \tau \\ [1\ \dot{0}] \\ i \end{matrix} \right\rangle \to \sum_j \left\langle \begin{matrix} \tau \\ [1\ \dot{0}] \\ j \end{matrix} \right\rangle u_{ji}, \tag{3.2}$$

commuting sets are transformed into commuting sets.

Using these two properties, we can now prove the following relation:

$$g^{\Delta}\left\langle \begin{matrix} (\Gamma) \\ [p\ \dot{0}] \\ (M) \end{matrix} \right\rangle = \left[\prod_{i=1}^{n} (\Delta_i)!\,(W_i)!\right]^{1/2} \times \sum_{\boxed{\alpha}} \prod_{i=1}^{n} \left\langle \begin{matrix} n \\ [1\ \dot{0}] \\ i \end{matrix} \right\rangle^{\alpha_i^n} \cdots \prod_{i=1}^{n} \left\langle \begin{matrix} 2 \\ [1\ \dot{0}] \\ i \end{matrix} \right\rangle^{\alpha_i^2} \times \prod_{i=1}^{n} \left\langle \begin{matrix} 1 \\ [1\ \dot{0}] \\ i \end{matrix} \right\rangle^{\alpha_i^1} \Big/ \prod_{i,j=1}^{n} (\alpha_i^j)!, \tag{3.3}$$

where $g^{(\Delta)}$ is an invariant operator. Note that the right-hand side of this expression may be obtained from the right-hand side of Eq. (2.19) by first *ordering* the bosons and then making the mapping

$$a_i^j \to \left\langle \begin{matrix} j \\ [1\ \dot{0}] \\ i \end{matrix} \right\rangle.$$

Let us also note that the column and row sums in the matrix $\boxed{\alpha}$ are now specified to be $[\Delta]$ and $[W]$, respectively.

The proof of Eq. (3.3) is quite simple: One first notes that the operator on the right-hand side effects the shift $[m] \to [m] + [\Delta]$ on the irrep labels $[m]$ of an arbitrary state vector. One next notes that under the unitary transformation, Eq. (3.2), the right-hand side of Eq. (3.3) undergoes the transformation appropriate to an irreducible tensor operator labeled by irrep labels $[p\ \dot{0}]$. [This second property follows immediately from the corresponding property of the boson polynomials (Eq. 2.19) under the transformation $a_i^k \to \sum_{j=1}^{n} a_j^k u_{ji}$.] Together these two properties assure that the right-hand side of Eq. (3.3) can differ from the totally symmetric Wigner operator

$$\left\langle \begin{matrix} (\Gamma) \\ [p\ \dot{0}] \\ . \end{matrix} \right\rangle$$

by at most a multiplicative invariant operator.

The eigenvalues $g^{\Delta}([m] + [\Delta])$ of the invariant operator g^{Δ} {the number which appears in Eq. (3.3) when it acts on an arbitrary state vector having irrep labels $[m]$} can now be evaluated by setting $(m) = (0)$ and using the explicit matrix elements given in Ref. 24. The result of this calculation is

$$g^{\Delta}([m] + [\Delta]) = \times \frac{\left[\dfrac{p!}{\prod_i (\Delta_i)!}\right]^{1/2} D\begin{pmatrix} [\Delta] \\ [p\ \dot{0}] \end{pmatrix}([m])}{\left\{D\begin{pmatrix} [\Delta_n \Delta(n)] \\ [\Delta_n\ \dot{0}] \end{pmatrix} \cdots D\begin{pmatrix} [\Delta_2 \Delta(2)] \\ [\Delta_2\ \dot{0}] \end{pmatrix} D\begin{pmatrix} [\Delta_1 \Delta(1)] \\ [\Delta_1\ \dot{0}] \end{pmatrix}\right\}([m])}$$

$$= \left[\prod_{i>j=1}^{n} \binom{p_{in} - p_{jn} + \Delta_i}{\Delta_j} \Big/ \binom{p_{in} - p_{jn}}{\Delta_j}\right]^{1/2}, \tag{3.4}$$

where $\binom{x}{a}$ denotes a binomial coefficient (defined for arbitrary variable x and nonnegative integers a). [Note that $p_{in} - p_{jn} < 0$ in Eq. (3.4) for $i > j$ and one must use

$$\binom{-x}{a} = (-1)^a \binom{x + a - 1}{a}$$

before writing the factors in terms of factorials.]

Equation (3.3) is now a fully explicit relation which may, in principle, be used to calculate all the totally symmetric Wigner coefficients (since the fundamental Wigner coefficients are known). It is, however, not very practical to calculate explicit matrix elements in this manner. Nonetheless, from the viewpoint of structure, Eq. (3.3) is interesting in that it exhibits a quite general object in terms of its elementary constituents.

It is evident from the above derivation of Eq. (3.3) that it is not necessary to collect together the commuting sets of operators in the order $n \ldots 21$. A similar form must hold for each order $\tau = (\tau_1 \tau_2 \ldots \tau_n)$, where τ is any permutation belonging to S_n:

$$\mathcal{G}^{\Delta}_{\tau}\left\langle \begin{matrix}(\Gamma)\\ [p\ \ 0]\\ (M)\end{matrix}\right\rangle = \left[\prod_{i=1}^{n}(\Delta_i)!(W_i)!\right]^{1/2} \times \sum_{[\alpha]}\prod_{i=1}^{n}\left\langle \begin{matrix}\tau_1\\ [1\ \ 0]\\ i\end{matrix}\right\rangle^{\alpha_i^{\tau_1}} \prod_{i=1}^{n}\left\langle \begin{matrix}\tau_2\\ [1\ \ 0]\\ i\end{matrix}\right\rangle^{\alpha_i^{\tau_2}} \cdots \prod_{i=1}^{n}\left\langle \begin{matrix}\tau_n\\ [1\ \ 0]\\ i\end{matrix}\right\rangle^{\alpha_i^{\tau_n}} \Bigg/ \prod_{i,j=1}^{n}(\alpha_i^j)! \ . \tag{3.5}$$

The eigenvalue of the invariant operator is now determined to be

$$\mathcal{G}^{\Delta}_{\tau}([m]+[\Delta]) = \left[\prod_{i<j=1}^{n}\begin{pmatrix} p_{\tau_i n} - p_{\tau_j n} + \Delta_{\tau_i} \\ \Delta_{\tau_j}\end{pmatrix}\Bigg/\begin{pmatrix} p_{\tau_i n} - p_{\tau_j n} \\ \Delta_{\tau_j}\end{pmatrix}\right]^{1/2}. \tag{3.6}$$

It is interesting to note that, depending on which order τ one chooses, the matrix elements of the right-hand side of Eq. (3.5) will have very different appearances. Thus, Eq. (3.5) leads to a variety of different summation expressions representing the same Wigner coefficient. In particular, we point out that for $n = 2$ the ordering $(\tau_1 \tau_2) = (12)$ leads *directly* to Racah's form[41] of the $SU(2)$ Wigner coefficients. [Using the pattern calculus rules, the evaluation of the matrix elements of the string of Wigner operators appearing in Eq. (3.6) for $n = 2$ is quite easy.]

The form (3.5) exhibits clearly that the structure of the totally symmetric Wigner operators is essentially just that of the irrep functions themselves where the domain of definition of these functions is now the set of n^2 fundamental Wigner operators. Observe, in particular, that whenever (Γ) is an extremal pattern, i.e., the Δ pattern is a permutation of the irrep labels $[p\ \ 0]$, then in consequence of the fact that only commuting Wigner operators enter into the right-hand side of Eq. (3.5), the Wigner operator is precisely the corresponding irrep function defined on the fundamental Wigner operators.

We next consider briefly the results of applying the preceding analysis to the general Eq. (2.29). The bosons appearing in right-hand side of this expression are first ordered into sets having a common superscript, followed by the substitution

$$a_i^j \to \left\langle \begin{matrix} j\\ [1\ \ 0]\\ i\end{matrix}\right\rangle .$$

(For notational convenience, we also change all lower case m's to upper case M's.) The conclusion that the resulting operator is an irreducible tensor operator characterized by the Gel'fand pattern

$$\begin{pmatrix}[M]\\ (M)\end{pmatrix}$$

goes through just as before, as does the shift property $[m] \to [m] + [\Delta]$. However, because there are—in general—several canonical Wigner operators possessing this shift property, one can only infer the following general form:

$$\sum_{(\Gamma')}\mathcal{G}_{\tau}\begin{pmatrix}(\Gamma')\\ [M]\\ (\Gamma)\end{pmatrix}\left\langle \begin{matrix}(\Gamma')\\ [M]\\ (M)\end{matrix}\right\rangle = [\mathfrak{M}([M])]^{1/2}\sum_{[\alpha]} C\begin{pmatrix}(\Gamma)\\ [M]\\ (M)\end{pmatrix}(\alpha)\Bigg/\left[\prod_{i,j=1}^{n}(\alpha_i^j)!\right]^{1/2} \times \prod_{i=1}^{n}\left\langle \begin{matrix}\tau_1\\ [1\ \ 0]\\ i\end{matrix}\right\rangle^{\alpha_i^{\tau_1}} \prod_{i=1}^{n}\left\langle \begin{matrix}\tau_2\\ [1\ \ 0]\\ i\end{matrix}\right\rangle^{\alpha_i^{\tau_2}} \cdots \prod_{i=1}^{n}\left\langle \begin{matrix}\tau_n\\ [1\ \ 0]\\ i\end{matrix}\right\rangle^{\alpha_i^{\tau_n}}, \tag{3.7}$$

where the

$$\mathcal{G}_{\tau}\begin{pmatrix}(\Gamma')\\ [M]\\ (\Gamma)\end{pmatrix} \tag{3.8}$$

denote invariant operators, and the sum is over all operator patterns (Γ') which belong to the multiplicity set determined by the Δ pattern of (Γ).

A knowledge of the invariants (3.8) (for any specified τ) would constitute the solution to the construction of the canonical Wigner operators (assuming that the set of equations (3.7) corresponding to the various operator patterns (Γ) is invertible). We can relate the eigenvalues of these invariants to a certain Wigner coefficient in the case of the ordering $\tau = (n \ldots 21)$. Our method for obtaining this result constitutes a rederivation of Eq. (3.7) [for $\tau = (n \ldots 21)$] directly from the factorization lemma (illustrating once more the power of this lemma for obtaining abstract results). We sketch the technique. Using

$$B\begin{pmatrix}(M')\\ [M]\\ (M)\end{pmatrix}(A) = \mathfrak{M}^{1/2}\sum_{\Gamma}\left\langle \begin{matrix}(\Gamma)\\ [M]\\ (M)\end{matrix}\right\rangle_{l}\left\langle \begin{matrix}(\Gamma)\\ [M]\\ (M')\end{matrix}\right\rangle_{u}\mathfrak{M}^{-1/2} \tag{3.9}$$

in the left-hand side of Eq. (2.29) (after a suitable change of notation) and

$$a_i^j = \mathfrak{M}^{1/2}\sum_{\tau=1}^{n}\left\langle \begin{matrix}\tau\\ [1\ \ 0]\\ i\end{matrix}\right\rangle_{l}\left\langle \begin{matrix}\tau\\ [1\ \ 0]\\ j\end{matrix}\right\rangle_{u}\mathfrak{M}^{-1/2} \tag{3.10}$$

in the right-hand side, we obtain the following identity:

$$\sum_{(\Gamma)}\left\langle \begin{matrix}(\Gamma)\\ [M]\\ (M)\end{matrix}\right\rangle_{l}\left\langle \begin{matrix}(\Gamma)\\ [M]\\ (M')\end{matrix}\right\rangle_{u} = \mathfrak{M}^{1/2}([M])\sum_{[\alpha]} C\begin{pmatrix}(M')\\ [M]\\ (M)\end{pmatrix}(\alpha)\Bigg/\left[\prod_{i,j}(\alpha_i^j)!\right]^{1/2} \times \prod_{i,j=1}^{n}\left(\sum_{\tau=1}^{n}\left\langle \begin{matrix}\tau\\ [1\ \ 0]\\ i\end{matrix}\right\rangle_{l}\left\langle \begin{matrix}\tau\\ [1\ \ 0]\\ j\end{matrix}\right\rangle_{u}\right)^{\alpha_i^j}. \tag{3.11}$$

In the next step, we order the fundamental Wigner operator factors in Eq. (3.11) according to the value of j in the following manner:

(all $j = n$ terms) ... (all $j = 2$ terms) (all $j = 1$ terms).

Now we take matrix elements between the states

$$\left\langle\!\!\left\langle \begin{pmatrix} (\max) \\ [m] + [\Delta] \\ (m') \end{pmatrix} \right| \cdots \left| \begin{pmatrix} (\max) \\ [m] \\ (m) \end{pmatrix} \right\rangle\!\!\right\rangle . \tag{3.12}$$

One readily sees that only matrix elements of the maximal type arise for the "upper" fundamental Wigner operators, i.e., all intermediate states which arise from the "upper space" are again maximal. Using repeatedly (for various initial labels $[m]$) the relation

$$\left\langle\!\!\left\langle \begin{pmatrix} [m] + \Delta(\tau) \\ (\max) \end{pmatrix} \right| \left\langle \begin{matrix} \tau \\ [1 \ \ \dot{0}] \\ j \end{matrix} \right| \left| \begin{pmatrix} [m] \\ (\max) \end{pmatrix} \right\rangle\!\!\right\rangle$$

$$= \delta^{\tau}_{j} \left[\prod_{s=1}^{j-1} \frac{(p_{jn} - p_{sn} + 1)}{(p_{jn} - p_{sn})} \right]^{1/2}, \qquad j = 2, 3, \ldots, n \tag{3.13}$$

$$= \delta^{\tau}_{1} \qquad \text{for} \quad j = 1,$$

we obtain from Eq. (3.11) a matrix element relation which is just the statement of the abstract relation (3.7). Now, however, we also obtain the following additional relation for the eigenvalue of the invariant (3.8) in the case $\tau = (n \ldots 21)$:

$$\mathcal{G}_{(n \ldots 21)} \begin{pmatrix} (\Gamma') \\ [M] \\ (\Gamma) \end{pmatrix} ([m] + [\Delta])$$

$$= \left[\prod_{i<j=1}^{n} \frac{(p_{in} - p_{jn} + \Delta_i)}{(p_{in} - p_{jn} + \Delta_i - \Delta_j)} \right]^{1/2} \tag{3.14}$$

$$\times \left\langle\!\!\left\langle \begin{pmatrix} [m] + [\Delta] \\ (\max) \end{pmatrix} \right| \left\langle \begin{pmatrix} (\Gamma') \\ [M] \\ (\Gamma) \end{pmatrix} \right| \left| \begin{pmatrix} [m] \\ (\max) \end{pmatrix} \right\rangle\!\!\right\rangle .$$

Equation (3.14) expresses the unknown invariants in Eq. (3.7) for the ordering $\tau = (n \ldots 21)$ in terms of the matrix elements of a Wigner operator between maximal states.[42] These latter matrix elements are, of course, unknown in the general case, but Eqs. (3.7) and (3.14) do demonstrate clearly that the canonical determination of these *particular matrix elements* would constitute a complete solution to the problem of constructing the *general canonical Wigner operators*. (This important structural property of the factorization lemma has been emphasized repeatedly, and it is not surprising that it reappears in the context above.)

In practice it has thus far proved difficult to use Eqs. (3.7) and (3.14) as a direct method for obtaining explicit results. Even in the case of $U(3)$, where the canonical splitting of the multiplicity is known completely, there appears to be no correspondingly simple structure to the matrix elements (3.14)

In contrast to the similarity in form between Eqs. (2.19) and (3.3), the forms of the left-hand sides of Eqs. (2.29) and (3.7) have quite different appearances. This distinction arises because of the intrinsically different character of Gel'fand versus operator patterns. This difference, in turn, may be attributed directly to the lack of a "subgroup law" for relating operator patterns in a Racah operator (the objects which are used to couple operator patterns). For totally symmetric operators, we have a one-to-one correspondence between operator patterns and Δ patterns; hence, a subgroup-like law always prevails. In the general case, however, we have a many-to-one correspondence between operator patterns and Δ patterns; the further lack of a subgroup significance for the entries in the operator pattern now admits the construction of objects (Wigner operators) more general than structures which are isomorphic to the boson polynomials. The appearance of the invariants in the left-hand side of Eq. (3.7) may be considered to be a consequence of these general structural distinctions between Gel'fand and operator patterns. Our principal reason for presenting Eq. (3.7) has been to emphasize these distinctions.

IV. THE RELATIONSHIP BETWEEN THE EXTENDED PROJECTIVE OPERATORS AND THE RACAH FUNCTIONS

It is our purpose in the present section to demonstrate that there exists a close structural relationship between two types of objects in the Racah-Wigner calculus[8]: the extended projective operators and the Racah invariants—despite the fact that these two structures entered the calculus in very different ways. This relationship is particularly close for $U(2)$—the two objects prove in fact to be identical—and one obtains in this case new insight into the structure of the Racah function of angular momentum theory.

Let us briefly recall the way in which the extended projective operators were introduced.[8] The projective operators themselves are the Wigner operators of $U(n)$ projected onto the $U(n-1)$ subgroup; accordingly, the projective operators evaluated on $U(n)$: $U(n-1)$ states become functions evaluated on the variables $(m_{1n}, m_{2n}, \ldots m_{nn}; m_{1,n-1}, \ldots, m_{n-1,n-1})$. The pattern calculus evaluates the elementary projective operators by the pattern calculus rules associated with the pattern:

● ● ● ... ● ● n dots

● ● ... ● $n-1$ dots,

each dot being associated with a partial hook, $p_{ij} \equiv m_{ij} + j - i$.

It was observed that the pattern calculus rules were meaningful if applied to an "extended" pattern where each row had n dots. (This introduced an additional parameter $m_{n,n-1}$.) In this way one defines *extended projective operators*, initially for elementary irrep labels of the form $[\dot{1}_k \ \dot{0}_{n-k}]$, and then generalizes (by Racah coupling) to all irrep labels.

Clearly this introduction of extended operators is completely *ad hoc*; why should one expect such a concept to be of any use? The answer lies in the way that the projective operators of $U(n)$ contain the projective operators of all lower subgroups $U(k), k < n$, very much in the manner of "nested Chinese boxes": under the *two* limit operators: first, lim: $m_{n,n} \to -\infty$, followed by lim: $m_{n-1,n-1} \to -\infty$ the projective operators of $U(n)$: $U(n-1)$ go over into projective operators of $U(n-1)$: $U(n-2)$. (In terms of patterns, this is just the removal of the last pair of dots.)

The motivation behind the extended projective operators is now clear: They are functions defined by a

single limit ($m_{nn} \to -\infty$), and are to be "intermediate" in the set of "nested" projective operators. The fact that the $U(2)$ extended projective operators turn out to be Racah functions is remarkable—even though qualitatively this is just what might have been suspected (or hoped!) from the known limit relations between Racah and Wigner coefficients[43,44] in $U(2)$.

We will now demonstrate this identity, explicitly for the $U(2)$ case, but in a manner general enough to encompass the $U(n)$ case subsequently. As mentioned earlier, the definition of the extended projective operators is first given for the set of elementary irreps, and then extended to all irreps by the (symbolic) coupling law:

$$\begin{bmatrix}(\Gamma)\\ [M]\\ (\Gamma')\end{bmatrix}_{\text{ext}} = \begin{bmatrix}\cdot\\ [a]\\ \cdot\end{bmatrix}_{\text{ext}} \{R\} \begin{bmatrix}\cdot\\ [b]\\ \cdot\end{bmatrix}_{\text{ext}}. \tag{4.1}$$

Equation (4.1) denotes the Racah coupling of operator patterns in both upper and lower operator space of the two extended operators on the right-hand side. (The dots signify operator patterns summed over in the Racah coupling.)

The symbolic coupling of Eq. (4.1) corresponds to the following fully explicit form[45]:

$$\begin{aligned}
&\left\langle\left\langle \begin{pmatrix}[x]+[\Delta]\\ [y]+[\Delta']\end{pmatrix} \middle\| \begin{bmatrix}(\Gamma)\\ [M]\\ (\Gamma')\end{bmatrix}_{\text{ext}} \middle\| \begin{pmatrix}[x]\\ [y]\end{pmatrix} \right\rangle\right\rangle \\
&= \sum_{\substack{(\alpha)(\beta)\\ (\alpha')(\beta')}} \left\{\begin{pmatrix}[M]\\ (\Gamma)\end{pmatrix}\begin{pmatrix}(\gamma)\\ [a]\\ (\alpha)\end{pmatrix}\begin{pmatrix}[b]\\ (\beta)\end{pmatrix}\right\}([x]+[\Delta]) \\
&\quad\times \left\{\begin{pmatrix}[M]\\ (\Gamma')\end{pmatrix}\begin{pmatrix}(\gamma)\\ [a]\\ (\alpha')\end{pmatrix}\begin{pmatrix}[b]\\ (\beta')\end{pmatrix}\right\}([y]+[\Delta']) \\
&\quad\times \left\langle\left\langle \begin{pmatrix}[x]+[\Delta]\\ [y]+[\Delta']\end{pmatrix} \middle\| \begin{bmatrix}(\alpha)\\ [a]\\ (\alpha')\end{bmatrix}_{\text{ext}} \middle\| \begin{pmatrix}[x]+[\Delta(\beta)]\\ [y]+[\Delta(\beta')]\end{pmatrix} \right\rangle\right\rangle \\
&\quad\times \left\langle\left\langle \begin{pmatrix}[x]+[\Delta(\beta)]\\ [y]+[\Delta(\beta')]\end{pmatrix} \begin{bmatrix}[\beta]\\ [b]\\ (\beta')\end{bmatrix}_{\text{ext}} \middle| \begin{pmatrix}[x]\\ [y]\end{pmatrix} \right\rangle\right\rangle .
\end{aligned} \tag{4.2}$$

The symbolic coupling law of Eq. (4.1) is reminiscent of a similar coupling law for Racah functions. To obtain this coupling law let us begin with a familiar identity for the Racah coefficients[43,44]

$$\begin{aligned}
&W(a\alpha b\beta; c\gamma)\, W(a'\alpha b'\beta; c'\gamma) \\
&\quad= \sum_{\lambda} (2\lambda+1)\, W(a'\lambda\alpha c; ac') \\
&\qquad\times W(b\lambda\beta c'; b'c) \\
&\qquad\times W(a'\lambda\gamma b; ab').
\end{aligned} \tag{4.3}$$

Using the orthonormality of the Racah functions, we may transfer one of the two W's on the left-hand side to the right-hand side. This yields

$$\begin{aligned}
W(a\alpha b\beta; c\gamma) = \sum_{\lambda, b'} &(2\lambda+1)(2\alpha+1)(2b'+1) \\
&\times W(a'\alpha b'\beta; c'\gamma) \\
&\times W(a'\lambda\alpha c; ac') \\
&\times W(b\lambda\beta c'; b'c) \\
&\times W(a'\lambda\gamma b; ab').
\end{aligned} \tag{4.4}$$

The content of the explicit coupling law given in Eq. (4.4) is much more easily grasped in a symbolic form emphasizing the coupling aspects. Let us introduce the operator pattern notation for Racah coefficients which emphasizes couplings[9]:

$$\left\{\begin{pmatrix}c_1 & c_2\\ \rho_1+\sigma_1\end{pmatrix}\begin{pmatrix}c_1-b_1\\ a_1 \quad a_2\\ \rho_1\end{pmatrix}\begin{pmatrix}b_1 & b_2\\ \sigma_1\end{pmatrix}\right\}(m_1 \quad m_2)$$

$$= [(2c+1)(2j-2\rho+1)]^{1/2} W(j-\rho-\sigma, b, j, a; j-\rho, c), \tag{4.5}$$

where $j = \frac{1}{2}(m_1 - m_2)$, $a = \frac{1}{2}(a_1 - a_2)$, $b = \frac{1}{2}(b_1 - b_2)$, $c = \frac{1}{2}(c_1 - c_2)$, $\sigma = \sigma_1 - \frac{1}{2}(b_1 + b_2)$, $\rho = \rho_1 - \frac{1}{2}(a_1 + a_2)$, and $c_1 + c_2 = a_1 + a_2 + b_1 + b_2$. With this notation Eq. (4.4) becomes

$$\begin{aligned}
&\left\{\begin{pmatrix}e\\ \epsilon\end{pmatrix}\begin{pmatrix}\delta''\\ d\\ \delta\end{pmatrix}\begin{pmatrix}c\\ \tau\end{pmatrix}\right\}(m) \\
&\quad= \sum_{\substack{\rho'\sigma'\\ \rho\sigma}} \left\{\begin{pmatrix}d\\ \delta''\end{pmatrix}\begin{pmatrix}\rho''\\ a\\ \rho'\end{pmatrix}\begin{pmatrix}b\\ \sigma'\end{pmatrix}\right\}(e)\left\{\begin{pmatrix}d\\ \delta\end{pmatrix}\begin{pmatrix}\rho''\\ a\\ \rho\end{pmatrix}\begin{pmatrix}b\\ \sigma\end{pmatrix}\right\}(m) \\
&\qquad\times \sum_{\delta'} \left\{\begin{pmatrix}e\\ \epsilon\end{pmatrix}\begin{pmatrix}\rho'\\ a\\ \rho\end{pmatrix}\begin{pmatrix}c+\Delta(\sigma')\\ \delta'\end{pmatrix}\right\}(m) \\
&\qquad\times \left\{\begin{pmatrix}c+\Delta(\sigma')\\ \delta'\end{pmatrix}\begin{pmatrix}\sigma'\\ b\\ \sigma\end{pmatrix}\begin{pmatrix}c\\ \tau\end{pmatrix}\right\}(m-\Delta(\rho)),
\end{aligned} \tag{4.6}$$

where, for example,

$$\begin{pmatrix}e\\ \epsilon\end{pmatrix} = \begin{pmatrix}e_1 & e_2\\ \epsilon\end{pmatrix}.$$

For $SU(2)$ the summation over δ' is redundant in Eq. (4.6) but has been included so that this equation becomes the correct relation between $U(n)$ Racah coefficients.[9]

It will be observed that the recursion relation for $U(2)$ extended projective coefficients Eq. (4.2) and the Racah coefficients Eq. (4.6) are very similar. Indeed, under the proper correspondence, they become exactly the same [cf. Eq. (4.16) for n = 2]. Thus, it will follow that the coefficients themselves are identical *if they agree for the fundamental operators*.

To establish this latter point it is only necessary to make a direct comparison. Consider, for example, the Racah coefficient $W(a, b, e+\frac{1}{2}, \frac{1}{2}; e, b-\frac{1}{2})$. From standard tables, one finds the explicit algebraic form:

$$[(2e+1)(2b)]^{1/2} W(a, b, e+\tfrac{1}{2}, \tfrac{1}{2}; e, b-\tfrac{1}{2}) = (-)\left[\frac{(a+b-e)(a+e+1-b)}{(2e+2)(2b+1)}\right]^{1/2}. \tag{4.7}$$

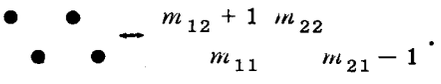
Next consider the extended projective operator denoted by

$$\begin{bmatrix} 0 \\ 1 \quad 0 \\ \quad 1 \end{bmatrix}_{\text{ext}} .$$

We will evaluate this operator on the (extended) Gel'fand pattern

$$\begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & & m_{21} \end{pmatrix}$$

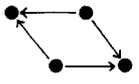
by using the pattern calculus rules. These labels are assigned to the dots of the pattern calculus rules according to



Next we note that the operator

$$\begin{bmatrix} 0 \\ 1 \quad 0 \\ \quad 1 \end{bmatrix}_{\text{ext}}$$

corresponds to the shift pattern $\Delta = [01]$, $\Delta' = [10]$, and hence we get the arrow pattern



The two arrows between rows imply the numerator factor:

$$N^2 = |[m_{11} + 1 - (m_{12} + 1)][m_{22} - (m_{21} - 1)]|. \tag{4.8}$$

The two arrows within rows imply the denominator factor:

$$D^2 = |[m_{12} - (m_{12} + 1)][m_{11} + 1 - (m_{21} - 1)]|. \tag{4.9}$$

Using the sign taken from the pattern calculus rules (a minus sign in the present example) we get the result:

$$\begin{bmatrix} 0 \\ 1 \quad 0 \\ \quad 1 \end{bmatrix}_{\text{ext}} \begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & & m_{21} \end{pmatrix} = (-)\left[\frac{(m_{12} - m_{11})(m_{22} + 1 - m_{21})}{(m_{12} + 1 - m_{22})(m_{11} + 2 - m_{21})}\right]^{1/2}. \tag{4.10}$$

The problem is now to identify the symbols so as to achieve an identity. The correct identification is this [cf. Eqs. (4.16) and (4.5)]:

$$\begin{bmatrix} f-b+d \\ 2d \quad\quad 0 \\ \quad c-e+d \end{bmatrix}_{\text{ext}} \begin{pmatrix} 2b & & 0 \\ & b-a+e & & b-a-e \end{pmatrix} = [(2e+1)(2f+1)]^{1/2} W(abcd;ef). \tag{4.11}$$

(Note that we can add a constant integer to either pattern on the left so as to eliminate the zeroes.)

Using this identification we see that the evaluation given in Eq. (4.10) becomes (for $c = e + \frac{1}{2}, d = \frac{1}{2}$, $f = b - \frac{1}{2}$):

$$[(2e+1)(2f+1)]^{1/2} W(abcd;ef) = (-)\left[\frac{(a+b-e)(a+e-b+1)}{(2b+1)(2e+2)}\right]^{1/2}, \tag{4.12}$$

which is seen to be precisely the result from standard Racah tabulations. In a similar way one may verify that the remaining spin-$\frac{1}{2}$ Racah functions also coincide with their extended projective function counterparts. *Thus, we have established the identity for the fundamental coefficients, and therefore (from the recursion formula) in general.* Equation (4.11) expresses this identity.

We would like to remark that this identity is really quite astonishing for it asserts that *the pattern calculus rules suffice to encompass in SU(2) not only the Wigner operators, but the Racah operators as well.*

To illustrate the content of this remark let us recall that for totally symmetric operators [those of the form $[p\ 0]$, which is the general case for $SU(2)$], if either the upper or the lower pattern is extremal, we have a *monomial* coefficient. *The identity establishes this same monomial property for "extremal" Racah coefficients* (a result, of course, previously known, but now interpretable uniformly).

The identity may also be used, conversely, to extend to the pattern calculus, symmetries and results previously proven for the $SU(2)$ Racah coefficients. In this way one may obtain the following extension of the numerator pattern calculus rules.

We obtain the following result:

$$\begin{bmatrix} \Delta_1 \\ p \quad\quad 0 \\ \quad \Delta_1' \end{bmatrix}_{\text{ext}} \begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & & m_{21} \end{pmatrix} = \left[\frac{\mathfrak{D}(m_{11}\, m_{21})}{\mathfrak{D}(m_{11}'\, m_{21}')}\right]^{1/2} \times N\begin{pmatrix} \Delta_1 & & 0 \\ & p & & \Delta_1' \end{pmatrix} \begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & & m_{21} \end{pmatrix} \times 1\Big/D\begin{pmatrix} \Delta_1' & \Delta_2' \\ p & 0 \end{pmatrix}(m_{11}\, m_{21}) D\begin{pmatrix} \Delta_1 & \Delta_2 \\ p & 0 \end{pmatrix}(m_{12}\, m_{22}), \tag{4.13a}$$

where $\mathfrak{D}$ denotes the dimension operator and the denominators are given by the pattern calculus rules[1] and have the explicit forms:

$$1\Big/D\begin{pmatrix} \Delta_1 & \Delta_2 \\ p & 0 \end{pmatrix}(m_{12}\, m_{22}) = \left[\frac{p!(m_{12} - m_{22} + \Delta_1 - \Delta_2 + 1)(m_{12} - m_{22} - \Delta_2)!}{(\Delta_1)!(\Delta_2)!(m_{12} - m_{22} + \Delta_1 + 1)!}\right]^{1/2}, \tag{4.13b}$$

$$1\Big/D\begin{pmatrix} \Delta_1' & \Delta_2' \\ p & 0 \end{pmatrix}(m_{11}\, m_{21}) = \left[\frac{p!(m_{11} - m_{21} + \Delta_1' - \Delta_2' + 1)(m_{11} - m_{21} - \Delta_2')!}{(\Delta_1')!(\Delta_2')!(m_{11} - m_{21} + \Delta_1' + 1)!}\right]^{1/2}. \tag{4.13c}$$

The values of the numerator function are given by

$$N\begin{pmatrix}\Delta_1 & \\ p & 0 \\ & \Delta'_1\end{pmatrix}\begin{pmatrix}m_{12} & m_{22} \\ & m_{11} & m_{21}\end{pmatrix} = \frac{(-1)^{\Delta_1}}{p!}\prod_{i=1}^{2}(\Delta_i)!(\Delta'_i)!$$
$$\times\left[\frac{(p'_{11}-p'_{22})!}{(p_{11}-p_{22})!}\prod_{i<j=1}^{2}\frac{(p_{i2}-p_{j1}-1)!}{(p'_{i2}-p'_{j1}-1)!}\right]^{1/2}$$
$$\times\sum_{[\alpha]}\begin{pmatrix}p_{11}-p_{12}+\alpha^2_1 \\ \alpha^1_2\end{pmatrix}\begin{pmatrix}p_{11}-p_{22} \\ \alpha^2_2\end{pmatrix}$$
$$\times\begin{pmatrix}p_{21}-p_{12} \\ \alpha^1_1\end{pmatrix}\begin{pmatrix}p_{21}-p_{22}+\alpha^1_2 \\ \alpha^2_1\end{pmatrix} \tag{4.13d}$$

in which the sum over $[\alpha]$ is over all square matrices with nonnegative integral entries with row sums and column sums given by $[\Delta']$ and $[\Delta]$:

$$\begin{array}{|cc|c}\hline \alpha^1_1 & \alpha^2_1 & \Delta'_1 \\ \alpha^1_2 & \alpha^2_2 & \Delta'_2 \\ \hline \Delta_1 & \Delta_2 & \end{array} \quad . \tag{4.13e}$$

(The following notations have also been used in writing out the above results: $\Delta_1+\Delta_2=\Delta'_1+\Delta'_2=p, m'_{12}=m_{12}+\Delta_1, m'_{11}=m_{11}+\Delta'_1, m'_{21}=m_{21}+\Delta'_2$, and $p'_{ij}=m'_{ij}+j-i$.)

Let us note explicitly that this generalization of the numerator function, Eqs. (4.13d, e), has a value which depends *only on the differences* $p_{i2}-p_{j1}, i,j=1,2$.

Let us now generalize—as far as possible—the results obtained above for $SU(2)$. For the extended projective operators the results given in (4.1) and (4.2) are already in general form. The Racah coefficient relation, Eq. (4.6) is also already in general form when we interpret the operator patterns to be general ones. Recall that this relation is derivable directly from the associativity property of the Racah–Wigner calculus.[9] Comparing the general forms (4.2) and (4.6), one observes a similarity of structure, except that there occurs in Eq. (4.6) an extra summation over the patterns δ' [which is but a single term for $U(2)$]. The idea now is to select the operator patterns in this expression such that the summation over δ' is forced to a single term. To this end we write out the following relation from Eq. (4.6), making, however, some notational changes to enhance the similarity to Eq. (4.2):

$$\left\{\begin{pmatrix}[x]_n+[\Delta]_n \\ [y]_{n-1}+[\Delta']_{n-1} \\ (\max)\end{pmatrix}\begin{pmatrix}(\Gamma) \\ [M] \\ (\Gamma')\end{pmatrix}\begin{pmatrix}[x]_n \\ [y]_{n-1} \\ (\max)\end{pmatrix}\right\}([y]_n+[\Delta']_n) = \sum_{\substack{(\alpha)(\beta)\\(\alpha')(\beta')}}\left\{\begin{pmatrix}[M] \\ (\Gamma)\end{pmatrix}\begin{pmatrix}(\gamma) \\ [a] \\ (\alpha)\end{pmatrix}\begin{pmatrix}[b] \\ (\beta)\end{pmatrix}\right\}([x]+[\Delta])$$
$$\times\left\{\begin{pmatrix}[M] \\ (\Gamma')\end{pmatrix}\begin{pmatrix}(\gamma) \\ [a] \\ (\alpha')\end{pmatrix}\begin{pmatrix}[b] \\ (\beta')\end{pmatrix}\right\}([y]+[\Delta'])\sum_{(\gamma')}\left\{\begin{pmatrix}[x]_n+[\Delta]_n \\ [y]_{n-1}+[\Delta']_{n-1} \\ (\max)\end{pmatrix}\begin{pmatrix}(\alpha) \\ [a] \\ (\alpha')\end{pmatrix}\begin{pmatrix}[x]_n+[\Delta(\beta)]_n \\ (\gamma')\end{pmatrix}\right\}([y]_n+[\Delta']_n)$$
$$\times\left\{\begin{pmatrix}[x]_n+[\Delta(\beta)]_n \\ (\gamma')\end{pmatrix}\begin{pmatrix}(\beta) \\ [b] \\ (\beta')\end{pmatrix}\begin{pmatrix}[x]_n \\ [y]_{n-1} \\ (\max)\end{pmatrix}\right\}([y]_n+[\Delta(\beta')]_n). \tag{4.14}$$

This result still contains a summation over γ', and an essential step in the proof is to establish the following subsidiary relation:

$$\left\{\begin{pmatrix}[x]_n+[\Delta(\beta)]_n \\ (\gamma')\end{pmatrix}\begin{pmatrix}(\beta) \\ [b] \\ (\beta')\end{pmatrix}\begin{pmatrix}[x]_n \\ [y]_{n-1} \\ (\max)\end{pmatrix}\right\}([y]_n+[\Delta(\beta')]_n$$
$$=0 \text{ unless } (\gamma')=\begin{pmatrix}[y]_{n-1}+[\Delta(\beta')]_{n-1} \\ (\max)\end{pmatrix}. \tag{4.15}$$

Let us assume for the moment that this result is correct. Then under the identification,

$$\left\langle\left\langle\begin{pmatrix}[x]+[\Delta] \\ [y]+[\Delta']\end{pmatrix}\right|\left|\begin{bmatrix}(\Gamma) \\ [M] \\ (\Gamma')\end{bmatrix}_{\text{ext}}\right|\left|\begin{pmatrix}[x] \\ [y]\end{pmatrix}\right\rangle\right\rangle$$
$$=\left\{\begin{pmatrix}[x]_n+[\Delta]_n \\ [y]_{n-1}+[\Delta']_{n-1} \\ (\max)\end{pmatrix}\begin{pmatrix}(\Gamma) \\ [M] \\ (\Gamma')\end{pmatrix}\begin{pmatrix}[x]_n \\ [y]_{n-1} \\ (\max)\end{pmatrix}\right\}([y]_n+[\Delta']_n), \tag{4.16}$$

Eqs. (4.2) and (4.14) become precisely the same. Since either Eq. (4.2) or (4.14) may be used to generate recursively the corresponding general coefficients (for $M_{nn}\geq 0$) from the fundamental ones, it follows that Eq. (4.16) is correct in general ($M_{nn}\geq 0$) if it is correct for the fundamental operators, i.e., if

$$\left\langle\left\langle\begin{pmatrix}[x]+\Delta(\tau) \\ [y]+\Delta(\rho)\end{pmatrix}\right|\left|\begin{bmatrix}\tau \\ [1\ \dot{0}] \\ \rho\end{bmatrix}_{\text{ext}}\right|\left|\begin{pmatrix}[x] \\ [y]\end{pmatrix}\right\rangle\right\rangle$$
$$=\left\{\begin{pmatrix}[x]_n+\Delta_n(\tau) \\ [y]_{n-1}+\Delta_{n-1}(\rho) \\ (\max)\end{pmatrix}\begin{pmatrix}\tau \\ [1\ \dot{0}] \\ \rho\end{pmatrix}\begin{pmatrix}[x]_n \\ [y]_{n-1} \\ (\max)\end{pmatrix}\right\}([y]_n+\Delta_n(\rho)). \tag{4.17}$$

The proofs of both Eqs. (4.15) and (4.17) are given in Appendix B, and hence establish the desired identity, Eq. (4.16).

Equation (4.16) expresses the fact that the extended projective coefficients are, in the general case, identical to a subset of the set of general Racah coefficients.

V. CONCLUDING REMARKS

A number of new and interesting results have been obtained. These include: (a) the derivation of explicit expressions for the general boson polynomials, where the rich structure of these forms has been illustrated by showing how they yield as particular cases the Gel'fand–Graev beta functions and even the irreps of the symmetric group; (b) the derivation of corresponding forms for the canonical tensor operators; and (c) the

proof of a relation between extended projective coefficients and certain Racah coefficients. While none of these developments completes our principal goal of actually constructing the set of general canonical tensor operators, each suggests new approaches to this task.

Consider first the boson polynomial expressions. These may now be used directly in the factorization lemma[3,4,7,8] to evaluate the boson scalar product part of the lemma (Appendix B contains an example of this technique). One might hope then to see how to take the boson matrix element apart to determine the canonical Wigner coefficients appearing in the lemma. This is a difficult undertaking, but Holman[11] has made considerable progress in this direction, particularly, for $U(3)$, and his efforts still continue (private communication). There are, of course, also other ways[1,4,8] of using the factorization lemma.

The results of Sec. III already give the complete formal answer for the totally symmetric Wigner operators. They also demonstrate how the general problem might be solved by directing ones attention to the determination of the particular Wigner coefficients appearing in Eq. (3. 14). More significant, perhaps, is the reminder that one must really deal effectively with the noncommutative properties of the algebra—hence, with the Racah coefficients.

It becomes increasingly clear that determining the fundamental Racah coefficients is the key to the whole problem of defining the canonical tensor operators. This remark becomes more meaningful when one realizes that Eq. (4. 6) may be interpreted as a relation which generates (recursively) the general Racah coefficient from the fundamental ones. It takes on an added significance if one accepts the conjecture[8] that Racah coefficients limit to Wigner coefficients—hence, the whole Racah–Wigner calculus becomes a structure which is uniquely defined through its algebraic relations (coupling laws, orthonormality relations, etc.) and the fundamental Racah coefficients. (The Racah coefficients seem to play the rôle in defining the Racah–Wigner calculus which the structure constants play in a Lie algebra.)

In view of the significance of Racah coefficients, the results of Sec. IV become more prominent. Furthermore, the fact that the $U(2)$ extended projective coefficient takes the structural form exhibited by the right-hand side of Eq. (4. 13a) strongly suggests the existence of further generalizations of the pattern calculus rules.

It is our plan to pursue these new approaches. The problem of giving an explicit determination of the fundamental Racah coefficients in $U(3)$ is particularly challenging (and already solved, in principle, in I and II), and we have already made considerable progress with this. The intricate relationship of Racah coefficients (subgroup independent objects) to null spaces (a subgroup independent concept) of Wigner operators, and the related concept of indecomposability[46] are problems of considerable complexity, yet the unexpected elegant structure of results[2] seems to justify the effort to understand them.

APPENDIX A

Commuting sets of elementary Wigner operators

The purpose of this appendix is to prove the mutual commutivity of the Wigner operators belonging to the set of elementary Wigner operators having a common operator pattern, i.e., to prove the commutivity of the *components* of the elementary tensor operator which transforms like the irrep $[\dot{1}_k\ \dot{0}_{n-k}]$ and effects the shift

$$\Delta = \sum_{s=1}^{k} \Delta_n(\tau_s), \tag{A1}$$

where the τ_i are any integers satisfying $1 \leq \tau_1 < \tau_2 < \ldots < \tau_k \leq n$. The commutation properties of the elementary projective operators have been discussed previously,[4] but it is not at all obvious from those results which elementary Wigner operators commute.

The notation[4]

$$\left\langle \begin{matrix} (\tau)_k \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle, \tag{A2}$$

where $(\tau)_k = (\tau_1\tau_2 \ldots \tau_k)$, $1 \leq \tau_1 < \tau_2 < \ldots < \tau_k \leq n$, and $(i)_k = (i_1 i_2 \ldots i_k)$, $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, denotes the elementary Wigner operator which has the unique operator pattern determined by the shift (A1) and which has the unique Gel'fand pattern determined by the weight

$$\sum_{s=1}^{k} \Delta_n(i_s).$$

The result proved in this appendix can now be expressed in the following form:

$$\left[\left\langle \begin{matrix} (\tau)_k \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle, \left\langle \begin{matrix} (\tau)_k \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (j)_k \end{matrix} \right\rangle \right] = 0 \tag{A3}$$

for each $k = 1, 2, \ldots, n$, for each operator pattern $(\tau)_b$, and for each pair $(i)_k, (j)_k$ of Gel'fand patterns.

The proof can be given directly by using the product law[8] for two Wigner operators. For the case of interest, this law becomes the following relation:

$$\begin{aligned} &\left\langle \begin{matrix} (\tau)_k \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle \left\langle \begin{matrix} (\tau)_k \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (j)_k \end{matrix} \right\rangle \\ &= \left\{ \begin{pmatrix} [\dot{2}_k\ \dot{0}_{n-k}] \\ (\tau)_k \end{pmatrix} \begin{pmatrix} (\max) \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (\tau)_k \end{pmatrix} \begin{pmatrix} [\dot{1}_k\ \dot{0}_{n-k}] \\ (\tau)_k \end{pmatrix} \right\} \\ &\quad \times \sum_{(m)} \left\langle \begin{matrix} [\dot{2}_k\ \dot{0}_{n-k}] \\ (m) \end{matrix} \middle| \left\langle \begin{matrix} (\max) \\ [\dot{1}_k\ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle \middle| \left\langle \begin{matrix} [\dot{1}_k\ \dot{0}_{n-k}] \\ (j)_k \end{matrix} \right\rangle \right. \\ &\quad \times \left\langle \begin{matrix} (\tau)_k \\ [\dot{2}_k\ \dot{0}_{n-k}] \\ (m) \end{matrix} \right\rangle . \end{aligned} \tag{A4}$$

In this result, the notation

$$\begin{pmatrix} [\dot{2}_k\ \dot{0}_{n-k}] \\ (\tau)_k \end{pmatrix} \tag{A5}$$

designates the operator pattern having the Δ pattern

$$2 \sum_{s=1}^{k} \Delta_n(\tau_s).$$

The crucial point to note in obtaining Eq. (A4) is that

choosing the two operator patterns $(\tau)_k$ and $(\rho)_k$ to be the same in the product law for

$$\left\langle \begin{matrix} (\tau)_k \\ [\dot{1}_k \ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle \left\langle \begin{matrix} (\rho)_k \\ [\dot{1}_k \ \dot{0}_{n-k}] \\ (j)_k \end{matrix} \right\rangle$$

has the effect of forcing the admissible final operator patterns in the Racah invariant to the single pattern (A5); hence, the irrep labels of the Wigner coefficient are also forced to be $[\dot{2}_k \ \dot{0}_{n-k}]$.

The proof of property (A3) now follows immediately from Eq. (A4) and the following symmetry property of the Wigner coefficients:

$$\left\langle \begin{matrix} [\dot{2}_k \ \dot{0}_{n-k}] \\ (m) \end{matrix} \middle| \left\langle \begin{matrix} (\max) \\ [\dot{1}_k \ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle \middle| \begin{matrix} [\dot{1}_k \ \dot{0}_{n-k}] \\ (j)_k \end{matrix} \right\rangle = \left\langle \begin{matrix} [\dot{2}_k \ \dot{0}_{n-k}] \\ (m) \end{matrix} \middle| \left\langle \begin{matrix} (\max) \\ [\dot{1}_k \ \dot{0}_{n-k}] \\ (j)_k \end{matrix} \right\rangle \middle| \begin{matrix} [\dot{1}_k \ \dot{0}_{n-k}] \\ (i)_k \end{matrix} \right\rangle \tag{A6}$$

for $k = 1, 2, \ldots, n$. This symmetry relation can be proved by several methods—one such method would be to evaluate these matrix elements directly through the use of the subgroup reduction law and the pattern calculus rules. A simpler method (since we do not require the explicit value of the coefficient) is to use induction on n.

We assume that Eq. (A6) is correct for n replaced by $n-1$ and for each $k = 1, 2, \ldots, n-1$, and then prove its validity for n and each $k = 1, 2, \ldots, n$. Since the relation is correct for $n = 2$, its general validity for all n then follows.

To prove the validity of Eq. (A6) under the induction on n, we note that row $n-1$ in the pattern (m) must have one of the following three forms: (a) $[m]_{n-1} = [\dot{2}_k \ \dot{0}_{n-k-1}]$ or $[\dot{2}_{k-1} \ \dot{0}_{n-k}]$; (b) $[m]_{n-1} = [\dot{2}_{k-1} \ 1 \ \dot{0}_{n-k-1}]$. In case (a) the patterns $(i)_k$ and $(j)_k$ must have identical labels in their $(n-1)$st rows given by $[\dot{1}_k \ \dot{0}_{n-k-1}]$ or $[\dot{1}_{k-1} \ \dot{0}_{n-k}]$, respectively. Using the subgroup reduction law on both sides of Eq. (A6) and the assumed validity for $n-1$, we easily establish the validity of Eq. (A6) for case (a). For case (b) there are two possible entries in the $(n-1)$st rows of the patterns $(i)_k$ and $(j)_k$, namely, $[\dot{1}_k \ \dot{0}_{n-k-1}]$ in $(i)_k$ and $[\dot{1}_{k-1} \ \dot{0}_{n-k}]$ in $(j)_k$, or we may make the opposite choice. It is, however, no restriction to make the first choice since the second is equivalent to interchanging the left- and right-hand sides of Eq. (A6). We next observe that

$$\left| \begin{matrix} [\dot{2}_k \quad \dot{0}_{n-k}] \\ [\dot{2}_{k-1} \ 1 \ \dot{0}_{n-k-1}] \\ (m)_{n-2} \end{matrix} \right\rangle = \# \, E_{n,n-1} \left| \begin{matrix} [\dot{2}_k \ \dot{0}_{n-k}] \\ [\dot{2}_k \ \dot{0}_{n-k-1}] \\ (m)_{n-2} \end{matrix} \right\rangle, \tag{A7}$$

where # denotes a nonzero number, and $E_{n,n-1}$ is one of the generators of $U(n)$. We now use Eq. (A7) in each side of Eq. (A6) and perform the following operations on each side: transfer $E_{n,n-1}$ to the position just left of the Wigner operator by replacing it by its Hermitian conjugate $E_{n-1,n}$; evaluate the action of $E_{n-1,n}$ on the Wigner operator and the initial state vector. The result of these straightforward steps is: Equation (A6) is valid for $[m]_{n-1} = [\dot{2}_{k-1} \ 1 \ \dot{0}_{n-k-1}]$ if it is valid for $[m]_{n-1} = [\dot{2}_k \ \dot{0}_{n-k-1}]$. Since the latter result has already been proved [case (a)], the validity of Eq. (A6) for case (b) is also proved. The induction loop has now been closed, and the general validity of the symmetry relation (A6) established.

APPENDIX B

The relation of the extended fundamental projective coefficients and Racah coefficients

The purpose of this appendix is to prove Eqs. (4.15) and (4.17) of Sec. IV. We begin with the proof of Eq. (4.17). The key relation for giving the proof is Eq. (2.25b) of Sec. II. Using this relation and properties (2.26b) and (2.27), we easily obtain the following result which is valid for $i, j < n$:

$$\begin{aligned} &\left\langle \begin{pmatrix} (\nu')_{n-1} \\ [\nu']_{n-1} \ 0 \\ (m')_{n-1} \end{pmatrix} \middle| a^j_i \middle| \begin{pmatrix} (\nu)_{n-1} \\ [\nu]_{n-1} \ 0 \\ (m)_{n-1} \end{pmatrix} \right\rangle \\ &= \sum_{\substack{(\kappa)_{n-2}(\mu)_{n-2} \\ (\mu')_{n-2}}} \left\langle \begin{pmatrix} [\nu]_{n-1} \\ (\nu)_{n-2} \end{pmatrix} \middle| \begin{pmatrix} (\gamma) \\ [k \ \dot{0}]_{n-1} \\ (\kappa)_{n-2} \end{pmatrix} \middle| \begin{pmatrix} [m]_{n-1} \\ (\mu)_{n-2} \end{pmatrix} \right\rangle \\ &\times \left\langle \begin{pmatrix} [\nu']_{n-1} \\ (\nu')_{n-2} \end{pmatrix} \middle| \begin{pmatrix} (\gamma') \\ [k \ \dot{0}]_{n-1} \\ (\kappa)_{n-2} \end{pmatrix} \middle| \begin{pmatrix} [m']_{n-1} \\ (\mu')_{n-2} \end{pmatrix} \right\rangle \\ &\times \left\langle \begin{pmatrix} (\mu')_{n-2} \\ [m']_{n-1} \\ (m')_{n-2} \end{pmatrix} \middle| a^j_i \middle| \begin{pmatrix} (\mu)_{n-2} \\ [m]_{n-1} \\ (m)_{n-2} \end{pmatrix} \right\rangle \end{aligned} \tag{B1}$$

where $[\nu]_{n-1} = [\nu_{1,n-1} \nu_{2,n-1} \cdots \nu_{n-1,n-1}] = [m_{1n} m_{2n} \cdots m_{n-1,n}]$, $[m]_{n-1} = [m_{1,n-1} m_{2,n-1} \cdots m_{n-1,n-1}]$, and similar definitions are made for the labels with primes. The number k as well as the operator patterns (γ) and (γ') are uniquely determined in terms of the specified labels.

We now proceed to simplify Eq. (B1) by effecting the following four operations: (1) Set $[\nu']_{n-1} = [\nu]_{n-1} + \Delta_{n-1}(\tau)$ and $[m']_{n-1} = [m]_{n-1} + \Delta_{n-1}(\rho)$ for some τ and ρ selected from 1 to $n-1$. (2) Apply the factorization lemma to each side, i.e., the factorization lemma for $U(n)$ in the left-hand side and the factorization lemma for $U(n-1)$ in the right-hand side. (3) Apply the $U(n) : U(n-1)$ subgroup reduction law to the fundamental $U(n)$ Wigner operators which appear at step (2) in the left-hand side. (4) Use the orthonormality of the fundamental $U(n-1)$ Wigner coefficients to move all $U(n-1)$ Wigner coefficients to the right-hand side of the equation. The result of these operations is the following expression:

$$\begin{aligned} &\left\langle \begin{pmatrix} [[\nu]_{n-1} \ 0] + \Delta_n(\tau) \\ [m]_{n-1} + \Delta_{n-1}(\rho) \end{pmatrix} \middle| \begin{bmatrix} \tau \\ [1 \ \dot{0}]_n \\ \rho \end{bmatrix} \middle| \begin{pmatrix} [\nu]_{n-1} \ 0 \\ [m]_{n-1} \end{pmatrix} \right\rangle \\ &= \left[\frac{\mathfrak{M}([\nu]_{n-1}) \mathfrak{M}([m]_{n-1} + \Delta_{n-1}(\rho))}{\mathfrak{M}([m]_{n-1}) \mathfrak{M}([\nu]_{n-1} + \Delta_{n-1}(\tau))} \right]^{1/2} \\ &\times \sum_{\substack{j,(\nu)_{n-2}(\kappa)_{n-2} \\ (\mu)_{n-2}(\mu')_{n-2}}} \left\langle \begin{pmatrix} [\nu]_{n-1} + \Delta_{n-1}(\tau) \\ (\nu')_{n-2} \end{pmatrix} \middle| \begin{matrix} \tau \\ [1 \ \dot{0}]_{n-1} \\ j \end{matrix} \middle| \begin{pmatrix} [\nu]_{n-1} \\ (\nu)_{n-2} \end{pmatrix} \right\rangle \end{aligned}$$

$$\times \left\langle \begin{pmatrix} [\nu]_{n-1} \\ (\nu)_{n-2} \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma) \\ [k\ \dot{0}]_{n-1} \\ (\kappa)_{n-2} \end{matrix} \right\rangle \middle| \begin{pmatrix} [m]_{n-1} \\ (\mu)_{n-2} \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [\nu]_{n-1} + \Delta_{n-1}(\tau) \\ (\nu')_{n-2} \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma') \\ [k\ \dot{0}]_{n-1} \\ (\kappa)_{n-2} \end{matrix} \right\rangle \middle| \begin{pmatrix} [m]_{n-1} + \Delta_{n-1}(\rho) \\ (\mu')_{n-2} \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [m]_{n-1} + \Delta_{n-1}(\rho) \\ (\mu')_{n-2} \end{pmatrix} \middle| \left\langle \begin{matrix} \rho \\ [1\ \dot{0}]_{n-1} \\ j \end{matrix} \right\rangle \middle| \begin{pmatrix} [m]_{n-1} \\ (\mu)_{n-2} \end{pmatrix} \right\rangle \tag{B2}$$

in which the pattern $(\nu')_{n-2}$ is fixed, but arbitrary.

We next perform the operations as follows on Eq. (B2): (1) Add m_{nn} to all of the initial and all of the final irrep labels in the left-hand side (this does not change the value of the matrix element). (2) Replace m_{in} by $m_{in} - m_{nn}$ and $m_{i,n-1}$ by $m_{i,n-1} - m_{nn}$ for $i = 1, 2, \ldots, n-1$ in both sides of the equation resulting from step (1), noting that the summation part of the right-hand side is unaltered by these substitutions. (3) Separate off the terms of the right-hand side which contain m_{nn}, noting that this factor is $[(p_{\rho n} - p_{nn} + 1)/(p_{\tau n} - p_{nn})]^{1/2}$ (originating entirely from the measure factors preceding the summation). (4) Take the limit of the resulting equation as $m_{nn} \to -\infty$. The result of these four operations is an equation which relates an extended fundamental projective coefficient in $U(n-1)$ (we know these limit relations are valid for the fundamental operators) to a certain sum over $U(n-1)$ Wigner coefficients. Finally, we perform the last operation: (5) Replace n by $n+1$ and make a change of notation so that the left-hand side becomes the left-hand side of Eq. (4.17). The result of these operations is the following expression:

$$\left\langle \begin{pmatrix} [x] + \Delta(\tau) \\ [y] + \Delta(\rho) \end{pmatrix} \middle| \begin{bmatrix} \tau \\ [1\ \dot{0}] \\ \rho \end{bmatrix}_{\text{ext}} \middle| \begin{pmatrix} [x] \\ [y] \end{pmatrix} \right\rangle$$

$$= \left[\frac{\mathfrak{D}([y])\,\mathfrak{D}([x] + \Delta(\tau))}{\mathfrak{D}([x])\,\mathfrak{D}([y] + \Delta(\rho))} \right]^{1/2}$$

$$\times \sum_{j,(\alpha)(\kappa)(\beta)(\beta')} \left\langle \begin{pmatrix} [x] + \Delta(\tau) \\ (\alpha') \end{pmatrix} \middle| \left\langle \begin{matrix} \tau \\ [1\ \dot{0}] \\ j \end{matrix} \right\rangle \middle| \begin{pmatrix} [x] \\ (\alpha) \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [x] \\ (\alpha) \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma) \\ [k\ \dot{0}] \\ (\kappa) \end{matrix} \right\rangle \middle| \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [x] + \Delta(\tau) \\ (\alpha') \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma') \\ [k\ \dot{0}] \\ (\kappa) \end{matrix} \right\rangle \middle| \begin{pmatrix} [y] + \Delta(\rho) \\ (\beta') \end{pmatrix} \right\rangle$$

$$\times \left\langle \begin{pmatrix} [y] + \Delta(\rho) \\ (\beta') \end{pmatrix} \middle| \left\langle \begin{matrix} \rho \\ [1\ \dot{0}] \\ j \end{matrix} \right\rangle \middle| \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \right\rangle , \tag{B3}$$

in which (α') is a "free" pattern, and $\mathfrak{D}$ is the dimension operator.

We have now expressed an extended fundamental projective coefficient in terms of a summation over four $U(n)$ Wigner coefficients. It remains to be demonstrated that the right-hand side of Eq. (B3) is just the special Racah coefficient appearing on the right-hand side of Eq. (4.17). This requires further manipulation of the Wigner coefficients to which we now turn.

We use the identity:

$$\left\langle \begin{matrix} (\Gamma) \\ [k\ \dot{0}] \\ (M) \end{matrix} \right\rangle = (-1)^{\phi(\Gamma)-\phi(M)}\,\mathfrak{D}^{-1/2} \left\langle \begin{matrix} (\bar{\Gamma}) \\ [\dot{0}\ -k] \\ (\bar{M}) \end{matrix} \right\rangle \mathfrak{D}^{1/2}, \tag{B4}$$

to obtain the following relations:

$$\left\langle \begin{pmatrix} [x] \\ (\alpha) \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma) \\ [k\ \dot{0}] \\ (\kappa) \end{matrix} \right\rangle \middle| \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \right\rangle$$

$$= \left\langle \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \middle| \left\langle \begin{matrix} (\gamma) \\ [k\ \dot{0}] \\ (\kappa) \end{matrix} \right\rangle^{\dagger} \middle| \begin{pmatrix} [x] \\ (\alpha) \end{pmatrix} \right\rangle$$

$$= \left[\frac{\mathfrak{D}([x])}{\mathfrak{D}([y])} \right]^{1/2} (-1)^{\phi(\gamma)-\phi(\kappa)} \left\langle \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \middle| \left\langle \begin{matrix} (\bar{\gamma}) \\ [\dot{0}\ -k] \\ (\bar{\kappa}) \end{matrix} \right\rangle \middle| \begin{pmatrix} [x] \\ (\alpha) \end{pmatrix} \right\rangle$$

$$= \left[\frac{\mathfrak{D}([x])}{\mathfrak{D}([y])} \right]^{1/2} (-1)^{\phi(\gamma)-\phi(\kappa)} (-1)^{\delta([x];[y])}$$

$$\times \left\langle \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \middle| \left\langle \begin{matrix} (\max) \\ [y]_{n-1} \\ [x] \\ (\alpha) \end{matrix} \right\rangle \middle| \begin{pmatrix} [\dot{0}\ -k] \\ (\bar{\kappa}) \end{pmatrix} \right\rangle . \tag{B5}$$

The last step in this result requires justification.

In order to justify the last step in Eq. (B5), we consider the following couplings of irreps:

$$[\dot{0}\ -k] \otimes [x] \to [y], \tag{B6a}$$

$$[x] \otimes [\dot{0}\ -k] \to [y], \tag{B6b}$$

where

$$k = \sum_{i=1}^{n} (x_i - y_i) \geq 0. \tag{B6c}$$

In the case of the coupling (B6a), the Wigner coefficients are the ones designated by

$$\left\langle \begin{pmatrix} [y] \\ (\beta) \end{pmatrix} \middle| \left\langle \begin{matrix} (\bar{\gamma}) \\ [\dot{0}\ -k] \\ (\bar{\kappa}) \end{matrix} \right\rangle \middle| \begin{pmatrix} [x] \\ (\alpha) \end{pmatrix} \right\rangle . \tag{B7a}$$

These coefficients are zero unless the pattern

$$\begin{pmatrix} [x] \\ [y] \end{pmatrix} \tag{B7b}$$

is lexical, i.e., unless $x_1 \geq y_1 \geq x_2 \geq y_2 \geq \cdots \geq x_n \geq y_n$. This nontrivial property is verified by direct examination of the reduced matrix elements given in Refs. 1 and 24. In the case of the coupling (B6b), the Wigner coefficients are the matrix elements denoted by

$$\left\langle\left(\begin{matrix}[y]\\(\beta)\end{matrix}\right)\right| \left\langle\begin{matrix}(\Gamma)\\[x]\\(\alpha)\end{matrix}\right\rangle \left|\left(\begin{matrix}[\dot{0}\ \ -k]\\(\bar{\kappa})\end{matrix}\right)\right\rangle, \tag{B8}$$

where (Γ) denotes an operator pattern belonging to the *multiplicity set* determined by the following Δ pattern belonging to $[x]$:

$$[\Delta] = [y_1 y_2 \cdots y_{n-1} y_n + k]. \tag{B9}$$

However, since the multiplicity of $[y]$ in the direct product (B6b) is the same as that in (B6a), namely, *one if conditions* (B7b) *hold or zero otherwise*, the coefficients (B8) must vanish except for precisely *one operator pattern* in the multiplicity set defined by the Δ pattern (B9), and even then the coefficient must vanish unless conditions (B7b) hold. We assert, and later justify, that the unique operator pattern for which the coefficients (B8) are not identically vanishing is

$$\begin{pmatrix}[x]\\(\Gamma)\end{pmatrix} = \begin{pmatrix}[x]\\[y]_{n-1}\\(\max)\end{pmatrix}. \tag{B10}$$

Since the ordering of the coupling of two irreps in which there is no multiplicity in the direct product can at most change the overall phase, the following relation must hold:

$$\left\langle\left(\begin{matrix}[y]\\(\beta)\end{matrix}\right)\right| \left\langle\begin{matrix}(\bar{\gamma})\\[\dot{0}\ \ -k]\\(\bar{\kappa})\end{matrix}\right\rangle \left|\left(\begin{matrix}[x]\\(\alpha)\end{matrix}\right)\right\rangle$$

$$= (-1)^{\delta([x];[y])} \left\langle\left(\begin{matrix}[y]\\(\beta)\end{matrix}\right)\right| \left\langle\begin{matrix}(\max)\\[y]_{n-1}\\[x]\\(\alpha)\end{matrix}\right\rangle \left|\left(\begin{matrix}[\dot{0}\ \ -k]\\(\bar{\kappa})\end{matrix}\right)\right\rangle, \tag{B11}$$

where $\delta([x];[y])$ is an integer which depends at most on the x_i and y_i. Equation (B11) is the result which we have used in making the last step in Eq. (B5).

One may still question whether or not the identification (B10) is correct, or whether or not it is a question of choice. Let us defer this problem for the moment, noting, however, that the identification is unambiguous in the case of $U(2)$ *and in the case of* $U(3)$. [The $U(3)$ case may be verified directly from the intertwining number–null space diagram given in Ref. 2.]

Equation (B3) may now be brought to the following form by using Eq. (B5) to replace the second factor in the right-hand side and a similar expression to replace the third factor:

$$\left\langle\left(\begin{matrix}[x]+\Delta(\tau)\\[y]+\Delta(\rho)\end{matrix}\right)\right| \left[\begin{matrix}\tau\\[1\ \ \dot{0}]\\\rho\end{matrix}\right]_{\text{ext}} \left|\left(\begin{matrix}[x]\\[y]\end{matrix}\right)\right\rangle$$

$$= (\text{phase})\, \mathfrak{D}([x]+\Delta(\tau))/\mathfrak{D}([y]+\Delta(\rho))$$

$$\times \sum_{(\beta')j(\alpha)} \left\langle\left(\begin{matrix}[x]+\Delta(\tau)\\(\alpha')\end{matrix}\right)\right| \left\langle\begin{matrix}\tau\\[1\ \ \dot{0}]\\j\end{matrix}\right\rangle \left|\left(\begin{matrix}[x]\\(\alpha)\end{matrix}\right)\right\rangle$$

$$\times \left\langle\left(\begin{matrix}[y]+\Delta(\rho)\\(\beta')\end{matrix}\right)\right| \left\langle\begin{matrix}\rho\\[1\ \ \dot{0}]\\j\end{matrix}\right\rangle \left\langle\begin{matrix}(\max)\\[y]_{n-1}\\[x]\\(\alpha)\end{matrix}\right\rangle$$

$$\times \left\langle\begin{matrix}(\max)\\[y]_{n-1}+\Delta_{n-1}(\rho)\\[x]+\Delta(\tau)\\(\alpha')\end{matrix}\right\rangle^{\dagger} \left|\left(\begin{matrix}[y]+\Delta(\rho)\\(\beta')\end{matrix}\right)\right\rangle, \tag{B12}$$

where in obtaining this result we have recognized that certain of the summations appearing in the right-hand side of Eq. (B3) became sums over intermediate states. The phase in this result is given by

$$(-1)^{\delta_{\tau\rho}([x];[y])}, \tag{B13a}$$

where

$$\delta_{\tau\rho}([x];[y]) = \tau - \rho + \delta([x]+\Delta(\tau),[y]+\Delta(\rho)) - \delta([x];[y]). \tag{B13b}$$

We are now ready to make the last step in the proof of Eq. (4.17) of Sec. IV. This is accomplished by summing both sides of Eq. (B12) over all patterns (α'). The right-hand side then becomes [see Eq. (2.46) of Ref. 8] (phase) $\mathfrak{D}([x])+\Delta(\tau))$ (the Racah coefficient of Eq. (4.17)].

Since summing Eq. (B12) over (α') also puts the dimension factor $\mathfrak{D}([x]+\Delta(\tau))$ in the left-hand side of Eq. (B12), we obtain the desired result, Eq. (4.17), except possibly for phase. We argue that the phase on Eq. (4.17) already agrees with the convention set forth in Ref. 8, namely, that in the limit $y_n \to -\infty$, the fundamental Racah coefficient should become a square-bracket coefficient of the same labels. This is precisely the case for Eq. (4.17) since the extended fundamental projective coefficient then becomes an ordinary reduced Wigner coefficient which is, in turn, just a fundamental square-bracket coefficient because the following $U(n-1)$ Racah coefficient has value unity (also part of our phase convention):

$$\left\langle\left(\begin{matrix}[y]_{n-1}+\Delta_{n-1}(\rho)\\(\max)\end{matrix}\right)\right| \left(\begin{matrix}\rho\\[1\ \ \dot{0}]_{n-1}\\\rho\end{matrix}\right) \left|\left(\begin{matrix}[y]_{n-1}\\(\max)\end{matrix}\right)\right\rangle ([y]_{n-1}+\Delta_{n-1}(\rho)) = 1. \tag{B14}$$

To complete the proof of Eq. (4.17) we still must demonstrate that the operator pattern assignment (B10) is correct. There is, of course, no question that *there is but a single operator pattern to be determined*. Suppose we had identified the pattern simply by the symbol (Γ_0). Then the Δ pattern $[\Delta(\Gamma_0)]_n$ is still given by Eq. (B9). One easily sees that the net effect of this unspecified identification on our previous proof is to yield the following modified results:

$$\left\langle\left(\begin{matrix}[x]+\Delta(\tau)\\[y]+\Delta(\rho)\end{matrix}\right)\right| \left[\begin{matrix}\tau\\[1\ \ \dot{0}]\\\rho\end{matrix}\right]_{\text{ext}} \left|\left(\begin{matrix}[x]\\[y]\end{matrix}\right)\right\rangle$$

$$= (\text{phase}) \left\{ \begin{pmatrix} [x] + \Delta(\tau) \\ (\Gamma_0') \end{pmatrix} \begin{pmatrix} \tau \\ [1\ \dot{0}] \\ \rho \end{pmatrix} \begin{pmatrix} [x] \\ (\Gamma_0) \end{pmatrix} \right\} ([x] + \Delta(\rho)), \tag{B15}$$

where

$$[\Delta(\Gamma_0)]_{n-1} = [y_1 y_2 \cdots y_{n-1}],$$

$$[\Delta(\Gamma_0')]_{n-1} = [y_1 y_2 \cdots y_{n-1}] + \Delta_{n-1}(\rho).$$

But the limit $y_n \to -\infty$ of the left-hand side of this result is unambiguously given by the square-bracket coefficient

$$\left[\begin{pmatrix} [x]_n + \Delta_n(\tau) \\ [y]_{n-1} + \Delta_{n-1}(\rho) \\ (\max) \end{pmatrix} \begin{pmatrix} \tau \\ [1\ \dot{0}] \\ \rho \end{pmatrix} \begin{pmatrix} [x]_n \\ [y]_{n-1} \\ (\max) \end{pmatrix} \right] ([y]_{n-1} + \Delta_{n-1}(\rho)). \tag{B16}$$

In accordance with our previous discussions[8,2] regarding the identification of operator patterns through the use of limits, we see that the pattern (Γ_0) must be identified as in Eq. (B10) with a similar identification for (Γ_0').

Finally, we must prove Eq. (4. 15) to complete the full proof of the important relation, Eq. (4. 16). This result is, however, now easily proved directly by evaluating the general Racah invariant as given by Eq. (2. 46) in Ref. 8 on the labels indicated, *noting the pattern identification of Eq.* (B10). We omit these details.

We conclude this appendix by noting two subsidiary results.

Equation (B14) is a special case of the general result

$$\left\{ \begin{pmatrix} [m] + [\Delta(\Gamma)] \\ (\max) \end{pmatrix} \begin{pmatrix} (\Gamma) \\ [M] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [m] \\ (\max) \end{pmatrix} \right\} [m] + [\Delta(\Gamma)] = \delta_{(\Gamma)(\Gamma')}, \tag{B17}$$

unless $[m]$ belongs to the null space of the Wigner operator

$$\begin{pmatrix} (\Gamma) \\ [M] \\ \cdot \end{pmatrix}$$

in which case the Racah coefficient is zero even for $(\Gamma) = (\Gamma')$. This result follows easily from the product rule for two Wigner operators [see Eq. (15) of Ref. 9] upon operating on the state

$$\left| \begin{pmatrix} [0] \\ (0) \end{pmatrix} \right\rangle$$

and using the property

$$\left\langle \begin{pmatrix} (\gamma) \\ [m] \\ (m) \end{pmatrix} \right| \left| \begin{pmatrix} [0] \\ (0) \end{pmatrix} \right\rangle = \delta_{(\gamma)(\max)} \left| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle. \tag{B18}$$

The phase of Eq. (B11) can be obtained in the following manner: Since the phase in Eq. (B12) has been demonstrated to be + 1, we may choose $\delta_{\tau\rho}([x];[y]) = 0$. Furthermore, we see that $\delta([x];[x]) = 0$. Equation (B13b) now becomes a recursion formula for generating $\delta([x];[y])$. The result of iterating this equation is

$$\delta([x];[y]) = \sum_{i=1}^{n} (n-i)(x_i - y_i). \tag{B19}$$

*Supported in part by the U. S. Army Research Office (Durham), the National Science Foundation, and the U. S. Atomic Energy Commission.

[1]L. C. Biedenharn, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. **13,** 1957 (1972).
[2]L. C. Biedenharn and J. D. Louck, J. Math. Phys. **13,** 1985 (1972).
[3]L. C. Biedenharn. A. Giovannini, and J. D. Louck, J. Math. Phys. **8,** 691 (1967).
[4]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8,** 89 (1968).
[5]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5,** 1723 (1964); J. Math. Phys. **5,** 1730 (1964); J. Math. Phys. **6,** 1847 (1965).
[6]L. C. Biedenharn, in *Spectroscopic and Group Theoretical Methods in Physics (Racah Memorial Volume),* edited by F. Block *et al.* (North-Holland, Amsterdam, 1968), p. 59.
[7]J. D. Louck, Am. J. Phys. **38,** 3 (1970).
[8]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11,** 2368 (1970).
[9]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **12,** 173 (1971).
[10]W. J. Holman and L. C. Biedenharn, in *Group Theory and Its Applications,* Vol. II, edited by E. M. Loebl (Academic, New York, 1971), p. 1.
[11]W. J. Holman, Nuovo Cimento A **4,** 904 (1971).
[12]I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR Mekh. Mashinostr. **29,** 1329 (1965) [Am. Math. Soc. Transl. **65,** 116 (1967), ser. 2].
[13]See Refs. 1, 2, 7, 8, and 9 for a complete explanation of the notations used in the present paper.
[14]H. Weyl, *The Structure and Representations of Continuous Groups,* Lectures at the Inst. for Adv. Study, Princeton, N. J., 1934-35 (notes by R. Brauer). Some of this material also appears in: H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, N. J., 1946).
[15]I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71,** 825 (1950).
[16]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4,** 1499 (1963). A purely algebraic construction was also sketched in this paper.
[17]For a more complete set of references relating to the development of the boson polynomials as well as the representations of the Lie algebra of $U(n)$, see the review articles, Refs. 7 and 10.
[18]In Ref. 16, the polynomials $B([m]^{\max}_{\text{semimax}})$ were constructed in closed form in this way.
[19]Further explicit results using combinatorial methods were obtained for $U(3)$ and $U(4)$ by M. Ciftan and L. C. Biedenharn, J. Math. Phys. **10,** 221 (1969); M. Ciftan, J. Math. Phys. **10,** 1635 (1969).
[20]J. G. Nagel and M. Moshinsky, J. Math. Phys. **6,** 682 (1965).
[21]J. D. Louck, J. Math. Phys. **6,** 1786 (1965).
[22]Note that this decomposition fails for the set of $(n+1) \times (n+1)$ nonsingular matrices whose first n rows and columns form a singular $n \times n$ matrix. These exceptional submanifolds may, however, simply be ignored (cf. Ref. 12).
[23]$[m]$ always designates a row vector, (m) and (m') Gel'fand triangular arrays. It is usually possible to deduce the "size" of these symbols from the context. However, it occasionally becomes necessary to indicate this explicitly, in which case $[m]_k$ designates the row of length k. In Eq. (2.25a) the symbol $[m']_{n-1}$ denotes row $n-1$ of the pattern (m') which has an implied number of rows equal to $n-1$. Despite the occasional confusion these conventions may cause, it seems superior to the alternative of having a profusion of subscripts. As a further example, we note that $\begin{pmatrix} [m']_{n-1}\ 0 \\ (m') \end{pmatrix}$ denotes a Gel'fand array of n rows comprised of the $n-1$ rows of (m') and an nth row which is itself made up of the top row of (m') and 0.
[24]E. Chacón, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13,** 577 (1972).
[25]S. J. Ališauskas, A.-A. A. Jucys, and A. P. Jucys, J. Math. Phys. **13,** 1349 (1972).
[26]This requires use of the orthogonality of Wigner operators expressed by Eq. (2.14) of Ref. 8.

[27]T. Regge, Nuovo Cimento **10,** 544 (1958).
[28]A. M. Bincer, J. Math. Phys. **11,** 1835 (1970).
[29]E. P. Wigner, Z. Phys. **43,** 624 (1927); Z. Phys. **45,** 601 (1927).
[30]This point was brought to the attention of one of us (JDL) during a series of lucid lectures given at the Los Alamos Scientific Laboratory by Professor Gian-Carlo Rota in May of 1972 and taken from: *On the Combinatorics of the Symmetric Group* by P. Doubilet and Gian-Carlo Rota (to be published by the American Math. Society for the Conference Board of the Mathematical Sciences). We both wish to thank Professor Rota for the benefit of several stimulating conversations.
[31]D. E. Knuth, Pac. J. Math. **34,** 709 (1970).
[32]A very readable account of the application of these concepts to symmetric function theory and other problems has been given by: R. P. Stanley, Studies in Applied Math. **50,** 1967 (1971); Studies in Applied Math. **50,** 259 (1971). (We are indebted to Professor Rota for bringing this reference and Ref. 31 to our attention.)
[33]G. de B. Robinson, *Representations Theory of the Symmetric Group* (Univ. of Toronto Press, Toronto, 1961).
[34]We are indebted to Dr. Eeltje de Vries for his remark, and for helpful discussions.
[35]M. Moshinsky, J. Math. Phys. **7,** 691 (1966).
[36]A. Young, Proc. Lond. Math. Soc. **28,** 255 (1928); Proc. Lond. Math. Soc. **31,** 253 (1930); Proc. Lond. Math. Soc. **3,** 273 (1930).
[37]A. Young, Proc. Lond. Math. Soc. **34,** 196 (1932).
[38]T. Yamanouchi, Proc. Phys.-Math. Soc. Jap. **18,** 623 (1936); Proc. Phys.-Math. Soc. Jap. **19,** 436 (1937).
[39]R. M. Thrall, Duke Math. J. **8,** 611 (1941).
[40]Cf. M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), pp. 214–223 and Refs. cited there.
[41]G. Racah, Phys. Rev. **62,** 438 (1942).
[42]The factorization lemma appears not to lead easily to a corresponding result for an arbitrary permutation τ.
[43]L. C. Biedenharn, J. Math. Phys. **31,** 287 (1953).
[44]J. P. Elliott, Proc. Roy. Soc. Lond. A **218,** 370 (1953).
[45]For economy of expression we use the following notation in this section and Appendix B: $[x] = [x_1 x_2 \cdots x_n]$, where $x_i = m_{in}$, and $[y] = [y_1 y_2 \cdots y_n]$, where $y_i = m_{i\,n-1}$.
[46]In all honesty we must admit that we have not yet fully established the validity of conjecture 3 (stated in I) for all $U(3)$ Wigner operators. It appears that this may become possible with the complete determination of the $U(3)$ fundamental Racah coefficients.

Scattering by a spin-dependent spheroidal potential*

Ming Chiang Li

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
(Received 1 December 1972; revised manuscript received 5 March 1973)

The scattering by a spin-dependent spheroidal potential is studied. For this potential the scattering amplitude is considerably simple. The simplification is accomplished through the use of "reccurence relations" among the spheroidal functions. An iteration procedure is used for solving the infinitely coupled integral equations of the radial functions. The convergence of the solutions is proved.

1. INTRODUCTION

The importance of spheroidal scattering in classical physics and quantum mechanics was pointed out in two previous papers.[1] Practically, it is a physical problem which one has to deal with in the acoustic and electromagnetic scattering[2] by circular disks and apertures, and in the scattering of slow electrons by diatomic molecules.[3] Physically, it denotes the scattering between a particle and the simplest composite system which can be formed by two particles with a fixed separation. It also denotes the simplest nonspherical scattering.

In the existing spheroidal scattering theory, one often neglects the vector character of the incident electromagnetic waves, or the spin of the incident particles. The reason for such an approach is the mathematical difficulties inherent to these problems. With the inclusion of these features, the wave equation is no longer separable in spheroidal coordinates[4]; and the solutions have to be expressed in terms of infinitely coupled integral equations. The importance of spheroidal potentials in nuclear physics was noticed quite early. Rainwater, Granger, and Spence[5] pointed out the necessity for introducing such a potential in the determination of nuclear energy levels. The spheroidal potential also plays a role in the droplet model of elementary particles. Without the inclusion of the spin contribution, the spheroidal scattering theory only has a very limited applicability in particle and nuclear physics.

In the present paper, the spheroidal scattering by a simple spin dependent potential is discussed. In Sec. 2, we discuss the relation between the proposed potential and the spin-dependent potential from the relativistic Dirac equation, and the spheroidal nuclear potential. In Sec. 3, the integral equations for the scattered radial wave functions are given. They are infinitely coupled. A simplification of these equations is accomplished through the study of the 'recurrence relations' for the spheroidal functions in Sec 4. The scattering amplitude is expressed in terms of radial wave functions in Sec. 5. In Sec. 6, an iteration procedure is used for solving the coupled integral equations of the radial function; and the convergence of the solutions is proved.

2. THE POTENTIAL

A spin $\frac{1}{2}$ particle with mass m and incident momentum $\mathbf{k}$ is scattered by a prolate spheroidal potential $V(\mathbf{r})$

$$V(\mathbf{r}) = V_1(\mathbf{r}) + \boldsymbol{\sigma}\cdot\mathbf{L}V_2(\mathbf{r}) \equiv \frac{2\hbar^2}{md^2}\left(\frac{U_1(\xi)}{\xi^2-\eta^2} + \boldsymbol{\sigma}\cdot\mathbf{L}\,\frac{U_2(\xi)}{\xi^2-\eta^2}\right). \tag{2.1}$$

The potentials $V_1(\mathbf{r})$ and $V_2(\mathbf{r})$ are short ranged, namely

$$U_1(\xi) \to O(\xi^{-\epsilon}),$$

$$U_2(\xi) \to O(\xi^{-\epsilon}) \quad \text{for } \xi \to \infty, \tag{2.2}$$

where ϵ is a positive number. The prolate spheroidal coordinates of the position vector $\mathbf{r}$ are the variables ξ, η, ϕ with $(1 \le \xi \le \infty, -1 \le \eta \le 1, 0 \le \phi \le 2\pi)$ and d the interfocal distance for the prolate spheroidal coordinate system. $\boldsymbol{\sigma}$ is the Pauli spin operator and $\mathbf{L}$ the orbital angular momentum. The first term in Eq. (2.1) denotes the static spheroidal potential, which was used by Stier and Fisk[3] to describe the scattering of slow electrons by diatomic molecules. It has been used repeatedly by many other workers. In nuclear physics such a potential has been used by Rainwater, Granger, and Spence[5] in the determination of nuclear energy levels. The second term in Eq. (2.1) denotes the spin-orbital spheroidal potential. In the limit, when the interfocal distance d becomes zero, and $\frac{1}{2}d\xi$ takes on a finite value, the nonspherical nature of the spheroidal potential in Eq. (2.1) disappears. The resultant spherical spin-orbital potential has been widely used to describe the interaction of an electron with its orbital angular momentum in an atom, and the interaction of a scattered nucleon by atomic nuclei. For the electric interaction, the spherical spin-orbital potential is a consequence of the relativistic Dirac equation and is related to the static potential. In nuclear interactions, the spherical spin-orbital potential is phenomenologically introduced, and its presence still can not be accounted for in terms of a completely satisfactory theoretical explanation.[6] This potential is not of the first order in the velocity and is present even in the nonrelativistic approximation. The recognition of its existence led to very significant advances in the theory of nuclear structure.

It is interesting to compare the nonspherical potential in Eq. (2.1) and the electric potential in the presence of spin. The later has the form

$$V_e(\mathbf{r}) = V_1(\mathbf{r}) + (e/4m^2)\boldsymbol{\sigma}\cdot\nabla V_1(\mathbf{r}) \times \mathbf{p}, \tag{2.3}$$

where $\mathbf{p}$ is the momentum operator of an electron and the velocity of light is taken as unity. The two potentials in Eqs. (2.1) and (2.3) differ from each other but only in a specific region. For a spheroidal potential $V_1(\mathbf{r})$, the spin-dependent potentials in Eqs. (2.1) and (2.3) have the same singularity behavior at $\xi = 1$, and $\eta = \pm 1$ with the form

$$O\{[(\xi^2-1)(1-\eta^2)]^{1/2}(\xi^2-\eta^2)^{-3}\}. \tag{2.4}$$

They are essentially equivalent in the region $\xi > 1$ which is away from singularities. The region, for which they are different, is between both singularities. The choice of the potential $V_2(\mathbf{r})$ is able to make this discrepancy small.

A rigorous treatment of the scattering problem with a spheroidal potential given in Eq. (2.3) is, at the present, beyond our theoretical capabilities. Because of the small discrepancy between Eq. (2.1) and Eq. (2.3), a

study of the scattering problem with the potential in Eq. (2.1) will give some insight into the scattering with potential (2.3).

The nuclear interaction between two nucleons is another interesting problem. The interaction, one often thinks, is mediated by a pion exchange. The Yukawa potential is associated with such an exchange. It has an exponential tail at large distance and an r^{-1} singularity at the origin. The Coulomb potential for a single charged particle has a similar singularity behavior. It is well known that the Coulomb potential produced by two charged particles is spheroidal symmetric. Now an interesting consequence occurs, if two nucleons are separated by a small distance, which is less than the range of the Yukawa potential. For instance, let these nucleons be located at coordinates $\mathbf{r}_1' = (0, 0, d/2)$ and $\mathbf{r}_2' = (0, 0, -d/2)$. The combined potential $V_p(\mathbf{r})$ may be written as

$$V_p(\mathbf{r}) = \frac{g}{r_1} e^{-\mu r_1} + \frac{g}{r_2} e^{-\mu r_2} = \frac{4g\xi \exp[-(\mu d\xi/2)]}{d(\xi^2 - \eta^2)} \left\{ 1 + O\left[\left(\frac{\mu d}{2}\eta\right)^2\right]\right\}, \quad (2.5)$$

where g is a strength constant and μ^{-1} the range of Yukawa potential. This potential is not spherical symmetric, but is predominantly spheroidal symmetric.

The importance of a spheroidal potential in nuclear physics is illustrated by other considerations. Many nuclei have large nuclear quadrupole moments. This indicates that the basic shape for these nuclei is not spherical, but corresponds to a considerable distortion of the whole nucleus into a spheroidal shape. On the basis of a single-particle model, each nucleon inside these nuclei would move in a spheroidal potential. However, the difficulties in handling the spheroidal scattering problem has prevented the full development of the above considerations. The basic difficulty is created by the presence, even in the nonrelativistic approximation, of the spin-orbit term. Firstly, there are not firm theoretical criteria that will determine the exact form of this term.[6] Secondly, were the spin-orbit interaction known exactly, its very nature would prevent mathematical tractability of the problem. In spite of these difficulties, the importance of the spheroidal potential in nuclear physics cannot be ignored. The present attempt, with its simplifications, may be viewed as a natural effort to study spheroidal nuclei.

3. SCHRÖDINGER EQUATION

The Schrödinger equation for describing a spin $\frac{1}{2}$ particle scattered by the spheroidal potential in Eq. (2.1) has the form

$$-(\hbar^2/2m)\nabla^2\psi + [V_1(\mathbf{r}) + \boldsymbol{\sigma}\cdot\mathbf{L}V_2(\mathbf{r})]\psi = (\hbar^2k^2/2m)\psi, \quad (3.1)$$

where ψ is a two-component spinor. The incident particle with momentum $\mathbf{k}$ moves in the direction $\mathbf{n}_0 = \mathbf{k}/k$, specified by the polar angle $\theta_0 \equiv \cos^{-1}\eta_0$ and azimuthal angle ϕ_0. The incident wave can be expressed in terms of the spheroidal functions[4] $S_{mn}(c, \eta)$ and $R_{mn}^{(1)}(c, \xi)$

$$e^{ikr\cos\Theta}|\chi_i\rangle = 2\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} \frac{i^n}{N_{mn}(c)} S_{mn}(c, \eta_0) \times S_{mn}(c, \eta)R_{mn}^{(1)}(c, \xi)e^{im(\phi-\phi_0)}|\chi_i\rangle, \quad (3.2)$$

where

$$c = \tfrac{1}{2}kd,$$

$$\int_{-1}^{1} S_{mn}(c, \eta)S_{mn'}(c, \eta)d\eta = \delta_{nn'}N_{mn}(c),$$

$$\cos\Theta = \cos\theta\cos\theta_0 + \sin\theta\sin\theta_0\cos(\phi - \phi_0), \quad (3.3)$$

$N_{mn}(c)$ is the normalization constant for the function $S_{mn}(c, \eta)$, Θ is the angle between the position vector $\mathbf{r}$ and momentum vector $\mathbf{k}$. The spinor $|\chi_i\rangle$ denotes the spin polarization of the initial spin direction. It is well known from the general scattering theory that the scattered wave with the final spin polarization χ_f satisfies the integral equation

$$\begin{aligned}\psi_k^{(+)}(\mathbf{r}; \chi_f, \chi_i) &= e^{ikr\cos\Theta}\langle\chi_f|\chi_i\rangle \\ &- \frac{m}{2\pi\hbar^2}\sum_h \int \frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} \\ &\times \langle\chi_f|V(\mathbf{r}')|\chi_h\rangle\psi_k^{(+)}(\mathbf{r}'; \chi_h, \chi_i)d\mathbf{r}' \\ &= e^{ikr\cos\Theta}\langle\chi_f|\chi_i\rangle \\ &- \frac{m}{2\pi\hbar^2}\int \frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} V_1(\mathbf{r}')\langle\chi_f|\chi_i\rangle \\ &\times \psi_k^{(+)}(\mathbf{r}'; \chi_f, \chi_i)d\mathbf{r}' \\ &+ \frac{m}{2\pi\hbar^2}\sum_h \int \left\langle\chi_f|\boldsymbol{\sigma}\cdot\mathbf{L}'\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}|\chi_h\right\rangle \\ &\times V_2(\mathbf{r}')\psi_k^{(+)}(\mathbf{r}'; \chi_h, \chi_i)d\mathbf{r}' \\ &+ i\frac{m}{2\pi\hbar}\sum_h \int \left\langle\chi_f|\boldsymbol{\sigma}\cdot(\mathbf{r}'\times\mathbf{n}')\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}|\chi_h\right\rangle \\ &\times V_2(\mathbf{r}')\psi_k^{(+)}(\mathbf{r}'; \chi_h, \chi_i)ds', \end{aligned} \quad (3.4)$$

where $\mathbf{n}'$ is the unit vector in the normal direction to the surface element ds', and S is a surface at infinity. In arriving at the final form of Eq. (3.4), the Green's divergence theorem has been used. The differentiation in the third term of Eq. (3.4) is only with respect to the Green's function $|\mathbf{r}-\mathbf{r}'|^{-1}\exp(ik|\mathbf{r}-\mathbf{r}'|)$. The last term in Eq. (3.4) is an integration over the surface at infinity S. On account of the short-range nature of the potential $V_2(\mathbf{r})$ this term is zero.

The spheroidal potentials $V_1(\mathbf{r})$ and $V_2(\mathbf{r})$ both have a pair of singularities at $\xi = 1, \eta = 1$ and $\xi = 1, \eta = -1$, respectively. The spherical potential has only one singularity which is located at the origin. This is the characteristic difference between these two types of potential. Such a difference has prevented the approximation of the spheroidal potential by any noncentral form of a spherical potential, and therefore excluded the application of the standard procedure for handling spherical scatter problems.

The Schrödinger equation in Eq. (3.1) is not separable in the spheroidal coordinates. Thus the spheroidal phase shift analysis method is not directly applicable here. An important step in this analysis is the proper handling of the pair of singularities of the spheroidal potential. This procedure is the only one known capable of handling such singularities and will be adopted in the present problem.

The scattered wave $\psi_k^{(+)}(\mathbf{r}; \chi_f, \chi_i)$ is expanded in terms of prolate spheroidal functions

$$\psi_k^{(+)}(\mathbf{r};\chi_f,\chi_i) = 2\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty}\frac{i^n}{N_{mn}(c)}S_{mn}(c,\eta) \times T_{mn}(c,\xi;\mathbf{n}_0;\chi_f,\chi_i)e^{im\phi}. \quad (3.5)$$

The prolate spheroidal expansion of the free-space Green's function is written as

$$\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|} = \frac{ik}{2\pi}\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty}\frac{1}{N_{mn}(c)} \times R_{mn}^{(1)}(c,\xi_<)R_{mn}^{(3)}(c,\xi_>)S_{mn}(c,\eta) \times S_{mn}(c,\eta')e^{im(\phi-\phi')}, \quad (3.6)$$

where

$$R_{mn}^{(1)}(c,\xi_<)R_{mn}^{(3)}(c,\xi_>) = \begin{cases} R_{mn}^{(1)}(c,\xi)R_{mn}^{(3)}(c,\xi'), & \xi<\xi' \\ R_{mn}^{(1)}(c,\xi')R_{mn}^{(3)}(c,\xi), & \xi>\xi' \end{cases}. \quad (3.7)$$

The vector $\mathbf{r}'$ has the spheroidal coordinates ξ',η', and ϕ'. Equations (3.2), (3.5), and (3.6) permit us to express the integral equation Eq. (3.4) in the following form:

$$T_{mn}(c,\xi;\mathbf{n}_0;\chi_f,\chi_i) = S_{mn}(c,\eta_0)e^{-im\phi_0}R_{mn}^{(1)}(c,\xi)\langle\chi_f|\chi_i\rangle - ic\int_1^{\infty}d\xi' R_{mn}^{(1)}(c,\xi_<)R_{mn}^{(3)}(c,\xi_>)U_1(\xi') \times T_{mn}(c,\xi';\mathbf{n}_0;\chi_f,\chi_i) + \sum_j\sum_k\langle\chi_f|\sigma_j|\chi_k\rangle W_{mn}^j(c,\xi;\mathbf{n}_0;\chi_k,\chi_i), \quad (3.8)$$

where

$$W_{mn}^j(c,\xi;\mathbf{n}_0;\chi_k,\chi_i) = \frac{ic}{(2\pi)^2}(-i)^n\int_{-1}^{1}d\eta\int_0^{2\pi}d\phi S_{mn}(c,\eta)e^{-im\phi} \times \sum_{m'=-\infty}^{\infty}\sum_{n'=|m'|}^{\infty}\sum_{m''=-\infty}^{\infty}\sum_{n''=|m''|}^{\infty}i^{n'}\frac{1}{N_{m''n''}(c)}\frac{1}{N_{m'n'}(c)} \times \int_1^{\infty}d\xi'\int_{-1}^{1}d\eta'\int_0^{2\pi}d\phi' S_{m'n'}(c,\eta')e^{im'\phi'} \times T_{m'n'}(c,\xi';\mathbf{n}_0;\chi_k,\chi_i)U_2(\xi') \times L'^jR_{m''n''}^{(1)}(c,\xi_<)R_{m''n''}^{(3)}(c,\xi_>) \times S_{m''n''}(c,\eta)S_{m''n''}(c,\eta')e^{im''(\phi-\phi')}, \quad \text{for } j=+,-,0. \quad (3.9)$$

The spin matrices and angular momentum operators in Eqs. (3.9), (3.10), and (3.11) are defined in the usual fashion

$$\sigma_{\pm} = \tfrac{1}{2}(\sigma_x \pm i\sigma_y), \quad \sigma_0 = \sigma_z, \qquad L'^{\pm} = L'_x \pm iL'_y, \quad L'^0 = L'_z. \quad (3.10)$$

The integrations in Eq. (3.9) with respect to the variables η,ϕ,η', and ϕ' will be discussed in the next section and are performed by using the "recurrence relations" for the spheroidal functions.

4. "RECURRENCE RELATIONS"

The "recurrence relations" for the spheroidal functions have been studied by Marx and Chako.[7] The most well-known recurrence relations are those among the three contiguous hypergeometric or confluent hypergeometric functions. The differential equations satisfied by the spheroidal functions has an irregular point at infinity in addition to regular points for the values ± 1 of the respective variables, and is not a hypergeometric type. The coefficients in the "recurrence relations" for the spheroidal functions can not be reduced to simple expressions in the variable and in the parameters, as is the case for functions of the hypergeometric type. These coefficients involve products of the functions themselves. The relations are obtained through a method suggested by Whittaker[8] in his derivation of the "recurrence relations" for Mathieu functions, which are also not hypergeometric. In that method an integral representation of the spheroidal radial functions $R_{mn}^{(i)}(c,\xi)$ in terms of the spheroidal angle functions $S_{mn}(c,\xi)$ is used

$$R_{mn}^{(i)}(c,\xi) = \int_a^b K_m(\eta,\xi)S_{mn}(c,\eta)d\eta. \quad (4.1)$$

The function $K_m(\eta,\xi)\exp(-im\phi)$ satisfies Helmholtz's equation, provided that

$$(1-\eta^2)\left(S_{mn}(c,\eta)\frac{\partial}{\partial\eta}K_m(\xi,\eta) - K_m(\xi,\eta)\frac{\partial}{\partial\eta}S_{mn}(c,\eta)\right) \quad (4.2)$$

takes the same value at both ends of the integration limit a and b. The detailed derivation of "recurrence relations" might be found in the papers by Marx and Chako.[7] We will give a brief discussion of these "recurrence relations," which are related to the evaluation of the integrations in Eq. (3.9), in the Appendix.

The angular momentum operators in the spheroidal coordinates have the forms

$$L^{\pm} = \pm\hbar\left[\frac{\rho}{\xi^2-\eta^2}\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right)e^{\pm i\phi} \pm \frac{ie^{\pm i\phi}}{\rho}\xi\eta\frac{\partial}{\partial\phi}\right.,$$

$$L_0 = -i\hbar\frac{\partial}{\partial\phi}, \quad (4.3)$$

where

$$\rho = [(\xi^2-1)(1-\eta^2)]^{1/2}. \quad (4.4)$$

The "recurrence relations" are expressed as

$$\hbar A_{mn,n'}^{(i,\pm)}R_{mn}^{(i)}(c,\xi)e^{-im\phi} = \int_{-1}^{1}d\eta S_{mn}(c,\eta)L^{\pm}R_{m\pm1,n'}^{(i)}(c,\xi) \times S_{m\pm1,n'}(c,\eta)e^{-i(m\pm1)\phi}, \quad \text{for } i=1,2,3,4. \quad (4.5)$$

The constants quantities $A_{mn,n'}^{(i,\pm)}$ are determined from the asymptotic behaviors of the spheroidal radial functions $R_{mn}^{(i)}(c,\xi)$ along with their relations

$$R_{mn}^{(1)}(c,\xi) = \tfrac{1}{2}[R_{mn}^{(3)}(c,\xi) + R_{mn}^{(4)}(c,\xi)], \qquad R_{mn}^{(2)}(c,\xi) = (1/2i)[R_{mn}^{(3)}(c,\xi) - R_{mn}^{(4)}(c,\xi)]. \quad (4.6)$$

One obtains thus

$$A_{mn,n'}^{(i,\pm)} = e^{i(n-n')\pi/2}\int_{-1}^{1}d\eta S_{mn}(c,\eta) \times\left(\frac{(m\pm1)\eta}{(1-\eta^2)^{1/2}}S_{m\pm1,n'}(c,\eta) \mp (1-\eta^2)^{1/2}\frac{d}{d\eta}S_{m\pm1,n'}(c,\eta)\right), \quad \text{for } n-n' = \text{even}, \quad (4.7)$$

$$A_{mn,n'}^{(i,\pm)} = 0, \quad \text{for } n-n' = \text{odd}. \quad (4.8)$$

Equations (4.7) and (4.8) express the fact that the constants $A^{(i,\pm)}_{mn,n}$ do not depend on the index i, which denotes the kind of the spheroidal radial function $R^{(i)}_{mn}(c,\xi)$ in Eq. (4.5). For simplicity we write

$$A^{\pm}_{mn,n} = A^{(i,\pm)}_{mn,n'}. \tag{4.9}$$

After tedious but straightforward manipulations, the "recurrence relations" in Eq. (4.5) yield the following set of equations:

$$\begin{aligned}&\int_{-1}^{1} d\eta \int_{-1}^{1} d\eta' \int_0^{2\pi} d\phi \int_0^{2\pi} d\phi' S_{mn}(c,\eta)e^{-im\phi} \\ &\quad\times S_{m'n'}(c,\eta')e^{im'\phi'}L'^{\pm}R^{(i)}_{m''n''}(c,\xi') \\ &\quad\times S_{m''n''}(c,\eta)S_{m''n''}(c,\eta')e^{im''(\phi-\phi')} \\ &= (2\pi)^2\hbar\phi_{m'',\,m'\pm1}\delta_{mm''}\delta_{nn''}A^{\pm}_{m'n',n}N_{mn}(c)R^{(i)}_{m'n'}(c,\xi')\end{aligned} \tag{4.10}$$

In obtaining Eq. (4.10) one also uses the orthogonality relations in Eq. (3.3) of the spheroidal angle functions $S_{mn}(c,\eta)$.

The above integrations appear in Eq. (3.9) and have been performed through the utilization of the "recurrence relations." The other integrations in Eq. (3.9) are similar to these, with the operators $L'^{\pm}$ replaced by L'^{0}. The evaluation of these integrations is simpler and is accomplished by means of the orthogonality relations in Eq. (3.3):

$$\begin{aligned}&\int_{-1}^{1} d\eta \int_{-1}^{1} d\eta' \int_0^{2\pi} d\phi \int_0^{2\pi} d\phi' S_{mn}(c,\eta)e^{-im\phi}S_{m'n'}(c,\eta')e^{im'\phi'} \\ &\quad\times L'^{0}R^{(i)}_{m''n'}(c,\xi')S_{m''n''}(c,\eta)S_{m''n''}(c,\eta')e^{im''(\phi-\phi')} \\ &= -(2\pi)^2 m\hbar\delta_{nn'}\delta_{nn''}\delta_{mm'}\delta_{mm''}N^2_{mn}(c)R^{(i)}_{m'n'}(c,\xi').\end{aligned} \tag{4.11}$$

By using the above relations the functions $W^{j}_{mn}(c,\xi;\mathbf{n}_0;\chi_h,\chi_i)$ in Eq. (3.9) can be written as

$$\begin{aligned}&W^{j}_{mn}(c,\xi;\mathbf{n}_0;\chi_h,\chi_i) \\ &\quad= ic\sum_{m'=-\infty}^{\infty}\sum_{n'=|m'|}^{\infty}\int_1^{\infty} d\xi' R^{(1)}_{mn}(c,\xi_<)R^{(3)}_{m'n'}(c,\xi_>)U_2(\xi') \\ &\quad\times D^{j}_{mn,m'n'}T_{m'n'}(c,\xi';\mathbf{n}_0;\chi_h,\chi_i),\end{aligned} \tag{4.12}$$

where

$$D^{\pm}_{mn,\,m'n'} = (-i)^{n-n'}\hbar N^{-1}_{m'n'}(c)\delta_{m,m'\pm1}A^{\pm}_{m'n',n} \tag{4.13}$$

and

$$D^{0}_{mn,m'n'} = -m\hbar\delta_{mm'}\delta_{nn'}. \tag{4.14}$$

The recurrence relations among the associated Legendre functions have played an essential role in arriving at the simple form of the spherical scattering formulation. Due to their complexity, the "recurrence relations" among the spheroidal angle functions only play a limited role. These differences may be seen as follows. Each of the quantities $D^{j}_{mn,\,m'n'}$ contain at least a Kronecker delta. Then, only one of the two infinite sums is present in Eq. (4.12). In the spherical case, i.e., $c\to 0$, each of the quantities $D^{j}_{mn,\,m'n'}$ contains two Kronecker deltas. The additional one comes from the "recurrence" coefficients $A^{\pm}_{mn,\,n'}$,

$$\begin{aligned}A^{-}_{mn,\,n'} &\xrightarrow{c\to 0} \frac{2}{2n+1}\frac{(n+m)!}{(n-m)!}\delta_{nn'}, \\ A^{+}_{mn,\,n'} &\xrightarrow{c\to 0} \frac{2}{2n+1}\frac{(n+m+1)!}{(n-m-1)!}\delta_{nn'}.\end{aligned} \tag{4.15}$$

That is to say that in the spherical case the infinite sum in Eq. (4.12) is not present and the scattering formulation is greatly simplified.

5. SCATTERING AMPLITUDE

The radial functions $T_{mn}(c,\xi;\chi_f,\chi_i)$ of the scattering wave satisfy the coupled integral equations in Eq. (3.8). For the scattering problem one has to find their asymptotic forms and the scattering amplitude. Suppose these functions have the following asymptotic forms:

$$\begin{aligned}&T_{mn}(c,\xi;\mathbf{n}_0;\chi_f,\chi_i) \xrightarrow{c\xi\to\infty} \\ &S_{mn}(c,\eta_0)e^{-m\phi_0}(1/c\xi)\cos[c\xi-\tfrac{1}{2}(n+1)\pi]\langle\chi_f|\chi_i\rangle \\ &+(1/c\xi)\exp i[c\xi-\tfrac{1}{2}(n+1)\pi]K_{mn}(\mathbf{n}_0;\chi_f,\chi_i).\end{aligned} \tag{5.1}$$

Quantities $S_{mn}(c,\eta)$ and $K_{mn}(\chi_f,\chi_i)$ are obtained through Eq. (3.8) with the aid of the asymptotic forms[4] of the spheroidal radial function $R^{(1)}_{mn}(c,\xi)$ and $R^{(3)}_{mn}(c,\xi)$:

$$R^{(1)}_{mn}(c,\xi) \xrightarrow{c\xi\to\infty} (1/c\xi)\cos[c\xi-\tfrac{1}{2}(n+1)\pi], \tag{5.2}$$

$$R^{(3)}_{mn}(c,\xi) \xrightarrow{c\xi\to\infty} (1/c\xi)\exp i[c\xi-\tfrac{1}{2}(n+1)\pi]. \tag{5.3}$$

These constants can be expressed as

$$\begin{aligned}&K_{mn}(\mathbf{n}_0;\chi_f,\chi_i) \\ &\quad= -ic\int_1^{\infty} d\xi' R^{(1)}_{mn}(c,\xi')U_1(\xi')T_{mn}(c,\xi';\mathbf{n}_0;\chi_f,\chi_i) \\ &\quad+ ic\sum_j\sum_h\langle\chi_f|\sigma_j|\chi_h\rangle\sum_{m'=-\infty}^{\infty}\sum_{n'=|m'|}^{\infty}\int_1^{\infty} d\xi' R^{(1)}_{m'n'}(c,\xi')U_2(\xi') \\ &\quad\times D^{j}_{mn,\,m'n'}T_{m'n'}(c,\xi';\mathbf{n}_0;\chi_h,\chi_i).\end{aligned} \tag{5.4}$$

The scattering amplitude $f_k(\mathbf{n},\mathbf{n}';\chi_f,\chi_i)$ is related to the asymptotic form of the scattered wave $\psi^{(+)}_k(\mathbf{r};\chi_f,\chi_i)$.

$$\psi^{(+)}_k(\mathbf{r};\chi_f,\chi_i) \xrightarrow{r\to\infty} e^{i\mathbf{k}\cdot\mathbf{r}} + (1/r)e^{ikr}f_k(\mathbf{n},\mathbf{n}_0;\chi_f,\chi_i), \tag{5.5}$$

where the unit vector $n = \mathbf{r}/r$ denotes the direction for the particle after scattering. Using Eqs. (3.5), (5.1), and (5.2) we obtain

$$\begin{aligned}&f_k(\mathbf{n},\mathbf{n}_0;\chi_f,\chi_i) \\ &\quad= \frac{2}{ik}\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty}N^{-1}_{mn}(c)S_{mn}(c,\eta)e^{im\phi}K_{mn}(\mathbf{n}_0;\chi_f,\chi_i).\end{aligned} \tag{5.6}$$

The above expression of the scattering amplitude involves four summations with respect to indices m and n. Two of these are explicit. The other two are from the expression in Eq. (5.4) for the quantity $K_{mn}(\chi_f,\chi_i)$. These four summations can be reduced to two by noting the following representations for the "recurrence" constants $A^{\pm}_{mn,\,n'}$ and the normalization constant $N_{mn}(c)$:

$$\begin{aligned}A^{\pm}_{mn,\,n'}\delta_{m\pm1,\,m'} &= \frac{i^{n-n'}}{2\pi\hbar}\int_{-1}^{1} d\eta\int_0^{2\pi} d\phi S_{mn}(c,\eta)e^{im\phi} \\ &\quad\times L^{\pm}_S S_{m'n'}(c,\eta)e^{-im'\phi},\end{aligned} \tag{5.7}$$

$$\begin{aligned}mN_{mn}(c)\delta_{mm'}\delta_{nn'} &= \frac{-1}{2\pi\hbar}\int_{-1}^{1} d\eta\int_0^{2\pi} d\phi S_{mn}(c,\eta)e^{im\phi} \\ &\quad\times L^{0}_S S_{m'n'}(c,\eta)e^{-im'\phi},\end{aligned} \tag{5.8}$$

where $L^{\pm}_S$ and L^{0}_S are the angular momentum operators $\mathbf{L}$ in the spherical coordinates with the polar angle $\theta =$

$\cos^{-1}\eta$ and azimuth angle ϕ. From Eqs. (5.7) and (5.8), we rewrite quantities $D^{j}_{mn,\,m'n'}$ in Eqs. (4.13) and (4.14) as follows:

$$D^{j}_{mn,\,m'n'} = [2\pi N_{m'n'}(c)]^{-1}\int_{-1}^{1} d\eta \int_{0}^{2\pi} d\phi \times S_{m'n'}(c,\eta)e^{im'\phi}L_S^j S_{mn}(c,\eta)e^{-im\phi}. \quad (5.9)$$

A direct evaluation of the integration with respect to angle ϕ will lead from Eqs. (5.7), (5.8), and (5.9) to Eqs. (4.7), (3.3), and (4.13), respectively. The spheroidal angle functions satisfy the completeness relation

$$\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty}[2\pi N_{mn}(c)]^{-1}S_{mn}(c,\eta)S_{mn}(c,\eta')e^{im(\phi-\phi')} = \delta(\eta-\eta')\delta(\phi-\phi'). \quad (5.10)$$

Substituting Eqs. (5.4), (5.5), (5.9), and (5.10) into Eq. (5.7) one arrives at

$$f_k(\mathbf{n},\mathbf{n}_0;\chi_f,\chi_i) = -d\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty}N_{mn}^{-1} \times \sum_h \int_1^{\infty} d\xi R^{(1)}_{mn}(c,\xi)\langle\chi_f|U_1(\xi)+U_2(\xi)\sigma\cdot\mathbf{L}_S|\chi_h\rangle \times S_{mn}(c,\eta)e^{im\phi}T_{mn}(c,\xi;\mathbf{n}_0;\chi_h,\chi_i). \quad (5.11)$$

Comparing with Eq. (5.6) the expression in Eq. (5.11) for the scattering amplitude $f_k(\mathbf{n};\mathbf{n}_0;\chi_f,\chi_i)$ is considerably simplified.

6. RADIAL FUNCTION

A further investigation of the scattering amplitude as expressed in Eq. (5.11) leads us to a study of the radial function $T_{mn}(c,\xi;\mathbf{n}_0;\chi_h,\chi_i)$ and its integral equation in Eq. (3.8). A formal representation for the function $T_{mn}(c,\xi;\mathbf{n}_0;\chi_h,\chi_i)$ can be obtained through Eq. (3.8) by a direct iteration procedure resulting in the usual Born series. The convergence problem, for a static potential, of Born expansion has been studied by a number of authors.[9] To insure proper convergence they have imposed a condition on potential $V(\mathbf{r})$

$$|V(\mathbf{r})| \le Mr^{-2}, \quad (6.1)$$

where M is a finite constant. The above potential has only one singularity which is located at the origin. The spheroidal potential in Eq. (2.1) has a pair of singularities, which are located at the foci of the spheroidal coordinates. This discrepancy clouds the formal representation for the spheroidal radial function. In this section the problem will be clarified.

The Born series has the form

$$T_{mn}(c,\xi;\mathbf{n}_0;\chi_f,\chi_i) = R^{(1)}_{mn}(c,\xi)S_{mn}(c,\eta_0)e^{-im\phi_0}\delta_{fi} + \sum_j \sum_{m_1=-\infty}^{\infty}\sum_{n_1=|m_1|}^{\infty}\sum_{\chi_1}\int_1^{\infty}d\xi_1\cdots\sum_{m_j=-\infty}^{\infty}\sum_{n_j=|m_j|}^{\infty}\int_1^{\infty}d\xi_j \times K_{mn,\,m_1n_1}(c;\xi,\xi_1;\chi_f,\chi_1)\cdots \times K_{m_{j-1}n_{j-1},\,m_jn_j}(c;\xi_{j-1},\xi_j;\chi_{j-1},\chi_i) \times R^{(1)}_{m_jn_j}(c,\xi_j)S_{m_jn_j}(c,\eta_0)e^{-im_j\phi_0}, \quad (6.2)$$

where

$$K_{m_ln_l,\,m_kn_k}(c;\xi_l,\xi_k;\chi_l,\chi_k) = -icR^{(1)}_{m_ln_l}(c,\xi_<^{lk})R^{(3)}_{m_kn_k}(c,\xi_>^{lk}) \times [U_1(\xi_k)\delta_{m_lm_k}\delta_{n_ln_k}\delta_{\chi_l\chi_k} + U_2(\xi_k)\langle\chi_l|\sigma\cdot\mathbf{D}_{m_ln_l,\,m_kn_k}|\chi_k\rangle]. \quad (6.3)$$

The symbol $\xi_<^{lk}(\xi_>^{lk})$ denotes the smaller (larger) of the variables ξ_l and ξ_k. To prove the convergence of the Born series in Eq. (6.2), at first, one has to verify the existence of each term in the series. This is accomplished through the study of each iterated kernel $K_{m_ln_l,\,m_kn_k}(c;\xi_l,\xi_k;\chi_l,\chi_k)$. In Eq. (6.3) the spheroidal radial function $R^{(1)}_{mn}(c,\xi)$ is regular at $\xi=1$ and has the asymptotic form

$$R^{(1)}_{mn}(c,\xi) \xrightarrow[\xi\to 1]{} O[(\xi^2-1)^{m/2}], \quad (6.4)$$

$$R^{(1)}_{mn}(c,\xi) \xrightarrow[\xi\to\infty]{} (1/c\xi)\cos[c\xi-\tfrac{1}{2}(n+1)\pi]. \quad (6.5)$$

The function $R^{(3)}_{mn}(c,\xi)$ has a singularity at $\xi=1$.

$$R^{(3)}_{mn}(c,\xi) \xrightarrow[\xi\to 1]{} O[(\xi^2-1)^{-m/2}] \quad (6.6)$$

and the symptotic form

$$R^{(3)}_{mn}(c,\xi) \xrightarrow[\xi\to\infty]{} (1/c\xi)\exp i[c\xi-\tfrac{1}{2}(n+1)\pi]. \quad (6.7)$$

In region between $\xi=1$ to $\xi=\infty$, these two functions $R^{(1)}_{mn}(c,\xi)$ and $R^{(3)}_{mn}(c,\xi)$ are regular. The coefficient $\mathbf{D}_{mn,\,m'n'}$ in Eq. (6.3) depends on the spheroidal angle function $S_{mn}(c,\eta)$ through Eq. (5.9). For larger order of n, the later function can be approximately expressed as

$$S_{mn}(c,\eta) \sim P_n^m(\eta) \quad \text{at } n \gg c. \quad (6.8)$$

The associated Legendre function $P_n^m(\eta)$ satisfies the following recurrence relations

$$L_S^+P_n^m(\eta)e^{-im\phi} = \hbar(n+m)(n-m+1)P_n^{m-1}(\eta)e^{-i(m-1)\phi},$$
$$L_S^-P_n^m(\eta)e^{-im\phi} = \hbar P_n^{m+1}(\eta)e^{-i(m+1)\phi}, \quad (6.9)$$
$$L_S^0P_n^m(\eta)e^{-im\phi} = -m\hbar P_n^m(\eta)e^{im\phi}.$$

The spheroidal angle function $S_{mn}(c,\eta)$ is usually expanded as an infinite sum of the associated Legendre function[4]

$$S_{mn}(c,\eta) = \sum_{r=0,1}^{\infty}{}' d_r^{mn}(c)P^{m}|m| + r(\eta), \quad (6.10)$$

where the prime over the summation sign indicates that the summation is only over even (odd) values of r when $n-m$ is even (odd). As r approaches infinity the ratio d_r^{mn}/d_{r-2}^{mn} goes to zero as $-c^2/(4r^2)$. It implies the absolute convergence of the expansion in Eq. (6.10) for all finite η. The orthogonality relation among the associated Legendre functions yields the coefficient $d_r^{mn}(c)$

$$d_r^{mn}(c) = \frac{2|m|+2r+1}{2}\frac{(|m|+r-m)!}{(|m|+r+m)!}\int_{-1}^{1}S_{mn}(c,\eta)P^m_{|m|+r}(\eta)d\eta. \quad (6.11)$$

Taking into account Eqs. (6.8), (6.9), (6.10), and (6.11), one arrives at the approximate expressions for coefficients $\mathbf{D}_{mn,m'n'}$ at large orders of n':

$$D^{+}_{mn,m'n'} \sim \hbar(n'+m'+1)(n'-m')\delta_{m,m'+1}d^{mn}_{n'-|m|},$$
$$D^{-}_{mn,m'n'} \sim \hbar\delta_{m,m'-1}d^{mn}_{n'-|m|}, \quad (6.12)$$
$$D^{0}_{mn,m'n'} \sim -m\hbar\delta_{m,m'}\delta_{nn'}, \quad \text{for } n' \gg c.$$

The existence of the following series can be understood as follows:

$$\sum_{m_j=-\infty}^{\infty}\sum_{n_j=|m_j|}^{\infty}\sum_{\chi_j}\int_1^{\infty}d\xi_j K_{m_{j-1}n_{j-1},\,m_jn_j}(c;\xi_{j-1},\xi_j;\chi_{j-1},\chi_j). \quad (6.13)$$

The ξ_j dependence of the kernel in Eq. (6.13) is through the spheroidal radial functions $R^{(1)}_{m_jn_j}(c,\xi_j)$ and

$R^{(3)}_{m_j n_j}(c, \xi_j)$, and the short range potentials $U_1(\xi_j)$ and $U_2(\xi_j)$. Then Eqs. (6.4), (6.5), (6.6), (6.7), (6.12), and (2.2) guarantee the absolute and uniform convergence of the improper integral with respect to ξ_j. The finite sum $\sum_{\chi_j}$ does not create a problem. The main dependence on the indices m_{j-1}, n_{j-1}, m_j, and n_j in the kernel is through the coefficients $\mathbf{D}_{m_{j-1}n_{j-1},m_j n_j}$. The asymptotic behavior of the infinite series with respect to the indices m_j and n_j in Eq. (6.13). We shall denote the uniform bound of the series in Eq. (6.13) by $B(c, V)$, which is a monotonic function of the spheroidal parameter c and the strength V of the potential $V(r)$ in Eq. (2.1). Then the Born series is bounded by the series

$$A[1 + \sum_j B^j(c, V)], \tag{6.14}$$

where A is the least upper bound

$$A = \text{l.u.b.}\,[R^{(1)}_{mn}(c, \xi) S_{mn}(c, \eta_0) e^{-im\phi_0}].$$

For $B(c, V) < 1$, the Born series is absolutely and uniformly convergent.

APPENDIX

To prove "recurrence relations" in Eq. (4.5), we first choose function $K_m(\eta, \xi) \exp(-im\phi)$ in Eq. (4.2) as $R^{(i)}_{mn'}(c, \xi) S_{mn'}(c, \eta) \exp(-im\phi)$, which obviously satisfies Helmholtz's equation. The angular momentum operators $L^{\pm}$ commute with the Laplace operator. This allows us to choose $K_m(\eta, \xi) \exp(-im\phi)$ as $L^{\pm} R^{(i)}_{m\pm1,n'}(c, \xi) S_{m\pm1,n'}(c, \xi) \exp[-i(m \pm 1)\phi]$. From Eq. (4.1) or from Eq. (5.1.5) of the book by Flammer,[4] we obtain

$$\tilde{R}^{(i,+)}_{mn}(c, \xi) e^{-im\phi} = \int_{-1}^{1} d\eta S_{mn}(c, \eta) L^{\pm} R^{(i)}_{m\pm1,n'}(c, \xi) S_{m\pm1,n'}(c, \eta) e^{-i(m\pm1)\phi}, \tag{A1}$$

where $\tilde{R}^{(i,\pm)}_{mn}(c, \xi) e^{-im\phi}$ is a spheroidal radial function. In Eq. (A1) The integration limits are taken as -1 and 1, for which the condition imposed by Eq. (4.2) is satisfied. The spheroidal radial function $\tilde{R}^{(i)}_{mn}(c, \xi)$ can be expressed in terms of two independent spheroidal radial functions

$$\tilde{R}^{(i,\pm)}_{mn}(c, \xi) = a^{(i,\pm)}_{mn,n'} R^{(3)}_{mn}(c, \xi) + b^{(i,\pm)}_{mn,n'} R^{(4)}_{mn}(c, \xi), \tag{A2}$$

where constants $a^{(i,\pm)}_{mn,n'}$ and $b^{(i,\pm)}_{mn,n'}$ will be determined from Eq. (A1) through the asymptotic forms of spheroidal functions

$$R^{(1)}_{mn}(c, \xi) \xrightarrow{\xi \to \infty} (1/c\xi) \cos[c\xi - \tfrac{1}{2}(n+1)\pi],$$
$$R^{(2)}_{mn}(c, \xi) \xrightarrow{\xi \to \infty} (1/c\xi) \sin[c\xi - \tfrac{1}{2}(n+1)\pi],$$
$$R^{(3),(4)}_{mn}(c, \xi) \xrightarrow{\xi \to \infty} (1/c\xi) \exp \pm i[c\xi - \tfrac{1}{2}(n+1)\pi]. \tag{A3}$$

For the index $i = 3$, the asymptotic limit of Eq. (A1) has the form

$$\{a^{(3,\pm)}_{mn,n'}(1/c\xi) \exp i[c\xi - \tfrac{1}{2}(n+1)\pi] + b^{(3,\pm)}_{nm,n'}(1/c\xi) \exp -i[c\xi - \tfrac{1}{2}(n+1)\pi]\} e^{-im\phi} = \{(1/c\xi) \exp i[c\xi - \tfrac{1}{2}(n+1)\pi]\} \times \int_{-1}^{1} d\eta S_{mn}(c, \eta) L^{(\pm)}_S S_{m\pm1,n'}(c, \eta) e^{-i(m\pm1)\phi}, \tag{A4}$$

where

$$L^{(\pm)}_S = \pm \hbar e^{\pm i\phi} \left(-(1-\eta^2)^{1/2} \frac{\partial}{\partial \eta} \pm \frac{i\eta}{(1-\eta^2)^{1/2}} \frac{\partial}{\partial \phi} \right). \tag{A5}$$

are the operators in Eq. (5.8). From Eq. (A4) we obtain

$$a^{(3,\pm)}_{mn,n'} = e^{i(n-n')\pi/2} \hbar \int_{-1}^{1} d\eta S_{mn}(c, \eta) \left(\frac{(m \pm 1)\eta}{(1-\eta^2)^{1/2}} S_{m\pm1,n'}(c, \eta) \mp (1-\eta^2)^{1/2} \frac{d}{d\eta} S_{m\pm1,n'}(c, \eta) \right),$$

$$b^{(3,\pm)}_{mn,n'} = 0. \tag{A6}$$

The spheroidal angle function $S_{mn}(c, \eta)$ is even or odd.

$$S_{mn}(c, -\eta) = (-)^{n-m} S_{mn}(c, \eta). \tag{A7}$$

Then we have

$$a^{(3,\pm)}_{mn,n'} = 0, \quad \text{for } n - n' = \text{odd}. \tag{A8}$$

By using the above method, one obtains

$$a^{(4,\pm)}_{mn,n'} = 0,$$

$$b^{(4,\pm)}_{mn,n'} = e^{i(n'-n)\pi/2} \hbar \int_{-1}^{1} d\eta S_{mn}(c, \eta) \times \left(\frac{m \pm 1)\eta}{(1-\eta^2)^{1/2}} S_{m\pm1,n'}(c, \eta) \mp (1-\eta^2)^{1/2} \frac{d}{d\eta} S_{m\pm1,n'}(c, \eta) \right). \tag{A9}$$

The constants $a^{(3,\pm)}_{mn,n'}$, $b^{(4,\pm)}_{mn,n'}$, and $\hbar A^{\pm}_{mn,n}$ in Eq. (4.9) are the same. From Eqs. (A2), (A7), and (A9), we have

$$\tilde{R}^{(3,\pm)}_{mn}(c, \xi) = \hbar A^{\pm}_{mn,n'} R^{(3)}_{mn}(c, \xi), \tag{A10}$$
$$\tilde{R}^{(4,\pm)}_{mn}(c, \xi) = \hbar A^{\pm}_{mn,n'} R^{(4)}_{mn}(c, \xi).$$

For the index $i = 1$, Eq. (A1) can be written as

$$\tilde{R}^{(1,\pm)}_{mn} = \tfrac{1}{2} e^{im\phi} \int_{-1}^{1} S_{mn}(c, \eta) L^{\pm} \times [R^{(3)}_{m\pm1,n'}(c, \xi) + R^{(4)}_{m\pm1,n'}(c, \xi)] S_{m\pm1,n'}(c, \eta) e^{-i(m\pm1)\phi} = \tfrac{1}{2}[\tilde{R}^{(3,\pm)}_{mn} + \tilde{R}^{(4,\pm)}_{mn}] = \hbar A^{\pm}_{mn,n'} R^{(1)}_{mn}(c, \xi). \tag{A11}$$

In arriving at Eq. (A11), one uses Eqs. (4.6), (A1), and (A10). Now we have proved the "recurrence relations" in Eq. (4.5) and expressions in Eqs. (4.7) and (4.8).

ACKNOWLEDGMENTS

In conclusion, the author wishes to express his appreciation to Professor D. Kaplan and Mr. J. R. Malebranche for several discussions. He also wishes to thank Mr. Malebranche for reading the manuscript.

*Work supported in part by the National Science Foundation.

[1] M. C. Li, J. Math. Phys. **12**, 936 (1971); J. Math. Phys. 13, 1381 (1972).
[2] J. Meixner and W. Andrejewski, Ann. Phys. (N.Y.) **7**, 157 (1950); C. Flammer, J. Appl. Phys. **24**, 1218 (1953), J. Appl. Phys. **24**, 1224 (1953); J. Meixner, Ann. Phys. (N.Y.) **12**, 227 (1953).
[3] H. Stier, Z. Phys. **76**, 439 (1932); J. B. Fisk, Phys. Rev. **49**, 167 (1936); K. Takayanagi, Prog. Theor. Phys. Suppl. **40**, 216 (1967); K. Takayanagi and Y. Itikawa, J. Phys. Soc. Jap. **24**, 160 (1968).
[4] C. Flammer, *Spheroidal Wave Functions* (Stanford U. P., Palo Alto, CA, 1957); J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbato, *Spheroidal Wave Function* (Wiley, New York, 1956).
[5] J. Rainwater, Phys. Rev. **79**, 432 (1950); S. Granger and R. D. Spence, Phys. Rev. **83**, 460 (1951).
[6] M. A. Preston, *Physics of the Nucleus* (Addison-Wesley, Reading, MA, 1962).
[7] I. Marx, J. Math. & Phys. **32**, 269 (1954); Mich. Math. J. 2, 45 (1953); 2, 99 (1954); N. Chako, J. Math. & Phys. **36**, 62 (1957).
[8] E. T. Whittaker, J. Lond. Math. Soc. **4**, 88 (1929).
[9] R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951); W. Kohn, Rev. Mod. Phys. **26**, 292 (1954).

Wave propagation in a random lattice. II*

Pao-Liu Chow

Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 3 July 1972)

Propagation of the mean waves in a simple cubic crystal with small, random mass defects is studied using the results of a previous paper. The dispersion equations for two- and three-dimensional crystals are analyzed. For uncorrelated mass defects, the effective propagation constants are obtained explicitly in the limits of low frequency and near cut-off frequency. The mean reflection and transmission coefficients are determined for a plane wave incident in a crystal formed by joining a semi-infinite perfect crystal to a semi-infinite defective crystal or by connecting two semi-infinite perfect crystals to a finite imperfect crystal. We also construct the mean Green's functions for an infinite crystal lattice containing defective masses distributed throughout it or confined to a half-space.

INTRODUCTION

This is a continuation of the first part of our study on wave propagation in a random lattice,[1] which we shall refer to as Paper I. In Sec. 1 of this paper we attempt to solve the dispersion equations for the mean waves in two and three dimensions. For uncorrelated mass defects, asymptotic results are obtained at low frequency and near the cut-off frequency. Here we need Green's functions for a perfect lattice in higher dimensions. For lattice vibration problems, the approximate evaluation of Green's functions can be found in an excellent book by Maradudin *et al.*[2] An approximate evaluation of Green's functions corresponding to a lattice wave in two and three dimensions is carried out in the Appendix. These results were not given in Ref. 2.

The scattering problem for lattice waves by point defects was first analyzed by Lifshitz,[3,4] and by Koster[5] in a different physical context. This theory has been further developed by Klein,[6] Krumhans,[7] and Callaway.[8] However, as mentioned in I, Rubin[9] seems to be the first to have investigated the transmission of a lattice wave in a one-dimensional lattice with randomly distributed mass defects (for related references, see I). He defined the transmission coefficient T in terms of the mean value of the logarithm of the amplitude of the transmitted wave. In contrast, we shall define the reflection coefficient Γ and the transmission coefficient T based on the mean (or coherent) waves. In our definition, the coefficients Γ and T may be complex. One could have defined $|\Gamma|^2$ and $|T|^2$ to be the reflection and the transmission coefficients respectively, as done by some authors, then they become real quantities. Physically the constants Γ and T can be interpreted as the effective parameters, in the sense of Yvon[10] and Kirkwood[11] who determined the electromagnetic properties of a polarizable medium based upon the average field. That is, if the random lattice is replaced by a deterministic one whose propagation property is characterized by the mean wave, the reflection and transmission coefficients would be given by what we have defined. The advantage of our approach lies in the simplification of the mathematical analysis. This makes it possible to apply the present method to different boundary problems associated with nonsimple lattices, which would, otherwise, be difficult to treat analytically.

In Sec. 2, we formulate the problem of reflection and transmission of a plane incident wave. It is shown in Sec. 3 that the higher dimensional problems are reducible to that in one dimension if the correlation function is homogeneous and isotropic. Three special cases are solved in the following section. They correspond to the problems of reflection and transmission by a semi-infinite imperfect lattice with uncorrelated mass defects in one and three dimensions, and exponentially correlated mass defects in one dimension. A similar problem is solved when the plane wave is incident from the imperfect crystal. In Sec. 6 we determine the reflection and transmission coefficients in the presence of a finite imperfect crystal for various cases. For arbitary correlation functions the problems of reflection and transmission by a semi-infinite imperfect crystal can be solved by adapting the Wiener–Hopf method to the case of a discrete variable. This is done in Sec. 7. In the last two sections we construct the mean Green's functions for an infinite random lattice and a semi-infinite random lattice joining a perfect one.

In an earlier paper,[12] we treated the problem of reflection and transmission by slabs of continuous random media. By an appropriate limiting process from a lattice to a continuum, our current results can be shown to yield the corresponding results obtained previously in Ref. 12.

1. PROPAGATION CONSTANT FOR A TWO- OR THREE-DIMENSIONAL CRYSTAL

Let us consider a simple cubic crystal in three dimensions with random mass defects. It is convenient to set $q = \mathbf{n} = (n_1, n_2, n_3)$, in (I. 1. 14) and (I. 1. 15) where n_i, $i = 1, 2, 3$, are integers. Then the position of the lattice site $\mathbf{n}$ is given by $\mathbf{x}(\mathbf{n}) = d(n_1\mathbf{e}_2 + n_2\mathbf{e}_2 + n_3\mathbf{e}_3)$. Here d denotes the lattice spacing, and $\mathbf{e}_i$ are mutually orthogonal unit vectors. For nearest-neighbor interaction with spring constants α_j, ϕ takes the form

$$\begin{aligned}\phi(\mathbf{n},\mathbf{n}') &= -\alpha_j \quad \text{if } n_j' = n_j \pm 1, \quad n_i' = n_i \text{ for } i \neq j,\\ &= 2(\alpha_1 + \alpha_2 + \alpha_3) \quad \text{if } \mathbf{n}' = \mathbf{n}, \qquad (1.1)\\ &= 0 \quad \text{otherwise.}\end{aligned}$$

Then (I. 1. 14) and (I. 1. 15) become

$$L_0 = -\sum_{j=1}^{3} \alpha_j \Delta_j^2 - m\omega^2, \qquad (1.2)$$

$$V = -\epsilon m\omega^2 \mu(\mathbf{n}). \qquad (1.3)$$

Here Δ_j^2 is the second central difference operator with respect to n_j. In view of (1.1), (I. 2. 2) becomes

$$\tilde{\phi}(\mathbf{k}) = 2\sum_{j=1}^{3} (\alpha_j \cos k_j d - 1). \qquad (1.4)$$

The Green's function of L_0 defined by (I. 4. 2) is

given by the principal value integral

$$G(\mathbf{n},\mathbf{n}',\omega) = \lim_{\epsilon\to 0^+} \frac{-1}{(2\pi)^3} \int\int_0^{2\pi} \int \times \frac{e^{i\theta\cdot(\mathbf{n}-\mathbf{n}')}}{[m\omega^2 - 2\sum_{j=1}^{3} \alpha_j(1-\cos\theta_j) \pm i\epsilon]} d\theta_1 d\theta_2 d\theta_3, \quad |\omega| < \omega_c \tag{1.5}$$

the choice of the upper or the lower sign of $i\epsilon$ is determined by an outgoing wave condition and ω_c is the cut-off frequency defined by

$$\omega_c = 2\left(\frac{1}{m}\sum_{j=1}^{3}\alpha_j\right)^{1/2}. \tag{1.6}$$

For uncorrelated defects, the dispersion equation (I. 3. 3) yields, when (1.4) is used,

$$\omega^2 m + \epsilon^2\omega^4 m^2 \langle\mu^2\rangle\, g(\omega) - 2\sum_{j=1}^{3}(\alpha_j \cos k_j d - 1) = 0, \tag{1.7}$$

where $g(\omega) = G(0,0,\omega)$.

The solution of (1.7) for $\mathbf{k}$ corresponding to (I. 3. 6), is found to be

$$k(\omega,\epsilon^2) = \hat{\mathbf{k}}_0 k_0(\omega)\left[1 + \epsilon^2\left(2d\sum_{j=1}^{3}\alpha_j k_{0_j}\sin k_{0_j} d\right)^{-1} \times \omega^4\langle\mu^2\rangle g(\omega) + O(\epsilon^4)\right]. \tag{1.8}$$

According to (I. 3. 8), the vibration frequency ω, which is shifted from ω_0 is given by

$$\omega(\mathbf{k},\epsilon^2) = \omega_0(k) - \epsilon^2 m\langle\mu^2\rangle\omega_0^3(\mathbf{k})\, g(\omega_0) + O(\epsilon^4). \tag{1.9}$$

For the two-dimensional case, the expression for $g(\omega)$ is given by (1.5) with $\mathbf{n} = \mathbf{n}' = 0$ and the triple integral replaced by a double integral. It is shown in the Appendix that g can be expressed in terms of the hypergeometric function F as follows:

$$g(\omega) = \left(\frac{1}{2\pi}\right)^2 \frac{1}{m\omega(\omega_c^2 - \omega^2)^{1/2}} F(\tfrac{1}{2},\tfrac{1}{2};1;z), \tag{1.10}$$

in which

$$z(\omega) = 4\sqrt{\alpha_1\alpha_2}/m\omega\,(\omega_c^2 - \omega^2)^{-1/2}. \tag{1.11}$$

It is clear that, when ω is small or near ω_c, z is large.

Therefore, we can expand F for large argument z either for low frequency or near cut-off frequency. The leading term in this expansion reads (see p. 560 of Ref. 2)

$$F(\tfrac{1}{2},\tfrac{1}{2};1;z) = \pm (i/\pi)z^{-1/2}[(\ln z + i\pi) + 4\ln 2] \pm (i/\sqrt{\pi})z^{-1/2} + O(z^{-3/2}\ln z). \tag{1.12}$$

When (1.10) and (1.12) are used in (1.8), we see that the upper branch of F must be chosen in order to have positive attenuation.

Then (1.8) yields

$$\mathbf{k}(\omega,\epsilon^2) \sim \hat{\mathbf{k}}_0 k_0(\omega)\left[1 - \epsilon^2\langle\mu^2\rangle\left(\frac{1}{2\pi}\right)^3 \left(d\sum_{j=1}^{2}\alpha_j k_{0_j}\sin k_{0_j} d\right)^{-1} \frac{\omega^4 m^2}{4\sqrt{\alpha_1\alpha_2}} z^{1/2}(\omega)[\pi - i(\ln z(\omega) + 4\ln 2 + \sqrt{\pi})]\right] \tag{1.13}$$

Similarly, when $|\omega_0|$ is small or near ω_c, (1.9) becomes

$$\omega(\mathbf{k},\epsilon^2) \sim \pm\,\omega_0(\mathbf{k})\left[1 \pm \epsilon^2\left(\frac{1}{2\pi}\right)^3\langle\mu^2\rangle \frac{\omega_0^3(\mathbf{k})m^2}{2\sqrt{\alpha_1\alpha_2}} z^{1/2} \times (\omega_0)[\pi - i(\ln z(\omega_0) + 4\ln 2 + \sqrt{\pi})]\right]. \tag{1.14}$$

From (1.13) it is seen that $\mathrm{Re}\{k\} < k_0$ and $\mathrm{Im}\{k\} > 0$. Therefore, the phase speed increases and the mean wave attenuates. Both of these effects, of $O(\omega^{7/2}\ln\omega)$, are negligible at low frequency, while they are noticeable, of $O[(\omega_c - \omega)/\ln(\omega_c - \omega)^{1/4}]$, when ω is near ω_c.

In three dimensions we assume $\alpha_2 = \alpha_3$ for computational convenience. Then, as shown in the Appendix, $g(\omega)$ can be represented by a contour integral along a unit semicircle C with a hypergeometric function as its integrand:

$$g(\omega) = \left(\frac{1}{2\pi}\right)^5\left(\frac{2}{\alpha_1\alpha_2}\right)^{1/2}\int_C \{[\zeta^2 + 2(\delta_1 - 1)\zeta + 1] \times [\zeta^2 + 2(\delta_2 - 1)\zeta + 1]\}^{1/2} \times F[\tfrac{1}{2},\tfrac{1}{2},1;W(\zeta)]d\zeta. \tag{1.15}$$

Here we have set

$$\delta_1 = m\omega^2/2\alpha_1, \tag{1.16}$$

$$\delta_2 = [m(\omega_c^2 - \omega^2)/2\alpha_2], \tag{1.17}$$

and

$$W(\zeta) = \left(\frac{\zeta(\zeta^2+1)}{[\zeta^2 + 2(\delta_1 - 1)\zeta + 1][\zeta^2 + 2(\delta_2 - 1)\zeta + 1]}\right)^{1/2}. \tag{1.18}$$

At low frequency δ_1 is small and near the cut-off frequency. When δ_1 is small, the main contribution to the integral (1.15) comes from the branch integrals along the line segment connecting the branch points ζ_1 and ζ_2, which are the roots of the quadratic equation $\zeta^2 + 2(\delta_1 - 1)\zeta + 1 = 0$. This is so because, along these paths, $|W|$ is uniformly large. Therefore, we can use the asymptotic formula (1.12) to evaluate those integrals approximately. When these results are used in (1.15), it yields (for details, see the Appendix)

$$g(\omega) \sim \frac{i}{4}\left(\frac{1}{\pi}\right)^6\left(\frac{\delta_2}{\alpha_1\alpha_2}\right)^{1/2}(\delta_1)^{1/4}\left[(4\ln 2 - \ln\delta_1 + \sqrt{\pi} + i\pi)\int_0^{\pi/2}\cos^{1/2}\theta d\theta + \int_0^{\pi/2}\cos^{1/2}\theta\ln\sec\theta d\theta\right], \quad \delta_1 \to 0. \tag{1.19}$$

By the symmetry of δ_1 and δ_2 in (1.15), the expression for $g(\omega)$ as $\delta_2 \to 0$ is given by (1.19) with δ_1 and δ_2 interchanged. When this result and (1.19) are inserted respectively in (1.8) and (1.9), we obtain

$$\mathbf{k}(\omega,\epsilon) \sim \hat{\mathbf{k}}_0 k_0(\omega)\left[1 + i\epsilon^2(1/\pi)^6\left(8d\sum_{l=1}^{3}\alpha_l k_{0l}\sin k_{0l} d\right)^{-1} \times \omega^4 m^2\left(\frac{\delta_j}{\alpha_1\alpha_2}\right)^{1/2}(\delta_k)^{1/4}\left[(4\ln 2 - \ln\delta_k + \sqrt{\pi} + i\pi) \times \int_0^{\pi}\cos^{1/2}\theta d\theta + \int_0^{\pi/2}\cos^{1/2}\theta\ln\sec\theta d\theta\right]\right], \tag{1.20}$$

where $j = 2, k = 1$ when $\omega \approx \omega_c$. The corresponding

result for the frequency shift (1.9) reads

$$\omega(\mathbf{k},\epsilon) \sim \pm\,\omega_0(\mathbf{k})\Big[1 - i\epsilon^2(1/\pi)^6\langle\mu^2\rangle\omega_0^3(\mathbf{k})m \times \Big(\frac{\delta_j(\omega_0)}{\alpha_1\alpha_2}\Big)^{1/2}[\delta_k(\omega_0)]^{1/4} \times \Big\{[4\ln 2 - \ln\delta_k(\omega_0) + \sqrt{\pi} + i\pi]\int_0^{\pi/2}\cos^{1/2}\theta d\theta + \int_0^{\pi/2}\cos^{1/2}\theta \ln\sec\theta d\theta\Big\}\Big]. \quad (1.21)$$

Here, as above, $j = 2$, $k = 1$ for $\omega_0 \approx 0$ and $j = 1$, $k = 2$ for $\omega_0 \approx \omega_c$. Separation of the real and the imaginary parts of (1.20) and (1.21) yields

$$\mathrm{Re}\{k(\omega,\epsilon^2)\} \sim k_0(\omega)\Big[1 - \epsilon^2(1/\pi)^5\Big(8d\sum_{l=1}^{3}\alpha_l k_{0l}\sin k_{0l}d\Big)^{-1} \times \omega^4 m^2\Big(\frac{\delta_j}{\alpha_1\alpha_2}\Big)^{1/2}(\delta_k)^{1/4}\int_0^{\pi}\cos^{1/2}\theta d\theta\Big], \quad (1.22)$$

$$\mathrm{Im}\{k(\omega,\epsilon^2)\} \sim \epsilon^2 k_0(\omega)(1/\pi)^6\Big(8d\sum_{l=1}^{3}\alpha_l k_{0l}\sin k_{0l}d\Big)^{-1}\omega^4 m^2 \times \Big(\frac{\delta_j}{\alpha_1\alpha_2}\Big)^{1/2}(\delta_k)^{1/4}\Big[(4\ln 2 - \ln\delta_k + \sqrt{\pi})\int_0^{\pi}\cos^{1/2}\theta d\theta + \int_0^{\pi}\cos^{1/2}\theta\ln\sec\theta d\theta\Big], \quad (1.23)$$

$$\mathrm{Re}\{\omega(\mathbf{k},\epsilon)\} \sim \pm\,\omega_0(\mathbf{k})\Big[1 + \epsilon^2(1/\pi)^5\langle\mu^2\rangle\omega_0^3(\mathbf{k})m \times \Big(\frac{\delta_j(\omega_0)}{\alpha_1\alpha_2}\Big)^{1/2}[\delta_k(\omega_0)]^{1/4}\int_0^{\pi/2}\cos^{1/2}\theta d\theta\Big], \quad (1.24)$$

$$\mathrm{Im}\{\omega(\mathbf{k},\epsilon)\} \sim \pm\,\epsilon^2(1/\pi)^{1/6}\langle\mu^2\rangle\omega_0^4(\mathbf{k})m\Big(\frac{\delta_j(\omega_0)}{\alpha_1\alpha_2}\Big)^{1/2} \times [\delta_k(\omega_0)]^{1/4}\Big\{[4\ln 2 - \ln\delta_k(\omega_0) + \sqrt{\pi}]\int_0^{\pi/2}\cos^{1/2}\theta d\theta + \int_0^{\pi/2}\cos^{1/2}\theta\ln\sec\theta d\theta\Big\}. \quad (1.25)$$

Formulas (1.22) and (1.23) provide the effective phase speed and the attenuation coefficient, and (1.24) and (1.25) gives the frequency shift and the decay rate, respectively. The essential features resemble those of the two-dimensional problem, that is, the mean wave attenuates and its phase speed increases due to the scattering by the uncorrelated mass defects. It is interesting to note that, in two and three dimensions the uncorrected mass defects introduce a change in the phase speed and attenuation of the mean wave. However, in one-dimension, the same mass defects yield, to order ϵ^2, attenuation only.

2. FORMULATION OF REFLECTION AND TRANSMISSION PROBLEMS

We now consider the reflection and transmission of a plane wave by a slab of imperfect crystal of finite thickness, or by a semi-infinite imperfect crystal in r dimensions, $r = 1, 2, 3$. For the finite imperfect crystal, the "imperfect" lattice points are contained in the slab, $0 \le n_1 \le N$, where n_1 denotes the first component of n so that $\mathbf{n} = (n_1, \mathbf{n}_T)$. As in the previous section, we confine ourselves to simple cubic crystals. Then the operators L_0 and V are given by (1.2) and (1.3) with $\mu(\mathbf{n}) = 0$ if $n_1 < 0$ or $n_1 > N$. From Eq. (1.9) in I, a scalar mean wave $\langle u(\mathbf{n})\rangle$ satisfies the equation

$$\Big(\sum_{j=1}^{3}\alpha_j\Delta_j^2 + m\omega^2\Big)\langle u(\mathbf{n})\rangle + \epsilon^2\omega^4 m^2\sum_{n_i=0}^{N} \times G(\mathbf{n},\mathbf{n}',\omega)R(\mathbf{n},\mathbf{n}')\langle u(\mathbf{n}')\rangle = 0, \quad |\mathbf{n}| \ge 0, \quad (2.1)$$

where G is the Green's function of L_0 defined by (1.16) and R is the two point correlation function of μ

$$R(\mathbf{n},\mathbf{n}') = \langle\mu(\mathbf{n})\mu(\mathbf{n}')\rangle, \quad 0 \le n_1 \le N \text{ and } 0 \le n_1' \le N, \quad = 0, \quad \text{otherwise.} \quad (2.2)$$

Let $e^{i\mathbf{k}_0\cdot\mathbf{n}d}$ be a plane incident wave from the left semi-infinite lattice, $n_1 < 0$. Then we can write the mean fields for $n_1 \le 0$ and $n_1 \ge N$, respectively, as

$$\langle u(\mathbf{n})\rangle = e^{i\mathbf{k}_0\cdot\mathbf{n}d} + \langle u_R(\mathbf{n})\rangle, \quad n_1 \le 0 \quad (2.3)$$

and

$$\langle u(\mathbf{n})\rangle = \langle u_T(\mathbf{n})\rangle, \quad n_1 \ge N, \quad (2.4)$$

where $\langle u_R\rangle$ and $\langle u_T\rangle$ are the reflected and the transmitted waves which satisfy the outgoing conditions in their respective domains of definition. Our problem is to determine the mean fields $\langle u_R\rangle$ and $\langle u_T\rangle$ from (6.1). For a semi-infinite imperfect crystal, $N = \infty$, the transmitted wave defined by (2.4) is replaced by

$$\langle u(\mathbf{n})\rangle = \langle u_T(\mathbf{n})\rangle, \quad n_1 \ge 0. \quad (2.5)$$

3. REDUCTION TO A ONE-DIMENSIONAL PROBLEM

We will show that the problem posed is reducible to a one-dimensional problem. Splitting the wave vector $\mathbf{k}_0$ into the component $\mathbf{k}_{01}$ along the one or the longitudinal direction and $\mathbf{k}_{0T}$ transverse to this direction, we can write the incident plane wave as $e^{i\mathbf{k}_0\cdot\mathbf{n}d} = e^{i\mathbf{k}_{01}\cdot\mathbf{n}_1 d}\, e^{i\mathbf{k}_{0T}\cdot\mathbf{n}d}$, where $\mathbf{k}_0 = \mathbf{k}_{01} + \mathbf{k}_{0T}$. Let us seek a solution of the form

$$\langle u(\mathbf{n})\rangle = v(n_1)e^{i\mathbf{k}_{0T}\cdot\mathbf{n}d}. \quad (3.1)$$

Inserting (3.1) into (2.1) yields, with $n = |\mathbf{n}_1|$,

$$(\Delta^2 + \beta^2)v(n) + \epsilon^2\frac{\omega^4 m^2}{\alpha_1}\sum_{n'=0}^{N}\widehat{GR}(|n-n'|,\mathbf{k}_{0T},\omega)v(n'), \quad (3.2)$$

where

$$\beta = (1/\sqrt{\alpha_1})\{m\omega^2 + 2[\alpha_2(\cos k_{02}d - 1) + \alpha_3(\cos k_{03}d - 1)]\}^{1/2} \quad (3.3)$$

and $\widehat{GR}$ is the transverse discrete Fourier transform of GR, defined by

$$\widehat{GR}(|n-n'|,\mathbf{k}_{0T},\omega) = \sum_{\mathbf{n}_T'} G(\mathbf{n},\mathbf{n}',\omega)R(\mathbf{n},\mathbf{n}') \times e^{i\mathbf{k}_{0T}\cdot(\mathbf{n}'-\mathbf{n})}. \quad (3.4)$$

Due to the invariance of G and R under lattice translations, $\widehat{GR}$ is independent of $\mathbf{n}_T$. For the one-dimension problem, we must set $\beta = k_0 = |\mathbf{k}_{01}|$, $\widehat{GR} = GR$, and $\langle u\rangle = v$. When (2.1) is used in (2.3) and (2.4), we find that

$$v(n) = e^{ik_0 nd} + \Gamma e^{-ik_0 nd}, \quad n \ge 0, \quad (3.5)$$

$$v(n) = Te^{ik_0(n-N)d}, \quad n \ge N, \quad (3.6)$$

where Γ and T are independent of the transverse vari-

able $\mathbf{n}_T$ because the reduced equation (3.2) is one-dimensional. Γ and T, which may be complex valued, are defined to be the reflection and transmission coefficients for the mean wave. For a semi-infinite imperfect crystal, in view of (2.5), Eq. (3.6) is replaced by

$$v(n) \text{ outgoing}, \quad n > 0. \tag{3.7}$$

In this case, as we shall see, there may be more than one transmitted wave. Then we define the coefficient of each attenuated plane wave to be a transmission coefficient for that particular transmitted wave. To simplify the analysis, we rewrite Eq. (3.2) as

$$(\Delta^2 + \beta^2)v(n) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|n-n'|, \mathbf{k}_{0T}, \omega)v(n') = 0, \quad 1 \le n \le N-1 \tag{3.8}$$

$$(\Delta^2 + \beta^2)v(n) = 0, \quad n \le -1 \text{ or } n \ge N+1 \tag{3.9}$$

and

$$(\Delta^2 + \beta^2)v(0) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|n'|, \mathbf{k}_{0T}, \omega)v(n') = 0, \tag{3.10}$$

$$(\Delta^2 + \beta^2)v(N) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|N-n'|, \mathbf{k}_{0T}, \omega)v(n') = 0. \tag{3.11}$$

We note that $v(n) = e^{ik_0 nd} + O(\epsilon^2)$ for all n. Therefore, if we replace $v(n')$ in (3.10) and (3.11) by $e^{ik_0 n'd}$, these equations are still valid up to $O(\epsilon^2)$. That is, correct to $O(\epsilon^2)$

$$(\Delta^2 + \beta^2)v(0) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|n'|, \mathbf{k}_{0T}, \omega)e^{ik_0 n'd} = 0. \tag{3.12}$$

$$(\Delta + \beta^2)v(N) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|N-n'|, \mathbf{k}_{0T}, \omega)e^{ik_0 n'd} = 0. \tag{3.13}$$

However, we should not use such a replacement in (3.8) because this procedure will give rise to an unphysical result—the growth of the mean wave as N becomes large. Evaluating (3.5) at $n = 0$, (3.6) at $n = N$, and noting (3.12), (3.13), we obtain a set of boundary conditions at the interfaces

$$v(0) = 1 + \Gamma, \tag{3.14}$$

$$v(1) = \mathrm{b}_1 + \mathrm{b}_2 \Gamma, \tag{3.15}$$

and

$$v(N) = T, \tag{3.16}$$

$$v(N-1) = \mathrm{b}_3 + \mathrm{b}_2 T, \tag{3.17}$$

where

$$\mathrm{b}_1 = (2 - \beta^2 - e^{-ik_0 d}) - \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|n'|, \mathbf{k}_{0T}, \omega)e^{ik_0 n'd}, \tag{3.18}$$

$$\mathrm{b}_2 = (2 - \beta^2 - e^{ik_0 d}), \tag{3.19}$$

$$\mathrm{b}_3 = -\epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{N} \widehat{GR}(|N-n'|, \mathbf{k}_{0T}, \omega) \times e^{ik_0 n'd}. \tag{3.20}$$

Since (3.5), (3.6) automatically satisfy (3.9), our task is to solve Eq. (3.8) subject to the set of conditions (3.14)–(3.17). In the process, when $N = \infty$, the unknown coefficients Γ and T will be determined. The conditions (3.16) and (3.17) are replaced by an outgoing condition.

4. SPECIAL CASES FOR A SEMI-INFINITE IMPERFECT CRYSTAL

Case A: Uncorrelated mass defects in one-dimension

In one dimension, (3.8) becomes

$$(\alpha\Delta^2 + m\omega^2)v(n) + \epsilon^2\omega^4 m^2 \sum_{n'=0}^{\infty} \times G(|n-n'|, \omega)R(|n-n'|)v(n') = 0, \quad n > 0 \tag{4.1}$$

where the Green's function G was defined by (I. 4. 5). For uncorrelated mass defects, we have $R(|n|) = 0$ for $n \neq 0$, $R(0) = \langle \mu^2 \rangle$. Then (4.1) reduces to

$$(\alpha\Delta^2 + m\omega^2)v(n) - \epsilon^2 \frac{\langle \mu^2 \rangle \omega^4 m^2}{2i\alpha \sin k_0 d} v(n) = 0, \quad n > 0. \tag{4.2}$$

The outgoing solution of (4.2) is

$$v(n) = Te^{iknd}, \quad n > 0. \tag{4.3}$$

Here $k(\omega, \epsilon^2)$ is a solution of the dispersion equation (I. 4. 15) with the real and the imaginary parts given by (I. 4. 16) and (I. 4. 17). Substituting (4.3) into Eqs. (3.14) and (3.15) yields a pair of equations for T and Γ. Their solutions are

$$T = \frac{\mathrm{b}_2 - \mathrm{b}_1}{\mathrm{b}_2 - e^{ikd}} = 1 - i\epsilon^2 \frac{\langle \mu^2 \rangle}{8} \frac{1}{[(\omega_c/\omega)^2 - 1]} \times \left(\frac{(\omega_c/\omega)^2 - 2}{(\omega_c/\omega)^2 - 1} + 3i \right) + 0(\epsilon^4), \tag{4.4}$$

$$\Gamma = -\frac{\mathrm{b}_1 - e^{ikd}}{\mathrm{b}_2 - e^{ikd}} = -i\epsilon^2 \frac{\langle \mu^2 \rangle}{8} \frac{1}{[(\omega_c/\omega)^2 - 1)]^{1/2}} \times \left(\frac{(\omega_c/\omega)^2 - 2}{(\omega_c/\omega)^2 - 1} + 3i \right) + 0(\epsilon^4). \tag{4.5}$$

Case B: Uncorrelated mass defects in three-dimensions

In this case, the Green's function is given by (1.5) and the correlation function R is given by $R(|\mathbf{n}|) = 0$ if $|\mathbf{n}| \neq 0$ and $R(0) = \langle \mu^2 \rangle$. When these are used in (3.4), we find after simplification, that

$$\widehat{GR}(0, \mathbf{k}_{0T}, \omega) = g(\omega), \tag{4.6}$$

$$\widehat{GR}(|\mathbf{n}|, \mathbf{k}_{0T}, \omega) = 0, \quad \text{if } n \neq 0. \tag{4.7}$$

Here we have set $\alpha_2 = \alpha_3$ and the expression for $g(\omega)$ is given by (1.15). In view of (4.6) and (4.7), Eq. (4.8) becomes

$$(\Delta^2 + \beta^2)v(n) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \langle \mu^2 \rangle g(\omega)v(n) = 0, \quad n > 0. \tag{4.8}$$

Let $\lambda = k_1$ be a solution of the equation

$$2\cos\lambda d + (\beta^2 - 2) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \langle \mu^2 \rangle g(\omega) = 0, \tag{4.9}$$

so that $0 < \mathrm{Re}\{k_1\} < \pi/d$ and $\mathrm{Im}\{k_1\} > 0$. Then the outgoing solution of (4.9) takes the form

$$v(n) = Te^{ik_1nd}, \quad n > 0. \tag{4.10}$$

It is clear that the transmission and reflection coefficients T and Γ can be obtained from (4.4) and (4.5) in their unexpanded forms by changing k to k_1. That is,

$$T = \frac{b_2 - b_1}{b_2 - e^{ik_1d}} = 1 - \frac{\epsilon^2\omega^4m^2}{2\alpha_1\beta(4-\beta^2)^{1/2}}\left(\frac{(2-\beta^2)}{\beta(4-\beta^2)^{1/2}} + 3i\right) \times g(\omega) + 0(\epsilon^4), \tag{4.11}$$

$$\Gamma = -\frac{b_1 - e^{ik_1d}}{b_2 - e^{ik_1d}} = -\frac{\epsilon^2\omega^4m^2}{2\alpha_1\beta(4-\beta^2)^{1/2}}\left(\frac{2-\beta^2}{\beta(4-\beta^2)^{1/2}} + 3i\right) \times g(\omega) + 0(\epsilon^4), \tag{4.12}$$

where use has been made of Eqs. (4.18), (3.19), and (4.9). The definition of β is given by (3.3). At low-frequency and near cut-off frequency, asymptotic forms for $g(\omega)$ are given by (1.19).

Case C: Exponential correlation in one-dimension

To be explicit, we shall treat the one-dimensional case only. Our reduction procedure discussed in Sec. 3 allows the subsequent results to be extended to that in higher dimensions.

Let the correlation function R be given by (I.4.21). Then the one-dimensional version of (3.4) reads

$$GR(|n-n'|,\omega) = \langle\mu^2\rangle \frac{e^{iK|n-n'|d}}{2i\sin k_0d}, \tag{4.13}$$

where

$$K = (k_0 + ia^{-1}). \tag{4.14}$$

Substituting (4.14) into (3.8) yields

$$(\Delta^2 + \omega^2/\alpha)v(n) + i\epsilon^2\frac{\langle\mu^2\rangle\omega^4m^2}{2\alpha\sin k_0d}\sum_{n'=0}^{\infty} \times e^{iK|n-n'|d}v(n') = 0, \quad n > 0. \tag{4.15}$$

If we define β_e^2 to be

$$\beta_e^2 = 2(1 - \cos Kd), \tag{4.16}$$

the following identity holds

$$(\Delta^2 + \beta_e^2)e^{iK|n-n'|d} = 2i\delta_{nn'}\sin\gamma d, \tag{4.17}$$

where $\delta_{nn'}$ is the Kronecker delta symbol. Now, applying the second order difference operator $(\Delta^2 + \beta_e^2)$ to (4.16), one obtains a fourth order difference equation

$$(\Delta^2 + \beta_e^2)(\Delta^2 + \omega^2m/\alpha)v(n) - \epsilon^2\langle\mu^2\rangle\frac{\omega^4m^2}{\alpha} \times \frac{\sin Kd}{\sin k_0d}v(n) = 0, \quad n > 0. \tag{4.18}$$

Its general solution can be written as

$$v(n) = \sum_{j=1}^{4} T_je^{ik_jnd}, \quad n > 0. \tag{4.19}$$

For arbitrary constants T_j, (3.19) cannot satisfy the original second order equation (4.16) where only two arbitrary constants are involved. The wavenumbers k_j, $j = 1, 2, 3, 4$, are the perturbed roots, about $\lambda = \pm k_0$ and $\lambda = \pm K$, of the equation

$$(\cos\lambda d - \cos k_0d)(\cos\lambda d - \cos Kd) - \epsilon^2\langle\mu^2\rangle\frac{\omega^4m^2}{\alpha} \times \frac{\sin Kd}{\sin k_0d} = 0. \tag{4.20}$$

Two of these roots, say k_1 and k_3, have a positive imaginary part. They are found to be

$$k_1(\omega,\epsilon^2) = k_0 + 2\epsilon^2\langle\mu^2\rangle\left(\frac{\omega_c^2}{\omega^2} - 1\right)^{-1} \times \frac{\sin 2k_0d(1 - \cosh d/a) + 2i(\sinh d/a - \cos^2 k_0d)}{d(\cos^2 k_0d(\cosh d/a - 1)^2 + \sin^2 k_0d\sinh^2 d/a]} + O(\epsilon^4), \tag{4.21}$$

$$k_3(\omega,\epsilon^2) = (k_0 + i/a) - 4\epsilon^2\langle\mu^2\rangle\left[\left(\frac{\omega_c^2}{\omega^2}\right) - 1\right]^{-1}\sin k_0d\frac{(\sin k_0d\cosh d/a - i\cos k_0d)}{d(\sin k_0d\cosh^2 d/a + \cos^2 k_0d\sinh d/a)} \times \frac{\sin k_0d\cos k_0d(1 - \cosh d/a) + i\sinh d/a(\cosh a/d - \cos^2 k_0d)}{[\cos^2 k_0d(\cosh d/a - 1)^2 + \sin^2 k_0d\sinh^2 d/a]} + O(\epsilon^4). \tag{4.22}$$

The constants T_2 and T_4 must be set equal to zero for $v(n)$ shown by (4.19) to be an outgoing solution. Then we insert (4.19), with $\Gamma_2 = \Gamma_4 = 0$, into (4.16) and invoke (4.20) to obtain

$$\frac{T_1}{1 - e^{i(k_1 - K)d}} + \frac{T_3}{1 - e^{i(k_3 - K)d}} = 0. \tag{4.23}$$

This additional condition adjointed to (3.14) and (3.15) makes it possible to determine T_1, T_3 and Γ uniquely. The results are

$$T_1 = i\frac{2\sin k_0d - \epsilon^2[2\sin k_0d(1 - e^{i(K+k_0)d})]^{-1}}{(e^{ik_1d} - \Omega e^{ik_3d}) - (1-\Omega)e^{-ik_0d}}, \tag{4.24}$$

$$T_3 = -\Omega T_1, \tag{4.25}$$

$$\Gamma = (1 - \Omega)T_1 - 1, \tag{4.26}$$

where

$$\Omega = \frac{1 - e^{i(k_3 - K)d}}{1 - e^{i(k_1 - K)d}}. \tag{4.27}$$

In terms of these constants, the solution (4.19) becomes

$$v(n) = T_1(e^{ik_1nd} - \Omega e^{ik_3nd}), \quad n > 0. \tag{4.28}$$

Since k_1 and k_3 have positive imaginary parts, (4.28) shows that both of the transmitted waves are attenuated. The k_3 wave decays within a transition layer containing

$[a/d]$ lattice points, where $[x]$ means the least integer greater than x. The corresponding transmission coefficient T_3 is of order ϵ^2. This can be verified by expanding Ω in (4.27) in a power series with respect to the small parameter ϵ. Therefore, the major contribution to transmitted field is due to the k_1 wave. By comparing (4.21) with (I.4.15), we see that k_1 is approximately equal to the effective propagation constant defined there only when $(a^{-1}d) \gg 1$, i.e., when the correlation length is small compared with the lattice spacing. This means that we cannot, in general, replace the random lattice in the half-space by a perfect lattice with an effective propagation constant.

5. REFLECTION AND TRANSMISSION OF A WAVE INCIDENT FROM A RANDOM LATTICE

We only consider a one-dimensional lattice with an exponential correlation. We suppose that the incident wave from the random lattice for $n > 0$ is e^{-ik_1nd} where k_1 is given by (4.21). We seek a solution of the form

$$v(n) = e^{ik_1nd} + \Gamma_{11}e^{ik_1nd} + \Gamma_{13}e^{ik_3nd}, \quad n \geq 0, \tag{5.1}$$

$$v(n) = T_{10}e^{-ik_0nd}, \quad n \leq 0. \tag{5.2}$$

The three constants Γ_{11}, Γ_{13}, and T_{10} are as yet to be determined. To this end, we first equate (5.1) to (5.2) at $n = 0$ and then substitute (5.1) and (5.2) into (3.10) to obtain

$$\Gamma_{11} + \Gamma_{13} - T_{10} + 1 = 0, \tag{5.3}$$

$$\Gamma_{11}(1 + e^{ik_1d} - e^{ik_0d}) + \Gamma_{13}(1 + e^{ik_3d} - e^{ik_0d}) + (1 + e^{-ik_1d} - e^{ik_0d}) = i\epsilon^2 \frac{\omega^4 m^2}{2\alpha \sin k_0 d}(1 - e^{-d/a}). \tag{5.4}$$

Inserting (5.1) and (5.2) into (4.16) gives an additional equation

$$(1 - e^{-i(k_1+K)d})^{-1}\Gamma_{11} + (1 - e^{-i(k_3+K)d})^{-1} \times \Gamma_{13} + (1 - e^{i(k_1-K)d})^{-1} = 0. \tag{5.5}$$

Equations (5.3) to (5.5), when solved, yield solutions for Γ_{11}, Γ_{13}, and T_{10} given by

$$\Gamma_{11} = E_{11}/D, \tag{5.6}$$

$$\Gamma_{13} = E_{13}/D, \tag{5.7}$$

$$T_{10} = 1 + (E_{11} + E_{13})/D. \tag{5.8}$$

Here, for brevity, we have defined E_{11}, E_{13}, and D as

$$E_{11} = \frac{1 + e^{-ik_1d} - e^{ik_0d}}{1 - e^{-i(k_3+K)d}} - \frac{1 + e^{ik_3d} - e^{ik_0d}}{1 + e^{i(k_1-K)d}} - i\epsilon^2 \frac{\omega^4 m^2(1 - e^{-d/a})}{2\alpha \sin k_0 d(1 - e^{-2ikd})}, \tag{5.9}$$

$$E_{13} = \frac{1 + e^{ik_1d} - e^{ik_0d}}{1 + e^{i(k_1-K)d}} - \frac{1 + e^{-ik_1d} - e^{ik_0d}}{1 - e^{-i(k_1+K)d}} + i\epsilon^2 \frac{\omega^4 m^2(1 - e^{-d/a})}{2\alpha \sin k_0 d[1 - e^{-i(k_0+K)d}]}, \tag{5.10}$$

$$D = \frac{1 + e^{ik_3d} - e^{ik_0d}}{1 - e^{-i(k_1+K)d}} - \frac{1 + e^{ik_1d} - e^{ik_0d}}{1 - e^{-i(k_3+K)d}}. \tag{5.11}$$

Results (5.6)-(5.8) can be made explicit by expanding (5.9)-(5.11) in powers of ϵ^2. However, we shall not do so because the expanded forms are rather lengthy.

If the incident wave is e^{-ik_3nd}, the corresponding solution is of the form

$$v(n) = e^{-ik_3nd} + \Gamma_{31}e^{ik_1nd} + \Gamma_{33}e^{ik_3nd}, \quad n \geq 0, \tag{5.12}$$

$$v(n) = T_{30}e^{-ik_0nd}, \quad n \leq 0. \tag{5.13}$$

The coefficients Γ_{31}, Γ_{33}, and T_{30} are obtainable from Γ_{11}, Γ_{13}, and T_{10}, respectively, by interchanging k_1 and k_3 in (5.9)-(5.11).

6. REFLECTION AND TRANSMISSION BY A FINITE IMPERFECT CRYSTAL

Let us consider an infinite lattice in which only the point masses at nd, $n = 0, 1, 2, \ldots, N$, are defective, while the remaining ones are uniform. For the reflection and transmission problem, we have to solve Eqs. (3.8) and (3.9) subject to boundary conditions (3.14) to (3.17).

Case A: Uncorrelated mass defects

This is a modification of Case B in Sec. 4 to a finite N. Equation (3.8) reads

$$(\Delta^2 + \beta^2)v(n) + \epsilon^2 \frac{\langle\mu^2\rangle\omega^4 m^2}{\alpha_1} g(\omega)v(n) = 0, \quad 0 < n < N. \tag{6.1}$$

The solution of (6.1) can be written as

$$v(n) = T_1e^{ik_1nd} + T_2e^{ik_2nd}, \quad 0 \leq n \leq N, \tag{6.2}$$

where k_1 and k_2 are two roots of the dispersion Eq. (4.9)

To satisfy the boundary conditions, we use (6.2) in (3.14) to (3.17) to get

$$T_1 + T_2 = 1 + \Gamma, \tag{6.3}$$

$$T_1e^{ik_1d} + T_2e^{ik_2d} = b_1 + b_2\Gamma, \tag{6.4}$$

$$T_1e^{iNk_1d} + T_2e^{iNk_2d} = T, \tag{6.5}$$

$$T_1e^{i(N-1)k_1d} + T_2e^{i(N-1)k_2d} = b_3 + b_2T. \tag{6.6}$$

After solving the system of Eqs. (6.3) to (6.6) and simplifying the results, we obtain, with $L = Nd$

$$T_1 = 1 - \epsilon^2 \frac{\omega^4 m^2\langle\mu^2\rangle}{2\alpha_1\beta(4 - \beta^2)^{1/2}}\left(\frac{2 - \beta^2}{\beta(4 - \beta^2)^{1/2}} + 3i\right) \times g(\omega) + O(\epsilon^4), \tag{6.7}$$

$$T_2 = i\epsilon^2 \frac{\omega^4 m^2\langle\mu^2\rangle}{2\alpha_1\beta(4 - \beta^2)^{1/2}}\left(1 + i\frac{2 - \beta^2}{\beta(4 - \beta^2)^{1/2}}\right) \times g(\omega)e^{2ik_1L} + O(\epsilon^4), \tag{6.8}$$

$$\Gamma = -\frac{\epsilon^2\omega^4 m^2\langle\mu^2\rangle}{2\alpha_1\beta(4 - \beta^2)^{1/2}}\left(\frac{2 - \beta^2}{\beta(4 - \beta^2)^{1/2}}(1 + e^{2ik_1L}) + i(3 - e^{2ik_1L})\right)g(\omega) + O(\epsilon^4), \tag{6.9}$$

$$T = \left[1 - \epsilon^2 \frac{\omega^4 m^2\langle\mu^2\rangle}{\alpha_1\beta(4 - \beta^2)^{1/2}}\left(\frac{2 - \beta^2}{2\beta(4 - \beta^2)^{1/2}} - i\right) \times g(\omega)\right]e^{ik_1L} + O(\epsilon^4), \tag{6.10}$$

We note that, as $L \to \infty$, T_2 and T approach zero, and T_1 and Γ yield, respectively, the transmission and reflection coefficients given by (4.11) and (4.12) for a semi-infinite imperfect crystal. This is a consequence of the positivity of $\mathrm{Im}\{k_1\}$ so that the exponential functions e^{ik_1L} and e^{2ik_1L} decay in L.

If we specialize our results (6.7)–(6.10) to the one-dimensional case, they become

$$T_1 = 1 - i\epsilon^2 \frac{\langle \mu^2 \rangle}{8}\left[\left(\frac{\omega_c}{\omega}\right)^2 - 1\right]^{-1}\left(\frac{(\omega_c/\omega)^2 - 2}{(\omega_c/\omega)^2 - 1} + 3i\right) + O(\epsilon^4), \quad (6.11)$$

$$T_2 = -\epsilon^2 \frac{\langle \mu^2 \rangle}{8}\left[\left(\frac{\omega_c}{\omega}\right)^2 - 1\right]^{-1}\left(1 + i\frac{(\omega_c/\omega)^2 - 2}{(\omega_c/\omega)^2 - 1}\right) \times e^{2ikL} + O(\epsilon^4), \quad (6.12)$$

$$T = -i\epsilon^2 \frac{\langle \mu^2 \rangle}{8}\left[\left(\frac{\omega_c}{\omega}\right)^2 - 1\right]^{-1}\left(\frac{(\omega_c/\omega)^2 - 2}{(\omega_c/\omega)^2 - 1} \times (1 + e^{2ikL}) + i(3 - e^{2ikL})\right) + O(\epsilon^4), \quad (6.13)$$

$$T = 1 - i\epsilon^2 \frac{\langle \mu^2 \rangle}{8}\left[\left(\frac{\omega_c}{\omega}\right)^2 - 1\right]^{-1}\left(\frac{(\omega_c/\omega)^2 - 2}{(\omega_c/\omega)^2 - 1} - \frac{i}{2}\right) \times e^{ikL} + O(\epsilon^4), \quad (6.14)$$

Case B: Exponentially correlated mass defects in one-dimension

This is the finite version of Case C in Sec. 4. We may write down the solution as follows:

$$v(n) = e^{ik_0nd} + \Gamma e^{-ik_0nd}, \quad n \le 0, \quad (6.15)$$

$$v(n) = \sum_{j=1}^{4} T_j e^{ik_jnd}, \quad 0 \le n \le N, \quad (6.16)$$

$$v(n) = Te^{ik_0nd}, \quad n \ge N. \quad (6.17)$$

Imposing the boundary conditions (3.14) to (3.17) gives four equations for the unknown constants $\Gamma, T_1, \cdots, T_4, T$

$$\sum_{j=1}^{4} T_j = 1 + \Gamma, \quad (6.18)$$

$$\sum_{j=1}^{4} T_j e^{ik_jd} = b_1 + b_2\Gamma, \quad (6.19)$$

$$\sum_{j=1}^{4} T_j e^{ik_jL} = T, \quad (6.20)$$

$$\sum_{j=1}^{4} T_j e^{ik_j(L-d)} = b_3 + b_2T. \quad (6.21)$$

Since there are six unknowns, we need two more equations. The additional two equations can be obtained by inserting (6.16) into (4.18) when $0 < n < N$. This procedure yields

$$\sum_{j=1}^{4} T_j(1 - e^{i(k_j - K)d})^{-1} = 0, \quad (6.22)$$

$$\sum_{j=1}^{4} T_j e^{i(k_j + K)(L+d)}(1 - e^{i(k_j - K)d})^{-1} = 0. \quad (6.23)$$

Now Eqs. (6.18) to (6.23) determine the six constants uniquely. To solve them, let D_1 denote the determinant of a 4×4 matrix $[d_{ij}]$ with elements d_{ij}, $i, j = 1, 2, 3, 4$; defined by

$$d_{1j} = e^{ik_0d} - e^{-ik_0d}, \quad (6.24)$$

$$d_{2j} = e^{ik_j(L-d)} - e^{-ik_0d}, \quad (6.25)$$

$$d_{3j} = (1 - e^{i(k_j - K)d})^{-1}, \quad (6.26)$$

$$d_{4j} = e^{i(k_j + K)(L+d)}(1 - e^{i(k_j - K)d})^{-1}. \quad (6.27)$$

Further, let D_{ij} designate the cofactors of the matrix $[d_{ij}]$; E_j the determinants with jth column of D_1 replaced by ones; F_k, the determinant obtained from D_1 with the kth column replaced by e^{ik_jL}, $j = 1, 2, 3, 4$. Then the solutions of Eqs. (6.18) to (6.23) read

$$\Gamma = \frac{i}{D}\left(2E_1 \sin k_0d - \epsilon^2(E_1 + E_2) \times \frac{\omega^2 m\langle \mu^2 \rangle(1 - e^{ik(L+d)})}{2\alpha \sin k_0d(1 - e^{iKd})}\right) - 1, \quad (6.28)$$

$$T_j = \frac{i}{D}\left(2D_{1j} \sin k_0d - \epsilon^2(D_{1j} + D_{2j}) \times \frac{\omega^2 m\langle \mu^2 \rangle(1 - e^{iK(L+d)})}{2\alpha \sin k_0d(1 - e^{iKd})}\right), \quad j = 1, 2, 3, 4, \quad (6.29)$$

$$T = \frac{i}{D}\left(2F_1 \sin k_0d - \epsilon^2(F_1 + F_2)\frac{\omega^2 m\langle \mu^2 \rangle(1 - e^{iK(L+d)})}{2\alpha \sin k_0d(1 - e^{iKd})}\right). \quad (6.30)$$

Here we shall not expand these formulas in ϵ for compactness.

7. THE SCATTERING BY A SEMI-INFINITE IMPERFECT CRYSTAL WITH AN ARBITRARY CORRELATION FUNCTION

Let us generalize the problems considered in Sec. 8 to that of an arbitrary, homogeneous and istropic correlation function. To solve this problem, we shall adapt the Weiner–Hopf method to the case of a discrete variable. We first introduce a sequence $h'(n)$ defined as

$$h'(n) = (\Delta^2 + \beta^2)v(n) + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1} \sum_{n'=0}^{\infty} \widehat{GR}(|n - n'|, \omega)V(n'), \quad -\infty < n < \infty, \quad (7.1)$$

where $\widehat{GR}(|n - n'|, \omega)$ was given by (3.8). We note that

$$h'(n) = 0, \quad n > 0. \quad (7.2)$$

In order to apply the conventional Wiener–Hopf method, we extend the definitions of functions of discrete variable $f(n)$ (or sequences) to that of functions of a continuous variable $f(x)$ as follows

$$f(x) = f(n), \quad (n - \tfrac{1}{2})d < x \le (n + \tfrac{1}{2})d, \quad n = 0, \pm 1, \pm 2, \cdots. \quad (7.3)$$

With the aid of (7.3), Eqs. (7.1) and (7.2) can be rewritten as

$$h(x) = (\Delta_x^2 + \beta^2)v(x) + \frac{\epsilon^2\omega^4 m^2}{\alpha_1 d}\int_0^{\infty} \widehat{GR}(|x - y|, \omega)v(y)dy, \quad -\infty < x < \infty. \quad (7.4)$$

Here, we have put

$$h(x) = h'(x) - \tfrac{1}{2}\epsilon^2 \frac{\omega^4 m^2}{\alpha_1 d} \widehat{GR}(|x|,\omega)v(0), \quad x < 0, \tag{7.5}$$

$$= -\tfrac{1}{2}\epsilon^2 \frac{\omega^4 m^2}{\alpha_1 d} \widehat{GR}(|x|,\omega)v(0), \quad x > 0, \tag{7.6}$$

and

$$\Delta_x^2 v(x) = v(x+d) - 2v(x) + v(x-d). \tag{7.7}$$

Let the Fourier transform of a function $f(x)$ be denoted by $\tilde{f}(\lambda)$ which is given by

$$\tilde{f}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} f(x)dx \tag{7.8}$$

and by $\tilde{f}_+(\lambda)$ or $\tilde{f}_-(\lambda)$ if the range of integration in (7.8) is over $(0,\infty)$ or over $(-\infty,0)$. Taking the Fourier transform of (7.4) yields

$$\tilde{h}(\lambda) = [2(\cos\lambda d - 1) + \beta^2]\tilde{v}_+(\lambda) + \frac{\epsilon^2\omega^4 m^2}{\alpha_1 d} \times [\widetilde{GR}](\lambda,\omega)v_+(\lambda) + r(\lambda), \tag{7.9}$$

where

$$\begin{aligned} r(\lambda) = {} & 4\frac{\sin\lambda d/2\,(\cos\lambda d - 1)}{\lambda}[(1 - e^{i(\lambda - k_0)d})^{-1} \\ & + \Gamma(1 - e^{i(\lambda + k_0)d})^{-1} - i\frac{(b_1 + b_2\Gamma)}{\lambda}(1 - e^{i\lambda d/2}) \\ & + i(e^{-ik_0 d} + \Gamma e^{ik_0 d})\frac{e^{-i\lambda d/2} - 1}{\lambda} + i2\frac{(1+\Gamma)}{\lambda} \\ & \times [(\cos k_0 d - 1)(1 - e^{i\lambda d/2}) + (\cos\lambda d - \cos\lambda d/2)]. \end{aligned} \tag{7.10}$$

In obtaining the above equations, use has been made of the boundary conditions (3.14) and (3.15). We assume that the coefficient of $v_+(\lambda)$ can be factored into the form

$$2(\cos\lambda d - 1) + \beta^2 + \epsilon^2 \frac{\omega^4 m^2}{\alpha_1 d}[\widetilde{GR}](\lambda,\omega) = \frac{m_+(\lambda)}{m_-(\lambda)}, \tag{7.11}$$

where m_+ and m_- are analytic functions of λ in the respective regions $\mathrm{Im}\{\lambda\} \geq 0$ and $\mathrm{Im}\{\lambda\} \leq 0$. Then Eq. (6.9) becomes

$$m_+(\lambda)\tilde{v}_+(\lambda) = m_-(\lambda)\{\tilde{h}_-(\lambda) - r(\lambda)\}. \tag{7.12}$$

Since by definition, $\tilde{v}_+(\lambda)$, is analytic for $\mathrm{Im}\{\lambda\} \geq 0$ and $\tilde{h}_-(\lambda)$ is analytic for $\mathrm{Im}\{\lambda\} \leq 0$, the left-hand side and the right-hand side of (6.12) have domains of analyticity similar to that of m_+ and m_-. As the real axis is the common domain of analyticity, each side can be continued analytically into the whole λ plane. Furthermore, each side must be an entire function $e(\lambda)$. Therefore, we have $m_+\tilde{v}_+ = e$ and

$$\tilde{v}_+(\lambda) = e(\lambda)/m_+(\lambda). \tag{7.13}$$

The entire function $e(\lambda)$ can be determined by regularity conditions at infinity and the boundary conditions (3.14) and (3.15). The solution $v(n)$ is thus obtained by taking the inverse transform of (6.13) and setting $x = nd$.

8. MEAN GREEN'S FUNCTION FOR AN INFINITE LATTICE

We wish to find the mean Green's function H_0 which satisfies the equation

$$\Big(\sum_{j=1}^{3} \alpha_j \Delta_j^2 + \omega^2 m\Big)H_0(n,n') + \epsilon^2\omega^4 m^2 \sum_1 G(n,l)R(n,l)H_0(l,n') = \delta_{nn'} \tag{8.1}$$

and the outgoing condition at infinity. Here $\delta_{nn'}$ is an abbreviation for the triple product of Kronecker deltas $\Pi_{j=1}^3 \delta_{n_j n_j'}$.

We first consider a one-dimensional case with an exponential correlation function, then indicate how to treat the general problem. In this case, Eq. (7.1) reduces to

$$\Big(\Delta^2 + \frac{\omega^2 m}{\alpha}\Big)H_0(n,n') - \epsilon^2 \frac{\omega^4 m^2\langle\mu^2\rangle}{2i\alpha\,\sin k_0 d} \sum_{l=-\infty}^{\infty} \times e^{iK|n-l|d}H_0(l,n') = \frac{1}{\alpha}\delta_{nn'}. \tag{8.2}$$

To get an outgoing solution of (8.2), we try a solution of the form

$$H_0(n,n') = \Omega(e^{ik_1|n-n'|d} - \Theta e^{ik_3|n-n'|d}), \tag{8.3}$$

where k_1 and k_3 were defined by Eq. (4.20) and Ω, Θ are constants to be determined. For (8.3) to satisfy (8.2), we substitute the former into the latter, for $n \neq n'$, to obtain

$$\begin{aligned} & \Lambda\Big[\Big(2(\cos k_1 d - 1) + \frac{\omega^2 m}{\alpha} - \epsilon^2\langle\mu^2\rangle \\ & \times \frac{\omega^4 m^2 \sin k_1 d}{2\alpha\,\sin k_0 d\,(\cos k_1 d - \cos Kd)}\Big)e^{ik_1|n-n'|d}\Big] \\ & - \Theta\Big[\Big(2(\cos k_3 d - 1) + \frac{\omega^2 m}{\alpha} - \epsilon^2\langle\mu^2\rangle \\ & \times \frac{\omega^4 m^2 \sin k_1 d}{2\alpha\,\sin k_0 d\,(\cos k_3 d - \cos Kd)} e^{ik_3|n-n'|d} \\ & + \epsilon^2 \frac{\omega^4 m^2}{\alpha\,\sin k_0 d}\Big(\frac{\sin k_1 d}{\cos k_1 d - \cos Kd} - \Theta\frac{\sin k_3 d}{\cos k_3 d - \cos Kd}\Big) \\ & \times e^{iK|n-n'|d}\Big] = 0. \end{aligned} \tag{8.4}$$

Invoking (4.20), (8.4) can be reduced to

$$\frac{\sin k_1 d}{\cos k_1 d - \cos Kd} - \Theta\frac{\sin k_3 d}{\cos k_3 d - \cos Kd} = 0. \tag{8.5}$$

For $n = n'$, the above substitution yields

$$\begin{aligned} & \Lambda\Big(2(e^{ik_1 d} - \Theta e^{ik_3 d}) - (1 - \Theta)\Big(2 - \frac{\omega^2 m}{\alpha}\Big) \\ & + \epsilon^2 \frac{\omega^4 m^2\langle\mu^2\rangle}{2i\alpha^2\,\sin k_0 d} i\frac{\sin k_1 d - \sin Kd}{\cos k_1 d - \cos Kd} \\ & - i\Theta\frac{\sin k_3 d - \sin Kd}{\cos k_3 d - \cos Kd}\Big) = \frac{1}{\alpha}. \end{aligned} \tag{8.6}$$

Upon solving (8.5) and (8.6), we obtain

$$\Theta = \sin k_1 d\,(\cos k_1 d - \cos Kd)/2i\alpha\,(\cos k_1 d - \cos k_3 d), \tag{8.7}$$

$$\Lambda = \frac{\cos k_1 d - \cos Kd}{2i\alpha\,(\cos k_1 d - \cos k_3 d)}. \tag{8.8}$$

In obtaining (8.8), use has been made of Eq. (4.20)

which holds for an infinite lattice as well. This completes the construction of the mean Green's function H_0.

For an arbitrary correlation function in three dimensions, the Fourier transform technique may be employed to solve (8.1). The answer will be explicit if the inversion is possible.

9. MEAN GREEN'S FUNCTION FOR A SEMI-INFINITE LATTICE

We shall now construct the mean Green's function $H(n,n')$ corresponding to the special Case (C) in Sec. 4. To this end, it is found convenient to treat two cases separately according to the location of a unit source in the right or in the left half-planes. For a source placed at $n'd$ with $n' < 0$, the mean wave excited must satisfy the equations

$$\left(\Delta^2 + \frac{\omega^2 m}{\alpha}\right) H(n,n') - \epsilon^2 \frac{\omega^4 m^2 \langle \mu^2 \rangle}{2i\alpha \sin k_0 d} \sum_{l=0}^{\infty} \times e^{iK|l-n|d} H(l,n') = 0, \quad n > 0, \quad n' < 0, \tag{9.1}$$

$$\left(\Delta^2 + \frac{\omega^2 m}{\alpha}\right) H(n,n') = \frac{1}{\alpha}\delta_{nn'}, \quad n,n' < 0. \tag{9.2}$$

Also, we require $H(n,n')$ to conform with an analog of boundary conditions (3.14) and (3.15). For $n < 0$, we write $H(n,n')$ as

$$H(n,n') = \frac{e^{ik_0|n-n'|d}}{2i\alpha \sin k_0 d} + U(n,n'), \quad n,n' \le 0, \tag{9.3}$$

where $U(n,n')$ is a solution to the homogeneous Eq. of (9.2).

Since a unit source at $n'd$ produces a plane wave incident upon the random lattice, it must be reflected and transmitted. U is therefore a reflected plane wave, so that

$$H(n,n') = \frac{1}{2i\alpha \sin k_0 d}\left(e^{ik_0|n-n'|d} + \Gamma e^{-ik_0(n+n')d}\right), \quad n,n' \le 0 \tag{9.4}$$

where Γ is the reflection coefficient given by (4.26). With this interpretation, one sees clearly that the transmitted waves are

$$H(n,n') = \frac{1}{2i\alpha \sin k_0 d} T_1 e^{in'd}\left(e^{ik_1 nd} - \Omega e^{ik_3 nd}\right), \quad n \ge 0, \quad n' \le 0. \tag{9.5}$$

Here T_1 and Ω are the same constants found before [see (4.24) and (4.27)].

When the source lies in the right half-plane, we have to interchange the right-hand sides of (9.1) and (9.2). Let us express $H(n,n')$ in the form

$$H(n,n') = H_0(n,n') + V(n,n'), \quad n,n' \ge 0, \tag{9.6}$$

in which H_0 is the Green's function computed in the previous section and $V(n,n')$ is a solution to the homogeneous equation (9.1). In view of (8.3), H_0 constitutes two plane waves incident from the right half-plane. Again this may be viewed as a reflection and transmission problem which was solved in Sec. 5. The function V in (9.6) represents the reflected field. Making use of the results obtained before, we have

$$H(n,n') = H_0(n,n') + \Lambda[(\Gamma_{11}e^{ik_1 nd} + \Gamma_{13}e^{ik_3 nd})e^{ik_1 n'd} - \Theta(\Gamma_{31}e^{ik_1 nd} + \Gamma_{33}e^{ik_3 nd})e^{ik_3 n'd}], \quad n,n' \ge 0. \tag{9.7}$$

The corresponding transmitted wave is given by

$$H(n,n') = \Lambda(T_{10}e^{ik_1 n'd} - \Theta T_{30}e^{ik_3 n'd})e^{-ik_0 nd}, \quad n \le 0, \quad n' \ge 0. \tag{9.8}$$

Equations (9.4), (9.5), (9.7), and (9.8), when taken all together, specify the mean Green's function $H(n,n')$ completely.

It is noted that, in general, the mean Green's can be obtained by the Wiener–Hopf method in a similar manner to that we presented in Sec. 7.

APPENDIX: APPROXIMATE EVALUATIONS OF $g(\omega)$ IN TWO AND THREE DIMENSIONS

The Green's function (1.5) can be rewritten as a single integral involving an exponential function and a triple product of Bessel functions (see Ref. 14):

$$G(\mathbf{n},\mathbf{m}) = \left(\frac{1}{2\pi}\right)^N (i)^{\sum_{j=1}^N l_j + N} \lim_{\epsilon \to 0^+} \int_0^\infty \exp\left[it\left(M_0\omega^2 - 2 \times \sum_{j=1}^N K_j \pm i\epsilon\right)\right] \prod_{j=1}^N J_{l_j}(2K_j t)dt, \tag{A1}$$

where N is the dimensions of the lattice and $l_j = n_j - m_j$, $j = 1,2,3$.

When $\mathbf{n} = \mathbf{m}$, it yields

$$g(\omega) = \left(\frac{i}{2\pi}\right)^N \lim_{\epsilon \to 0^+} \int_0^\infty \exp\left[it\left(M_0\omega^2 - 2\sum_{j=1}^N K_j \pm i\epsilon\right)\right] \times \prod_{j=1}^N J_0(2K_j t)dt. \tag{A2}$$

For $N = 2$, $g(\omega)$ can be expressed in terms of the complete elliptic integral of the first kind K as follows (p. 314, Ref. 13):

$$g(\omega) = (-1/\sqrt{K_1 K_2})(1/2\pi)^3 zK(z) \tag{A3}$$

in which z was defined by (1.11).

Since K is related to the hypergeometric function F by $K(z) = (\pi/2)F(\frac{1}{2},\frac{1}{2};1;z)$ (see p. 591, Ref. 13) for z real and $z < 1$, it follows by the principle of analytic continuation, that the expression (1.10) holds for other values of z in the domain of analyticity of F in the complex z plane.

For $N = 3$ and $K_2 = K_3$, (A2) yields

$$g(\omega) = -i\left(\frac{1}{2\pi}\right)^3 \lim_{\epsilon \to 0^+} \int_0^\infty \exp\left[it\left(M_0\omega^2 - 2\sum_{j=1}^3 K_j \pm i\epsilon\right)\right] \times J_0(2K_1 t)J_0^2(2K_2 t)dt. \tag{A4}$$

By using the identity (p. 294, Ref. 14)

$$J_0^2(2K_2 t) = \left(\frac{2}{\pi}\right)\int_0^{\pi/2} J_0(4K_2 t\cos\phi)d\phi, \tag{A5}$$

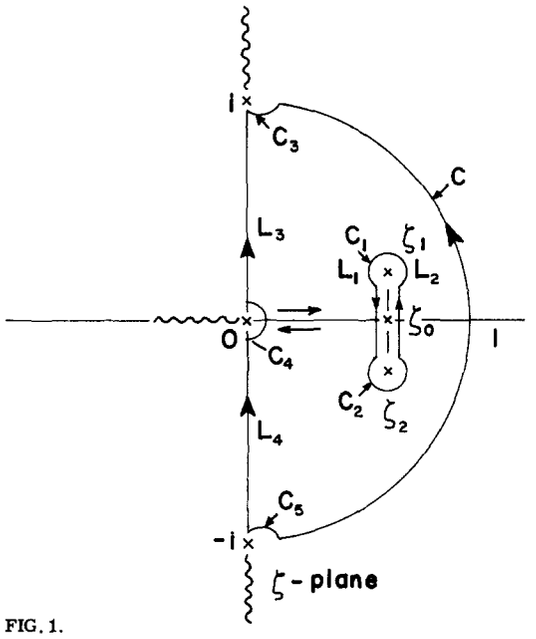


FIG. 1.

(A4) becomes after an interchange of order of integrations

$$g(\omega) = \left(\frac{1}{2\pi}\right)^4 \int_0^{\pi/2} \int_0^{\infty} \exp\left[it\left(M_0\omega^2 - 2\sum_{j=1}^{3} K_j\right)\right] \times J_0(2K_1 t) J_0(4K_2 t \cos\phi) dt d\phi. \quad \text{(A6)}$$

The integral with respect to t in (A6) is similar to (A2) for $N = 2$; therefore, we may apply the previous results to obtain

$$g(\omega) = -i\left(\frac{1}{2\pi}\right)^4 \frac{1}{\sqrt{K_1 K_2}} \int_0^{\pi/2} (\cos\phi)^{-1/2} W^{1/2} F\left(\tfrac{1}{2}, \tfrac{1}{2}; 1; W^{1/2}\right) d\phi. \quad \text{(A7)}$$

Here we have defined W to be

$$W = 4(2K_1 K_2 \cos\phi)^{1/2} \{[M_0\omega^2 + 4K_1(\cos\phi - 1)] \times [M_0(\omega_c^2 - \omega^2) + 4K_2(\cos\phi - 1)]\}^{-1/2}. \quad \text{(A8)}$$

By letting $\zeta = e^{i\theta}$, and extending the range of integration in (A7) to $-\pi/2$, we arrive at the expression (1.12) for $g(\omega)$ as a contour integral along a semicircle C as shown. By closing the contour and applying Cauchy's theorem, we conclude that

$$\int_C = \sum_{j=1}^{4} \int_{L_j} + \sum_{k=1}^{5} \int_{c_k}. \quad \text{(A9)}$$

Along the path and inside the contour, the singular (branch) points are $0, \pm i, \zeta_1, \zeta_2$, where ζ_j are the roots of $\zeta^2 + 2(\delta_1 - 1)\zeta + 1 = 0$ when δ_1 is small. It is found that $\int_{c_k} = 0$ and the main contributions are due to $\int_{L_1}$ and $\int_{L_2}$. Since $|\zeta_1 - \zeta_2| \approx 2\sqrt{2\delta_1}$ is small, in the integration, we approximate the regular parts of integrand $f(\zeta)$ by $f(\zeta_0)$, where $\zeta_0 = 1 - \delta_1 \approx 1$. In this way we obtain

$$\int_{L_1} \sim \int_{-(2\delta_1)^{1/2}}^{(2\delta_1)^{1/2}} [2\delta_2(\eta^2 - 2\delta_1)]^{-1/2} F\left\{\tfrac{1}{2}, \tfrac{1}{2}; 1; 2[2\delta_2 (\eta^2 - 2\delta_1)]^{-1/2}\right\} d\eta. \quad \text{(A10)}$$

By using the asymptotic formula (1.12) and effecting a trigonometric substitution, we arrive at the result

$$\int_{L_1} + \int_{L_2} \sim \pm\, 2i(2\delta_1)^{1/4}(2\delta_2)^{1/2}(1/\pi)[(4\ln 2 + \ln(1/\delta_1) + \sqrt{\pi + \pi i}) \int_0^{\pi/2} \cos^{1/2}\theta d\theta + \int_0^{\pi/2} \cos^{1/2}\theta \ln\sec\theta d\theta]. \quad \text{(A11)}$$

When (A11) is used in (A9) and (1.15), the result (1.19) follows.

ACKNOWLEDGMENT

This work is a continuation of our previous work, Ref. 1, on this subject which was initiated at the suggestion of Professor J. B. Keller. The author is grateful to Professor Keller for reading a copy of the original manuscript and his valuable comments which lead to an improvement of this paper.

*This work was supported in part by NSF Grant GP-32996X2.

[1] P. L. Chow and J. B. Keller, J. Math. Phys. **13**, 1404 (1972).
[2] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation, Solid State Physics, Suppl.* 3 (Academic, New York, 1971), 2nd ed.
[3] I. M. Lifshitz, Zh. Eksp. Teor. Fiz. **18**, 293 (1948).
[4] I. M. Lifshitz, Nuovo Cimento Suppl. **3**, 716 (1956).
[5] G. F. Koster, Phys. Rev. **95**, 1436 (1954).
[6] M. V. Klein, Phys. Rev. **131**, 1500 (1963).
[7] J. A. Khrumhansl in *Lattice Dynamics*, edited by R. F. Wallis (Pergamon, Oxford, 1965), p. 523.
[8] J. Callaway, J. Math. Phys. **5**, 783 (1964).
[9] R. J. Rubin, J. Math. Phys. **9**, 2252 (1968).
[10] Y. Yvon, C.R. Acad. Sci. (Paris) **202**, 35 (1936).
[11] J. Kirkwood, J. Chem. Phys. **4**, 592 (1936).
[12] I. Kupiec, L. B. Felson, S. Rosenbaum, J. B. Keller, and P. L. Chow, Radio Sci. **4**, 1067 (1969).
[13] *Handbook of Mathematical Functions*, edited by M. Abramovitz and I. A. Stegun (Natl. Bur. Stand., Washington, D. C., 1964).
[14] Y. E. Luke, *Integrals of Bessel Functions* (McGraw-Hill, New York, 1962).

Entropy production in the functional random-walk model

I. Hosokawa

Department of Mechanical Engineering, Iwate University, Morioka 020, Japan

(Received 20 November 1972)

An information-theoretical background is presented here for the functional random-walk model of a many-particle system, which was recently proposed to simulate nonequilibrium statistical mechanics in a certain coarse-graining sense. Next, entropy productivity and the maximum-entropy state in the model dynamics are studied with the new definition of entropy, which turns out to be a natural extension of the original Boltzmann entropy.

1. INTRODUCTION

It has been a central problem to derive the macroscopically observed, irreversible development of a many-particle system from the exact mechanical basis represented by the Liouville equation. For example, the ergodic theory may say how to deal with an equilibrium state of matters, but "gives no indication how quickly systems approach equilibrium" (Prigogine[1]). Therefore, it is natural to consider that some kind of operation independent of mechanics should be added in order to obtain a satisfactory explanation of such an irreversible development.[2] Such an operation, however, does not seem to have been generally established yet, although we have seen many operations, case by case, that might be called "ad hoc approximations." (E.g., Stosszahlansatz. A very general one was recently discussed by Prigogine,[1] but the whole dynamics including this is far from simplifying the Liouville dynamics.) It is for this reason that the author has explored a new irreversible dynamics in as general and as simple a way as possible to filter the irreversible behavior out of the many-particle motion based on the exact dynamics. This dynamics was recently formulated into the functional random-walk model of a many-particle system.[3]

This model dynamics looks quite simple in structure. We treat, rather than 6N-dimensional phase points, *stochastic* particle-number density fields in the one-body phase space. The equation for the probability distribution functional of the field is written in the Fokker-Planck form, which guarantees the irreversibility in question only if there exists any particle interaction. Such an operation is clearly defined in the theory to make it link to the exact dynamics, though some ambiguity was left in its true physical meaning despite our attempt to relate it with coarse graining. Here, the physical meaning of the operation will be discussed again from another angle, i.e., from the information-theoretical point of view (in Sec. 2).

If we have only a macroscopic (mechanically very imperfect) knowledge of a system at the initial time, we will probably take the most *unbiased* initial condition for this knowledge in the exact dynamics. (This idea was adopted in plasma physics.[4] Such a limitation of initial conditions may result in prohibiting reversed processes, though any entropy production is not included.) However, when we macroscopically observe the system at some time later, is the information supplied by the exact dynamics still the most unbiased for the observed quantities? The answer will be "no." But why "no"? If we consider this time as a new initial time, the most unbiased condition for the macroscopic knowledge at hand should then be taken. So long as the information-theoretical idea is consistently pursued, we must discontinuously change the condition on the exact dynamics at every macroscopic observation. Since the information entropy for the most unbiased condition is larger than any other and since the exact dynamics conserves the total entropy of a closed system, the information entropy must jump at every macroscopic observation. Then, what happens, if we observe the system macroscopically in a continuous manner? It is in this limit that a new entropy-productive dynamics, the functional random-walk model, may replace the exact dynamics. We may consider this idea as a dynamic version of Jaynes' statistical theory.[5]

Validity of the new dynamics is, however, to be finally judged by whether it can be well connected with the known equilibrium statistical mechanics of a steady state. Therefore, after entropy productivity of the dynamics is proved, the state with the maximum entropy is presented explicitly (in Sec. 3). This was left as a future problem in the previous work.[3] It is further shown that the mode of the stochastic particle-number density field for this state is equal to the Maxwell-Boltzmann distribution with the particle-interaction energy included in a *self-consistent* way (in Sec. 4). This result is of the same quality as Jaynes' in his information-theoretical approach to an equilibrium system, and can be well compared with the Boltzmann and Gibbs equilibrium states. It is finally verified that our information entropy reduces to the original Boltzmann entropy when the state is in equilibrium and if the number density of particles is large (compared to one).

For simplicity, the discussion will be given for a closed, single-component many-particle system governed by classical mechanics.

2. INFORMATION-THEORETICAL BACKGROUND

According to the previous work,[3] the basic Fokker-Planck equation in the functional random-walk model is formally written as

$$\frac{\partial \tilde{\rho}}{\partial t} = -\int_X \frac{\delta}{\delta z(x)}[Qz(x)\tilde{\rho}]dx + \frac{1}{2}\int_X\int_X \frac{\delta^2}{\delta z(x)\delta z(x')} \times \{\mathcal{P}[\phi(|q-q'|);z(x)z(x')]\tilde{\rho}\}dxdx', \quad (2.1)$$

which governs the probability distribution functional $\tilde{\rho}$ of the particle-number density field $nz(x)$ in the one-body phase space X ($\ni x$). Here, n is the average number density, t the time variable, $\delta/\delta z(x)$ the functional derivative with respect to z at x, $\mathcal{P}$ the operator which makes the subsequent symmetric function positive definite, $[\ ;\ ]$ the Poisson bracket, $\phi(q)$ the interaction potential between particles, q being the displacement vector part of x, and Q is the nonlinear operator defined as

$$Qz(x) = [H_1(x);z(x)] + n\int_X[\phi(|q-q'|);z(x)z(x')]dx', \quad (2.2)$$

in which $H_1(x)$ is the one-body Hamiltonian; Q may be called the Vlasov operator, since $\partial f/\partial t = Qf$ gives the self-consistent Vlasov equation.[2,3]

Associated with (2.1), there are three conditions. First,

$$\int_B \tilde{\rho}\delta z = 1; \tag{2.3}$$

this is the normalization for $\tilde{\rho}$ to be a probability distribution, and $\int_B \delta z$ denotes a functional integral over the function space B. In order to prescribe B, we have

$$\lim_{V\to\infty} \int_X z(x)dx/V = 1 \tag{2.4}$$

and

$$\lim_{V\to\infty} \{\int_X nH_1(x)z(x)dx + \int_X\int_X \tfrac{1}{2}n^2\phi(|q-q'|)z(x)z(x')dxdx'\}/V = \text{const}, \tag{2.5}$$

where V is the volume containing the whole system; the latter gives the average energy of the closed system in question.

The previous work shows that if the operator $\mathcal{P}$ is removed from (2.1), the equation becomes perfectly equivalent to the exact dynamics, the Liouville equation. In this case, however, the physical meaning of $\tilde{\rho}$ and z as described above should be completely lost. In other words, our (macroscopic) picture of the system as an entirety of random motion of particle-number density fields is tied to a slight mathematical modification of the basic equation by $\mathcal{P}$. We will then, in the exact dynamics, distinguish the dependent variable as ρ from $\tilde{\rho}$. Originally, $\rho(z, t)$ was introduced as the functional Fourier transform of a *state-functional*.[3]

A general solution for $\tilde{\rho}$ may be given in the form of a repeated multiple integral. That is,

$$\tilde{\rho}(z,t) = \lim_{\Delta t\to 0} \int_B \cdots \int_B P_{\Delta t}(z^L/z^{L-1}) \cdots P_{\Delta t}(z^2/z^1)\tilde{\rho}(z^1,0)\prod_{k=1}^{L-1}\delta z^k, \tag{2.6}$$

where $z = z^L$, $\Delta t = t/L$, and the superscripts indicate the order of the time subintervals. The infinitesimal kernel $P_{\Delta t}$ is explicitly obtained from (2.1) as

$$P_{\Delta t}(z^{k+1}/z^k) = \int_B \exp\{i\int_X y^k(x)[z^k(x) - z^{k+1}(x) + \Delta t Q z^k(x)]dx - \tfrac{1}{2}\Delta t\int_X\int_X y^k(x)y^k(x') \mathcal{P}[\phi; z^k(x)z^k(x')]dxdx'\}\delta y^k. \tag{2.7}$$

In parallel with this, we can write a general solution for ρ in the exact dynamics; namely

$$\rho(z,t) = \lim_{\Delta t\to 0} \int_B \cdots \int_B K_{\Delta t}(z^L/z^{L-1}) \cdots K_{\Delta t}(z^2/z^1)\rho(z^1,0)\prod_{k=1}^{L-1}\delta z^k. \tag{2.8}$$

$K_{\Delta t}$ is also obtained, using the expression (2.7), as

$$K_{\Delta t}\left(\frac{z^{k+1}}{z^k}\right) = P_{\Delta t}\left(\frac{z^{k+1}}{z^k}\right) + \tfrac{1}{2}\Delta t\int_X\int_X \frac{\delta^2}{\delta z^{k+1}(x)\delta z^{k+1}(x')} \times (1-\mathcal{P})[\phi; z^k(x)z^k(x')]P_{\Delta t}(z^{k+1}/z^k)\,dxdx'. \tag{2.9}$$

From this relation we can know to what degree the two dynamics differ from each other.

For instance, let us see the development in Δt of F_1 and $\tilde{F}_1$, the one-body distribution functions in both dynamics, remembering that[3]

$$F_1(x,t) = \int_B z(x)\rho(z,t)\delta z, \tag{2.10}$$

$$\tilde{F}_1(x,t) = \int_B z(x)\tilde{\rho}(z,t)\delta z. \tag{2.11}$$

If $F_1 = \tilde{F}_1$ and $\rho = \tilde{\rho}$ at time t^k, it is easy to prove still that $F_1 = \tilde{F}_1$ but $\rho \neq \tilde{\rho}$ at $t^{k+1} = t^k + \Delta t$, by performing the functional integrations [derived from (2.10), (2.11)],

$$F_1(x,t^{k+1}) = \int_B\int_B z^{k+1}(x)K_{\Delta t}(z^{k+1}/z^k)\rho(z^k,t^k)\delta z^k\delta z^{k+1}, \tag{2.12}$$

$$\tilde{F}_1(x,t^{k+1}) = \int_B\int_B z^{k+1}(x)P_{\Delta t}(z^{k+1}/z^k)\tilde{\rho}(z^k,t^k)\delta z^k\delta z^{k+1}, \tag{2.13}$$

respectively, while such a perfect equality does not occur with all the other many-body distribution functions (F_s for $s \geq 2$). Thus, it is obvious that the model dynamics is not able to compete with the exact dynamics in predicting a change of the microscopic state, but that macroscopically we would hardly discriminate the predictions from both of a change of the state in Δt, if almost all macroscopic information comes from the one-body distribution function.

If this is the case, which is more suitable for the macroscopic description of a system, ρ or $\tilde{\rho}$? According to the information-theoretical argument for statistical physics,[5] reality of a distribution may be measured by the magnitude of information entropy for the same amount of restricted knowledge of the system. It is well known that the Liouville equation conserves total entropy. Therefore, if the new dynamics produces total information entropy, $\tilde{\rho}$ will take a relative advantage over ρ in describing the system which has developed only with the restricted knowledge $\tilde{F}_1$ $(= F_1)$ from t_k to t^{k+1}. The same thing can be said for any time-subinterval.[6] And then, in the limit $\Delta t \to 0$ the new dynamics for $\tilde{\rho}$ will completely replace the dynamics for ρ. Entropy productivity of the functional random-walk model is to be proved in the next section. Thus, so long as we observe a many-particle system in a macroscopic way, i.e., with a very restricted knowledge like, at most, the one-body distribution function, it may be reasonable that we rather adopt the functional random-walk model than pursue the exact dynamics with a large amount of complete knowledge (i.e., F_s for all s) most of which is never observed.

In other words, our standpoint is different than that of the traditional mechanics which deals with a complete knowledge of the state of all particles; the Liouville equation is based on this mechanics. However, we need not approach the reaility in only this way, unless we really have such a complete (microscopic) knowledge of the system from the beginning to the end. We may have a definite reason to introduce an information-theoretical approach, when the problem is treated with very inadequate knowledge.

3. ENTROPY

A. Definition of H function

According to Shannon,[7] it might seem natural in our case to define the H function (per degree of freedom) as[8]

$$H_S = \lim_{M\to\infty} \int_B \tilde{\rho}^M \log\tilde{\rho}^M\delta z^M/M, \tag{3.1}$$

where $\tilde{\rho}^M(z) = \tilde{\rho}(z^M)$ and

$$z^M(x) = \sum_{j=1}^{M} b_j s_j(x), \quad b_j = \int_X z(x) s_j(x) dx; \tag{3.2}$$

$\{s_j(x)\}$ is an orthonormal function set in X; in other words, $\tilde{\rho}^M$ is the cylinder functional of $\tilde{\rho}$ in M dimensions. However, we should take into account that there are many possible permutations of particles which yield the same density field $z(x)$. Then, the total number of the permutations makes the weight for the field $z(x)$, which is calculated as

$$\frac{N!}{\prod_j [nz(x_j)\Delta x]!} = \exp\left(-\sum_j nz(x_j) \log z(x_j) \Delta x + \text{const}\right), \tag{3.3}$$

using the Stirling formula for $N = nV \to \infty$.

If the weight now calculated is considered, the suitable form of our H function should be

$$H^M = \int_B \tilde{\rho}^M \log\left[\tilde{\rho}^M \exp\left(n \int_X z^M \log z^M dx\right)\right] \delta z^M / M \tag{3.4}$$

in the cylinder functional approach. $\lim_{M\to\infty} H^M$ is assumed to exist. For simplicity, all M in (3.4) may be omitted with the understanding that $\tilde{\rho}$ was originally introduced as a cylinder functional.[3]

B. Entropy production

With the aid of (2.1), the time rate of H is calculated to be

$$\frac{dH}{dt} = \frac{1}{2} \int_B \int_X \int_X D \left(\frac{\delta^2 \tilde{\rho}}{\delta z(x) \delta z(x')} - \frac{1}{\tilde{\rho}} \frac{\delta \tilde{\rho}}{\delta z(x)} \frac{\delta \tilde{\rho}}{\delta z(x')} + \frac{n\delta(x-x')}{z(x)} \tilde{\rho} \right) dx dx' \delta z, \tag{3.5}$$

where $D = \mathcal{P}[\phi; z(x)z(x')]$. It is because of the following equality that the effect of the Q term in (2.1) vanishes in (3.5):

$$\int_B \int_X \frac{\delta}{\delta z(x)} (Qz\tilde{\rho}) dx \log\left[\tilde{\rho} \exp\left(n \int_X z \log z dx\right)\right] \delta z = \int_B \left(\int_X \frac{\delta}{\delta z(x)} (Qz) dx - n \int_X (\log z + 1) Qz dx \right) \tilde{\rho} \delta z. \tag{3.6}$$

The two terms in the big bracket vanish on account of the boundary condition on $z(x)$ as well as of the form of Q: (2.2). If we make the substitution

$$\tilde{\rho} = A \exp[-n \int_X z \log z dx - \psi(z)] \tag{3.7}$$

with A as the normalization factor, (3.5) reduces to

$$\frac{dH}{dt} = -\frac{1}{2} \int_B \int_X \int_X D \frac{\delta^2 \psi}{\delta z(x) \delta z(x')} dx dx' \tilde{\rho} \delta z. \tag{3.8}$$

ψ can be considered as an analytic functional:

$$\psi(z) = \tfrac{1}{2} \int_X \int_X A_2(x_1, x_2)[z(x_1) - z_0(x_1)] \times [z(x_2) - z_0(x_2)] dx_1 dx_2 + \text{higher-order terms in } z - z_0 \text{ with } A_n(x_1, \ldots, x_n). \tag{3.9}$$

Here $z_0, A_2, \ldots, A_n$ are certain definite functions in $X, X^2, \ldots X^n$, respectively. $\exp(-\psi)$ may be understood as a functional Gram-Charlier series apart from the normalization factor, if A_n for $n \geq 3$ are small. A_n is symmetric with respect to interchange of arguments without loss of generality, and A_2 and $\int_X \int_X A_n (z - z_0) \ldots (z - z_0) dx_3 \ldots dx_n$ are naturally assumed to be nonnegative definite functions of x_1 and x_2 in order for ρ to be functional-integrable [to give (2.3)]. Accordingly, $\delta^2\psi/\delta z(x)\delta z(x')$ must be nonnegative definite.

Thus, the double integral over X^2 in the integrand of (3.8) makes the trace of the product of two nonnegative-definite functions, which can be proved to be nonnegative. (See Appendix.) Hence, it is clear that

$$\frac{dH}{dt} \leq 0 \text{ (the general } H \text{ theorem)}. \tag{3.10}$$

It is concluded that entropy as the minus H function is always produced except at the only one state, where all $A_n = 0$ so that $\psi = 0$.

C. Steady state to be expected

$\tilde{\rho}$ for that state is expressed as

$$\tilde{\rho}_\infty = A_\infty \exp\left(-n \int_X z \log z dx\right). \tag{3.11}$$

If this is not a steady-state solution of (2.1), this state should readily be succeeded by another state at the next instant, so that entropy should increase endlessly so long as the assumption made on A_n holds. We could not expect any steady state to exist in this case, since entropy must be constant in a steady state. However, this fact contradicts a general property of the Fokker–Planck equation: There should be a unique steady state which all states go towards.[3,9] Hence, it may be reasoned that $\tilde{\rho}_\infty$ should be the steady-state solution of (2.1) finally approached by all $\tilde{\rho}$, and that the entropy with $\tilde{\rho}_\infty$ should be maximum.[10] Since our system is conditioned by (2.5), (3.11) may be considered to correspond to the Gibbs microcanonical distribution,[11] although all points in the B space are never equally realizable.

4. MAXIMUM-ENTROPY STATE

It is desirable to know the most probable value, i.e., the mode of $z(x)$ at the state with the maximum entropy. It is given as the $z(x)$ in the (Riemannian) B space such that the exponent of (3.11) is maximum, and then it should be solved by the standard variation method applied to the quantity $-n \int_X z \log z dx$ under the subsidiary conditions (2.4) and (2.5). It is easy to see that the implicit solution is a Maxwell-Boltzmann distribution with the particle interaction included in a self-consistent way:

$$z_B(x) = \exp\{-\lambda - \beta[H_1(x) + n \int_X \phi(|q - q'|) z_B(x') dx']\}, \tag{4.1}$$

where $\exp(-\lambda)$ and β are the constants to be determined by the conditions (2.4) and (2.5). It must be noted that $z_B(x)$ is also a particular (Maxwellian-type) solution of the steady Vlasov equation, as is easily verified.[2]

The exponent of (3.11) may be rewritten in terms of the functional Taylor expansion around the maximum in B;

$$\tilde{\rho}_\infty = A_\infty \exp\left(-n \int_X z_B \log z_B dx - \frac{n}{2} \int_X \frac{1}{z_B} (z - z_B)^2 dx - \ldots\right). \tag{4.2}$$

When n is large enough, as is usually the case (even with

such a rarefied gas as Boltzmann considered), the Gaussian distribution in (4.2) is a good approximation to $\tilde{\rho}_\infty$. Then, the mode z_B may actually be considered as the average. From this result, the significance of the particular steady-state solution of the Vlasov equation is clear. One can say that $z_B(x)$ is irreversibly approached by the one-body distribution $\tilde{F}_1(x,t)$ [cf. (2.11)], despite the fact that the Vlasov equation itself is time-reversible! If so, this solution must be widely applicable to the equilibrium state of any closed (classical) system: gas, liquid, etc., so long as we view the system macroscopically (through the model dynamics).

Now, comparison with the known theories of Boltzmann and Gibbs is possible at this equilibrium state. Our steady solution (4.1) for $\tilde{F}_1$ is obviously an extension of the Boltzmann distribution to the case with the particle interaction present. Equation (4.1) is canonical type, so that $\tilde{F}_1$ is conjectured to be also close to the one-body distribution to be derived from the Gibbs canonical distribution. In fact, all the theories are coincident in the limit when the particle interaction tends to vanish. It is interesting to see that our theory derives a canonical-type distribution of $\tilde{F}_1$ so naturally without appealing to the ergodic theory: indeed, calculation of the mode z_B is obviously on the same basis as the information-theoretical approach of Jaynes.[5] However, we could hardly expect a complete accord between our equilibrium state and the Gibbs state, since our macroscopic viewpoint is not so precise microscopically (as was described in Sec. 2).

As is obvious in (3.4), H is $\log A_\infty$ at $\tilde{\rho} = \tilde{\rho}_\infty$. In order to compare our entropy with Boltzmann's original form, a new definition, changed in level,

$$S = -kH + k\int_B \tilde{\rho}_\infty \log\tilde{\rho}_\infty \delta z \tag{4.3}$$

is most suitable, where k is the Boltzmann constant. In fact, at $\tilde{\rho} = \tilde{\rho}_\infty$ this entropy is related to the original Boltzmann entropy, as follows [cf. (4.2)]:

$$S_\infty = -kn\int_X \overline{z\ \log z}dx \lesssim -kn\int_X z_B \log z_B dx, \tag{4.4}$$

where the bar on top means the average with the measure $\tilde{\rho}_\infty$. The Boltzmann entropy gives an upper bound of S_∞, which is asymptotically approached as $n \to \infty$. By virtue of (4.1), we have

$$S_\infty \lesssim k[N\lambda + n\beta\int_X E(x)z_B(x)dx] \tag{4.5}$$

with

$$E(x) = H_1(x) + n\int_X \phi(|q-q'|)z_B(x')dx'. \tag{4.6}$$

If $N\lambda$ is read as thermodynamic potential, β as $1/kT$ (T: absolute temperature), and $n\int_X E(x)z_B(x)dx$ as total energy, the rhs of (4.5) gives the usual thermodynamic entropy.[11] It should be noted, however, that the particle-interaction energy is counted twice in the total energy so read, so that the correspondence of rhs of (4.5) with thermodynamic entropy is not exact by that amount. Anyhow, the consideration given here is meaningful in indicating that S_∞ is fairly close to equilibrium thermodynamic entropy.

5. CONCLUSION

We began with investigating the information-theoretical background of the functional random-walk model, which is useful for understanding the physical meaning of the operator $\mathcal{P}$ in (2.1). Next, the general explicit form of $\tilde{\rho}$ with the maximum entropy was obtained with the aid of a proper H function newly introduced. The general H theorem was given together. It was proved that at the maximum-entropy state the one-body distribution function $\tilde{F}_1$ in the model dynamics is canonical with the smoothed-out interaction potential included, while the solution $\tilde{\rho}_\infty$ at the same state corresponds to the micro-canonical distribution. Then, as far as we treat maximum-entropy states, an analytical approach to the model dynamics is possible and we do not need any such Monte Carlo quadrature as was discussed in some detail in previous work.[3] The relation of our information entropy and the original Boltzmann entropy was also clarified.

APPENDIX

The trace of the product of symmetric nonnegative-definite functions $A(x,x')$ and $B(x,x')$, i.e., $\int_X\int_X A(x,x')B(x',x)dxdx'$ is treated in the approximate form: $\sum_{i,k} A_{ik}B_{ki}$, when the functions are approximated by finite Fourier series made from the orthonormal function set $\{s_i(x); i = 1,\ldots,M\}$ in X as

$$A(x,x') = \sum_{i,k} A_{ik}s_i(x)s_k(x'), \quad \text{etc.} \tag{A1}$$

Now, consider the orthogonal transformation (T_{ik}) which diagonalizes both matrices (A_{ik}) and (B_{ik}) simultaneously. Then, we have

$$A_{ik} = \sum_j T_{ji}\Lambda_j T_{jk}, \tag{A2}$$

$$B_{ki} = \sum_l T_{lk}M_l T_{li}, \tag{A3}$$

where (Λ_i) and (M_i) are the eigenvalues of (A_{ik}) and (B_{ik}), respectively, and we note, by definition,

$$\text{all } \Lambda_i \geq 0, \tag{A4}$$

$$\text{all } M_i \geq 0. \tag{A5}$$

Hence,

$$\sum_{i,k} A_{ik}B_{ki} = \sum_{i,k}\sum_{j,l}(T_{ji}T_{li})(T_{jk}T_{lk})\Lambda_j M_l = \sum_j \Lambda_j M_j \geq 0 \tag{A6}$$

on account of (A4), (A5). Thus, the proof of

$$\int_X\int_X A(x,x')B(x',x)dxdx' \tag{A7}$$

is concluded, if the lhs of (A6) converges to that of (A7) in the limit $M \to \infty$.

[1] I. Prigogine, in *A Critical Review of Thermodynamics*, edited by E. B. Stuart, B. Gal-Or, and A. J. Brainard (Mono, Baltimore, 1970), p. 1.

[2] T-Y. Wu, *Kinetic Equations of Gases and Plasmas* (Addison-Wesley, Reading, Mass., 1966).

[3] I. Hosokawa, J. Phys. Soc. Japan Suppl. **26**, 224 (1969); J. Math. Phys. **11**, 657 (1970).

[4] S. Gartenhaus, *Elements of Plasma Physics* (Holt, Rinehart and Winston, New York, 1964), Chap. 3.

[5] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); in *Statistical Physics*, 1962 Brandeis Lectures (Benjamin, New York, 1963), Vol. 3, p. 181.

[6] The reason that we can set $\rho = \tilde{\rho}$ initially at $t = t^k$ in any time subinterval $[t^k, t^{k+1}]$ is that our knowledge is restricted to F_1 ($=\tilde{F}_1$) so that any ρ which gives the F_1 in question is suitable if it has an entropy large enough.

[7] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1959), p. 58.

[8] The statement of entropy in the previous work[3] is erroneous in sign. It should be corrected as presented here.

[9] A. H. Gray, Jr., J. Math. Phys. **6**, 644 (1965).

[10] A direct proof of this fact is given only by taking the functional variation of H under the condition (2.3).

[11] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).

Renormalized oscillator equations

Charles N. Friedman

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 19 April 1973)

Let $\{\mathcal{H} = L^2(R, dx),\ q = x, p = -id/dx\}$ be the standard Schrödinger representation of the canonical commutation relations for one degree of freedom. For a, b, c, real scalars, put $H_{abc} = \frac{1}{2}p^2 + q^4 + aq^3 + bq^2 + cq$ and $q(t) = \exp(itH_{abc})q \exp(-itH_{abc})$. Then $q(t)$ satisfies the operator equation $\ddot{q}(t) = -4q^3(t) - 3aq^2(t) - 2bq(t) - c$. For $n = 0, 1, 2, 3$ we define renormalized (or Wick-ordered) nth powers of q, denoted $q^{(n)}$; these are polynomials of degree n in q and are characterized by the conditions:

$$\left\{\begin{array}{c} q^{(0)} = 1 \\ [q^{(n+1)}, p] = i(n+1)q^{(n)} \\ \text{for } n > 0, \quad \int G^* q^{(n)} G dx = 0, \text{ where } G = \text{ground state of } H_{abc} \end{array}\right\}$$

In terms of these powers the equation for $q(t)$ can be written $\ddot{q}^{(1)} = -4q^{(3)} - Aq^{(2)} - Bq^{(1)} - C$, for some real A, B, C which are functions of a, b, c. This has the form of a much studied prototypical nonlinear quantum field equation in the somewhat (physically) trivial case of one space-time dimension. Basically, we prove two results concerning this equation which we feel are of interest, because they may provide some basis for conjecture about the behavior of nonlinear field equations in a higher number of dimensions. First we determine a nontrivial condition which the renormalization constants A, B, C must satisfy and which implies that the set of points (A, B, C) assumed as a, b, c vary is of measure zero in R^3. We study in somewhat more detail the situation when $a = c = 0$. The associated renormalized equation then has the form $\ddot{q}^{(1)} + Bq^{(1)} = -4q^{(3)}$, where $B = B(b)$. We determine the qualitative behavior of the function $B(b)$ as $b \to \pm\infty$ and show for example that $B(b)$ is not one-to-one. As a corollary, for many values of B, there exist at least two equations of the form $\ddot{q}^{(1)} + Bq^{(1)} = -4q^{(3)}$ with the same B, but which are not unitarily equivalent in a sense, to be precised. Such nonunicity cannot occur for linear equations, as has been known for some time.

I. INTRODUCTION AND MATHEMATICAL PRELIMINARIES

Let $\{\mathcal{H}, p, q\}$ be the standard Schrödinger representation of the canonical commutation relations $[p, q] = pq - qp = -i$ for one degree of freedom; i.e., $\mathcal{H} = L^2(R, dx)$ is the Hilbert space of square-integrable functions on the real line R with respect to Lebesgue measure, q is the operator of multiplication by x, and $p = -id/dx$, with the usual domains which make these (unbounded) self-adjoint operators in $\mathcal{H}$:

$$D(q) = \{u \in \mathcal{H}: xu(x) \in \mathcal{H}\},$$
$$D(p) = \{u \in \mathcal{H}: du/dx \in \mathcal{H}\} = \{u \in \mathcal{H}: k\hat{u}(k) \in L^2(R, dk)\},$$

where du/dx is taken in the sense of distributions and $u(k)$ is the Fourier transform of $u(x)$. For $n \geq 1$ we have the commutation relations

$$[q^n, p] = inq^{n-1}, \quad [p^n, q] = -inp^{n-1} \quad ([X, Z] = XZ - ZX). \tag{I.1}$$

(In such equations we interpret the equalities to hold on the common domain of the operators on both sides of the equations.)

The proof of the following theorem is contained in the results of Ref. 1 (with some changes in notation.)

Theorem I.1: The operator $H_{abc} = \frac{1}{2}p^2 + q^4 + aq^3 + bq^2 + cq$ with a, b, c real is self-adjoint on the domain $D = D(p^2) \cap D(q^4)$. The spectrum of H_{abc} consists solely of real eigenvalues, bounded below, with a minimum eigenvalue. Each eigenvalue corresponds to exactly one (linearly independent) eigenvector in $\mathcal{H}$.

We denote the inner product on $\mathcal{H}$ by $\langle \cdot, \cdot \rangle$ which we take to be conjugate-linear in the first slot: $\langle u, v \rangle = \int_R u(x)^* v(x)\, dx$. It follows easily from Theorem I.1 that, up to sign (± 1), there is a unique real-valued, normalized function $G_{abc}(x) \in \mathcal{H}$ with $H_{abc}G_{abc}(x) = w_{abc}G_{abc}(x)$, where w_{abc} is the minimum eigenvalue of H_{abc}. We call $G_{abc}(x)$ and w_{abc} the ground state and ground state energy, respectively. In all future equations or relations involving $G_{abc}(x)$ we mean to allow for a possible ambiguity in sign; this ambiguity disappears if only even powers of $G_{abc}(x)$ appear which will often be the case. For $n = 0, 1, 2, 3$ we put $E_{abc}(q^n) = \langle q^n G_{abc}, G_{abc} \rangle$ and similarly for $E_{abc}(p^n)$, $n = 0, 1, 2$. The following theorem will be of use in a later section.

Theorem I.2: Suppose $a = c = 0$, so that $H_{abc} = H_b = \frac{1}{2}p^2 + q^4 + bq^2$. Then, for positive odd integral m, $E_b(q^m)$ and $E_b(p^m)$ are zero if they exist at all.

Proof: The transformation $x \to -x$ is unitary on $\mathcal{H}$ and leaves H_b invariant; thus $G_{abc}(-x) = \pm G_{abc}(x)$. Since this transformation takes $q \to -q$ and $p \to -p$, it follows immediately from the definitions of $E_b(q^m)$, $E_b(p^m)$ that these are zero for all odd positive m for which they exist.

We put $q(t) = \exp(itH_{abc})\, q \exp(-itH_{abc})$. Then an easy calculation using the commutation relations (I.1) yields

$$\ddot{q}(t) = -4q(t)^3 - 3aq(t)^2 - 2bq(t) - c, \tag{I.2}$$

where the double dots denote d^2/dt^2 and where the equality is to be interpreted on a suitable domain on which both sides of the equation make sense. (The exact nature of such a domain will be unimportant in the present work.) We view (I.2) as the one space-time dimen-

sion form of the heuristic quantum field equation:

$$\Box\varphi = P(\varphi), \quad \text{where } \Box = d^2/dt^2 - \nabla^2,$$
$$\varphi = \varphi(\mathbf{x}, t) \text{ is a quantum field, } P \text{ a polynomial.}$$

Actually this last equation will not make sense unless the application of P to φ is interpreted via suitable renormalization. In the next section we define certain renormalized powers of q and p which are analogous to the usual Wick-ordered powers of the field φ.

II. THE RENORMALIZED POWERS OF q AND p

Let $v \in \mathcal{K}$ be in the domain of q^N for some positive integer N (i.e., $q^N v \in \mathcal{K}$) We define renormalized powers of q, denoted $q^{(n)}$ $(n \leq N)$ by the conditions:

$$\begin{aligned} &q^{(0)} = q^0 = 1, \\ &q^{(n)} = q^n + a_{n-1}^{(n)} q^{n-1} + \cdots + a_1^{(n)} q + a_0^{(n)}, \\ &[q^{(n+1)}, p] = i(n+1) q^{(n)}, \\ &\langle q^{(n)} v, v\rangle = 0 \quad (n \geq 1), \end{aligned} \tag{II. 1}$$

where the $a_j^{(i)}$ are real scalars and equality is as usual interpreted on suitable domains. The $a_j^{(i)}$ can be determined recursively, and we find

$$q^{(n)} = \sum_{s=0}^{n} \sum_{j_1+2j_2+\cdots+(n-s)j_{n-s}=n-s} C(j_1,,,j_{n-s},n,s) \times E(q)^{j_1} E(q^2)^{j_2} \cdots E(q^{n-s})^{j_{n-s}} q^s, \tag{II. 2}$$

where

$$C(j_1,,,j_{n-s},n,s) = \frac{(-1)^{j_1+\cdots+j_{n-s}}}{j_1!\cdots j_{n-s}!} \frac{n!}{s!} \frac{(j_1+\cdots+j_{n-s})!}{(2!)^{j_2}\cdots((n-s)!)^{j_{n-s}}}$$

and $E(q^m) = \langle q^m v, v\rangle$.

We define $p^{(n)}$ by the same formula with p replacing q everywhere.

The verification of (II. 2) is somewhat tedious and we shall omit the details, since we will not need to make use of it except for small values of n which can be dealt with by using (II. 1) directly. For example,

$$\begin{aligned} &q^{(0)} = 1, \\ &q^{(1)} = q - E(q), \\ &q^{(2)} = q^2 - 2E(q)q + 2E(q)^2 - E(q^2), \\ &q^{(3)} = q^3 - 3E(q)q^2 + (6E(q)^2 - 3E(q^2))q \\ &\qquad + 6E(q)E(q^2) - 6E(q)^3 - E(q^3). \end{aligned} \tag{II. 3}$$

From now on, by $q^{(n)}$, we shall always mean the renormalized powers of q with respect to $v = G_{abc}$. [Notice that no sign ambiguity occurs here, because only ${G_{abc}}^2$ enters in the definition of $E_{abc}(q^m)$.] Now putting $q^{(n)}(t) = \exp(itH_{abc})\, q^{(n)} \exp(-itH_{abc})$, we write (I. 2) in the following form:

$$\ddot{q}^{(1)}(t) = -4q^{(3)}(t) - Aq^{(2)}(t) - Bq^{(1)}(t) - C, \tag{II. 4}$$

and we may then determine A, B, C as functions of a, b, c so that (II. 4) is the same as (I. 2). Using (II. 3), we find after a short calculation

$$\begin{aligned} A &= 3a + 12E_{abc}(q), \\ B &= 2b + 12E_{abc}(q^2) + 6aE_{abc}(q), \\ C &= c + 3aE_{abc}(q)^2 + 2bE_{abc}(q) + 12E_{abc}(q)^3 \\ &\quad + 4E_{abc}(q^3) - 12E_{abc}(q)E_{abc}(q^2). \end{aligned} \tag{II. 5}$$

It would be of interest to know exactly which real values A, B, C assume as a, b, c range over all real values. (An analogous question for the case of field equations in a larger number of dimensions has been raised in Ref. 2. We do not answer this question in any great detail, but we will show that many values are not assumed. Later we will consider the case $a = c = 0$ (which forces $A = C = 0$) and will determine the qualitative behavior of B (b) for large $|b|$.

Theorem II. 1: Let t be any real number. Then, as functions of a, b, c, A, B, and C are invariant under the transformation:

$$\begin{aligned} &a \to a' = a + 4t, \\ &b \to b' = b + 3ta + 6t^2, \\ &c \to c' = c + 2tb + 3t^2 a + 4t^3. \end{aligned} \tag{II. 6}$$

Proof: If we make the substitution $q \to q + t$, then $H_{abc} \to H_{a'b'c'} + \text{const} = \frac{1}{2}p^2 + q^4 + a'q^3 + b'q^2 + c'q + \text{const}$. It follows easily that $G_{abc}(x + t) = G_{a'b'c'}(x)$. Let A', B', C' be given by (II. 5) with a, b, c replaced by a', b', c'. We may verify directly that $A = A', B = B', C = C'$. For example, we have

$$\begin{aligned} A' &= 3a' + 12E_{a'b'c'}(q) = 3(a + 4t) + 12\int_R x\, G_{a'b'c'}(x)^2\, dx \\ &= 3a + 12t + 12\int_R (x - t) G_{a'b'c'}(x - t)^2\, dx \\ &= 3a + 12t + 12\int_R (x - t)\, G_{abc}(x)^2\, dx \\ &= 3a + 12E_{abc}(q) = A. \end{aligned}$$

The calculations for $B = B'$ and $C = C'$ are similar, and we shall not reproduce them here.

Remark: A more elegant and less computational proof of the theorem may be based on certain invariance properties of the renormalized powers under translations. However the proof depends on certain "unicity" results for the $q^{(n)}$ and these are difficult to formulate without expressing the relations (II. 1) in bounded form [i.e., in terms of the $\exp(itq^{(n)})$, $\exp(itp)$, etc.] in order to avoid troublesome domain considerations; we have deliberately sidestepped such questions here. (These delicate questions have been investigated in great detail in Refs. 2, 3, 4, for example.)

Corollary II. 1: The set of points: $\{(A, B, C) : (a, b, c) \in R^3\}$ is of measure zero in R^3.

Proof: It is known that the $E_{abc}(q^k)$ are smooth (real analytic) functions of a, b, c. For the proof we refer to Ref. 1, where analogous results are proved; the same methods of proof apply in the present case. The theorem shows that the mapping $(a, b, c) \to (A, B, C)$ is continuum to one from R^3 to R^3, and then Sard's theorem[5] implies that the image of this mapping is of measure zero in R^3.

In the next section we will prove that, in the case $a = c = 0$, the image of the mapping $b \to B(b)$ is not all of R, and this mapping is not one to one.

III. THE EQUATION $\ddot{q}^{(1)} + B(b)q^{(1)} = -4q^{(3)}$

In this section we will consider the special case $a = c = 0$, so that $H_b = H_{0b0} = \frac{1}{2}p^2 + q^4 + bq^2$. We denote $E_{0b0}, w_{0b0}, G_{0b0}$ by E_b, w_b, G_b.

Theorem I. 2 together with (II. 5) implies that

$$A(0, b, 0) = C(0, b, 0) = 0 \quad \text{and} \quad B(b) = B(0, b, 0) = 2b + 12E_b(q^2). \tag{III. 1}$$

From (III. 1) we have immediately

$$\lim_{b\to+\infty} B(b) = +\infty \tag{III. 2}$$

since $E_b(q^2) \geq 0$. We will show that $\lim_{b\to-\infty} B(b) = +\infty$ also. This requires a series of estimates to follow.

Proposition III. 1: For $b < 0$, $E_b \geq b^{-1}w_b$.

Proof: $w_bG_b = H_bG_b = H_0G_b + bq^2G_b$ so that $w_b = b\langle q^2G_b, G_b\rangle + \langle H_0G_b, G_b\rangle$. Since $H_0 = \frac{1}{2}p^2 + q^4 \geq 0$, we have $w_b \geq bE_b(q^2)$ and dividing by b gives the result.

Proposition III. 2: For $b < 0$, $w_b \leq 1 - b - b^2/4$.

Proof: For any $u \in D(H_b)$, with $\langle u, u\rangle = 1$, we have $\langle H_bu, u\rangle \geq w_b$, because w_b is the minimum of the spectrum of H_b.

Now by taking $u(x) = (2/\pi)^{1/4}\exp\{-[x - (|b|/2)^{1/2}]^2\}$ this last estimate yields $w_b \leq 11/16 - b/2 - b^2/4$.

Combining Proposition III. 1 and Proposition III. 2 yields

$$E_b(q^2) \geq -1 + b^{-1} - b/4. \tag{III. 3}$$

Since $B(b) = 2b + 12E_b(q^2)$, we have

$$B(b) \geq 2b + 12(-1 + b^{-1} - b/4) = -12 - 12b^{-1} - b \qquad (b < 0)$$

and thus $\lim_{b\to-\infty} B(b) = +\infty$. (III. 4)

We shall not discuss here the question of whether $B(b)$ assumes the value zero.

We now turn to the question of unicity (or lack of it) for the equation $\ddot{q}^{(1)}(t) + Bq^{(1)}(t) = -4q^{(3)}(t)$. We have $H_b = \frac{1}{2}p^2 + q^4 + bq^2$, and we define a "renormalized" Hamiltonian by $H_{\text{ren}} = \frac{1}{2}p^{(2)} + q^{(4)} + \frac{1}{2}B(b)q^{(2)}$. (This may also be denoted by $H_{\text{ren},b}$ when confusion is likely to occur.) Using the explicit form of the $q^{(k)}$ and $p^{(k)}$ together with Theorem I. 2, a routine calculation yields

$$H_{\text{ren}} = \tfrac{1}{2}p^{(2)} + q^{(4)} + \tfrac{1}{2}B(b)q^{(2)} = H_b - w_b \geq 0. \tag{III. 5}$$

Thus G_b is the ground state of H_{ren} and $H_{\text{ren}}G_b = 0$.

For real b, consider the system

$$\left.\begin{aligned} &\mathcal{H} = L^2(R, dx), q^{(k)}, p^{(j)} \\ &H_{\text{ren}} = \tfrac{1}{2}p^{(2)} + q^{(4)} + \tfrac{1}{2}B(b)q^{(2)} \geq 0 \\ &\ddot{q}^{(1)}(t) + B(b)q^{(1)}(t) = -4q^{(3)}(t) \\ &G_b, \text{ the ground state of } H_{\text{ren}}; H_{\text{ren}}G_b = 0 \end{aligned}\right\}. \tag{b}$$

Here $B(b) = 2b + 12E_b(q^2)$. Also we have $q^{(1)} = q$, $p^{(1)} = p$ by an application of Theorem I. 2. If it is necessary to distinguish the system (b) to which certain operators or objects belong, we will use the subscript "b" for this purpose.

We will say that two systems (b) and (b') are unitarily equivalent if there is a unitary operator U from $\mathcal{H}_b$ to $\mathcal{H}_{b'}$ with $UG_b = G_{b'}$, $Uq_b^{(k)}U^{-1} = q_{b'}^{(k)}$ $Up_b^{(j)}U^{-1} = p_{b'}^{(j)}$, $UH_{\text{ren},b}U^{-1} = H_{\text{ren},b'}$.

Theorem III. 1: Let b and b' be two distinct real numbers for which $B(b) = B(b')$. [These exist by (III. 2) and (III. 4).] Then the systems (b) and (b') are not unitarily equivalent.

Proof: If a unitary operator U existed giving such a unitary equivalence, then because $UqU^{-1} = q$, it follows easily that $G_b(x) = G_{b'}(x)$ (with a proper choice of sign for each). But this is impossible, for

$$H_bG_b = H_0G_b + bq^2G_b = w_bG_b,$$

$$H_{b'}G_{b'} = H_0G_{b'} + b'q^2G_{b'} = w_{b'}G_{b'},$$

and if $G_b = G_{b'}$, we would then have

$$[(b - b')q^2 + w_{b'} - w_b]G_b = 0, \qquad \text{which forces } G_b = 0,$$

since $b \neq b'$. This is absurd, because G_b is the ground state for H_b.

Remarks: (i) In view of the preceding result, we feel that it is not unlikely that a similar situation exists in the case of nonlinear field equations in, say, two space-time dimensions of the form

$$(\Box + m^2)\varphi^{(1)} = -\lambda\varphi^{(3)} \quad (\lambda > 0), \tag{III. 6}$$

where $\varphi^{(1)}$ and $\varphi^{(3)}$ are Wick-ordered products with respect to the physical vacuum.[2] However, (III. 6) is largely heuristic and difficult to interpret as a bona fide equation unless, for example, spacial cutoffs are introduced, and this introduces additional terms in the equation. The question of the unicity of (III. 6) is much deeper than any of the questions we have discussed in the present work.

(ii) The unicity of positive energy *free* fields (corresponding to equations of the form $(\Box + m^2)\varphi = 0$) is well known.[6] In the context of the present work, a special case of this result appears in the following terms:

If we put $H_c = \frac{1}{2}p^2 + \frac{1}{2}q^2 + cq = \frac{1}{2}p^2 + \frac{1}{2}(q + c)^2 - c^2/2$, then the ground state of H_c is $G_c(x) = (\pi)^{-1/4}\exp[-(x + c)^2/2]$, with $H_cG_c = w_cG_c$, $w_c = \frac{1}{2}(1 - c^2)$. We have also $H_{\text{ren}} = \frac{1}{2}p^{(2)} + \frac{1}{2}q^{(2)} = H_c - w_c$, and $q^{(1)} = q - E_q(q) = q + c$. If $q(t) = \exp(itH_c)q\exp(-itH_c)$, then $\ddot{q}(t) + q(t) + c = 0$, and in terms of $q^{(1)}$ this equation can be written $\ddot{q}^{(1)}(t) + q^{(1)}(t) = 0$. However, it is easily seen that for any real c, the systems

$$\left.\begin{aligned} &\mathcal{H} = L^2(R, dx), q^{(k)}, p^{(j)} \\ &H_{\text{ren}} = \tfrac{1}{2}p^{(2)} + \tfrac{1}{2}q^{(2)} \geq 0 \\ &\ddot{q}^{(1)}(t) + q^{(1)}(t) = 0 \\ &G_c \quad \text{with } H_cG_c = 0 \end{aligned}\right\} \tag{c}$$

are all unitarily equivalent; explicitly for $c \neq c'$, the unitary operator $U_{c,c'}: \mathcal{H} \to \mathcal{H}$ given by $(U_{c,c'}f)(x) = f(x + c' - c)$ for $f \in \mathcal{H}$, transforms (c) into (c') in the sense discussed above.

ACKNOWLEDGMENT

The author would like to express his gratitude to Irving Segal for suggesting the problems treated in this article and for several enlightening discussions concerning the results.

[1] B. Simon, Ann. Phys. (N.Y.) **58**, 76 (1970).
[2] I. Segal, Inventiones Math. **14**, 211 (1971).
[3] I. Segal, J. Funct. Anal. **4**, 404 (1969); 6, 29 (1970).
[4] I. Segal, Topology **7**, 147 (1968).
[5] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1964), p. 47.
[6] I. Segal, Ill. J. Math. **6**, 500 (1962).

Application of Kraichnan's direct interaction approximation to kinematic dynamo theory. I. Incompressible isotropic turbulence and a singular integral equation

I. Lerche

Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637

(Received 15 January 1973)

Using Kraichnan's direct interaction approximation in the induction equations we set up the kinematic equations governing behavior of an ensemble average magnetic field under homogeneous, incompressible, mirror symmetric isotropic velocity turbulence. We demonstrate that the normal modes of the field depend on the solutions to a nonlinear single integral equation. For a simple form of the velocity turbulence we have investigated some of the properties of the integral equation. In particular, we have been able to construct a particular solution. We also point out what remains to be done if we are to obtain all the modes of the ensemble average magnetic field. We have done this calculation so that the behaviour and normal modes of the field can be investigated at arbitrary magnetic Reynolds' numbers. This is in contrast to customary approximations (like first-order smoothing theory) which normally are valid only for very small Reynolds numbers, if at all, and which therefore omit large regimes of considerable physical interest.

I. INTRODUCTION

In a remarkable paper published in 1961 Kraichnan[1] developed a powerful method (now called the direct interaction approximation—we shall refer to it as DIA hereinafter) for handling turbulence problems. The main thrust of DIA is to replace the true turbulence problem by a model, or models, that lead, *without approximation,* to closed equations for correlation functions and Green's functions. The model solutions are *exact* descriptions of possible dynamical systems. Consequently, as Frisch (1968)[2] has noted, the "the exact model solutions are approximate solutions of the true turbulence problem." And as such they are both physically realizable and acceptable.

In the decade or so since the introduction of DIA, Kraichnan has applied this method to the problem of hydrodynamic turbulence (see numerous papers in the Physics of Fluids from years 1962-1972). There are no free parameters in the theory and the method reproduces the observed turbulence spectrum. The model results agree closely with the observations at both large and small Reynolds' number indicating that the range of validity of DIA is over all possible values. This is in contrast to so-called "closure approximation" schemes like first-order smoothing theory (FOST), for example, which neglect certain correlation functions and whose range of validity, if any, is therefore restricted to small Reynolds' number.

Using Kraichnan's DIA we shall give here the calculation of isotropic turbulent kinematic dynamo activity valid for *all* Reynolds' numbers. We refer the interested reader to Kraichnan's (1961)[1] original paper for the prescription used in obtaining the DIA model turbulence kinematic induction equations. We shall quote here only those results which are pertinent to our particular problem.

II. FORMULATION AND REDUCTION OF THE EQUATIONS

Consider an infinite medium, of constant resistivity η, devoid of any large scale velocity shear but possessing a turbulent velocity $\mathbf{v}(\mathbf{x}, t)$. Then the induction equation describing the magnetic field behavior is

$$\frac{\partial B_i}{\partial t} - \eta\nabla^2 B_i = \epsilon_{ijk}\epsilon_{klm}\frac{\partial}{\partial x_j}(v_l B_m), \tag{1}$$

with $\nabla\cdot\mathbf{B} = 0$.

Under Kraichnan's DIA the equations for the ensemble average magnetic field $\mathbf{B}$ and the ensemble average Green's tensor, $\underline{\underline{G}}$, are

$$\left(\frac{\partial}{\partial t} - \eta\nabla^2\right)B_i = \epsilon_{ijk}\epsilon_{klm}\epsilon_{aJK}\epsilon_{KLM}\frac{\partial}{\partial x_j}\int_{-\infty}^{t} dt'd^3\mathbf{x}'$$
$$\times\left(G_{ma}(\mathbf{x}, t|\mathbf{x}', t')\frac{\partial}{\partial x_J'}[U_{lL}(\mathbf{x}, t|\mathbf{x}', t')B_M(\mathbf{x}', t')]\right) \tag{2}$$

and

$$\left(\frac{\partial}{\partial t} - \eta\nabla^2\right)G_{iu}(\mathbf{x}, t|\mathbf{x}', t') = \delta_{iu}\delta(\mathbf{x}-\mathbf{x}')\delta(t-t')$$
$$+ \epsilon_{ijk}\epsilon_{klm}\epsilon_{abK}\epsilon_{KLM}\frac{\partial}{\partial x_j}\int_{t'}^{t} dt''d^3\mathbf{x}''$$
$$\times\left(G_{ma}(\mathbf{x}, t|\mathbf{x}'', t'')\frac{\partial}{\partial x_b''}[U_{lL}(\mathbf{x}, t|\mathbf{x}'', t'')G_{Mu}(\mathbf{x}'', t''|\mathbf{x}', t')]\right), \tag{3}$$

where

$$U_{lL}(\mathbf{x}, t|\mathbf{x}', t') = \langle v_l(\mathbf{x}, t)v_L(\mathbf{x}', t')\rangle, \tag{4}$$

with $G_{ij}(\mathbf{x}, t|\mathbf{x}', t') = 0$ for $t < t'$,

since only the forward-going (in time) Green's tensor is physically permissible.

For homogeneous, stationary velocity turbulence (and we shall restrict our attention to just this form of turbulence for the remainder of the paper), we have

$$U_{lL} = U_{lL}(\mathbf{x} - \mathbf{x}', t - t'). \tag{5}$$

Then by inspection of Eq. (3) we see that the Green's tensor must be homogeneous and stationary, so that

$$G_{iu}(\mathbf{x}, t|\mathbf{x}', t') = G_{iu}(\mathbf{x} - \mathbf{x}', t - t'). \tag{6}$$

It then follows that with

$$B_i(\mathbf{x}, t) = \int B_i(\mathbf{k}, \omega)\exp[i(\mathbf{k}\cdot\mathbf{x} - \omega t)]d^3\mathbf{k}d\omega, \tag{7a}$$

and

$$[U_{ij}(\mathbf{x} - \mathbf{x}', t - t'), G_{ij}(\mathbf{x} - \mathbf{x}', t - t')] \tag{7b}$$
$$= \int[U_{ij}(\mathbf{k}, \omega), G_{ij}(\mathbf{k}, \omega)]\exp\{i[\mathbf{k}\cdot(\mathbf{x} - \mathbf{x}') - \omega(t - t')]\}$$
$$\times d^3\mathbf{k}d\omega,$$

Eq. (2) becomes

$$(\eta k^2 - i\omega)B_i(\mathbf{k},\omega) = -(2\pi)^4 \int d^3\mathbf{K} d\Omega [k_j K_b B_b(\mathbf{k},\omega) \times [-U_{ia}(\mathbf{K},\Omega)G_{ja}(\mathbf{k}-\mathbf{K},\omega-\Omega) + U_{ja}(\mathbf{K},\Omega)G_{ia}(\mathbf{k}-\mathbf{K},\omega-\Omega)] + k_j k_b B_a(\mathbf{k},\omega)[U_{jb}(\mathbf{K},\Omega)G_{ia}(\mathbf{k}-\mathbf{K},\omega-\Omega) - U_{ib}(\mathbf{K},\Omega)G_{ja}(\mathbf{k}-\mathbf{K},\omega-\Omega)]], \quad (8)$$

while Eq. (3) gives

$$(\eta k^2 - i\omega)G_{iu}(\mathbf{k},\omega) = (2\pi)^{-4}\delta_{iu} - (2\pi)^4 \int d^3\mathbf{K} d\Omega k_j (k_b - K_b) \times [U_{ia}(\mathbf{K},\Omega)G_{ja}(\mathbf{k}-\mathbf{K},\omega-\Omega)G_{bu}(\mathbf{k},\omega) + U_{jb}(\mathbf{K},\Omega) \times G_{ia}(\mathbf{k}-\mathbf{K},\omega-\Omega)G_{au}(\mathbf{k},\omega) - U_{ib}(\mathbf{K},\Omega)G_{ja}(\mathbf{k}-\mathbf{K},\omega-\Omega)G_{au}(\mathbf{k},\omega) - U_{ja}(\mathbf{K},\Omega)G_{ia}(\mathbf{k}-\mathbf{K},\omega-\Omega)G_{bu}(\mathbf{k},\omega)], \quad (9)$$

where use has been made of $k_i B_i = 0$ to eliminate some of the terms in Eq. (8).

Note that Eq. (8) is linear in **B**. So a solution to it exists if, and only if, a dispersion relation is satisfied. Our task is to obtain that dispersion relation and to see if it possesses any growing modes. If so we then have regenerative dynamo action under kinematic velocity turbulence. In order to obtain the dispersion relation from Eq. (8) we must do two things: first we must specify the tensor form of $U_{ij}(\mathbf{k},\omega)$; second we must then solve Eq. (9) *exactly* for $G_{iu}(\mathbf{k},\omega)$. Armed with this information we can then substitute for U_{ij} and G_{iu} in Eq. (8) to obtain the dispersion relation.

III. INCOMPRESSIBLE, ISOTROPIC VELOCITY TURBULENCE

In view of the debate [LII, L[2], LV, Krause and Roberts (1973)[6]] that has arisen concerning the contribution of incompressible, isotropic velocity turbulence to kinematic dynamo action it seems appropriate here to use the DIA equations (8) and (9) to attempt to settle the controversy once and for all. As we have remarked earlier all previous attempts to construct kinematic dynamo action using incompressible, isotropic velocity turbulence have been limited by invocation of FOST. Under FOST the Reynolds' number must be small; and when the Reynolds' number is small there is no possibility of dynamo action as we shall presently demonstrate using the DIA equations.

For incompressible, isotropic, mirror symmetric velocity turbulence we have

$$U_{ij}(\mathbf{k},\omega) = E(k,\omega)(\delta_{ij} - k_i k_j k^{-2}), \quad (10)$$

so that $k_i U_{ij} = k_j U_{ij} = 0$.

Further $E(k,\omega) \geq 0$ for all real k and ω by Cramér's theorem (1940).[7]

We define the two basic integrals

$$I_{iajb} \equiv \int U_{ia}(\mathbf{K},\Omega)G_{jb}(\mathbf{k}-\mathbf{K},\omega-\Omega)d^3\mathbf{K}d\Omega \quad (11a)$$

and

$$J_{iajb\lambda} \equiv \int U_{ia}(\mathbf{K},\Omega)G_{jb}(\mathbf{k}-\mathbf{K},\omega-\Omega)K_\lambda d^3\mathbf{K}d\Omega. \quad (11b)$$

In terms of these integrals Eqs. (8) and (9) become

$$B_a(\mathbf{k},\omega)\{\delta_{ia}(\eta k^2 - i\omega) + (2\pi)^4[k_j(J_{jbiba} - J_{ibjba}) + k_j k_b(I_{jbia} - I_{ibja})]\} = 0 \quad (12)$$

and

$$G_{iu}(\mathbf{k},\omega)(\eta k^2 - i\omega) = (2\pi)^{-4}\delta_{iu} - (2\pi)^4 k_j \times [G_{bu}(\mathbf{k},\omega)k_b(I_{iaja} - I_{jaia}) + G_{au}(\mathbf{k},\omega)k_b(I_{jbia} - I_{ibja}) - G_{bu}(\mathbf{k},\omega)(J_{iajab} - J_{jaiab})], \quad (13)$$

where we have used the incompressibility condition $k_i U_{ij} = 0$ to eliminate several of the terms in Eqs. (8) and (9).

Now note by inspection of Eqs. (10)–(13) that since both $U_{ij}(-\mathbf{k}) = U_{ij}(\mathbf{k})$ and $U_{ij}(\mathbf{k}) = U_{ji}(\mathbf{k})$ it follows that $G_{iu}(\mathbf{k}) = G_{iu}(-\mathbf{k})$ and $G_{iu}(\mathbf{k}) = G_{ui}(\mathbf{k})$, so that G_{iu} must have the form

$$G_{iu}(\mathbf{k},\omega) = R(k,\omega)\delta_{iu} + S(k,\omega)k_i k_u, \quad (14)$$

where R and S are (as yet) unknown scalar functions of arguments k, ω.

We can use the symmetry conditions (10) and (14) to simplify Eqs. (12) and (13).

First note that we require only $J_{iaja\lambda}$ in both Eqs. (12) and (13). From the definition (11b) and using the symmetry conditions (10) and (14) we must have

$$J_{iaja\lambda}(\mathbf{K},\Omega) = A\delta_{ij}K_\lambda + B\delta_{i\lambda}K_j + C\delta_{j\lambda}K_i + DK_i K_j K_\lambda, \quad (15)$$

where A, B, C, D are scalar functions of arguments K, Ω.

It then follows that

$$J_{iaja\lambda} - J_{jaia\lambda} = (B - C)(\delta_{ib}K_j - \delta_{jb}K_i). \quad (16)$$

We also have

$$I_{iajb}(\mathbf{k},\omega) = \alpha\delta_{ia}\delta_{jb} + \beta(\delta_{ij}\delta_{ab} + \delta_{ib}\delta_{ja}) + d\delta_{ia}k_j k_b + \epsilon\delta_{jb}k_i k_a + \zeta(\delta_{ib}k_j k_a + \delta_{ab}k_i k_j + \delta_{ja}k_i k_b + \delta_{ij}k_a k_b) + \mu\delta_{ia}k_j k_b + \nu k_i k_j k_a k_b, \quad (17)$$

where $\alpha, \beta, d, \epsilon, \zeta, \mu, \nu$ are scalar functions of arguments k, ω.

It follows that

$$k_b k_j(I_{jbia} - I_{ibja}) = [\alpha - \beta + k^2(\mu - \epsilon)](k^2\delta_{ia} - k_i k_a). \quad (18)$$

Use of Eqs. (16) and (18) in Eq. (12) and (13) gives

$$G_{iu}(\mathbf{k},\omega)(\eta k^2 - i\omega) = (2\pi)^{-4}\delta_{iu} - (2\pi)^4(\delta_{ia}k^2 - k_i k_a) \times G_{au}(\mathbf{k},\omega)\{[\alpha - \beta + k^2(\mu - \epsilon)] - (B - C)\} \quad (19)$$

and

$$B_a(\mathbf{k},\omega)\{(\eta k^2 - i\omega)\delta_{ia} + (2\pi)^{-4}\delta_{ia}k^2[\alpha - \beta + k^2(\mu - \epsilon) - (B - C)]\} = 0. \quad (20)$$

Now

$$B - C = -(2k^2)^{-1}k_i J_{iajaj} \quad (21a)$$

and

$$\alpha - \beta + k^2(\mu - \epsilon) = (2k^2)^{-1}k_a(k_i I_{iajj} - k_b I_{iaib}). \quad (21b)$$

Using Eqs. (10) and (14) in Eq. (21) gives

$$M(\mathbf{k},\omega) \equiv k^2[\alpha - \beta + k^2(\mu - \epsilon) - (B - C)] = \int E(K,\Omega)(k^2 - (\mathbf{k}.\mathbf{K})^2K^{-2})R(|\mathbf{k}-\mathbf{K}|, \times \omega - \Omega)d^3\mathbf{K}d\Omega. \quad (22)$$

Now use Eq. (22) in Eq. (19) to obtain

$$R(k,\omega)(\eta k^2 - i\omega + (2\pi)^4 M) = (2\pi)^{-4}, \tag{23}$$

$$k^2 S(k,\omega)(\eta k^2 - i\omega) = (2\pi)^4 M R(k,\omega). \tag{24}$$

Write $Q(k,\omega) = (2\pi)^4 R(k,\omega)$ when Eq. (23) becomes

$$Q(k,\omega)[\eta k^2 - i\omega + \int E(K,\Omega)(k^2 - (\mathbf{k}.\mathbf{K})^2 K^{-2}) \times Q(|\mathbf{k}-\mathbf{K}|, \omega - \Omega) d^3\mathbf{K} d\Omega] = 1. \tag{25}$$

So the immediate task before us is to solve Eq. (25) for $Q(k,\omega)$ with k *real*. Having obtained the solution we then note that the dispersion relation obtained from Eq. (20) is just[8]

$$\frac{1}{Q} \equiv \eta k^2 - i\omega + (2\pi)^4 M = 0, \tag{26}$$

for *complex* ω and real k.

Knowing the solution of Eq. (25) for Q, we look for the zeros of Q^{-1} for real k in the complex ω plane. These are the normal modes of the large-scale magnetic field.

Note that Eq. (25) is a nonlinear four dimensional integral equation for Q. Note further that the structure of the solution depends on the form chosen for $E(k,\omega)$. It is, therefore, difficult to solve in general.

IV. THE FIRST-ORDER SMOOTHING THEORY RESULT

Assume *a priori* that

$$|\eta k^2 - i\omega| \gg |\int E(K,\Omega)(k^2 - (\mathbf{k}.\mathbf{K})^2 K^{-2}) \times Q(|\mathbf{k}-\mathbf{K}|, \omega - \Omega) d^3\mathbf{K} d\Omega|. \tag{27}$$

Then a first iteration of Eq. (25) gives

$$\frac{1}{Q} \simeq \eta k^2 - i\omega + \int \frac{E(K,\Omega)(k^2 - (\mathbf{k}.\mathbf{K})^2 K^{-2})}{[\eta(\mathbf{k}-\mathbf{K})^2 - i(\omega - \Omega)]} d^3\mathbf{K} d\Omega. \tag{28}$$

With $Q^{-1} = 0$, Eq. (28) is *precisely* the dispersion relation obtained using FOST (see, e.g., LV, Appendix C).

The zeros of Eq. (28) for complex ω and real k then determine the normal modes of the large-scale magnetic field under first-order smoothing theory. The requirement that inequality (27) be valid then gives (for k, $\omega \to 0$) about

$$|\int E(K,\Omega) K dK d\Omega| \ll \eta^2. \tag{29}$$

Now $\langle V^2 \rangle = \int E(K,\Omega) d^3\mathbf{K} d\Omega$.

Let E be characterized by a scale-length L. Then equality (29) demands that

$$L\langle V^2 \rangle^{1/2} \ll \eta. \tag{30}$$

And Eq. (30) is the requirement that the Reynolds' number $L\langle V^2 \rangle^{1/2}/\eta$ should be small. Further, it can be shown (LV, Appendix C) that when $E(k,\omega) \geq 0$ then Eq. (28) has solutions with $\omega = \omega_R - i\sigma$ where $\sigma > 0$. So the normal modes of **B**, which were chosen to have the dependence $\exp(-i\omega t)$, are decaying and no regenerative dynamo action results under the approximation of FOST.

However, as Hammerstein (1930)[9] has pointed out, a uniformly convergent approximation of a nonlinear integral equation of Hammerstein's normal form[10] exists if, and only if, $|\int_0^Q f(y,u)du| \leq \frac{1}{2}C_1 Q^2 + C_2 Q$ for all Q, where, in our case, $C_1 (> 0)$ is the lowest positive eigenvalue of the real symmetric, positive definite kernel

$$K^2 \int_{-1}^{+1} (1 - \mu^2) E(|\mathbf{k} + \mathbf{K}|, \omega - \Omega) |\mathbf{k} + \mathbf{K}|^{-2} d\mu.$$

In the present case where $f \alpha 1/u$, this condition becomes

$$|\ln Q| \leq \tfrac{1}{2}C_1 Q^2 + C_2 Q$$

which is not satisfied for all Q. Accordingly any approximate solution to Eq. (25), like that given by FOST, is suspect.[11]

In order to ensure that an accurate solution is obtained one must first solve the nonlinear integral equation exactly and then look at the behavior of the solution for various values of the Reynolds' number. We shall first show how to solve Eq. (25) for particular forms of the turbulence spectrum and then we will point out what remains to be done in order to obtain solutions for arbitrary turbulence spectra. Consider then the exact solution to Eq. (25).

V. AN EXACT SOLUTION OF EQ. (25) FOR STATIC TURBULENCE

Let the velocity turbulence be independent of time when

$$E(K,\Omega) = \xi(K)\delta(\Omega), \tag{31}$$

with $\xi(K) \geq 0$.

Write $\varphi(k,\omega) = Q(k,\omega)^{-1}$ when Eq. (25) can be cast in the form

$$\varphi(k,\omega) - (\eta k^2 - i\omega) = \int [k^2 - (\mathbf{k}\cdot\mathbf{K})^2 K^{-2}] \xi(K) \times \varphi(|\mathbf{k}-\mathbf{K}|, \omega)^{-1} d^3K. \tag{32}$$

The immediate task before us is to solve Eq. (32) given $\xi(k)$ and to then find the zeros $\varphi(k,\omega)$ on the real k axis for complex ω.

The form of Eq. (32) can be simplified with the following reductions.

Let $\xi(k)$ be characterized by a scale-length L (the correlation length) and an "intensity" V^2 so that with $k \to kL$, $\omega \to \eta L^{-2}\omega$, $\varphi \to \eta k^2 \Phi$ we have[12]

$$\Phi(k,\omega) - (1 - i\omega k^{-2}) = R^2 \int_0^\infty \kappa^2 J(k,\kappa)\Phi(\kappa,\omega)^{-1} d\kappa, \tag{33}$$

where

$$J(k,\kappa) = 2\pi \int_{-1}^{+1} d\mu (1 - \mu^2) \xi(|k^2 + \kappa^2 + 2k\kappa\mu|^{1/2}) \times (k^2 + \kappa^2 + 2k\kappa\mu)^{-1}, \tag{34}$$

and where the magnetic Reynolds' number $R = LV/\eta$. In Eq. (34) $\xi(k)$ is, of course, the dimensionless energy density of the velocity turbulence per mode k.

Thus we are to solve the nonlinear integral Eq. (33); the normal modes of the ensemble average magnetic field are then given by the zeros of Φ. Note that since the mode dependence is $\exp(i\mathbf{k}.\mathbf{x})$, the zeros of Φ *must* occur on the real k axis in order to preserve physical sense. Zeros in the complex k plane would lead to spatially unbounded magnetic fields, which are physically absurd.

Note that $J(-k,\kappa) = J(k,\kappa) = J(\kappa,k)$, so that $\Phi(-k,\omega) = \Phi(k,\omega)$. So if Φ possesses a zero at $k = m$, it also possesses one at $k = -m$.

Supposing that $\Phi(k,\omega)$ does possess a zero at $k=m$ the first question that arises is how to interpret the singular integral occurring in Eq. (33). Here we follow a prescription given by Landau (1946)[13] and move the singularities from $k=\pm m$ to $k=\pm(m+i\epsilon)$ and then let $\epsilon\to+0$. This is done so that only physically causal solutions to the DIA equations arise. [For an elegant, and excellent, discussion of this point under wider conditions than we are concerned with here see Jackson (1960).[14]]

Now let $\xi(k)$ be chosen so that $J(k,\kappa)$ is analytic in both k and κ with an essential singularity on the circle at infinity. [For example, $\xi(k)\alpha k^2\exp(-k^2)$ is one such function.]

The method we give below for constructing solutions to Eq. (33) splits into two parts. First, there is consideration of a particular integral; second, we show that a limit form of the integral satisfies Eq. (33).

A. Consideration of an integral

Consider

$$I=\int_{-\infty}^{\infty}\kappa^2 J(k,\kappa)[1-i\omega k^{-2}+\Lambda_1 J(m,\kappa)+\Lambda_2 J(M,\kappa)]^{-1}d\kappa, \tag{35}$$

where $\pm(m+i\epsilon)$ and $\pm(M+i\epsilon)$ $(M>m)$ are zeros of $1-i\omega k^{-2}+\Lambda_1 J(m,\kappa)+\Lambda_2 J(M,\kappa)$ which, as $\epsilon\to+0$, lie on the real k axis at points $\kappa=\pm m$, $\kappa=\pm M$.

What we wish to show is that, apart from functions of ω, m, and M the integral I has the general form

$$I=aJ(m,k)+bJ(M,k).$$

If we can do so then by assuming

$$\Phi=1-i\omega k^{-2}+\Lambda_1 J(m,k)+\Lambda_2 J(M,k),$$

we can obtained *closed* expressions for Λ_1 and Λ_2 and so effect a solution to Eq. (33).

Consider then the integral path along the real κ axis from $-\infty$ to $+\infty$ and path closure by a large semicircle in the upper half complex κ plane. On the semicircle the integral of I will converge for $k<M$ at least as fast as $\exp(-|M-k|\,|\kappa|)$ [we have in mind $\xi(k)\sim k^n\exp(-k^2)$ at large k].

And then

$$I=2\pi i\sum_{m_i} m_i^2 J(k,m_i)[2i\omega m_i^{-3}+\Lambda_1 J'(m,m_i)+\Lambda_2 J'(M,m_i)]^{-1}, \tag{36}$$

where the sum m_i is over the zeros of the denominator $1-i\omega k^{-2}+\Lambda_1 J(m,\kappa)+\Lambda_2 J(M,\kappa)$ occurring for $\kappa>0$. Further $J'(m,m_i)\equiv(\partial/\partial k)J(m,k)|_{k=m_i}$.

As $M\to\infty$ (m finite) it is a simple matter to show that $J(k,M)=0$ $(\exp-(k-M)^2)$ and that $\Lambda_2\leq 0(M^n)$, so that

$$I(M\to\infty)=2\pi i m^2 J(k,m)[2i\omega m^{-3}+\Lambda_1 J'(m,m)]^{-1} \tag{37}$$

[and again we have in mind that $\xi(k)\sim k^n\exp(-k^2)$ for large k].

B. Application to Eq. (35)

Assume that $\Phi(k,\omega)$ has only two zeros, both on the positive real κ axis at $\kappa=m$ and $\kappa=M$, respectively.

Then write

$$\int_0^{\infty}\kappa^2 J(k,\kappa)\Phi(\kappa,\omega)^{-1}d\kappa=\frac{1}{2}\int_{-\infty}^{\infty}\kappa^2 J(k,\kappa)\Phi(\kappa,\omega)^{-1}d\kappa. \tag{38}$$

Now assume

$$\Phi(k,\omega)=1-i\omega k^{-2}+\Lambda_1 J(k,m)+\Lambda_2 J(k,M) \tag{39}$$

to obtain (as $M\to\infty$)

$$\Phi(k,\omega)=1-i\omega k^{-2}+i\pi m^2 R^2 J(k,m)\Big/\left.\frac{\partial\Phi(k,\omega)}{\partial k}\right|_{k=m}. \tag{40}$$

Equation (40) can be solved simply, for $\Phi(m,\omega)=0$, by definition and this implies

$$\left.\frac{\partial\Phi(k,\omega)}{\partial k}\right|_{k=m}\equiv\Phi'=-i\pi R^2 m^2 J(m,m)(1-i\omega m^{-2})^{-1}. \tag{41}$$

Also by differentiating Eq. (40) we obtain

$$\Phi'=2i\omega m^{-3}+i\pi R^2 m^2(\Phi')^{-1}\left(\frac{\partial J(k,m)}{\partial k}\right)_{k=m}. \tag{42}$$

Equations (41) and (42) are compatible if, and only if, a dispersion relation is satisfied.

The dispersion relation is

$$2(1+\tfrac{1}{2}mJ'/J)(1-i\omega m^{-2})=1\pm[1+2i\pi R^2 m^3\times(J+\tfrac{1}{2}mJ')]^{1/2}, \tag{43}$$

where $J\equiv J(m,m)$, $J'\equiv(\partial J(k,m)/\partial k)_{k=m}$.

It is a simple matter to show from Eq. (34) that $2J+mJ'>0$ for all m as long as $\xi(k)\geq 0$ for all k.

Now both modes of Eq. (43) are not permitted. For if $R\to 0$ (i.e., $V\to 0$) the upper sign predicts growing modes even in the absence of velocity turbulence. This is physically absurd. So the upper sign must be rejected on physical grounds. Then with $i\omega=-\sigma+iA$ (and with growing modes if $\sigma>0$) we obtain

$$2(1+\tfrac{1}{2}mJ'/J)(1+\sigma/m^2)=1-2^{-1/2}[(1+\alpha^2)^{1/2}+1]^{1/2}, \tag{44a}$$

$$2(1+\tfrac{1}{2}mJ'/J)A=2^{-1/2}m^2[(1+\alpha^2)^{1/2}-1]^{1/2}, \tag{44b}$$

where $\alpha=2\pi R^2 m^3(J+\tfrac{1}{2}mJ')$.

Equation (44a) gives $\sigma<0$ for all values of R. So as long as we restrict our attention to two zeros of $\Phi(k,\omega)$ (one at m and the second at $M\to\infty$) for incompressible, isotropic and static velocity turbulence, no growing modes exist provided that the turbulence spectrum makes $J(k,\kappa)$ analytic in the finite complex k and κ planes.

Author's note. The reader who is primarily interested in the physics may skip Secs. VI and VII and go directly to Sec. VIII. The reader who is more interested in the mathematical development should first read Secs. VI and VII before reading Sec. VIII.

VI. UNIQUENESS OF THE SOLUTION TO EQ. (33) AND A VARIATIONAL PRINCIPLE

In deriving the above results from Eq. (33) we assumed that it possessed only two zeros on the positive real κ axis at m and M, and we then let $M\to\infty$. It should be obvious that this procedure can be generalized.

A. Uniqueness of the solution

Let $\Phi(k,\omega)$ have simple zeros at $m_1, m_2, \ldots, m_n, M$ with $m_1<m_2<m_3\cdots<m_n<M$ on the positive real k

axis. Then by analogy with Sec. V we have (as $M \to \infty$)

$$\Phi(k,\omega) = 1 - i\omega k^{-2} + i\pi R^2 \sum_{i=1}^{n} m_i^2 J(k,m_i)/\Phi'(m_i,\omega). \quad (45)$$

In Eq. (45) set $k = m_\alpha$ ($\alpha = 1,\ldots,n$) with $\Phi(m_\alpha,\omega) = 0$ by definition. Then

$$1 - i\omega m_\alpha^{-2} = -i\pi R^2 \sum_{i=1}^{n} m_i^2 J(m_\alpha,m_i)/\Phi'(m_i,\omega), \quad (\alpha = 1,2,\ldots,n). \quad (46)$$

Also by differentiating Eq. (45) with respect to k and then setting $k = m_\alpha$ we obtain

$$\Phi'(m_\alpha,\omega) = 2i\omega m_\alpha^{-3} + i\pi R^2 \sum_{i=1}^{n} m_i^2 J'(m_\alpha,m_i)/\Phi'(m_i,\omega), \quad (\alpha = 1,2,\ldots,n). \quad (47)$$

Consider then the implications of Eqs. (46) and (47). We have $2n$ equations for the n unknowns $\Phi'(m_\alpha,\omega)$ which seems contradictory *if* the m_α are specified *a priori*. However, if we regard the m_α as arbitrary (except for m_1), then Eqs. (46) and (47) give $2n$ equations for the $2n$ unknowns $\Phi'(m_\alpha,\omega), m_2, m_3 \cdots m_n$ and ω, all of which can then be uniquely determined in terms of m_1.

Of the many dispersion relations for $\omega = \omega(m_1)$ which result only those which yield $i\omega = m_1^2$ as $R \to 0$ are physically permissible.

To illustrate the point let us consider in some detail Eqs. (46) and (47). Let $\Phi'(m_\alpha,\omega)m_\alpha \equiv u_\alpha$. Then from Eq. (47) we have

$$m_\alpha^2 u_\alpha - m_\beta^2 u_\beta = i\pi R^2 \sum_i m_i^3 u_i^{-1}(m_\alpha^3 J'_{\alpha i} - m_\beta^3 J'_{\beta i}), \quad (48)$$

where $J'_{\alpha i} \equiv (\partial/\partial k)J(k,m_i)\big|_{k=m_\alpha}$.

From Eq. (46) we have

$$m_\alpha^2 u_\alpha - m_\beta^2 u_\beta = 2(m_\alpha^2 - m_\beta^2) + i\pi R^2 \sum_i m_i^3 u_i^{-1} \times (m_\alpha^3 J'_{\alpha i} - m_\beta^3 J'_{\beta i} + 2m_\alpha^2 J_{\alpha i} - 2m_\beta^2 J_{\beta i}). \quad (49)$$

From Eqs. (48) and (49) it follows that if $m_\alpha \neq m_\beta$, then

$$2 = -i\pi R^2 \sum_i m_i^3 u_i^{-1}(m_\beta^2 J_{\beta i} - m_\alpha^2 J_{\alpha i})/(m_\beta^2 - m_\alpha^2). \quad (50)$$

Now all the m_i and the J's are positive. So from Eq. (50)

$u_i = iV_i$ with V_i real.

But if $u_i = iV_i$, it follows from Eq. (48) that

$$V_\alpha/V_\beta = m_\beta^2/m_\alpha^2 \quad (51)$$

and that

$$\sum_i m_i^5 (m_\alpha^3 J'_{\alpha i} - m_\beta^3 J'_{\beta i}) = 0 \quad (\alpha \neq \beta). \quad (52)$$

Equations (50) and (52) must be valid for all values of α and β with $\alpha \neq \beta$. But Eqs. (50) and (52), must be true for arbitrary choices of the turbulence spectrum $\xi(k)$ which give analytic J.

Hence Eq. (50) is not valid in general. The only other possibility is then $m_\alpha = m_\beta$ (all α,β). Thus there are no solutions to Eq. (33) which consist of more than one simple zero at *finite* m values.

What about solutions involving double, triple, etc., zeros of Φ?

Suppose there exists a solution to Eq. (33) with $J(k,\kappa)$ analytic which consists of (a) $m_1,\ldots,m_n$ being simple zeros of Φ, (b) $m_{r+1},\ldots,m_p$ being double zeros of Φ, and (c) $m_{p+1}\cdots$ being triple zeros of Φ, etc., with altogether N zeros being at finite m values; further let there be a last simple zero at $k = M$.

Then, as $M \to \infty$,

$$\Phi(k,\omega) = 1 - i\omega k^{-2} + i\pi R^2 \left\{ \sum_{i=1}^{r} m_i^2 \frac{J(k,m_i)}{\Phi'(m_i,\omega)} + \sum_{r+1}^{p} \frac{1}{\Phi''(m_i,\omega)} \frac{\partial}{\partial\kappa}(\kappa^2 J(k,\kappa))\Big|_{k=m_i} + \cdots \right\}. \quad (53)$$

Now Φ must vanish at $m = m_1 \cdots m_N$, Φ' must vanish at $m_{r+1},\ldots,m_p$, Φ'' must vanish at $m_{p+1}\cdots$, etc.

From Eq. (53) the vanishing of Φ gives N equations. By differentiating Φ we obtain N further equations. This is enough to give a unique solution, for we have $2N$ unknowns $[\omega, m_2\cdots m_n; \Phi'(m_1)\cdots\Phi'(m_r); \Phi''(m_{r+1})\cdots \Phi''(m_p)$; etc.]. But we have more equations available—for we can differentiate Eq. (53) twice to give $p - (r+1)$ further equations, etc. Thus the system is much overdetermined and no solution exists. Altogether then as long as $J(k,\kappa)$ is analytic in the finite domains of the complex k and κ planes, the only solution to Eq. (33) is when Φ has a simple zero at $k = m$ and a second simple zero at $k = M \to \infty$. And this solution does *not* admit of growing modes for any form of the static velocity turbulence [which preserves the analyticity of $J(k,\kappa)$].

What about choices of $\xi(k)$ which give nonanalytic $J(k,\kappa)$ [e.g., $\xi(k)\alpha k^n \exp(-k)$ or $\xi(k)\alpha k^2(k^2+1)^{-\alpha}$ give nonanalytic $J(k,\kappa)$ as can be verified directly]?

In such cases we have been unable to solve Eq. (33) analytically, but we have been able to construct a variational principle which enables solutions to be found, *in principle*.

B. The variational principle

Consider the expression

$$L = \int_0^\infty dk \left\{ \varphi^+(k,\omega)[\varphi(k,\omega) - (1 - i\omega k^{-2})] - R^2\varphi^+(k,\omega)\int_0^\infty d\kappa\kappa^2 J(k,\kappa)\varphi(\kappa,\omega)^{-1} \right\}. \quad (54)$$

If L is varied extremally with respect to the adjoint function φ^+ we recover Eq. (33). If L is varied extremally with respect to φ we obtain the adjoint equation

$$\varphi^+ = -R^2 k^2 \varphi^{-2} \int_0^\infty d\kappa J(k,\kappa)\varphi^+(\kappa,\omega). \quad (55)$$

Now multiply Eq. (33) by φ^+, Eq. (55) by φ, integrate the results over $0 \leq k \leq \infty$ to obtain

$$2\int_0^\infty \varphi\varphi^+ dk + \int_0^\infty dk(1 - i\omega k^{-2})\varphi^+ = 0, \quad (56a)$$

$$\int_0^\infty dk(1 - i\omega k^{-2})\varphi^+ - 2R^2\int_0^\infty dk\int_0^\infty d\kappa k^2 J(k,\kappa) \times \varphi^+(\kappa,\omega)\varphi(k,\omega)^{-1}. \quad (56b)$$

Suppose then that we had been able to solve Eqs. (33) and (55). Then upon using equations (56) in equation (54) we obtain

$$L = 0. \quad (57)$$

Thus we can use

$$\rho^2 = \int_0^\infty dk\varphi^+(k,\omega)[\varphi(k,\omega) - (1 - i\omega k^{-2})] \times \left[\int_0^\infty dk\varphi^+(k,\omega)\left[\int_0^\infty d\kappa\kappa^2 J(k,\kappa)\varphi(\kappa,\omega)^{-1}\right]\right]^{-1}, \quad (58)$$

as a variational principle for determining approximate solutions to Eqs. (33) and (55) when we are unable to perform the integrals analytically. Now $R^2 > 0$, so the variational statement for Eq. (58) is that for trial functions φ and φ^+; we have

$$\mathrm{Re}\rho^2 = R^2 > 0, \quad \mathrm{Im}\rho^2 = 0. \quad (59)$$

The "boundary conditions" on trial functions to be used in Eq. (58) are $\varphi(k \to \infty) = 1; \varphi^+(k \to \infty) = 0$; $\varphi(k \to 0) \to 1 - i\omega/k^2; \varphi^+(k \to 0) \to 0$. Further if $\varphi(k,\omega)$ is taken to have a zero of degree r at $k = m$, then φ^+ has a pole of degree $2r$ at $k = m$.

In view of the complexity of the variational statement (58), it is clear that some considerable effort is necessary in order to obtain even an approximate solution. However, in view of our inability to solve Eq. (33) directly when $J(k,\kappa)$ is *not* analytic in the finite domains of the complex k and κ planes, such effort may be worthwhile.

VII. TIME DEPENDENT VELOCITY TURBULENCE

All the results of Secs. V and VI are based on Eq. (33) and this, in turn, was derived from Eq. (25) using static velocity turbulence [see Eq. (31)].

More generally when $E(\kappa, \Omega)$ depends on Ω we have to solve

$$\Phi(k,\omega) = \eta k^2 - i\omega + \int E(\kappa,\Omega)[k^2 - (\mathbf{k}\cdot\boldsymbol{\kappa})^2 K^{-2}] \times \Phi(|\mathbf{k} - \boldsymbol{\kappa}|, \omega - \Omega)^{-1} d^3\kappa d\Omega, \quad (60)$$

where the normal modes of the ensemble average magnetic field are given through the dispersion relation

$$\Phi(k,\omega) = 0, \quad (61)$$

which determines the complex frequency ω as a function of real wavenumber k.

With

$$\Phi(k,\omega) = \eta k^2 \Psi(k,\omega), \quad (62)$$

equation (55) can be cast in the form

$$\Psi(k,\omega) = 1 - i\omega k^{-2} + \int_0^\infty \kappa^2 \frac{J(k,\kappa,\omega - \Omega)}{\Psi(\kappa,\Omega)} d\kappa d\Omega, \quad (63)$$

where

$$J(k,\kappa,\omega - \Omega) = 2\pi\eta^{-2}\int_{-1}^{+1}(1 - \mu^2)|\mathbf{k} + \boldsymbol{\kappa}|^{-2} \times E(|\mathbf{k} + \boldsymbol{\kappa}|, \omega - \Omega)d\mu, \quad (64)$$

with $|\mathbf{k} + \boldsymbol{\kappa}| = (k^2 + \kappa^2 + 2k\kappa\mu)^{1/2}$.

We have not been able to solve Eq. (63) but we have found a variational principle which, *in principle*, enables approximate solutions to be obtained to Eq. (58).

Consider

$$L = \int dkd\omega\{\varphi^+(k,\omega)[\Psi(k,\omega) - (1 - i\omega k^{-2})] - \varphi^+(k,\omega)\int \kappa^2 J(k,\kappa,\omega - \Omega)d\kappa d\Omega\Psi(\kappa,\Omega)^{-1}\}. \quad (65)$$

Extremal variation of L with respect to the adjoint function φ^+ gives Eq. (58), while extremal variation of L with respect to Ψ gives the adjoint equation

$$\varphi^+(k,\omega) = -k^2\Psi^{-2}\int J(k,\kappa,\omega - \Omega)\varphi^+(\kappa,\Omega)d\kappa d\Omega. \quad (66)$$

Then by multiplying Eq. (66) by Ψ, Eq. (58) by φ^+ and integrating the results over k and ω it follows that

$$2\int \Psi\varphi^+ dkd\omega = \int \varphi^+(1 - i\omega k^{-2})dkd\omega \quad (67)$$

and

$$\int \varphi^+(1 - i\omega/k^2)dkd\omega = -2\int dkd\omega d\kappa d\Omega k^2\Psi(k,\omega)^{-1}\varphi^+(\kappa,\Omega)J(k,\kappa,\omega - \Omega). \quad (68)$$

Suppose then that we had somehow obtained solutions to Eqs. (63) and (66). Then by using Eqs. (67) and (68) in Eq. (65) we see that

$$L = 0. \quad (69)$$

Altogether then

$$\rho^2 = \int dkd\omega\varphi^+(k,\omega)[\Psi(k,\omega) - (1 - i\omega k^{-2})] \times [\int dkd\omega\varphi^+(k,\omega)[\int d\kappa d\Omega\kappa^2 J(k,\kappa,\omega - \Omega)\Psi(\kappa,\Omega)^{-1}]]^{-1} \quad (70)$$

provides a variational principle for obtaining approximate solutions to Eqs. (63) and (66) for trial functions Ψ and φ^+. The variational statement is

$$\mathrm{Re}\rho^2 = 1, \quad \mathrm{Im}\rho^2 = 0. \quad (71)$$

Note that if we characterize Ψ by an amplitude A, then

$$|\rho^2| \alpha |A| (A \to 0), \quad |\rho^2| \alpha |A|^2 (A \to \infty). \quad (72)$$

So there must exist at least one amplitude A_* such that $|\rho^2| = 1$.

Then the real and imaginary parts of ω are to be chosen so that on $A = A_*$, $\mathrm{Re}\rho^2 = 1$, $\mathrm{Im}\rho^2 = 0$.

VIII. REMARKS ON THE DECAYING SOLUTION (44a)

Here we shall consider in a little more detail the behavior of the decay rate σ, and the oscillating frequency A, of the mode given by Eqs. (44a) and (44b).[15]

From Eq. (44a) we have

$$\sigma = -m^2 + \tfrac{1}{2}m^2(1 + \tfrac{1}{2}mJ'/J)^{-1}[1 - 2^{-1/2} \times [(1 + \alpha^2)^{1/2} + 1]^{1/2}], \quad (73a)$$

$$A = 2^{-3/2}m^2(1 + \tfrac{1}{2}mJ'/J)^{-1}[(1 + \alpha^2)^{1/2} - 1]^{1/2}, \quad (73b)$$

where $\alpha = 2\pi R^2 m^3(J + \tfrac{1}{2}mJ')$.

As $m \to \infty$, $J \to cm^{-4}$, $J' \to -mJ$, where c is a positive constant, so that $\alpha \to 2\pi cR^2/m$.

Then for $R \lesssim (m/c)^{1/2}$ we have $\alpha \ll 1$ and then

$$\sigma \simeq -m^2(1 + \tfrac{1}{16}\alpha^2), \quad (74a)$$

while for $R \gtrsim (m/c)^{1/2}$ we have

$$\sigma \simeq -2^{-3/4}m^2\alpha^{1/4}. \quad (74b)$$

Note further from Eq. (73a) that the decay rate of the field is at a rate greater than the free decay rate $-m^2$ for all values of the Reynolds number, R.

Note also that for any finite Reynolds' number the mode is decaying but oscillatory. This is in contrast to the situation $R = 0$ when the mode is purely decaying with no oscillatory component.

We do not yet completely understand the physics underlying this oscillatory behavior.

IX. DISCUSSION AND CONCLUSION

A. Detailed comments on the calculation

Using Kraichnan's direct interaction approximation we have set up the singular nonlinear integral equation describing evolution of a magnetic field under incompressible isotropic velocity turbulence. For a particular form of velocity turbulence [i.e., static and giving $J(k, \kappa)$ analytic in the finite domains of the complex k and κ planes] which satisfies Cramer's theorem we have given a method of solving the equation. By restricting our attention to the solution possessing two zeros (one at a finite value of k and the other at a value of k tending to infinity), we demonstrated that there existed two normal modes of the ensemble average magnetic field.

One of these modes is oscillatory but degenerative no matter where the finite zero of Eq. (33) is placed and no matter how large the Reynolds' number. The other mode predicts a growing field even in the absence of a turbulent velocity field. It must therefore by excluded on physical grounds.

When the velocity turbulence is static, but gives rise to a nonanalytic $J(k, \kappa)$, we have been unable to solve the nonlinear singular integral equation. However, we have been able to construct a variational principle which enables approximate solutions to be obtained, *in principle*. Unfortunately, the labor involved in using the variational principle is considerable, but unless a direct method can be found for solving the equation when $J(k, \kappa)$ is nonanalytic, recourse to such a variational principle is inevitable.

Further, when the velocity turbulence is not static we again have been unable to solve the resulting nonlinear singular equation. Once again we have found a variational principle which may be of use in obtaining approximate solutions if a direct method of solving the equation is not found.

We point out that the possibility of incompressible, isotropic, homogeneous and stationary velocity turbulence regenerating a large-scale magnetic field is still a wide open question. There remain to be investigated the cases of (i) nonanalytic $J(k, \kappa)$ for static velocity turbulence, and (ii) velocity turbulence which is not static. We do not yet possess the mathematical tools to investigate these situations.

We would, of course, be highly interested in seeing calculations relating to the above unanswered points. Until such time as these computations are forthcoming about all we can do is to emphasize that the question of whether isotropic velocity turbulence can regenerate a large scale field is as wide open now as it ever has been. What we have done is two things: First, we have shown that for static velocity turbulence which gives rise to analytic $J(k, \kappa)$ the answer is that no regeneration is possible; second, we have set up the equations that have to be solved when either $J(k, \kappa)$ is not analytic or the velocity turbulence is not static. We have not yet succeeded in solving them.

B. General comments on the physics

The generation of magnetic fields by turbulence has had a long history.

Batchelor (1950)[16] pointed out that there is a strong analogy between the vorticity ($\boldsymbol{\omega} = \nabla \times \mathbf{v}$) of a fluid field and an imbedded magnetic field, for the vorticity in a hydrodynamic fluid satisfies

$$\frac{\partial \omega_i}{\partial t} = \frac{\partial}{\partial x_j}(v_i \omega_j - v_j \omega_i), \tag{75}$$

while a magnetic field in a perfectly conducting fluid satisfies

$$\frac{\partial B_i}{\partial t} = \frac{\partial}{\partial x_j}(v_i B_j - v_j B_i). \tag{76}$$

On this basis it has been argued (Batchelor, 1950[16]; Chandrasekhar, 1950; 1951[17]; Moffatt, 1961[18]) that there should be equilibrium between the kinetic energy contained in the small eddies of a turbulent fluid and the magnetic energy no matter whether the turbulence is isotropic or not, with

$$\langle B^2/8\pi \rangle \approx \tfrac{1}{2}\rho \langle v^2 \rangle_{\text{small eddies}} \ll \tfrac{1}{2}\rho \langle v^2 \rangle_{\text{all eddies}}. \tag{77}$$

On the other hand it has been pointed out (Biermann and Schluter, 1951[19]; Biermann, 1953[20]; Chandrasekhar, 1955[21]) that in a perfectly conducting inviscid incompressible fluid the equations admit of an alternative symmetry:

$$\rho\left(\frac{\partial U_i}{\partial t} + V_j \frac{\partial U_i}{\partial x_j}\right) = -\frac{\partial}{\partial x_i}(p + B^2/8\pi), \tag{78}$$

$$\rho\left(\frac{\partial V_i}{\partial t} + U_j \frac{\partial V_i}{\partial x_j}\right) = -\frac{\partial}{\partial x_i}(p + B^2/8\pi), \tag{79}$$

where

$$U_i = v_i + B_i(4\pi\rho)^{-1/2}, \qquad V_i = v_i - B_i(4\pi\rho)^{-1/2}.$$

The Eqs. (78) and (79) treat U_i and V_i with equal symmetry. It has then been argued that it is reasonable to expect equipartition of energy between U_i and V_i, and between the Reynolds and Maxwell stresses of v_i and B_i, respectively. Thus in a turbulent system one expects

$$\langle B^2/8\pi \rangle \approx \tfrac{1}{2}\rho \langle v^2 \rangle_{\text{all eddies}}, \tag{80}$$

no matter how the turbulence is distributed (i.e., no matter whether it is isotropic or not).[22]

And in any event whichever argument appears more plausible to the reader, the main point is that a seed field should grow to a finite value.

Opposing these arguments are the mathematical results of first order smoothing theory and the diffusion argument (Parker, 1972[23]).

The results of FOST refer only to *small* Reynolds number ($R \ll 1$), and in this limit isotropic turbulence does not regenerate a large scale magnetic field (although it may still regenerate a completely turbulent magnetic field—this problem has not yet been solved).

The diffusion argument of Parker (1972) notes that Eq. (76) implies that the magnetic field is frozen into the fluid and so it is dragged around by the turbulence. It is diffused by the chaotic, jumbling motion just as smoke diffuses in the turbulent city air. This argument assumes only that there is a largest eddy size so that magnetic fields of larger scale cannot have energy transferred to them from even larger scale eddies, for there are none.

There are several points we do not understand in the solution of the Kraichnan DIA equations. First, we do not know what the physical reason is for the fact that both the growing and decaying modes are propagating. Second, while we are guaranteed that the ensemble of turbulent systems described by the Kraichnan DIA equations is physically realizable, we are not guaranteed that it describes the ensemble system that Nature provides.[24]

We should, perhaps, point out that there does not appear to be a better statistical description currently available than Kraichnan's DIA equations. Accordingly they are the best that we can do at the present time.

We believe that the DIA equations represent a powerful method of handling turbulent kinematic dynamo problems. In particular, for example, it has not escaped our attention that the inclusion of turbulent velocity possessing a net helicity is a problem which can now be tackled with complete generality using the Kraichnan DIA equations. We shall discuss this problem in the next paper in this series.

ACKNOWLEDGMENTS

We are grateful to Professor E. N. Parker and Dr. B. C. Low for many fruitful discussions concerning this paper. We are also grateful to Dr. R. H. Kraichnan for commenting on an earlier draft of this paper and for his critical and helpful comments on the present draft. We also wish to thank Professor P. H. Roberts who informs us that he has carefully checked and verified the mathematics.

The work reported here was supported by the United States Air Force, Air Force Systems Command under contract F-19628-72-C-0056.

[1]R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[2]U. Frisch in *Probablistic Methods and Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), Vol. 1.
[3]I. Lerche, Astrophys. J. **166**, 627 (1971) (LI); Astrophys. J. **166**, 639 (1971) (LII); Astrophys. J. **168**, 115 (1971) (LIII); Astrophys. J. **168**, 123 (1971) (LIV).
[4]L. Lerche and B. C. Low, Astrophys. J. **168**, 503 (1972) (L^2).
[5]I. Lerche, Ap. J. **181**, 993 (1973) (LV).
[6]F. Krause and P. H. Roberts, Ap. J. **181**, 977 (1973).
[7]H. Cramér, Ann. Math. **41**, 215 (1940).
[8]R. H. Kraichnan (private communication, 1973) has pointed out that when Eq. (26) is satisfied then $G_{ij} \propto (k^2\delta_{ij} - k_i k_j)$.
[9]A. Hammerstein, Acta Math. **54**, 117 (1930).
[10]A nonlinear integral equation is said to be in Hammerstein's normal form if it can be written $\psi(x) - \int K(x,y)\, f(y,\psi(y))\, dy = 0$.
[11]For example, the nonlinear integral equation $\psi(x) = \int_0^1 \alpha(x)\alpha(y)\psi(y) \times \sin[\psi(y)/\alpha(y)]dy$, with $\alpha(x) > 0$ in $0 \le x \le 1$ and $\int_0^1 \alpha(x)^2 dx = R^{-2}$, has the solution $\psi(x) = \Lambda\alpha(x)$, with $\sin\Lambda = R^2$. For $R^2 \le 1$ there exists a countable infinite of real values of Λ satisfying $\sin\Lambda = R^2$; while for $R^2 > 1$ there are no real values of Λ satisfying $\sin\Lambda = R^2$, and then $\Lambda = \Lambda_0 + i\gamma$ with $\Lambda_0 = (n + 1/2)\pi$, $\cosh\gamma = R^2$, $n = 0, \pm 2, \pm 4, \cdots$. This gives *two* values of γ for each value of Λ_0. If we had approximated the equation by $\psi(x) = \alpha(x)\int_0^1 \psi(y)^2 dy$, we would find $\psi(x) = \Lambda\alpha(x)$ with $\Lambda = R^2$ which demands $R^2 \ll 1$ in order to be valid. The other real solutions which also exist when $R^2 \ll 1$ and are given by $\Lambda_n \simeq n\pi + (-1)^n R^2$, and the complex Λ values (for $R^2 > 1$), are not obtained by approximation.
[12]First order smoothing theory sets $\Phi = 1 - i\omega k^{-2}$ on the right-hand of Eq. (33) and $\Phi = 0$ on the left-hand side. This appears to be a mutually inconsistent pair of approximations.
[13]L. D. Landau, J. Phys. USSR **10**, 25 (1946).
[14]J. D. Jackson, J. Nucl. Energy 1, 171 (1960).
[15]As Kraichnan, Ref. 8 above, has pointed out the fact that normal modes of $\langle B \rangle$ decay does not, in general, imply that $\langle B \cdot B \rangle$ decays—in fact, the converse may obtain. But if $\langle B \cdot B \rangle$ decays then this does imply that $\langle B \rangle$ decays.
[16]G. K. Batchelor, Proc. R. Soc. A **201**, 405 (1950).
[17]S. Chandrasekhar, Proc. R. Soc. A **204**, 435 (1950), Proc. R. Soc. A **207**, 301 (1951).
[18]K. Moffat, J. Fluid Mech. **11**, 625 (1961).
[19]L. Biermann and Schluter, Phys. Rev. **82**, 863 (1951).
[20]L. Biermann, in *Kosmische Strahlung*, edited by W. Heisenberg (Springer-Verlag, Berlin, 1953).
[21]S. Chandrasekhar, Proc. R. Soc. A **229**, 1 (1955); Proc. R. Soc. A **233**, 322 (1955); Proc. R. Soc. A **233**, 330 (1955).
[22]See Ref. 15.
[23]E. N. Parker, Astrophys. J. **157**, 1119 (1969); Astrophys. J. **160**, 383 (1970); Astrophys. J. **174**, 499 (1972).
[24]For example, in playing roulette there is an ensemble of systems in which every other game is random and in which the remaining games favor the house by more than a two to one margin. Hopefully this is not the ensemble of systems provided in Las Vegas.

Perturbation method for a nonlinear wave modulation. III

Tosiya Taniuti

Department of Physics, Nagoya University, Nagoya, Japan

Nobuo Yajima

Research Institute for Applied Mechanics, Kyusyu University, Fukuoka, Japan
(Received 22 October 1971; revised manuscript received 3 January 1973)

The perturbation method for the nonlinear, slow modulation of a rapidly oscillatory plane wave, which was given in the first paper of this series for a class of systems of nonlinear partial differential equations, is now established for a general system of nonlinear integro-partial differential equations. It is shown that the system can be reduced to simpler nonlinear equations which in certain cases become the nonlinear Schrödinger equation. The reduction proceeds as in the first paper, and the result is then applied to nonlinear optics.

1. INTRODUCTION

Using a multiscale perturbation method, Taniuti and Washimi[1] have shown that the nonlinear modulation of a quasimonochromatic whistler wave propagating in a cold plasma can be described by the nonlinear Schrödinger equation. In the first paper (I) of this series,[2] it was demonstrated that the perturbation method of these authors applies to a wide class of nonlinear partial differential equations which are linearly dispersive and that in the general case the nonlinear modulation can be governed by a nonlinear equation of the Schrödinger type. Also, an intuitive, entirely different derivation of the nonlinear Schrödinger equation was given by Karpman and Krushkal[3] for a quasimonochromatic dispersive wave with frequency depending on a single nonlinear parameter—the slowly varying amplitude. Furthermore, conspicuous properties of the nonlinear Schrödinger equation such as the soliton–soliton interaction were first observed numerically by Yajima and Outi,[4] and recently it was shown by Zakharov and Shabat[5] that the nonlinear Schrödinger equation is solvable analytically. Hence, it may be stated that for a general set of partial differential equations of the dispersive type, the characteristic behavior of the nonlinear wave modulation can be found in the asymptotic way.

In the following sections of this paper, we shall consider a more general system of equations which are integro-differential, and it will be shown that a similar reduction to a nonlinear equation of the Schrödinger type is possible. This generalization was motivated by recent developments in nonlinear optics. For illustration, let the nonlinear Maxwell equation for the electric field vector $\mathbf{E}$ be given by

$$-(1/c^2)(n^2 + \alpha|\mathbf{E}|^2)\frac{\partial^2\mathbf{E}}{\partial t^2} + \frac{\partial^2\mathbf{E}}{\partial x^2} = 0,$$

where n and α are real constants. This equation can be brought into the matrix form (I.1). [Here (I.1) denotes Eq. (1) in the first paper, I, of this series. In what follows, similar notation will be used for any equation in I.] However, in this case, the linear dispersion relation (I.4) becomes

$$\omega^2 - (c/n)^2k^2 = 0;$$

consequently, there is no dispersion and our method is not applicable. [Note that the condition (I.4′) is not valid for any $|l| \neq 1$]. Rather, the above equation should be regarded as a special case of the system of equations considered by Taniuti and Wei[6] for long wavelength, and it can be reduced[7] to a single nonlinear equation of the form, $\partial\varphi/\partial\tau + \bar{\alpha}\varphi\,\partial\varphi/\partial\xi = 0$, where $\bar{\alpha}$ is constant. (Here we note that the equation can be solved exactly by means of the method of characteristics, but because of characteristic crossing, the waves break after a finite time.[8,9]) This implies that our method in I does not apply to nonlinear optics, if (color) dispersion in the linear approximation is neglected. However, physically speaking, as waves steepen, color dispersion does become effective in preventing characteristic crossing and would allow us to describe behaviors of longer time scales. In order to take into account dispersion, we may consider coupling between the Maxwellian equations and a polarization field which is assumed to be governed by a set of differential equations. For example, Ostrovskii[10] modeled the polarization field by an anharmonic oscillator coupled with the electromagnetic field. Our method in the first paper can be applied to his model and will lead to a similar result, and appropriate modifications will then enable us to apply it to more realistic models. (Also the method could be extended so as to be applicable to systems with relaxation,[11] such as those involving envelope shocks[12]). However, as an alternative way to take account of the dispersion, we may introduce a general dielectric tensor which is nonlocal in the coordinate space so that the electric displacement is given as a convolution of the dielectric tensor and the electric field vector.[13] Then, the Maxwell equations become integro-differential equations. This leads us to consider a general system of integro-differential equations. In the next section, a simple model equation is considered to explain the outline of the general theory which will be developed in Sec. 3. Then the general theory is illustrated in Sec. 4 by an application to nonlinear optics.

2. ILLUSTRATION BY MODEL EQUATION

To begin, we first consider the following integral equation as a model equation:

$$\int w(x_1,t_1)u(x-x_1,t-t_1)\,dx_1dt_1 + \int \mathrm{X}(x_1,t_1;x_2,t_2)u(x-x_1,t-t_1) \times u(x-x_2,t-t_2)\,dx_1dt_1dx_2dt_2 = 0.$$

Here u is a scalar, unknown, real variable of the space and time coordinates x and t, with the first term representing a linear response and the second a nonlinear one. Consequently, w and X may be regarded as a linear and a nonlinear dielectric function, respectively, and it will be assumed that w and X are real and fall off sufficiently rapidly as their arguments $|x_1|$, $|t_1|$, $|x_2|$, $|t_2|$ go to infinity, while the integration is carried

from $-\infty$ to $+\infty$ for each argument. For infinitesimal disturbances, we have the linear integral equation,

$$\int w(x_1, t_1)u(x - x_1, t - t_1)\,dx_1 dt_1 = 0,$$

which immediately yields the dispersion relation

$$\tilde{w}(k, \omega) = 0,$$

where $\tilde{w}$ is the Fourier transform of w,

$$\int w(x, t)\,\exp[-i(kx - \omega t)]\,dx dt.$$

Corresponding to the condition (I. 4'), we assume

$$\tilde{w}(lk, l\omega) \neq 0 \quad \text{for } |l| \neq 1, \quad \text{i.e., for } |l| = 0, 2, 3, \ldots,$$

and consider the wave modulation given by Eq. (I. 5),

$$u = \sum_{\alpha=1} \sum_{l=-\infty}^{\infty} \epsilon^{\alpha} u_l^{(\alpha)}(\xi, \tau) Z_l.$$

Here the notation is the same as that of I, i.e., $\xi = \epsilon(x - \lambda t)$, $\tau = \epsilon^2 t$, $\lambda = \partial\omega/\partial k$, and Z_l stands for $\exp[il(kx - \omega t)]$. Then, Taylor-expanding $u(x - x_1, t - t_1)$ with respect to the slow variables ξ_1 and τ_1 and integrating with respect to x_1 and t_1 gives the following expression for the linear response:

$$\begin{aligned}
&\int w(x_1, t_1)u(x - x_1, t - t_1)\,dx_1 dt_1 \\
&= \sum_{\alpha} \sum_{l} \epsilon^{\alpha} \left[\tilde{w}_l u^{(\alpha)} - i \frac{\epsilon}{l} \frac{d\tilde{w}_l}{dk} \frac{\partial u_l^{(\alpha)}}{\partial \xi} + i \frac{\epsilon^2}{l} \frac{\partial \tilde{w}_l}{\partial \omega} \frac{\partial u_l^{(\alpha)}}{\partial \tau} \right. \\
&\quad \left. - \frac{1}{2} \frac{\epsilon^2}{l^2} \left(\frac{d\tilde{w}_l}{dk} - \frac{d^2\omega}{dk^2} \frac{\partial \tilde{w}_l}{\partial \omega} \right) \frac{\partial^2 u_l^{(\alpha)}}{\partial \xi^2} + \cdots \right] Z_l.
\end{aligned}$$

In deriving this equation we have used the following identities and notations:

$$\int w(x_1, t_1) Z_l^*(x_1, t_1)\,dx_1 dt_1 = \tilde{w}(lk, l\omega) \equiv \tilde{w}_l,$$

$$\begin{aligned}
\int w(x_1, t_1)\xi_1 Z_l^*(x_1, t_1)\,dx_1 dt_1 &= i \frac{\epsilon}{l} \left(\frac{\partial}{\partial k} + \lambda \frac{\partial}{\partial \omega} \right) \tilde{w}_l \\
&\equiv i \frac{\epsilon}{l} \frac{d\tilde{w}_l}{dk},
\end{aligned}$$

$$\int w(x_1, t_1)\tau_1 Z_l^*(x_1, t_1)\,dx_1 dt_1 = -i \frac{\epsilon^2}{l} \frac{\partial}{\partial \omega} \tilde{w}_l,$$

$$\frac{\partial \tilde{w}}{\partial \omega} \equiv \frac{\partial \tilde{w}}{\partial \omega}\bigg|_{\tilde{w}(k,\omega)=0}.$$

Similarly, for the nonlinear response, we have an expansion in powers of ϵ, which begins with order ϵ^2. Combining these two expansions, one sees easily that the terms to first order in ϵ yield

$$\tilde{w}_l u_l^{(1)} = 0,$$

i.e., $u_l^{(1)} = 0$, if $|l| \neq 1$, while for $|l| = 1$ the equation is satisfied automatically so that $u_{\pm 1}^{(1)}$ remains undetermined in this order. However, also in the second order, the coefficient of $Z_{\pm 1}$ of the linear response vanishes identically because of the dispersion relation and its derivative, $d\tilde{w}_{\pm 1}/dk = 0$. Since the nonlinear response in this order is given by the products of the first order terms $u_1^{(1)}$ and $u_{-1}^{(1)}$, it does not contain the factors $Z_{\pm 1}$. Hence for $|l| = 1$, the second order term does not appear. However, for $|l| \neq 1$, we have equations to determine $u_l^{(2)}$ in terms of $u_{\pm 1}^{(1)}$, i.e.,

$$u_2^{(2)} = -(\tilde{X}_{1,1}/\tilde{w}_2) u_1^{(1)} u_1^{(1)},$$

$$u_{-2}^{(2)} = -(\tilde{X}_{-1,-1}/\tilde{w}_{-2}) u_{-1}^{(1)} u_{-1}^{(1)},$$

$$u_0^{(2)} = -(2\tilde{X}_{00}/\tilde{w}_0) |u_1^{(1)}|^2;$$

otherwise,

$$u_l^{(2)} = 0,$$

where

$$\tilde{X}_{ll'} \equiv \int X(x_1, t_1; x_2, t_2) Z_l^*(x_1, t_1) Z_{l'}^*(x_2, t_2)\,dx_1 dx_2 dt_1 dt_2.$$

(Note that $u_{\pm 1}^{(2)}$ and $u_{\pm 1}^{(1)}$ are not yet determined.)

We now proceed to third order in ϵ. For the coefficient of Z_1 of the linear response, it is again readily seen that the coefficients of $u_1^{(3)}$ and $u_1^{(2)}$ vanish, and we have

$$i \frac{\partial \tilde{w}}{\partial \omega} \frac{\partial u_1^{(1)}}{\partial \tau} + \frac{\omega''}{2} \frac{\partial \tilde{w}}{\partial \omega} \frac{\partial^2 u_1^{(1)}}{\partial \xi^2},$$

while the nonlinear response yields

$$\begin{aligned}
&\int X(x_1, t_1; x_2, t_2) \{ u_2^{(2)}(\xi, \tau) u_{-1}^{(1)}(\xi, \tau) [Z_2^*(x_1, t_1) Z_{-1}^*(x_2, t_2) \\
&\quad + Z_2^*(x_2, t_2) Z_{-1}^*(x_1, t_1)] \\
&\quad + u_0^{(2)}(\xi, \tau) u_1^{(1)}(\xi, \tau) [Z_0^*(x_1, t_1) Z_1^*(x_2, t_2) \\
&\quad + Z_0^*(x_2, t_2) Z_1^*(x_1, t_1)] \} dx_1 dt_1 dx_2 dt_2 \\
&= [-(\tilde{X}_{2,-1} + \tilde{X}_{-1,2})(\tilde{X}_{1,1}/\tilde{w}_2) \\
&\quad - 2(\tilde{X}_{0,1} + \tilde{X}_{1,0})(\tilde{X}_{00}/\tilde{w}_0)] |u_1^{(1)}|^2 u_1^{(1)}.
\end{aligned}$$

Therefore we obtain a nonlinear equation of the Schrödinger type

$$i \frac{\partial u_1^{(1)}}{\partial \tau} + \frac{\omega''}{2} \frac{\partial^2 u_1^{(1)}}{\partial \xi^2} + Q |u_1^{(1)}|^2 u_1^{(1)} = 0,$$

in which Q is given by

$$\begin{aligned}
Q = \left(\frac{\partial \tilde{w}}{\partial \omega} \right)^{-1} \Bigg[&(\tilde{X}_{2,-1} + \tilde{X}_{-1,2}) \left(\frac{\tilde{X}_{1,1}}{\tilde{w}_2} \right) \\
&+ 2(\tilde{X}_{0,1} + \tilde{X}_{1,0}) \left(\frac{\tilde{X}_{00}}{\tilde{w}_0} \right) \Bigg].
\end{aligned}$$

Q is, in general, complex; however, if the linear and nonlinear dielectric functions w and X are even with respect to their arguments, the Fourier transforms $\tilde{w}_l$, $\tilde{X}_{l,l'}$ are real, and consequently Q becomes real.

Though we have considered the quadratic nonlinear response, in nonlinear optics the nonlinear response is usually cubic and takes a slightly different convolution form,

$$\begin{aligned}
&\int X(x_1, t_1; x_2, t_2; x_3, t_3) u(x - x_1, t - t_1) \\
&\quad \times u(x - x_1 - x_2, t - t_1 - t_2) \\
&\quad \times u(x - x_1 - x_2 - x_3, t - t_1 - t_2 - t_3) \\
&\quad \times dx_1 dx_2 dx_3 dt_1 dt_2 dt_3.
\end{aligned}$$

The reduction for this case is essentially similar, and if X is even, Q is real.

3. GENERAL THEORY

The extension to the system of integro-differential equations can be achieved by incorporating the derivation for the single integral equation into the method in I for the system of differential equations, and it is anticipated that reduction will yield a nonlinear equation of the same Schrödinger type. [The coefficient of the second order derivative p in Eq. (I. 16) can be reduced easily to $\frac{1}{2}d^2\omega/dk^2$ by a simple, algebraic manipulation, which was first pointed out by Asano.][11,14]

In the following discussion, in order to include the Maxwell equations in a dispersive medium as a special case, we shall consider the following system of equations for n real unknowns $u_1, u_2, \ldots, u_n$,

$$\frac{\partial F^\mu}{\partial x^\mu} + G = 0. \tag{1}$$

Here, the Greek superscript μ runs from 0 to 3, the x^μ's are the space-time coordinates, namely, x^0 denotes time t while x^1, x^2, and x^3 are the space coordinates, and a dummy index will be used throughout unless otherwise stated. The F^μ's are n-component column vectors (f_i^μ, $i = 1, 2, \ldots, n$) given by

$$\begin{aligned} f_i^\mu = {} & \int K_{ij}^{\mu 1}(x_1) u_j(x - x_1)\,(dx_1)^4 \\ & + \iint K_{ijk}^{\mu 2}(x_1, x_2) u_j(x - x_1) u_k(x - x_1 - x_2) \\ & \times (dx_1)^4 (dx_2)^4 + \cdots + \int \cdots \int K_{i,j,\ldots,s}^{\mu N}(x_1 \cdots x_N) \\ & \times u_j(x - x_1) \cdots u_s\Big(x - \sum_{\nu=1}^{N} x_\nu\Big)\,(dx_1)^4 \cdots (dx_N)^4. \end{aligned} \tag{2}$$

Here $K^{\mu r}$ ($r = 1, \ldots N$) is a real tensor of rank $r + 1$ and its arguments $x_1, x_2, \ldots, x_r$ denote that $K^{\mu r}$ depends on the r-space-time points, $x_1(x_1^\mu), x_2(x_2^\mu), \ldots, x_r(x_r^\mu)$; also, $u_j(x - x_1 - x_2 \cdots - x_m)$ denotes that u_j is a function of the four variables $x^\mu - x_1^\mu - x_2^\mu - \cdots - x_m^\mu$ ($\mu = 0, 1, 2, 3$); $(dx_r)^4$ is the four-dimensional volume element $dx_r^0 dx_r^1 dx_r^2 dx_r^3$; the domain of integration ranges from $-\infty$ to $+\infty$ for each variable; and the dummy indices $i, j, \ldots, s$, run from 1 to n. Equation (2) will be symbolically designated as

$$F^\mu = \sum_{r=1}^{N} K^{\mu r} * U^r, \tag{2'}$$

where U is a column vector of the n components $u_1, u_2, \ldots, u_n$, and G is given likewise,

$$G = \sum_{r=1}^{N'} M^r * U^r, \tag{3}$$

in which M^r is a real tensor of rank $r + 1$ and depends on the points $x_1, x_2, \ldots, x_r$. It will be assumed that all the tensors $K^{\mu r}, M^r$ damp out sufficiently rapidly as their arguments go to plus and minus infinity; i.e.,

$$\int \cdots \int K^{\mu r} \cdot (x^\nu)^p \cdot (dx_1)^4 \cdots (dx_r)^4 \quad (r = 1, 2, \ldots, N)$$

and

$$\int \cdots \int M^r \cdot (x^\nu)^p \cdot (dx_1)^4 \cdots (dx_r)^4 \quad (r = 1, 2, \ldots, N')$$

are bounded for $p = 0, 1, 2, \cdots$. Also, necessary analyticities with respect to their arguments will be assumed. We now introduce a crucial assumption for the Fourier transform of the tensors $K^{\mu 1}$ and M^1:

Let $\tilde{K}_l^{\mu r}$ and $\tilde{M}_l^r$ be the Fourier transforms of $K^{\mu r}$ and M^r, i.e.,

$$\tilde{K}_l^{\mu r} = \int \cdots \int K^{\mu r}(x_1, \ldots, x_r) \exp\Big(-il \sum_{s=1}^{r} k_s \cdot x_s\Big) \times (dx_1)^4 \cdots (dx_r)^4,$$

$$\tilde{M}_l^r = \int \cdots \int M^r(x_1, \ldots, x_r) \exp\Big(-il \sum_{s=1}^{r} k_s \cdot x_s\Big) \times (dx_1)^4 \cdots (dx_r)^4,$$

where $k_s \cdot x_s = k_s^\mu x_s^\mu$.

Define the matrix W_l by

$$W_l = ilk^\mu \tilde{K}_l^{\mu 1} + \tilde{M}_l^1.$$

Then, we assume that the equation

$$\det W_{\pm 1} = 0 \tag{4}$$

admits at least one real root k^0 for a given set of the vectors $\mathbf{k}(k^1, k^2, k^3)$; and, moreover, when (4) is valid,

$$\det W_l \neq 0 \quad \text{for } |l| \neq 1, \tag{5}$$

i.e., for $l = 0$ and for any integer l which is not equal to ± 1.

It is readily seen that Eq. (4) is the dispersion relation of the linearized system of (1), corresponding to Eq. (I. 4), while the inequality (5) corresponds to the condition (I. 4′).

Following the expansion method used in I, we consider a solution about $U = 0$:

$$U = \sum_{\alpha=1} \epsilon^\alpha U^{(\alpha)}, \tag{6a}$$

$$U^{(\alpha)} = \sum_{l=-\infty}^{\infty} U_l^{(\alpha)}(\xi, \tau) \exp(ilk^\mu x^\mu). \tag{6b}$$

Here ϵ is a small parameter, k^μ is a set of the frequency and the wavenumber satisfying Eq. (4), and τ and ξ are stretched coordinates introduced by

$$\tau = \epsilon^2 t, \tag{7a}$$

$$\xi = \epsilon \bar{k}^\mu x^\mu, \tag{7b}$$

where $\bar{k}^\mu$ is defined as,

$$\bar{k}^0 = \mathbf{k} \cdot \frac{\partial k^0}{\partial \mathbf{k}} \quad \left(\equiv \sum_{i=1}^{3} k_i \frac{\partial k^0}{\partial k_i}\right), \tag{8a}$$

$$\bar{k}^i = k^i = k_i \quad (i = 1, 2, 3). \tag{8b}$$

Here we note that

$$\bar{k}^\mu x^\mu = \mathbf{k} \frac{\partial}{\partial \mathbf{k}} (k^\mu x^\mu). \tag{8c}$$

Then we have

$$\begin{aligned} K^{\mu 1} * U &= \sum_\alpha \sum_l \epsilon^\alpha K^{\mu 1} * U_l^{(\alpha)} \exp(ilk \cdot x) \\ &= \sum_{\alpha=1} \sum_i \epsilon^\alpha \left\{ \tilde{K}_l^{\mu 1} + \left[\epsilon \tilde{K}_{\xi,l}^{\mu 1} \frac{\partial}{\partial \xi} + \epsilon^2 \left(\tilde{K}_{\tau,l}^{\mu 1} \frac{\partial}{\partial \tau} + \tilde{K}_{\xi\xi,l}^{\mu 1} \frac{\partial^2}{\partial \xi^2} \right) \right. \right. \\ &\quad \left. \left. + \cdots \right] U_l^{(\alpha)}(\xi, \tau) \right\} \exp(ilk \cdot x) \\ &\equiv \sum_l \langle K^{\mu 1} * U \rangle_l \exp(ilkx), \end{aligned} \tag{9}$$

in which $\tilde{K}^{\mu 1}_{\xi,l}$, etc., are given in Appendix A, and $k\cdot x$ denotes $k^\mu x^\mu$.

The reductions of other terms $K^{\mu\tau} * U^\tau$ go parallel; for example,

$$K^{\mu 2} * U^2 = \sum_{\alpha,\beta}\sum_{l,l'} \epsilon^{\alpha+\beta} K^{\mu 2} * U_l^{(\alpha)} U_{l'}^{(\beta)} \exp[i(l+l')k\cdot x]$$
$$= \sum_{\alpha,\beta}\sum_{l,l'} \epsilon^{\alpha+\beta}\Big\{ \tilde{K}^{\mu 2}_{l+l';l,l'} U_l^{(\alpha)} U_{l'}^{(\beta)} + \epsilon\Big[\tilde{K}^{\mu 2}_{\xi,l+l';l'}\Big(\frac{\partial U_l^{(\alpha)}}{\partial\xi} U_{l'}^{(\beta)} + U_l^{(\alpha)}\frac{\partial U_{l'}^{(\beta)}}{\partial\xi}\Big) + \tilde{K}^{\mu 2}_{l+l';\xi,l'} U_l^{(\alpha)}\frac{\partial U_{l'}^{(\beta)}}{\partial\xi}\Big] + \cdots\Big\} \exp[i(l+l')k\cdot x]$$
$$\equiv \sum_{l,l'} \langle K^{\mu 2} * U^2\rangle_{l,l'} \cdot \exp[i(l+l')k\cdot x] \tag{10}$$

in which $\tilde{K}^{\mu 2}_{l+l';l'}$ etc. are defined in Appendix A,

$$K^{\mu 3} * U^3 = \sum_{\alpha,\beta,\gamma}\sum_{l,l',l''} \epsilon^{\alpha+\beta+\gamma}[\tilde{K}^{\mu 3}_{l+l'+l'';l'+l'';l''} \times U_l^{(\alpha)} U_{l'}^{(\beta)} U_{l''}^{(\gamma)} + \cdots] \exp[i(l+l'+l'')k\cdot x]$$
$$\equiv \sum_{l,l',l''} \langle K^{\mu 3} * U^3\rangle_{l,l',l''} \exp[i(l+l'+l'')k\cdot x], \tag{11}$$

where $K^{\mu 3}_{l+l'+l'';l'+l'';l''}$ is given in Appendix A.

Similarly, replacing $\tilde{K}^{\mu 1}_l$, etc., by $\tilde{M}^1_l$, etc., defined in correspondence to equations in the Appendix A, we have the expression for G.

In view of Eqs. (9), (10), (11), etc., one can perform the differentiation with respect to x^μ in Eq. (1):

$$\frac{\partial F^\mu}{\partial x^\mu} = \sum_l (ilk^\mu)\langle K^{\mu 1} * U\rangle_l \exp(ilk\cdot x) + \sum_{l,l'} i(l+l')k^\mu \langle K^{\mu 2} * U^2\rangle_{l,l'} \exp[i(l+l')k\cdot x] + \sum_{l,l',l''} i(l+l'+l'')k^\mu \langle K^{\mu 3} * U^3\rangle_{l,l',l''} \times \exp[i(l+l'+l'')k\cdot x] + \cdots$$
$$+ \sum_l \left(\epsilon\bar{k}^\mu\frac{\partial}{\partial\xi} + \epsilon^2\delta_{\mu 0}\frac{\partial}{\partial\tau}\right)\cdot\langle K^{\mu 1} * U\rangle_l \cdot \exp(ilk\cdot x) + \sum_{l,l'}\left(\epsilon\bar{k}^\mu\frac{\partial}{\partial\xi} + \epsilon^2\delta_{\mu 0}\frac{\partial}{\partial\tau}\right)\langle K^{\mu 2} * U\rangle_{l,l'} \times \exp[i(l+l')k\cdot x] + \cdots. \tag{12}$$

Then, equating the various powers of ϵ of the same harmonics to zero, we get for $O(\epsilon)$

$$(ilk^\mu \tilde{K}^{\mu 1}_l + \tilde{M}^1_l) U_l^{(1)} = 0; \tag{13}$$

for $O(\epsilon^2)$

$$(ilk^\mu \tilde{K}^{\mu 1}_l + \tilde{M}^1_l) U_l^{(2)} + ilk^\mu \tilde{K}^{\mu 1}_{\xi,l}\frac{\partial}{\partial\xi} U_l^{(1)} + \bar{k}^\mu\frac{\partial}{\partial\xi}(\tilde{K}^{\mu 1}_l U_l^{(1)}) + \tilde{M}^1_{\xi,l}\frac{\partial}{\partial\xi} U_l^{(1)} + \sum_{l'} (ilk^\mu \tilde{K}^{\mu 2}_{l;l'} + \tilde{M}^2_{l;l'}) U_{l-l'}^{(1)} U_{l'}^{(1)} = 0;$$

and for $O(\epsilon^3)$

$$(ilk^\mu \tilde{K}^{\mu 1}_l + \tilde{M}^1_l) U_l^{(3)} + (ilk^\mu \tilde{K}^{\mu 1}_{\xi,l} + \bar{k}^\mu \tilde{K}^{\mu 1}_l + \tilde{M}^1_{\xi,l})\frac{\partial}{\partial\xi} U_l^{(2)} + \left[ilk^\mu \tilde{K}^{\mu 1}_{\tau,l}\frac{\partial}{\partial\tau} + ilk^\mu \tilde{K}^{\mu 1}_{\xi\xi,l}\frac{\partial^2}{\partial\xi^2} + \bar{k}^\mu \tilde{K}^{\mu 1}_{\xi,l}\frac{\partial^2}{\partial\xi^2} + \tilde{K}^{01}_l\frac{\partial}{\partial\tau} + \left(\tilde{M}^1_{\tau,l}\frac{\partial}{\partial\tau} + \tilde{M}^1_{\xi,\xi,l}\frac{\partial^2}{\partial\xi^2}\right)\right] U_l^{(1)}$$
$$+ \sum_{l'} ilk^\mu \tilde{K}^{\mu 1}_{l;l'}(U_{l-l'}^{(2)} U_{l'}^{(1)} + U_{l-l'}^{(1)} U_{l'}^{(2)}) + \sum_{l'} [ilk^\mu \tilde{K}^{\mu 2}_{\xi,l;l'}(U_{l-l'}^{(1)}{}_{,\xi} U_{l'}^{(1)} + U_{l-l'}^{(1)} U_{l'}^{(1)}{}_{,\xi}) + \tilde{K}^{\mu 2}_{l;\xi,l'}(U_{l-l'}^{(1)} U_{l'}^{(1)}{}_{,\xi})] + \sum_{l'} \bar{k}^\mu \tilde{K}^{\mu 2}_{l;l'}(U_{l-l'}^{(1)}{}_{,\xi} U_{l'}^{(1)} + U_{l-l'}^{(1)} U_{l'}^{(1)}{}_{,\xi}) + \sum_{l'} \tilde{M}^2_{l;l'}(U_{l-l'}^{(2)} U_{l'}^{(1)} + U_{l-l'}^{(1)} U_{l'}^{(2)})$$
$$+ \sum_{l'} [\tilde{M}^2_{\xi,l;l'}(U_{l-l'}^{(1)}{}_{,\xi} U_{l'}^{(1)} + U_{l-l'}^{(1)} U_{l'}^{(1)}{}_{,\xi}) + \tilde{M}^2_{l;\xi,l'}(U_{l-l'}^{(1)} U_{l'}^{(1)}{}_{,\xi})] + \sum_{l'l''} ilk^\mu \tilde{K}^{\mu 3}_{l;l'+l'';l''} U_{l-l'-l''}^{(1)} U_{l'}^{(1)} U_{l''}^{(1)} + \sum_{l'l''} \tilde{M}^3_{l;l'+l'';l''} U_{l-l'-l''}^{(1)} U_{l'}^{(1)} U_{l''}^{(1)} = 0.$$

(For $U_l^{(1)}{}_{,\xi}$ and $U_{l-l'}^{(1)}{}_{,\xi}$, the subscript ξ denotes partial differentiation.)

By means of Eqs. (4) and (5), it is readily seen that Eq. (13) for $O(\epsilon)$ results in

$$U_l^{(1)} = 0 \quad \text{for } |l| \neq 1,$$
$$U_1^{(1)} = \varphi^{(1)}(\xi,\tau)\mathbf{R}, \tag{14}$$

where $\varphi^{(1)}$ is a scalar function of ξ and τ to be determined later and $\mathbf{R}$ is the right eigenvector given by

$$W_1\mathbf{R} = 0, \tag{15a}$$

with its complex conjugate

$$W_{-1}\mathbf{R}^* = 0. \tag{15b}$$

The corresponding left eigenvector $\mathbf{L}$ and $\mathbf{L}^*$ introduced by

$$\mathbf{L}W_1 = 0, \tag{15c}$$

$$\mathbf{L}^* W_{-1} = 0 \tag{15d}$$

will also be used.

Then, substituting Eq. (14) into the equations for $O(\epsilon^2)$, we have, for $l = 1$,

$$W_1 U_1^{(2)} + (ik^\mu \tilde{K}^{\mu 1}_{\xi,1} + \tilde{M}^1_{\xi,1} + \bar{k}^\mu \tilde{K}^{\mu 1}_1)\mathbf{R}\frac{\partial\varphi^{(1)}}{\partial\xi} = 0. \tag{16}$$

The second term on the left-hand side of this equation can be reduced further. That is, using Eq. (8c), we find

$$\mathbf{k}\frac{\partial}{\partial\mathbf{k}} W_1 = i\bar{k}^\mu \tilde{K}^{\mu 1}_1 + ik^\mu \mathbf{k}\frac{\partial}{\partial\mathbf{k}}\tilde{K}^{\mu 1}_1 + \mathbf{k}\frac{\partial}{\partial\mathbf{k}}\tilde{M}^1_1$$
$$= i\bar{k}^\mu \tilde{K}^{\mu 1}_1 + ik^\mu \int K^{\mu 1}(-i\bar{k}x_1) e^{-ikx_1} d^4x_1 + \int M^1(-i\bar{k}x_1) e^{-ikx_1} d^4x_1$$
$$= i(\bar{k}^\mu \tilde{K}^{\mu 1}_1 + ik^\mu \tilde{K}^{\mu 1}_{\xi,1} + \tilde{M}^1_{\xi,1}).$$

Hence Eq. (16) can be written as

$$iW_1 U_1^{(2)} + \left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{k}} W_1\right)\mathbf{R}\frac{\partial\varphi^{(1)}}{\partial\xi} = 0. \tag{17}$$

Multiplying this equation by $\mathbf{L}$ from left yields the compatibility condition

$$\mathbf{L}\left(\mathbf{k}\frac{\partial}{\partial\mathbf{k}} W_1\right)\mathbf{R}\frac{\partial\varphi^{(1)}}{\partial\xi} = 0,$$

which is satisfied automatically by means of the relation

$$\mathbf{L}\left(\mathbf{k}\frac{\partial}{\partial \mathbf{k}}W_1\right)\mathbf{R} = \mathbf{L}\mathbf{k}\frac{\partial}{\partial \mathbf{k}}(W_1\mathbf{R}) - \mathbf{L}W_1\mathbf{k}\frac{\partial}{\partial \mathbf{k}}\mathbf{R} = 0.$$

Since

$$[\mathbf{k}(\partial/\partial\mathbf{k})W_1]\mathbf{R} = \mathbf{k}(\partial/\partial\mathbf{k})(W_1\mathbf{R}) - W_1\mathbf{k}(\partial/\partial\mathbf{k})\mathbf{R}$$
$$= -W_1\mathbf{k}(\partial/\partial\mathbf{k})\mathbf{R},$$

Eq. (17) reduces finally to

$$W_1\left(U_1^{(2)} - \frac{1}{i}\mathbf{k}\frac{\partial}{\partial\mathbf{k}}\mathbf{R}\frac{\partial\varphi^{(1)}}{\partial\xi}\right) = 0.$$

Therefore we obtain

$$U_1^{(2)} = \varphi^{(2)}\mathbf{R} + \frac{1}{i}\left(\mathbf{k}\frac{\partial}{\partial\mathbf{k}}\mathbf{R}\right)\frac{\partial\varphi^{(1)}}{\partial\xi}, \tag{18}$$

where $\varphi^{(2)}$ is also a scalar function of ξ, τ and will be determined in a higher order approximation.

The other equations for $O(\epsilon^2)$ yield $U_l^{(2)}$ ($|l| \neq 1$) in terms of $\varphi^{(1)}$.

For $l = 0$, we have

$$W_0U_0^{(2)} + \tilde{M}^2_{0;1}U_{-1}^{(1)}U_1^{(1)} + \tilde{M}^2_{0;-1}U_1^{(1)}U_{-1}^{(1)} = 0,$$

which gives

$$U_0^{(2)} = -W_0^{-1}(\tilde{M}^2_{0;1}\mathbf{R}^*\mathbf{R} + \tilde{M}^2_{0;-1}\mathbf{R}\mathbf{R}^*)|\varphi^{(1)}|^2. \tag{19}$$

For $l = 2$, we get

$$W_2U_2^{(2)} + (2ik^\mu\tilde{K}^{\mu\,2}_{2;1} + \tilde{M}^2_{2;1})U_1^{(1)}U_1^{(1)} = 0,$$

i.e., $$U_2^{(2)} = -W_2^{-1}(2ik^\mu\tilde{K}^{\mu\,2}_{2;1} + \tilde{M}^2_{2;1})\mathbf{R}\mathbf{R}\varphi^2, \tag{20}$$

where $\varphi^{(1)}$ is written as φ, and likewise for $l = -2$

$$U_{-2}^{(2)} = -W_{-2}^{-1}(-2ik^\mu\tilde{K}^{\mu\,2}_{-2;-1} + \tilde{M}^2_{-2;-1})\mathbf{R}^*\mathbf{R}^*\varphi^{*2}, \tag{21}$$

while, for $|l| > 3$, $U_l^{(2)}$'s vanish.

Introducing these results into the equation of $O(\epsilon^3)$ for $l = 1$, we obtain

$$W_1U_1^{(3)} + \frac{1}{i}\left(\mathbf{k}\frac{\partial}{\partial\mathbf{k}}W_1\right)\frac{\partial}{\partial\xi}U_1^{(2)} + (ik^\mu\tilde{K}^{\mu\,1}_{\tau,1} + \tilde{K}^{01}_1 + \tilde{M}^1_{\tau,1})\frac{\partial}{\partial\tau}U_1^{(1)}$$
$$+ (ik^\mu\tilde{K}^{\mu\,1}_{\xi\xi,1} + \bar{k}^\mu\tilde{K}^{\mu\,1}_{\xi,1} + \tilde{M}^1_{\xi\xi,1})\frac{\partial^2}{\partial\xi^2}U_1^{(1)}$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;1} + \tilde{M}^2_{1;1})U_0^{(2)}U_1^{(1)}$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;0} + \tilde{M}^2_{1;0})U_1^{(1)}U_0^{(2)}$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;-1} + \tilde{M}^2_{1;-1})U_2^{(2)}U_{-1}^{(1)}$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;2} + \tilde{M}^2_{1;2})U_{-1}^{(1)}U_2^{(2)}$$
$$+ \sum_{l'l''}(ik^\mu\tilde{K}^{\mu\,3}_{1;l';l''} + \tilde{M}^3_{1;l';l''})U_{l'}^{(1)}U_{l-l'}^{(1)}U_{l''}^{(1)} = 0; \tag{22}$$

by means of Eq. (14), the nonvanishing terms in the summation of the last term are those for ($l' = 0, l'' = \pm 1$) and ($l' = 2, l'' = 1$).

Multiplying Eq. (22) by $\mathbf{L}$ from left, we obtain the equation for φ (see Appendix B),

$$i\frac{\partial\varphi}{\partial\tau} + P\frac{\partial^2\varphi}{\partial\xi^2} + Q|\varphi|^2\varphi = 0. \tag{23}$$

Here, P and Q are given by

$$P = -\frac{k_jk_m}{2}\frac{\partial^2k^0}{\partial k_j\partial k_m}, \tag{24a}$$

$$Q = \mathbf{L}[(ik^\mu\tilde{K}^{\mu\,2}_{1;1} + \tilde{M}^2_{1;1})W_0^{-1}(\tilde{M}^2_{0;1}\mathbf{R}^*\mathbf{R}\mathbf{R} + \tilde{M}^2_{0;-1}\mathbf{R}\mathbf{R}^*\mathbf{R})$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;0} + \tilde{M}^2_{1;0})\mathbf{R}W_0^{-1}(\tilde{M}^2_{0;1}\mathbf{R}^*\mathbf{R} + \tilde{M}^2_{0;-1}\mathbf{R}\mathbf{R}^*)$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;-1} + \tilde{M}^2_{1;-1})W_2^{-1}(2ik^\mu\tilde{K}^{\mu\,2}_{2;1} + \tilde{M}^2_{2;1})\mathbf{R}\mathbf{R}\mathbf{R}^*$$
$$+ (ik^\mu\tilde{K}^{\mu\,2}_{1;2} + \tilde{M}^2_{1;2})\mathbf{R}^*W_2^{-1}(2ik^\mu\tilde{K}^{\mu\,2}_{2;1} + \tilde{M}^2_{2;1})\mathbf{R}\mathbf{R}$$
$$- (ik^\mu\tilde{K}^{\mu\,3}_{1;0;1} + \tilde{M}^3_{1;0;1})\mathbf{R}\mathbf{R}^*\mathbf{R}$$
$$- (ik^\mu\tilde{K}^{\mu\,3}_{1;0;-1} + \tilde{M}^3_{1;0;-1})\mathbf{R}\mathbf{R}\mathbf{R}^*$$
$$- (ik^\mu\tilde{K}^{\mu\,3}_{1;2;1} + \tilde{M}^3_{1;2;1})\mathbf{R}^*\mathbf{R}\mathbf{R}]\Big/\left(\mathbf{L}\frac{\partial W_1}{\partial k^0}\mathbf{R}\right). \tag{24b}$$

4. APPLICATION TO NONLINEAR OPTICS IN LOSSLESS MEDIA

Consider the Maxwell equations in Gaussian units,

$$\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} + \nabla\times\mathbf{E} = 0, \tag{25a}$$

$$\frac{1}{c}\frac{\partial\mathbf{D}}{\partial t} - \nabla\times\mathbf{H} = 0, \tag{25b}$$

and assume that the electric displacement $\mathbf{D}$ is related to the electric field vector $\mathbf{E}$ by the equation[13,15]

$$D_i(x,t) = \int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}d\mathbf{x}_1\epsilon_{ij}(\mathbf{x}_1,t_1)\mathbf{E}_j(\mathbf{x}-\mathbf{x}_1,t-t_1)$$
$$+ 4\pi\int_{-\infty}^{\infty}\iiint dt_1dt_2dt_3\int_{-\infty}^{\infty}\iiint d\mathbf{x}_1d\mathbf{x}_2d\mathbf{x}_3$$
$$\times\mathbf{X}_{ijkl}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3;t_1,t_2,t_3)$$
$$\times E_j(\mathbf{x}-\mathbf{x}_1,t-t_1)E_k(\mathbf{x}-\mathbf{x}_1-\mathbf{x}_2,t-t_1-t_2)$$
$$\times E_l(\mathbf{x}-\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3,t-t_1-t_2-t_3), \tag{26a}$$

while the magnetic inductance $\mathbf{B}$ is related to the magnetic field vector $\mathbf{H}$ by

$$\mathbf{B} = \mu_0\mathbf{H}. \tag{26b}$$

Here, $\epsilon_{ij}(\mathbf{x},t)$, $\mathbf{X}_{ijkl}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3;t_1,t_2,t_3)$ are components of tensors, and μ_0 is the magnetic permeability, assumed to be a scalar constant. For simplicity, we shall assume further that ϵ_{ij} is symmetric, i.e., $\epsilon_{ij} = \epsilon_{ji}$. Introducing F^μ by the equations

$$F^0 = \begin{pmatrix}(\mu_0/c)D_1\\(\mu_0/c)D_2\\(\mu_0/c)D_3\\(1/c)B_1\\(1/c)B_2\\(1/c)B_3\end{pmatrix},\quad F^1 = \begin{pmatrix}0\\B_3\\-B_2\\0\\-E_3\\E_2\end{pmatrix},\quad F^2 = \begin{pmatrix}-B_3\\0\\B_1\\E_3\\0\\-E_1\end{pmatrix},\quad F_3 = \begin{pmatrix}B_2\\-B_1\\0\\-E_2\\E_1\\0\end{pmatrix}, \tag{27}$$

one can rewrite Eqs. (25) in the form of Eq. (1) (with $G = 0$),

$$\frac{\partial}{\partial x^\mu} F^\mu = 0. \tag{28}$$

The equations

$$\nabla \cdot \mathbf{D} = \nabla \cdot \mathbf{B} = 0 \tag{25c}$$

are imposed as the subsidiary conditions: It is easy to prove that they perpetuate if they are valid initially.

Let the column vector U be given by

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_6 \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ B_1 \\ B_2 \\ B_3 \end{pmatrix}. \tag{29}$$

Then the tensors $K^{\mu r}$'s introduced by Eq. (2) take the forms

$$K^{01}_{ij}(x) = \left(\begin{array}{c|c} (\mu_0/c)\epsilon_{ij}(x) & 0 \\ \hline 0 & \delta(x)/c \end{array}\right),$$

$$K^{21}_{ij}(x) = \left(\begin{array}{ccc|ccc} & & & 0 & 0 & -1 \\ & 0 & & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & & & \\ 0 & 0 & 0 & & 0 & \\ -1 & 0 & 0 & & & \end{array}\right) \delta(x),$$

$$K^{11}_{ij}(x) = \left(\begin{array}{ccc|ccc} & & & 0 & 0 & 0 \\ & 0 & & 0 & 0 & 1 \\ & & & 0 & -1 & 0 \\ \hline 0 & 0 & 0 & & & \\ 0 & 0 & -1 & & 0 & \\ 0 & 1 & 0 & & & \end{array}\right) \delta(x),$$

$$K^{31}_{ij}(x) = \left(\begin{array}{ccc|ccc} & & & 0 & 1 & 0 \\ & 0 & & -1 & 0 & 0 \\ & & & 0 & 0 & 0 \\ \hline 0 & -1 & 0 & & & \\ 1 & 0 & 0 & & 0 & \\ 0 & 0 & 0 & & & \end{array}\right) \delta(x), \tag{30}$$

$$K^{\mu}_{ijk}{}^{3} = (4\pi\mu_0/c)\mathrm{X}_{ijkl} \quad \text{for } \mu = 0, 1 \le i,j,k,l \le 3,$$
$$= 0 \quad \text{otherwise}, \tag{31}$$

whereas if r is not equal to 1 or 3, the $K^{\mu r}$'s vanish,

$$K^{\mu r} = 0 \quad \text{for } r \neq 1, r \neq 3.$$

Because of the symmetry of ϵ_{ij}, we have

$$K^{\mu 1}_{ij} = K^{\mu 1}_{ji}. \tag{30'}$$

Hereafter, we consider an isotropic medium so that

$$\epsilon_{ij} = \epsilon^d \delta_{ij}, \qquad \epsilon^d(x_1, t_1) = \epsilon^d(-x_1, -t_1). \tag{32}$$

Then $W_1 = ik^\mu \tilde{K}^{\mu 1}_1$ is represented by

$$W_1 = i \begin{pmatrix} \mu_0\tilde{\epsilon}^d_1 k^0/c & 0 & 0 & 0 \\ 0 & \mu_0\tilde{\epsilon}^d_1 k^0/c & 0 & -k_3 \\ 0 & 0 & \mu_0\tilde{\epsilon}^d_1 k^0/c & k_2 \\ 0 & -k_3 & k_2 & k^0/c \\ k_3 & 0 & -k_1 & 0 \\ -k_2 & k_1 & 0 & 0 \end{pmatrix} \begin{pmatrix} k_3 & -k_2 \\ 0 & k_1 \\ -k_1 & 0 \\ 0 & 0 \\ k^0/c & 0 \\ 0 & k^0/c \end{pmatrix}, \tag{33}$$

where $\tilde{\epsilon}^d_1 = \int \epsilon^d \exp(ikx)\,(dx)^4$, and

$$\det W_1 = -\left(\frac{\mu_0\tilde{\epsilon}^d_1}{c^2}(k^0)^2 - \mathbf{k}^2\right)^2 \frac{\mu_0\tilde{\epsilon}^d_1}{c^2}(k^0)^2. \tag{33'}$$

Consequently, $\det W_1 = 0$ yields

$$(\mu_0\tilde{\epsilon}^d_1/c^2)\,(k^0)^2 = k^2. \tag{34}$$

Isotropy implies that $\tilde{\epsilon}^d_1$ depend $\mathbf{k}^2$ and k^0. In particular, without temporal dispersion, $\tilde{\epsilon}^d_1$ is a function of $|\mathbf{k}|$ only; hence in this case Eq. (34) can be solved explicitly to give

$$(k^0)^2 = \mathbf{k}^2 c^2/[\mu_0\tilde{\epsilon}^d_1(|\mathbf{k}|)]. \tag{34'}$$

In the remainder of this section, a lossless isotropic medium will be considered, and the temporal dispersion may not be neglected. Then, without loss of generality we may assume that $\tilde{\epsilon}^d_1 > 0$ so that Eq. (34) gives at least one real k^0 for any real $\mathbf{k}$, and in addition that $\tilde{\epsilon}^d_1$ is not constant but depends on $\mathbf{k}$ and k^0 to ensure the condition

$$\det W_l \neq 0 \quad \text{for } |l| \ge 2.$$

(In this regard, linear dispersion is essential for the application of our method of solution.)

Since G, and consequently $M^{\mu r}$, vanish identically, W_0 also vanishes identically; hence the condition (5), $\det W_l \neq 0$, is violated for $l = 0$. (As will be seen later,

however, this causes no trouble because of the cubic nonlinearity, $K^{\mu 2} = 0$). In an isotropic medium, we may assume without loss of generality that $k_1 = k_2 = 0$; hereafter, for simplicity, we put $|\mathbf{k}| = k_3 > 0$. Then, the eigenvectors for the eigenmode (34) can be obtained easily. As is shown in Eq. (33′), the state is doubly degenerate, and there exist two linearly independent eigenvectors,

$$\mathbf{R}_1 = \begin{pmatrix} k_3 \\ 0 \\ 0 \\ 0 \\ -\mu_0 \tilde{\epsilon}_1^d k^0/c \\ 0 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} 0 \\ k_3 \\ 0 \\ \mu_0 \tilde{\epsilon}_1^d k^0/c \\ 0 \\ 0 \end{pmatrix}. \tag{35}$$

Since W_1 is symmetric, the corresponding left eigenvectors are given by

$$\mathbf{L}_1 = \mathbf{R}_1^T, \quad \mathbf{L}_2 = \mathbf{R}_2^T, \tag{35′}$$

where the superscript T of $\mathbf{R}_1$ and $\mathbf{R}_2$ designate the transpose. {Note that the degeneracy corresponds to that of $[\mu_0 \tilde{\epsilon}_1^d k^{0 2}/c^2 - k^2]^2$ in Eq. (33′) and does not correspond to that of Eq. (34), which is in general a transcendental equation for k^0 and may admit an infinite number of roots. Physically, it represents the sense of polarization.}

From (35) and (35′), we have

$$(\mathbf{L}_1 \cdot \mathbf{R}_1) = (\mathbf{L}_2 \cdot \mathbf{R}_2) = k_3^2 + (\mu_0^2 \tilde{\epsilon}_1^2/c^2) k^{0 2} = (1 + \mu_0 \tilde{\epsilon}_1^d) k_3^2,$$
$$(\mathbf{L}_1 \cdot \mathbf{R}_2) = (\mathbf{L}_2 \cdot \mathbf{R}_1) = 0, \tag{36}$$

and

$$\left(\mathbf{L}_1 \frac{\partial W_1}{\partial k^0} \mathbf{R}_1\right) = \left(\mathbf{L}_2 \frac{\partial W_1}{\partial k^0} \mathbf{R}_2\right) = i \frac{\mu_0 \tilde{\epsilon}_1^d}{c} \left(2 + \frac{k^0}{\tilde{\epsilon}_1^d} \frac{\partial \tilde{\epsilon}_1^d}{\partial k^0}\right) k_3^2, \tag{37}$$

in which $\partial W_1/\partial k^0$ is partial differentiation with k_3 kept constant.

We now evaluate the nonlinear effect. Since an isotropic medium is being considered, it is consistent to assume for $\mathbf{X}_{ijkl}$ that

$$\tilde{\mathbf{X}}_{ijkl}(l_1, l_2, l_3) = \tilde{\mathbf{X}}_0(l_1, l_2, l_3)(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{kj}) \quad (1 \le i, j, k, l \le 3). \tag{38}$$

Then noting that $\mathbf{R}^* = \mathbf{R}$, and omitting the subscripts of $\mathbf{L}$ and $\mathbf{R}$, one has

$$\mathbf{L}(ik^\mu \tilde{K}^{\mu 3}_{1;0;1} \mathbf{RR^*R} + ik^\mu \tilde{K}^{\mu 3}_{1;0;-1} \mathbf{RRR^*} + ik^\mu \tilde{K}^{\mu 3}_{1;2;1} \mathbf{R^*RR}) = 12\pi i k^0 (\mu_0/c)[\tilde{\mathbf{X}}_0(1,0,1) + \tilde{\mathbf{X}}_0(1,0,-1) + \tilde{\mathbf{X}}_0(1,2,1)](L'R')^2, \tag{39}$$

in which L' and R' are defined by the equations,

$$(\mathbf{L}')_i = \begin{cases} (\mathbf{L})_i & 1 \le i \le 3 \\ 0 & 4 \le i \le 6 \end{cases}, \qquad (\mathbf{R}')_i = \begin{cases} (R)_i & 1 \le i \le 3 \\ 0 & 4 \le i \le 6 \end{cases}; \tag{39′}$$

consequently from Eqs. (35), (35′) it follows that

$$\mathbf{L}' \cdot \mathbf{R}' = k_3^2. \tag{39″}$$

Using Eqs. (37), (38), (39), and (39″), we have

$$\mathbf{L}(ik^\mu \tilde{K}^{\mu 3}_{1;0,1} \mathbf{RR^*R} + ik^\mu \tilde{K}^{\mu 3}_{1;0,-1} \mathbf{RRR^*} + ik^\mu \tilde{K}^{\mu 3}_{1;2;1} \mathbf{R^*RR}) \Big/ \left(\mathbf{L} \frac{\partial W_1}{\partial k^0} \mathbf{R}\right) = \frac{12\pi}{\tilde{\epsilon}_1^d} \frac{k^0 k_3^2}{(2 + k^0 (\partial \tilde{\epsilon}_1^d/\partial k^0) \tilde{\epsilon}_1^d)} [\tilde{\mathbf{X}}_0]. \tag{40}$$

where $[\tilde{\mathbf{X}}_0]$ stands for $\tilde{\mathbf{X}}_0(1,0,1) + \tilde{\mathbf{X}}_0(1,0,-1) + \tilde{\mathbf{X}}_0(1,2,1)$. On the other hand, by means of the cubic nonlinearity, the $K^{\mu 2}$ vanish and hence, in Eq. (24a), the terms containing W_0^{-1} and W_2^{-1} also vanish. (We thus find that $\det W_0 = 0$ does not cause trouble.) Therefore, it is readily seen that Q is given by Eq. (40), and we obtain

$$i \frac{\partial \varphi}{\partial \tau} - \frac{k_3^2}{2} \frac{\partial^2 k^0}{\partial k_3^2} \frac{\partial^2 \varphi}{\partial \xi^2} - \frac{12\pi}{\tilde{\epsilon}_1^d} \frac{k^0 k_3^2}{(2 + k^0 (\partial \tilde{\epsilon}_1^d/\partial k^0) \tilde{\epsilon}_1^d)} \times [\tilde{\mathbf{X}}_0] |\varphi|^2 \varphi = 0. \tag{41}$$

If the temporal dispersion and the dispersion in the nonlinear polarization are neglected (i.e., $[\tilde{\mathbf{X}}_0] = 3\mathbf{X}$), Eq. (41) reduces to

$$i \frac{\partial \psi}{\partial \tau} + \frac{1}{2} \frac{\partial^2 \omega}{\partial k^2} \frac{\partial^2 \psi}{\partial \xi^2} + \left(18\pi \frac{\omega}{\tilde{\epsilon}}\right) \mathbf{X} |\psi|^2 \psi = 0, \tag{41′}$$

in which $k_3, k^0, \tilde{\epsilon}_1^d$ are denoted by $k, -\omega, \tilde{\epsilon}$, respectively, and $k_3 \varphi = E_{11}^{(1)}$ by ψ, while ξ denotes $\epsilon[x_3 - (\partial\omega/\partial k)t]$. Equation (41′) is identical with that derived by Karpman's[3] method, in which the dispersion relation is given by

$$\omega^2(\tilde{\epsilon} + 36\pi \mathbf{X} |\psi|^2) - (c^2 k^2/\mu_0) = 0.$$

Here we note that under the condition $|W_0| \equiv 0$, even if the nonlinear Schrödinger equation is derived as above without ambiguity, special considerations are required to determine $U_0^{(\alpha)}$. Since we are considering the nonlinear modulation of the wave $U_1^{(1)} e^{ikx}$ + complex conjugate, $U_0^{(1)}$ may be equated to zero as a physical requirement. In order to determine $U_0^{(2)}$, we must use the equation of third order for $l = 0$, which, in this case, becomes

$$\bar{k}^\mu \tilde{K}_0^{\mu 1} \frac{\partial U_0^{(2)}}{\partial \xi} = 0, \tag{42}$$

and hence yields

$$\frac{\partial U_0^{(2)}}{\partial \xi} = 0, \tag{42′}$$

provided that $\det |\bar{k}^\mu \tilde{K}_0^{\mu 1}|$ does not vanish,

$$[\mu_0 \tilde{\epsilon}_0^d (\bar{k}^0/c)^2 - \mathbf{k}^2] \mu_0 \tilde{\epsilon}_0^d (\bar{k}^0/c)^2 \neq 0,$$

that is,

$$\tilde{\epsilon}_0^d \neq 0 \quad \text{and} \quad \bar{k}^{0 2} \neq c^2 k^2/\mu_0 \tilde{\epsilon}_0^d. \tag{43}$$

Consequently, when the condition (43) is valid and $U_0^{(2)}$ vanishes at a boundary, say $x \to \infty$, Eq. (42′) gives $U_0^{(2)} =$

0 for all time. In higher orders, the equation of the order of $\epsilon^{\alpha+1}$ for $l = 0$ will allow us to determine $U_0^{(\alpha)}$.

It may be of interest to note that the latter condition of (43) implies that if $(d\omega/dk)^2 - c^2/\mu_0\tilde{\epsilon}_0^d \approx O(\epsilon)$, we might encounter a different case in which a slow (d.c.) mode coexists with the carrier wave.

We also remark that the coefficient Q of the nonlinear term of Eq. (41) becomes infinite if $2 + k^0(\partial\tilde{\epsilon}_1^d/\partial k^0)/\tilde{\epsilon}_1^d = 0$, namely, that if there exists a wavenumber k which makes $2 + \omega(\partial\tilde{\epsilon}_1^d/\partial\omega)/\tilde{\epsilon}_1^d \approx O(\epsilon)$, the present asymptotic expansion is not valid at this wavenumber and a markedly different solution may be expected. [Note that this singularity corresponds to $\mathbf{L}(\partial W_1/\partial k^0)\mathbf{R} = 0$].

Finally, we mention that if $K^{\mu 2}$ does not vanish, in other words, if a quadratic nonlinear polarization exists, Q depends on $U_0^{(2)}$. [Note that the terms with W_0^{-1} in Eq. (24b) must be given in terms of $U_0^{(2)}$.] On the other hand, $U_0^{(2)}$ is determined by an equation corresponding to Eq. (44) which, however, contains extra terms dependent on $K^{\mu 2}$. Therefore, in general, we shall have to consider a coupled system of equations for φ, φ^*, and $U_0^{(2)}$.

APPENDIX A: DERIVATION OF EQS. (9), (10), AND (11)

$$K^{\mu 1} * U = \sum_\alpha \sum_l \epsilon^\alpha K^{\mu 1} * U_l^{(\alpha)} \exp(ilk\cdot x)$$

$$= \sum_\alpha \sum_l \epsilon^\alpha \int K^{\mu 1}(x_1) U_l^{(\alpha)}(\tau - \tau_1, \xi - \xi_1)$$

$$\exp[ilk\cdot(x - x_1)](dx_1)^4 = \sum_\alpha \sum_l \epsilon^\alpha \int K^{\mu 1}(x_1)$$

$$\times \left[U_l^{(\alpha)}(\tau,\xi) - \tau_1 \frac{\partial U_l^{(\alpha)}(\tau,\xi)}{\partial\tau} - \xi_1 \frac{\partial U_l^{(\alpha)}}{\partial\xi}\right.$$

$$+ \frac{1}{2!}\left(\frac{\partial^2 U_l^{(\alpha)}}{\partial\tau^2}\tau_1^2 + 2\frac{\partial^2 U_l^{(\alpha)}}{\partial\xi\partial\tau}\tau_1\xi_1 + \frac{\partial^2 U_l^{(\alpha)}}{\partial\xi^2}\xi_1^2\right)$$

$$\left. + \cdots + \frac{(-1)^m}{m!}\left(\frac{\partial}{\partial\tau}\tau_1 + \frac{\partial}{\partial\xi}\xi_1\right)^m U_l^{(\alpha)}(\tau,\xi) + \cdots\right]$$

$$\times \exp[ilk(x - x_1)](dx_1)^4$$

$$= \sum_\alpha \sum_l \epsilon^\alpha \int K^{\mu 1}(x_1)\left[1 + \left(\epsilon^2(-t_1)\frac{\partial}{\partial\tau} + \epsilon(-\bar{k}^\nu x_1^\nu)\frac{\partial}{\partial\xi}\right)\right.$$

$$\left. + \frac{1}{2}\left(\epsilon^4 t_1^2\frac{\partial^2}{\partial\tau^2} + 2\epsilon^3 t_1(\bar{k}^\nu x_1^\nu)\frac{\partial^2}{\partial\xi\partial\tau} + \epsilon^2(\bar{k}^\nu x_1^\nu)^2\frac{\partial^2}{\partial\xi^2}\right) + \cdots\right]$$

$$\times U_l^{(\alpha)}(\xi,\tau)\exp[ilk\cdot(x - x_1)](dx_1)^4$$

$$= \sum_{\alpha=1}\sum_l \epsilon^\alpha \left\{\tilde{K}_l^{\mu 1} + \left[\epsilon\tilde{K}_{\xi,l}^{\mu 1}\frac{\partial}{\partial\xi} + \epsilon^2\left(\tilde{K}_{\tau,l}^{\mu 1} + K_{\xi\xi,l}^{\mu 1}\frac{\partial^2}{\partial\xi^2}\right)\right]\right.$$

$$\left. + \cdots\right\} U_l^{(\alpha)}(\xi,\tau)\exp(ilk\cdot x),$$

where $\tilde{K}_{\xi,l}^{\mu 1}$, etc. are defined by

$$\tilde{K}_{\xi,l}^{\mu 1} = \int(-\bar{k}x_1)K^{\mu 1}\exp(-ilk\cdot x_1)(dx_1)^4,$$

$$\tilde{K}_{\xi\xi,l}^{\mu 1} = \frac{1}{2!}\int(\bar{k}x_1)^2 K^{\mu 1}\exp(-ilk\cdot x_1)(dx_1)^4,$$

$$\tilde{K}_{\tau,l}^{\mu 1} = \int(-t_1)K^{\mu 1}\exp(-ilk\cdot x_1)(dx_1)^4,$$

$$\tilde{K}_{\tau\xi,l}^{\mu 1} = \int t_1(\bar{k}^\nu x_1^\nu)K^{\mu 1}\exp(-ilk\cdot x_1)(dx_1)^4.$$

Similar calculations yield Eqs. (10) and (11) with

$$\tilde{K}_{l+l';l'}^{\mu 2} = \iint K^{\mu 2}(x_1,x_2)$$

$$\times \exp[-i(l + l')k\cdot x_1 - il'k\cdot x_2](dx_1)^4(dx_2)^4,$$

$$\tilde{K}_{\xi,l+l';l'}^{\mu 2} = \iint(-\bar{k}x_1)K^{\mu 2}(x_1,x_2)$$

$$\times \exp[-i(l + l')k\cdot x_1 - il'k\cdot x_2](dx_1)^4(dx_2)^4,$$

$$\tilde{K}_{l+l';\xi,l'}^{\mu 2} = \iint(-\bar{k}x_2)K^{\mu 2}(x_1,x_2)$$

$$\times \exp[-i(l + l')k\cdot x_1 - il'k\cdot x_2](dx_1)^4(dx_2)^4\cdots,$$

$$\tilde{K}_{l+l'+l'';l'+l'';l''}^{\mu 3} = \iiint K^{\mu 3}(x_1,x_2,x_3)$$

$$\times \exp\{-i[(l + l' + l'')k\cdot x_1$$

$$+ (l' + l'')k\cdot x_2 + l''k\cdot x_3]\}(dx_1)^4(dx_2)^4(dx_3)^4.$$

APPENDIX B: DERIVATION OF EQS. (23), (24a), AND (24b)

Noting that

$$\tilde{K}_{\tau,l}^{\mu 1} = -i\frac{\partial}{\partial k^0}\tilde{K}_1^{\mu 1},$$

$$\tilde{M}_{\tau,1}^{1} = -i\frac{\partial}{\partial k^0}\tilde{M}_1^{1},$$

$$\tilde{K}_1^{01} = -i\left(\frac{\partial}{\partial k^0}ik^\mu\right)\tilde{K}_1^{\mu 1},$$

one can rewrite the coefficient of $\partial U_1^{(1)}/\partial\tau$ in Eq. (22) as follows:

$$ik^\mu\tilde{K}_{\tau,1}^{\mu 1} + \tilde{K}_1^{01} + \tilde{M}_{\tau,1}^{1} = -i\frac{\partial}{\partial k^0}(ik^\mu\tilde{K}_1^{\mu 1} + \tilde{M}_1^{1})$$

$$= -i\frac{\partial}{\partial k^0}W_1,$$

where $\partial/\partial k^0$ is differentiation keeping $\mathbf{k}$ constant. Similarly, using the relations

$$\tilde{K}_{\xi\xi,1}^{\mu 1} = \tfrac{1}{2}\left(-k_j k_m\frac{\partial^2}{\partial k_j\partial k_m} + k_j k_m\frac{\partial^2 k^0}{\partial k_j\partial k_m}\frac{\partial}{\partial k^0}\right)\tilde{K}_1^{\mu 1},$$

$$\tilde{M}_{\xi\xi,1}^{1} = \tfrac{1}{2}\left(-k_j k_m\frac{\partial^2}{\partial k_j\partial k_m} + k_j k_m\frac{\partial^2 k^0}{\partial k_j\partial k_m}\cdot\frac{\partial}{\partial k^0}\right)\tilde{M}_1^{1},$$

$$\bar{k}^\mu\tilde{K}_{\xi,1}^{\mu 1} = -ik_j k_m\left(\frac{\partial}{\partial k_j}k^\mu\right)\frac{\partial}{\partial k_m}\tilde{K}_1^{\mu 1},$$

we find that the coefficient of $\partial^2 U_1^{(1)}/\partial\xi^2$ can be rewritten as

$$(ik^\mu\tilde{K}_{\xi\xi,1}^{\mu 1} + \bar{k}^\mu\tilde{K}_{\xi,1}^{\mu 1} + \tilde{M}_{\xi\xi,1}^{1}) =$$

$$-\tfrac{1}{2}k_j k_m\frac{\partial^2}{\partial k_j\partial k_m}W_1 + \tfrac{1}{2}k_j k_m\frac{\partial^2 k^0}{\partial k_j\partial k_m}\frac{\partial}{\partial k^0}W_1.$$

Multiplying Eq. (22) by $\mathbf{L}$ from left and using the relations

$$\tfrac{1}{2}\mathbf{L}k_j k_m\left(\frac{\partial^2}{\partial k_j\partial k_m}\right)W_1\ \mathbf{R} = \tfrac{1}{2}k_j k_m\left[\mathbf{L}\frac{\partial}{\partial k_j}\left(\frac{\partial W_1}{\partial k_m}\mathbf{R}\right)\right.$$

$$\left. - \mathbf{L}\frac{\partial W_1}{\partial k_m}\frac{\partial\mathbf{R}}{\partial k_j}\right]$$

$$= \tfrac{1}{2} k_j k_m \left[\frac{\partial}{\partial k_j} \left(\mathbf{L} \frac{\partial W_1}{\partial k_m} \mathbf{R} \right) - \frac{\partial \mathbf{L}}{\partial k_j} \frac{\partial W_1}{\partial k_m} \mathbf{R} - \mathbf{L} \frac{\partial W_1}{\partial k_m} \frac{\partial \mathbf{R}}{\partial k_j} \right]$$

$$= \tfrac{1}{2} k_j k_m \left(- \frac{\partial \mathbf{L}}{\partial k_j} \frac{\partial}{\partial k_m} (W_1 \mathbf{R}) + \frac{\partial \mathbf{L}}{\partial k_j} W_1 \frac{\partial \mathbf{R}}{\partial k_m} - \frac{\partial}{\partial k_m} (\mathbf{L} W_1) \frac{\partial \mathbf{R}}{\partial k_j} + \frac{\partial \mathbf{L}}{\partial k_m} W_1 \frac{\partial \mathbf{R}}{\partial k_j} \right)$$

$$= \mathbf{k} \frac{\partial \mathbf{L}}{\partial \mathbf{k}} \cdot W_1 \cdot \mathbf{k} \frac{\partial \mathbf{R}}{\partial \mathbf{k}},$$

$$= - \mathbf{L} \mathbf{k} \frac{\partial}{\partial \mathbf{k}} W_1 \cdot \mathbf{k} \frac{\partial}{\partial \mathbf{k}} \mathbf{R},$$

we can obtain Eqs. (23) and (24).

[1] T. Taniuti and H. Washimi, Phys. Rev. Lett. **21,** 209 (1968).
[2] T. Taniuti and N. Yajima, J. Math. Phys. **10,** 1369 (1969).
[3] V. I. Karpman and E. M. Krushkal, Zh. Eksp. Teor. Fiz. **55,** 530 (1968) [Sov. Phys.-JETP **28,** 277 (1969)].
[4] N. Yajima and A. Outi, Prog. Theor. Phys. **45,** 1997 (1971).
[5] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61,** 118 (1971) [Sov. Phys.-JETP **34,** 62 (1972)].
[6] T. Taniuti and C. C. Wei, J. Phys. Soc. Jap. **24,** 941 (1968).
[7] N. Asano and T. Taniuti, J. Phys. Soc. Jap. **27,** 1059 (1969).
[8] K. Ikuta and T. Taniuti, Report of the Institute of Plasma Physics, Nagoya University, IPPJ-31, 1964 (unpublished); A. Jeffrey and T. Taniuti, *Nonlinear Wave Propagation* (Academic, New York, 1964), §2.1.
[9] G. Rosen, Phys. Rev. **139,** A539 (1965).
[10] L. A. Ostrovskii, Zh. Eksp. Teor. Fiz. **51,** 1189 (1966) [Sov. Phys.-JETP **24,** 797 (1967)].
[11] N. Asano, J. Phys. Soc. Japan (to be published).
[12] F. De Martini, C. H. Townes, T, K. Gustafson, and P. L. Kelly, Phys. Rev. **164,** 312 (1967).
[13] S. A. Akhmanov, A. P. Sukhorukov, and R. V. Khokhlov, Usp. Fiz. Nauk **93,** 19 (1967) [Sov. Phys.-Usp. **10,** 609 (1968)].
[14] N. Asano (private communication).
[15] The usual dielectric tensors are given by ϵ_{ij} and X_{ijk} multiplied by the step function $\theta(t)$: $\theta(t) = 1$ for $t > 0$, $\theta(t) = 0$ for $t < 0$.

Modified Lippmann–Schwinger equations for two-body scattering theory with long-range interactions

E. Prugovečki* and J. Zorbas

Department of Mathematics, University of Toronto, Toronto M5S 1A1, Canada
(Received 6 March 1973)

Two kinds of modified Lippmann–Schwinger equations are derived for the case of long-range potentials. The equations of the first kind are homogeneous and are a direct result of the fact that the standard Lippmann–Schwinger equations do not hold when long-range forces are present. The equations of the second kind depend on the existence of an operator Z such that $W_\pm = \text{s-lim}\ \exp(iHt)Z\exp(-iH_0t)$. A general recipe for constructing Z is given and its computation is carried through for the case of asymptotically Coulombic potentials. The resulting equations are used to compare the long-range theory with the theory with a space cutoff (i.e., screened potential) in the limit in which that cutoff is being removed.

1. INTRODUCTION

When two particles are interacting via long-range potentials, such as the Coulomb potential or potentials decreasing at infinity slower than the Coulomb potential, the "strong" wave operators of time-dependent theory

$$W_\pm^{(s)} = \underset{t\to\pm\infty}{\text{s-lim}}\ \exp(iHt)\exp(-iH_0t) \tag{1.1}$$

do not exist. It has been recently observed[1-4] that a time-dependent scattering theory can be formulated if based on a more general definition of wave operators, namely the "renormalized" wave operators

$$W_\pm^{(r)} = \underset{t\to\pm\infty}{\text{s-lim}}\ \exp(iHt)\exp[-iG(H_0;t)]\exp(-iH_0t) \tag{1.2}$$

for adequate choices of the self-adjoint function $G(H_0;t)$. The introduction of $W_\pm^{(r)}$ in (1.2) can be physically justified[4] by showing that the asymptotic conditions

$$\lim_{t\to\pm\infty} |\langle \exp(-iH_0t)\Psi \mid A\exp(-iH_0t)\Psi\rangle - \langle \exp(-iHt)W_\pm^{(r)}\Psi \mid A\exp(-iHt)W_\pm^{(r)}\Psi\rangle| = 0 \tag{1.3}$$

are satisfied for each bounded observable A.

Stationary scattering theory for short range potentials can be based[5] on the expression (1.1) which leads to the following time-independent representations for the strong wave operators:

$$W_\pm^{(s)} = \underset{\epsilon\to+0}{\text{s-lim}} \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H-\lambda'\pm i\epsilon}\, d_{\lambda'}E_\lambda^{H_0} = \underset{\epsilon\to+0}{\text{s-lim}} \int_{-\infty}^{+\infty} d_\lambda E_\lambda^H \frac{\pm i\epsilon}{\lambda - H_0 \pm i\epsilon}, \tag{1.4}$$

where $E_\lambda^{H_0}$ and E_λ^H are the spectral functions of H_0 and H, respectively. The above representations of the strong wave operators in terms of strong limits of the above Riemann-Stieltjes integrals[5-6] allows one to derive Hilbert-space versions of the Lippmann-Schwinger equations and T operator[5]:

$$\Psi_\pm = \Psi + \underset{\epsilon\to+0}{\text{s-lim}} \int_{-\infty}^{+\infty} \frac{1}{\lambda - H_0 \pm i\epsilon}(\lambda - H_0)d_\lambda E_\lambda^H \Psi_\pm, \tag{1.5}$$

$$T = \frac{1}{\pi}\underset{\epsilon\to+0}{\text{s-lim}} \int_{-\infty}^{+\infty} d_{\lambda'}E_{\lambda'}^{H_0}\left((H-\lambda')\frac{\epsilon}{(H-\lambda')^2+\epsilon^2}\right)W_-^{(s)} \tag{1.6}$$

where $\Psi_\pm = W_\pm^{(s)}\Psi$.

The stationary scattering theory outlined above can be generalized[7] to include long-range potentials by basing the theory on the expressions (1.2) for the renormalized wave operators. The expressions for the renormalized wave operator and T operator take the following form:

$$W_\pm^{(r)} = \underset{\epsilon\to+0}{\text{s-lim}} \int_{-\infty}^{+\infty} \Lambda(H,\lambda',\pm\epsilon)d_{\lambda'}E_{\lambda'}^{H_0}, \tag{1.7}$$

$$T = \underset{\epsilon\to+0}{\text{s-lim}} \frac{1}{2\pi i}\int_{-\infty}^{+\infty} d_{\lambda'}E_{\lambda'}^{H_0}[\Lambda^*(H,\lambda',-\epsilon) - \Lambda^*(H,\lambda',+\epsilon)]W_-^{(r)}, \tag{1.8}$$

where $\Lambda(\lambda,\lambda',\pm\epsilon)$ is defined by the following integral:

$$\Lambda(\lambda,\lambda',\pm\epsilon) = \pm\epsilon\int_0^{\pm\infty} \exp[i(\lambda-\lambda')t - iG(\lambda';t) \mp \epsilon t]dt. \tag{1.9}$$

Due to the complicated structure of (1.9) the derivation of Lippmann-Schwinger equations and the explicit taking of the limit $\epsilon \to +0$ for the above T operator expressed in terms of eigenfunctions of H_0 (which would lead to expressions for the T matrix) is a very difficult undertaking. It is instead convenient to introduce (Sec.3) time-independent operators $Z_\pm^\Delta$ for which $W_\pm^{(r)}E^{H_0}(\Delta)$ can be written as

$$W_\pm^{(r)}E^{H_0}(\Delta) = \underset{t\to\pm\infty}{\text{s-lim}}\ \exp(iHt)Z_\pm^\Delta\exp(-iH_0t) \tag{1.10}$$

for any Δ which is a compact subset of the spectrum σ_0 of H_0.

A general procedure for the construction of the operators $Z_\pm^\Delta$ is given in Sec. 3. It is convenient to use two-Hilbert space theory in order to derive Hilbert space versions of the Lippmann-Schwinger equations and T operator involving the operators $Z_\pm^\Delta$. Existence and boundedness of $Z_\pm^\Delta$ is proven for the case of asymptotically Coulombic potentials (i.e., potentials which are the sum of the Coulomb potential and a short-range potential) in Sec. 4.

In Sec. 2 another version of the Lippmann-Schwinger equations is derived. These equations are independent of the operators $Z_\pm^\Delta$ and are valid for potentials which behave asymptotically as $r^{-\kappa}$, $\frac{1}{2} < \kappa \le 1$ and whose renormalized wave operators $W_\pm^{(r)}$ are complete. Since they do not contain explicitly inhomogeneous terms related to the plane waves, they are not computationally suitable for calculating distorted waves for long-range potentials. However, these equations turn out to be extremely useful in understanding the long-range theory as a limit of a theory with a space cutoff.

In Sec. 5 we compare the equations for distorted waves for a Coulomb potential (obtained in Sec. 4) with the corresponding equations for the same potential with a space cut-off. This enables us to study the behaviour of the cutoff theory in the limit when the cutoff is being

removed. We arrive at conclusions which partially confirm earlier findings pertaining to the pure Coulomb case $V(\mathbf{r}) = cr^{-1}$. However, these results also explicitly display the limitations of any method for long-range potentials based on solving the problem for the short-range case obtained by introducing a cutoff, and then removing that cutoff. It turns out that such a procedure is suitable off the energy shell but it can be misleading *on* the energy shell. This is due to the fact that on the energy shell the leading terms obtained in the modified Lippmann-Schwinger equations *off* the energy shell become zero by virtue of the equations derived in Sec. 2. In particular, we conclude that the distorted waves for long-range potentials *cannot* be obtained by *first* removing the cut-off and eliminating a "renormalization" factor while working off the energy shell with the standard Lippmann-Schwinger equations, and *afterwards* making the transition to the energy shell.

2. HOMOGENEOUS MODIFIED LIPPMANN-SCHWINGER EQUATIONS FOR DISTORTED WAVES

In this section we derive a set of integral equations satisfied by distorted waves for scattering in long-range potentials. The equations do not contain the operators $Z_{\pm}^{\Delta}$ appearing in (1.10) and are quite distinct from the equations for distorted waves obtained by taking (1.10) as a starting point. The procedure used in deriving the equations is a direct extension of the approach taken for the case of short-range potentials. Essential use is made of the fact that for long-range potentials we have

$$\text{w-lim}_{t\to\pm\infty} \exp(iHt)\exp(-iH_0t) = 0.$$

This fact has been first established by Dollard[8] for the Coulomb potential. The resulting equations

$$\Psi_{\pm} = \text{w-lim}_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{\lambda - H_0 \mp i\epsilon}\, V dE_{\lambda}^{H}\Psi_{\pm} \tag{2.1}$$

correspond to Hilbert space versions of the Lippmann-Schwinger equations for distorted waves, derived for the Coulomb potential by West[9] (who used a different procedure).

The following two lemmas are required:

Lemma 2.1: Suppose the wave operators

$$W_{\pm}^{(\tau)} = \text{s-lim}_{t\to\pm\infty} \exp(iHt)\exp\{-i[H_0t + G(H_0;t)]\} \tag{2.2}$$

exist on the Hilbert space $\mathfrak{K}$ and are complete, where $G(H_0;t)$ satisfies

$$\text{w-lim}_{t\to\pm\infty} \exp[iG(H_0;t)] = 0 \tag{2.3}$$

and H_0 has an absolutely continuous spectrum. Then we have

$$\text{w-lim}_{t\to\pm\infty} \exp(iHt)\exp(-iH_0t) = 0 \tag{2.4}$$

and

$$\text{w-lim}_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{\mp i\epsilon}{\lambda - H_0 \mp i\epsilon}\, d_{\lambda}E_{\lambda}^{H} = \text{w-lim}_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H - \lambda \pm i\epsilon}\, d_{\lambda}E_{\lambda}^{H_0} = 0. \tag{2.5}$$

If $\mathfrak{D}_H = \mathfrak{D}_{H_0}$ and $V = H - H_0$, then

$$\Phi = \text{w-lim}_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{\lambda - H_0 \mp i\epsilon}\, V d_{\lambda}E_{\lambda}^{H}\Phi \tag{2.6}$$

for all $\Phi \in \mathfrak{D}_H$ and

$$\Psi = \text{w-lim}_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{H - \lambda \mp i\epsilon}\, V d_{\lambda}E_{\lambda}^{H_0}\Psi \tag{2.7}$$

for all $\Psi \in \mathfrak{D}_{H_0}$.

Proof: We first prove that

$$\lim_{t\to\pm\infty} \langle g|\exp(iHt)\exp(-iH_0t)f\rangle = 0 \quad \text{for every } g \in \mathfrak{K}.$$

If g is a bound state, i.e., $Hg = -\eta g$, we have

$$\langle g|\exp(iHt)\exp(-iH_0t)f\rangle = \exp(-i\eta t)\langle g|\exp(-iH_0t)f\rangle.$$

Since the spectrum σ_0 of H_0 is absolutely continuous,

$$\langle g|\exp(-iH_0t)f\rangle = \int_{\sigma_0} \exp(-i\lambda t)\,\frac{d\langle g|E_{\lambda}^{H_0}f\rangle}{d\lambda}\, d\lambda.$$

The above expression converges to zero as $t \to \pm\infty$ by the Riemann-Lebesgue lemma.

If g is not a bound state we have

$$\begin{aligned}
&|\langle g|\{\exp(iHt)\exp(-iH_0t) - W_{\pm}^{(\tau)}\exp[iG(H_0;t)]\}f\rangle| \\
&\quad = |\langle\{\exp(iH_0t)\exp[iG(H_0;t)]\exp(-iHt) \\
&\qquad - W_{\pm}^{(\tau)*}\}g|\exp[iG(H_0;t)]f\rangle| \\
&\quad \le \|\{\exp(iH_0t)\exp[iG(H_0,t)]\exp(-iHt) \\
&\qquad - W_{\pm}^{(\tau)*}\}g\|\,\|f\|,
\end{aligned}$$

with the last expression going to zero as $t \to \pm\infty$ due to the completeness of $W_{\pm}^{(\tau)}$. Thus

$$\lim_{t\to\pm\infty} \langle g|\exp(iHt)\exp(-iH_0t)f\rangle = \lim_{t\to\pm\infty} \langle g|W_{\pm}^{(\tau)}\exp[iG(H_0;t)]f\rangle = 0$$

by (2.3) and therefore (2.4) is true.

We note that for any two self-adjoint operators H_0 and H we have[5]

$$\left(\int_{-\infty}^{+\infty} \frac{\mp i\epsilon}{\lambda - H_0 \mp i\epsilon}\, d_{\lambda}E_{\lambda}^{H}\right)^{*} = \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H - \lambda \pm i\epsilon}\, d_{\lambda}E_{\lambda}^{H_0} = \pm\epsilon \int_{0}^{\pm\infty} \exp([\mp\epsilon + iH]t)\exp(-iH_0t)dt.$$

Now if (2.4) is true then (cf. Ref. 5, p. 437)

$$\text{w-lim}_{\epsilon\to+0}(\pm\epsilon)\int_{0}^{\pm\infty} \exp([\mp\epsilon + iH]t)\exp(-iH_0t)dt = 0.$$

Thus (2.5) is verified. To prove (2.6) we use the following identity

$$\int_{-\infty}^{+\infty} \frac{\lambda - H_0}{\lambda - H_0 \mp i\epsilon}\, d_{\lambda}E_{\lambda}^{H} + \int_{-\infty}^{+\infty} \frac{\mp i\epsilon}{\lambda - H_0 \mp i\epsilon}\, d_{\lambda}E_{\lambda}^{H} = \mathbf{1},$$

the result following from (2.5) and the following relation (cf. Ref. 10, Lemma 2.1)

$$\int_{-\infty}^{+\infty} \frac{\lambda}{\lambda - H_0 \mp i\epsilon}\, d_{\lambda}E_{\lambda}^{H} = \int_{-\infty}^{+\infty} \frac{1}{\lambda - H_0 \mp i\epsilon}\, H d_{\lambda}E_{\lambda}^{H}. \tag{2.8}$$

The validity of (2.7) follows from the symmetric roles played in (2.5) by the operators H and H_0.

In order to apply Lemma 2.1 we must first examine under what conditions (2.3) is true.

Lemma 2.2: Let H have an absolutely continuous spectrum and suppose $-G(\lambda;t) = g_1(\lambda) + f_1(\lambda)f_2(t)$, where $\lim_{t\to\pm\infty} f_2(t) = \infty$ and $g_1(\lambda)$ is real. Also assume $u = f_1(\lambda)$ is such that $\lambda = f_1^{-1}(u)$ and $df_1(\lambda)/d\lambda$ exist almost everywhere on the set $\sigma = f_1((-\infty, +\infty))$. Then (2.3) is true.

Proof: We consider the following expression for $f, g \in \mathcal{H}$

$$\langle f|\exp[-iG(H_0;t)]g\rangle = \int_{-\infty}^{+\infty} \exp[if_1(\lambda)f_2(t)]\exp[ig_1(\lambda)]\frac{d\langle f|E_\lambda^{H_0}g\rangle}{d\lambda}d\lambda. \quad (2.9)$$

By performing the substitution of variables $u = f_1(\lambda)$ and denoting $\phi(\lambda) = d\langle f|E_\lambda^{H_0}g\rangle/d\lambda$, we obtain

$$\int_\sigma \exp[iuf_2(t)]\exp\{ig_1[f_1^{-1}(u)]\} \times \phi[f_1^{-1}(u)]\left(\frac{df_1(\lambda)}{d\lambda}\bigg|_{\lambda=f_1^{-1}(u)}\right)^{-1} du. \quad (2.10)$$

Now since the above integrand is an element of $L^1(\sigma)$ we obtain (2.3) from (2.10) by an application of the Riemann-Lebesgue lemma.

Thus whenever $G(H_0;t)$ satisfies the requirements of Lemma 2.2 we conclude that (2.6) is true provided $W(\frac{r}{\pm})$ are complete and $\mathfrak{D}_H = \mathfrak{D}_{H_0}$. In particular, for $\Phi = \Psi_\pm = W(\frac{r}{\pm})\Psi$, where $\Psi_\pm$ are the interacting states having the free state Ψ as an incoming or outgoing asymptotic state, respectively, the following integral equations are valid

$$\Psi_+ = \text{w-}\lim_{\epsilon\to+0}\int_{-\infty}^{+\infty}\frac{1}{\lambda - H_0 \pm i\epsilon}\,VdE_\lambda^H\Psi_+,$$
$$\Psi_- = \text{w-}\lim_{\epsilon\to+0}\int_{-\infty}^{+\infty}\frac{1}{\lambda - H_0 \pm i\epsilon}\,VdE_\lambda^H\Psi_-, \quad (2.11)$$

They represent one alternative set of integral equations which for long-range potentials can be considered to replace (1.5).

We can verify explicitly from the renormalization term that (2.11) holds for a certain class of asymptotically Coulombic potentials. For Coulomb-like potentials the renormalization term takes the form[11,12]

$$G(\lambda;t) = \pm\left\{\frac{q_0}{\sqrt{\lambda}}\log(4\lambda) + \frac{q_0}{\sqrt{\lambda}}\log|t|\right\}, \quad q_0 = \left(\frac{m}{2}\right)^{1/2}e_1e_2 \quad (2.12)$$

where the positive and negative sign are valid for $t > 0$ and $t < 0$, respectively. Then we can make the following substitution in Lemma 2.2:

$$f_1(\lambda) = \pm\frac{1}{\sqrt{\lambda}}, \qquad f_2(t) = q_0\log|t|,$$
$$g_1(\lambda) = \pm\frac{q_0}{\sqrt{\lambda}}\log(4\lambda).$$

Thus, in order to have (2.11) hold for Coulomb-like potential we require the wave operators $W(\frac{r}{\pm})$ to be complete. A recent study[13] establishes completeness for a large class of asymptotically Coulombic potentials.

For potentials which decrease asymptotically as r^{-k} the renormalization term is known[1,14] for all $\frac{1}{2} < k \leq 1$. It can easily be shown that Lemma 2.2 holds for such potentials. Thus if the renormalized wave operators are complete the interacting state satisfies Eqs. (2.11).

We note that (2.11) are the Hilbert space analogs of the integral equations

$$\phi_{\mathbf{k}}^{\pm}(\mathbf{r}) \equiv -\lim_{\epsilon\to+0}\int_{\mathcal{R}^3} G_0\left(\mathbf{r},\mathbf{r}';\frac{k^2}{2m}\pm i\epsilon\right)V(\mathbf{r}')\phi_{\mathbf{k}}^{\pm}(\mathbf{r}')d\mathbf{r}' \quad (2.13)$$

$$G_0(\mathbf{r},\mathbf{r}';\zeta) = \frac{m}{2\pi|\mathbf{r}-\mathbf{r}'|}\exp(i\sqrt{2m\zeta}\,|\mathbf{r}-\mathbf{r}'|), \quad \text{Im}\sqrt{\zeta} > 0, \quad (2.14)$$

with similar equations holding for $\phi_{\mathbf{k}}^{-}(\mathbf{r})$, where the above square root is chosen with positive imaginary part and the limit for $\epsilon \to +0$ exists at least in the weak sense of distributions. For the Coulomb case $V_c(\mathbf{r}) = \sqrt{2}/m \times q_0|\mathbf{r}|^{-1}$ the above equations become the equations derived by West [cf. Ref. 9, Eq. (44)].

To rigorously derive equations of the form (2.13) from (2.11) we have to resort to some specification of eigenfunction expansions for H_0 and H. The most general formulation[15] can be achieved by equipping $\mathcal{H}$ with spaces of positive and negative norm. However, in potential scattering eigenfunction expansions of the operator H_0 can be introduced directly[5] as functions $\phi_\gamma(\beta)$ of the variables $\beta, \gamma \in \mathcal{R}^l$ which correspond to the complete sets $\mathcal{B}$ and $\mathcal{C}$ of observables (such as position and momentum observables for spinless particles) if H_0 can be expressed as a function of one of these two sets, say $\mathcal{C}$ [i.e., $(H_0\Psi)(\gamma) = h(\gamma)\Psi(\gamma)$ for any $\Psi \in \mathfrak{D}_{H_0}$] in the spectral representation space[5] $L_\mu^2(\mathcal{R}^l)$ of $\mathcal{C}$. The precise mathematical statements embodying these assumptions lie in the request that $h(\gamma)$ be Borel measurable and that for any Borel set $\Delta \subset \mathcal{R}^l$ and for any $f(\beta)$ in the spectral representation space $L_\nu^2(\mathcal{R}^l)$ of $\mathcal{B}$

$$[E^{H_0}(\Delta)f](\beta) = \text{l.i.m.}\int_{h^{-1}(\Delta)}\phi_\gamma(\beta)\tilde{f}(\gamma)d\mu(\gamma), \quad (2.15)$$

$$\tilde{f}(\gamma) = \text{l.i.m.}\int_{\mathcal{R}^l}\overline{\phi_\gamma(\beta)}f(\beta)d\nu(\beta). \quad (2.16)$$

If we assume that $\phi_\gamma^{\pm}(\beta)$ are the eigenfunction expansions of H, then for f in the ranges of $W(\frac{r}{\pm})$ we have

$$[E^H(\Delta)f](\beta) = \text{l.i.m.}\int_{h^{-1}(\Delta)}\phi_\gamma^{\pm}(\beta)\tilde{f}(\gamma)d\mu(\gamma). \quad (2.17)$$

Combining (2.11) with (2.17), we get as a consequence of the definition of weak Riemann-Stieltjes integrals (cf. Ref. 5, p. 472, Lemma 4.1)

$$\int_{-\infty}^{+\infty} d_\lambda\langle g|\frac{1}{\lambda - H_0 \mp i\epsilon}\,VE_\lambda^H W(\frac{r}{\pm})f\rangle = \int_{\mathcal{R}^l} d\nu(\beta)\overline{g(\beta)}\int_{\mathcal{R}^l}\left(\beta\left|\frac{1}{h(\gamma) - H_0 \mp i\epsilon}\,V\right|\beta'\right) \times \int_{\mathcal{R}^l}\phi_\gamma^{\pm}(\beta')\tilde{f}(\gamma)d\mu(\gamma). \quad (2.18)$$

If we restrict ourselves to functions $g(\beta)$ and $\tilde{f}(\gamma)$ with supports of finite ν and μ measure, respectively, and assume that $|\phi_\gamma^{\pm}(\beta')| \leq$ constant and that $(\beta|[1/(h(\gamma) - H_0 \mp i\epsilon)]V|\beta')$ is ν-integrable in β', we can use Tonelli's and Fubini's theorems[5] to state that as a consequence of (2.11), (2.15), and (2.18)

$$\int d\nu(\beta)\overline{g(\beta)}\int d\mu(\gamma)\phi_\gamma^{\pm}(\beta)\tilde{f}(\gamma) = \lim_{\epsilon\to+0}\int_{\mathcal{R}^l} d\nu(\beta)\overline{g(\beta)}\int_{\mathcal{R}^l} d\mu(\gamma)\tilde{f}(\gamma) \times \int_{\mathcal{R}^l}\left(\beta\left|\frac{1}{h(\gamma) - H_0 \mp i\epsilon}\,V\right|\beta'\right)\phi_\gamma^{\pm}(\beta')d\nu(\beta'). \quad (2.19)$$

If, in addition, we assume

$$\lim_{\epsilon\to+0}\int_{\mathcal{R}^l}\left(\beta\left|\frac{1}{h(\gamma)-H_0\mp i\epsilon}V\right|\beta'\right)\phi_\gamma^\pm(\beta')d\nu(\beta')$$

exists and

$$|\overline{g(\beta)}|\,|\bar{f}(\gamma)|\left|\int_{\mathcal{R}^l}\left(\beta\left|\frac{1}{h(\gamma)-H_0\mp i\epsilon}V\right|\beta'\right)\phi_\gamma^\pm(\beta')d\nu(\beta')\right|$$

is bounded by an element of $L^1(\mu\times\nu,\mathcal{R}^{2l})$ independent of ϵ, we conclude by the Lebesgue dominated convergence theorem

$$\phi_\gamma^\pm(\beta)=\lim_{\epsilon\to+0}\int_{\mathcal{R}^l}\left(\beta\left|\frac{1}{h(\alpha)-H_0\mp i\epsilon}V\right|\beta'\right)\phi_\gamma^\pm(\beta')d\nu(\beta') \tag{2.20}$$

for ν-almost all $\beta\in\mathcal{R}^l$ and μ-almost all $\gamma\in\mathcal{R}^l$. However, in general we can expect that (2.20) holds only in the sense of the preceding Eq. (2.19), i.e., only after a smearing in β or γ or both is carried out (cf. Sec. 5). We obtain (2.13) by specializing (2.20) to the potential scattering case where $\beta=\mathbf{r}$, $\gamma=\mathbf{k}$, and $h(k)=k^2/2m$.

It should be emphasized that the solution of (2.13) is not unique since both the outgoing and incoming distorted waves satisfy exactly the same equations. We note that a very important implication of (2.20) is that the ordinary Lippmann-Schwinger equations

$$\phi_\gamma^\pm(\beta)=\phi_\gamma(\beta)+\lim_{\epsilon\to0}\int_{\mathcal{R}^l}\left(\beta\left|\frac{1}{h(\alpha)-H_0\pm i\epsilon}V\right|\beta'\right)\phi_\gamma^\pm(\beta')d\nu(\beta') \tag{2.21}$$

cannot hold in the case of long-range potentials. That would lead to the self-contradictory conclusion that the free wave $\phi_\gamma(\beta)$ is identically equal to zero.

3. THE CONSTRUCTION OF $Z_\pm^\Delta$ FOR LONG-RANGE INTERACTIONS

The existence of operators $Z_\pm^\Delta$ for which (1.10) is satisfied has been established in Ref. 16 for s-level two-body scattering with spherically symmetric potential $V(r)$ for which $|(d^\kappa/dr^\kappa)V(r)|\le c\,|1+r|^{-\epsilon-\kappa}$ for some $\epsilon>0$ and $\kappa=0,1,2$. The method of proof is analytic and not applicable to larger classes of long-range potentials, including those which are asymptotically Coulombic, but locally nonspherically symmetric and without outstanding smoothness properties. In this section we present a procedure which is applicable to such cases.

To formulate a general procedure for constructing the operators $Z_\pm^\Delta$ when $G(H_0;t)$ is given, it is convenient to introduce an auxiliary set of observables $\{A_1,\ldots,A_n\}$ which are independent of H_0 but which together with H_0 form a complete set.[5] Then one can work in the spectral representation space[5] $L^2_{\mu_0}(\mathcal{R}^{n+1})$ of $\{H_0,A_1,\ldots,A_n\}$ in which $H_0,A_1,\ldots,A_n$ are represented by

$$(H_0'\psi)(\lambda,\alpha)=\lambda\psi(\lambda,\alpha),\quad \mathfrak{D}_{H_0}=[\psi|\textstyle\int\lambda^2|\psi(\lambda,\alpha)|^2d\mu_0<\infty],$$
$$(A'\psi)(\lambda,\alpha)=\alpha_j\psi(\lambda,\alpha),\quad \alpha=(\alpha_1,\ldots,\alpha_n)\in\mathcal{R}^n. \tag{3.1}$$

If we further assume that the spectrum of H_0 is absolutely continuous (as is always the case in potential scattering), then $\mu_0=\mu_1\times\mu$, where Radon-Nikodym derivative $d\mu_1(\lambda)/d\lambda$ of μ_1 exists. Hence if $\mathfrak{L}^2$ denotes the space $L^2(\mathcal{R}^{n+1},d\lambda d\mu)$ of functions $\psi(\lambda,\alpha)$ square integrable with respect to the Cartesian product of the Lebesgue measure on $\mathcal{R}^1$ with the measure μ on $\mathcal{R}^n$, we can choose $\mathfrak{L}^2(\sigma_0)=P_{\sigma_0}\mathfrak{L}^2$ as a spectral representation space for $\{H_0,A_1,\ldots,A_n\}$ in which (3.1) holds, where

$$(P_\Delta\psi)(\lambda,\alpha)=\chi_\Delta(\lambda)\psi(\lambda,\alpha),\quad \psi\in\mathfrak{L}^2, \tag{3.2}$$

$\chi_\Delta(\lambda)$ being the characteristic function of the set Δ, which in the present case is the spectrum σ_0 of H_0.

According to the theory of Fourier-Plancherel transforms, the mapping which takes $\psi\in\mathfrak{L}^2$ into

$$\hat\psi(t,\alpha)=(2\pi)^{-1/2}\ \text{l.i.m.}\int_{-\infty}^{+\infty}\exp(i\lambda t)\psi(\lambda,\alpha)d\lambda \tag{3.3}$$

is a unitary operator on $\mathfrak{L}^2$. If we set $G(\lambda,t)=0$ for $\lambda\notin\sigma_0$ at any $t\in\mathcal{R}^1$, the operators $Z_\pm$ can be introduced as

$$Z_\pm=P_{\sigma_0}Z_\pm', \tag{3.4}$$

where, if the following limit in the mean exists, we have

$$(Z_\pm'\psi)(\lambda,\alpha)=\pm(2\pi)^{-1/2}\times\text{l.i.m.}\int_0^{\pm\infty}\exp[-i\lambda t-iG(\lambda;t)]\hat\psi(t,\alpha)dt. \tag{3.5}$$

It is clear that the above operators $Z_\pm'$ are not going to be bounded or even densely defined in $\mathfrak{L}^2$ for arbitrary choices of $G(\lambda;t)$ [such as $G(\lambda;t)=-\lambda t$]. In fact, we shall see that $Z_\pm'$ is not bounded even for $G(\lambda;t)$ chosen in Sec. 4 to satisfy the existence requirements of particular renormalized wave operators such as those for asymptotically Coulombic potentials. Now, boundedness is desirable if we want to apply the formulas of two-Hilbert space theory[10] to the present case. Hence the following operators are introduced: $Z_\pm^\Delta=E^{H_0}(\Delta)Z_\pm E^{H_0}(\Delta)$, where Δ is a compact subset of σ_0. The boundedness of this operator can be investigated for each particular class of potentials.

A study of the boundedness of $Z_\pm^\Delta$ can be based on the following representation of $Z_\pm$,

$$(Z_\pm\psi)(\lambda,\alpha)=\lim_{\epsilon\to+0}\int_{-\infty}^{+\infty}Z_{\pm\epsilon}(\lambda,\lambda')\psi(\lambda',\alpha)d\lambda' \tag{3.6}$$

valid at least for all $\psi\in\mathfrak{L}^2$ which are integrable as well as square-integrable, where

$$Z_{\pm\epsilon}(\lambda,\lambda')=\pm(2\pi)^{-1}\int_0^{\pm\infty}\exp[\mp\epsilon t-i(\lambda-\lambda')t-iG(\lambda;t)]dt. \tag{3.7}$$

The expression (3.6) follows from the relation

$$(Z_\pm\psi)(\lambda,\alpha)=\pm(2\pi)^{-1/2}\lim_{\epsilon\to+0}\int_0^{\pm\infty}\exp(\mp\epsilon t)\times\exp[-i\lambda t-iG(\lambda;t)]\hat\psi(t,\alpha)dt, \tag{3.8}$$

(which holds by the Lebesgue dominated convergence theorem) by substituting (3.3) for $\hat\psi(t,\alpha)$ and using Tonelli's and Fubini's theorems[5] to reverse the orders of integration in t and λ.

In order to prove that the operators $Z_\pm^\Delta$ are the required operators for which (1.10) holds true, we note first that the limit in (1.2) is equal to the limit in (1.10) for $\psi\in\mathfrak{L}^2(\Delta)=P_\Delta\mathfrak{L}^2=L^2(\Delta)\otimes L^2_\mu(\mathcal{R}^n)$ if and only if

$$\text{s-lim}_{t\to\pm\infty}\exp[iH_0t+iG(H_0;t)]Z_\pm^\Delta\exp(-iH_0t)=P_\Delta. \tag{3.9}$$

To prove (3.9) we have to take advantage of the fact that from the physical as well as the mathematical point of view the functions $G(H_0;t)$ have to satisfy the conditions

$$\text{s-lim}_{t\to\pm\infty}\exp[iG(H_0;t+\tau)-iG(H_0;t)]=1 \tag{3.10}$$

for any $\tau\in\mathcal{R}^1$. In fact, the above relation is a necessary and sufficient prerequisite[4] for the intertwining pro-

perties of the wave operators $W_{\pm}^{(\tau)}$, which are essential from both a physical and mathematical point of view. These properties imply the following result, which is required in the sequel.

Lemma 3.1: If the spectrum σ_0 of H_0 is absolutely continuous, then (3.10) is true if and only if

$$\lim_{t\to\pm\infty} |1 - \exp[iG(\lambda; t+\tau) - iG(\lambda; t)]| = 0 \tag{3.11}$$

at each fixed $\tau \in \mathcal{R}^1$ for all $\lambda \in \sigma_0$, with the possible exception of a set of Lebesgue measure zero.

Proof: If (3.11) is true then (3.10) follows from

$$\|\psi - \exp[iG(\lambda; t+\tau) - iG(\lambda; t)]\psi\|^2 = \int_{\sigma_0} |1 - \exp[iG(\lambda; t+\tau) - iG(\lambda; t)]|^2 \frac{d\|E_\lambda^{H_0}\psi\|^2}{d\lambda} d\lambda$$

by applying Lebesgue's bounded convergence theorem. Conversely, if (3.10) is true for some fixed $\tau \in \mathcal{R}^1$, then by choosing $\psi_n \in \mathcal{H}$ such that

$$\frac{d\|E_\lambda^{H_0}\psi_n\|^2}{d\lambda} = \chi_{\sigma_0 \cap [n, n+1)}(\lambda)$$

we conclude that (3.11) is satisfied almost everywhere on $\sigma_0 \cap [n, n+1)$. Hence (3.11) is true almost everywhere on

$$\sigma_0 = \bigcup_{n=-\infty}^{+\infty} \sigma_0 \cap [n, n+1).$$

An easy computation based on (3.6) and (3.7) leads to the result that for any $\psi(\lambda, \alpha) \in \mathfrak{L}^2(\Delta)$ which belongs to $L^1(\Delta)$ for μ-almost all $\alpha \in \mathcal{R}^n$

$$\|\psi - \exp[iH_0 t + iG(H_0; t)]Z_{\pm}^{\Delta} \exp(-iH_0 t)\psi\|^2 = \int_\Delta B_\pm(\lambda, t, \psi) d\lambda, \tag{3.12}$$

where in terms of the Heavyside function $\theta(s)$, for which $\theta(s) = 0$ for $s < 0$ and $\theta(s) = 1$ for $s > 0$, we have

$$B_\pm(\lambda, t, \psi) = \Big| \int_{-\infty}^{+\infty} \{1 - \theta(\pm[t+\tau]) \exp[iG(\lambda; t) - iG(\lambda; t+\tau)]\} \exp(-i\lambda\tau)\hat{\psi}(\tau, \alpha) d\tau d\mu(\alpha) \Big|^2. \tag{3.13}$$

In view of (3.11) and the fact that the integrand in (3.13) is bounded by the integrable function $|2\hat{\psi}(\tau, \alpha)|$, we conclude by Lebesgue's dominated convergence theorem[5] that

$$\lim_{t\to\pm\infty} B_\pm(\lambda, t, \psi) = 0. \tag{3.14}$$

We can apply this theorem again to (3.12) since we are dealing with a finite region of integration, thus arriving at (3.9) for ψ in the considered dense set of functions in $\mathfrak{L}^2(\Delta)$. Hence if $Z_{\pm}^{\Delta}$ is bounded on $\mathfrak{L}^2(\Delta)$ we can extend (3.9) to every $\psi \in \mathfrak{L}^2(\Delta)$.

Finally, let us point out that the above arguments leading to (3.9) also lead to the conclusion that

$$\text{s-}\lim_{t\to\pm\infty} \exp[iH_0 t + iG(H_0; t)]Z_{\mp}^{\Delta} \exp(-iH_0 t) = 0. \tag{3.15}$$

Hence the operator

$$Z^{\Delta} = Z_{+}^{\Delta} + Z_{-}^{\Delta} \tag{3.16}$$

has the properties

$$\text{s-}\lim_{t\to\pm\infty} \exp[iH_0 t + iG(H_0; t)]Z^{\Delta} \exp(-iH_0 t) = P_\Delta.$$

Thus we can deal with

$$W_{\pm}^{(\tau)}E^{H_0}(\Delta) = \text{s-}\lim_{t\to\pm\infty} \exp(iHt)Z^{\Delta} \exp(-iH_0 t) \tag{3.17}$$

instead of (1.10). Note that if $Z_\pm$ defined by (3.4) and (3.5) are bounded operators on $\mathfrak{L}^2(\sigma_0)$, then a similar argument as given above implies (3.17) holds with $\Delta = \sigma_0$.

We can summarize the above results in the following lemma.

Lemma 3.2: Suppose that the operators $Z_\pm$ defined on the spectral representation space $\mathfrak{L}^2(\sigma_0)$ by (3.4) and (3.5) are such that $Z_{\pm}^{\Delta} = E^{H_0}(\Delta)Z_\pm E^{H_0}(\Delta)$ are bounded for some compact subset Δ of the spectrum σ_0 of H_0. If $G(\lambda, t)$ satisfies (3.10) for almost all $\tau \in \mathcal{R}^1$ and $\lambda \in \sigma_0$, then $W_{\pm}^{(\tau)}E^{H_0}(\Delta)$ can be represented by (1.10) and (3.17).

The above lemma allows us to apply two-Hilbert space theory[10] to the operators $W_{\pm}^{(\tau)\Delta} = W_{\pm}^{(\tau)}E^{H_0}(\Delta)$ considered as wave operators from $\mathcal{H}_\Delta = E^{H_0}(\Delta)\mathcal{H}$ to $\mathcal{H}$, if we let Z^{Δ} play the role of the "identification" operator. Hence, all the results of Ref. 10, Sec. 2 can be restated for the present case, and in particular the following theorem is true.

Theorem 3.1: If (3.17) is true for some Borel subset Δ of σ_0, then for any $\Psi = E^{H_0}(\Delta)\Psi$ and $\Psi_\pm = W_{\pm}^{(\tau)}\Psi$

$$\Psi_\pm = Z^{\Delta}\Psi - \text{s-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{H - \lambda \pm i\epsilon} (HZ^{\Delta} - Z^{\Delta}H_0) d_\lambda E_\lambda^{H_0}\Psi, \tag{3.18}$$

$$Z^{\Delta *}\Psi_\pm = \Psi - \text{w-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{H_0 - \lambda \pm i\epsilon} \times (Z^{\Delta *}H - H_0 Z^{\Delta *}) d_\lambda E_\lambda^{H}\Psi_\pm, \tag{3.19}$$

and the T operator satisfies the relation

$$E^{H_0}(\Delta)TE^{H_0}(\Delta) = \frac{1}{\pi} \text{w-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} d_\lambda E_\lambda^{H_0}(Z^{\Delta *}H - H_0 Z^{\Delta *})\Omega_- \times \frac{\epsilon}{(H_0 - \lambda)^2 + \epsilon^2}. \tag{3.20}$$

We will now rewrite (3.18) as integral equations relating the free eigenfunctions $\phi_\gamma(\beta)$, $\gamma, \beta \in \mathcal{R}^l$, to the full eigenfunctions $\phi_\gamma^{\pm}(\beta)$. In a completely analogous derivation as that given for (2.18) we can write

$$\int_{-\infty}^{+\infty} d_\lambda \Big\langle g \Big| \frac{1}{H - \lambda \pm i\epsilon} (HZ^{\Delta} - Z^{\Delta}H_0) E_\lambda^{H_0} f \Big\rangle = \int_{\mathcal{R}^l} d\nu(\beta) \overline{g(\beta)} \int_{\mathcal{R}^l} \Big(\beta \Big| \frac{1}{H - h(\gamma) \pm i\epsilon} \times (HZ^{\Delta} - Z^{\Delta}H_0) \Big| \beta' \Big) d\nu(\beta') \int_{\mathcal{R}^l} \phi_\gamma(\beta') f(\gamma) d\mu(\gamma).$$

Under the same conditions on $(\beta|[1/(H - h(\gamma) \pm i\epsilon)] \times (HZ^{\Delta} - Z^{\Delta}H_0)|\beta')$ as stated in Sec. 2 for $(\beta|[1/h(\gamma) - H_0 \mp i\epsilon)]V|\beta')$ and the additional assumptions that $(\beta|Z^{\Delta}|\beta')$ be ν-integrable in β' we conclude that, under the implicit agreement that $h(\gamma) \in \Delta$,

$$\phi_\gamma^{\pm}(\beta) = \int_{\mathcal{R}^l} (\beta|Z^{\Delta}|\beta')\phi_\gamma(\beta') d\nu(\beta') - \lim_{\epsilon\to+0} \int_{\mathcal{R}^l} \Big(\beta \Big| \frac{1}{H - h(\gamma) \pm i\epsilon} (HZ^{\Delta} - Z^{\Delta}H_0) \Big| \beta' \Big) \times \phi_\gamma(\beta') d\nu(\beta') \tag{3.21}$$

is valid for ν-almost all β and μ-almost γ. In particular, if we choose $\beta = \mathbf{r}$ and $\gamma = (k, l, m)$, where k corresponds to $\sqrt{H_0}$, l to angular momentum $\mathbf{L}^2$, and m to the z component L_z of $\mathbf{L}$, then $\phi_\gamma(\beta) = \Phi_{klm}(\mathbf{r})$ is the outgoing free spherical wave[5] and (3.21) can be written as

$$\Phi^{(\pm)}_{klm}(\mathbf{r}) = \int_{\mathbb{R}^3} (\mathbf{r} \mid Z^\Delta \mid \mathbf{r}')\Phi_{klm}(\mathbf{r}')d\mathbf{r}' - \lim_{\epsilon \to +0} \int_{\mathbb{R}^3} d\mathbf{r}' G\left(\mathbf{r}, \mathbf{r}'; \frac{k^2}{2m} \mp i\epsilon\right) \times \int_{\mathbb{R}^3} (\mathbf{r}' \mid HZ^\Delta - Z^\Delta H_0 \mid \mathbf{r}'')\Phi_{klm}(\mathbf{r}'')d\mathbf{r}'', \tag{3.22}$$

where $G(\mathbf{r}, \mathbf{r}'; \zeta)$ denotes the full Green's function corresponding to the resolvent $(H - \zeta)^{-1}$. Note that (3.22) is valid only when $k^2/2m$ belongs to the set Δ!

4. EXISTENCE AND BOUNDEDNESS OF $Z^\Delta_\pm$ FOR ASYMPTOTICALLY COULOMB POTENTIALS

In this section we compute $Z^\Delta_\pm$ for asymptotically Coulomb potentials and show that it is a bounded operator for any compact subset Δ of $\sigma_0 = [0, \infty)$. We shall limit ourselves to sets Δ of the form $[c, d]$, where $0 \leq c < d < \infty$, since the result for the more general case follows from the result for such sets $\Delta = [c, d]$.

For asymptotically Coulomb potentials we can express the integral kernel in (3.7) in terms of gamma and logarithmic functions. By inserting the expression (2.12) in (3.7) and performing the integration in the complex t plane, we obtain

$$Z_{+\epsilon}(\lambda, \lambda') = \frac{1}{2\pi i}[(\lambda - \lambda') - i\epsilon]^{-1} \times \exp\left(\frac{iq_0}{\sqrt{\lambda}} \log[(\lambda - \lambda') - i\epsilon]\right) \times \exp\left(-\frac{\pi q_0}{2\sqrt{\lambda}}\right)(4\lambda)^{-iq_0/\sqrt{\lambda}}\, \Gamma\left(1 - \frac{iq_0}{\sqrt{\lambda}}\right), \tag{4.1}$$

$$Z_{-\epsilon}(\lambda, \lambda') = -\frac{1}{2\pi i}[(\lambda - \lambda') + i\epsilon]^{-1} \times \exp\left(-\frac{iq_0}{\sqrt{\lambda}} \log[(\lambda - \lambda') + i\epsilon]\right) \times \exp\left(-\frac{\pi q_0}{2\sqrt{\lambda}}\right)(4\lambda)^{iq_0/\sqrt{\lambda}}\, \Gamma\left(1 + \frac{iq_0}{\sqrt{\lambda}}\right), \tag{4.2}$$

where we have chosen the branches of the logarithmic function with $-3\pi/2 < \arg[(\lambda - \lambda') - i\epsilon] \leq \pi/2$ and $-\pi/2 < \arg[(\lambda - \lambda') + i\epsilon] \leq 3\pi/2$, respectively.

We will first evaluate the expression (3.6) for functions $\psi(\lambda'; \alpha) = \chi_{[a,b)}(\lambda')\psi_1(\alpha)$, where $\chi_{[a,b)}$ is the characteristic function of the interval $[a, b)$ with $0 < a < b < \infty$. The integration for the computation of $Z_\pm(\chi_{[a,b)} \otimes \psi_1) = (\hat{Z}_\pm \chi_{[a,b)}) \otimes \psi_1$ can be carried out by using (4.1), (4.2), and the limit $\epsilon \to +0$ can be taken explicitly. We give the results for the Z_+ case since the Z_- case is quite analogous:

$$(\hat{Z}_+ \chi_{[a,b)})(\lambda) = \frac{\sqrt{\lambda}}{2\pi q_0} \exp\left(-\frac{\pi q_0}{2\sqrt{\lambda}}\right)(4\lambda)^{-iq_0/\sqrt{\lambda}}\, \Gamma\left(1 - \frac{iq_0}{\sqrt{\lambda}}\right) \times \left[\exp\left(\frac{iq_0}{\sqrt{\lambda}} \log|\lambda - b|\right) - \exp\left(\frac{iq_0}{\sqrt{\lambda}} \log|\lambda - a|\right)\right], \quad a < b < \lambda;$$

$$(\hat{Z}_+ \chi_{[a,b)})(\lambda) = \frac{\sqrt{\lambda}}{2\pi q_0}(4\lambda)^{-iq_0/\sqrt{\lambda}}\, \Gamma\left(1 - \frac{iq_0}{\sqrt{\lambda}}\right) \times \left[\exp\left(\frac{iq_0}{\sqrt{\lambda}} \log|\lambda - b|\right)\exp\left(\frac{\pi q_0}{2\sqrt{\lambda}}\right) - \exp\left(\frac{iq_0}{\sqrt{\lambda}} \log|\lambda - a|\right)\exp\left(-\frac{\pi q_0}{2\sqrt{\lambda}}\right)\right], \quad a < \lambda < b;$$

$$(\hat{Z}_+ \chi_{[a,b)})(\lambda) = \frac{\sqrt{\lambda}}{2\pi q_0} \exp\left(\frac{\pi q_0}{2\sqrt{\lambda}}\right)(4\lambda)^{-iq_0/\sqrt{\lambda}}\, \Gamma\left(1 - \frac{iq_0}{\sqrt{\lambda}}\right) \times \left[\exp\left(\frac{iq_0}{\sqrt{\lambda}} \log|\lambda - b|\right) - \exp\left(\frac{iq_0}{\sqrt{\lambda}} \log|\lambda - a|\right)\right], \quad \lambda < a < b. \tag{4.3}$$

An inspection of the above equations reveals that infrared divergencies are not present since $|\Gamma(1 - iq_0/\sqrt{\lambda})\exp(\pi q_0/2\sqrt{\lambda})|$ stays finite as $\lambda \to +0$. However, the factor $\lambda^{1/2}$ in (4.3) shows that ultraviolet divergencies are present. Hence Z_+ is certainly not a bounded operator from $\mathfrak{L}^2(\sigma_0)$ to $\mathfrak{L}^2(\sigma_0)$.

However,

$$Z^\Delta_\pm(\chi_{[a,b)} \otimes \psi_1) = E^{H_0}(\Delta)Z_\pm(\chi_{[a,b)} \otimes \psi_1) \tag{4.4}$$

is a well-defined element of $\mathfrak{L}^2(\Delta)$ for $\Delta = [c, d]$ with $0 \leq c < a < b < d < \infty$. The boundness of $Z^\Delta_\pm$ obviously depends on whether the mapping

$$(D\chi_{[a,b)})(\lambda) = \sqrt{\lambda}\left[1 - \cos\left(\frac{1}{\sqrt{\lambda}} \log\left|\frac{\lambda - b}{\lambda - a}\right|\right)\right] \tag{4.5}$$

determines a bounded linear operator on $L^2(\Delta)$.

A study of the operator D is made in the following theorem.

Theorem 4.1: For fixed $0 \leq c < d < \infty$ the mapping $\chi_{[a,b)}(\lambda) \to (D\chi_{[a,b)})(\lambda)$ is defined for each $a, b \in (c, d)$, where $c < a < b < d$, and acts as a bounded linear operator on functions $\Psi(\lambda) \in \mathfrak{L}^2([c, d])$ which are of the form

$$\Psi(\lambda) = \sum_{j=1}^{n} \delta_j \chi_{[a_j, b_j)}(\lambda), \quad n = 1, 2, \cdots. \tag{4.6}$$

To prove the above theorem we require the following lemma:

Lemma 4.1: There is a constant $C = C(c, d)$ such that for any $c < a < b < d$ with $x = b - a$ we have

$$\int_c^d |\chi_{[a-2x, b+x)}(\lambda)(D\chi_{[a,b)})(\lambda)|^2 d\lambda \leq 16dx, \tag{4.7}$$

$$\int_c^d |1 - \chi_{[a-2x, b+x)}(\lambda)|^2 |(D\chi_{[a,b)})(\lambda)|^2 d\lambda \leq Cx^2. \tag{4.8}$$

Proof of Lemma 4.1: The inequality (4.7) follows immediately from the relation $|(D\chi_{[a,b)})(\lambda)| \leq 2\sqrt{d}$. The integral in (4.8) is equal to

$$I(x) = \left(\int_c^{a-2x} + \int_{b+x}^d\right)\left[1 - \cos\left(\frac{1}{\sqrt{\lambda}} \log\left|\frac{b - \lambda}{a - \lambda}\right|\right)\right]^2 \lambda d\lambda = \left(\int_c^{a-2x} + \int_{b+x}^d\right)\left[2\sin^2\left(\frac{1}{2\sqrt{\lambda}} \log\left|\frac{b - \lambda}{a - \lambda}\right|\right)\right]\lambda d\lambda.$$

Now

$$\frac{I(x)}{x^2} \leq \left(\int_c^{a-2x} + \int_{b+x}^d\right)\frac{1}{2}\left(\frac{\log|1 + [x/(a - \lambda)]|}{x}\right)^2 d\lambda.$$

Thus for $\lambda \leq a - 2x$ and $\lambda \geq a + 2x$ we have

$$I(x) \leq \frac{1}{2}(d-c)\left(\sum_{s=1}^{\infty}\frac{1}{2^s}\right)^2 x^2 = \frac{1}{2}(d-c)x^2.$$

Hence (4.8) holds and the lemma has been proven.

Proof of Theorem 4.1: Consider $\Psi(\lambda)$ given by (4.6), where we assume $(a_i, b_i) \cap (a_j, b_j) = \emptyset$ if $i \neq j$ and $x_{\max}/x_{\min} \leq 2$ for every n, where $x_{\min} = \min_{1\leq i\leq n}(b_i - a_i)$, $x_{\max} = \max_{1\leq i\leq n}(b_i - a_i)$. This last requirement can always be achieved by subdividing the partition in an appropriate manner.

We now decompose the action of D on $\chi_{[a_j,b_j)}$ into two parts:

$$(D\chi_{[a_j,b_j)})'(\lambda) = \delta_j \chi_{[a_j-2x_j,\,b_j+x_j)}(\lambda)(D\chi_{[a_j,b_j)})(\lambda),$$

$$(D\chi_{[a_j,b_j)})''(\lambda) = \delta_j (D\chi_{[a_j,b_j)})(\lambda) - (D\chi_{[a_j,b_j)})'(\lambda).$$

Thus we can write

$$\|D\Psi\| \leq \|\sum_{j=1}^{n}(D\chi_{[a_j,b_j)})'\| + \sum_{j=1}^{n}\|(D\chi_{[a_j,b_j)})''\|$$

$$\leq \|\sum_{j=1}^{n}(D\chi_{[a_j,b_j)})'\| + C^{1/2}(d-c)^{1/2}\|\Psi\|, \quad (4.9)$$

where Lemma 4.1 was used to obtain that

$$\sum_{j=1}^{n}\|(D\chi_{[a_j,b_j)})''\| \leq \sum_{j=1}^{n} C^{1/2}\delta_j(b_j - a_j)$$

$$\leq C^{1/2}\left(\sum_{j=1}^{n}(b_j - a_j)\right)^{1/2}\left(\sum_{j=1}^{n}\delta_j^2(b_j - a_j)\right)^{1/2}$$

$$= C^{1/2}(d-c)^{1/2}\|\Psi\|.$$

Now we write the first term in (4.9) as

$$\|\sum_{i=1}^{n}(D\chi_{[a_i,b_i)})'\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\langle(D\chi_{[a_i,b_i)})'\,|\,(D\chi_{[a_j,b_j)})'\rangle \quad (4.10)$$

and consider any fixed value of i. The only terms in the above sum over j that contribute are those for which

$$[a_i - 2x_i, b_i + x_i) \cap [a_j - 2x_j, b_j + x_j) \neq \emptyset.$$

According to an earlier made assumption $x_{\max}/x_{\min} \leq 2$. Hence we conclude that the maximum number of terms which contribute in the sum over j are ten for $j \leq i$ and eight for $j > i$. Thus (4.10) reduces to

$$\sum_{i=1}^{n}\sum_{j=i-10}^{i+8}|\langle(D\chi_{[a_i,b_i)})'\,|\,(D\chi_{[a_j,b_j)})'\rangle|$$

$$\leq \sum_{i=1}^{n}\sum_{j=i-10}^{i+8}\|(D\chi_{[a_i,b_i)})'\|\,\|(D\chi_{[a_j,b_j)})'\|$$

$$\leq (16)(19)d\|\Psi\|^2,$$

where (4.7) was used to obtain the bound on $\|(D\chi_{[a_i,b_i)})'\|$.

Since any function whose support is contained in $[c,d]$ can be approximated by a sequence of functions of the form (4.6), we conclude that the operator D is bounded on $L^2([c,d])$. This completes the proof of Theorem 4.1.

Due to the boundedness of D and the form of Z_+^{Δ} given in (4.3) we conclude that Z_+^{Δ} is a bounded operator for $\Delta = [c,d]$, where $0 \leq c < d < \infty$. The same is true for Z_-^{Δ} by a similar argument. Thus Z^{Δ} is a bounded operator for Coulomb-like potentials. This implies that (3.17) is true.

5. REMOVAL OF THE SPACE CUTOFF IN SCATTERING THEORIES WITH LONG-RANGE POTENTIALS

In this section we shall compare the stationary scattering theory for long-range potential $V_\infty(\mathbf{r})$ with the theory in which a space cut-off (or screening) $v_R(\mathbf{r})$ is introduced, i.e., the theory with the potential

$$V_R(\mathbf{r}) = v_R(\mathbf{r})V_\infty(\mathbf{r}), \quad (5.1)$$

where v_R is some Borel measurable function for which $v_R(\mathbf{r}) = 1$ for $|\mathbf{r}| \leq R$ and $v_R(\mathbf{r}) = O(r^{2+\epsilon})$, $\epsilon > 0$, when $r \gg R$. For $R < \infty$ we are dealing with a theory with short-range interaction for which the strong wave-operators

$$W_\pm = \text{s-}\lim_{t\to\pm\infty}\exp(iH_R t)\exp(-iH_0 t), \qquad H_R = H_0 + V_R, \quad (5.2)$$

exist. It has been shown by Dollard[12] for the case of Coulomb-like potentials that

$$\text{w-}\lim_{R\to+\infty} W_\pm^R = \text{w-}\lim_{R\to+\infty} S^R = 0, \qquad S^R = (W_+^R)^* W_-^R. \quad (5.3)$$

On the other hand, it has been shown by several authors[17,18] that the distorted waves and the S-matrix calculated from the off-energy-shell Lippmann-Schwinger equations for the cut-off Coulomb potential converge to the Coulomb distorted wave and the Coulomb T matrix if a certain "renormalization" energy dependent factor is divided out. Using the equations derived in Sec. 2, we shall show that when the energy-shell limit is taken in the solution to the cut-off Lippmann–Schwinger equations we obtain zero at any fixed energy if the limit $R \to +\infty$ is taken first; a zero result is obtained, however, also if the limit $R \to +\infty$ is taken last provided (since the point-wise limit does not exist) that a smearing in the energy (achieved by integration in k) after multiplication with an arbitrary function on the energy spectrum is first carried out. A similar result has been obtained via another approach in the special case of the pure Coulomb T matrix.[19,20]

We first note that the second resolvent equation

$$\frac{1}{H_R - \zeta} = \frac{1}{H_0 - \zeta} - \frac{1}{H_0 - \zeta}V_R\frac{1}{H_R - \zeta} \quad (5.4)$$

holds for all values of $R \geq 0$ when ζ is in the resolvent set of $H_R = H_0 + V_R$. Hence the Green's function $G_R(\mathbf{r}, \mathbf{r}'; \zeta)$ of H_R satisfies the same type of integral equation

$$G_R(\mathbf{r}, \mathbf{r}'; \zeta) = G_0(\mathbf{r}, \mathbf{r}'; \zeta) - \int_{\mathbb{R}^3} G_0(\mathbf{r}, \mathbf{r}''; \zeta)V_R(\mathbf{r}'')G_R(\mathbf{r}'', \mathbf{r}'; \zeta)d\mathbf{r}'' \quad (5.5)$$

for all $R \geq 0$ including $R = \infty$.

Lemma 5.1: Suppose $\lim_{R\to\infty} G_R(\mathbf{r}, \mathbf{r}'; \zeta)$ exists and $G_R(\mathbf{r}, \mathbf{r}'; \zeta)$ is bounded by a function $f(\mathbf{r}, \mathbf{r}'; \zeta)$ which is square integrable in each neighborhood of $\mathbf{r}'$ and bounded by a constant in the complement of each neighborhood of $\mathbf{r}'$. Then for almost all $\mathbf{r}$ and $\mathbf{r}'$

$$\lim_{R\to\infty} G_R(\mathbf{r}, \mathbf{r}'; \zeta) = G(\mathbf{r}, \mathbf{r}'; \zeta). \quad (5.6)$$

Furthermore, if $|\Phi(\mathbf{r})| \leq$ const for large $|\mathbf{r}|$ (such as in the case of plane waves or spherical waves) and $f(\mathbf{r}, \mathbf{r}'; \zeta)$ is square integrable then

$$\lim_{R\to\infty} \int d\mathbf{r}' G_R(\mathbf{r}, \mathbf{r}'; \zeta) V_R(\mathbf{r}')\Phi(\mathbf{r}') = \int d\mathbf{r}' G(\mathbf{r}, \mathbf{r}'; \zeta) V(\mathbf{r}')\Phi(\mathbf{r}'). \quad (5.7)$$

Proof: The statement (5.6) follows from Eq. (5.5) by application of the Lebesgue dominated convergence theorem and from the uniqueness of the solutions of Eq. (5.5). The application of the Lebesgue dominated convergence theorem also yields (5.7).

From Lemma 3.1 and an analogous argument as given in the derivation of Eq. (2.20) we conclude

$$\Phi_{\mathbf{k}}(\mathbf{r}) = \lim_{\zeta\to\mathbf{k}^2/2m} \int d\mathbf{r} G(\mathbf{r}, \mathbf{r}'; \zeta) V(\mathbf{r}')\Phi_{\mathbf{k}}(\mathbf{r}'). \quad (5.8)$$

Thus if we consider the off-energy shell Lippmann-Schwinger equations for the screened potential,

$$\Phi^{\pm}_{\mathbf{k},R}(\mathbf{r}; \zeta) = \Phi_{\mathbf{k}}(\mathbf{r}) - \int_{\mathbb{R}^3} d\mathbf{r}' G_R(\mathbf{r}, \mathbf{r}'; \zeta) V_R(\mathbf{r}')\Phi_{\mathbf{k}}(\mathbf{r}') \quad (5.9)$$

we conclude by the above lemmas that

$$\lim_{\zeta\to\mathbf{k}^2/2m} \lim_{R\to+\infty} \Phi^{\pm}_{\mathbf{k},R}(\mathbf{r}; \zeta) = 0. \quad (5.10)$$

This result implies that for long-range potentials which satisfy Lemmas 3.1 and 5.1 the approach in which one starts with a cut-off potential and the corresponding off-energy shell Lippmann-Schwinger equations, and then removes the cut-off by first taking the limit $R \to +\infty$ and then the energy shell limit, will not lead to the correct result for the distorted wave. Due to (5.3) we arrive at the same conclusion when the limit $R \to +\infty$ is taken last.

We now examine (3.22) to see exactly how the equations involving the operator Z^{Δ}_{+} overcome the above problem. For asymptotically Coulomb interactions we will obtain explicitly the operators Z^{Δ}_{+}, $\Delta = [k^2 - \eta, k^2 + \eta]$, in configuration space and examine the Eq. (3.22) as $\eta \to +0$. This will provide insight as to the relationship of (3.22) to the standard Lippmann-Schwinger equations. The behaviour of each term of (3.22) is given in the subsequent three lemmas.

To simplify the notation we shall set $\lambda = k^2$ in (4.1), i.e., in the sequel λ is a variable in the spectrum of $2mH_0$ rather than of H_0. This means that the operators $2mH_0$ and $2mH$ now assume the roles played before by the operators H_0 and H, respectively.

Lemma 5.2: Let $Z^{(\eta)}_{+}$ be the operator defined in Sec. 4 for asymptotically Coulomb potentials corresponding to the interval $\Delta = [k^2 - \eta, k^2 + \eta]$, with $\eta > 0$, in the spectrum of $2mH_0$ and let U be the unitary[5] transformation

$$(U\psi)(k, l, m) = \text{l.i.m.} \int \overline{\Phi_{klm}(\mathbf{r})}\psi(\mathbf{r})d\mathbf{r},$$

$$\Phi_{klm}(\mathbf{r}) = \left(\frac{2k^2}{\pi}\right)^{1/2}(-i)^l j_l(k\gamma)Y^m_l\left(-\frac{\mathbf{r}}{\gamma}\right)$$

of $L^2(R^3)$ onto $\mathfrak{L}^2(\sigma_0)$ [we are taking $\alpha = (l, m)$ in (3.1)].

Then we have

$$(U^{-1}Z^{(\eta)}_{+}U\Phi_{klm})(\mathbf{r}) = P_{klm}(\mathbf{r}; \eta) + Q_{klm}(\mathbf{r}; \eta) \quad (5.11)$$

where $Q_{klm}(\mathbf{r}; \eta)$ is given in (A2)-(A6) and converges to zero as $\eta \to +0$, while

$$P_{klm}(\mathbf{r}; \eta) = \frac{k}{2\pi q_0}\left[\Phi_{(k^2+\xi)^{1/2}lm}(\mathbf{r})[4(k^2 + \xi)]^{-iq_0/(k^2+\xi)^{1/2}} \times \Gamma\left(1 - \frac{iq_0}{(k^2 + \xi)^{1/2}}\right)\exp\left(\frac{q_0}{k^2 + \xi)^{1/2}} \times [-(\pi/2)\,\mathrm{sgn}\xi + i\log|\xi|]\right)\right]^{\xi=+\eta}_{\xi=-\eta}. \quad (5.12)$$

Proof In terms of U, U^{-1} and (4.1) we have

$$(U^{-1}Z^{(\eta)}_{+}U\Phi_{klm})(\mathbf{r}) = \frac{1}{2\pi i}\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk'\Phi_{k'lm}(\mathbf{r}) \times \exp\left(-\frac{\pi q_0}{2k'}\right)(2k')^{-2iq_0/k'}\Gamma\left(1 - \frac{iq_0}{k'}\right) \times \lim_{\gamma\to+0}\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk''2k'' \exp\left[\left(\frac{iq_0}{k'} - 1\right) \times \log(k'^2 - k''^2 - i\gamma)\right]\left\{\int \overline{\Phi_{k''lm}(\mathbf{r}')}\Phi_{klm}(\mathbf{r}')r'^2 \times \sin\theta dr'd\theta d\phi\right\},$$

where here, as well as in the sequel, we choose the branch of the logarithmic function with $-3\pi/2 < \arg(k'^2 - k''^2 - i\gamma) \leq \pi/2$. The normalization of $\Phi_{klm}(\mathbf{r})$ is so chosen that the term in curly brackets is equal to $\delta(k - k'')$. To do the computation we shall replace that term by the function $(2\rho)^{-1}\chi_{[k-\rho,k+\rho]}(k'')$ (this amounts to "smearing" $\Phi_{klm}(\mathbf{r})$ so as to obtain a function of $\mathbf{r}$ which belongs to $L^2(\mathbf{R}^3)$ and on which $U^{-1}Z^{(\eta)}_{+}U$ is well defined). After integrating in k'' and taking the limit $\gamma \to +0$, we obtain

$$-\frac{1}{2\pi q_0}\left(\frac{1}{2\rho}\right)\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk'k'\Phi_{k'lm}(\mathbf{r})\exp\left(-\frac{\pi q_0}{2k'}\right) \times (2k')^{-2iq_0/k'}\Gamma\left(1 - \frac{iq_0}{k'}\right)\left[\exp\left(\frac{iq_0}{k'}\log(k'^2 - (k + \rho)^2)\right) - \exp\left(\frac{iq_0}{k'}\log(k^2 - (k - \rho)^2)\right)\right]. \quad (5.13)$$

To complete the proof we integrate by parts in (5.13). This allows us to take the limit $\rho \to +0$ and then arrive at (5.11) and (5.12). This is done explicitly in the Appendix.

Lemma 5.3: Assume that

$$\int_{\mathbb{R}^3} d\mathbf{r}'|G(\mathbf{r}, \mathbf{r}'; k^2 \pm i\epsilon)V(\mathbf{r}')\Phi_{k'lm}(\mathbf{r}')| \leq M_{k'lm}(\mathbf{r}; k) \quad (5.14)$$

for all $\epsilon > 0$, where $M_{k'lm}(\mathbf{r}; k)$ as a function in k' belongs to $L^1([(k^2 - \eta)^{1/2}, (k^2 + \eta)^{1/2}])$ for all $\eta > 0$. Then for $Q_{klm}(\mathbf{r}'; \eta)$ and $Q^{(1)}_{klm}(\mathbf{r}; \eta)$ defined in (A2) and (A6), we have

$$\lim_{\eta\to+0}\int_{\mathbb{R}^3} d\mathbf{r}'G(\mathbf{r}, \mathbf{r}'; k^2 \pm i\epsilon) \times V(\mathbf{r}')[Q_{klm}(\mathbf{r}'; \eta) - Q^{(1)}_{klm}(\mathbf{r}'; \eta)] = 0 \quad (5.15)$$

uniformly with respect to all $\epsilon > 0$.

Proof: From (A3)-(A5) we see that $Q_{klm}(\mathbf{r}; \eta) - Q^{(1)}_{klm}(\mathbf{r}; \eta)$ is of the form

$$p_0(k, k'; \eta)\Phi_{k'lm}(\mathbf{r})\Big|^{k'=(k^2+\eta)^{1/2}}_{k'=(k^2-\eta)^{1/2}} + \int_{k^2-\eta}^{k^2+\eta} d(k'^2)p_1(k, k')\Phi_{k'lm}(\mathbf{r}),$$

where the contributions to $p_0(k,k';\eta)$ originate from the first term in (A4) and from integrating by parts the term in the integrand in (A4) containing the $df_0/d(k'^2)$ factor (after the differentiation in ρ followed by setting $\rho = 0$ had been carried out in both cases). A detailed analysis shows that an η factor can be extracted from $p_0(k,k';\eta)$ leaving the function p_0/η which is uniformly bounded in η. Hence this term behaves in accordance with the statement of the lemma.

To deal with the term containing p_1, we apply first Tonelli's and then Fubini's theorem to interchange the order of integration in k' and $\mathbf{r}'$. This application is warranted by (5.14) and the fact that p_1 is bounded in $k' \in [(k^2-\eta)^{1/2}, (k^2+\eta)]$ when $\eta < k^2$. The desired result follows by using (5.14) to majorize the integrand of the resulting integral in k'.

The assumption (5.14) of the above lemma is true for the case of the pure Coulomb potential. This can easily be seen by using the explicit expression[22] of the Green's function. Hence, it can be expected that due to the second resolvent equation relating the Green's functions for the Coulomb-like and pure Coulomb cases, this assumption is also satisfied by any Coulomb-like potential.

Lemma 5.4: If $Z_+^{(\eta)}$ is as in Lemma 5.2, then

$$[U^{-1}(H_0 Z_+^{(\eta)} - Z_+^{(\eta)} H_0) U\Phi_{klm}](\mathbf{r}) = \frac{-k}{2\pi i m}\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk'\,\Phi_{k'lm}(\mathbf{r})(2k')^{-2iq_0/k'} \times \Gamma\left(1-\frac{iq_0}{k'}\right)\exp\left(-\frac{\pi q_0}{2k'}+\frac{iq_0}{k'}\log(k'^2-k^2)\right). \tag{5.16}$$

Proof: Adopting the same procedure as in Lemma 5.2, we get

$$[U^{-1}(H_0 Z_+^{(\eta)} - Z_+^{(\eta)} H_0) U\Phi_{klm}](\mathbf{r}) = \frac{1}{4\pi i m}\lim_{\rho\to+0}\left(\frac{1}{2\rho}\right)\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk'\,\Phi_{k'lm}(\mathbf{r})(2k')^{-2iq_0/k'} \times \Gamma\left(1-\frac{iq_0}{k'}\right)\exp\left(-\frac{\pi q_0}{2k'}\right) \times \lim_{\sigma\to+0}\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} (k'^2-k''^2-i\sigma)^{(iq_0/k')-1} \times (k'^2-k''^2)\chi_{[k-\rho,\,k+\rho]}(k'')2k''dk''.$$

We note that we can replace the $(k'^2-k''^2)$ factor by $(k'^2-k''^2-i\sigma)$ without changing the value of the limit as $\sigma\to+0$. Then the integration in $(k'')^2$ can be explicitly performed and afterwards the limit $\sigma\to+0$ can be computed to obtain

$$\frac{1}{4\pi i m}\lim_{\rho\to+0}\left(\frac{1}{2\rho}\right)\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk'\,\Phi_{k'lm}(\mathbf{r})(2k')^{-2iq_0/k'} \times \Gamma\left(1-\frac{iq_0}{k'}\right)\exp\left(-\frac{\pi q_0}{2k'}\right)\left(\frac{iq_0}{k'}+1\right)^{-1}\left\{\exp\left[\left(\frac{iq_0}{k'}+1\right)\times\log[k'^2-(k+\rho)^2]\right] - \exp\left[\left(\frac{iq_0}{k'}+1\right)\log[k'^2-(k-\rho)^2]\right]\right\}.$$

Noting that the above limit in $\rho\to+0$ represents differentiation with respect to ρ under the integral sign (cf. Ref. 21, p. 124) we arrive at (5.16).

We can now apply the preceding three technical lemmas to the study of the expression

$$\Phi_{klm}^{\pm\epsilon}(\mathbf{r};\eta) = (U^{-1}Z_\pm^{(\eta)}U\Phi_{klm})(\mathbf{r}) - \int_{\mathbb{R}^3} d\mathbf{r}'G(\mathbf{r},\mathbf{r}';k^2\mp i\epsilon)V(\mathbf{r}')(U^{-1}Z_\pm^{(\eta)}U\Phi_{klm})(\mathbf{r}') - \int_{\mathbb{R}^3} d\mathbf{r}'G(\mathbf{r},\mathbf{r}';k^2\mp i\epsilon)(U^{-1}[H_0,Z_\pm^{(\eta)}]U\Phi_{klm})(\mathbf{r}'). \tag{5.17}$$

Comparison with (3.17) and (3.22) shows that

$$\Phi_{klm}^{\pm}(\mathbf{r}) = (W_\pm^{(r)}\Phi_{klm})(\mathbf{r}) = \{W_\pm^{(r)}E^{2mH_0}([(k^2-\eta)^{1/2},(k^2+\eta)^{1/2}])\Phi_{klm}\}(\mathbf{r}) = \lim_{\epsilon\to+0}\Phi_{klm}^{\pm\epsilon}(\mathbf{r};\eta)$$

i.e., the limit of (5.17) for $\epsilon\to+0$ (which exist at least in the weak sense of distributions) leads to the η-independent distorted wave for the long-range potential.

By applying Lemma 5.2 to the first two terms on the right-hand side we arrive at the conclusion that their contribution is of the form

$$F_k^{\pm}(\eta)\left(\Phi_{(k^2+\eta)^{1/2}lm}(\mathbf{r}) - \int d\mathbf{r}'G(\mathbf{r},\mathbf{r}';k^2\mp i\epsilon)V(\mathbf{r}')\Phi_{(k^2-\eta)^{1/2}lm}(\mathbf{r}')\right) - F_k^{\pm}(-\eta)\left(\Phi_{(k^2-\eta)^{1/2}lm}(\mathbf{r}) - \int d\mathbf{r}'G(\mathbf{r},\mathbf{r}';k^2\mp i\epsilon)V(\mathbf{r}')\Phi_{(k^2-\eta)^{1/2}lm}(\mathbf{r}')\right) + N_{klm}^{\pm\epsilon}(\mathbf{r};\eta), \tag{5.18}$$

where $N_{klm}^{\pm\epsilon}(\mathbf{r};\eta)\to 0$ as $\eta\to+0$ and by (5.12)

$$F_k^{+}(\xi) = \overline{F_k^{-}(\xi)} = \frac{k}{2\pi q_0}[4(k^2+\xi)]^{-iq_0/(k^2+\xi)^{1/2}} \times \Gamma\left(1-\frac{iq_0}{(k^2+\xi)^{1/2}}\right)\exp\left(-\frac{\pi q_0\,\mathrm{sgn}\,\xi}{2(k^2+\xi)^{1/2}} + \frac{iq_0}{(k^2+\xi)^{1/2}}\log|\xi|\right), \tag{5.19}$$

where ξ takes the values $\pm\eta$. Since at fixed $\epsilon>0$ the last term in (5.17) obviously vanishes in the limit $\eta\to+0$, we arrive at a conclusion concurrent with that of other authors,[17,18,23] namely that off the energy shell the distorted wave for the long-range case behaves like the distorted wave for the theory with cut-off at R in the limit $R\to\infty$ [cf. the terms in large parentheses in (5.18)] times the oscillating factors $F_k(\pm\eta)$.

The picture changes, however, if we take the limit $\epsilon\to+0$ in (5.17) for the purpose of computing the distorted spherical wave $\Phi_{klm}^{\pm}(\mathbf{r})$. In analyzing the behavior of the two terms in the large parentheses of (5.18) we cannot apply the result of Sec. 2 since the free spherical waves $\Phi_{k'lm}$ which occur correspond to the energies $k' = (k^2\pm\eta)^{1/2}$ rather than k. Since no estimates uniform in ϵ could be derived for the difference between these two cases, we shall assume for the moment that we are dealing with a pure Coulomb case. In this case we can resort to a formula by Ford [cf. Ref. 17, Eq. (75)] to compute the terms in large parentheses in (5.18), thus obtaining[24]

$$\Phi_{(k^2+\xi)^{1/2}lm}(\mathbf{r}) - \int d\mathbf{r}' G(\mathbf{r},\mathbf{r}';k^2 \pm i\epsilon) V(\mathbf{r}') \Phi_{(k^2+\xi)^{1/2}lm}(\mathbf{r}')$$
$$= \exp\left(\frac{iq_0}{(k^2 \pm i\epsilon)^{1/2}} \log(\xi \mp i\epsilon)\right)$$
$$\times [4(k^2+\xi)^{-iq_0/(k^2\pm i\epsilon)^{1/2}} h(k,\mathbf{r};\pm\epsilon), \tag{5.20}$$

where

$$h(k,\mathbf{r};\pm\epsilon) = \frac{1}{2}\exp\left(-\frac{\pi q_0}{2(k^2 \pm i\epsilon)^{1/2}}\right)\Gamma^{1/2}\left(1 - \frac{iq_0}{2(k^2 \pm i\epsilon)^{1/2}}\right)$$
$$\times \Gamma^{1/2}\left(1 + \frac{iq_0}{2(k^2 \pm i\epsilon)^{1/2}}\right)\Phi^{\pm}_{(k^2\pm i\epsilon)^{1/2}lm}(\mathbf{r})$$

and $\Phi^{\pm}_{(k^2\pm i\epsilon)^{1/2}lm}(\mathbf{r})$ converges to the Coulomb distorted spherical wave as $\epsilon \to +0$. Using the above result we see that the limit $\epsilon \to +0$ of (5.18) is well defined and yields

$$F^{\pm}_k(\eta)\exp\left(\mp\frac{iq_0}{k}\log|\eta|\right)[4(k^2+\eta)]^{\pm iq_0/k} h(k,\mathbf{r};\mp 0)$$
$$- F^{\pm}_k(-\eta)\exp\left(\mp\frac{iq_0}{k}(\log|\eta| \mp i\pi)\right)$$
$$\times [4(k^2-\eta)]^{\pm iq_0/k} h(k,\mathbf{r};\mp 0). \tag{5.21}$$

Due to the form of $F^{\pm}_k(\xi)$ given by (5.19) we conclude that in the limit $\eta \to +0$ we obtain zero:

$$\lim_{\eta\to+0}\frac{k}{2\pi q_0}\left\{[4(k^2+\eta)]^{\mp iq_0/(k^2+\eta)^{1/2}}\right.$$
$$\times [4(k^2+\eta)]^{\pm iq_0/k}\Gamma\left(1 \mp \frac{iq_0}{(k^2+\eta)^{1/2}}\right)$$
$$\times \exp\left(-\frac{\pi q_0}{2(k^2+\eta)^{1/2}} + iq_0\log|\eta|\left\{\pm\frac{1}{(k^2+\eta)^{1/2}} \mp \frac{1}{k}\right\}\right)$$
$$\times h(k,\mathbf{r};\mp 0) - [4(k^2-\eta)]^{\mp iq_0/(k^2-\eta)}[4(k^2-\eta)]^{\pm iq_0/k}$$
$$\times \Gamma\left(1 \mp \frac{iq_0}{(k^2-\eta)^{1/2}}\right)\exp\left\{\left(\frac{\pi q_0}{2(k^2-\eta)^{1/2}} - \frac{\pi q_0}{k}\right)\right.$$
$$\left.\left. + iq_0\log|\eta|\left\{\pm\frac{1}{(k^2-\eta)} \mp \frac{1}{k}\right\}\right\} h(k,\mathbf{r};\mp 0)\right\}. \tag{5.22}$$

Now if we examine (5.17) and take first the limit $\epsilon \to +0$ and then $\eta \to +0$, we conclude from (5.22) and Lemma 5.3 that the terms which contribute to the Coulomb distorted wave are those containing the expression $Q^{(1)}$ given in (A2) and the commutator term in (5.17). Thus

$$\Phi^{\pm}_{klm}(\mathbf{r}) = \lim_{\eta\to+0}\lim_{\epsilon\to+0}\left\{\left(\frac{k}{2m\pi i}\right)\right.$$
$$\times \int_{\mathbb{R}^3} d\mathbf{r}' G(\mathbf{r},\mathbf{r}';k^2 \mp i\epsilon)\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk'$$
$$\times \Phi_{k'lm}(\mathbf{r}')(2k')^{-2iq_0/k'}\Gamma\left(1 - \frac{iq_0}{k'}\right)$$
$$\times \exp\left[\frac{q_0}{k'}\left(-\frac{\pi}{2} + i\log(k'^2 - k^2)\right)\right]$$
$$+ \left(\frac{k}{2\pi q_0}\right)\int_{\mathbb{R}^3} d\mathbf{r}' G(\mathbf{r},\mathbf{r}';k^2 \mp i\epsilon)$$
$$\times V(\mathbf{r}')\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}} dk' \frac{\partial\Phi_{k'lm}(\mathbf{r}')}{\partial k'}$$
$$\times (2k')^{-2iq_0/k'}\Gamma\left(1 - \frac{iq_0}{k'}\right)$$
$$\left.\times \exp\left[\frac{q_0}{k'}\left(-\frac{\pi}{2} + i\log(k'^2 - k^2)\right)\right]\right\}, \tag{5.23}$$

where the above limits are to be taken only after a smearing in the energy is first carried out. Hence on the energy shell it is the term containing $[H_0, Z^{(\eta)}_+]$ and a term from the integral of the product of the Green's function, potential and free spherical wave which contribute most to the distorted Coulomb wave $\Phi^{\pm}_{klm}(\mathbf{r})$ if $\eta > 0$ is sufficiently small.

While the above result is valid only for the pure Coulomb potential we expect that it is true also for Coulomb-like potentials. We should point out now that the fact that (5.10) is true (in the weak sense of distributions) regardless of the order in which the two limits $R \to +\infty$ and $\zeta \to \mathbf{k}^2/2m$ are taken is due to the presence of cutoff dependent factors in $\Phi^{\pm}_{\mathbf{k},R}$ which oscillate as $R \to +\infty$. These factors were first extracted in the pure Coulomb case,[18] but they are actually present in the Coulomb-like case.[25] We show elsewhere[25] that once these factors are extracted from the distorted wavefunction corresponding to the cutoff case at real energies (i.e., after the limit $\zeta \to \mathbf{k}^2/2m$ has been taken, then we can remove the cutoff (i.e., take the limit $R \to \infty$) and recover the distorted wavefunction for the Coulomb-like case. This procedure turns out to be valid not only for the two-body case but also for the general n-body case in all channels.[25]

Finally, let us point out that the above considerations extend immediately to the T matrix since on the energy-shell for any $\eta > 0$ [cf. (3.20)]

$$(klm|\,T\,|kl'm') = \int_{\mathbb{R}^3}\overline{\Phi_{klm}(\mathbf{r})}(Z^{(\eta)*}V)(\mathbf{r})\Phi^{-}_{kl'm}(\mathbf{r})d\mathbf{r}$$
$$+ \int_{\mathbb{R}^3}\overline{\Phi_{klm}(\mathbf{r})}([Z^{(\eta)*}, H_0]\Phi^{-}_{kl'm})(\mathbf{r})d\mathbf{r}$$

[naturally, the above expression contains a $\delta_{ll'}$ factor if $V(\mathbf{r})$ is spherically symmetric]. Note that the normalization of the spherical waves $\Phi_{klm}(\mathbf{r})$ was chosen in Lemma 5.2 in such a manner that the T operator in (1.8) is related to the above T matrix by

$$(Tf)(k,l,m) = \sum_{l'=0}^{\infty}\sum_{m'=-l}^{+l}(klm|\,T\,|kl'm')f(k,l',m'),$$

if we work in the space $\bigoplus_{l=0}^{\infty}\bigoplus_{m=-l}^{+l} L^2([0,\infty))$, i.e., if we let T act on f with $f(k,l,m) \in L^2([0,\infty))$ for each (l,m). It should be emphasized that the above T matrix has the usual physical interpretation despite the long-range nature of the potential $V(\mathbf{r})$ since the adopted choice of $G(H_0;t)$ in (1.2) makes the renormalized wave-operators $W^{(r)}$ physical[4] and hence preserves (cf. Ref. 4) the customary relationship between the T matrix and differential cross section.

APPENDIX

We make the following substitution in (5.13)

$$\exp\left(\frac{iq_0}{k'}\log(k'^2 - (k \pm \rho)^2)\right)$$
$$= \left(\frac{iq_0}{k'} + 1\right)^{-1}\left\{\frac{d}{d(k'^2)}\exp\left[\left(\frac{iq_0}{k'} + 1\right)\log(k'^2 - (k \pm \rho)^2)\right]\right.$$

$$+\frac{iq_0}{2k'^3}\log(k'^2-(k\pm\rho)^2)$$
$$\times\exp\left[\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k\pm\rho)^2)\right]\Bigg\}$$

and then integrate the first term by parts. This gives the following expression

$$-\frac{1}{\pi q_0}\left(\frac{1}{2\rho}\right)\Bigg(\Phi_{k'lm}(\mathbf{r})\exp\left(-\frac{\pi q_0}{2k'}\right)(2k')^{-2iq_0/k'}\Gamma\left(1-\frac{iq_0}{k'}\right)$$
$$\times\left(\frac{iq_0}{k'}+1\right)^{-1}\Bigg\{\exp\left[\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k+\rho)^2)\right]$$
$$-\exp\left[\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k-\rho)^2)\right]\Bigg\}\Bigg)\Bigg|_{k'=(k^2+\eta)^{1/2}}^{k'=(k^2-\eta)^{1/2}}$$
$$+\frac{1}{4\pi q_0}\left(\frac{1}{2\rho}\right)\int_{k^2-\eta}^{k^2+\eta}d(k'^2)$$
$$\times\Bigg\{\exp\left[\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k+\rho)^2)\right]$$
$$-\exp\left[\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k-\rho)^2)\right]\Bigg\}\frac{d}{d(k'^2)}$$
$$\times\Bigg[\Phi_{k'lm}(\mathbf{r})\exp\left(-\frac{\pi q_0}{2k'}\right)$$
$$\times(2k')^{-2iq_0/k'}\Gamma\left(1-\frac{iq_0}{k'}\right)\left(\frac{iq_0}{k'}+1\right)^{-1}\Bigg]$$
$$-\frac{i}{8\pi}\left(\frac{1}{2\rho}\right)\int_{k^2-\eta}^{k^2+\eta}d(k'^2)\left(\frac{1}{k'}\right)\Phi_{k'lm}(\mathbf{r})$$
$$\times\exp\left(-\frac{\pi q_0}{2k'}\right)(2k')^{-2iq_0/k'}$$
$$\times\Gamma\left(1-\frac{iq_0}{k'}\right)\left(\frac{iq_0}{k'}+1\right)^{-1}\Bigg\{\log(k'^2-(k+\rho)^2)$$
$$\times\exp\left[\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k+\rho)^2)\right]$$
$$-\log(k'^2-(k-\rho)^2)\exp\left[\left(\frac{iq_0}{k'}+1\right)\right.$$
$$\left.\times\log(k'^2-(k-\rho)^2)\right]\Bigg\}. \tag{A1}$$

We can now take the limit $\rho\to+0$ explicitly in the first term of the above expression. This yields the expression for $P_{klm}(\mathbf{r};\eta)$ in (5.12). The limits $\rho\to+0$ in the second and third terms of (A1) are equivalent to differentiating the integral with respect to the parameter ρ. Differentiating the integral with respect to the parameter ρ is justified if the derivative in ρ of the integrand can be bound by an integrable function which is independent of ρ (cf. Ref. 21, p. 124). For the second term of (A1) this is easily verified to be true. Hence, by differentiating that term we obtain $Q^{(1)}_{klm}(\mathbf{r};\eta)+Q^{(2)}_{klm}(\mathbf{r};\eta)$, where

$$Q^{(1)}_{klm}(\mathbf{r};\eta)=-\frac{k}{2\pi q_0}\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}}dk'$$
$$\times\left(\frac{\partial}{\partial k'}\Phi_{k'lm}(\mathbf{r})\right)\exp\left(\frac{iq_0}{k'}\log(k'^2-k^2)\right)$$
$$\times\exp\left(-\frac{\pi q_0}{2k'}\right)(2k')^{-2iq_0/k'}\Gamma\left(1-\frac{iq_0}{k'}\right), \tag{A2}$$

$$Q^{(2)}_{klm}(\mathbf{r};\eta)=-\frac{k}{2\pi q_0}\int_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}}$$
$$\times dk'\Phi_{k'lm}(\mathbf{r})\left(\frac{iq_0}{k'}+1\right)\exp\left(\frac{iq_0}{k'}\log(k'^2-k^2)\right)$$
$$\times\frac{d}{dk'}\left[\exp\left(-\frac{\pi q_0}{2k'}\right)(2k')^{-2iq_0/k'}\left(\frac{iq_0}{k'}+1\right)^{-1}\right.$$
$$\left.\times\Gamma\left(1-\frac{iq_0}{k'}\right)\right]. \tag{A3}$$

Differentiation with respect to ρ of the integrand in the third term in (A1) does not yield a function which can be bounded by a ρ-independent Lebesgue integrable function of $k'\in[(k^2-\eta)^{1/2},(k^2+\eta)^{1/2}]$. Hence, in order to take the limit $\rho\to+0$ of the third term of (A1) we integrate that term by parts with respect to the function $\log[k'^2-(k\pm\rho)^2]$. This yields the following expression:

$$\sum_{\pm\rho}\left(\frac{1}{\pm 2\rho}\right)f_0(k')\Bigg\{\Bigg(\log[k'^2-(k\pm\rho)^2]-1\Bigg)$$
$$\times\exp\left[\left(\frac{iq_0}{k'}+2\right)\log[k'^2-(k\pm\rho)^2]\right]\Bigg\}\Bigg|_{(k^2-\eta)^{1/2}}^{(k^2+\eta)^{1/2}}$$
$$-\sum_{\pm\rho}\left(\frac{1}{\pm 2\rho}\right)\int_{k^2-\eta}^{k^2+\eta}d(k'^2)\Bigg\{\{\log[k'^2-(k\pm\rho)^2]-1\}$$
$$\times\exp\left[\left(\frac{iq_0}{k'}+2\right)\log(k'^2-(k\pm\rho)^2)\right]$$
$$\times\frac{df_0(k')}{d(k'^2)}-\frac{iq_0}{2k'^3}\{\log[k'^2-(k\pm\rho)^2]-1\}$$
$$\times\log[k'^2-(k\pm\rho)^2]\exp\left[\left(\frac{iq_0}{k'}+2\right)\right.$$
$$\left.\times\log(k'^2-(k\pm\rho)^2)\right]f_0(k')-\left(\frac{iq_0}{k'}+1\right)$$
$$\times\{\log[k'^2-(k\pm\rho)^2]-1\}\exp\left[\left(\frac{iq_0}{k'}+1\right)\right.$$
$$\left.\times\log(k'^2-(k\pm\rho)^2)\right]f_0(k')\Bigg\}, \tag{A4}$$

where

$$f_0(k')=\frac{-i}{8\pi k'^3}\Phi_{k'lm}(\mathbf{r})\exp\left(-\frac{\pi q_0}{2k'}\right)(2k')^{-2iq_0/k'}$$
$$\times\Gamma\left(1-\frac{iq_0}{k'}\right)\left(1+\frac{iq_0}{k'}\right)^{-1}$$

and $\sum_{\pm\rho}$ refers to the sum of the two terms following the sign $\sum$ and corresponding to $(+\rho)$ and $(-\rho)$, respectively. The limit $\rho\to+0$ can be taken by Ref. 21 in each term of the integrand of the above expression except the last one. This term is of the same form as the integrand of the last term in (A1) multiplied by the factor $[-(iq_0/k')+1]$. If we add and subtract the following two expressions

$$\frac{iq_0}{(k\pm\rho)}\left(\frac{1}{2\rho}\right)\int_{k^2-\eta}^{k^2+\eta}d(k'^2)f_0(k')\log(k'^2-(k\pm\rho)^2)$$
$$\times\exp\left(\left(\frac{iq_0}{k'}+1\right)\log(k'^2-(k\pm\rho)^2)\right),$$

we see that the last term of (A1) is equal to

$$\pm\left(2+\frac{iq_0}{k\pm\rho}\right)^{-1}\Big\{g_0(k,\eta;\rho)+\frac{iq_0}{2\rho}\int_{k^2-\eta}^{k^2+\eta}d(k'^2)f_0(k')\times\left(\frac{k\pm\rho-k'}{k'(k\pm\rho)}\right)\log[k'^2-(k\pm\rho)^2]\times\exp\left[\left(\frac{iq_0}{k'}+1\right)\log[k'^2-(k\pm\rho)^2]\right]\Big\},\tag{A5}$$

where $g_0(k,\eta;\rho)$ consists of the terms in (A4) for which the limit $\rho\to+0$ can either be taken explicitly or by using the theorem in Ref. 21, p. 124. Now, we can apply this theorem to the second term of the integrand in (A5). Hence we conclude that the limit $\rho\to+0$ can be taken in (A5) and is obtained by differentiating the integrand with respect to ρ and afterwards setting $\rho=0$. Thus the limit $\rho\to+0$ of (A5) is given by a term $Q^{(3)}_{klm}(\mathbf{r};\eta)$ which goes to zero as $\eta\to+0$.

The above results imply that (5.11) is true, where

$$Q_{klm}(\mathbf{r};\eta)=Q^{(1)}_{klm}(\mathbf{r};\eta)+Q^{(2)}_{klm}(\mathbf{r};\eta)+Q^{(3)}_{klm}(\mathbf{r};\eta).\tag{A6}$$

*Supported in part by the National Research Council of Canada.

[1]V. S. Buslaev and V. B. Matveev, Teor. Mat. Fiz. **2**, 367 (1970).
[2]W. Amrein, Ph. Martin and B. Misra, Helv. Phys. Acta **43**, 313 (1970).
[3]R. Lavine, J. Funct. Anal. **5**, 368 (1970).
[4]E. Prugovečki, Nuovo Cimento B **4**, 105 (1971).
[5]E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971).
[6]E. Prugovečki, SIAM J. Math. Anal. **3**, 183 (1972).
[7]E. Prugovečki, Nuovo Cimento B **4**, 124 (1971).
[8]J. Dollard, J. Math. Phys. **7**, 802 (1966).
[9]G. West, J. Math. Phys. **8**, 942 (1967).
[10]E. Prugovečki, J. Math. Phys. **14**, 957 (1973).
[11]J. Dollard, J. Math. Phys. **5**, 729 (1964).
[12]J. Dollard, J. Math. Phys. **9**, 620 (1968).
[13]D. Marchesin and M. L. O'Carroll, J. Math. Phys. **13**, 982 (1972).
[14]E. B. Manoukian and E. Prugovečki, Can. J. Phys. **49**, 102 (1971).
[15]E. Prugovečki, J. Math. Phys., **14**, 1410 (1973).
[16]V. B. Matveev and M. M. Skriganov, Teor. Mat. Fiz. **10**, 238 (1972).
[17]W. F. Ford, J. Math. Phys. **7**, 626 (1966).
[18]V. G. Gorshkov, Sov. Phys.-JETP **13**, 1037 (1961).
[19]G. L. Nutt, J. Math. Phys. **9**, 796 (1968).
[20]W. W. Zazhary, J. Math. Phys. **12**, 1379 (1971).
[21]J. Dieudonné, *Treatise on Analysis* (Academic, New York, 1970).
[22]L. Hostler, J. Math. Phys. **5**, 591 (1964).
[23]S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960);
[24]Note that in Ref. 17 opposite conventions from our own are adopted for the resolvent $R_A(\zeta)$ of A [i.e., $(\zeta-A)^{-1}$ instead of $(A-\zeta)^{-1}$] and for the wave operators ($\Omega_\pm$ for $t\to\mp\infty$ instead of $t\to\pm\infty$).
[25]E. Prugovečki and J. Zorbas, "Many-body modified Lippmann-Schwinger equations for Coulomb-like potentials," (to be published).

The bra and ket formalism in extended Hilbert space

Eduard Prugovečki*

Department of Mathematics, University of Toronto, Toronto, Canada M5S 1A1
(Received 13 March 1973)

The concept of bra-vector is endowed with a precise mathematical meaning by taking a Hilbert-space $\mathcal{H}$ (whose elements are the ket-vectors) and embedding it into a larger bra-space $\mathcal{L}$. The space $\mathcal{L}$ is constructed by fitting together all the different spaces with "negative" norm corresponding to equipping operators D in $\mathcal{H}$. It is shown that the formal manipulations of Dirac's formalism become theorems on the resulting extended Hilbert space structure $\mathcal{H} \subset \mathcal{L}$. On $\mathcal{L}$ we define topologies and bra-adjoints of operators in $\mathcal{H}$ and investigate their properties. We show how these results can be used in deriving rigorous versions of the typical relations involving distorted waves in time-independent scattering theory. By applying this formalism to Fock space we obtain an extended Fock space framework suitable for the rigorous formulation of the concept of a field at a point in configuration or momentum space.

1. INTRODUCTION

In the present paper we adopt the theory of equipped Hilbert spaces[1] recently applied[2] to stationary quantum scattering theory as the basis of a general rigorous formulation of Dirac's bra and ket formalism.[3] There have been in recent years a number of attempts[4-12] (see the Refs. 7–9 for surveys of such work) in this direction, mostly based on rigged[6-11] or nested[12] Hilbert spaces $\Phi \subset \mathcal{H} \subset \Phi'$. The triplets constituting these spaces consist in general of a Hilbert space $\mathcal{H}$, a dense subset Φ (the "ket" space) of $\mathcal{H}$ which is endowed with a finer topology (usually countably normed) than the strong topology of $\mathcal{H}$ and is nuclearly embedded in $\mathcal{H}$, and of the dual Φ' of Φ (the "bra" space). Their scope in applications to quantum mechanics is limited by the relative intricacy of these topologies as well as by the fact that the generalized expansions of different noncommuting observables might lie[9] in different spaces Φ' if the corresponding operators do not have a common domain of essential self-adjointness.

An equipped Hilbert space $\mathcal{H}_+ \subset \mathcal{H} \subset \mathcal{H}_-$ also consists of a triplet in which $\mathcal{H}_+$ (called the space with positive norm[1] in the triplet) is a dense linear subset of $\mathcal{H}$, but it is supplied with an inner product which makes it into a Hilbert space and its embedding in $\mathcal{H}$ is quasinuclear.[13] This implies the existence of an operator D (to be called the *equipping* operator) which maps isometrically $\mathcal{H}_+$ onto $\mathcal{H}$ and has an inverse D^{-1} which is Hilbert–Schmidt as an operator from $\mathcal{H}$ to $\mathcal{H}$. When D is self-adjoint its extension $\mathbb{D}$ to $\mathcal{H}$ provides a unitary transformation of $\mathcal{H}$ onto a larger Hilbert space $\mathcal{H}_-$, $\mathcal{H}_- \supset \mathcal{H}$ (called the space with negative norm in the triplet $\mathcal{H}_+ \subset \mathcal{H} \subset \mathcal{H}_-$). The quasinuclearity of the embedding $\mathcal{H}_+ \to \mathcal{H}$ turns out to be a necessary as well as a sufficient condition[1] for having the space $\mathcal{H}_-$ contain generalized expansions for arbitrary self-adjoint operators in $\mathcal{H}$.

In most of Sec. 2 we present some of the basic features of the theory of equipped Hilbert spaces from a point of view which puts the onus on the equipping operator D rather than on $\mathcal{H}_+$. We require such an approach since one of the main goals we are after is to give a universal meaning to the concept of bra-adjoint $A^\dagger$ of arbitrary bounded or densely defined unbounded operators A. It is clear from the outset that this will not be possible if working with a fixed triplet $\mathcal{H}_+ \subset \mathcal{H} \subset \mathcal{H}_-$. By means of the procedure in Sec. 2 we assign a triplet $\mathcal{H}_D \subset \mathcal{H} \subset \mathcal{H}_D^\dagger$ to every equipping operator D. Then in Sec. 3 we show how to "fit together," so to speak, all such triplets. By fitting together spaces with negative norm $\mathcal{H}_D^\dagger$ we arrive at a larger space $\mathcal{L}$, which we call the bra-space.

The Hilbert space $\mathcal{H}$ can be identified with a linear subspace $\mathcal{L}_{\text{ket}}$ of $\mathcal{L}$ ($\mathcal{L}$ is not linear but vector operations can be defined for some of the elements of $\mathcal{L}$). We call the resulting ordered pair $\mathcal{L}_{\text{ket}} \subset \mathcal{L}$ an extended Hilbert space. The construction of this space is purely algebraic, but different topologies can be introduced, two of which are explicitly considered in Sec. 3. The inner product $\langle \cdot | \cdot \rangle$ on $\mathcal{H}$ is extended to a sesquilinear form $(\cdot | \cdot)$ in $\mathcal{L} \times \mathcal{L}_{\text{ket}}$, with respect to which the concept of bra-adjoint $A^\dagger$ of an operator A in $\mathcal{L}_{\text{ket}} \equiv \mathcal{H}$ is then defined. The key ingredient in the construction of $\mathcal{L}$ which enables the introduction of the concept of the bra-adjoint $A^\dagger$ for a much larger class of operator than it is possible when we restrict ourselves to a single equipped Hilbert space is contained in the fact that $\mathcal{L}$ is constructed out of sequences $\{f_n\}$ of vectors f_n in $\mathcal{H}$. Hence $\{A^* f_n\}$ can be always defined and although in general it does not belong to the same space with negative norm as $\{f_n\}$ did, it will belong to another such space, i.e., it will determine an element of $\mathcal{L}$. Since such a construction is not based directly on duality, it bypasses the restrictions usually imposed[4] on A in constructing $A^\dagger$. A slightly different construction of the bra-space is presented in Sec. 7.

In Sec. 4 we find necessary and sufficient conditions for existence of Green's functions in spectral representation spaces of complete sets of observables. We show in Sec. 5 how the introduced framework can be used to endow with precise mathematical meaning some of the key formal manipulations encountered in time-independent scattering theory for free and distorted waves. The last section is presented only as an illustration of the applicability of the present formalism to quantum field theory. It is clear, however, that in this context the scope of the present formalism extends beyond the physically uninteresting case of free quantum fields. We intend to elaborate on this point on a future occasion.

In conclusion we would like to emphasize that we have not attempted to attach any physical meaning to those bra-vectors which are not also ket-vectors. Thus, we adopt the attitude that the physical interpretation of the theory is based exclusively on "wave-packets" and that observables are associated with self-adjoint operators in the Hilbert space $\mathcal{H}$ (i.e., the ket-space). The objects in bra-space are only auxiliary mathematical objects which the extensive experience accumulated by physicists has established as being of great computational value.

2. EXPANSIONS IN GENERALIZED EIGENVECTORS OF A COMPLETE SET OF OBSERVABLES

Let $\mathcal{H}$ be an infinite-dimensional complex separable

Hilbert space with an inner product $\langle \cdot | \cdot \rangle$ which is linear in the right-hand side variable and determines the vector norm $\| \cdot \|$. We shall present in this section a rigorous theory of eigenfunction expansions for arbitrary self-adjoint operators in $\mathcal{K}$ which embodies all the essential features of Dirac's formalism. The starting point is the notion of equipment[1] for $\mathcal{K}$. Since we shall find it necessary to deal simultaneously with different equipments, we base this notion on first introducing an equipping operator D and then building the spaces with positive and negative norms corresponding to D, rather than on the diametrically opposed procedure adopted in Ref. 1, in which these spaces are introduced first and the existence of an equipping operator D is deduced afterwards. The advantage of the first approach will become apparent in the next two sections when we discuss the construction of a bra-space $\mathcal{L}$ and the existence of adjoints in $\mathcal{L}$ of operators acting in $\mathcal{K}$.

Lemma 2.1: Let D be a densely defined operator in $\mathcal{K}$ with range equal to $\mathcal{K}$ and having an inverse D^{-1} of bound equal to one. The domain of definition $\mathfrak{D}(D)$ of D is a separable Hilbert space $\mathcal{K}_D$ (called the *space with positive norm corresponding to* D) with inner product $\langle \cdot | \cdot \rangle_D$,

$$\langle f | g \rangle_D = \langle Df | Dg \rangle, \quad f, g \in \mathcal{K}_D, \tag{2.1}$$

and vector norm $\| \cdot \|_D$ for which

$$\| f \|_D \geq \| f \|, \quad f \in \mathcal{K}_D. \tag{2.2}$$

Proof: The functional $\langle \cdot | \cdot \rangle_D$ introduced in (2.1) is obviously a sesquilinear form in $\mathfrak{D}(D) \times \mathfrak{D}(D)$. If $f \neq \mathbf{0}$ then $\langle f | f \rangle_D = \langle Df | Df \rangle > 0$ since $\langle f | f \rangle_D = 0$ implies $Df = \mathbf{0}$ and consequently $f = D^{-1}(Df) = \mathbf{0}$. Thus $\langle \cdot | \cdot \rangle$ is an inner product on the vector space $\mathfrak{D}(D)$. This space provided with this inner product is a unitary space denoted by $\mathcal{K}_D$. To prove the completeness of this unitary space, consider a Cauchy sequence $f_1, f_2, \cdots$ in $\mathcal{K}_D$. Since

$$\| Df_m - Df_n \| = \| D(f_m - f_n) \| = \| f_m - f_n \|_D,$$

we conclude that $Df_1, Df_2, \cdots$ converges in the $\| \cdot \|$-norm to a limit $g \in \mathcal{K}$. In view of the fact that the range of D is $\mathcal{K}$, there is a vector $f \in \mathcal{K}_D$ for which $g = Df$. We have

$$\| f_n - f \|_D = \| Df_n - Df \| \to 0,$$

thus establishing that $\mathcal{K}_D$ is a Hilbert space.

The inequality (2.2) is an immediate consequence of the requirement that $\| D^{-1} \| = 1$:

$$\| f \| = \| D^{-1}(Df) \| \leq \| Df \| = \| f \|_D.$$

Finally, the separability of $\mathcal{K}_D$ follows from the separability of $\mathcal{K}$ since the existence in $\mathcal{K}$ of a countable dense set $\{f_1, f_2, \cdots\}$ implies that the countable set $\{D^{-1}f_1, D^{-1}f_2, \cdots\}$ is dense in $\mathcal{K}_D$.

In Ref. 1, p. 36, Berezanskiĭ proves the converse to the above lemma, namely that to any Hilbert space $\mathcal{K}$ which is a dense subset of $\mathcal{K}$ and for which (2.2) holds on all $f \in \mathcal{K}$ corresponds a positive-definite operator D_0 mapping $\mathcal{K}$ into $\mathcal{K}$ and having a left inverse D_0^{-1} of bound one. Naturally, once the existence of such a D_0 is established, any operator of the form $D = UD_0$ with U unitary can be substituted for D_0, and in general these substituted operators are not positive-definite.

Lemma 2.2: Let the adjoint D^* of D be a densely defined operator in $\mathcal{K}$ with range $\mathcal{K}$ and having an inverse D^{*-1} of bound one, and denote by $(\cdot | \cdot)_D$ the sesquilinear form

$$(f | g)_D = \langle D^{*-1}f | D^{*-1}g \rangle \tag{2.3}$$

on $\mathcal{K} \times \mathcal{K}$. This sesquilinear form is an inner product on $\mathcal{K}$, with the property

$$\| f \|_{\mathbb{D}^{-1}} = (f | f)_D^{1/2} \leq \| f \|, \quad f \in \mathcal{K}. \tag{2.4}$$

Let $\mathcal{K}_D^\dagger$ denote the completion[14] of $\mathcal{K}$ with respect to this inner product ($\mathcal{K}_D^\dagger$ is called[1] the *space with negative norm corresponding to* D). If D^* is considered as a mapping from $\mathcal{K}$ to $\mathcal{K}_D^\dagger$, then it can be extended uniquely to a unitary operator $\mathbb{D}$ mapping $\mathcal{K}$ onto $\mathcal{K}_D^\dagger$,

$$(\mathbb{D}f | \mathbb{D}g)_D = \langle f | g \rangle, \quad f, g \in \mathcal{K}. \tag{2.5}$$

$\mathcal{K}_D^\dagger$ is a separable Hilbert space.

Proof: The sesquilinear form $(\cdot | \cdot)_D$ is positive definite. In fact, if $\| f \|_{\mathbb{D}^{-1}} = 0$ then $D^{*-1}f = \mathbf{0}$. Since D^* has range $\mathcal{R}(D^*) = \mathcal{K}$, there is a vector $g \in \mathfrak{D}(D^*)$ for which $f = D^*g$. Thus $g = D^{*-1}D^*g = D^{*-1}f = \mathbf{0}$ and consequently $f = \mathbf{0}$.

To prove the existence and unitarity of $\mathbb{D}$ note that for any $f, g \in \mathcal{K}_D$

$$(D^*f | D^*g)_D = \langle D^{*-1}D^*f | D^{*-1}D^*g \rangle = \langle f | g \rangle.$$

Hence, by the extension principle,[14] D^* can be extended in a unique manner to an isometric linear mapping $\mathbb{D}$ of $\mathcal{K}$ into $\mathcal{K}_D^\dagger$. To show that the range of $\mathbb{D}$ is $\mathcal{K}_D^\dagger$, note that $\mathcal{K}$ is dense in $\mathcal{K}_D^\dagger$ and therefore for any $f \in \mathcal{K}_D^\dagger$ there is a sequence $D^*g_1, D^*g_2, \cdots \in \mathcal{K}$ converging in the $\| \cdot \|_{\mathbb{D}^{-1}}$-norm to f. Furthermore, as

$$\| g_m - g_n \| = \| D^{*-1}(D^*g_m - D^*g_n) \| = \| D^*g_m - D^*g_n \|_{\mathbb{D}^{-1}},$$

the sequence $g_1, g_2, \cdots \in \mathfrak{D}(D^*)$ converges in the $\| \cdot \|$-norm to a limit $g \in \mathcal{K}$. Hence $f = \mathbb{D}g$.

Finally, the separability of $\mathcal{K}_D^\dagger$ follows from the separability of $\mathcal{K}$ and the unitary equivalence of $\mathcal{K}$ and $\mathcal{K}_D^\dagger$ expressed by (2.5).

In the sequel we shall consider only operators D which together with their adjoints D^* satisfy the conditions stipulated in Lemma 2.1. We shall refer to the pair $(\mathcal{K}_D, \mathcal{K}_D^\dagger)$ by the name "*equipment* of $\mathcal{K}$ corresponding to D". The ordered triple $\mathcal{K}_D \subset \mathcal{K} \subset \mathcal{K}_D^\dagger$ is called an *equipped Hilbert space.*

In the preceding two lemmas we have shown that any choice of an operator D which satisfies the conditions stipulated in Lemma 2.1 but is not necessarily possitive-definite or even self-adjoint, leads to a particular choice of equipment. Due to the nature of the later considerations in which many different equipments of $\mathcal{K}$ are simultaneously considered, this approach is more convenient for the applications we have in mind than the approach advocated in Ref. 1, in which $\mathcal{K}$ is chosen first and D is constructed afterwards as a positive-definite operator. In fact, this positive definiteness is a restriction which is exceedingly cumbersome in constructing specific operators D well-suited to scattering theory.[2] Moreover, the present approach also emphasizes the generalized inner product[15] aspect of an equipped Hilbert space, in which $\mathcal{K}_D^\dagger$ is the enveloping space, $\mathcal{K}$ the nucleus and $\{\mathbb{D}\}$ the "adequate" family of operators

from $\mathcal{K}_D^\dagger$ to $\mathcal{K}$ determining the various topologies[15] in such spaces.

Theorem 2.1: Suppose that D and D^* satisfy the conditions of Lemmas 2.1 and 2.2. Then there is a unique sesquilinear form $(f|g\rangle_D$ defined for all $f \in \mathcal{K}_D^\dagger$ and all $g \in \mathcal{K}_D$ which is continuous in both arguments with respect to the norm topology of the respective Hilbert spaces and such that

$$(f|g\rangle_D = \langle f|g\rangle, \quad f \in \mathcal{K}, \quad g \in \mathcal{K}_D, \tag{2.6}$$

$$(\mathbb{D}f|g\rangle_D = \langle f|Dg\rangle, \quad f \in \mathcal{K}, \quad g \in \mathcal{K}_D, \tag{2.7}$$

for all $f \in \mathcal{K}$ and all $g \in \mathcal{K}_D$. Furthermore, the mapping $M: \mathcal{K}_D^\dagger \to \mathcal{K}_D'$ which takes an element f of $\mathcal{K}_D^\dagger$ into the functional $(f|\cdot\rangle_D$ on $\mathcal{K}_D$ is a one-to-one antilinear mapping of $\mathcal{K}_D^\dagger$ onto the dual $\mathcal{K}_D'$ of $\mathcal{K}_D$. Any bounded operator A on $\mathcal{K}_D$ has a unique adjoint $A^\dagger$ in $\mathcal{K}_D^\dagger$,

$$(A^\dagger f|g\rangle_D = (f|Ag\rangle_D, \quad f \in \mathcal{K}_D^\dagger, \quad g \in \mathcal{K}_D, \tag{2.8}$$

which is a bounded operator on $\mathcal{K}_D^\dagger$, and coincides with A^* on the domain $\mathfrak{D}(A^*)$ of A^*, i.e.,

$$A^\dagger f = A^* f, \quad f \in \mathfrak{D}(A^*). \tag{2.9}$$

Proof: For any $f \in \mathcal{K}$ and $g \in \mathcal{K}_D$

$$\langle f|g\rangle = \langle f|D^{-1}Dg\rangle = \langle (D^{-1})^* f|Dg\rangle. \tag{2.10}$$

Since $(DD^{-1})^*$ is the identity operator on $\mathcal{K}$, we have $(D^{-1})^* = D^{*-1}$. We recall that $\mathbb{D}$ is the extension of D^*. Hence (2.10) assumes the form

$$\langle f|g\rangle = \langle \mathbb{D}^{-1}f|Dg\rangle, \quad f \in \mathcal{K}, \quad g \in \mathcal{K}_D, \tag{2.11}$$

which makes it obvious that an extension of the above sesquilinear form to all $f \in \mathcal{K}_D^\dagger$ is possible since $\mathcal{K}$ is dense in $\mathcal{K}_D^\dagger$ and $\mathbb{D}^{-1}$ is a bounded operator from $\mathcal{K}_D^\dagger$ to $\mathcal{K}$. Moreover, (2.6) becomes true by the definition

$$(f|g\rangle_D = \langle \mathbb{D}^{-1}f|Dg\rangle, \quad f \in \mathcal{K}_D^\dagger, \quad g \in \mathcal{K}_D, \tag{2.12}$$

while (2.7) is obtained from (2.11) and (2.12) by setting $f = \mathbb{D}h$ with $h \in \mathcal{K}$.

Since for any fixed $f \in \mathcal{K}_D^\dagger$ the linear functional $L_f(\cdot) = (f|\cdot\rangle_D$ is bounded on $\mathcal{K}_D$,

$$|(f|g\rangle_D| \le \|\mathbb{D}^{-1}f\| \, \|Dg\| = \|\mathbb{D}^{-1}f\| \, \|g\|_D,$$

it represents an element of $\mathcal{K}_D'$. The mapping $M: f \to L_f$ of $\mathcal{K}_D^\dagger$ into $\mathcal{K}_D'$ is obviously antilinear. Furthermore, if $L_f(g) \equiv 0$ then by (2.11) $\mathbb{D}^{-1}f = 0$ and consequently $f = \mathbb{D}\mathbb{D}^{-1}f = \mathbf{0}$. Hence M is one-to-one. In order to see that the range of M is $\mathcal{K}_D'$ note that if $L(\cdot) \in \mathcal{K}_D'$, then by Riesz' theorem there is a vector $h \in \mathcal{K}_D$ such that for any $g \in \mathcal{K}_D$

$$L(g) = \langle h|g\rangle_D = \langle Dh|Dg\rangle = \langle \mathbb{D}^{-1}(\mathbb{D}Dh)|Dg\rangle.$$

Hence $L = M(\mathbb{D}Dh)$.

According to the preceding result, for each bounded functional $(f|A(\cdot)\rangle$ corresponding to a given bounded operator A on $\mathcal{K}_D$ and for given $f \in \mathcal{K}_D^\dagger$ there is a unique vector $f^\dagger \in \mathcal{K}_D^\dagger$ such that

$$(f|Ag\rangle_D = (f^\dagger|g\rangle_D, \quad g \in \mathcal{K}_D.$$

The mapping $f \to f^\dagger = A^\dagger(f)$ is obviously linear. The operator $A^\dagger$ is bounded by the closed graph theorem[14] since it is defined on the entire space $\mathcal{K}_D^\dagger$ and it is closed. In fact, if $f_1, f_2, \cdots$ and $f_1^\dagger, f_2^\dagger, \cdots$ converge in the $\|\cdot\|_{D^{-1}}$ norm to f and h, respectively, then

$$\begin{aligned}(A^\dagger f|g\rangle_D &= (f|Ag\rangle_D = \lim_{n\to\infty} (f_n|Ag\rangle_D \\ &= \lim_{n\to\infty} (f_n^\dagger|g\rangle_D = (h|g\rangle_D\end{aligned}$$

for all vectors $g \in \mathcal{K}_D$ and therefore $h = A^\dagger f$.

Finally, to prove (2.9) we only have to note that if $f \in \mathfrak{D}(A^*)$ then $A^* f \in \mathcal{K}$. Hence by (2.6)

$$(A^* f|g\rangle = \langle A^* f|g\rangle = \langle f|Ag\rangle$$

for all $g \in \mathcal{K}_D$.

We shall introduce now the concept of generalized expansion associated with any finite number of commuting operators. This concept generalizes that of eigenfunction expansion.[14]

Definition 2.1: Let $A_1, \ldots, A_\nu$ be a set of commuting self-adjoint operators in $\mathcal{K}$, and $E^{A_1,\ldots,A_\nu}(\Delta)$, $\Delta \subset \mathbb{R}^\nu$, their joint spectral measure defined on all Borel sets of $\mathbb{R}^\nu$, i.e., if $\alpha = (\alpha_1, \ldots, \alpha_\nu) \in \mathbb{R}^\nu$

$$A_i = \int_{-\infty}^{+\infty} \alpha_i E^{A_1,\ldots,A_\nu}(\mathbb{R}^1 \times \cdots \times d\alpha_i \times \cdots \times \mathbb{R}^1), \quad i = 1, \ldots, \nu. \tag{2.13}$$

Denote by ${}_D\langle\cdot|\cdot\rangle$ the sesquilinear form

$${}_D\langle f|g\rangle = \overline{(g|f\rangle_D}, \quad f \in \mathcal{K}_D, \quad g \in \mathcal{K}_D^\dagger, \tag{2.14}$$

on $\mathcal{K}_D^\dagger \times \mathcal{K}_D$. An α-dependent family of mutually orthogonal vectors $\phi_1(\alpha), \ldots, \phi_{N_\alpha}(\alpha)$ in $\mathcal{K}_D^\dagger$, normalized so that

$$\sum_{j=1}^{N_\alpha} \|\phi_j(\alpha)\|_{D^{-1}}^2 = 1, \quad N_\alpha = 1, 2, \ldots, +\infty \tag{2.15}$$

is a *generalized expansion* for $\{A_1, \ldots, A_\nu\}$ if there is a measure $\rho(\Delta)$ on the Borel subsets Δ of the Cartesian product $\sigma(A_1) \times \cdots \times \sigma(A_\nu)$ of the spectra of $A_1, \ldots, A_\nu$ for which

$$\langle f|E^{A_1,\ldots,A_\nu}(\Delta)g\rangle = \int \sum_{j=1}^{N_\alpha} {}_D\langle f|\phi_j(\alpha)\rangle\,(\phi_j(\alpha)|g\rangle_D \, d\rho(\alpha) \tag{2.16}$$

for all $f, g \in \mathcal{K}_D$.

In the above definition of a generalized expansion we have not required $\phi_j(\alpha)$ to be generalized eigenvectors[1] of A_i in the sense that

$$(\phi_j(\alpha)|(A_i - \alpha_i \mathbb{1})f\rangle_D = 0, \quad j = 1, \ldots, N_\alpha \tag{2.17}$$

for a dense subset of vectors f in $\mathcal{K}_D$. By basing the definition on spectral measures we avoid the domain questions inherent in (2.17) and common to both the equipped[1] and the rigged[7-9] Hilbert space approach. Of course, whenever $A_i f \in \mathcal{K}_D$ for some $f \in \mathcal{K}_D$, we can easily derive from (2.16) (cf. the method used in proving Theorem 4.2) that (2.17) has to be satisfied for ρ—almost all $\alpha \in \mathbb{R}^\nu$. However, we cannot generally expect this to happen for a dense set of vectors $f \in \mathfrak{D}(A_i) \cap \mathcal{K}_D$ even when $A_1, \ldots, A_\nu$ are bounded in $\mathcal{K}$. On the other hand, a generalized version of (2.17) will hold true after we extend $(\cdot|\cdot\rangle_D$ to a larger space than $\mathcal{K}_D^\dagger$.

It is shown in Ref. 1 (Chap. V, Sec. 2) that generalized expansions in the sense of Definition 2.1 exist in $\mathcal{K}_D^\dagger$ for

an *arbitrary* set $\{A_1,\ldots,A_\nu\}$ of commuting self-adjoint operators in $\mathcal{H}$ if an only if D^{-1} is a Hilbert–Schmidt operator. This fact motivates the introduction of the following concept.

Definition 2.2: Let D be an operator for which both D and D^* are densely defined and have range $\mathcal{H}$, i.e., $\mathcal{R}(D)=\mathcal{R}(D^*)=\mathcal{H}$. The operator D is an *equipping operator* iff D^{-1} and D^{*-1} exist, have bound one, and are Hilbert–Schmidt, i.e., $\mathrm{Tr}\{D^{*-1}D^{-1}\}<\infty$.

The case of principal interest in quantum mechanics[3] when dealing with generalized expansions occurs when $\{A_1,\ldots,A_\nu\}$ is a complete[14] set of self-adjoint operators. In that case, $N_\alpha\equiv 1$ and the following result is valid.

Theorem 2.2: Suppose $\{A_1,\ldots,A_\nu\}$ is a complete set of self-adjoint operators, D an equipping operator and $\mathcal{H}_D\subset\mathcal{H}\subset\mathcal{H}_D^\dagger$ the corresponding equipped Hilbert space. Then there is a generalized expansion $\phi(\alpha)$ for $\{A_1,\ldots,A_\nu\}$ such that

$$\langle f|E^{A_1,\ldots,A_\nu}(\Delta)g\rangle=\int_\Delta {}_D\langle f|\phi(\alpha)\rangle\langle\phi(\alpha)|g\rangle_D\,d\rho_D(\alpha), \tag{2.18}$$

for any $f,g\in\mathcal{H}_D$, where

$$\rho_D(\Delta)=\mathrm{Tr}[D^{-1*}E^{A_1,\ldots,A_\nu}(\Delta)D^{-1}] \tag{2.19}$$

for any Borel set $\Delta\subset\mathbb{R}^\nu$. The mapping

$$U_D: f\to\hat f(\alpha)\equiv(\phi(\alpha)|f\rangle_D \tag{2.20}$$

of $\mathfrak{D}(D)$ in $L^2_{\rho_D}(\mathbf{R}^\nu)$ is isometric and can be uniquely extended to a unitary transformation of $\mathcal{H}$ onto $L^2_{\rho_D}(\mathbb{R}^\nu)$.

Proof: The first part of the theorem is an immediate consequence of Theorem 2.2 in Ref. 2.

The fact that U_D maps isometrically $\mathfrak{D}(D)$ into $L^2_{\rho_D}(\mathbb{R}^\nu)$ follows from (2.18) by setting $\Delta=\mathbb{R}^\nu$

$$\langle f|g\rangle=\int_{\mathbf{R}^\nu}{}_D\langle f|\phi(\alpha)\rangle\langle\phi(\alpha)|g\rangle_D\,d\rho_D(\alpha),\qquad f,g\in\mathcal{H}_D. \tag{2.21}$$

Since $\mathfrak{D}(D)$ is dense in $\mathcal{H}$, we can extend U_D to all of $\mathcal{H}$.

To show that the range of the so defined isometry is $L^2_{\rho_D}(\mathbb{R}^\nu)$, let f_0 be a cyclic vector[14] for $\{A_1,\ldots,A_\nu\}$ and consider the linear manifold spanned in $L^2_{\rho_D}(\mathbb{R}^\nu)$ by all functions

$$\hat f_\Delta(\alpha)=(\phi(\alpha)|E^{A_1,\ldots,A_\nu}(\Delta)f_0\rangle_D$$

corresponding to all Borel sets Δ in $\mathbf{R}^\nu$. By (2.18) $\hat f_0(\alpha)\neq 0$ for ρ_D-almost all $\alpha\in\mathbb{R}^\nu$. Using this fact and the relation

$$\int_{\Delta_1}|\hat f_0(\alpha)|^2\,d\rho_D(\alpha)=\int_{\Delta\cap\Delta_1}\overline{\hat f_0(\alpha)}\hat f_\Delta(\alpha)\,d\rho_D(\alpha)$$

which by (2.18) holds for any Borel set $\Delta_1\subset\mathbb{R}^\nu$, we infer that $\hat f_\Delta(\alpha)=\chi_\Delta(\alpha)\hat f_0(\alpha)$. Since the linear manifold spanned by all such functions as Δ varies over the Borel sets in the support of ρ_D is dense[14] in $L^2_{\rho_D}(\mathbb{R}^\nu)$ if $\hat f_0(\alpha)\neq 0$, ρ_D-almost everywhere, the last statement in the theorem is true.

The concept of "smearing" a generalized expansion $\phi(\alpha)$ with a function $\hat h(\alpha)$,

$$\phi(\hat h)=\int_{\mathbf{R}^\nu}\hat h(\alpha)\,\phi(\alpha)\,d\rho_D(\alpha), \tag{2.22}$$

will prove to be very useful in defining smeared fields and other objects constructed from the fields at a point $\alpha\in\mathbf{R}^\nu$ in quantum field theory (cf. Sec. 6). In case that $\hat h\in L^2_{\rho_D}(\mathbb{R}^\nu)$ one can proceed as in defining Bochner integrals[14] in $\mathcal{H}$ by taking

$$(\phi(\hat h)|f\rangle_D=\int_{\mathbb{R}^\nu}(\hat h(\alpha)\,\phi(\alpha)|f\rangle_D\,d\rho_D(\alpha) \tag{2.23}$$

as a defining relationship. In fact, since

$$|(\phi(\hat h)|f\rangle_D|=|\int_{\mathbf{R}^\nu}\overline{\hat h(\alpha)}\hat f(\alpha)\,d\rho_D(\alpha)|\le\|\hat h\|_{L^2}\|f\|$$

we conclude that $(\phi(\hat h)|\cdot\rangle_D$ is a bounded linear functional on $\mathcal{H}$. Consequently, by Riesz' theorem[14] and (2.6) we conclude that there is a unique vector $\phi(\hat h)$ in $\mathcal{H}$ which satisfies (2.23), and such that

$$\hat h(\alpha)=(U_D\phi(\hat h))(\alpha). \tag{2.24}$$

In quantum field theory we require (cf. Sec. 6) a definition of (2.22) for a much larger class of functions than those in $L^2_{\rho_D}(\mathbb{R}^\nu)$.

Theorem 2.3: Let $K_{D^{-1}}(\alpha,\beta)$ be the kernel[2] of the operator $U_D D^{-1}U_D^{-1}$,

$$(U_D D^{-1}f)(\alpha)=\int_{\mathbb{R}^\nu}K_{D^{-1}}(\alpha,\beta)\hat f(\beta)\,d\rho_D(\beta), \tag{2.25}$$

for a given equipping operator D. Denote by $\mathcal{P}_D(\mathbb{R}^\nu)$ the linear space of all locally ρ_D-square-integrable functions $\hat g(\alpha)$ defined on $\mathbf{R}^\nu$ which are such that the function

$$\text{l.i.m.}\int_{\mathbf{R}^\nu}\overline{K_{D^{-1}}(\beta,\alpha)}\hat h(\beta)\,d\rho_D(\beta) \tag{2.26}$$

is defined for ρ_D-almost all $\alpha\in\mathbb{R}^\nu$ and belongs to $L^2_{\rho_D}(\mathbb{R}^\nu)$. For each $\hat h\in\mathcal{P}_{D^*}(\mathbb{R}^\nu)$ there is a unique element $\phi(\hat h)$ in $\mathcal{H}_D^\dagger$ which satisfies (2.23) for all $f\in\mathcal{H}_D$. This vector belongs to $\mathcal{H}$ if and only if $\hat h\in L^2_{\rho_D}(\mathbf{R}^\nu)$, and in that case $h=U_D^{-1}\hat h$.

Proof: For compact $\Delta_n\subset\mathbb{R}^\nu$ the existence of

$$\phi(\hat h\cdot\chi_{\Delta_n})=\int_{\Delta_n}\hat h(\alpha)\phi(\alpha)\,d\rho_D(\alpha) \tag{2.27}$$

as a vector in $\mathcal{H}$ has been established in the discussion preceding the theorem since $\hat h\cdot\chi_{\Delta_n}\in L^2_{\rho_D}(\mathbb{R}^\nu)$.

If $\Delta_1\subset\Delta_2\subset\cdots$ covers $\mathbb{R}^\nu$ and $m\ge n$ then we infer from the existence of (2.26) that

$$\|D^{*-1}[\phi(\hat h\cdot\chi_{\Delta_m})-\phi(\hat h\cdot\chi_{\Delta_n})]\|^2=\|D^{*-1}\phi(\hat h\cdot\chi_{\Delta_m-\Delta_n})\|^2$$

$$=\int_{\mathbf{R}^\nu}d\rho_D(\alpha)\,|\int_{\Delta_m-\Delta_n}K_{D^{*-1}}(\alpha,\beta)\hat h(\beta)\,d\rho_D(\beta)|^2$$

and that $\{\phi(\hat h\cdot\chi_{\Delta_n})\}$ is a Cauchy sequence in $\mathcal{H}_D^\dagger$. Hence, $\phi(\hat h)$ exists as a limit in the norm of $\mathcal{H}_D^\dagger$. Its uniqueness in $\mathcal{H}_D^\dagger$ is a consequence of the existence of the one-to-one map M of $\mathcal{H}_D^\dagger$ onto $\mathcal{H}_D'$ introduced in Theorem 2.1.

We have already seen that $\phi(\hat h)\in\mathcal{H}$ when $\hat h\in L^2_{\rho_D}(\mathbb{R}^\nu)$. Conversely, if $\phi(\hat h)\in\mathcal{H}$ we infer from (2.23) that $\int\overline{\hat h(\alpha)}\hat f(\alpha)\,d\rho_D(\alpha)$ is a bounded linear functional in $\hat f\in L^2_{\rho_D}(\mathbb{R}^\nu)$. Hence $\hat h(\alpha)\in L^2_{\rho_D}(\mathbb{R}^\nu)$ by Riesz' theorem.

In conclusion, let us point out that $\mathcal{P}_D(\mathbb{R}^\nu)$ contains $L^2_{\rho_D}(\mathbb{R}^\nu)$, but in general it contains many other additional

functions. As an example, consider $\mathcal{H} = L^2(\mathbb{R}^3)$ and take $D = (-\nabla^2 + 1)\,V_0^{-1}$, where V_0 is multiplication by $V_0(\mathbf{r}) \in L^2(\mathbb{R}^3)$, and $V_0(\mathbf{r})$ is real and nonzero almost everywhere in $\mathbb{R}^3$. It is well-known that in this case D^{-1} and D^{-1*} are Hilbert–Schmidt since[14]

$$K_{D^{-1}}(\mathbf{r},\mathbf{r}') = V_0(\mathbf{r})\,\frac{\exp(-|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|}$$

is square-integrable on $\mathbb{R}^6$. Hence $D^{*-1} = D^{-1*}$ and the ranges of D and D^* are obviously $L^2(\mathbb{R}^3)$. Consequently, D is an equipping operator. By a judicious choice of $V_0(\mathbf{r})$ we can have $\mathcal{P}_D(\mathbb{R}^3)$ contain even locally square-integrable functions which diverge polynomially when $r \to \infty$. Such a choice would be, for example, $V_0(\mathbf{r}) = \exp(-r)$.

3. TOPOLOGIES AND OPERATORS IN EXTENDED HILBERT SPACE

Let D be an equipping operator and A an operator mapping $\mathcal{H}_D$ into itself and bounded on $\mathcal{H}_D$, i.e.,

$$\|A\|_D = \sup_{f\in\mathcal{H}_D} \|f\|_D^{-1}\|Af\|_D < \infty. \tag{3.1}$$

According to Theorem 2.1, $A^\dagger$ exists as an operator on $\mathcal{H}_D^\dagger$. Thus if $\phi(\alpha)$ is an expansion corresponding to a complete set of observables and $f, g \in \mathcal{H}_D$ then by (2.18)

$$\langle f|Ag\rangle = \int_{\mathbb{R}^\nu} {}_D\langle f|\phi(\alpha)\rangle(A^\dagger\phi(\alpha)|g\rangle_D\, d\rho_D(\alpha). \tag{3.2}$$

If $A^\dagger\phi(\alpha) \in \mathcal{H}_D$ for ρ-almost all $\alpha \in \mathbb{R}^\nu$ then we define

$$(\phi(\alpha)|A|\phi(\beta)) = {}_D\langle A^\dagger\phi(\alpha)|\phi(\beta)\rangle. \tag{3.3}$$

Furthermore, we have $(A^\dagger\phi(\alpha)|g\rangle_D = \langle A^\dagger\phi(\alpha)|g\rangle$ by (2.6), and we can apply again (2.18) to write

$$\langle f|Ag\rangle = \int_{\mathbb{R}^{2\nu}} {}_D\langle f|\phi(\alpha)\rangle(\phi(\alpha)|A|\phi(\beta))\times(\phi(\beta)|g\rangle_D\, d\rho_D(\alpha)\, d\rho_D(\beta), \tag{3.4}$$

since the once iterated integral over $\mathbb{R}^\nu$ is equal to the above single integral over $\mathbb{R}^{2\nu}$ by Tonelli's and Fubini's theorems.[14]

The above relation (3.4) is a rigorous version of the basic formula for expansions of operators in Dirac's formalism.[3] This type of procedure for deriving the formula (3.2) has been also used in rigged Hilbert space and other alternative approaches to Dirac's formalism.[7,8] It shares, however, with this approach a number of features which are serious deficiencies from the point of view of applications. First of all, the derivation of (3.4) requires f, g and $A^\dagger\phi(\alpha)$ to belong to $\mathcal{H}_D$. Moreover, we must have $\|A\|_D < \infty$. This last condition can be difficult to verify in practice. If there are equipping operators D for which D^{-1} commutes with A, then the condition is satisfied by any A which is bounded in $\mathcal{H}$ and leaves $\mathfrak{D}(D)$ invariant, since then

$$\|Af\|_D = \|DAf\| = \|ADf\| \le \|A\|\,\|Df\| = \|A\|\,\|f\|_D. \tag{3.5}$$

However, if A has a nonempty continuous spectrum, then (cf. Ref. 16, Lemma 1) there are no Hilbert–Schmidt operators D^{-1} commuting with A for which $\mathcal{R}(D) = \mathcal{H}$. In that case this possibility is closed to us.

The condition that (3.1) be fulfilled can be relaxed somewhat by making the more modest requirement that A maps a dense domain of $\mathcal{H}_D$ into $\mathcal{H}_D$ and therefore that its adjoint $A^\dagger$ in $\mathcal{H}_D^\dagger$ is unique (provided it exists). However, even such a condition can be quite difficult to establish in practice. For example, this is the case with the derivation[2] of the formula relating the T operator to the T matrix in scattering theory. Hence a more general approach is desirable. We base such an approach on the following observation.

Lemma 3.1: Let X and Y be bounded linear operators on $\mathcal{H}$ which have bounded inverses together with their adjoints and are such that $\mathcal{R}(X^*) = \mathcal{R}(Y^*) = \mathcal{H}$. If for some equipping operator D the sets $X^{-1}\mathfrak{D}(D)$ and $Y^{-1}\mathfrak{D}(D^*)$ are dense in $\mathcal{H}$, then $\|(Y^*DX)^{-1}\|\,Y^*DX$ is also an equipping operator. Conversely, suppose that D and D' are equipping operators for which $d_1 < d_2 < \cdots$ and $d_1' < d_2' < \cdots$ are the eigenvalues of $(D^*D)^{1/2}$ and $(D'^*D')^{1/2}$, respectively. If $C = \sup\{d_n' d_n^{-1}\}$ is finite then there are pairs $\{X, Y\}$ of operators with the above properties for which $D' = [Y^*DX]_0$, where $[Y^*DX]_0 = \|(Y^*DX)^{-1}\|\,Y^*DX$.

Proof: The conditions imposed on X and Y imply that Y^*DX and $(Y^*DX)^* \supseteq X^*D^*Y$ map dense domains of $\mathcal{H}$ onto $\mathcal{H}$. Furthermore, $(Y^*DX)^{-1} = X^{-1}D^{-1}Y^{*-1}$ and $(Y^*DX)^{*-1} = Y^{**-1}D^{*-1}X^{*-1}$ are Hilbert–Schmidt since D^{-1} and D^{*-1} are Hilbert–Schmidt and X^{-1}, Y^{-1}, X^{*-1} are bounded.[14]

Conversely, if D and D' are given, we can decompose D^{-1} and D'^{-1} in the canonical form[14]

$$\begin{aligned} D^{-1} &= \sum_{n=1}^{\infty} |u_n\rangle\lambda_n\langle v_n|, \quad \sum_{n=1}^{\infty}\lambda_n^2 < \infty, \\ &\lambda_1 \ge \lambda_2 \ge \cdots > 0, \\ D'^{-1} &= \sum_{n=1}^{\infty} |u_n'\rangle\lambda_n'\langle v_n'|, \quad \sum_{n=1}^{\infty}\lambda_n'^2 < \infty, \\ &\lambda_1' \ge \lambda_2' \ge \cdots > 0, \end{aligned} \tag{3.6}$$

where $\{u_n\}, \ldots, \{v_n'\}$ are orthonormal systems of vectors in $\mathcal{H}$. The condition $\mathcal{R}(D) = \mathcal{R}(D') = \mathcal{R}(D^*) = \mathcal{R}(D'^*) = \mathcal{H}$ implies that these systems are actually orthonormal bases in $\mathcal{H}$. Any linear operators defined on the linear manifolds spanned by $\{u_n'\}$ and $\{v_n'\}$ so that

$$Xu_n' = \xi_n u_n, \qquad Yv_n' = \eta_n v_n, \tag{3.7}$$

satisfy the conditions stipulated in the lemma if

$$\xi_n\eta_n = \lambda_n/\lambda_n' = d_n'/d_n, \qquad \sup|\xi_n|, \sup|\eta_n| < \infty.$$

The above lemma indicates that for given bounded A it is possible to define an extension $(A-\zeta)^\dagger$ of $(A-\zeta)^*$ as an operator from $\mathcal{H}_D^\dagger$ to $\mathcal{H}_{[D(A-\zeta)]_0}^\dagger$ by choosing $|\zeta| > \|A\|$. In order to have such operators act in one and the same space, we shall construct a larger space $\mathfrak{L}$ which in a certain sense contains both $\mathcal{H}_D^\dagger$ and $\mathcal{H}_{[D(A-\zeta)]_0}^\dagger$. This will make the derivation of (3.4) (in which actually $\mathcal{H}_D^\dagger \equiv \mathcal{H}_{[D(A-\zeta)]_0}^\dagger$) a special case of the present derivation.

Lemma 3.2: Let $\mathcal{E}$ be a family of equipping operators. Denote by $\tilde{\mathfrak{L}}$ the set whose general element is a D-convergent sequence $\{f_n\}$ of vectors in $\mathcal{H}$ for some $D \in \mathcal{E}$, i.e., a sequence for which $D^{-1}f_1, D^{-1}f_2, \cdots$ is norm-convergent in $\mathcal{H}$. Let us write $\{f_n\} \sim \{f_n'\}$ for $\{f_n\}, \{f_n'\} \in \tilde{\mathfrak{L}}$ iff $\|D^{-1}(f_n - f_n')\| \to 0$ for all $D \in \mathcal{E}$. The relationship $\sim$ is an equivalence relationship. If $\{g_n\} \sim \{g_n'\}$ and $\{af_n + bg_n\} \in \mathfrak{L}$ for some $a, b \in \mathbb{C}^1$ then $\{af_n' + bg_n'\} \in \mathfrak{L}$ and $\{af_n + bg_n\} \sim \{af_n' + bg_n'\}$.

Proof: The above relationship is obviously reflexive and symmetric. It is also transitive since if $\{f_n'\} \sim \{f_n''\}$ as well, then for any $D \in \mathcal{E}$

$$\|D^{-1}(f_n - f_n'')\| \leq \|D^{-1}(f_n - f_n')\| + \|D^{-1}(f_n' - f_n'')\| \to 0.$$

In the same manner we verify that

$$\lim_{n\to\infty} \|D^{-1}(af_n + bg_n) - D^{-1}(af_n' + bg_n')\| = 0$$

for all $D \in \mathcal{E}$. Moreover if $\{h_n\} \in \tilde{\mathcal{L}}$ and $\|D^{-1}(h_n - h_n')\| \to 0$ for all $D \in \mathcal{E}$, then it follows that $\{h_n'\} \in \tilde{\mathcal{L}}$. Thus the last statement of the lemma is true.

Remark: The family $\mathcal{E}$ will be presumed to be fixed in the sequel. Hence the dependence of $\mathcal{L}$ on $\mathcal{E}$ is not explicitly exhibited. We note that if $\mathcal{E}$ contains only one element D then $\mathcal{L} \equiv \mathcal{K}_{D*}^{\dagger}$. To avoid reverting to the case considered in Sec. 2, we tacitly assume that in general $\mathcal{E}$ contains two or more distinct elements. The other extreme case occurs when $\mathcal{E}$ consists of all equipping operators in $\mathcal{K}$. We believe that in general the choice of $\mathcal{E}$ should be dictated by the problem under consideration.

Definition 3.1: Let $\mathcal{K}$ be a separable complex Hilbert space, $\mathcal{E}$ a family of equipping operators, and $\mathcal{L}$ the family of equivalence classes of sequences $\{f_n\}$ from $\tilde{\mathcal{L}}$ with respect to the equivalence relation introduced in Lemma 3.2. Denote by $\mathcal{L}_{\text{ket}}$ the image of $\mathcal{K}$ in $\mathcal{L}$ under the embedding which identifies each element f of $\mathcal{K}$ with the equivalence class in $\mathcal{L}$ containing the sequence $\{f_n\}$ with $f_n = f$ for $n = 1, 2, \cdots$. The ordered pair $\mathcal{L}_{\text{ket}} \subset \mathcal{L}$ will be called the *extended Hilbert space* over $\mathcal{E}$ corresponding to $\mathcal{K}$; $\mathcal{L}$ is the *bra-space* and $\mathcal{L}_{\text{ket}}$ the *ket-space* of this pair (see also Sec. 7 for an alternative definition).

We note that the above definition is purely set-theoretical and does not involve any explicit topological concepts. However, many topologies can be introduced in a natural manner, one of the most useful being the following.

Definition 3.2: For each equipping operator $D \in \mathcal{E}$ define the surjective mapping l_D from $\mathcal{L}$ onto $\mathcal{K}$ by setting

$$l_D(\{f_n\}) = \mathrm{s}-\lim_{n\to\infty} D^{-1}f_n \tag{3.8}$$

for any $\{f_n\} \in \tilde{\mathcal{L}}$ for which the above limit exists. Denote by $\mathfrak{N}_f$ the family of all nonempty sets

$$\mathcal{U}(f; D_1, \ldots, D_n; \epsilon) = \{g \mid \|l_{D_i}(f) - l_{D_i}(g)\| < \epsilon, \quad i = 1, \ldots, n\} \tag{3.9}$$

corresponding to all $\epsilon > 0$ and any possible choice of $D_1, \ldots, D_n \in \mathcal{E}$, $n = 1, 2, \cdots$, for which $l_{D_i}(f)$ is defined. The topology in which $\mathfrak{N}_f$ is a neighborhood basis for each $f \in \mathcal{L}$ will be called the *strong topology* in the bra-space $\mathcal{L}$.

We note that topology induced in $\mathcal{L}_{\text{ket}}$ by the strong topology in $\mathcal{L}$ does not coincide with the Hilbert-norm topology of $\mathcal{L}_{\text{ket}}$ obtained by the identification of $\mathcal{L}_{\text{ket}}$ with $\mathcal{K}$ (i.e., the topology which makes this identification into a homeomorphism), the second being finer than the first. On the other hand, the above topology deserves the name of strong topology since it is in fact the strong topology[17] coinduced by the maps $l_D^{-1}, D \in \mathcal{E}$, of $\mathcal{L}_{\text{ket}}$ supplied with the norm-topology of $\mathcal{K}$ to $\mathcal{L}$. Since these maps cover $\mathcal{L}$, this topology can be also viewed as a quotient topology[17] induced by the map $\times_{D\in\mathcal{E}} l_D^{-1}$ from the disjoint union $\{(\mathcal{L}_{\text{ket}}, D) \mid D \in \mathcal{E}\}$ to $\mathcal{L}$.

We can introduce in $\tilde{\mathcal{L}}$ the operations

$$a\{f_n\} = \{af_n\}, \quad a \in \mathbb{C}^1, \tag{3.10}$$

$$\{f_n\} + \{g_n\} = \{f_n + g_n\}. \tag{3.11}$$

The operation (3.10) is defined for all $a \in \mathbb{C}^1$ and $\{f_n\} \in \tilde{\mathcal{L}}$. However, (3.11) is not defined if $\{f_n\}$ and $\{g_n\}$ have no common D for which $\{f_n\}$ and $\{g_n\}$ are both D-convergent sequences. Thus, in general, $\mathcal{L}$ cannot be made into a vector space in a natural manner. On the other hand, $\mathcal{L}$ can be decomposed into maximal families of vector spaces $\mathcal{L}_D$ (which are not disjoint!).

Theorem 3.1: For any fixed equipping operator $D \in \mathcal{E}$ the subset $\mathcal{L}_D$ of $\mathcal{L}$ consisting of the domain of definition of the mapping l_D in (3.8) is a vector space under the vector operations (3.10) and (3.11). The multiplication by scalar operation (3.10) is continuous, while the vector addition operation (3.11) is not a generally continuous operation in the strong topology of $\mathcal{L}_D$. The ket space $\mathcal{L}_{\text{ket}}$ is a vector subspace of $\mathcal{L}_D$ and is dense in $\mathcal{L}$ in the strong topology of $\mathcal{L}$.

Proof: First of all, we note that l_D is well-defined for each equivalence class containing a sequence $\{f_n\}$ for which (3.8) exists, since if $\{f_n'\} \sim \{f_n\}$ then

$$\mathrm{s}-\lim D^{-1}f_n' = \mathrm{s}-\lim D^{-1}(f_n' - f_n) + \mathrm{s}-\lim D^{-1}f_n = \mathrm{s}-\lim D^{-1}f_n.$$

Due to the last statement in Lemma 3.2, the vector operations (3.10) and (3.11) are unambiguously defined on all of $\mathcal{L}_D$. The continuity of (3.10) is obvious. However, if for some $f \in \mathcal{L}_D - \mathcal{L}_{D'}$, where $D \neq D', D, D' \in \mathcal{E}$, we choose an element h of $\mathcal{L}_D \cap \mathcal{L}_{D'}$, then vector addition $(f, g) \to f + g$ is not continuous at $g = f - h$ since for the neighborhood $\mathcal{U}(h; D'; \epsilon)$ we cannot find neighborhoods $\mathcal{U}(f)$ and $\mathcal{U}(g)$ such that $\mathcal{U}(f) + \mathcal{U}(g) \subset \mathcal{U}(h; D'; \epsilon)$.

The embedding of $\mathcal{K}$ in $\mathcal{L}$ obviously preserves the vector structure of $\mathcal{K}$. To see that the image $\mathcal{L}_{\text{ket}}$ of the embedding is dense in $\mathcal{L}$, recall that any $f \in \mathcal{L}$ is represented by a sequence $\{f_n\} \in \tilde{\mathcal{L}}$ and for any given $\mathcal{U}(f; D_1; \epsilon)$ the vector f is in the domain of definition of l_{D_1}. Hence $\|l_{D_1}f - f_{n_0}\| < \epsilon$ for large n_0, and therefore for these values of n_0 the equivalence class containing the sequence $f_{n_0}, f_{n_0}, \cdots$ belongs to $\mathcal{U}(f; D_1; \epsilon)$.

Remark: The vector space $\mathcal{L}_D$ is generally not a topological vector space[18] in the induced strong topology since vector addition is not continuous. By extending each mapping l_D to a mapping l_D' of $\mathcal{L}$ into $\mathcal{K}$ which is linear on each $\mathcal{L}_{D_1}, D_1 \in \mathcal{E}$, we would obtain many possible topological structures compatible with the vector operations on $\mathcal{L}_D$ by noting that the ordered pair $\mathcal{L}_{\text{ket}} \subset \mathcal{L}_D$ supplied with the family $\mathcal{E}'$ of all such mappings l_D' is a generalized inner product space.[15] By using Zorn's lemma we can prove the existence of a "vector base" $\mathfrak{B}_0$ in $\mathcal{L}$ in the same manner[18] as for vector spaces (the only difference being that linear independence in $\mathcal{L}$ can be sometimes due to the absence of a universally defined vector-addition operation) and then define such extensions l_D' of l_D by assigning to $l_D'f$ an arbitrary value in $\mathcal{K}$ (such as the zero vector) for any $f \in \mathfrak{B}_0 - \mathcal{L}_D$. This would lead to $\mathcal{E}'$ and make $\mathcal{L}$ into a composite generalized inner-product space[15] with components $(\mathcal{L}_D, \mathcal{L}_{\text{ket}}, \mathcal{E}')$ dependent on $D \in \mathcal{E}$. However, a topological vector structure is not essential in deriving the results in this paper, and therefore we shall not pursue this approach any further.

Lemma 3.3: The mapping m_D of $\mathcal{L}$ into $\mathcal{K}_{D*}^{\dagger}$, which

assigns to an equivalence class f in $\tilde{\mathcal{L}}$ representing an element of $\mathcal{L}_D$ the larger equivalence class $m_D f$ in $\tilde{\mathcal{L}}$ representing an element of $\mathcal{K}^\dagger_{D*}$ containing f(i.e., $f \subset m_D f \subset \tilde{\mathcal{L}}$ for all $f \in \mathcal{L}_D$ and $m_D f \in \mathcal{K}^\dagger_{D*}$), maps $\mathcal{L}_D$ linearly onto $\mathcal{K}^\dagger_{D*}$. This mapping is continuous if $\mathcal{L}_D$ and $\mathcal{K}^\dagger_{D*}$ are supplied with their respective strong topologies (but generally it does not have an inverse).

Proof: The mapping m_D is obviously uniquely defined for each $f \in \mathcal{L}_D$. Linearity and continuity are also evident from the definitions. The typical element of $\mathcal{K}^\dagger_{D*}$ consists of some $\{f_n\} \in \tilde{\mathcal{L}}$ and all other $\{f'_n\} \in \tilde{\mathcal{L}}$ such that $\|D^{-1}(f_n - f'_n)\| \to 0$. Since some of these $\{f'_n\}$ might be such that $\{D_1^{-1}(f_n - f'_n)\}$ does not strongly converge to zero in $\mathcal{K}$ for all $D_1 \neq D$, each element of $\mathcal{K}^\dagger_{D*}$ will contain many elements of $\mathcal{L}_D$. Hence m_D is not generally invertible.

The above lemma shows that in general $\mathcal{L}_D$ is much larger than $\mathcal{K}^\dagger_{D*}$. This is due to the fact that the equivalence classes of $\mathcal{L}_D$ (determined by the equivalence relation of Lemma 3.2) are much smaller than those of $\mathcal{K}^\dagger_{D*}$. As a consequence some objects (such as bra-expansions) which are uniquely or almost everywhere uniquely defined in $\mathcal{K}^\dagger_{D*}$ are not unique in $\mathcal{L}_D$. Hence, it might turn out to be more desirable to make $\mathcal{L}$ smaller by increasing the size of the equivalence classes in $\tilde{\mathcal{L}}$ determining its elements to insure such uniqueness. However, we adopt the attitude that systematic applications should provide the guidelines for such a procedure. At the present stage we consider that the study of the largest of the conceivable choices for $\mathcal{L}$ at least avoids the most undesirable alternative, namely that of having key operations (such as taking bra-adjoints) undefinable because of the restricted size of $\mathcal{L}$.

Theorem 3.2: There is a unique complex functional $(\cdot|\cdot)$ defined in $\mathcal{L} \times \mathcal{L}_{\text{ket}}$ with the property

$$(f|g\rangle = \langle f|g\rangle, \quad f,g \in \mathcal{L}_{\text{ket}}, \tag{3.12}$$

for all pairs of ket-vectors f and g, which is strongly continuous on $\mathcal{L}$ at each fixed $g \in \mathcal{L}_{\text{ket}}$. The restriction of $(\cdot|\cdot)$ to $\mathcal{L}_{D*} \times \mathcal{K}_D$ is continuous in both its arguments in the respective strong topologies. In $\mathcal{L}_0 = \cap_{D\in\mathcal{E}} \mathcal{L}_D$ the relation $(f|g\rangle = 0$ is satisfied for all $g \in \mathcal{L}_{\text{ket}}$ only by the zero vector $f = \mathbf{O}$.

Proof: We set

$$(f|g\rangle = (m_{D*}f|g\rangle_D, \quad f \in \mathcal{L}_{D*}, \quad g \in \mathcal{K}_D \tag{3.13}$$

for all $f \in \mathcal{L}_{D*} \subset \mathcal{L}$ and all $g \in \mathcal{K}_D \subset \mathcal{L}_{\text{ket}}$. The continuity and sesquilinearity of the above form on $\mathcal{L}_{D*} \times \mathcal{K}_D$ are features inherited from the corresponding properties of $(\cdot|\cdot)_D$ (cf. Theorem 2.1) and of m_{D*} (cf. Lemma 3.3).

To establish that (3.13) defines a single-valued complex functional on $\mathcal{L} \times \mathcal{L}_{\text{ket}}$ as D varies over $\mathcal{E}$, we note first the functional is single-valued on $\mathcal{L}_{\text{ket}} \times \mathcal{L}_{\text{ket}}$ since by (2.6) we have $(m_{D*}f|g\rangle_D = \langle f|g\rangle$ for any $f \in \mathcal{L}_{\text{ket}}$ and for any $D \in \mathcal{E}$. This expression is independent of D [and incidentally it displays that (3.12) is satisfied by our candidate]. Let us assume now that $f \in \mathcal{L} - \mathcal{L}_{\text{ket}}$ and also $f \in \mathcal{L}_{D_1^*} \cap \mathcal{L}_{D_2^*}$ for some $D_1, D_2 \in \mathcal{E}$. Then there is a sequence $f_1, f_2, \cdots \in \mathcal{L}_{\text{ket}}$ which converges strongly to f, and therefore $m_{D_i^*}f_1, m_{D_i^*}f_2, \cdots$ converges strongly to $m_{D_i^*}f$ in $\mathcal{K}^\dagger_{D_i^*}$ $(i = 1, 2)$. Ambiguities in the definition (3.13) can arise only for $g \in \mathcal{K}_{D_1} \cap \mathcal{K}_{D_2}$. But for any such g

$$(m_{D_1^*}f_k|g\rangle_{D_1} = \langle f_k|g\rangle = (m_{D_2^*}f_k|g\rangle_{D_2}$$

and, consequently,

$$(m_{D_i^*}f|g\rangle_{D_i} = \lim_{k\to\infty} (m_{D_i^*}f_k|g\rangle_{D_i}$$

has the same value for $i = 1, 2$.

It is obvious from the construction that $(\cdot|g\rangle$ is strongly continuous on $\mathcal{L}_{D*}$ for each $g \in \mathcal{K}_D$. The uniqueness property follows from (3.12) and the fact that $\mathcal{L}_{\text{ket}}$ is dense in $\mathcal{L}$ (Theorem 3.1).

Let us assume now that $(f_0|g\rangle = 0$ for all $g \in \mathcal{L}_{\text{ket}}$. In view of (3.12) and the positive-definiteness of inner products, no $f_0 \neq \mathbf{O}$ from $\mathcal{L}_{\text{ket}}$ could satisfy such a relation. To be more general, suppose that $f_0 \in \mathcal{L}_0$. Then we infer by using (3.13) that for each $D \in \mathcal{E}$ for which $f_0 \in \mathcal{L}_D$ we have $(m_{D*}f|g\rangle_D = 0$. This implies that $m_{D*}f$ is the zero vector in $\mathcal{K}^\dagger_{D*}$ and, consequently, for any sequence $\{f_n\} \in \tilde{\mathcal{L}}$ in the equivalence class representing f we have $\text{s} - \lim D^{*-1}f_n = \mathbf{O}$. Since this is true for any $D \in \mathcal{E}$ when $f_0 \in \mathcal{L}_0$, it follows that $f_0 \in \mathbf{O}$.

We note that the conclusion $f_0 = \mathbf{O}$ does not follow if $f_0 \in \mathcal{L} - \mathcal{L}_0$. In that case we shall have that $\text{s} - \lim D^{*-1}f_n = \mathbf{O}$ for D in some proper subset $\mathcal{E}_{f_0}$ of $\mathcal{E}$, but there might be other D's for which this limit does not exist. In fact, for such an f_0 the functional $(f_0|g\rangle$ is defined only for $g \in \cup_{D\in\mathcal{E}_{f_0}} \mathcal{K}_D$.

Another significant feature of $(\cdot|\cdot\rangle$ is that for fixed $f \in \mathcal{L} - \mathcal{L}_{\text{ket}}$ the functional $(f|\cdot\rangle$ is not continuous on $\mathcal{L}_{\text{ket}}$ in its norm-topology (inherited from $\mathcal{K}$). In fact, if that were the case then $(f|\cdot\rangle$ would be in the dual $\mathcal{K}'$ of $\mathcal{K}$. This is not possible since by Riesz' theorem[8] we would have that $f \in \mathcal{L}_{\text{ket}}$.

Definition 3.3: For $f \in \mathcal{L}$ let $\mathfrak{N}_f$ denote the family of nonempty sets

$$\mathfrak{W}(f; h_1, \ldots, h_n) = \{g \| (f - g|h_i\rangle| < 1, \quad i = 1, \ldots, n\}$$

corresponding to any possible choice of $h_1, \ldots, h_n \in \mathcal{L}_{\text{ket}}$ for which $(f|h_i\rangle$ is defined. The *weak topology* in $\mathcal{L}$ is the topology in which $\mathfrak{N}_f$ is a neighborhood base of $f \in \mathcal{L}$.

We note that while the strong topology in $\mathcal{L}$ is Hausdorff, the weak topology is not since there are no weak neighborhoods to separate the zero-vector from a zero-like vector f, i.e., any element f of $\mathcal{L} - \mathcal{L}_0$ represented by $\{f_n\} \in \tilde{\mathcal{L}}$ for which $D^{-1}f_1, D^{-1}f_2, \cdots$ converges to zero for all those D for which it converges at all.

Theorem 3.3: Suppose that A is a linear operator for which $(DA)^*$ ranges over the entire set $\mathcal{E}$ as D^* varies over $\mathcal{E}$ (we shall refer to a family $\mathcal{E}$ with this property as being A-*stable*). Then A has a *bra-adjoint* $A^\dagger$ in $\mathcal{L}$ satisfying

$$(A^\dagger f|g\rangle = (f|Ag\rangle, \quad f \in \mathcal{L}, \quad g \in \mathcal{L}_{\text{ket}}, \tag{3.14}$$

for all $f \in \mathcal{L}$ and $g \in \mathcal{L}_{\text{ket}}$. $A^\dagger$ is continuous in the strong topology of $\mathcal{L}$ and is a linear operator from $\mathcal{L}_{D*}$ into $\mathcal{L}_{(DA)_0^*}$ for any $D \in \mathcal{E}$. Furthermore, the relation (3.14) determines $A^\dagger$ uniquely as a mapping of $\mathcal{L}$ into $\mathcal{L}_0$ (but not otherwise).

Proof: If $\mathcal{E}$ is A-stable and $\{f_n\} \in \mathcal{L}$ represents an

element f in $\mathfrak{L}_{D^*}$, $\{A^*f_n\}$ belongs to $\tilde{\mathfrak{L}}$ and represents an element of $\mathfrak{L}_{(DA)_0^*}$, which we shall denote by $A^\dagger f$. We have assumed that $(DA)_0^*$ varies over $\mathcal{E}$ as D^* varies over $\mathcal{E}$ (e.g., this will be the case if A^{-1} and A^{*-1} exist on $\mathcal{H}$ and are bounded and $\mathcal{E}$ contains all equipping operators). Hence, $\{f_n\} \sim \{f_n'\}$ implies $\{A^*f_n\} \sim \{A^*f_n'\}$. If $f \in \mathfrak{L}_{\text{ket}}$ then clearly $A^\dagger f$ is also a ket-vector. Thus, by (3.12) we have

$$\langle A^\dagger f|g\rangle = (A^\dagger f|g\rangle = (f|Ag\rangle = \langle f|Ag\rangle$$

whenever $f, g \in \mathfrak{L}_{\text{ket}}$. Hence $A^\dagger$ is an extension of A^*. Moreover, if $A_1^\dagger$ is another adjoint of A in $\mathfrak{L}$, then

$$(A_1^\dagger f - A^\dagger f|g\rangle = 0$$

for all $g \in \mathfrak{L}_{\text{ket}}$ if $A_1^\dagger f, A^\dagger f \in \mathfrak{L}_0$, and consequently by the last part of Theorem 3.2 we have $A_1^\dagger f = A^\dagger f$. However, in general we can alter arbitrarily the definition of $A^\dagger$ on a zero-like vector in $\mathfrak{L}$ without violating (3.14).

The linearity of $A^\dagger$ on $\mathfrak{L}_{D^*}$ is an obvious consequence of the linearity of A^* on $\mathcal{H}$ and the definition of $A^\dagger$. The continuity of $A^\dagger$ follows from the observation that the inverse image of $\mathcal{U}(A^\dagger f; D; \epsilon)$ under $A^\dagger$ contains $\mathcal{U}(f; A^{*-1}D; \epsilon)$.

We note that if $\mathcal{E}$ consists of all equipping operators in $\mathcal{H}$ and the bounded operator A is such that A^{-1} and A^{*-1} exist on $\mathcal{H}$ and are bounded then by Lemma 3.1 the conditions of the theorem are satisfied. More generally, if either A^{-1} or A^{*-1} do not exist or are not bounded, the operator $A_\zeta = A - \zeta$ is such that $(A - \zeta)^{-1}$ and $(A^* - \bar{\zeta})^{-1}$ exist and are bounded if ζ is chosen in the resolvent set of A. Hence the family $\mathcal{E}_0$ of all equipping operators is $(A - \zeta)$-stable if ζ is in the resolvent set. Thus, we can always define the *bra-adjoint* $A^\dagger$ of A by

$$A^\dagger = (A - \zeta)^\dagger + \zeta^* \mathbb{1}^\dagger,$$

where $\mathbb{1}^\dagger$ is the identity operator in $\mathfrak{L}$ over $\mathcal{E}_0$. However, while $\mathbb{1}^\dagger$ maps $\mathfrak{L}_D$ into itself, $(A - \zeta)^\dagger$ maps it into $\mathfrak{L}_{(DA-\zeta D)_0^*}$ so that $A^\dagger$ might not contain all vectors in $\mathfrak{L}$ in its domain of definition (see also Sec. 7).

4. BRA-EIGENVECTOR EXPANSIONS AND EXISTENCE OF GREEN'S FUNCTIONS

We shall convert the results of Sec. 2 so as to apply to extended Hilbert spaces. This will enable us to formulate and solve with ease problems which could be solved only under additional restrictive conditions if a single equipped Hilbert space were employed.

Theorem 4.1: Let $\{A_1, \ldots, A_\nu\}$ be a complete set of self-adjoint operators in $\mathcal{H}$. For any $D \in \mathcal{E}$ there is a bra-expansion $\phi_D(\alpha) \in \mathfrak{L}_D$ such that

$$\langle f|E^{A_1,\ldots,A_\nu}(\Delta)g\rangle = \int_\Delta \langle f|\phi_D(\alpha))(\phi_D(\alpha)|g\rangle d\rho_D(\alpha) \qquad (4.1)$$

for any $f, g \in \mathcal{H}_D$ [where ρ_D is defined in (2.19)]. The functional $\langle \phi_D(\alpha)|\cdot\rangle$ can be extended to $\mathfrak{L}_{\text{ket}}$ in such a manner that $(\phi_D(\alpha)|f\rangle$ is ρ_D-square-integrable in $\alpha \in R^\nu$ for any $f \in \mathfrak{L}_{\text{ket}}$ and (4.1) holds for any f, $g \in \mathfrak{L}_{\text{ket}}$. Moreover, if D' is any other equipping operator, then ρ_D and $\rho_{D'}$ are equivalent measures and there is ρ_D-almost everywhere in $\mathbb{R}^\nu$ a bra-expansion $\rho_{D'}(\alpha)$ such that

$$\langle f|\phi_D(\alpha)) = \langle f|\phi_{D'}(\alpha))\left(\frac{d\rho_{D'}(\alpha)}{d\rho_D(\alpha)}\right)^{1/2} \qquad (4.2)$$

for all $f \in \mathfrak{L}_{\text{ket}}$.

Proof: We can define $\phi_D(\alpha)$ by applying the axiom of choice to select from each equivalence class in $\tilde{\mathfrak{L}}$, representing an element of the expansion $\phi(\alpha)$ in $\mathcal{H}_{D^*}^\dagger$ of Theorem 2.2, a representative element which would determine $\phi_D(\alpha)$ as an element of $\mathfrak{L}_D$. [The use of the axiom of choice can be avoided by going back to the construction of $\phi_D(\alpha)$ from the spectral measure $E^{A_1,\ldots,A_\nu}(\Delta)$ and selecting at each α a specific sequence of contracting sets $\Delta_1 \supset \Delta_2 \supset \cdots \to \{\alpha\}$ in the limiting procedure determining $\phi_D(\alpha)$.] In view of (2.16) and (3.13) it is obvious that with such a choice (4.1) is true for any $f, g \in \mathcal{H}_D$. Since $\mathcal{H}_D$ is dense in $\mathfrak{L}_{\text{ket}}$ in the $\|\cdot\|$-norm, we can find for any given $h \in \mathfrak{L}_{\text{ket}}$ a sequence $h_1, h_2, \cdots \in \mathcal{H}_D$ converging in that norm to h. By inserting in (4.1) $f = g = h_m - h_n$ and $\Delta = \mathbb{R}^\nu$, we get

$$\|h_m - h_n\|^2 = \int_{\mathbb{R}^\nu} |(\phi_D(\alpha)|h_m\rangle - (\phi_D(\alpha)|h_n\rangle|^2 d\rho_D(\alpha). \qquad (4.3)$$

This indicates that

$$(\phi_D(\alpha)|h\rangle = \rho_D - \text{l.i.m.}\,(\phi_D(\alpha)|h_n\rangle \qquad (4.4)$$

exists and is ρ_D-almost everywhere unique. The so defined functional in $h \in \mathfrak{L}_{\text{ket}}$ is clearly linear ρ_D-almost everywhere and extends the validity of (4.1) to any $f, g \in \mathfrak{L}_{\text{ket}}$.

By using the canonical decomposition (3.6), we can rewrite (2.19) in the form

$$\rho_D(\Delta) = \sum_{n=1}^{\infty} \lambda_n^2 \|E^{A_1,\ldots,A_\nu}(\Delta)u_n\|^2. \qquad (4.5)$$

This relationship explicitly displays the fact that $\rho_D(\Delta) = 0$ for some Borel set Δ in $\mathbb{R}^\nu$ if and only if $E^{A_1,\ldots,A_\nu}(\Delta) = 0$. Hence, for any two $D, D' \in \mathcal{E}$ the measures ρ_D and $\rho_{D'}$ are equivalent. By applying (4.1) to two specific choices of expansions ϕ_D and ϕ_D', we infer that

$$\langle f|\phi_D(\alpha))(\phi_D(\alpha)|u_n\rangle = \langle f|\phi_{D'}'(\alpha))(\phi_{D'}'(\alpha)|u_n\rangle \frac{d\rho_{D'}(\alpha)}{d\rho_D(\alpha)},$$

ρ_D-almost everywhere in $\mathbb{R}^\nu$. According to (4.5), the set

$$\bigcap_{n=1}^{\infty} \Delta_n' = \left(\bigcup_{n=1}^{\infty} \Delta_n\right)', \qquad \Delta_n = \{\alpha|\langle u_n|\phi_D(\alpha))\neq 0\} \qquad (4.6)$$

has ρ_D-measure zero. Hence, the expansion

$$\phi_{D'}(\alpha) = \frac{(\phi_{D'}'(\alpha)|u_n\rangle}{(\phi_D(\alpha)|u_n\rangle}\left(\frac{d\rho_{D'}(\alpha)}{d\rho_D(\alpha)}\right)^{1/2}\phi_{D'}'(\alpha) \qquad (4.7)$$

is well-defined, ρ_D-almost everywhere, and satisfies (4.2) for all $f \in \mathfrak{L}_{\text{ket}}$.

From the point of view of notational convenience we could rewrite (4.1) for $\Delta = \mathbb{R}^\nu$ in the form

$$\int_{\mathbb{R}^\nu} |\phi_D(\alpha))d\rho_D(\alpha)(\phi_D(\alpha)| = \mathbf{1}. \qquad (4.8)$$

We have to emphasize, however, that the above identity operator is defined only on the ket-space $\mathfrak{L}_{\text{ket}}$ since the functional $(\cdot|\cdot)$ cannot be extended to $\mathfrak{L} \times \mathfrak{L}$.

Definition 4.2: Let $\phi_{D_1}(\alpha)$ and $\phi_{D_2}(\alpha)$ be two bra-expansions corresponding to the same complete set $\{A_1, \ldots, A_\nu\}$ and let $A_1^\dagger, A_2^\dagger$ be two operators in the bra-space $\mathfrak{L}$. If $(A_1^\dagger\phi_{D_1}(\alpha)|f\rangle = (A_2^\dagger\phi_{D_2}(\alpha)|f\rangle$ at each $f \in \mathfrak{L}_{\text{ket}}$ for ρ_{D_i}-almost all $\alpha \in \mathbb{R}^\nu$, then we shall write $A_1^\dagger\phi_{D_1} \overset{w}{=} A_2^\dagger\phi_{D_2}$ and say that $A_1^\dagger\phi_{D_1}$ and $A_2^\dagger\phi_{D_2}$ are *weakly equivalent* in $\mathfrak{L}$.

We note that the weak equivalence of $A_1^\dagger\phi_{D_1}$ and $A_2^\dagger\phi_{D_2}$ implies that for ρ_{D_i}-almost all $\alpha \in \mathbb{R}^\nu$ the two bra-vectors $A_1^\dagger\phi_{D_1}(\alpha)$ and $A_2^\dagger\phi_{D_2}(\alpha)$ share the same weak neighborhoods if they belong to a common $\mathcal{L}_D$. The fact that this will not always be the case is part of the obstacle which has stopped us from introducing the concept of weak equivalence for any two elements of $\tilde{\mathcal{L}}$, thus basing the construction of the bra-space $\mathcal{L}$ on weak rather than strong equivalence classes.

Theorem 4.2: If the closed operator B in $\mathcal{L}_{\text{ket}}$ is a function of the complete set $\{A_1, \ldots, A_\nu\}$, i.e., $B = F(A_1, \ldots, A_\nu)$, and $\phi_D(\alpha)$ is a bra-expansion for $\{A_1, \ldots, A_\nu\}$, then

$$(\phi_D(\alpha)|Bf\rangle = F(\alpha)(\phi_D(\alpha)|f\rangle, \tag{4.9}$$

ρ_D-almost everywhere in $\mathbb{R}^\nu$ for any $f \in \mathcal{L}_{\text{ket}}$.

Proof: As a consequence of the definition[14] of a function $F(A_1, \ldots, A_\nu)$ of a set $\{A_1, \ldots, A_\nu\}$ of commuting self-adjoint operators we have

$$E^{A_1,\ldots,A_n}(\Delta)F(A_1, \ldots, A_\nu)h = \int_\Delta F(\alpha)\, dE_\alpha^{A_1,\ldots,A_\nu}h, \tag{4.10}$$

for any $h \in \mathfrak{D}(B)$. Hence, the application of (4.1) yields

$$\begin{aligned}\int_\Delta \langle f|\phi_D(\alpha))(\phi_D(\alpha)|F(A_1, \ldots, A_\nu)h\rangle\, d\rho_D(\alpha) &= \langle f|E^{A_1,\ldots,A_\nu}(\Delta)F(A_1, \ldots, A_\nu)h\rangle \\ &= \int_\Delta F(\alpha)\, d\langle f|E_\alpha^{A_1,\ldots,A_\nu}h\rangle \\ &= \int_\Delta F(\alpha)\langle f|\phi_D(\alpha))(\phi_D(\alpha)|h\rangle\, d\rho_D(\alpha),\end{aligned}$$

where the last step follows by the Radon–Nikodym theorem. Since the above relation holds for any Borel set Δ in $\mathbb{R}^\nu$, the integrands on both sides must be ρ_D-almost everywhere equal. By replacing f with u_n and recalling that the set in (4.6) is of zero ρ_D-measure, we arrive at the conclusion that (4.10) is true.

In the special case $F(A_1, \ldots, A_\nu) = A_k$, we get from Theorem 4.2 that

$$(\phi_D(\alpha)|A_kf\rangle = \alpha_k(\phi_D(\alpha)|f\rangle. \tag{4.11}$$

Hence, if we define a *bra-eigenvector* $h \in \mathcal{L}_{D^*}$ of an operator A with eigenvalue λ by the requirement that it satisfy the relation

$$(h|Af\rangle = \lambda(h|f\rangle, \quad f \in \mathcal{K}_D, \tag{4.12}$$

we can state that the elements of a bra-expansion of $\{A_1, \ldots, A_\nu\}$ are simultaneous bra-eigenvectors of all the operators $A_1, \ldots, A_\nu$.

In light of the definition of the bra-adjoint $A^\dagger$ in Theorem 3.3 and (3.15), the relation (4.10) is equivalent to

$$F(A_1, \ldots, A_\nu)^\dagger\phi_D(\alpha) = \overline{F(\alpha)}\,\phi_D(\alpha) \tag{4.13}$$

if B is a bounded operator.

Definition 4.2: Let $\{A_1, \ldots, A_\nu\}$ be a complete set of observables in $\mathcal{K}$ and let $L^2_\mu(\mathbb{R}^\nu)$ be a spectral representation space[14] for this set, i.e., we assume the existence of a unitary transformation U of $\mathcal{K}$ onto $L^2_\mu(\mathbb{R}^\nu)$ such that

$$(UA_kf)(\alpha) = \alpha_k(Uf)(\alpha), \quad f \in \mathfrak{D}(A_k). \tag{4.14}$$

If A is a closed operator in $\mathcal{K}$, we shall say that A has an *integral representation* in the spectral representation space $L^2_\mu(\mathbb{R}^\nu)$ with kernel $K_A(\alpha, \beta)$ if and only if

$$[UAg](\alpha) = \int_{\mathbb{R}^\nu} K_A(\alpha, \beta)(Ug)(\beta)\, d\mu(\beta) \tag{4.15}$$

for any $g \in \mathfrak{D}(A)$, and the function $K_A(\alpha, \beta)$ belongs to $L^2_\mu(\mathbb{R}^\nu)$ for each fixed $\alpha \in \mathbb{R}^\nu$.

We should emphasize that in the above definition $K_A(\alpha, \alpha')$ is assumed to be a bona fide function. Thus, A_k itself does not have an integral representation in $L^2_\mu(\mathbb{R}^\nu) \equiv L^2(\mathbb{R}^\nu)$ despite the fact that formally (cf. Sec. 7)

$$(UA_kf)(\alpha) = \int_{\mathbb{R}^\nu} \delta(\alpha - \beta)\beta_k(Uf)(\beta)\, d^\nu\beta.$$

Theorem 4.3: Suppose that $\mathcal{E}$ is A-stable. Then A has an integral representation (in the sense of Definition 4.2) in a spectral representation space $L^2_\mu(\mathbb{R}^\nu)$ of the complete set $\{A_1, \ldots, A_\nu\}$ if and only if $A^\dagger\phi_D(\alpha)$ is a ket-vector for μ-almost all $\alpha \in \mathbb{R}^\nu$ and at such $\alpha \in \mathbb{R}^\nu$

$$K_A(\alpha, \beta) = \langle A^\dagger\phi_D(\alpha)|\phi_D(\beta))\frac{d\rho_D(\alpha)}{d\mu(\alpha)} \tag{4.16}$$

at μ-almost all $\beta \in \mathbb{R}^\nu$ for at least one bra-expansion $\phi_D(\alpha)$.

Proof: Since $A^\dagger$ exists by Theorem 3.3, we can use (4.1) to write

$$\langle f|Ag\rangle = \int_{\mathbb{R}^\nu} \langle f|\phi_D(\alpha))(A^\dagger\phi_D(\alpha)|g\rangle\, d\rho_D(\alpha). \tag{4.17}$$

In view of the assumption $A^\dagger\phi_D(\alpha) \in \mathcal{L}_{\text{ket}}$, we can apply (4.1) again:

$$\langle A^\dagger\phi_D(\alpha)|g\rangle = \int_{\mathbb{R}^\nu} \langle A^\dagger\phi_D(\alpha)|\phi_D(\beta))(\phi_D(\beta)|g\rangle\, d\rho_D(\beta). \tag{4.18}$$

The relations (4.14) imply that

$$(UE^{A_1,\ldots,A_\nu}(\Delta)f)(\alpha) = \chi_\Delta(\alpha)(Uf)(\alpha). \tag{4.19}$$

This fact in conjunction with (4.1) implies that μ and ρ_D are equivalent measures. Hence, we can deduce from the relation

$$\begin{aligned}\int_{\mathbb{R}^\nu} \overline{(Uf)(\alpha)}(Ug)(\alpha)\, d\mu(\alpha) &= \langle f|g\rangle \\ &= \int_{\mathbb{R}^\nu} \langle f|\phi_D(\alpha))(\phi_D(\alpha)|g\rangle\, d\rho_D(\alpha)\end{aligned} \tag{4.20}$$

that there is a real μ-measurable function $\omega_D(\alpha)$ such that for any $f \in \mathcal{L}_{\text{ket}}$

$$(Uf)(\alpha) = \left[\frac{d\rho_D(\alpha)}{d\mu(\alpha)}\right]^{1/2} \exp[i\omega_D(\alpha)](\phi_D(\alpha)|f\rangle, \tag{4.21}$$

μ-almost everywhere. Combining (4.17) and (4.18) and comparing the result with (4.15), in light of (4.21) we arrive at the conclusion that

$$\begin{aligned}\int_{\mathbb{R}^\nu} K_A(\alpha, \beta)(Ug)(\beta)d\mu(\beta) &= \left(\frac{d\rho_D(\alpha)}{d\mu(\alpha)}\right)^{1/2} \\ &\times \int_{\mathbb{R}^\nu} (A^\dagger\phi_D(\alpha)|g\rangle(Ug)(\beta)\left(\frac{d\rho_D(\beta)}{d\mu(\beta)}\right)^{1/2} d\mu(\beta)\end{aligned}$$

holds for ρ_D-almost all $\alpha \in \mathbb{R}^\nu$ if among the different bra-expansions corresponding to $D \in \mathcal{E}$ we choose the one for which $\omega_D(\alpha) \equiv 0$. This implies that the above relation is also satisfied by all vectors g in any fixed

countable set $\mathfrak{D}_0 \subset \mathfrak{D}(A)$ at all $\alpha \in \Delta_0$, where $\mu(\mathbb{R}^\nu - \Delta_0) = 0$. By choosing $\mathfrak{D}_0$ dense in $\mathcal{H}$, we can extend the validity of this conclusion to all $g \in \mathcal{H}$ since $K_A(\alpha, \cdot) \in \mathcal{L}^2_\mu(\mathbb{R}^\nu)$, and thus conclude that (4.16) holds.

Conversely, if $K_A(\alpha, \beta)$ exists in the sense of Definition 4.2, then we can combine (4.15) with (4.21) to arrive at the relation

$$(\phi_D(\alpha)|Af\rangle = \left[\frac{d\mu(\alpha)}{d\rho_D(\alpha)}\right]^{1/2} \exp[-i\omega_D(\alpha)] \times \int_{\mathbb{R}^\nu} K_A(\alpha,\beta)\,(Uf)\,(\beta)d\mu(\beta).$$

By applying the Schwarz–Cauchy inequality in $L^2_\mu(\mathbb{R}^\nu)$, we obtain

$$|(A^\dagger\phi_D(\alpha)|f\rangle| \leq \left[\frac{d\mu(\alpha)}{d\rho_D(\alpha)}\right]^{1/2} \left(\int |K_A(\alpha,\beta)|^2\, d\mu(\beta)\right)^{1/2} \|f\|.$$

By Riesz' theorem the above inequality implies that $A^\dagger\phi_D(\alpha) \in \mathcal{L}_{\text{ket}}$. Comparison of (4.15) with (4.18) shows that (4.16) is satisfied.

Let us agree to write for $f, g \in \mathcal{L}$

$$(f|A|g) = \langle A^\dagger f|g\rangle \tag{4.22}$$

for some operator in $\mathcal{L}_{\text{ket}}$ if and only if $A^\dagger f \in \mathcal{L}_{\text{ket}}$. In that case, (4.16) assumes the form

$$K_A(\alpha,\beta) = (\phi_D(\alpha)|A|\phi_D(\beta)) \left\{\frac{d\rho_D(\alpha)}{d\mu(\alpha)}\,\frac{d\rho_D(\beta)}{d\mu(\beta)}\right\}^{1/2}. \tag{4.23}$$

The above results have an immediate bearing to the question of the existence of a Green's function $G_A(\alpha, \beta; \zeta)$ for an operator A at the point ζ in the resolvent set of A. Substituting $R_A(\zeta) = (A - \zeta)^{-1}$ for A in the above considerations, we can apply Theorem 4.3 and arrive at the representation

$$G_A(\alpha,\beta;\zeta) = (\phi_D(\alpha)|R_A(\zeta)|\phi_D(\beta)) \left(\frac{d\rho_D(\alpha)}{d\mu(\alpha)}\,\frac{d\rho_D(\beta)}{d\mu(\beta)}\right)^{1/2} \tag{4.24}$$

for the Green's function of $U(A - \zeta)U^{-1}$ in $L^2_\mu(\mathbb{R}^\nu)$.

5. APPLICATIONS TO TIME-INDEPENDENT SCATTERING THEORY

Let us consider now two self-adjoint operators H and H_0 in the separable Hilbert space $\mathcal{H}$. Physically H could be the total Hamiltonian and H_0 the free Hamiltonian or one of the cluster Hamiltonians (in case that we are dealing with multichannel scattering problem). To avoid inconsequential notational complications, we assume that the wave-operators

$$\Omega_\pm = \operatorname{s}\!-\!\lim_{\epsilon\to+0} \Omega_{\pm\epsilon} \tag{5.1}$$

exist as partial isometries with initial domains $\mathcal{H}$. However, in order to allow for the possibility that long-range forces are present, we assume that[19]

$$\Omega_{\pm\epsilon} = \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H - \lambda \pm i\epsilon} Z_\pm E^{H_0}(d\lambda) = \int_{-\infty}^{+\infty} E^H(d\lambda)\, Z_\pm \frac{\pm i\epsilon}{\lambda - H_0 \pm i\epsilon}, \tag{5.2}$$

where $Z_\pm$ are bounded operators determined[20] by the interactions acting between the particles of the system and

$$H_0 = \int_{-\infty}^{+\infty} \lambda dE^{H_0}_\lambda, \qquad H = \int_{-\infty}^{+\infty} \lambda dE^H_\lambda. \tag{5.3}$$

Suppose that $\{A_1, \ldots, A_\nu\}$ is a complete set of observables and $H_0 = h(A_1, \ldots, A_\nu)$. If $\phi_D(\alpha)$ is a bra-expansion for $\{A_1, \ldots, A_\nu\}$ for which at any $f \in \mathcal{H}$

$$\lim_{\lambda\to h(\alpha)} (\phi_D(\alpha)|Z^*_\pm(H - \lambda \pm i\epsilon)^{-1}f\rangle = (\phi_D(\alpha)|Z^*_\pm(H - h(\alpha) \pm i\epsilon)^{-1}f\rangle \tag{5.4}$$

at ρ_D-almost all $\alpha \in \mathbb{R}^\nu$, then

$$(\phi_D(\alpha)|\Omega^*_{\pm\epsilon}f\rangle = \mp\, i\epsilon(\phi_D(\alpha)|Z^*_\pm(H - \lambda \mp i\epsilon)^{-1}f\rangle. \tag{5.5}$$

The proof of the above relation (5.5) is essentially the same as the proof of Theorem 4.1 in Ref. 2 supplemented by the observation that the restriction of having $\Omega^*_{\pm\epsilon}$ map some $\mathcal{H}_D$ into itself can be dropped when working in $\mathcal{L}$ rather than in $\mathcal{H}^\dagger_D$ since we can work with f in $\mathfrak{D}(D(\Omega^*_{\pm\epsilon} + \zeta))$ at fixed $|\zeta| > \|Z_\pm\|$.

In view of Definition 4.1, the relation (5.5) states that the following relation of weak equivalence holds:

$$\Omega^{*\dagger}_{\pm\epsilon}\phi_D(\alpha) \overset{\text{w}}{=} \pm\, i\epsilon[Z^*_\pm(H - \lambda \mp i\epsilon)^{-1}]^\dagger\phi_D(\alpha). \tag{5.6}$$

Hence, (5.1) implies that

$$\phi^\pm_D(\alpha) = \Omega^{*\dagger}_\pm\phi_D(\alpha) = \mp\, \operatorname{w}\!-\!\lim_{\epsilon\to+0} \text{l.i.m.}\,[i\epsilon Z^*_\pm(H - \lambda \mp i\epsilon)]^\dagger\phi_D(\alpha) \tag{5.7}$$

if we agree on the following notation: if $A^\dagger_1, A^\dagger_2, \cdots$ is a sequence of operators in $\mathcal{L}$, we write

$$A^\dagger\phi_D(\alpha) = \text{w} - \text{l.i.m.}\ A^\dagger_n\phi_D(\alpha) \tag{5.8}$$

if and only if for any $f \in \mathcal{L}_{\text{ket}}$

$$(A^\dagger\phi_D(\alpha)|f\rangle = \rho_D - \lim_{n\to\infty} \text{l.i.m.}\,(A^\dagger_n\phi_D(\alpha)|f\rangle. \tag{5.9}$$

We note that the concept of convergence contained in (5.8) is related to convergence in the weak topology of $\mathcal{L}$ to the extent that (5.9) implies weak convergence of some subsequence of $\{A^\dagger_n\phi_D(\alpha)\}$ to $A^\dagger\phi_D(\alpha)$ for ρ_D-almost all $\alpha \in \mathbb{R}^\nu$.

The equations (5.7) are equivalent to the type II Lippmann–Schwinger equations for distorted waves modified[20] so as to apply also when long-range forces are present. Similar considerations yield the equations

$$\phi_D(\alpha) = \Omega^\dagger_\pm\phi^\pm_D(\alpha) = \mp\, \operatorname{w}\!-\!\lim_{\epsilon\to+0} \text{l.i.m.}\,[i\epsilon Z_\pm(H_0 - \lambda \mp i\epsilon]^\dagger\phi^\pm_D(\alpha). \tag{5.10}$$

An application of the extended Hilbert space formalism also provides a very straightforward method for deciding when the T operator is an integral operator in the spectral representation space of some complete set $\{A_1, \ldots, A_\nu\}$ of observables. In fact, after writing[2]

$$\langle f|Tg\rangle = \lim_{\epsilon\to+0} \frac{\epsilon}{\pi} \int_{\mathbb{R}^\nu} \langle f|\phi_D(\alpha))\,(\phi_D(\alpha)|H_I\Omega_- \times [(H_0 - h(\alpha))^2 + \epsilon^2]^{-1}g\rangle\, d\rho_D(\alpha), \tag{5.11}$$

we conclude by applying Theorem 4.3 [assuming that $\mathcal{E}$ has been chosen $(H_I\Omega_-)$-stable; cf. also Sec. 7] that a further expansion involving g will be possible if and only if $(H_I\Omega_-)^\dagger\phi_D(\alpha) \in \mathcal{L}_{\text{ket}}$ for ρ_D-almost all $\alpha \in \mathbb{R}^\nu$.

Using Theorem 4.2 we arrive at the expression

$$\langle f|Tg\rangle = \lim_{\epsilon\to+0} \frac{\epsilon}{\pi} \int_{\mathbb{R}^\nu} d\rho_D(\alpha)\overline{\hat{f}(\alpha)} \int_{\mathbb{R}^\nu} d\rho_D(\beta) \times [(h(\alpha)-h(\beta))^2+\epsilon^2]^{-1} \langle (H_I\Omega_-)^\dagger\phi_D(\alpha)|\phi_D(\beta))\hat{g}(\beta) \tag{5.12}$$

for the transition operator T, relating this operator to the T matrix

$$T(\alpha,\beta) = (\phi_D(\alpha)|H_I\Omega_-|\phi_D(\beta)) = \langle (H_I\Omega_-)^\dagger\phi_D(\alpha)|\phi_D(\beta)). \tag{5.13}$$

6. FIELDS AT A POINT AS OPERATORS IN EXTENDED FOCK SPACE

Another area of quantum physics in which the extended Hilbert space formalism can be a very handy mathematical tool is quantum field theory. In this section we shall use the concept of expansion in bra-eigenvectors to define the concept of free field at a point. Rigorous formulations of this concept in the context of nested Hilbert spaces[21] and the $\mathcal{L}$ formalism[22] have been previously considered, but their practical applicability is limited by the relative intricacy of the constructions. In contradistinction, the fields at a point to be defined in the sequel have all the desired basic properties and lead to straightforward definitions of practically important objects, such as normally ordered polynomials in the fields, integrals of such polynomials, etc.

We shall work in extended Fock space. Fock space itself is constructed by means of direct sums and tensor products from the single-particle spaces. Hence, if $\mathcal{L}_k$ denotes the bra-space of $\mathcal{K}_k$, let us simply agree to denote by $\oplus_k\mathcal{L}_k$ and $\otimes_k\mathcal{L}_k$ the bra-spaces associated with $\oplus_k\mathcal{K}_k$ and $\otimes_k\mathcal{K}_k$, respectively.

To avoid cumbersome notation, we consider the case when only one kind of particle is involved. Let $\mathcal{F}_1$ denote the one particle space, and $\mathcal{F}_0$ the one-dimensional complex Hilbert space describing the Fock vacuum. We shall let S_n stand for the symmetrizer or antisymmetrizer, depending on whether the particle is a boson or fermion, respectively. Thus, the n-particle component $\mathcal{F}_n$ of the Fock space $\mathcal{F}$ will contain as a dense set the linear manifold of all vectors $f=\oplus_{n=0}^\infty f^{(n)}$ with zero $f^{(m)}$ components for $m\neq n$ and

$$f^{(n)} = S_n(f_1\otimes\cdots\otimes f_n) = \frac{1}{n!}\sum_s (\pm 1)^{\pi(s)} f_{i_1}\otimes\cdots\otimes f_{i_n}, \tag{6.1}$$

where s varies over the permutation group of n objects and $\pi(s)$ is 0 for an even and +1 for an odd permutation $s=(i_1,\ldots,i_n)$. We define the n-particle component $\mathcal{G}_n$ of the Fock bra-space $\mathcal{G}$ to be the closure of the ket-space $\mathcal{F}_n$ in the strong topology of the extended Hilbert space $\oplus_{n=0}^\infty \mathcal{F}_1^{\otimes n}\subset\oplus_{n=0}^\infty \mathcal{G}_1^{\otimes n}$. This definition is consistent with the fact that by Theorem 3.1 the ket-space $\mathcal{F}_1^{\otimes n}$ is dense in the bra-space $\mathcal{G}_1^{\otimes n}$ in the strong topology of that space.

Let $\{A_1,\ldots,A_\nu\}$ be a complete set of observables in the one-particle space $\mathcal{F}_1$ (such as spin and momentum or spin and position, etc.) and let $D\in\mathcal{E}_1$ be an equipping operator for $\mathcal{F}_1$. Denote by $\phi_D(\alpha)$ the corresponding bra-expansion and let $(\phi_D(\alpha)|f\rangle$ be a given extension of the functionals $(\phi_D(\alpha)|\cdot\rangle$, $\alpha\in\mathbb{R}^\nu$, originally defined only for $f\in\mathcal{K}_D\subset\mathcal{F}_1$ to all $f\in\mathcal{F}_1$. We define then a corresponding annihilator $\psi_D(\alpha)$ on vectors in $\mathcal{F}_n$ of the form (6.1) by

$$\psi_D(\alpha)S_n(f_1\otimes\cdots\otimes f_n) = n^{1/2}\sum_s (\pm 1)^{\pi(s)} \times (\phi_D(\alpha)|f_{i_1}\rangle f_{i_2}\otimes\cdots\otimes f_{i_n}. \tag{6.2}$$

The restriction $\psi_D^{(n)}(\alpha)$ of this operator to $\mathcal{F}_n$ can be defined on all of $\mathcal{F}_n$ by setting

$$\psi_D^{(n)}(\alpha) = n^{1/2}n!(\phi_D(\alpha)|\otimes \mathbf{1}_{n-1}, \tag{6.3}$$

where $\mathbf{1}_{n-1}$ is the identity operator on $\mathcal{F}_{n-1}$, i.e., by (4.8)

$$\mathbf{1}_{n-1} = \int |\phi_D(\alpha_1)\otimes\cdots\otimes\phi_D(\alpha_{n-1}))d\rho(\alpha_1)\cdots d\rho(\alpha_{n-1}) \times (\phi_D(\alpha_1)\otimes\cdots\otimes\phi_D(\alpha_{n-1})|. \tag{6.4}$$

It should be emphasized, however, that $\psi_D^{(n)}(\alpha)$ is not a closed linear operator (except when all observables $A_1,\ldots,A_\nu$ have only point spectra) since by the closed graph theorem that would imply that it is bounded, and therefore that $\psi_D(\alpha)$ has an adjoint $\psi_D^*(\alpha)$ in $\mathcal{F}$. It is well-known that this is not the case. The reason why we could extend $\psi_D(\alpha)$ to $\mathcal{F}_n$ is that in defining the functionals $(\phi_D(\alpha)|f\rangle$ for all of $f\in\mathcal{F}_1$ we have isolated one of the many possible extensions of the original functionals in which f varied over $\mathcal{K}_D$.

The bra-adjoint $\psi_D^\dagger(\alpha)$ is easily established to be the creator which acts on vectors in $\mathcal{G}_n$ of the form (6.1) in the following manner:

$$\psi_D^\dagger(\alpha)S_nf_1\otimes\cdots\otimes f_n = (n+1)^{-1/2}S_{n+1}\phi_D(\alpha)\otimes f_1\otimes\cdots\otimes f_n. \tag{6.5}$$

Hence, if so desired, the restriction $\psi_D^{\dagger(n)}(\alpha)$ of $\psi_D^\dagger(\alpha)$ to $\mathcal{G}_n$ can be described as an operator with range in $\mathcal{G}_{n+1}$ by writing

$$\psi_D^{\dagger(n)}(\alpha) = (n+1)^{-1/2}S_{n+1}|\phi_D(\alpha))\otimes\mathbf{1}_n. \tag{6.6}$$

We note that normally ordered polynomials:Φ: of creators and annihilators are perfectly well-defined operators from the ket-space $\mathcal{F}$ to the bra-space $\mathcal{G}$. Moreover, the concept of "smearing" such normal products as

$$:\Phi(\alpha_1,\ldots,\alpha_l,\alpha_{l+1},\ldots,\alpha_{l+m}): = \psi_D^\dagger(\alpha_1)\cdots\psi_D^\dagger(\alpha_l)\psi_D(\alpha_{l+1})\cdots\psi_D(\alpha_{l+m}) \tag{6.7}$$

with any function $\hat{f}_0$ in $L^2_{\rho_D\times(l+m)}(\mathbb{R}^{\nu(l+m)})$ which is appropriately symmetrized, i.e., for which $S_{l+m}\hat{f}_0=\hat{f}_0$,

$$\Phi(\hat{f}_0) = \int_{\mathbb{R}^{\nu(l+m)}}:\Phi(\alpha_1,\ldots,\alpha_{l+m}):\hat{f}_0(\alpha_1,\ldots,\alpha_{l+m})d\rho_D(\alpha_1)\times\cdots d\rho_D(\alpha_{l+m}) \tag{6.8}$$

is defined by Theorem 2.3. The restriction of $\Phi(\hat{f}_0)$ to $\mathcal{F}_l$ is a bounded operator from $\mathcal{F}_n$ to $\mathcal{F}_{l+n-m}$. To see that this last statement is correct, note that by (6.2) and (6.5) if $n\geq m$ and $f_0\in\mathcal{F}_n$ has its component $f^{(n)}$ in $\mathcal{F}_n$ given by (6.1), then

$$\begin{aligned}:\Phi(\alpha_1,\ldots,\alpha_{l+m}):f^{(n)} &= \left(\frac{n(n-1)\cdots(n-m+1)}{(l+n-m+1)\cdots(n-m+1)}\right)^{1/2} \\ &\times S_{l+n-m+1}\cdots S_{n-m+1}\sum_s(\pm 1)^{\pi(s)} \\ &\times (\phi_D(\alpha_{l+1})|f_{i_1}\rangle\cdots(\phi_D(\alpha_{l+m})|f_{i_m}\rangle \\ &\times \phi_D(\alpha_1)\otimes\cdots\otimes\phi_D(\alpha_l)\otimes f_{i_{m+1}}\otimes\cdots\otimes f_{i_n}.\end{aligned} \tag{6.9}$$

The notation is such that each of the above symmetrizers S_j acts on the last j vectors of the tensor product in (6.9).

Let U_1 denote the unitary transformation

$$\hat{f}(\alpha) = (U_1 f)(\alpha) = (\phi_D(\alpha)|f\rangle \tag{6.10}$$

of $\mathfrak{F}_1$ onto $L^2_{\rho_D}(\mathbb{R}^\nu)$. If $U_n = U_1^{\otimes n}$ is the corresponding unitary transformation of $\mathfrak{F}_n$ onto $S_n L^2_{\rho_D{}^{\times n}}(\mathbb{R}^{\nu n})$ and $\hat{g} = U_l g, \hat{h} = U_m h$ for $g \in \mathfrak{F}_l, h \in \mathfrak{F}_m$, then by using (6.9) and (2.23) we easily get

$$\begin{aligned} :\Phi(\hat{g}\otimes\hat{\bar{h}}):\, f^{(n)} &= \left\{\frac{n(n-1)\cdots}{(l+n-m+1)\cdots}\right\}^{1/2} \\ &\times S_{l+n-m+1}\cdots S_{n-m+1}\sum_s (\pm 1)^{\pi(s)} \\ &\times (h|f_{i_1}\otimes\cdots\otimes f_{i_m}\rangle_m\, g\otimes f_{i_{m+1}}\otimes\cdots\otimes f_{i_n}, \end{aligned} \tag{6.11}$$

where $(\cdot|\cdot)_m$ is the form (3.12) in $\mathfrak{F}_m \subset \mathcal{G}_m$. Since this form becomes the inner product on $\mathfrak{F}_m$ when $h \in \mathfrak{F}_m$, the relation (6.11) establishes that $:\Phi(\hat{g}^*\otimes\hat{h}):$ is bounded when restricted to $\mathfrak{F}_n$. Note that the "smeared" fields $\psi_D(\hat{f}_0)$ and $\psi_D^\dagger(\hat{f}_0)$, $\hat{f}_0 \in L^2_{\rho_D}(\mathbb{R}^\nu)$, are special cases of the general expression in (6.8).

In light of the above remarks, this means that, as expected, the smeared fields are bona fide operators in $\mathfrak{F}$ if the smearing function $\hat{f}_0$ belongs to $S_{l+m}L^2_{\rho_D{}^{\times(l+m)}}(\mathbb{R}^{\nu(l+m)})$. Theorem 2.3 enables us to define $\Phi(\hat{f}_0)$ as a unique element of $\mathcal{K}_{D^*}^\dagger$ for general case of $\hat{f} \in \mathcal{P}_{D^{\otimes(l+m)}}(\mathbb{R}^{\nu(l+m)})$. However, in general $m_D^{-1}\Phi(\hat{f}_0)f^{(n)}$ contains more than one element of $\mathcal{G}_D$. By (6.11) we still get uniqueness for $m_D^{-1}\Phi(\hat{f}_0)f^{(n)}$ as long as $f^{(n)} \in \mathcal{K}_D$ and $g \in \mathcal{G}_0 = \cap_D \mathcal{G}_D$. But if $f \in \mathcal{G} - \mathfrak{F}$ the integral in (6.8) will in general depend not only on $\hat{f}$ but on the expanding sequence $\Delta_1 \subset \Delta_2 \subset \cdots$ used in defining $\Phi(\hat{f})f^{(n)}$ as in (2.27). Only in case that $\Phi(\hat{f}_0)f^{(n)} \in \mathcal{G}_{0-} = \cap\, \mathcal{G}_D$ can we be assured of the uniqueness of $\Phi(\hat{f}_0)$ in $\mathcal{G}_D$ by Theorem 3.3.

Since $\psi_D(\alpha)$ is an operator on $\mathfrak{F}$ and not on the entire bra-space $\mathcal{G}$, objects which are not normally ordered [e.g., $\psi_D(\alpha)\psi_D^\dagger(\beta)$] are not well-defined as operators from $\mathcal{G}$ to $\mathcal{G}$. This is to be expected since (anti-) commutators such as $[\psi_D(\alpha), \psi_D^\dagger(\beta)]_\pm$ contain δ functions rather than only bona fide functions. In fact, we easily infer from (6.2) and (6.5) by the usual algebraic manipulations that

$$[\psi_D(\alpha), \psi_D^\dagger(\hat{f})]_\pm = (\phi_D(\alpha)|f\rangle = \hat{f}(\alpha) \tag{6.12}$$

for any $\hat{f} \in L^2_{\rho_D}(\mathbb{R}^\nu)$. The above relation reflects the δ-function nature of the above (anti-) commutator, which is, however, well defined for smeared-out creators since $\psi_D^\dagger(\hat{f})$ is an operator from $\mathfrak{F}$ to $\mathfrak{F}$.

Let $F(A_1, \ldots, A_\nu)$ be a self-adjoint operator-valued function of $A_1, \ldots, A_\nu$. Using (6.3) and (6.6) we easily establish that

$$\int_{\mathbb{R}^\nu} F(\alpha)\psi^\dagger(\alpha)\psi(\alpha)d\rho_D(\alpha) = \bigoplus_{n=0}^{\infty}\sum_{k=1}^{n+1} 1_k \otimes F(A_1, \ldots, A_\nu) \otimes 1_{n-k}. \tag{6.13}$$

For example, the particle-number operator is obtained as a special case for $F(\alpha) \equiv 1$. Other objects, such as the global energy, momentum, angular momentum, etc., are well-known special cases of (6.13).

7. DISCUSSION

The main purpose of introducing the concept of extended Hilbert space is to supply a mathematical framework which is rigorous and at the same time simple and easy to apply to problems in quantum physics. The facility in applying a formalism depends to a great extent on whether certain key operations can be performed with ease under a variety of circumstances without being restricted by a load of conditions which have to be met to validate such applications and which might be difficult or impossible to verify in practice. In our case we deem that the operation of taking the bra-adjoint of an operator is such a key operation. We note that Theorem 3.3 which deals with this operation does require the condition of A-stability to be met as a precondition for its guaranteeing the existence of $A^\dagger$. We have shown how in the problems treated in this paper that condition can be easily satisfied. However, in the long run it might turn out that this condition severely hampers the applicability of the formalism. Therefore, we present here an alternative definition of an extended Hilbert space (see also Ref. 23).

It is easily seen that the relation $\{f_n\} \cong \{g_n\}$ by which two sequences from $\tilde{\mathcal{L}}$ (for given $\mathcal{E}$) are identified iff $\|f_n - g_n\| \to 0$ as $n \to \infty$ is an equivalence relation. Let us denote by $\mathcal{B}$ the family of all corresponding equivalence classes. After identifying each element f in $\mathcal{K}$ with the equivalence class containing the sequence $f, f, \cdots$, we obtain a subset $\mathcal{K}$ of $\mathcal{B}$ as the image of this identification. The *alternative extended Hilbert space* consists now of the ordered pair $\mathcal{K} \subset \mathcal{B}$, where $\mathcal{K}$ and $\mathcal{B}$ play the role of alternative ket and bra spaces, respectively.

It is easy to see that now for any operator A which has an adjoint A^* and for which $\{A^*f_n\}$ and $\{A^*g_n\}$ belong to $\tilde{\mathcal{L}}$, we shall have that $\{A^*f_n\} \cong \{A^*g_n\}$ whenever $\{f_n\} \cong \{g_n\}$ due to the fact that A^* is closed.[14] Hence the bra-adjoint $A^\dagger$ of A will be always unambiguously defined so that the A-stability of $\mathcal{E}$ is not required.

Strong and weak topologies can be introduced in $\mathcal{K} \subset \mathcal{B}$ in the same manner as in $\mathcal{L}_{\text{ket}} \subset \mathcal{L}$, and the construction of $(\cdot|\cdot\rangle$ proceeds in the same manner. All the theorems of Secs. 3 and 4 stay true without even having to require the A-stability of $\mathcal{E}$. Hence, from the point of view of applications, $\mathcal{K} \subset \mathcal{B}$ seems to have marked advantages over $\mathcal{L}_{\text{ket}} \subset \mathcal{L}$.

The disadvantages of $\mathcal{K} \subset \mathcal{B}$ in comparison with $\mathcal{L}_{\text{ket}} \subset \mathcal{L}$ seem to be more of a pure mathematical nature. In $\mathcal{K} \subset \mathcal{B}$ the strong topology is not Hausdorff. Furthermore, $\mathcal{K} \subset \mathcal{B}$ does not have the nice feature of reducing to $\mathcal{K} \subset \mathcal{K}_D^\dagger$ in the special case when $\mathcal{E} = \{D^*\}$, so that equipped Hilbert spaces are not special cases of alternative extended Hilbert cases.

One object which plays a prominent role in the formal manipulations with the bra and ket formalism but has not been given a rigorous counterpart in the preceding sections is the "inner product" $(f|g)$ of two bra-vectors f, g in $\mathcal{L}$ (or $\mathcal{B}$). The obvious reason for this intentional omission is that *in general* such an object cannot have a meaning as a *complex number* [e.g., witness the "inner product" of a plane wave $\exp(i\mathbf{kr})$ with itself, which yields a divergent integral]. However, a rigorous meaning can be given to the symbol $(\phi_D^{(1)}(\alpha)|\phi_D^{(2)}(\beta))$, where $\phi_D^{(1)}(\alpha)$ and $\phi_D^{(2)}(\beta)$ are any two bra-expansions, since such an object can be treated as a *functional* on $L^2_{\rho_D(1)} \times L^2_{\rho_D(2)}$. We simply set by definition

$$\int \overline{(U_D^{(1)}f)(\alpha)}\,(\phi_D^{(1)}(\alpha)|\,\phi_D^{(2)}(\beta))\,(U_D^{(2)}g)(\beta)d\rho_D{}^{(1)}(\alpha)d\rho_D{}^{(2)}(\beta) = \langle f|g\rangle, \tag{7.1}$$

where $U_D^{(1)}$ and $U_D^{(2)}$ are defined by (2.20). Such a definition will obviously allow the above-mentioned formal manipulations, but might be easily deemed to be of a trivial nature. It has to be remembered, however, that any bra-vector is a limit of ket-vectors in the strong topology of the extended Hilbert space. Hence (7.1) is equal to

$$\lim_{n\to\infty} \int \overline{(U_D f)(\alpha)} \langle \phi_D^{(1)}(\alpha;n) | \phi_D^{(2)}(\beta)) \times (U_D^{(2)} g)(\beta) d\rho_D^{(1)}(\alpha) d\rho_D^{(2)}(\beta), \quad (7.2)$$

where $\{\phi_D^{(1)}(\alpha;n)\}_{n=1}^{\infty}$ is some sequence approximating $\phi_D^{(1)}(\alpha)$. We note that each

$$\Gamma_n(\alpha,\beta) = \langle \phi_D^{(1)}(\alpha;n) | \phi_D^{(2)}(\beta)) \quad (7.3)$$

is a perfectly well-defined complex function, which thus provides an approximation of the functional (7.1). Moreover, in some cases it might happen that

$$\Gamma_\infty(\alpha,\beta) = \lim_{n\to\infty} \Gamma_n(\alpha,\beta) \quad (7.4)$$

exists $\rho_D^{(1)} \times \rho_D^{(2)}$ almost everywhere (as, for example, in the case of the expansions for position and momentum in quantum mechanics).

We can generalize (7.1) to give meaning to the symbol $(\phi_D^{(1)}(\alpha)|A|\phi_D^{(2)}(\beta))$ by setting

$$\int (\phi_D^{(1)}(\alpha)|A|\phi_D^{(2)}(\beta))(U^{(2)}g)(\beta) d\rho_D^{(2)}(\beta) = \lim_{n\to\infty} \int \langle A^* \phi_D^{(1)}(\alpha;n) | \phi_D^{(2)}(\beta)) (U^{(2)}g)(\beta) d\rho_D^{(2)}(\beta) \quad (7.5)$$

for $\rho_D^{(1)}$-almost all α. The above definition extends the scope of the notation encountered in (4.23), which now corresponds to the special case when $\phi_D^{(1)} \equiv \phi_D^{(2)}$.

*Supported in part by the National Research Council of Canada.

[1]Ju. M. Berezanskiĭ, *Expansions in Eigenfunctions of Self-Adjoint Operators*(American Mathematical Society, Providence, Rhode Island, 1968).

[2]E. Prugovečki, "Eigenfunction expansions for stationary scattering theory in spaces with negative norm" (University of Toronto preprint).

[3]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U.P., Oxford, 1930).

[4]C. Foiaş, Rev. Math. Pures et Appl. **7**, 241 (1962); Rev. Math. Pures et Appl. **7**, 571 (1962); Acta Szeged. **20**, 117 (1959).

[5]M. E. Mayer, *Proceedings Seminar on Unified Theories of Elementary Particles.* (Max-Planck Institute, Munich, 1965).

[6]J. E. Roberts, J. Math. Phys. **7**, 1097 (1966); Commun. Math. Phys. **3**, 98 (1966).

[7]D. Böhm, "Rigged Hilbert space and mathematical description of physical systems" in *Lectures in Theoretical Physics*, edited by W. E. Brittin, A. O. Barut, and M. Guenin (Gordon and Breach, New York, 1967), Vol. 9A.

[8]J.-P. Antoine, J. Math. Phys. **10**, 53 (1969).

[9]O. Melsheimer, "Nuclear spaces, rigged Hilbert spaces and applications to quantum systems," Habilitationsschrift, Philipps-Universität, Marburg/L., 1971; see also papers to appear in J. Math. Phys.

[10]N. Bogoliubov, V. Lagunov, and I. Todorov, *The Axiomatic Approach to Quantum Field Theory* (Izdatel'stoo 'Nauka, Moskow, 1969) (in Russian).

[11]D. Babbitt, Rep. on Math. Phys. **3**, 1 (1972).

[12]A. Grossmann, Commun. Math. Phys. **2**, 1 (1966).

[13]In the Russian literature "rigged" and "equipped" Hilbert spaces are referred to by the same term, namely "osnashchennoe gil'bertovoe prostranstoo". However, the Gel'fand-Vilenkin definition of "rigged" Hilbert space is distinct from Berezanskiĭ's definition of "equipped" Hilbert space. In the present work this distinction is very essential and therefore we completely divorce these two concepts from one another.

[14]E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971).

[15]E. Prugovečki, Can. J. Math. **21**, 158 (1969).

[16]E. Prugovečki and A. Tip, "On ergodic limits of normal states in quantum statistical mechanics" (preprint).

[17]S. Willard, *General Topology* (Addison-Wesley, Reading, Massachusetts, 1970).

[18]A. P. Robertson and W. J. Robertson, *Topological Vector Spaces* (Cambridge, U. P., Cambridge, 1964).

[19]E. Prugovečki, J. Math. Phys. **14**, 957 (1973).

[20]E. Prugovečki and J. Zorbas, J. Math. Phys. **14**, 1398 (1973).

[21]A. Grossmann, Commun. Math. Phys. **3**, 203 (1967).

[22]E. Prugovečki, J. Math. Phys. **10**, 933 (1969).

[23]E. Prugovečki, "Extended Hilbert space formulation of Dirac's bra and ket formalism and its applications to abstract stationary scattering theory" in *Proceedings of the Advanced Study Institute at Marburg, March-April, 1973*, Springer Lecture Notes, edited by A. Hartkämper and H. Neumann, (Springer, New York) (to appear).

Tensor operators of finite groups*

A. J. van Zanten[†] [‡] and E. de Vries[†]

Physics Department, Duke University, Durham, North Carolina 27706
(Received 22 March 1973)

Irreducible tensor operators of a finite group G are discussed as elements of the group algebras of $G^{\otimes n}$ ($n = 1,2,3, \cdots$). Formulas are given for the number of times that an irreducible tensor operator of a certain rank can be constructed in these algebras and more specifically how often they can be constructed from the elements of a fixed class of conjugate elements in $G^{\otimes n}$ ($n = 1,2,3, \cdots$). Some of results are interpreted in the framework of the duality between classes and irreducible representations in finite groups.

1. INTRODUCTION

Recently there has been much discussion in the literature concerning the possibility of constructing irreducible tensor operators in the group algebra of a finite group (Refs. 1-6). In Ref. 1 Gamba gave a preliminary exploration of the possibility of constructing such tensor operators. In Refs. 2-4 Killingbeck exploited this idea somewhat further, mainly in view of applications in crystal field theory. In Ref. 5 De Vries and Van Zanten showed that the Dirac matrices can be considered as irreducible tensor operators of the Dirac matrix group. This idea was used to study Fierz transformations, which play a role in weak interaction theory. Yet another application of this same basic concept can be given in treating the quantitative splitting of a degenerate energy level under the action of a symmetry-breaking Hamiltonian, which was recently discussed in group theoretical terms by Biedenharn and Gamba in Ref. 7.

Most of the discussions of tensor operators constructed in the group algebra are intuitive and not exhaustive. In Sec. 2 of this paper we shall discuss the precise relationship between these tensor operators and the conventional tensor operators as used, for example, in the theory of angular momentum, or, more generally, in the theory of $SU(n)$ (cf. Ref. 8). It will be shown that tensor operators constructed in the group algebra of a finite group $\mathcal{G}$ suffer from a lack of generality. In the first place it is not possible in general to construct irreducible tensor operators of an arbitrary rank in the group algebra and secondly the nondiagonal reduced matrix elements of those tensor operators which can be constructed in the group algebra all vanish. To avoid these restrictions one can embed $\mathcal{G}$ in a larger group and try to construct tensor operators of $\mathcal{G}$ in the group algebra of this larger group. In Sec. 3 we take for such larger group direct products $\mathcal{G}^{\otimes n}$ ($n = 2, 3, 4, \ldots$) of $\mathcal{G}$ with itself. This turns out to be a natural and powerful embedding, which gives rise to more general tensor operators, which have nondiagonal reduced matrix elements unequal to zero.

In Sec. 4 we discuss problems concerning the number of times that an irreducible tensor operator of a certain rank can be constructed in the group algebras of $\mathcal{G}^{\otimes n}$ ($n = 1, 2, 3, \cdots$) and more specifically how often they can be constructed from the elements of a fixed class of conjugate elements in these groups. Furthermore, we prove that the irreducible representations of $\mathcal{G}$ which correspond to tensor operators constructible in the above mentioned group algebras, are just the irreducible representation of the quotient group $\mathcal{G}/\mathcal{Z}$, where $\mathcal{Z}$ is the centre of $\mathcal{G}$.

In Sec. 5 we give some explicit constructions, to illustrate the techniques and some of the properties discussed in Secs. 2, 3 and 4.

Finally, in Sec. 6, we interpret some results of the previous sections in the framework of the duality between classes and irreducible representation in finite groups, as recently discussed by Gamba, Robinson, De Vries and Van Zanten (see Refs. 1, 9, 10, and 11). It will be shown that the properties of irreducible tensor operators constructible in the group algebras of $\mathcal{G}^{\otimes n}$ ($n = 1, 2, 3, \ldots$) can be cast in a form, strikingly dual to properties of commutators in $\mathcal{G}$. A list of dual items concerning irreducible tensor operators and commutators has been added at the end of Sec. 6.

2. PRELIMINARY REMARKS

In the literature one can find the concept of a tensor operator defined with various degrees of precision. In this paper we shall use the following rigorous definition (cf. Refs. 8, 12-15). Let $\mathcal{G}$ be a compact group with elements $R, S, \ldots$. Let $R \to U_R$ be a representation of $\mathcal{G}$ by operators on a Hilbert space $\mathcal{U}$. A tensor operator T with respect to $\mathcal{G}$ is a set of operators

$$T = \{T_\alpha : \alpha = 1, 2, \ldots, q\}, \tag{1}$$

the elements of which are transformed into linear combinations of themselves according to the transformation law

$$U_R T_\alpha U_R^{-1} = \sum_{\beta=1}^{q} D_{\beta\alpha}(R) T_\beta, \tag{2}$$

for all $\alpha = 1, 2, \ldots, q$ and for all $R \in \mathcal{G}$. The matrices $D(R)$ provide then a matrix representation of $\mathcal{G}$. We suppose that the product of the operators on the lhs of Eq. (2) and the linear combination of the operators on the rhs of Eq. (2) have a meaning in the usual sense as products and sums of operators acting in the Hilbert space $\mathcal{U}$. If the representation matrices $D(R)$ happen to be the matrices of an irreducible representation (j) the tensor operator is called an irreducible tensor operator

$$T^j = \{T^j_m : m = 1, 2, \ldots, [j]\}, \tag{3}$$

where $[j]$ is the dimension of the irreducible representation (j). Instead of Eq. (2) one has now

$$U_R T^j_m U_R^{-1} = \sum_{m'} D^{(j)}_{m'm}(R) T^j_{m'}. \tag{4}$$

If one considers irreducible tensor operators which are elements of the group algebra (Ref. 16, p. 108), one uses the following definition (Ref. 1-6):

$$R T^j_m R^{-1} = \sum_{m'} D^{(j)}_{m'm}(R) T^j_{m'}. \tag{5}$$

for all elements $R \in \mathcal{G}$ and all m.

The product on the lhs and the summation on the rhs of Eq. (5) now have sense in the group algebra $A(\mathcal{G})$. It is

obvious that this definition differs from Eq. (4). If one takes for R and T^j_m of Eq. (5) representation matrices of an arbitrary representation of the group (which gives, of course, also a representation of the group algebra), we arrive at a formula which is just a special case of the conventional definition (Eq. (4)) of irreducible tensor operators. However, manipulating in this way we shall never obtain the most general form of an irreducible tensor operator. The tensor operators which we get by using Eq. (5) with T^j_m taken from the group algebra always possess the property that their representations are reduced to block form as soon as one reduces the representation of the elements R to block form, owing to the fact that the representation of the group algebra is reduced simultaneously with the representation of the group. It is then apparent that the reduced matrix elements $\langle j' \| T^j \| j'' \rangle$ are equal to zero for $(j') \neq (j'')$. This is what we meant by saying that a tensor operator which is an element of the group algebra never gives rise to the most general form of a tensor operator as defined by Eq. (4) (cf. also Ref. 2).

If we restrict ourselves to tensor operators which are elements of the group algebra, the matrix elements $\langle j'm' | T^j_m | j''m'' \rangle$ vanish identically unless $(j') = (j'')$ and $\{j'jj'\} \neq 0$. [The $3j$ symbol $\{j_1 j_2 j_3\}$ denotes the number of times that the irreducible representation (j_3) is contained in the Kronecker product $(j_1) \otimes (j_2)$ of the irreducible representations (j_1) and (j_2), see Ref. 16, p. 148.] From this we see that only such tensor operators T^j_m can possibly occur in the group algebra if there exists a representation (j') such that

$$(j) \subset (j') \otimes (j'^*), \tag{6}$$

where (j'^*) denotes the complex conjugate representation of (j'). Equation (6) is a necessary condition for T^j to be constructed in the group algebra $A(\mathcal{G})$. In Sec. 4 we prove that Eq. (6) is also a sufficient condition.

3. CONSTRUCTIBILITY OF IRREDUCIBLE TENSOR OPERATORS

In Eq. (5) the T^j_m are elements of the group algebra and thus linear combinations of the group elements. The coefficients in these expressions depend on the matrix elements $D^{(j)}_{m'm}(R)$, but are unknown *a priori*. In order to study how to determine these coefficients, we take instead of a linear combination just one group element C from a class of conjugate elements $\mathcal{C}_k$ and see how it is transformed under the operator

$$\rho_R(C) = RCR^{-1}. \tag{7}$$

If R runs through the set of group elements, RCR^{-1} runs through the elements of the class $\mathcal{C}_k$. From this it is clear that the elements of a fixed class transform among each other under the operations ρ_R. As is known, the ρ_R form a group (the group of inner automorphisms) if one defines the product $\rho_R \rho_S$ as the operator ρ_{RS} (see Ref. 17). This group is isomorphic to the quotient group $\mathcal{G}/\mathcal{Z}$, where $\mathcal{Z}$ denotes the center of $\mathcal{G}$ ($\mathcal{Z} \in \mathcal{Z}$ if $RZR^{-1} = \mathcal{Z}$, $\forall R \in \mathcal{G}$).

We see that under the transformations of Eq. (7) the g_k elements of the class $\mathcal{C}_k$ can be considered as the basis elements of a representation space of the group of inner automorphisms. This representation is a permutation representation. (cf. Ref. 18, § 187), i.e., every row and every column contains one element equal to 1, whereas all others equal 0. Because of the fact that $\mathcal{G}$ is homorphic to $\mathcal{G}/\mathcal{Z}$, this permutation representation is at the same time a representation of $\mathcal{G}$. In general this representation is reducible. If we decompose it into irreducible constituents then in the representation space we get linear combinations of the elements of the class $\mathcal{C}_k$, which are the basis elements of an irreducible representation space under the operations ρ_R of Eq. (7). It is clear that we get irreducible tensor operators in $A(\mathcal{G})$ in this way. From these considerations it is also evident that the tensor operator T^j_m can only appear as tensor operator in $A(\mathcal{G})$, if the irreducible representation (j) of $\mathcal{G}$ is at the same time an irreducible representation of the group $\mathcal{G}/\mathcal{Z}$. This implies

$$\chi^{(j)}(Z) = [j], \quad \forall Z \in \mathcal{Z}, \tag{8}$$

where $\chi^{(j)}(Z)$ is the character of Z in (j).

This again is a necessary condition for T^j to be an irreducible tensor operator in $A(\mathcal{G})$. However, Eq. (8) is not a sufficient condition although it is difficult to find a counter example. The only counter example we could find is a finite group of order 6048, called $U_3(3)$ in the mathematical literature. (Cf. Ref. 19, where a character table of this group is given.) For this group Eq. (8) is satisfied, but one can verify by using Eq. (6) or Eq. (13) that the six-dimensional irreducible representation of this group does not correspond to an irreducible tensor operator constructible in $A(\mathcal{G})$. In Ref. 6 an example is given illustrating how the irreducible tensor operators of the symmetric group S_3 can explicitly be found by reducing the permutation representation generated by the elements of a class.

We already pointed out that the tensor operators in the group algebra $A(\mathcal{G})$ are of a restrictive nature, because it is not possible to construct tensor operators corresponding to all (j) and those that can be constructed have the property that only their diagonal reduced matrix elements can be nonzero. To avoid these restrictions one can try to embed the group $\mathcal{G}$ in a larger group $\overline{\mathcal{G}}$ and construct tensor operators of $\mathcal{G}$ in the group algebra of $\overline{\mathcal{G}}$.

A natural and powerful embedding, which readily suggests itself is to embed $\mathcal{G}$ in $\mathcal{G} \otimes \mathcal{G}$, the direct product of $\mathcal{G}$ with itself and more generally in $\mathcal{G} \otimes \mathcal{G} \otimes \ldots \otimes \mathcal{G} = \mathcal{G}^{\otimes n}$ $(n = 2, 3, \ldots)$. In particular, the isomorphism $\mathcal{G} \cong \mathcal{G}^*$ $\mathcal{G}^* \ldots {}^*\mathcal{G} = \mathcal{G}^{*n}$ is considered, where $\mathcal{G}^{*n}$ denotes the diagonal subgroup of $\mathcal{G}^{\otimes n}$, consisting of the elements $(R, R, \ldots, R)$ of $\mathcal{G}^{\otimes n}$. We now consider tensor operators T^j of the group $\mathcal{G}$, which are elements of the group algebra $A(\mathcal{G} \otimes \mathcal{G} \otimes \ldots \otimes \mathcal{G}) = A(\mathcal{G}^{\otimes n})$. To this end we generalize the defining relation (5) as follows:

$$(R, R, \ldots, R)\, T^j_m\, (R, R, \ldots, R)^{-1} = \sum_{m'} D^{(j)}_{m'm}(R) T^j_{m'}, \tag{9}$$

where $(R, R, \ldots, R)$ is an element of $\mathcal{G}^{*n}$ and $D^{(j)}(R)$ is again the representation matrix of the element R of $\mathcal{G}$ in the representation (j). In the next section we shall see that in this way one can construct *all* tensor operators transforming as irreducible representations of the group $\mathcal{G}/\mathcal{Z}$ (i.e., those representations for which Eq. (8) holds). Moreover, the nondiagonal reduced matrix elements $\langle j' \| T^j \| j'' \rangle$ will in general no longer be zero.

Another aspect of the embedding of $\mathcal{G}$ in $\mathcal{G}^{\otimes n}$ is that it leads one to consider a certain kind of duality between classes and irreducible representations. This will be discussed in detail in Sec. 6. To construct tensor operators corresponding to an irreducible representation (j) of $\mathcal{G}$ which is not a representation of $\mathcal{G}/\mathcal{Z}$, one has to embed $\mathcal{G}$ in other groups than $\mathcal{G}^{\otimes n}$. We shall give an

example of such an embedding in Sec. 5. We have not been able to solve the general problem of finding a group $\bar{\mathcal{G}}$ in which $\mathcal{G}$ can be embedded such that all tensor operators of $\mathcal{G}$ can be constructed in the group algebra of $\bar{\mathcal{G}}$.

Because of the fact that $\mathcal{G}/\mathcal{Z}$ is isomorphic to the group of inner automorphisms we might call tensor operators which can be constructed in $A(\mathcal{G}^{\otimes n})$ inner tensor operators. Tensor operators which cannot be constructed in $A(\mathcal{G}^{\otimes n})$, we might call outer tensor operators. Whether there is any relationship with outer automorphisms is not known to us.

4. NUMBER OF TIMES TENSOR OPERATORS OCCUR

We now want to discuss the problem whether an irreducible tensor operator of rank j can actually be constructed in $A(\mathcal{G})$ from the elements of a fixed class $\mathcal{C}_k$ and more specifically how often it can be constructed from these elements (cf. also Ref. 1). We define $a_1(j;k)$ to be the number of independent ways that a tensor operator of rank j can be constructed from the elements of $\mathcal{C}_k$. Then one has

$$a_1(j;k) = \frac{1}{g}\sum_{R\in\mathcal{G}} \chi^{(j)}(R)\chi^{\mathcal{C}_k^*}(R) = \frac{1}{g}\sum_i g_i \chi_i^{(j)} \chi_i^{\mathcal{C}_k^*}. \quad (10)$$

In this equation $\chi^{\mathcal{C}_k}(R)$ denotes the character of the element R in the permutation representation generated from the class $\mathcal{C}_k$, whereas g is the number of elements of $\mathcal{G}$. It can readily be shown that $g_i\chi_i^{\mathcal{C}_k}$ is equal to the number of commuting pairs of elements from $\mathcal{C}_k$ and $\mathcal{C}_i$. We shall call this number $N_{ik} = N_{ki}$. Equation (10) now becomes

$$a_1(j;k) = \frac{1}{g}\sum_i N_{ik}\chi_i^{(j)}. \quad (11)$$

This equation was already given by Gamba (Ref. 1).

Now for $(j) = (1_1)$ (the trivial representation) the summation in Eq. (11) can be performed and gives

$$a_1(1_1;k) = \frac{1}{g}\sum_i N_{ik} = \frac{1}{g}\left(g_k\frac{g}{g_k}\right) = 1. \quad (12)$$

From this we see that a tensor operator of rank (1_1), can be constructed exactly once from the class $\mathcal{C}_k$. (This tensor operator is the class sum.)

From Eq. (11) it follows more generally that an irreducible tensor operator of rank j occurs at most $[j]$ times in a class $\mathcal{C}_k$, because

$$a_1(j;k) = \frac{1}{g}\sum_i N_{ik}\chi_i^{(j)} \leqq [j]\frac{1}{g}\sum_i N_{ik} = [j]. \quad (13)$$

In the case of simply reducible groups (Refs. 12 and 16) we can express the number N_{ik} in terms of $6j$ symbols. One finds

$$N_{ik} = \frac{1}{g}\sum_{j'j_1j_2} (-1)^{j_1+j_2+2j'}[j']\begin{Bmatrix} j'j'j_1 \\ j'j'j_2 \end{Bmatrix} g_i g_k \chi_i^{(j_1)}\chi_k^{(j_2)}. \quad (14)$$

[See Ref. 20, Eq. (33).]

Hence

$$a_1(j;k) = \frac{g_k}{g}\sum_{j'j_2} (-1)^{j+j_2+2j'}[j']\begin{Bmatrix} j'j'j \\ j'j'j_2 \end{Bmatrix}\chi_k^{(j_2)}. \quad (15)$$

From this formula we see again that tensor operators of rank j can occur in $A(\mathcal{G})$ only if there exists a representation (j') with the property that $(j) \in (j')\otimes(j')$– note that for a simply reducible group $(j'^*) = (j')$. In particular, tensor operators corresponding to half integer representations (see Refs. 12 and 16) of a simply reducible group do not occur in the group algebra $A(\mathcal{G})$. For the case of the Dirac matrix group and the quaternion group this was already observed in Ref. 5. For an arbitrary one-dimensional representation (1_i) we can apply Eq. (1) of Ref. 21 giving

$$a_1(1_i;k) = \frac{g_k}{g}\sum_{jj_2} (-1)^{(1_i)+j_2+(1_i)\otimes j_2}\{jj1_i\}\{jjj_2\}\chi_k^{(j_2)}. \quad (16)$$

We shall now give a criterion of a somewhat different nature with which one can check whether a tensor operator of rank j can be constructed from the elements of a class $\mathcal{C}_k$. First we present the following theorem.

Theorem 1: Let $\mathfrak{N}(C)$ be the normalizer of the element C from the class $\mathcal{C}_k$. The number of times that an irreducible tensor operator of rank j can be constructed from the elements of $\mathcal{C}_k$ is equal to the number of times that the trivial representation (1_1) of $\mathfrak{N}(C)$ is contained in the irreducible representation (j) of $\mathcal{G}$ restricted to the subgroup $\mathfrak{N}(C)$.

For the proof of this theorem we refer to p. 207 of Ref. 18. Our Theorem 1 is in fact part of the contents of Theorem II of that paragraph. If one takes for the subgroup H of Theorem II the normalizer $\mathfrak{N}(C)$ (i.e., the subgroup of elements of $\mathcal{G}$, which commute with C) the above theorem follows.

Theorem 1 gives rise to the following corollary.

Corollary: A tensor operator of rank j can be constructed from the elements of the class $\mathcal{C}_k$ if and only if the irreducible representation (j) of $\mathcal{G}$ restricted to the subgroup $\mathfrak{N}(C)$ contains the trivial representation (1_1) of $\mathfrak{N}(C)$ or

$$\sum_{R\in\mathfrak{N}(C)} \chi^{(j)}(R) > 0.$$

This criterion can be applied easily for practical calculations. A direct proof of this corollary can be given by using Eq. (39) of Sec. 5. (cf. Ref. 6).

Next we shall derive an expression for the total number of times that a tensor operator of rank j can be constructed in the group algebra $A(\mathcal{G})$. If one sums Eq. (11) over k, one gets

$$a_1(j) = \sum_k a_1(j;k) = \frac{1}{g}\sum_{i,k} N_{ik}\chi_i^{(j)} = \sum_i \chi_i^{(j)} \quad (17)$$

[see also Eq. (12)].

This equation is very useful for practical applications. To decide whether an irreducible tensor operator of rank j appears in the group algebra one only has to check whether $\sum_i \chi_i^{(j)} > 0$.

Equation (17) can be brought into a different form, which has some advantages from a theoretical point of view:

$$\begin{aligned} a_1(j) &= \sum_i \chi_i^{(j)} = \sum_i\left(\frac{1}{g}\sum_{j'} g_i\chi_i^{(j')*}\chi_i^{(j')}\right)\chi_i^{(j)} \\ &= \sum_{j'}\frac{1}{g}\sum_R \chi^{(j')*}(R)\chi^{(j')}(R)\chi^{(j)}(R) \\ &= \sum_{j'}\{j'jj'\} = \sum_{j'}\{j'j'^*j\}. \end{aligned} \quad (18)$$

We see from Eq. (18) that the necessary condition of Eq. (6) for constructing the tensor operator T^j in the group algebra $A(\mathcal{G})$ is sufficient as well.

Similarly, we can derive expressions for the total number of times that a tensor operator of rank j can be constructed in the group algebra $A(\mathcal{G}^{\otimes n})$ $(n = 2, 3, \ldots)$. In the case of $\mathcal{G} \otimes \mathcal{G}$, e.g., one has as classes the sets of elements (R, S), where R is from $\mathcal{C}_k$ and S from $\mathcal{C}_l$. We shall denote such a class by the symbol $(\mathcal{C}_k, \mathcal{C}_l)$. Once again the elements of the set $(\mathcal{C}_k, \mathcal{C}_l)$ can be considered as the basis vectors of a permutation representation of the group $\mathcal{G}$. The character of the element R of $\mathcal{G}$ in this representation is equal to

$$\chi^{(\mathcal{C}_k, \mathcal{C}_l)}(R) = \chi^{\mathcal{C}_k}(R)\chi^{\mathcal{C}_l}(R). \tag{19}$$

If R is from $\mathcal{C}_i$ one has

$$\chi^{(\mathcal{C}_k, \mathcal{C}_l)}(R) = \frac{1}{g_i^2} N_{ik}N_{il} \tag{20}$$

[cf. the derivation of Eq. (12)].

From this it follows that the number of times that a tensor operator of rank j can be constructed from the class $(\mathcal{C}_k, \mathcal{C}_l)$ in $A(\mathcal{G} \otimes \mathcal{G})$ equals

$$a_2(j;k,l) = \frac{1}{g}\sum_i g_i \chi_i^{(j)} \chi_i^{(\mathcal{C}_k, \mathcal{C}_l)} = \frac{1}{g}\sum_i \frac{1}{g_i} \chi_i^{(j)} N_{ik}N_{il},$$

and the total number of times that it can be constructed in $A(\mathcal{G} \otimes \mathcal{G})$ is

$$a_2(j) = \frac{1}{g}\sum_{ikl} \frac{1}{g_i} \chi_i^{(j)} N_{ik}N_{il} = \sum_i \frac{g}{g_i} \chi_i^{(j)}. \tag{21}$$

We remark that for the six-dimensional irreducible representation of the finite group $U_3(3)$, mentioned in Sec. 3, the rhs of Eq. (21) is larger than zero, which means that a tensor operator corresponding to this representation can be constructed in $A(\mathcal{G} \otimes \mathcal{G})$, although as we saw before it cannot be constructed in $A(\mathcal{G})$.

Analogous to Eq. (18) we have

$$\begin{aligned} a_2(j) &= \sum_i \frac{g}{g_i} \chi_i^{(j)} \\ &= \frac{1}{g}\sum_R \sum_{j'j''} \chi^{(j)}(R)\chi^{(j')}(R)\chi^{(j'')}(R)\chi^{(j')*}(R)\chi^{(j'')*}(R) \\ &= \sum_{j',j''} \{j' \otimes j''j'^* \otimes j''^* j\}. \end{aligned} \tag{22}$$

In Eq. (22) we introduced a generalization of the $3j$ symbol, $\{j' \otimes j''j'^* \otimes j''^* j\}$, which gives the number of times that the representation (j) is contained in the Kronecker product of the representation $(j') \otimes (j'')$ times its complex conjugate representation.

Along the same line one can derive that the number of times that a tensor operator of rank j can be constructed in $A(\mathcal{G}^{\otimes n})$ is equal to

$$a_n(j) = \sum_i \left(\frac{g}{g_i}\right)^{n-1} \chi_i^{(j)} = \sum_{J^{(n)}} \{J^{(n)}J^{(n)*}j\}. \tag{23}$$

In this equation the symbol $J^{(n)}$ denotes a representation of the type

$$J^{(n)} = (j_1) \otimes (j_2) \otimes \cdots \otimes (j_n), \tag{24}$$

where the (j_i) are irreducible representations of $\mathcal{G}$. The summation in Eq. (23) runs over all representations $J^{(n)}$, with n fixed. The symbol $\{J^{(n)}J^{(n)*}j\}$ gives the number of times that the representation (j) is contained in the Kronecker product $J^{(n)} \otimes J^{(n)*}$.

Now we introduce sets of irreducible representations. We shall say that an irreducible representation (j) belongs to the set M_n if $a_n(j) > 0$. In the following we shall need two lemmas dealing with these sets.

Lemma A: Let (j_1) be an irreducible representation belonging to the set M_1 and let (j_2) be an irreducible representation belonging to the set M_n. If (j_3) is an irreducible representation contained in the Kronecker product $(j_1) \otimes (j_2)$ then (j_3) belongs to M_{n+1}.

Lemma B: If (j) is an irreducible representation belonging to the set M_n then (j) belongs also to M_{n+1} and so one has $M_n \subseteq M_{n+1}$, $n = 1, 2, 3, \ldots$.

Proof: From the assumption of Lemma A and from Eq. (23) it follows that there exists an irreducible representation (j) with the property that

$$\{jj^*j_1\} > 0. \tag{25}$$

Furthermore, it follows that there exists a representation $J^{(n)}$ with

$$\{J^{(n)}J^{(n)*}j_2\} > 0. \tag{26}$$

Now we consider the representation $J^{(n+1)} = J^{(n)} \otimes (j)$. One has

$$\begin{aligned} &\{J^{(n)} \otimes j\, J^{(n)*} \otimes j^* j_3\} \\ &\quad = \sum_{j'j''} \{J^{(n)}J^{(n)*}j'\}\{jj^*j''\}\{j'j''j_3\} \\ &\quad \geq \{J^{(n)}J^{(n)*}j_2\}\{jj^*j_1\}\{j_1j_2j_3\} > 0, \end{aligned} \tag{27}$$

because of Eqs. (25) and (26) and the assumption that (j_3) is contained in $(j_1) \otimes (j_2)$. This proves Lemma A. Lemma B is an immediate consequence of Lemma A [take $(j_1) = (1_1)$ and $(j_2) = (j)$ in Lemma A].

If one takes two irreducible representations (j_1) and (j_2) from M_1 and if (j_3) is contained in $(j_1) \otimes (j_2)$ then (j_3) does not necessarily belong to M_1 in general. However, from Lemma A it follows that (j_3) does belong to M_2. Now we consider all possible Kronecker products of irreducible representations from M_1. They define a subalgebra in the algebra of representations (or a subring in the ring of characters). This subalgebra may coincide with the whole algebra of representations. From the remark made above and from Lemma A it is clear that the irreducible representation belonging to this subalgebra are just representations from the union of the sets $M_1 \cup M_2 \cup \cdots \equiv M$. According to a theorem of Burnside (cf. Ref. 18, p. 299) the irreducible representations (j) belonging to a subalgebra define a quotient group $\mathcal{G}/\mathcal{K}$ of $\mathcal{G}$ of which they are the irreducible representations. [We might say that $\mathcal{G}/\mathcal{K}$ is generated by the irreducible representations (j) of M_1.] In all representations (j) of the subalgebra all elements of $\mathcal{K}$ are represented by $D^{(j)}(E)$, whereas this is not the case for any (j) not belonging to the subalgebra, i.e., for (j)'s with $(j) \notin M$. Now we know already that all irreducible representations from M_1 are also irreducible representations of $\mathcal{G}/\mathcal{J}$. Moreover, if (j_1) and (j_2) are from M_1 and if (j_3) is contained in $(j_1) \otimes (j_2)$ then (j_3) is again an irreducible representation of $\mathcal{G}/\mathcal{J}$. From this we have that

$\mathcal{G}/\mathcal{J} \subseteq \mathcal{G}/\mathcal{H}$ or that $\mathcal{H} \supseteq \mathcal{K}$. We shall prove now that in fact $\mathcal{G}/\mathcal{K} = \mathcal{G}/\mathcal{H}$ or equivalently:

Theorem: The set of all irreducible representations of $\mathcal{G}/\mathcal{H}$ coincides with M.

Proof: We saw that $\mathcal{G}/\mathcal{K} \subseteq \mathcal{G}/\mathcal{H}$. Let (j) be an irreducible representation of $\mathcal{G}$, such that (j) is also an irreducible representation of $\mathcal{G}/\mathcal{H}$. Then

$$\chi^{(j)}(R) = [j], \quad \forall\, R \in \mathcal{H}. \tag{28}$$

One has

$$\begin{aligned} a_n(j) &= \sum_i \left(\frac{g}{g_i}\right)^{n-1} \chi_i{}^{(j)} \\ &= \sum_{R\in\mathcal{H}} g^{n-1}\chi^{(j)}(R) + \sum_{i\notin\mathcal{H}} \left(\frac{g}{g_i}\right)^{n-1} \chi_i{}^{(j)} \\ &\geqslant g^{n-1}[j]z - [j]\sum_{i\notin\mathcal{H}} \left(\frac{g}{g_i}\right)^{n-1} \\ &\geqslant g^{n-1}[j]z - \left(\frac{g}{2}\right)^{n-1}[j](s-z) \\ &> 0, \end{aligned} \tag{29}$$

if n is such that $2^{n-1} > (s-z)/z$.

In Eq. (29) $\sum_{i\notin\mathcal{H}}$ denotes the summation over all classes which do not belong to $\mathcal{H}$, whereas z is the order of $\mathcal{H}$ and s is the total number of classes of $\mathcal{G}$.

This shows that $a_n(j) > 0$ if n is sufficiently large and so $(j) \in M$. Q.E.D.

Now we come back to our remark in Sec. 3 on the nonvanishing of the nondiagonal matrix elements $\langle j_1 \| T^j \| j_2 \rangle$. As an example, we consider a tensor T^j from $A(\mathcal{G}\otimes\mathcal{G})$. Furthermore, we take the irreducible representation $(j_1)\otimes(j_2)$ of the group $\mathcal{G}\otimes\mathcal{G}$. With respect to this representation the tensor operator T^j of Eq. (9) has matrix elements of the form

$$\langle j'm_1', j''m_1'' | T^j_m | j'm_2', j''m_2'' \rangle. \tag{30}$$

If one reduces the irreducible representation $(j')\otimes(j'')$ of $\mathcal{G}\otimes\mathcal{G}$ into irreducible representations of the diagonal subgroup $\mathcal{G}*\mathcal{G}$, the matrix elements of T^j_m take the form

$$\langle j'j''j_1 m_1 | T^j_m | j'j''j_2 m_2 \rangle. \tag{31}$$

This makes it clear that a necessary condition for the nonvanishing of the reduced matrix element $\langle j_1 \| T^j \| j_2 \rangle$ is, that there exist irreducible representations (j') and (j'') of $\mathcal{G}$ with the property that

$$\{j'j''j_1\} > 0, \quad \{j'j''j_2\} > 0. \tag{32}$$

Because of the Wigner–Eckart theorem one also has the condition

$$\{j_2 j j_1\} > 0. \tag{33}$$

The conditions (32) and (33) are not sufficient conditions for the nonvanishing of $\langle j_1 \| T^j \| j_2 \rangle$. We shall give an example of this in the next section.

5. EXPLICIT CONSTRUCTIONS AND EXAMPLES

If an explicit (unitary) form of an irreducible representation (j) of a group $\mathcal{G}$ is known one can calculate the tensor operators in $A(\mathcal{G})$ by applying the following rules. Let $\mathcal{C}_k$ be a fixed class of the finite group $\mathcal{G}$. Let C be an arbitrary element of $\mathcal{C}_k$. Then the components of an irreducible tensor operator of rank j formed from the elements of $\mathcal{C}_k$ is given by

$$T^j_m = \sum_{G\in\mathcal{G}} D^{(j)}_{im}(G)\, G^{-1}CG = \sum_{G\in\mathcal{G}} D^{(j)*}_{mi}(G)\, GCG^{-1}, \tag{34}$$

where i is a fixed, but otherwise arbitrary row or column index, respectively.

Equation (34) can be proved by substitution in Eq. (5). It can also be proved that by varying i and C this procedure provides us with all tensor operators of rank j in the class $\mathcal{C}_k$.

Note that the character projection operators of Refs. 1 and 4 do not give explicitly the components T^j_m of a tensor operator T^j. Of course, Eq. (34) shall give $T^j_m = 0$ if the condition formulated in the corollary of Theorem 1 is not satisfied.

One can construct more tensor operators by coupling known tensor operators by means of Clebsch–Gordan coefficients. For this construction one has two possibilities (cf. Refs. 13 and 14):

(a) The tensor operators operate on the same system, which means in our case that both operators are from the same algebra $A(\mathcal{G})$ and so is their resultant.

(b) The tensor operators operate on different systems, which means here that the constituting tensor operators operate in isomorphic but different algebras $A(\mathcal{G})$ and their resultant operators in $A(\mathcal{G}\otimes\mathcal{G})$.

The first possibility does not give anything which is not already covered by Eq. (34). However, the second alternative gives new tensor operators which lie in $A(\mathcal{G}\otimes\mathcal{G})$. This second construction can be iterated to get tensor operators in $A(\mathcal{G}^{\otimes n})$ $(n = 2, 3, \ldots)$.

To illustrate the various procedures for the construction of tensor operators of finite groups we shall now give some simple examples. First we consider the symmetric group in three variables S_3. This group has three classes $\mathcal{C}_1 = \{E\}$, $\mathcal{C}_2 = \{P, Q\}$, and $\mathcal{C}_3 = \{R, S, T\}$. The character table is

	1_1	1_2	2
$\mathcal{C}_1$	1	1	2
$\mathcal{C}_2$	1	1	−1
$\mathcal{C}_3$	1	−1	0

. (35)

An irreducible two-dimensional representation can be given by the (unitary) matrices

$$\begin{aligned} D^{(2)}(E) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D^{(2)}(P) = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}, \\ D^{(2)}(Q) &= \begin{bmatrix} c & s \\ -s & c \end{bmatrix}, \quad D^{(2)}(R) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \\ D^{(2)}(S) &= \begin{bmatrix} c & -s \\ -s & -c \end{bmatrix}, \quad D^{(2)}(T) = \begin{bmatrix} c & s \\ s & -c \end{bmatrix}, \end{aligned} \tag{36}$$

where $c = -\frac{1}{2}, s = \frac{1}{2}\sqrt{3}$. (This representation is faithful.)

The center of S_3 consists of the unit element only and hence all tensor operators occur in some algebra $A(\mathcal{G}^{\otimes n})$. Because in this example $a_1(j) = \sum_i \chi_i{}^{(j)} > 0$ for all (j), every tensor operator occurs in $A(\mathcal{G})$ already.

The elements of the classes can be combined to irreducible tensor operators in the following way:

Class $\mathcal{C}_1$ gives

$$T_1^{(1_1)} = E; \tag{37}$$

Class $\mathcal{C}_2$ gives

$$T_1^{(1_1)} = P + Q, \tag{38}$$

$$T_1^{(1_2)} = P - Q; \tag{39}$$

and class $\mathcal{C}_3$ gives

$$T_1^{(1_1)} = R + S + T, \tag{40}$$

$$T_1^{(2)} = 2R - S - T, \quad T_2^{(2)} = \sqrt{3}\ (S - T). \tag{41}$$

To find Eq. (41) one can use Eq. (34).

Next we give an example of a tensor operator in $A(\mathcal{G} \otimes \mathcal{G})$ and show that some nondiagonal matrix elements are nonzero.

Consider

$$T_1^{(2)} = \frac{1}{3}(E, 2R - S - T), \quad T_2^{(2)} = \frac{1}{3}(E, \sqrt{3}S - \sqrt{3}T) \tag{42}$$

(the factor 1/3 is for convenience).

Here, the tensor operators of Eqs. (37) and (41) have been coupled. An explicit matrix representation can be found by calculating the direct products $D^{(2)}(E) \otimes D^{(2)}(R)$, $D^{(2)}(E) \otimes D^{(2)}(S)$, etc. One finds

$$D(T_1^{(2)}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \quad D(T_2^{(2)}) = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}. \tag{43}$$

The rows and columns are numbered by $(m_1, m_2) = (1,1), (1,2), (2,1)$ and $(2,2)$.

Next we shall reduce the representation $(2) \otimes (2)$ of $S_3 \otimes S_3$ into irreducible representations of S_3 itself. To this end we make use of the following unitary transformation matrix U, the elements of which are Clebsch–Gordan coefficients of S_3 (see Ref. 22):

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 \end{bmatrix}. \tag{44}$$

The columns in U are numbered as in Eq. (43), whereas the rows are numbered by $(j,m) = (1_1,1), (1_2,1), (2,1)$ and $(2,2)$. By applying U one finds, e.g.,

$$UD(T_1^{(2)})U^\dagger = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{45}$$

We see from this equation that the reduced nondiagonal matrix elements $\langle 1_1 \| T^{(2)} \| 2 \rangle$ and $\langle 1_2 \| T^{(2)} \| 2 \rangle$ are nonzero. At the same time we observe that the reduced matrix element $\langle 2 \| T^{(2)} \| 2 \rangle$ vanishes, although Eqs. (32) and (33) are satisfied. This can be explained by the accidental zero of the 9j symbol

$$\begin{Bmatrix} 2 & 2 & 2 \\ 2 & 1_1 & 2 \\ 2 & 2 & 2 \end{Bmatrix} = 0 \tag{46}$$

[cf. Ref. 13, Eq. (7.1.5) and Ref. 22].

Finally, we shall consider the cyclic group $\mathcal{C}_3 = \{E, P, Q\}$ with character table

	1_1	1_2	1_3	
$\mathcal{C}_1 = \{E\}$	1	1	1	
$\mathcal{C}_2 = \{P\}$	1	ω	ω^2	
$\mathcal{C}_3 = \{Q\}$	1	ω^2	ω,	$\omega = e^{2\pi i/3}$.

(47)

The center of this (Abelian) group coincides with the group itself, which means that only the tensor operator $T^{(1_1)}$ can be constructed from the elements of $A(C_3{}^{\otimes n})$, $n = 1, 2, 3, \ldots$).

To obtain the irreducible tensor operators of rank (1_2) and rank (1_3) we embed C_3 in the symmetric group S_3 of the previous example. It is easy to check that the following elements of $A(S_3)$ transform as the required tensor operators of C_3:

$$T_1^{(1_2)} = R + \omega^2 S + \omega T,$$

$$T_1^{(1_3)} = R + \omega S + \omega^2 T. \tag{48}$$

6. TENSOR OPERATORS AND DUALITY

As one knows in finite groups many properties of irreducible representations are similar to properties of classes. Such properties are called dual properties. Recently this kind of duality has been discussed somewhat in the literature (see Refs. 1, 9-11). Although one cannot give an exact definition of this duality, it is shown in Ref. 10 and 11, that it can be used as a heuristic principle to suggest new theorems.

In this section we shall show, that it is possible to give notions and properties dual to the notions and properties dealing with irreducible tensor operators constructible in the group algebras $A(\mathcal{G}^{\otimes n})$ $(n = 1, 2, 3, \ldots)$. To start with we show that $a_1(j)$, the number of times that an irreducible tensor operator of rank j can be constructed in $A(\mathcal{G})$, is dual to the number $b_1(i)$ which denotes the number of times that a group element A of class $\mathcal{C}_i$ of $\mathcal{G}$ can be written as a commutator (cf. Ref. 18, p. 319, Ex. 7) or stated otherwise the number of solutions of the equation

$$XYX^{-1}Y^{-1} = A, \tag{49}$$

where X and Y have to be elements of $\mathcal{G}$. It is easy to see that this number $b_1(i)$ equals (cf. Ref. 20)

$$\begin{aligned} b_1(i) &= \frac{1}{g} \sum_j \sum_{R,S} \chi_i^{(j)} \chi^{(j)}(RSR^{-1}S^{-1}) \\ &= \frac{1}{g} \sum_j \sum_{R,S} \chi_i^{(j)} D_{\alpha\beta}^{(j)}(R) D_{\beta\gamma}^{(j)}(S) D_{\gamma\epsilon}^{(j)}(R^{-1}) D_{\epsilon\alpha}^{(j)}(S^{-1}) \\ &= \frac{1}{g} \sum_j \chi_i^{(j)} \left(\frac{g}{[j]}\right)^2 \delta_{\alpha\epsilon}\delta_{\beta\gamma}\delta_{\beta\alpha}\delta_{\gamma\epsilon} = \sum_j \frac{g}{[j]} \chi_i^{(j)}. \end{aligned} \tag{50}$$

More generally one can derive that the number of times that an element A of a class $\mathcal{C}_i$ can be written as a product of n commutators is equal to

$$b_n(i) = \sum_j \left(\frac{g}{[j]}\right)^{2n-1} \chi_i^{(j)}. \tag{51}$$

We observe that the rhs of Eqs. (50) and (51) are dual to the rhs of Eqs. (17) and (23), especially if we take into account that $\sqrt{g_i}$ can be considered as dual to $[j]$ (see also Ref. 10).

We shall present now a table of dual items. In the left column items are enumerated concerning tensor operators and in the right column concerning commutators.

TABLE I. A listing of dual items.

Representations	*classes*
Irreducible tensor operators of rank j constructible in $A(\mathcal{G}^{*n})$. Number of times $a_n(j)$ that an irreducible tensor operator of rank j can be constructed in $A(\mathcal{G}^{*n})$.	Elements of class $\mathcal{C}_i$ expressible as the product of n commutators. Number of times $b_n(i)$ that an element of a class $\mathcal{C}_i$ can be expressed as the product of n commutators.
$a_n(j) = \sum_i \left(\frac{g}{g_i}\right)^{n-1} \chi_i^{(j)}$.	$b_n(i) = \sum_j \left(\frac{g}{[j]}\right)^{2n-1} \chi_i^{(j)}$.
If $a_{n_0}(j) > 0$ then $a_{n_0+1}(j) > 0$.	If $b_{n_0}(i) > 0$ then $b_{n_0+1}(i) > 0$.
$M_n = \{(j): a_n(j) > 0\}$.	$N_n = \{\mathcal{C}_i; b_n(i) > 0\}$.
$M_1 \subseteq M_2 \subseteq M_3 \subseteq \ldots$.	$N_1 \subseteq N_2 \subseteq N_3 \subseteq \ldots$.
$M = \bigcup_i M_i$ defines the quotient group $\bar{\mathcal{G}} = \mathcal{G}/\mathcal{J}$.	$N = \bigcup_i N_i$ defines the (commutator) subgroup $\mathcal{G}'$.
M_1 contains the 'generating irreducible representations' of $\bar{\mathcal{G}}$.	N_1 contains the generators of $\mathcal{G}'$.
The number of classes with only one element equals $g/\bar{g}$ ($\bar{g}$ is the order of $\bar{\mathcal{G}}$).	The number of irreducible representation of dimension one equals g/g' (g' is the order of $\mathcal{G}'$).
An example of a group which has a representation (j) for which $a_1(j) = 0$ and $a_2(j) > 0$ is the finite group $U_3(3)$ of order 6048 (Ref. 19).	An example of a group which has a class $\mathcal{C}_i$ for which $b_1(i) = 0$ and $b_2(i) > 0$ is a group or order 256 (Ref. 17, p. 39, Ex. 30).

The properties of tensor operators have been derived in the previous sections of this paper. The derivation of the properties of commutators are either well known or left to the reader.

ACKNOWLEDGMENTS

We wish to thank Professor L. C. Biedenharn for some fruitful discussions and for help in preparing the manuscript. One of us (A.J.v.Z.) would like to thank the Netherlands American Commission for Educational Exchange for financial support.

*Research supported in part by the National Science Foundation and the Army Research Office (Durham).

†On leave from the University of Groningen, The Netherlands.

‡Fulbright Scholar.

[1] A. Gamba, J. Math. Phys. **9**, 186 (1968).
[2] J. Killingbeck, J. Math. Phys. **11**, 2268 (1970).
[3] J. Killingbeck, Rep. Prog. Phys. **33**, 533 (1970).
[4] J. Killingbeck, J. Phys. A **5**, 1131 (1972).
[5] E. de Vries and A. J. van Zanten, Commun. Math. Phys. **17**, 322 (1970).
[6] E. de Vries, "Some Applications of the Representation Theory of Finite Groups: Racah Algebraic Methods," thesis, University of Groningen, the Netherlands, 1972.
[7] L. C. Biedenharn and A. Gamba, Rev. Brasil. Fís. 2, 319 (1972).
[8] J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[9] G. de B. Robinson, J. Math. Phys. **9**, 186 (1968).
[10] E. de Vries and A. J. van Zanten, Physica **49**, 532 (1970).
[11] A. J. van Zanten, "Some Applications of the Representation Theory of Finite Groups: A Partial Reduction Method," thesis, University of Groningen, the Netherlands, 1972.
[12] E. P. Wigner in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), p. 87.
[13] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1957).
[14] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[15] J. S. Lomont, *Applications of Finite Groups* (Academic, New York, 1959).
[16] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962).
[17] R. D. Carmichael, *Introduction to the Theory of Groups of Finite Order* (Dover, New York, 1956).
[18] W. Burnside, *Theory of Groups of Finite Order* (Dover, New York, 1955), 2nd ed.
[19] M. Hall, Jr. and D. Wales, J. Algebra **9**, 441 (1968).
[20] A. J. van Zanten and E. de Vries, Physica **49**, 536 (1970).
[21] E. de Vries and A. J. van Zanten, J. Math. Phys. **13**, 862 (1972).
[22] W. T. Sharp, Report CRT-935 At. En. Comm. of Canada, Chalk River, 1960, p. 141.

Similarity transformation for the compressible Jeffery-Hamel flow of dissipative plasmas*

H. E. Wilhelm

Department of Physics, Colorado State University, Fort Collins, Colorado 80521
(Received 2 April 1973)

A similarity transformation is presented which transforms the nonlinear, partial differential equations describing a compressible plasma flow across an azimuthal magnetic field between plane inclined walls (Jeffery–Hamel flow with viscous momentum transfer, Ohmic and viscous heating, and thermal heat conduction) into one linear and two nonlinear ordinary, coupled differential equations. By elimination, the nonlinear boundary-value problem of this compressible plasma flow is reduced to a boundary-value problem for (i) two ordinary, coupled differential equations or (ii) a single functional integro-differential equation. The characteristic parameter "m" of the similarity transformation is shown to be an eigenvalue of the nonlinear boundary-value problem.

1. INTRODUCTION

Nonlinear boundary-value problems can frequently be solved exactly[1] or quasilinearized[2] by means of nonlinear transformations. In incompressible fluid dynamics, a single, exact, nonlinear solution has been obtained by Jeffery and Hamel[3] for the flow between inclined plane walls by means of a similarity transformation. It is shown herein that a similarity transformation exists also for the general compressible Jeffery–Hamel flow of partially ionized gases or plasmas with viscous momentum transfer, viscous and Ohmic heating, and heat conduction. In the limit of vanishing electrical conductivity, the similarity transformation reduces to that for the compressible flow of a dissipative gas in a diffuser, a previously unsolved problem of ordinary gas dynamics.

The similarity transformation is of the form $F(r, \theta) = r^{-N}G(\theta)$ [with different power N and function $G(\theta)$ for different plasma fields]. It transforms the nonlinear partial differential equations describing the plasma flow into two coupled, ordinary nonlinear differential equations of second order, in which a single similarity parameter $N = m$ occurs. It is demonstrated that the similarity parameter $N = m$ is an eigenvalue of the nonlinear boundary-value problem, if the Reynolds (R), Mach (M), Eckert (E), Prandtl (P), Hartmann (H), adiabatic (γ) numbers, and duct angle (θ_0) are prescribed. In the general case, the determination of the eigenvalue m requires a (standard) numerical integration of the ordinary nonlinear differential equations.

As an illustration, the theory is applied to isothermal flows, which have the eigenvalue $m = 0$. For this case, a closed form analytical solution is presented.

2. THEORETICAL PRINCIPLES

The plasma flow under consideration is bounded in the planes ($\theta = +\theta_0$, $r_1 \le r \le r_2$) and ($\theta = -\theta_0$, $r_1 \le r \le r_2$) by isolating walls, and quasi-unbounded in the directions parallel to the z-axis (Fig. 1). The latter assumption is applicable to a finite diffuser with electrode plates at $z = \pm z_\infty$, where $z_\infty \gg \frac{1}{2}(r_1 + r_2)\theta_0$. The injection ($r = r_1$) and removal ($r = r_2$) of the plasma occurs in a self-similar way (Jeffery–Hamel flow).[3,4]

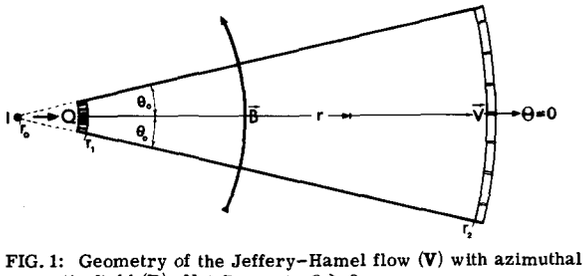


FIG. 1: Geometry of the Jeffery–Hamel flow (**V**) with azimuthal magnetic field (**B**). Net flow rate $Q > 0$.

The magnetic field has its sources in an electric current I flowing through a conducting rod ($0 \le r \le r_o$, $-\infty \le z \le \infty$, $r_o < r_1$). In accordance with Stokes' law, $\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0 I$, the magnetic field is azimuthal (μ_0 = permeability of vacuum),

$$\mathbf{B} = \frac{\mu_0}{2\pi}\frac{I}{r}\mathbf{e}_\theta, \qquad r_o \le r \le \infty. \tag{1}$$

In absence of flow sources or sinks at the inclined walls, the velocity field is radial, if the accelerating force fields are radial,

$$\mathbf{v} = u\mathbf{e}_r. \tag{2}$$

The flow of the plasma (conductivity σ) across the magnetic field induces an axial current density field (Hall-effect neglected),[5]

$$\mathbf{j} = \sigma(E_z + uB_\theta)\mathbf{e}_z. \tag{3}$$

The resulting Lorentz force density is a purely radial field which opposes the inducting flow,

$$\mathbf{j} \times \mathbf{B} = -\sigma(E_z + uB_\theta)B_\theta\mathbf{e}_r. \tag{4}$$

Because of $\nabla \times \mathbf{E} = 0$, $\nabla\cdot\mathbf{j} = \sigma(\nabla\cdot\mathbf{E} + \mathbf{B}\cdot\nabla\times\mathbf{v} - \mathbf{v}\cdot\nabla\times\mathbf{B}) = \sigma\nabla\cdot\mathbf{E} = 0$, and the boundary conditions at $z = z_{\pm\infty}$, the electric field vanishes:

$$\mathbf{E} = E_z\mathbf{e}_z = 0, \qquad \text{for } \mathbf{E}_{z=z_{\pm\infty}} = 0. \tag{5}$$

The Eqs. (1)–(5) are based on the assumption that the induced magnetic field is small compared to the external magnetic field, which implies small magnetic Reynolds numbers,[5]

$$R_B = \mu_0\sigma u(r, \theta)r \ll 1, \qquad r_1 \le r \le r_2, \qquad |\theta| \le \theta_0. \tag{6}$$

The nonlinear partial differential equations describing the radial velocity [$u = u(r, \theta)$], density [$\rho = \rho(r, \theta)$], and pressure [$p = p(r, \theta)$] fields of the steady-state plasma flow between the inclined ($-\theta_0 \le \theta \le \theta_0$) walls are in cylindrical coordinates[5]:

$$\rho u\frac{\partial u}{\partial r} = -\frac{\partial p}{\partial r} - \sigma\left(\frac{\mu_0 I}{2\pi}\right)^2\frac{u}{r^2} + \frac{4}{3}\mu\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}(ur)\right) + \frac{\mu}{r^2}\frac{\partial^2 u}{\partial\theta^2}, \tag{7}$$

$$0 = -\frac{1}{r}\frac{\partial p}{\partial \theta} + \frac{\mu}{3}\frac{1}{r}\frac{\partial}{\partial \theta}\left(\frac{1}{r}\frac{\partial}{\partial r}(ru)\right) + 2\mu\frac{1}{r^2}\frac{\partial u}{\partial \theta}, \tag{8}$$

$$\frac{\partial}{\partial r}(\rho u) + \frac{\rho u}{r} = 0, \tag{9}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(\rho r c_v T u) = \lambda\left(\frac{\partial^2 T}{\partial^2 r^2} + \frac{1}{r}\frac{\partial T}{\partial r} + \frac{1}{r^2}\frac{\partial^2 T}{\partial \theta^2}\right) - p\left(\frac{\partial u}{\partial r} + \frac{u}{r}\right) + \sigma\left(\frac{\mu_0 I}{2\pi}\right)^2\frac{u^2}{r^2} - \frac{2}{3}\mu\left(\frac{1}{r}\frac{\partial}{\partial r}(ru)\right)^2 + \mu\left[2\left(\frac{\partial u}{\partial r}\right)^2 + \left(\frac{\partial u}{r\partial \theta}\right)^2 + 2\left(\frac{u}{r}\right)^2\right], \tag{10}$$

where

$$p = \rho \mathcal{R} T. \tag{11}$$

The transport coefficients are regarded as constant; μ is the viscosity, λ the heat conductivity, c_v the specific heat at constant volume, and $\mathcal{R}$ the ideal gas constant.

3. SIMILARITY TRANSFORMATION

The Eqs. (7)–(11) are brought into nondimensional form by means of the substitutions

$$\frac{r}{r_0} \to r, \quad \frac{\rho}{\rho_0} \to \rho, \quad \frac{u}{u_0} \to u, \quad \frac{p}{p_0} \to p, \quad \frac{T}{T_0} \to T, \tag{12}$$

where

$$\rho_0 = \rho(r_0, \bar{\theta}) > 0, \quad u_0 = u(r_0, \bar{\theta}) > 0,$$
$$p_0 = p(r_0, \bar{\theta}) > 0, \quad T_0 = T(r_0, \bar{\theta}) > 0, \quad |\bar{\theta}| < \theta_0 \tag{13}$$

are known reference values of the flow fields, e.g.,

$$\rho_0 = \rho(r_1, 0), \quad u_0 = u(r_1, 0), \quad p_0 = p(r_1, 0), \quad T_0 = T(r_1, 0)$$

is a suitable choice for applications, since the fields at $r = r_1$ of the central streamline $\bar{\theta} = 0$ are most likely to be measured in experiments.

By means of the similarity ansatz (following from group theoretical considerations) for the dimensionless plasma fields,

$$u(r, \theta) = r^{-m} g(\theta), \tag{14}$$

$$p(r, \theta) = r^{-n} h(\theta), \tag{15}$$

$$\rho(r, \theta) = r^{m-1} f(\theta)/g(\theta), \tag{16}$$

$$T(r, \theta) = r^{-m-n+1}\phi(\theta), \tag{17}$$

where [$n = m + 1$ by Eqs. (7), (8), (10)]

$$\phi(\theta) = \frac{g(\theta)h(\theta)}{f(\theta)}, \quad n = m + 1, \tag{18}$$

the nondimensional partial differential equations corresponding to Eqs. (7)–(10) are reduced to the ordinary differential equations ($' = d/d\theta$):

$$g'' - \left[\tfrac{4}{3}(1 - m^2) + H^2\right]g + (m + 1)\frac{R}{\gamma M^2}h = -mRfg, \tag{19}$$

$$h' = \frac{7 - m}{3}\frac{\gamma M^2}{R}g', \tag{20}$$

$$\frac{1}{PE}(\phi'' + 4m^2\phi) = -\left[2(1 + m^2) - \tfrac{2}{3}(1 - m)^2 + H^2\right]g^2 - (g')^2 - \frac{R}{\gamma}\left(\frac{2m}{E} - \frac{1 - m}{M^2}\right)gh, \tag{21}$$

where

$R = \rho_0 u_0 r_0/\mu$ = Reynolds number,
$M = u_0/(\gamma p_0/\rho_0)^{1/2}$ = Mach number,
$E = u_0^2/c_p T_0$ = Eckert number,
$P = \mu c_p/\lambda$ = Prandtl number,
$H = (\sigma/\mu)^{1/2}(\mu_0 I/2\pi)$ = Hartmann number,
$\gamma = c_p/c_v$ = ratio of specific heats,

and

$$g(\bar{\theta}) = 1, \quad h(\bar{\theta}) = 1, \quad f(\bar{\theta}) = 1, \quad \phi(\bar{\theta}) = 1. \tag{22}$$

[Equation (22) has to be replaced by other relations if reference values different from those in Eq. (13) are chosen for the nondimensionalization.]

The similarity transformation in Eqs. (14)–(17) separates the dependence of the plasma fields on the radial (r) and azimuthal (θ) coordinates, and reduces the partial differential equations [Eqs. (7)–(10)] into ordinary ones [Eqs. (19)–(21)]. The pressure $h(\theta)$ and velocity $g(\theta)$ amplitudes are interrelated by the linear Eq. (20) which is readily integrated:

$$h = \frac{7 - m}{3}\gamma\frac{M^2}{R}g + C_0, \tag{23}$$

where

$$C_0 = 1 - \frac{7 - m}{3}\gamma\frac{M^2}{R} \tag{24}$$

by Eq. (22). The integration constant C_0 determines the pressure at the walls $\theta = \pm\theta_0$. It is $C_0 \geq 0$ for $p(r, \theta = \pm\theta_0) \geq 0$, since $g(\theta = \pm\theta_0) = 0$. By Eqs. (23)-(24),

$$h = 1 + \frac{7 - m}{3}\gamma\frac{M^2}{R}(g - 1). \tag{25}$$

4. BOUNDARY-VALUE PROBLEM

The plasma flow has boundaries in the planes $r = r_1$ and $r = r_2$ ($-\theta_0 \leq \theta \leq \theta_0$) and $\theta = \pm\theta_0$ ($r_1 \leq r \leq r_2$). The boundary conditions at $r = r_1$ and $r = r_2$ are automatically satisfied for a self-similar injection ($r = r_1$) and removal ($r = r_2$) of the material and thermal flow.[3,4] The velocity $v(r, \theta)$ vanishes at the side walls $\theta = \pm\theta_0$ ($r_1 \leq r \leq r_2$). The temperature $T(r, \theta)$ can be subject to various boundary conditions, e.g., those for thermally insulated side walls. Accordingly, the boundary conditions for the velocity and temperature fields are

$$[u(r, \theta)]_{\theta = \pm\theta_0} = 0, \quad [\partial T(r, \theta)/\partial\theta]_{\theta = \pm\theta_0} = 0, \quad r_1 \leq r \leq r_2.$$

Upon elimination of $h(\theta)$ from Eqs. (19) and (21) by means of Eq. (25), one arrives at the following nonlinear boundary-value problem for the coupled fields $g(\theta) \geq 0$ and $\phi(\theta) > 0$ [$T(r, \theta) > 0$]:

$$g'' + [(1 + m)^2 - H^2]g + mR\left(1 + \frac{7 - m}{3}\gamma\frac{M^2}{R}(g - 1)\right)\frac{g^2}{\phi} + (1 + m)\left(\frac{R}{\gamma M^2} - \frac{7 - m}{3}\right) = 0, \tag{26}$$

$$\phi'' + (2m)^2\phi = -PE\mathfrak{F}(g,g'), \tag{27}$$

$$\mathfrak{F}(g,g') \equiv \left(1 - \frac{7-m}{3}\gamma\frac{M^2}{R}\right)\frac{R}{\gamma}\left(\frac{2m}{E} - \frac{1-m}{M^2}\right)g + \left[\frac{4}{3}(1+m+m^2) + H^2 + \frac{7-m}{3}M^2\left(\frac{2m}{E} - \frac{1-m}{M^2}\right)\right]g^2 + (g')^2,$$

where

$$[g(\theta)]_{\theta=\pm\theta_0} = 0, \tag{28}$$

$$[\partial\phi(\theta)/\partial\theta]_{\theta=\pm\theta_0} = 0. \tag{29}$$

The boundary-value problem defined in Eqs. (26)–(29) can be formulated as well for other boundary conditions of the temperature field, e.g.,

$$[\phi(\theta)]_{\theta=\pm\theta_0} = \phi_\pm, \quad \phi_\pm > 0, \tag{29'}$$

if the temperature is prescribed at the side walls, $T(r, \theta = \pm\theta_0) = \phi_\pm r^{-2m}$ [Eq. (17)].

5. FUNCTIONAL INTEGRO-DIFFERENTIAL EQUATION

The Eqs. (26)–(27) permit a more condensed representation. Formal integration of Eq. (27) gives $\phi(\theta)$ as a functional of $g(\theta), g'(\theta)$, and θ:

$$\phi[g(\theta),g'(\theta),\theta] = C_1\cos 2m\theta + C_2\sin 2m\theta + \frac{PE}{2m}\int_0^\theta \mathfrak{F}[g(\theta'),g'(\theta')]\sin 2m(\theta'-\theta)d\theta'. \tag{30}$$

The constants C_1 and C_2 are determined by the respective boundary conditions for the temperature field. As an example, C_1 and C_2 are given for the boundary conditions in Eq. (29):

$$C_1 = -\frac{1}{2}PE\frac{ctg2m\theta_0}{2m}\int_{-\theta_0}^{+\theta_0}\mathfrak{F}[g(\theta),g'(\theta)]\cos 2m\theta d\theta - \frac{1}{2}\frac{PE}{2m}\int_0^\theta\{\mathfrak{F}[g(\theta),g'(\theta)] + \mathfrak{F}[g(-\theta),g'(-\theta)]\}\sin 2m\theta d\theta \tag{31}$$

and

$$C_2 = +\frac{1}{2}PE\frac{tg2m\theta_0}{2m}\int_{-\theta_0}^{+\theta_0}\mathfrak{F}[g(\theta),g'(\theta)]\sin 2m\theta d\theta + \frac{1}{2}\frac{PE}{2m}\int_0^{\theta_0}\{\mathfrak{F}[g(\theta),g'(\theta)] - \mathfrak{F}[g(-\theta),g'(-\theta)]\}\cos 2m\theta d\theta. \tag{32}$$

Hence, e.g.,

$$C_1 \neq 0 \text{ and } C_2 = 0, \text{ for } g(+\theta) = g(-\theta),$$

by Eq. (27). Certainly, symmetrical solutions exist,[6] $g(+\theta) = g(-\theta)$, due to the symmetry of the boundary conditions in Eqs. (28)–(29). On the other hand, if $\phi_+ \neq \phi_-$, the boundary conditions in Eq. (29)′ result exclusively in asymmetrical solutions, $g(+\theta) \neq g(-\theta)$.

Substitution of Eq. (30) into Eq. (26) leads to the nonlinear, functional integro-differential equation:

$$g'' + [(1+m)^2 - H^2]g + mR\left(1 + \frac{7-m}{3}\gamma\frac{M^2}{R}(g-1)\right)g^2\phi^{-1}[g(\theta),g'(\theta),\theta] + (1+m)\left(\frac{R}{\gamma M^2} - \frac{7-m}{3}\right) = 0. \tag{33}$$

The boundary conditions for Eq. (33) have already been stated in Eq. (28).

6. SIMILARITY PARAMETER

In the formulation of the boundary-value problem, Eqs. (26)–(29) [Eq. (29′)], the constants $R, M, E, P, H, \gamma, \theta_0$ are given, whereas the similarity parameter, m, is not known *a priori*. In accordance with the theory of boundary-value problems,[7] the Eqs. (26)–(29) [Eq. (29′)] have only solutions for certain characteristic values of m, the so-called eigenvalues of m. Since the pressure is necessarily positive, $h(\theta) > 0$, and $g(\theta = \pm\theta_0) = 0$, Eq. (25) gives $1 - (7-m)\gamma M^2/3R > 0$. Hence, m satisfies, for physical reasons, the inequality

$$m > 7 - 3R/\gamma M^2. \tag{34}$$

Since powers $|m| \gg 1$ are physically not meaningful [Eq. (14)] and $|R/\gamma M^2| \gg 1$ in actual flows, it follows that the similarity transformation describes pure outflows:

$$g(\theta) \geq 0, \quad R > 0. \tag{35}$$

In the general case, the eigenvalue m has to be determined by numerical integration. To obtain an analytical estimate of m, Eq. (10) can be replaced by the polytropic energy integral, $p/p_0 = (\rho/\rho_0)^\beta$ (β = polytropic coefficient). This approximation gives for the order of magnitude of the similarity parameter[8]

$$m \cong (\beta-1)/(\beta+1), \quad \text{i.e.,} \quad -1 \lesssim m \lesssim +1.$$

7. APPLICATION TO ISOTHERMAL FLOWS

For isothermal flows, $p/p_0 = \rho/\rho_0$, the eigenvalue is $m \equiv 0$. In this special case, the boundary-value problem in Eqs. (26)–(29) becomes linear:

$$g'' + (1-H^2)g + \left(\frac{R}{M^2} - \frac{7}{3}\right) = 0 \tag{36}$$

where

$$[g(\theta)]_{\theta=\pm\theta_0} = 0. \tag{37}$$

Hence

$$g(\theta) = \left(1 - \frac{1-\cosh(H^2-1)^{1/2}\theta}{1-\cosh(H^2-1)^{1/2}\theta_0}\right)g(0) \tag{38}$$

and

$$M^2 = R/\left(\frac{7}{3} + \frac{(H^2-1)\cosh(H^2-1)^{1/2}\theta_0}{\cosh(H^2-1)^{1/2}\theta_0 - 1}\,g(0)\right). \tag{39}$$

$g(0)$ is given by the normalization, e.g., $g(0) = 1$ for $\bar{\theta} = 0$ by Eqs. (13).

The boundary conditions in Eq. (28) interlates R, M, H, and θ_0 as shown in Eq. (39). Accordingly, if R, H, and θ_0 are prescribed, isothermal flows ($m = 0$) exist only for Mach numbers M satisfying Eq. (39). It is seen that isothermal flows with standard Mach numbers, $M \cong 1$, are possible for Hartmann numbers $H^2 \cong R \gg 1$ as common in experiments. On the other hand, in absence of magnetic fields ($H = 0$), isothermal flows with reasonable Mach numbers, $M < 10$, occur only at small duct angles, $\theta_0 \ll 1$, since $R \gg 1$.

In Eqs. (38)–(39), it should be noted that $\cosh(H^2 - 1)^{1/2}\theta = \cos(1 - H^2)^{1/2}\theta$. The Eqs. (38) and (39) become for $H^2 = 1$

$$g(\theta) = [1 - (\theta/\theta_0)^2]g(0) \tag{40}$$

and

$$M^2 = R[(7/3) + (2/\theta_0^2)g(0)]^{-1} \ll R, \quad \theta_0 \ll 1. \tag{41}$$

*Supported in part by the U. S. Office of Naval Research.

[1]E. W. Montroll in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Britton (Gordon and Breach, New York, 1968).
[2]H. E. Wilhelm, J. Math. Phys. **13**, 252 (1972).
[3]G. B. Jeffery, Philos. Mag. **27**, 455 (1915).
[4]K. Millsaps and K. Pohlhausen, J. Aeron. Sci. **20**, 187 (1953); J. Nikuradse, Z. Angew. Math. Mech. **8**, 424 (1928).
[5]L. E. Kalikhman, *Magnetogasdynamics* (Saunders, Philadelphia, PA, 1967).
[6]In spite of the symmetry of the boundary conditions, also asymmetrical solutions exist for the incompressible Jeffery–Hamel flow.[3–4]
[7]A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[8]In the polytropic case, the similarity transformation exists by Eq. (8) only if $m = (\beta - 1)/(\beta + 1)$.

A symmetry of Petrov-type N spaces*

R. Sigal

Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada
(Received 1 March 1973; revised manuscript received 20 April 1973)

It is shown that for empty Petrov-type N spaces the Lie derivative in the principal null direction of the Weyl conformal tensor and its dual is equal to a conformal transformation and a duality rotation of the original tensor. This type of symmetry is studied for general relativity and electromagnetism.

1. INTRODUCTION

In this paper the Lie derivatives of the Weyl conformal tensor and the Maxwell electro-magnetic tensor are examined for principal null directions of these tensors.[1] For the case of empty Petrov-type N spaces, it is proven that

$$\underset{\xi}{\pounds}\, C_{\rho\sigma\mu\nu} = A[C_{\rho\sigma\mu\nu}\cos\theta + {}^*C_{\rho\sigma\mu\nu}\sin\theta], \tag{1}$$

$$\underset{\xi}{\pounds}\, {}^*C_{\rho\sigma\mu\nu} = A[{}^*C_{\rho\sigma\mu\nu}\cos\theta - C_{\rho\sigma\mu\nu}\sin\theta], \tag{2}$$

where $C_{\rho\sigma\mu\nu}\xi^\nu = 0$, ${}^*C_{\rho\sigma\mu\nu}$ is the dual of the Weyl tensor $C_{\rho\sigma\mu\nu}$, and A and θ are given by directional derivatives of ξ^ν. A similar result also holds for the source free null Maxwell tensor, and its dual. When equations of the form (1) or (2) hold, this will be called a null-type symmetry.

Properties of these equations are examined, and they are shown to be invariant under conformal changes of the metric. The Lie derivatives of the nonnull Maxwell tensor are examined in some detail.

2. THE NULL MAXWELL CASE

Before establishing Eqs. (1) and (2) it is interesting to look at the corresponding situation for the null Maxwell field. It is well known that $F_{\alpha\beta}$ represents a null field if and only if its Lorentz invariants $\mathbf{E}\cdot\mathbf{B}$ and $\mathbf{E}^2 - \mathbf{B}^2$ are everywhere zero. Plane waves are, of course, of this type. For simple Maxwellian plane waves, the complex tensor

$F^\dagger_{\alpha\beta} \equiv F_{\alpha\beta} + i{}^*F_{\alpha\beta}$ takes on the form

$$F^\dagger_{\alpha\beta} = C(k_\mu x^\mu) l_{[\alpha} m_{\beta]},$$

where l_α is a real constant null vector giving the direction of propagation of the wave, and m_α is a complex constant null vector normal to l_α. l^α is the principal null vector of $F_{\alpha\beta}$.

The Lie derivative of $F^\dagger_{\alpha\beta}$ in the ξ^σ direction is given by

$$\underset{\xi}{\pounds}\, F^\dagger_{\alpha\beta} = F^\dagger_{\alpha\beta,\sigma}\xi^\sigma + F^\dagger_{\alpha\sigma}\xi^\sigma_{,\beta} + F^\dagger_{\sigma\beta}\xi^\sigma_{,\alpha}. \tag{3}$$

For $\xi^\sigma = \xi l^\sigma$, $F_{\alpha\sigma}\xi^\sigma_{,\beta} = F_{\sigma\beta}\xi^\sigma_{,\alpha} = 0$, and thus

$$\underset{\xi}{\pounds}\, F^\dagger_{\alpha\beta} = \left(\frac{\xi C_{,\sigma} l^\sigma}{C}\right) F^\dagger_{\alpha\beta}, \qquad \frac{\xi C_{,\sigma} l^\sigma}{C} = Ae^{-i\theta}. \tag{4}$$

Equation (4) is equivalent to the Maxwell analog of equations (1) and (2) both holding. In terms of the E and B fields (4) just says that when E and B are Lie transported along the principal null direction they only change by having their amplitude multiplied by a common factor A, and being rotated through an angle θ. This is all that could be expected for a plane wave since $\mathbf{B}\cdot\mathbf{E}$ and $B^2 - E^2$ must stay zero and B and E normal to the spacial propagation direction so that Eq. (4) uses up their complete freedom to change.

For the generic type N field satisfying Maxwell's source free equations, $F^\dagger_{\alpha\beta}$ can be put in the form[2]

$$F^\dagger_{\alpha\beta} = B l_{[\alpha} m_{\beta]},$$

where

$$l_{\alpha;\beta} l^\beta = m_{\alpha;\beta} l^\beta = 0, \qquad m_\sigma l^\sigma_{;\alpha} = a l_\alpha + b m_\alpha.$$

It follows from the above that for $\xi^\sigma = \xi l^\sigma$

$$\underset{\xi}{\pounds}\, F^\dagger_{\alpha\beta} = \xi\left[(B_{,\sigma} l^\sigma/B) + b\right] F^\dagger_{\alpha\beta}. \tag{5}$$

If $F^{\alpha\beta}$ and ${}^*F^{\alpha\beta}$ are studied instead, an equation of the form of (5) would be obtained. On the other hand, if

$$\underset{\xi}{\pounds}\, F^{\dagger\beta}_{\alpha}$$

is examined an equation of the form of (5) will not always hold.

3. THE NONNULL MAXWELL CASE[4]

For the remainder of this paper it will be more convenient to work in spinors.[2] Then in the standard way

$$F_{\alpha\beta} \leftrightarrow \epsilon_{AB}\bar{\Phi}_{A'B'} + \epsilon_{A'B'}\Phi_{AB},$$

$$\epsilon_{A'B'}\Phi_{AB} \leftrightarrow F^\dagger_{\alpha\beta}.$$

For the nonnull Maxwell case Φ_{AB} can be put in the form

$$\Phi_{AB} = 2\Phi\iota_{(A}{}^\circ{}_{B)}, \qquad {}^\circ{}_A\iota^A = 1, \tag{6}$$

where its principal null directions are given by ι_A and ${}^\circ{}_B$. A change in $\iota_A \to \Lambda\iota_A = \hat{\iota}_A$ forces ${}^\circ{}_A \to \Lambda^{-1}{}^\circ{}_A = \hat{}^\circ{}_A$ so that Φ is not changed. Φ in fact gives the two Lorentz scalars $F_{\alpha\beta}F^{\alpha\beta} = 2(B^2 - E^2)$, and $F_{\alpha\beta}{}^*F^{\alpha\beta} = 4\mathbf{E}\cdot\mathbf{B}$. If S is defined by

$$S \equiv \Phi_{AB}\Phi^{AB} = -2\Phi^2,$$

then

$$\mathrm{Re}\, S = \tfrac{1}{4} F_{\alpha\beta}F^{\alpha\beta},$$

$$\mathrm{Im}\, S = \tfrac{1}{4} F_{\alpha\beta}{}^*F^{\alpha\beta}.$$

Thus if Φ is real, $\mathbf{B}\cdot\mathbf{E} = 0$ and $B^2 - E^2 < 0$, and if Φ is pure imaginary, $B\cdot E = 0$ and $B^2 - E^2 > 0$.

The condition for $F_{\mu\nu}$ to allow a null-type symmetry is for $\underset{\xi}{\pounds}\, F_{\mu\nu}$ expressed as a spinor to be of the form

$$\underset{\xi}{\pounds}\, F_{UU'VV'} = \Omega\epsilon_{UV}\Phi_{U'V'} + \bar{\Omega}\epsilon_{U'V'}\Phi_{UV}, \tag{7}$$

where Ω is on arbitrary complex scalar field. $\underset{\xi}{\pounds}\, F_{UV'VV'}$ is a formal symbol for the spinor representation of the tensor $\underset{\xi}{\pounds}\, F_{\mu\nu}$.

Introducing a spin dyad (ι^A, o^A), $J_0^A = o^A$, $J_1^A = \iota^A$ in the standard Newman–Penrose way,[5] and taking $\xi^{AA'} = \iota^A \bar{\iota}^{A'}$, Eq. (7) becomes in components

$$\Phi[-2\tau + \beta + \bar{\alpha} - \bar{\pi} - \delta(\ln\Lambda + \ln\bar{\Lambda})] + \bar{\Phi}[\beta + \bar{\alpha} + \bar{\pi} - \delta(\ln\Lambda + \ln\bar{\Lambda})] = 0, \quad (8)$$

$$\Lambda\bar{\Lambda}[\Delta(\ln\Lambda + \ln\bar{\Lambda}) + \gamma + \bar{\gamma}][\Phi + \bar{\Phi}] - \Delta[\Phi + \bar{\Phi}] = -[\bar{\Omega}\Phi + \Omega\bar{\Phi}], \quad (9)$$

$$(\Lambda\bar{\Lambda})(\mu + \bar{\mu})(\Phi - \bar{\Phi}) + \Delta(\Phi - \bar{\Phi}) = -[-\bar{\Omega}\Phi + \Omega\bar{\Phi}], \quad (10)$$

$$(\Phi + \bar{\Phi})\nu = 0. \quad (11)$$

The terms in Λ and $\bar{\Lambda}$ enter by taking $\Phi_{AB} = 2\Phi \hat{o}_{(A}\hat{\iota}_{B)}$, $\hat{\xi}^{AA'} = \hat{\iota}^A\hat{\bar{\iota}}^{A'}$ which is physically equivalent to the original choice. The condition $\nu = 0$ says that ι^A is a tangent vector field to a geodesic congruence.

For ${}^*F_{\mu\nu}$ the equations analogous to Eqs. (8)–(11) are obtained by replacing Φ by $-\Phi$, and leaving $\bar{\Phi}$ unchanged. If both ${}^*F_{\mu\nu}$ and $F_{\mu\nu}$ admit a null-type symmetry then the set of equations obtained are

$$\beta + \bar{\alpha} + \bar{\pi} - \delta(\ln\Lambda + \ln\bar{\Lambda}) = 0, \quad (12)$$

$$\tau = -\bar{\pi}, \quad (13)$$

$$\nu = 0, \quad (14)$$

$$\Delta(\ln\Lambda + \ln\Lambda) + \gamma + \bar{\gamma} = -(\mu + \bar{\mu}). \quad (15)$$

These equations unlike the null case, can only be satisfied by one value of Λ. In flat space there are a large number of trivial solutions to these equations. The most trivial is that of an everywhere constant electric field. For this case any constant null ξ^α will do. For the Coulomb solution the principal null vectors (in a coordinate system where the change is at rest) are given by $l^\mu = [(x^2 + y^2 + z^2)^{1/2}, x, y, z]$ and $n^\mu = [-(x^2 + y^2 + z^2)^{1/2}, x, y, z)]$, and it is easy to see $\pounds_l F_{\mu\nu} = F_{\mu\nu}$. Similar equations hold for n^μ and ${}^*F_{\mu\nu}$, and thus it has a null-type symmetry in both its principal null directions.

If $F^{\mu\nu}$ and ${}^*F^{\mu\nu}$ are studied the conditions for a null-type symmetry are not the same as those of $F_{\mu\nu}$ and ${}^*F_{\mu\nu}$ separately, but if the symmetry is to hold for $F^{\mu\nu}$ and ${}^*F^{\mu\nu}$ the same equations (12)–(15) are obtained.

4. THE GENERIC WEYL TENSOR CASE

The Lie derivative of the Weyl conformal tensor expressed in a spinor representation is given by defining $E_{RR'SS'UU'VV'}$ as

$$E_{RR'SS'UU'VV'} = \xi^{AA'}[\epsilon_{RS}\epsilon_{UV}\bar{\Psi}_{R'S'U'V'})_{;AA'} + [\epsilon_{RS}\epsilon_{UA}\Psi_{R'S'U'A}]\xi^{AA'}_{;VV'} + [\epsilon_{RS}\epsilon_{AV}\bar{\Psi}_{R'S'A'V'}]\xi^{AA'}_{;UV'} + [\epsilon_{RA}\epsilon_{UV}\bar{\Psi}_{R'A'U'V'}]\xi^{AA'}_{;SS'} + [\epsilon_{AS}\epsilon_{UV}\bar{\Psi}_{A'S'U'V'}]\xi^{AA'}_{;RR'},$$

then

$$E_{RR'SS'UU'VV'} + \overline{E_{RR'SS'UU'VV'}} = \pounds_\xi C_{RR'SS'UU'VV'}, \quad (16)$$

where

$$C_{RR'SS'UU'VV'} = \epsilon_{RS}\epsilon_{UV}\bar{\Psi}_{R'S'U'V'} + \epsilon_{R'S'}\epsilon_{U'V'}\bar{\Psi}_{RSUV}. \quad (17)$$

${}^*C_{RR'SS'UU'VV'}$ is obtained from $C_{RR'SS'UU'VV'}$ by replacing $\bar{\Psi}_{R'S'U'V'}$ by $i\bar{\Psi}_{R'S'U'V'}$ and Ψ_{RSUV} by $-i\Psi_{RSUV}$. $\pounds_\xi\, {}^*C_{RR'SS'UU'VV'}$ is then given by

$$\pounds_\xi\, {}^*C_{RR'SS'UU'VV'} = iE_{RR'SS'UU'VV'} - i\overline{E_{RR'SS'UU'VV'}}. \quad (18)$$

Equations (1) and (2) can be written in spinor form as

$$E_{RR'SS'UU'VV'} + \overline{E_{RR'SS'UU'VV'}} = \epsilon_{RS}\epsilon_{UV}\Omega\bar{\Psi}_{R'S'U'V'} + \epsilon_{R'S'}\epsilon_{U'V'}\bar{\Omega}\Psi_{RSUV}, \quad (19)$$

$$E_{RR'SS'UU'VV'} - \overline{E_{RR'SS'UU'VV'}} = \epsilon_{RS}\epsilon_{UV}\Omega\bar{\Psi}_{R'S'I'V'} - \epsilon_{R'S'}\epsilon_{U'V'}\bar{\Omega}\Psi_{RSUV}, \quad (20)$$

where $\Omega = Ae^{i\theta}$. Equations (19) and (20) together give the single spinor equation

$$E_{RR'SS'UU'VV'} = \Omega\epsilon_{RS}\epsilon_{UV}\bar{\Psi}_{R'S'U'V'}. \quad (21)$$

Since $\pounds_\xi\, C_{RR'SS'UU'VV'}$ has the same symmetries as the Riemann tensor it can be written in the form[2]

$$\pounds_\xi\, C_{RR'SS'UU'VV'} \equiv B_{RR'SS'UU'VV'} = \epsilon_{R'S'}\epsilon_{U'V'}B_{(RSUV)} + \epsilon_{RS}\epsilon_{UV}\bar{B}_{(R'S'U'V')} + \tfrac{1}{6}(\epsilon_{RV}\epsilon_{SU} + \epsilon_{RU}\epsilon_{SV}) \times \epsilon_{A'B'}\epsilon_{U'V'}B + \tfrac{1}{6}(\epsilon_{R'V'}\epsilon_{S'U'} + \epsilon_{R'U'}\epsilon_{S'V'})\epsilon_{RS}\epsilon_{UV}\bar{B} + \epsilon_{UV}\epsilon_{R'S'}D_{RSU'V'} + \epsilon_{RS}\epsilon_{U'V'}D_{UVR'S'}, \quad (22)$$

$$B_{ABCD} = \tfrac{1}{4}B_{AP'B}{}^{P'}{}_{CQ'D}{}^{Q'},$$

$$B = B_{AB}{}^{AB} = \bar{B},$$

$$D_{ABY'Z'} = \tfrac{1}{4}B_{AP'B}{}^{P'}{}_{GY'Z'}{}^{G}.$$

Equation (1) holding for a given Weyl tensor is thus equivalent to

$$B = 0,$$

$$D_{RSU'V'} = 0,$$

$$\bar{B}_{(R'S'U'V')} = \Omega\bar{\Psi}_{R'S'U'V'} = \Omega\bar{\alpha}_{(R'}\bar{\beta}_{S'}\bar{\gamma}_{U'}\bar{\delta}_{V')},$$

where $\bar{\alpha}_{R'}\bar{\beta}_{S'}\bar{\gamma}_{U'}\bar{\delta}_{V'}$ are the principal null directions of $C_{\rho\sigma\mu\nu}$. That is $B_{\rho\sigma\mu\nu}$ has the same algebraic structure and principal null directions as $C_{\rho\sigma\mu\nu}$, and differs from it only in the complex factor Ω.

The proof of the invariance of Eqs. (1) and (2) under the transformations $g_{\mu\nu} \to \sigma g_{\mu\nu}$ $\hat{g}_{\mu\nu}$ follow from $\hat{C}_{\rho\sigma\mu\nu} = \Omega C_{\rho\sigma\mu\nu}$, and ${}^*\hat{C}_{\rho\sigma\mu\nu} = \Omega {}^*\hat{C}_{\rho\sigma\mu\nu}$ along with the nonmetric dependence of the $\pounds_\xi$ operator. Thus

$$\pounds_\xi\, \hat{C}_{RR'SS'UU'VV'} = (\pounds_\xi\sigma)C_{RR'SS'UU'VV'} + \sigma\pounds_\xi\, C_{RR'SS'UU'VV'} = \{[(\pounds_\xi\sigma) + \bar{\Omega}\sigma]/\sigma\}[\epsilon_{R'S'}\epsilon_{U'V'}\hat{\Psi}_{RSUV}] + \{[(\pounds_\xi\sigma) + \Omega\sigma]/\sigma\}[\epsilon_{RS}\epsilon_{UV}\hat{\Psi}_{R'S'U'V'}]. \quad (23)$$

The proof for (2) is essentially the same.

Equations (1) and (2) immediately admit an integrability condition. As $g^{\rho\nu}C_{\rho\sigma\mu\nu} = g^{\rho\nu}{}^*C_{\rho\sigma\mu\nu} = 0$, it follows from these and Eq. (1) that

$$h^{\rho\nu}C_{\rho\sigma\mu\nu} = 0, \quad \pounds_\xi\, g^{\rho\nu} = h^{\rho\nu}. \quad (24)$$

From Eq. (2) it follows that

$$h^{\rho\nu}{}^*C_{\rho\sigma\mu\nu} = 0.$$

If both equations hold, then the process can be continued indefinitely to get the set of equations

$$
\begin{aligned}
&i^{\rho\nu}C_{\rho\sigma\mu\nu} = 0,\\
&i^{\rho\nu} = \pounds_\xi h^{\rho\nu} = \pounds_\xi^2 g^{\rho\nu},\\
&i^{\rho\nu}{}^*C_{\rho\sigma\mu\nu} = 0,\\
&j^{\rho\nu}C_{\rho\sigma\mu\nu} = 0,\\
&\pounds_\xi\, i^{\rho\nu} = j^{\rho\nu},\\
&j^{\rho\nu}{}^*C_{\rho\sigma\mu\nu} = 0.\\
&\quad\vdots
\end{aligned}
\tag{25}
$$

Also if Eq. (1) or (2) holds with Ω real a set of equations of the form of (25) just involving $C_{\rho\sigma\mu\nu}$ or ${}^*C_{\rho\sigma\mu\nu}$ alone must hold.

These integrability conditions are trivial in the case where ξ^α is a conformal motion

$$(\pounds_\xi\, g_{\rho\nu} = \Phi g_{\rho\nu}).$$

In all cases it is only the trace-free parts of $h^{\rho\nu}$, $\iota^{\rho\nu}$, etc. that enter into the integrability conditions.

5. THE NULL WEYL CASE

For $C_{\rho\sigma\mu\nu}$ of type N,

$$\Psi_{ABCD} = {}^\circ_A{}^\circ_B{}^\circ_C{}^\circ_D,$$

and $\xi^{AA'}$ is taken to be of the form

$$\xi^{AA'} = \xi^\circ{}^A{}^\circ{}^{A'}.$$

The easiest way to establish (1) and (2) is to use the spin dyad method. Expanding Eq. (19) in dyad components gives the independent set of equations

$$\xi(4\epsilon - 2\rho) = \overline{\Omega}, \tag{26a}$$

$$\kappa = 0, \tag{26b}$$

$$\sigma + \overline{\sigma} = 0. \tag{26c}$$

All other components are identically zero, or just the complex conjugate or negative of one of these equations. Notice that the set of equations (26) can be satisfied for any value of ξ. This is unlike the case of a conformal motion, where two different conformal Killing vectors cannot have the same streamlines.[6] For ξ^α to generate a conformal motion it is necessary that Ω be real, since then Eq. (1) is an integrability condition of the conformal Killing equation.

From the Goldberg–Sachs theorem,[5] or directly from the vacuum Bianchi identities, κ and σ must be zero for Ricci-flat spaces. $\kappa = 0$ is the condition that the principal null vector field is tangent to a geodesic congruence, and $\sigma = 0$ the condition that the congruence is shear free. The real part of ρ measures the expansion of the congruence, while its imaginary part gives its twist. When ϵ is nonzero the tangent vectors ${}^\circ{}^A$ are generated by a nonaffine paramaterization of the geodesics. Since Eq. (26a) can be always satisfied Eq. (1) has now been shown to hold for all conformally Ricci-flat-type N spaces. The same method will work for the case of the dual tensor, ${}^*C_{\rho\sigma\mu\nu}$.

If the Lie derivative of $C^{\rho\sigma\mu\nu}$ with respect to a principal null direction is examined equations of the same form are obtained, and the same type of theorem can be proved. In fact all of this paper could have been done for the Maxwell and Weyl tensors with all indices up. On the other hand, if mixed index tensors are studied (F^μ_ν or $C^\rho_{\sigma\mu\nu}$), equations of the type of (5) or (1) will not in general be true. This presents a difficult problem in interpretating these results.

ACKNOWLEDGMENTS

I wish to thank R. Geroch, W. Israel, and H. Schiff for helpful discussion on this work.

*Work supported in part by the National Research Council of Canada.

[1] Spaces for which $\pounds_\xi R^\rho_{\sigma\mu\nu} = 0$ have been studied by G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).

[2] F. A. E. Pirani, *Lectures on General Relativity* (Prentice-Hall, Englewood Cliffs, N. J., 1965).

[3] The basic content of these theorems is the same as that contained in propagation equations for null type tensors; see Ref. 2, pp. 360–363.

[4] For similar considerations see G. C. Debney and J. D. Zund, Tensor **25**, 53 (1972).

[5] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[6] K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1955), p. 157.

Kinetic theory of a one-dimensional model

G. R. Anstis and H. S. Green

Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia

D. K. Hoffman*

Department of Chemistry and Ames Laboratory, USAEC, Iowa State University, Ames, Iowa 50010
(Received 6 December 1972)

The kinetic theory of a model previously studied by Jepsen and others is reformulated in terms of the theory of the grand partition function and the hierachy of equations satisfied by the distribution functions derived therefrom. It is shown that the principle of molecular chaos implies an exact factorization of the two- and many-body distribution functions, which allows the equations of the hierachy to be solved exactly, with the recovery of results obtained by other methods. The methods of the present paper also allow examination of the origin of the divergences generated by well-known approximations, and suggest alternative approximations in which such divergences do not appear.

INTRODUCTION

The development of the theory of transport properties of dense gases has met with certain difficulties. The usual method of obtaining approximations to the correct formulation of the principle of molecular chaos is to consider the dynamics of small numbers of temporarily isolated particles in the fluid environment. One then obtains the generalization of Boltzmann's equation due to Bogoliubov[1] and Green[2] in one approximation and the Choh–Uhlenbeck equation[3] in another. However, higher order approximations lead to divergences in the density expansion of transport coefficients.

A further major problem is that inconsistencies arise when Boltzmann's equation is applied to a system where bound states can occur. Attempts have been made by Green and Hoffman[4] to rectify this problem.

The purpose of the present paper is to examine the divergences associated with the exactly soluble model of a system of one-dimensional hard rods. This model has previously been discussed by Jepsen[5] and Lebowitz and Percus.[6] We show here that it is possible to give an exact formulation of the principle of molecular chaos which leads to the known exact value of the self-diffusion coefficient. However, if the usual series of approximations to the principle of molecular chaos is made, it is found that the first approximation gives the exact result but that the Choh–Uhlenbeck term is not even zero but divergent. The observation that the exact result is obtained by just considering the first approximation has been made by Blum and Lebowitz,[7] who, however, conjectured that higher order terms would be zero.

In more complicated systems the principle of molecular chaos cannot be stated so simply, and the problem is to find a series of successive approximations which will give convergent results at each stage. We consider a simple generalization of the hard sphere model and indicate how the Choh–Uhlenbeck term may be modified to give finite results. While this is a far from general treatment of the divergence difficulties for two- and three-dimensional models with more general forces it may give some indication of the sources of the difficulties and how it may be remedied.

In Sec. 1 we derive the appropriate kinetic equations for a one-dimensional system of hard rods of arbitrary thickness.

In Sec. 2 we give the exact formulation of the principle of molecular chaos and solve the closed set of equations which result from its use.

In Sec. 3 it is shown that the Choh–Uhlenbeck approximation to the exact formulation of molecular chaos is divergent even though the lowest order term gives the exact result. Finally a one-dimensional model in which there is a nonzero probability of penetration of particles on collision is considered. It is possible to develop a series of approximations which are convergent at each stage.

1. THE DISTRIBUTION FUNCTIONS AND THEIR HIERARCHICAL EQUATIONS

In this section we wish to examine the behavior in time of a labelled "test" particle in a one-dimensional system and also to study the response of the background particles. We assume that the test particle has a specified initial position and velocity and that the particles which make up the background initially have an equilibrium distribution. The ensemble which we consider consists of subensembles of systems which have M particles to the left and N particles to the right of the test particles, where M and N range from 0 to ∞. The particles in any given system can be labelled according to their initial position relative to the test particle (which is taken to be particle zero). In the subensemble with M particles to the left and N particles to the right of the test particle, the distribution of particles is given by the distribution function $F^{M,N}$, which obeys the $M + N + 1$ particle Liouville equation

$$\left(\frac{\partial}{\partial t} + \kappa^{M,N}\right) F^{M,N} = 0, \tag{1}$$

where $\kappa^{M,N}$ is the Liouville operator for the system. The ensemble average $\langle G\rangle$ of any property G is then given by

$$\langle G\rangle = \sum_{M=0}^{\infty}\sum_{N=0}^{\infty}\left(\prod_{i=-M}^{N}\int d\chi_i(t)\right) F^{M,N}\, G^{M,N}(\chi_{-M}(t),\ldots,\chi_N(t)),$$

$$g_{-M} < \cdots < g_N, \tag{2}$$

where $\chi_i(t) = g_i(t)$, $v_i(t)$ denotes the position and velocity of particle i at time t and $\chi_i = g_i, v_i$ denotes the initial position and velocity of particle i. Since the integral limits in (2) are expressed in terms of the initial particle distribution, it is somewhat more convenient to compute $\langle G\rangle$ from the initial distribution as follows:

$$\langle G\rangle = \sum_{M=0}^{\infty}\sum_{N=0}^{\infty}\left(\prod_{i=-M}^{N} d\chi_i\right) F_0^{M,N}\, G_t^{M,N},$$

$$g_{-M} < \cdots < g_N. \tag{3}$$

Here $F_0^{M,N}$ is the initial distribution which is given by

$$F_0^{M,N} = \delta(\chi_0 - \chi') \exp\{\beta[\mu(M+N) - H^{M,N})]\}/\Xi_L \Xi_R, \quad (4)$$

where $\chi' = g', v'$ specifies the initial position and velocity of the test particles, $\beta = 1/kT$ is the statistical temperature, μ is the chemical potential, and $H^{M,N}$ is the $(M+N+1)$-particle Hamiltonian. The quantity Ξ_L is the equilibrium grand ensemble partition function for the particles to the left of the test particle and Ξ_R is similarly defined for the particles to the right. Finally,

$$G_t^{M,N} = G^{M,N}(\chi_{-M}(t), \ldots, \chi_N(t)).$$

The probability that the jth particle is in the phase volume $d\chi_\alpha$ about the phase point $\chi_\alpha = g_\alpha, v_\alpha$ is $f_j^{(1)}(\chi_\alpha, t)\, d\chi_\alpha$, where

$$f_j^{(1)}(\chi_\alpha, t) = \langle \delta(\chi_\alpha - \chi_j(t)) \rangle. \quad (5)$$

The function $f_j^{(1)}$ (as well as the other distribution functions which we later introduce), of course, also depends on $\chi' = g', v'$, the initial position and velocity of the test particle. The pair distribution function corresponding to the jth particle being at χ_α and the kth particle at χ_β is given by

$$f_{j,k}^{(2)}(\chi_\alpha, \chi_\beta, t) = \langle \delta(\chi_\alpha - \chi_j(t)) \delta(\chi_\beta - \chi_k(t)) \rangle. \quad (6)$$

A closely related quantity is $f_j^{(2)}(\chi_\alpha, \chi_\beta, t)$ which gives the probability that particle j is at χ_α and any other particle is at χ_β. It is defined

$$f_j^{(2)}(\chi_\alpha, \chi_\beta, t) = \sum_{k \neq j} f_{j,k}^{(2)}(\chi_\alpha, \chi_\beta, t). \quad (7)$$

By obvious extension of these definitions, a whole family of distribution functions can be defined which serve to completely define the state of the ensemble at any time. These functions obey a hierarchy of coupled equations which are very similar to the BBGKY equations. For instance, from (3), (5) and the fact that $(\partial/\partial t)\, \delta(\chi_\alpha - \chi_j(t)) = \kappa^{M,N} \delta(\chi_\alpha - \chi_j(t))$ it is easy to show that

$$\left(\frac{\partial}{\partial t} + v_\alpha \frac{\partial}{\partial g_\alpha}\right) f_j^{(1)}(\chi_\alpha, t) = \frac{1}{m} \int d\chi_\beta \frac{\partial \phi_{\alpha,\beta}}{\partial g_\alpha} \frac{\partial}{\partial v_\alpha} f_j^{(2)}(\chi_\alpha, \chi_\beta, t), \quad (8)$$

where m is the particle mass, and $\phi_{\alpha,\beta} = \phi(g_\alpha - g_\beta)$ is the interaction potential between a particle at g_α and one at g_β. Similarly we find that $f_{j,k}^{(2)}$ obeys the equation

$$\left(\frac{\partial}{\partial t} + v_\alpha \frac{\partial}{\partial g_\alpha} + v_\beta \frac{\partial}{\partial g_\beta} - \frac{1}{m} \frac{\partial \phi_{\alpha,\beta}}{\partial g_\alpha} \frac{\partial}{\partial v_\alpha} - \frac{1}{m} \frac{\partial \phi_{\alpha,\beta}}{\partial g_\alpha} \frac{\partial}{\partial v_\beta}\right) \times f_{j,k}^{(2)}(\chi_\alpha, \chi_\beta, t) = \int d\chi_\gamma \left(\frac{\partial \phi_{\alpha,\gamma}}{\partial g_\alpha} \frac{\partial}{\partial v_\alpha} + \frac{\partial \phi_{\beta,\gamma}}{\partial g_\beta} \frac{\partial}{\partial v_\beta}\right) f_{j,k}^{(3)}(\chi_\alpha, \chi_\beta, \chi_\gamma, t) \quad (9)$$

and that the other distribution functions obey coupled equations of this kind.

We now consider the specific case of a gas of impenetrable points (i.e., rods of zero length) which are initially constrained to lie in a box of length $2L$ with the test particle at the origin of the box center (i.e., $g' = 0$). Those particles to the right of the origin, as well as the ones to the left, are initially uniformly distributed with a Maxwellian velocity distribution. Although the initial total particle density is uniform (except for the particle at the origin), the distribution of labelled particles is, of course, a function both of position and particle label.

The initial distribution is found from (3), (4), and (5) to be

$$\begin{aligned} &\text{for } j = 0 \quad f_0^{(1)}(\chi_\alpha, 0) = \delta(g_\alpha)(v_\alpha - v'), \\ &\text{for } j > 0 \quad f_j^{(1)}(\chi_\alpha, 0) = \frac{\rho(\rho g_\alpha)^{j-1}}{(j-1)!} \exp(-\rho g_\alpha) h_0(v_\alpha), \\ &\qquad 0 \leq g_\alpha \leq L \\ &\qquad\qquad = 0, \quad \text{otherwise}, \\ &\text{for } j < 0 \quad f_j^{(1)}(\chi_\alpha, 0) = f_{-j}^{(1)}(-g_\alpha, v_\alpha, 0), \end{aligned} \quad (10)$$

where ρ is the initial particle density and $h_0(v) = (m\beta/2\pi)^{1/2} \exp(-\frac{1}{2}\beta m v^2)$. The restraining walls are removed at time $t = 0$, and the system is allowed to evolve in time without external constraint.

Before proceeding we wish to show that the distribution functions for the more general case where the rods are of length $a \neq 0$ are trivially related to those for the $a = 0$ case. For convenience in examining this question, we use a canonical ensemble of systems with N particles to the right and N particles to the left of the test particles. In the thermodynamic limit the choice of ensembles is, of course, immaterial. The independent intensive thermodynamic variables for the ensemble in its initial state are the temperature and the density, $\rho = N/L$. We now define a new set of initial position coordinates by

$$\bar{g}_j = g_j - ja, \quad j = -N \text{ to } N.$$

Clearly $\bar{g}_j - \bar{g}_i$ is the "unoccupied" distance between the jth particle and the ith particle. From a study of the dynamics of the system of rods, it is apparent that $\bar{g}_j(t) = g_j(t) - ja$ is identical to the position of the jth particle at time t in a system of rods of zero length which were initially in the state $\bar{g}_i, v_i$, $i = -N$ to N. Also, taking into account the impenetrability of the rods (which does not allow the centers of neighboring rods to approach closer than within a distance a of one another), the integration limits of (3) can be written

$$L - Na \geq \bar{g}_N \geq \bar{g}_{n-1} \geq \cdots \geq \bar{g}_1 \geq 0 \geq \bar{g}_{-1} \geq \cdots \geq \bar{g}_{-N+1} \geq \bar{g}_{-N} \geq -L + Na.$$

Taking these points into account, we see that the system of rods of length $a \neq 0$ is essentially equivalent to a system of impenetrable points in a box of length $2(L - Na)$. Thus

$$f_j^{(1)}(g_\alpha, v_\alpha, t \mid a, \rho) = f_j^{(1)}(g_\alpha - ja, v_\alpha, t \mid 0, \rho/(1-\rho a)),$$

where we have indicated the explicit dependence of $f_j^{(1)}$ on the length of the rods and the initial density. Similarly

$$f_{j,k}^{(2)}(g_\alpha, v_\alpha, g_\beta, v_\beta, t \mid a, \rho) = f_{j,k}^{(2)}(g_\alpha - ja, v_\alpha, g_\beta - ka, v_\beta, t \mid 0, \rho/(1-\rho a))$$

and the other distribution functions for $a \neq 0$ can be related to their $a = 0$ counterparts by obvious extension.

For the case of impenetrable point particles, (8) can be written in a somewhat more convenient form. We can obtain from (9) an equation for $f_j^{(2)}$ of the form

$$\left((v_\alpha - v_\beta)\frac{\partial}{\partial(g_\alpha - g_\beta)} - \frac{1}{m}\frac{\partial\phi_{\alpha,\beta}}{\partial g_\alpha}\frac{\partial}{\partial v_\alpha} - \frac{1}{m}\frac{\partial\phi_{\alpha,\beta}}{\partial g_\beta}\frac{\partial}{\partial v_\beta}\right) \times f_j^{(2)}(\chi_\alpha, \chi_\beta, t) = \text{other terms.} \quad (11)$$

Now for the interaction which we are considering it is seen that the integration range of g_β in (8) can be limited to an arbitrarily small line segment containing g_α and, in the limit that this range became vanishingly small, only δ-function terms in the integrand contribute to the integral. The δ-function terms in the $f_j^{(2)}$ equation are explicitly displayed in (11). Substitution of (11) into (8) then leads rigorously to the equation

$$\begin{aligned}\left(\frac{\partial}{\partial t} + v_\alpha\frac{\partial}{\partial g_\alpha}\right)f_j^{(1)}(\chi_\alpha, t) &= \lim_{\epsilon\to 0}\int_{g_\alpha - \epsilon}^{g_\alpha+\epsilon} dg_\beta \int dv_\beta (v_\beta - v_\alpha)\frac{\partial}{\partial(g_\beta - g_\alpha)} f_j^{(2)}(\chi_\alpha, \chi_\beta, t)\\ &= \lim_{\epsilon\to 0}\int dv_\beta\, |v_\beta - v_\alpha|\,[f_j^{(2)}(g_\alpha, v_\alpha, g_\alpha^+, v_\beta, t)\\ &\quad - f_j^{(2)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, t)], \end{aligned} \quad (12)$$

where $g_\alpha^+ = g_\alpha + \epsilon\,\mathrm{sgn}(v_\beta - v_\alpha)$ and $g_\alpha^- = g - \epsilon\,\mathrm{sgn}(v_\beta - v_\alpha)$. The variables $g_\alpha, v_\alpha, g_\alpha^+, v_\beta$ are postcollision variables whereas $g_\alpha, v_\alpha, g_\alpha^-, v_\beta$ are precollision variables. But from (11)

$$\lim_{\epsilon\to 0} f_j^{(2)}(g_\alpha, v_\alpha, g_\alpha^+, v_\beta, t) = \lim_{\epsilon\to 0} f_j^{(2)}(g_\alpha^-, v_\beta, g_\alpha, v_\alpha, t), \quad (13)$$

where we have again made use of the fact that the "other terms" of (11) can be ignored in the $\epsilon \to 0$ limit [which reduces (11) to a Liouville equation in the pair space]. We have also made use of the fact that, for this system, two particles simply exchange velocities upon collision. Combining (12) and (13) yields

$$\left(\frac{\partial}{\partial t} + v_\alpha\frac{\partial}{\partial g_0}\right)f_j^{(1)}(\chi_\alpha, t) = \lim_{\epsilon\to 0}\int dv_\beta\,|v_\beta - v_\alpha| \times [f_j^{(2)}(g_\alpha^-, v_\beta, g_\alpha, v_\alpha, t) - f_j^{(2)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, t)]. \quad (14)$$

We can now make use of the fact that the particles are impenetrable [and hence the jth particle is always next to the $(j+1)$th and $(j-1)$th particles] to write

$$\lim_{\epsilon\to 0} f_j^{(2)}(g_\alpha^-, v_\beta, g_\alpha, v_\alpha, t) = \lim_{\epsilon\to 0} f_j^{(2)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, t),$$

where

$$(j) = j + \mathrm{sgn}(v_\beta - v_\alpha).$$

Equation (14) can then be written in the form

$$\left(\frac{\partial}{\partial t} + v_\alpha\frac{\partial}{\partial g_\alpha}\right)f_j^{(1)}(\chi_\alpha, t) = \lim_{\epsilon\to 0}\int dv_\beta\,|v_\beta - v_\alpha| \times [f_{(j)}^{(2)}(g_\alpha, v_\beta, g_\alpha^-, v_\beta, t) - f_j^{(2)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, t)], \quad (15)$$

which is equivalent to (8) for impenetrable point particles.

2. SOLUTION OF EQUATIONS OF THE HIERARCHY

Equation (15) governs the time evolution of $f_j^{(1)}$ and involves $f_j^{(2)}$. By an almost identical procedure, we can derive an equation for $f_j^{(2)}$ involving $f_j^{(3)}$ and one for $f_j^{(3)}$ involving $f_j^{(4)}$, etc. In general, in these equations, it is not necessary to take into account any interactions other than those of the jth particle with the unlabelled particles since $f_j^{(n)}$ is symmetric with respect to unlabelled particle interchange. (Since, for the system we consider, two particles simply exchange momenta upon collision, the unlabelled particles behave dynamically like an ideal gas.) The equations governing $f_j^{(n)}, n \geq 1$, then are

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \kappa_j^{(n)}\right)f_j^{(n)}(g_\alpha, v_\alpha, \ldots, t) &= \lim_{\epsilon\to 0}\int dv_\beta\,|v_\beta - v_\alpha|\,[f_{(j)}^{(n+1)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, \ldots, t)\\ &\quad - f_j^{(n+1)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, \ldots, t)], \end{aligned} \quad (16)$$

where $\kappa_j^{(n)}$ is the n-particle Liouville operator which includes only interactions between particle j and the unlabelled particles. These equations form an infinite coupled hierarchy which can be exactly solved. We now solve the equations to the extent of finding an explicit expression for $f_j^{(1)}$.

We first consider the hierarchy of (16) for $n \geq 2$ by introducing a new function

$$\begin{aligned} g_j^{(n)}(g_\alpha, v_\alpha, g_\gamma, v_\gamma, \ldots, t) &= f_j^{(n)}(g_\alpha, v_\alpha, g_\gamma, v_\gamma, \ldots, t)\\ &\quad - f_{j^*}^{(n-1)}(g_\alpha, v_\alpha, \ldots, t)h(g_\gamma, v_\gamma, t), \end{aligned} \quad (17)$$

where

$$\begin{aligned} j^* &= j-1, \quad 0 < g_\gamma - v_\gamma t < g_\alpha - v_\alpha t,\\ &= j+1, \quad g_\alpha - v_\alpha t < g_\gamma - v_\gamma t < 0,\\ &= j, \quad \text{otherwise,} \end{aligned}$$

and

$$h(g_\gamma, v_\gamma, t) = \rho h_0(v_\gamma) + \delta(g_\gamma - v't)\delta(v_\gamma - v')$$

for $-L < g_\gamma - \gamma_\gamma t < L$ and is zero otherwise. The function $h(g_\gamma, v_\gamma, t)$ is just the distribution of particles on the line at time t. The function $g_j^{(n)}$ is manifestly symmetric in the particle coordinates for the $n-2$ particles whose coordinates are not explicitly shown in (17).

We now confine our attention to the precollision regions of the n-particle phase space in which $g_j^{(n)}$ is defined. A phase point in a precollision region corresponds to an n-particle system for which particle j has suffered no collision in its past history under the n-particle motion. (Note that we have not excluded the possibility of the crossing of unlabelled particle trajectories in the past history of the system.)

Using the fact that

$$\left(\frac{\partial}{\partial t} + v_\gamma\frac{\partial}{\partial g_\gamma}\right)h(g_\gamma, v_\gamma, t) = 0,$$

we have from (16) that in precollision regions

$$\begin{aligned}\frac{d^{(n)}}{dt^{(n)}}\, g_j^{(n)}(g_\alpha, v_\alpha, g_\gamma, v_\gamma, \ldots, t) &= \lim_{\epsilon\to 0}\int dv_\beta\,|v_\beta - v_\alpha|\,[g_{(j)}^{(n+1)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, g_\gamma, v_\gamma, \ldots, t)\\ &\quad - g_j^{(n+1)}(g_\alpha, v_\alpha, g_\alpha^-, v_\beta, g_\gamma, v_\gamma, \ldots, t)], \end{aligned} \quad (18)$$

where

$$\frac{d^{(n)}}{dt^{(n)}} = \frac{\partial}{\partial t} + v_\alpha \frac{\partial}{\partial g_\alpha} + v_\gamma \frac{\partial}{\partial g_\gamma} + \cdots .$$

It should be remarked that the functions $g_j^{(n)}$ appearing on both sides of equations of this hierarchy are evaluated only in precollision regions of their phase spaces. Then, since as is easy to show $g_j^{(n)} = 0$ at $t = 0$, it follows that $g_j^{(n)} = 0$ in all precollision regions of the n-particle phase space. Thus, from (17), we have that

$$f_j^{(n)}(g_\alpha, v_\alpha, g_\gamma, v_\gamma, \ldots, t) = f_{j*}^{(n-1)}(g_\alpha, v_\alpha, \ldots, t)h(g_\gamma, v_\gamma, t) \tag{19}$$

in precollision regions of the n-particle phase space. This relation can be viewed as an exact statement of the principle of molecular chaos for the systems which we consider.

Since (16) for $n = 1$ involves $f_j^{(2)}$ only in precollision regions of the two particle phase space, the factorization of (19) is directly applicable in this equation. Substitution of (19) for $n = 2$ into (16) for $n = 1$ yields the equation

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial g}\right) f_j^{(1)}(\chi, t) = \gamma_0[f_{j-1}^{(1)}(\chi, t) - f_j^{(1)}(\chi, t)] + \beta_0[f_{j+1}^{(1)}(\chi, t) - f_j^{(1)}(\chi, t)]$$

$$+ \lim_{\xi\to 0}\delta(g^- - v't)\begin{Bmatrix}\epsilon(g - vt)(v' - v)[f_{j+1}^{(1)}(\chi, t) - f_j^{(1)}(\chi, t)] \\ + \\ \epsilon(vt - g)(v - v')[f_{j-1}^{(1)}(\chi, t) - f_j^{(1)}(\chi, t)]\end{Bmatrix}, \tag{20}$$

where $\chi = (g, v)$, $g^- = g - \operatorname{sgn}(v' - v)$, $\epsilon(\ )$ is the unit step function,

$$\beta_0 = \int_{g/t}^{(g+L)/t} d\omega(\omega - v)h_0(\omega)$$

and

$$\gamma_0 = \int_{(g-L)/t}^{g/t} d\omega(v - \omega)h_0(\omega).$$

Because of the δ-function term on the rhs of this equation, $f_j^{(1)}$ is discontinuous at $g = v't$ and consequently $\lim_{\xi\to 0}\delta(g^- - v't)$ cannot be simply replaced by $\delta(g - vt)$. In the immediate neighborhood of the discontinuity (20) can be simplified to

$$(v - v')\frac{\partial}{\partial(g - v't)} f_j^{(1)} = \lim_{\xi\to 0}\delta(g^- - v't)\begin{Bmatrix}\epsilon(g - vt)(v' - v)[f_{j+1}^{(1)} - f_j^{(1)}] \\ + \\ \epsilon(vt - g)(v - v')[f_{j-1}^{(1)} - f_j^{(1)}]\end{Bmatrix}.$$

Integrating this equation across the discontinuity yields the following relation between the values of $f_j^{(1)}$ on either side of the discontinuity:

$$\lim_{\xi\to 0} f_j^{(1)}(v't - \xi, v, t) = \lim_{\xi\to 0} f_{j+1}^{(1)}(v't + \xi, v, t). \tag{21}$$

Using these relations and those of (10) as boundary conditions, we can solve the equation

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial g}\right) f_j^{(1)}(\chi, t) = \gamma_0[f_{j-1}^{(1)}(\chi, t) - f_j^{(1)}(\chi, t)] + \beta_0[f_{j+1}^{(1)}(\chi, t) - f_j^{(1)}(\chi, t)], \tag{22}$$

which is valid everywhere except at the previously mentioned point of discontinuity.

To solve this equation, we define the transform function $\eta(\theta)$ by

$$\eta(\theta) = \sum_{j=-\infty}^{\infty} e^{\hat{\imath}\theta j} f_j^{(1)}, \tag{23}$$

which has the inverse transform

$$f_j^{(1)} = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{-\hat{\imath}\theta j}\eta(\theta). \tag{24}$$

From (22) and (23) we find that $\eta(\theta)$ obeys the equation

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial g}\right)\ln\eta(\theta) = \gamma_0(e^{\hat{\imath}\theta} - 1) + \beta_0(e^{-\hat{\imath}\theta} - 1), \tag{25}$$

which can be immediately integrated to obtain

$$\eta(\theta, g, v, t) = \eta(\theta, g - v(t - t_{0+}), v, t_{0+})\exp\left(\int_{t_{0+}}^{t} dt[\gamma_0(e^{\hat{\imath}\theta} - 1) + \beta_0(e^{-\hat{\imath}\theta} - 1)]\right), \quad t > t_0 > 0,$$
$$= \eta(\theta, g - vt, v, 0)\exp\left(\int_0^t dt[\gamma_0(e^{\hat{\imath}\theta} - 1) + \beta_0(e^{-\hat{\imath}\theta} - 1)]\right), \quad \text{otherwise}, \tag{26}$$

where $t_0 = (g - vt)/(v' - v)$ and t_{0+} is infinitesimally larger than t_0. These results follow from integrating (25) down to the point of discontinuity at $t' = t_0$ or to $t' = 0$ if t_0' does not lie between 0 and t. It is easily verified that

$$\int_s^t dt'\beta_0 = \left(\int_{g/t}^{(g+L)/t} d\omega(\omega t - g)h_0(\omega) - \int_{Q/S}^{(Q+L)/S} d\omega(\omega S - Q)h_0(\omega) - L\int_{(Q+L)/S}^{(g+L)/t} d\omega h_0(\omega)\right) \tag{27}$$

and

$$\int_s^t dt'\gamma_0 = \rho\left(\int_{(g-L)/t}^{g/t} d\omega(g - \omega t)h_0(\omega) - \int_{(Q-L)/S}^{Q/S} d\omega(Q - \omega S)h_0(\omega) - L\int_{(g-L)/t}^{(Q-L)/S} d\omega h_0(\omega)\right), \tag{28}$$

where S is arbitrary and $Q = g - vt + vS$. It is also readily established from (10) and (23) that

$$\eta(\theta, g - vt, v, 0) = \delta(g - vt)\delta(v - v') + e^{\hat{\imath}\theta}\rho h_0(v)\epsilon(g - vt)\epsilon(L - g + vt) \times \exp[\rho(g - vt)(e^{\hat{\imath}\theta} - 1)] + e^{-\hat{\imath}\theta}\rho h_0(v)\epsilon(L + g - vt)\exp[\rho(vt - g)(e^{-\hat{\imath}\theta} - 1)] \tag{29}$$

and from (21) that

$$\eta(\theta, g - v(t - t_{0+}), v, t_{0+}) = \exp[i\theta\ \operatorname{sgn}(v - v')]\eta(\theta, g - v(t - t_{0-}), v, t_{0-}), \tag{30}$$

where t_{0-} is infinitesimally less than t_0. Finally, from (24), (26), (27), (28), (29), and (30), we have that

$$f_j^{(1)}(\chi, t) = \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{-\hat{\imath}\theta j}\{\delta(g - vt)\delta(v - v') + \rho h_0(v)\epsilon(t_0)\epsilon(t - t_0)\epsilon(L - g + vt)\epsilon(L + g - vt) + \rho h_0(v)[1 - \epsilon(t - t_0)\epsilon(t_0)][e^{\hat{\imath}\theta}\epsilon(L - g + vt)\epsilon(g - vt) + e^{-\hat{\imath}\theta}\epsilon(L + g - vt)\epsilon(vt - g)]\} \times \exp[(e^{\hat{\imath}\theta} - 1)\nu(t) + (e^{-\hat{\imath}\theta} - 1)\mu(t)], \tag{31}$$

where

$$\mu(t) = \int_{g/t}^{(g+L)/t} d\omega(\omega t - g)\rho h_0(\omega) + \rho L \int_{(g+L)/t}^{\infty} d\omega h_0(\omega)$$

and

$$\nu(t) = \int_{(g-L)/t}^{g/t} d\omega(g - \omega t)\rho h_0(\omega) + \rho L \int_{-\infty}^{(g-L)/t} d\omega h_0(\omega).$$

In the limit $L \to \infty$ and for $j = 0$ this reduces to the results previously obtained by Jepsen[5] and Lebowitz and Percus.[6]

Equation (31) can be expressed in terms of modified Bessel functions[8] by means of the identity

$$(1/2\pi) \int_0^{2\pi} d\theta\, e^{-i\theta k} \exp[(e^{i\theta} - 1)\nu(t) + e^{-i\theta} - 1)\mu(t)] = (\mu/\nu)^{k/2} I_k[2(\mu\nu)^{1/2}] \exp[-\mu(t) - \nu(t)] \equiv A_k(t) \quad (32)$$

$$f_j^{(1)} = [\delta(g - vt)\delta(v - v') + \rho h_0(v)\epsilon(t_0)\epsilon(t - t_0)]A_j + \rho h_0(v)[1 - \epsilon(t_0)\epsilon(t - t_0)][\epsilon(L - g + vt)\epsilon(g - vt)A_{j-1} + \epsilon(L + g - vt)\epsilon(vt - g)A_{j+1}]. \quad (33)$$

3. CONCLUSIONS

In the usual development of density corrections to Boltzmann's equation, an expression for the pair distribution function is derived by formally solving the BBGKY hierarchy subject to some assumed initial conditions about the ensemble (usually that there is no initial correlation). The pair distribution function is then given as a functional of the singlet distribution and is expressed in a density expansion. When this result for the pair distribution function is substituted into the first BBGKY equation, a closed and formally exact equation for the singlet distribution function is obtained. However, as is well known, all but the lowest order terms in the density expansion of the pair distribution are ill behaved.

Our exact solution of the hierarchy of Eqs. (16) for $n \geq 2$ closely parallels this procedure. To illustrate this, we write (16) in the form

$$\left(\frac{\partial}{\partial t} + \kappa^{(n)}\right) f_j^{(n)} = \lim_{\epsilon \to 0} \int d\omega\, |v - \omega| \left(f_{(j)}^{(n+1)} - f_j^{(n+1)}\right) \equiv J(f_j^{(n+1)}), \quad (34)$$

where $\kappa^{(n)}$ is the n-body Liouville operator. This equation can be formally integrated to yield

$$f_j^{(n)}(t) = \exp(-t\kappa^{(n)}) f_{j*}^{(n-1)}(0)h(0) + \int_0^t dt' \exp[-(t - t')\kappa^{(n)}] J(f_j^{(n+1)})(t'), \quad (35)$$

where we have made use of the exact, initial factorization condition

$$\exp(-t\kappa^{(n)}) f_j^{(n)}(0) = \exp(-t\kappa^{(n)}) f_{j*}^{(n-1)}(0)h(0).$$

Here j^* is defined as before. By iterative substitution of the members of the hierarchy of (35) into one another an expression for $f_j^{(n)}(t)$ as a functional of $f_k^{(1)}(t)$ can be obtained. (Various values of k occur in the expansion.) In particular, for the pair distribution function in precollision regions of the pair phase space, we obtain the result

$$f_j^{(2)}(t) = f_{j*}^{(1)}(t)h(t) + \text{C–U} + \cdots,$$

where the analog of the Choh–Uhlenbeck (C–U) term and higher correction terms are found to be identically zero. The fact that the first term in the expansion yields the exact result without correction results from the fact that if $f_j^{(2)}$ on the lhs of (34) or (35) is evaluated in a precollision configuration (in the sense previously defined) this is also the case for the distribution functions on the rhs. This is not a result of the simple dynamics of the system per se, but results from use of the identity

$$\lim_{\epsilon \to 0} f_j^{(n)}(g^-, \omega, g, v, \ldots, t) = \lim_{\epsilon \to 0} f_{(j)}^{(n)}(g, v, g^-, \omega, \ldots, t), \quad (36)$$

which is necessary to establish (16).

To illustrate this point, we now obtain a formal expression for $f_j^{(2)}$ by formal solution of the BBGKY hierarchy equations

$$\left(\frac{\partial}{\partial t} + \kappa^{(n)}\right) f_j^{(n)} = \frac{1}{m} \int d\chi_\gamma \frac{\partial \phi_{j\gamma}}{\partial g} \frac{\partial}{\partial v} f_j^{(n+1)}, \quad (37)$$

where $\phi_{j\gamma}$ is the interaction potential between the jth particle and an unlabelled particle whose coordinates are $\chi_\gamma = g_\gamma, v_\gamma$. (Here and in the subsequent discussion we use Greek letters to denote the coordinates of unlabelled particles.) It is easy to show that

$$F_j^{(n)}(t) = \exp(-t\kappa^{(n)})\, F_j^{(n)}(0),$$

where

$$F_j^{(n)} = f_j^{(n)} - \int d\chi_\alpha f_j^{(n+1)} + \tfrac{1}{2} \int d\chi_\alpha \int d\chi_\beta f_j^{(n+2)} + \cdots.$$

After some manipulation it can be deduced that

$$f_{j,\beta}^{(2)} = S_{j\beta}^{(t)} f_{j*}^{(1)} h_\beta + \int d\chi_\gamma (S_{j\beta\gamma}^{(t)} - S_{j\beta}^{(t)} S_{j\gamma}^{(t)}) f_{[j*]}^{(1)} h_\beta h_\gamma + \cdots, \quad (38)$$

where

$$S_{j\beta\gamma}^{(t)} = \exp(-t\kappa^{(3)}) \exp\left[t\left(v \frac{\partial}{\partial g} + v_\beta \frac{\partial}{\partial g_\beta} + v_\gamma \frac{\partial}{\partial g_\gamma}\right)\right]$$

and

$$S_{j\beta}^{(t)} = \exp(-t\kappa_{j\beta}^{(2)}) \exp\left[t\left(v \frac{\partial}{\partial g} + v_\beta \frac{\partial}{\partial g_\beta}\right)\right].$$

The S operators have the effect of permuting particle coordinates.

In (37), j^* is determined from the exact factorization of $\exp(-t\kappa_{j\beta}^{(2)}) f_{j,\beta}^{(2)}(0)$ whereas the value of $[j]$ depends on the S operator that acts on $f_{[j]}^{(1)} h_\beta h_\gamma$. For $S_{j\beta\gamma}$, $[j]$ is determined from the exact factorization of $\exp(-t\kappa^{(3)}) f_j^{(3)}(0)$ and, for $S_{j\beta} S_{j\gamma}$, $[j]$ is determined from the factorization of $\exp(-t\kappa_{j\gamma}^{(2)}) f_{j*,\gamma}^{(2)}(0)$. In precollision regions of the j, β phase space $S_{j\beta}^{(t)} = 1$ and hence

$$f_{j,\beta}^{(2)} = f_{j*}^{(1)} h_\beta + \int d\chi_\gamma (S_{j\beta\gamma}^{(t)} - S_{j\gamma}^{(t)}) f_{[j]}^{(1)} h_\beta h_\gamma + \cdots. \quad (39)$$

Now, even though the first term in this expansion is rigorously equal to $f_{j,\beta}^{(2)}$ the Choh–Uhlenbeck term and higher terms are nonzero. To be precise, the integrand of the Choh–Uhlenbeck term in (39) is nonzero in regions where the effect of $S_{j\beta\gamma}^{(t)}$ is a permutation of coordinates which places particle j on the β trajectory. The domain of g_γ for which this is the case grows linearly with t (i.e., the length of time interval between the initial time and time t). As a result this term is ill behaved in exactly the same sense as the usual divergent terms in the density expansion of Boltzmann's equation. The advantage of the hierarchy (34) to that of (37) is then obvious.

The divergent terms arise from the fact that the behavior of the whole system is analysed in terms of the dynamics of small numbers of isolated particles. It is quite straightforward to develop a series of approximations to the one-particle distribution function $f_j^{(1)}$, which takes into account the interaction between a small group of particles and the rest of the fluid.

Consider the system of equations (20). Using the notation $(f_j^{(1)})_k$ for the kth approximation to $f_j^{(1)}$, we can solve (20) along with the boundary conditions (10) by expressing $(f_j^{(1)})_k$ in terms of $(f_j^{(1)})_{k-1}$ by means of the following equations.

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial g}\right)(f_0^{(1)})_k = \gamma_0[(f_{-1}^{(1)})_{k-1} - (f_0^{(1)})_k] + \beta_0[(f_1^{(1)})_{k-1} - (f_0^{(1)})_k] + D[(f_{-1}^{(1)})_{k-1}, (f_0^{(1)})_k, (f_1^{(1)})_k]$$

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial g}\right)(f_j^{(1)})_k = \gamma_0[(f_{j-1}^{(1)})_k - (f_j^{(1)})_k] + \beta_0[(f_{j+1}^{(1)})_{k-1} - (f_j^{(1)})_k] + D[(f_{j-1}^{(1)})_k, (f_j^{(1)})_k, (f_{j+1}^{(1)})_{k-1}], \quad j > 0,$$

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial g}\right)(f_j^{(1)})_k = \gamma_0[(f_{j-1}^{(1)})_{k-1} - (f_j^{(1)})_k] + \beta_0[(f_{j+1}^{(1)})_k - (f_j^{(1)})_k] + D[(f_{j-1}^{(1)})_{k-1}, (f_j^{(1)})_k, (f_{j+1}^{(1)})_k], \quad j < 0, \quad (40)$$

where

$$D[f_{j-1}^{(1)}, f_j^{(1)}, f_{j+1}^{(1)}] = \lim_{\xi\to 0}[\delta(g - \xi - v't)\epsilon(g - vt)(v' - v) \times (f_{j+1}^{(1)} - f_j^{(1)}) + \delta(g + \xi - v't)\epsilon(vt - g)(v - v') \times (f_{j-1}^{(1)} - f_j^{(1)})]$$

and $(f_j^{(1)})_0 = 0$.

The kth approximation to $f_j^{(1)}$ is related to its exact value as given by (33) by expanding the modified Bessel function in a power series by means of the identity[8]

$$I_j(z) = (\tfrac{1}{2}z)^j \sum_{l=0}^{\infty} \frac{(\frac{1}{4}z^2)^l}{l!\,\Gamma(j + l + 1)}.$$

Hence

$$(f_j^{(1)})_k = [\delta(g - vt)\delta(v - v') + \rho h_0(v)\epsilon(t_0)\epsilon(t - t_0)]A_j^{(k)} + \rho h_0(v)[1 - \epsilon(t_0)\epsilon(t - t_0)][\epsilon(L - g + vt)\epsilon(g - vt)A_{j-1}^{(k)} + \epsilon(L + g - vt)\epsilon(vt - g)A_{j+1}^{(k)}], \quad (41)$$

where

$$A_j^{(k)} = \left(\frac{\mu}{\nu}\right)^{j/2}(\mu\nu)^{|j|/2}\sum_{l=0}^{k}(\mu\nu)^l[l!(|j| + l)!]^{-1} \times \exp[-\nu(t) - \mu(t)].$$

The retention of the term $\exp[-\nu(t) - \mu(t)]$ in its unexpanded form ensures that, to every order of approximation, $f_j^{(1)}$ is finite for all times and yields a finite approximation to the coefficient of self-diffusion which is the only important transport coefficient for this system. Each approximation approaches zero for large times. However, the limit of the sequence of approximations is correctly normalized for all times. This procedure therefore differs from that suggested, e.g., by Dorfman and his associates[9] and Kawasaki and Oppenheim.[10]

The success of this method of approximation in yielding finite transport coefficients gives some hope that a similar method of approximation could be devised for models with short-ranged forces of a more general nature, and in two or three dimensions. The important features are the correct formulation of the principle of molecular chaos and the adoption of a method of successive approximations which takes account of the scattering by particles of the environment of the particles involved in three-, four-, or many-body collisions.

The foregoing considerations do in fact suggest a perturbation technique for solving more complicated problems. As a simple example we consider the case of point particles which have a velocity independent probability R of being reflected upon collisions where $0 \leq R \leq 1$. For the case of impenetrable points $R = 1$ and for an ideal gas $R = 0$.

Although this system is not dynamically deterministic (except in the $R = 1$ or $R = 0$ cases), it well illustrates the techniques which can also be applied for deterministic systems. The hierarchy equations in this case are

$$\left(\frac{\partial}{\partial t} + \kappa^{(n)}\right)f_j^{(n)} = R\int_{-\infty}^{\infty} d\omega\,|\omega - v|\left(Pf_j^{(n+1)} - f_j^{(n+1)}\right) = R\int_{-\infty}^{\infty} d\omega\,|\omega - v|\left(f_{(j)}^{(n+1)} - f_j^{(n+1)}\right) + \xi R\int_{-\infty}^{\infty} d\omega\,|\omega - v|\left(Pf_j^{(n+1)} - f_{(j)}^{(n+1)}\right). \quad (42)$$

Here P is an operator which permutes the coordinates of the jth particle (i.e., g, v) with those of the unlabelled particles with which it is about to collide (i.e., g^-, ω).

The quantity $\xi = 1$ is a marking parameter which indicates that the term it multiplies is presumed to be small. (This term vanishes in both the $R = 0$ and $R = 1$ limits.) We assume a solution to (42) of the form

$$f_j^{(n)} = \sum_k \xi^k (f_j^{(n)})_k. \quad (43)$$

The ξ^0 equation is

$$\left(\frac{\partial}{\partial t} + \kappa^{(n)}\right)(f_j^{(n)})_0 = R\int_{-\infty}^{\infty} d\omega\,|\omega - v|[(f_{(j)}^{(n+1)})_0 - (f_j^{(n+1)})_0], \quad (44)$$

which can be solved for $n \geq 2$ in precollision regions as before to obtain

$$(f_j^{(n)})_0 = (f_{j*}^{(n-1)})_0 h, \quad (45)$$

where it is assumed that at $t = 0$ $(f_j^{(n)})_0 = f_j^{(n)}$. A closed equation for $(f_j^{(1)})_0$ can be obtained by substituting (45) for $n = 2$ into (44) for $n = 1$. This equation can be solved in exactly the manner previously discussed to obtain

$$(f_j^{(1)})_0 = \frac{1}{2\pi}\int_0^{2\pi} d\theta\; e^{-ij\theta}\eta(\theta, g, v, t), \quad (46)$$

where

$$\eta(\theta, g, v, t) = \{\delta(g - v't)\delta(v - v') + e^{i\theta}\rho h_0(v)\epsilon(g - vt)\exp[\rho\,|g - vt|(e^{i\theta} - 1)(1 - R)] + e^{-i\theta}\rho h_0(v)\epsilon(vt - g)\exp[\rho\,|vt - g|(e^{i\theta} - 1)(1 - R)]\} \times \{1 + \epsilon(t - t_0)\epsilon(t_0)(R - 1)\exp[i\theta\ \mathrm{sgn}(v - v')]\} \times \exp[(e^{i\theta} - 1)Rt\gamma(g/t) + (e^{-i\theta} - 1)Rt\beta(g/t)]. \quad (47)$$

Here

$$\beta(v) = \rho \int_v^\infty d\omega(\omega - v)h_0(\omega)$$

and

$$\gamma(v) = \rho \int_{-\infty}^v d\omega(v - \omega)h_0(\omega).$$

This result for η assumes a box of infinite length.

As is to be expected, (46) provides a rigorously correct expression for $f_j^{(1)}$ in the two limiting cases $R = 0$ and $R = 1$. Furthermore, the factors $\exp[\rho|g - vt|(e^{\pm i\theta} - 1)(1 - R)]$ assure that $f_j^{(1)})_0$ is exponentially damped in the long time limit if $R \neq 1$. This leads us to the conclusion that the higher order perturbation terms of (43), which from (42) obey equations of the form

$$\left(\frac{\partial}{\partial t} + \kappa^{(n)}\right)(f_j^{(n)})_k = R \int_{-\infty}^\infty d\omega|\omega - v|[(f_{(j)}^{(n+1)})_k - (f_j^{(n+1)})_k] + R \int_{-\infty}^\infty d\omega|\omega - v| \times [P(f_j^{(n+1)})_{k-1} - (f_{(j)}^{(n+1)})_{k-1}],$$

will not exhibit secular time behavior.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance of a grant from the Australian Research Grants Committee.

*Work done while an Australian Research Grants Committee Fellow on leave from Iowa State University, 1971-72 academic year.

[1]N. N. Bogoliubov, *Studies in Statistical Mechanics*, edited by J. deBoer and G. E. Uhlenbeck (Interscience, New York, 1962), Vol. 1, Chap. 1.

[2]H. S. Green, *Molecular Theory of Fluids* (North-Holland, Amsterdam, 1952).

[3]S. T. Choh and G. E. Uhlenbeck, thesis (University of Michigan, 1958).

[4]H. S. Green and D. Hoffman, J. Chem. Phys. **49,** 2600 (1968).

[5]D. W. Jepsen, J. Math. Phys. **6,** 405 (1965).

[6]J. L. Lebowitz and J. K. Percus, Phys. Rev. **155,** 122 (1967).

[7]L. Blum and J. L. Liebowitz, Phys. Rev. **185,** 273 (1969).

[8]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

[9]J. R. Dorfman and E. G. D. Cohen, J. Math. Phys. **8,** 282 (1967); M. H. Ernst, L. K. Haines, and J. R. Dorfman, Rev. Mod. Phys. **41,** 296 (1969).

[10]K. Kawasaki and I. Oppenheim, Phys. Rev. **139**,A1763 (1965).

Generalized isoperimetric inequalities. II*

J. M. Luttinger

Department of Physics, Columbia University, New York, New York 10027
(Received 19 April 1973)

Continuing the development of a previous paper on generalized isoperimetric inequalities (i.e., rearrangement inequalities for Green's functions), we extend the theory to the case of Green's functions for a potential which approaches zero at infinity. Specialization to domain potentials and long times gives Pólya and Szegö's isoperimetric inequality for the electrostatic capacity. Long times and a more general potential give a new isoperimetric inequality (for the "scattering length" of a potential). We also obtain from another specialization a curious isoperimetric inequality for the trace of the phase shift operator of scattering theory (for a given energy).

1. INTRODUCTION

In a previous paper,[1] a generalization of the usual isoperimetric inequalities of geometry and physics was given. The new inequalities may be expressed as inequalities for the Green's function of a particle in a potential φ (in quantum language), or that of diffusion in the presences of distributed sources and absorbers of particles (proportional to φ). To simplify the discussion in Paper I, we only treated the case where φ becomes infinitely large at infinity, so that the Hamiltonian has a purely discrete spectrum. This not only has the disadvantage of incompleteness, but also does not include (in a direct way) the important isoperimetric inequalities for the electrostatic capacity.[2] Now in I, the inequalities were based on a certain rearrangement inequality for multiple integrals[3] which, when translated into Green's function inequalities, is not sufficiently general to discuss all cases where φ approaches zero (say) at infinity. It proves not to be difficult to obtain another rearrangement inequality[4] which enables us to treat this case more generally. The resulting Green's function inequalities are analogous to those of I. When specialized for long "times", and φ such that the particle stays *outside* a certain domain, we find the usual isoperimetric inequalities for the capacity of that domain. If the potential is such that there are no bound states, the general inequality yields isoperimetric inequalities for the "scattering length" and for the "phase shift operator" which occur in the quantum theory of scattering. (These are new results.)

The paper is organized as follows. In Sec. 2, the basic inequalities are derived in one or more dimensions. In Sec. 3, various specializations are made and new isoperimetric inequalities are obtained.

2. GENERAL FORM OF INEQUALITIES

We first discuss one-dimensional problems. Consider a particle with the Hamiltonian operator ($\hbar = m = 1$)

$$H = \frac{1}{2}\frac{d^2}{dx^2} + \varphi(x). \tag{2.1}$$

In this paper we shall limit ourselves to the case where φ approaches zero as (x) approaches infinite. The Green's function $G_t(x, x'|\varphi)$ is defined by

$$HG_t(x, x'|\varphi) + \frac{\partial G_t(x, x'|\varphi)}{\partial t} = 0 \quad (t > 0) \tag{2.2}$$

with

$$\lim_{t\to 0} G_t(x, x'|\varphi) = \delta(x - x'). \tag{2.3}$$

We may write G as a Wiener integral[5]:

$$G_t(x, x'|\varphi) = \lim_{n\to 0} \int_{-\infty}^{\infty} dx_2 dx_3 \cdots dx_{n-1} P(x - x_2) e^{-\Delta_n \varphi(x_2)} \times P(x_2 - x_3) e^{-\Delta_n \varphi(x_3)} P(x_{n-1} - x') e^{-\Delta_n \varphi(x')}, \tag{2.4}$$

where

$$\Delta_n = t/n - 1, \tag{2.5}$$

$$P(x) \equiv \frac{1}{(2\pi\Delta_n)^{1/2}} e^{-x^2/2\Delta_n}. \tag{2.6}$$

It is necessary to distinguish three cases:

(a) $\varphi \leqslant 0$ (purely attractive potential),
(b) $\varphi \geqslant 0$ (purely repulsive potential),
(c) φ has both positive and negative regions.

Case (a): Since φ has a symmetrically increasing rearrangement (which we call $*[\varphi]$) reasoning identical to that of I yields the inequality[6]

$$\int_{-\infty}^{\infty} dx\, dx' G_t(x, x'|\varphi)\alpha(x')\beta(x)\Gamma(x' - x) \leqslant \int_{-\infty}^{\infty} dx\, dx' G_t(x, x'|*[\varphi])\,[\alpha(x)]^*[\beta(x)]^*[\Gamma(x' - x)]^*, \tag{2.7}$$

where α, β, Γ are nonnegative, and $[\alpha]^*$ is the symmetrically decreasing rearrangement of $\alpha(x)$. This only makes use of the rearrangement inequality (2.9) of I. It is possible to put (2.7) in another form which is often more convenient. The special case of (2.7) where α, β, Γ are already symmetrically decreasing functions may be written

$$\int_{-\infty}^{\infty} dx\, dx'[G_t(x, x'|0) - G_t(x, x'|\varphi)]\,\alpha(x')\beta(x)\Gamma(x' - x) \geqslant \int_{-\infty}^{\infty} dx\, dx'[G_t(x, x'|0) - G_t(x, x'|*[\varphi])] \times \alpha(x')\beta(x)\Gamma(x' - x), \tag{2.8}$$

where

$$G_t(x, x'|0) = \frac{1}{\sqrt{2\pi t}} \exp[-(x - x')^2/2t] \equiv G_t^{(0)}(x - x'). \tag{2.9}$$

This form is particularly interesting if we let α and β approach unity and Γ approach unity or a Dirac δ-function. These cases give the formulas

$$\int_{-\infty}^{\infty} dx\, dx'[G_t^{(0)}(x - x') - G_t(x, x'|\varphi)] \geqslant \int_{-\infty}^{\infty} dx\, dx'[G_t^{(0)}(x - x') - G_t(x, x'|*[\varphi])] \tag{2.10}$$

and

$$\int_{-\infty}^{\infty} dx[G_t^{(0)}(0) - G_t(x, x|\varphi)] \geqslant \int_{-\infty}^{\infty} dx[G_t^{(0)}(0) - G_t(x, x|*[\varphi])] \tag{2.11}$$

to which we shall return later.

Case (b): In this case φ does not have a symmetrically increasing rearrangement, and (2.9) of I becomes meaningless. However, the following inequality[4] may be proved. Let $F_i(x)$ satisfy $0 \leq F_i(x) \leq 1$ and let $H_i(x)$ be a symmetrically decreasing function of x. Then

$$\int_{-\infty}^{\infty} \left(1 - \prod_i^n [1 - F_i(x_i)]\right) \prod_i^n H_i(x_i - x_{i+1})\, dx_i \geq \int_{-\infty}^{\infty} \left(1 - \prod_i^n \{F_i(x_i)]^*\}\right) \prod_i^n H_i(x_i - x_{i+1})\, dx_i, \quad (2.12)$$

where $x_{n+1} \equiv x_1$.

Now suppose we define

$$1 - F_i(x) \equiv f_i(x). \quad (2.13)$$

Clearly,

$$1 - [F_i(x)]^* = {}^*[f_i(x)], \quad (2.14)$$

so that we may also write (2.12) as

$$\int_{-\infty}^{\infty} \left(\prod_i H_i(x_i - x_{i+1}) - \prod_i f_i(x_i) H_i(x_i - x_{i+1})\right) \prod_i dx_i \geq \int_{-\infty}^{\infty} \left(\prod_i H_i(x_i - x_{i+1}) - \prod_i {}^*[f_i(x_i)] H_i(x_i - x_{i+1})\right) \prod_i dx_i. \quad (2.15)$$

Applying (2.15) to (2.4), we obtain

$$\int_{-\infty}^{\infty} [G_t^0(x - x') - G_t(x, x' | \varphi)]\, \Gamma(x' - x)\, dx \geq \int_{-\infty}^{\infty} [G_t^0(x - x') - G_t(x, x' | {}^*[\varphi])]\, \Gamma(x' - x)\, dx, \quad (2.16)$$

where $\Gamma(x)$ is a symmetrically decreasing function of x, and we have made use of the fact that if

$$f(x) = e^{-\Delta_n(x)}, \quad (2.17)$$

then

$${}^*[f(x)] = e^{-\Delta_n[\varphi(x)]^*}. \quad (2.18)$$

If we again specialize $\Gamma(x)$ to be unity or a Dirac δ function, we obtain

$$\int_{-\infty}^{\infty} [G_t^{(0)}(x - x') - G_t(x, x' | \varphi)]\, dx\, dx' \geq \int_{-\infty}^{\infty} [G_t^{(0)}(x - x') - G_t(x, x' | [\varphi]^*)]\, dx\, dx' \quad (2.19)$$

and

$$\int_{-\infty}^{\infty} [G_t^{(0)}(0) - G_t(x, x | \varphi)]\, dx \geq \int_{-\infty}^{\infty} [G_t^{(0)}(0) - G_t(x, x | [\varphi]^*)]\, dx. \quad (2.20)$$

Case (c): In this case φ has neither a symmetrically increasing nor a symetrically decreasing rearrangement. We can, however, establish inequalities which go over smoothly into (2.10), (2.11), (2.19), and (2.20) for Case (a) or Case (b), as follows. Write

$$\varphi(x) = \varphi^+(x) + \varphi^-(x), \quad (2.21)$$

where

$\varphi^+(x) = \varphi(x)$ if $\varphi(x) \geq 0$ and $\varphi^-(x) = \varphi(x)$ if $\varphi(x) \leq 0$

$= 0$ if $\varphi(x) \leq 0$ $= 0$ if $\varphi(x) \geq 0$.

Now the following inequality holds:

$$G_t(x, x' | \varphi) + G_t(x, x' | 0) - G_t(x, x' | \varphi^+) - G_t(x, x' | \varphi^-) \leq 0. \quad (2.22)$$

The proof of (2.22) is trivial if we use Kac' representation[5] for the Green's function

$$G_t(x, x' | \varphi) = G_t^{(0)}(x - x') \left\langle \exp\left(-\int_0^t [x(\tau))d\tau \,\middle|\, \begin{matrix} x(t) = x \\ x(0) = x' \end{matrix}\right\rangle \quad (2.23)$$

The left hand side of (2.22) then becomes

$$G_t^{(0)}(x - x') \left\langle 1 - \exp\left(-\int_0^t \varphi^+[x(\tau)]d\tau\right) \times \left[1 - \exp\left(-\int_0^t \varphi^-[x(\tau)]d\tau\right)\right] \,\middle|\, \begin{matrix} x(t) = x \\ x(0) = x' \end{matrix}\right\rangle. \quad (2.24)$$

However, (2.24) is clearly nonpositive since the first factor we are averaging is nonnegative ($\varphi^+ \geq 0$), while the second factor is nonpositive ($\varphi^- \leq 0$).[7]

We may write (2.22) as

$$G_t^{(0)}(x - x') - G_t(x, x' | \varphi) \geq [G_t^{(0)}(x - x') - G_t(x, x' | \varphi^+)] + [G_t^{(0)}(x - x') - G_t(x, x' | \varphi^-)]. \quad (2.25)$$

Multiplying both sides of (2.25) with $\Gamma(x' - x)$ [$\Gamma(x)$ being a nonnegative symmetrically decreasing function) and making use of (2.8) and (2.16), we have

$$\int_{-\infty}^{\infty} [G_t^{(0)}(x - x') - G_t(x, x' | \varphi)]\, \Gamma(x' - x)\, dx\, dx' \geq \int_{-\infty}^{\infty} [G_t^{(0)}(x - x') - G_t(x, x' | [\varphi^+]^*)] \Gamma(x' - x)\, dx\, dx' + \int_{-\infty}^{\infty} [G_t^{(0)}(x - x') - G_t(x, x' | {}^*[\varphi^-])] \Gamma(x' - x)\, dx\, dx. \quad (2.26)$$

The result (2.26) is the most general rearrangement inequality (of the type considered here) for the Green's function, when the potential goes to zero at infinity.

In more than one dimension (d dimensions, say) we may proceed just as in I. That is, we use the d-dimensional analog of (2.4) and integrate over a single variable (say z) first, holding the others constant. This at once yields the analog of (2.26)

$$\int_{-\infty}^{\infty} [G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | \varphi)] \Gamma(\mathbf{r}' - \mathbf{r})\, d\mathbf{r}\, d\mathbf{r}' \geq \int_{-\infty}^{\infty} [G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | [\varphi^+]_z^*)] \Gamma(\mathbf{r}' - \mathbf{r})\, d\mathbf{r}\, d\mathbf{r}' + \int_{-\infty}^{\infty} [G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | {}_z^*[\varphi^-])] \Gamma(\mathbf{r}' - \mathbf{r})\, d\mathbf{r}\, d\mathbf{r}', \quad (2.27)$$

where

$$G_t^{(0)}(\mathbf{r} - \mathbf{r}') \equiv \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}'|^2}{2t}\right). \quad (2.28)$$

$[f(\mathbf{r})]_z^*$ is the symmetrically decreasing rearrangement of $f(\mathbf{r})$ viewed as a function of z with the other variables held constant; ${}_z^*[f(\mathbf{r})]$ is the symmetrically increasing rearrangement of $f(\mathbf{r})$ viewed as a function of z with the other variables held constant, and $\Gamma(\mathbf{r})$ is a symmetrically decreasing function of z for the other variables fixed. The integrations are over the entire d-dimensional Euclidean space.

The symmetrization process leading to (2.27) may be applied again to some other direction (rather than the z direction). By continuing this procedure with respect to "all possible directions" we will finally obtain (2.27) with $[\varphi^+]_z^*$ replaced by $[\varphi^+]_s^*$ (the decreasing rearrangement of φ^+ which is a function of $|\mathbf{r}|$ alone), ${}_z^*[\varphi^-]$ replaced by ${}_s^*[\varphi^-]$ (the increasing rearrangement of φ^+ which is a function of $|\mathbf{r}|$ alone), and $\Gamma(\mathbf{r})$ a decreasing function of $|\mathbf{r}|$ alone.

3. APPLICATIONS

We now consider applications of the inequalities of the last section. We shall restrict ourselves to three-dimensional considerations.

(A) Suppose we have a potential φ and a finite domain D such that

$$\varphi(\mathbf{r}) = \infty \quad \text{if } \mathbf{r} \text{ belongs to } D$$
$$= 0 \quad \text{otherwise.} \tag{3.1}$$

This is a purely repulsive potential, so that the inequality (2.27) applies with $\varphi^- = 0$. For this we need the symmetrically decreasing rearrangement of $\varphi(\mathbf{r})$. Just as in I, this is given by

$$[\varphi(\mathbf{r})]_z^* = \infty \quad \text{if } \mathbf{r} \text{ belongs to } D_z^*$$
$$= 0 \quad \text{otherwise,} \tag{3.2}$$

where D_z^* is the Steiner symmetrization of the domain D with respect to the plane $z = 0$. Using obvious notation, (2.27) becomes

$$\int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | D))\Gamma(\mathbf{r}' - \mathbf{r})\, d\mathbf{r}\, d\mathbf{r}'$$
$$\geq \int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | D_z^*)\Gamma(\mathbf{r}' - \mathbf{r})\, d\mathbf{r}\, d\mathbf{r}'. \tag{3.3}$$

For the moment, let us consider the case $\Gamma = 1$. Then (3.3) becomes

$$\int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | D))\, d\mathbf{r}\, d\mathbf{r}'$$
$$\geq \int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | D_z^*))\, d\mathbf{r}\, d\mathbf{r}'. \tag{3.4}$$

It is interesting to evaluate this formula in the limit of small and large t. In the limit of small t, a completely elementary calculation yields for the leading terms

$$\int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | D))\, d\mathbf{r}\, d\mathbf{r}'$$
$$= \Omega(D) + \sqrt{2/\pi}\ S(D)\sqrt{t} + \cdots, \tag{3.5}$$

where $\Omega(D)$ is the volume of the domain D, and $S(D)$ is the surface area of D. Therefore, since $\Omega(D) = \Omega(D_z^*)$, (3.4) reduces to

$$S(D) \geq S(D_z^*) \tag{3.6}$$

which is the usual result for the effect of Steiner symmetrization on the surface area of a domain.

In the limit of large t, the calculation is a little more difficult, though still straightforward. Fortunately, it has already been carried out by Spitzer.[8] The first two terms are given by

$$\int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | D))\, d\mathbf{r}\, d\mathbf{r}'$$
$$= 2\pi C(D)t + 4\sqrt{2\pi}C^2(D)\sqrt{t} + \cdots, \tag{3.7}$$

where $C(D)$ is the electrostatic capacity of the domain D. Therefore, for large t (3.4) becomes

$$C(D) \geq C(D_z^*) \tag{3.8}$$

which is Pólya and Szegö's well-known result[9] for the effect of Steiner symmetrization on the electrostatic capacity of a domain.

(B) Suppose we have a purely repulsive potential φ, and again consider (2.27) for $\Gamma = 1$. The case of small t and a sufficiently smooth potential is easily worked out; (2.27) reduces again to a trivial rearrangement inequality. For large t the calculation is again somewhat more involved. It can be shown[10] that

$$a(\varphi) = \lim_{t \to 0} \frac{1}{2\pi t} \int_{-\infty}^{\infty} (G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' | \varphi))\, d\mathbf{r}\, d\mathbf{r}', \tag{3.9}$$

where $a(\varphi)$ is the *scattering length* for the potential φ. (In particular this means that the scattering length for a potential of the form (3.1), is just given by the electrostatic capacity of the domain.) Therefore, (2.27) reduces to

$$a(\varphi) \geq a([\varphi]_z^*), \tag{3.10}$$

a new isoperimetric inequality.

(C) Now we consider (2.27) for a repulsive potential and $\Gamma(\mathbf{r}) = \delta(\mathbf{r})$, so (2.27) becomes

$$B(\varphi) \geq B([\varphi]_z^*), \tag{3.11}$$

where

$$B(\varphi) = \int_{-\infty}^{\infty} (G_t^{(0)}(0) - G_t(\mathbf{r}, \mathbf{r} | \varphi))\, d\mathbf{r}. \tag{3.12}$$

This type of object $[B(\varphi)]$ arises naturally in several problems of statistical mechanics. It was first studied by Beth and Uhlenbeck[14] (in connection with the quantum treatment of the second virial coefficient), who limited themselves to the case where $\varphi(\mathbf{r})$ is a function of $|\mathbf{r}|$ alone. They found that the result could be expressed as a sum of energy integrals over the scattering phase shifts for different angular momenta. The result for a general (i.e., nonspherically symmetric) potential may be written[12]

$$B = t \int_0^{\infty} d\epsilon e^{-\epsilon t} \operatorname{Tr}(\eta_\epsilon), \tag{3.13}$$

where η_ϵ is the phase shift operator of scattering theory (to be defined below) and Tr is the operation of trace (also to be defined below). Let the scattering operator[13] (for the potential φ) taken between momentum states $\mathbf{k}$ and $\mathbf{k}'$ be denoted by $(\mathbf{k}|S|\mathbf{k}')$. The scattering operator conserves energy, i.e., $(\mathbf{k}|S|\mathbf{k}')$ contains as a factor the energy conserving δ function $\delta(k^2/2 - k'^2/2)$. To take explicit advantage of this, let us describe the states $\mathbf{k}, \mathbf{k}'$ in spherical coordinates, k, θ, ϕ and k', θ', ϕ' and denote k by $\sqrt{2\epsilon}$, k' by $\sqrt{2\epsilon'}$. Then we may write

$$(\mathbf{k}|S|\mathbf{k}') = \frac{\delta(\epsilon - \epsilon')}{\sqrt{2\epsilon}}\ (\mu\phi\,|S_\epsilon|\,\mu'\phi'), \tag{3.14}$$

where $\mu = \cos\theta, \mu' = \cos\theta'$. Now the unitary property of S, i.e.,

$$\int d\mathbf{k}''(\mathbf{k}|S|\mathbf{k}'')\,\overline{(\mathbf{k}'|S|\mathbf{k}'')} = \delta(\mathbf{k} - \mathbf{k}') \tag{3.15}$$

(bar denoting complex conjugate), implies that S_ϵ is unitary in the sense

$$\int_{-1}^{1} d\mu'' \int_0^{2\pi} d\phi''(\mu\phi\,|S_\epsilon|\,\mu''\phi'')\,\overline{(\mu'\phi'|S_\epsilon|\,\mu''\phi'')}$$
$$= \delta(\mu - \mu')\delta(\phi - \phi'). \tag{3.16}$$

Since S_ϵ is unitary, it must have a representation of the form

$$S_\epsilon = e^{-2\pi i \eta_\epsilon}, \tag{3.17}$$

where η_ϵ is an Hermitian operator. η_ϵ as defined by (3.17) is the η_ϵ of (3.13) and the operation of Tr is simply

$$\mathrm{Tr}(\eta_\epsilon) \equiv \int_{-1}^{1} \int_{0}^{2\pi} d\mu d\phi (\mu\phi \,|\, \eta_\epsilon \,|\, \mu\phi) \tag{3.18}$$

since η_ϵ is hermitian, $\mathrm{Tr}(\eta_\epsilon)$ is real.

The inequality (3.11) with $B(\varphi)$ given by (3.13) is a quite general isoperimetric inequality for the trace of the phase shift operator. Nothing new comes from considering the simple cases of very small and very large t. That is, they give the same inequalities as Cases A and B in the respective limits.

ACKNOWLEDGMENTS

I would like to thank Professor R. Friedberg of Columbia University and Professor M. Kac of Rockefeller University for very helpful discussions.

*This work was supported in part by the National Science Foundation.

[1]J. M. Luttinger, J. Math. Phys. 14, 586 (1973). Henceforth we shall refer to this paper as I.

[2]Actually it *is* possible to transform the problem of determining the capacity into a problem of the type dealt with in I, but it seems to us somewhat artificial compared to the method used in this paper.

[3]J. M. Luttinger and R. Friedberg, "A New Rearrangement Inequality for Multiple Integrals," (unpublished).

[4]The proof is also due to R. Friedberg and myself, and will be published shortly. The theorem is a special case of a more general theorem which is also of use in discussing symmetrization procedures different from Steiner Symmetrization.

[5]M. Kac, *Probability and Related Topics in Physical Science, I* (Interscience, New York, 1959), p. 161 ff.

[6]It is assumed the α, β, Γ are such that the integrals on both sides exist.

[7]The inequality (2.22) is actually a special case of much more general "superadditive" inequalities for Green's functions, which have been noticed by R. Friedberg and myself. A paper on these general inequalities has been prepared, and is in the course of publication.

[8]F. Spitzer, Z. Wahrscheinlichkeitrechnung 3, 110 (1964).

[9]G. Pólya and G. Szegö,Am. J. Math. **67,** 1 (1945).

[10]M. Kac and J. M. Luttinger (unpublished). For a purely repulsive potential, the result has been established rigorously. For a general potential without bound states it is probably also true, but we have not as yet established it rigorously.

[11]E. Beth and G. E. Uhlenbeck, Physica (Utr.) **4,** 915 (1937).

[12]V. S. Buslaev, Dokl. Akad. Nauk SSSR **143,** 1067 (1962) [Sov. Phys.-Dokl. **7,** 295 (1962)].

[13]See, for example, M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964), p. 287.

Generalized isoperimetric inequalities. III*

J. M. Luttinger

Department of Physics, Columbia University, New York, New York 10027
(Received 16 May 1973)

The generalized isoperimetric inequalities for rearranged Green's functions, which have previously been discussed for a rearrangement process analogous to Steiner symmetrization, are obtained for a type of rearrangement analogous to circular symmetrization. It is also shown how rearrangements analogous to spherical and Schwarz symmetrization may be defined and considered as limiting cases of previously discussed rearrangements.

1. INTRODUCTION

In previous papers[1] we have found and investigated a class of "rearrangement" inequalities for certain Green's functions (which could be regarded as the Green's functions for a quantum mechanical particle in a potential, or for the diffusion of particles in the presence of absorbers and creators of particles). These were called 'generalized isoperimetric inequalities' because they reduced in very special cases to the type of isoperimetric inequalities long known in mathematics and physics.[2] These latter inequalities are all of the type which involve the "Steiner symmetrization" of a domain. Now other types of symmetrization procedures which also give rise to isoperimetric inequalities are known.[3] The question arises as to whether these other symmetrization procedures have analogous generalizations to Green's function inequalities, similar to the ones established in Papers I and II.

In this short note we shall deal in detail with only one important case, the analogue of "circular" symmetrization.[4] We shall show that this does have a very simple generalization to Green's function inequalities. Just as in the previous work, the results are an almost trivial consequence of another new rearrangement inequality for multiple integrals. This inequality—different from the ones used in I and II—will be used here, but the proof (due to R. Friedberg and myself) will be published elsewhere. (This is partly due to the complexity of the proof, but also because we believe that the inequality may be of some interest in other contexts.)

2. NEW GREEN'S FUNCTION INEQUALITIES

We begin our discussion with two-dimensional problems, the type of symmetrization we are discussing not being present in one dimension. The generalization to higher dimensions will be trivial.) Consider a particle with the Hamiltonian operator

$$H = -\frac{1}{2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \varphi(x,y). \tag{2.1}$$

The Green's function $G_t(\mathbf{r},\mathbf{r}'\,|\,\varphi)$ is defined by

$$HG_t(\mathbf{r},\mathbf{r}'\,|\,\varphi) + \frac{\partial G_t(\mathbf{r},\mathbf{r}'\,|\,\varphi)}{\partial t} = 0 \quad (t>0) \tag{2.2}$$

with

$$\lim_{t\to 0} G_t(\mathbf{r},\mathbf{r}'\,|\,\varphi) = \delta(\mathbf{r}-\mathbf{r}'). \tag{2.3}$$

$\mathbf{r}$ stands for the vector (x,y) and $\delta(\mathbf{r})$ is the Dirac δ function in two-dimensional space.

Let us represent G_t as a Wiener integral[5]

$$G_t(\mathbf{r},\mathbf{r}'\,|\,\varphi) = \lim_{n\to\infty} \int d\mathbf{r}_2 d\mathbf{r}_3 \cdots d\mathbf{r}_n P(\mathbf{r}-\mathbf{r}_2) \times e^{-\Delta_n \varphi(\mathbf{r}_2)} P(\mathbf{r}_2-\mathbf{r}_3) e^{-\Delta_n\varphi(\mathbf{r}_3)} \cdots P(\mathbf{r}_{n-1}-\mathbf{r}'), \tag{2.4}$$

where

$$\Delta_n = t/n \tag{2.5}$$

and

$$P(\mathbf{r}) = \frac{1}{2\pi\Delta_n}\exp\left(-\frac{x^2+y^2}{2\Delta_n}\right). \tag{2.6}$$

(The integrals are over the entire space, unless otherwise indicated.)

Now consider the quantity $I(\varphi)$ defined by $(\alpha,\beta,\Gamma \geqslant 0)$

$$I(\varphi) = \int d\mathbf{r}d\mathbf{r}' G_t(\mathbf{r},\mathbf{r}'\,|\,\varphi)\alpha(\mathbf{r})\beta(\mathbf{r}')\Gamma(\mathbf{r}',\mathbf{r}). \tag{2.7}$$

Using (2.4) and rearranging the notation slightly, we have

$$I(\varphi) = \lim_{n\to\infty} \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_n\, \alpha(\mathbf{r}_1) P(\mathbf{r}_1-\mathbf{r}_2) e^{-\Delta_n\varphi(\mathbf{r}_2)} \times P(\mathbf{r}_2-\mathbf{r}_3) e^{-\Delta_n\varphi(\mathbf{r}_3)} \cdots P(\mathbf{r}_{n-1}-\mathbf{r}_n)\beta(\mathbf{r}_n)\Gamma(\mathbf{r}_n,\mathbf{r}_1). \tag{2.8}$$

Let us introduce polar coordinates for the $\mathbf{r}_i$ integrations $d\mathbf{r}_i = r_i dr_i d\phi_i (-\pi \leqslant \phi_i \leqslant \pi)$, and do the ϕ_i integrations first. Note that

$$P(\mathbf{r}_i - \mathbf{r}_{i+1}) = \frac{1}{2\pi\Delta_n}\exp\left(-\frac{r_i^2 + r_{i+1}^2 - 2r_i r_{i+1}\cos(\phi_i-\phi_{i+1})}{2\Delta_n}\right) \tag{2.9}$$

is a monotonically nondecreasing function of $\cos(\phi_i - \phi_{i+1})$. We assume also that as far as the ϕ, ϕ' dependence of $\Gamma(\mathbf{r},\mathbf{r}')$ goes, it is a monotonically nondecreasing function of $\cos(\phi-\phi')$. [This is achieved most easily by choosing $\Gamma(\mathbf{r},\mathbf{r}')$ to be a monotonically nonincreasing function of $|\mathbf{r}-\mathbf{r}'|$.] We now make use of the following rearrangement inequality for multiple integrals.[6]

Let $f_i(\phi)(1 \leqslant i \leqslant n)$ be a periodic function of ϕ with period 2π. Let $[f_i(\phi)]^*$ be the symmetrically decreasing rearrangement of f_i in $(-\pi,\pi)$ about the origin $\phi = 0$. Finally $h_i(a)(-1 \leqslant a \leqslant 1)$ is a monotonically nondecreasing function of a. Then

$$\int_{-\pi}^{\pi} \prod_{i=1}^{n} [f_i(\phi_i) h_i(\cos(\phi_i - \phi_{i+1}))d\phi_i] \leqslant \int_{-\pi}^{\pi} \prod_{i=1}^{n} \left[[f_i(\phi_i)]^* h_i(\cos(\phi_i-\phi_{i+1}))d\phi_i\right], \tag{2.10}$$

where $\phi_{n+1} \equiv \phi_1$.

Applying (2.10) to the ϕ_i integrations in (2.8), we obtain at once the inequality

$$\int d\mathbf{r}d\mathbf{r}' G_t(\mathbf{r},\mathbf{r}'\,|\,\varphi)\alpha(\mathbf{r})\beta(\mathbf{r}')\Gamma(\mathbf{r}',\mathbf{r}) \leqslant \int d\mathbf{r}d\mathbf{r}' G_t(\mathbf{r},\mathbf{r}'\,|\,{}^*_\phi[\varphi])[\alpha(\mathbf{r})]^*_\phi[\beta(\mathbf{r}')]^*_\phi(\mathbf{r}',\mathbf{r}), \tag{2.11}$$

where ${}^*_\phi[\varphi(\mathbf{r})]$ is the symmetrically increasing rearrangement of $\varphi(\mathbf{r})$ (viewed as a function of ϕ for

fixed r) in $(-\pi, \pi)$ about the origin $\phi = 0$, $[\alpha(\mathbf{r})]^*_\varphi$ the analogous symmetrically decreasing rearrangement of $\alpha(\mathbf{r})$ {We note in passing that $[\alpha(r,\theta,\phi)]^*_\phi = {}^*_\varphi[\alpha(r,\theta,\pi-\phi)]$.}

Similarly, we obtain[7]

$$\int \prod_{k=1}^{m} [G_{t_k}(\mathbf{r}_k, \mathbf{r}_{k'} \mid \varphi)\alpha_k(\mathbf{r}_k)\beta_k(\mathbf{r}'_k)\Gamma_k(\mathbf{r}'_k, \mathbf{r}_k)d\mathbf{r}_k d\mathbf{r}'_k] \leq \int \prod_{k=1}^{m} [G_{t_k}(\mathbf{r}_k, \mathbf{r}'_k \mid {}^*_\varphi[\varphi])[\alpha_k(\mathbf{r}_k)]^*_\varphi[\beta_k(\mathbf{r}_k)]^*_\varphi \times \Gamma_k(\mathbf{r}'_k, \mathbf{r}_k)d\mathbf{r}_k d\mathbf{r}'_k], \tag{2.12}$$

where $\mathbf{r}_m = \mathbf{r}_1$, $\alpha_k, \beta_k, \Gamma_k$ are nonnegative, $\Gamma_k(\mathbf{r}'_k, \mathbf{r}_k)$ depends on ϕ_k, ϕ'_k only through $\cos(\phi_k - \phi'_k)$ and is a nondecreasing function of $\cos(\phi_k - \phi'_k)$ for r_k, r'_k held fixed.

Again, various specializations of (2.11) are of particular interest.

A. ϕ approaches infinity at infinity.

Specialization of α, β, Γ yields

$$\int d\mathbf{r}d\mathbf{r}' G_t(\mathbf{r}, \mathbf{r}' \mid \varphi) \leq \int d\mathbf{r}d\mathbf{r}' G_t(\mathbf{r}, \mathbf{r}' \mid {}^*_\varphi[\varphi]), \tag{2.13}$$

$$\int d\mathbf{r} G_t(\mathbf{r}, \mathbf{r} \mid \varphi) \leq \int d\mathbf{r} G_t(\mathbf{r}, \mathbf{r} \mid {}^*_\varphi[\varphi]). \tag{2.14}$$

(2.14) is an inequality for the partition function (Cf. I). For large t, it yields the result that the smallest characteristic value of the Hamiltonian of a particle in a potential φ is greater than or equal to the smallest characteristic value of a particle in the potential ${}^*_\varphi[\varphi]$. For when t is small, the limiting result depends on the nature of φ. For a sufficiently smooth potential (2.14) reduces, after a little manipulation [Cf. I, Eq. (3.25)ff] to a special case of (2.10) for $n = 2$. On the other hand, suppose φ is a "domain potential," i.e.,

$$\varphi(\mathbf{r}) = \begin{cases} 0 & \mathbf{r} \text{ inside a domain } D \\ \infty & \mathbf{r} \text{ outside } D. \end{cases} \tag{2.15}$$

Then from the definition we have at once that

$${}^*_\varphi[\varphi(\mathbf{r})] = \begin{cases} 0 & \mathbf{r} \text{ inside } D^*_\varphi \\ \infty & \mathbf{r} \text{ outside } D^*_\varphi, \end{cases} \tag{2.16}$$

where D^*_φ is just the circular symmetrization of the domain D.[4] For small t, (2.14) yields in this case the usual result that circular symmetrization does not increase the length of the boundary of D. [Large t for the "domain" case yields the result that circular symmetrization does not increase the principal frequency of a membrane, while integration of (2.13) over t gives the result that the torsional rigidity of a domain is not decreased by circular symmetrization.]

B. φ approaches zero at infinity

In this case (as in II) we take α, β to be symmetrically decreasing functions of ϕ in $(-\pi, \pi)$ around the origin. Then we may write (2.11) as

$$\int d\mathbf{r}d\mathbf{r}'[G_t(\mathbf{r}, \mathbf{r}' \mid 0) - G_t(\mathbf{r}, \mathbf{r}' \mid \varphi)\alpha(\mathbf{r})\beta(\mathbf{r}')\Gamma(\mathbf{r}', \mathbf{r}) \geq \int d\mathbf{r}d\mathbf{r}'[G_t(\mathbf{r}, \mathbf{r}' \mid 0) - G_t(\mathbf{r}, \mathbf{r}' \mid {}^*_\varphi[\varphi])]\alpha(\mathbf{r})\beta(\mathbf{r}')\Gamma(\mathbf{r}', \mathbf{r}), \tag{2.17}$$

where

$$G_t(\mathbf{r}, \mathbf{r}' \mid 0) \equiv G_t^{(0)}(\mathbf{r} - \mathbf{r}') = \frac{1}{2\pi t} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}'|^2}{2t}\right). \tag{2.18}$$

Again, two limiting cases are of interest:

$$\int d\mathbf{r}d\mathbf{r}'[G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' \mid \varphi)] \geq \int d\mathbf{r}d\mathbf{r}'[G_t^{(0)}(\mathbf{r} - \mathbf{r}') - G_t(\mathbf{r}, \mathbf{r}' \mid {}^*_\varphi[\varphi])] \tag{2.19}$$

and

$$\int d\mathbf{r}[G_t^{(0)}(0) - G_t(\mathbf{r}, \mathbf{r} \mid \varphi)] \geq \int d\mathbf{r}[G_t^{(0)}(0) - G_t(\mathbf{r}, \mathbf{r} \mid {}^*_\varphi[\varphi])]. \tag{2.20}$$

Before we discuss limiting cases of (2.19) and (2.20), let us consider the generalization of all these formulas to $p + 2$ dimensions. Take any two-dimensional planar subspace and call rectangular coordinates on it x and y; call the remaining rectangular coordinates $(x_1, x_2, \cdots, x_p) \equiv \mathbf{p}$. [A point $\mathbf{r}$ has the coordinates $(x, y, x_1, x_2 \ldots, x_p)$.] Now introduce polar coordinates in the (x, y) plane, $x = a\cos\phi, y = a\sin\phi$. The volume element becomes

$$d\mathbf{r} = ad\mathbf{p}da d\phi \tag{2.21}$$

and the distance between two points $\mathbf{r}$ and $\mathbf{r}'$ is

$$|\mathbf{r} - \mathbf{r}'|^2 = |\mathbf{p} - \mathbf{p}'|^2 + a^2 + a'^2 - 2aa'\cos(\phi - \phi'). \tag{2.22}$$

The entire analysis is identical to the two-dimensional case, except that $P(\mathbf{r})$ is replaced by

$$P(\mathbf{r}) = \frac{1}{(2\pi\Delta_n)^{(p+2)/2}} \exp\left(-\frac{r^2}{2\Delta_n}\right) \tag{2.23}$$

and $G_t^{(0)}(\mathbf{r} - \mathbf{r}')$ by

$$G_t^{(0)}(r - r') = \frac{1}{(2\pi t)^{(p+2)/2}} \exp\left(\frac{|\mathbf{r} - \mathbf{r}'|^2}{2t}\right). \tag{2.24}$$

The important point is that because of (2.21) and (2.22) the ϕ integrals are of the same form as in the two-dimensional case, and (2.10) applies. Therefore, all the inequalities that we had for the two-dimensional case remain valid in an arbitrary number of dimensions (greater than two) if we interpret ${}^*_\varphi[\varphi]$ as the symmetrically increasing rearrangement of φ viewed as a function of $\phi, \mathbf{p}$ and a being held fixed.

Just as in II, the most immediately interesting consequences of (2.19) and (2.20) come from the large t limit. If we call the scattering length for a particle in the potential φ (with no bound states) $a(\varphi)$, then (2.19) reduces to (Cf. II, Eq. (3.9)]

$$a(\varphi) \geq a({}^*_\varphi[\varphi]). \tag{2.25}$$

if φ becomes a domain potential of the form

$$\varphi(\mathbf{r}) = \begin{cases} \infty & \mathbf{r} \text{ belongs to a domain } D \\ 0 & \mathbf{r} \text{ not in } D, \end{cases} \tag{2.26}$$

then one sees immediately that

$${}^*_\varphi[\varphi(\mathbf{r})] = \begin{cases} \infty & \mathbf{r} \text{ belongs to } D^*_\varphi, \\ 0 & \mathbf{r} \text{ not in } D^*_\varphi, \end{cases} \tag{2.27}$$

where D^*_φ is again the circular symmetrization of the domain D. In this case, as in II, the scattering length becomes the electrostatic capacity $C(D)$ of the domain D (in three dimensions) and we obtain the result

$$C(D) \geq C(D^*_\phi). \tag{2.28}$$

So far we have only considered the process of circular symmetrization for one fixed coordinate system. We can now continue the process (in more than two dimensions) by considering symmetrization with respect to a new coordinate system related to the old by an orthogonal transformation of coordinate axes. Suppose we continue this process by symmetrizing with respect to "all possible" rotated sets of coordinate axes. We shall assume that for a reasonable potential this process converges to a unique limit. Consider the three-dimensional case (for simplicity only). Then $\varphi(\mathbf{r}) = \varphi(r,\theta,\phi)$, $a(\mathbf{r}) = \alpha(r,\theta,\phi)$, etc., are rearranged by the process only on the variables θ, ϕ. We would expect on intuitive grounds (though we have not proved it) that just as repeated Steiner symmetrization with respect to different axes of a function of x and y leads to a function of $x^2 + y^2$ alone, that this process would lead to a dependence on θ alone. Define ${}^*_\theta[\varphi]$ to be the (unique) rearrangement of $\varphi(r,\theta,\phi)$ (r held fixed) which is a symmetrically increasing function of θ alone, and define $[\alpha]^*_\theta$ to be the (unique) rearrangement of $\alpha(r,\theta,\phi)$ (r held fixed) which is a symmetrically decreasing function of θ alone. That is,

$$\int_0^\pi\int_0^{2\pi} \Theta(\alpha(r,\theta,\phi) - g)\,\sin\theta d\theta d\phi = \int_0^\pi\int_0^{2\pi} \Theta([\alpha(r,\theta)]^*_\theta - g)\,\sin\theta d\theta d\phi \qquad (2.29)$$

and

$$\int_0^\pi\int_0^{2\pi} \Theta(g - \varphi(r,\theta,\phi))\,\sin\theta d\theta d\phi = \int_0^\pi\int_0^{2\pi} \Theta\,(g - \theta^*[\varphi(r,\theta)])\,\sin\theta d\theta d\phi, \qquad (2.30)$$

where g is a real number and Θ is the step-function

$$\Theta(g) = \begin{cases} 1 & g \geqslant 0 \\ 0 & g < 0 \end{cases}. \qquad (2.31)$$

{We mention in passing that

$${}^*_\theta[\alpha(r,\theta)] = [\alpha(r,\pi - 0)]^*_\theta.\}$$

From what has been said above, it is clear that we would then expect results like (2.11) to still be valid if the symmetrization processes ${}^*_\phi[\varphi], [\alpha]^*_\phi$, etc., be replaced by ${}^*_\theta[\varphi], [\alpha]^*_\theta$, etc. On the other hand, if φ is a domain potential such as is given by (2.15), we see at once that

$${}^*_\theta[\varphi(r)] = \begin{cases} 0 & \mathbf{r} \text{ inside } D^*_\theta \\ \infty & \mathbf{r} \text{ outside } D^*_\theta \end{cases}, \qquad (2.32)$$

where D^*_θ is the "spherical" symmetrization[8] of the domain D. Thus, in the nonrigorous sense of this discussion we have shown that spherical symmetrization has an analog for Green's function rearrangement inequalities.

Finally, we mention that in the same spirit as the above discussion another important symmetrization procedure, namely "Schwarz" symmetrization,[9] may be considered as a limiting case of the Steiner type of symmetrization used in I and II. That is, if we consider the symmetrization process of I ${}^*_x[\varphi(x,y,z)]$ repeated with respect to "all possible" coordinate systems rotated from the original one around the z axis by different amounts, we would expect ultimately to end up with a rearranged function which only depends on $p = (x^2 + y^2)^{1/2}$. Call this ${}^*_p[\varphi(p,z)]$. Again, when φ is specialized to a domain potential like (2.15), we find that ${}^*_p[\varphi]$ is reduced to the domain potential of a domain D^*_p which is just the Schwarz symmetrization of D.

ACKNOWLEDGMENT

I would like to thank Professor R. Friedberg of Columbia University for many very helpful discussions.

*This research was supported in part by the National Science Foundation.

[1] J. M. Luttinger, J. Math. Phys. **14**, 586, 1444 (1973). We shall refer to these papers as I and II, respectively.

[2] See, for example, G. Pólya and G. Szegö, *Isoperimetric Inequalities in Mathematical Physics* (Princeton U.P., Princeton, N. J., 1951).

[3] A paper including many modern references is L. E. Payne, SIAM Rev. (Soc. Ind. Appl. Math.) **9**, 453 (1967). See also Ref. 2.

[4] G. Pólya, C.R. Acad. Sci. (Paris) **230**, 25 (1950); Ref. 2, p. 193 ff.

[5] M. Kac, *Probability and Related Topics in Physical Science, I* (Interscience, New York, 1959), p. 161 ff.

[6] J. M. Luttinger and R. Friedberg (unpublished).

[7] Inequalities of the type given by (2.12) also hold for the rearranged Green's functions of I, but we did not write them out explicitly in that paper.

[8] See Ref. 2, p. 205 ff.

[9] See Ref. 2, p. 189 ff.

A condition for the applicability of Biedenharn's lemma*

Robert E. Beck

Department of Mathematics, Villanova University, Villanova, Pa. 19085
(Received 24 January 1973)

Biedenharn's lemma is used to compute the outer multiplicities of the irreducible components of the tensor product of two irreducible representations of a complex simple Lie algebra L with highest weights λ and λ'. The lemma is applicable if for all weights μ of λ, we have that $\lambda' + \mu$ is a dominant integral form. This paper deduces a condition written in terms of the root system of L and the coordinates of λ and λ' which is equivalent to $\lambda' + \mu$ being a dominant integral form for all weights μ of λ.

1. INTRODUCTION

One tool which is available to compute the outer multiplicities of the tensor product of two irreducible representations of a complex simple Lie algebra is the Biedenharn lemma.[1] When this lemma is applicable, it says the outer multiplicities are equal to the respective inner multiplicities. To state the lemma precisely we introduce the following notation. Let L be a complex simple Lie algebra, and let ρ and ρ be irreducible representations of L with highest weights λ and λ', respectively.

Definition: The highest weight λ' dominates λ (written $\lambda' \gg \lambda$) if for all weights $\mu \in \Lambda(\lambda)$, the weight system of $\lambda, \lambda' + \mu$ is a dominant integral form.

Lemma (Biedenharn): If $\lambda' \gg \lambda$, then the outer multiplicity of the representation with highest weight $\lambda' + \mu$, $m_{\lambda,\lambda'}^{\lambda'+\mu} = n_{\mu}^{\lambda}$ where μ is a weight of ρ and n_{μ}^{λ} is the inner multipliticity of μ in ρ. Furthermore, any irreducible component of $\rho' \otimes \rho$ has a highest weight of the form $\lambda' + \mu$.

In this paper we develop an equivalent condition to $\lambda' \gg \lambda$ which is computationally oriented. This result extends those given by MacFarlane, O. Raifeartaigh, and Rao.[2] Furthermore, our method shows the underlying reasons for their results and does not necessitate case-by-case checking.

For a more thorough discussion of representation theory and facts from Lie algebras, the reader is referred to one of the recent standard books.[3] The conventions and notations which we use are given below:

$\alpha_1, \ldots, \alpha_l$	are the simple roots of the Lie algebra,
$(,)$	denotes the inner product induced by the Killing form,
S_β	denotes the Weyl reflection defined by the root β,
$\lambda_1, \ldots, \lambda_l$	are the fundamental weights,
$\Lambda(\lambda)$	denotes the weight system of a representation with highest weight λ,
$\mu = \sum_{i=1}^{l} m_i \lambda_i$	is a dominant integral form if for each i, m_i is a nonnegative integer.

2. FACTS ABOUT ROOTS AND WEIGHTS

For $j = 1, \ldots, l$ let γ_j be the highest root of L such that $(\gamma_j, \gamma_j) = (\alpha_j, \alpha_j)$ and write

$$\gamma_j = \sum_{i=1}^{l} c_i^{(j)} \alpha_i .$$

The $c_i^{(j)}$, $1 \le j \le l$, are positive integers.

Lemma 1: For each j, $1 \le j \le l$, and for every positive root β, we have $(\beta, \gamma_j) \ge 0$.

Proof: Since each element of the Weyl group preserves the inner product, we have $(S_\beta \gamma_j, S_\beta \gamma_j) = (\gamma_j, \gamma_j)$. Also $S_\beta \gamma_j$ is a root so that

$$S_\beta(\gamma_j) = \gamma_j - \frac{2(\beta, \gamma_j)}{(\beta, \beta)} \beta \le \gamma_j$$

by the maximality of γ_j.

Consequently, $(\beta, \gamma_j) \ge 0$.

Lemma 2: For any j, $1 \le j \le l$ and for any highest weight λ we have $(\gamma_j, \mu) \le (\gamma_j, \lambda)$ for each $\mu \in \Lambda(\lambda)$.

Proof: We may write $\mu = \lambda - \sum_{\beta>0} c_\beta \beta$, where c_β is a nonnegative integer and β is a positive root.[4] Hence $(\gamma_j, \mu) = (\gamma_j, \lambda) - \sum_{\beta>0} c_\beta (\gamma_j, \beta)$ and the result follows by using Lemma 1.

Lemma 3: Let $\mu = \sum_{i=1}^{l} p_i \lambda_i$. For any j, $1 \le j \le l$, we have

$$\frac{2(\gamma_j, \mu)}{(\gamma_j, \gamma_j)} = \sum_{i=1}^{l} c_i^{(j)} p_i \frac{(\alpha_i, \alpha_i)}{(\gamma_j, \gamma_j)} .$$

Proof: By definition, $2(\alpha_i, \lambda_k)/(\alpha_i, \alpha_i) = \delta_{ik}$, the Kroneker delta.

Consequently,

$$\frac{2(\gamma_j, \mu)}{(\gamma_j, \gamma_j)} = \sum_{i,k=1}^{l} c_i^{(j)} p_i \frac{2(\alpha_i, \lambda_k)}{(\alpha_i, \alpha_i)} \frac{(\alpha_i, \alpha_i)}{(\gamma_j, \gamma_j)} = \sum_{i=1}^{l} c_i^{(j)} p_i \frac{(\alpha_i, \alpha_i)}{(\gamma_j, \gamma_j)} .$$

3. THE EQUIVALENT CONDITION

Let

$$\lambda = \sum_{j=1}^{l} m_j \lambda_j \quad \text{and} \quad \lambda' = \sum_{j=1}^{l} m_j' \lambda_j .$$

Theorem: $\lambda' \gg \lambda$ if and only if for each j,

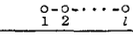
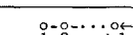
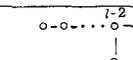
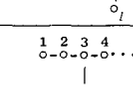
$$m_j' \ge \frac{2(\gamma_j, \lambda)}{(\gamma_j, \gamma_j)} = \sum_{i=1}^{l} c_i^{(j)} m_i \frac{(\alpha_i, \alpha_i)}{(\gamma_j, \gamma_j)} .$$

The computational condition for each type of complex simple Lie algebra is summarized in Table I.

Proof of theorem: Let $\mu \in \Lambda(\lambda)$ and write $\mu = \sum_{j=1}^{l} p_j \lambda_j$.

We must show that for all $\mu \in \Lambda(\lambda)$, for all j, $1 \le j \le l$ we have $m_j' + p_j \ge 0$ or equivalently $m_j' \ge -p_j$.

TABLE I. Equivalent condition in computational form.

Type	Compact form	Dynkin diagram	γ_j	Condition
A_l	$SU(l+1)$		$\alpha_1 + \cdots + \alpha_l$	$m'_j \geq m_1 + \cdots + m_l$
B_l	$SO(2l+1)$		$\alpha_1 + 2(\alpha_2 + \cdots + \alpha_l), j = 1, \ldots, l-1$ $\alpha_1 + \cdots + \alpha_l, j = l$	$m'_j \geq m_1 + 2(m_2 + \cdots + m_{l-1}) + m_l$ $m'_j \geq 2(m_1 + \cdots + m_{l-1}) + m_l$
C_l	$Sp(2l)$		$\alpha_1 + 2(\alpha_2 + \cdots + \alpha_{l-1}) + \alpha_l, j = 1, \ldots, l-1$ $2(\alpha_1 + \cdots + \alpha_{l-1}) + \alpha_l, j = l$	$m'_j \geq m_1 + 2(m_2 + \cdots m_l)$ $m'_j \geq m_1 + \cdots + m_l$
D_l	$SO(2l)$		$\alpha_1 + 2(\alpha_2 + \cdots + \alpha_{l-2}) + \alpha_{l-1} + \alpha_l$	$m'_j \geq m_1 + 2(m_2 + \cdots + m_{l-2}) + m_{l-1} + m_l$
E_l			$l = 6, \alpha_1 + 2\alpha_3 + 3\alpha_3 + 2\alpha_4 + \alpha_5 + 2\alpha_6$ $l = 7, 2\alpha_1 + 3\alpha_2 + 4\alpha_3 + 3\alpha_4 + 2\alpha_5 + \alpha_6 + 2\alpha_7$ $l = 8, 2\alpha_1 + 4\alpha_2 + 6\alpha_3 + 5\alpha_4 + 4\alpha_5 + 3\alpha_6 + 2\alpha_7 + 3\alpha_8$	$l = 6, m'_j \geq m_1 + 2m_2 + 3m_3 + 2m_4 + m_5 + 2m_6$ $l = 7, m'_j \geq 2m_1 + 3m_2 + 4m_3 + 3m_4 + 2m_5 + m_6 + 2m_7$ $l = 8, m'_j \geq 2m_1 + 4m_2 + 6m_3 + 5m_4 + 4m_5 + 3m_6 + 2m_7 + 3m_8$
F_4			$2\alpha_1 + 3\alpha_2 + 4\alpha_3 + 2\alpha_4, j = 1, 2$ $\alpha_1 + 2\alpha_2 + 3\alpha_3 + 2\alpha_1, j = 3, 4$	$m'_j \geq 2m_1 + 3m_2 + 2m_3 + m_4$ $m'_j \geq 2m_1 + 4m_2 + 3m_3 + 2m_4$
G_2			$2\alpha_1 + 3\alpha_2, j = 1$ $\alpha_1 + 2\alpha_2, j = 2$	$m'_1 \geq 2m_1 + m_2$ $m'_2 \geq 3m_1 + 2m_2$

Now $p_j = 2(\alpha_j, \mu)/(\alpha_j, \alpha_j)$. Also $\Lambda(\lambda)$ is invariant under the action of the Weyl group. Furthermore, since $(\gamma_j, \gamma_i) = (\alpha_i, \alpha_i)$, there is an element of the Weyl group, w, such that $w(-\alpha_j) = \gamma_j$.

Using these facts we may write for each j

$$\{-p_j \mid \mu \in \Lambda(\lambda)\} = \left\{ \frac{2(-\alpha_j, \mu)}{(\alpha_j, \alpha_j)} \,\middle|\, \mu \in \Lambda(\lambda) \right\} = \left\{ \frac{2(\gamma_j, \mu)}{(\gamma_j, \gamma_j)} \,\middle|\, \mu \in \Lambda(\lambda) \right\}$$

Finally, for each j, $m'_j \geq -p_j$ for all $\mu \in \Lambda(\lambda)$ is equivalent to

$$m'_j \geq \max_{\mu \in \Lambda(\lambda)} \{-p_j\} = \max_{\mu \in \Lambda(\lambda)} \left\{ \frac{2(\gamma_j, \mu)}{(\gamma_j, \gamma_j)} \right\} = \frac{2(\gamma_j, \lambda)}{(\gamma_j, \gamma_j)}.$$

The last equality was found by using Lemma 2.

*This work was partially supported by the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under AFOSR Grant No. 71-2060.

[1]L. C. Biedenharn, Phys. Lett. **3**, 254 (1963); Bertrand Kostant, Trans. Am. Math. Soc. **93**, 53 (1959).

[2]A. J. MacFarlane, L. O'Raifeartaigh, and P. S. Rao, J. Math. Phys. **8**, 523 (1967).

[3]For example, Johan G. F. Belinfante and Bernard Kolman, *A Survey of Lie Groups and Lie Algebras* (SIAM, Philadelphia, 1972); James E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972); Ernest M. Loebl, *Group Theory and Its Applications* (Academic, New York, 1968); Hans Samelson, *Notes on Lie Algebras* (Van Nostrand Reinhold, New York, 1969).

[4]This form of μ is implicit in the Dynkin algorithm. See Belinfante and Kolman, p. 106. Note that there may be several such expressions for a single μ.

On perturbations of a Kerr black hole*

Robert M. Wald

Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 5 April 1973)

Gravitational perturbations of a Kerr black hole are analyzed using the Newman–Penrose formalism. Teukolsky has obtained decoupled wave equations for the perturbed Weyl tensor components ψ_0 and ψ_4. In this paper we prove that for well-behaved perturbations ψ_0 and ψ_4 uniquely determine each other, i.e., $\psi_0 = 0$ if and only if $\psi_4 = 0$. Then we solve the Kerr perturbation equations with $\psi_0 = \psi_4 = 0$ and show that the only well-behaved solutions are the trivial perturbations to other Kerr solutions via an infinitesmal change in the mass and angular momentum parameters. These results prove that either of the quantities ψ_0 or ψ_4 alone uniquely specifies the nontrivial part of a gravitational perturbation of a Kerr black hole. Consequences of this result are discussed.

1. INTRODUCTION

It is well established that sufficiently massive stars cannot support themselves after they have exhausted their thermonuclear fuel and must undergo catastrophic gravitational collapse. From the dynamical analysis of Price[1] and others and from the black hole uniqueness theorems of Israel,[2] Carter,[3] and Hawking[4] (see also Wald[5]), there is now considerable theoretical evidence that—at least for small deviations from spherical symmetry and very likely in all cases—when complete gravitational collapse does occur, the final state will be a Kerr black hole. Thus, there has been a great deal of interest in the properties of Kerr black holes and the astrophysical effects they might produce.

Although important upper limits on energy extraction from black holes have been derived by Hawking[6] using the exact field equations, one must resort to perturbation theory to obtain the detailed and quantitative answers to the questions of black hole astrophysics. For a nonrotating, Schwarzschild black hole, the standard metric perturbation approach has proved very successful.[7] However, for a rotating, Kerr black hole—where energy extraction mechanisms[8] and other interesting phenomena become possible—the metric perturbation approach yields a complicated system of coupled partial differential equations. As yet no significant progress has been made with these equations except in the stationary, axisymmetric case where the uniqueness of well-behaved gravitational[3] and electromagnetic[9] perturbations has been proven. A variational principle approach, which may prove useful for a number of problems, has been developed by Chandrasekhar and Friedman.[10] However, this approach applies only to stationary or to axisymmetric perturbations.

On the other hand, the Newman–Penrose[11] (NP) formalism has been extremely successful for treating perturbations of a rotating black hole. One reason why this is so (although perhaps not the only reason) can be understood as follows: One of the keys to a successful perturbation analysis is to choose to work with variables such that as many quantities vanish in zeroth order as is possible. When this is done, the perturbation equations, in general, will simplify and will be more likely to decouple. In addition, the perturbation quantities will be more likely to be gauge invariant (and hence more physically meaningful). The NP formalism satisfies this criterion beautifully for the Kerr metric in that of the five basic quantities describing the gravitational field (namely, the components $\psi_0, \psi_1, \psi_2, \psi_3, \psi_4$ of the Weyl tensor with respect to an appropriately chosen complex null tetrad) only one of them (ψ_2) is nonvanishing.

Using the NP formalism, Teukolsky[12] has obtained decoupled partial differential equations for the quantities $\dot{\psi}_0$ and $\dot{\psi}_4$. (Here and throughout this paper, dots denote perturbation quantities.) Furthermore, Teukolsky has shown that these equations can be solved by separation of variables. Thus, the problem of obtaining the perturbation quantities $\dot{\psi}_0$ or $\dot{\psi}_4$ reduces to the very managable task of solving ordinary differential equations. Since the energy and angular momentum radiated to infinity by gravitational waves can be calculated immediately from $\dot{\psi}_4$ alone, a large number of important astrophysical problems can already be solved.

However, one important question which remains is whether the complete gravitational perturbation can be recovered (at least in principle) once one has solved for $\dot{\psi}_0$ or $\dot{\psi}_4$. This question is of particular importance for the stability analysis of a Kerr black hole since, although one can analyze the modes of $\dot{\psi}_0$ or $\dot{\psi}_4$, one must determine if there are any unstable modes which do not excite $\dot{\psi}_0$ or $\dot{\psi}_4$.

Since a gravitational perturbation of Kerr should have two degrees of freedom and since the complex quantity $\dot{\psi}_4$ contains the full information on the two polarizations of outgoing radiation, one might expect that $\dot{\psi}_4$ (or $\dot{\psi}_0$) alone should completely specify a gravitational perturbation. However, there are two perturbations about which neither $\dot{\psi}_0$ nor $\dot{\psi}_4$ can contain any information. These are the trivial perturbations obtained by making infinitesimal variations $\dot{m}, \dot{a}$ of the mass and angular momentum parameters of the Kerr solution. (It is easy to show that $\dot{\psi}_0 = \dot{\psi}_4 = 0$ for these perturbations.) Nevertheless, the following result in the electromagnetic case encourages one to conjecture that either one of $\dot{\psi}_0$ or $\dot{\psi}_4$ may uniquely determine the complete gravitational perturbation of a Kerr black hole up to the trivial Kerr perturbations: As Teukolsky[12] has pointed out, using the NP form of Maxwell's equations[13] it is not difficult to show that either of the quantities $\dot{\phi}_0$ and $\dot{\phi}_2$—which are analogous, respectively, to $\dot{\psi}_0$ and $\dot{\psi}_4$—uniquely determines an electromagnetic perturbation of Kerr up to the perturbation which corresponds to adding charge (or magnetic monopole moment) to the black hole.

In this paper we prove that either of the quantities $\dot{\psi}_0$ or $\dot{\psi}_4$ does indeed uniquely specify a well-behaved gravitational perturbation of a Kerr black hole up to the trivial $\dot{m}, \dot{a}$ perturbations. By "well-behaved" we mean that at some initial "time" (i.e., on an initial spacelike hypersurface which intersects the future event horizon) the perturbation (1) vanishes sufficiently rapidly at infinity, (2) has no angular singularities, and (3) is regular on the future event horizon. Specifically, property (3) requires that $\Delta^2\psi_0$ and $\Delta^{-2}\psi_4$ (where $\Delta = r^2 - 2mr + a^2$) be regular functions of coordinates which cover the Kerr horizon in a nonsingular fashion; in particular, we require that these functions (and their deri-

vatives) remain bounded as one approaches the future horizon. (The quantities $\Delta^2\dot{\psi}_0$ and $\Delta^{-2}\dot{\psi}_4$ are gauge invariant components of the perturbed Weyl tensor with respect to a tetrad which is nonsingular on the future horizon of Kerr.)

By taking the difference of two solutions with the same $\dot{\psi}_0$ (or $\dot{\psi}_4$), it is easily seen that the statement that $\dot{\psi}_0$ (or $\dot{\psi}_4$) uniquely specifies a perturbation is equivalent to showing that there are no nontrivial, well-behaved perturbations of Kerr with $\dot{\psi}_0 = 0$ or $\dot{\psi}_4 = 0$. The first step of the proof of this result consists of showing that $\dot{\psi}_0$ and $\dot{\psi}_4$ uniquely determine each other, i.e., $\dot{\psi}_0 = 0$ implies $\dot{\psi}_4 = 0$ and vice versa. We prove this by solving the NP equations for the radial dependence of $\dot{\psi}_4$ when $\dot{\psi}_0 = 0$ and showing that there are no well-behaved solutions compatible both with this radial dependence and with Teukolsky's equation.[12] (The converse result that $\dot{\psi}_4 = 0$ implies $\dot{\psi}_0 = 0$ then follows immediately by interchanging the tetrad vectors l^μ and n^μ.) An independent proof that $\dot{\psi}_0$ and $\dot{\psi}_4$ uniquely determine each other has been obtained by Press and Teukolsky.[14]

The problem of determining the radial dependence of $\dot{\psi}_4$ when $\dot{\psi}_0 = 0$ is mathematically very closely related to the problem of determining the radial dependence of the NP variables for type II algebraically special exact solutions. (A Kerr perturbation with $\dot{\psi}_0 = 0$ is, in effect, an algebraically special perturbation.) This problem is of interest in itself since it is the first step toward the explicit determination of all type II exact solutions. Since in any case we need to carry through the analysis for its applications to Kerr perturbations, in Sec. 2 we derive the radial dependence of all the NP variables for type II exact solutions. The derivation relies heavily on the methods and results of Kinnersley's[15] analysis of type D exact solutions.

In Sec. 3 we discuss the general question of transcribing results on exact solutions into results on perturbations. In Sec. 4 we use the results of Sec. 2 to show that for well-behaved Kerr perturbations we have $\dot{\psi}_0 = 0$ if and only if $\dot{\psi}_4 = 0$. The Kerr perturbation equations with $\dot{\psi}_0 = \dot{\psi}_4 = 0$ are then explicitly solved in Sec. 5 by paralleling Kinnersley's[15] analysis of type D exact solutions. (A Kerr perturbation with $\dot{\psi}_0 = \dot{\psi}_4 = 0$ is, in effect, a type D perturbation.) It is found that after all tetrad and coordinate freedom is used up, there are precisely four linearly independent solutions of the perturbation equations when $\dot{\psi}_0 = \dot{\psi}_4 = 0$. Two are the trivial $\dot{m}$, $\dot{a}$ Kerr perturbations, one is a perturbation toward the Kerr–NUT solution,[16] and the last is a perturbation toward Kinnersley's "rotating C metric."[15] These last two perturbations are physically unacceptable, i.e., they would be excluded by the boundary conditions one would impose in any physical problem. Thus, the analysis of Secs. 4 and 5 yields the result that either of the quantities $\dot{\psi}_0$ or $\dot{\psi}_4$ fully specifies the nontrivial part of a Kerr perturbation. Some consequences of this result are discussed in Sec. 6.

2. RADIAL DEPENDENCE OF TYPE II SOLUTIONS

In this section we shall derive the radial dependence of all type II algebraically special solutions using the Newman–Penrose (NP) formalism. The results of this analysis will be applied to Kerr perturbations in Sec. 4. Using a different tetrad choice, Talbot[17] has independently derived the radial dependence of the NP variables for algebraically special solutions. Hence, the results of this section could also be obtained from Talbot's results by making a null rotation of his tetrad. However, for completeness and because in Secs. 4 and 5 we will need the definitions of the quantities introduced here, we give the full derivation of the radial dependence of the NP variables below.

For the benefit of the reader who is totally unfamiliar with the NP approach, we make the following brief remarks, referring the reader to the original NP paper[11] for a more adequate discussion (including the spinor motivation of the approach). In the NP formalism, one chooses a tetrad at each point in the space-time consisting of two real null vectors, l^μ and n^μ, a complex linear combination of spacelike vectors, m^μ, and its complex conjugate $\bar{m}^\mu$. The tetrad vectors are chosen so that $l_\mu n^\mu = -m_\mu \bar{m}^\mu = 1$ (where the metric signature is $+---$) and all other dot products of the tetrad vectors vanish. The spin coefficients are defined in terms of the components with respect to the tetrad of the covariant derivatives of the tetrad vectors. Rather than work with the components of the metric tensor as the basic variables, in the NP approach one chooses to work with the spin coefficients and the tetrad components of the Weyl and Ricci tensors. In the Appendix, we give the precise definitions of the spin coefficients and Weyl tensor components which are used in the NP formalism and we describe the allowed tetrad transformations which preserve the above orthogonality relations.

At each point in a space–time there are four principal null directions of the Weyl tensor (solutions of $l_{[\mu}C_{\alpha]\beta\gamma[\delta}l_{\nu]}l^\beta l^\gamma = 0$). If two or more of these principal null directions coincide everywhere, the space–time is called algebraically special. If precisely two principal null directions coincide everywhere, the space–time is said to be of type II. If we choose the tetrad vector l^μ to point along the direction of the double principal null vector of a type II space–time, the Weyl tensor components ψ_0 and ψ_1 vanish. By the Goldberg–Sachs theorem[11] we also obtain $\kappa = \sigma = 0$. Under a null rotation of the tetrad with l^μ fixed, (see Appendix) ψ_3 transforms as

$$\psi_3 \to \psi_3 + 3\bar{a}\psi_2 + 3\bar{a}^2\psi_1 + \bar{a}^3\psi_0. \tag{2.1}$$

Since $\psi_2 \neq 0$ or the space–time would be more special than type II, we can perform a null rotation with l^μ fixed to set $\psi_3 = 0$. In addition, we can make $\epsilon = 0$ by a tetrad rotation, and we can choose the coordinate $x^2 = r$ so that $l^\mu = \partial/\partial r$. Then, the NP tetrad takes the form,

$$l^\mu = (0, 1, 0, 0), \tag{2.2}$$

$$n^\mu = (X^1, U, X^3, X^4), \tag{2.3}$$

$$m^\mu = (\xi^1, \omega, \xi^3, \xi^4). \tag{2.4}$$

Since $\kappa = \sigma = \epsilon = 0$, the NP commutation relations for the derivative operators $D \equiv l^\mu\partial/\partial x^\mu = \partial/\partial r$, $\Delta \equiv n^\mu\partial/\partial x^\mu$, $\delta \equiv m^\mu\partial/\partial x^\mu$, and $\bar{\delta} \equiv \bar{m}^\mu\partial/\partial x^\mu$ become

$$\Delta D - D\Delta = (\gamma + \bar{\gamma})D - (\tau + \bar{\pi})\bar{\delta} - (\bar{\tau} + \pi)\delta, \tag{2.5}$$

$$\delta D - D\delta = (\bar{\alpha} + \beta - \bar{\pi})D - \bar{\rho}\delta, \tag{2.6}$$

$$\delta\Delta - \Delta\delta = -\bar{\nu}D + (\tau - \bar{\alpha} - \beta)\Delta + \bar{\lambda}\bar{\delta} + (\mu - \gamma + \bar{\gamma})\delta, \tag{2.7}$$

$$\bar{\delta}\delta - \delta\bar{\delta} = (\bar{\mu} - \mu)D + (\bar{\rho} - \rho)\Delta - (\bar{\alpha} - \beta)\bar{\delta} - (\bar{\beta} - \alpha)\delta. \tag{2.8}$$

In terms of the tetrad components, these relations yield

$$DU = (\bar{\tau} + \pi)\omega + (\tau + \bar{\pi})\bar{\omega} - (\gamma + \bar{\gamma}), \tag{2.9}$$

$$DX^i = (\bar\tau + \pi)\xi^i + (\tau + \bar\pi)\bar\xi^i, \tag{2.10}$$

$$D\omega = \bar\rho\omega + (\bar\pi - \bar\alpha - \beta), \tag{2.11}$$

$$D\xi^i = \bar\rho\xi^i, \tag{2.12}$$

$$\delta U - \Delta\omega = -\bar\nu + (\tau - \bar\alpha - \beta)U + \bar\lambda\bar\omega + (\mu - \gamma + \bar\gamma)\omega, \tag{2.13}$$

$$\delta X^i - \Delta\xi^i = (\tau - \bar\alpha - \beta)X^i + \bar\lambda\bar\xi^i + (\mu - \gamma + \bar\gamma)\xi^i, \tag{2.14}$$

$$\bar\delta\omega - \delta\bar\omega = (\bar\mu - \mu) + (\bar\rho - \rho)U - (\bar\alpha - \beta)\bar\omega - (\bar\beta - \alpha)\omega, \tag{2.15}$$

$$\bar\delta\xi^i - \delta\bar\xi^i = (\bar\rho - \rho)X^i - (\bar\alpha - \beta)\bar\xi^i - (\bar\beta - \alpha)\xi^i. \tag{2.16}$$

The vacuum NP equations give

$$D\rho = \rho^2, \tag{2.17}$$

$$D\beta = \bar\rho\beta, \tag{2.18}$$

$$D\alpha = \rho(\alpha + \pi), \tag{2.19}$$

$$D\tau = \rho(\tau + \bar\pi), \tag{2.20}$$

$$D\gamma = (\tau + \bar\pi)\alpha + (\bar\tau + \pi)\beta + \tau\pi + \psi_2, \tag{2.21}$$

$$D\mu - \delta\pi = \bar\rho\mu + \pi\bar\pi - \pi(\bar\alpha - \beta) + \psi_2, \tag{2.22}$$

$$\delta\rho = \rho(\bar\alpha + \beta) + (\rho - \bar\rho)\tau, \tag{2.23}$$

$$\delta\tau = \rho\bar\lambda + \tau(\tau + \beta - \bar\alpha), \tag{2.24}$$

$$\delta\alpha - \bar\delta\beta = \mu\rho + \alpha\bar\alpha + \beta\bar\beta - 2\alpha\beta + \gamma(\rho - \bar\rho) - \psi_2, \tag{2.25}$$

$$D\nu - \Delta\pi = (\pi + \bar\tau)\mu + (\bar\pi + \tau)\lambda + (\gamma - \bar\gamma)\pi, \tag{2.26}$$

$$\Delta\alpha - \bar\delta\gamma = \rho\nu - (\tau + \beta)\lambda + (\bar\gamma - \bar\mu)\alpha + (\bar\beta - \bar\tau)\gamma, \tag{2.27}$$

$$\Delta\beta - \delta\gamma = -\gamma(\tau - \bar\alpha - \beta) - \mu\tau + \beta(\gamma - \bar\gamma - \mu) - \alpha\bar\lambda, \tag{2.28}$$

$$\delta\nu - \Delta\mu = \mu^2 + \lambda\bar\lambda + \mu(\gamma + \bar\gamma) - \bar\nu\pi + \nu(\tau - 3\beta - \bar\alpha), \tag{2.29}$$

$$D\lambda - \bar\delta\pi = \rho\lambda + \pi(\pi + \alpha - \bar\beta), \tag{2.30}$$

$$\delta\lambda - \bar\delta\mu = (\rho - \bar\rho)\nu + (\mu - \bar\mu)\pi + \mu(\alpha + \bar\beta) + \lambda(\bar\alpha - 3\beta), \tag{2.31}$$

$$\Delta\rho - \bar\delta\tau = -\rho\bar\mu + (\bar\beta - \alpha - \bar\tau)\tau + (\gamma + \bar\gamma)\rho - \psi_2, \tag{2.32}$$

$$\Delta\lambda - \bar\delta\nu = -(\mu + \bar\mu)\lambda - (3\gamma - \bar\gamma)\lambda - \psi_4 + (3\alpha + \bar\beta + \pi - \bar\tau)\nu. \tag{2.33}$$

The Bianchi identities yield

$$D\psi_2 = 3\rho\psi_2, \tag{2.34}$$

$$\bar\delta\psi_2 = -3\pi\psi_2, \tag{2.35}$$

$$\delta\psi_2 = 3\tau\psi_2, \tag{2.36}$$

$$\Delta\psi_2 = -3\mu\psi_2, \tag{2.37}$$

$$D\psi_4 = \rho\psi_4 - 3\lambda\psi_2, \tag{2.38}$$

$$\delta\psi_4 = (\tau - 4\beta)\psi_4 - 3\nu\psi_2. \tag{2.39}$$

The key point to notice is that with our choice of tetrad all of the equations used by Kinnersley[15] *to determine the radial dependence of the NP variables for the case of type D solutions also appear without modification in the type II case.* Thus, in the type II case we can immediately obtain the radial dependence of all the NP variables except those which vanish in the type D case, namely ν, λ, and ψ_4. For completeness and because later we will need the definitions of the variables introduced here, we outline Kinnersley's derivation of the radial dependence below.

The case $\rho = 0$ is not of interest here and will not be considered (although the radial dependence of the NP variables can be found without difficulty in that case). If $\rho \neq 0$, integration of Eq. (2.17) gives

$$\rho = -1/(r + i\rho^0), \tag{2.40}$$

where the superscript "0" indicates a "constant" of integration of an equation involving radial derivatives, i.e., a quantity which is independent of r. By re-adjusting r is necessary (keeping $l^\mu = \partial/\partial r$) we make ρ^0 real. Application of the commutator $(\bar\delta D - D\bar\delta)$ to ρ yields

$$D\bar\delta\rho - 3\rho\bar\delta\rho = \rho^2(\pi - \alpha - \bar\beta) \tag{2.41}$$

with the general solution [using Eqs. (2.17)–(2.19)]

$$\bar\delta\rho = \rho(\alpha + \bar\beta) - 2\rho^3\bar\tau^0. \tag{2.42}$$

Application of the commutator $(\bar\delta D - D\bar\delta)$ to ψ_2 yields

$$D\pi = \rho(\alpha + \bar\beta) - \bar\delta\rho = 2\rho^3\bar\tau^0 \tag{2.43}$$

or

$$\pi = \pi^0 + \rho^2\bar\tau^0. \tag{2.44}$$

Integration of Eqs. (2.9)–(2.12), (2.18)–(2.21), and (2.34) then yields

$$\beta = \bar\rho\beta^0, \tag{2.45}$$

$$\alpha = \rho\alpha^0 - \pi^0 + \rho^2\bar\tau^0, \tag{2.46}$$

$$\tau = \rho\eta^0 + \rho\bar\rho\tau^0 - \bar\pi^0, \tag{2.47}$$

$$\omega = \bar\rho\omega^0 + \bar\alpha^0 + \beta^0 - \bar\pi^0/\bar\rho, \tag{2.48}$$

$$\xi^i = \bar\rho\xi^{0i}, \tag{2.49}$$

$$X^i = X^{0i} + \rho\bar\rho(\bar\tau^0\xi^{0i} + \tau^0\bar\xi^{0i}) + \rho\eta^0\bar\xi^{0i} + \bar\rho\bar\eta^0\xi^{0i}, \tag{2.50}$$

$$\psi_2 = \psi_2^0\rho^3, \tag{2.51}$$

$$\gamma = \gamma^0 + \rho(\eta^0\alpha^0 - \bar\tau^0\bar\pi^0) + \bar\rho(\bar\eta^0\beta^0 - \tau^0\pi^0) - r\pi^0\bar\pi^0 + \rho^2(\tfrac{1}{2}\psi_2^0 + \bar\tau^0\eta^0) + \rho\bar\rho(\tau^0\alpha^0 + \bar\tau^0\beta^0) + \rho^2\bar\rho\tau^0\bar\tau^0, \tag{2.52}$$

$$U = U^0 - r(\gamma^0 + \bar\gamma^0 + \eta^0\pi^0 + \bar\eta^0\bar\pi^0) + r^2\pi^0\bar\pi^0 + \rho[\bar\tau^0(\bar\alpha^0 + \beta^0) - \bar\tau^0\eta^0 + \eta^0\bar\omega^0 - \tfrac{1}{2}\psi_2^0] - \rho\bar\tau^0\bar\pi^0/\bar\rho + \bar\rho[\tau^0(\alpha^0 + \bar\beta^0) - \tau^0\bar\eta^0 + \bar\eta^0\omega^0 - \tfrac{1}{2}\bar\psi_2^0] - \bar\rho\tau^0\pi^0/\rho + \rho\bar\rho[\bar\tau^0\omega^0 + \tau^0\bar\omega^0 - \tau^0\bar\tau^0]. \tag{2.53}$$

Next we apply the commutator $\Delta D - D\Delta$ to ρ and get

$$D\Delta\rho - 2\rho\Delta\rho = -\rho^2(\gamma + \bar\gamma) + (\tau + \bar\pi)\bar\delta\rho + (\bar\tau + \pi)\delta\rho \tag{2.54}$$

which, using Eqs. (2.23), (2.42), (2.44), (2.47), and (2.52) can be integrated to yield

$$\Delta\rho = -\rho^2M^0 + \rho^2\eta^0(\alpha^0 + \bar\beta^0) + \bar\rho^0\bar\eta^0\bar\pi^0 + \rho(\gamma^0 + \bar\gamma^0 + \eta^0\pi^0) - \rho^3(\tfrac{1}{2}\psi_2^0 + \bar\tau^0\eta^0)$$

$$+ \rho\bar{\rho}[\bar{\eta}^0(\bar{\alpha}^0 + \beta^0) - \tau^0\pi^0 + \bar{\tau}^0\bar{\pi}^0 - \eta^0\bar{\eta}^0]$$
$$- \rho^2\bar{\rho}[\tfrac{1}{2}\bar{\psi}_2^0 + \bar{\tau}^0\eta^0 - \tau^0(\alpha^0 + \bar{\beta}^0) - \bar{\tau}^0(\bar{\alpha}^0 + \beta^0)]$$
$$- \rho^3\bar{\rho}\tau^0\bar{\tau}^0 + r^2\rho^2\pi^0\bar{\pi}^0. \tag{2.55}$$

Finally, we apply the commutator $\Delta D - D\Delta$ to ψ_2 and get

$$D\mu = -\Delta\rho + \rho(\gamma + \bar{\gamma}) + \pi\bar{\pi} - \tau\bar{\tau} \tag{2.56}$$

which can be integrated to yield

$$\mu = \mu^0 + \rho(M^0 - \bar{\tau}^0\bar{\pi}^0) + \bar{\rho}\tau^0\pi^0$$
$$+ \rho^2(\tfrac{1}{2}\psi_2^0 + \bar{\tau}^0\eta^0) + \tfrac{1}{2}\rho\bar{\rho}\psi_2^0 + \rho^2\bar{\rho}\tau^0\bar{\tau}^0$$
$$- r^2\rho\pi^0\bar{\pi}^0. \tag{2.57}$$

This completes Kinnersley's derivation of the radial dependence of all the NP variables except ψ_4, λ, and ν. Substituting the above results into Eq. (2.30), we obtain

$$D\lambda = \rho\lambda + K^0\rho + L^0\rho^3, \tag{2.58}$$

where

$$K^0 = \xi^{0i}\pi^0_{,i} + \pi^0(\alpha^0 - \bar{\beta}^0), \tag{2.59}$$

$$L^0 = \bar{\xi}^{0i}\tau^0_{,i} + \bar{\tau}^0(3\bar{\alpha}^0 + \beta^0). \tag{2.60}$$

Integration of Eq. (2.58) yields,

$$\lambda = \lambda^0\rho - K^0 + L^0\rho^2. \tag{2.61}$$

Substituting this result and Eq. (2.51) into Eq. (2.38), we get

$$D\psi_4 = \rho\psi_4 - 3\psi_2^0\rho^3(\lambda^0\rho - K^0 + L^0\rho^2) \tag{2.62}$$

which can be integrated to yield

$$\psi_4 = \psi_4^0\rho + 3K^0\psi_2^0\rho^2 - \tfrac{3}{2}\lambda^0\psi_2^0\rho^3 - L^0\psi_2^0\rho^4. \tag{2.63}$$

Equation (2.63) is the key result of this section. It shows that $\rho^{-4}\psi_4$ is a cubic polynomial in r. In Sec. 4 we will use the linearized version of this result to prove that $\dot{\psi}_4 = 0$ for well-behaved Kerr perturbations with $\dot{\psi}_0 = 0$.

Finally, the radial dependence of ν can be determined, e.g., from Eq. (2.39), now that the radial dependence of all the other NP variables is known. The next step in a systematic analysis of type II solutions is to substitute the radial dependences determined above into all the remaining NP equations. One then equates coefficients of the various powers of $1/r$, thus eliminating r and reducing the problem to solving for functions of three variables only. Using a different tetrad choice, Talbot[17] has proceeded with this analysis.

3. LINEARIZATION OF RESULTS ON EXACT EQUATIONS

In the next two sections, we shall see that with an appropriate tetrad choice the Kerr perturbation equations with $\dot{\psi}_0 = 0$ and with $\dot{\psi}_0 = \dot{\psi}_4 = 0$ are precisely the linearizations off Kerr of the exact NP equations for type II and type D solutions, respectively. We shall then apply the results of Sec. 2 and much of Kinnersley's analysis[15] to Kerr perturbations. The purpose of this section is to warn the reader that such a carrying over of results on exact equations to results on perturbations need not always be possible. This is so because there are a number of operations which can be performed on exact equations which cannot be paralleled in the linearized case. Most notorious among these operations is the taking of square roots or other fractional powers: The exact equation $f^2 = 0$ implies $f = 0$, but if f vanishes in zeroth order the linearized equation corresponding to $f^2 = 0$ is merely $0 = 0$, which does not imply $f = 0$. Thus, in applying the results of any analysis on exact equations to perturbations, one must carefully check that all the corresponding steps can be performed with the linearized equations; one cannot automatically linearize the final conclusions.

Hawking's theorem[4] provides an interesting and relevant illustration of the difficulties one can encounter in attempting to apply results from exact theory to perturbations. Hawking has proven that a stationary, rotating black hole must be axisymmetric. One might expect that this result would exclude the existence of well-behaved, stationary, nonaxisymmetric perturbations of Kerr. (Such perturbations are of interest since they would indicate a neutral mode of oscillation and would be of strong evidence for the instability of Kerr black holes.) However, Hawking's theorem does not carry over to the linearized case. Early in his argument Hawking uses the condition $\rho = 0$ and the equation

$$D\rho = \rho^2 + \sigma\bar{\sigma} + (\epsilon + \bar{\epsilon})\rho \tag{3.1}$$

to conclude that $\sigma = 0$ for the generators of the horizon. But for Kerr perturbations, Eq. (3.1) yields no information on $\dot{\sigma}$ to first order. Thus one cannot conclude that $\dot{\sigma} = 0$ and the attempted proof breaks down. However, as a consequence of the results of Secs. 4 and 5 we will prove in Sec. 6 that, in fact, there exist no well-behaved stationary, nonaxisymmetric perturbations of Kerr.

4. KERR PERTURBATIONS WITH $\dot{\psi}_0 = 0$

We now turn our attention to perturbations of a Kerr black hole. The notation throughout the rest of the paper is as follows: Dotted quantities (e.g., $\dot{\psi}_2$) will denote perturbations of Kerr; undotted quantities (e.g., ψ_2) will denote the zeroth order Kerr values. In zeroth order, we use Kinnersley's tetrad.[15] Then we have $\psi_0 = \psi_1 = \psi_3 = \psi_4 = \kappa = \sigma = \lambda = \nu = \epsilon = 0$. With a suitable choice of coordinates, given in terms of Kerr coordinates by $x^1 = u$, $x^2 = r$, $x^3 = x = \cos\theta$, $x^4 = \varphi$, the Kerr values of the "constants" of integration introduced in Sec. 2 become

$$\rho^0 = -ax, \tag{4.1a}$$

$$\tau^0 = -ia(1 - x^2)^{1/2}/\sqrt{2}, \tag{4.1b}$$

$$M^0 = -U^0 = \tfrac{1}{2}, \tag{4.1c}$$

$$\psi_2^0 = m, \tag{4.1d}$$

$$\beta^0 = -\alpha^0 = -(\sqrt{2}/4)x/(1 - x^2)^{1/2}, \tag{4.1e}$$

$$X^{0i} = \delta_{i1}, \tag{4.1f}$$

$$\xi^{01} = -iax^2(1 - x^2)^{-1/2}/\sqrt{2}, \tag{4.1g}$$

$$\xi^{03} = (1 - x^2)^{1/2}/\sqrt{2}, \tag{4.1h}$$

$$\xi^{04} = i\sqrt{2}/(1 - x^2)^{1/2}, \tag{4.1i}$$

$$\pi^0 = \omega^0 = \eta^0 = \gamma^0 = \mu^0 = 0 \tag{4.1j}$$

(and, of course, $\lambda^0 = K^0 = L^0 = \psi_4^0 = 0$).

From the behavior of the Weyl tensor components under tetrad transformations, it may be easily verified that the perturbation quantities $\dot{\psi}_0$ and $\dot{\psi}_4$ are gauge invariant, i.e., their values are unaffected to first order by all allowed infinitesimal tetrad transformations and all infinitesimal coordinate transformations. We shall prove below that for well-behaved perturbations, if $\dot{\psi}_0$ is given, then $\dot{\psi}_4$ is uniquely determined. By taking the difference between two perturbations with the same given $\dot{\psi}_0$, this is easily seen to be equivalent to proving that $\dot{\psi}_0 = 0$ implies $\dot{\psi}_4 = 0$. Once this is proven, the converse result that $\dot{\psi}_4 = 0$ implies $\dot{\psi}_0 = 0$ (i.e., $\dot{\psi}_4$ uniquely determines $\dot{\psi}_0$) follows immediately by interchanging the tetrad vectors l^μ and n^μ.

Consider a Kerr perturbation with $\dot{\psi}_0 = 0$. By an infinitesimal null rotation of the tetrad with l^μ fixed, we may set $\dot{\psi}_3 = 0$ [see Eq. (2.1)]. Similarly, by an infinitesimal null rotation with n^μ fixed, we set $\dot{\psi}_1 = 0$. (Neither of these transformations affects $\dot{\psi}_0 = 0$, since $\dot{\psi}_0$ is gauge invariant). The linearized Bianchi identities involving $\dot{\psi}_0$ and $\dot{\psi}_1$ then immediately yield $\dot{\kappa} = \dot{\sigma} = 0$ (i.e., we obtain the linearized version of the Goldberg–Sachs theorem[11]). By an infinitesimal tetrad rotation we can set $\dot{\epsilon} = 0$. If we choose the coordinate r for the perturbed space-time to maintain $l^\mu = \partial/\partial r$, we find that the *Kerr perturbation equations for the case* $\dot{\psi}_0 = 0$ *are precisely the linearizations off Kerr of the NP equations for type II exact solutions given in Sec. 2.*

It may be verified that all the integrations and algebraic manipulations performed with the exact equations in Sec. 2 can be performed with equal validity for the linearized equations. Thus, by a precise paralleling of steps, we find that when $\dot{\psi}_0 = 0$, the radial dependence of $\dot{\psi}_4$ is given by the linearized version of Eq. (2.63),

$$\dot{\psi}_4 = \dot{\psi}_4^0\rho + 3\dot{K}^0\psi_2^0\rho^2 - \tfrac{3}{2}\dot{\lambda}^0\psi_2^0\rho^3 - \dot{L}^0\psi_2^0\rho^4. \quad (4.2)$$

Thus, $\rho^{-4}\dot{\psi}_4$ is a cubic polynomial in r. On the other hand, $\dot{\psi}_4$ must also satisfy Teukolsky's equation.[12] [This equation is derived by solving the linearized version of Eqs. (2.38) and (2.39) for $\dot{\lambda}$ and $\dot{\nu}$ in terms of $\dot{\psi}_4$ and substituting the result into the linearized version of Eq. (2.33). Since Eq. (2.33) was not used in the analysis of Sec. II, Teukolsky's equation can be expected to yield new information on $\dot{\psi}_4$.] Teukolsky has shown that in Kerr coordinates the general solution of his equation takes the form

$$\rho^{-4}\dot{\psi}_4 = \int d\omega \sum_{l,m} R_{lm}(r, \omega)S_{lm}(\theta, \omega)e^{-i\omega u}e^{im\varphi}, \quad (4.3)$$

where the S_{lm} are spin weighted spheroidal harmonics[12] and R_{lm} satisfies Teukolsky's radial equation. In our case, R_{lm} must be a cubic polynomial. However, Teukolsky (private communication) has found that when one substitutes a cubic polynomial into the radial equation one gets four homogeneous, linear equations for the four coefficients of the polynomial. Hence, the necessary and sufficient condition for the existence of a cubic polynomial solution of Teukolsky's equation is that the determinant of this linear system vanish. Explicitly, this condition is

$$0 = \det = \Lambda^2(\Lambda - 2)^2 - 12\Lambda(\Lambda - 2)\Gamma\omega - 8\Lambda^2\Gamma\omega + 36(\Gamma^2 + 4M^2)\omega^2 - 96\Lambda a^2\omega^2, \quad (4.4)$$

where

$$\Gamma = 2\omega a^2 - 2am, \quad (4.5)$$

$$\Lambda = 2am\omega - a^2\omega^2 - A, \quad (4.6)$$

and $A = A(l, m, \omega)$ is the eigenvalue associated with the spin weighted spheroidal harmonic $S_{lm}(\theta, \omega)$. In the case of a Schwarzschild black hole ($a = 0$), we have $A = (l + 2)(l - 1)$ (independent of ω) and the determinant, Eq. (4.4), reduces to a quadratic polynomial in ω. It vanishes when

$$\omega = \pm\, i(l + 2)(l + 1)l(l - 1)/12M. \quad (4.7)$$

This agrees with the time dependence of the algebraically special perturbations of Schwarzschild investigated[18] by Kinnersley and Couch and Newman. In the general Kerr case ($a \neq 0$), Eq. (4.4) is considerably less simple to solve, owing mainly to the fact that the value of A depends on ω and must be determined by solving the angular eigenfunction equation. The important point to note, however, is that solutions can occur only for special values of ω. In particular, when $\omega = 0$ we again have $A = (l + 2)(l - 1)$ and Eq. (4.4) is not satisfied, i.e., there are no stationary, algebraically special perturbations of Kerr with $\dot{\psi}_4 \neq 0$. Since the determinant, Eq. (4.4), is an analytic function of ω, it follows that solutions of Eq. (4.4) can occur only at isolated values of ω (because if the determinant had an accumulation point of zeros it would have to vanish everywhere and this would contradict the fact that it doesn't vanish at $\omega = 0$). It appears likely that there are no solutions of Eq. (4.4) for real ω. In that case, we see from Eq. (4.3) (recalling that R_{lm} has cubic polynomial dependence on r) that on an initial spacelike slice which intersects the future horizon, the imaginary part of ω will cause $\dot{\psi}_4$ to blow up exponentially either at the horizon (where $u \to +\infty$) or at spacial infinity (where $u \to -\infty$). Hence, these perturbations will be physically unacceptable. But even if solutions with real ω occur, each individual ω mode is not a physically acceptable solution by itself since it does not behave properly at the horizon: As can be seen from Eq. (4.3) the real part of ω will cause $\dot{\psi}_4$ to undergo infinite oscillation in a finite region as one approaches the future horizon ($u \to \infty$). If solutions over a continuous range of ω were available, one could cancel the singularities of the real ω modes by forming wave packets. However, in our case ω can take on only the discrete values satisfying Eq. (4.4). Hence, we conclude that there exist no physically acceptable cubic polynomial solutions of Teukolsky's equation. Thus the only physically acceptable perturbations with $\dot{\psi}_0 = 0$ must also satisfy

$$\dot{\psi}_4 = 0. \quad (4.8)$$

Strictly speaking to draw this conclusion in an entirely rigorous fashion one should prove that (for fixed l, m) either (1) there exist only a finite number of solutions of Eq. (4.4) for ω (which is equivalent to showing that solutions do not occur for arbitrarily large ω since only isolated solutions can occur), or (2) in the unlikely event that there are infinitely many solutions, that the perturbations corresponding to these solutions cannot be added in such a way as to exactly cancel the bad behavior of each of the individual solutions on the horizon. This last step has not been carried out, but an independent proof that $\dot{\psi}_0 = 0$ implies $\dot{\psi}_4 = 0$ has been obtained by Press and Teukolsky.[14]

5. KERR PERTURBATIONS WITH $\dot{\psi}_0 = \dot{\psi}_4 = 0$

In this section we shall prove that the only well-behaved perturbations of a Kerr black hole with $\dot{\psi}_0 = \dot{\psi}_4 = 0$ are the trivial perturbations to other Kerr solutions via an infinitesimal change in the mass and angular momentum parameters. Together with the results of

Sec. 4, this shows that a perturbation of a Kerr black hole is uniquely specified up the trivial $\dot{m}, \dot{a}$ perturbations by giving either of the quantities $\dot{\psi}_0$ or $\dot{\psi}_4$ alone.

Consider a Kerr perturbation with $\dot{\psi}_0 = \dot{\psi}_4 = 0$. As in Sec. 4 above we may set $\dot{\psi}_1 = \dot{\psi}_3 = 0$ (without affecting $\dot{\psi}_0 = \dot{\psi}_4 = 0$) by infinitesimal null rotations of the tetrad. Then we also obtain $\dot{\sigma} = \dot{\kappa} = \dot{\nu} = \lambda = 0$. Again, by an infinitesimal tetrad rotation we set $\dot{\epsilon} = 0$ and we choose the coordinate r for the perturbed space-time so as to maintain $l^\mu = \partial/\partial r$. *The Kerr perturbation equations then precisely reduce to the equations one gets by linearizing the NP equations for exact type D solutions analyzed by Kinnersley*[15] *about Kerr background.*

As in Sec. 4, we may parallel Kinnersley's[15] derivation of the radial dependence of the NP variables to get the linearized versions of the equations given in Sec. 2 above. The next step in Kinnersley's analysis is the substitution of the radial dependence of the NP variables in the remaining NP equations. In the exact case, this yields equations of the form

$$\sum_n a_n^0(1/r^n) = 0, \tag{5.1}$$

where a_n^0 is independent of r, which implies

$$a_n^0 = 0. \tag{5.2}$$

For the case of Kerr perturbations, the analogous procedure leads to the linearized version of Eq. (5.1), i.e.,

$$\sum_n \dot{a}_n^0(1/r^n) = 0 \tag{5.3}$$

which similarly implies

$$\dot{a}_n^0 = 0. \tag{5.4}$$

Thus, by paralleling Kinnersley's analysis, we obtain the linearized versions of all of Kinnersley's equations given in Sec. IIIB of his paper. Furthermore, as in the exact case, we still have the freedom to make infinitesimal tetrad rotations which are independent of r. (In addition, the coordinates x^1, x^3, and x^4 have yet to be specified for the perturbed space-time.) Following Kinnersley, we make an infinitesimal tetrad rotation $l^\mu \to (1 + \dot{A}^0)^{-1} l^\mu$, $n^\mu \to (1 + \dot{A}^0) n^\mu$ to keep $\psi_2^0 \bar{\psi}_2^0$ a constant to first order. As in the case of exact type D solutions, this tetrad choice can be shown to imply the following relations:

$$\dot{\bar{\alpha}}^0 + \dot{\beta}^0 = 2i\rho^0 \dot{\bar{\pi}}^0, \tag{5.5a}$$

$$\dot{\gamma}^0 + \dot{\bar{\gamma}}^0 + \dot{\mu}^0 = 0. \tag{5.5b}$$

Below, we give the NP Kerr perturbation equations for the case $\dot{\psi}_0 = \dot{\psi}_4 = 0$. These equations are obtained by linearizing the equations of Sec. IIIB of Kinnersley's paper about the Kerr background, and making the simplifications Eq. (5.5), given by our tetrad choice, and some simplifications resulting from the simple form of the zeroth order Kerr solution, Eq. (4.1):

$$\dot{\omega}^0 = -2(\rho^0)^2 \dot{\bar{\pi}}^0 \tag{5.6}$$

$$\dot{\eta}^0 = 2i\rho^0 \dot{\bar{\pi}}^0, \tag{5.7}$$

$$\dot{\mu}^0 = \dot{\bar{\mu}}^0 \tag{5.8}$$

$$\dot{M}^0 - \dot{\bar{M}}^0 = 2i\rho^0 \dot{\mu}^0 + 4\bar{\tau}\dot{\bar{\pi}}^0, \tag{5.9}$$

$$\tau^0 \dot{\pi}^0 + \bar{\tau}^0 \dot{\bar{\pi}}^0 = 0, \tag{5.10}$$

$$\dot{U}^0 = -\tfrac{1}{2}(\dot{M}^0 + \dot{\bar{M}}^0), \tag{5.11}$$

$$\tau^0 \dot{\mu}^0 = i\rho^0 \dot{\bar{\pi}}^0, \tag{5.12}$$

$$\dot{\xi}^{0i}\rho^0_{,i} + \xi^{0i}\dot{\rho}^0_{,i} = -i\dot{\tau}^0 + 2i(\rho^0)^2 \dot{\bar{\pi}}^0, \tag{5.13}$$

$$\dot{X}^{0i}\rho^0_{,i} + X^{0i}\dot{\rho}^0_{,i} = 0, \tag{5.14}$$

$$\xi^{0i}\dot{\psi}^0_{2,i} = 6i\rho^0 \psi_2^0 \dot{\bar{\pi}}^0, \tag{5.15}$$

$$\bar{\xi}^{0i}\dot{\psi}^0_{2,i} = 6i\rho^0 \psi_2^0 \dot{\bar{\pi}}^0, \tag{5.16}$$

$$X^{0i}\dot{\psi}^0_{2,i} = 0, \tag{5.17}$$

$$\xi^{0i}\dot{\bar{\pi}}^0_{,i} = -\dot{\mu}^0 + \dot{\bar{\pi}}^0(\bar{\alpha}^0 - \beta^0), \tag{5.18}$$

$$\bar{\xi}^{0i}\dot{\bar{\pi}}^0_{,i} = \dot{\bar{\pi}}^0(\bar{\beta}^0 - \alpha^0), \tag{5.19}$$

$$\dot{\xi}^{0i}\tau^0_{,i} + \xi^{0i}\dot{\tau}^0_{,i} = -\dot{\tau}^0(3\bar{\alpha}^0 + \beta^0) - \tau^0(3\dot{\bar{\alpha}}^0 + \dot{\beta}^0), \tag{5.20}$$

$$\dot{\bar{\xi}}^{0i}\tau^0_{,i} + \bar{\xi}^{0i}\dot{\tau}^0_{,i} = -\dot{\tau}^0(\alpha^0 + 3\bar{\beta}^0) - \tau^0(\dot{\alpha}^0 + 3\dot{\bar{\beta}}^0) - i\dot{\rho}^0 + \tfrac{1}{2}(\dot{\psi}_2^0 - \dot{\bar{\psi}}_2^0) - 4i\tau^0\rho^0\dot{\pi}^0 - 2i\rho^0\dot{\bar{M}}^0, \tag{5.21}$$

$$\dot{\xi}^{0i}\alpha^0_{,i} + \xi^{0i}\dot{\alpha}^0_{,i} - \dot{\bar{\xi}}^{0i}\beta^0_{,i} - \bar{\xi}^{0i}\dot{\beta}^0_{,i} = 2\dot{\beta}^0(\bar{\beta}^0 - \alpha^0) + 2\beta^0(\dot{\bar{\beta}}^0 - \dot{\alpha}^0) + \dot{M}^0 + 2i\rho^0(\dot{\gamma}^0 + \dot{\mu}^0 + \alpha^0\dot{\bar{\pi}}^0 + \beta^0\dot{\pi}^0), \tag{5.22}$$

$$\bar{\xi}^{0i}\dot{\mu}^0_{,i} = \dot{\pi}^0, \tag{5.23}$$

$$\bar{\xi}^{0i}\dot{M}^0_{,i} = 2i\rho^0\dot{\pi}^0 + (\psi_2^0 + 2\bar{\psi}_2^0)\dot{\pi}^0 - 2\bar{\tau}^0\dot{\mu}^0, \tag{5.24}$$

$$X^{0i}\dot{\pi}^0_{,i} = 0, \tag{5.25}$$

$$\dot{X}^{0i}\tau^0_{,i} + X^{0i}\dot{\tau}^0_{,i} = -\tau^0(2\dot{\bar{\gamma}}^0) - i\rho^0\dot{\bar{\pi}}^0, \tag{5.26}$$

$$\dot{X}^{0i}\alpha^0_{,i} + X^{0i}\dot{\alpha}^0_{,i} - \xi^{0i}\dot{\gamma}^0_{,i} = -\alpha^0(\dot{\mu}^0 + \dot{\gamma}^0) + \bar{\beta}^0\dot{\gamma}^0 + \tfrac{1}{2}\dot{\pi}^0, \tag{5.27}$$

$$\dot{X}^{0i}\beta^0_{,i} + X^{0i}\dot{\beta}^0_{,i} - \xi^{0i}\dot{\gamma}^0_{,i} = -\beta^0(\dot{\mu}^0 - \dot{\gamma}^0 + 2\dot{\bar{\gamma}}^0) + \bar{\alpha}^0\dot{\gamma}^0 + \tfrac{1}{2}\dot{\bar{\pi}}^0, \tag{5.28}$$

$$X^{0i}\dot{\mu}^0_{,i} = 0, \tag{5.29}$$

$$X^{0i}\dot{M}^0_{,i} = 0, \tag{5.30}$$

$$\xi^{0i}\dot{\mu}^0_{,i} = \dot{\bar{\pi}}^0, \tag{5.31}$$

$$\xi^{0i}\dot{M}^0_{,i} = 4i\rho^0\dot{\bar{\pi}}^0 + 3\bar{\psi}_2^0\dot{\bar{\pi}}^0, \tag{5.32}$$

$$\xi^{0i}\dot{X}^{0j}_{,i} - \dot{X}^{0i}\xi^{0j}_{,i} - X^{0i}\dot{\xi}^{0j}_{,i} = (\dot{\mu}^0 + 2\dot{\bar{\gamma}}^0)\xi^{0j}, \tag{5.33}$$

$$\dot{\xi}^{0i}\xi^{0j}_{,i} - \bar{\xi}^{0i}\dot{\xi}^{0j}_{,i} - \dot{\xi}^{0i}\bar{\xi}^{0j}_{,i} - \xi^{0i}\dot{\bar{\xi}}^{0j}_{,i} = -\xi^{0j}(2\dot{\bar{\beta}}^0 + 2i\rho^0\dot{\bar{\pi}}^0) - 2\bar{\beta}^0\dot{\xi}^{0j} + \bar{\xi}^{0j}(2\dot{\beta}^0 - 2i\rho^0\dot{\bar{\pi}}^0) + 2\beta^0\dot{\bar{\xi}}^{0j} - 2i\rho^0\dot{X}^{0j} - 2i\dot{X}^{0j}\dot{\rho}^0. \tag{5.34}$$

At this stage in the analysis of exact type D solutions, Kinnersley resolves the solutions into various cases according to whether or not $\pi^0 = 0$. Unfortunately, we cannot parallel his analysis exactly here because we can neither assume that $\dot{\pi}^0 = 0$ nor can we follow his steps for the case $\dot{\pi}^0 \neq 0$ (since, in effect, he divides by π^0 and takes square roots). However, we can still proceed to solve the equations, following closely the methods used in Kinnersley's Ph.D. thesis.[15]

Reality of $\dot{U}^0, \dot{X}^{0i}$, and $\dot{\rho}^0$ is guaranteed by their definitions. The quantity $\dot{\mu}^0$ is real by Eq. (5.8), and using the zeroth order Kerr values, Eq. (4.1), we find that Eq. (5.10) shows that $\dot{\pi}^0$ is real. Equations (5.18) and (5.19) then show that

$$\dot{\mu}^0 = -4\beta^0\dot{\pi}^0. \tag{5.35}$$

(This is consistent with Eq. (5.12) since $4\beta^0 = -i\rho^0/\tau^0$ in zeroth order.) Equations (5.9) and (5.11) imply

$$\dot{M}^0 = -\dot{U}^0 + i\rho^0\dot{\mu}^0 + 2\bar{\tau}^0\dot{\pi}^0. \tag{5.36}$$

Following Kinnersley,[15] we define the quantities $\dot{b}^0, \dot{t}^0$, and $\dot{m}^0$ by the equations

$$\dot{\beta}^0 = \dot{b}^0 + i\rho^0\dot{\pi}^0, \tag{5.37}$$

$$\dot{\tau}^0 = 2(\rho^0)^2\dot{\pi}^0 + i\dot{t}^0, \tag{5.38}$$

$$\dot{\psi}_2^0 = \dot{m}^0 + i(-2\dot{\rho}^0U^0 - 2\rho^0\dot{U}^0 + 4\dot{b}^0(-i\tau^0) + 4\beta^0\dot{t}^0). \tag{5.39}$$

For zeroth order Kerr, we write $b^0 = \beta^0$, $t^0 = -i\tau^0$, and $m^0 = \psi_2^0$. Then b^0, t^0 and m^0 are real. We now make use of the infinitesimal tetrad freedom $m^\mu \to e^{i\dot{\theta}}m^\mu$ (where $\dot{\theta}$ is independent of r) to make $\dot{t}^0$ real, i.e., $\mathrm{Re}(\dot{\tau}^0) = 2(\rho^0)^2\dot{\pi}^0$. (This can be done since $\mathrm{Im}(\tau^0) \neq 0$ in zeroth order.) Using this tetrad choice, the real part of Eq. (5.26) together with Eq. (5.25) and reality of $\dot{\pi}^0$ shows that $\dot{\gamma}^0$ is real. Then, from Eqs. (5.5b) and (5.35) we obtain,

$$\dot{\gamma}^0 = 2b^0\dot{\pi}^0. \tag{5.40}$$

Next, adding Eq. (5.20) to the complex conjugate of Eq. (5.21) and using Eq. (5.5a), the tetrad condition $\mathrm{Re}(\dot{\tau}^0) = 2(\rho^0)^2\dot{\pi}^0$ and the definition of $\dot{b}^0$, Eq. (5.37), we obtain,

$$4\xi^{0i}((\rho^0)^2\dot{\pi}^0)_{,i} = 4i(b^0\dot{t}^0 + \dot{b}^0t^0) + 2i\rho^0\dot{M}^0 + \tfrac{1}{2}(\dot{\bar{\psi}}_2^0 - \dot{\psi}_2^0) + i\dot{\rho}^0. \tag{5.41}$$

Using Eqs. (5.18), (5.35), (5.36) and the known zeroth order Kerr values, the real part of this equation yields the result that $\dot{b}^0$ is real. Finally, the imaginary part of Eq. (5.41) shows that $\dot{m}^0$ is real, where $\dot{m}^0$ is defined by Eq. (5.39). Thus, we now have reduced the problem of solving the Kerr perturbation equations with $\psi_0 = \psi_4 = 0$ to task of obtaining $\dot{\xi}^{0i}$, and the real quantities $\dot{X}^{0i}, \dot{\rho}^0, \dot{\pi}^0, \dot{b}^0, \dot{t}^0, \dot{U}^0$, and $\dot{m}^0$; all the NP variables can be obtained immediately by algebraic substitutions once these quantities are known.

We now specify the coordinates x^1, x^3, and x^4 for the perturbed space-time, following closely the choices made in Kinnersley's Ph.D. thesis.[15] The coordinate $x^2 = r$ was previously specified by the condition $l^\mu = \delta^\mu{}_2$ and ρ^0 real. Since X^{0i} is independent of r, we may choose x^1 so that $X^{0i} = \delta_{i1}$ (i.e., $X^0 = \partial/\partial x^1$) is maintained for the perturbed space-time. This yields $\dot{X}^{0i} = 0$. Since ρ^0 and $\dot{\rho}^0$ are independent of r and [by Eq. (5.14)] independent of x^1, we may choose the coordinate $x^3 = x$ so that the relation

$$\rho^0 + \dot{\rho}^0 = -(a + \dot{a})x + \dot{l} \tag{5.42}$$

holds for the perturbed space-time, where $\dot{a}$ and $\dot{l}$ are constants. (This equation defines x up to an infinitesimal linear transformation, $x \to x + \dot{K}x + \dot{L}$, to be fixed later.) Then Eq. (5.13) shows that $\dot{\xi}^{03}$ must be real.

Aside from the linear freedom in x, we still have the infinitesimal coordinate freedom

$$x^1 \to x^1 + \dot{f}_1(x, x^4), \tag{5.43a}$$

$$x^4 \to x^4 + \dot{f}_4(x, x^4). \tag{5.43b}$$

Under these transformations, $\dot{\xi}^{01}$ and $\dot{\xi}^{04}$ transform as follows:

$$\dot{\xi}^{01} \to \dot{\xi}^{01} + \xi^{03}\frac{\partial \dot{f}_1}{\partial x} + \xi^{04}\frac{\partial \dot{f}_1}{\partial x^1}, \tag{5.44a}$$

$$\dot{\xi}^{04} \to \dot{\xi}^{04} + \xi^{03}\frac{\partial \dot{f}_4}{\partial x} + \xi^{04}\frac{\partial \dot{f}_4}{\partial x^4}. \tag{5.44b}$$

Following Kinnersley,[15] we use this coordinate freedom to make $\dot{\xi}^{01}$ and $\dot{\xi}^{04}$ pure imaginary.

We now note the following remarkable simplifications: Equations (5.17), (5.25), (5.26), (5.28), (5.30), and (5.33), respectively, show that $\dot{m}^0, \dot{\pi}^0, \dot{t}^0, \dot{b}^0, \dot{U}^0$, and $\dot{\xi}^{0j}$ are independent of x^1. Equations (5.15) and (5.16), (5.24) and (5.32), (5.18), and (5.20), respectively, then show that $\dot{m}^0$, $\dot{U}^0, \dot{\pi}^0$, and $\dot{t}^0$ are also independent of x^4. Thus, all variables are functions of x alone, with the possible exception of $\dot{b}^0$ and $\dot{\xi}^{0i}$ which may also depend on x^4. In terms of these variables, the equations remaining to be solved [namely, Eqs. (5.13), (5.15), (5.18), (5.20), (5.22), (5.24), and (5.34)] become

$$a\dot{\xi}^{03} + \dot{a}\xi^{03} = -\dot{t}^0, \tag{5.45}$$

$$\xi^{03}\dot{m}^0_{,x} = 0, \tag{5.46}$$

$$\xi^{03}\dot{\pi}^0_{,x} = 2b^0\dot{\pi}^0, \tag{5.47}$$

$$\dot{\xi}^{03}t^0_{,x} + \xi^{03}\dot{t}^0_{,x} = 2b^0\dot{t}^0 + 2t^0\dot{b}^0, \tag{5.48}$$

$$2\dot{\xi}^{03}b^0_{,x} + 2\xi^{03}\dot{b}^0_{,x} = -8b^0\dot{b}^0 + \dot{U}^0, \tag{5.49}$$

$$\xi^{03}\dot{U}^0_{,x} = -3m^0\dot{\pi}^0, \tag{5.50}$$

$$\dot{\xi}^{03}\xi^{01}_{,x} + \xi^{03}\dot{\xi}^{01}_{,x} = -2\dot{b}^0\xi^{01} - 2b^0\dot{\xi}^{01} - i(-\dot{a}x + \dot{l}), \tag{5.51}$$

$$\dot{\xi}^{03}\xi^{04}_{,x} + \xi^{03}\dot{\xi}^{04}_{,x} = -2\dot{b}^0\xi^{04} - 2b^0\dot{\xi}^{04}. \tag{5.52}$$

Equations (5.45) and (5.48) show that $\dot{\xi}^{03}$ and $\dot{b}^0$ are also functions of x only. The solution of Eq. (5.46) is

$$\dot{m}^0 = \dot{m}, \tag{5.53}$$

where $\dot{m}$ is a real constant. Using the zeroth order Kerr values, Eq. (4.1), the solution of Eq. (5.47) is found to be

$$\dot{\pi}^0 = \dot{p}(1 - x^2)^{1/2}, \tag{5.54}$$

where $\dot{p}$ is a real constant. Integration of Eq. (5.50) yields

$$\dot{U}^0 = -3\sqrt{2}m\dot{p}x + \dot{B}, \tag{5.55}$$

where $\dot{B}$ is a real constant. We use the final remaining infinitesimal tetrad freedom $l^\mu \to (1 + \dot{A})^{-1}l^\mu$, $n^\mu \to (1 + \dot{A})n^\mu$ with $\dot{A}$ constant to set $\dot{B} = 0$. The combination of Eqs. (5.45), (5.48), and (5.49) gives

$$[(1 - x^2)^{1/2}\dot{\xi}^{03}]_{,xx} = \sqrt{2}\dot{U}^0 = -6m\dot{p}x \tag{5.56}$$

which implies

$$\dot{\xi}^{03} = (1 - x^2)^{-1/2}[-m\dot{p}x^3 + \dot{C}x + \dot{E}], \tag{5.57}$$

where $\dot{C}$ and $\dot{E}$ are constants. We make use of the remaining freedom $x \to x + \dot{K}x + \dot{L}$ to set $\dot{C} = \dot{E} = 0$. Having solved for $\dot{\xi}^{03}$, the quantities $\dot{t}^0$ and $\dot{b}^0$ are then obtained immediately from Eqs. (5.45) and (5.48). Finally, integration of Eqs. (5.51) and (5.52) yields the solutions

$$\xi^{03}\dot{\xi}^{01} = i\dot{g}_1(x^4) + \tfrac{1}{2}i\dot{a}x^2 - i\dot{l}x - \xi^{01}\dot{\xi}^{03}, \qquad (5.58)$$

$$\xi^{03}\dot{\xi}^{04} = i\dot{g}_4(x^4) - \xi^{04}\dot{\xi}^{03}, \qquad (5.59)$$

where $\dot{g}_1$ and $\dot{g}_4$ are arbitrary real functions. We now use the final coordinate freedom, Eq. (5.43) with $\dot{f}_1$ and $\dot{f}_4$ functions of x^4 only, to set $\dot{g}_1 = \dot{g}_4 = 0$.

We have now solved the Kerr perturbation equations in the case $\dot{\psi}_0 = \dot{\psi}_4 = 0$. We emphasize that no symmetries have been assumed for the solutions and, as yet, we have imposed no boundary conditions.

Having used up every last drop of tetrad and coordinate freedom, we see that the general perturbation of Kerr with $\dot{\psi}_0 = \dot{\psi}_4 = 0$ is fully specified by the four parameters $\dot{m}, \dot{a}, \dot{l}$, and $\dot{p}$. The perturbations generated by $\dot{m}$ and $\dot{a}$ are easily recognized to be the trivial perturbations to other Kerr solutions, i.e., the perturbations one obtains by making infinitesimal variations of the mass and angular momentum parameters of the Kerr solution. The $\dot{l}$ perturbation is the NUT perturbation, i.e., the perturbation one obtains by linearizing the Kerr–NUT solution[16] about Kerr. Finally, the perturbation generated by $\dot{p}$ is the perturbation toward Kinnersley's "rotating C metric."[15] (This can be seen by making the coordinate and parameter substitutions given by Kinnersley in his Ph.D. thesis[15] to recover Kerr as a limiting case of the rotating C metric, and linearizing the resulting metric about Kerr). Both the NUT and C metric perturbations are physically unacceptable, i.e., they would be excluded by the boundary conditions one would impose in any physical problem. (This can be seen without a detailed study of the pathology of these solutions by noting that they are independent of the coordinates x^1 and x^4 and thus are stationary and axisymmetric. If they were well-behaved, they would violate Carter's theorem[3] on stationary, axisymmetric perturbations of a black hole.) Thus, we are left with the final conclusion: *The only well-behaved Kerr perturbations with* $\dot{\psi}_0 = \dot{\psi}_4 = 0$ *are the trivial* $\dot{m}, \dot{a}$ *perturbations to other Kerr solutions.*

6. CONSEQUENCES

The results of the previous two sections prove that the nontrivial part of a perturbation of a Kerr black hole is fully specified by giving the quantity $\dot{\psi}_0$ (or $\dot{\psi}_4$) alone. Namely, if we take the difference between two perturbations with the same $\dot{\psi}_0$, we obtain a perturbation with $\dot{\psi}_0 = 0$. Then, by the results of Sec. 4, this difference perturbation must also have $\dot{\psi}_4 = 0$, and by the results of of Sec. 5 it therefore must be a trivial $\dot{m}, \dot{a}$ perturbation to other Kerr solutions.

The main importance of this result is that it shows that one does not lose any essential information by working with only the quantity $\dot{\psi}_0$ (or $\dot{\psi}_4$) alone. Once one has solved Teukolsky's equation[12] to get $\dot{\psi}_0$ (or $\dot{\psi}_4$), the complete gravitational perturbation can be recovered (at least in principle) except for the trivial $\dot{m}, \dot{a}$ parts. This is of particular importance for stability analyses of a Kerr black hole. Since the Kerr geometry is stationary, all Kerr perturbations can be decomposed into modes with time dependence $\exp(\alpha t + i\omega t)$. If there exist any modes with $\alpha > 0$ obeying the appropriate boundary conditions at $t = 0$, an initially small perturbation will grow without bound, and the Kerr black hole will be unstable. However, the analysis of Press and Teukolsky[19] indicates that there are no such exponentially blowing-up solutions of Teukolsky's equation for $\dot{\psi}_4$. The results proven above rule out the remaining possibility that there might be exponentially growing modes which do not excite $\dot{\psi}_4$ (i.e., for which $\dot{\psi}_4 = 0$).

As a final application of the results of this paper, we reprove Carter's theorem[3] on stationary Kerr perturbations and generalize it to the nonaxisymmetric case. (As described in Sec. 3 above, Hawking's theorem[4] cannot exclude well-behaved stationary nonaxisymmetric perturbations of Kerr). Carter[3] has proven that there can be at most two linearly independent well-behaved stationary, axisymmetric perturbations of a Kerr black hole (or any other stationary, axisymmetric black hole, if others besides Kerr exist). The generalization of this result for Kerr is proven as follows: In the stationary case, Teukolsky's equation can be solved in closed form, and Teukolsky[12] has found that there are no solutions which are well-behaved both at infinity and on the horizon. Thus, $\dot{\psi}_0 = \dot{\psi}_4 = 0$ for well-behaved stationary perturbations of Kerr. Then, by the results of this paper, it follows that the only well-behaved stationary perturbations of Kerr are the trivial $\dot{m}, \dot{a}$ perturbations to other Kerr solutions. In particular, there are no well-behaved stationary, nonaxisymmetric perturbations of Kerr.

ACKNOWLEDGMENTS

I wish to thank William Kinnersley for helpful comments on an outline of the results of this paper. Special thanks go to Saul Teukolsky for helpful discussions and for performing some of the algebraic computations required to complete the proof of Sec. 4.

APPENDIX: NP VARIABLES AND TETRAD TRANSFORMATIONS

In the NP formalism one chooses at each point in the space–time two real null vectors l^μ and n^μ, a complex combination of spacelike vectors m^μ and its complex conjugate $\overline{m}^\mu$. The vectors are required to satisfy $l^\mu n_\mu = -m^\mu \overline{m}_\mu = 1$, and all other dot products are required to vanish. The spin coefficients and Weyl tensor components used in the formalism are defined as follows:

$$\kappa = l_{\mu;\nu} m^\mu l^\nu, \qquad \nu = -n_{\mu;\nu}\overline{m}^\mu n^\nu,$$

$$\rho = l_{\mu;\nu} m^\mu \overline{m}^\nu, \qquad \mu = -n_{\mu;\nu}\overline{m}^\mu m^\nu,$$

$$\sigma = l_{\mu;\nu} m^\mu m^\nu, \qquad \lambda = -n_{\mu;\nu}\overline{m}^\mu \overline{m}^\nu,$$

$$\tau = l_{\mu;\nu} m^\mu n^\nu, \qquad \pi = -n_{\mu;\nu}\overline{m}^\mu l^\nu,$$

$$\epsilon = \tfrac{1}{2}(l_{\mu;\nu} n^\mu l^\nu - m_{\mu;\nu}\overline{m}^\mu l^\nu),$$

$$\gamma = \tfrac{1}{2}(l_{\mu;\nu} n^\mu n^\nu - m_{\mu;\nu}\overline{m}^\mu n^\nu),$$

$$\alpha = \tfrac{1}{2}(l_{\mu;\nu} n^\mu \overline{m}^\nu - m_{\mu;\nu}\overline{m}^\mu \overline{m}^\nu),$$

$$\beta = \tfrac{1}{2}(l_{\mu;\nu} n^\mu m^\nu - m_{\mu;\nu}\overline{m}^\mu m^\nu),$$

$$\psi_0 = -C_{\alpha\beta\gamma\delta} l^\alpha m^\beta l^\gamma m^\delta,$$

$$\psi_1 = -C_{\alpha\beta\gamma\delta} l^\alpha n^\beta l^\gamma m^\delta,$$

$$\psi_2 = -\tfrac{1}{2}C_{\alpha\beta\gamma\delta}(l^\alpha n^\beta l^\gamma n^\delta - l^\alpha n^\beta m^\gamma \overline{m}^\delta),$$

$$\psi_3 = C_{\alpha\beta\gamma\delta} l^\alpha n^\beta n^\gamma \overline{m}^\delta,$$

$$\psi_4 = -C_{\alpha\beta\gamma\delta} n^\alpha \overline{m}^\beta n^\gamma \overline{m}^\delta.$$

The tetrad transformations which preserve the above orthogonality relations of the tetrad vectors can be decomposed into a product of the following three kinds of transformations.

(1) Null rotations with l^μ fixed:

$$l^\mu \to l^\mu,$$

$$m^\mu \to m^\mu + a l^\mu,$$

$$n^\mu \to n^\mu + a\bar{m}^\mu + \bar{a} m^\mu + a\bar{a} l^\mu.$$

(2) Null rotations with n^μ fixed:

$$n^\mu \to n^\mu,$$

$$m^\mu \to m^\mu + b n^\mu,$$

$$l^\mu \to l^\mu + b\bar{m}^\mu + \bar{b} m^\mu + b\bar{b} n^\mu.$$

(3) Tetrad rotations:

$$l^\mu \to A^{-1} l^\mu, \quad n^\mu \to A n^\mu,$$

$$m^\mu \to e^{i\theta} m^\mu.$$

(Here a and b are complex, while A and θ are real.) The behavior of the NP variables under these tetrad transformations can be computed in a straightforward manner from their definitions. Explicit formulas are given in Kinnersley's thesis[15] and by Aronson and Newman.[20]

*Work supported by the University of Maryland Center for Theoretical Physics.

[1]R. Price, Phys. Rev. D **5**, 2419 (1972); Phys. Rev. D **5**, 2439 (1972).
[2]W. Israel, Phys. Rev. **164**, 1776 (1967); Commun. Math. Phys. **8**, 245 (1968).
[3]B. Carter, Phys. Rev. Lett. **26**, 331 (1971).
[4]S. W. Hawking, Commun. Math. Phys. **25**, 152 (1972).
[5]R. M. Wald, J. Math. Phys. **13**, 490 (1972).
[6]S. W. Hawking, Phys. Rev. Lett. **26**, 1344 (1971).
[7]T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957); C. V. Vishveshwara, Phys. Rev. D **1**, 2870 (1970); F. Zerilli, Phys. Rev. D **2**, 2141 (1970); Phys. Rev. Lett. **24**, 737 (1970).
[8]R. Penrose, Riv. Nuovo Cimento **1**, 252 (1969).
[9]R. M. Wald, Ph.D. thesis, Princeton University, 1972 (unpublished). See also J. Ipser, Phys. Rev. Lett. **27**, 529 (1971).
[10]S. Chandrasekhar and J. L. Friedman, Astrophys. J. **175**, 379 (1972); Astrophys. J. **176**, 745 (1972); Astrophys. J. **177**, 745 (1972).
[11]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[12]S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972); Phys. Rev. (to be published).
[13]E. D. Fackerell and J. R. Ipser, Phys. Rev. D **5**, 2455 (1972).
[14]W. H. Press and S. A. Teukolsky (to be published).
[15]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969); Ph.D. thesis, California Institute of Technology, 1968 (unpublished).
[16]M. Demianski and E. T. Newman, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astron. Phys. **14**, 653 (1966).
[17]C. J. Talbot, Commun. Math. Phys. **13**, 45 (1969). See also R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963) and R. Lind (to be published).
[18]W. Kinnersley (unpublished); W. E. Couch and E. T. Newman, J. Math. Phys. **14**, 285 (1973). See also J. Foster and E. T. Newman, J. Math. Phys. **8**, 189 (1967).
[19]W. H. Press and S. A. Teukolsky (to be published).
[20]B. Aronson and E. T. Newman, J. Math. Phys. **13**, 1847 (1972).

Facial aspect of superposition principle in algebraic quantum theory

Etang Chen*

Department of Mathematics, University of California, Los Angeles, California 90024
(Received 30 March 1973)

A facial version of superposition principle is introduced in algebraic quantum theory; similar to the quantum logic approach it exhibits characteristic features of quantum theory. It is shown that any pair of primary states satisfy the superposition principle if and only if they are quasiequivalent. A coherent sector is then defined as a quasi-equivalent class of primary states, which holds for both pure states in quantum theory and pure thermodynamic phases in quantum statistical mechanics.

One of the basic assumptions in quantum theory is the superposition principle of quantum states. In the traditional framework of quantum mechanics, it is transcribed mathematically as the linearity of states, which is similar to the linearity of some equations of motion describing physical properties in classical mechanics. However, the physical background of superposition principles in quantum systems is extremely different from classical systems, in fact, it contains the essential properties of quantum mechanics. One cannot see such implication from the superposition principle formulated in conventional quantum mechanics. Even in the recent algebraic approach, the superposition of two pure states formulated by Roberts and Roestorff[1] does not exhibit any further properties of quantum mechanics.

On the other hand, in the quantum logic approach, superposition principle implies essential features of quantum mechanics. The superposition principle is formulated for atomic proposition by Jauch[2] as follows: For any pair of atomic propositions e_1 and e_2 with $e_1 \neq e_2$, there exists an atomic proposition e_3 such that $e_3 \neq e_1, e_3 \neq e_2$ and satisfying

$$e_1 \vee e_2 = e_1 \vee e_3 = e_2 \vee e_3.$$

An immediate consequence is that the sublattice generated by e_1, e_2 and e_3, satisfying the above relation, cannot be Boolean. Therefore, superposition principle in the above form implies the non-Boolean structure of proposition system, it is why the superposition principles is so characteristic in quantum systems. We refer to Jauch's book[2] for detailed definitions of propositions and physical contents of the above formulation for superposition principle.

Jauch's formulation is indeed the superposition for (atomic) propositions but not for states. Hence a question arises; whether there is a similar form for the superposition of states, say, in algebraic approach. Furthermore, in C^*-algebraic framework[1], the superposition is defined only for pure states; and the coherent sectors, within which superposition principle holds unrestrictedly, are also defined only for irreducible representations. It is inadequate to the states in quantum statistical mechanics, e.g., the equilibrium states in an infinite system can never be pure states at finite temperature. Therefore, another question: Is there another version of the superposition principle and the coherent sector which is adequate to both pure states and equilibrium states?

In the present note, we introduce in the algebraic approach a facial form of the superposition principle for states, which, similar to the quantum logic approach, exhibits the characteristic features of quantum systems, and which is applicable to states both in quantum theory and statistical mechanics. It is shown that two primary states satisfy the superposition principle if and only if they are quasi-equivalent. A coherent sector is then defined as a quasi-equivalent class of primary states; it is adequate to both pure states and pure thermodynamic phases. Several examples of physical interest are given to illustrate some properties of coherent sectors. We also compare our superposition principle with other versions of superposition in C^*-algebraic formulation and quantum logic approach.

Before we begin our formulation within the framework of von Neumann algebras, we recall some elementary definitions from the theory of compact convex sets.

Given a partially ordered vector space X with generating cone X^+. A subset F of X^+ is said to be a *face* if and only if it is closed under addition and multiplication by positive scalars and is such that if $x \in F$ and $y \in X$ with $0 \leq y \leq x$, then $y \in F$. A *complementary set* F' of F is the union of all faces H of X^+ such that $H \cap F = \{0\}$. F is a *split face* if the complementary set F' is a face and each element $x \in X^+$ has a unique decomposition $x = x_1 + x_2$ with $x_1 \in F$ and $x_2 \in F'$. In fact, we are interested in the faces and split faces on the predual of a von Neumann algebra, in particular, those generated by normal states.

Let $\mathfrak{M}$ be a von Neumann algebra acting on a Hilbert space $\mathfrak{H}$, and $\mathfrak{M}_*^+$ the positive cone of the predual $\mathfrak{M}_*$ of $\mathfrak{M}$. If φ is a normal state of $\mathfrak{M}$, then there is a unique projection e_φ in $\mathfrak{M}$, the support of φ, such that $\mathfrak{M}(1 - e_\varphi) = \{x \in \mathfrak{M}; \varphi(x^*x) = 0\}$. e_φ is the smallest projection in $\mathfrak{M}$ such that $\varphi(e_\varphi) = \varphi(1)$ and $\varphi_{e_\varphi} = \varphi$, where $\varphi_x(\cdot) = \varphi(x^* \cdot x)/\varphi(x^*x)$ for $x \in \mathfrak{M}$, with $\varphi(x^*x) \neq 0$.

Given two normal states φ_1, φ_2 of $\mathfrak{M}$, the norm-closure of the smallest face of $\mathfrak{M}_*^+$ containing φ_1 (resp. $\{\varphi_1, \varphi_2\}$) is denoted by $F(\varphi_1)$ [resp. $F(\varphi_1, \varphi_2)$]. Similarly, the norm-closure of the smallest split face of $\mathfrak{M}_*^+$ containing φ_1 (resp. $\{\varphi_1, \varphi_2\}$) is denoted by $H(\varphi_1)$ [resp. $H(\varphi_1, \varphi_2)$]. The norm-closed faces and split faces on $\mathfrak{M}_*^+$ can be characterized by the projections and central projections in $\mathfrak{M}$.[3,4] In fact, $F(\varphi_1)$ and $F(\varphi_1, \varphi_2)$ are characterized by e_{φ_1} and $e_{\varphi_1} \vee e_{\varphi_2}$, respectively.[3] And, $H(\varphi)$ is by the central support $c(e_\varphi)$ of e_φ.[4]

Furthermore, e_φ can be considered as the *decision effect* of the microscopic object in the quantum mechanical measurement.[3] It has been shown that the set of decision effects $\mathfrak{G}$ forms a complete orthocomplemented lattice,[3] which is similar to the proposition system in quantum logic approach. By considering e_φ as a decision effect, one can apply the same physical illustrations given by Jauch,[2] and formulate a similar version of superposition principle as follows:

For any pair of normal states φ_1, φ_2 with $e_{\varphi_1} \wedge e_{\varphi_2} = 0$ there exists a normal state φ_3 such that $e_{\varphi_3} \wedge e_{\varphi_1} = e_{\varphi_3} \wedge e_{\varphi_2} = 0$ and satisfying

$$e_{\varphi_1} \vee e_{\varphi_2} = e_{\varphi_1} \vee e_{\varphi_3} = e_{\varphi_2} \vee e_{\varphi_3}. \tag{1}$$

From the above arguments, this relation is equivalent to the following form:

$$F(\varphi_1, \varphi_2) = F(\varphi_1, \varphi_3) = F(\varphi_2, \varphi_3). \tag{2}$$

This is why we may call the above version the *facial superposition principle*. φ_3 is the superposition of φ_1 and φ_2. We note that our hypothesis is different from the proposition system, however, $e_{\varphi_1} \wedge e_{\varphi_2} = 0$ implies $e_{\varphi_1} \neq e_{\varphi_2}$. We shall see later that there is indeed an equivalent class of states satisfying the relation (1). Before showing this assertion, let us make some remarks:

Remark 1: Similar to propositions,[2] it is easy to see that the lattice generated by $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3} are not Boolean whenever they satisfy the relation (1); in fact if they were Boolean, then $e_{\varphi_3} \wedge (e_{\varphi_1} \vee e_{\varphi_2}) = (e_{\varphi_3} \wedge e_{\varphi_1}) \vee (e_{\varphi_3} \wedge e_{\varphi_2})$. This relation is not true, since by the assumption, $e_{\varphi_3} \wedge e_{\varphi_1} = 0$, $e_{\varphi_3} \wedge e_{\varphi_2} = 0$, hence the right-hand side is equal to 0. On the other hand, $e_{\varphi_3} < e_{\varphi_1} \vee e_{\varphi_2}$, the left-hand side is equal to e_{φ_3}.

Remark 2: Suppose $[e_{\varphi_1}, e_{\varphi_3}] = [e_{\varphi_2}, e_{\varphi_3}] = 0$, then $e_{\varphi_1} \vee e_{\varphi_3} = e_{\varphi_1} + e_{\varphi_3} - e_{\varphi_1} e_{\varphi_3}$, and $e_{\varphi_2} \vee e_{\varphi_3} = e_{\varphi_2} + e_{\varphi_3} - e_{\varphi_2} e_{\varphi_3}$. If they satisfy the relation (1) then by an immediate calculation $e_{\varphi_1} = e_{\varphi_2}$ which is impossible by the hypothesis. Hence, e_{φ_3} cannot commute with e_{φ_1} and e_{φ_2}, respectively. Similarly, one can show that e_{φ_1} and e_{φ_2} are not commutative.

Remark 3: If $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3} are mutually orthogonal, then $e_{\varphi_1} \vee e_{\varphi_2} = e_{\varphi_1} + e_{\varphi_2}$, $e_{\varphi_1} \vee e_{\varphi_3} = e_{\varphi_1} + e_{\varphi_3}$, and $e_{\varphi_2} \vee e_{\varphi_3} = e_{\varphi_2} + e_{\varphi_3}$. Obviously, they cannot satisfy the facial superposition principle.

Remark 4: If we consider e_φ is a decision effect, then the set of all decision effects $\mathfrak{G}$ forms a lattice.[3] Similar to the proposition system,[2] if each pair of decision effects satisfy the relation (1), then $\mathfrak{G}$ is irreducible in the sense that it is impossible to reduce $\mathfrak{G}$ into the direct union of two lattices.[2]

The above remarks show that the facial superposition principle exhibits some essential properties of quantum mechanics.

We recall that two states φ_1 and φ_2 of a C^*-algebra $\mathfrak{A}$ are quasi-equivalent (resp. disjoint) if their cyclic representations π_{φ_1} and π_{φ_2} induced by φ_1 and φ_2, respectively, are quasi-equivalent (resp. disjoint); φ is primary if $\pi_\varphi(\mathfrak{A})''$ is a factor. This notion can be transcribed in the norm-closed split faces in our case as follows; φ_1 and φ_2 are quasi-equivalent if and only if $H(\varphi_1) = H(\varphi_2)$, which is equivalent to $c(e_{\varphi_1}) = c(e_{\varphi_2})$, the central supports of e_{φ_1} and e_{φ_2}, respectively. φ_1 and φ_2 are disjoint if and only if $H(\varphi_1) \cap H(\varphi_2) = \{0\}$, which is corresponding to $c(e_{\varphi_1}) \wedge c(e_{\varphi_2}) = 0$. However, $c(e_{\varphi_1})$ and $c(e_{\varphi_2})$ commute, hence $c(e_{\varphi_1}) \wedge c(e_{\varphi_2}) = c(e_{\varphi_1})c(e_{\varphi_2}) = 0$, which implies $c(e_{\varphi_1})$ and $c(e_{\varphi_2})$ are orthogonal. We are now able to characterize those states satisfying the facial superposition principle.

Let φ_1, φ_2, and φ_3 be primary normal states with supports $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3} satisfying the relation (1). Suppose that φ_1, φ_2, and φ_3 are mutually disjoint, then by the preceding argument, $c(e_{\varphi_1})$, $c(e_{\varphi_2})$, and $c(e_{\varphi_3})$ are mutually orthogonal. Hence $c(e_{\varphi_1}) \vee c(e_{\varphi_2}) = c(e_{\varphi_1}) + c(e_{\varphi_2})$, $c(e_{\varphi_1}) \vee c(e_{\varphi_3}) = c(e_{\varphi_1}) + c(e_{\varphi_3})$. And by the definition of central support, $c(e_{\varphi_i}) > e_{\varphi_i}$ $(i = 1, 2, 3)$, thus $c(e_{\varphi_1}) + c(e_{\varphi_2}) > e_{\varphi_1} \vee e_{\varphi_2}$. It follows that

$$[c(e_{\varphi_1}) + c(e_{\varphi_2})](e_{\varphi_1} \vee e_{\varphi_2}) = e_{\varphi_1} \vee e_{\varphi_2}. \tag{3}$$

Similarly,

$$[c(e_{\varphi_1}) + c(e_{\varphi_3})](e_{\varphi_1} \vee e_{\varphi_3}) = e_{\varphi_1} \vee e_{\varphi_3}. \tag{4}$$

By the relation (1), (3) equals (4), hence

$$c(e_{\varphi_1}) + c(e_{\varphi_2}) = c(e_{\varphi_1}) + c(e_{\varphi_3}).$$

Thus, $c(e_{\varphi_2}) = c(e_{\varphi_3})$, which is impossible by the assumption that φ_2 and φ_3 are primary and disjoint. Therefore, φ_2 and φ_3 are quasi-equivalent. Similarly, φ_1 and φ_2 are quasi-equivalent. Hence φ_1, φ_2, and φ_3 belong to a quasi-equivalent class.

Conversely, if φ_1, φ_2, and φ_3 are quasi-equivalent, then $c(e_{\varphi_1}) = c(e_{\varphi_2}) = c(e_{\varphi_3})$. But, $c(e_{\varphi_1}) \vee c(e_{\varphi_2}) > e_{\varphi_1} \vee e_{\varphi_2}$, hence $c(e_{\varphi_1}) > e_{\varphi_1} \vee e_{\varphi_2}$. It follows that $1 - c(e_{\varphi_1}) < 1 - (e_{\varphi_1} \vee e_{\varphi_2})$, so that

$$[1 - (e_{\varphi_1} \vee e_{\varphi_2})][1 - c(e_{\varphi_1})] = 1 - c(e_{\varphi_1}). \tag{5}$$

Similarly

$$[1 - (e_{\varphi_1} \vee e_{\varphi_3})][1 - c(e_{\varphi_3})] = 1 - c(e_{\varphi_3}). \tag{6}$$

By assumption, $c(e_{\varphi_1}) = c(e_{\varphi_3})$, thus (5) equals (6), and $1 - (e_{\varphi_1} \vee e_{\varphi_2}) = 1 - (e_{\varphi_1} \vee e_{\varphi_3})$, hence

$$e_{\varphi_1} \vee e_{\varphi_2} = e_{\varphi_1} \vee e_{\varphi_3}.$$

Similarly, one can show that $e_{\varphi_1} \vee e_{\varphi_2} = e_{\varphi_2} \vee e_{\varphi_3}$. Thus $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3} satisfy the relation (1). Therefore we have proved the following:

Proposition 1: Let φ_1, φ_2, and φ_3 be primary normal states with supports $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3}. Then e_{φ_1}, e_{φ_2}, and e_{φ_3} satisfy the relation (1) if and only if φ_1, φ_2, and φ_3 are quasi-equivalent.

If we consider a C^*-algebra $\mathfrak{A}$ with identity, then the double dual $\mathfrak{A}^{**}$ of $\mathfrak{A}$ is a von Neumann algebra, and the dual $\mathfrak{A}^*$ of $\mathfrak{A}$ coincides with the predual of $\mathfrak{A}^{**}$. Moreover, $\mathfrak{A}$ can be regarded as a C^*-subalgebra embedded into $\mathfrak{A}^{**}$. Therefore the above formulation and result can be transferred to the states on a C^*-algebra.

From Proposition 1, it is natural to define a *coherent sector* as a quasi-equivalent class of primary states on a C^*-algebra. This definition meets the physical requirement of a coherent sector, where superposition of states can be performed unrestrictedly. Some remarks about coherent sectors are given:

Remark 5: We note that our notion of coherent sector is slightly different from the coherence of proposition system given by Jauch.[2]

Remark 6: For pure states, quasi-equivalence is coincided with unitary equivalence, hence our definition of coherent sector is equivalent to the sector in C^*-algebraic approach.[1] Moreover, Proposition 1, and hence, coherent sectors are interesting in particular for the case of equilibrium states in quantum statistical

mechanics, where pure phases are expressed mathematically as primary states.

Remark 7: Let $\mathfrak{F}$ be the set of all primary states of a C^*-algebra $\mathfrak{A}$, then the coherent sectors Q_α form a disjoint covering of $\mathfrak{F}$. Since each coherent sector Q_α corresponds to a central projection e_α in $\mathfrak{A}^{**}$, regarded as a von Neumann algebra acting on a Hilbert space $\mathcal{H}$, hence the disjointness of $\{Q_\alpha\}$ implies the orthogonality of $\{e_\alpha\}$. Therefore, the corresponding Hilbert space $\mathcal{H}$ can be decomposed into the direct sum of coherent subspaces $\mathcal{H}_\alpha = e_\alpha \mathcal{H}$, i.e., $\mathcal{H} = \oplus_\alpha \mathcal{H}_\alpha$, which amounts to a system with commutative discrete superselection rules in the conventional quantum mechanics.

There are many examples for quasi-equivalent classes of primary states in mathematical literature. We mention here only some of those having physical interest.

Example 1: Given a C^*-algebra $\mathfrak{A}$ with identity, let G be a group represented as a large group of automorphism of $\mathfrak{A}$,[5] and $\mathcal{S}^G$ be the set of all G-invariant states on $\mathfrak{A}$. If the set of extremal points of $\mathcal{S}^G$ is denoted by $E(\mathcal{S}^G)$, and $\varphi_1, \varphi_2 \in E(\mathcal{S}^G) \cap \mathfrak{F}$, then $\varphi_1, \varphi_2 \in Q_\alpha$ if and only if $\|\varphi_1 - \varphi_2\| < 2$.[6] Indeed, if $Q_\alpha^G = E(\mathcal{S}^G) \cap Q_\alpha$, then $Q_\alpha^G = \{\varphi_\alpha\}$, whenever $Q_\alpha^G \neq \emptyset$.

Example 2: Let $\mathfrak{A}$ be a C^*-algebra with net $\{\mathfrak{A}_\alpha\}$ of sub-C^*-algebras in the sense of Ref. 7, then a coherent sector Q of $\mathfrak{A}$ has the following property: For any pair $\varphi_1, \varphi_2 \in Q$ and each $\epsilon > 0$, there is an $\mathfrak{A}_\alpha$ in the net such that

$$\|\varphi_1|\mathfrak{A}_\alpha^c - \varphi_2|\mathfrak{A}_\alpha^c\| \leq \epsilon,$$

where $\mathfrak{A}_\alpha^c$ is the relative commutant of $\mathfrak{A}_\alpha$ in $\mathfrak{A}$.[7]

In the case of a quasilocal algebra $\mathfrak{A}$ of a quantum spin system, a coherent sector of states with short-range correlation has the same property as above.[8]

Example 3: For the equilibrium states on an infinite system at a given temperature, two KMS states for two different temperatures are disjoint.[9] Hence pure phases, which are primary KMS states, for different temperature belong to different coherent sectors.

Example 4: In the case of macroscopic observables in the quantum mechanical measurement as discussed by Hepp,[10] two primary states belong to two different coherent sectors whenever there exist two different expectation values of macroscopic observables in these states. For the details about coherence and classical observables, we refer to Ref. 10. We note that the notation of coherence given by Hepp is also slightly different from our coherent sector, however both of them are coincided for primary states.

As we have noted in Remark 6, our definition of coherent sectors are equivalent to the sectors of pure states given by Roberts and Roepstorff.[1] In fact, their superposition of states is defined as follows[1]: Let φ be a pure state of a C^*-algebra $\mathfrak{A}$, and $x_1, x_2 \in \mathfrak{A}$ be linearly independent such that $\varphi(x_1^* x_1) \neq 0$, $\varphi(x_2^* x_2) \neq 0$, then $\varphi_{x_1+x_2}$ is called a superposition of φ_{x_1} and φ_{x_2}, where $\varphi_x(\cdot) = \varphi(x^* \cdot x)/\varphi(x^* x)$ for $x \in \mathfrak{A}$. We shall show that this notion of superposition for primary states is also related to ours; more precisely, $\varphi_{x_1+x_2}$, φ_{x_1}, and φ_{x_2} all belong to the same coherent sector. This can be seen from the following.

Proposition 2: If φ is a primary state, then φ_x and φ belong to the same coherent sector.

Proof: At first we have to show that φ_x is also primary. Let us recall a partial ordering on the state space of $\mathfrak{A}$; two states φ and ψ on $\mathfrak{A}$ are in the relation $\varphi \lesssim \psi$ whenever $\tilde{\varphi} \subseteq \tilde{\psi}$, where $\tilde{\varphi}$ is the norm-closure of the set $\{\varphi_x; x \in \mathfrak{A}\}$. And, $\varphi \lesssim \psi$ if and only if $\varphi \in \tilde{\psi}$. Moreover, φ is primary if and only if $\lesssim$ totally orders $\tilde{\varphi}$. We refer to Emch's book[11] for a detailed exposition of this ordering defined by Kadison. Applying to our special case: $\varphi_x \in \tilde{\varphi}$, hence $\lesssim$ also totally orders $\tilde{\varphi}_x$, so that φ_x is also primary.

Let $H(\varphi)$ and $H(\varphi_x)$ be the norm-closures of the smallest split faces containing φ and φ_x, respectively. As $H(\varphi_x) \subseteq H(\varphi)$, $H(\varphi_x) \cap H(\varphi) \neq \{0\}$, hence φ_x and φ are not disjoint. But, φ and φ_x are primary, by assumption and the above argument, therefore φ and φ_x are quasi-equivalent, which implies that φ_x and φ belong to the same coherent sector.

From Proposition 2, it follows that $\varphi_{x_1+x_2}, \varphi_{x_1}, \varphi_{x_2}$ all belong to the coherent sector containing φ.

We give a final remark, to compare our version of superposition with another one defined by Varadarajan in quantum logic approach.[12] A lattice $\mathfrak{L}$ with zero element 0 and unit element 1, equipped with an orthocomplementation $a \to a'$, is called a *logic*, if (i) for any countably infinite sequence $a_1, a_2, \ldots$ of elements of $\mathfrak{L}$, $\vee_n a_n$ and $\wedge_n a_n$ exist in $\mathfrak{L}$, (ii) if $a_1, a_2 \in \mathfrak{L}$ and $a_1 < a_2$, there exists an element $a_3 \in \mathfrak{L}$ such that $a_3 < a_1'$ and $a_3 \vee a_1 = a_2$. Furthermore, a state on a logic $\mathfrak{L}$ is a real-valued function $p: a \to p(a)$ for $a \in \mathfrak{L}$, such that (i) $0 \leq p(a) \leq 1$ for all $a \in \mathfrak{L}$, (ii) $p(0) = 0$, $p(1) = 1$, and (iii) if $a_1, a_2, \cdots$ is a sequence of mutually orthogonal elements of $\mathfrak{L}$ and $a = \vee_n a_n$, then $p(a) = \sum_n p(a_n)$.

It has been shown[3] that the set of all decision effects $\mathfrak{G}$ is a logic. It is not difficult to see that a normal state restricting on $\mathfrak{G}$ satisfying the above conditions of a state on a logic, indeed (iii) holds in particular for a normal state.

Let p_1 and p_2 be two states on a logic $\mathfrak{L}$. A state p_0 on $\mathfrak{L}$ is called a superposition of p_1 and p_2 if the following property is satisfied[12]:

$$a \in \mathfrak{L}, p_1(a) = p_2(a) = 0 \quad \text{implies } p_0(a) = 0. \tag{7}$$

Given three normal states φ_1, φ_2, and φ_3 with supports $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3}, respectively, if $e_{\varphi_1}, e_{\varphi_2}$, and e_{φ_3} satisfy the relation (1), then $e_{\varphi_3} < e_{\varphi_1} \vee e_{\varphi_2}$ which implies

$$F(\varphi_3) \subseteq F(\varphi_1, \varphi_2). \tag{8}$$

If we denote by $\mathfrak{N}$ (resp. $\mathfrak{M}_+$) the set of all normal states (resp. the positive portion) of a von Neumann algebra $\mathfrak{M}$. Let $\mathfrak{D}$ (resp. $\mathcal{S}$) be a subset of $\mathfrak{N}$ (resp. $\mathfrak{M}_+$), and $\mathfrak{D}^\perp = \{x \in \mathfrak{M}_+; \varphi(x) = 0 \text{ for all } \varphi \in \mathfrak{D}\}$ $\mathcal{S}^\perp = \{\varphi \in \mathfrak{N}; \varphi(x) = 0 \text{ for all } x \in \mathcal{S}\}$. Then $F(\varphi_3) = \{\varphi_3\}^{\perp\perp}$, and $F(\varphi_1, \varphi_2) = \{\tau\}^{\perp\perp}$, where $\tau = \lambda\varphi_1 + (1 - \lambda)\varphi_2$ with $\lambda \in (0, 1)$.[3] Hence, the relation (8) implies $\{\tau\}^\perp \subseteq \{\varphi_3\}^\perp$, as $\{\varphi_1, \varphi_2\}^\perp \subseteq \{\tau\}^\perp$. Therefore,

$$\{\varphi_1, \varphi_2\}^\perp \subseteq \{\varphi_3\}^\perp. \tag{9}$$

It is easy to see that for those $e \in \mathfrak{G}$ such that $\varphi_1(e) = \varphi_2(e) = 0$, then $e \in \{\varphi_1, \varphi_2\}^\perp$, hence by (9), $\varphi_3(e) = 0$, which satisfies the condition (7). Therefore we have shown that our superposition φ_3 is a superposition of φ_1 and φ_2 on logic $\mathfrak{G}$, whenever φ_3 is restricted on $\mathfrak{G}$.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor M. Takesaki and the Department of Mathematics at UCLA for their kind hospitality during the preparation of this work. The author is also grateful to the referee for calling my attention to the paper by G. G. Emch and C. Piron [J. Math. Phys. 4, 469 (1963)].

*Present address: Department of Mathematics, University of Rochester, Rochester, N.Y. 14627.

[1]J. E. Roberts and G. Roepstorff, Commun. Math. Phys. **11,** 321 (1969).
[2]J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[3]E. Chen, J. Math. Phys. **12,** 2364 (1971).
[4]C. M. Edwards, Commun. Math. Phys. **20,** 26 (1971); Commun. Math. Phys. **24,** 260 (1972).
[5]E. Størmer, Commun. Math. Phys. **5,** 1 (1967).
[6]E. Størmer, J. Funct. Anal. **3,** 48 (1969).
[7]R. Haag, R. V. Kadison, and K. Kastler, Commun. Math. Phys. **16,** 13 (1970).
[8]O. E. Landford and D. Ruelle, Commun. Math. Phys. **13,** 194 (1969).
[9]M. Takesaki, Commun. Math. Phys. **17,** 33 (1970).
[10]K. Hepp, Helv. Phys. Acta **45,** 237 (1972).
[11]G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley–Interscience, New York, 1972).
[12]V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, N.J., 1968), Vol. I.

The mathematical reflection problem

B. Chakraborty

Department of Mathematics, Jadavpur University, Calcutta-32, India
(Received 30 March 1972)

The one-dimensional mathematical problem of finding the scattering amplitude for weak reflection by slowly and continuously varying plane layered media was defined and solved by an extension of the WKB method by Chakraborty. A further study of the solution of this problem near the zeros of the parameters is considered in this note. We find quantities proportional to the energy density and energy flux density. In the last section some speculative remarks about possible lines of future work are made.

I. INTRODUCTION

The one-dimensional mathematical problem of finding the scattering amplitude for weak reflection by slowly and continuously varying plane layered media was defined and solved by an extension of the WKB method by Chakraborty.[1] The investigation should be of interest to a wide range of physicists dealing with certain mathematical aspects of wave propagation. In the present note we further extend the study of the pair of first-order differential equations

$$\alpha^2 u + iv' = 0, \tag{1.1}$$

$$\beta^2 v + iu' = 0, \tag{1.2}$$

where α^2, β^2 are the slowly varying parameters and a primed quantity means derivative of the corresponding unprimed quantity with respect to the independent variable ξ. The solution of these equations in the neighborhood of the singularities of the solutions of one of u and v or of both will now be considered. For the physical problem of propagation of waves through a slowly varying medium, the region containing these singularities is identified with the region where the direction of propagation is parallel to the layers of equal value of the slowly varying parameter. A special case of this solution can be identified with the solution of the TM field equations (Landau and Lifshitz[2]) for slowly varying dielectric media in the neighborhood of the zero of the dielectric constant.

Broer and Van Vroonhoven[3] have defined two quantities which can be taken to be proportional to the energy density and the energy flux density in any actual physical problem. These two quantities were defined in terms of the field variable ψ which satisfies the second-order one-dimensional wave equation

$$\frac{\partial^2\psi}{\partial\xi^2} - \frac{1}{c^2(\xi)}\frac{\partial^2\psi}{\partial t^2} = 0. \tag{1.3}$$

We have extended the ideas of Broer and Van Vroonhouen by finding the values of energy density and energy flux density in terms of the two field variables u and v which occur in our pair of first-order equations (1.1), (1.2). But we have found that the ideas of Broer and Van Vroonhouen cannot be uniquely extended to the one-dimensional Klein—Gordon equation. The Klein—Gordon equation is usually obtained from the self-consistent system of plasma field equations having first-order partial deviatives with respect to time and space which yield for $\partial/\partial t = -iw$ the pair of equations of the type (1.1), (1.2).

We define some three-dimensional reflection problems. The n-dimensional generalized multivariable reflection problem in which n field variables and their n first order derivatives are linearly related by n linear simultaneous algebraic equations is stated.

II. SOLUTION NEAR A ZERO OF THE PARAMETERS

The second-order equations for u and v obtained from (1.1), (1.2) are

$$u'' - (\log\beta^2)'u' + \alpha^2\beta^2 u = 0, \tag{2.1}$$

$$v'' - (\log\alpha^2)'v' + \alpha^2\beta^2 v = 0. \tag{2.2}$$

To show the behavior of these equations in the neighborhood of a zero of $\alpha^2\beta^2$ on traditional lines, we put

$$\alpha^2\beta^2 = -A_0(\xi - \xi_0), \tag{2.3}$$

$$\rho = A_0^{1/3}(\xi - \xi_0) \tag{2.4}$$

and get

$$[D^2 - \beta(1/\beta)'' - \rho](u/\beta) = 0,$$

$$[D^2 - \alpha(1/\alpha)'' - \rho](v/\alpha) = 0, \tag{2.5}$$

where $D = d/d\rho$, $x' = dx/d\rho$, etc. The value of ρ obtained from the equation $\rho + \beta(1/\beta)'' = 0$ is not in general identical to ρ obtained from $\rho + \alpha(1/\alpha)'' = 0$ because α and β do not necessarily have identical functional dependence on ρ or on ξ. Hence the point at which the direction of propagation of u is perpendicular to the ξ axis is not identical to the similar point for v. In other words, the position of the mirror for u does not coincide with that for v if α and β are slowly varying with respect to ξ. Budden (Ref. 4, Sec. 9.2) pointed out this peculiar behavior in the Maxwell equations in a medium in which the dielectric constant varies slowly in space. He commented that the solutions do not represent progressive waves because in a progressive wave field the variation in ξ is contained in the factor $\exp\varphi(\xi)$ which should be same for all of the associated field variables. The physically interesting point is that the motion of a surface of constant u is not identical with that for v.

The pair of equations (1.1), (1.2) for electromagnetic wave propagation in slowly varying dielectric medium (Bremmer[5]), in plasma (Chakraborty, 1970)[6] or for acoustic type of propagation in slowly varying warm plasma (chakraborty[7]) are such that α^2, β^2 are connected in a specific way, for example, in one or the other of the following forms

$$\beta^2 = a_0^2 + \alpha^2, \quad \beta^2 = b_0^2\alpha^2, \quad \alpha^2 = \alpha_0^2 - \frac{\alpha_1^2}{\beta^2}, \tag{2.6}$$

where a_0^2, b_0^2, α_0^2, α_1^2 are constant. We consider the relation

$$\alpha^2=\alpha_0^2-\alpha_1^2/\beta^2 \tag{2.7}$$

for an interesting solution of (1.1), (1.2) which should not be identified with the material given beneath equation (2.2). In this case, the equation $\rho+\alpha(1/\alpha)''=0$ yields the equation

$$\rho+\beta\left(\frac{1}{\beta}\right)''=\rho\left(2-\frac{\alpha_0^2\beta^2}{\alpha_1^2}\right)-\frac{\beta'^2}{\beta^2}\frac{(\alpha_0^2+2\alpha_1^2/\beta^2)}{\alpha_0^2-\alpha_1^2/\beta^2},$$

which is in general different from the equation $\rho+\beta(1/\beta)''=0$ for a wide range of continuous functions β of ρ. Writing (2.1) as

$$(u'/\beta^2)'+\alpha^2u=0, \tag{2.8}$$

putting $\beta^2=a(\xi-\xi_0)$, α^2 from (2.7) into (2.8) we get

$$u''-u'/\rho+(\alpha_0^2a\rho-\alpha_1^2)u=0, \tag{2.9}$$

where now $\rho=\xi-\xi_0$, $x'=dx/d\rho$, etc. This equation for u is identical in form to the equation for the induced magnetic field H of Landau and Lifshitz (Ref. 2, p. 286, Problem 1) [which is Hy of the TM field of Chakraborty (1970)]. Following Landau and Lifshitz, therefore, for $\alpha_1\neq0$ the two independent solutions,

$$u_1(\rho)=\rho^2+\alpha_1^2\rho^4/8-a\alpha_0^2\rho^5/15+\cdots, \tag{2.10}$$

$$u_2(\rho)=u_1(\rho)\log(\alpha_1\rho)+2/\alpha_1^2-2\alpha_0^2a/3\alpha_1^2\rho^3+\cdots. \tag{2.11}$$

As $\rho\to0$ we find that $u_1\to0$, $u_2\to2/\alpha_1^2=u_0$ (say). Our v is identical to E_x of Landau and Lifshitz and so

$$v\approx-(i\alpha_1^2u_0/a)\log(\alpha_1\rho). \tag{2.12}$$

For oblique incidence α_1 corresponds to the component of the wave vector in the plane of incidence and parallel to the layers of equal value of the slowly varying parameters (see Ref. 2, Problem 1). When $\alpha_1^2=0$ the solutions (2.10) to (2.12) do not hold good because the wave field equations for oblique incidence cannot be identified with the model equations (1.1) and (1.2). In this case the relation (2.7) gives

$$(\alpha^2-\alpha_0^2)\beta^2=0.$$

If $\alpha^2=\alpha_0^2$, $\beta^2\neq0$, simple solutions of (2.1) and (2.2) can be easily found out because they correspond to the normally incident field solutions. In the neighborhood of $\beta^2=0$ the original equations can be easily integrated.

When $\beta^2=a_0^2\alpha^2$, up to lowest nonvanishing order we have

$$\alpha^2\beta^2\approx a_0^2a(\xi-\xi_0).$$

This case has been treated above in Eqs. (2.3) to (2.5).

When $\beta^2=b_0^2\alpha^2$ both α^2 and β^2 have their zeros at the same point and so near $\xi=\xi_0$

$$\alpha^2\beta^2=\text{const}(\xi-\xi_0)^2.$$

The consequent field equations can not be reduced to Stokes equations of the type (2.5) by the substitution (2.4). For this reason this case is not discussed here and may be considered elsewhere in future.

III. THE CONSERVATION CONDITION

The exact solutions of (1.1), (1.2) should satisfy the conservation law

$$\frac{\partial W}{\partial t}+\text{div}\mathbf{S}=0, \tag{3.1}$$

where W and S are density and flux of some conserved quantity which can be the energy density. We note that in place of W and S quantities which are proportional to them also can be used in (3.1). The expressions for the energy density and the energy flux density are ultimately proportional to linear combination of squares of the solutions for u and v. Broer and Van Vroonhouen[3] have considered these questions in general terms for solutions of the second order wave equation (1.3). They have defined W and S as

$$W=\frac{1}{2}\left(\frac{1}{c^2}\frac{\partial\psi^*}{\partial t}\frac{\partial\psi}{\partial t}+\frac{\partial\psi}{\partial\xi}\frac{\partial\psi^*}{\partial\xi}\right), \tag{3.2}$$

$$S=-\frac{1}{2}\left(\frac{\partial\psi^*}{\partial t}\frac{\partial\psi}{\partial\xi}+\frac{\partial\psi^*}{\partial\xi}\frac{\partial\psi}{\partial t}\right), \tag{3.3}$$

where ψ^* is the complex conjugate of ψ. The three-dimensional generalizations of these quantities can be defined as

$$W=\frac{1}{2}\left(\frac{1}{c^2}\frac{\partial\psi}{\partial t}\frac{\partial\psi^*}{\partial t}+(\text{grad}\psi\text{grad}\psi^*)\right),$$

$$\mathbf{S}=-\frac{1}{2}\left(\frac{\partial\psi^*}{\partial t}\text{grad}\psi+\frac{\partial\psi}{\partial t}\text{grad}\psi^*\right).$$

The form of the positive definite W shows that it is independent of the time variable because ψ can be written as a product of a function of position multiplied by $\exp(-i\omega t)$. Hence in one-dimensional space problems S must be a constant. The ratio S/W is the group velocity of propagation.

Denoting $\partial x/\partial t$ by $\dot{x}$, $\partial x/\partial\xi$ by x', we get

$$W\mp S/c=\frac{1}{2}\left(\frac{\dot{\psi}^*}{c}\pm\psi^{*\prime}\right)\left(\frac{\dot{\psi}}{c}\pm\psi'\right), \tag{3.4}$$

where ψ satisfies Eq. (1.3).

The equation for u obtained from (2.1) can be written in the normal form

$$(qu)''+(qu)(1-q''/q)=0, \tag{3.5}$$

where $q=(\alpha/\beta)^{1/2}$, $x'=dx/d\rho$, $\rho=\int\alpha\beta d\xi$.

Putting $\psi=(qu)$ we therefore get

$$S=\left((qu)\frac{\partial}{\partial\rho}(qu)^*-(qu)^*\frac{\partial}{\partial\rho}(qu)\right)\frac{i\omega}{2}$$

$$=\frac{\omega}{2}(uv^*+u^*v),$$

$$2W\mp\frac{2S}{c}=uu^*\left(\frac{\alpha}{\beta}-\frac{\alpha''\beta-\alpha\beta''}{2\alpha^2\beta^2}+\frac{(\alpha'\beta-\alpha\beta')^2}{\alpha^3\beta^5}\right)$$

$$+\frac{i(\alpha'\beta-\alpha\beta')(u^*v-uv^*)}{2\alpha^2\beta^2}+\frac{vv^*\beta}{\alpha}$$

$$\mp\frac{\omega(uv^*+u^*v)}{c},$$

where $\alpha'=d\alpha/d\xi$, etc. In this way the physically significant quantities W and S are defined in terms of u, v.

IV. SOME GENERAL REMARKS

The more general pair of first order equations (1.1), (1.2) then are two linearly independent linear combinations equated to zero of the four quantities u, v, u', v', for example, the pair of equations

$$\alpha^2 u + iv' + ia_1^2 u' + b_1^2 v = 0, \tag{4.1}$$

$$\beta^2 v + iu' + ia_2^2 v' + b_2^2 u = 0, \tag{4.2}$$

where the parameters α^2, β^2, a_1^2, b_1^2, a_2^2, b_2^2 must be slowly varying functions of ξ if the two equations are nonhomogeneous.

In the case of two semi-infinite uniform media on the two sides of the plane $\xi = \xi_0$ the boundary conditions satisfying these equations can be written as

$$v_1 + a^2_{11} u_1 = v_2 + a^2_{12} u_2, \tag{4.3}$$

$$u_1 + a^2_{21} v_1 = u_2 + a^2_{22} v_2, \tag{4.4}$$

where a_{11}^2 is the value of a_1^2 on $\xi > \xi_0$, a_{12}^2 its value on $\xi < \xi_0$ and similarly a_{21}^2, a_{22}^2 are the values of a_2^2 on the two sides of $\xi = \xi_0$. Since a_1^2, a_2^2 can have independent values the solution of this problem should be a two parameter boundary value problem.

The boundary conditions imposed on the solution of the Eqs. (4.1) and (4.2) are derived by integrating them over an infinitesinal distance ϵ surrounding the point $\xi = \xi_0$ and requiring that the integrals involving u and v approach zero as $\epsilon \to 0$. It is thus not necessary to restrict the boundary conditions only to the case of two semi-infinite uniform media, that is, one only requires nonsingular behavior of α^2, β^2, b_1^2, and b_2^2 on both sides of the interface, but there can be inhomogeneity and also as many interfaces as one wishes.

We can even formulate the problem of solution of the n field variables u, v, $w, \cdots$ from n linear combination equated to zero of the $2n$ field variables u, v, $w, \cdots u'$, v', $w', \cdots$. The solution of these equations in the sense conceived earlier by the author[1] may be called the solution of the generalized multivariable reflection problem.

Instead of the one-dimensional pair of first order equations (1.1), (1.2), we can conceive of solution of the more general three-dimensional pair of equations

$$\alpha^2 \mathbf{A} + i\,\mathrm{grad}\varphi = 0, \quad \beta^2 \varphi + i\,\mathrm{div}\mathbf{A} = 0$$

or the pair of equations

$$\alpha^2 \mathbf{A} + i\,\mathrm{curl}\mathbf{B} = 0, \quad \beta^2 \mathbf{B} + i\,\mathrm{curl}\mathbf{A} = 0,$$

where A, B are vector fields, φ a scalar field, α^2, β^2 are the slowly varying parameters.

The one dimensional Klein—Gordon equation

$$\frac{\partial^2 \psi}{\partial \xi^2} - \frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} - \frac{\omega_0^2}{c^2}\psi = 0$$

is usually obtained from the plasma field equations having first-order partial derivatives with respect to time and space which yield for $\partial/\partial t = -iw$ the pair of first-order equations of the type (1.1), (1.2). This equation can be written as

$$\frac{\partial^2 \psi}{\partial \xi^2} - \frac{1}{c^2}\frac{\partial}{\partial T}\frac{\partial \psi}{\partial T^*} = 0,$$

where $\partial/\partial T = \partial/\partial t + iw_0$, $\partial/\partial T^* = \partial/\partial t - iw_0$. We can then define the quantities

$$2W = \frac{1}{c^2}\frac{\partial \psi}{\partial T}\frac{\partial \psi^*}{\partial T^*} + \frac{\partial \psi^*}{\partial \xi}\frac{\partial \psi}{\partial \xi},$$

$$-2S = \frac{\partial \psi^*}{\partial T^*}\frac{\partial \psi}{\partial \xi} + \frac{\partial \psi^*}{\partial \xi}\frac{\partial \psi}{\partial T}$$

and obtain

$$W = \frac{(w - w_0)^2}{2C^2}\psi\psi^* + \psi^{*\prime}\psi',$$

$$S = \frac{i(w - w_0)}{2}(\psi\psi^{*\prime} - \psi^*\psi'),$$

$$I = \frac{S}{w - w_0} = \frac{i}{2}(\psi\psi^{*\prime} - \psi^*\psi').$$

But this correspondence is not unique because we can analogously define

$$2W = \frac{1}{C^2}\frac{\partial \psi}{\partial T^*}\frac{\partial \psi^*}{\partial T} + \frac{\partial \psi^*}{\partial \xi}\frac{\partial \psi}{\partial \xi},$$

$$-2S = \frac{\partial \psi^*}{\partial T}\frac{\partial \psi}{\partial \xi} + \frac{\partial \psi^*}{\partial \xi}\frac{\partial \psi}{\partial T^*},$$

and get

$$W = \frac{(w - w_0)^2}{C^2}\psi\psi^* + \psi^{*\prime}\psi',$$

$$S = \frac{i(w + w_0)}{2}(\psi\psi^{*\prime} - \psi^*\psi'),$$

$$I = \frac{S}{w + w_0}.$$

Both the definitions of W and S yield the results of Broer and Van Vroonhoven for $w_0 = 0$.

Convergence of the Bremmer (1949) solution is discussed by these authors and earlier by other authors. In this connection we can say that the purely mathematical question of the boundedness and convergence of the sequence of solutions developed by the extended WKB method by Chakraborty[1] remains open.

In conclusion, the author acknowledges his indebtedness to the referee for taking unusual interest in this note and in the preceding paper, and for making some useful suggestions.

[1]B. Chakraborty, J. Math. Phys. **14**, 188 (1973).
[2]L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1960).
[3]L. J. F. Broer and J. B. Van Vroonhoven, Physica (Utr.) **52**, 441 (1971).
[4]K. G. Budden, *Radio Waves in the Ionosphere* (Cambridge U. P., Cambridge, 1961).
[5]H. Bremmer, Physica (Utr.) **15**, 593 (1949).
[6]B. Chakraborty, J. Math. Phys. **11**, 2570 (1970).
[7]B. Chakraborty, "Radiation by perturbation in slowly varying warm plasma," (unpublished).

The field of real numbers in axiomatic quantum mechanics

M. J. Maczyński

Institute of Mathematics, Technical University, Warsaw, Poland
(Received 13 April 1972)

Consider a system of states and observables based on the logic $L(V, D)$ of all ϕ-closed subspaces of a vector space V over a division ring D with a definite Hermitian form ϕ. Gudder and Piron have shown that if the system admits a special kind of observable called the smooth maximal observable, then the division ring D must be an extension of the real field. In this paper it is shown that the same conclusion can be obtained from weaker and more physical assumptions which are in principle experiment verifiable.

I. INTRODUCTION

Jauch and Piron[1] have shown that physical arguments can be found to deduce that the collection of propositions for a physical system forms a complete, atomic, semimodular, orthomodular lattice L called the logic of the system. In a previous paper[2] we have shown that these arguments can be given an exact mathematical form with the help of the so-called probability function p. Namely, we have given a list of sufficient and necessary conditions (called axioms) to be satisfied by a map $p: \mathscr{O}\times S\times B(R)\to[0,1]$ (where $\mathscr{O}$ and S are interpreted, respectively, as the set of all observables and the set of all states of the system, and $p(A,\alpha,E)$ is the probability that a measurement of observable A for the system in state α will lead to a value in a Borel set E on the real line) in order that p be representable in the form $p(A,\alpha,E)=\varphi_\alpha\mu_A(E)$, where μ_A is an L-valued measure and φ_α is a probability measure on L, L being some (fixed for the whole system) irreducible complete orthomodular atomistic lattice with the covering property, of length greater than four.[3] It was shown by Mac Laren[4] and Piron[5] that this type of lattice can be identified with the lattice $L(V,D)$ of all ϕ-closed subspaces of a vector space V over a division ring D with a definite Hermitian form ϕ. To pass to the usual Hilbert space formalism of quantum mechanics one has to assume that D is the field of complex numbers (in some theories the field of reals and the field of quaternions have also been used). There remains an open question: What physical arguments can be given to support the assumption that D is the field of complex numbers? This problem has been partly solved by Gudder and Piron[6] who showed that if the system admits a special kind of observable (called smooth maximal observable), then the division ring D must contain a subring isomorphic to the real field. Since the only finite extensions of the reals are reals, complexes, and quaternions, this result strongly supports our selection for D as one of these fields. However, not all the conditions given by Gudder and Piron can be termed as physically basic or experimentally verifiable. Our point of view is that conditions that can be termed as physically basic should be expressible in terms of the probability function $p(A,\alpha,E)$, since this function constitutes a direct link between the theory and experimental practice. In other words, every statement expressed in terms of p can be, at least in principle, experimentally verified. Some of the assumptions of Gudder and Piron do not satisfy this requirement. For example, they assume that every probability measure on $R(X)$ [where $R(X)$ is the range of an observable X] has a unique extension to a pure state. The range of an observable is not an experimentally accessible object and the more so is an arbitrary probability measure on $R(X)$. The same objection can be raised about the second assumption that arbitrary signed measures on $R(X)$ satisfy some condition. Physically accessible objects are states, observables, and the probability function p, as well as notions derived from them [e.g., a probability measure induced on $B(R)$ by an observable and a state, but not an arbitrary probability measure on $B(R)$]. Most ideally, all assumptions about a physical theory should be expressible in such terms.

In this paper we would like to show that the main conclusion of Gudder and Piron's theorem can be obtained under weaker and more physical assumptions satisfying the above discussed requirement that all the involved notions are derived from the probability function with the help of states and observables. We will not only obtain the same conclusion under weaker assumptions, but also the proof of the theorem will be simplified.

II. BASIC DEFINITIONS

Let $\mathscr{O}$ be the set of all observables and S the set of all states associated with a fixed physical system F. For $A\in\mathscr{O}$, $\alpha\in S$, and a Borel set E on the real line, let $p(A,\alpha,E)$ be the probability that a measurement of A in state α will lead to a value in E. We assume that the function $p: \mathscr{O}\times S\times B(R)\to[0,1]$ [where $B(R)$ is the Borel σ algebra of the real line] satisfies the axioms listed in a paper of the author,[7] so there is a vector space V over a division ring D with a definite Hermitian form such that each $A\in\mathscr{O}$ corresponds uniquely to an L-valued measure μ_A and each $\alpha\in S$ corresponds uniquely to a probability measure m_α on L, and we have $p(A,\alpha,E)=m_\alpha\mu_A(E)$ (here L denotes the σ-orthocomplemented lattice of all ϕ-closed subspaces of V and is called the logic of the system). In the sequel for simplicity we shall identify an observable A with the L-valued measure μ_A and a state α with the probability measure m_α on L writing $\mu_A=A$ and $m_\alpha=\alpha$.

According to the construction described in the author's paper,[8] each member of $L=L(V,D)$ can be identified with an equivalence class $|(A,E)|=\{(B,F): B\in\mathscr{O}, F\in B(R), p(A,\alpha,E)=p(B,\alpha,F)$ for every $\alpha\in S\}$. An element $|(A,E)|$ is an atom of L if it is not equivalent to (A,ϕ), i.e., if there exists $\alpha\in S$ such that $p(A,\alpha,E)\neq 0$, and if for every (B,F) not equivalent to (B,ϕ) we have either $p(B,\alpha,F)=p(A,\alpha,E)$ for all $\alpha\in S$ [then $|(B,F)|=|(A,E)|$], or else there is $\alpha\in S$ such that $p(B,\alpha,F)>p(A,\alpha,E)$. Thus the definition of an atom of L can be expressed in physical terms and we may use it in further construction.

A state $\alpha \in S$ is said to be pure if

$$p(A,\alpha,E)=\sum_{i=1}^{\infty} c_i p(A,\alpha_i,E)$$

for all $A \in \mathcal{O}$, $E \in B(R)$ implies $c_i = 1$ for some i (we assume $c_i \geq 0$, $\Sigma c_i = 1$). We shall assume that our system of states based on logic L satisfies the so-called Gleason axiom. Namely, we assume that for each pure state α there is a unique atom $|(A,E)| \in L$ such that $p(A,\alpha,E) = 1$. We shall denote this atom by $a(\alpha)$.

Let M be a set of pure states $M \leq S$ and let A be a fixed observable. Each pure state $\alpha \in M$ determines a probability measure α^A on $B(R)$, namely $\alpha^A(E) = p(A,\alpha,E)$ for $E \in B(R)$. Assume that the observable A has the following two properties with respect to M: (1) $p(A,\alpha_1,E) = p(A,\alpha_2,E)$ for α_1, $\alpha_2 \in M$ and for all $E \in B(R)$ implies $\alpha_1 = \alpha_2$ (i.e., $\alpha_1^A = \alpha_2^A$ for α_1, $\alpha_2 \in M$ implies $\alpha_1 = \alpha_2$), and (2) for every α_1, $\alpha_2 \in M$ and every c_1, $c_2 \geq 0$, $c_1 + c_2 = 1$, there exists $\alpha \in M$ such that $p(A,\alpha,E) = c_1 p(A,\alpha_1,E) + c_2 p(A,\alpha_2,E)$ for all $E \in B(R)$ (i.e., the set $\{\alpha^A : \alpha \in M\}$ is convex). Note that conditions (1) and (2) are expressed in the admissible language of the function p.

We say that two pure states α_1, α_2 are singular with respect to A, $\alpha_1 \perp^A \alpha_2$, if there exist Borel sets E_1, $E_2 \in B(R)$, $E_1 \cap E_2 = \phi$, $E_1 \cup E_2 = R$, such that $p(A,\alpha_1,E_1) = p(A,\alpha_2,E_2) = 0$. If α_1, α_2 are singular pure states, we can form, for any c_1, $c_2 \geq 0$, $c_1 + c_2 = 1$, a signed measure $c_1\alpha_1^A - c_2\alpha_2^A$. Such signed measures are said to be induced by states.

If there are a set of pure states M and an observable A with properties (1) and (2), we can associate with every signed measure φ induced by states from M an atom $a(\varphi) \in L$ as follows:

If $\varphi = \pm\alpha^A$, then $a(\varphi) = a(\alpha)$.

If $\varphi = c_1\alpha_1^A - c_2\alpha_2^A$, $\alpha_1 \perp^A \alpha_2$, $c_1, c_2 > 0$,

$c_1 + c_2 = 1$, then $a(\varphi) = [a(\alpha_1) \vee a(\varphi^*)] \wedge a(\varphi^*)'$,

where $\varphi^* = c_2\alpha_1^A + c_1\alpha_2^A$.

Observe that by (1) α^A uniquely determines α and by (2) there is a unique α such that $\alpha^A = \varphi^*$, so $a(\varphi^*)$ is well defined. It follows from a lemma proved by Gudder and Piron[9] that for any atoms $p \neq q \in L$, $[p \vee q] \wedge q'$ is an atom, so $a(\varphi)$ is always an atom.

Let μ be a fixed measure on $B(R)$. In most applications it will be the Lebesgue measure or some discrete probability measure. We say that a pure state φ is absolutely continuous relative to μ with respect to an observable A, $\varphi \ll^A \mu$, if φ^A is absolutely continuous relative to μ ($\varphi^A \ll \mu$). A signed measure $\varphi = c_1\alpha_1^A - c_2\alpha_2^A$ is absolutely continuous relative to μ, $\varphi \ll^A \mu$, if both $\alpha_1 \ll^A \mu$ and $\alpha_2 \ll^A \mu$. If M is a set of pure states, the set of all signed measures induced by states in M which are absolutely continuous relative to μ with respect to an observable A will be denoted by M_μ^A.

Observe that the notion of absolute continuity relative to a measure μ with respect to an observable A can be expressed in terms of the probability function p, since $\varphi \ll^A \mu$ means that $\mu(E) = 0$ implies $p(A,\varphi,E) = 0$.

If $\varphi \ll^A \mu$, then $d\varphi/d\mu = d\varphi^A/d\mu$ denotes the Radon–Nikodym derivative of φ^A with respect to μ. We have $(d\varphi/d\mu) \in L_1(R,\mu)$ and it is easy to see that an experimental procedure can be devised to calculate the values of the function $d\varphi/d\mu$. Consequently, the function $d\varphi/d\mu$ also belong to our admissible language and may be used to formulate assumptions about the considered system.

Finally, we assume the following convention. If $r \in R$ is a real number (or a real-valued function), we define $r^{1/2} = r^{1/2}$ for $r \geq 0$ and $r^{1/2} = -(-r)^{1/2}$ for $r < 0$. For example, if $\varphi = c_1\alpha_1^A - c_2\alpha_2^A (\alpha_1 \perp^A \alpha_2, c_1, c_2 > 0, c_1 + c_2 = 1)$, then $\varphi^{1/2} = (c_1\alpha_1^A)^{1/2} - (c_2\alpha_2^A)^{1/2}$. Note that $(\varphi^{1/2})^2 = c_1\alpha_1^A + c_2\alpha_2^A = |\varphi|$.

III. THE THEOREM

We can now state and prove a modified version of Gudder and Piron's theorem.

Theorem: Assume that the system of observables $\mathcal{O}$ and states S based on a logic $L = L(V,D)$ satisfies the Gleason axiom and admits a set of pure states $M \leq S$ and an observable A with the following properties:

(1) If for α_1, $\alpha_2 \in M$ $p(A,\alpha_1,E) = p(A,\alpha_2,E)$ for all $E \in B(R)$, then $\alpha_1 = \alpha_2$.

(2) For every α_1, $\alpha_2 \in M$ and every $c_1, c_2 \geq 0$, $c_1 + c_2 = 1$, there is $\alpha \in M$ such that $p(A,\alpha,E) = c_1 p(A,\alpha_1,E) + c_2 p(A,\alpha_2,E)$ for all $E \in B(R)$.

(3) Let φ_1, φ_2 be any distinct ($\varphi_1 \neq \pm\varphi_2$) signed measures induced by states in M which are absolutely continuous relative to the measure μ on $B(R)$ with respect to A, φ_1, $\varphi_2 \in M_\mu^A$. For every pair of non-zero real numbers $r_1, r_2 \in R$ there is $\varphi \in M_\mu^A$ such that

$$(*)\quad r\left(\frac{d\varphi}{d\mu}\right)^{1/2} = r_1\left(\frac{d\varphi_1}{d\mu}\right)^{1/2} + r_2\left(\frac{d\varphi_2}{d\mu}\right)^{1/2}$$

for some $r \in R$, and there exists $\varphi \in M_\mu^A$ such that (*) does not for any $r \in R$ and any $r_1, r_2 \in R$, $r_i \neq 0$. (We assume that M_μ^A has at least two distinct elements.) If (*) holds (with $\varphi_1 \neq \pm\varphi_2$ and $r_i \neq 0$), then $a(\varphi) < a(\varphi_1) \vee a(\varphi_2)$. If (*) holds with $r_1 = r_2 = 1$, $r = \pm\sqrt{2}$, then $a(\varphi_1) \perp a(\varphi_2)$.

Then there is a real vector space $W \leq L_2(R,\mu)$ of dimension ≥ 3 and a lattice monomorphism from the lattice of finite-dimensional subspaces of W to the lattice of finite elements of L that maps atoms to atoms. Consequently, by a theorem due to Baer,[10] the division ring D is an extension of the field of real numbers.

Observe that all the assumptions of the theorem are, or can be, expressed in the physically basic language of the probability function p. This includes the assumptions about atoms in L, since, as we have shown,[11] the lattice operations (and the relation of orthogonality) in L can also be expressed in terms of the function p.

It is easy to see that if A is a smooth maximal observable with respect to the measure space $[R, B(R), \mu]$ (in the sense of Gudder and Piron), then A satisfies the assumptions of the theorem. In particular, the system of states and observables of quantum mechanics based on the complex Hilbert space $L_2(R,\mu)$, where μ is the Lebesgue measure, also admits a set of pure states M and an observable A satisfying conditions (1), (2), and (3). Namely, in this case M can be taken to be the set

of pure states corresponding to nonnegative unit vectors and A the usual position observable. As pointed out by Gudder and Piron,[12] the position observable is not maximal with respect to the set of all pure states of $L_2(R,\mu)$, so in general M is not the set of all pure states of the system but only a subset of it.

Proof of the theorem: We define

$$W=\left\{r\left(\frac{d\varphi}{d\mu}\right)^{1/2}:\varphi\in M_\mu^A,\ r\in R\right\}.$$

Since for any $f_1,f_2\in W$ we have $r_1f_1+r_2f_2\in W$ for all $r_1,r_2\in R$ [if $\varphi_1\neq\pm\varphi_2$ this follows from (3), otherwise this is obvious], W is a real vector space. Since $(d\varphi/d\mu)\in L_1(R,\varphi)$ implies $(d\varphi/d\mu)^{1/2}\in L_2(R,\mu)$, we have $W\leq L_2(R,\mu)$. Condition (3) also implies that $\dim W\geq 3$.

For $0\neq f\in W$, let $[f]$ denote the one-dimensional subspace of W generated by f. Let $L_0(W)$ denote the set of all one-dimensional subspaces of W [the set of atoms of $L(W,R)$], and let $L_0(V)$ denote the set of all atoms of $L=L(V,D)$. We define a map ψ: $L_0(W)\to L_0(V)$ as follows: if $f=r(d\varphi/d\mu)^{1/2}$, $r\neq 0$, then $\psi[f]=a(\varphi)$. We will show that this map is well defined. Assume that $r_1(d\varphi/d\mu)^{1/2}=r_2(d\varphi_2/d\mu)^{1/2}$, $r_1\neq 0$. This implies $O(d\varphi/d\mu)^{1/2}=r_1(d\varphi_1/d\mu)^{1/2}-r_2(d\varphi_2/d\mu)^{1/2}$ for all $\varphi\in M_\mu^A$. Hence $\varphi_1=\pm\varphi_2$ by (3). If $\varphi_1=\alpha^A$, $\alpha\in M$, then $a(\varphi_1)=a(\varphi_2)=a(\alpha)$ by definition. Let $\varphi_1=c_1\alpha_1^A-c_2\alpha_2^A$, $\alpha_1\perp^A\alpha_2$, $c_1,c_2>0$, $c_1+c_2=1$, and let $\varphi_2=-\varphi_1=c_2\alpha_2^A-c_1\alpha_1^A$. We have $\varphi_1^*=\varphi_2^*=c_2\alpha_1^A+c_1\alpha_2^A$. By definition, $a(\varphi_1)=[a(\alpha_1)\vee a(\varphi_1^*)]\wedge a(\varphi_1^*)'$ and $a(\varphi_2)=[a(\alpha_2)\vee a(\varphi_2^*)]\wedge a(\varphi_2^*)$, thus to show that $a(\varphi_1)=a(\varphi_2)$ we must show that $a(\alpha_1)\vee a(\varphi_1^*)=a(\alpha_2)\vee a(\varphi_1^*)$. The equality $\varphi_1^*=c_2\alpha_1^A+c_1\alpha_2^A$ implies

$$\left(\frac{d\varphi_1^*}{d\mu}\right)^{1/2}=(c_2)^{1/2}\left(\frac{d\alpha_1}{d\mu}\right)^{1/2}+(c_1)^{1/2}\left(\frac{d\alpha_2}{d\mu}\right)^{1/2}\qquad(1)$$

(note that $\alpha_1\perp^A\alpha_2$). Hence by (3) we conclude that $a(\varphi_1^*)<a(\alpha_1)\vee a(\alpha_2)$. Consequently $a(\alpha_1)\vee a(\varphi_1^*)\leq a(\alpha_1)\vee a(\alpha_2)$. From Eq. (1) we also obtain

$$(c_1)^{1/2}\left(\frac{d\alpha_2}{d\mu}\right)^{1/2}=\left(\frac{d\varphi_1^*}{d\mu}\right)^{1/2}-(c_2)^{1/2}\left(\frac{d\alpha_1}{d\mu}\right)^{1/2},$$

which implies $a(\alpha_2)<a(\varphi_1^*)\vee a(\alpha_1)$, and consequently $a(\alpha_1)\vee a(\alpha_2)\leq a(\varphi_1^*)\vee a(\alpha_1)$. This shows that $a(\alpha_1)\vee a(\varphi_1^*)=a(\alpha_1)\vee a(\alpha_2)$. Similarly we show that $a(\alpha_2)\vee a(\varphi_1^*)=a(\alpha_1)\vee a(\alpha_2)$, and we obtain $a(\alpha_1)\vee a(\varphi_1^*)=a(\alpha_2)\vee a(\varphi_1^*)$. Hence $a(\varphi_1)=a(\varphi_2)$ and the map ψ is well defined. We now show that it is one-to-one. Let $\psi[f_1]=\psi[f_2]$, where $f_i=r_i(d\varphi_i/d\varphi)^{1/2}$, $r_i\neq 0$. Be the definition of ψ we have $a(\varphi_1)=a(\varphi_2)$. By (3) there are $\varphi\in M_\mu^A$ and $r\in R$ such that (*) holds. If $\varphi_1\neq\pm\varphi_2$, we conclude that $a(\varphi)<a(\varphi_1)\vee a(\varphi_2)=a(\varphi_1)$, a contradiction. Hence $\varphi_1=\pm\varphi_2$, which implies $[f_1]=[f_2]$ and shows that ψ is one-to-one. We will show that ψ preserves orthogonality in $W\leq L_2(R,\mu)$. Let for $f_i=r_i(d\varphi_i/d\mu)^{1/2}$, $i=1,2$, $[f_1]\perp[f_2]$, i.e.,

$$\int\left(\frac{d\varphi_1}{d\mu}\right)^{1/2}\left(\frac{d\varphi_2}{d\mu}\right)^{1/2}d\mu=0.$$

By (3), there exist $\varphi\in M_\mu^A$ and $r\in R$ such that (*) holds with $r_1=r_2=1/\sqrt{2}$. We will show then $r=\pm 1$. If we square (*) putting $r_1=r_2=1/\sqrt{2}$, we get

$$r^2\frac{d|\varphi|}{d\mu}=\frac{1}{2}\frac{d|\varphi_1|}{d\mu}+\left(\frac{d\varphi_1}{d\mu}\right)^{1/2}\left(\frac{d\varphi_2}{d\mu}\right)^{1/2}+\frac{1}{2}\frac{d|\varphi_2|}{d\varphi}.$$

After integrating and applying the Radon–Nikodym theorem ($|\varphi|$ and $|\varphi_i|$ are probability measures) we obtain

$$r^2=\frac{1}{2}+\int\left(\frac{d\varphi_1}{d\mu}\right)^{1/2}\left(\frac{d\varphi_2}{d\mu}\right)^{1/2}d\mu+\frac{1}{2}=1,$$

that is, $r=\pm 1$. By (3) we infer that $a(\varphi_1)\perp a(\varphi_2)$, i.e., $\psi[f_1]\perp\psi[f_2]$ and ψ preserves orthogonality.

Denote the lattice of finite-dimensional subspaces of W by $P(W)$. We extend ψ to $P(W)$ as follows: if $a\in P(W)$ and $\{e_\alpha\}$ is the set of atoms (one-dimensional subspaces) included in a, we define $\psi(a)=V_\alpha\psi(e_\alpha)$. We have to show that the so-extended ψ preserves the lattice operations. It is evident that $\psi(a\vee b)\geq\psi(a)\vee\psi(b)$. Suppose that $[e]$ is an atom and $[e]\leq a\vee b$. Then there are atoms $[f]$ and $[g]$ such that $[f]\leq a$, $[g]\leq b$ and $[e]\leq[f]\vee[g]$. If $e=r(d\varphi/d\mu)^{1/2}$, $f=r_1(d\varphi_1/d\mu)^{1/2}$, $g=r_2(d\varphi_2/d\mu)^{1/2}$, this implies that condition (*) holds, and consequently by (3) $\psi[e]<\psi[f]\vee\psi[g]\leq\psi(a)\vee\psi(b)$. Hence $\psi(a\vee b)\leq\psi(a)\vee\psi(b)$, and finally $\psi(a\vee b)=\psi(a)\vee\psi(b)$. Since for $a\leq b\in P(W)$ we have $b-a\in P(W)$ where $a\perp(b-a)$ and $a\vee(b-a)=b$, and ψ preserves orthogonality, we conclude that for $a\leq b$ $\psi(b-a)=\psi(b)-\psi(a)$, i.e., ψ preserves relative orthogonal complementation. Since for $a,b\in P(W)$ $a\wedge b=(a^0\vee b^0)^0$, where for $x\leq a\vee b$ $x^0=(a\vee b)-x$ is the relative orthogonal complement in the interval $[0,\ a\vee b]$, we see that $\psi(a\wedge b)=\psi(a)\wedge\psi(b)$. Hence ψ preserves the lattice operations. To show that ψ is a monomorphism, observe that $\psi(a)=0$ implies $a=0$, and if $\psi(a)=\psi(b)$, then $\psi(a-a\wedge b)=\psi(a)-\psi(a)\wedge\psi(b)=0$. Since $a=(a\wedge b)\vee(a-a\wedge b)$ by orthomodularity, we see that $a=a\wedge b$, i.e., $a\leq b$. Similarly, $b\leq a$, so $a=b$. This ends the proof of the theorem.

Concluding this paper we would like to note that although the theorem of this paper is simpler than that of Gudder and Piron and accomplishes the main purpose of the latter by showing that D is an extension of R, the result is not as strong since their monomorphism is from the lattice of all finite dimensional subspaces of $L_2(R,\mu)$.

[1] J. Jauch and C. Piron, Helv. Phys. Acta **42**, 842 (1969).
[2] M. J. Maczynski, Repts. Math. Phys. 3, (1972), in print
[3] For the definitions of the involved lattice-theoretical terms see F. Maeda and S. Maeda, *Theory of Symmetric Lattices* (Springer, Berlin, 1970).
[4] M. D. MacLaren, Pac. J. Math. 14, 597 (1964).
[5] C. Piron, Helv. Phys. Acta **47**, 439 (1964).
[6] S. Gudder and C. Piron, J. Math. Phys. **12**, 1583 (1971).
[7] See Ref. 2.
[8] See Ref. 2.
[9] See Ref. 6.
[10] R. Baer, *Linear Algebra and Projective Geometry* (Academic, New York, 1952). For a detailed proof, see V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, N.J., 1968), Vol. I., pp. 36–42. See also Theorem 3 in Ref. 6.
[11] See Ref. 2.
[12] See Ref. 6.

Operational statistics. II. Manuals of operations and their logics

C. H. Randall and D. J. Foulis

Department of Mathematics, University of Massachusetts, Amherst, Massachussetts 01002
(Received 12 January 1972)

In the first paper of this series, the notion of a manual of operations and its (generalized) sample space was introduced. For each such manual of operations, a so-called logic was constructed. In the present paper, some of the algebraic properties of such logics are identified with the corresponding operational properties of the manual.

I. INTRODUCTION

In the conventional theory of quantum phenomena[1-3] the connections with actual physical operations are, at best, vague. One need only examine the recent literature concerning the theory of measurement to appreciate the extent to which this is so[4-6,7]; hence, it is not surprising that there have been numerous attempts to correct this situation.[8-12] The notion of a (generalized) sample space and its affiliated logics introduced in Part I of the present series[13] provides a framework within which such attempts can be developed and analyzed.

The theory we have built in (Ref. 13) not only provides for the formal representation of the underlying physical operations, but indeed, casts them in a fundamental role in the sense that one may actually start with a so-called manual of operations and construct from it a (generalized) sample space and thence a logic of propositions. We shall attempt in the present paper to trace down the operational significance of some of the more important structural features that the logic of a sample space might possess. In this paper, we shall assume that the reader is familiar with the ideas presented in (Ref. 13).

In order to emphasize the fundamental role played by coherent collections or manuals of operations, we shall, as promised above, formally abstract the notion of a manual from that of a sample space. Thus, let $\mathscr{A}$ be a nonempty collection of nonempty sets and let $X=\cup\mathscr{A}$. For $x,y\in X$, say that x is *orthogonal* to y and write $x\perp y$ provided that $x\neq y$ and there exists $E\in\mathscr{A}$ with $x,y\in E$. If $(X,\perp,\mathscr{A})$ is a sample space, we agree to call $\mathscr{A}$ a *manual*. If $\mathscr{A}$ is a manual, we refer to $(x,\perp,\mathscr{A})$ as the sample space corresponding to $\mathscr{A}$. Since this is a one-to-one correspondence between manuals and sample spaces, we shall, for convenience, frequently fail to distinguish between a manual and its corresponding sample space as far as terminology is concerned. For instance, the outcomes and events for the sample space $(X,\perp,\mathscr{A})$ will also be referred to as $\mathscr{A}$-outcomes and $\mathscr{A}$-events; the manual $\mathscr{A}$ of a Dacey sample space will be called a Dacey manual, and so on. Moreover, we shall even denote the logic $\Pi(X,\perp,\mathscr{A})$ of the manual $\mathscr{A}$ by $\Pi(\mathscr{A})$.

II. CONJUNCTION AND DISJUNCTION

The operational interpretation of the infimum and supremum of propositions in quantum logics is a difficult matter that has engaged the attention of many authors.[14-16] In the logic $\Pi(\mathscr{A})$ of a manual $\mathscr{A}$, such an interpretation is always available for the infimum and, when the logic $\Pi(\mathscr{A})$ is closed under the negation mapping $p(A)\to(p(A))'$, the supremum also admits an operational explanation.

Let $\mathscr{A}$ be a manual and let $\{A_j\mid j\in J\}$ be a family of events for this manual. If A is an event for $\mathscr{A}$, the notation $p(A)=\wedge_j p(A_j)$ will be understood to mean that the family of propositions $\{p(A_j)\mid j\in J\}$ has an infimum (greatest lower bound) in the partially ordered set $(\Pi(\mathscr{A}),\leq)$ and that $p(A)$ is this infimum. Likewise, the notation $p(B)=\vee_j p(A_j)$ will mean that the family of propositions $\{p(A_j)\mid j\in J\}$ has a supremum (least upper bound) in $(\Pi(\mathscr{A}),\leq)$ and that $p(B)$ is this supremum. Finally, the notation $p(B)=\oplus_j p(A_j)$ will be understood to mean that the propositions in the family $\{p(A_j)\mid j\in J\}$ are pairwise othogonal and that $p(B)=\vee_j p(A_j)$.

Let X be the outcome set for the manual $\mathscr{A}$ and let E be any operation in $\mathscr{A}$. Clearly, $p(\phi)=(\phi,X)$ is the infimum of all of the propositions in $\Pi(\mathscr{A})$, while $p(E)=(X,\phi)$ is the supremum of all of these propositions. Thus, we define $0=p(\phi)$ and $1=p(E)$ and note that $(\Pi(\mathscr{A}),\leq)$ is a bounded poset in that $0\leq p(A)\leq 1$ holds for all $p(A)\in\Pi(\mathscr{A})$.

Theorem 1: The necessary and sufficient condition that $p(A)=\wedge_j p(A_j)$ is that $A^{\perp\perp}=\cap_j(A_j)^{\perp\perp}$.

Proof: Suppose first that $p(A)=\wedge_j p(A_j)$. Since $p(A)\leq p(A_j)$ for all j, then $A^{\perp\perp}\subseteq(A_j)^{\perp\perp}$ for all j; hence, $A^{\perp\perp}\subseteq\cap_j(A_j)^{\perp\perp}$. Suppose $x\in\cap_j(A_j)^{\perp\perp}$. Then $p(\{x\})\leq p(A_j)$ for all j, whence, $p(\{x\})\leq\wedge_j p(A_j)=p(A)$, so $x\in\{x\}^{\perp\perp}\subseteq A^{\perp\perp}$. This proves that $A^{\perp\perp}=\cap_j(A_j)^{\perp\perp}$ as desired.

Conversely, suppose that $A^{\perp\perp}=\cap_j(A_j)^{\perp\perp}$. Evidently, then, $p(A)$ is a lower bound in $(\Pi(\mathscr{A}),\leq)$ for the family $\{p(A_j)\mid j\in J\}$. Suppose that $p(D)$ is another proposition in $\Pi(\mathscr{A})$ which is also a lower bound for this family. Then $D^{\perp\perp}\subseteq\cap_j(A_j)^{\perp\perp}=A^{\perp\perp}$; hence, $p(D)\leq p(A)$. This shows that $p(A)$ is indeed effective as the greatest lower bound of $\{p(A_j)\mid j\in J\}$ and completes the proof.

In conventional logic, the conjunction of a set of propositions is generally understood to be a proposition that is true if and only if each of the propositions in the set are true. As a consequence of Theorem 1, we see that in the logic $\Pi(\mathscr{A})$, $\wedge_j p(A_j)$, if it exists, is precisely that proposition which is confirmed by an outcome if and only if each of the propositions $p(A_j)$ are confirmed by this outcome. Consequently, we feel entitled to call $\wedge_j p(A_j)$, when it exists, the *conjunction* of the propositions in the family $\{p(A_j)\mid j\in J\}$. In conventional logic, it is a theorem that the conjunction of a set of propositions is false if and only if at least one of the propositions in the set is false. Suppose that we have $p(A)=\wedge_j p(A_j)$ in our logic $\Pi(\mathscr{A})$. As a consequence of Theorem 1, we see that

$p(A)$ is refuted precisely by those outcomes in the set $(\cap_j (A_j)^{\perp\perp})^{\perp} = (\cup_j (A_j)^{\perp})^{\perp\perp}$. Since for any particular $i \in J$, $(A_i)^{\perp} \subseteq (\cup_j (A_j)^{\perp})^{\perp\perp}$, then any outcome that refutes any one of the propositions in the family $\{p(A_j) \mid j \in J\}$ will refute the conjunction $p(A)$. However, our conjunction need not behave classically in that there may be outcomes in $(\cup_j (A_j)^{\perp})^{\perp\perp}$ that refute $p(A)$ but do not belong to $\cup_j (A_j)^{\perp}$ and, hence, do not refute any of the propositions in the family $\{p(A_j) \mid j \in J\}$.

Consider, for instance, the propositions $p(\{b,n\}) = (p(\{a\}))' = (\{b,n\},\{a\})$ and $p(\{c,n\}) = (p(\{d\}))' = (\{c,n\},\{d\})$ in the logic $\Pi(\mathscr{A})$ of the manual $\mathscr{A} = \{\{a,b,n\},\{c,d,n\}\}$ of (Ref. 13, Example III). The conjunction $p(\{b,n\}) \wedge p(\{c,n\})$ is given by $p(\{b,n\}) \wedge p(\{c,n\}) = p(\{n\}) = (\{n\}, \{a,b,c,d\})$. Notice that this conjunction is refuted not only by outcomes a and d, as it would be classically, but also by outcomes b and c, neither of which refute either $p(\{b,n\})$ or $p(\{c,n\})$. If we consider the physical situation described in Example III,[13] we find that the above departure from classical logic is virtually forced upon us.

In conventional logic, the disjunction of a set of propositions is generally understood to be a proposition which is true if and only if at least one of the propositions in the set is true. This suggests the following definition: Let $\{A_j \mid j \in J\}$ be a family of events for the manual $\mathscr{A}$ and let A be an event for $\mathscr{A}$. We shall say that the proposition $p(A)$ is the disjunction of the propositions in the family $\{p(A_j) \mid j \in J\}$ if and only if $p(A)$ is confirmed precisely by those outcomes that confirm at least one of the propositions in the family $\{p(A_j) \mid j \in J\}$. This simply requires that $A^{\perp\perp} = \cup_j (A_j)^{\perp\perp}$.

Theorem 2: Let A_j be an event for the manual $\mathscr{A}$ for each $j \in J$ and let A be an event for $\mathscr{A}$ such that $p(A)$ is the disjunction of the propositions in the family $\{p(A_j) \mid j \in J\}$. Then $p(A) = \vee_j p(A_j)$.

Proof: We have $A^{\perp\perp} = \cup_j (A_j)^{\perp\perp}$, from which it is clear that $p(A)$ is an upper bound in $(\Pi(\mathscr{A}), \leq)$ for the family $\{p(A_j) \mid j \in J\}$. Suppose that B is any event for $\mathscr{A}$ such that $p(B)$ is also an upper bound for this family. Then, for each $j \in J$, $(A_j)^{\perp\perp} \subseteq B^{\perp\perp}$; hence, $A^{\perp\perp} = \cup_j (A_j)^{\perp\perp} \subseteq B^{\perp\perp}$, that is, $p(A) \leq p(B)$. This shows that $p(A)$ is effective as the supremum of the propositions in the family $\{p(A_j) \mid j \in J\}$ and completes the proof.

If we glance again at the manual $\mathscr{A} = \{\{a,b,n\},\{c,d,n\}\}$ of (Ref. 13, Example III), we see, for instance, that $p(\{b,n\})$ is the disjunction of the two propositions $p(\{b\})$ and $p(\{n\})$; hence, by Theorem 2, $p(\{b,n\}) = p(\{b\}) \vee p(\{n\})$. However, note carefully that $p(\{a,b\}) = (\{a,b,c,d\},\{n\}) = p(\{a\}) \vee p(\{b\})$; but, $p(\{a,b\})$ is clearly not the disjunction of the propositions $p(\{a\})$ and $p(\{b\})$. Thus, every disjunction in the logic of a manual is a supremum, but the converse is false.

Suppose that $\mathscr{A}$ is a manual whose logic is closed under the negation map $p(A) \to (p(A))'$. Then, the negation map is an antiautomorphism of period two on the structure $(\Pi(\mathscr{A}), \leq)$; hence, trivially, it converts suprema into infima and vice-versa. Consequently, we have the conventional deMorgan laws: $(\vee_j p(A_j))' = \wedge_j (p(A_j))'$ and $(\wedge_j p(A_j))' = \vee_j (p(A_j))'$. In particular, if $p(B) = \vee_j p(A_j)$, then $p(B) = (\wedge_j (p(A_j))')'$. Since we already have an operational interpretation for the infimum as a conjunction, this provides us with an operational interpretation for the supremum. Specifically, we have the following theorem:

Theorem 3: Suppose that $\mathscr{A}$ is a manual whose logic is closed under the negation mapping. Let $\{A_j \mid j \in J\}$ be a family of events for $\mathscr{A}$. Then, the necessary and sufficient condition that $p(B) = \vee_j p(A_j)$ is that $p(B)$ is refuted by precisely those outcomes for $\mathscr{A}$ that refute every proposition $p(A_j)$, $j \in J$.

As the following example shows, one has, in general, no guarantee of the existence of suprema and infima in the logic $\Pi(\mathscr{A})$ of a manual $\mathscr{A}$. We should like to thank M. F. Janowitz for calling this simple example to our attention.

Example IV: Let $\mathscr{A} = \{\{a,b,c,\},\{c,d,e\},\{e,f,g\},\{g,h,a\}\}$. Then $\mathscr{A}$ is a Dacay manual, but the supremum of $p(\{a\})$ and $p(\{e\})$ does not exist in the logic $\Pi(\mathscr{A})$. Also, the infimum of $p(\{b,c\})$ and $p(\{c,d\})$ does not exist in this logic.

We shall say that a manual $\mathscr{A}$ is *conjunctive* provided that given any two propositions $p(A)$ and $p(B)$ in the logic $\Pi(\mathscr{A})$, there exists a proposition $p(C) \in \Pi(\mathscr{A})$ such that $p(C) = p(A) \wedge p(B)$. If every countable family (respectively, every family) of propositions in $\Pi(\mathscr{A})$ has an infimum, we shall say that the manual $\mathscr{A}$ is *σ-conjunctive* (respectively, *completely conjunctive*). Notice that if the logic $\Pi(\mathscr{A})$ is closed under the negation map $p(A) \to (p(A))'$, then $\mathscr{A}$ is conjunctive if and only if every pair of propositions in $\Pi(\mathscr{A})$ have a supremum in $\Pi(\mathscr{A})$. (This follows immediately from the deMorgan laws.) Similar remarks hold for the σ-conjunctive and completely conjunctive cases.

Although many authors[15,17,18] impose conjunctivity axiomatically on quantum logics, Example IV shows that, in general, the operational logic $\Pi(\mathscr{A})$ of a manual need not enjoy this property. Moreover, as the following theorem shows, the imposition of conjunctivity—even in the case of a completely coherent Dacey manual—is equivalent to a rather technical operational condition:

Theorem 4: Let $\mathscr{A}$ be a completely coherent manual. Then, the necessary and sufficient condition that $\mathscr{A}$ be a conjunctive Dacey manual is the following: If D is an event such that $p(D)$ is neither confirmed nor refuted by an outcome x, then, there exists an outcome y which refutes $p(D)$ and which operationally rejects every outcome z that refutes both $p(D)$ and $p(\{x\})$.

Proof: The given condition is equivalent to the following: Given an orthogonal set D and an outcome $x \notin D^{\perp} \cup D^{\perp\perp}$, it follows that $D^{\perp} \cap (\{x\}^{\perp} \cap D^{\perp})^{\perp} \neq \phi$. Reference to (Ref. 19, Theorem 1) completes the proof.

Similarly we have the following theorem:

Theorem 5: If $\mathscr{A}$ is a conjunctive, completely coherent Dacey manual, then it is completely conjunctive and every closed generalized proposition belongs to the logic $\Pi(\mathscr{A})$.

As Theorem 4 shows, the existence of infima and suprema, in general, is a delicate matter. However, the question of the existence of orthogonal suprema is more

tractible. In fact, it is clear from the discussion preceeding Lemma 6 in (Ref. 13) that the coherence of the manual $\mathcal{A}$ guarantees the existence of finite orthogonal suprema. We shall say that the logic $\Pi(\mathcal{A})$ is *orthocomplete* (respectively, *σ-orthocomplete*) if and only if every orthogonal family (respectively, every countable orthogonal family) of propositions in $\Pi(\mathcal{A})$ has a supremum. Evidently, $\Pi(\mathcal{A})$ is orthocomplete (respectively, σ-orthocomplete) if $\mathcal{A}$ is completely coherent (respectively, σ-coherent). In fact, if $\{A_j | j \in J\}$ is an orthogonal family of events for $\mathcal{A}$ such that $A = \cup_j A_j$ is also an event for $\mathcal{A}$, then $p(A) = \oplus_j p(A_j)$. Notice, however there exist manuals $\mathcal{A}$ that are not σ-coherent, but whose logics $\Pi(\mathcal{A})$ are even σ-complete Boolean algebras.

III. NEGATION AND DACEY MANUALS

Conventionally, if a logic contains a proposition, then it also contains the negation of that proposition. In Example V we shall see that this need not be the case for an operational logic. To begin with, let D be an event for a manual $\mathcal{A}$; then $p(D) = (D^{\perp\perp}, D^{\perp})$ is a proposition in its logic $(\Pi(\mathcal{A}), \leq, \perp)$. The negation of such a proposition $p(D)$ has been defined, in (Ref. 13), to be $p(D)' = (D^{\perp}, D^{\perp\perp})$. Evidently, if $p(D)'$ is to be a member of $\Pi(\mathcal{A})$, there must exist another event B such that $B^{\perp\perp} = D^{\perp}$.

A natural place to look for such an event B is among the "local complements" of the event D; that is, look among those $E \setminus D$ for which E is an operation in $\mathcal{A}$ containing D. Notice that $p(E \setminus D)$ is a complement of $p(D)$ in $(\Pi(\mathcal{A}), \leq)$ and in fact an orthogonal complement; that is $p(D) \perp p(E \setminus D)$ and $p(D) \wedge p(E \setminus D) = 0$, $p(D) \vee p(E \setminus D) = p(E) = 1$. Moreover, all orthogonal complements of $p(D)$ are of this form. If $p(D) \oplus p(B) = 1$, then $p(D \cup B) = 1$, $(D \cup B)^{\perp} = \phi$ and thus $(D \cup B) \in \mathcal{A}$ and we have the following lemma:

Lemma 1: Let B and D be events for a manual $\mathcal{A}$. Then $p(D) \oplus p(B) = 1$ if and only if $B = E \setminus D$ for an operation $E \in \mathcal{A}$ containing D.

We shall now show that when $p(D)'$ is in $\Pi(\mathcal{A})$, it has the prescribed form. Suppose $p(D)' = p(B)$ for some event B and consequently that $D^{\perp} = B^{\perp\perp}$. It follows that $D \perp B$ and by coherence that $D \cup B$ is an event for $\mathcal{A}$. Moreover $(D \cup B)^{\perp} = D^{\perp} \cap B^{\perp} = B^{\perp\perp} \cap B^{\perp} = \phi$ and hence $(D \cup B) \in \mathcal{A}$. Thus we have proved the following lemma:

Lemma 2: Let D be an event for a manual $\mathcal{A}$ such that $p(D)' \in \Pi(\mathcal{A})$. Then $p(D)' = p(E \setminus D)$ for some operation E containing D.

In spite of the apparent good behavior of these local complements, we can still have manuals with logics that are not closed under negation. Consider the following example:

Example V: Let $\mathcal{A} = \{\{a, b\}, \{b, c\}, \{c, d\}, \{d, e\}\}$. Since $\{c\}$ is an event for the manual $\mathcal{A}$, $p(\{c\}) = (\{c\}, \{b, d\}) \in \Pi(\mathcal{A})$. However, as is now easily verified, the negation $p(\{c\})' = (\{b, d\}, \{c\})$ is not a member of $\Pi(\mathcal{A})$.

If a logic $\Pi(\mathcal{A})$ is closed under the negation map, then the restriction of this map to $\Pi(\mathcal{A})$ is effective as an orthocomplementation, (Ref. 20, p. 52). Indeed, Lemma 5 shows such a map to be a complementation and it has been noted that the negation map is an involutory order antiautomorphism. Thus, when $\Pi(\mathcal{A})$ is closed under negation, the system $(\Pi(\mathcal{A}), \leq, ')$ is an orthocomplemented poset (Ref. 23, p. 70). As a consequence, we shall define a manual to be an orthocomplemented manual when its logic is closed under negation.

Naturally we would prefer to have all of the local negations $p(E \setminus D)$, that arise from the local complements of an event D, equal to the global negation $p(D)'$. As the following example shows, this need not be the case even for an orthocomplemented manual.

Example VI: Let $\mathcal{A} = \{\{a, b\}, \{b, c\}, \{c, d\}\}$. It is easy to verify that $\mathcal{A}$ is an orthocomplemented manual. Evidently, we have $p(\{b\})' = (\{a, c\}, \{b\}) = p(\{a\})$. On the other hand, $\{c\}$ is a local complement of $\{b\}$, but $p(\{c\}) = (\{c\}, \{b, d\}) \neq p(\{b\})'$.

The following theorem states that these desired equalities are in fact equivalent to a number of other interesting and familiar conditions; for instance condition (ii) is satisfied whenever $\mathcal{A}$ allows sufficiently many complete stochastic models, that is, whenever $\Pi(\mathcal{A})$ admits a full set of regular states. Moreover, most of the proposed "quantum logics"[12,15–18,21,22] are required to be orthomodular posets (Ref. 23, p. 70) and thus satisfy condition (v).

Theorem 6: Let $\mathcal{A}$ be a manual. Then the following are equivalent:

(i) If A, B and C are events for $\mathcal{A}$ and $p(A) \oplus p(B) = 1$ and $p(A) \oplus p(C) = 1$, then $p(B) = p(C)$.

(ii) $\mathcal{A}$ is a Dacey Manual.

(iii) If A and B are events for $\mathcal{A}$ for which $A \subseteq B^{\perp}$ and $A^{\perp} \cap B^{\perp} = \phi$, then $A^{\perp} = B^{\perp\perp}$.

(iv) If A and B are events for $\mathcal{A}$ such that $A \subseteq B^{\perp}$, then there exists an event C for $\mathcal{A}$ for which $A \subseteq C \subseteq C^{\perp\perp} = B^{\perp}$.

(v) The system $(\Pi(\mathcal{A}), \leq, ')$ is an orthomodular poset.

Proof: (i) $\Rightarrow$ (ii): Assume (i) and let $E \in \mathcal{A}$ and $x, y \in X$ be such that $E \subseteq \{x\}^{\perp} \cup \{y\}^{\perp}$. We must show $x \perp y$. Set $A = E \cap \{x\}^{\perp}$ and $B = E \setminus A$ so that $p(A) \oplus p(B) = p(E) = 1$, $x \in A^{\perp}$ and $y \in B^{\perp}$. Coherence forces the existence of an operation $G \in \mathcal{A}$ with $A \cup \{x\} \subseteq G$. Set $C = G \setminus A$ and note that $p(A) \oplus p(C) = p(G) = 1$ and $x \in C$. By (i), $p(C) = p(B)$, so $y \in B^{\perp} = C^{\perp} \subseteq \{x\}^{\perp}$, proving that $x \perp y$ as required.

(ii) $\Rightarrow$ (iii): Assume (ii) and the hypothesis of (iii). By coherence, there exists an operation $E \in \mathcal{A}$ with $A \cup B \subseteq E$. Since $E \setminus (A \cup B) \subseteq (A \cup B)^{\perp} = A^{\perp} \cap B^{\perp} = \phi$, then $A \cup B = E$. Also, since $A \subseteq B^{\perp}$, we have $B^{\perp\perp} \subseteq A^{\perp}$. To show that $A^{\perp} \subseteq B^{\perp\perp}$, we assume that $x \in A^{\perp}$ and that $y \in B^{\perp}$ and argue that $x \perp y$ as follows: $E = A \cup B \subseteq \{x\}^{\perp} \cup \{y\}^{\perp}$, hence, $x \perp y$ by (ii). Thus, we have $A^{\perp} = B^{\perp\perp}$ as desired.

(iii) $\Rightarrow$ (iv): Assume (iii) and the hypothesis of (iv). By coherence, there exists an operation $E \in \mathcal{A}$ with $A \cup B \subseteq E$. Set $C = E \setminus B$, noting that $C \subseteq B^{\perp}$ and $C^{\perp} \cap B^{\perp} = E^{\perp} = \phi$; hence, that $C^{\perp\perp} = B^{\perp}$ by (iii). Clearly, $A \subseteq C \subseteq C^{\perp\perp} = B^{\perp}$.

(iv) $\Rightarrow$ (v): If B is any event, then, since $\phi \subseteq B^{\perp}$, (iv) forces the existence of an event C with $C^{\perp\perp} = B^{\perp}$, so that $(p(B))' = p(C) \in \Pi(\mathcal{A})$. As we have seen, this implies that $(\Pi(\mathcal{A}), \leq, ')$ is an orthocomplemented poset.

Suppose now that A and B are events with $p(A) \leq p(B)$. We must prove the orthomodular identity $p(B) = p(A) \oplus [p(A) \oplus (p(B))']'$, that is, $p(B) = p(A) \oplus [(p(A))' \wedge p(B)]$. Choose an event C such that $p(C) = p(A) \oplus [(p(A))' \wedge p(B)]$ and note that $p(C) \leq p(B)$. By an argument given below, we can assume without loss of generality that $C \subseteq B$. We now have

$$p(C) \oplus (p(B))' = [p(A) \oplus (p(B))'] \oplus [p(A) \oplus (p(B))']' = 1;$$

hence, $(p(C))' \wedge p(B) = 0$ by deMorgan's law. This yields $C^{\perp} \cap B^{\perp\perp} = \phi$, which, together with $C \subseteq B$, implies that $C = B$, so that $p(C) = p(B)$ as desired.

For the promised argument, suppose that $p(A) \leq p(D)$ for events A and D, so that $A \subseteq A^{\perp\perp} \subseteq D^{\perp\perp}$. Since $\Pi(\mathscr{A})$ is closed under negation, $D^{\perp\perp} = K^{\perp}$ for some event K. Condition (iv) now requires the existence of an event C with $A \subseteq C \subseteq C^{\perp\perp} = K^{\perp} = D^{\perp\perp}$, so we have $p(D) = p(C) \geq p(A)$ and actually $A \subseteq C$.

(v) $\Rightarrow$ (i): Assume (v) and let A and B be events for which $p(A) \oplus p(B) = 1$. Thus, from (v), $p(B) \leq (p(A))'$ and $(p(A))' = p(B) \oplus [p(A) \oplus p(B)]' = p(B) \oplus 1' = p(B) \oplus 0 = p(B)$. Consequently, (i) holds, and the proof is complete.

It now follows that the logic of a conjunctive Dacey manual is an orthomodular lattice (Ref. 20, p. 53); hence, the logic of a σ-coherent conjunctive Dacey space is a σ-complete orthomodular lattice. It follows from (24) that if $\mathscr{A}$ is a σ-coherent conjunctive Dacey manual in which every operation is at most countably infinite, then the logic $\Pi(\mathscr{A})$ is a complete orthomodular lattice.

IV. COMMUTATIVITY AND TESTABILITY

In the study of quantum logics, the notion of commutativity (or compatibility) plays a crucial role. For example, in the orthomodular lattice of all orthogonal projection operators on a Hilbert space, two elements P and Q commute in the sense presently to be defined if and only if $PQ - QP = 0$. In this section, we shall show how commutativity in the logic $\Pi(\mathscr{A})$ of a manual $\mathscr{A}$ is intimately connected with the notion of testability introduced in Ref. 13.

Let $\mathscr{A}$ be a given manual. Following Mackey (Ref. 21, p. 70), we shall say that two propositions $p(A)$ and $p(B)$ in $\Pi(\mathscr{A})$ commute, and write $p(A)$ **C** $p(B)$, provided that there exist three mutually orthogonal events A_1, B_1 and D with $p(A) = p(A_1) \oplus p(D)$ and $p(B) = p(B_1) \oplus p(D)$. Recall from Ref. 13 that two events A and B are said to be *compatible* if there exists an operation $E \in \mathscr{A}$ such that $A \cup B \subseteq E$. Clearly, if A and B are compatible events, then $p(A)$ **C** $p(B)$, for it is only necessary to take $A_1 = A \setminus B$, $B_1 = B \setminus A$ and $D = A \cap B$ in the above definition.

Theorem 7: Let A and B be events for the manual $\mathscr{A}$. Then, the necessary and sufficient condition for $p(A)$ **C** $p(B)$ to hold is that there exist compatible events A_0 and B_0 such that $p(A) = p(A_0)$ and $p(B) = p(B_0)$.

Proof: The remark preceding the theorem establishes the sufficiency. Conversely, suppose that $p(A) = p(A_1) \oplus p(D)$ and $p(B) = p(B_1) \oplus p(D)$, where A_1, B_1 and D are mutually orthogonal events. By coherence, there exists an operation $E \in \mathscr{A}$ with $A_1 \cup B_1 \cup D \subseteq E$. Put $A_0 = A_1 \cup D$, $B_0 = B_1 \cup D$, and note that A_0 and B_0 have the requisite properties.

Corollary 1: Let $\mathscr{A}$ be a manual and let $p(A)$, $p(B) \in \Pi(\mathscr{A})$ with $p(A)$ **C** $p(B)$. Then, the operational propositions $p(A)$ and $p(B)$ are simultaneously testable.

Proof: By Theorem 7, we can suppose, without loss of generality, that A and B are compatible, so that there exists an operation $E \in \mathscr{A}$ with $A \cup B \subseteq E$. By (Ref. 13, Lemma 5), E is effective as a test operation for both $p(A)$ and $p(B)$, and the proof is complete.

The converse of Corollary 1 is, in general, false. For instance, consider the manual $\mathscr{A} = \{\{a, b\}, \{b, c\}, \{c, d\}\}$ of Example VI. A simple computation reveals, that the operation $E = \{b, c\}$ is effective as a simultaneous test operation for *all* of the propositions in the logic $\Pi(\mathscr{A})$; however, $p(\{a\})$ fails to commute with $p(\{d\})$ in the logic $\Pi(\mathscr{A})$. The next theorem shows, among other things, that such anomalous situations do not arise for a Dacey manual.

Theorem 8: Let A and B be events for a Dacey manual $\mathscr{A}$. *Then*:

(i) An operation $E \in \mathscr{A}$ tests the proposition $p(A)$ if and only if there exists an event $D \subseteq E$ with $p(A) = p(D)$.

(ii) $p(A)$ **C** $p(B)$ if and only if $p(A)$ and $p(B)$ are simultaneously testable.

(iii) $p(A) \perp p(B)$ if and only if $p(A)$ and $p(B)$ are simultaneously testable and disjoint.

Proof: (i) If there exists $D \subseteq E$ with $p(A) = p(D)$, then it is clear from (Ref. 13, Lemma 5) that E tests $p(A)$. Conversely, suppose that E tests $p(A)$, that is, $E \subseteq A^{\perp\perp} \cup A^{\perp}$. Put $D = A^{\perp\perp} \cap E$, and note that $D \subseteq E$ and that $p(D) \leq p(A)$. To prove that $p(D) = p(A)$, it will suffice to prove that $D^{\perp} \subseteq A^{\perp}$. To this end, we shall assume that $x \in D^{\perp}$ and that $a \in A$ and argue that $x \perp a$. Since $\mathscr{A}$ is a Dacey manual, it will be enough to show that $E \subseteq \{x\}^{\perp} \cup \{a\}^{\perp}$. But, $E \subseteq D \cup A^{\perp} \subseteq \{x\}^{\perp} \cup \{a\}^{\perp}$, and our argument is complete.

(ii): This follows immediately from Corollary 1, Theorem 7 and part (i) of the present theorem.

(iii): Recall that $p(A)$ and $p(B)$ are said to be disjoint[13] if $A^{\perp\perp} \cap B^{\perp\perp} = \phi$, that is, if $p(A) \wedge p(B) = 0$ in $\Pi(\mathscr{A})$. Obviously, orthogonal propositions are disjoint and, by coherence and (Ref. 13, Lemma 5), orthogonal propositions are simultaneously testable. Conversely, if $p(A)$ and $p(B)$ are disjoint and simultaneously testable, then, by part (ii) of the present theorem, $p(A)$ **C** $p(B)$ and $p(A) \wedge p(B) = 0$. It is well-known, and easy to check, that in an orthomodular poset, the latter two conditions imply $p(A) \perp p(B)$. Since, by part (v) of Theorem 6, $\Pi(\mathscr{A})$ is an orthomodular poset, our proof is complete.

We have shown in (Ref. 13, Lemma 5) that every proposition in $\Pi(\mathscr{A})$ is closed and testable. Suppose that $\mathscr{A}$ is a Dacey manual and that (A, B) is a closed and testable operational proposition for $\mathscr{A}$. Then $A^{\perp} = B$, $B^{\perp} = A$ and there exists an operation $E \in \mathscr{A}$ such that $E \subseteq A \cup B$. If we now put $D = E \cap B^{\perp}$, and argue as in Part (i) of Theorem 8, we conclude that $p(D) = (A, B)$, so that $(A, B) \in \Pi(\mathscr{A})$. Thus, we have the following result:

Theorem 9: The logic $\Pi(\mathscr{A})$ of a Dacey manual consists precisely of the closed and testable operational propositions over $\mathscr{A}$.

Let $\mathscr{A}$ be any manual. If Q is any subset of $\Pi(\mathscr{A})$, we define the *commutant* of Q, in symbols $\mathbf{C}(Q)$, to be the set of all those propositions in $\Pi(\mathscr{A})$ that commute with all of the propositions in Q. In particular, the *center* of $\Pi(\mathscr{A})$ is defined to be $\mathbf{C}(\Pi(\mathscr{A}))$. Notice that 0 and 1 always belong to the center of $\Pi(\mathscr{A})$. Barbara Jeffcott[25] has shown that the center of $\Pi(\mathscr{A})$ is always closed under the negation mapping $p(A) \to (p(A))'$ and has proved the following theorem:

Theorem 10: If $\mathscr{A}$ is any manual, then the center of $\Pi(\mathscr{A})$ forms a Boolean algebra

$$(\mathbf{C}(\Pi(\mathscr{A})), \leq, ').$$

Following customary usage, we shall say that the manual $\mathscr{A}$ [or the logic $\Pi(\mathscr{A})$] is *irreducible* provided that $\mathbf{C}(\Pi(\mathscr{A})) = \{0, 1\}$. We note that the existence of quantum mechanical superselection rules is equivalent to the condition that the logic of quantum mechanics is reducible (that is, not irreducible), (Ref. 21, p. 136). In a later paper, we shall show that a manual $\mathscr{A}$ is reducible if and only if it can be "factored" as a "product" of "simpler manuals". In the extreme case in which $\mathbf{C}(\Pi(\mathscr{A})) = \Pi(\mathscr{A})$, that is when $\Pi(\mathscr{A})$ is a Boolean algebra itself, we shall say that $\mathscr{A}$ is a *Boolean manual*. Notice that the manual given in Example I of (13) is a Boolean manual. As a corollary of Theorem 10, we have the following:

Corollary 2: Let $\mathscr{A}$ be a manual. Then, the following conditions are mutually equivalent:

(i) $\mathscr{A}$ is a Boolean manual.

(ii) $\Pi(\mathscr{A})$ is closed under the negation mapping and the system $(\Pi(\mathscr{A}), \leq, ')$ forms a Boolean algebra.

(iii) Any two propositions in $\Pi(\mathscr{A})$ commute.

(iv) $\mathscr{A}$ is a Dacey manual and any two propositions in $\Pi(\mathscr{A})$ are simultaneously testable.

V. SUBLOGICS AND SUBMANUALS

In the following theorem we collect the characteristic properties of the logic of a manual:

Theorem 11: Let $\mathscr{A}$ be a manual. Then the logic $(\Pi(\mathscr{A}), \leq, \perp)$ has the following properties:

(i) $\Pi(\mathscr{A})$ is partially ordered by $\leq$ and, for every $p(A) \in \Pi(\mathscr{A})$, $0 \leq p(A) \leq 1$.

(ii) The relation $\perp$ is symmetric on $\Pi(\mathscr{A})$ and, for $p(A) \in \Pi(\mathscr{A})$, $p(A) \perp p(A)$ implies that $p(A) = 0$.

(iii) If $p(A)$, $p(B) \in \Pi(\mathscr{A})$ with $p(A) \perp p(B)$, then the supremum $p(A) \oplus p(B)$ exists in $(\Pi(\mathscr{A}), \leq)$.

(iv) If $p(A)$, $p(B)$, $p(C) \in \Pi(\mathscr{A})$ with $p(A) \perp p(B)$, $p(A) \perp p(C)$ and $p(B) \perp p(C)$, then $p(A) \perp (p(B) \oplus p(C))$.

(v) If $p(A) \in \Pi(\mathscr{A})$, there exists $p(B) \in \Pi(\mathscr{A})$ with $p(B) \perp p(A)$ and $p(A) \oplus p(B) = 1$.

(vi) If $p(A)$, $p(B) \in \Pi(\mathscr{A})$, then $p(A) \leq p(B)$ if and only if, for every $p(C) \in \Pi(\mathscr{A})$, the condition $p(C) \perp p(B)$ implies the condition $p(C) \perp p(A)$.

One can actually show that an abstract system $(L, \leq, \perp)$ satisfying conditions (i)—(vi) above is isomorphic to the logic of a suitable manual.

Motivated by the above considerations, we define a *sublogic* of the logic $\Pi(\mathscr{A})$ to be a subset L of $\Pi(\mathscr{A})$ satisfying the following conditions:

(i) $0, 1 \in L$.

(ii) If $p(A)$, $p(B) \in L$ with $p(A) \perp p(B)$, then $p(A) \oplus p(B) \in L$.

(iii) If $p(A) \in L$, there exists $p(B) \in L$ with $p(A) \perp p(B)$ and $p(A) \oplus p(B) = 1$.

(iv) If $p(A)$, $p(B) \in L$ and if, for every $p(C) \in L$, the condition $p(C) \perp p(B)$ implies the condition $p(C) \perp p(A)$, then $p(A) \leq p(B)$.

Notice that if L is a sublogic of $\Pi(\mathscr{A})$, then the system $(L, \leq, \perp)$ inherits properties (i)—(vi) of Theorem 15. We now have the following easily proved result:

Lemma 3: Let $\mathscr{A}$ be a Dacey manual. Then a subset L of $\Pi(\mathscr{A})$ is a sublogic if and only if it satisfies the following conditions:

(a) $0, 1 \in L$.

(b) If $p(A)$, $p(B) \in L$ with $p(A) \perp p(B)$, then $p(A) \oplus p(B) \in L$.

(c) If $p(A) \in L$, then $(p(A))' \in L$.

In particular, if L is a sublogic of $\Pi(\mathscr{A})$, then $(L, \leq, ')$ is an orthomodular poset.

Since properties (a), (b), and (c) of Lemma 3 are closure properties, then if $\mathscr{A}$ is a Dacey manual, the intersection of any family of sublogics of $\Pi(\mathscr{A})$ is again a sublogic of $\Pi(\mathscr{A})$.

Let $\mathscr{A}$ be any manual. A sublogic L of $\Pi(\mathscr{A})$ will be called a *Boolean sublogic* of $\Pi(\mathscr{A})$ if it satisfies the following condition: Given $p(A)$, $p(B) \in L$, there exist $p(A_1)$, $p(B_1)$, $p(D) \in L$ with $p(A) = p(A_1) \oplus p(D)$, $p(b) = p(B_1) \oplus p(D)$ and $p(A_1) \perp p(B_1)$. Note that this condition implies that $p(A)$ commutes with $p(B)$ in the logic $\Pi(\mathscr{A})$; however, in general, elements of a sublogic L of $\Pi(\mathscr{A})$ can commute in $\Pi(\mathscr{A})$ without satisfying this condition as the following example shows:

Example VII: Let $\mathscr{A}$ be the classical manual consisting of the single operation $E = \{a, b, c, d\}$. Let $L = \{0, p(\{a, d\}), p(\{b, c\}), p(\{c, d\}), p(\{d, a\}), 1\}$. Here $\mathscr{A}$ is a Boolean manual, $\Pi(\mathscr{A})$ is the 16-element Boolean algebra and L is evidently a sublogic of $\Pi(\mathscr{A})$. However, L is isomorphic to the "horizontal sum" of two Boolean algebras of order four; hence it is not a Boolean algebra.

Even though a sublogic of the logic of a Boolean manual need not be a Boolean sublogic, it does follow, from Corollary 2, that a Boolean sublogic L of the logic of any manual is a Boolean algebra in its own right. Moreover, as the next easily proven lemma shows, there is one such Boolean sublogic naturally associated with each operation E in the manual $\mathscr{A}$.

Lemma 4: Let E be an operation in the manual $\mathscr{A}$. Then $\{p(D) \mid D \subseteq E\}$ is a Boolean sublogic of $\Pi(\mathscr{A})$ which is isomorphic to the Boolean algebra of all subsets of E.

It follows from Lemma 4 that, for any nonempty subcollection $\mathfrak{M}$ of the manual $\mathscr{A}$, the subset $\Pi(\mathfrak{M}, \mathscr{A})$

$=\{p(D) \mid D \subseteq E$ for some $E \in \mathfrak{M}\}$ is the union of a collection of Boolean sublogics woven together within the logic $\Pi(\mathscr{A})$. However, $\Pi(\mathfrak{M},\mathscr{A})$ need not be a sublogic—let alone a Boolean sublogic—of $\Pi(\mathscr{A})$. As we shall see below, certain mild and natural assumptions on $\mathscr{A}$ and $\mathfrak{M}$ will guarantee that $\Pi(\mathfrak{M},\mathscr{A})$ is at least a sublogic of $\Pi(\mathscr{A})$.

Let $\mathscr{A}$ be a fixed manual. We naturally define a *submanual* of $\mathscr{A}$ to be a nonempty subcollection $\mathfrak{M}$ of $\mathscr{A}$ which is a manual in its own right. Let X and M denote the outcome sets of $\mathscr{A}$ and $\mathfrak{M}$ respectively and let $\perp$ and $\#$ denote the relations of operational rejection induced on X and M respectively by $\mathscr{A}$ and $\mathfrak{M}$. Evidently, any $\mathfrak{M}$-event $D \subseteq M$ is automatically an $\mathscr{A}$-event and for $x, y \in M$, $x \# y \Rightarrow x \perp y$. If D is an $\mathfrak{M}$-event, we denote the proposition $(D^{\#\#}, D^{\#}) \in \Pi(\mathfrak{M})$ corresponding to D by $p_0(D) = (D^{\#\#}, D^{\#})$. As usual, if A is an $\mathscr{A}$-event, we denote the proposition $(A^{\perp\perp}, A^{\perp}) \in \Pi(\mathscr{A})$ by $p(A)$.

It might be expected that if $\mathfrak{M}$ is a submanual of $\mathscr{A}$, then $\Pi(\mathfrak{M},\mathscr{A})$ would be a sublogic of $\Pi(\mathscr{A})$ isomorphic to $\Pi(\mathfrak{M})$ under a correspondence $p_0(D) \leftrightarrow p(D)$. However, this does not obtain in general because of the weak connection between $\perp$ and $\#$. Thus, let us say that a submanual $\mathfrak{M}$ of $\mathscr{A}$ is an *induced* submanual provided that, for $x, y \in M$, $x \# y \Leftrightarrow x \perp y$. For our purposes, it is not quite sufficient merely to assume that $\mathfrak{M}$ is an induced submanual of $\mathscr{A}$; we must also impose on the parent manual $\mathscr{A}$ a condition that we have found desirable in other connections, namely, that $\mathscr{A}$ be a Dacey manual. We now have the following theorem:

Theorem 12: Let $\mathfrak{M}$ be an induced submanual of the Dacey manual $\mathscr{A}$ and let A and B be $\mathfrak{M}$-events. Let $M = \cup \mathfrak{M}$ and let $\#$ be the relation of operational rejection induced on M by $\mathfrak{M}$. Then:

(i) $\mathfrak{M}$ is a Dacey manual.

(ii) $A^{\#} = A^{\perp} \cap M$.

(iii) $A^{\perp\perp} = A^{\#\perp}$.

(iv) $A^{\#\#} = A^{\perp\perp} \cap M$.

(v) $A^{\perp} = A^{\#\#\perp}$.

(vi) $P_0(A) \leq P_0(B) \Leftrightarrow p(A) \leq p(B)$.

(vii) $\Pi(\mathfrak{M},\mathscr{A})$ is a sublogic of $\Pi(\mathscr{A})$ isomorphic to $\Pi(\mathfrak{M})$ under the correspondence $P_0(A) \leftrightarrow p(A)$.

Proof: (i) Let $E \in \mathfrak{M}$ and suppose that $x, y \in M$ with $E \subseteq \{x\}^{\#} \cup \{y\}^{\#}$. Since $\{x\}^{\#} \subseteq \{x\}^{\perp}$ and $\{y\}^{\#} \subseteq \{y\}^{\perp}$, it follows that $E \subseteq \{x\}^{\perp} \cup \{y\}^{\perp}$. Since $\mathscr{A}$ is a Dacey manual and $E \in \mathfrak{M} \subseteq \mathscr{A}$, the latter condition forces $x \perp y$; hence, since $\mathfrak{M}$ is an induced submanual of $\mathscr{A}$, we have $x \# y$. Thus, $\mathfrak{M}$ is a Dacey manual.

(ii): Since $\mathfrak{M}$ is an induced submanual of $\mathscr{A}$, then $\#$ is precisely the restriction to M of the relation $\perp$ on X; hence, (ii) follows immediately.

(iii): Since $\mathfrak{M}$ is a Dacey manual by (i) above, then, by Theorem 6, there exists an $\mathfrak{M}$-event C such that $C^{\#\#} = A^{\#}$. It follows that $A \# C$ and that $A \cup C \in \mathfrak{M}$; hence, that $A \perp C$ and $A \cup C \in \mathscr{A}$. Since $\mathscr{A}$ is a Dacey manual, Theorem 6 implies that $A^{\perp\perp} = C^{\perp}$. Since, by (ii) above, $A^{\#} \subseteq A^{\perp}$, then $A^{\perp\perp} \subseteq A^{\#\perp} = C^{\#\#\perp} \subseteq C^{\perp} = A^{\perp\perp}$; hence, $A^{\perp\perp} = A^{\#\perp}$.

(iv): By (iii) above, we have $A^{\#\perp} = A^{\perp\perp}$, so $A^{\#\#} = A^{\#\perp} \cap M = A^{\perp\perp} \cap M$.

(v): By (iv) above, $A^{\#\#} \subseteq A^{\perp\perp}$; hence, $A^{\perp} = A^{\perp\perp\perp} \subseteq A^{\#\#\perp} \subseteq A^{\perp}$, so $A^{\perp} = A^{\#\#\perp}$.

(vi): Using (ii) and (iii) above, we produce the following chain of implications: $p_0(A) \leq p_0(B) \Rightarrow B^{\#} \subseteq A^{\#} \Rightarrow A^{\#\perp} \subseteq B^{\#\perp} \Rightarrow A^{\perp\perp} \subseteq B^{\perp\perp} \Rightarrow p(A) \leq p(B) \Rightarrow B^{\perp} \subseteq A^{\perp} \Rightarrow B^{\perp} \cap M \subseteq A^{\perp} \cap M \Rightarrow B^{\#} \subseteq A^{\#} \Rightarrow p_0(A) \leq p_0(B)$.

(vii): By (vi), the correspondence $p_0(A) \leftrightarrow p(A)$ is a well-defined order isomorphism which clearly preserves orthogonality. By the argument in part (iii), there exists an $\mathfrak{M}$-event C such that $A^{\perp\perp} = C^{\perp}$; hence, $(p(A))' = p(C) \in \Pi(\mathfrak{M},\mathscr{A})$, that is $\Pi(\mathfrak{M},\mathscr{A})$ is closed under the negation map and it easily follows that $\Pi(\mathfrak{M},\mathscr{A})$ is a sublogic of $\Pi(\mathscr{A})$. The proof is complete. Notice that conditions (ii) and (iv) of Theorem 12 specify that the proposition $p_0(A)$ is confirmed (respectively, refuted) by precisely those outcomes in M that confirm (respectively, refute) the corresponding proposition $p(A)$.

Given a manual $\mathscr{A}$, it is now natural to inquire whether every sublogic of $\Pi(\mathscr{A})$ is of the form $\Pi(\mathfrak{M},\mathscr{A})$ for some induced submanual $\mathfrak{M}$ of $\mathscr{A}$. Consideration of the classical case in which $\mathscr{A}$ consists of a single operation quickly reveals that the answer is no. However, for the Boolean manual $\mathscr{A}$ of example I[13] it does turn out that every sublogic is so induced. As we shall see, this is a consequence of the fact that every nonzero proposition in the logic $\Pi(\mathscr{A})$ of the manual $\mathscr{A}$ of example I is *elementary* in sense that it can be written in the form $p(\{x\})$ for some outcome x. Thus, let us define an *elementary manual* to be a manual $\mathscr{A}$ such that every nonzero proposition in $\Pi(\mathscr{A})$ is elementary. It turns out, as we shall see in Example IX in the next section, that the condition that a manual be elementary can always be satisfied in practice.

Theorem 13: Let $\mathscr{A}$ be an elementary manual and let L be a sublogic of $\Pi(\mathscr{A})$. Let $\mathscr{A}_L = \{E \in \mathscr{A} \mid$ for every $D \subseteq E$, $p(D) \in L\}$. Then $\mathscr{A}_L$ is an induced submanual of $\mathscr{A}$, $\mathscr{A}_L$ is an elementary manual in its own right, and $\Pi(\mathscr{A}_L,\mathscr{A}) = L$.

Proof: Since $1 \in L \subseteq \Pi(\mathscr{A})$ and $\mathscr{A}$ is elementary, there exists an outcome e for $\mathscr{A}$ with $p(\{e\}) = 1$. Hence, $\{e\} \in \mathscr{A}_L$, so $\mathscr{A}_L$ is not empty. We now claim that if $A \subseteq E \in \mathscr{A}_L$ and $B \subseteq F \in \mathscr{A}_L$ with $A \perp B$, then there exists $G \in \mathscr{A}_L$ with $A \cup B \subseteq G$. In fact, since $p(A) \perp p(B)$ with $p(A), p(B) \in L$, then, since L is a sublogic of $\Pi(\mathscr{A})$, we must have $p(A \cup B) = p(A) \oplus p(B) \in L$. Moreover, there exists an event C with $p(C) \in L$ and $p(A \cup B) \oplus p(C) = 1$. Since $\mathscr{A}$ is elementary, there exists an outcome c for $\mathscr{A}$ with $p(C) = p(\{c\})$. Put $G = A \cup B \cup \{c\}$, and note that $G \in \mathscr{A}$. In order to show that $G \in \mathscr{A}_L$, we choose any $D \subseteq G$. Then $p(D) = p(D \cap A) \oplus p(D \cap B) \oplus p(D \cap \{c\})$. Since $D \cap \{c\}$ is either empty or equal to $\{c\}$, then $p(D \cap \{c\})$ belongs to L along with $p(D \cap A)$ and $p(D \cap B)$. Since L is closed under finite orthogonal joins, it follows that $p(D) \in L$; hence, $G \in \mathscr{A}_L$ as required. Put $X_L = \cup \mathscr{A}_L$ and define, for $x, y \in X_L$, $x \# y \Leftrightarrow x \neq y$ and there exists $E \in \mathscr{A}_L$ with $x, y \in E$. Note that $x \# y \Rightarrow x \perp y$; hence, for $A, B \subseteq X_L$, $A \# B \Rightarrow A \perp B$. The coherence of $\mathscr{A}_L$ is now an immediate consequence of the above claim, so $\mathscr{A}_L$ is a submanual of $\mathscr{A}$. To show that $\mathscr{A}_L$ is an induced submanual of $\mathscr{A}$, we choose $x, y \in X_L$ with $x \perp y$. In the above

claim, again, let $A=\{x\}$, $B=\{y\}$ and conclude that $x\#y$. Now, to show that $\mathscr{A}_L$ is elementary, let D be an $\mathscr{A}_L$-event. Thus, there exists $E\in\mathscr{A}_L$ with $D\subseteq E$. Since $\mathscr{A}$ is elementary, there exist outcomes x and y for $\mathscr{A}$ with $p(\{x\})=p(D)\in L$ and $p(\{y\})=p(E\setminus D)\in L$. Evidently, $\{x,y\}\in\mathscr{A}$. As in the proof of the above claim, $\{x,y\}\in\mathscr{A}_L$, so $x\in X_L$. Since $\mathscr{A}_L$ is an induced submanual of $\mathscr{A}$, then $x^{\#}=x^{\perp}\cap X_L=D^{\perp}\cap X_L=D^{\#}$. Consequently, $\mathscr{A}_L$ is an elementary manual.

Clearly, $\Pi(\mathscr{A}_L,\mathscr{A})\subseteq L$. Conversely, choose $p(A)\in L$, where A is an $\mathscr{A}$-event. There exists an $\mathscr{A}$-event B with $p(B)\in L$ and $p(A)\oplus p(B)=1$. As above, there exist outcomes a and b for $\mathscr{A}$ with $p(\{a\})=p(A)$ and $p(\{b\})=p(B)$ and so $\{a,b\}\in\mathscr{A}_L$. Hence, $p(A)=p(\{a\})\in\Pi(\mathscr{A}_L,\mathscr{A})$, and the proof is complete.

VI. REFINEMENT AND COARSENING OF OPERATIONS

As we pointed out in (Ref. 13), a manual of physical operations is often treated classically by imagining that there exists a single "grand canonical operation" that—in some sense—simultaneously refines all of the operations in the given manual. "Hidden variable theories" could be interpreted as attempts to secure such an operation.[26-28] The notion of refinement, as well as the converse notion of coarsening, will be introduced formally in the present section and some of its basic properties will be explored.

Let E and F be operations in the manual $\mathscr{A}$. We shall say that F *refines* E relative to the manual $\mathscr{A}$ and write $E\sqsubseteq F$ provided that for each outcome $e\in E$ there exists an event $D\subseteq F$ such that $p(\{e\})=p(D)$. If $E\sqsubseteq F$, we shall also say that E is a *coarsening* of F.

Example VIII: Let Z be a nonempty set. By a *partition* of Z, we mean, as usual, a collection $\{Z_j\mid j\in J\}$ of nonempty subsets of Z such that $Z_i\cap Z_j=\emptyset$ for $i,j\in J$ with $i\neq j$ and such that $Z=\cup\{Z_j\mid j\in J\}$. Let $\mathscr{A}$ be the set of all partitions of Z. Then $\mathscr{A}$ is a Boolean manual and, for $E,F\in\mathscr{A}$, $E\sqsubseteq F$ if and only if the partition F refines the partition E in the customary sense that every set in E is a union of sets in F. Notice that the operation $Z^*=\{\{z\}\mid z\in Z\}\in\mathscr{A}$ is effective as a common refinement of all of the operations in the manual $\mathscr{A}$.

Theorem 14: Let $\mathscr{A}$ be a manual and let $E,F\in\mathscr{A}$. Consider the following conditions:

(i) $E\sqsubseteq F$.

(ii) $A\subseteq E\Rightarrow$ there exists $B\subseteq F$ with $p(A)=p(B)$.

(iii) $\Pi(\{E\},\mathscr{A})\subseteq\Pi(\{F\},\mathscr{A})$.

(iv) F tests $p(D)$ for all $D\subseteq E$.

(v) F tests $p(\{e\})$ for all $e\in E$.

(vi) $f\in F\Rightarrow$ there exists $e\in E$ with $p(\{f\})\leq p(\{e\})$.

(vii) $e\in E\Rightarrow$ there exists $f\in F$ with $p(\{f\})\leq p(\{e\})$.

Then, conditions (i), (ii) and (iii) are mutually equivalent. Also, conditions (iv), (v) and (vi) are mutually equivalent. Condition (iii) implies condition (iv) and condition (vi) implies condition (vii). If $\mathscr{A}$ is a Dacey manual, then conditions (i) through (vi) are mutually equivalent.

Proof: Let $A\subseteq E$ and assume that condition (i) holds. For each $e\in E$, choose $F_e\subseteq F$ such that $p(\{e\})=p(F_e)$. If we put $B=\cup\{F_e\mid e\in A\}$, we will have $B\subseteq F$ and $p(A)=p(B)$; hence, (i)$\Rightarrow$(ii). That (ii)$\Rightarrow$(iii) and that (iii)$\Rightarrow$(i) is clear. Hence, (i)$\Leftrightarrow$(ii)$\Leftrightarrow$(iii). We shall now show that (iii)$\Rightarrow$(iv). Assume (iii) and let $D\subseteq E$. By (iii), there exists $B\subseteq F$ with $p(D)=p(B)$. Since $B\subseteq F$, then F tests $p(B)=p(D)$; hence, (iii)$\Rightarrow$(iv).

Evidently, (iv)$\Rightarrow$(v). We now show that (v)$\Rightarrow$(vi). Assume (v) and let $f\in F$. Since $f\notin\emptyset=E^{\perp}$, there exists $e\in E$ with $f\notin\{e\}^{\perp}$. But, by (v), $f\in F\subseteq\{e\}^{\perp\perp}\cup\{e\}^{\perp}$; hence, $f\in\{e\}^{\perp\perp}$, $p(\{f\})\leq p(\{e\})$. This proves that (v)$\Rightarrow$(vi). We now claim that (vi)$\Rightarrow$(iv). To see this, assume (vi) and let $D\subseteq E$. We must prove that $F\subseteq D^{\perp\perp}\cup D^{\perp}$. Thus, choose an arbitrary $f\in F$. By (vi), there exists $e\in E$ with $f\in\{e\}^{\perp\perp}$. If $e\in D$, then $f\in\{e\}^{\perp\perp}\subseteq D^{\perp\perp}$, while if $e\in E\setminus D$, then, $f\in\{e\}^{\perp\perp}\subseteq D^{\perp}$. This proves that (vi)$\Rightarrow$(iv). Hence, (iv)$\Leftrightarrow$(v)$\Leftrightarrow$(vi). We shall now show that (vi)$\Rightarrow$(vii). Assume (vi) and suppose that $e\in E$. Since $e\notin\emptyset=F^{\perp}$, there exists $f\in F$ with $f\notin\{e\}^{\perp}$. By (vi), there exists $e_0\in E$ with $f\in\{e_0\}^{\perp\perp}$. If $e_0\neq e$, then we would have $f\in\{e_0\}^{\perp\perp}\subseteq\{e\}^{\perp}$, a contradiction. Hence, $e_0=e$ and $f\in\{e\}^{\perp\perp}$, that is, $p(\{f\})\leq p(\{e\})$. Thus, (vi)$\Rightarrow$(vii).

Now, assume that $\mathscr{A}$ is a Dacey manual. In order to show that conditions (i) through (vi) are mutually equivalent it will suffice to show that (iv)$\Rightarrow$(ii). Thus, assume (iv) and let $A\subseteq E$. By (iv), F tests $p(A)$. By theorem 8, part (i), there exists $B\subseteq F$ with $p(B)=p(A)$. This proves (iv)$\Rightarrow$(ii) when $\mathscr{A}$ is a Dacey manual, and the proof of our theorem is complete.

The equivalence of conditions (i) and (ii) in Theorem 14 shows immediately that the relation $\sqsubseteq$ is transitive on $\mathscr{A}$, that is, if $E,F,G\in\mathscr{A}$ with $E\sqsubseteq F$ and $F\sqsubseteq G$, then $E\sqsubseteq G$. Since it is clear that any operation is a refinement of itself, then $(\mathscr{A},\sqsubseteq)$ is a *quasiordered set* in the sense of (Ref. 20, p. 20).

In the following example, we shall show that any manual can be replaced by a manual $\mathscr{A}^*$ that contains all possible coarsenings of the operations in $\mathscr{A}$. In order to effect this, it is only necessary to promote the nonempty $\mathscr{A}$-events to the status of outcomes.

Example IX: Given a manual $\mathscr{A}$, we define the *event saturation* of $\mathscr{A}$ to be the manual $\mathscr{A}^*$ consisting of all partitions of the operations in $\mathscr{A}$. Then, each operation $E\in\mathscr{A}$ can be canonically identified with the corresponding $E^*=\{\{x\}\mid x\in E\}\in\mathscr{A}^*$. The operations in $\mathscr{A}^*$ accordingly represent coarsenings of the operations in $\mathscr{A}$ and every such coarsening appears explicitly in $\mathscr{A}^*$. Furthermore, $\mathscr{A}^*$ is an elementary manual and $\Pi(\mathscr{A}^*)$ is canonically isomorphic to $\Pi(\mathscr{A})$. Notice that the outcomes for $\mathscr{A}^*$ are precisely the nonempty $\mathscr{A}$-events. The manual $\mathscr{A}^*$ has a number of interesting and significant induced submanuals. In particular, if we are concerned only with finite (respectively, countable) coarsenings of the operations in $\mathscr{A}$, we can look at the submanual $\mathscr{A}_0^*$ (respectively, $\mathscr{A}_\infty^*$) consisting only of the finite (respectively, finite or countable) operations in $\mathscr{A}^*$. Note that $\mathscr{A}_0^*$ and $\mathscr{A}_\infty^*$ are also elementary and that $\Pi(\mathscr{A}_0^*,\mathscr{A}^*)=\Pi(\mathscr{A}_\infty^*,\mathscr{A}^*)=\Pi(\mathscr{A}^*)$.

In the passage from $\mathscr{A}$ to $\mathscr{A}^*$ we have lost nothing of physical significance. Moreover, $\mathscr{A}^*$ is the "elementary

version" of the manual $\mathscr{A}$ promised in the preceding section; clearly, in practice, it is always available.

Let $\mathscr{A}$ be a manual and let $\sqsubseteq$ be the refinement relation on $\mathscr{A}$. If $\mathfrak{M}$ is a subset of $\mathscr{A}$ and if there exists an operation E *in* $\mathfrak{M}$ such that, for every operation $F \in \mathfrak{M}$, $F \sqsubseteq E$, then we would surely be prepared to say that E is effective as a "grand canonical operation" for the set $\mathfrak{M}$. In practice, however, such a stringent operational condition can rarely be achieved. A more realistic condition is obtained if we merely require that, given any two operations E and F in $\mathfrak{M}$ there exists an operation G, *also in* $\mathfrak{M}$, such that $E, F \sqsubseteq G$. If $\mathfrak{M}$ is a nonempty subset of $\mathscr{A}$ satisfying the latter condition, we shall follow customary mathematical usage and say that $\mathfrak{M}$ is *upward directed* by refinement. Given such an upward directed set $\mathfrak{M}$, nothing is lost if we adjoin to $\mathfrak{M}$ all of the available coarsenings of the operations already in $\mathfrak{M}$. These considerations lead us to define a *refinement ideal* in the manual $\mathscr{A}$ to be a nonempty subset $\mathfrak{M}$ of $\mathscr{A}$ which is upward directed by refinement and which has the property that if $F \in \mathfrak{M}$ and if $E \in \mathscr{A}$ with $E \sqsubseteq F$, then $E \in \mathfrak{M}$.

Theorem 15: Let $\mathfrak{M}$ be a refinement ideal in the manual $\mathscr{A}$. Then, $\mathfrak{M}$ is an induced Boolean submanual of $\mathscr{A}$ and the Boolean algebra $\Pi(\mathfrak{M})$ is isomorphic to the Boolean sublogic $\Pi(\mathfrak{M},\mathscr{A})$ of $\Pi(\mathscr{A})$.

Proof: Let $E, F \in \mathfrak{M}$ and let $A \subseteq E$, $B \subseteq F$. Since $\mathfrak{M}$ is a refinement ideal, there exists an operation $G \in \mathfrak{M}$ with $E, F \sqsubseteq G$. By Theorem 14, there exist $A_1, B_1 \subseteq G$ such that $p(A) = p(A_1)$ and $p(B) = p(B_1)$. Put $A_0 = A_1 \setminus B_1$, $B_0 = B_1 \setminus A_1$ and $D = A_1 \cap B_1$, so that $p(A_0)$, $p(B_0)$ and $p(D)$ are mutually orthogonal propositions in $\Pi(\mathfrak{M},\mathscr{A})$, $p(A) = p(A_0) \oplus p(D)$ and $p(B) = p(B_0) \oplus p(D)$.

Suppose first that $A \perp B$. Then, $p(A) \oplus p(B) = p(A_1) \oplus p(B_1)$. Let $C = G \setminus (A_1 \cup B_1)$, $H = A \cup B \cup C$, noting that, by the coherence of $\mathscr{A}$, H is an $\mathscr{A}$-event. But, $p(H) = p(A) \oplus p(B) \oplus p(C) = p(A_1) \oplus p(B_1) \oplus p(C) = p(G) = 1$; hence, $H \in \mathscr{A}$. We claim that $H \sqsubseteq G$. To establish this claim, we choose $h \in H$ and prove the existence of an event $D_h \subseteq G$ such that $p(\{h\}) = p(D_h)$. If $h \in A$, we use the fact that $A \subseteq E \sqsubseteq G$ to produce $D_h \subseteq G$ with $p(\{h\}) = p(D_h)$. A similar argument applies if $h \in B$, while, if $h \in C$, we merely put $D_h = \{h\}$. Thus, $H \sqsubseteq G \in \mathfrak{M}$. Thus, since $\mathfrak{M}$ is a refinement ideal, $H \in \mathfrak{M}$. This proves that $\mathfrak{M}$ is actually an induced submanual of $\mathscr{A}$.

We now drop the assumption that $A \perp B$. Let Y be the outcome set for the manual $\mathfrak{M}$ and let $\#$ be the relation of operational rejection induced on Y by $\mathfrak{M}$. Since $\mathfrak{M}$ is an induced submanual of $\mathscr{A}$, then, for $M \subseteq Y$ we have $M^{\#} = M^{\perp} \cap Y$. We now suppose that the two propositions $(A^{\#\#}, A^{\#})$ and $(B^{\#\#}, B^{\#})$ in the logic $\Pi(\mathfrak{M})$ coincide, that is, we suppose that $A^{\#} = B^{\#}$. Since $p(A) = p(A_0) \oplus p(D)$ and $p(B) = p(B_0) \oplus p(D)$, we have $A^{\perp} = (A_0)^{\perp} \cap D^{\perp}$ and $B^{\perp} = (B_0)^{\perp} \cap D^{\perp}$. Intersection of both sides of the latter equations with Y produces $A^{\#} = (A_0)^{\#} \cap D^{\#}$ and $B^{\#} = (B_0)^{\#} \cap D^{\#}$, so that $(A_0)^{\#} \cap D^{\#} = (B_0)^{\#} \cap D^{\#}$. Since $A_0 \subseteq (B_0)^{\#} \cap D^{\#}$, we have $A_0 \subseteq (A_0)^{\#}$, which forces $A_0 = \phi$. Similarly, $B_0 = \phi$, so $p(A) = p(D) = p(B)$. This allows us to define a mapping $\phi\colon \Pi(\mathfrak{M}) \to \Pi(\mathfrak{M},\mathscr{A})$ by $\phi((A^{\#\#}, A^{\#})) = p(A)$ for every $\mathfrak{M}$-event A. That ϕ is an isomorphism is clear. The above considerations also show immediately that $\Pi(\mathfrak{M},\mathscr{A})$ is a Boolean sublogic of $\Pi(\mathscr{A})$. Since $\Pi(\mathfrak{M})$ is isomorphic to the Boolean algebra $\Pi(\mathfrak{M},\mathscr{A})$, it too is a Boolean algebra. The proof is complete.

The preceding theorem naturally raises the following question: If $\mathfrak{M}$ is an induced submanual of the manual $\mathscr{A}$, and if there exists an operation $E \in \mathscr{A}$ which simultaneously refines all of the operations in $\mathfrak{M}$, then is $\mathfrak{M}$ a Boolean submanual? An obvious modification of Example VII shows that, in general, the answer is no.

Suppose now that $\mathscr{A}$ is an elementary manual and that L is a Boolean sublogic of $\Pi(\mathscr{A})$. By Theorem 13, $\mathscr{A}_L = \{E \in \mathscr{A} \mid p(D) \in L \text{ for every } D \subseteq E\}$ is an induced elementary submanual of $\mathscr{A}$ and $\Pi(\mathscr{A}_L,\mathscr{A}) = L$. Furthermore, it is clear that if $F \in \mathscr{A}_L$ and if $E \in \mathscr{A}$ with $E \sqsubseteq F$, then $E \in \mathscr{A}_L$. In view of Theorem 15, it is natural to ask if $\mathscr{A}_L$ is a refinement ideal. Suitable examples show that the answer, in general, is no. If we assume that $\mathscr{A}$ is not only elementary, but *locally finite* in the sense that every operation in $\mathscr{A}$ is finite, then we can prove that $\mathscr{A}_L$ is indeed a refinement ideal.

Theorem 16: Let $\mathscr{A}$ be an elementary and locally finite manual, let L be a Boolean sublogic of $\Pi(\mathscr{A})$ and let $\mathscr{A}_L = \{E \in \mathscr{A} \mid p(D) \in L \text{ for every } D \subseteq E\}$. Then, $\mathscr{A}_L$ is a refinement ideal in $\mathscr{A}$ and $L = \Pi(\mathscr{A}_L,\mathscr{A})$.

Proof: By Theorem 13, we know that $\mathscr{A}_L$ is an induced elementary submanual of $\mathscr{A}$ and that $\Pi(\mathscr{A}_L,\mathscr{A}) = L$. It is easy to see that if $F \in \mathscr{A}_L$ and if $E \in \mathscr{A}$ with $E \sqsubseteq F$, then $E \in \mathscr{A}_L$. We have only to show that if $E, F \in \mathscr{A}_L$, there exists $G \in \mathscr{A}_L$ with $E, F \sqsubseteq G$. Thus, let $E, F \in \mathscr{A}_L$ and choose, for each $e \in E$ and each $f \in F$, an $\mathscr{A}$-event $D(e,f)$ such that $p(D(e,f))$ is the infimum, calculated in the Boolean algebra L, of $p(\{e\})$ and $p(\{f\})$. Because of the distributive laws in the Boolean algebra L, we have $p(\{e\}) = \oplus\{p(D(e,f)) \mid f \in F\}$ for each $e \in E$.

Let $J = \{(e,f) \mid e \in E,\ f \in F \text{ and } D(e,f) \neq \phi\}$. Since $\mathscr{A}$ is elementary, we can choose an $\mathscr{A}$-outcome $x(e,f)$ for each $(e,f) \in J$ such that $p(\{x(e,f)\}) = p(D(e,f))$. Let $G = \{x(e,f) \mid (e,f) \in J\}$. Since $\mathscr{A}$ is locally finite, then G is a finite orthogonal set; hence, by coherence, G is an $\mathscr{A}$-event. We are going to show that $G \in \mathscr{A}$. In order to do this, it will suffice to show that there exists no $\mathscr{A}$-outcome x with $x \in G^{\perp}$. But, if $x \in G^{\perp}$, then $x \in (D(e,f))^{\perp}$ for every $e \in E$ and every $f \in F$. From the equation $p(\{e\}) = \oplus\{p(D(e,f)) \mid f \in F\}$, it follows that $x \perp e$ for every $e \in E$, contradicting (Ref. 13, Lemma 3). We conclude that $G \in \mathscr{A}$.

Since $\mathscr{A}$ is locally finite, then G is a finite set. Thus, if $A \subseteq G$, $p(A) = \oplus\{p(\{x\}) \mid x \in A\} \in L$; hence, $G \in \mathscr{A}_L$. Suppose that $e \in E$. Let $A = \{x(e,f) \mid (e,f) \in J,\ f \in F\}$ and note that $A \subseteq G$ with $p(A) = \oplus\{p(\{x(e,f)\}) \mid (e,f) \in J,\ f \in F\} = \oplus\{p(D(e,f)) \mid f \in F\} = p(\{e\})$. This proves that $E \sqsubseteq G$. A similar argument shows that $F \sqsubseteq G$, and the proof is complete.

Theorem 17: Let $\mathscr{A}$ be a locally finite elementary manual. Then there is a one-to-one correspondence $L \longleftrightarrow \mathscr{A}_L$ between Boolean sublogics L of $\Pi(\mathscr{A})$ and refinement ideals $\mathscr{A}_L$ in $\mathscr{A}$.

Proof: For each Boolean sublogic L of $\Pi(\mathscr{A})$, let $\mathscr{A}_L = \{E \in \mathscr{A} \mid p(D) \in L \text{ for all } D \subseteq E\}$. By Theorem 16, $\mathscr{A}_L$ is a refinement ideal in $\mathscr{A}$ and $\Pi(\mathscr{A}_L,\mathscr{A}) = L$. Let $\mathfrak{M}$ be a refinement ideal in $\mathscr{A}$. We must prove that there exists

a Boolean sublogic L of $\Pi(\mathscr{A})$ with $\mathfrak{M}=\mathscr{A}_L$. By Theorem 15, $L=\Pi(\mathfrak{M},\mathscr{A})$ is a Boolean sublogic of $\Pi(\mathscr{A})$. Furthermore, it is clear that $\mathfrak{M}\subseteq\mathscr{A}_L$. Suppose that $E\in\mathscr{A}_L$. Then, for each $e\in E$, $p(\{e\})\in L=\Pi(\mathfrak{M},\mathscr{A})$; hence, there exists an $\mathfrak{M}$-event D_e such that $p(\{e\})=p(D_e)$. By Theorem 15, $\mathfrak{M}$ is an induced submanual of $\mathscr{A}$. By coherence, and the fact that E is finite, it follows that $\cup\{D_e\mid e\in E\}$ is an $\mathfrak{M}$-event. Put $F=\cup\{D_e\mid e\in E\}$ and note that $p(F)=\oplus\{p(D_e)\mid e\in E\}=\oplus\{p(\{e\})\mid e\in E\}=p(E)=1$; hence, $F\in\mathfrak{M}$. Evidently, $E\subseteq F$; hence, since $\mathfrak{M}$ is a refinement ideal, $E\in\mathfrak{M}$. Thus, $\mathfrak{M}=\mathscr{A}_L$, and the proof is complete.

If we wish to invoke Theorem 17 in order to study the Boolean sublogics of a manual $\mathscr{A}$ which is not necessarily elementary or locally finite, we can always pass to the manual $\mathscr{A}_0^*$ of Example IX, noting that this entails no essential loss of physical content and that the logic $\Pi(\mathscr{A}_0^*)$ is canonically isomorphic to the logic $\Pi(\mathscr{A})$. Thus, suppose that $\mathscr{A}$ is a locally finite elementary manual. Let D be an $\mathscr{A}$-event and choose any operation $E\in\mathscr{A}$ with $D\subseteq E$. Let $\mathfrak{M}=\{H\in\mathscr{A}\mid H\subseteq E\}$, noting that $\mathfrak{M}$ is a refinement ideal in $\mathscr{A}$ and that, by Zorn's lemma, $\mathfrak{M}$ can be extended to a maximal refinement ideal $\mathscr{N}$ of $\mathscr{A}$. By Theorem 17, the corresponding Boolean sublogic $B=\Pi(\mathscr{N},\mathscr{A})$ is a maximal Boolean sublogic of $\Pi(\mathscr{A})$ and $p(D)\in B$. Following Greechie,[29] we refer to a maximal Boolean sublogic of $\Pi(\mathscr{A})$ as a block. Thus, we have the following result:

Theorem 18: Let $\mathscr{A}$ be any manual and let $p(D)$ be any proposition in the logic $\Pi(\mathscr{A})$. Then, there exists at least one block B in $\Pi(\mathscr{A})$ with $p(D)\in B$.

Thus the structure of the logic $\Pi(\mathscr{A})$ of a manual $\mathscr{A}$ is determined by the structure of its blocks and the manner in which they intertwine. The notion of an observable will be introduced in a forthcoming paper in this series, and it will be seen that each observable "lives on" some block. Moreover, the set of all observables that "live on" a given block will provide a maximal classical view of the experimental universe of discourse implicit in the manual of operations under consideration. As a consequence of Theorem 17, it will be seen that the operations that measure these observables are precisely those in the refinement ideal corresponding to the block. The maximal classical views provided by the various blocks are, of course, complementary in the sense of N. Bohr.[30]

[1]P. A. M. Dirac, *The Principles of Quantum Mechanics*, International Series of Monographs on Physics (Oxford U. P., Oxford, London, 1958), 4th ed.
[2]W. Heisenberg, *The Physical Principles of Quantum Theory* (University of Chicago Press, Chicago, 1930).
[3]J. von Neuman, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1955).
[4]J. M. Jaunch, Helv. Phys. Acta **37**, 293 (1964).
[5]Henry Marganeau, Ann. Phys. (Leipz.) **23** (3), 469 (1963).
[6]M. M. Yanase, Am. J. Phys. **32** (3), 208 (1964).
[7]E. P. Wigner, Am. J. Phys. **31**, 6 (1963).
[8]Paul A. Benioff, J. Math. Phys. **13**, 231 (1972).
[9]E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970).
[10]G. Ludwig, Z. Phys. **181**, 233 (1964).
[11]Bogdan Mielnik, Commun. Math. Phys. **15**, 1 (1969).
[12]J. C. T. Pool, Commun. Math. Phys. **9**, 118 (1968).
[13]D. J. Foulis and C. H. Randall, J. Math. Phys. **13**, 1667 (1972).
[14]Garret Birkhoff, *Lattices in Applied Mathematics*, Proceedings Symposia in Pure Mathematics II, Lattice Theory (Amer. Math. Soc., Providence, R. I., 1961), pp. 155–184.
[15]Garret Birkhoff and John von Neuman, Ann. Math. **37**, 823 (1936).
[16]M. D. MacLaren, Argonne National Laboratory Report No. ANL 7065, 1965.
[17]J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[18]V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, 1968), Vol. I.
[19]C. H. Randall and D. J. Foulis, J. Comb. Theory **11**, 157 (1971).
[20]Garret Birkhoff, *Lattice Theory*, American Mathematics Society Colloquium Publication (Amer. Math. Soc., Providence, R. I., 1967), 3rd ed.
[21]G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[22]C. H. Randall and D. J. Foulis, Amer. Math. Monthly 77 (4), 363 (1970).
[23]D. J. Foulis, Portugal. Math. **27**, 65 (1962).
[24]S. S. Holland, Jr., Proc. Am. Math. Soc. **24**, 716 (1970).
[25]B. Jeffcott, "The center of an orthologic," J. Symbolic Logic, **37**, 641 (1972).
[26]John S. Bell, Rev. Mod. Phys. **38**, 447 (1966).
[27]D. Bohm and J. Bub, Rev. Mod. Phys. **38**, 453 (1966).
[28]J. M. Jauch and C. Piron, Helv. Phys. Acta **36**, 827 (1963).
[29]R. J. Greechie, J. Comb. Theory **4**, 210 (1968).
[30]Niels Bohr, *Essays 1958-1962 on Atomic Physics and Human Knowledge* (Interscience, New York, 1963).

Electron in the field of two monochromatic electromagnetic waves

I. Berson and J. Valdmanis

Physics Institute, Latvian SSR, Academy of Sciences, Riga, Salaspils, USSR
(Received 19 January 1972)

The motion of the electron in the classical as well as in the quantized field of two circularly polarized waves which move in opposite directions is investigated. The spin characteristics of the electron are neglected and it is assumed that the electron and one of the waves move in the same direction. In the classical case the problem is reduced to the solution of the general Mathieu equation, but in the quantum case to the solution of an ordinary second-order differential equation with two irregular points. For particular values of the parameters the analytic expressions of the solution are found. The character of the solution essentially depends on the mutual polarization of waves. The conserved quantity for opposite polarization of waves is the energy of the electron, but for the same polarization it is the momentum of the electron. In the latter case the effective potential depends on the time, and the electron–positron pair creation is allowed.

I. INTRODUCTION

In view of the increased experimental possibilities of producing strong electromagnetic fields, there has been more interest in finding fields for which the exact solution of the Dirac equation can be obtained (or of the Klein—Gordon equation in the case of scalar electrodynamics). Apart from the previously known solutions of the Dirac equation (the bibliography is mentioned in Ref. 1), in recent years solutions have also been found for other potentials.[2–4]

On the other hand, quantization of an electromagnetic field involves such difficulties of a mathematical and fundamental nature that beside the traditional perturbation theory there are but few exactly solvable models.[5,6] One of the authors[7,8] has found a solution for the model which, of all the quantized electromagnetic field, allowed only for the photons propagating in one direction. The model is valid at large numbers of photons and the solution obtained is a quantum generalization of Volkov's solution for the electron in the classical field of a plane wave.

The present paper considers the motion of an electron in the field of two monochromatic electromagnetic waves. A similar problem has already been considered in connection with the Kapitza—Dirac effect. In this process one usually proceeds from the Schrödinger equation in which the vector potential is time-averaged.[9,10] For Dirac's equation this problem was considered by Sen Gupta.[11] We shall treat both the case of the classical field of two electromagnetic waves and that of the quantized field. The latter leads to equations which are of interest from the point of view of mathematics. For simplicity, we shall neglect the spin properties of the electron, using the Klein—Gordon equation throughout.

II. ELECTRON IN THE CLASSICAL FIELD OF TWO ELECTROMAGNETIC WAVES

The Klein—Gordon equation has the form (the metrics and the system of units are the same as in Ref. 12)

$$\left[\left(i\frac{\partial}{\partial x_\mu}+eA_\mu\right)^2+m^2\right]\Psi=0, \tag{1}$$

where e and m are the charge and mass of the electron, A_μ is the vector potential of the electromagnetic field.

We shall use the coordinate system in which the waves propagate in opposite directions with equal frequencies. Suppose that this direction coincides with the z axis. One can always pass to such a coordinate system with the help of a Lorentz transformation, excepting the case when both waves propagate in the same direction. Further, we shall confine ourselves only to the case when the electron moves along the z axis and both waves have circular polarization. The vector potential of such waves can be written as

$$A^{\pm}_\mu=a_1[e_{1\mu}\cos(k_1x)+e_{2\mu}\sin(k_1x)]+a_2[e_{1\mu}\cos(k_2x) \pm e_{2\mu}\sin(k_2x)], \tag{2}$$

where a_1 and a_2 are the wave amplitudes and the unit polarization vectors $e_{1\mu}$ and $e_{2\mu}$ and the wave vectors k_1 and k_2 satisfy the relations $(e_1e_2)=(e_ik_j)=k_1^2=k_2^2=0$. The "+" and "−" signs indicate, respectively, opposite and similar circular polarizations of both waves.

Now if expression (2) is substituted in Eq. (1), we shall find that with opposite polarization of the waves the wave function is $\Psi=\varphi_1(z)\exp(-iEt)$, where φ_1 satisfies the equation

$$\left[\frac{d^2}{dz^2}+E^2-m^2-e^2(a_1^2+a_2^2)-2e^2a_1a_2\cos2\omega z\right]\varphi_1=0. \tag{3}$$

With similar circular polarization the wave function is $\Psi=\varphi_2(t)\exp(ip_zz)$ and φ_2 satisfies the equation

$$\left[\frac{d^2}{dt^2}+p_z^2+m^2+e^2(a_1^2+a_2^2)+2e^2a_1a_2\cos2\omega t\right]\varphi_2=0. \tag{4}$$

Here E and p_z are the energy and the momentum of the electron, ω is the frequency of the electromagnetic waves, t is the time.

It is noteworthy that in both cases the variables are separable. This property vanishes if one passes to linearly polarized fields and takes into account that component of the momentum of the electron which is perpendicular to the direction of propagation of the waves. As is seen, the electron energy is conserved and the effective potential is periodic in z in the first case, and in the second case the momentum of electron is conserved along the z axis and the effective potential is periodic in time. Equations (3) and (4) are Mathieu equations.[13] We are concerned with stationary solutions in the form of propagating waves, so that

$$\varphi_1 = \exp(ip_z' z) \sum_{n=-\infty}^{\infty} a_n \exp(i\omega n z),$$
$$\varphi_2 = \exp(-iE't) \sum_{n=-\infty}^{\infty} b_n \exp(-i\omega n t), \tag{5}$$

where p_z' and E' are real characteristic exponents and the sums are periodic functions in z and t with period of $2\pi/\omega$. However, the stationary solutions (5) exist not for arbitrary values of the parameters, but only in the so-called stability zones. Along with these zones there exist forbidden zones with respect to energy for Eq. (3) and with respect to momentum for Eq. (4). In the first case the physical picture corresponds qualitatively to the motion of an electron in a one-dimension lattice in a solid. For the second case a physically analogous problem is difficult to imagine. It should be noted that since in case (5) n changes from $-\infty$ to ∞ the function φ_2 will always contain terms both with positive and negative frequency. This means that pair production is observed in the second case. Apparently, it will be most intense when E' becomes a multiple of ω. In practically produced fields, however, $e^2 a_1 a_2/m^2 \ll 1$; therefore, in the second case the electron motion is almost free, and only weakly perturbed by the field. Besides, for optical frequencies $\omega \ll 2m$ pairs will be produced only in a very high order of perturbation theory, since to produce them many photons must be absorbed from both waves.

III. ELECTRON IN THE QUANTIZED FIELD OF TWO ELECTROMAGNETIC WAVES

Quantization of the electromagnetic field leads to the vector potential A_μ becoming an operator and instead of expression (2) we have[12]

$$A_\mu^{\pm} = \frac{1}{\sqrt{2\omega\Omega}} \{e_\mu^- c_1 \exp[i(k_1 x)] + e_\mu^+ c_1^+ \exp[-i(k_1 x)]\} + \frac{1}{\sqrt{2\omega\Omega}} \{e_\mu^{\mp} c_2 \exp[i(k_2 x)] + e_\mu^{\pm} c_2^+ \exp[-i(k_2 x)]\}, \tag{6}$$

where

$$e_\mu^- = \frac{e_{1\mu} - ie_{2\mu}}{\sqrt{2}}, \quad e_\mu^+ = \frac{e_{1\mu} + ie_{2\mu}}{\sqrt{2}}.$$

Ω is the normalization volume. For the annihilation operator c_i and the creation operator c_i^+ for photons we shall choose the following representation:

$$c_i = (1/\sqrt{2})\left(\xi_i + \frac{\partial}{\partial \xi_i}\right), \quad c_i^+ = (1/\sqrt{2})\left(\xi_i - \frac{\partial}{\partial \xi_i}\right). \tag{7}$$

The upper (lower) signs on the vectors e_μ in the second set of parentheses of expression (6) correspond to opposite (similar) circular polarization of both waves.

Now, let us substitute expressions (6) and (7) in Eq. (1) and apply to the latter the canonical transformation used by one of the authors in Ref. 7. Then it can be easily shown that the wave function Ψ has the form

$$\Psi = \exp\left[i(qx) + \frac{i(k_1 x)}{2}\left(\frac{\partial^2}{\partial \xi_1^2} - \xi_1^2 + 1\right) + \frac{i(k_2 x)}{2}\left(\frac{\partial^2}{\partial \xi_2^2} - \xi_2^2 + 1\right)\right]\varphi, \tag{8}$$

where $q_\mu(0, 0, q_z, iq_0)$ is the total momentum of the system under consideration and the function φ depends only on the variables ξ_1 and ξ_2 and satisfies the equation

$$\left[\frac{(k_1 k_2)}{2}\left(\frac{\partial^2}{\partial \xi_1^2} - \xi_1^2 + 1\right)\left(\frac{\partial^2}{\partial \xi_2^2} - \xi_2^2 + 1\right) + [(qk_1) - h]\left(\frac{\partial^2}{\partial \xi_1^2} - \xi_1^2 + 1\right) + [(qk_2) - h]\left(\frac{\partial^2}{\partial \xi_2^2} - \xi_2^2 + 1\right) + 2h\left(\xi_1 \xi_2 \mp \frac{\partial^2}{\partial \xi_1 \partial \xi_2}\right) + q^2 + m^2 + 2h\right]\varphi = 0. \tag{9}$$

Here $h = e^2/2\omega\Omega$.

In deriving Eq. (9) the electron and photons are assumed to propagate along the z axis. In order not to deal with the oscillator's zero-point energy we have introduced the term $i(k_1 x)/2 + i(k_2 x)/2$ in the index of the exponent (8). Below we shall consider the two cases of polarization separately.

A. Opposite circular polarization

In this case we have the "+" sign in the last round brackets of Eq. (9) and it is easily seen that the operator for the total number of photons S commutes with the entire square bracket of Eq. (9):

$$S = -\frac{1}{2}\left(\frac{\partial^2}{\partial \xi_1^2} - \xi_1^2 + 1\right) - \frac{1}{2}\left(\frac{\partial^2}{\partial \xi_2^2} - \xi_2^2 + 1\right). \tag{10}$$

We shall denote the eigenvalues and eigenfunctions of the operator S by s and φ_s. The quantity s can be a nonnegative integer. Since the total energy and the energy of electromagnetic field ωs are conserved, the electron energy is conserved too, which fully agrees with the case of nonquantized waves having opposite polarization.

Since the total number of photons is a good quantum number, the solution of Eq. (9) decreasing for $\xi_1, \xi_2 \to \pm\infty$ can be sought as a finite series of undisturbed oscillator wave functions

$$\varphi_s = \exp\left(-\frac{\xi_1^2}{2} - \frac{\xi_2^2}{2}\right) \sum_{n=0}^{s} C_n^s H_n(\xi_1) H_{s-n}(\xi_2), \tag{11}$$

where H_n are Hermite polynomials. If one substitutes expression (11) in Eq. (9) and takes into account known relations between adjacent Hermite polynomials the following recurrence relations for the coefficients C_n^s are obtained:

$$[n(s-n) + d_1 n + \lambda_s] C_n^s - b[(s-n+1)C_{n-1}^s + (n+1)C_{n+1}^s] = 0, \tag{12}$$

$n = 0, 1, \ldots, s$, $C_n^s = 0$ if $n < 0$ or $n > s$.

Here the following notation is introduced

$$d_1 = \frac{(qk_2) - (qk_1)}{(k_1 k_2)} = \frac{q_z}{\omega}, \quad b = -\frac{h}{(k_1 k_2)} = \frac{h}{2\omega^2},$$
$$\lambda_s = \frac{q_0^2 - q_z^2 - m^2 - 2h - 2s[\omega(q_z + q_0) + h]}{4\omega^2}. \tag{13}$$

In order to obtain nontrivial values for the $s+1$ coefficients C_n^s, it is necessary to equate to zero the determinant composed of the coefficients at C_n^s. Hence for any s one can obtain $s+1$ real and different roots $\lambda_{s\nu}$ and construct the corresponding orthonormalized functions $\varphi_{s\nu}(\xi_1, \xi_2)$ where $\nu = 0, 1, \ldots$. Knowing $\lambda_{s\nu}(q_z, \omega, h)$ and taking into account (13) we find that the total energy of the system is equal to

$$q_0 = s\omega \pm [(q_z + s\omega)^2 + m^2 + 2h(s+1) + 4\omega^2\lambda_{s\nu}]^{1/2}, \quad (14)$$

i.e., the sum of the photon energy and the electron energy. Due to the interaction with photons the latter is different from the expression for energy of a free electron.

We shall give expressions for $\lambda_{s\nu}$, the total energy of the system q_0 and normalized eigenfunctions $\varphi_{s\nu}$ at $s=0$ and $s=1$:

$s=0$, $\lambda_{00}=0$,

$$q_0 = \pm(m^2 + q_z^2 + 2h)^{1/2},$$

$$\varphi_{00} = \frac{1}{\sqrt{\pi}} \exp\left(-\frac{\xi_1^2}{2} - \frac{\xi_2^2}{2}\right);$$

$s=1$, $\nu=0;1$, $\lambda_{1\nu} = -\frac{d_1}{2} + (-1)^\nu\left(\frac{d_1^2}{4} + b^2\right)^{1/2}$,

$$q_0 = \omega \pm [m^2 + q_z^2 + \omega^2 + 4h + (-1)^\nu 2(\omega^2 q_z^2 + h^2)^{1/2}]^{1/2},$$

$$\varphi_{1\nu} = \frac{2(b\xi_2 - \lambda_{1\nu}\xi_1)}{[2\pi(b^2 + \lambda_{1\nu}^2)]^{1/2}} \exp\left(-\frac{\xi_1^2}{2} - \frac{\xi_2^2}{2}\right).$$

We note the following circumstance. One can construct from the coefficients C_n^s a polynomial of degree s in a certain variable x

$$F_s = \sum_{n=0}^{s} C_n^s x^n \quad (15)$$

and taking into account the recurrence relations (12) one can derive the differential equation which is satisfied by this polynomial:

$$\left(x^2\frac{d^2}{dx^2} - [bx^2 + (d_1 + s - 1)x - b]\frac{d}{dx} + bsx - \lambda_s\right)F_s = 0. \quad (16)$$

Polynomial solutions of the type of Eq. (16) were recently investigated by Pham Ngoc Dinh.[14] He also proved that the polynomials $F_{s\nu}(x)$ are orthogonal on a unit circle.

B. Similar circular polarizations

In this case we have the "+" sign in the last round brackets of Eq. (9) and the operator L describing the difference in the number of photons propagating in opposite directions commutes with the whole bracket of Eq. (9).

$$L = \frac{1}{2}\left(\frac{\partial^2}{\partial\xi_2^2} - \xi_2^2\right) - \frac{1}{2}\left(\frac{\partial^2}{\partial\xi_1^2} - \xi_1^2\right). \quad (17)$$

We shall denote the eigenvalues and eigenfunctions of this operator by l and φ_l. The quantity l may be any integer. Since the total momentum of the system and the momentum of photons ωl are conserved, the momentum of the electron along the z axis is conserved too, which is in agreement with the case of nonquantized waves with similar polarization.

The solution of Eq. (9) will be sought again as a series in undisturbed oscillator functions. However, now this series will be infinite because now it is not the total number of photons that is limited, but rather the difference in the number of photons propagating in opposite directions. At $l \geq 0$ we have

$$\varphi_l = \exp\left(-\frac{\xi_1^2}{2} - \frac{\xi_2^2}{2}\right) \sum_{n=0}^{\infty} C_n^l H_n(\xi_1) H_{l+n}(\xi_2). \quad (18)$$

Substitution of expression (18) in Eq. (9) leads to the following recurrence relations for the coefficients

$$[n(n+l) + d_2 n + \lambda_l]C_n^l - b[\tfrac{1}{2}C_{n-1}^l + 2(n+1)(n+l+1)C_{n+1}^l] = 0, \quad (19)$$

where $n=0,1,\cdots$, $C_n^l = 0$ if $n<0$,

$$d_2 = -\frac{(qk_1) + (qk_2) + 2h}{(k_1k_2)} = -\frac{q_0}{\omega} - 2b, \quad (20)$$

and λ_l has the same form as λ_s with l instead of s. At $l \leq 0$ the sum (18) begins with $|l|$, but the recurrence relations obtained are the same as (19) with $|l|$ instead of l and the opposite sign before q_z.

Equating to zero the infinite determinant of the coefficients at C_n^l, one can find the eigenvalues of $\lambda_{l\nu}$ where $\nu = 0, 1, \cdots$. Now expression (14) is valid with l instead of s. But since d_2 and consequently $\lambda_{l\nu}$ depends on q_0, it is in fact the equation determining the total energy of the system in terms of the other parameters.

If instead of C_n^l we introduce

$$r_n^l = C_n^l 2^n(n+1), \quad (21)$$

the recurrence relations will become more symmetrical:

$$[n(n+l) + d_2 n + \lambda_l]r_n^l - b[(n+l)r_{n-1}^l + (n+1)r_{n+1}^l] = 0. \quad (22)$$

As in the first section one can introduce the functions F_l and G_l of a certain variable x,

$$F_l = \sum_{n=0}^{\infty} C_n^l x^n, \quad G_l = \sum_{n=0}^{\infty} r_n^l x^n, \quad (23)$$

and taking into account the relations (19) and (22) one can derive the differential equations which they satisfy:

$$\left((x^2 - 2bx)\frac{d^2}{dx^2} + [(d_2 + l + 1)x - 2b(l+1)]\frac{d}{dx} - \frac{bx}{2} + \lambda_l\right)F_l = 0, \quad (24)$$

$$\left(x^2\frac{d^2}{dx^2} - [bx^2 - (d_2 + l + 1)x + b]\frac{d}{dx} - b(l+1)x + \lambda_l\right)G_l = 0. \quad (25)$$

Equation (24) has two regular singular points at $x=0$ and $x=2b$ and one essential singularity at infinity. Equation (25) has two essential singularities at zero and infinity and differs from Eq. (16) only in signs of some coefficients. We shall dwell only on Eq. (25). According to the general theory of equations with two essential singular points[15] Eq. (25) belongs to the type of equations which have one essential singular solution and the second solution, which is just that we are interested in, can be represented as a series (23). This series is asymptotic and becomes convergent only when λ_l is equal to the eigenvalue $\lambda_{l\nu}$.

At small b the eigenvalues and eigenfunctions can be sought as

$$\lambda_l = \sum_{m=0}^{\infty} b^m \lambda_l^m, \quad G_l = \sum_{m=0}^{\infty} b^m g_l^m(x). \quad (26)$$

Accurate to within b^2,

$$\lambda_{l\nu} = -\nu(\alpha - \nu) + b^2\left(\frac{\nu(\nu+l)}{1-\alpha} + \frac{(\nu+1)(\nu+l+1)}{1+\alpha}\right) + \cdots,$$

$$G_{l\nu}=x^{\nu}+b\left(\frac{\nu x^{\nu-1}}{1-\alpha}+\frac{\nu+l+1}{1+\alpha}\,x^{\nu+1}\right)$$
$$+b^2\left(\frac{\nu(\nu-1)}{2(1-\nu)(2-\nu)}\,x^{\nu-2}+\frac{(\nu+l-1)(\nu+l+2)}{2(1+\alpha)(2+\alpha)}\,x^{\nu+2}\right)+\cdots, \tag{27}$$

where $\alpha=2\nu+d_2+l$, $\nu=0,1,\cdots$. We shall give one more expression for λ_{l0} at $\nu=0$ accurate to within b^6:

$$\lambda_{l0}=b^2\frac{(l+1)}{d_2+l+1}+b^4\frac{d_2(l+1)}{(d_2+l+1)^3(d_2+l+2)}$$
$$+b^6\frac{2d_2(l+1)(d_2-l-1)}{(d_2+l+1)^5(d_2+l+2)(d_2+l+3)}+\cdots. \tag{28}$$

Here d_2 is determined according to (20) and consequently depends on b too.

In conclusion, we shall consider some exact solutions of Eq. (25) at special values of parameter d_2 without being concerned with the physical meaning of these solutions. It can be easily verified that at $d_2=0$ the solution of Eq. (25) is

$$G_{l0}=e^{bz},\quad \lambda_{l0}=b^2. \tag{29}$$

That this solution corresponds to the state with $\nu=0$ is seen from the fact that at $d_2=0$ the series (28) is finite and becomes equal to b^2. It is noteworthy that these eigenvalues are the same for all l. If one takes into account the relations (18), (21), and (23) one can obtain the following expression for the wave function

$$\varphi_{l0}=N\exp\left(-\frac{\xi_1^2}{2}-\frac{\xi_2^2}{2}\right)\sum_{n=0}^{\infty}\frac{(b/2)^n}{n!\,(n+l)!}\,H_n(\xi_1)H_{n+l}(\xi_2), \tag{30}$$

where the normalization constant N is

$$N=[b^l/\pi 2^l I_l(2b)]^{1/2}. \tag{31}$$

$I_l(2b)$ is the modified Bessel function of the first kind. If further $1/(n+l)!$ is represented as circuital integral round zero and the generalized expression of Mehler[16] for Hermite polynomials is used, we shall get an integral presentation for the function

$$\varphi_{l0}=N\exp\left(-\frac{\xi_1^2}{2}-\frac{\xi_2^2}{2}\right)\frac{b^{-l}}{2\pi i}\int_c d\tau\,\frac{H_l[(\xi_2\tau-\xi_1)/(\tau^2-1)^{1/2}]}{(\tau^2-1)^{(l+1)/2}}$$
$$\times\exp\left(b\tau+\frac{2\xi_1\xi_2\tau-\xi_1^2-\xi_2^2}{\tau^2-1}\right). \tag{32}$$

Integration here is over a closed circuit round both essential singular points $\tau=\pm 1$.

Solutions in closed form which correspond to another ν at $d_2=0$ failed to be found. But if d_2 is a negative integer, for instance $d_2=-k$, one can find $k+1$ solutions which are the exponents (29) multiplied by the polynomial of k degree. The eigenvalues of $\lambda_{1\nu}$ are determined from the $(k+1)$th order determinant similarly to the previous section.

Under even more special assumptions $l=1$ and $d_2=\lambda_1$ the recurrence relations (22) pass into the recurrence relations of the Bessel functions. All of these except the first are satisfied if we put $r_n=J_{\lambda_1+n}(2b)$. To satisfy the recurrence relation at $n=0$ one should set $J_{\lambda_1-1}(2b)=0$. This equation serves to determine the eigenvalues of $\lambda_{1\nu}$.

[1]G. N. Stanciu, J. Math. Phys. **8**, 2043 (1967).
[2]L. Lam, Phys. Lett. A **31**, 406 (1970).
[3]L. Lam, J. Math. Phys. **12**, 299 (1971).
[4]A. E. Shabad, Yad. Fiz. **7**, 685 (1968).
[5]F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).
[6]N. G. van Kampen, K. Dan. Vidensk. Selsk. Mat.-Fys. Medd. **26**, 15 (1951).
[7]I. Berson, Zh. Eksp. Teor. Fiz. **56**, 1627 (1969).
[8]I. Berson, Latv. PSR Zinat. Akad. Vestis Fiz. Teh. Zinat. Ser. 3, 3 (1970).
[9]M. V. Fedorov, Zh. Eksp. Teor. Fiz. **52**, 1434 (1967).
[10]H. Ezawa and H. Namaizawa, J. Phys. Soc. Jap. **26**, 458 (1969).
[11]N. D. Sen Gupta, Z. Phys. **200**, 13 (1967).
[12]A. J. Akhiezer and V. B. Berestecki, *Quantum Electrodynamics* (Wiley Interscience, New York, 1969).
[13]*Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vols. 2 and 3.
[14]Pham Ngoc Dinh, C.R. Acad. Sci. A **270**, 650 (1970).
[15]P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[16]L. Carlitz, Boll. Unione Mat. Ital. **3**, 43 (1970).

Erratum: On the convergence of separable expansions for the t matrix [J. Math. Phys. 14, 373 (1973)]

T. A. Osborn

Department of Physics, Brooklyn College of the City University, Brooklyn, New York
(Received 11 May 1973)

It has been pointed out to the author, by I. Sloan and J. Gray[1] that the proof of Lemma 1 is incorrect. We present here a corrected proof of Lemma 1.

Proof: From Eq. (7), we can show $t(z_1) - t(z_2)$ is compact if $(z_2 - z_1)t(z_1)g_0(z_1)g_0(z_2)t(z_2)$ is compact. The only case of interest is when one of the two complex z lie on the positive real axis. The situation where both z_1 and z_2 are bounded away from the positive real axis is dealt with by Proposition 1. The kernel corresponding to the rhs of Eq. (7) is

$$K(\mathbf{p}_1, \mathbf{p}_2; z_1, z_2) = (z_2 - z_1) \int \frac{t(\mathbf{p}_1, \mathbf{p}_3; z_1)t(\mathbf{p}_3, \mathbf{p}_2; z_2)d^3p_3}{(p_3^2 - z_1)(p_3^2 - z_2)}.$$

The proof of this Lemma is complete if we can show K is Hilbert-Schmidt in $\mathcal{H}$. We effect this demonstration by employing Faddeev's lemma on singular integrals.[2] This lemma states that if a function f is bounded and Holder continuous with respect to an estimating function M with indices μ_i and ν_i, viz.

$$|f(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3; z_1)| \leq CM(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3; z_1)$$

and

$$\begin{aligned}|f(\mathbf{p}_1 + \mathbf{h}_1; \mathbf{p}_2 + \mathbf{h}_2; \mathbf{p}_3 + \mathbf{h}_3; z_1 + \Delta_1) - f(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, z_1)| \\ \leq CM(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3; z_1) \\ = [|h_1|^{\mu_1} + |h_2|^{\mu_2} + |h_3|^{\mu_3} + |\Delta_1|^{\nu_1}],\end{aligned}$$

where

$$|\mathbf{h}_i| \leq 1, \quad i = 1, 2, 3, \quad |\Delta_i| \leq 1,$$

then the following result holds. Let $N(\mathbf{p}_1, \mathbf{p}_2; z_1)$ be such that

$$\int d\Omega_{\hat{p}_3} M(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3; t) \leq CN(\mathbf{p}_1, \mathbf{p}_2; z_1)(1 + |p_3|)^{-1-0},$$

then f_1 defined by

$$f_1(\mathbf{p}_1, \mathbf{p}_2; z_1, z_2) = \int \frac{f(\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_3; z_1)}{p_3^2 - z_2} d^3p_3$$

is bounded and Holder continuous with indicies $\mu_i' < \mu_i$, $\nu_j' < \nu_j$ relative to an estimating function M_1 given by

$$M_1(\mathbf{p}_1, \mathbf{p}_2; z_1, z_2) = N(\mathbf{p}_1, \mathbf{p}_2; z_1)(1 + |z_1|)^{-\theta'/2}$$

where $\theta' < \min(1, \theta)$ and μ_i' and ν_j' may be chosen as close to μ_i and ν_j as desired.

This lemma is easy to use here since the kernel of K is bounded by an estimating function of the form

$$\begin{aligned}&\frac{t(\mathbf{p}_1, \mathbf{p}_3; z_1)t(\mathbf{p}_3, \mathbf{p}_2; z_2)}{(p_3^2 - z_1)} \\ &\leq \frac{C}{|p_3^2 - z_1|} \frac{1}{(1 + |\mathbf{p}_3 - \mathbf{p}_1|)^{1+\theta}(1 + |\mathbf{p}_3 - \mathbf{p}_2|)^{1+\theta}},\end{aligned}$$

where $\theta > 1/2$ and index for Holder continuity in p_3 is $\mu' < \mu$. Here μ is the index given in condition B. We need to calculate the angular integral over $d\Omega_{\hat{p}_3}$ to find the N that results from the inequality immediately above. An elementary estimate gives

$$\begin{aligned}\int \frac{1}{|p_3^2 - z_1|} &\frac{d\Omega_{\hat{p}_3}}{(1 + |\mathbf{p}_3 - \mathbf{p}_1|)^{1+\theta}(1 + |\mathbf{p}_3 - \mathbf{p}_2|)^{1+\theta}} \\ &\leq \frac{C}{|p_3^2 - z_1|(1 + |\mathbf{p}_1|)^{1+\theta}(1 + |\mathbf{p}_2|)^{1+\theta}}.\end{aligned}$$

Thus N can take the form

$$N(\mathbf{p}_1, \mathbf{p}_2; z_1, z_2) = \frac{1}{(1 + |p_1|)^{1+\theta}(1 + |p_2|)^{1+\theta}}$$

which used in Faddeev's lemma for singular integrals gives

$$\begin{aligned}&|K(\mathbf{p}_1, \mathbf{p}_2; z_1, z_2)| \\ &\leq \frac{|z_1 - z_2|C}{(1 + |p_1|)^{1+\theta}(1 + |p_2|)^{1+\theta}(1 + |z_2|)^{\theta'/2}}.\end{aligned}$$

Thus for $\theta > 1/2$, the kernel K is square integrable. This completes the proof.

[1] I. Sloan and J. Grey (private communication, 1973).

[2] L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering Theory* (Davey, New York, 1965), Chap. IV, Proposition III.

Erratum: Off-energy-shell t matrix for local potentials with nonlocal square well core interaction [J. Math. Phys. 14, 205 (1973)]

O. Zohni

Institut für Theoretische Physik der Universität Frankfurt/M., Frankfurt am Main, West Germany

(Received 20 April 1973)

The following corrections should be noted:

1. In Eq. (2.31) the factor

$$\frac{U_0\beta^2}{(k^2+\beta^2)\sinh\beta r_0}$$

should be deleted from the right-hand side.

2. In Eq. (2.32) d_2 and d_3 should read:

$$d_2 = -U_1\int_{r_0}^{r_1} dr r j_0(k'r) r h_0^+(\alpha_1 r),$$

$$d_3 = -U_1\int_{r_0}^{r_1} dr r j_0(k'r) r h_0^-(\alpha_1 r).$$

3. In Eq. (2.33) the right-hand side should read:

$$\chi(k',\beta;r) = \frac{U_0\beta^2}{(k'^2+\beta^2)\sinh\beta r_0} \times \{\sin k'r\,[\sinh\beta(r-r_0) - \sinh\beta r - \cosh\beta(r-r_0) \times \sinh\beta r] + \sinh\beta r \sin k r_0\}.$$

4. In Eq. (2.34) the factor $(1q^2|)^{1/2}$ after the curly bracket on the right-hand side should read $(|q^2|)^{1/2}$.