Global Existence of Spin Structures for Gravitational Fields

KLAUS BICHTELER*
Southwest Center for Advanced Studies, Dallas, Texas

(Received 8 October 1967)

The notion of a spin structure for a gravitational field is defined, and it is shown that a spin structure exists if and only if the second Stiefel–Whitney class of the space–time M vanishes. The number of different spin structures is then equal to the number of elements in $H^1(M, Z_2)$.

I. INTRODUCTION

Spinors have been used in general relativity for quite a while, proving to be of considerable advantage in facilitating and shortening calculations. They have been introduced by "affixing a spin dyad to every point of a sufficiently small set in space–time in a differentiable manner."[1-3] To my knowledge, no attempts have been made to find a proper notion of "spin structure" for a gravitational field and to study under what conditions it exists. There are, on the other hand, satisfactory results for the analogous questions in the case of a Riemannian manifold with positive-definite metric, e.g., Milnor.[4]

We shall define the notion of spin structures for a gravitational field in a suggestive manner, and express their number in terms of topological invariants of the space–time manifold M.

The result is literally the same as Milnor's in the positive-definite case and can be derived from it, using the fact that the group $SO(3, \mathbf{R})$ is a strong deformation retract of the proper orthochronous Lorentz group $L_+^\uparrow$. We shall, however, give a proof along Milnor's lines, with the appropriate alterations and filling in arguments only hinted at in Milnor's paper. We hope to achieve in this way that a physicist acquainted with the textbooks, e.g., of Spanier on algebraic topology[5] and of Husemoller on fiber bundles[6] gets a fair idea of the proof.

II. THE SITUATION

Let (M, Q) be a gravitational filed; i.e., M is a four-dimensional real paracompact Hausdorff manifold, connected and of class C^r, $(r \geq 2)$, and Q is a nondegenerate metric on M of signature $(+ - - -)$ and of class C^{r-1}.

We assume, furthermore, that (M, Q) is oriented and time-oriented; i.e., there exists a four-form ϵ and a timelike tangent field X_0 $[Q(X_0, X_0) > 0]$, both defined all over M, continuous and nowhere zero. There is no natural notion of a spin structure for fields (M, Q) which do not satisfy both of these conditions. We call a field (M, Q) with these properties an *orientable field.*

Two pairs (ϵ, X_0) and (ϵ', X_0'), as above, are *equivalent* if $\epsilon' = f\epsilon$, where f is an everywhere-positive function, and if $Q(X_0, X_0') > 0$. An orientation of an orientable field (M, Q) is the choice of an equivalence class $\{\epsilon, X_0\}$, which is called an orientation. A pair $[(M, Q), \{\epsilon, X_0\}]$ is an oriented field. (M, Q) has four distinct orientations.

We choose now a fixed orientation $\{\epsilon, X_0\}$ for (M, Q) and define a tangent Y to be *positive*, if it is timelike and $Q(X_0, Y) > 0$ and a quadruple (Y_0, Y_1, Y_2, Y_3) of tangents to be *positively oriented*, if Y_0 is positive and $\epsilon(Y_0, \cdots, Y_3) > 0$. The set Z^0 of all positively oriented orthonormal tetrads has then a natural structure as a differentiable principal fiber bundle over M with structure group $L_+^\uparrow$, the identity component of the Lorentz group, and projection p^0. There is a soldering form θ^a, a tensorial one-form on Z with values in $\mathbf{R}^4$ and of type *id* (cf., e.g., Ref. 3.)

We thus have associated with the oriented field a differentiable principal fiber bundle $\xi^0 = (Z^0, p^0, M, L_+^\uparrow)$. ξ^0 together with θ^a in turn fixes the metric and orientation structures on M. We may, thus, talk about the oriented field ξ^0. The choice of another orientation gives another principal bundle over M, different as a principal bundle, but equal as a fibration over M. We shall see that only the topology of the fibration $p^0: Z^0 \to M$ enters the considerations.

III. DEFINITION OF SPIN STRUCTURE FOR (M, Q)

Let Σ be the universal covering group of $L_+^\uparrow$ and $\dot{\sigma}: \Sigma \to L_+^\uparrow$ a covering homomorphism. It is known that $\dot{\sigma}$ is a twofold covering and that every twofold covering with connected covering space is equivalent

* On leave of absence from the University of Heidelberg, Germany.

[1] P. Jordan, J. Ehlers, and G. R. Sachs, Akad. Wiss. Lit (Mainz) Abhandl. Math Nat. Kl. **1** (1961).
[2] R. Penrose, Ann. Phys. **10**, 171 (1960).
[3] K. Bichteler, Z. Physik **178**, 488 (1964).
[4] J. Milnor, L'Enseignement Math. **9**, 198 (1963).
[5] E. Spanier, *Algebraic Topology* (McGraw-Hill Book Co., New York, 1966).
[6] K. Husemoller, *Fibre Bundles* (McGraw-Hill Book Co., New York, 1966).

to $\dot{\sigma}$. Σ is isomorphic to $SL(2, \mathbf{C})$, the group of complex 2×2 matrices with determinant 1.

Definition. (See Ref. 4). A spin structure for the oriented field ξ^0 is a differentiable principal fiber bundle $\xi^s = (Z^s, p^s, M, \Sigma)$ over M with group Σ and projection p^s, together with a differentiable map $S: Z^s \to Z^0$ such that the following diagram is commutative:

$$\begin{array}{ccc} Z^s \times \Sigma & \xrightarrow{S \times \dot{\sigma}} & Z^0 \times L_+^\uparrow \\ \downarrow & \leftarrow \text{right multiplication} \rightarrow & \downarrow \\ Z^s & \xrightarrow{S} & Z^0 \\ & {\scriptstyle p^s} \searrow \quad \swarrow {\scriptstyle p^0} & \\ & M & \end{array}$$

Remarks. Fix a point $s \in Z^s$ and its image $z = S(s) \in Z^0$. Because of the commutativity of the lower triangle, s and z lie over the same point $m \in M$. The fiber Z^s_m of ξ^s over m is $s \cdot \Sigma$ and the fiber Z^0_m of ξ^0 over m is $z \cdot L_+^\uparrow$. The commutativity of the upper square implies that the restriction of S to Z^s_m is equivalent to the covering $\dot{\sigma}: \Sigma \to L_+^\uparrow$, hence S is a double covering $Z^s \to Z^0$ which upon restriction to a fiber Z^s_m induces the nontrivial covering of the fiber Z^s_m.

The following proposition reduces the question of the existence of a spin structure to the question of the existence of such a twofold covering.

Proposition. Let Z^s be a topological space and $S: Z^s \to Z^0$ a twofold covering of Z^0 which upon restriction to $S^{-1}(Z^0_m)$ is the unique nontrivial (connected) covering of Z^0_m. Then there exists a unique structure $\xi^s = (Z^s, p^s, M, \Sigma)$ of differentiable principal fiber bundle for Z^s such that $S: Z^s \to Z^0$ is a spin structure for ξ^0.

Proof. The differentiable structure on Z^s is uniquely given by requiring S to be a local diffeomorphism; $p^s := p^0 \circ S$ is differentiable. In order to define the right action of Σ on Z^s, we proceed as follows: choose an open set $W \subset M$, a cross section $\lambda: W \to Z^0$, and a cross section $\Lambda: W \to Z^s$ such that $S \circ \Lambda = \lambda$. Abbreviate $(p^s)^{-1}(W) =: Z^s_W$, $(p^0)^{-1}(W) = Z^0_W$. According to the assumptions, there is a commutative differentiable diagram:

$$\begin{array}{ccc} Z^s_W & \xrightarrow{J_\Lambda \times p^s} & \Sigma \times W \\ \downarrow {\scriptstyle S} & & \downarrow {\scriptstyle \dot{\sigma} \times id} \\ Z^0_W & \xrightarrow{J_\lambda \times p^0} & L_+^\uparrow \times W \end{array}$$

with $J_\lambda(z \cdot L) = J_\lambda(z) \cdot L$ for $L \in L_+^\uparrow$. J_λ is some differentiable function $Z^s_w \to \Sigma$ with $J_\Lambda \circ \Lambda(m) = e \in \Sigma$. Define $s \cdot U$ for $s \in Z^s_W$, $U \in \Sigma$ by $p^s(s \cdot U) = p^s(s)$ and $J_\Lambda(s \cdot U) = J_\Lambda(s) \cdot U$. One can immediately check that $(S, U) \to SU$ is a differentiable right action $[S(UV) = (SU) \cdot V]$ which acts transitively and effectively on the fibers Z^s_m, $m \in W$. Furthermore, the first five-cornered diagram is commutative, when restricted to W.

It remains to be shown that the action so defined does not depend on the choice of W, Λ, and λ. Thus, let W', Λ', λ' be another triple and denote by $(S, U) \to S!\, U$ the corresponding action. Let $m \in W \cap W'$ and $s \in Z^s_m$. Write $\Lambda'(m) = \Lambda(m)V(m)$ and $v(m) = \dot{\sigma}(V(m))$. Then $\lambda'(m) = \lambda(m)v(m)$. With $u = \dot{\sigma}(U)$, we have $S(\Lambda'(m)!\, U) = \lambda'(m)u = \lambda(m)v(m) \cdot u = S(\Lambda(m)V(m)U) = S(\Lambda'(m)U)$. Thus $\Lambda'(m)!\, U = \Lambda'(m)U$ for small U, hence for all $U \in \Sigma$. Finally, ξ^s is locally trivial with the cross sections Λ as above.

The uniqueness of ξ^s is implicit in the above considerations.

IV. THE EXISTENCE OF SPIN STRUCTURES FOR (M, Q)

The proposition permits us to replace the definition of a spin structure for ξ^0 by the following one:

A spin structure for ξ^0 is a cohomology class c of $H^1(Z^0, Z_2)$ with coefficients the ring Z_2 of integers modulo 2 such that $i^*_m(c)$ is the nonzero element of $H^1(Z^0_m, Z_2)$ for all injections $i_m: Z^0_m \to Z^0$ (see Ref. 4). [$H^1(Z^0, Z_2)$ counts the twofold coverings $Z^s \to Z^0$, and the condition $i^*(c) \neq 0$ is the requirement that the induced covering $Z^s_m \to Z^0_m$ be the nontrivial one.]

We are now ready to state the theorem on the existence and number of global spin structures:

Theorem. Let ξ^0 be an oriented gravitational field. There exists a spin structure for ξ^0 if and only if the second Stiefel–Whitney class $w_2 \in H^2(M, Z_2)$ of the manifold M vanishes. If $w_2 = 0$, the number of different spin structures is given by the number of elements in $H^1(M, Z_2)$.

Corollary. If M is simply connected and $\pi_2(M) = 0$ (two-spheres can be contracted) then there is a unique spin structure. This follows from Hurewicz's theorem and universal coefficient theorems.

It is not too difficult to prove the corollary directly by constructing a structure of principal bundle on the twofold covering Z^s of Z^0.

Remarks. If one removes n parallel two-planes from $\mathbf{R}^4$ and takes the Minkowski metric, one is left

with a connected flat manifold M with $H^2(M, Z_2) = 0$ and 2^n elements in $H^1(M, Z_2)$. So there exist 2^n different spin structures.

If we choose another oriented field $\xi^{0'}$ for the orientable field (M, Q), it has the same number of spin structures as ξ^0, according to the theorem. The proof below shows that there is, in fact, a one-to-one correspondence between the spin structures for ξ^0 and $\xi^{0'}$.

We need the following:

Lemma: $SO(3, \mathbf{R})$ is a deformation retract of $L_+^\uparrow$.

Proof: As is well known [a local isomorphism $SL(2, \mathbf{C}) \to L_+^\uparrow$ is given, e.g., in Ref. 3, and $SL(2, \mathbf{C})$ is simply connected; see Ref. 6, p. 93], $SL(2, \mathbf{C})$ is isomorphic to Σ, so let $\dot{\sigma}: \Sigma = SL(2, \mathbf{C}) \to L_+^\uparrow$ be a covering homorphism. The subgroup $SU(2, \mathbf{C})$ of Σ of unitary matrices is mapped onto $SO(3, \mathbf{R}) \subset L_+^\uparrow$ under $\dot{\sigma}$. If we show that $SU(2, \mathbf{C})$ is a deformation retract of $SL(2, \mathbf{C})$, the lemma is proved.

To show this, we use the well-known fact that every element $A \in SL(2, \mathbf{C})$ can be written as a product $A = U \cdot \Delta$ of a unitary and a triangular matrix $\Delta = \left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right)$. This representation can be made unique by requiring Δ to have positive entities in the diagonal and determinant 1. Thus, $SL(2, \mathbf{C})$ is topologically the product of $SU(2, \mathbf{C})$ and the set T of triangular matrices with determinant one and positive diagonal elements. T is obviously contractible, and the lemma is proven. [$A = U \cdot \Delta$ is the Iwasawa decomposition (see Ref. 7) with respect to the maximal compact subgroup $SU(2, \mathbf{C})$, and the proof consists in the remark that the solvable factor in it is always contractible.]

Proof of the theorem. (a) The fibration $p^0: Z^0 \to M$ is orientable over any coefficient ring. In fact, the action of the fundamental group on each fiber is homotopically trivial: let ω be a closed loop starting and ending in $m \in M$. One element in the homotopy class $h[\omega]: Z^0_m \to Z^0_m$ belonging to the homotopy class of ω is the right translation of Z^0 by the element of the holonomy group corresponding to ω (see Refs. 8 and 9). As $L_+^\uparrow$ is connected, this right translation is homotopic to the identity $h[\omega] = id$.

(b) We derive another fibration $p^r: Z^r \to M$ from p^0 in the following manner: choose a positive tangent field X_0 and consider the set Z^r of all tetrads z in Z^0 which have X_0 as its first tangent [this is equivalent to requiring that $\xi^a(z) = (1, 0, 0, 0)$ for the vector field ξ^a corresponding to X_0 by the soldering form θ^a; see Ref. 3]. There is a principal fiber bundle $\xi^r = [Z^r, p^r, M, SO(3, \mathbf{R})]$ with Z^r as bundle space, which is a restriction of ξ^0. ξ^r is orientable over Z_2.

(c) Consider the universal $SO(3, \mathbf{R})$ bundle $\gamma^3 = (\pi: \Gamma^3 \to G^3)$, where G^3 is the Grassmann manifold of three-spaces $s \subset \mathbf{R}^\infty$ and Γ^3_s is the set of orthogonal three-frames of s. π is orientable over Z_2. There is a classifying map $F: M \to G^3$ such that $F^*(\gamma^3) = \xi^r$.

(d) Consider the diagram (coefficients are Z_2):

$$
\begin{array}{ccccccc}
0 \to H^1(M) & \xrightarrow{p^{0*}} & H^1(Z^0) & \xrightarrow{i^*} & H^1(L_+^\uparrow) & \xrightarrow{\tau^0} & H^2(M) \\
\downarrow \approx & & \downarrow (2) & & \downarrow (1) & & \downarrow \approx \\
0 \to H^1(M) & \xrightarrow{p^{r*}} & H^1(Z^r) & \longrightarrow & H^1(SO(3, \mathbf{R})) & \xrightarrow{\tau^r} & H^2(M) \\
\uparrow F^* & & \uparrow F^* & & \approx \uparrow & & \uparrow F^* \\
0 \to M^1(G_3) & \xrightarrow{\pi^*} & H^1(\Gamma^3) & \longrightarrow & H^1(SO(3, \mathbf{R})) & \xrightarrow{\tau\phi} & H^2(G^3)
\end{array}
$$

Each row is an exact sequence, which can be extracted from the spectral sequence of the corresponding fibration. As the spectral sequence and "extracting" are functorial, one obtains the vertical arrows (from the restriction $\xi^r \to \xi^0$ and the classifying map F). In the third row of groups, the fibers are substituted by their homeomorphs, the groups. Now (1) is an isomorphism, by the lemma. Then so is (2), by the five lemma. As γ^3 is universal, $H^1(\Gamma^3) = 0$ and $\tau\phi$ is a monomorphism.

$H^2(G^3)$ consists of the elements 0 and $W_2 = w_2(\gamma^3)$, and the second Stiefel–Whitney class $w_2(\xi^r)$ is defined to be $F^*(W_2) \in H^2(M)$. We see that $\tau^r\{H^1[SO(3, \mathbf{R})]\} = [0, w_2(\xi^r)] \in H^2(M)$. Now the tangent bundle T to M is the Whitney sum of the vector bundle associated to ξ^r and the trivial bundle $\mathbf{R}X_0(m) \to m$. Hence, the second Stiefel–Whitney class $w_2(M)$ of T equals $w_2(\xi^r)$, by the Whitney formula. Hence $\tau^0[H^1(L_+^\uparrow)] = [0, w_2(M)] \subset H^2(M)$.

An element $c \in H^1(Z^0)$ is mapped into the nonzero element $\dot{\sigma}$ of $H^1(L_+^\uparrow)$ by the injection i: Fibre $\to Z^0$, if and only if $\tau^0(\dot{\sigma}) = 0$, i.e., if and only if $w_2(M) = 0$.

According to Sec. III above, there exists a spin structure if and only if $w_2(M) = 0$.

If c, c' are spin structures, they differ by an element of $p^{0*}(H^1(M))$, and as p^{0*} is mono, this proves that $H^1(M)$ corresponds to the different spin structures.

ACKNOWLEDGMENT

This work was supported by the National Aeronautics and Space Administration under Grant NGR 44-044-042, and by the Air Force Office of Scientific Research under Grant AF-AFOSR-903-67.

[7] S. Helgason, *Differential Geometry and Symmetric Spaces*, (Academic Press Inc., New York, 1962).

[8] A. Lichnerowicz, *Théorie globale des connexions et des groupes d'holonomie* (Edizioní Cremonese, Roma, 1955).

[9] W. Beiglboeck, Z. Physik **179**, 148 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Lie-Group and Lie-Algebra Inhomogenizations*

VITTORIO BERZI
Istituto di Scienze Fisiche dell'Università-Milano,
Istituto Nazionale di Fisica Nucleare-Sezione di Milano

AND

VITTORIO GORINI
Istituto di Scienze Fisiche dell'Università-Milano

(Received 29 June 1967)

A systematic formulation of the concept of inhomogenization is given both for Lie groups and for Lie algebras, and the connection between the two structures is clarified in terms of the notion of semidirect product. Special emphasis is devoted to the classification of the inhomogenizations of semisimple Lie algebras. As an application, a lemma due to O'Raifeartaigh is generalized to a wide class of inhomogeneous structures.

1. INTRODUCTION

Recently, some interest has been devoted to the study of the structure and of certain classes of representations of inhomogenizations of Lie groups and Lie algebras.[1] The importance of this study is motivated by the role that inhomogenizations seem to play in connection with the problem of understanding the symmetries displayed by quantum-mechanical systems, with special reference to the field of elementary-particle physics.[2] In this respect, it will be enough to mention as representative the almost classical examples of the inhomogeneous ISL $(6, C)$ and ISU $(6, 6)$, which have been proposed as relativistic generalizations of SU (6).[2(a)–2(g)]

Furthermore, inhomogenizations appear naturally as contractions of "homogeneous" Lie algebras. In fact, the process of contraction defined by Wigner and Inönü[3] leads from the initial algebra to an inhomogenization of the subalgebra with respect to which the contraction is made. As an example, the inhomogeneous pseudo-orthogonal and pseudounitary groups IO (p, q) and $U(1) \otimes IU(p, q)$, respectively, arise as suitable contractions of corresponding homogeneous groups $O(p, q + 1)$ [or $O(p + 1, q)$] and, respectively, $U(p, q + 1)$ [or $U(p + 1, q)$].[1(i)]

The aim of the present paper is to give a systematic formulation of the concept of inhomogenization both for real Lie groups and real Lie algebras and to clarify the structural connection thereof. The fundamental concept which is set as a basis of our definition is the abstract notion of semidirect product. In Sec. 2 we recall the definition of semidirect product both for abstract groups and Lie algebras. In Sec. 3 we apply some standard results of the theory of Lie groups to prove the theorem that the Lie algebra of the semidirect product of two Lie groups G and $\tilde{G}$ corresponding to a given homomorphism σ of $\tilde{G}$ into Aut G[4] is the semidirect product of the Lie algebras $\Lambda(G)$ and $\Lambda(\tilde{G})$ of the two factors, corresponding to the homomorphism Δ, induced by σ, of $\Lambda(\tilde{G})$ into Der $\Lambda(G)$.[5] In Sec. 4 we introduce the definition of inhomogenization of a real Lie group G as the semidirect product $R^n{}_\sigma \times G$ of G by the Abelian real Lie group R^n relative to a representation σ of G acting in the vector space R^n. The exactly parallel definition is given for Lie algebras. In connection with the results of Sec. 3, it is shown that an inhomogenization of a Lie algebra Λ is the Lie algebra of a definite inhomogenization of the simply connected Lie group of which Λ is the Lie algebra. In Sec. 5 we look at the problem of the classification of all inhomogenizations of a given semisimple Lie algebra. Some recent statements of Rosen concerning this matter[1(h)] are subjected to critical analysis and generalized through the introduction of the notion of quasiequivalent representations. As an application of our results, we consider in Sec. 6 the problem of the generalization of a lemma

* Supported in part by the Ministero della Pubblica Istruzione.

[1] (a) H. Bacry and A. Kihlberg, Commun. Math. Phys. **1**, 150 (1965); (b) H. Bacry, Ann. Henri Poincaré **11**, 327 (1965); (c) E. Angelopoulos, Compt. Rend. **263**, A400 (1966); (d) Y. Ne'eman, Commun. Math. Phys. **3**, 181 (1966); (e) R. Mirman, J. Math. Phys. **8**, 57, (1967); (f) A. Kihlberg, "On the Unitary Irreducible Representations of the Strong Coupling Group $[SU(2) \otimes SU(2)] \times T_\sigma$," CERN preprint, 67/419/5-TH 760, March 1967; (g) C. George and M. Lévy-Nahas, J. Math. Phys. **7**, 980 (1966); (h) J. Rosen, Nuovo Cimento **45A**, 234 (1966); (i) J. Rosen, Nuovo Cimento **46B**, 1 (1966); (j) P. Roman and J. Rosen, J. Math. Phys. **7**, 2072 (1966); (k) T. Cook and B. Sakita, J. Math. Phys. **8**, 708 (1967).

[2] (a) B. Sakita, Phys. Rev. **136B**, 1756 (1964); (b) T. Fulton and J. Wess, Phys. Letters **14**, 57 (1965); (c) W. Rühl, Nuovo Cimento **37**, 301 (1965); (d) J. S. Bell and H. Ruegg, Nuovo Cimento **39**, 1166 (1965); (e) W. Rühl, Nuovo Cimento **38**, 675 (1965); (f) S. Coleman, Phys. Rev. **138**, B1262 (1965); (g) W. Rühl, Nuovo Cimento **37**, 1629 (1965); (h) L. O'Raifeartaigh, Phys. Rev. **139**, B1052 (1965); (i) M. Flato and D. Sternheimer, J. Math. Phys. **7**, 1932 (1966); (j) M. Flato, P. Hillion, and D. Sternheimer, Compt. Rend. **264**, A82 (1967); (k) A. Böhm, N. Mukunda, and E. C. G. Sudarshan, Phys. Letters **24B**, 301 (1967).

[3] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953).

[4] By Aut G we denote the group of automorphisms of G.

[5] By Der $\Lambda(G)$ we denote the Lie algebra of derivations of $\Lambda(G)$ (see Ref. 6, p. 8).

[6] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962).

due to O'Raifeartaigh, on which the proof of the well-known mass-splitting theorem is founded.[7,8] We show that the validity of this lemma is not bound to specific properties of the Poincaré algebra, but extends to a wide class of inhomogeneous algebras. This generalization is interesting because it provides a good example of the utilization of some common properties of inhomogenizations to derive a result previously established for a particular inhomogeneous structure, namely the Poincaré algebra.

2. SEMIDIRECT PRODUCT OF LIE GROUPS AND OF LIE ALGEBRAS

We shall recall the notion of semidirect product of two groups.[9]

Consider two groups G and $\tilde{G}$. Let $\sigma:\tilde{x} \to \sigma_{\tilde{x}}$ be a homomorphism of $\tilde{G}$ into the group Aut G of automorphisms of G. Consider the Cartesian product $G \times \tilde{G}$ of G by $\tilde{G}$, equipped with the following internal composition law:

$$(x, \tilde{x})(y, \tilde{y}) = (x\sigma_{\tilde{x}}(y), \tilde{x}\tilde{y}). \tag{1}$$

With respect to this law the set $G \times \tilde{G}$ becomes a group which we shall denote by $G_{\sigma} \times \tilde{G}$ and which is referred to as the semidirect product of G by $\tilde{G}$ corresponding to the homomorphism σ. If both G and $\tilde{G}$ are Lie groups[10] we further require that σ be an analytic homomorphism of $\tilde{G}$ into Aut G.[11] In this way, $G_{\sigma} \times \tilde{G}$ is a Lie group. For the sake of clarity, the preceding definition of semidirect product will be called "external."

Parallel to this definition, one introduces another definition of semidirect product, that will be called "internal" and which is formulated as follows. A group M is said to be the semidirect product of two subgroups H and $\tilde{H}$ if:

(1) M is generated by $H \cup \tilde{H}$.

(2) $H \cap \tilde{H} = 1$.

(3) One of the two subgroups, H say, is normal in M.

From (1), (2), and (3) there follows that any element $a \in M$ can be written in a unique way in the form $a = x\tilde{x}$ with $x \in H$, $\tilde{x} \in \tilde{H}$. Consequently, the product of any two elements of M, $a = x\tilde{x}$, $b = y\tilde{y}$, reads $ab = x\tilde{x}y\tilde{y} = (x\tilde{x}y\tilde{x}^{-1})(\tilde{x}\tilde{y})$, $x\tilde{x}y\tilde{x}^{-1} \in H$, $\tilde{x}\tilde{y} \in \tilde{H}$. Define $\sigma_{\tilde{x}}:y \to \sigma_{\tilde{x}}(y) = \tilde{x}y\tilde{x}^{-1}$. Since H is normal in M, $\sigma_{\tilde{x}}$ is an element of Aut H. The mapping $\sigma:\tilde{x} \to \sigma_{\tilde{x}}$ is a homomorphism of $\tilde{H}$ into Aut H. It follows that the mapping $x\tilde{x} \to (x, \tilde{x})$ is an isomorphism of M onto $H_{\sigma} \times \tilde{H}$ ($M \simeq H_{\sigma} \times \tilde{H}$). Conversely, observe that, starting from the "external" definition of semidirect product and setting $G' = \{(x, 1) \mid x \in G\}$ and $\tilde{G}' = \{(1, \tilde{x}) \mid \tilde{x} \in \tilde{G}\}$, G' and $\tilde{G}'$ are both subgroups of $G_{\sigma} \times \tilde{G}$ for which (1), (2), and (3) hold, G' being normal in $G_{\sigma} \times \tilde{G}$. Further, one has $(\sigma_{\tilde{x}}(y), 1) = (1, \tilde{x})(y, 1)(1, \tilde{x})^{-1}$. This proves the equivalence of the two definitions.

As to Lie algebras, one usually says that a Lie algebra B is the semidirect product of two subalgebras A and $\tilde{A}$ if:

(1) the vector space underlying B is the direct sum of the vector spaces underlying A and $\tilde{A}$, and

(2) one of the two subalgebras, A say, is an ideal.

Then, any element $s \in B$ can be uniquely decomposed in a sum $s = x + \tilde{x}$ with $x \in A$, $\tilde{x} \in \tilde{A}$. Consequently, the commutator of any two elements $s = x + \tilde{x}$ and $t = y + \tilde{y}$ reads

$$\begin{aligned}[s, t] &= [x + \tilde{x}, y + \tilde{y}] \\ &= ([x, y] + [\tilde{x}, y] - [\tilde{y}, x]) + [\tilde{x}, \tilde{y}] \\ &= ([x, y] + \operatorname{ad} \tilde{x}(y) - \operatorname{ad} \tilde{y}(x)) + [\tilde{x}, \tilde{y}]. \end{aligned} \tag{2}$$

Since A is an ideal, the mapping $\tilde{x} \to \operatorname{ad} \tilde{x}$ is a homomorphism of $\tilde{A}$ into the Lie algebra Der A of derivations of A. This suggests introducing also for Lie algebras an "external" definition of semidirect product as follows.[13] Let two Lie algebras A and $\tilde{A}$ be given, and let $D:\tilde{x} \to D_{\tilde{x}}$ be a homomorphism of $\tilde{A}$ into Der A. Consider the product vector space $A \times \tilde{A}$ of the two vector spaces underlying A and $\tilde{A}$, respectively, equipped with the following bilinear composition law:

$$\begin{aligned}&[(x, \tilde{x}), (y, \tilde{y})] \\ &\qquad = ([x, y] + D_{\tilde{x}}(y) - D_{\tilde{y}}(x), [\tilde{x}, \tilde{y}]). \end{aligned} \tag{3}$$

One can verify that, with respect to this law, the vector space $A \times \tilde{A}$ becomes a Lie algebra.[14] This Lie algebra will be denoted by $A_{D} \times \tilde{A}$ and referred to as

---

[7] L. O'Raifeartaigh, Phys. Rev. Letters **14**, 575 (1965).

[8] (a) R. Jost, Helv. Phys. Acta **39**, 369 (1966); (b) I. Segal, J. Functional Anal. **1**, 1 (1967). Since completing the present manuscript, we have learned that a result similar to the one we derive in this section has been obtained by Segal in the reference cited.

[9] L. C. Biedenharn, "Group Theoretical Approaches to Nuclear Spectroscopy" in *Lectures in Theoretical Physics 1962* (Interscience Publishers Inc. New York, 1962), Vol. V.

[10] Throughout in the following we shall be concerned with real Lie groups, i.e., with Lie groups of which the underlying analytic manifold is real. Similarly, all Lie algebras will be over the field R of real numbers.

[11] In the following we shall be concerned only with the case when G is connected (actually, simply connected). Then Aut G is a Lie group (see Ref. 12, p. 138) and we can speak of an analytic homomorphism of $\tilde{G}$ into Aut G.

[12] C. Chevalley, *Theory of Lie groups* (Princeton University Press, Princeton, N.J., 1946).

[13] N. Bourbaki, *Algèbres de Lie* (Hermann & Cie., Paris, 1960).

[14] The proof of this statement relies essentially on the two following properties: (1) $[D_{\tilde{x}}, D_{\tilde{y}}] = D_{[\tilde{x}, \tilde{y}]}$, which is a consequence of D being a homorphism, and (2) $D_{\tilde{x}}([x, y]) = [D_{\tilde{x}}(x), y] + [x, D_{\tilde{x}}(y)]$ which follows from the definition of derivation.

the semidirect product of A by $\tilde{A}$, corresponding to the homomorphism D.

Setting $\tilde{A}' = \{(0, \tilde{x}) \mid \tilde{x} \in \tilde{A}\}$ and $A' = \{(x, 0) \mid x \in A\}$, we readily see that: (1) the vector space underlying $A \,{}_D\!\times \tilde{A}$ is the direct sum of the vector spaces underlying $\tilde{A}'$ and A', respectively; (2) both $\tilde{A}'$ and A' are subalgebras and A' is an ideal.

Proof:

$$\forall (x, \tilde{x}) \in A_D \times \tilde{A},$$
$$(x, \tilde{x}) = (x, 0) + (0, \tilde{x}), (x, 0) \in A', (0, \tilde{x}) \in \tilde{A}',$$

and the decomposition is obviously unique. This proves (1).

$$[(0, \tilde{x}), (0, \tilde{y})] = (0, [\tilde{x}, \tilde{y}]).$$

Thus $\tilde{A}'$ is a subalgebra.

$$[(x, 0), (y, \tilde{y})] = ([x, y] - D_{\tilde{y}}(x), 0),$$

i.e., A' is an ideal.

Further, one has

$$\begin{aligned}(D_{\tilde{x}}(y), 0) &= [(0, \tilde{x}), (y, 0)] \\ &= \text{ad } (0, \tilde{x})\{(y, 0)\}.\end{aligned}$$

This proves that the "external" definition of semidirect product of Lie algebras is equivalent to the usual one (which, analogous to the group case, we shall call "internal").

We conclude this section with a remark about the close link existing between semidirect products and extensions of Lie algebras. As is well known, if a Lie algebra Q is a homomorphic image of a Lie algebra E, one says that E is an extension of Q by S, S being the kernel of the homomorphism. If, as a vector space, $E = S \oplus T$ and T is a subalgebra, one speaks of an inessential extension of Q by S. If this is the case, E is clearly the semidirect product of S by T in the internal sense. Hence, as $Q \simeq E/S \simeq T$, E is isomorphic to the semidirect product $S \,{}_{\alpha\circ\beta}\!\times Q$ where β is an isomorphism of Q onto T, and α is the homomorphism $x \to \text{ad}_S x$ of T into Der S. Conversely, any semidirect product $S \,{}_D\!\times Q$ is an inessential extension of Q by S.

3. LIE ALGEBRA OF THE SEMIDIRECT PRODUCT OF TWO LIE GROUPS

In this section we examine the structure of the Lie algebra of the semidirect product of two Lie groups. For this purpose we need some standard results of the theory of Lie groups.

(1) Let G be a simply connected Lie group. Then (see Ref. 12, pp. 137, 138) Aut G is a Lie group and there exists an (analytic) isomorphism $\alpha: \rho \to \alpha_\rho$ of Aut G onto the Lie group Aut $\Lambda(G)$ of automorphisms of the Lie algebra $\Lambda(G)$ of G such that

$$\rho(\exp X) = \exp \alpha_\rho(X). \tag{4}$$

(2) Let $\varphi: \tilde{x} \to \varphi_{\tilde{x}}$ be an analytic homomorphism of a Lie group $\tilde{G}$ into a Lie group $\tilde{G}'$. φ induces a homomorphism $\Delta^\varphi: \tilde{X} \to \Delta^\varphi_{\tilde{X}}$ between the corresponding Lie algebras such that (see Ref. 12, p. 118)

$$\varphi_{\exp \tilde{X}} = \exp \Delta^\varphi_{\tilde{X}}. \tag{5}$$

Suppose now an analytic homomorphism $\sigma: \tilde{x} \to \sigma_{\tilde{x}}$ of $\tilde{G}$ into Aut G be given. Then, if α is the isomorphism of Aut G onto Aut $\Lambda(G)$ defined in (1), $\alpha \circ \sigma: \tilde{x} \to (\alpha \circ \sigma)_{\tilde{x}}$ is an analytic homomorphism of $\tilde{G}$ into Aut $\Lambda(G)$. Noting that the Lie algebra of Aut $\Lambda(G)$ is the derivation algebra Der $\Lambda(G)$ of $\Lambda(G)$ (see Ref. 6, p. 8), we can apply (2) to deduce the existence of a homomorphism

$$\Delta^{(\alpha\circ\sigma)}: \tilde{X} \to \Delta^{(\alpha\circ\sigma)}_{\tilde{X}}$$

of $\Lambda(\tilde{G})$ into Der $\Lambda(G)$ such that

$$(\alpha \circ \sigma)_{\exp \tilde{X}} = \exp \Delta^{(\alpha\circ\sigma)}_{\tilde{X}}. \tag{6}$$

From this basis we can deduce the following:

Theorem: Let $G \,{}_\sigma\!\times \tilde{G}$ be the semidirect product (in the external sense) of a simply connected Lie group G by a Lie group $\tilde{G}$, corresponding to an analytic homomorphism $\sigma: \tilde{x} \to \sigma_{\tilde{x}}$ of $\tilde{G}$ into Aut G. Then the Lie algebra of $G \,{}_\sigma\!\times \tilde{G}$ is the semidirect product (in the external sense) $\Lambda(G) \,{}_{\Delta(\alpha\circ\sigma)}\!\times \Lambda(\tilde{G})$ of $\Lambda(G)$ by $\Lambda(\tilde{G})$, corresponding to the homomorphism $\Delta^{(\alpha\circ\sigma)}$ of $\Lambda(\tilde{G})$ into Der $\Lambda(G)$.

Proof: Note first that the vector space underlying the Lie algebra of $G \,{}_\sigma\!\times \tilde{G}$ can be identified with the direct product of the vector spaces underlying the Lie algebras $\Lambda(G)$ and $\Lambda(\tilde{G})$ respectively, i.e., with the set $\{(X, \tilde{X}) \mid X \in \Lambda(G), \tilde{X} \in \Lambda(\tilde{G})\}$, where sum and multiplication by a scalar are defined by

$$(X, \tilde{X}) + (Y, \tilde{Y}) = (X + Y, \tilde{X} + \tilde{Y})$$

and

$$k(X, \tilde{X}) = (kX, k\tilde{X}).$$

As to the commutator, we shall determine it by means of the relation (see Ref. 15, Chap. VI)

$$\begin{aligned}&(x(\epsilon), \tilde{x}(\epsilon))(y(\epsilon), \tilde{y}(\epsilon))(x(\epsilon), \tilde{x}(\epsilon))^{-1}(y(\epsilon), \tilde{y}(\epsilon))^{-1} \\ &= \exp\{\epsilon^2[(X, \tilde{X}), (Y, \tilde{Y})] + O(\epsilon^3)\} \\ &= 1 + \epsilon^2[(X, \tilde{X}), (Y, \tilde{Y})] + O(\epsilon^3),\end{aligned} \tag{7}$$

where $\epsilon \to x(\epsilon) = \exp \epsilon X$, etc., are one-parameter

[15] P. M. Cohn, *Lie Groups* (Cambridge University Press, 1957.)

subgroups of G, etc., and ϵ is sufficiently small.[16] Making use of Eq. (1) we obtain the following expression for the commutator of two elements of $G_\sigma\times \tilde{G}$:

$$(x,\tilde{x})(y,\tilde{y})(x,\tilde{x})^{-1}(y,\tilde{y})^{-1} = (x\sigma_{\tilde{x}}(y)\sigma_{\tilde{x}\tilde{y}\tilde{x}^{-1}}(x^{-1})\sigma_{\tilde{x}\tilde{y}\tilde{x}^{-1}\tilde{y}^{-1}}(y^{-1}), \tilde{x}\tilde{y}\tilde{x}^{-1}\tilde{y}^{-1}). \quad (8)$$

We calculate each term up to the second order in the neighborhood of the identity

(1) $$x(\epsilon) = 1 + \epsilon X + \tfrac{1}{2}\epsilon^2 X^2 + \cdots$$

(2) $$\sigma_{\tilde{x}(\epsilon)}(y(\epsilon)) = \sigma_{\exp\epsilon\tilde{X}}(\exp\epsilon Y)$$

[by Eq. (4)]

$$= \exp\epsilon\alpha_{\sigma\exp\epsilon\tilde{X}}(Y)$$
$$= 1 + \epsilon(\alpha_{\sigma\exp\epsilon\tilde{X}}(Y)) + \tfrac{1}{2}\epsilon^2(\alpha_{\sigma\exp\epsilon\tilde{X}}(Y))^2 + \cdots$$
$$= 1 + \epsilon[(\alpha\circ\sigma)_{\exp\epsilon\tilde{X}}](Y) + \tfrac{1}{2}\epsilon^2\{[(\alpha\circ\sigma)_{\exp\epsilon\tilde{X}}](Y)\}^2 + \cdots$$

[by (6)]

$$= 1 + \epsilon(\exp\epsilon\Delta_{\tilde{X}}^{(\alpha\circ\sigma)}(Y)) + \tfrac{1}{2}\epsilon^2(\exp\epsilon\Delta_{\tilde{X}}^{(\alpha\circ\sigma)}(Y))^2 + \cdots$$
$$= 1 + \epsilon Y + \epsilon^2\Delta_{\tilde{X}}^{(\alpha\circ\sigma)}(Y) + \tfrac{1}{2}\epsilon^2 Y^2 + \cdots.$$

In the same way, and using the relations (see Ref. 15, Chap. VI)

$$\exp\epsilon\tilde{X}\exp\epsilon\tilde{Y}\exp(-\epsilon\tilde{X}) = \exp\{\epsilon\tilde{Y} + \epsilon^2[\tilde{X},\tilde{Y}] + O(\epsilon^3)\}$$

and

$$\exp\epsilon\tilde{X}\exp\epsilon\tilde{Y}\exp(-\epsilon\tilde{X})\exp(-\epsilon\tilde{Y}) = \exp\{\epsilon^2[\tilde{X},\tilde{Y}] + O(\epsilon^3)\},$$

we get

(3) $$\sigma_{\tilde{x}(\epsilon)\tilde{y}(\epsilon)\tilde{x}^{-1}(\epsilon)}(x^{-1}(\epsilon)) = 1 - \epsilon X + \tfrac{1}{2}\epsilon^2 X^2 - \epsilon^2\Delta_{\tilde{Y}}^{(\alpha\circ\sigma)}(X) + \cdots$$

and

(4) $$\sigma_{\tilde{x}(\epsilon)\tilde{y}(\epsilon)\tilde{x}^{-1}(\epsilon)\tilde{y}^{-1}(\epsilon)}(y^{-1}(\epsilon)) = 1 - \epsilon Y + \tfrac{1}{2}\epsilon^2 Y^2 + \cdots.$$

[16] The expansions of the exponential mappings that we make use of must be understood in the following way. Given a finite number of one-parameter subgroups $\epsilon\to x_1(\epsilon) = \exp\epsilon X_1, \cdots, \epsilon\to x_k(\epsilon) = \exp\epsilon X_k$, consider an analytic function $f(x)$ defined in a neighborhood Γ of the identity element. Then (see Ref. 15, Chap. III) we can find an ϵ sufficiently small to ensure that

$$\prod_{j=1}^{k} x_j(\epsilon)\in\Gamma$$

and that the following expansion holds:

$$f(\exp\epsilon X_1\cdot\exp\epsilon X_2\cdot\cdots\cdot\exp\epsilon X_k) = \sum_{n_i,\cdots,n_k=0}^{\infty}\left[\epsilon^{\sum_1^k n_i}\Big/\prod_{i=1}^{k} n_i!\right]\left(\prod_{i=1}^{k} X_i^{n_i} f\right)\Big|_{\epsilon=0}.$$

Then:

$$x(\epsilon)\sigma_{\tilde{x}(\epsilon)}(y(\epsilon))\sigma_{\tilde{x}(\epsilon)\tilde{y}(\epsilon)\tilde{x}^{-1}(\epsilon)}(x^{-1}(\epsilon)) \times \sigma_{\tilde{x}(\epsilon)\tilde{y}(\epsilon)\tilde{x}^{-1}(\epsilon)\tilde{y}^{-1}(\epsilon)}(y^{-1}(\epsilon)) = 1 + \epsilon^2([X,Y] + \Delta_{\tilde{X}}^{(\alpha\circ\sigma)}(Y) - \Delta_{\tilde{Y}}^{(\alpha\circ\sigma)}(X)) + O(\epsilon^3).$$

Further, we have

$$\tilde{x}(\epsilon)\tilde{y}(\epsilon)\tilde{x}^{-1}(\epsilon)\tilde{y}^{-1}(\epsilon) = 1 + \epsilon^2[\tilde{X},\tilde{Y}] + \cdots.$$

Hence

$$(x(\epsilon),\tilde{x}(\epsilon))(y(\epsilon),\tilde{y}(\epsilon))(x(\epsilon),\tilde{x}(\epsilon))^{-1}(y(\epsilon),\tilde{y}(\epsilon))^{-1} = 1 + \epsilon^2([X,Y] + \Delta_{\tilde{X}}^{(\alpha\circ\sigma)}(Y) - \Delta_{\tilde{Y}}^{(\alpha\circ\sigma)}(X),[\tilde{X},\tilde{Y}]) + \cdots$$

and comparing with (7)

$$[(X,\tilde{X}),(Y,\tilde{Y})] = ([X,Y] + \Delta_{\tilde{X}}^{(\alpha\circ\sigma)}(Y) - \Delta_{\tilde{Y}}^{(\alpha\circ\sigma)}(X),[\tilde{X},\tilde{Y}]). \quad (9)$$

Q.E.D.

4. INHOMOGENIZATIONS OF LIE GROUPS AND LIE ALGEBRAS

We shall be interested in the case when G is the additive group R^n (every Abelian, real, simply connected, n-dimensional Lie group is isomorphic to R^n). Here, Aut R^n is simply $GL(n,R)$.[17] Then, given any semidirect product $R^n{}_\sigma\times\tilde{G}$, σ is a (not necessarily faithful) representation of $\tilde{G}$ by linear transformations of the vector space R^n.

The Lie algebra $\Lambda(R^n) = \mathfrak{R}^n$ of R^n is the uniquely determined Abelian Lie algebra with underlying vector space R^n and we have

$$\Lambda(R^n{}_\sigma\times\tilde{G}) = \mathfrak{R}^n{}_{\Delta^\sigma}\times\Lambda(\tilde{G}),^{18} \quad (10)$$

where Δ^σ is a (not necessarily faithful) representation of $\Lambda(\tilde{G})$ by linear transformations of the vector space R^n. Actually, Δ^σ is a homomorphism of $\Lambda(\tilde{G})$ into the Lie algebra of derivations of $\mathfrak{R}^n$, which identifies with $gl(n,R)$, the Lie algebra of $GL(n,R)$.[19]

[17] In fact, an automorphism f of the additive Lie group R^n is a homomorphism of R^n satisfying the condition $f(x+y) = f(x) + f(y)$. For any integer n we then have $f(nx) = nf(x)$. For any rational $r = n/m$ set $y = (n/m)x$. Then $my = nx$, $f(my) = f(nx)$, $mf(y) = nf(x)$, $f[(n/m)x] = (n/m)f(x)$. For any real α there exists a sequence of rationals $\{\alpha_n\}$ such that $\alpha = \lim_{n\to\infty}\alpha_n$. Since f is a homeomorphism, we have $\lim_{n\to\infty} f(\alpha_n x) = f(\alpha x)$. On the other hand $\lim_{n\to\infty}\alpha_n f(x) = \alpha f(x)$. Hence $\alpha f(x) = f(\alpha x)$ so that f is a nonsingular linear transformation of the vector space R^n; that is to say, it is an element of $GL(n,R)$.

[18] The reason we have written Δ^σ in place of $\Delta^{(\alpha\circ\sigma)}$ is that Aut R^n = Aut $\Lambda(R^n)$ which enables us to take α as the identity mapping of Aut R^n.

[19] As is well known, $gl(n,R)$, the Lie algebra of $GL(n,R)$, is the Lie algebra of all linear transformations of $\mathfrak{R}^n$. One can arrive at this conclusion also by observing that, as a consequence of $\mathfrak{R}^n$ being Abelian, every linear transformation of its underlying vector space is a derivation.

Conversely, let $\mathcal{R}^n{}_\Delta \times \Lambda$ be the semidirect product of a Lie algebra Λ by $\mathcal{R}^n$ corresponding to an arbitrary representation Δ of Λ acting in the vector space R^n. Then, if $\tilde{G}$ is the simply connected group of which Λ is the Lie algebra, a well-known theorem (see Ref. 12, Theorem 2, p. 113) ensures the existence of a representation σ of G acting in R^n, such that Δ is the representation of $\Lambda(\tilde{G}) = \Lambda$ induced by σ. This enables us to conclude that

$$\mathcal{R}^n_\Delta \times \Lambda = \Lambda(R^n_\sigma \times \tilde{G}).$$

The reason we have required $\tilde{G}$ to be simply connected is that, if it were not so, it would be possible that, for a representation Δ of Λ, there would not be a representation σ of the group $\tilde{G}$, inducing Δ (as long as we confine ourselves to single-valued representations; see Ref. 12, p. 113).

According to the terminology which is used in physics, we call any semidirect product of the kind $R^n{}_\sigma \times \tilde{G}$ an *inhomogenization* of $\tilde{G}$. Note that the invariant Abelian subgroup R^n can be identified with the group of translations in the n-dimensional vector space on which the representation σ of $\tilde{G}$ acts. Similarly, we call any semidirect product of the kind $\mathcal{R}^n{}_\Delta \times \Lambda$ an *inhomogenization*[1(h)] of the Lie algebra Λ.

We note that the class of inhomogenizations of a Lie algebra Λ identifies with the class of inessential extensions of Λ by an Abelian kernel.

The results of this section show that every inhomogenization of an arbitrary Lie algebra Λ can be looked upon as the Lie algebra of an inhomogenization of a simply connected Lie group $\tilde{G}$ of which Λ is the Lie algebra. In the following, since Δ is a representation of Λ, we shall always speak of inhomogenization of a Lie algebra relative to one of its representations.[20]

If the representation, relative to which the inhomogenization is made, is faithful, it is easy to construct a faithful representation of the inhomogenization. Let $P = \mathcal{R}^n{}_\Delta \times \Lambda$ be an inhomogenization of an r-dimensional Lie algebra Λ corresponding to a faithful n-dimensional representation Δ.[22] We choose in P a basis $\{T_\mu, M_i\}$ where the T_μ's ($\mu = 1, 2, \cdots, n$) span $\mathcal{R}^n$ and the M_i's ($i = 1, 2, \cdots, r$) span Λ. Since Δ is $\mathrm{ad}_{\mathcal{R}^n}\Lambda$ the commutation relations are:

$$\begin{aligned} &[M_i, M_j] = c^k_{ij} M_k, \\ &[M_i, T_\mu] = \Delta_{\lambda\mu}(M_i) T_\lambda, \\ &[T_\mu, T_\nu] = 0, \end{aligned} \qquad \begin{array}{l}\text{(summation over repeated}\\ \text{indices is understood),}\end{array} \tag{11}$$

where the $\Delta(M_i)$'s are the images of the M_i's in the representation Δ and the $\{\Delta_{\lambda\mu}(M_i)\}$'s are their representative matrices relative to the basis T_μ. Define the mapping

$$a + t \to \begin{pmatrix} \Delta_{11}(a) & \Delta_{12}(a) & \cdots & \Delta_{1n}(a) & t_1 \\ \Delta_{21}(a) & \Delta_{22}(a) & \cdots & \Delta_{2n}(a) & t_2 \\ & & \cdots & & \\ \Delta_{n1}(a) & \Delta_{n2}(a) & \cdots & \Delta_{nn}(a) & t_n \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

$$a \in \Lambda, \quad t \in \mathcal{R}^n, \quad t = \sum_{\mu=1}^{n} t_\mu T_\mu. \tag{12}$$

Under (12)

$$T_\mu \to \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & 0 & 1 \\ & & \cdots & & \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \leftarrow \mu\text{th row},$$

$$M_i \to \begin{pmatrix} \Delta_{11}(M_i) & \cdots & \Delta_{1n}(M_i) & 0 \\ & \cdots & & \\ \Delta_{n1}(M_i) & \cdots & \Delta_{nn}(M_i) & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

It is easy to verify that the mapping (12) is a representation of P. This representation is faithful because of the faithfulness of Δ. If Δ where not faithful the mapping (12) would again define a representation of P. But this representation would be unfaithful since the $\Delta(M_i)$'s are no longer linearly independent; its kernel is the kernel of Δ.

5. CLASSIFICATION OF THE INHOMOGENIZATIONS OF SEMISIMPLE LIE ALGEBRAS

In this section we investigate the problem of the classification of all inhomogenizations of a given real semisimple Lie algebra. This problem has already been considered by Rosen in Ref. 1(h). His classification relies essentially on the statement that two inhomogenizations of a semisimple Lie algebra are isomorphic only if the defining representations are equivalent. There exist, however, cases where isomorphic inhomogenizations are produced by inequivalent representations as well. The following analysis is

[20] It is a consequence of our definition of inhomogenization of a real Lie algebra Λ that the representation Δ, relative to which the inhomogenization is made, is a real representation, i.e., a representation of Λ by linear transformations of a real vector space. Nevertheless, it is possible to inhomogenize a real Lie algebra also with respect to its essentially complex representations, i.e., representations by linear transformations of complex vector spaces, having the property that for one such representation there exists no basis with respect to which its elements are represented by matrices with real entries. As we show in another paper (Ref. 21), the inhomogenization of a real Lie algebra with respect to an essentially complex representation may be given a meaning as an inhomogenization with respect to a certain real representation of double dimension.

[21] V. Berzi and V. Gorini, J. Math. Phys. 9, 829 (1968) following paper.

[22] The existence of at least one finite-dimensional faithful representation of Λ is ensured by Ado's theorem.

devoted to the generalization of Rosen's statement in order to include these cases.

Definition: Two finite-dimensional representations Δ and Δ' of a Lie algebra Λ will be called *quasi-equivalent* if they have equal dimension and if there exists a nonsingular endomorphism A of the vector space on which they act and an automorphism I of Λ such that

$$A\Delta_l = \Delta'_{I(l)}A, \quad \forall l \in \Lambda. \tag{13}$$

Observation: Quasiequivalence is an equivalence relation.

We shall now prove the following:

Theorem: Two inhomogenizations $P = \mathfrak{R}^n{}_{\Delta} \times \Lambda$ and $P' = \mathfrak{R}^n{}_{\Delta'} \times \Lambda$ of a given semisimple Lie algebra Λ are isomorphic if and only if the defining representations Δ and Δ' are quasiequivalent.[23]

Observation: P and P' agree as vector spaces.

Proof: (a) Assume first that Δ and Δ' are quasiequivalent. We denote the Lie product in P by $[x, y]_{\Delta}$ and in P' by $[x', y']_{\Delta'}$. The mapping $\mathfrak{I}:(t, l) \to \mathfrak{I}(t, l) = (A(t), I(l))$ is an isomorphism of P onto P'. Indeed, it is obviously bijective and linear. Further,

$$\begin{aligned}
&\mathfrak{I}[(t, l), (s, m)]_{\Delta} \\
&\quad = \mathfrak{I}(\Delta_l(s) - \Delta_m(t), [l, m]) \\
&\quad = (A\Delta_l(s) - A\Delta_m(t), I[I(l), I(m)]) \\
&\quad = (\Delta'_{I(l)}A(s) - \Delta'_{I(m)}A(t), I[I(l), I(m)]) \\
&\quad = [(A(t), I(l)), (A(s), I(m))]_{\Delta'} = [\mathfrak{I}(t, l), \mathfrak{I}(s, m)]_{\Delta'}.
\end{aligned}$$

(b) Conversely, let $\mathfrak{I}$ be an isomorphism of P onto P'. We remark first that $\mathfrak{R}^n{}_{\Delta} \times \Lambda$ is a Levi decomposition of P (see Ref. 6, p. 91). Indeed, let S be the radical of P. $\mathfrak{R}^n$ is a solvable ideal, so $\mathfrak{R}^n \subseteq S$. In fact, if it were not so, we would have $S \cap \Lambda \neq 0$. Now, $S \cap \Lambda$ is a subalgebra of a solvable algebra (S), so it is solvable. But $S \cap \Lambda$ is also an ideal in Λ, and this contradicts the hypothesis of semisimplicity of Λ. In the same way, $\mathfrak{R}^n{}_{\Delta'} \times \Lambda$ is a Levi decomposition of P'.

By the uniqueness of the radical, it is evident that $\mathfrak{I}(\mathfrak{R}^n) = \mathfrak{R}^n$, i.e.,

$$\mathfrak{I}(t, 0) = (A(t), 0),$$

where $A:t \to A(t)$ is a nonsingular endomorphism of R^n (actually, it is an automorphism of $\mathfrak{R}^n$).

Set

$$\mathfrak{I}(0, l) = (t_l, I(l)).$$

The mapping $I:l \to I(l)$ is an automorphism of Λ. Indeed, observe first that it is an endomorphism, as can be proved by a trivial calculation. Further, if $m \in \operatorname{Ker} I$,

$$\begin{aligned}
&\mathfrak{I}(0, m) = (t_m, 0) = (A \circ A^{-1}(t_m), 0) = \mathfrak{I}(A^{-1}(t_m), 0) \\
&\qquad \Rightarrow (0, m) = (A^{-1}(t_m), 0) \Rightarrow m = 0.
\end{aligned}$$

From

$$\begin{aligned}
&\mathfrak{I}[(0, l), (t, 0)]_{\Delta} \\
&\quad = \mathfrak{I}(\Delta_l(t), 0) = (A\Delta_l(t), 0) \\
&\quad = [\mathfrak{I}(0, l), \mathfrak{I}(t, 0)]_{\Delta'} = [(t_l, I(l)), (A(t), 0)]_{\Delta'} \\
&\quad = (\Delta'_{I(l)}A(t), 0),
\end{aligned}$$

we get, by the arbitrariness of t,

$$A\Delta_l = \Delta'_{I(l)}A, \quad \forall l \in \Lambda.$$

Comparison with (13) shows that Δ and Δ' are quasiequivalent representations. The proof of the theorem is thus completed.[24]

Observation: Obviously, equivalence of two representations implies their quasiequivalence. The converse, however, is not true in general.[25] This will be shown by the following example. Consider the simple real Lie algebra $sl\,(n, R)$ (i.e., the Lie algebra of the $n \times n$ traceless real matrices), $n > 2$. The mapping

$$I:\, a \to I(a) = -a^T, \quad a \in sl\,(n, R) \tag{14}$$

is an automorphism of $sl\,(n, R)$.

By definition, the self-representation of $sl\,(n, R)$ and its contragredient (14) are quasiequivalent, thus giving rise to isomorphic inhomogenizations. But they are not equivalent, since there exists no nonsingular $n \times n$ matrix S with the property that

$$SaS^{-1} = -a^T, \quad \forall a \in sl\,(n, R).^{26}$$

However, it is interesting to note that, if Aut Λ is connected, quasiequivalence implies equivalence. In fact, we recall that the Lie algebra of Aut Λ is Der Λ. Then, the hypothesis of connectedness of Aut Λ implies that every $I \in$ Aut Λ can be written as a product of a finite number of automorphisms of the form $\exp D$, with $D \in$ Der Λ. Further, as Λ is semisimple, Der Λ = ad Λ (see Ref. 6, p. 74). So

$$I = \prod_{j=1}^{s} \exp(\operatorname{ad} a_j), \quad a_j \in \Lambda. \tag{15}$$

[23] V. Berzi and V. Gorini, Nuovo Cimento **51B**, 207 (1967).

[24] Note that the requirement of semisimplicity of Λ has been used only in the course of proving the necessity of the condition, and precisely to establish that $\mathfrak{I}$ maps $\mathfrak{R}^n$ onto $\mathfrak{R}^n$.

[25] It is precisely this point which seems to have been overlooked by Rosen. (see Ref. 1h).

[26] Take a to be symmetric. $SaS^{-1} = -a$ implies a and $-a$ to have the same characteristic roots which, for arbitrary a, is impossible.

Let now Δ and Δ' be two quasiequivalent representations. Then, by definition

$$\Delta'(x) = A\Delta\left(\prod_{j=1}^{s} \exp(\operatorname{ad} a_j)(x)\right)A^{-1}. \tag{16}$$

One has

$$\begin{aligned}\Delta(\exp(\operatorname{ad} a)(y)) &= \Delta\left(\sum_{n=0}^{\infty}\frac{1}{n!}(\operatorname{ad} a)^n(y)\right)\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}(\operatorname{ad}\Delta(a))^n(\Delta(y))\\ &= \{\exp\Delta(a)\}\Delta(y)\{\exp(-\Delta(a))\}.\end{aligned}$$

By iterated application of this formula, Eq. (16) becomes

$$\begin{aligned}\Delta'(x) &= A\prod_{j=1}^{s}[\exp\Delta(a_j)]\Delta(x)\\ &\quad\times\prod_{j=1}^{s}\{\exp[-\Delta(a_{s+1-j})]\}A^{-1}\\ &= B\Delta(x)B^{-1},\end{aligned}$$

i.e., Δ' and Δ are equivalent.

We are now in a position to give a solution to the problem of the classification of the inhomogenizations of a given real semisimple Lie algebra Λ. In fact, it is clear that by virtue of the theorem stated and proved above, there exists a one-to-one correspondence between the classes of isomorphisms of the inhomogenizations of Λ and the classes of quasiequivalence of its representations. The further observation that, as a consequence of semisimplicity, the representations of Λ are completely reducible, allows a classification of the inhomogenizations of Λ in terms of its irreducible representations[27] to be given in the manner explained in the following.

Given two finite sets Σ and Σ' of finite irreducible representations of Λ,[27] we shall say that Σ' is quasiequivalent to Σ if there exist a one-to-one mapping π of Σ onto Σ' and an automorphism I of Λ such that $\forall\sigma\in\Sigma$, $\pi(\sigma)\circ I$ is equivalent to σ.[28] It is easy to verify that the above defined relation between sets of irreducible representations of Λ is an equivalence relation. Further, it is clear that, given two quasiequivalent representations Δ and Δ', the corresponding systems $\Sigma(\Delta)$ and, respectively, $\Sigma(\Delta')$ of their irreducible components are quasiequivalent and, conversely, the representation built as a direct sum of the elements of a system Σ of irreducible representations is quasiequivalent to the representation formed as a direct sum of the elements of a system Σ', quasiequivalent to Σ. This establishes a one-to-one correspondence between the classes of quasiequivalence of the representations and the classes of quasiequivalence of the finite sets of finite irreducible representations, and, therefore, a one-to-one correspondence between these latter classes and the classes of isomorphism of the inhomogenizations. Thus, the inhomogenizations of Λ can be classified by means of the finite sets of its finite irreducible representations (up to quasiequivalence).

We shall conclude this section with a closer look at the structure of the inhomogenizations of a semisimple Lie algebra. This clarification is important for the problem of the generalization of O'Raifeartaigh's lemma, treated in Sec. 6.

Consider an inhomogenization $\mathcal{R}^n{}_\Delta\times\Lambda$ and let Λ be semisimple. The mapping

$$\operatorname{ad}_{\mathcal{R}^n} l : t \to [l, t], \quad l\in\Lambda, \quad t\in\mathcal{R}^n,$$

which is a derivation of $\mathcal{R}^n$, defines a representation

$$l \to \operatorname{ad}_{\mathcal{R}^n} l \tag{17}$$

of Λ on the vector space $\mathcal{R}^n$. By definition, this representation identifies with Δ. As Δ is completely reducible, the vector space $\mathcal{R}^n$ decomposes into a direct sum of invariant irreducible subspaces. These can be grouped in two sets, one consisting of one-dimensional subspaces and the other of the subspaces of dimension higher than one:

$$\begin{aligned}\mathcal{R}^n = U_1\oplus U_2\oplus\cdots\oplus U_m\oplus T_1\oplus T_2\oplus\cdots\oplus T_k\\ \dim U_j = 1, \quad \dim T_j > 1.\end{aligned} \tag{18}$$

Consider the representation induced in anyone of the U_i's. *A priori*,

$$\operatorname{ad}_{U_i} l = \alpha_l E_{U_i} \quad (E_{U_i} = \text{identity mapping of } U_i).$$

Hence

$$\operatorname{ad}_{U_i}[r, s] = 0.$$

But, as Λ is semisimple, $[\Lambda, \Lambda] = \Lambda$, so that $\operatorname{ad}_{U_i} l = 0$, $\forall l\in\Lambda$, i.e.,

$$[\Lambda, U_i] = 0. \tag{19}$$

[27] As has already been made clear, it is to be noted that, since we are dealing with inhomogenizations of real Lie algebras, the term "representation" means "real representation," i.e., representation by linear transformations of a real vector space. In this connection, irreducibility is intended with respect to real similarity transformations, which by no means implies irreducibility with respect to complex ones. Consider for instance an irreducible essentially complex n-dimensional matrix representaion ρ. The real $2n$-dimensional representation

$$\rho^R = \begin{pmatrix}\operatorname{Re}\rho & -\operatorname{Im}\rho\\ \operatorname{Im}\rho & \operatorname{Re}\rho\end{pmatrix},$$

uniquely determined by ρ, is irreducible with respect to real similarity transformations. On the other hand, by performing a suitable complex change of basis, it is possible to reduce it completely to the form

$$\begin{pmatrix}\rho & 0\\ 0 & \rho^*\end{pmatrix}$$

(see Ref. 21).

[28] We do not exclude that some of the σ's are equivalent.

Consider now one of the T_i's. Set $T_i' = [\Lambda, T_i]$. Since T_i is irreducible, it can be neither $T_i' = 0$ nor $T_i' \subset T_i$, $T_i' \neq 0$. For, in the first case, every one-dimensional subspace of T_i is an invariant subspace and this contradicts the hypothesis of irreducibility of T_i. As to the second case, note that it would imply $[\Lambda, T_i'] \subseteq T_i'$, which again contradicts the hypothesis. We conclude that

$$[\Lambda, T_i] = T_i. \tag{20}$$

We can express this result by saying that the vector space $\mathcal{R}^n$ is a direct sum

$$\mathcal{R}^n = U \oplus T, \tag{21}$$

where U and T are invariant, not necessarily irreducible, subspaces of $\mathcal{R}^n$, for which

$$[\Lambda, U] = 0, \quad [\Lambda, T] = T. \tag{22}$$

Observe that, as Λ is centerless, U is just the center of P. Thus, we can say that every inhomogenization of a semisimple Lie algebra Λ is of the form

$$P = U \oplus T \oplus \Lambda, \tag{23}$$

where U is the center and T is an Abelian ideal such that $[\Lambda, T] = T$. This analysis shows that the system Σ (defined up to quasiequivalence) of finite irreducible representations of Λ which classifies a given inhomogenization, can be uniquely specified through a pair $\{m, \Omega\}$ where m is the dimension of the center of the inhomogenization and Ω is the system of all nontrivial components of Σ (some of which may of course be equivalent). We can thus give a more refined formulation of the criterion of classification. To every inhomogenization $P = \mathcal{R}^n{}_\Delta \times \Lambda$ of a given real semisimple Lie algebra Λ we can uniquely associate a pair $\{m, \Omega\}$, where m is the dimension of the center of P and Ω is the system of nontrivial irreducible components of the defining representation.

Conversely, to every (up to quasiequivalence) pair $\{m, \Omega\}$ constructed as above, there corresponds a unique (up to isomorphism) inhomogenization.

A convenient notation for designating an inhomogenization $P = \mathcal{R}^n{}_\Delta \times \Lambda$ of a Lie algebra Λ is $I_\Delta\Lambda$. According to the above classification, if Λ is semisimple we shall write $P = I_{\{m,\Omega\}}\Lambda$. In writing down explicitly the elements of Ω, we agree to regroup together those which are equivalent under the same symbol, paired with an integer specifying the number of elements of the regroupment (i.e., the number of equivalent elements). Specifically, the notation is $I_\Delta\Lambda = I_{\{m,\Omega\}}\Lambda = I_{\{m,n_1\Delta_1,n_2\Delta_2,\cdots,n_s\Delta_s\}}\Lambda$, where Δ_1, $\Delta_2, \cdots, \Delta_s$ are the nontrivial inequivalent irreducible components[27] of Δ and n_j is the number of times the component Δ_j is contained, up to equivalence, in the reduction of Δ. With respect to complex similarity transformations, the generical irreducible component Δ_j may either be still irreducible, or reduce to the direct sum of two essentially complex irreducible representations which are one the complex conjugate of the other.[29] In this second case (which of course can occur only if the dimension of Δ_j is even) it is natural to indicate Δ_j by σ_j^R, σ_j denoting one of its two essentially complex irreducible components (the other being denoted by σ_j^*).[21]

In general, given a sequence $\{\Delta_1, \Delta_2, \cdots, \Delta_r\}$ of (not necessarily irreducible) inequivalent real representations[27] of a (not necessarily semisimple) Lie algebra Λ, we agree that $I_{\{n_1\Delta_1,\cdots,n_r\Delta_r\}}\Lambda$ denotes the inhomogenization of Λ relative to the representation formed as the direct sum of the Δ_i's where the ith component appears in the sum, up to equivalence, n_i times. If Δ_1, say, is trivial, we write simply n_1 in place of $n_1\Delta_1$. If two sequences $\{\Delta_1, \Delta_2, \cdots, \Delta_r\}$ and $\{\Delta_1', \Delta_2', \cdots, \Delta_r'\}$ are quasiequivalent (in the sense that there exists an automorphism I of Λ such that Δ_j' is equivalent to $\Delta_j \circ I$) the two inhomogenizations $I_{\{n_1\Delta_1,\cdots,n_r\Delta_r\}}\Lambda$ and $I_{\{n_1\Delta_1',\cdots,n_r\Delta_r'\}}\Lambda$ are isomorphic no matter whether Λ is semisimple or not (compare the theorem stated at the beginning of the section (see also Ref. 24).

6. GENERALIZED O'RAIFEARTAIGH'S LEMMA

As an application of the theory and results discussed in the foregoing, we shall consider in his section to what extent a well-known lemma proved by O'Raifeartaigh for the Poincaré algebra[7] can be generalized to arbitrary inhomogenizations of real simisimple Lie algebras.

The lemma under consideration can be stated as follows. Let $\mathfrak{P}$ denote the Lie algebra of the Poincaré group (the inhomogeneous Lorentz group) and T the Abelian ideal of the space–time translations. Then, for every finite dimensional Lie algebra G containing $\mathfrak{P}$, there exists a positive integer $k = k(G)$ such that, for an arbitrary sequence $\{t_1, t_2, \cdots, t_k\}$ of elements of T, the identity

$$[t_k, [t_{k-1}, [t_{k-2}, \cdots, [t_1, g] \cdots]]] = 0 \tag{24}$$

holds, for every $g \in G$.

This result is essential for the proof of O'Raifeartaigh's theorem[7,8] which, as is well known, prevents any reasonable mass splitting within the context of

[29] The proof of this proposition, which holds for nonsemisimple Lie algebras as well, is given in the Appendix.

finite-dimensional Lie algebras containing the Poincaré algebra $\mathfrak{P}$.

In order to clarify our purpose, we note that the Poincaré algebra $\mathcal{R}^4 {}_\Delta\times L$ is a particular example of inhomogenization of a semisimple (actually simple) Lie algebra, the algebra L of the homogeneous Lorentz group. Here, the defining representation is the self representation of L, which is irreducible and nontrivial. Our aim is precisely to show that the properties of the Poincaré algebra on which the proof of (24) relies, are only those which can be expressed by saying that we are dealing with an inhomogenization of a semisimple Lie algebra with respect to a representation which does not contain trivial components. In this way, the lemma can be shown to hold for a wide class of inhomogenizations which can be exactly specified when the homogeneous factor is semisimple.

We start by giving the following:

Definition: Let $P = \mathcal{R}^n {}_\Delta\times \Lambda$ be an inhomogenization of a Lie algebra Λ. A subset $\tau \subseteq \mathcal{R}^n$ will be called an *O subset* if for every finite-dimensional Lie algebra $G \supset P$, there exists a positive integer $k = k(G)$ such that (24) holds for an arbitrary sequence $\{t_1, t_2, \cdots, t_k\}$ of elements of τ and for every $g \in G$.

We immediately note that, if τ is an O subset, the subspace $[\tau]$ generated by τ is an O subset too. Thus we can confine ourselves to the consideration of O *subspaces* of $\mathcal{R}^n$.

Theorem 1: There exists a maximal O subspace which contains every O subspace.

This theorem is an immediate consequence of the following:

Lemma 1: The necessary and sufficient condition for a subset $\tau \subseteq \mathcal{R}^n$ to be an O subset is that $\mathrm{ad}_G t$ is nilpotent, $\forall t \in \tau$.

Proof: The necessity of the condition is trivial. To prove the sufficiency, we start by selecting in τ a maximal number $r\ (\geq n)$ of linearly independent vectors $\{\vartheta_1, \vartheta_2, \cdots, \vartheta_r\}$, so that every $t \in \tau$ is a linear combination of the ϑ_j's. By hypothesis, $(\mathrm{ad}_G \vartheta_j)^{k_j} = 0$ $(j = 1, 2, \cdots, r)$, where $k_1, k_2, \cdots, k_r$ are certain positive integers. Let $\bar{k} = \max\{k_j\}$ and $k = r\bar{k}$, and consider an arbitrary sequence $\{t_1, t_2, \cdots, t_k\}$, $t_j \in \tau$. Expanding each t_l over the ϑ_j's $(t_l = \sum_{j_l=1}^{r} \alpha_{j_l}^{(l)} \vartheta_{j_l})$ we have, writing simply ad t in place of $\mathrm{ad}_G t$:

$$\mathrm{ad}\, t_1\, \mathrm{ad}\, t_2 \cdots \mathrm{ad}\, t_k = \sum_{j_1, \cdots, j_k=1}^{r} \alpha_{j_1}^{(1)} \alpha_{j_2}^{(2)} \cdots \alpha_{j_k}^{(k)}\, \mathrm{ad}\, \vartheta_{j_1}\, \mathrm{ad}\, \vartheta_{j_2} \cdots \mathrm{ad}\, \vartheta_{j_k}.$$

Since $[\mathrm{ad}\, t, \mathrm{ad}\, s] = 0$, we can change freely the order of the factors in the product $\mathrm{ad}\, \vartheta_{j_1}\, \mathrm{ad}\, \vartheta_{j_2} \cdots \mathrm{ad}\, \vartheta_{j_k}$ to get

$$\mathrm{ad}\, \vartheta_{j_1}\, \mathrm{ad}\, \vartheta_{j_2} \cdots \mathrm{ad}\, \vartheta_{j_k} = (\mathrm{ad}\, \vartheta_1)^{l_1} (\mathrm{ad}\, \vartheta_2)^{l_2} \cdots (\mathrm{ad}\, \vartheta_r)^{l_r},$$

with $l_1 + l_2 + \cdots + l_r = r\bar{k}$. Since $\max\{l_j\} \geq \bar{k}$, we get $\mathrm{ad}\, \vartheta_{j_1} \cdots \mathrm{ad}\, \vartheta_{j_k} = 0$ and this in turn implies

$$\mathrm{ad}\, t_1\, \mathrm{ad}\, t_2 \cdots \mathrm{ad}\, t_k = 0,$$

which is equivalent to (24).

It is worthwhile to remark that, in the course of the proof, we did not have to introduce any hypothesis concerning the dimensionality of G so that the lemma is true even if G is infinite-dimensional.

At this point, Theorem 1 is readily proved. Indeed, from Lemma 1, we get that the union of two O subsets is an O subset. Then, if τ_1 and τ_2 are O subspaces, their sum $\tau_1 + \tau_2$ is an O subspace, since it is generated by $\tau_1 \cup \tau_2$. From this it follows immediately that there exists a maximal O subspace containing every O subspace, which was what we intended to prove.

We shall say that a generalized O'Raifeartaigh's lemma holds for an inhomogenization $P = \mathcal{R}^n {}_\Delta\times \Lambda$ if the maximal O subspace identifies with $\mathcal{R}^n$.

We shall now specify the class of inhomogenizations of a semisimple Lie algebra for which this condition is fulfilled. The relevant result, for this purpose, is embodied in the following:

Theorem 2: Let τ be a subspace of $\mathcal{R}^n$ with the property that

$$[\Lambda, \tau] = \tau, \tag{25}$$

then τ is an O subspace.[30]

We first prove the following:

Lemma 2: Let N be a nilpotent Lie algebra of linear transformations of a finite-dimensional vector space $\mathcal{M}$ over a field Φ of characteristic zero, and suppose that every $\rho \in N$ can be written as a finite sum:

$$\rho = \sum_j [\sigma^{(j)}, \rho^{(j)}], \tag{26}$$

with $\rho^{(j)} \in N$ and $\sigma^{(j)}$ some linear transformation of $\mathcal{M}$. Then every $\rho \in N$ is a nilpotent linear transformation.

Proof: Suppose first that Φ is algebraically closed. Then a well-known theorem (see Ref. 6, p. 50) states

[30] As the context of the proof will show, the theorem holds even if Λ is not semisimple.

that $\mathcal{M}$ breaks up into a direct sum $\mathcal{M} = \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_s$, where the $\mathcal{M}_i$'s are subspaces invariant under N, and in each $\mathcal{M}_i$ we can find a basis with respect to which the restriction ρ_i to $\mathcal{M}_i$ of any $\rho \in N$ is represented by a matrix of the form

$$\begin{pmatrix} \alpha^{(i)} & & & * \\ & \alpha^{(i)} & & \\ & & \ddots & \\ & & & \\ 0 & & & \alpha^{(i)} \end{pmatrix},$$

i.e., ρ_i is a sum

$$\rho_i = n^{(i)} + \alpha^{(i)} E, \tag{27}$$

where $n^{(i)}$ is a nilpotent linear transformation. We prove that $\alpha^{(i)} = 0$ for every subspace $\mathcal{M}_i$ and for every $\rho \in N$. To this purpose observe that, by (26), ρ_i is of the form:

$$\rho_i = \sum_j [\sigma_i^{(j)}, \rho_i^{(j)}], \tag{28}$$

where $\sigma_i^{(j)}$ is the restriction to $\mathcal{M}_i$ of $P_i \sigma^{(j)}$, P_i being the projection operator on $\mathcal{M}_i$. Equating (28) to (27) and taking the trace, we get

$$n_i \alpha^{(i)} = 0, \quad n_i = \dim \mathcal{M}_i \Rightarrow \alpha^{(i)} = 0. \tag{29}$$

Then there exists a basis in $\mathcal{M}$ relative to which the matrices of the ρ's are niltriangular, from which it follows that the ρ's are nilpotent linear operators.

If the base field Φ is not algebraically closed but, all the same, of zero characteristic, we can again reach the conclusion that the elements of N are nilpotent linear transformations by noting that the vector space $\mathcal{M}$ on which N acts can be uniquely extended to a vector space $\mathcal{M}^C$ obtained from $\mathcal{M}$ by extension of the base field Φ to its algebraic closure C.

A basis for $\mathcal{M}$ over Φ is also a basis for $\mathcal{M}^C$ over C and every linear transformation "a" of $\mathcal{M}$ can be uniquely extended in an obvious way to a linear transformation "a^C" of $\mathcal{M}^C$ so that

$$(a + b)^C = a^C + b^C, \quad (ab)^C = a^C b^C$$
$$[a, b]^C = [a^C, b^C], \quad \text{and} \quad a^C = 0 \Leftrightarrow a = 0.$$

It is then apparent that the set N^C of linear transformations of $\mathcal{M}^C$ of the form $\sum_i \alpha_i \rho^{(i)}$ with $\rho^{(i)} \in N$ and $\alpha_i \in C$ can be made canonically into a nilpotent Lie algebra of linear transformations of $\mathcal{M}^C$ for which the hypothesis of the lemma holds. Then, $\forall \rho \in N$, ρ^C is nilpotent, which in turn implies that ρ is nilpotent. This completes the proof of the lemma.

At this point the theorem can be easily demonstrated. In fact, if ρ is an arbitrary finite-dimensional (real) representation of a Lie algebra G containing P, we have that $\rho(\tau)$ is a nilpotent (actually Abelian) Lie algebra of linear transformations of the (real) vector space on which the representation acts. Further, we deduce from (25) that $\forall t \in \tau$, $\rho(t)$ is of the form $\rho(t) = \sum_j [\rho(l_j), \rho(t_j)]$, with $l_j \in \Lambda$ and $t_j \in \tau$. Then, as the field of the reals is of characteristic zero, the hypotheses of Lemma 2 are satisfied for $\rho(\tau)$ and we can conclude that $\rho(t)$ is a nilpotent linear transformation, $\forall t \in \tau$. This implies that the relation

$$[\rho(t_k), [\rho(t_{k-1}), \cdots, [\rho(t_1), \rho(g)] \cdots]] = 0 \tag{30}$$

holds for an arbitrary sequence $\{t_1, t_2, \cdots, t_k\}$ of elements of τ, with $k \geq 2n$, $n = \dim \rho$. Choosing ρ to be faithful, we see that Eq. (30) implies Eq. (24)[31] and this proves the statement of the theorem.

We have thus obtained the result that a sufficient condition for a generalized O'Raifeartaigh's lemma to hold for an inhomogenization $P = \mathcal{R}^n {}_\Delta\times \Lambda$ of a Lie algebra Λ is that

$$[\Lambda, \mathcal{R}^n] = \mathcal{R}^n. \tag{31}$$

We shall now show that if Λ is semisimple, this condition is necessary as well. Indeed, we have proved in Sec. 5 that, if Λ is semisimple, the most general form of an inhomogenization is given by (23) together with (22). These formulas tell us that T is an O subspace. We prove that it is also the maximal O subspace. For this purpose observe that, by definition, the property of a subspace τ of $\mathcal{R}^n$ being an O subspace implies (22) to hold no matter which finite-dimensional Lie algebra G we choose, subject to the condition of containing P. Thus, to prove that T is the maximal O subspace it is sufficient, by Lemma 1, to find a particular G on which ad U has not a nilpotent action. In other words, $\mathrm{ad}_G u$ must have a nonzero semisimple part, $\forall u \in U$ $(u \neq 0)$.[32] Define G as follows: as a vector space, it is a direct sum

$$G = A \oplus P = A \oplus U \oplus T \oplus \Lambda \tag{32}$$

and the subspace A is taken to be an Abelian ideal with the property that $\mathrm{ad}_A \lambda = 0$ and $\mathrm{ad}_A t = 0$, $\forall \lambda \in \Lambda$ and $\forall t \in T$, while $\mathrm{ad}_G u$ is required to have a nonzero semisimple part, $\forall u \in U$ $(u \neq 0)$. More specifically, the commutation relations are postulated

[31] The requirement that G be finite-dimensional enters here. In fact, if it were not so, G would not have any faithful finite-dimensional representation. On the other hand, the existence of at least one such representation when the dimension of G is finite is ensured by Ado's theorem.

Note that, alternatively, we could have reached the required result (24) by choosing for ρ the adjoint representation $\rho(t) = \mathrm{ad}\, t$, thus proving the nilpotency of ad t. Here again the condition that G be finite dimensional is essential, because the adjoint representation acts in the vector space underlying G.

[32] For the decomposition of a linear transformation into a semisimple plus a nilpotent part, see N. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press, New York, 1962), p. 131.

to be such that $U \oplus T \oplus \Lambda$ is the given inhomogenization, while the elements of A have the following brackets with the elements of P:

$$[\lambda, a] = 0, \quad [t, a] = 0; \quad \lambda \in \Lambda, \quad t \in T, \quad a \in A$$
$$[u_i, a_\alpha] = \sum_\beta D_{\beta\alpha}(u_i) a_\beta . \tag{33}$$

Here, $\{u_i\}$ and $\{a_\alpha\}$ denote bases for U and A, respectively, and the mapping $u \to D(u)$ is a representation of U by nonnilpotent linear transformations of A.[33] It is easy to verify that all the Jacobi identities are satisfied as well as the bilinearity of the commutators, so that G is a Lie algebra. Further, denoting by g_A the projection on A parallel to P of any element $g \in G$, we have $\mathrm{ad}_G u(g) = \mathrm{ad}_G u(g_A)$. This proves that $\mathrm{ad}_G u$ has a nonzero semisimple part, thus furnishing the required result, namely that T is the maximal O subspace relative to the particular G chosen above.[34]

Thus we have established that O'Raifeartaigh's lemma holds for an inhomogenization of a semisimple Lie algebra if and only if the inhomogenization is centerless. As to inhomogenizations of nonsemisimple Lie algebras, for the lemma to hold it is sufficient, for example, that the inhomogenization be relative to a completely reducible representation which does not contain trivial components [in this case, in fact, Eq. (31) is verified].

ACKNOWLEDGMENTS

We wish to express our sincere thanks to Professor P. Caldirola and to Professor A. Loinger for their kind interest. We are further indebted to Professor F. Duimio for many helpful discussions and constant encouragement.

APPENDIX

Let S be an n-dimensional vector space over the field R of real numbers. We denote by S^C the complexification of S, namely, the n-dimensional vector space over the field C of complex numbers which is obtained from S by extension of the base field from R to C. A basis $\{e_1, e_2, \cdots, e_n\}$ for S is also a basis for S^C. Any given linear transformation A of S can be canonically extended to a linear transformation A^C of S^C in such a way that if A_{ik} is the matrix representing A with respect to the basis $\{e_1, e_2, \cdots, e_n\}$ for S,

$$Ae_k = \sum_{i=1}^n A_{ik} e_i ,$$

and A^C_{ik} the matrix representing A^C with respect to $\{e_1, e_2, \cdots, e_n\}$ considered as a basis for S^C

$$A^C e_k = \sum_{i=1}^n A^C_{ik} e_i ,$$

one has $A^C_{ik} = A_{ik}$. Given a representation $\Delta: a \to \Delta(a)$ of a real Lie algebra Λ by linear transformations of the real vector space S (a real representation), the mapping $\Delta^C: a \to \Delta^C(a) = \Delta(a)^C$ is a representation of Λ by linear transformations of S^C which we refer to as the complex representation associated (canonically) to the real representation Δ. One has the following:

Theorem: Let $\Delta: a \to \Delta(a)$ be a real irreducible representation of a real Lie algebra Λ. Then, if the dimension of Δ is odd, the associated complex representation Δ^C is irreducible. If the dimension of Δ is even, Δ^C is either irreducible or it reduces completely to a direct sum of two irreducible essentially complex representations which are complex conjugates of each other (see Ref. 20 for the meaning of essentially complex representation).

Proof: Denote by S the real vector space on which Δ acts. By hypothesis, there is no proper subspace of S which is invariant under Δ. Suppose, on the other hand, that there is a proper subspace V of S^C which is invariant under Δ^C. Let $n = \dim S = \dim S^C$ and $k = \dim V < n$. The elements of V can be written as linear combinations

$$\sum_{i=1}^n \xi_i e_i ,$$

where $\{e_1, e_2, \cdots, e_n\}$ is a basis for S, $\xi_i \in C$ $(i = 1, 2, \cdots, n)$, and the n-tuples $\{\xi_1, \xi_2, \cdots, \xi_n\}$ span a k-dimensional subspace of the complex n-dimensional vector space C^n of all n-tuples of complex numbers.

Consider the following (antilinear) mapping of V into S^C:

$$\mathcal{C}: x = \sum_{i=1}^n \xi_i e_i \to x^* = \sum_{i=1}^n \xi_i^* e_i , \tag{A1}$$

where ξ^* denotes the complex conjugate of ξ.

The image $V^* = \mathcal{C}(V)$ of V by $\mathcal{C}$ is a k-dimensional subspace of S^C. Further, V^* is invariant under Δ^C. Indeed, the invariance of V under Δ^C implies that

[33] For example, we can choose $\dim A \geq \dim U$ and $D_{\alpha\beta}(u_i) = \delta_{\alpha_i}\delta_{\beta_i}$.

[34] Note, however, that in the $U \neq 0$ case the maximal O subspace relative to a given embedding algebra G may be greater than T. For example, let $\dim U = r > 1$ and consider a G of the following structure:

$$G = A \oplus P,$$

where the subspace A is taken to be one-dimensional and the commutation relations involving A are given by $[\Lambda, A] = 0$, $[T, A] = 0$, and $[U, A] = A$. In this case, the maximal O subspace is the direct sum of T plus the $(r-1)$-dimensional kernel of the representation $u \to \mathrm{ad}_A u$, $u \in U$.

$\Delta^C(a)x = y \in V$, $\forall a \in \Lambda$ and $\forall x \in V$. Writing

$$x = \sum_{i=1}^{n} \xi_i e_i,$$

one has

$$y = \sum_{i,k=1}^{n} \xi_i \Delta^C_{ki}(a) e_k = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \xi_i \Delta_{ki}(a) \right) e_k$$

and the $\Delta_{ki}(a)$ are real coefficients. Then,

$$\Delta^C(a)x^* = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \xi_i^* \Delta_{ki}(a) \right) e_k$$
$$= \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \xi_i \Delta_{ki}(a) \right)^* e_k = y^* \in V^*.$$

Invariance of V and of V^* implies invariance of their intersection $V \cap V^*$. Then, irreducibility of V requires either $V \cap V^* = V$ or $V \cap V^* = 0$. Consider the first alternative. Since dim V = dim V^*, this implies $V = V^*$. Then, if $\{a_1, a_2, \cdots, a_k\}$,

$$a_\mu = \sum_{j=1}^{n} c_{j\mu} e_j,$$

is a basis for V, the vectors $\{a_1^*, a_2^*, \cdots, a_k^*\}$, which form a basis for V^*, belong to V (actually, they form another basis for V) and we can write, for any $x \in V$,

$$x = \sum_{\mu=1}^{k} \alpha_\mu a_\mu = \sum_{\mu=1}^{k} i\alpha_\mu \left\{ \frac{i(a_\mu^* - a_\mu)}{2} \right\} + \sum_{\mu=1}^{k} \alpha_\mu \left\{ \frac{a_\mu^* + a_\mu}{2} \right\}. \tag{A2}$$

This formula shows that the vectors $a'_\mu = \frac{1}{2}i(a_\mu^* - a_\mu)$ and $a''_\mu = \frac{1}{2}(a_\mu^* + a_\mu)$ $(\mu = 1, 2, \cdots, k)$ generate V and since they are linear combinations of the e_j's with real coefficients [actually, $a'_\mu = \sum_{j=1}^{n} (\mathrm{Im}\ c_{j\mu}) e_j$ and $a''_\mu = \sum_{j=1}^{n} (\mathrm{Re}\ c_{j\mu}) e_j$], we get that V admits a basis formed by vectors all belonging to S. Let us denote by $\{\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_k\}$ one such basis

$$\hat{e}_\mu = \sum_{j=1}^{n} \hat{c}_{j\mu} e_j, \quad \hat{c}_{j\mu} \in R.$$

By the invariance of V we have

$$\Delta^C(a)\hat{e}_\mu = \sum_{\nu=1}^{k} \Delta_{\nu\mu}(a) \hat{e}_\nu,$$

the $\Delta_{\nu\mu}(a)$ being of course real. Then, regarding the $\hat{e}_\mu$'s as elements of S,

$$\Delta(a)\hat{e}_\mu = \sum_{\nu=1}^{k} \Delta_{\nu\mu}(a) \hat{e}_\nu,$$

which proves that the k-dimensional subspace V_R of S spanned by the $\hat{e}_\mu$'s is invariant under Δ. Since we have supposed $k < n$, this contradicts the hypothesis of irreducibility of Δ. The alternative $V = V^*$ is thus ruled out.

Consider the other possibility, namely, $V \cap V^* = 0$. We remark first that this implies $k \leq [n/2]$. Next, choose a basis $\{a_1, a_2, \cdots, a_k\}$,

$$a_\mu = \sum_{j=1}^{n} c_{j\mu} e_j,$$

for V. By the invariance of V, we have

$$\Delta^C(a)a_\mu = \sum_{l=1}^{n} \left\{ \sum_{j=1}^{n} \Delta_{lj}(a) c_{j\mu} \right\} e_l$$
$$= \sum_{\nu=1}^{k} \sigma_{\nu\mu}(a) a_\nu = \sum_{l=1}^{n} \left\{ \sum_{\nu=1}^{k} c_{l\nu} \sigma_{\nu\mu}(a) \right\} e_l, \tag{A3}$$

where $\{\sigma_{\nu\mu}(a)\}$ is the matrix representing the restriction $\sigma: a \to \sigma(a)$ of Δ^C to V, with respect to the basis $\{a_1, a_2, \cdots, a_k\}$. The vectors $\{a_1^*, a_2^*, \cdots, a_k^*\}$ form a basis for V^* and, by the invariance of V^*, we have

$$\Delta^C(a)a_\mu^* = \sum_{l=1}^{n} \left\{ \sum_{j=1}^{n} \Delta_{lj}(a) c_{j\mu}^* \right\} e_l$$
$$= \sum_{l=1}^{n} \left\{ \sum_{j=1}^{n} \Delta_{lj}(a) c_{j\mu} \right\}^* e_l$$
$$= \sum_{\nu=1}^{k} \tilde{\sigma}_{\nu\mu}(a) a_\nu^* = \sum_{l=1}^{n} \left\{ \sum_{\nu=1}^{k} c_{l\nu}^* \tilde{\sigma}_{\nu\mu}(a) \right\} e_l, \tag{A4}$$

where $\{\tilde{\sigma}_{\mu\nu}(a)\}$ is the matrix representing the restriction $\tilde{\sigma}: a \to \tilde{\sigma}(a)$ of Δ^C to V^*, with respect to the basis $\{a_1^*, a_2^*, \cdots, a_k^*\}$. Comparing (A3) and (A4) we get, by the linear independence of the e_j's,

$$\sum_{\nu=1}^{k} c_{l\nu}^* \tilde{\sigma}_{\nu\mu}(a) = \sum_{\nu=1}^{k} c_{l\nu}^* \sigma_{\nu\mu}^*(a).$$

Multiplying by e_l and summing over l from 1 to n, gives

$$\sum_{\nu=1}^{k} \tilde{\sigma}_{\nu\mu}(a) a_\nu^* = \sum_{\nu=1}^{k} \sigma_{\nu\mu}^*(a) a_\nu^*.$$

which, by the linear independence of the a_ν^*'s, implies

$$\tilde{\sigma}_{\nu\mu}(a) = \sigma_{\nu\mu}^*(a). \tag{A5}$$

Relation (A5) shows that the restriction $\rho^C: a \to \rho^C(a)$ of Δ^C to the subspace $V \oplus V^*$ is the direct sum of the representation σ plus its complex conjugate σ^*. The corresponding invariant subspaces are just V (acted upon by σ) and V^* (acted upon by σ^*), and the matrix representing $\rho^C(a)$ with respect to the basis $\{a_1, a_2, \cdots, a_k, a_1^*, a_2^*, \cdots, a_k^*\}$ for $V \oplus V^*$ is

$$\begin{pmatrix} \{\sigma_{\mu\nu}(a)\} & 0 \\ 0 & \{\sigma_{\mu\nu}^*(a)\} \end{pmatrix}. \tag{A6}$$

The vectors

$$a'_\mu = \frac{i}{2}(a^*_\mu - a_\mu) = \sum_{j=1}^{n}(\mathrm{Im}\, c_{j\mu})e_j,$$

$$a''_\mu = \tfrac{1}{2}(a^*_\mu + a_\mu) = \sum_{j=1}^{n}(\mathrm{Re}\, c_{j\mu})e_j, \quad \mu = 1, 2, \cdots, k, \tag{A7}$$

which belong to S, form a basis for $V \oplus V^*$ with respect to which the matrix representing $\rho^C(a)$ is real and given by

$$\begin{pmatrix} \{\mathrm{Re}\, \sigma_{\mu\nu}(a)\} & -\{\mathrm{Im}\, \sigma_{\mu\nu}(a)\} \\ \{\mathrm{Im}\, \sigma_{\mu\nu}(a)\} & \{\mathrm{Re}\, \sigma_{\mu\nu}(a)\} \end{pmatrix}. \tag{A8}$$

Then, the $2k$-dimensional subspace T_R of S spanned by the vectors (A7) is invariant under Δ. Therefore, by the hypothesis of irreducibility of Δ, we must have $T_R = S$ and this implies $2k = n$. If n is odd, this is impossible and the first statement of the theorem is proved. Let n be even. Then, $T_R = S$, $V \oplus V^* = S^C$ and we obtain the result that the condition of irreducibility of Δ together with the one that Δ^C is reducible implies Δ^C to split into the direct sum of a complex representation σ plus its complex conjugate. Further, σ is irreducible and essentially complex. It is irreducible because we have required V to be an invariant irreducible subspace. It is essentially complex because, if it were not so, we could find a basis $\{\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_k\}$,

$$\bar{a}_\mu = \sum_{\nu=1}^{k} A_{\nu\mu} a_\nu,$$

for V, with respect to which, $\forall a \in \Lambda$, $\sigma(a)$ is represented by a matrix $\{\bar{\sigma}_{\mu\nu}(a)\}$ with real entries. Consider then the basis $\{\bar{f}_1, \bar{f}_2, \cdots, \bar{f}_n\}$ for S defined by

$$\bar{f}_j = \sum_{l=1}^{n} A^R_{lj} f_l,$$

$$(f_1 = a'_1, f_2 = a'_2, \cdots, f_k = a'_k, f_{k+1} = a''_1, f_{k+2} = a''_2, \cdots, f_n = a''_k),$$

where

$$\{A^R_{il}\} = \begin{pmatrix} \{\mathrm{Re}\, A_{\mu\nu}\} & -\{\mathrm{Im}\, A_{\mu\nu}\} \\ \{\mathrm{Im}\, A_{\mu\nu}\} & \{\mathrm{Re}\, A_{\mu\nu}\} \end{pmatrix}.$$

One easily verifies that, with respect to this basis, $\Delta(a)$ is represented by the matrix

$$\begin{pmatrix} \{\bar{\sigma}_{\mu\nu}(a)\} & 0 \\ 0 & \{\bar{\sigma}_{\mu\nu}(a)\} \end{pmatrix},$$

so that Δ is completely reducible.

It remains to show that under the above conditions, namely that σ is irreducible and essentially complex, Δ is actually irreducible. For the proof of this fact we refer to Ref. 21.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Inhomogenizations and Complex Representations*

VITTORIO BERZI
Istituto di Scienze Fisiche dell'Università-Milano
Istituto Nazionale di Fisica Nucleare-Sezione di Milano,
AND
VITTORIO GORINI
Istituto di Scienze Fisiche dell'Universitá-Milano

(Received 29 June 1967)

Real Lie-algebra inhomogenizations are considered for complex defining representations and some emphasis is devoted to the case when the defining representation is irreducible. A theorem is given concerning the conditions under which nonequivalent complex representations give rise to isomorphic inhomogenizations and a classification is made of the complex inhomogenizations of semisimple Lie algebras.

1. INTRODUCTION

Inhomogenizations have recently been made the object of studies in view of the role that some of these generalized Poincaré structures seem to play in the symmetries displayed by elementary-particle interactions.[1]

In a preceding paper[2] (hereafter referred to as I) we have given a systematic formulation of the general concept of inhomogenization both for real Lie algebras and real Lie groups, and have clarified the connection between the two structures in terms of the notion of semidirect product. It was a basic point that, as a consequence of the definition, real Lie-algebra and real Lie-group inhomogenizations were made relative to real representations. However, since essentially complex representations occur as well and they are no less relevant than real ones, the question arises as to whether one can consistently define inhomogenizations also with respect to complex representations. In this paper we show that this is indeed possible in a very simple and natural way. This is not an unexpected fact, since one can add translations in a complex representation space as well as in a real one. Throughout this paper, we shall refer to inhomogenizations relative to complex representations as complex inhomogenizations.

A specific instance of complex inhomogenization, namely the group $SL(2, C)$ inhomogenized relative to its self-representation, has been recently considered by Flato, Hillion, and Sternheimer in connection with the relativistic covariance of the Dirac equation for zero-mass particles.[3] Another example is provided by $IU(p, q)$, the inhomogenization of the pseudo-unitary group $U(p, q)$ relative to its self-representation. The interest of this inhomogenization has been emphasized by Roman and Rosen who have discussed some aspects of its enveloping algebra.[4]

In this paper we shall work with Lie algebras rather than with Lie groups. The correspondence with Lie groups can be established in a straightforward way by means of the results of I.

In Sec. 2 we collect some standard definitions and results to be used in the following. In Sec. 3 we give from the outset the definition of inhomogenization of a real Lie algebra Λ with respect to a complex n-dimensional representation ρ as the inhomogenization $\mathfrak{R}^{2n}{}_{\rho^R} \times \Lambda$ of Λ relative to the $2n$-dimensional real representation ρ^R uniquely induced by ρ. We then justify this definition by showing that, regarding the group of translations in the complex space on which ρ acts as a real Lie group, the Abelian ideal $\mathfrak{R}^{2n}$ can naturally be interpreted as the Lie algebra of this group. In Sec. 4 we introduce the concept of irreducible inhomogenization for real defining representations and prove a theorem which allows an unambiguous extension of this concept to include essentially complex inhomogenizations as well. Section 5 is devoted to the characterization of the conditions on the defining representations under which different complex inhomogenizations of a given real semisimple Lie algebra are isomorphic. A theorem is proved which gives an answer to this question and allows us to give a complete classification of the complex inhomogenizations of any given semisimple real Lie algebra.

2. SOME STANDARD DEFINITIONS AND RESULTS

For the purpose of free use in the following we briefly sketch two formal processes which one introduces in connection with vector spaces and Lie

* Supported in part by the Ministero della Pubblica Istruzione.

[1] Extensive references can be found in our preceding paper (see Ref. 2).

[2] V. Berzi and V. Gorini, J. Math. Phys. **9**, 816 (1968), preceding paper.

[3] M. Flato, P. Hillion, and D. Sternheimer, Compt. Rend. **264**, A82 (1967).

[4] P. Roman and J. Rosen, J. Math. Phys. **7**, 2072 (1966).

algebras over the complex and real fields:

(a) the complexification of a real vector space (or Lie algebra) and,

(b) the construction of a real vector space (or Lie algebra) from a complex one by restriction of the base field.[5]

Process (a) is well known. A given vector space S over the real field R can be embedded in a standard way in a vector space S^C over the complex field C in such a manner that a basis $\{e_1, \cdots, e_n\}$ for S is also a basis for S^C. S^C is called the complexification of S. In an analogous way, a Lie algebra Λ over R can be embedded in a Lie algebra Λ^C over C, which is called the complexification of Λ. The vector space underlying Λ^C is the complexification of the vector space underlying Λ and the structure constants of Λ^C relative to a basis $\{e_1, \cdots, e_n\}$, which is also a basis for Λ, are real. Indeed, they are the same structure constants as those of Λ with respect to the same basis. As to process (b), consider a vector space S over C. By restricting the base field to R we deduce from S a vector space S_R over R which, as a set, coincides with S, but whose dimension is twice the dimension of S. Indeed, let $\{e_1, \cdots, e_n\}$ be a basis for S. Consider the decomposition

$$x = \sum_{i=1}^{n} x_i e_i$$

of a vector $x \in S$. Regarded as an element of S_R, x admits the decomposition

$$x = \sum_{i=1}^{n} (\mathrm{Re}\, x_i) e_i + \sum_{i=1}^{n} (\mathrm{Im}\, x_i) f_i, \quad (f_i = ie_i).$$

It is a trivial matter to show that the set $\{e_1, \cdots, e_n, f_1, \cdots, f_n\}$ is a basis for S_R.

Let now A be a linear transformation of S. Then A is automatically a linear transformation of S_R, which we shall denote by A^R. Let $\{A_{ik}\}$ be the matrix of A with respect to the basis $\{e_1, \cdots, e_n\}$ for S, i.e.,

$$Ae_k = \sum_{i=1}^{n} A_{ik} e_i.$$

Then the matrix of A^R with respect to the basis $\{e_1, \cdots, e_n, f_1, \cdots, f_n\}$ for S_R is

$$\begin{pmatrix} \{\mathrm{Re}\, A_{ik}\} & -\{\mathrm{Im}\, A_{ik}\} \\ \{\mathrm{Im}\, A_{ik}\} & \{\mathrm{Re}\, A_{ik}\} \end{pmatrix}. \tag{1}$$

If Λ is a Lie algebra over R and $\rho: a \to \rho(a)$, a representation of Λ by linear transformations of a complex vector space S, the mapping $\rho^R: a \to \rho^R(a)$ is a representation of Λ on S_R, i.e., a real representation. We shall refer to ρ^R as the real representation induced by the complex representation ρ.

[5] F. Gantmacher, Mat. Sb. 5 (47), 218 (1939).

Consider now a Lie algebra Σ over C. By restriction of the base field to R we obtain a Lie algebra Σ_R over R. Denote by S the vector space underlying Σ.

Then the vector space underlying Σ_R is S_R and if $\{M_1, \cdots, M_n\}$ is a basis for Σ, $\{M_1, \cdots, M_n, N_1 = iM_1, \cdots, N_n = iM_n\}$ is a basis for Σ_R and the structure constants of Σ_R relative to this basis are obtained automatically by taking real and imaginary parts of the structure constants c_{ij}^k of Σ relative to the basis $\{M_1, \cdots, M_n\}$. Specifically,[6]

$$\begin{aligned} [M_i, M_j] &= (\mathrm{Re}\, c_{ij}^k) M_k + (\mathrm{Im}\, c_{ij}^k) N_k, \\ [M_i, N_j] &= -(\mathrm{Im}\, c_{ij}^k) M_k + (\mathrm{Re}\, c_{ij}^k) N_k, \\ [N_i, N_j] &= -(\mathrm{Re}\, c_{ij}^k) M_k - (\mathrm{Im}\, c_{ij}^k) N_k. \end{aligned} \tag{2}$$

3. INHOMOGENIZATION OF A REAL LIE ALGEBRA RELATIVE TO A COMPLEX REPRESENTATION

We turn now to the problem of giving a consistent definition of inhomogenization of a real Lie algebra with respect to a complex representation. We recall the definition given in I of inhomogenization of a real Lie algebra Λ relative to a representation Δ on the real n-dimensional vector space R^n, as the semidirect product $\mathfrak{R}^n{}_{\Delta} \times \Lambda$, $\mathfrak{R}^n$ denoting the n-dimensional Abelian Lie algebra over R.

By definition, thus, the representation relative to which the inhomogenization is performed is a real representation. We shall extend as follows the concept of inhomogenization to include complex representations as well. Let $\rho: a \to \rho(a)$ be an n-dimensional complex representation of a real Lie algebra Λ, i.e., a representation of Λ by linear transformations of the complex n-dimensional vector space C^n. In Sec. 2 we have seen that ρ induces a representation ρ^R of Λ on the real vector space deduced from C^n by restriction to R of the base field. Identifying this vector space with R^{2n}, we define the inhomogenization of Λ relative to the complex representation ρ as the semidirect product $\mathfrak{R}^{2n}{}_{\rho^R} \times \Lambda$.

We shall now make clear the meaning of this definition. Given a real Lie algebra Λ, consider its complexification Λ^C (compare Sec. 2). A complex representation ρ of Λ on C^n can be extended in a unique way to a representation, which we again denote by ρ, of Λ^C on C^n. In agreement with the definition given in the real case, the semidirect product $\mathfrak{C}^n{}_{\rho} \times \Lambda^C$ ($\mathfrak{C}^n$ denoting the complex n-dimensional Abelian Lie algebra) will be called inhomogenization of Λ^C relative to the representation ρ. Indeed, $\mathfrak{C}^n$ is the Lie algebra of the complex Lie group of translations in the vector space C^n (this group will be denoted by the same

[6] Summation over repeated indices is understood.

symbol C^n as used for the vector space). We now show that $\mathfrak{R}^{2n}{}_{\rho^R}\times \Lambda$ is a subalgebra of the real Lie algebra obtained from ${\mathfrak{C}^n}_\rho \times \Lambda$ by restriction to R of the base field, whereby the Abelian ideal $\mathfrak{R}^{2n}$ can be interpreted as the Lie algebra of the group C^n regarded as a real Lie group. Thus we see that our definition is consistent, because the elements of $\mathfrak{R}^{2n}$ appear naturally as translationlike generators in the space on which the representation ρ of Λ acts.

To prove our statement, we choose a basis $\{M_1, \cdots, M_r\}$ ($r = \dim \Lambda$) for Λ. Then $\{M_1, \cdots, M_r\}$ is also a basis for Λ^C (compare Sec. 2) and the set of vectors $\{M_1, \cdots, M_r, T_1, \cdots, T_n\}$, where the T_μ's ($\mu = 1, 2, \cdots, n$) span $\mathfrak{C}^n$, is a basis for ${\mathfrak{C}^n}_\rho \times \Lambda^C$. Denoting by c_{ij}^k the structure constants of Λ relative to the basis $\{M_1, \cdots, M_r\}$, the commutation brackets for ${\mathfrak{C}^n}_\rho \times \Lambda^C$ are given by[6]:

$$\begin{aligned} [M_i, M_j] &= c_{ij}^k M_k, \\ [M_i, T_\mu] &= \rho_{\lambda\mu}(M_i) T_\lambda, \\ [T_\mu, T_\nu] &= 0, \end{aligned} \tag{3}$$

where $\{\rho_{\lambda\mu}(M_i)\}$ is the matrix representing the linear transformation $\rho(M_i)$ in the basis $\{T_1, \cdots, T_n\}$ for $\mathfrak{C}^n$.

We now go over to the real Lie algebra deduced from ${\mathfrak{C}^n}_\rho \times \Lambda^C$ by restriction to R of the base field. By choosing for this algebra the basis $\{M_i, iM_i, T_\mu, T_{\mu+n} = iT_\mu\}$ (compare Sec. 2) we note that, as a consequence of the c_{ij}^k being real, the subspace spanned by the set $\{M_i, T_\alpha\}$ ($i = 1, \cdots, r$; $\alpha = 1, \cdots, 2n$) is a subalgebra Γ. Recalling formula (1), we can write the commutation brackets for Γ in the form[6]:

$$\begin{aligned} [M_i, M_j] &= c_{ij}^k M_k, \\ [M_i, T_\alpha] &= \rho^R_{\beta\alpha}(M_i) T_\beta, \\ [T_\alpha, T_\beta] &= 0, \end{aligned} \tag{4}$$

and this shows that Γ can be identified with the Lie algebra $\mathfrak{R}^{2n}{}_{\rho^R}\times \Lambda$ which was defined as the inhomogenization of Λ relative to the complex representation ρ.

The symbol $I_\rho\Lambda$, which was introduced in I to denote the inhomogenization of a real Lie algebra Λ relative to a real representation ρ, will now be extended to designate inhomogenizations relative to complex representations as well. Thus, if ρ is real, $I_\rho\Lambda = {\mathfrak{R}^n}_\rho \times \Lambda$; if ρ is complex, $I_\rho\Lambda = \mathfrak{R}^{2n}{}_{\rho^R}\times \Lambda\ (= I_{\rho^R}\Lambda)$, $n = \dim \rho$. The ambiguity in writing $I_\rho\Lambda$ whenever ρ is not essentially complex,[7] which stems from the fact that ρ can be intended either as a real representation (acting on a real vector space S_R) or as a complex representation (acting on the complexification S_R of S_R), can be removed by keeping the notation $I_\rho\Lambda$ for the case when ρ is understood as real, while writing directly $I_{\rho^R}\Lambda$ when ρ is understood as complex. In the latter case, alternatively, as ρ^R is actually equivalent to the direct sum of two representations both equivalent to ρ (compare next section), one can as well use the notation of I, $I_{\{2\rho\}}\Lambda$.

Analogous to what has been done in the real case (compare I, Sec. 4) we note that, if the n-dimensional representation relative to which the inhomogenization is made is faithful, we can easily construct an $(n+1)$-dimensional faithful representation of the inhomogenization. This representation is complex and is defined by the mapping

$$a + t \to \begin{pmatrix} \rho_{11}(a) & \rho_{12}(a) & \cdots & \rho_{1n}(a) & t_1 + it_{n+1} \\ \rho_{21}(a) & \rho_{22}(a) & \cdots & \rho_{2n}(a) & t_2 + it_{n+2} \\ & & \cdots & & \\ \rho_{n1}(a) & \rho_{n2}(a) & \cdots & \rho_{nn}(a) & t_n + it_{2n} \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

$$a \in \Lambda, \quad t \in \mathfrak{R}^{2n}, \quad t = \sum_{\alpha=1}^{2n} t_\alpha T_\alpha, \quad n = \dim \rho. \tag{5}$$

Under (5),

$$T_\mu \to \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & 0 & 1 \\ & & \cdots & & \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \leftarrow \mu\text{th row}, \tag{5'}$$

$\mu = 1, 2, \cdots, n,$

$$T_{\mu+n} \to \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & 0 & i \\ & & \cdots & & \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \leftarrow \mu\text{th row}, \tag{5''}$$

$\mu = 1, 2, \cdots, n,$

$$M_i \to \begin{pmatrix} \rho_{11}(M_i) & \cdots & \rho_{1n}(M_i) & 0 \\ & \cdots & & \\ \rho_{n1}(M_i) & \cdots & \rho_{nn}(M_i) & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{5'''}$$

If ρ were not faithful, the mapping (5) would define an unfaithful representation of $I_\rho\Lambda$ of which the kernel is the kernel of ρ. Correspondences (5') and (5'') tell

[7] A complex representation ρ of a real Lie algebra Λ will be termed essentially complex if there exists no basis in the representation space with respect to which the operators of the representation are represented by matrices with real entries. If this is not the case, ρ is equivalent to a real representation (also denoted by ρ) and we shall say that it is not essentially complex.

us that the T_μ's (respectively, the $T_{\mu+n}$'s) are the generators of the real translations (respectively, of the purely imaginary translations) in the complex space on which ρ acts.

Besides (5), two real representations of $I_\rho\Lambda$ are also easily obtained. One is the real representation induced by (5) and it is $2(n+1)$-dimensional. The other is the $(2n+1)$-dimensional representation defined in I [Sec. 4, formula (12)] regarding $I_\rho\Lambda$ as $I_{\rho^R}\Lambda$.

4. IRREDUCIBLE INHOMOGENIZATIONS

Among the inhomogenizations of a real Lie algebra relative to its real representations, those for which the defining representations are irreducible[8] appear to have a particularly simple structure. Besides, all inhomogenizations which, to our knowledge, have been considered in physics up to now, are of this type.[1,3] In the following, we shall refer to inhomogenizations relative to irreducible real representations, as irreducible inhomogenizations. On the other hand, if the defining real representation of an inhomogenization is reducible, we shall speak of a reducible inhomogenization. In connection with irreducible inhomogenizations the following question then arises naturally: Given an inhomogenization $I_\rho\Lambda$ of a real Lie algebra Λ relative to a complex irreducible representation ρ, is $I_\rho\Lambda$ an irreducible inhomogenization when regarded as an inhomogenization relative to the real representation ρ^R, induced by ρ? And, conversely: Given an irreducible inhomogenization the defining representation Δ of which is induced by a complex representation ρ, is ρ irreducible? The two questions may be stated in other words, and independently of inhomogenizations, by asking if irreducibility of ρ implies irreducibility of ρ^R and vice versa. The answer is furnished by the following:

Theorem: Let ρ be an essentially complex n-dimensional representation of a real Lie algebra Λ and let ρ^R denote the real representation induced by ρ. Then ρ^R is irreducible if and only if ρ is irreducible. On the other hand, if ρ is irreducible but not essentially complex, ρ^R is completely reducible to a direct sum of two representations both equivalent to ρ.

For the proof we require the following:

Lemma: Let S_R denote the $2n$-dimensional real vector space deduced from an n-dimensional complex vector space S_C by restriction to R of the base field (compare Sec. 2). Let V_R be a k-dimensional subspace of S_R and let V_C denote the subspace of S_C generated by V_R. Then, dim $V_C \leq k$.

Proof: Every $x \in V_C$ may be expressed by a finite linear combination $x = \sum_i c_i x_i$, $c_i \in C$, $x_i \in V_R$. Let $\{e_1, \cdots, e_k\}$ be a basis for V_R. Write

$$x_i = \sum_{l=1}^{\beta} a_{il} e_l, \quad a_{il} \in R.$$

Then

$$x = \sum_{l=1}^{\beta} c_l' e_l, \quad c_l' = \sum_i c_i a_{il},$$

which shows that the set $\{e_1, \cdots, e_k\}$ generates V_C. Since the vectors $\{e_1, \cdots, e_k\}$ are not necessarily linearly independent in S_C, we conclude that dim $V_C \leq k$.

We now proceed to the proof of the theorem. Denote by S_C the complex space on which ρ acts. The real space on which ρ^R acts is the space S_R deduced from S_C by restriction to R of the base field. Let ρ^R be irreducible. Then ρ is also irreducible. In fact, suppose there is a proper subspace V_C of S_C which is invariant under ρ. Regarded as a set of elements of S_R, V_C is a proper subspace of S_R and it is invariant under ρ^R, which contradicts the hypothesis. Further, ρ is essentially complex. For, if it were not so, we could find a basis for S_C, $\{e_1, \cdots, e_n\}$, say, with respect to which the elements $\rho(a)$ $(a \in \Lambda)$ of the representation are represented by matrices with real entries

$$\rho e_i = \sum_{k=1}^{n} \rho_{ki} e_k, \quad \rho_{ki} \in R.^9$$

Then, formula (1) gives, for the matrix representing ρ^R with respect to the "canonical" basis $\{e_1, \cdots, e_n, f_1, \cdots, f_n\}$ $(f_i = ie_i)$ for S_R, the expression

$$\begin{pmatrix} \{\rho_{ik}\} & 0 \\ 0 & \{\rho_{ik}\} \end{pmatrix},$$

which shows that ρ^R is completely reducible to the direct sum of two representations both equivalent to ρ, and this again contradicts the hypothesis.

We shall now prove the less trivial result that irreducibility of ρ implies irreducibility of ρ^R, if ρ is essentially complex. For, let us suppose that ρ^R is reducible. This implies the existence, in S_R, of a proper invariant irreducible subspace V. We shall distinguish three cases, to be treated separately: (1) dim $V <$ $n =$ dim S_C, (2) dim $V = n$, and (3) dim $V > n$. We note that invariance of V under ρ^R obviously implies invariance under ρ of V_C, the complex closure of V.

[8] The term "irreducibility," referred to real representations, has to be intended here as "irreducibility with respect to real similarity transformations."

[9] For simplicity, we write ρ (ρ^R) in place of $\rho(a)$ $(\rho^R(a))$, $a \in \Lambda$.

Consider the first case. By the previously stated lemma we would get dim $V_C < n$, and this contradicts the hypothesis that ρ is irreducible.

In the second case, let $\{e_1, \cdots, e_n\}$ be a basis for V. Then $\{e_1, \cdots, e_n\}$ must be a basis for S_C as well, otherwise we would again obtain dim $V_C < n$, which is absurd. Let ρ_{ki} be the matrix of ρ relative to the basis $\{e_1, \cdots, e_n\}$ for S_C,

$$\rho e_i = \sum_{k=1}^{n} \rho_{ki} e_k .$$

Since, by hypothesis, V is invariant under ρ^R one has

$$\rho^R e_i = \sum_{k=1}^{n} \rho^R_{ki} e_k .$$

By comparison, one gets

$$\sum_{k=1}^{n} (\rho_{ki} - \rho^R_{ki}) e_k = 0,$$

which implies $\rho_{ki} = \rho^R_{ki}$. Hence ρ is real, at variance with the hypothesis.

Finally, consider case (3). Let $\{e_1, \cdots, e_s\}$, $s = \dim V > n$, be a basis for V. The set of vectors $\{f_1, \cdots, f_s\}$ $(f_i = ie_i)$ forms a basis for an s-dimensional subspace $\bar{V}$ of S_R, which is invariant under ρ^R. Irreducibility of ρ implies that we can select in V vectors $\bar{e}_1, \cdots, \bar{e}_n$ which form a basis for S_C. Then $\{\bar{e}_1, \cdots, \bar{e}_n, i\bar{e}_1, \cdots, i\bar{e}_n\}$ is a basis for S_R and, since the $i\bar{e}_j$'s belong to $\bar{V}$, we get that the union $V \cup \bar{V}$ generates S_R. Since both V and $\bar{V}$ are invariant under ρ^R this is true also for their intersection $I = V \cap \bar{V}$ which, by the hypothesis of irreducibility of V, implies $I = 0$ or $I = V$.

Let $I = 0$. This would imply $S_R = V \oplus \bar{V}$ which is impossible since dim $S_R = 2n$, dim V + dim $\bar{V}$ = $2s$ and, by hypothesis, $s > n$. The other possibility, $I = V$, implies $V \subseteq \bar{V}$. Compatibility with dim V = dim $\bar{V}$ gives $V = \bar{V}$. Hence, $V \cup \bar{V} = V$, which implies $V = S_R$ in contradiction with the hypothesis that V is a proper subspace of S_R. The proof of the theorem is thus completed.

This theorem allows us to extend without ambiguity the term irreducible inhomogenization to include inhomogenizations relative to complex irreducible representations as well, provided the latter representations are essentially complex.

Thus, when speaking of an irreducible inhomogenization, it has to be understood that we mean an inhomogenization relative to a real or essentially complex irreducible representation. An inhomogenization of a real Lie algebra Λ relative to a complex irreducible representation ρ is, in our sense, reducible, if ρ is not essentially complex.

5. CLASSIFICATION OF THE COMPLEX INHOMOGENIZATIONS OF SEMISIMPLE LIE ALGEBRAS

In this section we shall be concerned with the characterization of the isomorphism of two inhomogenizations $I_\rho\Lambda$ and $I_{\rho'}\Lambda$ of a given real semisimple Lie algebra Λ, in terms of the relation between the corresponding defining representations ρ and ρ'. In I we have proved that if ρ and ρ' are real representations, $I_\rho\Lambda$ is isomorphic to $I_{\rho'}\Lambda$ if and only if ρ is *quasiequivalent* to ρ'.[10] It is then evident that if ρ is complex and ρ' is real, the two inhomogenizations are isomorphic if and only if ρ^R is quasiequivalent to ρ'. As to the case when ρ and ρ' are both complex, the problem of establishing the conditions under which $I_\rho\Lambda$ and $I_{\rho'}\Lambda$ are isomorphic obviously reduces to the question of when ρ and ρ' give rise to quasiequivalent induced real representations. This question is answered by the following:

Theorem: The necessary and sufficient condition for two complex representations ρ and ρ' of a real semisimple Lie algebra Λ to induce quasiequivalent real representations ρ^R and ρ'^R is that the decomposition of ρ and ρ' into irreducible components be the same, up to quasiequivalence and up to complex conjugation of some of the components.

In other words (compare I, Sec. 5), there should exist a one-to-one mapping π of the set Σ whose elements are the irreducible components of ρ onto the set Σ' of the irreducible components of ρ', and an automorphism I of Λ such that, $\forall \sigma \in \Sigma$, $\pi(\sigma) \circ I$ is either equivalent to σ or it is equivalent to σ^*, the representation complex conjugate to σ.

Proof: We start by showing that the condition is necessary. Observe first that, as obviously ρ and ρ' are of equal dimension, the complex vector space S_C on which ρ acts can be identified with the one on which ρ' acts. Correspondingly, ρ^R and ρ'^R can both be made to act on the same space S_R, which is the real vector space obtained from S_C by restriction to R of the base field.

Let now ρ_{ik} be the matrix representing ρ (supposed to be n-dimensional) relative to a basis $\{e_1, e_2, \cdots, e_n\}$ for S_C: $\sum_{k=1}^{n} \rho e_k = \rho_{ik} e_i$.[9] By (1), the matrix representing ρ^R relative to the basis $\{e_1, e_2, \cdots, e_n, ie_1, ie_2, \cdots, ie_n\}$ for S_R is given by

$$\begin{pmatrix} \{\mathrm{Re}\ \rho_{ik}\} & -\{\mathrm{Im}\ \rho_{ik}\} \\ \{\mathrm{Im}\ \rho_{ik}\} & \{\mathrm{Re}\ \rho_{ik}\} \end{pmatrix}. \tag{6}$$

[10] We recall the definition of quasiequivalence of representations, which was introduced in I, Sec. 5. Two representations ρ and ρ' of a Lie algebra Λ are termed *quasiequivalent* if they have equal dimension and if a nonsingular endomorphism A of the vector space on which they act and an automorphism I of Λ exist such that $A\rho_l = \rho'_{I(l)}A = (\rho' \circ I)_l A$, $\forall l \in \Lambda$.

It is easy to see that ρ^R, considered as acting on the complexification S_R^C of S_R, is completely reducible to the direct sum of a representation equivalent to ρ plus a representation equivalent to ρ^*, the complex conjugate of ρ. This reduction is accomplished through the similarity transformation induced by the following block matrix:

$$T = \begin{pmatrix} iE & -iE \\ E & E \end{pmatrix}, \tag{7}$$

with E the $n \times n$ unit matrix.

Indeed,[11]

$$T^{-1}\begin{pmatrix} \operatorname{Re}\rho & -\operatorname{Im}\rho \\ \operatorname{Im}\rho & \operatorname{Re}\rho \end{pmatrix} T = \frac{1}{2}\begin{pmatrix} -iE & E \\ iE & E \end{pmatrix}\begin{pmatrix} \operatorname{Re}\rho & -\operatorname{Im}\rho \\ \operatorname{Im}\rho & \operatorname{Re}\rho \end{pmatrix}\begin{pmatrix} iE & -iE \\ E & E \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ 0 & \rho^* \end{pmatrix}. \tag{8}$$

Now, as Λ is semisimple, ρ is completely reducible to a direct sum of irreducible components

$$\rho \to \begin{pmatrix} \rho_1 & & & 0 \\ & \rho_2 & & \\ & & \ddots & \\ 0 & & & \rho_r \end{pmatrix}. \tag{9}$$

Hence, combining (8) and (9), the complete reduction of ρ^R is written

$$\rho^R \to \begin{pmatrix} \rho_1 & & & & & & \\ & \rho_2 & & & 0 & & \\ & & \ddots & & & & \\ & & & \rho_r & & & \\ & & & & \rho_1^* & & \\ & & & & & \rho_2^* & \\ 0 & & & & & & \ddots \\ & & & & & & & \rho_r^* \end{pmatrix}. \tag{10}$$

[11] We suppress matrix-element indices, in order to simplify the notation.

By virtue of the quasiequivalence of ρ^R to ρ'^R, there exists an automorphism I of Λ such that a basis can be found for S_R^C relative to which the matrix representing ρ'^R displays the completely reduced form

$$\rho'^R \to \begin{pmatrix} \rho_1 \circ I & & & & & & \\ & \rho_2 \circ I & & & 0 & & \\ & & \ddots & & & & \\ & & & \rho_r \circ I & & & \\ & & & & \rho_1^* \circ I & & \\ & & & & & \rho_2^* \circ I & \\ 0 & & & & & & \ddots \\ & & & & & & & \rho_r^* \circ I \end{pmatrix}. \tag{11}$$

On the other hand, since ρ'^R is induced by ρ', there exists another basis for S_R^C relative to which ρ'^R is completely reduced to a direct sum

$$\rho'^R \to \begin{pmatrix} \rho_1' & & & & & & \\ & \rho_2' & & & 0 & & \\ & & \ddots & & & & \\ & & & \rho_t' & & & \\ & & & & \rho_1'^* & & \\ & & & & & \rho_2'^* & \\ 0 & & & & & & \ddots \\ & & & & & & & \rho_t'^* \end{pmatrix}. \tag{12}$$

Now, as is well known, when a representation is completely reducible, its decomposition into irreducible components is unique except for order and except for equivalence. Hence $t = r$ and, by setting $\rho_j^* = \rho_{j+r}$, $\rho_j'^* = \rho_{j+r}'$ $(j = 1, 2, \cdots, r)$, we have that there exists a permutation π of the indices $1, 2, \cdots, 2r$ such that $\rho_{\pi(\lambda)} \circ I$ is equivalent to ρ_λ' $(\lambda = 1, 2, \cdots, 2r)$. This simply tells us that a given irreducible component ρ_i' of ρ' is either quasiequivalent to one of the

components of ρ (if $1 \leq \pi(i) \leq r$) or it is quasiequivalent to the complex conjugate of one of these components (if $r+1 \leq \pi(i) \leq 2r$).

We now turn to the proof that the condition is sufficient. Specifically, suppose that the decomposition of ρ and ρ' into irreducible components is the same up to quasiequivalence and up to complex conjugation of some of the components. This implies the existence of two bases $\{e_1^{(1)}, \cdots, e_{j_1}^{(1)}, e_1^{(2)}, \cdots, e_{j_2}^{(2)}, \cdots, e_1^{(r)}, \cdots, e_{j_r}^{(r)}\}$ and $\{e_1'^{(1)}, \cdots, e_{j_1}'^{(1)}, e_1'^{(2)}, \cdots, e_{j_2}'^{(2)}, \cdots, e_1'^{(r)}, \cdots, e_{j_r}'^{(r)}\}$ for S_C with respect to which ρ and ρ' respectively, display the completely reduced forms[12]

$$\rho \to \begin{pmatrix} \rho_1 & & & 0 \\ & \rho_2 & & \\ & & \ddots & \\ 0 & & & \rho_r \end{pmatrix} \tag{13}$$

and

$$\rho' \to \begin{pmatrix} \rho_1' & & & 0 \\ & \rho_2' & & \\ & & \ddots & \\ 0 & & & \rho_r' \end{pmatrix}. \tag{14}$$

Here, for any given $l = 1, 2, \cdots, r$, and denoting by I a suitable automorphism of Λ, the matrix ρ_l' is either equal to the matrix $\rho_l \circ I$, or it is equal to the complex conjugate matrix $\rho_l^* \circ I$. The matrix representing ρ^R with respect to the basis $\{e_1^{(1)}, \cdots, e_{j_1}^{(1)}, ie_1^{(1)}, \cdots, ie_{j_1}^{(1)}, \cdots, e_1^{(r)}, \cdots, e_{j_r}^{(r)}, ie_1^{(r)}, \cdots, ie_{j_r}^{(r)}\}$ for S_R is then

$$\rho^R \to \begin{pmatrix} \rho_1^R & & & 0 \\ & \rho_2^R & & \\ & & \ddots & \\ 0 & & & \rho_r^R \end{pmatrix} \tag{15}$$

with

$$\rho_l^R \to \begin{pmatrix} \operatorname{Re}\rho_l & -\operatorname{Im}\rho_l \\ \operatorname{Im}\rho_l & \operatorname{Re}\rho_l \end{pmatrix}. \tag{16}$$

Similarly, the matrix representing ρ'^R with respect to the basis $\{e_1'^{(1)}, \cdots, e_{j_1}'^{(1)}, ie_1'^{(1)}, \cdots, ie_{j_1}'^{(1)}, \cdots, e_1'^{(r)}, \cdots, e_{j_r}'^{(r)}, ie_1'^{(r)}, \cdots, ie_{j_r}'^{(r)}\}$ is

$$\rho'^R \to \begin{pmatrix} \rho_1'^R & & & 0 \\ & \rho_2'^R & & \\ & & \ddots & \\ 0 & & & \rho_r'^R \end{pmatrix}, \tag{17}$$

where

$$\rho_l'^R \to \begin{pmatrix} \operatorname{Re}\rho_l & -\operatorname{Im}\rho_l \\ \operatorname{Im}\rho_l & \operatorname{Re}\rho_l \end{pmatrix} \circ I \tag{18}$$

or

$$\rho_l'^R \to \begin{pmatrix} \operatorname{Re}\rho_l & \operatorname{Im}\rho_l \\ -\operatorname{Im}\rho_l & \operatorname{Re}\rho_l \end{pmatrix} \circ I, \tag{19}$$

according to whether $\rho_l' = \rho_l \circ I$ or $\rho_l' = \rho_l^* \circ I$.

Consider the $n \times n$ block diagonal matrix

$$R = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{pmatrix}, \tag{20}$$

where the lth block A_l is $2j_l$-dimensional and has the structure (E being the $j_l \times j_l$ unit matrix)

$$A_l = \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} \tag{21}$$

or

$$A_l = \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}, \tag{22}$$

according to whether $\rho_l'^R$ is given by (18) or by (19). One easily verifies that

$$\begin{pmatrix} \rho_1'^R & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \rho_r'^R \end{pmatrix} = R\left\{ \begin{pmatrix} \rho_1^R & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \rho_r^R \end{pmatrix} \circ I \right\} R^{-1}. \tag{23}$$

This establishes the quasiequivalence of $\rho^{R'}$ to ρ^R. The proof of the theorem is thus completed.

Complex inhomogenizations form a special subclass of the class of all inhomogenizations of a given real Lie algebra.

[12] The meaning of the labels on the basis vectors is self-explanatory: the vectors $\{e_1^{(l)}, \cdots, e_{j_l}^{(l)}\}$ (respectively, $\{e_1'^{(l)}, \cdots, e_{j_l}'^{(l)}\}$), $l = 1, 2, \cdots, r$, span the invariant irreducible subspace of S_C on which the irreducible component ρ_l (respectively, ρ_l') acts.

A complete classification of the complex inhomogenizations of a real semisimple Lie algebra can now be given which easily stems from the preceding theorem and from the analogous classification which was given in I for arbitrary inhomogenizations.

Specifically, to every (up to isomorphism) complex inhomogenization of a real semisimple Lie algebra we can uniquely associate a pair (m, Ω), where the nonnegative integer m is half the dimension of the center of the inhomogenization[13] and Ω is a system of complex irreducible representations which is fixed up to quasiequivalence and up to complex conjugation of some of its elements. Conversely, to every such pair there corresponds a unique (up to isomorphism) inhomogenization.

ACKNOWLEDGMENTS

We thank Professors P. Caldirola and A. Loinger for their kind interest and Professor F. Duimio for valuable discussions.

[13] Actually, m is the number of times the trivial one-dimensional representation appears in the reduction of the complex representation ρ relative to which the inhomogenization is made. Each trivial component acts on a complex one-dimensional space where complex translations are introduced, thus giving rise to two independent translation generators (one for real and the other for purely imaginary translations), both commuting with the "homogeneous" part. The dimension of the center of the inhomogenization is thus even, and equal to $2m$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Ising-Model Spin Correlations

ROBERT E. HARTWIG
Baker Laboratory, Cornell University, Ithaca, New York
AND
JOHN STEPHENSON
Department of Mathematics, University of Adelaide, Adelaide, South Australia

(Received 11 November 1967)

In this paper we study a class of planar Ising lattices which have two or more directions along which the pair-correlation function admits a Toeplitz determinant representation. On the basis of the recently developed theory of Toeplitz determinants, we discuss the asymptotic behavior of these correlations at the critical point.

1. INTRODUCTION

Recently Wu[1] has investigated the asymptotic behavior of the pair-correlation function[2] $\langle s_k s_{k+N}\rangle$ along a row of the infinite rectangular Ising lattice, as the separation distance N increases indefinitely at fixed temperatures. Wu's approach uses the similarity between the Toeplitz determinant $D_N = |a_{i-j}|_0^{N-1}$, representing the correlation,[3] and the corresponding Wiener–Hopf sum equation of its formal generating function $F(\theta) = \sum_{-\infty}^{\infty} a_n e^{in\theta}$, to obtain asymptotic expansions for the Toeplitz determinant. The relevant generating function $F(\theta)$ is given explicitly by

$$F(\theta) = \left[\frac{1 - Ae^{-i\theta}}{1 - Be^{-i\theta}} \cdot \frac{1 - Be^{i\theta}}{1 - Ae^{i\theta}}\right]^{\frac{1}{2}},$$
$$(T < T_c,\ 0 < B < A < 1)$$
$$= ie^{-\frac{1}{2}i\theta}\left[\frac{1 - Be^{i\theta}}{1 - Be^{-i\theta}}\right]^{\frac{1}{2}},$$
$$(T = T_c,\ 0 < B < A = 1)$$
$$= -e^{-i\theta}\left[\frac{(1 - A^{-1}e^{i\theta})(1 - Be^{i\theta})}{(1 - A^{-1}e^{-i\theta})(1 - Be^{-i\theta})}\right]^{\frac{1}{2}},$$
$$(T > T_c,\ 0 < B < A^{-1} < 1), \quad (1.1)$$

where the parameters A and B depend on the interaction energies between nearest-neighboring pairs of spins and on the temperature (see for example, the paper by Green[4]). The square root is taken such that for $T < T_c$, $F(\pi) > 0$ and $\ln F(2\pi) = \ln F(0)$, for $T = T_c$, $F(2\pi) = -F(0) = i$, and for $T > T_c$, $\ln F(2\pi) - \ln F(0) = -2\pi i$.

The purpose of this article is to point out that Wu's analysis can be applied directly to correlations along "axes of symmetry" of more general lattices, such as the triangular, hexagonal, and modified

[1] T. T. Wu, Phys. Rev. **149**, 380 (1966).
[2] Since we only deal with pair correlations we simply refer to them as correlations.
[3] E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. **4**, 308 (1963).
[4] H. S. Green, Z. Physik **171**, 129 (1963).

A complete classification of the complex inhomogenizations of a real semisimple Lie algebra can now be given which easily stems from the preceding theorem and from the analogous classification which was given in I for arbitrary inhomogenizations.

Specifically, to every (up to isomorphism) complex inhomogenization of a real semisimple Lie algebra we can uniquely associate a pair (m, Ω), where the nonnegative integer m is half the dimension of the center of the inhomogenization[13] and Ω is a system of complex irreducible representations which is fixed up to quasiequivalence and up to complex conjugation of some of its elements. Conversely, to every such pair there corresponds a unique (up to isomorphism) inhomogenization.

ACKNOWLEDGMENTS

We thank Professors P. Caldirola and A. Loinger for their kind interest and Professor F. Duimio for valuable discussions.

[13] Actually, m is the number of times the trivial one-dimensional representation appears in the reduction of the complex representation ρ relative to which the inhomogenization is made. Each trivial component acts on a complex one-dimensional space where complex translations are introduced, thus giving rise to two independent translation generators (one for real and the other for purely imaginary translations), both commuting with the "homogeneous" part. The dimension of the center of the inhomogenization is thus even, and equal to $2m$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Ising-Model Spin Correlations

ROBERT E. HARTWIG
Baker Laboratory, Cornell University, Ithaca, New York
AND
JOHN STEPHENSON
Department of Mathematics, University of Adelaide, Adelaide, South Australia

(Received 11 November 1967)

In this paper we study a class of planar Ising lattices which have two or more directions along which the pair-correlation function admits a Toeplitz determinant representation. On the basis of the recently developed theory of Toeplitz determinants, we discuss the asymptotic behavior of these correlations at the critical point.

1. INTRODUCTION

Recently Wu[1] has investigated the asymptotic behavior of the pair-correlation function[2] $\langle s_k s_{k+N}\rangle$ along a row of the infinite rectangular Ising lattice, as the separation distance N increases indefinitely at fixed temperatures. Wu's approach uses the similarity between the Toeplitz determinant $D_N = |a_{i-j}|_0^{N-1}$, representing the correlation,[3] and the corresponding Wiener–Hopf sum equation of its formal generating function $F(\theta) = \sum_{-\infty}^{\infty} a_n e^{in\theta}$, to obtain asymptotic expansions for the Toeplitz determinant. The relevant generating function $F(\theta)$ is given explicitly by

$$F(\theta) = \left[\frac{1 - Ae^{-i\theta}}{1 - Be^{-i\theta}} \cdot \frac{1 - Be^{i\theta}}{1 - Ae^{i\theta}}\right]^{\frac{1}{2}},$$
$$(T < T_c,\ 0 < B < A < 1)$$
$$= ie^{-\frac{1}{2}i\theta}\left[\frac{1 - Be^{i\theta}}{1 - Be^{-i\theta}}\right]^{\frac{1}{2}},$$
$$(T = T_c,\ 0 < B < A = 1)$$
$$= -e^{-i\theta}\left[\frac{(1 - A^{-1}e^{i\theta})(1 - Be^{i\theta})}{(1 - A^{-1}e^{-i\theta})(1 - Be^{-i\theta})}\right]^{\frac{1}{2}},$$
$$(T > T_c,\ 0 < B < A^{-1} < 1), \quad (1.1)$$

where the parameters A and B depend on the interaction energies between nearest-neighboring pairs of spins and on the temperature (see for example, the paper by Green[4]). The square root is taken such that for $T < T_c$, $F(\pi) > 0$ and $\ln F(2\pi) = \ln F(0)$, for $T = T_c$, $F(2\pi) = -F(0) = i$, and for $T > T_c$, $\ln F(2\pi) - \ln F(0) = -2\pi i$.

The purpose of this article is to point out that Wu's analysis can be applied directly to correlations along "axes of symmetry" of more general lattices, such as the triangular, hexagonal, and modified

[1] T. T. Wu, Phys. Rev. **149**, 380 (1966).
[2] Since we only deal with pair correlations we simply refer to them as correlations.
[3] E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. **4**, 308 (1963).
[4] H. S. Green, Z. Physik **171**, 129 (1963).

Yamamoto lattices, since these correlations can also be expressed as a single Toeplitz determinant $D_N = D_N(A, B)$, with different values of A and B depending on the lattice and the relative orientation of the spins. In order to show this, use is made of the transformations of the partition function as developed by Fisher[5] and Green.[4] Because certain spin sites in the basic unit cell remain invariant under these transformations (i.e., they are not summed over in the reduction process), the correlation between two of these spins can be reduced to a standard form, which can be expressed, using the Pfaffian approach, as the product of the eigenvalues of a homogeneous integral equation.[4] The integral equation is then identified with the eigenvalue equation of a Toeplitz determinant $D_N(A, B)$,[6] with the appropriate values for A and B. This includes in particular, as shown in Sec. 6, the known results for the row and diagonal correlations on the rectangular and triangular lattices which were derived by means of a slightly different Pfaffian approach.[3,7]

Taking the papers by Wu and Green as our starting point, we obtain from the known asymptotic expansion of $D_N(1, B)$, the asymptotic behavior of the correlations at the critical point for the various lattices, in terms of the ratio of the radial separation distance $r = |\mathbf{r}|$ to the first nearest-neighbor lattice spacing a_1.

2. THE RECTANGULAR LATTICE

For the infinite rectangular lattice with interaction energies $-J_i = -kTK_i$ between nearest-neighboring horizontal ($i = 1$) and vertical ($i = 2$) pairs of spins, the parameters A and B are given, in Green's notation,[4] by

$$A = x_2^*/x_1, \quad B = x_2^* x_1, \tag{2.1}$$

where $x_i = \text{th}\, K_i = e^{-2K_i^*}$ and its dual is given by $x_i^* = \text{th}\, K_i^* = e^{-2K_i}$ $(i = 1, 2)$.

For temperatures above and below the critical point the question of the asymptotic decay of the row correlation has been completely settled.[1,8] The rate at which the Toeplitz determinant $D_N = D_N(A, B)$ converges to its limit is

$$D_N \sim J^2 \frac{A^{-N}}{(\pi N)^{\frac{1}{2}}}\left[1 + O\left(\frac{1}{N}\right)\right], \quad (T > T_c) \tag{2.2}$$

and

$$D_N - I^2 \sim I^2 \frac{A^{2N}}{2\pi N^2}\left[\frac{A}{1 - A^2}\right]^2\left[1 + O\left(\frac{1}{N}\right)\right], \quad (T < T_c), \tag{2.3}$$

where

$$I = [(1 - A^2)(1 - B^2)/(1 - AB)^2]^{\frac{1}{8}}$$

is the spontaneous magnetization and

$$J = [(1 - B^2)/(1 - A^{-2})(1 - AB)^2]^{\frac{1}{8}}.$$

At the critical point, however, the decay has been calculated rigorously, only for the diagonal case where $B = 0$. With the aid of the theory of Cauchy–Toeplitz determinants, one finds[1,9–14]

$$D_N(1, 0) \sim \frac{E}{N^{\frac{1}{4}}}\left[1 - \frac{1}{64N^2} + \frac{17}{4096N^4}\right], \tag{2.4}$$

where

$$E = \exp\left[-\tfrac{1}{4}(1 + C) - \sum_{s=2}^{\infty} \frac{\zeta(2s - 1)}{s4^s}\right] = 0.645002448\cdots$$

and C is Euler's constant.

For nonzero values of B, Wu derived the following result:

$$D_N(1, B) \sim \frac{E}{N^{\frac{1}{4}}}\left(\frac{1 + B}{1 - B}\right)^{\frac{1}{4}}\left[1 - \frac{1}{8N^2}\left[\frac{1}{8} - \frac{B}{(1 - B)^2}\right]\right], \tag{2.5}$$

on the basis of a conjecture equivalent to[13]

$$\lim_{N\to\infty} \frac{D_N(1, B)}{D_N(1, 0)} = \lim_{N\to\infty} \frac{D_N^*(1, B)}{D_N^*(1, 0)}, \tag{2.6}$$

where $D_N^*(A, B)$ is the so-called *resultant* approximation of $D_N(A, B)$. The resultant approximations have been discussed elsewhere[6] and are readily seen to satisfy[13]

$$\lim_{N\to\infty} \frac{D_N^*(1, B)}{D_N^*(1, 0)} = \left[\frac{1 + B}{1 - B}\right]^{\frac{1}{4}}. \tag{2.7}$$

The assumption (2.6) seems to be correct but has not yet been proved rigorously. Suppose we now introduce the radial separation vector $\mathbf{r} = re^{i\theta}$ between the two spins, and the first, second, etc., nearest-neighbor lattice spacing $a_1, a_2, \cdots$, etc. Then in general the ratio r/a_1 will be a multiple of the separation index N, i.e., $r/a_1 = \gamma N$, where γ depends on the lattice and the orientation θ of the vector $\mathbf{r}$. Hence the correlation function $\omega^L(\mathbf{r}/a_1) \equiv D_N$, for a lattice L, can be expanded for large-spin separation as

$$\omega^L(r/a_1, \theta) \sim \frac{E_0^L(\theta)}{(r/a_1)^{\frac{1}{4}}}\left[1 + \frac{E_1^L(\theta)}{(r/a_1)^2}\right], \tag{2.8}$$

[5] M. E. Fisher, Phys. Rev. **113**, 969 (1959).
[6] R. E. Hartwig, Australian J. Math. (to be published).
[7] J. Stephenson, J. Math. Phys. **5**, 1009 (1964).
[8] L. P. Kadanoff, Nuovo Cimento **44**, 276 (1966).
[9] B. Kaufman and L. Onsager, Phys. Rev. **76**, 1244 (1949).
[10] M. E. Fisher, Physica **25**, 521 (1959).
[11] H. Stillinger and H. L. Frisch, Physica **27**, 751 (1961).
[12] C. Domb, Advan. Phys. **9**, Nos. 34, 35 (1960).
[13] R. E. Hartwig and J. Stephenson, "A Note on Ising Model Spin-Correlations" (unpublished, 1965), in which essentially the same conjecture (2.6) was made and the verification of the Ornstein–Zernike law (2.2) above the critical point was also indicated.
[14] R. E. Hartwig, J. Math. Phys. **7**, 286 (1966).

where

$$E_0^L(\theta) = \left(\gamma \frac{1+B}{1-B}\right)^{\frac{1}{4}} E,$$

$$E_1^L(\theta) = -\frac{\gamma^2}{8}\left[\frac{1}{8} - \frac{B}{(1-B)^2}\right], \qquad (2.9)$$

which are also direction- and lattice-dependent. In the special cases where $B = Ax_1^2$, which include the rectangular and triangular lattices as may be seen from (7.10), these coefficients can be simplified to

$$E_0^L(\theta) = (\gamma \operatorname{ch} 2K_1)^{\frac{1}{4}} E, \quad E_1^L(\theta) = \frac{\gamma^2}{64}(\operatorname{ch} 4K_1 - 2). \qquad (2.10)$$

The expression (2.8) contains a "divergent" factor (in the infinite-product sense) which is the same for all lattices and yields the $r^{-\frac{1}{4}}$ decay. The coefficients E_0^L and E_1^L have to be evaluated at the critical point for the relevant direction θ under consideration, from which some interesting relations between the decay of the various correlations can be obtained, indicating the θ independence of the correlation at T_c. The form of the asymptotic decay in (2.5) and (2.8) has been tested by numerical evaluation of the Toeplitz determinants D_N at the critical point for the diagonal and row correlations on the square and triangular lattices.[7] Estimates for E_0^L and E_1^L exist for these cases, and will be compared with the exact values obtained.

For the row correlation on the quadratic lattice at the critical point, we have $N = r/a_1$, $\theta = 0$, $x_c = \sqrt{2} - 1$, $B = 3 - 2\sqrt{2}$ and $\operatorname{ch} 2K = \sqrt{2}$. Thus on the basis of (2.5) and (2.8)

$$\omega^Q(r/a_1, 0) \sim \frac{2^{\frac{1}{8}}E}{(r/a_1)^{\frac{1}{4}}}\left[1 + \frac{1}{64(r/a_1)^2}\right], \qquad (2.11)$$

where E is defined as in (2.4). The asymptotic analysis of the determinant $D_N(A, B)$ can now be applied to other planar lattices which have two or more directions along which the correlation can be represented by a Toeplitz determinant. As a first example we consider in the next section the triangular lattice.

3. THE TRIANGULAR LATTICE

It was shown by Stephenson,[7] using the approach of Montroll, Potts, and Ward, that the row correlation on the infinite triangular lattice between spins s_k and s_{k+N} can be expressed as the Toeplitz determinant $D_N(A, B)$. If the interaction energies between adjacent spins in the direction i is $-J_i = -kTK_i$, $i = 1, 2, 3$ (labeled anticlockwise), then

$$A = x_1^+/x_1, \quad B = x_1^+ x_1, \qquad (3.1)$$

where as usual $x_i = \operatorname{th} K_i = e^{-2K_i^*}$, $x_i^+ = \operatorname{th} K_i^+ = e^{-2K_i^{+*}}$. K_i^+ is here the inversion transformation variable, its dual K_i^{+*} being obtained from K_i by a star-triangle transformation,[15] i.e.,

$$e^{-4K_1^+} = (x_1^{+*})^2 = \frac{(x_2 + x_3x_1)(x_3 + x_1x_2)}{(x_1 + x_2x_3)(1 + x_1x_2x_3)} \quad (+ \text{cyclic}). \qquad (3.2)$$

When $x_3 = 0$, (3.1) reduces to (2.1) and we obtain the rectangular-lattice row correlation. When $x_1 = 0$, x_1^+ and B also vanish, and A reduces to $(\operatorname{sh} 2K_3 \operatorname{sh} 2K_1)^{-1}$. In this case the correlation reduces to the diagonal correlation on the rectangular lattice. For the isotropic triangular lattice at the critical point $N = r/a_1$, $x_c = 2 - \sqrt{3}$, $B = 7 - 4\sqrt{3}$ and $\operatorname{ch} 2K = 2/\sqrt{3}$. Hence application of the asymptotic expansion (2.5) yields

$$\omega^T(r/a_1, \theta) \sim \left(\frac{2}{\sqrt{3}}\right)^{\frac{1}{4}} \frac{E}{(r/a_1)^{\frac{1}{4}}}\left[1 - \frac{1}{192(r/a_1)^2}\right], \qquad (3.3)$$

where $\theta = 0, \pi/3, 2\pi/3$. Similarly for the diagonal correlation on the quadratic lattice at $T = T_c$, $N = r/a_2$, $a_2 = a_1\sqrt{2}$ and $B = 0$, so that by (2.4)

$$\omega^Q(r/a_1, \pi/4) \sim \frac{2^{\frac{1}{8}}E}{(r/a_1)^{\frac{1}{4}}}\left[1 - \frac{1}{32(r/a_1)^2}\right]. \qquad (3.4)$$

Comparison of the asymptotic constants in (2.8) shows that

$$E_0^Q(0) = E_0^Q(\pi/4) = 2^{\frac{1}{8}}E = 0.703380156\cdots,$$
$$E_0^T(0) = (\tfrac{2}{3})^{\frac{1}{8}}E_0^Q(0) = 0.66861896\cdots, \qquad (3.5)$$

and

$$E_1^Q(0) = -\tfrac{1}{2}E_1^Q(\pi/4) = -3E_1^T(0) = \tfrac{1}{64}.$$

These as well as their ratios agree very closely with the "exact" numerical estimates of $E_0^T = 0.66865$ and $E_0^Q(\pi/4) = 0.70338$.[7]

Let us now also compare the asymptotic constant $E[(1+B)/(1-B)]^{\frac{1}{4}}$ in the expansion (2.5) of D_N, with the determinant of order one, namely

$$D_1 = D_1(1, B) = \frac{1+B}{\pi(B)^{\frac{1}{2}}} \arccos\left(\frac{1-B}{1+B}\right), \qquad (3.6)$$

which generally serves as a good first approximation. Along the row of the rectangular and triangular lattices $D_1(A, B)$ represents the *nearest*-neighbor correlation which equals the normalized energy $U(T)/U(0)$. For these two cases the critical values of $D_1 = 1/\sqrt{2} = 0.7071068$ and $D_1 = \frac{2}{3} = 0.6\bar{6}$ compare favorably with the values of $E_0^Q(0)$ and $E_0^T(0)$, respectively. For the diagonal correlation on the square net,

[15] J. Stephenson, J. Math. Phys. 7, 1123 (1966).

D_1 does not represent the critical energy but its value of $2/\pi = 0.636197\cdots$ is again close to the value of $E = 0.6450024\cdots$.

In the following two sections we derive from the triangular lattice a more general class of lattices S, which have three directions along which the correlation function can be expressed as a Toeplitz determinant. The transformations used will be the generalized decoration- and star-triangle transformations as given by Fisher,[5] which are applied such as to leave the "pure" triangle spins t_j, i.e., spins on the triangular sublattice, invariant. This enables us to reduce the row correlation between two of these "pure" triangle spins to the row correlation on the triangular lattice with different parameters A and B. Due to the invariance of the triangle spins the radial separation distance r between the spins t_k and t_l remains the same in all cases, whereas the first nearest-neighbor lattice spacing a_1 depends on the particular lattice. As a specific example we discuss first the case of the hexagonal- and diced-lattice row correlations.

4. THE HEXAGONAL LATTICE

The hexagonal lattice with m spins per row has two sublattices and therefore two spins per cell, a "pure" hexagonal spin h_j and a "pure" triangle spin t_j,[16] where j is an index referring to a lattice point (see Fig. 1). The correlation between two triangle spins t_k and t_l can be expressed as usual[3,4,16] as the ratio of a perturbed to an unperturbed partition sum or generating function, i.e.,

$$\langle t_k t_l \rangle = \frac{R_H^*(y_i)}{R_H(y_i)},$$

where[17]

$$R_H^*(y_i) = \sum_{\mathbf{h}} \sum_{\mathbf{t}} t_k t_l \prod_{j=1}^{M} (1 + y_1 h_j t_{j+m}) \times (1 + y_2 h_j t_{j+1})(1 + y_3 h_j t_j) \quad (4.1)$$

and $R_H(y_i)$ is given similarly except that the factor $t_k t_l$ has been omitted. M is here the total number of spins $(= M_h + M_t)$ and y_i are the weights of the bonds of the hexagonal lattice. In most cases, such as here, the generating function R is a simple multiple of the partition function Z; sometimes however, R may represent only a "pseudo" partition function (see below in Sec. 7). The star-triangle transformation, which "commutes" with the factor $t_k t_l$, is performed by summing over the ± 1 values of the hexagonal spins h_j

[16] H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience Publishers Inc., London, 1964).

[17] For convenience all the constant factors in the definition of R and R^* have been omitted.

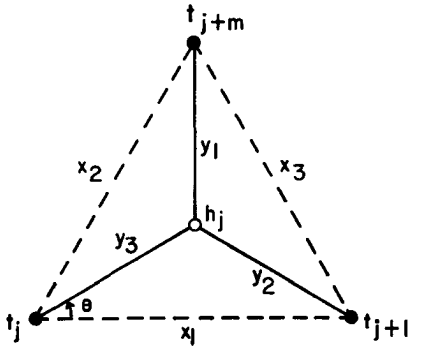


FIG. 1. The triangular and hexagonal lattices.

giving

$$R_T^*(x_i) = \left[\frac{1 + x_1x_2x_3}{2}\right]^{M_h} R_H^*(y_i) = \sum_{\mathbf{t}} t_k t_l \prod (1 + x_1 t_j t_{j+1})(1 + x_2 t_j t_{j+m}) \times (1 + x_3 t_{j+1} t_{j+m}) \quad (4.2)$$

and a similar expression for $R_T(x_i)$.

Hence

$$\langle t_k t_l \rangle = \frac{R_H^*(y_i)}{R_H(y_i)} = \frac{R_T^*(x_i)}{R_T(x_i)}, \quad (4.3)$$

where $y_i = x_i^{+*} = \text{th } K_i^{+*}$ is given by (3.2) in terms of x_i and conversely

$$x_1 = \frac{de_1 - f}{de_1 + f} \quad (+ \text{ cyclic}), \quad (4.4)$$

with

$$d = 1 + y_1y_2 + y_2y_3 + y_3y_1,$$
$$e_1 = 1 - y_1y_2 + y_2y_3 - y_3y_1 \quad (+ \text{ cyclic}),$$

and

$$f^2 = de_1e_2e_3.$$

Even though the critical points for the lattices in class S vary, at their individual critical points their parameters A all equal unity and their parameters B all take the *same* value. This follows from the definition of the critical point. Hence for these lattices all correlations which are derived from the triangular row correlation, have at their critical point the *same* asymptotic expansion

$$\omega^L(r/a_i, 0) \sim \frac{E_0^T}{(r/a_i)^{\frac{1}{4}}}\left[1 + \frac{E_i^T}{(r/a_i)^2}\right], \quad (4.5)$$

where $a_i = \gamma a_1$ is the appropriate higher-order lattice spacing. Since the factor γ will in general change from lattice to lattice, so will the asymptotic expansion in terms of (r/a_1). For the isotropic hexagonal lattice $N = r/a_2$ and $a_2 = a_1\sqrt{3}$, which when substituted in (3.3) yields at $T = T_c$,

$$E_0^H(0) = 3^{\frac{1}{8}}E_0^T = 2^{\frac{1}{4}}E,$$
$$E_1^H(0) = 3E_1^T = -\tfrac{1}{64}. \quad (4.6)$$

To test for radial symmetry we compare the ratios

$$\frac{\omega^H(r/a_1, \pi/6)}{\omega^H(r/a_2, 0)} \quad \text{and} \quad \frac{U_c/U_0}{D_1},$$

which should approximately be equal. Assuming this we have

$$\frac{E_0^H(\pi/6)}{E_0^T} \sim \frac{4.3^{-\frac{3}{2}}}{\frac{2}{3}} = \frac{2}{\sqrt{3}},$$

which shows that

$$E_0^H(\pi/6) \sim \frac{2}{\sqrt{3}} E_0^T = 0.772061\cdots, \tag{4.7}$$

which is reasonably consistent with the value of

$$E_0^H(0) = 3^{\frac{1}{2}} E_0^T = 0.767041\cdots.$$

5. THE DICED LATTICE

The diced lattice is obtained from the triangular lattice by a "double" star-triangle transformation.[5] It may be considered as having three spins h_j, s_j, and t_j per unit cell (see Fig. 2). The perturbed generating function is now given by

$$R_D^*(w_i, z_i) = \sum_{\mathbf{h}} \sum_{\mathbf{s}} \sum_{\mathbf{t}} t_k t_l \prod_{j=1}^{M} (1 + w_1 h_j t_{j+m})(1 + w_2 h_j t_{j+1}) \times (1 + w_3 h_j t_j)(1 + z_1 s_j t_{j+1})(1 + z_2 s_j t_{j+m}) \times (1 + z_3 s_j t_{j+m+1}), \tag{5.1}$$

where $w_i = u_i^{+*}$ and $z_i = v_i^{+*}$ are given by (3.2) in terms of u_i and v_i, respectively, which in turn satisfy

$$x_i = \frac{u_i + v_i}{1 + u_i v_i} \quad (i = 1, 2, 3). \tag{5.2}$$

On summing over the h_j and s_j spins, Eq. (5.1) yields

$$R_D^*(w_i, z_i) = \left[\frac{2}{1 + u_1 u_2 u_3}\right]^{M_h} \left[\frac{2}{1 + v_1 v_2 v_3}\right]^{M_s} R_{\mathrm{DT}}^*(u_i, v_i), \tag{5.3}$$

where $M = M_h + M_s + M_t$ and $R_{\mathrm{DT}}^*(u_i, v_i)$ is the perturbed generating function of the double-triangle lattice, which reduces to

$$R_{\mathrm{DT}}^*(u_i, v_i) = [(1 + u_1 v_1)(1 + u_2 v_2)(1 + u_3 v_3)]^{M_t} R_T^*(x_i). \tag{5.4}$$

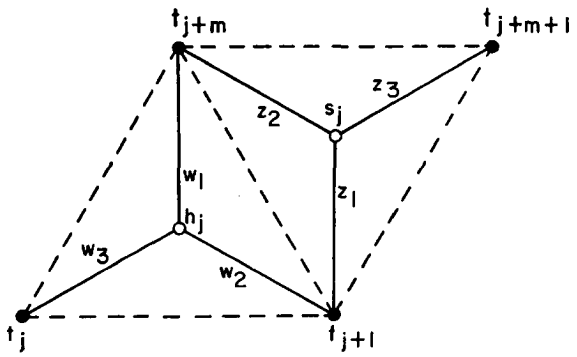


FIG. 2. The diced lattice.

The parameters A and B for the diced lattice are thus directly defined through (3.1), (3.2), and (5.2) in terms of w_i and z_i. Since in the isotropic case the lattice spacing a_1 is the same as for the hexagonal lattice, it is evident that at $T = T_c$, the asymptotic expansions (3.3) and (4.5) also hold. Unlike the partition function, the dual of the row-correlation function on the diced lattice, does *not* give the row correlation on the Kagomé lattice, but instead gives a Toeplitz-determinant representation of the so-called "misfit seam," separating regions of opposite spin on the Kagomé lattice. It was shown by Ferdinand[18] that the dual of a pair correlation on an arbitrary planar lattice is a misfit seam, and the present analysis of generating correlations which have a Toeplitz-determinant representation, can be carried over directly to this problem.[19]

6. THE GENERALIZED HEXAGONAL LATTICE

The hexagonal and diced lattices are only special cases of the class S of generalized hexagonal lattices, which are obtained from the hexagonal lattice by the generalized decoration- and star-triangle transformations.[5] The generalized decoration transformation leaves the spins s_j on the basic undecorated lattice invariant and hence it follows directly that

$$\langle s_k s_l \rangle = \frac{Z_{\mathrm{dec}}^*(K', L')}{Z_{\mathrm{dec}}(K', L')} = \frac{Z^*(K, L)}{Z(K, L)}, \tag{6.1}$$

where $Z_{\mathrm{dec}}^*(K', L')$ and $Z^*(K, L)$ are the perturbed partition functions of the decorated and the original undecorated lattice, respectively, together with their corresponding spin- and magnetic-interaction parameters. Similarly the generalized star-triangle transformation leaves the spins t_j on the triangular sublattice invariant and shows that

$$\langle t_k t_l \rangle = \frac{Z_{\mathrm{S}}^*(K', L_{\mathrm{S}}, L_{\mathrm{T}})}{Z_{\mathrm{S}}(K', L_{\mathrm{S}}, L_{\mathrm{T}})} = \frac{Z_{\mathrm{T}}^*(K, L_{\mathrm{T}})}{Z_{\mathrm{T}}(K, L_{\mathrm{T}})}, \tag{6.2}$$

where K and K' are the spin-interaction parameters, L_{S} and L_{T} are the magnetic parameters for the star and triangle vertices, respectively, and $L_{\mathrm{S}} \equiv 0$ by assumption.[5] A combination of the two generalized transformations makes it possible to derive from the basic hexagonal lattice in zero field, i.e., $L' = L_{\mathrm{T}} = 0$,

[18] A. E. Ferdinand, Ph.D. thesis, University of London (1967).
[19] M. E. Fisher and A. E. Ferdinand, Phys. Rev. Letters **19**, 169 (1967).

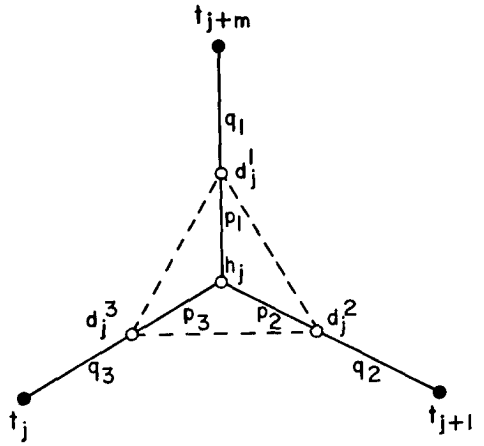


FIG. 3. The simply decorated hexagonal lattice.

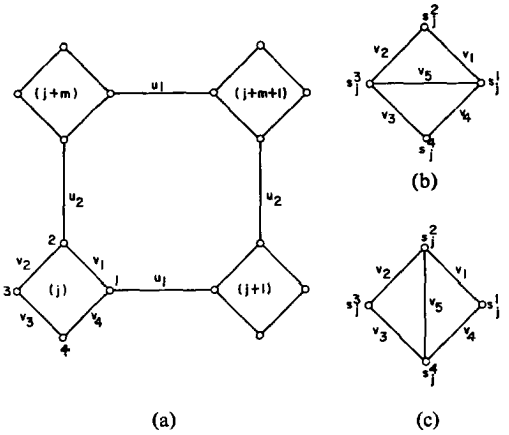


FIG. 4. (a) The simple bathroom-tile lattice. (b) and (c) Two simply decorated bathroom-tile lattices.

a class of lattices S, for which the row correlation between two basic triangle spins admits a Toeplitz-determinant representation

As an example of the combined transformations we consider the hexagonal lattice in zero field, in which each bond is decorated with a single spin d_j^i (see Fig. 3) and has associated weights $p_i = \text{th } L_i^{+*}$ and $q_i = \text{th } M_i^{+*}$, which satisfy

$$y_i = p_i q_i \quad (i = 1, 2, 3).$$

The perturbed partition sum becomes in this case

$$R_{DH}^*(p_i, q_i) = \sum_{\mathbf{t}} \sum_{\mathbf{h}} \sum_{\mathbf{d}} t_k t_l \prod_{j=1}^{M} (1 + p_1 h_j d_j^1)(1 + p_2 h_j d_j^2) \times (1 + p_3 h_j d_j^3)(1 + q_1 d_j^1 t_{j+m})(1 + q_2 d_j^2 t_{j+1}) \times (1 + q_3 d_j^3 t_j).$$

Summation over the decorated spins d_j^i in the perturbed generating function $R_{DH}^*(p_i, q_i)$ reduces it to $2^{M_d} R_H^*(y_i)$, whereas summation over the hexagonal spins h_j gives rise to the perturbed generating function of a "starred" lattice. Similar expressions obviously hold for the unperturbed partition sums and hence the invariance of the row correlation follows. We note that in this expression the pure triangle spins cannot be removed by summation as this does not leave the $t_k t_l$ spin-factor invariant, and thus we cannot deduce that the row correlation on the Kagomé lattice admits a Toeplitz-determinant representation. In the next section we examine several generalized rectangular lattices which have two or more directions along which the correlation has the required structure.

7. THE BATHROOM-TILE LATTICE

The bathroom-tile lattice or modified Yamamoto lattice is an example of a generalized rectangular lattice[20,21] with external weights $u_i = \text{th } K_i'$, in which each lattice point is replaced by a decorating cluster consisting of four spins and four or five weights $v_i = \text{th } K_i''$ (see Fig. 4). Both lattices are a special case of the modified Utiyama lattice,[22] in which the ladder has two or three rungs, respectively. When $v_5 = 0$ and $u_1 = u_2 = 1$, it reduces to the ordinary Yamamoto lattice,[23] and if in addition $v_3 = 1$ or $v_3 = 0$, then it collapses further to the triangular or hexagonal lattice, respectively. We will show that only when $u_1 = 1$ can the correlation $\langle s_k^3 s_l^3 \rangle$, with cell k and cell l in the same row or column (see Fig. 4), be expressed as a Toeplitz determinant $D_N(A, B)$, where A and B are functions of u_i and v_i. The appropriate Toeplitz determinant reduces to the known results for the triangular and rectangular lattices and also verifies the results of Sec. 4.

Let us consider the perturbed partition sum for the "3" direction of the decorated bathroom-tile lattice of Fig. 4(b),

$$R^*(u_i, v_i) = \sum_{\mathbf{s}} s_k^3 s_l^3 \prod_{j=1}^{M} (1 + u_1 s_j^1 s_{j+1}^3)(1 + u_2 s_j^2 s_{j+m}^4) E, \tag{7.1}$$

where

$$E = (1 + v_1 s_j^1 s_j^2)(1 + v_2 s_j^2 s_j^3)(1 + v_3 s_j^3 s_j^4) \times (1 + v_4 s_j^4 s_j^1)(1 + v_5 s_j^1 s_j^3), \tag{7.2}$$

m is the number of columns, M $(= mn)$ is the total number of clusters, and j is an index referring to a cluster. This may be reduced, when $u = 1$, to a multiple of the perturbed partition sum for the

[20] C. A. Hurst, J. Chem. Phys. **38**, 2558 (1963).
[21] See Ref. 16, Sec. 5.3.
[22] T. Utiyama, Progr. Theoret. Phys. (Kyoto) **6**, 907 (1951).
[23] T. Yamamoto, Progr. Theoret. Phys. (Kyoto) **6**, 533 (1951).

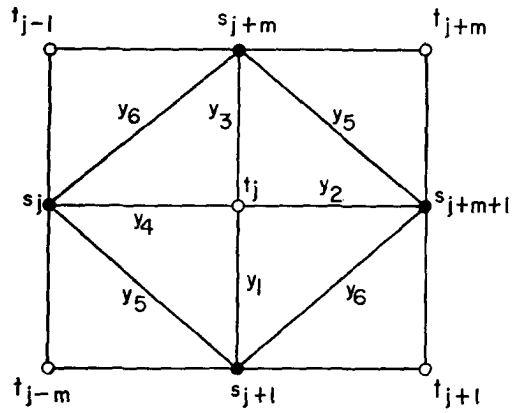


FIG. 5. The decorated Yamamoto or "crossed" lattice.

decorated Yamamoto lattice (see Fig. 5),

$$R^*(y_i) = \sum_s \sum_t s_k s_l \prod (1 + y_1 s_{j+1} t_j)(1 + y_2 s_{j+m+1} t_j) \times (1 + y_3 s_{j+m} t_j)(1 + y_4 s_j t_j)(1 + y_5 s_j s_{j+1}) \times (1 + y_6 s_j s_{j+m}), \quad (7.3)$$

which has two spins per cell, $s_j = s_j^3$ and $t_j = s_{j+m}^4$, two possible extra factors of $(1 + y_5 s_j s_{j+1})(1 + y_6 s_j s_{j+m})$ and where the weights y_i are functions of u_i and v_i. The decorated or "crossed" Yamamoto lattice of Fig. 5, has in general both first and second nearest-neighbor interactions, and is a generalization of the lattice considered by Vaks, Larkin, and Ovchinnikov.[24] The expression in (7.3) can now be reduced in the same way as for the Yamamoto lattice[4] to a multiple of the standard perturbed generating function

$$R^*(x_i) = \sum_s s_k s_l \prod (1 + x_1 s_j s_{j+1})(1 + x_3 s_j s_{j+m}) \times (1 + x_2 s_j s_{j+m+1} + x_4 s_{j+1} s_{j+m} - x_2 x_4 s_j s_{j+1} s_{j+m} s_{j+m+1}), \quad (7.4)$$

since the extra factors do not hinder the elimination, and the transformation leaves the spin factor $s_k s_l$ invariant. Because the reduction of (7.1) to (7.4) also holds for the unperturbed partition sum, we have

$$\langle s_k s_l \rangle = \frac{R^*(u_i, v_i)}{R(u_i, v_i)} = \frac{R^*(y_i)}{R(y_i)} = \frac{R^*(x_i)}{R(x_i)}, \quad (7.5)$$

where

$l = k + N \quad (N < m)$ or $l = k + Nm \quad (N < n)$.

It was shown by Green[4] that this standard ratio can be expressed, using the Pfaffian approach, as the product of the eigenvalues of an integral equation. If this homogeneous integral equation is written in the form

$$\lambda X(e^{i\phi}) = \frac{1}{2\pi} \int_0^{2\pi} F(\theta) \sum_{r=0}^{N-1} e^{i(\phi-\theta)r} X(e^{i\theta})\, d\theta, \quad (7.6)$$

it can be identified with the eigenvalue equation of the Toeplitz matrix C generated by the function $F(\theta)$,[6] which is given by (1.1). This can easily be seen on writing the eigenvalue equation $\lambda \mathbf{x} = C\mathbf{x}$ as

$$\lambda x_r = \sum_{s=0}^{N-1} c_{r-s} x_s = \frac{1}{2\pi} \int_0^{2\pi} F(\theta) e^{-ir\theta} \left(\sum_{s=0}^{N-1} x_s e^{is\theta} \right) d\theta \quad (0 \leq r < N), \quad (7.7)$$

multiplying it by $e^{ir\phi}$, and summing over r. Hence the analysis of Sec. 1 is applicable on substituting the appropriate values for A and B, which depend on the original lattice and the relative orientation of the two spins. It should be noted here that the partition sum $R(x_i)$ with $x_4 \neq 0$, is not a proper partition function, but merely a "pseudo" partition function with "pseudo" crossed-bonds, which does not correspond to an actual physically realizable planar lattice, but to an auxiliary "eight-terminal" lattice.[4] Its nonphysical character is further exemplified by its anomalous behavior at the isotropic critical point.

In terms of the weights $x_i = \text{th}\, K_i$ of this eight-terminal lattice, the parameters A and B are given by

$$A = \frac{a - G}{b - c}, \quad B = \frac{a - G}{b + c}, \quad (7.8)$$

where

$$a = S_1 C_2 C_3 C_4 + S_3 C_1 (S_2 + S_4), \quad b = C_1 (1 - S_2 S_4), \quad c = 1 + S_2 S_4, \quad G^2 = a^2 - b^2 + c^2, \quad (7.9)$$

and

$$S_i = \text{sh}\, 2K_i, \quad C_i = \text{ch}\, 2K_i.$$

Consequently

$$\frac{B}{A} = \frac{x_1^2 - S_2 S_4}{1 - x_1^2 S_2 S_4}. \quad (7.10)$$

When $|S_2 S_4| < 1$, the parameters are positive and satisfy at all temperatures the inequalities

$$0 < AB < 1, \quad 0 < B/A < 1,$$

while $A \gtrless 1$ according to whether $T \gtrless T_c$.

Moreover the critical point is also determined by the equality

$$a = b. \quad (7.10a)$$

In order to find the dependence of the weights x_i upon the original weights u_i and v_i, let us analyze the two stages of the transformation

$$R^*(u_i, v_i) \to R^*(y_i) \to R^*(x_i),$$

which will indicate how the y_i are obtained in terms

[24] V. G. Vaks, A. I. Larkin, and Yu. N. Ovchinnikov, Zh. Eksp. Teor. Fiz. 49, 1180 (1965) [Sov. Phys.—JETP 22, 820 (1966)].

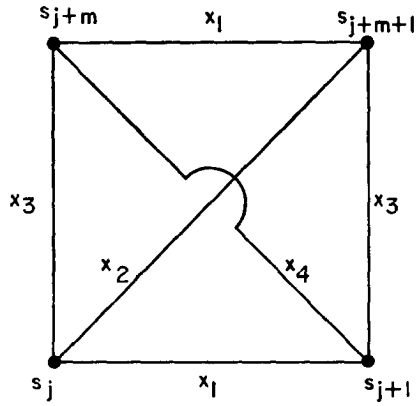


FIG. 6. The auxiliary eight-terminal lattice with "pseudo" crossed-bonds.

of u_i and v_i, and what effect the extra factors in (7.3) have upon the final weights x_i.

The first stage of the transformation is the reduction of (7.1) to (7.3). It is shown in Appendix A that this reduction is possible only when $u_i = 1$, and that in this case the transformation simply removes the s_j^2 spin by means of the star-triangle transformation and identifies the spins s_j^1 and s_{j+1}^3.

The second stage of the transformation reduces (7.3) to (7.4) and is executed using the so called "crossed-square" transformation discovered by Green.[4] In the "crossed-square" transformation a "cross" consisting of a central spin coupled to *four* neighboring spins can be transformed into a rectangle of four spins coupled to each other with either a simple bond or a "pseudo" crossed-bond (see Figs. 5, 6). A somewhat similar transformation was used by Fisher in his paper on hard-square lattice gases.[25] This purely algebraic transformation is not restricted to planar nets and includes the star-triangle transformation as a special case. The corresponding generalized "crossed-square" transformation is discussed in the next section.

In (7.3) the summation over the t_j spins leaves the spin factor $s_k s_l$ and the "bond" factors involving y_5 and y_6 invariant, and yields a kernel of the form

$$(1 + y_5 s_j s_{j+1})(1 + y_6 s_j s_{j+m})\psi,$$

where

$$\psi = (1 + y_1 y_3 s_{j+1} s_{j+m})(1 + y_2 y_4 s_j s_{j+m+1}) + (y_1 s_{j+1} + y_3 s_{j+m})(y_2 s_{j+m+1} + y_4 s_j). \quad (7.11)$$

Following Green,[4] the partition function ψ of a "cross" can be factorized in the form of (8.2), since the consistency condition (8.4) is identically satisfied. It is clear that the factorization is unaffected by the additional factors involving y_5 and y_6.

[25] M. E. Fisher, J. Math. Phys. **4**, 278 (1963).

Rearranging factors under the product Π_j shows further that

$$(1 + z_1 s_j s_{j+1})(1 + z_3 s_j s_{j+1})(1 + y_5 s_j s_{j+1}) = c_1(1 + x_1 s_j s_{j+1}),$$

$$(1 + z_2 s_j s_{j+m})(1 + z_4 s_j s_{j+m})(1 + y_6 s_j s_{j+m}) = c_2(1 + x_3 s_j s_{j+m}),$$

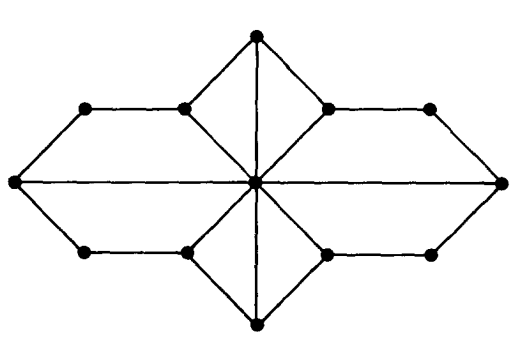
which yields

$$x_1 = \frac{(z_1 + z_3) + y_5(1 + z_1 z_3)}{(1 + z_1 z_3) + y_5(z_1 + z_3)},$$

$$x_3 = \frac{(z_2 + z_4) + y_6(1 + z_2 z_4)}{(1 + z_2 z_4) + y_6(z_2 + z_4)}. \quad (7.12)$$

Thus the only effect the additional factors in (7.3) have is to change x_1 and x_3 according to (7.12). Application of the generalized star-triangle and decoration transformations of Sec. 6, to the "crossed" or decorated Yamamoto lattice of Fig. 5, yields a more general class of lattices, which have two or more directions along which the correlation has the desired structure. For example see Fig. 7.

When $y_5 = y_6 = 0$ the correlation (7.5) becomes a diagonal correlation on the simple Yamamoto lattice, which reduces to the triangular row correlation when $y_4 = 1$, or to the rectangular diagonal correlation when $y_1 = y_3$ and $y_2 = y_4$. We note that basically both transformations reduce a Toeplitz determinant $D_N[A(y_i), B(y_i)]$ to a Toeplitz determinant $D_N[A(x_i)]$,



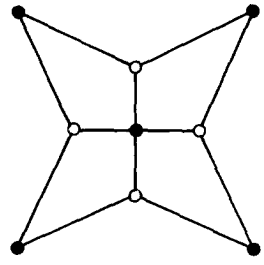


FIG. 7. Two lattices derived from the decorated Yamamoto lattice.

i.e., they transform the parameter B to zero. And these transformations are in fact the only ones possible within the Ising-model regime, as we will now show. Assuming that $|S_2S_4| < 1$, we see from (7.9) that $b > c > 0$ and hence by (7.8) the parameter B will vanish if and only if $a = G$, i.e., when $b = c$ or when

$$x_1^2 = S_2S_4. \tag{7.13}$$

Moreover when $b = c$ the parameter $A = [(b + c)/(a + G)]$ reduces to

$$A = \frac{b}{a} = \frac{C_1(1 - S_2S_4)}{S_1C_2C_3C_4 + S_3C_1(S_2 + S_4)}. \tag{7.14}$$

Let us now consider the simple Yamamoto lattice and suppose that $y_4 = 1$. Then this implies, as expected, that $x_4 = 0$ so that we are on the triangular lattice. If in addition we put $y_3 = 0$, then $x_1 = 0$ and we obtain the diagonal correlation on the rectangular lattice with (7.3) identically satisfied, i.e.,

$$B = 0, \quad A = \frac{1}{S_3S_2} = \frac{1}{S_1'S_2'},$$

where the prime refers to the original system with weights y_i. When $y_1 = y_3$ and $y_2 = y_4$ inspection of the factorization (8.2) shows that $x_1 = x_3$, ($x_2 \neq x_4 \neq 0$) and that (7.13) is again satisfied. Furthermore it can also be shown that

$$A = \frac{1 - S_2S_4}{S_1(C_2C_4 + S_2S_4)} = \frac{1}{S_1'S_2'}.$$

Since the transformation from (7.1) to (7.3) sums over all the t_j spins, we may regard it as effectively removing the t_j spins and replacing it by a set of new bonds between its four nearest neighbors. In this way it is evident that we may consider the *diagonal* correlation on the rectangular lattice, when $y_1 = y_3, y_2 = y_4$, and $y_5 = y_6 = 0$, as a *row* correlation on the associated tilted eight-terminal lattice with "pseudo" crossed-bonds and $B = 0$ (see Fig. 6).

We conclude this section with an examination of the isotropic critical point behavior of the "axial" correlations of the decorated bathroom-tile and Yamamoto lattices with $u_1 = 1$.

Since the simple Yamamoto lattice with equal interactions reduces to the square net, it is clear that the diagonal correlation derived for it reduces at the critical point to (3.4) for large spin separation. The critical point of the bathroom-tile lattice is determined by the equations[26]

$$(g_0 - u_1g_7) \pm u_2(g_5 + u_1g_6) = 0, \tag{7.15a}$$

$$(g_0 + u_1g_7) \pm u_2(g_5 - u_1g_6) = 0, \tag{7.15b}$$

where the g_i are given appropriately by (A2) or (A7). For the isotropic decorated bathroom-tile lattice of Fig. 4(b), with $u_1 = 1$, $v_1 = v_2 = v_3 = v_4 = u_2 = v$ and $v_5 = 0$ or v, this reduces to a single equation

$$(1 - v^2)^2(1 - v_5) = 4v^3(1 + v_5), \tag{7.15c}$$

which has in both cases only one positive zero in the interval (0, 1). Their respective numerical values are

$$v = v_c = 0.5136198300\cdots \qquad (v_5 = 0)$$

and (7.16)

$$v = v_c = \sqrt{2} - 1 = 0.414213563\cdots \quad (v_5 = v),$$

the latter of which is the same as for the square net. The row and column correlations $\langle s_ks_{k+N}\rangle$ and $\langle s_ks_{k+Nm}\rangle$ of this bathroom-tile lattice may be considered as row correlations on the anisotropic decorated Yamamoto lattice of Fig. 5. To obtain weights y_i for the row correlation the critical values of (7.16) have to be substituted in (4.4), giving with the aid of (A15)

$$y_1 = y_4 = w, \quad y_5 = \frac{w + v_5}{1 + wv_5},$$

$$y_2 = y_3 = v, \quad y_6 = 0, \tag{7.17}$$

where

$$w = \frac{d - (de)^{\frac{1}{2}}}{d + (de)^{\frac{1}{2}}} = \frac{(1 + 3v^2) - [(1 + 3v^2)(1 - v^2)]^{\frac{1}{2}}}{(1 + 3v^2) + [(1 + 3v^2)(1 - v^2)]^{\frac{1}{2}}}$$

$$= \begin{cases} 0.2187251054\cdots & (v_5 = 0) \\ 0.1497297975\cdots & (v_5 = v). \end{cases}$$

For the column correlation we simply interchange the weights y_1 with y_3 and y_5 with y_6, whilst the critical point remains the same. From the weights y_i the values of x_i and B can be found on using (8.2), (7.12), and (7.10). Even though the parameter B does not depend explicitly on y_6, it does so implicitly through the critical-point equation, which is affected by the value of y_6. With the aid of (2.8) and the geometry of the lattice, we can compute the required coefficients

$$E_0^L = E\left(\gamma \cdot \frac{1 + B}{1 - B}\right)^{\frac{1}{4}}$$

and

$$E_1^L = -\frac{\gamma^2}{8}[\tfrac{1}{8} - B(1 - B)^{-2}],$$

for the two correlations $\langle s_ks_l\rangle$ with $l = k + N, k + Nm$. Their numerical values are given in Table I.

[26] See Ref. 16, Eq. (5.70).

TABLE I. Numerical values for two correlations $\langle s_k s_l \rangle$.

	l	γ	B	$[(1+B)/(1-B)]^{\frac{1}{4}}$	E_0^L	E_1^L
$v_5 = 0$	$k+N$	$\sqrt{2}$	0.19142348	1.10175811	0.7749548	−0.04194694
$v_5 = v$	$k+N$	1	0.41421356	1.24650470	0.8039986	−0.13526335
$v_5 = 0$	$k+Nm$	$1+\sqrt{2}$	0	1	0.8039986	−0.09106917
$v_5 = v$	$k+Nm$	$1+\sqrt{3}$	0	1	0.8292463	−0.11662659

As a last example we consider the row and column correlation of the isotropic decorated Yamamoto lattice of Fig. 5, with $y_i = y$ $(1 \leq i \leq 6)$. Even though the lattice is metrically nonplanar, we can still carry through the analysis developed in this section. The isotropy implies that the spin states of (8.3) reduce to

$$\begin{aligned} p_1^2 &= (1+y^2)^2 + 4y^2, \\ p_2^2 &= p_3^2 = p_4^2 = (1-y^2)^2, \\ p_5^2 &= p_6^2 = p_7^2 = p_8^2 = (1-y^4), \end{aligned} \tag{7.18}$$

which shows that

$$z_1 = z_2 = z_3 = z_4$$

and

$$x_1 = x_3 = \frac{(p_1 - p_2) + y(p_1 + p_2)}{(p_1 + p_2) + y(p_1 - p_2)},$$

$$x_2 = x_4 = \frac{p_2(p_1 - p_2)}{p_2(p_1 + p_2) + 2p_5^2} \quad (p_i > 0). \tag{7.19}$$

To find its critical point we cannot use Eq. (7.15) since the corresponding bathroom-tile lattice has *two* vertical bonds connecting decorating clusters. Instead we employ (7.10a), which yields the following critical-point equation:

$$S_1 = \frac{1 - S_2}{1 + S_2}. \tag{7.20}$$

Comparison with (8.5) shows that this may be written as

$$\frac{2x_1}{1 - x_1^2} = \frac{p_3 p_4}{p_1 p_2} = \frac{p_2}{p_1}, \tag{7.21}$$

which reduces with the aid of (7.18) and (7.19) to

$$y^4 + 4y = 1. \tag{7.22}$$

This has exactly one positive zero in the interval (0, 1),

$$y = y_c = 0.249038377 \cdots.$$

Solving (7.21) for x_1, we obtain

$$x_1 = \frac{1}{p_2}\left[(p_1^2 + p_2^2)^{\frac{1}{2}} - p_1\right], \tag{7.23}$$

which shows, as required, that $0 < x_1 < 1$. The critical parameter B can now easily be obtained on substituting y_c in (7.18) and using (7.10), (7.20) and (7.23). Its numerical value is found to be

$$B = 0.1107310486 \cdots.$$

At the critical point the fact that the generating function $R(x_i)$ is only a "pseudo" partition function is exemplified in the symmetric case by the result that $B > 0$ if and only if $x_c^2 < \frac{1}{3}$, which need not be so in general. Hence the weights x_i $(1 \leq i \leq 4)$ can never all be equal.

8. THE GENERALIZED CROSSED-SQUARE TRANSFORMATION

In the generalized "crossed-square" transformation the central spin t_j of the "cross" in Fig. 5 is replaced by an arbitrary decorating cluster (or mechanical system), with energy states $E(\mathbf{i})$, depending on the four neighboring spins s_j, s_{j+1}, s_{j+m}, and s_{j+m+1}. The contribution to the total partition function from a "cross" is

$$\psi = \psi(s_j, s_{j+1}, s_{j+m}, s_{j+m+1}) = \sum_{\mathbf{i}} e^{-\beta E(\mathbf{i})} \qquad \left(\beta = \frac{1}{kT}\right), \tag{8.1}$$

which may be written in the form

$$\begin{aligned} \psi = b(1 + z_1 s_j s_{j+1})(1 + z_2 s_{j+1} s_{j+m+1}) \\ \times (1 + z_3 s_{j+m} s_{j+m+1})(1 + z_4 s_j s_{j+m}) \\ \times (1 + x_2 s_j s_{j+m+1} + x_4 s_{j+1} s_{j+m} \\ - x_2 x_4 s_j s_{j+1} s_{j+m} s_{j+m+1}), \end{aligned} \tag{8.2}$$

under the assumptions that the central decorating cluster is invariant under total spin inversion, i.e.,

$$\begin{aligned} \psi(+,+,+,+) &= \psi(-,-,-,-) = p_1^2, \\ \psi(+,-,-,+) &= \psi(-,+,+,-) = p_2^2, \\ \psi(-,+,-,+) &= \psi(+,-,+,-) = p_3^2, \\ \psi(-,-,+,+) &= \psi(+,+,-,-) = p_4^2, \\ \psi(+,+,+,-) &= \psi(-,-,-,+) = p_5^2, \\ \psi(-,+,+,+) &= \psi(+,-,-,-) = p_6^2, \\ \psi(+,-,+,+) &= \psi(-,+,-,-) = p_7^2, \\ \psi(+,+,-,+) &= \psi(-,-,+,-) = p_8^2, \end{aligned} \tag{8.3}$$

and satisfies the consistency condition

$$p_1^2 p_2^2 + p_3^2 p_4^2 = p_5^2 p_6^2 + p_7^2 p_8^2. \tag{8.4}$$

Otherwise the 16 equations have no unique solution for the seven independent parameters. Solving the equations on equating spin states, we find first of all

that

$$\frac{p_1p_2}{p_3p_4} = \frac{1 + x_2 + x_4 - x_2x_4}{1 - x_2 - x_4 - x_2x_4},$$
$$\frac{p_5p_6}{p_7p_8} = \frac{1 - x_2 + x_4 + x_2x_4}{1 + x_2 - x_4 + x_2x_4},$$

and

$$\frac{p_1p_2p_3p_4}{p_5p_6p_7p_8} = \frac{1 - S_2S_4}{1 + S_2S_4}, \tag{8.5}$$

which then yields[4]

$$x_2 = \frac{p_1p_2 + p_7p_8 - p_3p_4 - p_5p_6}{p_1p_2 + p_7p_8 + p_3p_4 + p_5p_6},$$
$$x_4 = \frac{p_1p_2 + p_5p_6 - p_3p_4 - p_7p_8}{p_1p_2 + p_5p_6 + p_3p_4 + p_7p_8}.$$

Furthermore one easily obtains[27]

$$z_i = \frac{\alpha_i - q^4}{\alpha_i + q^4} \quad (1 \leq i \leq 4) \tag{8.6}$$

and

$$b^{-1} = p_1^{-2}(1 + z_1)(1 + z_2)(1 + z_3)(1 + z_4) \times (1 + x_2 + x_4 - x_2x_4),$$

where

$$\alpha_1 = p_1p_4p_5p_8, \quad \alpha_3 = p_1p_4p_6p_7,$$
$$\alpha_2 = p_1p_3p_6p_8, \quad \alpha_4 = p_1p_3p_5p_7,$$

and q is a positive root of

$$q^8 = \prod_{i=1}^{8} p_i.$$

Substituting the expressions for z_i in (7.12) shows that

$$x_1 = \frac{(p_1p_4 - p_2p_3) + y_5(p_1p_4 + p_2p_3)}{(p_1p_4 + p_2p_3) + y_5(p_1p_4 - p_2p_3)},$$
$$x_3 = \frac{(p_1p_3 - p_2p_4) + y_6(p_1p_3 + p_2p_4)}{(p_1p_3 + p_2p_4) + y_6(p_1p_3 - p_2p_4)}. \tag{8.7}$$

The perturbed partition function Z_c^* for the "crossed" lattice in a magnetic field H, can be expressed as

$$Z_C^*(y_j, L_C, L_R) = \sum_{\mathbf{s}} \sum_{\mathbf{i}} s_k s_l \prod e^{-\beta \mathcal{H}(\mathbf{s}) - \beta E(\mathbf{i})}, \tag{8.8}$$

where $\mathcal{H}(\mathbf{s}) = \mathcal{H}(\mathbf{s}, K_j', L_R)$ is the usual Hamiltonian with interaction parameters $K_j' = \beta J_j'$, $y_j = \text{th } K_j'$ and L_C and L_R are the magnetic parameters for the "cross" and "rectangle" vertices, respectively. As before the assumed spin invariance (8.3) implies that $L_C = 0$. The summation over the energy levels of the decorating cluster may again be interchanged with the spin factor $s_k s_l$ so that

$$Z_C^*(y_j, L_C, L_R) = f^{N_c} R^*(x_j, L_R), \tag{8.9}$$

where R^* is the partition sum of the "eight-terminal" lattice with weights $x_j = \text{th } K_j$, $f = 2bc_1c_2$ and N_c is the number of "cross" vertices. Since the same reduction clearly also holds for the unperturbed partition function it follows that

$$\langle s_k s_l \rangle = \frac{Z_C^*(y_j, L_C, L_R)}{Z_C(y_j, L_C, L_R)} = \frac{R^*(x_j, L_R)}{R(x_j, L_R)}, \tag{8.10}$$

which reduces to Green's standard ratio when $L_R = 0$. Combining the generalized "crossed-square" transformation with the transformation of Sec. 6, yields a large class of lattices Y (which contains S), which have two or more directions along which the correlation has the required Toeplitz structure, and for which the derived asymptotic expansion in powers of (r/a_1) holds.

9. CONCLUSIONS

In this paper we have seen how a large class of correlations can be derived from the basic rectangular and triangular lattice correlations, on making suitable transformations which keep certain spins in a basic unit cell variant. These transformations do not change the Toeplitz character of the correlations, but merely change the two basic parameters A and B which characterize the lattice and the orientation of the spins. With the knowledge of this fact, all the known theory concerning the asymptotic expansions of Toeplitz determinants can then be applied. In particular it was shown on the basis of this theory, and a plausible conjecture, that at the critical point there is strong evidence of radial symmetry of the correlation as the spin separation becomes infinite. We have not endeavored to find the most general correlation which can still be represented by a *single* Toeplitz determinant, as this appears rather intractable. In the light of present developments, however, it seems that this question should be considered in the general framework of representations of correlations by block Toeplitz determinants and linear combinations of Toeplitz determinants, rather than single Toeplitz determinants.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the helpful advice given by Professor M. E. Fisher.

One of us (R. E. H.) gratefully acknowledges the support of the National Science Foundation and of the Advanced Research Projects Agency through the Materials Science Center at Cornell University.

APPENDIX: REDUCTION OF THE PERTURBED PARTITION SUM

In order to reduce the perturbed partition sum (7.1) to the form of (7.3), we have to perform the following four steps:

(i) First of all the product E in (7.2) is multiplied

[27] In Ref. 4, Eq. (80), z_1 and z_3 should be interchanged with z_2 and z_4, respectively, and in Eq. (17) the second term should read $g_1g_3\beta^2$.

out and expressed in the form

$$E = g_0 + g_1 s_j^1 s_j^2 + g_2 s_j^2 s_j^3 + g_3 s_j^3 s_j^4 + g_4 s_j^4 s_j^1 + g_5 s_j^2 s_j^4 + g_6 s_j^1 s_j^2 s_j^3 s_j^4 + g_7 s_j^1 s_j^3. \quad \text{(A1)}$$

The coefficients g_i are given by

$$g_i = a_i + v_5 b_i \quad (0 \leq i \leq 7), \quad \text{(A2)}$$

where

$$\begin{aligned} a_1 &= b_2 = v_1 + v_2 v_3 v_4, & a_5 &= b_6 = v_1 v_4 + v_2 v_3, \\ a_2 &= b_1 = v_2 + v_3 v_4 v_1, & a_6 &= b_5 = v_1 v_3 + v_2 v_4, \\ a_3 &= b_4 = v_3 + v_4 v_1 v_2, & a_7 &= b_0 = v_1 v_2 + v_3 v_4, \\ a_4 &= b_3 = v_4 + v_1 v_2 v_3, & a_0 &= b_7 = 1 + v_1 v_2 v_3 v_4. \end{aligned} \quad \text{(A3)}$$

It is easily verified that the constants a_i satisfy the three identities

$$a_1 a_2 + a_3 a_4 = a_0 a_7 + a_5 a_6, \quad \text{(A4a)}$$

$$a_1 a_3 + a_4 a_2 = a_0 a_6 + a_7 a_5, \quad \text{(A4b)}$$

$$a_1 a_4 + a_2 a_3 = a_0 a_5 + a_6 a_7, \quad \text{(A4c)}$$

and hence the two consistency conditions

$$g_1 g_3 + g_4 g_2 = g_0 g_6 + g_7 g_5, \quad \text{(A5a)}$$

$$g_1 g_4 + g_2 g_3 = g_0 g_5 + g_6 g_7, \quad \text{(A5b)}$$

which are necessary and sufficient to factorize an expression of the form (A1) in the form of (7.2), are also identically satisfied. There is no consistency condition corresponding to (A4a) unless one of the weights v_i ($1 \neq 5$) equals unity. We note here that if in (7.2) we replace E by

$$E' = (1 + v_1 s_j^1 s_j^2)(1 + v_2 s_j^2 s_j^3)(1 + v_3 s_j^3 s_j^4) \times (1 + v_4 s_j^4 s_j^1)(1 + v_5 s_j^2 s_j^4), \quad \text{(A6)}$$

then the coefficients g_i become

$$\begin{aligned} g_i &= a_i + v_5 a_{5-i} \quad (0 \leq i \leq 7), \\ a_{-i} &= a_{8-i}, \end{aligned} \quad \text{(A7)}$$

and (A5b) has to be replaced by

$$g_1 g_2 + g_3 g_4 = g_0 g_7 + g_5 g_6. \quad \text{(A8)}$$

(ii) On performing the summations over the s_j^2 and s_j^1 spins, (7.1) reduces to

$$R^*(u_i, v_i) = 2^{2M} \sum_{\mathbf{s}} \sum_{\mathbf{t}} s_k s_l \prod (g_0 + g_1 u_1 u_2 s_{j+1} t_j + g_2 u_2 s_j t_j + g_3 s_j t_{j-m} + g_4 u_1 s_{j+1} t_{j-m} + g_5 u_2 t_{j-m} t_j + g_6 u_1 u_2 s_j s_{j+m} t_{j-m} t_j + g_7 u_1 s_j s_{j+1}), \quad \text{(A9)}$$

where

$$s_j = s_j^3 \quad \text{and} \quad t_j = s_{j+m}^4.$$

With its kernel in this form the partition sum is easily seen to be invariant under the transformations

$$u_1 \to u_1/\alpha, \quad g_k \to \alpha g_k \quad (k = 1, 4, 6, 7), \quad \text{(A10a)}$$

$$u_2 \to u_2/\beta, \quad g_k \to \beta g_k \quad (k = 1, 2, 5, 6). \quad \text{(A10b)}$$

Moreover if one replaces the factor $(1 + u_2 s_j^2 s_{j+m}^4)$ by $(1 + u_2 s_{j-m}^2 s_j^4)$ under the product in (7.1) and sums over the spin s_j^4, then it follows that the perturbed partition sum is also invariant under the transformation

$$u_2 \to u_2/\beta, \quad g_k \to \beta g_k, \quad (k = 3, 4, 5, 6). \quad \text{(A10c)}$$

It is important to note here that due to the presence of the spin factor $s_k s_l$ one *cannot* sum over the spin s_j^3 to obtain further invariances of R^*, such as

$$u_1 \to u_1/\alpha, \quad g_k \to \alpha g_k \quad (k = 2, 3, 6, 7),$$

which does leave the unperturbed partition sum invariant. Applying the transformation (A10c) the new kernel becomes

$$g_0' + g_1' s_{j+1} t_j + g_2' s_j t_j + g_3' s_j t_{j-m} + g_4' s_{j+1} t_{j-m} + g_5' t_{j-m} t_j + g_6' s_j s_{j+1} t_{j-m} t_j + g_7' s_j s_{j+1}, \quad \text{(A11a)}$$

where

$$\begin{aligned} g_1' &= g_1 u_1 u_2/\beta, & g_3' &= \beta g_3, & g_5' &= g_5 u_2, & g_7' &= g_7 u_1 \\ g_2' &= g_2 u_2/\beta, & g_4' &= \beta g_4 u_1, & g_6' &= g_6 u_1 u_2, & g_0' &= g_0, \end{aligned} \quad \text{(A11b)}$$

and β is arbitrary.

(iii) In order for this kernel to be factorizable in the form of E or E', with $s_j^1 \to s_{j+1}$, $s_j^2 \to t_j$, $s_j^3 \to s_j$, $s_j^4 \to t_{j-m}$ and $v_i \to y_i$, it is necessary and sufficient that the coefficients g_i' satisfy the first two or the last two of the three consistency conditions

$$g_1' g_4' + g_2' g_3' = g_0' g_5' + g_6' g_7' \quad (E), \quad \text{(A12a)}$$

$$g_1' g_3' + g_4' g_2' = g_0' g_6' + g_7' g_5' \quad (E, E'), \quad \text{(A12b)}$$

$$g_1' g_2' + g_3' g_4' = g_0' g_7' + g_5' g_6' \quad (E'), \quad \text{(A12c)}$$

respectively. The factorization in the form E' is clearly not desirable since it gives rise to a factor of $(1 + y_5 t_{j-m} t_j)$, which represents a bond crossing the path of the correlation. Substituting (A11b) in (A12) we see with aid of (A5) that (A12b) is satisfied identically and that (A12a)–(A12c) reduce to

$$g_1 g_4 u_1^2 + g_2 g_3 = g_0 g_5 + g_6 g_7 u_1^2, \quad \text{(A13)}$$

and

$$g_1 g_2 u_2^2/\beta^2 + g_3 g_4 \beta^2 = g_0 g_7 + g_5 g_6 u_2^2. \quad \text{(A14)}$$

Hence on choosing the parameter β such that (A14) is satisfied, the factorization in the form E' is always possible, whereas the factorization in the required form E is only possible when (A13) is satisfied, i.e., when $u_1 \equiv 1$. The same difficulties arise for the

correlations along the other directions such as $\langle s_k^4 s_l^4 \rangle$. Retracing our steps we now see that in this case the transformation simply removes the s_j^2 spins by means of the star-triangle transformation and identifies the spins s_j^1 and s_{j+1}^3. Consequently

$$\begin{aligned} R^*(y_i) &= \left[\frac{1 + w_1w_2w_3}{4(1 + w_1v_5)}\right]^M R^*(u_i, v_i) \\ &= \sum_s \sum_t s_k s_l \prod (1 + y_1 s_{j+1} t_j) \\ &\quad \times (1 + y_2 t_j s_j)(1 + y_3 s_j t_{j-m}) \\ &\quad \times (1 + y_4 t_{j-m} s_{j+1})(1 + y_5 s_j s_{j+1}), \end{aligned} \tag{A15}$$

where

$$y_1 = w_3, \quad y_2 = w_2, y_3 = v_3, \quad y_4 = v_4,$$
$$y_5 = \frac{w_1 + v_5}{1 + w_1 v_5},$$

and w_1, w_2, w_3 are given by (4.4) in terms of u_2, v_1, and v_2.

(iv) Finally we change the factors $(1 + y_3 s_j t_{j-m}) \times (1 + y_4 t_{j-m} s_{j+1})$ to $(1 + y_3 s_{j+m} t_j)(1 + y_4 t_j s_{j+m+1})$ under the product $\prod_j$, which is permitted by the assumed helical lattice periodicity, and interchange the weights y_2 and y_4, thus completing the first stage of the transformation.

It is noteworthy that without difficulty one may introduce in Eq. (7.1) the additional factors of

$$(1 + u_3 s_j^1 s_{j+m}^4)(1 + u_4 s_j^3 s_{j+m}^4)(1 + u_5 s_j^3 s_{j+m}^3),$$

which simply change y_1 and y_2 to

$$y_1 = \frac{w_3 + u_3}{1 + w_3 u_3}, \quad y_2 = \frac{w_2 + u_4}{1 + w_2 u_4},$$

and add an extra term of $(1 + y_6 s_j s_{j+m})$ to Eq. (A15). The latter factor, however, gives rise to an additional vertical bond between the decorating clusters, so that the lattice is no longer a simple generalized rectangular lattice.

We note here that the reduction of the perturbed partition sum (7.1) for the "3" direction, to the form

$$\begin{aligned} R^*(y_i) &= \sum_s \sum_t s_k s_l \prod (1 + y_1 s_{j+1} t_j)(1 + y_2 t_j s_j) \\ &\quad \times (1 + y_3 s_j t_{j-m})(1 + y_4 t_{j-m} s_{j+1})(1 + y_5 t_{j-m} t_j), \end{aligned} \tag{A16}$$

does in fact hold for any generalized rectangular lattice in which the decorating clusters have no crossed (internal) bonds and are joined by single external bonds. Moreover the unperturbed partition sum can in this case also be reduced to Green's standard form (7.4) in which the spin factor $s_k s_l$ is omitted. Suppose we again denote by s_j^1, s_j^2, s_j^3, and s_j^4 the four "corner" spins of the jth cluster which are connected to the four neighboring clusters, then the necessary steps in the reduction are largely the same as for the bathroom-tile lattice. In (7.1) we have to replace E by the normalized energy of the cluster,

$$\begin{aligned} E &= \prod (1 + v_{ab} s_j^a s_j^b) \\ &= g_0 + \sum_{a_1 < a_2} g_{a_1 a_2} s_j^{a_1} s_j^{a_2} \\ &\quad + \sum_{a_1 < a_2 < a_3 < a_4} g_{a_1 a_2 a_3 a_4} s_j^{a_1} s_j^{a_2} s_j^{a_3} s_j^{a_4} + \cdots \\ &= g_0 + g_1 s_j^1 s_j^2 + g_2 s_j^2 s_j^3 + g_3 s_j^3 s_j^4 + g_4 s_j^4 s_j^1 \\ &\quad + g_5 s_j^2 s_j^4 + g_6 s_j^1 s_j^2 s_j^3 s_j^4 + g_7 s_j^1 s_j^3 + \Delta, \end{aligned} \tag{A17}$$

where the product is taken over all decorated bonds, only an even number of spins appears in each summation and Δ consists of all the terms which contain one or none of the four corner spins. Successive summations over the ± 1 values of all the spins in the jth cluster except s_j^3 and s_j^4, reduces $R^*(u_i, v_i)$ to the form (A9). Furthermore the same symmetries (A10) hold. The coefficients g_i may now be identified with the eight fundamental coefficients $f_0, f_{12}, \cdots, f_{1234}$, which were obtained by Green and Hurst,[20,21] using the slightly different technique of counting polygons on the decorating cluster.

The first consistency condition (A5a) or (A12b) is now a consequence of the absence of crossed bonds on the decorating cluster, whereas the second condition (A12c) can always be satisfied, by choosing the parameter β appropriately in (A14). For the unperturbed partition sum we may interchange the roles of the spins s_j and t_j in (A16), rearrange the factors under the product $\prod_j$, and perform the summation over the t_j spins in the same way as in (7.11).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Exact Quantization Conditions

C. ROSENZWEIG*† AND J. B. KRIEGER
Physics Department, Polytechnic Institute of Brooklyn, Brooklyn, New York

(Received 6 September 1967)

The method of Froman and Froman for proving exact quantization conditions is reviewed. This formalism, unlike the usual WKB approximation to which it bears a close resemblance, requires consideration of the behavior of the potential everywhere it is defined. This approach leads to proofs that certain quantization conditions are exact without having to compare the results to solutions of the Schrödinger equation obtained by other means. Using the formalism, we prove the correctness of all previously known exact quantization rules for the one-dimensional and radial cases. Furthermore, it is shown that exact quantization rules can be proved for two other potentials. For one of these, no analytic solutions to the Schrödinger equation are known. For the latter case, the proof is checked by numerical integration of the Schrödinger equation for a special case.

1. INTRODUCTION

One of the successes of the WKB approximation has been the justification for and improvement of the Bohr–Sommerfeld quantization condition. The WKB quantization rule has proved to be a very valuable tool when one seeks to find approximate eigenvalues for the one-dimensional or radial Schrödinger equation. By the use of higher-order correction terms to the standard Bohr–Sommerfeld condition highly accurate, although nonexact, energy eigenvalues have been obtained for a variety of potentials.[1] Furthermore, it has been known for many years that in the special case of the one-dimensional harmonic oscillator the standard WKB quantization condition gives rise to the exact values for the energy. Proofs of varying degrees of rigor have been advanced which demonstrate the exactness of the quantization condition. One method simply compares the WKB result to the eigenvalues obtained from an exact solution of the Schrödinger equation. Although this constitutes a completely rigorous demonstration it is obviously not useful in determining the correctness of the quantization condition in precisely those cases where it is of the most interest, i.e., when the Schrödinger equation cannot be solved in closed form, nor does it offer any deeper insight into the WKB method.

A second method proceeds by showing that all additional higher-order correction terms to the WKB integral vanish for the given potential.[2] These proofs, however, are not entirely rigorous since they fail to take cognizance of the fact that these correction terms are only asymptotically valid, as $\hbar \to 0$. Thus, when we prove that all higher-order terms are zero, we can only state that the correction to our quantization condition is asymptotically zero. It is still possible to have correction terms of order $e^{-|a|/\hbar}$, which are nonzero, but which have an asymptotic series representation consisting entirely of zeros, i.e., if we add to the harmonic oscillator potential a term which is small and negative in a region outside the classical turning points and zero elsewhere, then all the higher-order correction terms will be zero because they only involve the behavior of the potential between the turning points. However, if we calculate the expectation value of the Hamiltonian using the unperturbed ground-state wave function we find immediately that the perturbed ground-state energy must be lower than the unperturbed ground state by terms of $O(e^{-|a|/\hbar})$.

This situation is analogous to one in ordinary perturbation theory where it is possible to leave out terms of $O(e^{-1/\lambda})$ where λ is the coupling constant,[3] the best known example arising in calculating the binding energy of Cooper pairs in the theory of superconductivity.[4] Furthermore, there are known cases for which the WKB integral does not give rise to the exact eigenvalues, but a modified WKB integral does. In these cases it is not even clear what "correction" terms we must show are actually zero. Thus a different approach is necessary if we are to prove the exactness of the WKB or modified WKB quantization conditions.

Recently Froman and Froman[5] have devised a new method to handle this problem. They have developed an encompassing, rigorous approach to the entire theory of the WKB approximation. Part of their work deals with the WKB quantization condition and by use of their formalism they are able to prove rigorously

* Supported in part by NSF Undergraduate Research Grant No. GY752 and NSF Graduate Traineeship. This article is based on a thesis submitted, in partial fulfillment of the requirements for an M.S. degree, to the Polytechnic Inst. of Bklyn.

† Present address: Physics Department, Harvard University, Cambridge, Mass.

[1] J. B. Krieger, M. L. Lewis, and C. Rosenzweig, J. Chem. Phys. **47**, 2942 (1967).

[2] P. N. Argyes, Physics **2**, 131 (1965). See also a comment in J. L. Dunham's article, Phys. Rev. **41**, 713 (1932).

[3] J. B. Krieger, J. Math. Phys. **9**, 432 (1968).

[4] L. N. Cooper, Phys. Rev. **104**, 1189 (1956).

[5] N. Froman and P. O. Froman, *JWKB Approximation; Contributions to the Theory* (North-Holland Publ. Co., Amsterdam, 1965).

and unambiguously the exact nature of the quantization condition for the special case of the one-dimensional harmonic oscillator.

It is also well known that if we make a slight modification of the WKB quantization condition we are able to obtain exact-energy eigenvalues for the two radial problems corresponding to the Coulomb potential and the three-dimensional, isotropic, harmonic oscillator. Langer[6] has justified this modification for radial problems in general, and more recently the first two correction terms to the WKB integral have been derived and have been shown to be identically zero for the above two potentials.[7] Froman and Froman (referred to hereafter as FF) also examine this situation and are again able to provide a rigorous proof of the exactness of the modified quantization condition.

Until several years ago these three cases were the only known examples of potentials for which simple exact quantization rules existed. Bailey[8] then discovered several other potentials which give rise to exact quantization conditions by use of either the WKB or a modified WKB quantization condition. However, he could give no justification as to why these results should hold, other than that they provided the same results as obtained by an exact, analytical solution of the Schrödinger equation. It is the purpose of this paper to prove that all of Bailey's quantization rules can be obtained by applying the formalism of FF to the potentials under consideration. We thus avoid the direct recourse of comparing our results to the results of solving Schrödinger's equation in order to verify the exactness of the conditions. In addition we will also discuss other potentials for which no exact quantization conditions have previously been known to exist. For one of these potentials the energy eigenvalue can be obtained from an analytic solution of Schrödinger's equation. The other potential, however, is of special significance since it is the only potential known with an exact quantization condition whose Schrödinger equation cannot be solved in terms of known functions.

The potentials we treat in the following sections are slight modifications, or generalizations of those discussed by Bailey or by FF.

2. THE METHOD OF FROMAN AND FROMAN

Froman and Froman have developed a method for solving equations of the form

$$\frac{d^2\psi}{dz^2} + Q^2(z)\psi = 0, \tag{2.1}$$

which has a strong relationship to the standard WKB approximation employed in quantum mechanics. In this section we outline only the barest skeleton of the theory of FF in order to establish the tools which are necessary to prove the exactness of quantization conditions. We do not prove any of the statements which appear in this section since all proofs can be found in the monograph by FF or follow readily from their results.

Since we are interested only in quantization conditions some of the restraints imposed on the formulas and functions appearing in our discussion will be more specific than those employed in the general treatment in the Fromans's monograph.

For problems of interest in quantum mechanics $Q^2(z)$ appearing in Eq. (2.1) is

$$Q^2(z) = \frac{2m}{\hbar^2}[E - V(z)] \tag{2.2}$$

and is assumed to be an analytic and single-valued function of z in some region of the complex plane including the section of the real axis on which the physical potential is defined. We assume that, unless stated explicitly, the complex plane is cut in such a manner as to make all functions which appear in our formula analytic and single-valued throughout the region of interest. In the case when Eq. (2.1) is the radial Schrödinger equation we employ the same notation except that $V(z)$ will now be an effective potential

$$V_{\text{eff}}(z) = V_{\text{radial}}(z) + \frac{l(l+1)\hbar^2}{2mz^2}.$$

Our goal is to establish the exact validity of the following equation for the various potentials under consideration:

$$\int_{x'}^{x''} |q(x)|\,dx = (n + \tfrac{1}{2})\pi. \tag{2.3}$$

The function $q^2(x)$ is real on the x axis and is closely related to, although not necessarily identical to $Q^2(x)$ defined in Eq. (2.2). The exact choice of the functional relationship between q^2 and Q^2 will be determined by the nature of $V(z)$ in Eq. (2.2); x' and x'' are the two real zeros of $q^2(z)$. These zeros must be the only two zeros on the region of the real axis of physical interest.

Much of our attention will be focused on the function $q(z)$ and on functionals of $q(z)$. All theorems stated below are valid only if the phase of $q(x)$ is chosen as in Fig. 1.

The chief object of attention in the FF treatise is a certain matrix, called the F matrix, $\hat{F}$. Complete knowledge of this matrix is sufficient to determine the solutions of Eq. (2.1). Since we are not interested in the

[6] R. E. Langer, Phys. Rev. **51**, 669 (1937).
[7] J. B. Krieger and C. Rosenzweig, Phys. Rev. **164**, 171 (1967).
[8] P. B. Bailey, J. Math. Phys. **5**, 1293 (1964).

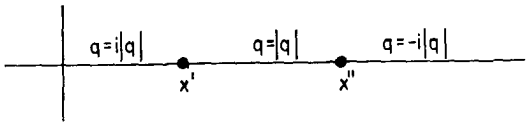


FIG. 1. Argument of $q(x)$ along the real axis.

wavefunctions we do not require detailed knowledge of the properties of $\hat{F}$, and a limited amount of information will suffice. The F matrix is a 2×2 matrix which is written as

$$\hat{F}(z, z_0) = \begin{pmatrix} F_{11}(z, z_0) & F_{12}(z, z_0) \\ F_{21}(z, z_0) & F_{22}(z, z_0) \end{pmatrix}. \tag{2.4}$$

$\hat{F}$ has the property that

$$\hat{F}(z_2, z_0) = \hat{F}(z_2, z_1)\hat{F}(z_1, z_0), \tag{2.5}$$

while the inverse matrix

$$\hat{F}^{-1}(z, z_0) = \hat{F}(z_0, z)$$

is given by

$$\begin{aligned} \hat{F}^{-1}(z, z_0) &= \begin{pmatrix} F_{11}(z_0, z) & F_{12}(z_0, z) \\ F_{21}(z_0, z) & F_{22}(z_0, z) \end{pmatrix} \\ &= \begin{pmatrix} F_{22}(z, z_0) & -F_{12}(z, z_0) \\ -F_{21}(z, z_0) & F_{11}(z, z_0) \end{pmatrix}. \end{aligned} \tag{2.6}$$

Before we proceed with our enumeration of certain properties of $\hat{F}$, we find it necessary to define several new functions. In particular we define the functions w, ϵ, μ.

$$w(z) \equiv \int_p^z q(\zeta)\, d\zeta \quad (p \text{ arbitrary point}), \tag{2.7}$$

$$\epsilon(z) \equiv \frac{Q^2(z) - q^2(z)}{q^2(z)} + \frac{1}{16q^2}\left\{5\left(\frac{d(q^2)}{dz}\right)^2 - 4q^2\frac{d^2q^2}{dz^2}\right\}, \tag{2.8}$$

$$\mu(z, z_0) \equiv \int_{z_0}^z |\epsilon q\, dz|. \tag{2.9}$$

The latter integration may be taken along any path λ, provided ϵq is analytic along the path (i.e., we cannot cross branch cuts). It should be noted, however, that μ is nonanalytic and hence path-dependent. Many times we will use the notation $\hat{F}(w, w_0)$ interchangeably with $\hat{F}(z, z_0)$ and the two expressions represent the same quantity but they are functionals of different, related functions.

The general quantization condition derived by FF for any arbitrary potential is

$$\int_{x'}^{x''} |q(x)|\, dx = (n + \tfrac{1}{2})\pi - \arg\frac{F_{22}(z, +\infty)}{F_{22}(z, -\infty)}, \tag{2.10}$$

where z is any arbitrary point in the complex plane provided that $|e^{iw(x)}|$ increases monotonically and tends to infinity as $x \to \pm\infty$, and that μ be convergent along a part of the real axis as $x \to \pm\infty$. For our particular choice of potentials we can show that $F_{22}(z, +\infty)/F_{22}(z, -\infty)$ is real and hence its argument will be 0 or π establishing the validity of the simple, exact quantization condition Eq. (2.3). We therefore need to know the properties of the F matrix which will enable us to establish the reality of the above specified ratio.

It is now necessary to examine the limiting properties of $\hat{F}$ since we make much use of these limits in our proofs. In particular, the limit of $\hat{F}(z, z_0)$ as $z \to Z$, where $|e^{iw(z)}| \to \infty$ as $z \to Z$, is of primary interest. If we can find a point z_1 such that

(a) all elements of $\hat{F}(z_1, z_0)$ are finite, (b) $\lim_{z\to Z} \mu(z_1, z)$ is finite, (c) $|e^{iw(z)}|$ increases monotonically and tends to infinity as $z \to Z$ along a path λ from z_1 to Z; then the following limiting properties hold:

$$\lim_{z\to Z} F_{11}(z, z_0) \quad \text{exists and is finite,} \tag{2.11a}$$

$$\lim_{z\to Z} F_{12}(z, z_0) \quad \text{exists and is finite,} \tag{2.11b}$$

$$\lim_{z\to Z} F_{21}(z, z_0)e^{-2iw(z)} = 0, \tag{2.11c}$$

$$\lim_{z\to Z} F_{22}(z, z_0)e^{-2iw(z)} = 0; \tag{2.11d}$$

$$\lim_{z\to Z} F_{22}(z_0, z) \quad \text{exists and is finite,} \tag{2.12a}$$

$$\lim_{z\to Z} F_{12}(z_0, z) \quad \text{exists and is finite,} \tag{2.12b}$$

$$\lim_{z\to Z} F_{21}(z_0, z)e^{-2iw(z)} = 0, \tag{2.12c}$$

$$\lim_{z\to Z} F_{11}(z_0, z)e^{-2iw(z)} = 0. \tag{2.12d}$$

All limits which appear in Eqs. (2.11) and (2.12) are to be taken along a path from z_0 to z_1, and then from z_1 to Z along the path λ described above. Furthermore if $\lim_{z\to Z} F_{22}(z_0, z)$ exists and is finite, this limit is independent of how one approaches Z.

In addition to the limiting values of $\hat{F}$, it will also be necessary for us to employ certain upper bounds on $\hat{F}(w, w_0)$. It is essential that when discussing these bounds a definite path λ, connecting w and w_0 in the complex plane be chosen. A number M is found such that

$$\tfrac{1}{2}|1 - e^{-2i(w_{\nu-1}-w_\nu)}| \leq M \tag{2.13a}$$

for any possible division of the path λ. The w_ν form an ordered division of the path λ from w_0 to w where w_ν can assume any of the values of x from w_0 to $w_{\nu-1}$ for $\nu > 1$ and for $\nu = 1$, w_1 is between w_0 and w. It can then be shown that

$$|F_{11}(w, w_0) - 1| \leq \frac{1}{2M}[e^{M\mu} - 1]. \tag{2.14a}$$

If the path λ is chosen such that $|e^{iw}|$ increases monotonically as one moves from w_0 to w,

$$|e^{-2iw_\nu}| \leq |e^{-2iw_0}|, \tag{2.13b}$$

$$|e^{2iw_1}| \leq |e^{2iw}|, \tag{2.13c}$$

$$|e^{2i(w_1-w_\nu)}| \leq |e^{2i(w-w_0)}|. \tag{2.13d}$$

Under this assumption for λ, the estimates involving the remaining elements of the F matrix are

$$|F_{12}(w, w_0)| \leq \frac{1}{2M}[e^{M\mu} - 1]\,|e^{-2iw_0}|, \tag{2.14b}$$

$$|F_{21}(w, w_0)| \leq \frac{1}{2M}[e^{M\mu} - 1]\,|e^{2iw}|, \tag{2.14c}$$

$$|F_{22}(w, w_0) - 1| \leq \tfrac{1}{2}\mu + \frac{1}{2M}[e^{M\mu} - 1 - M\mu]\,|e^{2i(w-w_0)}|. \tag{2.14d}$$

Care must be exercised if we choose a path λ where $|e^{iw}|$ does not increase monotonically from w_0 to w, and we want to employ estimates similar to (2.14). In these situations we must be certain that a finite M exists in order to employ Eq. (2.14a). To use Eqs. (2.14b)–(2.14d) additional caution is needed and we must find a w' such that

$$|e^{-2iw_\nu}| \leq |e^{-2iw'}|, \tag{2.15a}$$

for any w_ν; a w'' such that

$$|e^{2iw_1}| \leq |e^{2iw''}|, \tag{2.15b}$$

and then

$$|e^{2i(w_1-w_\nu)}| \leq |e^{2i(w''-w')}|. \tag{2.15c}$$

Thus $|e^{-2iw}|$ is a maximum along λ at $w = w'$ and $|e^{2iw}|$ is a maximum along λ at $w = w''$. Equations (2.14a)–(2.14d) become

$$|F_{11}(w, w_0) - 1| \leq \frac{1}{2M}[e^{M\mu} - 1], \tag{2.16a}$$

$$|F_{12}(w, w_0)| \leq \frac{1}{2M}[e^{M\mu} - 1]\,|e^{-2iw'}|, \tag{2.16b}$$

$$|F_{21}(w, w_0)| \leq \frac{1}{2M}[e^{M\mu} - 1]\,|e^{2iw''}|, \tag{2.16c}$$

$$|F_{22}(w, w_0) - 1| \leq \tfrac{1}{2}\mu + \frac{1}{2M}[e^{M\mu} - 1 - M\mu]\,|e^{2i(w''-w')}|. \tag{2.16d}$$

The final information which we need follows from the definition of the elements of $\hat{F}$. In Eq. (2.7) when we defined w, it was stated that the lower limit of integration was arbitrary. Naturally the choice of various different points alters the value of w and in general any arbitrary functional of w. However, it turns out that

$$F_{11}(w, w_0), \quad F_{22}(w, w_0), \quad F_{12}(w, w_0)e^{2iw}, \quad F_{21}(w, w_0)e^{-2iw} \tag{2.17}$$

are all independent of our choice of the lower limit in (2.7).

Finally, the definition of $F_{22}(w, w_0)$ involves only terms of the form

$$\int_{w_0}^{w_n} dw_{n+1}\, i\epsilon(w_{n+1})[1 - e^{-2i(w_n-w_{n+1})}]e^{2i(w_1-w_{n+1})}, \tag{2.18}$$

in such a manner that if this term is real, $F_{22}(w, w_0)$ itself will be real. Clearly this fact proves important in the following discussion.

The above equations, conditions, and discussions supply us with the tools with which to prove exact quantization conditions. With these theorems at our disposal we can proceed to establish the desired rules.

3. PROOF OF EXACT QUANTIZATION CONDITIONS

We open this section by listing the various potentials for which exact quantization conditions have been demonstrated. Along with the potentials we list the computed values of the energy eigenvalues when such energies can be written explicitly in closed form. All constants appearing in Table I are assumed to be positive. It is also assumed that the boundary conditions on ψ for each problem are such that ψ vanishes at the end points of the stated interval. In all cases n starts from $n = 0$ and assumes all positive integral values unless explicitly restricted.

The proofs of the exact quantization condition for the first three potentials are to be found in FF. Potential I is exactly the same as that employed in their monograph, while potentials II and III differ slightly because of the addition of a general term $b\hbar^2/2mr^2$ which breaks up the well-known l degeneracy of these potentials. However, the addition of this term does not effect the proof or validity of the quantization rules as presented by the Fromans.

The table shows that for cases II and III we must subtract $1/4r^2$ from Q^2 corresponding to the familiar replacement of $l(l+1)$ by $(l+\frac{1}{2})^2$. As we soon see this choice of q^2 different from Q^2 is typical of problems which contain second-order poles of Q^2 in the finite z plane. We now proceed to establish quantization rules for potentials IV–VIII. In order to simplify our equations somewhat we choose our system of units, such that $\hbar^2/2m = 1$.

TABLE I. Potentials possessing exact quantization conditions.

	Potential	Range of x	Energy eigenvalues	q^2	Comments
I.	$V = \frac{1}{2}kx^2$	$-\infty < x < +\infty$	$E = \hbar(k/m)^{\frac{1}{2}}(n + \frac{1}{2})$	Q^2	Linear harmonic oscillator
II.	$V = -\frac{V_0}{r} + \frac{b\hbar^2}{2mr^2} + \frac{l(l+1)\hbar^2}{2mr^2}$	$0 < r < \infty$	$E = \frac{-V_0^2 m}{2\hbar^2[n + \frac{1}{2} + ((l+\frac{1}{2})^2 + b)^{\frac{1}{2}}]^2}$	$Q^2 - \frac{1}{4r^2}$	Coulomb potential and degeneracy breaking term
III.	$V = \frac{1}{2}kr^2 + \frac{\hbar^2 b}{2mr^2} + \frac{l(l+1)\hbar^2}{2mr^2}$	$0 < r < \infty$	$E = \left(\frac{k}{m}\right)^{\frac{1}{2}}[2n + 1 + ((l+\frac{1}{2})^2 + b)^{\frac{1}{2}}]$	$Q^2 - \frac{1}{4r^2}$	Three-dimensional isotropic harmonic oscillator and degeneracy breaking term
IV.[a]	$V(x) = \frac{-V_0}{\cosh^2 \frac{x}{a}}$	$-\infty < x < +\infty$	$E = \frac{-\hbar^2}{2ma^2}\left[\left(\frac{2mV_0a^2}{\hbar^2} + \frac{1}{4}\right)^{\frac{1}{2}} - (n + \frac{1}{2})\right]^2$	$Q^2 + \frac{1}{4a^2\cosh^2\frac{z}{a}}$	$n \leq \left[\frac{2mV_0a^2}{\hbar^2} + \frac{1}{4}\right]^{\frac{1}{2}} - \frac{1}{2}$
V.[a]	$V(x) = V_0\cot^2\frac{\pi x}{a} = \frac{V_0}{\sin^2\frac{\pi x}{a}} - V_0$	$0 < x < a$	$E = -V_0 + \frac{\pi^2\hbar^2}{2ma^2}\left[\left(\frac{2mV_0a^2}{\pi^2\hbar^2} + \frac{1}{4}\right)^{\frac{1}{2}} + (n + \frac{1}{2})\right]^2$	$Q^2 - \frac{\pi^2}{4a^2\sin^2\frac{\pi z}{a}}$	
VI.[b]	$V(r) = \frac{-\lambda e^{-\alpha r}}{(1-e^{-\alpha r})} + \frac{be^{-\alpha r}}{(1-e^{-\alpha r})^2}$	$0 < r < \infty$	$E = -\frac{\{[n + \frac{1}{2} + \frac{1}{2}(1 + 4b/\alpha^2)^{\frac{1}{2}}]\alpha - \lambda\}^2}{4[n + \frac{1}{2} + \frac{1}{2}(1 + (4b/\alpha^2)^{\frac{1}{2}})]^2\alpha^2}$	$Q^2 - \frac{\alpha^2 e^{-\alpha r}}{4(1-e^{-\alpha r})^2}$	$n \leq \frac{\lambda}{\alpha} - \frac{1}{2} - \frac{1}{2}\left(1 + \frac{4b}{\alpha^2}\right)^{\frac{1}{2}}$
VII.[a]	$V(x) = Ae^{-2ax} - Be^{-ax}$	$-\infty < x < \infty$	$E = -\frac{B}{2}\left[\left(\frac{B}{2A}\right)^{\frac{1}{2}} - \left(\frac{a(n+\frac{1}{2})\hbar}{(mB)^{\frac{1}{2}}}\right)\right]^2$	Q^2	$n \leq \frac{B}{a\hbar}\left(\frac{m}{2A}\right)^{\frac{1}{2}} - \frac{1}{2}$
VIII.	$V(x) = Ae^{2ax} + Be^{-2ax}$	$-\infty < x < \infty$		Q^2	No closed expression for E exists, and we must rely on numerical integration of the quantization condition

[a] The exact quantization condition for these potentials was first proposed by Bailey (Ref. 8).
[b] For the choice $b = 0$ these energy levels agree with those derived by A. Bhattacharjie and E. C. G. Sudarshan, Nuovo Cimento 25, N4 864 (1962). For the functional form of the wavefunction see either Bhattacharjie and Sudarshan for the case $b = 0$, or A. H. Bose, Nuovo Cimento 32, N3 679 (1964) for a more general case.

Case IV: Due to the existence of poles at $z = a\pi i(n + \frac{1}{2})$ we must choose $q^2(z)$ different from $Q^2(z)$

$$q^2(z) = Q^2(z) + \frac{1}{4a^2\cosh^2(z/a)} = E + \frac{4a^2V_0 + 1}{4a^2\cosh^2(z/a)}. \tag{3.1}$$

The reason for this particular choice becomes evident when we examine the convergence of the μ integral.

Along the real axis the phase of $q(x)$ is chosen to agree with Fig. 1 and from Eq. (2.10) the exact quantization condition for $q^2(x)$ as above is

$$\int_{x'}^{x''} |q(x)|\,dx = (n + \tfrac{1}{2})\pi - \arg\frac{F_{22}(z, +\infty)}{F_{22}(z, -\infty)}. \tag{3.2}$$

As we previously noted, if we can establish that $F_{22}(z, +\infty)/F_{22}(z, -\infty)$ is real we will have arrived at our exact quantization rule given by Eq. (2.3). In order to show this we choose $z = a\pi i/2$ and proceed to prove that both $F_{22}(\pi ai/2, +\infty)$ and $F_{22}(\pi ai/2, -\infty)$ are real.

The first step is to cut the complex plane so that the phase of q agrees with Fig. 1. The branch points of q occur at the zeros and poles of q^2. The poles of q^2 occur when $\cosh^2 z/a \to 0$ or $z \to \pm(n + \frac{1}{2})a\pi i/2$ and they are poles of the second order. The zeros of q^2 occur when

$$\cosh\frac{z}{a} = \frac{1}{2a}\left[-\left(\frac{4a^2V_0 + 1}{E}\right)\right]^{\frac{1}{2}}.$$

(Only the positive square root gives rise to real zeros of q^2 and this restricts the possible values of E.) The zeros correspond to the values of z for which

$$z = \pm\cosh^{-1}\left\{\frac{1}{2a}\left[\frac{(4a^2V_0 + 1)}{-E}\right]^{\frac{1}{2}}\right\} + a\pi i.$$

The pole-zero diagram for the complex plane is depicted in Fig. 2. The Fig. 2, and all other diagrams of the branch cuts of the complex planes appearing in this section, conform to the following notation: x designates a pole of q^2, 0 designates a zero of q^2, ⌇⌇ designates a branch cut of q.

We also imagine the branch points to extend to infinity both in the $+i$ direction and the $-i$ direction for all five cases under study. As we soon see, this extra infinity of branch points makes essentially no contribution to the considerations of concern.

Region I consists of all points x to left of zero 4. Region II consists of all points x to the right of zero 4 and to the left of zero 5. Region III consists of all points to the right of zero 5. Region IV consists of all points $x + (a\pi i/2)$ with $x > 0$. Region V consists of all points $x + (a\pi i/2)$ with $x < 0$.

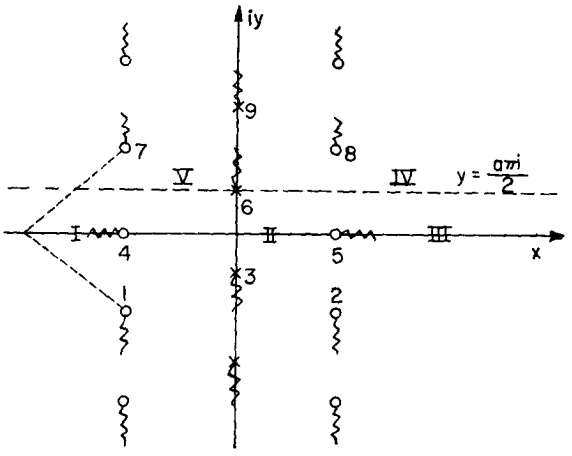


FIG. 2. Branch cuts in the complex plane for case IV.

All angles measured from branch points below the real axis range from $-\frac{1}{2}\pi < \theta < \frac{3}{2}\pi$, those measured from branch points lying above the axis range from $-\frac{3}{2}\pi < \theta < \frac{1}{2}\pi$, while θ_4 (angle measured from branch point 4) has the range $-\pi < \theta_4 < \pi$ and $0 < \theta_5 < 2\pi$. Because both the poles and zeros are symmetric with respect to the real axis we see that for the angles 1 and 7 depicted in Fig. 2, $\theta_7 = -\theta_1$. A similar result holds for the angles of any two symmetric branch points. Hence, the contributions to the phase from all the poles and zeros not on the real axis cancel each other out and the net contributions from the infinite set of branch points is zero. Only the branch points 4 and 5 contribute to the phase.

In region I $\quad \theta_4 + \theta_5 = \pi + \pi = 2\pi,$
region II $\quad \theta_4 + \theta_5 = 0 + \pi = \pi,$
region III $\quad \theta_4 + \theta_5 = 0 + 0 = 0.$

The phase for $q(z)$ is

$$q(z) = \exp(i\theta_0)\exp i\sum_n \frac{\theta_n}{2}|q|$$

[the $\frac{1}{2}$ arises from the fact that these poles and zeros are for q^2, while we are interested in the phase of q, where $q = (q^2)^{\frac{1}{2}}$] and in order to have the proper phase according to Fig. 1, $\theta_0 = -\frac{1}{2}\pi$ and

region I $\quad q = i|q|,$
region II $\quad q = |q|,$
region III $\quad q = -i|q|.$

We can also compute the phase of q on the line $y = a\pi i/2$ where again symmetry simplifies the choice of phase. For this line the only contributing branch point is 6.

In region IV $\theta_6 = 0$,
region V $\theta_6 = -\pi$,
so

$$q = \exp\left(-\frac{i\pi}{2}\right)\exp\left(-\sum\frac{2\theta_n}{2}\right)|q|$$

(-2 appears since 6 is a pole of second order) and in

$$\text{region IV} \quad q = -i|q|$$
$$\text{region V} \quad q = i|q|.$$

This choice of phase for q allows us to make the following observations about $|e^{2iw(z)}|$. We see that $|e^{2iw(z)}|$ increases monotonically as one goes from $x'' \to +\infty$, $x' \to -\infty$, $a\pi i/2 \to |x| + (a\pi i/2)$, and $a\pi i/2 \to -|x| + (a\pi i/2)$ where $|x|$ may be ∞. Since we have chosen to evaluate the elements of $\hat{F}$ between $a\pi i/2$ and $\pm\infty$ we must be confident that these elements exist. If we wish to make use of estimates (2.14) and (2.16) we must require that the μ integral converge in order to obtain finite estimates for the elements of the $\hat{F}$ matrix. Therefore, it is necessary for us to examine the properties of μ for the potential at hand. From Eqs. (2.8) and (3.1) we find that

$$\epsilon(z) = \frac{-1/4a^2\cosh^2(z/a)}{\left(E + \dfrac{4a^2V_0 + 1}{4a^2\cosh^2(z/a)}\right)} + \frac{1}{16\left(E + \dfrac{4a^2V_0 + 1}{4a^2\cosh^2(z/a)}\right)^3} \times \left\{\frac{2E(4a^2V_0 + 1)}{a^4\cosh^4(z/a)}\left(3 - 2\cosh^2\frac{z}{a}\right) + \frac{4a^2V_0 + 1}{4a^6\cosh^4(z/a)} + \frac{(4a^2V_0 + 1)^2}{4a^6\cosh^6(z/a)}\right\}. \quad (3.3)$$

As $z \to \pm(n + \frac{1}{2})a\pi i$, $\cosh z/a \to 0$ and a second-order pole appears for q^2. However, in this limit Eqs. (3.3) and (3.1) show that $\epsilon(z)q(z) \to 0$ and hence μ defined by Eq. (2.9) is finite as either z or $z_0 \to (n + \frac{1}{2})a\pi i$. It should be noted that only by choosing q^2 different from Q^2 in accordance with Eq. (3.1) were we able to eliminate the singularity from ϵq. It is this need to insure the corresponding convergence of the μ integral which dictates the choice of q^2 for all potentials which we discuss. In order to understand the motivation for the various specific choices of q^2 different from Q^2 which we employ, it is important that the manner in which the $(Q^2 - q^2)/q^2$ term cancels out the singularity in the remaining terms of ϵq be clearly understood. We should also check the behavior of μ as $|x| \to \infty$ ($z = x + iy$). For this situation $|\cosh z| \approx |\sinh z| \approx e^{|x|}$ and

$$\lim_{|x|\to\infty}|\epsilon q| \propto \frac{1}{|\cosh^2(x/a)|} \to \exp\left(-\frac{2|x|}{a}\right) \to 0.$$

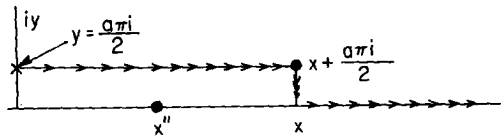


FIG. 3. The path λ used in evaluating $\hat{F}(a\pi i/2, +\infty)$.

Hence the μ integral is convergent. Moreover, if we evaluate $\mu(z, z_0)$ along a path λ chosen such that for the *entire* path λ in the complex plane $|\mathrm{Re}\, z| \to \infty$ then the above estimate shows that $\mu(z, z_0) \to 0$. This fact will also be of importance in our proof.

As we mentioned previously if we can prove that both $F_{22}(a\pi i/2, +\infty)$ and $F_{22}(a\pi i/2, -\infty)$ are real our exact quantization rule will have been established. In the following we only give the proof that $F_{22}(a\pi i/2, +\infty)$ is real, since the proof that $F_{22}(a\pi i/2, -\infty)$ is real, is completely analogous. In order to prove that $F_{22}(a\pi i/2, +\infty)$ is real we employ the 2,2 element from Eq. (2.5) with $z = a\pi i/2$, $z_1 = (a\pi i/2) + x$, $z_0 = +\infty$ and obtain

$$F_{22}\left(\frac{a\pi i}{2}, +\infty\right) = F_{21}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)F_{12}\left(x + \frac{a\pi i}{2}, +\infty\right) + F_{22}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)F_{22}\left(x + \frac{a\pi i}{2}, +\infty\right). \quad (3.4)$$

The path connecting $a\pi i/2$ and $+\infty$ in the complex plane is shown in Fig. 3. Since we shall shortly establish a result which is equivalent to proving that the elements of $\hat{F}[x + (a\pi i/2), x]$ are all finite and since we have already shown that $|e^{iw(z)}|$ increases monotonically between x, and $x' \to +\infty$ and that the μ integral converges, we are able to use (2.12a), (2.12b) to claim that $F_{12}[x + (\alpha\pi i/2), +\infty]$ and $F_{22}[x + (a\pi i/2), +\infty]$ exist and are finite. Equations (2.6) and (2.14a), (2.14c) combined with the finite value of $\mu[x + (a\pi i/2), a\pi i/2]$ establish the existence of finite values of the remaining two matrix elements in Eq. (3.4).

We now proceed to demonstrate that in the limit $x \to +\infty$,

$$F_{21}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)F_{21}\left(x + \frac{a\pi i}{2}, +\infty\right) = 0,$$

$$F_{22}\left(x + \frac{a\pi i}{2}, +\infty\right) = 1, \quad F_{22}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)$$

is real and hence that

$$F_{22}\left(\frac{a\pi i}{2}, +\infty\right) = F_{22}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)$$

is real. Equations (2.17) establish that $F_{21}F_{12}$ is independent of the lower limit in Eq. (2.7). Hence we can choose any convenient point for this lower limit.

Equations (2.6) and (2.14c) combine to give us the estimate for

$$\left|F_{21}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)\right| \leq \tfrac{1}{2}[e^{\mu} - 1]\left|\exp\left[2i\int_{p}^{x+(a\pi i)/2} q\,dx\right]\right|, \quad (3.5)$$

where the convergence of the μ integral and a convenient choice of P guarantees the finitude of F_{21}.

An estimate of $|F_{12}|$ cannot be arrived at so easily, however. Estimates (2.16b) must be employed, but first we must demonstrate that we can find an M and a w'. We can divide the path λ from $x + (a\pi i/2)$ to $+\infty$ shown in Fig. 3 into two parts, $x + (a\pi i/2)$ to x, and x to $+\infty$. Since $|e^{iw(x)}|$ increases monotonically from x to $+\infty$ we can choose $M_1 = 1$ and $w_1(z) = w(x)$ for this section of the path. If we can find an M_2 which will satisfy (2.13a) for all w on the path from $x + (a\pi i/2)$ to x, we can choose as the M to be used in (2.16) the maximum of M_1 and M_2. Similarly if we find a w_2' satisfying (2.15a) for the segment $x + (a\pi i/2)$ to x, we choose as our w' the w_1' or w_2' for which $|e^{-iw_1}|$ or $|e^{-iw_2'}|$ is a maximum. That such finite M_2 and w_2' exist is evident from the fact that we are dealing with a finite line segment along which q, $\int q\,dz$, and hence $|e^{iw(z)}|$ are all well-behaved with no singularities. This also guarantees the existence of finite $\hat{F}[x + (a\pi i/2), x]$. Our estimate for $F_{12}[x + (a\pi i/2), +\infty]$ is then

$$\left|F_{12}\left(x + \frac{a\pi i}{2}, +\infty\right)\right| \leq \frac{1}{2M}\left[\exp\left(M\mu\left(x + \frac{a\pi i}{2}, +\infty\right)\right) - 1\right] \times \left|\exp\left[-2i\int_{p}^{z'} q\,dz\right]\right|, \quad (3.6)$$

where $w'(z)$ discussed above is written as $\int_p^{z'} q\,dz$.

Combining (3.5) and (3.6), we have

$$\left|F_{21}\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)F_{12}\left(\frac{a\pi i}{2} + x, +\infty\right)\right| \leq \frac{1}{4M}\left[\exp\left(\mu\left(\frac{a\pi i}{2}, x + \frac{a\pi i}{2}\right)\right) - 1\right] \times \left[\exp\left(M\mu\left(x + \frac{a\pi i}{2}, +\infty\right)\right) - 1\right] \times \left|\exp\left(2i\int_{z'}^{x'+(a\pi i)/2} q\,dz\right)\right|. \quad (3.7)$$

In the limit of $x \to +\infty$ everything is finite and $\mu[x + (a\pi i/2), +\infty] \to 0$ so that $|F_{21}F_{12}| \to 0$ and our first contention has been shown to be correct.

By an essentially identical argument, in fact we use the same M and Eqs. (2.6) and (2.14a) to obtain

$$\left|F_{22}\left(x + \frac{a\pi i}{2}, +\infty\right) - 1\right| \leq \frac{1}{2M}\left[\exp\left(M\mu\left(x + \frac{a\pi i}{2}, +\infty\right)\right) - 1\right], \quad (3.8)$$

and as $x \to +\infty$, $F_{22}[x + (a\pi i/2), +\infty] \to 1$. Thus we only have to show that $F_{22}[(a\pi i/2), x + (a\pi i/2)]$ is real and our proof will be complete. We first note that by (2.17) we are free to choose the lower limit of (2.7) anywhere we please. If we chose a point in region IV and consider only values of w in this region we find that

$$w = \int_p^z q(z)\,dz = \int_p^{x+(a\pi i)/2} i\,|q|\,dx.$$

Hence w and dw are purely imaginary. Since $\epsilon(q^2, Q^2)$ assumes real values we see that expression (2.18) is real and by the discussion accompanying (2.18) we can claim that $F_{22}[(a\pi i/2), x + (a\pi i/2)]$ is real.

The same reasoning as employed above can also be used to prove that $F_{22}[(a\pi i/2), -\infty]$ is real. This establishes the validity of Bailey's quantization rule for the potential.

Case V:

$$q^2 = Q^2 - \frac{\pi^2}{4a^2\sin^2(\pi z/a)} = E + V_0 - \frac{a^2V_0 + \frac{1}{4}\pi^2}{a^2\sin^2(\pi z/a)}. \quad (3.9)$$

Before we begin analyzing this problem it is important to realize exactly what the physical situation is and what mathematical model we will be using to solve the problem. Physically we are dealing with a well-shaped $\cot^2$ potential whose walls are infinitely high at $x = 0$, $x = a$. We are not dealing with the periodic potential defined by letting x assume all real values. The boundary condition that this situation imposes on the Schrödinger equation is that $\psi(0) = \psi(a) = 0$. Mathematically when we set up the problem in order to use the apparatus of the $\hat{F}$-matrix approach, we consider the potential defined over the entire complex plane. In particular it will be periodic on the real axis. However, when we impose the boundary conditions on $\psi(x)$, i.e., $\psi(0) = \psi(a) = 0$, the problem for the region of physical interest $0 < x < a$ becomes mathematically identical to that of the infinite well. Thus the approach is justified.

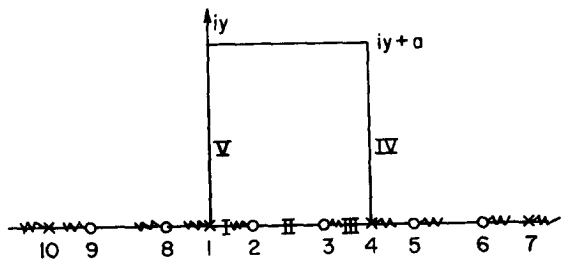


FIG. 4. Branch cuts in the complex plane for Case V.

Another consideration which arises for Case V, but was not necessary for Case IV, is the modification of the technique of Froman and Froman to deal with boundary conditions at $x = 0$ and $x = a$ instead of at $\pm\infty$. These difficulties are analogous to those treated in Chap. 11 of Ref. 5, when the radial problem is discussed. The resolution of the problem for the case presently under study completely parallels the solution outlined in this chapter. We assume $|e^{iw(x)}|$ increases monotonically to infinity as x goes to $+a$ from x'', or as x goes from x' to 0; the μ integral converges at 0 and a; and the phase of q is chosen properly. With these assumptions we are able to derive the analog of the one-dimensional and radial quantization conditions. In particular we have,

$$\int_{x'}^{x''} |q(x)|\,dx = (n + \tfrac{1}{2})\pi - \arg\frac{F_{22}(z, 0)}{F_{22}(z, a)}. \quad (3.10)$$

For the $\cot^2$ potential we can satisfy the above assumptions, and hence we can employ Eq. (3.10). In the following proof we choose $z = iy$ and show that in the limit $y \to +\infty$, both $F_{22}(iy, 0)$ and $F_{22}(iy, a)$ are real. Fig. 4 indicates the complex plane for our problem.

In Fig. 4 all angles to the left of and including point 2 have the range $-\pi < \theta < \pi$, while those to the right of point 2 have the range of $0 < \theta < 2\pi$. We find that the contributions from all the branch points other than 1, 2, 3, and 4 cancel each other out in groups of six for phases in regions I, II, and III. Thus, the contributions from points 5, 6, 8, and 9 is 2π, while that from 7 and 10 is -2π, and the total contribution is 0. In regions I, II, and III

$$q = |q|\,e^{i\theta}\exp\left[i\left\{\frac{\theta_2 + \theta_3 - 2(\theta_1 + \theta_4)}{2}\right\}\right].$$

We now find that $\theta_0 = \frac{1}{2}\pi$ so that in

$$\text{region I} \quad q = |q|\,e^{(i\pi/2)}\exp\left[i\left(\frac{2\pi - 2\pi}{2}\right)\right] = i|q|,$$

$$\text{region II} \quad q = |q|\,e^{(i\pi/2)}\exp\left[i\left(\frac{\pi - 2\pi}{2}\right)\right] = |q|,$$

$$\text{region III} \quad q = |q|\,e^{(i\pi/2)}\exp\left[i\left(\frac{-2\pi}{2}\right)\right] = -i|q|.$$

When we examine regions IV and V we see that only the branch points directly in each region contribute. For region IV $(\theta_1 + \theta_2 + \theta_3 + \theta_5 + \theta_6 + \theta_7) = 0$, while for region V $(\theta_2 + \theta_3 + \theta_4 + \theta_8 + \theta_9 + \theta_{10}) = 0$ and all other phases from branch points cancel out in sextuplets. In

$$\text{region IV} \quad q = e^{(i\pi/2)}e^{-(\pi i/2)}|q| = |q|,$$

$$\text{region V} \quad q = e^{(i\pi/2)}e^{-(i\pi/2)}|q| = |q|.$$

With the above phases established for q we see that $|e^{iw(z)}|$ increases monotonically as shown by the arrows in Fig. 5.

We now turn our attention to the properties of the μ integral. From Eqs. (3.9) and (2.8), we have

$$\epsilon = \frac{\pi^2}{[4a^2\sin^2(\pi z/a)]\left[E + V_0 - \left(\frac{a^2V_0 + \frac{1}{4}\pi^2}{a^2\sin^2(\pi z/a)}\right)\right]} + \frac{\pi^2\left[\frac{a^2V_0 + \frac{1}{4}\pi^2}{a^4\sin^4(\pi z/a)}\right]}{4\left[E + V_0 - \left(\frac{a^2V_0 + \frac{1}{4}\pi^2}{a^2\sin^2(\pi z/a)}\right)\right]} \times \left\{2(E + V_0)\left(3 - 2\sin^2\frac{\pi z}{a}\right) - \left[\frac{a^2V_0 + \frac{1}{4}\pi^2}{a^2\sin^2(\pi z/a)}\right][1 + \sin^2(\pi z/a)]\right\}. \quad (3.11)$$

In

$$\lim_{\substack{z\to 0\\ z\to a}} \epsilon q \to 0$$

so that $\mu(z_0, z)$ is convergent as z approaches a pole of q^2. For

$$z = x + iy, \quad y \to +\infty, \quad |\epsilon q| \propto \frac{1}{\sinh^2 y} \propto e^{-2y}$$

and hence the μ integral along a path λ such that $y \to +\infty$ for all points on the path, approaches 0.

We can now turn our attention to the quantity $F_{22}(iy, 0)/F_{22}(iy, a)$. From the above discussion concerning $|e^{iw(z)}|$, the μ integral, and Eq. (2.12) $F_{22}(iy, 0)$ exists. On lines IV and V

$$w = \int_p^z q\,dz = i\int_p^z |q|\,dy$$

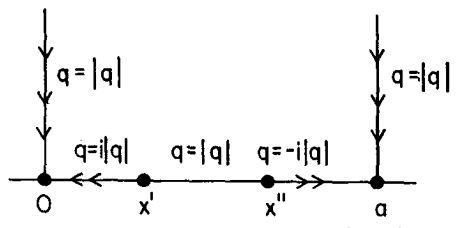


FIG. 5. Monotonic behavior of $|e^{iw(z)}|$.

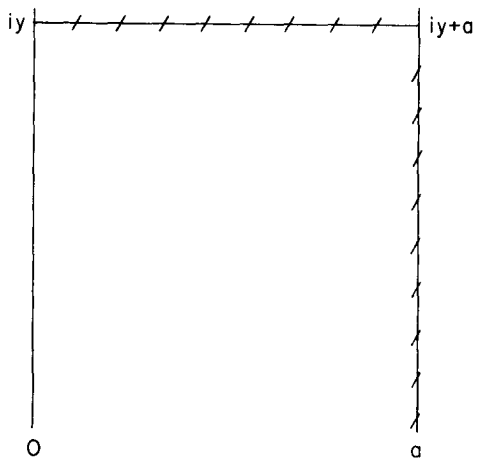


FIG. 6. The path λ used in evaluating $\hat{F}(iy, a)$.

is pure imaginary for p on line V (or IV) so that w is pure imaginary. Hence F_{22} connecting any two points on either line (both points must be on the same line, IV or V) is pure real for the proper choice of p. Since F_{22} is independent of the lower limit of the integral in (2.7) we can state that $F_{22}(iy, 0)$ is real.

By Eq. (2.5) with $z = iy$, $z_1 = iy + a$, $z_0 = a$; $F_{22}(iy, a)$ can be written

$$F_{22}(iy, a) = F_{21}(iy, iy + a)F_{12}(iy + a, a) + F_{22}(iy, iy + a)F_{22}(iy + a, a),$$

where our path connecting iy and a is shown in Fig. 6. The proof that $F_{22}(iy, a)$ is real closely resembles the proof that $F_{22}[(i\pi a/2), +\infty]$ is real. We will first show that $|F_{21}F_{12}| \to 0$ for $\lim y \to +\infty$ and then show that

$$F_{22}(iy, iy + a) \to 1.$$

Clearly (2.12b) is satisfied for $F_{12}(iy + a, a)$ and $|F_{12}|$ is finite. Since the line connecting $iy + a$ and iy is a finite segment along which $w(z)$, μ and $|e^{iw(z)}|$ are all well-behaved, the arguments employed for Case IV can be applied here to show the existence of an M and and a $w(z)$ such that Eq. (2.16c) holds and

$$|F_{21}(iy, iy + a)| \leq \frac{1}{2M}[e^{M\mu} - 1]|\exp[2iw'(iy + x')]|$$

We then find that

$$|F_{21}(iy, iy + a)F_{12}(iy + a, a)| \leq [e^{M\mu} - 1]C_1, \tag{3.12}$$

where C_1 is a finite constant. As we take the limit $y \to +\infty$ we find, since $\mu \to 0$ that $|F_{21}F_{12}| \to 0$. Similarly we obtain an estimate for $F_{22}(iy, iy + a)$ by using (2.16d),

$$|F_{22}(iy, iy + a) - 1| \leq [\exp(M\mu(iy, iy + a)) - 1]C_2,$$

where C_2 is a finite constant and thus

$$\lim_{y \to \infty} F_{22}(iy, iy + a) = 1.$$

Our final result is that

$$\lim_{y \to \infty} F_{22}(iy, a) = 0 + 1 \cdot F_{22}(iy + a, a) = F_{22}(iy + a, a). \tag{3.13}$$

But $F_{22}(iy + a, a)$ is real for the same reason that $F_{22}(iy, 0)$ is real, and so $F_{22}(iy, a)/F_{22}(iy, 0)$ is real and we have established an exact quantization condition.

Case VI:

$$q^2 = Q^2 - \frac{\alpha^2 e^{-\alpha r}}{4(1 - e^{-\alpha r})^2} = E + \frac{\lambda e^{-\alpha r}}{(1 - e^{-\alpha r})} - \frac{\alpha^2\left(\frac{b}{\alpha^2} + 1\right)e^{-\alpha r}}{(1 - e^{-\alpha r})^2}. \tag{3.14}$$

This problem must be treated as a radial problem with $r \geq 0$. The quantization condition for the radial problem is

$$\int_{r'}^{r''} |q|\, dr = (n + \tfrac{1}{2})\pi - \arg\frac{F_{22}(z, +\infty)}{F_{22}(z, +0)}, \tag{3.15}$$

we choose $z = -|x| + (\pi i/\alpha)$ and take the limit as $|x| \to \infty$. Poles of q^2 in the complex plane occur whenever $r = \pm(2n\pi i/\alpha)$ while the zeros are periodic with the same periodicity. In our standard notation and procedure we picture the complex z plane as shown in Fig. 7.

Again as we have already seen, all branch points symmetric with respect to the lines $y = 0$, or $y = (\pi i/\alpha)$ make no contribution to the phase of q. The z plane for this potential bears a closer resemblance to that for the sech^2 potential than for the cot^2 potential, and indeed we must have $\theta_0 = -\frac{1}{2}\pi$ as in Case

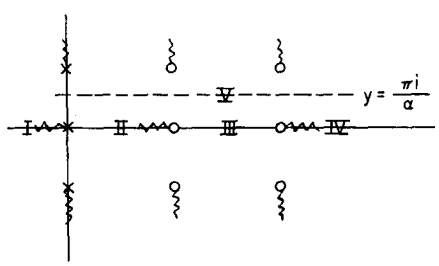


FIG. 7. Branch cuts in the complex plane for Case VI.

IV. The phase of q is now determined as follows:

region I

$$q = \exp\left(-\frac{i\pi}{2}\right)\exp\left[i\left(\frac{-2\pi+\pi+\pi}{2}\right)\right]|q| = -i|q|,$$

region II

$$q = \exp\left(-\frac{i\pi}{2}\right)\exp\left[i\left(\frac{-0+\pi+\pi}{2}\right)\right]|q| = i|q|,$$

region III

$$q = \exp\left(-\frac{i\pi}{2}\right)\exp\left[i\left(\frac{-0+0+\pi}{2}\right)\right]|q| = |q|,$$

region IV

$$q = \exp\left(-\frac{i\pi}{2}\right)\exp\left[i\left(\frac{0+0+0}{2}\right)\right]|q| = -i|q|,$$

region V

$$q = \exp\left(-\frac{i\pi}{2}\right)\exp\left[i(0)\right]|q| = -i|q|.$$

The directions in which $|e^{iw(z)}|$ increases monotonically are indicated in Fig. 8.

Now that the phase of q has been established we check to see if the properties of the μ integral are the desired ones.

From Eqs. (2.8) and (3.14) we find that

$$|\epsilon q| \to \frac{\alpha e^{-\alpha r}}{2(4b/\alpha^2+1)^{\frac{1}{2}}}$$

as $r \to 0$ so that μ is convergent when integrated over a region including the origin. We also find that $|\epsilon q| \to 0$ exponentially, as fast as $r \to +\infty$ so that we again have $\mu \to 0$ over a path λ which is entirely in a region of infinitely large $|r|$.

With these preliminaries out of the way we are now able to outline the desired proof. In order to prove that $F_{22}[(i\pi/\alpha) - |x|, +\infty]$ is real we notice that the conditions are equivalent to that for Case IV when we proved that $F_{22}(i\pi a/2, +\infty)$ was real. In fact our task is easier since there are no poles present on the

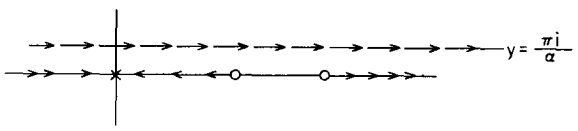


FIG. 8. Monotonic behavior of $|e^{iw(z)}|$ for Case VI.

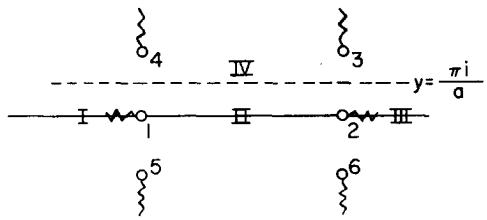


FIG. 9. Branch cuts in the complex plane for Case VII.

path we now consider. To prove that $F_{22}[-|x| + (i\pi/\alpha), 0]$ is real we essentially reproduce the proof used to show that $F_{22}(iy, a)$ was real in Case V. In the present situation the reality of $F_{22}(-|x|, 0)$ is established slightly differently, however, by making use of the proof that $F_{22}[(i\pi a/2), (i\pi a/2) + x]$ is real for Case IV.

Case VII:

$$q^2(z) = Q^2(z) = E - Ae^{-2az} - Be^{-az}. \quad (3.16)$$

We notice that there are no poles in the finite z plane and that $|\epsilon q| \to e^{-|x|a}$ as $|x| \to \infty$ $(z = x + iy)$. Hence $\mu(z, z_0) \to 0$ for a path λ with $|x| \to \infty$ for all $z = x + iy$ on λ. The zeros of q^2 are periodic with periodicity $2n\pi i/a$. Figure 9 is the branch point diagram for this problem. The convention on the range of angles is identical to Case IV, Fig. 2. In fact the situation here is totally analogous to that depicted in the latter diagram.

We find that θ_0 is again $-\frac{1}{2}\pi$ and in

$$\begin{aligned} \text{region I} \quad & q = i|q|, \\ \text{II} \quad & q = |q|, \\ \text{III} \quad & q = i|q|. \end{aligned}$$

Since line IV is symmetrically situated between the periodic zeros the phase contribution from the branch points is 0 and region IV $q = -i|q|$. The direction in which $|e^{iw}|$ increases monotonically is shown in the diagram of Fig. 10.

To prove that our quantization rule is exact we almost exactly reproduce the proof employed for Case IV. We choose as our z in the quantization condition (2.10) the point $z = -|x| + (i\pi/a)$ and must show that

$$\frac{F_{22}\left(-|x| + \frac{i\pi}{a}, +\infty\right)}{F_{22}\left(-|x| + \frac{i\pi}{a}, -\infty\right)}$$

is real. Because of the similarity in phase of q for line IV Fig. 9 and line IV Fig. 2 the proof that $F_{22}[(i\pi a/2), +\infty]$ is real can be exactly reproduced to establish

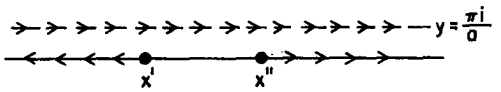


FIG. 10. Monotonic behavior of $|e^{iw(x)}|$ for Case VII.

that

$$F_{22}\left(-|x| + \frac{i\pi}{a}, +\infty\right)$$

is real. Again the task is made easier, since it is clear that all elements considered exist. It will be recalled that in order to prove that $F_{22}[(i\pi a/2), -\infty]$ was real in Case IV it was necessary to prove that

$$F_{22}\left(-|x| + \frac{i\pi a}{2}, -\infty\right)$$

was equal to 1 in $\lim x \to \infty$. This proof also serves to prove that

$$F_{22}\left(-|x| + \frac{i\pi}{a}, -\infty\right),$$

for the situation considered, is equal to 1. Hence

$$\lim_{x\to\infty} \frac{F_{22}\left(-|x| + \frac{i\pi}{a}, +\infty\right)}{F_{22}\left(-|x| + \frac{i\pi}{a}, -\infty\right)} = F_{22}\left(-|x| + \frac{i\pi}{a}, +\infty\right)$$

and

$$F_{22}\left(-|x| + \frac{i\pi}{a}, +\infty\right)$$

is real. The exact quantization condition for a modified Morse oscillator has now been proved.

Case VIII:

$$q^2(x) = Q^2(x) = E - Ae^{2\alpha a} - Be^{-2\alpha a}. \quad (3.17)$$

If we locate the zeros of q^2 we find that they are periodic with a periodicity of $\pm(n\pi i/a)$ in the complex plane. Since q^2 has no poles in the z plane the branch-point diagram for this case bears an extremely striking resemblance to Fig. 9. However, now there are twice as many zeros and instead of choosing line IV to be $y = \pi i/a$ we choose $y = \pi i/2a$. Since $|\epsilon q| \to e^{-a|x|}$ as $x \to \pm\infty$ the proof that quantization condition (2.3) is exact for Case VIII is identical to that of Case VII.

In order to provide an independent check on the validity of the exact quantization condition for Case VIII a numerical integration was performed of the quantization integral. The problem was made dimensionless by choosing $[ha/(2mA)^{\frac{1}{2}}] = 1$. The special case of $A = B$ (this corresponds to a $\sinh^2$ potential) was chosen and the energy value which satisfied the quantization equation for the ground state ($n = 0$) was found to be 2.32. The ground state, it is well known, provides the most severe test of the WKB quantization condition since for nonexact conditions the WKB eigenvalue is markedly different than the true eigenvalue for this state, even if the two types of energy values agree closely for excited states.[1] To check this value of E a numerical integration of Schrödinger's equation was performed. The procedure we employed was to find an E for which the corresponding wavefunction closely approached the x axis and then began to blow up rapidly, without ever having actually crossed the axis. We then found a value for the energy for which the wavefunction crossed the axis before it diverged. The true eigenvalue must lie between these two values of E, and our value of (2.32) has this property. The result then, confirms the validity of the exactness of our quantization condition.

We have now completed our proof that all the potentials listed in Table 1 have as an exact quantization condition Eq. (2.3).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Some Total Invariants of Asymptotically Flat Space-Times

J. WINICOUR
Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio

(Received 16 January 1967)

The total energy, momentum, supermomentum, and angular momentum of asymptotically flat space-times are calculated in terms of coordinate and conformally invariant expressions by taking the limit in an invariant way of the asymptotic symmetry linkages through a sequence of finite closed two-spaces which converge to a sphere at null infinity. The resulting expressions consist of integrals over the sphere at null infinity of cordinate and conformally invariant quantities. In the case of energy and momentum these integrals may be reduced to expressions previously proposed by Penrose.

1. INTRODUCTION

The importance of the Poincaré group is that it manifests the full group of motions leaving the Minkowski metric of special relativity invariant. Because these motions constitute a preferred group of symmetry transformations, their generators, energy, momentum, and angular momentum assume special significance. The infinitesimal motions of the Poincaré group are usually presented as infinitesimal transformations

$$y'^a = y^a + \xi^a\epsilon + O(\epsilon^2)$$

which leave invariant the Minkowski metric

$$\eta^{ab} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.1}$$

so that

$$y'^a{}_{,c}y'^b{}_{,d}\eta^{cd} = \eta^{ab}$$

and

$$\eta^{ac}\xi^b{}_{,c} + \eta^{bc}\xi^a{}_{,c} = 0. \tag{1.2}$$

The ten linearly independent solutions $\xi_Q{}^a$ of Eq. (1.2) constitute descriptors of infinitesimal Poincaré transformations. An observer referring to a particular Minkowski coordinate system[1] y^a associates the ten values of the label Q with his coordinate axes. Four descriptors with $Q = a$ $(a = 0, \cdots, 3)$ are associated with infinitesimal translations along the four coordinate axes y^a. Six descriptors with bivector labels $Q = [ab]$ are associated with infinitesimal Lorentz rotations in the $[y^a, y^b]$ plane. Observers referring to other Minkowski coordinate systems make the same canonical labeling of the descriptors *with respect to their own coordinate axes.* We write $\xi_{Q'}{}^{a''}(y'')$ to denote the descriptor of the Q transformation associated with the y' Minkowski frame and expressed in terms of some other y'' coordinate system. Here the y'' coordinate system is arbitrary. Indeed, we need not even restrict ourselves to Minkowski frames. The descriptors transform as vectors, so that under the curvilinear coordinate transformation $x^\mu = x^\mu(y'')$ we have

$$\xi_{Q'}{}^\mu(x) = x^\mu{}_{,a''}\xi_{Q'}{}^{a''}(y'').$$

Under such a transformation the Minkowski metric is no longer invariant and transforms into

$$g^{\mu\nu}(x) = x^\mu{}_{,a''}x^\nu{}_{,b''}\eta^{ab}.$$

Consequently, Eq. (1.2) does not hold in curvilinear frames, but as is well known its covariant form

$$2\xi_Q{}^{(\mu;\nu)} = g^{\mu\beta}\xi_Q{}^\nu{}_{,\beta} + g^{\nu\beta}\xi_Q{}^\mu{}_{,\beta} - g^{\mu\nu}{}_{,\beta}\xi_Q{}^\beta = 0 \tag{1.3}$$

is applicable. Equation (1.3) gives a general covariant meaning to the concept of a motion or symmetry of Minkowski space.

Let y and y' denote two Minkowski frames which differ infinitesimally,[2]

$$y'^a = y^a + \xi_Q{}^a\epsilon^Q + O(\epsilon^2). \tag{1.4}$$

Then in any coordinate system, the descriptors $\xi_Q{}^\mu(x)$ and $\xi_{Q'}{}^\mu(x)$ associated with these two Minkowski frames are related to first order

$$\xi_{Q'}{}^\mu(x) = \xi_Q{}^\mu(x) + C_{PQ}{}^R\xi_R{}^\mu(x)\epsilon^P, \tag{1.5}$$

where the $C_{PQ}{}^R$ are the structure constants of the Poincaré group. Suppose some linear functional L acts on the descriptors such that

$$L_Q \equiv L[\xi_Q{}^\mu]$$

is a scalar under curvilinear coordinate transformations. For example, if σ is a spacelike hypersurface and $T_\mu{}^\nu$ the energy-momentum tensor of some field then the scalar linear functional

$$L_Q(\sigma) = \int_\sigma \xi_Q{}^\mu T_\mu{}^\nu\, dS_\nu$$

[1] In a Minkowski coordinate system the metric has the form (1.1).

[2] Sum over the repeated index Q.

describes the total energy, momentum, and angular momentum of σ. From Eq. (1.5) we then have

$$L_{Q'} = L_Q + C_{PQ}{}^R L_R \epsilon^P. \tag{1.6}$$

Equation (1.6) describes the transformation of the L_Q induced by a transformation of the parameter space of the Poincaré group. Physically, this transformation law relates how the scalars L_Q are interpreted by observers referring to different Minkowski frames.

The foregoing discussion of the symmetries of Minkowski space while offering nothing new or original provides a familiar basis for understanding some recent results concerning symmetries in general relativity.[3–9] In a curved space global symmetries satisfying Eq. (1.3) do not in general exist. For asymptotically flat spaces there are, however, asymptotic symmetries which satisfy Eq. (1.3) at null infinity (in the limit of infinite luminosity distances along null hypersurfaces). Penrose[6,7] has devised a technique for studying null infinity in a covariant manner. He constructs a manifold conformal to the physical manifold in which null infinity J is a regular hypersurface consisting of two disjoint regions, future null infinity J^+ and past null infinity J^-, each having topology $S^2 \times E^1$. The possibility of such a construction is tantamount to the necessary conditions for space to be asymptotically flat. In this formalism, the physical-space metric now denoted by $\tilde{g}_{\mu\nu}$ is related to the metric $g_{\mu\nu}$ of the conformal manifold by

$$\tilde{g}_{\mu\nu} = \Omega^{-2} g_{\mu\nu}, \tag{1.7}$$

where $\Omega = 0$ at J. Asymptotic symmetries can then be defined in terms of conformal motions of J.[4–9] This leads to the Bondi–Metzner–Sachs (BMS) asymptotic symmetry group. There is obviously a double-valued nature to this group with respect to symmetries of both J^+ and J^-. We will not delve into this subject here and will restrict ourselves to considering symmetries of J^+. These may be expressed as infinitesimal conformal motions

$$x'^\mu = x^\mu + \xi^\mu \epsilon + O(\epsilon^2),$$

satisfying

$$\xi^{(\mu;\nu)} - \tfrac{1}{4} g^{\mu\nu} \xi^\beta{}_{;\beta} = O(J^+), \tag{1.8}$$

where $O(J^+)$ indicates terms which vanish on J^+. The transformations such that

$$\xi^\mu = O(J^+) \tag{1.9}$$

form an invariant subgroup. Their factor group is the BMS group.[9] Unlike the Minkowski-space solutions of Eq. (1.3), BMS descriptors satisfying Eq. (1.6) are defined only to within terms of the type indicated in Eq. (1.9). Given a null hypersurface Γ, however, a projection of Eq. (1.3) into Γ uniquely propagates BMS descriptors along Γ in terms of their values on the sphere Σ^+ in which Γ intersects J^+.[8,9] In terms of covariant derivatives $\tilde{\nabla}_\mu$ with respect to the physical-space metric, this propagation law takes the form

$$[\tilde{k}_\nu \tilde{\nabla}^{(\nu} \xi^{\mu)} - \tfrac{1}{2} \tilde{k}^\mu \tilde{\nabla}_\nu \xi^\nu]_\Gamma = 0, \tag{1.10}$$

where $\tilde{k}_\mu$ is the null normal to Γ. We can thus avoid the lack of uniqueness expressed by Eq. (1.9) by restricting our investigations to a single null hypersurface Γ and imposing the propagation law Eq. (1.10). For the remainder of the present paper this will be done. *We deal only with a single null hypersurface* Γ.

There are an infinite number of linearly independent descriptors $\xi_Q{}^\mu$ which satisfy Eq. (1.8) on J^+ and Eq. (1.10) on Γ. In the identical way that we have used Minkowski-coordinate systems to canonically label the descriptors of the Poincaré group, Sachs[4] has used null polar-coordinate systems on J^+ to canonically label the descriptors of the BMS group. Ten BMS descriptors are given labels $Q = a$ and $Q = [ab]$ analogous to the descriptors of the Poincaré group. The remaining BMS descriptors correspond to what have been called supertranslations[4] and are given the spherical-harmonic labels $Q = (lm)$ with $l \geq 2$. Just as observers referring to different Minkowski frames make the same canonical labeling of the Poincaré descriptors but with respect to their own coordinate axes, observers at null infinity referring to different (but isometric) null polar-coordinate systems on J^+ make the same canonical labeling of BMS descriptors but with respect to their own axes. In analogy with Eq. (1.5), BMS descriptors $\xi_Q{}^\mu$ and $\xi_{Q'}{}^\mu$ satisfying Eq. (1.10) which are associated with two infinitesimally differing null polar-coordinate systems on J^+,

$$y'^a = y^a + \xi_P{}^a \epsilon^P + O(\epsilon^2),$$

are related on Γ by

$$\xi_{Q'}{}^\mu(x) = \xi_Q{}^\mu(x) + C_{PQ}{}^R \xi_R{}^\mu(x) \epsilon^P, \tag{1.11}$$

where the $C_{PQ}{}^R$ are now the structure constants of the BMS group.[9]

Let $\tilde{n}^\mu$ be any vector field on Γ such that

$$\tilde{k}_\mu \tilde{n}^\mu = -1. \tag{1.12}$$

[3] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[4] R. K. Sachs, Phys. Rev. **128**, 2851 (1962).
[5] R. K. Sachs, in *Relativity, Groups, and Topology* (Gordon and Breach, Science Publishers, Inc., New York, 1964), p. 523.
[6] R. Penrose, Phys. Rev. Letters **10**, 66 (1963).
[7] R. Penrose, in *Relativity, Groups, and Topology* (Gordon and Breach, Science Publishers, Inc., New York, 1964), p. 565.
[8] J. Winicour and L. Tamburino, Phys. Rev. Letters **15**, 601 (1965).
[9] L. Tamburino and J. Winicour, Phys. Rev. **150**, 1039 (1966).

Then the scalar linear functional

$$L_Q(\Sigma) = \oint_\Sigma (\tilde{\nabla}^{[\nu}\xi_Q{}^{\mu]} - \tilde{k}^{[\mu}\tilde{n}^{\nu]}\tilde{\nabla}_\beta\xi^\beta)\, d\tilde{S}_{\mu\nu} \quad (1.13)$$

is defined for each finite closed two-space Σ lying on Γ.[8,9] Furthermore, this functional is independent of the choice of $\tilde{n}^\mu$ subject to Eq. (1.12). Corresponding to different values Q, this functional has been called the energy–momentum, angular-momentum, and supermomentum linkage through Σ.[8,9] Under the transformation of Eq. (1.11), these linkages transform as an adjoint representation of the BMS group,

$$L_{Q'}(\Sigma) = L_Q(\Sigma) + C_{P'Q}{}^R L_R(\Sigma)\epsilon^P. \quad (1.14)$$

This transformation law has the same physical interpretation (with respect to observers at null infinity) as Eq. (1.6). The most important feature of Eq. (1.14) is that the structure of the BMS group allows the identification of the energy–momentum linkage up to a Lorentz rotation.

In terms of a limiting process, the functional $L_Q(\Sigma^+)$ is also well defined.[8,9] It yields the proper value for the total mass of an asymptotically flat space, as defined by Bondi *et al.*[3] In Ref. 9, the limit $L_Q(\Sigma^+)$ was explicitly calculated in terms of a conformal Bondi coordinate system. The purpose of this paper is to calculate the limit in a covariant manner and to present $L_Q(\Sigma^+)$ as an integral over Σ^+ of coordinate and conformally invariant quantities. We strictly adhere to the notation of Ref. 9.

2. CONFORMAL SPACE FORMULAS

Equation (1.7) defines a conformal-space metric which is regular at J^+. The conformal factor Ω is also regular at J^+ and is subject to the conditions[6,7]

$$\Omega = O(J^+), \quad \Omega_{,\mu} \neq O(J^+). \quad (2.1)$$

In terms of conformal-space quantities, the vacuum Einstein equations are

$$\Omega^2 S_{\mu\nu} - 2\Omega\Omega_{;\mu\nu} + g_{\mu\nu}\Omega^{;\beta}\Omega_{;\beta} = 0, \quad (2.2)$$

where

$$S_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{6}g_{\mu\nu}R. \quad (2.3)$$

These equations and the topology of J^+ imply[6,7]

$$\Omega^{;\beta}\Omega_{;\beta} = O(J^+), \quad (2.4)$$

$$\Omega_{;\mu\nu} - \tfrac{1}{4}g_{\mu\nu}\Omega^{;\beta}{}_{\beta} = O(J^+), \quad (2.5)$$

and

$$C_{\mu\alpha\beta\nu} = O(J^+), \quad (2.6)$$

where our conventions for relating the Weyl, Riemann, and Ricci tensors are

$$C_{\mu\alpha\beta\nu} = R_{\mu\alpha\beta\nu} + g_{\beta[\mu}S_{\alpha]\nu} + g_{\nu[\alpha}S_{\mu]\beta} \quad (2.7)$$

and

$$R_{\mu\nu} = R^\alpha{}_{\mu\nu\alpha}.$$

We assign the conformal-transformation properties

$$\tilde{k}_\mu = k_\mu, \quad \tilde{n}^\mu = n^\mu \quad (2.8)$$

so as to preserve the normalization given in Eq. (1.12). The extensions of k_μ and n_μ are further restricted so that these vectors are regular at Σ^+. The surface element transforms into

$$d\tilde{S}_{\mu\nu} = \Omega^{-4}\, dS_{\mu\nu}. \quad (2.9)$$

The propagation law along Γ prescribed in Eq. (1.10) becomes

$$\xi^{(\mu;\nu)}k_\nu = \tfrac{1}{2}\xi^\beta{}_{;\beta}k^\mu - \Omega^{-1}\Omega_{;\beta}\xi^\beta k^\mu. \quad (2.10)$$

We immediately have the regularity condition

$$\Omega_{;\beta}\xi^\beta = O(J^+), \quad (2.11)$$

which states that asymptotically the descriptors must point along J^+. Finally, the linkage expression Eq. (1.13) becomes

$$L_\xi(\Sigma) = \oint_\Sigma \Omega^{-3}A^{\mu\nu}\, dS_{\mu\nu}, \quad (2.12)$$

where

$$A^{\mu\nu} = \Omega\xi^{[\mu;\nu]} - 2\xi^{[\mu}\Omega^{;\nu]} - \Omega\xi^\beta{}_{;\beta}k^{[\mu}n^{\nu]} + 4\xi^\beta\Omega_{;\beta}k^{[\mu}n^{\nu]}. \quad (2.13)$$

Equations (2.12) and (2.13) imply that the limit $L_\xi(\Sigma^+)$ is an integral of an indefinite form. In Ref. 9, however, it was proved that

$$(\Omega^{-3}A^{\mu\nu})_{;\nu}k_\mu = O(J^+).$$

With the help of the generalized Gauss theorem, this was then used to show that the limit

$$L_\xi(\Sigma^+) \equiv \lim_{\Sigma\to\Sigma^+} L_\xi(\Sigma)$$

exists and is furthermore independent of the choice of closed two-spaces Σ which converge to Σ^+ along Γ. We choose the particular family $\Sigma(\Omega)$ given by the intersections with Γ of the hypersurfaces $\Omega = \text{const}$.

The family $\Sigma(\Omega)$ defines a null tetrad on Γ (uniquely to within choice of extensions) by the equations

$$g_{\mu\nu} = -2k_{(\mu}m_{\nu)} + 2t_{(\mu}\bar{f}_{\nu)}, \quad (2.14)$$

with

$$\Omega_{;\mu}t^\mu = 0 \quad (2.15)$$

and

$$m_\mu = \lambda^{-1}\Omega_{;\mu} + \tfrac{1}{2}\lambda^{-2}\Omega^{;\beta}\Omega_{;\beta}k_\mu, \quad (2.16)$$

where

$$\lambda \equiv -\Omega_{;\mu}k^\mu. \quad (2.17)$$

The usual orthonormality conditions follow:

$$k_\mu k^\mu = k_\mu t^\mu = m_\mu m^\mu = m_\mu t^\mu = t_\mu t^\mu = 0, \quad -k_\mu m^\mu = t_\mu \bar{f}^\mu = 1. \quad (2.18)$$

We rewrite the surface element on $\Sigma(\Omega)$ as

$$dS_{\mu\nu} = k_{[\mu}m_{\nu]}\,dS, \tag{2.19}$$

where

$$dS = -2k^{[\mu}m^{\nu]}\,dS_{\mu\nu}. \tag{2.20}$$

Equations (2.12) and (2.13) then give

$$L_\xi(\Sigma^+) = \lim_{\Omega\to 0}\Omega^{-3}\oint_{\Sigma(\Omega)} A\,dS, \tag{2.21}$$

where

$$A = -\Omega\xi^{\mu;\nu}m_\mu k_\nu + \Omega\xi^{\mu;\nu}t_{(\mu}\bar{t}_{\nu)} - 3\xi^\mu\Omega_{;\mu} - \lambda^{-1}\xi^\mu k_\mu\Omega^{;\nu}\Omega_{;\nu}. \tag{2.22}$$

From this it is clear that $L_\xi(\Sigma)$ involves derivatives of ξ^μ only in directions lying in Γ.

In evaluating Eq. (2.21) we will use the following two formulas:

$$\frac{d}{d\Omega}\oint_{\Sigma(\Omega)} B\,dS = \oint_{\Sigma(\Omega)}[B_{,\nu}k^\nu + 2\rho B]\lambda^{-1}\,dS, \tag{2.23}$$

$$\oint_{\Sigma(\Omega)}[B^{\mu;\nu}t_{(\mu}\bar{t}_{\nu)} + \rho B^\mu m_\mu + m_{\alpha;\beta}t^{(\alpha}\bar{t}^{\beta)}B^\mu k_\mu]\,dS = 0, \tag{2.24}$$

where ρ measures the divergence of Γ:

$$\rho \equiv k_{\mu;\nu}t^{(\mu}\bar{t}^{\nu)}. \tag{2.25}$$

These equations are applications of the generalized Gauss theorem in a four-dimensional manifold

$$\oint_R B^{[\alpha\cdots\mu\nu]}{}_{;\nu}\,dS_{\alpha\cdots\mu} = (4-n)^{-1}\oint_{\partial R} B^{[\alpha\cdots\mu\nu]}\,dS_{\alpha\cdots\mu\nu},$$

where n equals the dimension of the boundary ∂R. To prove Eq. (2.23), write

$$\begin{aligned}\frac{d}{d\Omega}\oint_{\Sigma(\Omega)} B\,dS &= -2\lim(\Delta\Omega)^{-1}\left(\oint_{\Sigma(\Omega)} Bk^{[\mu}m^{\nu]}\,dS_{\mu\nu}\right)\Bigg|_{\Omega}^{\Omega+\Delta\Omega}\\ &= -4\lim(\Delta\Omega)^{-1}\oint_{\Sigma(\Omega)}^{\Sigma(\Omega+\Delta\Omega)}(Bk^{[\mu}m^{\nu]})_{;\nu}\,dS_\mu\\ &= -4\lim(\Delta\Omega)^{-1}\oint_{\Sigma(\Omega)}^{\Sigma(\Omega+\Delta\Omega)}(Bk^{[\mu}m^{\nu]})_{;\nu}\,dS_{\mu\nu}\,d_\Omega x^\nu,\end{aligned}$$

where

$$d\Omega = \Omega_{;\nu}\,d_\Omega x^\nu.$$

Performing a tetrad expansion then gives Eq. (2.23). Equation (2.24) is a generalization of the theorem that the integral of a curl over a closed two-dimensional surface vanishes. It follows from first applying the topological equation

$$\partial\Sigma(\Omega) = 0$$

to prove

$$\oint_{\Sigma(\Omega)}(B^{[\alpha}m^\mu k^{\nu]})_{;\nu}\,dS_{\alpha\mu} = 0$$

and then performing a tetrad expansion.

3. CALCULATION OF THE LIMIT

To evaluate the limit $L_\xi(\Sigma^+)$ we introduce a conformal frame which simplifies the description of Σ and Γ. There are two motivations for this. One is calculational ease. The other is the gained simplicity in recognizing the conformal invariance of the integrand appearing in the limit. The reason that introducing a special conformal frame is useful in this latter respect is that the initial integrand A is not conformally invariant. Furthermore, the use of Eqs. (2.23) and (2.24) in the limiting process introduces factors of ρ which also is not a conformal invariant. Consequently, it is helpful to transform away as many nonconformally invariant terms as possible. The justification for this procedure is given in the next section by showing that the surviving terms can be interpreted as conformal invariants.

We set

$$\rho = 0, \tag{3.1}$$

by choosing a conformal factor satisfying

$$\Omega_{,\nu}\tilde{k}^\nu = -\Omega\tilde{\rho}$$

along Γ. This defines Ω^{-1} to be a luminosity distance along Γ.[5] Furthermore, we use the angular freedom in the choice of inverse luminosity distance to set

$$\lambda = 1 + O(\Sigma^+). \tag{3.2}$$

This fixes Ω on Γ. To complete the specification of Ω we choose the hypersurfaces $\Omega = \text{const}$ to be the family of incoming null hypersurfaces intersecting Γ in $\Sigma(\Omega)$, so that

$$\Omega^{;\alpha}\Omega_{;\alpha} = 0. \tag{3.3}$$

In addition, we use the freedom in the extension of k_μ and the hypersurface orthogonality of k_μ to demand

$$k_{\mu;\nu}k^\nu = k_{[\mu;\nu]}t^\mu\bar{t}^\nu = 0. \tag{3.4}$$

In the next section we will show that our results do not depend upon this special choice of k_μ.

Even with the above simplifications, our calculation is still tedious. Taking the tetrad components of the propagation law Eq. (2.10) gives

$$\xi^{\mu;\nu}k_\mu k_\nu = 0, \tag{3.5}$$

$$\xi^{(\mu;\nu)}k_\mu t_\nu = 0, \tag{3.6}$$

$$\Omega\xi^{\mu;\nu}t_{(\mu}\bar{t}_{\nu)} = \Omega_{;\beta}\xi^\beta. \tag{3.7}$$

Combining Eqs. (2.22), (3.3), and (3.7) then gives

$$A = \Omega B, \tag{3.8}$$

where

$$B = -\lambda^{-1}\xi^{\mu;\nu}\Omega_{;\mu}k_\nu - 2\xi^{\mu;\nu}t_{(\mu}\bar{t}_{\nu)}. \tag{3.9}$$

From Eqs. (3.7), (3.3), and (2.3), we have

$$\xi^{\mu;\nu}\Omega_{;\mu}k_\nu = -\lambda\xi^{\mu;\nu}t_{(\mu}\bar{t}_{\nu)} - \tfrac{1}{2}\Omega S_{\mu\nu}k^\mu\xi^\nu + \Omega[\xi^{\alpha;\beta}t_{(\alpha}\bar{t}_{\beta)}]_{;\nu}k^\nu,$$

and from Eqs. (2.24), (3.1), (2.15), (2.16), (3.3), and (2.3), we have

$$\xi^{\mu;\nu}t_{(\mu}\bar{t}_{\nu)} = -\tfrac{1}{2}\Omega\lambda^{-1}\xi^\mu k_\mu S_{\alpha\beta}t^\alpha\bar{t}^\beta + C,$$

where C denotes terms whose integral over $\Sigma(\Omega)$ vanishes because of the curl theorem, Eq. (2.24). Combining these results with Eq. (3.9) gives

$$B = \Omega D + C, \tag{3.10}$$

where

$$D = \tfrac{1}{2}\lambda^{-1}S_{\mu\nu}k^\mu\xi^\nu + \tfrac{1}{2}\lambda^{-1}\xi^\mu k_\mu S_{\alpha\beta}t^\alpha\bar{t}^\beta - \lambda^{-1}[\xi^{\alpha;\beta}t_{(\alpha}\bar{t}_{\beta)}]_{;\nu}k^\nu. \tag{3.11}$$

Using Eqs. (2.18), (3.5), and (3.6), we find

$$[t_{(\alpha}\bar{t}_{\beta)}]_{;\nu}k^\nu\xi^{\alpha;\beta} = 0.$$

From Eqs. (2.2) and (3.3), we have

$$\Omega^{;\mu}S_{\mu\nu} = 0. \tag{3.12}$$

Using Eqs. (3.12) and (2.3) gives

$$S_{\mu\nu}k^\mu\xi^\nu + \xi^\mu k_\mu S_{\alpha\beta}t^\alpha\bar{t}^\beta = R_{\mu\nu}k^\mu\xi^\nu.$$

Combining these results with Eq. (3.11) then gives

$$D = -\lambda^{-1}\xi^{\alpha;\beta\nu}t_{(\alpha}\bar{t}_{\beta)}k_\nu + \tfrac{1}{2}\lambda^{-1}R_{\mu\nu}k^\mu\xi^\nu. \tag{3.13}$$

From the Ricci identities, we have

$$\lambda^{-1}\xi^{\alpha;\beta\nu}t_{(\alpha}\bar{t}_{\beta)}k_\nu = \lambda^{-1}\xi^\mu R_\mu{}^{\alpha\beta\nu}t_{(\alpha}\bar{t}_{\beta)}k_\nu + \lambda^{-1}\xi^{\alpha;\nu\beta}t_{(\alpha}\bar{t}_{\beta)}k_\nu.$$

Using Eqs. (2.7), (2.24), (3.1), and (3.5) then gives

$$\lambda^{-1}\xi^{\alpha;\beta\nu}t_{(\alpha}\bar{t}_{\beta)}k_\nu = \lambda^{-1}\xi^\mu C_\mu{}^{\alpha\beta\nu}t_{(\alpha}\bar{t}_{\beta)}k_\nu - \tfrac{1}{2}\lambda^{-1}\xi^\alpha t_{(\alpha}\bar{t}_{\beta)}R^{\beta\nu}k_\nu + \tfrac{1}{2}\lambda^{-1}\xi^\mu R_\mu{}^\nu k_\nu - \xi^{\alpha;\nu}t_{(\alpha}\bar{t}_{\beta)}(\lambda^{-1}k_\nu)^{;\beta} + C,$$

where C denotes additional curl terms whose integral over $\Sigma(\Omega)$ vanishes. Combining this with Eq. (3.13) gives

$$D = -\lambda^{-1}\xi^\mu C_\mu{}^{\alpha\beta\nu}t_{(\alpha}\bar{t}_{\beta)}k_\nu + \tfrac{1}{2}\lambda^{-1}\xi^\alpha t_{(\alpha}\bar{t}_{\beta)}R^{\beta\nu}k_\nu + \xi^{\alpha;\nu}t_{(\alpha}\bar{t}_{\beta)}(\lambda^{-1}k_\nu)^{;\beta} + C. \tag{3.14}$$

Next from Eqs. (2.17), (2.18), (2.16), (3.3), (2.2), and (3.1), we find

$$\xi^{\alpha;\nu}t_{(\alpha}\bar{t}_{\beta)}(\lambda^{-1}k_\nu)^{;\beta} = \mathrm{Re}\,[\tfrac{1}{2}\Omega\lambda^{-2}\xi^{\alpha;\nu}t_\alpha\bar{t}_\beta R_\mu{}^\beta k^\mu k_\nu + \lambda^{-1}\sigma\xi^{\alpha;\beta}\bar{t}_\alpha\bar{t}_\beta],$$

where Re indicates the real part, and

$$\sigma \equiv k_{\mu;\nu}t^\mu t^\nu \tag{3.15}$$

represents the shear of Γ. Combining these results with Eqs. (3.14), (3.10), (3.8), and (2.21) gives

$$L_\xi(\Sigma^+) = \lim_{\Omega\to 0}\Omega^{-1}\oint_{\Sigma(\Omega)} E\,dS, \tag{3.16}$$

where

$$E = \mathrm{Re}\,[-\lambda^{-1}\xi^\mu C_\mu{}^{\alpha\beta\nu}t_\alpha\bar{t}_\beta k_\nu + \tfrac{1}{2}\lambda^{-1}\xi^\alpha t_\alpha R^{\beta\nu}\bar{t}_\beta k_\nu + \lambda^{-1}\sigma\xi^{\alpha;\beta}\bar{t}_\alpha\bar{t}_\beta + \tfrac{1}{2}\Omega\lambda^{-2}\xi^{\alpha;\nu}t_\alpha k_\nu R^{\mu\beta}\bar{t}_\beta k_\mu], \tag{3.17}$$

or by Eq. (2.23)

$$L_\xi(\Sigma^+) = \oint_{\Sigma^+} G\,dS, \tag{3.18}$$

where

$$G = E_{,\nu}k^\nu.$$

Using the asymptotic properties indicated in Eqs. (2.6), (1.8), and (3.2) gives

$$G = \mathrm{Re}\,[-\xi^\nu C_\nu{}^{\alpha\beta\nu;\mu}t_\alpha\bar{t}_\beta k_\nu k_\mu + \sigma(\xi^{\alpha;\beta}\bar{t}_\alpha\bar{t}_\beta)_{;\nu}k^\nu + \tfrac{1}{2}\xi^\alpha t_\alpha\bar{t}_\beta R^{\beta\nu;\mu}k_\nu k_\mu] + O(\Sigma^+). \tag{3.19}$$

We must re-express the last term in this result before establishing its conformal invariance. Applying the Bianchi identities to Eq. (2.7) gives

$$C_{\mu\nu[\alpha\beta;\gamma]} = S_{\nu[\beta;\gamma}g_{\alpha]\mu} + S_{\mu[\alpha;\gamma}g_{\beta]\nu}. \tag{3.20}$$

Taking tetrad components and using Eq. (2.6) then gives

$$C_{\mu\nu\alpha\beta;\gamma}k^\mu\bar{t}^\nu t^\alpha\bar{t}^\beta k^\gamma = R_{\mu[\alpha;\nu]}k^\mu t^\alpha k^\nu + O(\Sigma^+). \tag{3.21}$$

Next, using the Ricci identities and a tetrad expansion gives

$$R^{\alpha\beta;\mu}k_\alpha k_\beta\bar{t}_\mu = (R^{\alpha\beta}k_\alpha k_\beta)^{;\mu}\bar{t}_\mu - 2R^{\alpha\beta}k_\alpha k_{\beta;\mu}\bar{t}^\mu = 2(\sigma\bar{\sigma})_{;\mu}\bar{t}^\mu - 2\bar{\sigma}R^{\alpha\beta}k_\alpha t_\beta + O(\Sigma^+). \tag{3.22}$$

Similarly,

$$\begin{aligned} R^{\alpha\beta}k_\alpha t_\beta &= -k^{\alpha;\beta}{}_\alpha t_\beta = -(k^{\alpha;\beta}t_\beta)_{;\alpha} + k^{\alpha;\beta}t_{\beta;\alpha} \\ &= (k^{\alpha;\beta}\Omega_{;\alpha}t_\beta k^\gamma)_{;\gamma} - (\sigma\bar{t}^\gamma)_{;\gamma} + \sigma t_{\alpha;\beta}\bar{t}^\alpha\bar{t}^\beta + O(\Sigma^+) \\ &= -\Omega_{;\alpha\beta\gamma}k^\alpha t^\beta k^\gamma + \sigma_{;\gamma}\bar{t}^\gamma + 2\sigma t_{\alpha;\beta}\bar{t}^\alpha\bar{t}^\beta + O(\Sigma^+) \\ &= \tfrac{1}{2}R^{\alpha\beta}k_\alpha t_\beta - \sigma_{;\gamma}\bar{t}^\gamma + 2\sigma t_{\alpha;\beta}\bar{t}^\alpha\bar{t}^\beta + O(\Sigma^+), \end{aligned}$$

so that

$$R^{\alpha\beta}k_\alpha t_\beta = -2\sigma_{;\alpha}\bar{t}^\alpha + 4\sigma t_{\alpha;\beta}\bar{t}^\alpha\bar{t}^\beta + O(\Sigma^+). \tag{3.23}$$

Combining Eqs. (3.18)–(3.23), we now have

$$L_\xi(\Sigma^+) = \mathrm{Re}\oint_{\Sigma^+}\lambda^{-2}\{\Xi + \xi^\mu k_\mu\Phi + \xi^\mu t_\mu[2\Psi + (\sigma\bar{\sigma})_{;\alpha}\bar{t}^\alpha + 2\bar{\sigma}\sigma_{;\alpha}\bar{t}^\alpha - 4\bar{\tau}\sigma\bar{\sigma} - 4\lambda^{-1}\sigma\bar{\sigma}\lambda_{;\alpha}\bar{t}^\alpha]\}\,dS, \tag{3.24}$$

where

$$\Xi = \sigma(\xi^{\alpha;\beta}\bar{t}_\alpha\bar{t}_\beta)_{;\nu}k^\nu, \tag{3.25}$$

$$\Phi = C^{\mu\alpha\beta\nu;\gamma}m_\mu t_\alpha\bar{t}_\beta k_\nu k_\gamma, \tag{3.26}$$

$$\Psi = C^{\mu\alpha\beta\nu;\gamma}t_\mu\bar{t}_\alpha\bar{t}_\beta k_\nu k_\gamma, \tag{3.27}$$

and

$$\tau = \bar{t}_{\alpha;\beta} t^\alpha t^\beta. \tag{3.28}$$

In the special conformal frame being used in this section, we may set $\lambda = 1$ and $\lambda_{;\mu} t^\mu = 0$ in Eq. (3.24). These terms have been explicitly introduced for future reference when considering the conformal invariance of Eq. (3.24).

It is also useful to express $L_\xi(\Sigma^+)$ in a form in which derivatives of the descriptor field do not appear. To do this we apply the curl theorem, Eq. (2.24), to the first term in Eq. (3.24). We have

$$\begin{aligned}
\Xi &= \sigma\xi^\mu R_\mu{}^{\alpha\beta\nu} \bar{t}_\alpha \bar{t}_\beta k_\nu + \sigma\xi^{\alpha;\nu\beta} \bar{t}_\alpha k_\nu \bar{t}_\beta + O(\Sigma^+) \\
&= \tfrac{1}{2}\sigma(\xi^\mu k_\mu R^{\alpha\beta} \bar{t}_\alpha \bar{t}_\beta - \xi^\mu \bar{t}_\mu R^{\alpha\beta} \bar{t}_\alpha k_\beta) \\
&\quad + 2\sigma\xi^{\alpha;\nu\beta} \bar{t}_\alpha k_\nu \bar{t}_{(\beta} t_{\mu)} t^\mu + O(\Sigma^+) \\
&= -\tfrac{1}{2}\sigma(2\xi^\mu k_\mu \Omega^{;\alpha\beta\gamma} \bar{t}_\alpha \bar{t}_\beta k_\gamma + \xi^\mu \bar{t}_\mu R^{\alpha\beta} \bar{t}_\alpha k_\beta) \\
&\quad - 2\xi^{\alpha;\nu} \bar{t}_{(\beta} t_{\mu)} (\sigma \bar{t}_\alpha k_\nu t^\mu)^{;\beta} + O(\Sigma^+) + C \\
&= -\tfrac{1}{2}\sigma[2\xi^\mu k_\mu (\Omega^{;\alpha\beta} \bar{t}_\alpha \bar{t}_\beta)_{;\gamma} k^\gamma + \xi^\mu \bar{t}_\mu R^{\alpha\beta} \bar{t}_\alpha k_\beta] \\
&\quad - \xi^{\alpha;\nu} \bar{t}_\alpha k_\nu (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma) \\
&\quad - \xi^{\alpha;\nu} \bar{t}_{[\alpha} t_{\nu]} \sigma\bar{\sigma} + O(\Sigma^+) + C \\
&= -\tfrac{1}{2}\sigma[2\xi^\mu k_\mu (\Omega^{;\alpha\beta} \bar{t}_\alpha \bar{t}_\beta)_{;\gamma} k^\gamma + \xi^\mu \bar{t}_\mu R^{\alpha\beta} \bar{t}_\alpha k_\beta] \\
&\quad - \xi^\alpha \bar{t}_\nu t_\mu [k_\alpha t^\mu (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma)]^{;\nu} \\
&\quad - \xi^{\alpha;\nu} \bar{t}_{[\alpha} t_{\nu]} \sigma\bar{\sigma} + O(\Sigma^+) + C \\
&= -\xi^\mu k_\mu [\sigma(\Omega^{;\alpha\beta} \bar{t}_\alpha \bar{t}_\beta)_{;\gamma} k^\gamma + (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma)_{;\alpha} \bar{t}^\alpha \\
&\quad - \bar{\tau}(\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma)] - \xi^{\alpha;\nu} \bar{t}_{[\alpha} t_{\nu]} \sigma\bar{\sigma} \\
&\quad + \tfrac{1}{2}(\xi^\alpha t_\alpha \bar{\sigma} R_{\mu\nu} k^\mu t^\nu - \xi^\alpha \bar{t}_\alpha \sigma R_{\mu\nu} k^\mu \bar{t}^\nu) + O(\Sigma^+) + C,
\end{aligned}$$

so that

$$\begin{aligned}
&\operatorname{Re} \oint_{\Sigma^+} \lambda^{-2} \Xi \, dS \\
&\quad = -\operatorname{Re} \oint_{\Sigma^+} \lambda^{-2} \xi^\mu k_\mu [\lambda^{-1}(\bar{N}\sigma - P\sigma\bar{\sigma}) \\
&\qquad + (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma - \lambda^{-1}\sigma\lambda_{;\beta} t^\beta)_{;\alpha} \bar{t}^\alpha - (\bar{\tau} + \lambda^{-1}\lambda_{;\alpha} \bar{t}^\alpha) \\
&\qquad\qquad \times (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma - \lambda^{-1}\sigma\lambda_{;\beta} t^\beta)] \, dS,
\end{aligned} \tag{3.29}$$

where

$$\bar{N} = (\Omega^{;\alpha\beta} \bar{t}_\alpha \bar{t}_\beta)_{;\gamma} k^\gamma, \tag{3.30}$$

$$P = \tfrac{1}{4}\Omega^{;\alpha}{}_\alpha, \tag{3.31}$$

and where we may set $\lambda = 1$, $\lambda_{;\mu} t^\mu = 0$ and $P = 0$ in the special conformal frame being used in this section. Combining Eqs. (3.24) and (3.29) now gives the desired form:

$$\begin{aligned}
L_\xi(\Sigma^+) = \operatorname{Re} \oint_{\Sigma^+} \lambda^{-2} \{ &-\xi^\mu k_\mu [\lambda^{-1}(\bar{N}\sigma - P\sigma\bar{\sigma}) - \Phi \\
&+ (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma - \lambda^{-1}\sigma\lambda_{;\beta} t^\beta)_{;\alpha} \bar{t}^\alpha \\
&- (\bar{\tau} + \lambda^{-1}\lambda_{;\alpha} \bar{t}^\alpha)(\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma - \lambda^{-1}\sigma\lambda_{;\beta} t^\beta)] \\
&+ \xi^\mu t_\mu [2\Psi' + (\sigma\bar{\sigma})_{;\alpha} \bar{t}^\alpha + 2\bar{\sigma}\sigma_{;\alpha} \bar{t}^\alpha \\
&- 4\bar{\tau}\sigma\bar{\sigma} - 4\lambda^{-1}\sigma\bar{\sigma}\lambda_{;\alpha} \bar{t}^\alpha] \} \, dS.
\end{aligned} \tag{3.32}$$

4. CONFORMAL INVARIANCE

To establish the conformal invariance of our expressions for $L_\xi(\Sigma^+)$ we first show that these expressions are invariant under the tetrad transformation

$$k'_\mu = Ak_\mu, \quad m'_\mu = A^{-1} m_\mu \tag{4.1}$$

which changes the extensions of the null tetrad vectors k_μ and m_μ while maintaining their normalization. The tetrad vector m_μ appears in $L_\xi(\Sigma^+)$ only through the surface element dS defined in Eq. (2.20). We have

$$dS' = dS.$$

The shear transforms into

$$\sigma' = A\sigma$$

and λ transforms into

$$\lambda' = A\lambda.$$

Using these results, it is easy to show that, because of our insertion of factors of λ, both Eq. (3.24) and Eq. (3.32) are invariant under the tetrad transformation of Eq. (4.1). Thus these expressions for $L_\xi(\Sigma^+)$ do not depend upon the special form of k_μ assumed in Eq. (3.4). Consequently, it is convenient to fix the extension of k_μ so that

$$\lambda = 1 + O(\Sigma^+). \tag{4.2}$$

Equations (3.24) and (3.32) then simplify to

$$\begin{aligned}
L_\xi(\Sigma^+) = \operatorname{Re} \oint_{\Sigma^+} \{ &\Xi + \xi^\mu k_\mu \Phi \\
&+ \xi^\mu t_\mu [2\Psi' + (\sigma\bar{\sigma})_{;\alpha} \bar{t}^\alpha + 2\bar{\sigma}\sigma_{;\alpha} \bar{t}^\alpha - 4\bar{\tau}\sigma\bar{\sigma}] \} \, dS
\end{aligned} \tag{4.3}$$

and

$$\begin{aligned}
L_\xi(\Sigma^+) = \operatorname{Re} \oint_{\Sigma^+} \{ &-\xi^\mu k_\mu [\bar{N}\sigma - P\sigma\bar{\sigma} - \Phi \\
&+ (\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma)_{;\alpha} \bar{t}^\alpha - \bar{\tau}(\sigma_{;\beta} t^\beta - 2\bar{\tau}\sigma)] \\
&+ \xi^\mu t_\mu [2\Psi' + (\sigma\bar{\sigma})_{;\alpha} \bar{t}^\alpha + 2\bar{\sigma}\sigma_{;\alpha} \bar{t}^\alpha - 4\bar{\tau}\sigma\bar{\sigma}] \} \, dS.
\end{aligned} \tag{4.4}$$

Under the conformal transformation

$$g'_{\mu\nu} = f^2 g_{\mu\nu}, \tag{4.5}$$

we have

$$\Omega' = f\Omega = O(J^+), \tag{4.6}$$

$$t'_\mu = f t_\mu, \tag{4.7}$$

$$k'_\mu = f k_\mu, \quad \text{and} \quad m'_\mu = f m_\mu. \tag{4.8}$$

The requirement that Eq. (4.2) be conformally invariant determines the conformal weights assigned in

Eq. (4.8). Using Eqs. (4.5)–(4.8), we then find

$$\{\Omega_{;\alpha\beta}\}' = f\Omega_{;\alpha\beta} + \Omega f_{;\alpha\beta} - 2\Omega f^{-1}f_{;\alpha}f_{;\beta} + g_{\alpha\beta}(\Omega f^{-1}f_{;\gamma}f^{;\gamma} + f_{;\gamma}\Omega^{;\gamma}), \quad f4.9)$$

$$\bar{N}' = f^{-2}\bar{N} - f^{-3}f_{;\alpha\beta}\bar{l}^{\alpha}\bar{l}^{\beta} + 2f^{-4}f_{;\alpha}\bar{l}^{\alpha}f_{;\beta}\bar{l}^{\beta} + O(\Sigma^{+}), \tag{4.10}$$

$$P' = f^{-1}P + (^{-2}f_{;\beta}\Omega^{;\beta} + O(\Sigma^{+}), \tag{4.11}$$

$$\sigma' = f^{-1}\sigma, \tag{4.12}$$

$$\bar{\tau}' = f^{-1}\bar{\tau} - f^{-2}f_{;\nu}\bar{l}^{\nu}, \tag{4.13}$$

$$\Xi' = f^{-2}\Xi, \tag{4.14}$$

$$\Phi' = f^{-3}\Phi + O(\Sigma^{+}), \tag{4.15}$$

$$\Psi' = f^{-3}\Psi + O(\Sigma^{+}), \tag{4.16}$$

and

$$dS' = f^{2}\,dS. \tag{4.17}$$

By a straightforward insertion of these results into Eqs. (4.3) and (4.4), we readily find that $L_{\xi}(\Sigma^{+})$ is conformally invariant; i.e.,

$$\{L_{\xi}(\Sigma^{+})\}' = L_{\xi}(\Sigma^{+}).$$

5. CONCLUDING RESULTS

We may appreciably simplify our expressions for $L_{\xi}(\Sigma^{+})$ by introducing a covariant form of the angular-differential operator which Newman and Penrose[10] have denoted by ð. A scalar η constructed out of tetrad vectors is said to be of spin weight s if under the spatial rotation of tetrad vectors

$$l'^{\mu} = e^{iC}l^{\mu} \tag{5.1}$$

η transforms into

$$\eta' = e^{isC}\eta. \tag{5.2}$$

For instance, σ is of spin weight 2, $\bar{\sigma}$ is of spin weight -2, and $\sigma\bar{\sigma}$ is of spin weight 0. If η is of spin weight s we then define

$$ð\eta \equiv \eta_{;\mu}l^{\mu} + s\tau\eta, \tag{5.3a}$$

$$\bar{ð}\eta \equiv \eta_{;\mu}\bar{l}^{\mu} - s\bar{\tau}\eta. \tag{5.3b}$$

(The definition of ð given in Ref. 10 is actually $\sqrt{2}$ times the difinition used here.) Note that

$$\overline{(ð\eta)} = \bar{ð}\bar{\eta}.$$

If η is of spin weight s, then $ð\eta$ is of spin weight $s+1$ and $\bar{ð}\eta$ is of spin weight $s-1$. Thus

$$\bar{ð}ð\eta = (ð\eta)_{;\mu}\bar{l}^{\mu} - (s-1)\bar{\tau}ð\eta,$$

and in particular

$$\bar{ð}ð\sigma = (\sigma_{;\beta}l^{\beta} - 2\bar{\tau}\sigma)_{;\gamma}\bar{l}^{\gamma} - \bar{\tau}(\sigma_{;\beta}l^{\beta} - 2\bar{\tau}\sigma). \tag{5.4}$$

[10] E. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966).

Using these results, we may rewrite Eqs. (4.3) and (4.4) as

$$L_{\xi}(\Sigma^{+}) = \mathrm{Re}\oint_{\Sigma^{+}}\{\Xi + \xi^{\mu}k_{\mu}\Phi + \xi^{\mu}t_{\mu}[2\Psi + \bar{ð}(\sigma\bar{\sigma}) + 2\bar{\sigma}\bar{ð}\sigma]\}\,dS \tag{5.5}$$

and

$$L_{\xi}(\Sigma^{+}) = \mathrm{Re}\oint_{\Sigma^{+}}\{-\xi^{\mu}k_{\mu}[\bar{N}\sigma - P\sigma\bar{\sigma} - \Phi + \bar{ð}^{2}\sigma] + \xi^{\mu}t_{\mu}[2\Psi + \bar{ð}(\sigma\bar{\sigma}) + 2\bar{\sigma}\bar{ð}\sigma]\}\,dS. \tag{5.6}$$

Penrose,[6,7] in his development of the conformal approach to null infinity, found it convenient to introduce a particular choice of conformal factor which maps the inner geometry of Σ^{+} into that of a unit sphere. In addition, he specified the behavior of this conformal factor along the null directions on J^{+} by requiring $P = O(J^{+})$. From Eq. (4.11) we see that this is always possible. These two conformal conditions completely determine Ω on J^{+}; i.e., the remaining conformal freedom in Eq. (4.5) is of the form $f = 1 + O(J^{+})$. From the results of the previous section we know that Eqs. (5.5) and (5.6) are form invariant under a transformation to this special conformal gauge. From the construction of this gauge, however, we may put $P = 0$ in Eq. (5.6). In addition, for the four translational descriptors we have $\xi_{a}{}^{\mu}t_{\mu} = 0$ and $\xi_{a}{}^{\mu}k_{\mu}$ contains angular dependence only of the type described by spherical harmonics with $l \leq 1$.[4,9] But $\bar{ð}^{2}\sigma$ contains angular dependence only of the type described by spherical harmonics with $l \geq 2$.[10] Hence, by the orthonormality of the spherical harmonics, we have

$$\oint_{\Sigma^{+}}\xi_{a}{}^{\mu}k_{\mu}\bar{ð}^{2}\sigma\,dS = 0. \tag{5.7}$$

This argument based upon the usual spherical coordinates (θ, φ) can be justified because of our special choice of conformal guage. Since each term in Eq. (5.6) is individually invariant under coordinate transformation we may transform into a (θ, φ) coordinate system on the unit sphere Σ^{+} to carry out the above proof of Eq. (5.7). In this special conformal frame Eq. (5.6) then gives for the total energy–momentum

$$P_{a}(\Sigma^{+}) \equiv L_{\xi_{a}}(\Sigma^{+}) = -\mathrm{Re}\oint_{\Sigma^{+}}\xi_{a}{}^{\mu}k_{\mu}(\bar{N}\sigma - \Phi)\,dS. \tag{5.8}$$

This is exactly the expression for the total energy–momentum initially proposed by Penrose[6,7] from quite a different point of view.

Electric-Field Penetration into a Plasma with a Fractionally Accommodating Boundary

RODNEY J. MASON*
Department of Aeronautics and Astronautics, and Center for Space Research, Massachusetts Institute of Technology, Cambridge Massachusetts

(Received 3 August 1967)

Landau's field penetration study is extended to a plasma with a boundary that reflects a fraction σ of the incident electrons specularly and the remainder diffusely. Exact solutions for specular and diffuse reflection, and series solutions for fractional accommodation are obtained. At great depths in the plasma the field is found to exhibit negligible σ dependence for ω near ω_p and weak dependence through a factor $(1 + \sigma)$ at other frequencies.

I. INTRODUCTION

The second half of Landau's paper[1] on plasma oscillations was concerned with the penetration of an external, oscillating, longitudinal electric field into a semi-infinite plasma. It was assumed that the plasma ions formed a uniform, smeared-out charge background and that the plasma was confined by a specularly reflecting boundary. This paper examines the law of field penetration in the more general case: A fraction σ of the electrons striking the boundary reflect specularly, while the remainder come off diffusely with a Maxwellian distribution at the background plasma temperature.

Landau's original treatment has been clarified and extended by many workers.[2-8] With insight gained from these studies and our own recent investigation of sound propagation forced by a fractionally accommodating piston,[9] our approach is to convert the physical field penetration problem to an equivalent full-space problem which we solve with the aid of half-range Fourier transforms. We rederive Landau's result in the specular limit ($\sigma = 1$) and use the Wiener–Hopf technique to find the solution for the diffuse limit ($\sigma = 0$). For the electric field at large distances from the boundary we develop an asymptotic solution exhibiting, in general, only a weak σ dependence. Finally, for arbitrary σ we develop formally exact series solutions to the problem by iterating our specular and diffuse results.

II. FORMULATION

We replace the given, semi-infinite plasma with an equivalent plasma filling the full space and divided at $x = 0$ by a plane conducting boundary. This is insulated from the medium and reflects electrons with fractional accommodation. An oscillating external charge is applied to the boundary so as to create at $x = 0^{\pm}$ the fields $\pm E_0 e^{-i\omega t}$, which by assumption then penetrate the entire plasma in accordance with Poisson's equation and the linearized Vlasov equation:

$$\partial E/\partial x = -4\pi e n_1(x, t), \quad n_1 = \int_{-\infty}^{+\infty} g_1(u, x, t)\, du, \tag{1}$$

$$\partial g_1/\partial t + u(\partial g_1/\partial x) - \frac{e n_0 E(x, t)}{m}(\partial g_0/\partial u) = 0, \qquad g_0(u) = e^{-u^2/2a_0^2}/(2\pi)^{\frac{1}{2}} a_0. \tag{2}$$

Our aim is to find $E(x > 0)$. The exponential time dependence is assumed throughout. We define the half-range Fourier transforms

$$g_+(u, k) = \int_0^{\infty} g_1(u, x) e^{-ikx}\, dx, \qquad \int_0^{\infty} \frac{\partial g_1}{\partial x} e^{-ikx}\, dx = ikg_+ - g_1(u, 0^+), \tag{3}$$

$$E_+ = \int_0^{\infty} E(x) e^{-ikx}\, dx, \qquad \int_0^{\infty} \frac{\partial E}{\partial x} e^{-ikx}\, dx = ikE_+ - E(0^+), \tag{4}$$

while "minus transforms" are obtained from corresponding integrals over the range $x < 0$. The presence of $\exp(-ikx)$ in the "plus transforms" implies that these functions are nonanalytic in Im $(k) > 0^-$ and fully analytic for Im $(k) < 0^-$. Similarly all the nonanalytic character of the "minus transforms" resides in Im $(k) < 0^+$.

---

* Present address: Bell Telephone Laboratories, Whippany, New Jersey.

[1] L. Landau, J. Phys. (USSR) **10**, 25 (1946).
[2] R. Gould, Phys. Rev. **136**, A991 (1964).
[3] M. Feix, Phys. Letters **9**, 123 (1964).
[4] F. Shure, Ph.D. thesis, University of Michigan (1963).
[5] R. E. Aamodt and K. M. Case, Ann. Phys. (N.Y.) **21**, 284 (1963).
[6] V. Silin, Sov. Phys.—JETP **13**, 430 (1961).
[7] V. Kurilko, Sov. Phys.—Tech. Phys. **6**, 50 (1961).
[8] V. Kurilko and V. Miroshnichenko, Sov. Phys.—Tech. Phys. **7**, 588 (1963).
[9] R. Mason, Proc. Intern. Symp. Rarefied Gas Dyn. I, 395 (1967).

Application of these transforms to (1) and (2) yields

$$ikE_{\pm} = -4\pi en_{\pm} + E_0, \quad n_{\pm} = \int_{-\infty}^{+\infty} g_{\pm}\, du, \quad (5)$$

$$n_{\pm} = [N_0\epsilon(k) \pm \mathcal{G}(k, 0^{\pm})]/D(k), \quad N_0 = -E_0/4\pi e, \quad (6)$$

$$\epsilon(k) = \frac{i\omega_p^2}{k}\int_{-\infty}^{+\infty} \frac{(\partial g_0/\partial u)\, du}{(-i\omega + iku)}, \quad D(k) = 1 - \epsilon(k), \quad (7)$$

and

$$\mathcal{G}(k, 0^{\pm}) = \int_{-\infty}^{+\infty} \frac{ug_1(u, 0^{\pm})}{(-i\omega + iku)}\, du, \quad \omega_p^2 = 4\pi e^2 n_0/m. \quad (8)$$

Fractional accommodation and particle flux conservation at the boundary $x = 0^+$ requires

$$g_1(u, 0^+) = \begin{cases} -(1-\sigma)\dfrac{n_0 s^-(0^+)}{c_0} g_0(u) + \sigma g_1(-u, 0^+), & u > 0 \\ g_1(u, 0^+), & u < 0, \end{cases} \quad (9)$$

where

$$c_0 = \int_0^{\infty} ug_0\, du, \quad n_0 s^-(0^+) = \int_{-\infty}^{0} ug_1(u, 0^+)\, du.$$

Similar considerations at $x = 0^-$ give the corresponding expressions for $g_1(u, 0^-)$. The substitution of these distributions into (8) produces

$$\mathcal{G}(k, 0^{\pm}) = \mp(1-\sigma)S_{\mp}(0)I_{\pm}(k) + \sigma S_{\pm}(k) + S_{\mp}(k), \quad (10)$$

$$I_+(k) = -I_-(-k) = \frac{-i\omega}{c_0}\int_0^{\infty} \frac{ug_0(u)}{(-i\omega + iku)}\, du, \quad (11a)$$

and

$$S_-(k) = -S_+(-k) = \int_{-\infty}^{0} \frac{ug_1(u, 0^+)}{(-i\omega + iku)}\, du, \quad (11b)$$

so that $S_-(0) = n_0 s^-(0^+)/(-i\omega)$. In the $S_{\pm}(k)$ definition we use $g_1(u, 0^+) = g_1(-u, 0^-)$, which follows from the antisymmetry of the applied fields at $x = 0^{\pm}$.

Finally, combining Eqs. (5) to (11),

$$n_{\pm} = \frac{N_0\epsilon(k) - (1-\sigma)S_{\mp}(0)I_{\pm}(k) \pm \sigma S_{\pm}(k) \pm S_{\mp}(k)}{D(k)}, \quad (12a)$$

$$E_+(k) = \frac{-iE_0 + i4\pi e[-(1-\sigma) \times S_-(0)I_+(k) + \sigma S_+(k) + S_-(k)]}{kD(k)}, \quad (12b)$$

and

$$E_-(k) = \frac{-iE_0 + i4\pi e[-(1-\sigma) \times S_+(0)I_-(k) - S_+(k) - \sigma S_-(k)]}{kD(k)}. \quad (12c)$$

Expression (12b) could now be directly Fourier inverted to yield $E(x > 0)$, if the source transforms $S_{\pm}(k)$ were known. But these depend directly on the return distribution to the boundary $g_1(u \lesssim 0, 0^{\pm})$ which must itself be found as part of the overall solution to the problem. Thus, to derive the law of field penetration one must first solve (12b) explicitly for E_+ and then invert, or, alternatively, one inverts a suitable combination of the E_+ and E_- transforms in which the unknown contribution of the $S_{\pm}$ terms to the electric field are cancelled.

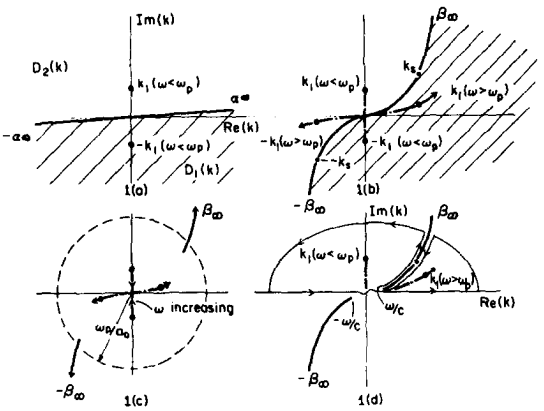


FIG. 1. The dispersion denominator $D(k)$: (a) Its branches $D_{1,2}(k)$. (b) Its analytic continuation to the steepest-descent cut. (c) The dominant asymptotic[1,2] behavior of this new $D(k)$. (d) The path for Fourier inversion for $x > 0$.

III. TRANSFORM-FUNCTION PROPERTIES

The functions $D(k)$, $I_{\pm}(k)$, and $S_{\pm}(k)$ are all defined for real k, if we ascribe a vanishingly small positive imaginary part to ω, corresponding to gradual growth of the external field. It is now useful to determine the analytic behavior of these functions in the remainder of the k plane.

According to (7), $D(k)$ originally defines a function cut along the line from $-\alpha\infty$ to $+\alpha\infty$ and $\alpha = \omega/|\omega|$, with branches $D_1(k)$ and $D_2(k)$ right and left of this cut, as shown in Fig. 1(a). Each branch can be analytically continued to the full k plane with the consequence that $D_2(k) = D_1(-k)$. Then, from (7) and Ref. 10

$$D_1(k) = 1 + s^2(2\omega_p^2/\omega^2)(1 + sZ(s)), \quad s = \omega/\sqrt{2}a_0 k,$$

$$Z(s) = i\pi^{\frac{1}{2}} \exp(-s^2) - s\sum_{n=0}^{\infty} \frac{(-s^2)^n\sqrt{\pi}}{(n+\frac{1}{2})!}, \quad (13a)$$

$$D_1(k) \sim 1 - \frac{\omega_p^2}{\omega^2}\left(1 + \frac{3k^2a_0^2}{\omega^2} + \frac{15k^4a_0^4}{\omega^4} + \cdots\right) + i\gamma\sqrt{\frac{\pi}{2}}\frac{\omega_p^2\omega}{k^3a_0^3} e^{-\omega^2/2a_0^2k^2}, \quad (13b)$$

with

$$\gamma = \begin{cases} 0 & < 0 \\ 1 & \text{Im } (k) \sim 0 \\ 2 & > 0. \end{cases}$$

[10] B. Fried and S. Conte, *The Plasma Dispersion Function* (Academic Press Inc., New York, 1961).

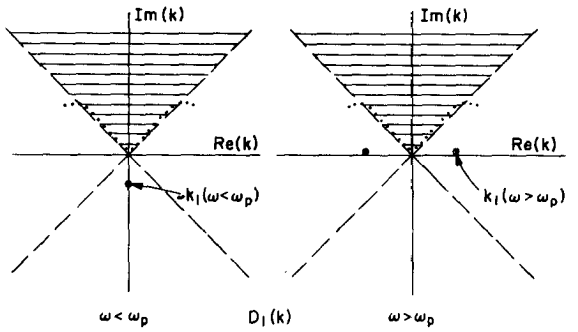


FIG. 2. Properties of $D_1(k)$.

The important features of $D_1(k)$ are indicated in Fig. 2. The function is analytic in the full k plane except for an essential singularity at $k = 0$ which produces divergence as the origin is approached within 45° of the positive imaginary axis. Outside this sector,

$$D_1(k \to 0) \to D(0) = 1 - \omega_p^2/\omega^2.$$

In the upper half-plane $D_1(k)$ has an infinity of zeros symmetrically distributed about the Im (k) axis. An asymptotic treatment[11] shows that these converge along rays at ±45° from the imaginary axis to a condensation point at the origin. Using (13b), we also find a pair of zeros near the origin and just above the real axis when ω is slightly in excess of ω_p. The right member of this pair $k_1(\omega > \omega_p)$ is Landau's least-damped root. The remaining zeros at large Im (k) must be located numerically. In the lower half-plane there is a single zero $-k_1(\omega < \omega_p)$ on the imaginary axis, if $\omega < \omega_p$; and no zeros, if $\omega > \omega_p$. From (13b) we derive

$$k_1(\omega \gtrless \omega_p) \sim \frac{\omega}{\sqrt{3}a_0} D^{\frac{1}{2}}(0) \times \left\{1 + i\gamma \frac{3}{2}\left(\frac{3\pi}{2}\right)^{\frac{1}{2}} \frac{1}{D^{\frac{5}{2}}(0)} e^{-3/[2D(0)]}\right\}, \quad (14)$$

with

$$\gamma = \begin{cases} 1 & \omega > \omega_p \\ 0 & \omega < \omega_p, \end{cases}$$

$\omega \approx \omega_p$, and

$$D^{\frac{1}{2}}(0) = i\,|1 - \omega_p^2/\omega^2|^{\frac{1}{2}} \quad \text{for} \quad \omega < \omega_p.$$

We can shift the branch cut for $D(k)$ to an arbitrary line dividing the plane and passing through the origin by simply requiring that D_1 apply to the right of this line and D_2 apply to the left. Across the new cut $D(k)$ suffers the jump

$$\Delta D = D_2 - D_1 = -(2\pi i)(\omega_p^2\omega/k^3a_0^2)g_0(\omega/k). \quad (15)$$

It is particularly useful to run this branch cut along the "steepest descent path" from, say, $-\beta_\infty$ to $+\beta_\infty$ through the saddle points $k_s = \pm(\omega^2/a_0^2x)^{\frac{1}{3}}e^{i\pi/6}$ of the expressions $\Delta De^{\pm ikx}$, as in Fig. 1(b).

[11] R. Hawks, Masters thesis, M.I.T. (1967).

This choice of cut constructs a new $D(k)$ which has only the Landau zeros of D_1 and D_2 near the origin. These zeros are on the imaginary axis for $\omega < \omega_p$ and in the first and third quadrants for $\omega_p < \omega < \omega_{\max}$. The 45° ray of D_1 zeros in Im $(k) > 0$ near $k = 0$ is excluded from the new $D(k)$ by the "steepest descent" cut, as is the corresponding ray of D_2 zeros in Im $(k) < 0$. Some zeros at large Im (k) may, however, be included, as must be investigated numerically. We must also expect the Landau zeros to cross the cut for some $\omega = \omega_{\max}$.

From (15) and (13b) we find that the cut in $D(k)$ is weak, i.e., $\Delta D/D_1 \ll 1$, for $|k| \ll \omega/a_0$. Then since $\pm k_1$ are close to the origin for $\omega \approx \omega_p$, these zeros dominate in $D(k)$ for small k, effectively yielding Fig. 1(c) and

$$D(k) \sim [D(0)/k_1^2](k_1^2 - k^2) + O[(ka_0/\omega)^2]. \quad (16)$$

Initially, the functions $I_+(k)$ and $S_+(k)$ are analytic in the full k plane except along the ray from $k = 0$ to $k = \alpha\infty$ in Im $(k) > 0$. This cut can be moved to coincide with the upper-half of the steepest descent cut for the new $D(k)$ by appropriately shifting the path of u plane integration in the definitions (11). The cut for $I_-(k)$ and $S_-(k)$ can be similarly shifted in Im $(k) <$ 0. Then, across the steepest descent path we get, for example,

$$\Delta I_+(k) = I_{+2} - I_{+1} = (2\pi i)\frac{\omega^2}{k^2c_0} g_0\left(\frac{\omega}{k}\right), \quad \text{Im}\,(k) > 0, \quad (17a)$$

$$\Delta S_-(k) = S_{-2} - S_{-1} = \frac{2\pi\omega}{k^2} g_1\left(\frac{\omega}{k}, 0^+\right), \quad \text{Im}\,(k) < 0, \quad (17b)$$

with subscripts "1" and "2" applying to the right and left of the cut, respectively.

The cut in $I_\pm$ and $S_\pm$ will also be weak for small k, if $g_1(u \lesseqgtr 0, 0^\pm)$ falls off rapidly for large u, as would be expected. Thus, for $|k| \ll \omega/a_0$ we can expand the denominators of the defining integrals (11) in powers of ku/ω and integrate term by term to produce the asymptotic estimates

$$I_\pm(k) \sim I_\pm(0) + I'_\pm(0)k + O\left[\left(\frac{ka_0}{\omega}\right)^2\right] \quad (18a)$$

$$= \pm 1 + \left(\frac{a_0^2}{c_0\omega}\right)k + O\left[\left(\frac{ka_0}{\omega}\right)^2\right], \quad (18b)$$

$$S_\pm(k) \sim S_\pm(0) + S'_\pm(0)k + O\left[\left(\frac{ka_0}{\omega}\right)^2\right] \quad (18c)$$

$$= \frac{\mp n_0 s^-(0^+)}{(-i\omega)} + \left(\frac{i}{\omega^2}\int_{-\infty}^{0} u^2 g_1(u, 0^+)\,du\right) \times k + O\left[\left(\frac{ka_0}{\omega}\right)^2\right]. \quad (18d)$$

Finally, we introduce a finite upper bound on all electron speeds, say $|u| < c$, $c \gg a_0$. This moves the branch points for $I_\pm$ and $S_\pm$ from the origin to $k = \pm\omega/c$, and erases the cut in $D(k)$ between these points, as in Fig. 1(d), thereby creating an analytic real axis strip, which allows us to analyze (12) with the aid of Wiener–Hopf technique.

IV. LIMITING SOLUTIONS

A. Specular Reflection

Let $\sigma = 1$ and add the $E_\pm$ transforms

$$\tilde{E} = E_+ + E_- = -2iE_0/kD(k). \tag{19}$$

The $S_\pm(k)$ cancel in (19) permitting a direct Fourier inversion which may be completed over the $1(d)$ contour when $x > 0$. Consistency with (13) requires that the integral pass below the E_+ pole at $k = 0$, and above the E_- pole there. This is equivalent to taking the principle value at the origin. Thus, from the poles at $k = 0$ and k_1 and the integral along the branch cut we obtain

$$E(x > 0, \sigma = 1) \doteq E_0\left[\frac{1}{D(0)} + \frac{2e^{ik_1x}}{k_1(\partial D/\partial k)|_{k_1}}\right] + B(1, x), \tag{20a}$$

$$B(1, x) = \frac{-iE_0}{\pi}\int_0^{\beta\infty} \frac{\Delta D e^{ikx}}{kD_1D_2}\,dk, \tag{20b}$$

and

$$E(x > 0, \sigma = 1) \sim \frac{E_0}{D(0)}\{1 - e^{ik_1x}\} + B(1, x). \tag{20c}$$

A dotted equal sign appears in (20a) since the possible contribution of D_1 zeros at large $|\mathrm{Im}\,(k)|$ has been neglected. The integral $B(1, x)$ is estimated in Sec. V. The exponential excitation in (20c) is computed with (16) for $\omega \approx \omega_p$ and assumed to hold for arbitrary ω, since away from ω_p the cut contribution $B(1, x)$ dominates the net asymptotic disturbance. This is essentially Landau's result.

B. Diffuse Reflection

Now set $\sigma = 0$. Equation (12b) becomes

$$E_+(k) = \{-iE_0 + i4\pi e[-S_-(0)I_+(k) + S_-(k)]\}/kD(k). \tag{21}$$

Since $S_+(k)$ is absent, (21) can be solved explicitly for E_+ and S_- by means of the Wiener–Hopf technique.

To proceed we require the operators $[\ ]_\pm$,

$$[f(t)]_+ \equiv \frac{-1}{2\pi i}\oint_{-\infty(+)}^{+\infty} \frac{f(t)\,dt}{t - k}\,dt = f_+(k), \tag{22a}$$

$$[f(t)]_- \equiv \frac{1}{2\pi i}\oint_{-\infty(-)}^{+\infty} \frac{f(t)}{t - k}\,dt = f_-(k), \tag{22b}$$

which separate $f(k)$ into parts[12] which are nonanalytic in just the upper or lower half-plane, respectively, provided that $f(k)$ is analytic on a real axis strip, and that $f(k) \sim k^{-p}$, $p > 0$ for $k \to \pm\infty$ in this strip. The $[\ ]_+$ contour passes along the real t axis and above the pole at $t = k$, k real. The $[\ ]_-$ contour passes below this pole. Thus, $f(k) = [f(t)]_+ + [f(t)]_-$.

In particular, the application of these operations to $\ln D(k)$ produces factors for $D(k)$, i.e.,

$$D^\pm(k) = \exp\{[\ln D(t)]_\pm\}, \quad D(k) = D^+(k)D^-(k), \tag{23}$$

such that $D^+(k)$ and $D^-(k)$ are zero and/or nonanalytic in only $\mathrm{Im}\,(k) > 0$, and $\mathrm{Im}\,(k) < 0$, respectively, and thus

$$\left.\begin{aligned} D^+(k_1) = D(k_1) = 0, \quad & D^-(k_1) \neq 0 \\ (\Delta D^+(k))D^-(k) = \Delta D, \quad & \Delta D^-(k) = 0 \end{aligned}\right\} \ \mathrm{Im}\,(k) > 0, \tag{24}$$

while corresponding relations apply in $\mathrm{Im}\,(k) < 0$.

Consequently, (21) can be written

$$kD^+E_+ = -iE_0/D^- - i4\pi e[S_-(0)(I_+/D^-) - S_-/D^-]. \tag{25}$$

Let $C(k) = I_+(k)/D^-(k)$, produce $C_\pm(k) = [I_+(t)/D^-(t)]_\pm$, and rearrange (25) to

$$kD^+(k)E_+(k) + i4\pi eS_-(0)C_+(k) = -iE_0 \tag{26a}$$

$$= \frac{-iE_0}{D^-(k)} - i4\pi e[S_-(0)C_-(k) - S^-(k)/D^-(k)]$$

$$= -iE_0, \tag{26b}$$

$$S_-(0) = -E_0[D^-(0) - 1]/4\pi eD^-(0)C_+(0). \tag{26c}$$

The right-hand sides of (26a, b) follow from the standard Wiener–Hopf arguments and the fact that, as $k \to \pm\infty$, $C_+ \to C_- \to S_- \to 0$, while $D^+ \to D^- \to 1$. To get (26c) we solve (26b) for $S_-(k)$, set $k = 0$, and note that $C_+(0) + C_-(0) = 1/D^-(0)$. Finally, putting (26c) in (26a) and using (23)

$$E_+(k) = -iE_0\left[1 + \frac{(1 - D^-(0))C_+(k)}{D^-(0)C_+(0)}\right]\frac{D^-(k)}{kD(k)}. \tag{27}$$

We can readily shift the (22a) integral path for $C_+(k)$ to the upper-half of the steepest descent curve in the t plane, since $D^-(t)$ is analytic and nonzero in $\mathrm{Im}\,(t) > 0$. Thus

$$C_+(k) = \frac{-1}{2\pi i}\oint_{-\infty(+)}^{+\infty} \frac{I_+(t)\,dt}{D^-(t)(t - k)}$$

$$= \frac{1}{2\pi i}\int_0^{\beta\infty} \frac{\Delta I_+(t)\,dt}{D^-(t)(t - k)}, \tag{28}$$

[12] B. Noble, *Methods Based on the Wiener–Hopf Technique* (Pergamon Press, Inc., New York, 1958).

so that $C_+(k)$ is analytic everywhere, but on the corresponding cut in the k plane. It follows that $E_+(k)$ is nonanalytic on this cut, at the zeros of $D(k)$ in Im $(k) > 0$, and at $k = 0$. Inverting over the 1(d) contour with an integral below $k = 0$, we therefore obtain

$$E(x > 0, \sigma = 0) \doteq E_0\left\{\frac{1}{D(0)} + \left[1 + \frac{(1 - D^-(0))C_+(k_1)}{D^-(0)C_+(0)}\right]\frac{D^-(k_1)}{k_1(\partial D/\partial k)|_{k_1}} \times e^{ik_1x}\right\} + B(0, x), \quad (29a)$$

$$B(0, x) = \frac{-iE_0}{2\pi}\int_0^{\beta\infty}\frac{D^-}{kD_1D_2}\left[\Delta D + \frac{(1 - D^-(0))}{D^-(0)C_+(0)} \times (\Delta DC_{+1} - \Delta C_+D_1)\right]e^{ikx}\,dk. \quad (29b)$$

The Appendix shows that for $\omega \approx \omega_p$ and $|k| \ll \omega_p/a_0$,

$$D^{\pm}(k) \sim \frac{D^{\pm}(0)}{k_1}(k_1 \mp k) + O\left[\left(\frac{ka_0}{\omega}\right)\right], \quad (30)$$

which means, for example, that D^- is dominated for small k by the lower half-plane zero of D, while the product D^+D^- from (30) equals our estimate (16) for $D(k)$ to $O[(ka_0/\omega)]$. Also from (28) $\Delta C_+(k) = \Delta I_+(k)/D^-(k)$, Im $(k) > 0$, and since this jump is negligible for $|k| \ll \omega/a_0$, (28) expands approximately to

$$C_+(k) \sim C_+(0) + C'_+(0)k + O\left[\left(\frac{ka_0}{\omega}\right)^2\right]. \quad (31)$$

With these expansions (29a) becomes

$$E(x > 0, \sigma = 0) \sim \frac{E_0}{D(0)}(1 - e^{ik_1x}) + B(0, x), \quad (32)$$

which agrees, except for the branch-cut term, with our specular result. Again, since $B(0, x)$ dominates the net asymptotic field for ω away from ω_p, (32) is good for all ω at great depths in the plasma.

V. ASYMPTOTIC SOLUTION (ARBITRARY σ)

For arbitrary σ we formally invert the sum $E_+ + \sigma E_-$ with an integral running along real axis below the E_+ pole at $k = 0$, but above the E_- pole there, and closing over the 1(d) contour. There results

$$E(x > 0, \sigma) \doteq \frac{E_0}{D(0)} + \frac{iN(k_1)}{k_1(\partial D/\partial k)|_{k_1}}e^{ik_1x} + B(\sigma, x), \quad (33a)$$

$$N(k) = -iE_0 + i4\pi e[-(1 - \sigma)S_-(0)I_+(k) + \sigma S_+(k) + S_-(k)], \quad (33b)$$

$$B(\sigma, x) = \frac{-i(1 + \sigma)E_0}{2\pi}\int_0^{\beta\infty}\frac{\Delta D}{kD_1D_2}e^{ikx}\,dk + \frac{i(1 - \sigma)}{2\pi}\int_0^{\beta\infty}\frac{(M_1\Delta D - D_1\Delta M)}{kD_1D_2}e^{ikx}\,dk, \quad (33c)$$

and

$$M(k, \sigma) = 4\pi e[-S_-(0)(I_+(k) - \sigma I_-(k)) + (1 + \sigma)S_-(k)]. \quad (33d)$$

The function $N(k)$ is the numerator of E_+ in (12b). Note that E_- makes no contribution to the residue at k_1, since this function is analytic in Im $(k) > 0$. Similarly, $N(-k_1) = 0$, since E_+ is analytic in Im $(k) < 0$.

When $\omega \approx \omega_p$, $|k_1| \ll \omega_p/a_0$. For such small k we can asymptotically expand $N(k)$ by virtue of (18). Thus

$$N(-k_1) = 0 = N(0) + N'(0)(-k_1) + O\left[\left(\frac{k_1a_0}{\omega}\right)^2\right]. \quad (34)$$

Then, since (33b) and the $S_\pm$, $I_\pm$ definitions show that $N(0) = -iE_0$, (34) yields $N'(0)$, and thus

$$N(k) \sim -iE_0(1 + k/k_1) + O\left[\left(\frac{k_1a_0}{\omega}\right)^2\right]. \quad (35)$$

Using this and (16), (33a) becomes for *all* σ

$$E(x > 0, \sigma) \sim \frac{E_0}{D(0)}\{1 - e^{+ik_1x}\} + B(\sigma, x). \quad (36)$$

Last, we employ the method of steepest descent to estimate $B(\sigma, x)$. By construction $M(k, \sigma)$ is independent of $S_+(k)$. Thus, $\Delta M(k) = -4\pi eS_-(0)\Delta I_+(k)$, Im $(k) > 0$, is known to within the constant $S_-(0)$. Also the path of integration has already been chosen so as to pass through the saddle point k_s of the exponential factor $g_0(\omega/k)e^{+ikx}$, which dominates each term in the (33c) integrands. Hence, we simply evaluate the nonexponential factors at k_s, shift them outside the integral sign, and complete the integration in the usual manner. Then, since $\Delta I_+/\Delta D \sim k_s$ and $M(k, \sigma) \sim k_s$ as $x \to \infty$ and $k_s \to 0$, the first integral in (33c) dominates the second for large x and yields

$$B(\sigma, x) \sim \frac{(1 + \sigma)}{\sqrt{3}}\frac{E_0}{D^2(0)}\left(\frac{\omega_p^2}{\omega^2}\right)\tau^{\frac{2}{3}} \times \exp\left(-\tfrac{3}{2}\tau^{\frac{2}{3}}e^{-i\pi/3}\right)e^{+i2\pi/3}, \quad (37)$$

with $0 \le \sigma \le 1$ and $\tau = \omega x/a_0 \gg 1$.

Near ω_p the exponential term is slowly damped, so the σ dependence of the net asymptotic field is negligible, except at the greatest depths. Away from ω_p, the weak σ dependence of $B(\sigma, x)$ which dominates asymptotically, will prevail.

VI. THE GENERAL SOLUTION

Equation (12a) can be expressed in the vector form

$$\begin{pmatrix} u_+ \\ S_+ \end{pmatrix} = \mathsf{G}(k,\sigma)\begin{pmatrix} n_- \\ S_- \end{pmatrix} + \begin{pmatrix} \dfrac{(1+\sigma)N_0\epsilon(k) - (1-\sigma)S_-(0)[I_+(k) - \sigma I_-(k)]}{D(k)} \\ N_0\epsilon(k) - (1-\sigma)S_+(0)I_-(k) \end{pmatrix}, \tag{38a}$$

$$\mathsf{G}(k,\sigma) = \begin{pmatrix} -\sigma & \dfrac{(1-\sigma^2)}{D(k)} \\ -D(k) & -\sigma \end{pmatrix}. \tag{38b}$$

If $\mathsf{G}(k)$ can now be factored into two nonsingular matrices $\mathsf{G}(k) = \mathsf{G}^+(k)\mathsf{G}^-(k)$ with $\mathsf{G}^+(k)$ and $\mathsf{G}^-(k)$ nonanalytic in Im $(k) > 0$ and Im $(k) < 0$, respectively, then (38a) can be solved for $(n_\pm, S_\pm)$ by means analogous to those for our earlier diffuse solution. Equation (5) then yields $E_+(k)$, which can be inverted over 1(d).

For $\sigma = 1$ and $\sigma = 0$ possible factorizations are

$$\mathsf{G}(k,1) = \begin{pmatrix} -1 & 0 \\ -1+\epsilon_+ & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\epsilon_- & 1 \end{pmatrix} = \mathsf{G}^+(k,1)\mathsf{G}^-(k,1),$$

$$\mathsf{G}(k,0) = \begin{pmatrix} 0 & 1/D^+ \\ -D^+ & 0 \end{pmatrix}\begin{pmatrix} D^- & 0 \\ 0 & 1/D^- \end{pmatrix} = \mathsf{G}^+(k,0)\mathsf{G}^-(k,0),$$

with $D(k) = 1 - \epsilon(k)$, $\epsilon_\pm = [\epsilon(t)]_\pm$. Use of these matrices reproduces our earlier specular and diffuse solutions. For arbitrary σ, however, no general factorization is available. We, therefore, re-examine (12a) and seek separate series solutions in powers of $\alpha = 1 - \sigma$ and σ. This is equivalent to finding factorization for $\mathsf{G}(k,\sigma)$ by iterating our limiting results.

A. Specular Series

Assume that $n_\pm = \sum \alpha^q n_\pm^{(q)}$, $S_\pm = \sum \alpha^q S_\pm^{(q)}$ with $\alpha = 1 - \sigma$, substitute these expansions into (12a), and add the resultant transforms. To each order in α one finds

$$n^{(0)} = n_+^{(0)} + n_-^{(0)} = 2N_0\epsilon(k)/D(k), \tag{39a}$$

$$n_{(q\geq 1)}^{(q)} = \left\{\frac{-S_-(0)^{(q-1)}I_+ - S_-(0)^{(q-1)}I_- - S_+(k)^{(q-1)} + S_-(k)^{(q-1)}}{D(k)}\right\} \equiv \frac{2N_0 S(k)^{(q)}}{D(k)}, \tag{39b}$$

$$S_\pm^{(0)} = \mp[D(t)n_\pm^{(0)}(t)]_\pm \pm N_0\epsilon_\pm(k), \tag{39c}$$

$$S_{\pm\,(q\geq 1)}^{(q)} = \mp[D(t)n_\mp^{(q)}(t)]_\pm, \tag{39d}$$

where $n_\pm^{(q)} = [n^{(q)}(t)]_\pm$ and (39c, d) are established by applying the separation operators to $D(k)$ times the expanded (12a) equations. The function $S(k)^{(q)}$, $q \geq 1$ is defined by (39b). We also set $S(k)^{(0)} \equiv 1$.

Add the field transform equations (5), and substitute $n = \sum \alpha^q n^{(q)}$, as derived from (39). Then, assuming uniform convergence for α sufficiently small, invert the series term by term over the 1(d) contour, taking the principle value at $k = 0$. Since $S(0)^{(q)} = 0$, $q \geq 1$, there results

$$E(x>0,\alpha) \doteq E_0\left(\frac{1}{D(0)} + \frac{2e^{ik_1x}}{k_1(\partial D/\partial k)|_{k_1}}\sum_0^\infty \alpha^q S(k_1)^{(q)}\right) + B(\alpha,x), \tag{40}$$

$$B(\alpha,x) = \frac{-iE_0}{\pi}\int_0^{\beta\infty}\left[\sum_0^\infty \alpha^q(\Delta D S_1^{(q)} - \Delta S^{(q)}D_1)\right] \times \frac{e^{ikx}}{kD_1D_2}\,dk.$$

B. Diffuse Series

Substitute $n_\pm = \sum \sigma^j n_\pm^{(j)}$ and $S_\pm = \sum \sigma^j S_\pm^{(j)}$ into (12a). To each order in σ we get equations which can be solved by the Wiener–Hopf technique. Accordingly,

$$n_+^{(0)} = \frac{1}{D^+}\left[\frac{N_0\epsilon(k) - S_-(0)^{(0)}I_+}{D^-}\right]_+, \tag{41a}$$

$$n_{+\,(j\geq 1)}^{(j)} = \frac{1}{D^+}\left[\frac{S_+^{(j-1)} + (S_-(0)^{(j-1)} - S_-(0)^{(j)})I_+}{D^-}\right]_+ \equiv \frac{N_0(1-D^-(0))}{D^-(0)C_+(0)}\frac{C_+^{(j)}(k)}{D^+(k)}, \tag{41b}$$

$$S_\pm^{(0)} = \pm D^\pm\left[\frac{N_0\epsilon(k) - S_\pm(0)^{(0)}I_\pm}{D^\pm}\right]_\pm, \tag{41c}$$

and

$$S_{\pm\,(j\geq 1)}^{(j)} = -D^\pm\left[\frac{S_\mp^{(j-1)} \mp (S_\pm(0)^{(j-1)} - S_\pm(0)^{(j)})I_\mp}{D^\pm}\right]_\pm. \tag{41d}$$

Equation (41b) defines $C_+^{(j)}(k)$. We also let $C_+^{(0)}(k) \equiv C_+(k)$ of (28). Then combining (5) with the n_+ series determined from (41), integrating term by term over the 1(d) contour, and noting that $n_+^{(j)}(0) = 0$, $j \geq 1$, we can derive

$$E(x>0,\sigma) \doteq E_0\left\{\frac{1}{D(0)} + \left[1 + \frac{(1-D^-(0))}{D^-(0)C_+(0)}\sum_0^\infty \sigma^j C_+^{(j)}(k_1)\right] \times \frac{D^-(k_1)}{k_1(\partial D/\partial k)|_{k_1}}e^{ik_1x}\right\} + B(\sigma,x), \tag{42}$$

and

$$B(\sigma, x) = \frac{-iE_0}{2\pi} \times \int_0^{\beta_\infty} \left[\Delta D + \frac{(1 - D^-(0))}{D^-(0)C_+(0)} \sum_0^\infty \sigma^j (\Delta D C_{+1}^{(j)} - \Delta C_+^{(j)} D_1) \right] \times \frac{e^{ikx} D^-}{k D_1 D_2} dk.$$

C. DISCUSSION

To solve the problem of field penetration into a semi-infinite plasma, we have examined a corresponding medium filling the full space and divided at $x = 0$ by a sheet of oscillating external charge. This charge gives rise to antisymmetric fields and symmetric density disturbances, so that for each electron approaching the sheet or boundary from the right there is a "twin" approaching from the left. Hence, our specularly reflected electrons effectively pass unaltered through the boundary, while the diffusely reflected ones return with a Maxwellian distribution to their original half-space. Thus for $\sigma = 1$ the disturbances in the two half-planes are completely coupled, while for $\sigma = 0$, they are thoroughly isolated.

Our specular series solution assumes in zero order that all electrons pass freely through the boundary. To $O(\alpha)$ it converts the incident zero-order distribution to Maxwellian form at $x = 0$, and produces first-order disturbances which themselves freely penetrate the boundary. To $O(\alpha^2)$ these are converted to Maxwellian form, etc. Alternatively, our diffuse series assumes complete isolation of the two half planes in zero order, and allows for "leakage" of zero-order electrons between the two half-planes to $O(\sigma)$, etc. Since our asymptotic solution for large x exhibits only first-order dependence on σ, it would seem that for an accurate description of the field only the first few terms of each series may be required, and thus rapid convergence is expected for both iterative schemes.

We note that (40) and (42) can be made more amenable to numerical calculation by shifting the branch cuts for all transform functions from the steepest descent path back to a line neighboring the real axis, as in Fig. 1(a). This constructs a new $D(k)$ which has no zeros for $\omega > \omega_p$ and only the Landau zeros for $\omega < \omega_p$. Expressions (40) and (42) will then have, at most, a single exponential mode for $\omega < \omega_p$, while $B(\alpha, x)$ and $B(\sigma, x)$ can readily be computed as integrals along the positive real k axis.

ACKNOWLEDGMENTS

The author is pleased to acknowledge a number of helpful discussions with Professor J. E. McCune and Professor L. Trilling.

This work was supported by National Aeronautics and Space Administration Grant No. NSG-496.

APPENDIX

To find $D^-(k)$ for small k we make the asymptotic expansion

$$D^-(k) = \exp\left\{ \frac{1}{2\pi i} \oint_{-\infty(-)}^{+\infty} \frac{\ln D(t)\, dt}{t - k} \right\} \sim \exp\{A_0 + A_1 k + A_2 k^2 + O(k^3)\} \sim D^-(0)\left\{ 1 + A_1 k + \left(\frac{A_1^2 + 2A_2}{2} \right) k^2 + O(k^3) \right\}, \quad \text{(A1)}$$

with

$$A_n = \frac{1}{2\pi i} \oint_{-\infty(-)}^{+\infty} \frac{\ln D(t)}{t^{n+1}}\, dt = \frac{1}{2\pi i} \oint_{-\infty(-)}^{+\infty} \frac{D'(t)}{n D(t)} \frac{dt}{t^n}, \quad n \geq 1, \quad \text{(A2)}$$

where the final equality follows from integration by parts. Evaluating this over the contour that mirrors Fig. 1(d) in Im $(k) < 0$, we pick up a residue at the $D(t)$ zero $-k_1$, and a contribution from the branch cut, i.e.,

$$A_n = \frac{1}{n(-k_1)^n} + \frac{1}{2\pi i} \int_{-\beta_\infty}^{0} \frac{\Delta\left[\frac{D'(t)}{D(t)}\right]}{nt^n}\, dt. \quad \text{(A3)}$$

With the substitution $t = \omega/\sqrt{2}a_0 s$, (A3) becomes

$$A_n = \frac{-1}{n(-k_1)^n} + \frac{B_n}{(\omega/a_0)^n}\left(\frac{\omega_p^2}{\omega^2}\right), \quad n \geq 1,$$

$$B_n = \frac{-2}{\sqrt{\pi}} \frac{2^{n/2}}{n} \oint_0^\infty \left[\frac{(2s^2 - 3)D_2(s) + sD_2'(s)}{D_1(s)D_2(s)} \right] \times s^{n+2} e^{-s^2}\, ds, \quad \text{(A4)}$$

where the path for $\oint$ is the s-plane mapping of the usual steepest-descent curve.

For small n and $\omega \approx \omega_p$ we expect that B_n is $O(1)$, since the $D_{1,2}(s)$ are slowly varying and nonzero along the integral path. Thus,

$$D^-(k) = D^-(0)\left\{ 1 + \left(\frac{1}{k_1} + \frac{B_1}{\omega/a_0} \right) k + \left(O\left[\frac{B_1}{k_1 \omega/a_0} \right] \right) k^2 + O(k^3) \right\} = [D^-(0)/k_1](k_1 + k) + O(ka_0/\omega).$$

Further, by substituting $-k$ in (22b) and (23) one can show that $D^+(k) = D^-(-k)$, and thus

$$D^\pm(k) = \frac{D^-(0)}{k_1}(k_1 \mp k) + O\left(\frac{ka_0}{\omega}\right), \quad \text{(A5)}$$

where, since $D^+(k)D^-(k) = D(k) = D^-(-k)D^-(k)$, we have

$$D^-(0) = D^{\frac{1}{2}}(0) = (1 - \omega_p^2/\omega^2)^{\frac{1}{2}}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Generalized Ornstein-Zernike Approach to Critical Phenomena

MELVILLE S. GREEN
National Bureau of Standards, Washington, D.C.

(Received 5 December 1967)

A generalization of the Ornstein–Zernike integral equation is derived and suggestions are made about a possible application to an improved theory of critical phenomena. A fundamental maximum principle of statistical mechanics is used to place the generalized equation in the context of phase transitions and critical points. The equation is a relationship between a generalized correlation matrix by means of which the average fluctuation product of any two sum functions may be expressed and a generalized direct-correlation matrix which is the second functional derivative of the functional in the maximum principle. The existence of a critical eigenvector of the direct-correlation matrix is proposed and three physical meanings of this vector are given. An explicit formula for the direct-correlation matrix is given and is used to derive two asymptotic properties. This formula exhibits an unexpected relationship between the generalized Ornstein–Zernike equation and the Percus–Yevick equation.

Of the many anomalous phenomena which take place in a fluid as its critical point is approached perhaps the most striking is the sudden development of a dense white opacity in the otherwise transparent fluid. A theory of this phenomenon, critical opalescence, was given a half century ago by Ornstein and Zernike in a famous series of papers[1] which illuminated not only this particular problem but some of the outstanding fundamental questions of statistical mechanics of the time. Somewhat earlier, van der Waals had presented his well-known molecular theory of thermodynamic properties near the critical point.[2] In his second published paper, Gibbs[3] gave a beautiful and thought-provoking diagram which expresses the essential qualitative features for the van der Waals theory with characteristic economy and clarity. The Ornstein–Zernike theory and the qualitative theory of Gibbs provide an explanation for both fluctuation and thermodynamic phenomena near critical points which is in many ways very satisfactory and which has not been challenged until very recently.

This challenge has come about through a closer investigation, both theoretical and experimental, of various critical phenomena motivated by the evidently deep-going analogy among such disparate and important phenomena as the λ transition in liquid helium, ferromagnetic and antiferromagnetic transitions, and order–disorder phenomena in alloys as well as the liquid-vapor critical point.[4] Although these two theories give the major qualitative features of the phenomena, they fail quantitatively more and more as the critical point is approached. It is characteristic of the Ornstein–Zernike theory as well as of the Gibbs theory that a fundamental element (in the one case, the direct-correlation function, in the other, the energy–entropy–volume relationship) has a simple analytic behavior. Anomalies, infinities, discontinuities in light scattering, compressibility, heat capacity, and the like come about through the singularity of the relationship between the measured quantity and the assumed simple behavior of the fundamental element. The success of these theories consists in the fact that anomalies are predicted which agree qualitatively with actual observed behavior. Their failure consists in the fact that they predict only poles, square-root branch points, or jump discontinuities, while experiment indicates more complicated singularities.[5]

It is the purpose of this paper to present a formal relationship which is a natural generalization of the integral equation connecting the correlation function with the direct-correlation function in the Ornstein–Zernike theory, to derive certain properties of the mathematical objects appearing in this equation, and to place the generalized equation in the context of phase transitions and critical phenomena through an appropriately generalized version of the Gibbs picture. At the end of the paper we present a prospect suggesting how the developed formalism together with the derived properties of the mathematical objects might be exploited to yield a theory of critical phenomena. An earlier version of some of the ideas of the present paper have been given in a series of lectures by the

[1] L. S. Ornstein and F. Zernike, Proc. Roy. Acad. Amsterdam **17**, 793 (1914); F. Zernike, Proc. Roy. Acad. Amsterdam **18**, 1520 (1916).

[2] J. D. van der Waals, *Die Continuität des Gas-förmigen und Flüssigen Zustandes* (Verlag Johann Ambrosius Barth, Leipzig, 1881).

[3] J. Willard Gibbs, *The Collected Works of J. Willard Gibbs* (Longmans Green and Co. Inc., New York, 1931), Vol. I, p. 44.

[4] M. S. Green and J. V. Sengers, Eds. *Critical Phenomena: Proceedings of a Conference, Washington, D.C., 1965* (N.B.S. Misc. Publ. 223) (National Bureau of Standards, Washington, D.C., 1966).

[5] G. E. Uhlenbeck, *Critical Phenomena: Proceedings of a Conference Washington, D.C., 1965*, M. S. Green and J. V. Sengers, Eds. (N.B.S. Misc. Publ. 223) (National Bureau of Standards, Washington, D.C., 1966), p. 3.

author delivered at the Cargese Summer School, 1965.[6]

In Sec. I, we present the Ornstein–Zernike theory in a form suitable to generalization. In Secs. II, III, and IV we define the generalized correlation matrix and exhibit its relationship to the average fluctuation product of arbitrary sum functions. The fundamental maximum principle for equilibrium-statistical mechanics is stated in Sec. V and used to define the direct-correlation matrix in Sec. VI. The relationship between the direct-correlation matrix and the transformation from the formulation of statistical mechanics in terms of potentials to the formulation in terms of molecular-distribution functions is given in Sec. VII and the generalized Ornstein–Zernike relation is presented in Sec. VIII. In Sec. IX a formula for the direct correlation matrix, which is derived in Appendix B, as well as certain of its asymptotic properties, which are derived in Appendix C, are stated, while an unexpected relationship between the generalized Ornstein–Zernike relation and the Percus–Yevick equation is presented in Sec. X. In Secs. XI and XII the generalized equation is placed in the context of phase transitions and critical points with the help of the fundamental maximum principle. In Sec. XII, the concept of the critical eigenvector of the direct-correlation matrix is introduced. Section XIII is a prospect suggesting directions of future research.

I. THE ORNSTEIN–ZERNIKE THEORY OF CRITICAL FLUCTUATIONS

Before presenting the theory, which is the main subject of the present paper, we will present the Ornstein–Zernike theory in a form most parallel to our generalized theory. This is not the original version but is a combination of several recent points of view.[7,8] Since the purpose of this section is to provide a pattern for the generalized theory we use heuristic arguments freely. In Ref. 6 it has been shown that for a system in contact with a heat reservoir of temperature $T = 1/k\beta$ and with a reservoir of molecules of chemical potential μ and subject to a single-particle potential $v(x)$, the equilibrium-density distribution is that one for which the functional

$$\log \Xi = -\beta A(\rho(x), T) + \int \varphi(x)\rho(x)\, dx \qquad (1)$$

is a maximum, where

$$\varphi(x) = \beta(-v(x) + \mu), \qquad (2)$$

and $A(\rho(x), T)$ is the Helmholtz free energy given as a functional of the number density $\rho(x)$ and as a function of the temperature. The form of this functional as a series of cluster integrals is given, for instance, in Ref. 9. In a one-phase region the equilibrium-density distribution corresponding to a given $\varphi(x)$ is given by the variational equation

$$\frac{\delta \log \Xi}{\delta \rho(x)} = -\beta \frac{\delta A}{\delta \rho(x)} + \varphi(x) = 0,$$

$$\text{or} \quad \varphi(x) = \beta \frac{\delta A}{\delta \rho(x)}. \qquad (3)$$

In case there is no external potential Eq. (3) simply states

$$\mu = \frac{\delta A}{\delta \rho(x)}. \qquad (4)$$

In the neighborhood of the density distribution which satisfies Eq. (3), $\log \Xi$ may be represented as

$$\log \Xi = \log \tilde{\Xi} - \frac{\beta}{2} \int \delta\rho(x)\delta\rho(x') \times \frac{\delta^2 A}{\delta\rho(x)\delta\rho(x')}\, dx\, dx', \qquad (5)$$

where $\log \tilde{\Xi}$ is the equilibrium value of $\log \Xi$. In the absence of an external potential

$$\log \tilde{\Xi} = \beta PV. \qquad (6)$$

The quadratic functional in Eq. (5) represents the deviation of $\log \Xi$ from its maximum value due to a density fluctuation $\delta\rho(x)$. Since $\log \Xi$ is a maximum, the quadratic functional must be positive-definite or semidefinite. Thus the exponential of this quadratic form, i.e., the distribution of density fluctuations, may be approximated as a generalized Gaussian. From a theorem on multivariate normal distributions,[10] the average value of the product $\langle \delta\rho(x)\delta\rho(x') \rangle$ may be determined as the reciprocal of the kernel of the quadratic form $\delta^2 A/\delta\sigma(x)\delta\rho(x')$:

$$\int \langle \delta\rho(x)\delta\rho(x') \rangle \beta\delta^2 A/\delta\rho(x')\delta\rho(x'')dx' = \delta(x - x''). \qquad (7)$$

In the case of no external potential the average density-fluctuation product may be presented in the form

$$\rho\delta(x - x') + \rho^2 G(|x - x'|), \qquad (8)$$

[6] M. S. Green, *Cargese Lectures in Theoretical Physics; Statistical Mechanics*, B. Jancovici, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1966), p. 59. The ideas on the thermodynamics of critical phenomena of L. Tisza have given much insight into these lectures, as well as the present paper. See L. Tisza, *Phase Transformations in Solids*, R. Smoluchoeski, Ed. (John Wiley & Sons, Inc., New York, 1957), p. 1; and more extensively in Ann. Phys. (N.Y.) **13**, 1 (1961).

[7] J. L. Lebowitz and J. K. Percus, J. Math. Phys. **4**, 218 (1963).

[8] L. Landau and E. M. Lifshitz, *Statistical Physics* (Addison-Wesley Publ. Co. Inc., Reading, Mass., 1958), pp. 363, *et seq.*

[9] T. Morita and K. Hiroike, Progr. Theor. Phys. **25**, 532 (1961).

[10] H. Cramèr, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, N.J., 1951), pp. 310, *et seq.*

where $G(r)$ is the so-called pair-correlation function. In terms of the molecular-distribution functions $f(1 \cdots n)$, $G(r)$ is defined by the formula:

$$f(1)f(2)[G(r) + 1] = f(12). \tag{9}$$

$G(r)$ goes to zero for large r.

An examination of the explicit formula for $A(\rho(x), T)$ shows that its second functional derivative for a spatially uniform density may be written in the form

$$\beta\delta^2 A/\delta\rho(x)\delta\rho(x') = \frac{1}{\rho}\delta(x - x') - T(|x - x'|), \tag{10}$$

where $T(r)$ is a function of r which goes to zero for large r. For reasons connected with the original derivation of Ornstein–Zernike, $T(r)$ is called the direct-correlation function. If we substitute Eqs. (8) and (10) into Eq. (7) we obtain a relation between G and T.

$$G(x) = T(x) + \int G(x')T(|x - x'|)\,dx'. \tag{11}$$

This is the Ornstein–Zernike integral equation.

If we are dealing with a fluctuation from a spatially uniform state, the kernel T in Eq. (11) is translationally invariant. Its eigenfunctions are of the form $\exp(ik \cdot x)$, where k is an arbitrary wave vector, and its eigenvalues are

$$\lambda(k) = 1 - \rho t(k), \tag{12}$$

where $t(k)$ is the Fourier transform of $T(r)$.

The critical point corresponds to a zero of the eigenvalue $\lambda(k)$ for $k = 0$:

$$\lambda(0) = 1 - \rho t(0) = 0.$$

The eigenvalues (or Fourier components) of the kernel $\langle\delta\rho(x)\delta\rho(x')\rangle$ are, by Eq. (7), $\lambda^{-1}(k)$. Thus, near the critical point the small-k, or large-wavelength, components of the density fluctuations will have large amplitudes.

The vanishing of the eigenvalue, Eq. (12), may be viewed from another point of view. Equation (3) gives the means for determining for any density distribution the corresponding $\varphi(x)$ which will maintain this distribution in equilibrium. The small change $\delta\varphi(x)$ which corresponds to a change $\delta\rho(x)$ in the density distribution is given by taking the differential of Eq. (3)

$$\delta\varphi(x) = \int \beta\delta^2 A/\delta\rho(x)\delta\rho(x')\,dx'. \tag{13}$$

The vanishing of the eigenvalue $\lambda(0)$ means that there is a (spatially uniform) $\delta\rho$ which corresponds to no change $\delta\varphi$, i.e., in the absence of an external potential, to no change in the chemical potential. Thus we have

$$\left(\frac{\partial\mu}{\partial\rho}\right)_T = 0. \tag{14}$$

This is one of the thermodynamic conditions for a critical point. Thus at the critical point we have a singularity in the transformation from chemical potential to density. (We will discuss the complete thermodynamic characterization of the critical point in connection with the generalized theory.) Although we have only given a heuristic discussion of them, all the aspects of the Ornstein–Zernike theory, which we have presented up to this point, may be made rigorous. There is one feature of the original theory which cannot be made rigorous and is presumably not true. This is the assumption that $T(r)$ is short ranged, even at the critical point. It is this assumption, which is responsible for the well-known prediction of the Ornstein–Zernike theory that near the critical point

$$G(r) \sim \frac{\exp - \alpha r}{r}, \tag{15}$$

where $\alpha \to 0$ at the critical point.

Among the aspects of this presentation of the Ornstein–Zernike theory for which we will find generalizations in what follows are: density fluctuation, correlation function, maximum principle, direct-correlation function, transformation from chemical potential to density, and zero eigenvalue.

II. FLUCTUATIONS OF A SUM FUNCTION FROM ITS EQUILIBRIUM AVERAGE VALUE

Let us consider the fluctuations of phase functions of the form

$$\begin{aligned} A(1 \cdots N) &= \sum_{\text{all singlets}} a(i) + \sum_{\text{all pairs}} a(ij) \\ &\quad + \sum_{\text{all triples}} a(ijk) \\ &= \sum_{[\nu]\subseteq[N]} a([\nu]), \end{aligned} \tag{16}$$

where $[\nu]$ represents a subset of the particles $(1 \cdots N)$ containing ν particles and the sum is over all subsets of $1 \cdots N$.[11] The functions $a([\nu])$ have the cluster property. We may assume $a([\nu]) = 0$ for $\nu > \nu_0$.

[11] We use the following notation here and in subsequent sections: $[\nu]$ signifies the coordinates and momenta of a set of ν molecules, $d[\nu]$ signifies the phase-space volume for the ν molecules. We use the usual set-theoretic symbols to indicate conditions on summations over sets:

$[\nu] \subset [\mu]$, the set $[\nu]$ is properly included in the set $[\mu]$,
$[\nu] \subseteq [\mu]$, proper inclusion or identity,
$[\nu] \cap [\mu]$, the set consisting of all elements common to $[\nu]$ and $[\mu]$,
$[\nu] \cup [\mu]$, all elements belonging either to $[\nu]$ or to $[\mu]$,
$[\nu] \oplus [\mu]$ is defined only when $[\nu]$ and $[\mu]$ are disjoint and is then their union,
$[\mu] \ominus [\nu]$ is defined only when $[\nu] \subseteq [\mu]$ and is then the complement of $[\nu]$ in $[\mu]$,
$i \in [\nu]$, i is a member of $[\nu]$,
ν means the number of molecules in the set $[\nu]$.

Sometimes we use a set symbol as an exponent of (-1) in which case the number of particles in the set is meant.

A summation over sets is indicated by the usual summation sign and means summation over all sets satisfying the conditions which appear below the summation sign.

Such functions have been called sum functions because the value of such a function for the union of two widely separated sets of molecules is just the sum of its value for each group separately.

If the functions $a([\nu])$, for example, are different from zero only when all the particles of the set $[\nu]$ lie within a certain region of ordinary space, A represents a local property of the system. Thus for example if $a(i)$ is simply $(1/\Omega_R)E_R(x_i)$, where $E_R(x)$ is the characteristic function of a region R of ordinary space and Ω_R its volume, A represents the number density of particles in R. If the $a([\nu])$ are translationally invariant, A represents a global property of the system. Thus for

$$\begin{aligned} a(i) &= \frac{p_i^2}{2m}, \\ a(ij) &= v(|x_i - x_j|), \\ a(ijk), \cdots \text{etc.} &= 0, \end{aligned} \tag{17}$$

A is the total energy of the system under the assumption of only pairwise-intermolecular forces. If for example,

$$\begin{aligned} a(i) &= \frac{1}{\Omega_R}\frac{p_i^2}{2m}E_R(x_i), \\ a(ij) &= \frac{1}{\Omega_R}v(|x_i - x_j|)E_R(x_i)E_R(x_j), \\ a(ijk)\,\text{etc.} &= 0, \end{aligned} \tag{18}$$

A represents the energy density in the region R.

The equilibrium average value of A may be easily expressed in terms of the reduced equilibrium molecular-distribution functions

$$\langle A\rangle = \sum_{n=1}^{\infty}\frac{1}{n!}\int a(1\cdots n)f(1\cdots n)\,d(1)\cdots d(n). \tag{19}$$

The mean-square fluctuation of a sum function A, $\langle(A - \langle A\rangle)^2\rangle$, or the mean product of the fluctuation of A and the fluctuation of another sum function B, $\langle(A - \langle A\rangle)(B - \langle B\rangle)\rangle$,

$$B(1\cdots N) = \sum_{[\nu]\subseteq(1\cdots N)} b([\nu]), \tag{20}$$

may also be expressed in terms of the molecular-distribution functions as a bilinear form in the a's and b's:

$$\langle(A - \langle A\rangle)(B - \langle B\rangle)\rangle = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{n!}\frac{1}{m!}\int d[n]\,d[m]a([n])b([m])\Gamma([n],[m]), \tag{21}$$

$$\Gamma([n],[m]) = f([n]\oplus[m]) - f([n])f([m]) + \sum_{\substack{0\subset[\kappa]\subseteq[n]\\ 0\subset[\kappa']\subseteq[m]}}\delta([\kappa],[\kappa'])f([n]\cup[m]). \tag{22}$$

In Eq. (22), $\delta([\kappa],[\kappa'])$ is a delta function which contributes only when the set of particles $[\kappa]$ coincides with the set $[\kappa']$. The delta function $\delta([\kappa],[\kappa'])$ may be defined by the property that for any sequence of functions $K([\kappa])$:

$$K([\kappa]) = \sum_{\kappa=1}^{\infty}\frac{1}{\kappa!}\int\delta([\kappa],[\kappa'])K([\kappa'])\,d[\kappa']. \tag{23}$$

Note that for configuration in which none of the particles of $[m]$ coincides with a particle of $[n]$, which is in some sense most configurations, only the first two terms of Eq. (22) contribute to Γ. When some of the particles coincide, Γ is given by a delta function corresponding to the coincident group of particles multiplied by the distribution function $f([\mu]\cup[\nu])$ of all distinct particles. An explicit representation of $\delta([\kappa],[\kappa'])$ is given in Appendix A.

The matrix $\Gamma([m],[n])$ may be considered to be one generalization of the average density-fluctuation product $\langle\Delta\rho(x)\Delta\rho(x')\rangle$, Eq. (8). Indeed, if we suppose that a fluctuation in the value of a sum function ΔA from equilibrium arises as a consequence of a fluctuation Δf in the distribution functions from their equilibrium average values we may write

$$\Delta A = \sum_{n=1}^{\infty}\frac{1}{n!}\int a([n])\Delta f([n])\,d[n]. \tag{24}$$

If we form the average $\langle\Delta A\Delta B\rangle$, we see that we have the interpretation

$$\Gamma([m],[n]) = \langle\Delta f[m]\Delta f([n])\rangle. \tag{25}$$

III. POTENTIALS OF AVERAGE FORCE

In the generalized Ornstein–Zernike theory of the present paper another matrix arises rather naturally which is closely related to $\Gamma([m],[n])$ and which may be considered to be the average fluctuation product of the potentials of average force. It will be to this matrix, rather than Γ, which we will give the name correlation matrix, and which we will take to be the analog of the correlation function of the Ornstein–Zernike theory. Before we consider this matrix in detail let us digress to discuss the representation of the molecular-distribution functions in terms of the potentials of average force.

The system of molecular-distribution functions $f([n])$ represent a complete statistical description of any system of a large (even indefinite) number of identical molecules in phase space. For a system consisting of a single phase in thermodynamic equilibrium these functions may not be chosen arbitrarily but must be assumed to have the product property

$$f([n]\oplus[m]) \to f([n])f([m]), \tag{26}$$

when the group of molecules $[m]$ is removed far from the group $[n]$. The molecular-distribution functions do not represent a mutually independent set of functions. The information contained in any molecular-distribution function redundantly expresses information already contained in the distribution function for a group of fewer molecules. One way of representing the information contained in the molecular-distribution functions nonredundantly is through the sequence of potentials of average force. This sequence of functions is defined by the formulas

$$\begin{aligned} \log f(1) &= \psi(1), \\ \log f(12) &= \psi(1) + \psi(2) + \psi(12), \\ \log f([n]) &= \sum_{[\nu]\subseteq[n]} \psi([\nu]). \end{aligned} \tag{27}$$

It is easily verified that, if the $f([n])$ have the product property, the $\psi([n])$ have the cluster property; i.e.,

$$\psi([n] \oplus [m]) \to 0$$

when the groups $[n]$ and $[m]$ are separated far apart. Conversely, when the $\psi([n])$ have the cluster property, the $f([n])$ have the product property. Since the limiting value of the $\psi([n])$ for widely separated groups is zero, the later members of the sequence do not contain information about the earlier members. The $\psi([n])$ are a mutually independent sequence of functions which contains the same information as the sequence of molecular-distribution functions. In order to assure that the $f([n])$ are positive we need only assume that the $\psi([n])$ are real.

Equation (27) may be inverted by the formulas

$$\begin{aligned} \psi(1) &= \log f(1), \\ \psi(12) &= -\log f(1) - \log f(2) + \log f(12), \\ \psi([m]) &= \sum_{[\nu]\subseteq[m]} (-1)^{m-\nu} \log f([\nu]). \end{aligned} \tag{28}$$

IV. THE CORRELATION MATRIX

The matrix $\Gamma([m], [n])$ has the property that an element is different from zero only if some of the members of the sets $[m]$ and $[n]$ are close to each other. It is, however, by no means necessary for a nonzero $\Gamma([m], [n])$ that all members of the sets $[m]$ and $[n]$ must be close together. Indeed the value of $\Gamma([m], [n])$ depends essentially on the subsets of $[m]$ and $[n]$, which are close enough to be mutually correlated to each other, and only trivially on the remaining molecules. Thus let $[m']$ and $[n']$ be the subsets of $[m]$ and $[n]$ which together contain all groups of molecules having members from both $[m]$ and $[n]$ which are mutually correlated and let $[m'']$ and $[n'']$ be the remaining molecules belonging to $[m]$ and $[n]$ respectively.

Then

$$f([m] \cup [n]) = f([m'] \cup [n'])f([m''])f([n'']), \tag{29}$$

since $[m'']$ and $[n'']$ may not be close to each other nor to $[m']$ or $[n']$, and since $[m'] \cap [n'] = [m] \cap [n]$.

Thus we have from Eqs. (22) and (29)

$$\begin{aligned} &\Gamma([m'] \oplus [m''], [n'] \oplus [n'']) \\ &\qquad = \Gamma([m'], [n'])f([m''])f([n'']). \end{aligned} \tag{30}$$

Equation (30) suggests that $\Gamma([m], [n])$ may be constructed from another matrix $G([m], [n])$ which is different from zero only for groups $[m]$ and $[n]$ such that all the members of $[m]$ are close enough to members of $[n]$ to be mutually correlated and vice versa. This is indeed the case.

Let us define $G([m], [n])$ by the relation

$$\Gamma([m], [n]) = \sum_{\substack{[\nu]\subseteq[n] \\ [\mu]\subseteq[m]}} f([m])f([n])G([\mu], [\nu]). \tag{31}$$

Noting the inverse relationship between the operations

$$\sum_{[\nu]\subseteq[n]} \quad \text{and} \quad \sum_{[\nu]\subseteq[n]} (-1)^{n-\nu},$$

we may write

$$\begin{aligned} &G([m], [n]) \\ &\quad = \sum_{\substack{[\nu]\subseteq[n] \\ [\mu]\subseteq[m]}} (-1)^{(n-\nu)+(m-\mu)}\Gamma([\mu], [\nu])f([\mu]f([\nu]). \end{aligned} \tag{32}$$

Now, if, in Eq. (32), the set $[n]$ has a subset $[n'']$, which is far from both $[m]$ and its complement $[n']$ in $[n]$, $G([m], [n])$ is zero, for, writing $[\nu'] \subseteq [n']$, $[\nu''] \subseteq [n'']$, and $[\nu'] \oplus [\nu''] = [\nu]$, we have by Eq. (30)

$$\Gamma([\mu], [\nu]) = \Gamma([\mu], [\nu'])f([\nu'']), \tag{33}$$

and by the product property Eq. (26)

$$f([\nu]) = f([\nu'])f([\nu'']).$$

Thus the summand in Eq. (32) depends only on the number of particles in $[\nu']$. Carrying out the summation over ν'', keeping $[\mu]$ and $[\nu']$ fixed, gives zero. The matrix $G([m], [n])$ is different from zero only if there is no set $[n'']$; i.e., unless all molecules of $[n]$ are correlated to $[m]$ and vice versa.

By Eq. (27), we may write for a small fluctuation in the distribution function

$$\Delta f([n]) = f([n]) \sum_{[\nu]\subseteq[n]} \Delta\psi([\nu]) \tag{34}$$

and, forming the average fluctuation product, we have

$$\langle \Delta f([n]) \Delta f([m]) \rangle = f([n]) f([m]) \sum_{\substack{[\nu] \subseteq [n] \\ [\mu] \subseteq [m]}} \langle \Delta\psi([\nu]) \Delta\psi([\mu]) \rangle; \quad (35)$$

i.e., by Eq. (31)

$$G([m], [n]) = \langle \Delta\psi([m]) \Delta\psi([n]) \rangle. \quad (36)$$

As we have remarked above, $G([m], [n])$ will be called the correlation matrix.

It will be useful to express the fluctuation of sum functions in terms of the fluctuations of the potentials of average force. This may be done with the help of Eq. (34). We have

$$\Delta A = \sum \frac{1}{n!} \int a([n]) \Delta f([n])\, d[n] = \sum \frac{1}{n!} \int a([n]) f([n]) \sum_{[\nu] \subseteq [n]} (\Delta\psi([\nu]))\, dn. \quad (37)$$

If we break up the integration over $[n]$ into an integration over the complement of $[\nu]$ in $[n]$, say $[h]$, and then over $[\nu]$ noting that there are $n!/\nu!\, h!$ ways of choosing the set $[\nu]$ from the set $[n]$ without changing the value of the integral, we obtain

$$\Delta A = \sum_{\nu} \sum_{h} \frac{1}{\nu!} \frac{1}{h!} \times \int a([\nu] \oplus [h]) f([\nu] \oplus [h]) \Delta\psi([\nu])\, d[\nu]\, d[h]. \quad (38)$$

This linear form in the $\Delta\psi([\nu])$ may be written

$$\Delta A = \sum \frac{1}{\nu!} \int \alpha(\nu) \Delta\psi([\nu]) f([\nu])\, d[\nu], \quad (39)$$

where the coefficient of $\Delta\psi([\nu])$ is defined by

$$\alpha([\nu]) f([\nu]) = \sum_{h=0}^{\infty} \frac{1}{h!} \int a([\nu] \oplus [h]) f([\nu] \oplus [h])\, d[h]. \quad (40)$$

The sequence of functions $\alpha([\nu])$ characterizes the sum function A just as the sequence $a([\nu])$ does. However, the $\alpha([\nu])$ depend on the $f([\nu])$ and are therefore especially suited to represent the fluctuations A around an equilibrium state described by the distributions $f([\nu])$.

If we form the average fluctuation product using Eq. (39), and the corresponding equations for ΔB, we obtain

$$\langle \Delta A \Delta B \rangle = r \sum_{\nu!\, \mu!} \frac{1}{\nu!} \frac{1}{\mu!} \times \int d[m]\, d[n] f([n]) f([m]) G([m], [n]) \alpha([n]) \beta([m]). \quad (41)$$

Thus the correlation matrix $G([m], [n])$ is the kernel of the bilinear form which expresses the average fluctuation produce $\langle \Delta A \Delta B \rangle$ in terms of the $\alpha([\nu])$ and $\beta([\nu])$.

V. FUNDAMENTAL MAXIMUM PRINCIPLE

Gibbs[12] showed that the equilibrium distribution in phase space is that normalized distribution which makes a certain functional attain its maximum value. In Ref. 6 it was shown that this maximum principle can be expressed in a form in which the independent variable is the system of molecular-distribution functions. In this section we will restrict ourselves to one-phase situations, i.e., to systems of molecular distributions which have the product property.

Let us denote by the sequence of functions $\varphi([\nu])$ the following quantities:

$$\varphi(1) = \beta\left[\mu - \frac{p_1^2}{2m} + v(x_1)\right], \quad \varphi(12) = -\beta v(x_1 x_2), \quad \varphi(123) = -\beta v(x_1 x_2 x_3), \quad \text{etc.}, \quad (42)$$

where $\beta = 1/kT$; μ, the chemical potential; $v(x_1)$, $v(x_1x_2)$, $v(x_1x_2x_3)$, etc. are the singlet, pair, triplet, etc. potentials, respectively. Like the functions ψ introduced earlier, the φ's have the cluster property. The fundamental maximum principle states that in equilibrium the quantity

$$\log \Xi = S(\psi) + \sum_{\nu=1}^{\infty} \frac{1}{n!} \int \psi([n]) f([n])\, d[n] \quad (43)$$

is a maximum among all systems of molecular distribution functions having the product property. In Eq. (43) $S(\psi)$ is the expression for the entropy as a functional of the distribution functions derived by Nettleton and Green.[13] The functional $S(\psi)$ is naturally expressed in terms of a set of potentials of average force having the cluster property. Although the fundamental maximum principle in the form given in Eq. (43) is only valid for values of the ψ's corresponding to single-phase states, we shall see, nevertheless, that it is very useful in describing phase transitions.

Assuming functional differentiability of $\log \Xi$ in the neighborhood of the values of the ψ's which correspond to the maxima, these latter must satisfy the

[12] J. Willard Gibbs, "Elementary Principles of Statistical Mechanics," *Collected Works of J. Willard Gibbs* (Longmans Green and Co., Inc., New York), Vol. II.

[13] R. E. Nettleton and M. S. Green, J. Chem. Phys. **29**, 1365 (1958).

Euler equations

$$\frac{\delta \log \Xi}{\delta \psi([n])} = 0. \tag{44}$$

Taking note of Eq. (27), we have

$$\frac{\delta S}{\delta \psi([n])} = -\sum_{k=0}^{\infty} \frac{1}{k!} \int \varphi([n] \oplus [k]) f([n] \oplus [k])\, d[k]. \tag{45}$$

This equation may be solved for the φ's:

$$-\varphi([n]) = \frac{1}{f([v])} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int \frac{\delta S}{\delta \psi([n+h])}\, d[h]. \tag{46}$$

For any fixed φ's there may be several sets of values of the ψ's which satisfy the Euler equations (45) or (46). Some of these may represent local maxima, while others may represent stationary points which are not maxima. Of these solutions to the Euler equations, the maximum principle picks out as the equilibrium solution the one corresponding to the largest local maximum. The value of $\log \Xi$ corresponding to the largest local maximum is the logarithm of the grand-partition function, i.e., βPV, where P is the equilibrium pressure.

VI. DIRECT CORRELATION MATRIX: DEFINITION

Taking a cue from the Ornstein–Zernike theory presented above, we may consider the quantity $\log \Xi$ to be a thermodynamic potential whose natural variables are the $\psi([v])$. The probability of a fluctuation $\delta\psi([v])$ from equilibrium may then be considered to be proportional to $\exp(\Delta \log \Xi)$, where $\Delta \log \Xi$ is the change in $\log \Xi$ corresponding to the $\delta\psi$'s. Recalling that the vanishing of the linear terms was our condition for equilibrium, Eq. (44), we have, up to terms of second order in the $\delta\psi$'s,

$$\begin{aligned} \Delta \log \Xi &= \sum_{n=m=1} \frac{1}{n!} \frac{1}{m!} \int \delta\psi([n]) \delta\psi([m]) \\ &\quad \times \frac{\delta^2 \log \Xi}{\delta\psi([n]) \delta\psi([m])}\, d[m]\, d[n] \\ &= \log \Xi - \log \tilde{\Xi}, \end{aligned} \tag{47}$$

where $\tilde{\Xi}$ is the equilibrium and therefore maximum value of $\log \Xi$ [compare Eq. (5)]. If we define the direct-correlation matrix to be

$$T([m], [n]) = -\frac{\delta^2 \log \Xi}{\delta\psi([n]) \delta\psi([m])}, \tag{48}$$

we may expect heuristically that the probability of a fluctuation $\delta\psi[n]$ is proportional to

$$\tilde{\Xi} \exp - \sum_{n,m=1}^{\infty} \frac{1}{n!} \frac{1}{m!} \times \int \delta\psi([n]) \delta\psi([m]) T([m], [n])\, d[m]\, d[n]. \tag{49}$$

Since Eq. (49) represents a generalized Gaussian distribution we expect, again heuristically, that the correlation matrix

$$\langle (\delta\psi([n]) \delta\psi([m])) \rangle$$

and the direct-correlation matrix

$$T([m], [n])$$

are reciprocals.[10]

VII. TRANSFORMATION BETWEEN PHYSICAL POTENTIALS AND POTENTIAL OF AVERAGE FORCE

In our discussion of the Ornstein–Zernike theory we pointed out that the direct-correlation function can be represented as the functional derivative $\delta\mu(x)/\delta\rho(x')$, i.e., as the derivative matrix of the transformation between chemical potential and density [Eq. (3)]. Equation (48) represents a much more general transformation between the physical potentials $\varphi([m])$ and the potentials of average force $\psi([m])$. The functional-derivative matrix of this transformation $\delta\varphi([m])/\delta\psi([n])$ is not, as might be expected from this analogy, the direct-correlation matrix. We will show, however, that the direct-correlation matrix is, nevertheless, very closely related to the derivative matrix of the transformation from φ's to ψ's.

If we take the first functional derivative of Eq. (43), we obtain

$$\frac{\delta \log \Xi}{\delta \psi([n])} = \frac{\delta S}{\delta \psi([n])} + \sum_{k=0}^{\infty} \frac{1}{k!} \times \int \varphi([n] \oplus [k]) f([n] \oplus [k])\, d[k]. \tag{50}$$

In Eq. (45) the conditions for equilibrium in physical potentials φ are obtained by setting the right-hand side of Eq. (50) equal to zero. We are also interested in nonequilibrium values of the ψ's. We may, however, define physical potentials φ corresponding to these nonequilibrium values of the ψ's through the equations

$$\frac{\delta S}{\delta \psi([n])} = -\sum_{k=0}^{\infty} \int \frac{1}{k!} f([n] \oplus [k]) \bar{\varphi}([n] \oplus [k])\, d[k], \tag{51}$$

where $\bar{\varphi} = \varphi$ when the ψ's are the equilibrium ψ's. The $\bar{\varphi}$'s will be related to the ψ's through the explicit

equation (46) with $\tilde{\varphi}$'s substituted for φ's. Substituting Eq. (51) in to Eq. (50), we obtain

$$\frac{\delta \log \Xi}{\delta\psi([n])} = \sum_{k=0}^{\infty} \frac{1}{k!} \int \{\tilde{\varphi}([n] \oplus [k]) - \varphi([n] \oplus [k])\} \times f([n] \oplus [k])\, d[k]. \quad (52)$$

Taking functional derivatives again we have

$$\frac{\delta^2 \log \Xi}{\delta\psi([n])\delta\psi([m])} = -\sum_{k=0} \frac{1}{k!} \int \left\{ \frac{\delta\tilde{\varphi}([n] \oplus [k])}{\delta\psi([m])} f([n] \oplus [k]) + \sum_{k=0}^{\infty} \frac{1}{k!} \int (\tilde{\varphi}([n] \oplus [k]) - \varphi([n] \oplus [k]) f([n] \oplus [k]) \right\} d[k]. \quad (53)$$

Now the direct-correlation matrix is defined to be minus the second functional derivative of log Ξ evaluated for the equilibrium ψ's. For these ψ's, $\varphi = \tilde{\varphi}$ and the second group of terms on the right vanish. Thus we have

$$T([m], [n]) = \sum_{k=0}^{\infty} \int \frac{1}{k!} \int d[k] \frac{\delta\varphi([n] \oplus [k])}{\delta\psi([m])} f([n] \oplus [k]), \quad (54)$$

where we have dropped the bar on the φ's understanding these to be related to the ψ's through Eq. (46). In spite of the apparently asymmetric form of the right-hand side, we know from its definition that T is symmetric in $[m]$ and $[n]$.

VIII. GENERALIZED ORNSTEIN–ZERNIKE RELATION

In Ref. 6 it was pointed out that the relationship

$$\sum_{n=0}^{\infty} \frac{1}{m!} \int \frac{\delta\varphi([m])}{\delta\psi([n])} \frac{\delta\psi([n'])}{\delta\varphi([m])} d[m] = \delta([n], [n']), \quad (55)$$

expressing the fact that the derivative matrices of the transformation from φ's to ψ's and from ψ's to φ's are reciprocal, can be considered to be a generalization of the Ornstein–Zernike integral equation. In the analogy presented there, $\delta\psi/\delta\varphi$ was the analog of the direct correlation function, Eq. (10), while $\delta\varphi/\delta\psi$ was the analog of the correlation function Eq. (8).

In this section we will show that Eq. (55) can be transformed into a relationship between the direct-correlation matrix and the correlation matrix. Indeed we will demonstrate the statement that was proposed heuristically above: that these two matrices are reciprocal. In appendix A, it is shown that the first and second functional derivatives of the functional of log $\tilde{\Xi}(\varphi)$, which is the maximum value of the functional log $\Xi(\psi, \varphi)$ for fixed φ, are respectively:

$$\delta \log \tilde{\Xi}/\delta\varphi(m) = f([m]), \quad (56)$$

$$\delta^2 \log \tilde{\Xi}/\delta\varphi([n])\delta\varphi([m]) = \Gamma([n], [m]), \quad (57)$$

where Γ is defined above in Eq. (22). Thus we may write

$$\frac{\delta f([n])}{\delta\varphi([m])} = \Gamma([n], [m]). \quad (58)$$

Now the left-hand side of Eq. (58) may be easily written in terms of $\delta\psi/\delta\varphi$ while the right hand side may be written in terms of

$$G([m], [n]) \equiv \langle \delta\psi([m])\delta\psi([n]) \rangle.$$

Differentiating Eq. (27) and using Eq. (31), we obtain

$$f([n]) \sum_{[\nu]\subseteq[n]} \frac{\delta\psi([\nu])}{\delta\varphi([m])} = f([m])f([n]) \sum_{\substack{[\nu]\subseteq[n] \\ [\mu]\subseteq[m]}} G([\nu], [\mu]). \quad (59)$$

Thus we have

$$\frac{\delta\psi([n])}{\delta\varphi([m])} = f([m]) \sum_{[\mu]\subseteq[m]} G([\mu], [n]). \quad (60)$$

Returning to Eq. (55), we have

$$\sum_{m=0}^{\infty} \frac{1}{m!} \int d[m] \frac{\delta\varphi[m]}{\delta\psi[n]} f([m]) \sum_{[\mu]\subseteq[m]} G([\mu,], [n']) = \delta([n], [n']). \quad (61)$$

If we break up the set $[m]$ into sets $[\mu]$ and $[k]$, we note that, for fixed μ and k,

$$\sum_{\mu=1}^{\infty} \frac{1}{\mu!} \int d[\mu] \left[\sum_{k=0} \frac{1}{k!} \int d[k] \frac{\delta\varphi([\mu] \oplus [k])}{\delta\psi([n])} f([\mu] \oplus [k]) \right] \times G([\mu], [n']) = \delta([n], [n']). \quad (62)$$

The expression in square brackets is nothing else than $T([n], [\mu])$, Eq. (54). Equation (62) simply states that the matrixes T and G are reciprocals.

$$\sum_{\mu=1}^{\infty} \frac{1}{\mu!} \int d[\mu] T([n], [\mu]) G([\mu][n']) = \delta([n], [n']) \quad (63)$$

We take Eq. (63) to be the appropriate generalization of the Ornstein–Zernike integral equation.

IX. DIRECT-CORRELATION MATRIX: PROPERTIES

If the generalized Ornstein–Zernike equation were the only result of the present approach we would have very little hope of going beyond the classical theory. It is possible, however, to derive at least for two-body forces a formula for the direct-correlation matrix from which a number of important properties can be

established. These properties make the direct-correlation matrix a much better known mathematical object, especially in its asymptotic properties, than the direct-correlation function of the Ornstein–Zernike theory.

In Appendix C we derive the following formula for $T([m], [n])$ for the case in which the φ's are zero for ν greater than two:

$$T([m], [n]) = \sum_{\substack{0\subseteq[\beta]\subseteq[m] \\ 0\subseteq[\beta']\subseteq[n]}} f([m] \cup [n])\delta([\beta], [\beta']) \times \sum_{\substack{0\subseteq[\alpha]\subseteq[\bar{m}] \\ 0\subseteq[\alpha']\subseteq[\bar{n}]}} (-1)^{\alpha+\alpha'} \exp - \sum_{\substack{i\in[\alpha] \\ j\in[\alpha']}} \varphi(i, j) \quad (64)$$

where $[\bar{m}] = [m] - [m] \cap [n]$,

$$[\bar{n}] = [n] - [m] \cap [n],$$

and where, in case $[\beta]$ or $[\beta']$ or both are empty, the exponential is to be replaced by unity.

First of all we note in Eq. (64) that each element of the matrix is given by a finite number of terms involving the physical potentials and distribution functions. This is in contrast to the direct-correlation function in the Ornstein–Zernike theory for which there exist several expressions in terms of f's and φ's which are, however, infinite series. Since $\log \Xi$ is a maximum for equilibrium, $T([m], [n])$ is a positive-definite matrix.

Secondly, whenever any single point j of the set $[n]$ is so far from each point of the set $[m]$ that all physical pair potentials are zero, $T([m], [n])$ is zero. This fact is proved in Appendix C. This means that $T([m], [n])$ is zero unless all parts of the set $[n]$ are within the range of pairwise forces of a part of $[m]$ and vice versa. This is analogous to a property of the correlation matrix $G([m], [n])$ [Eq. (32) *et seq.*]; i.e., for a nonzero value for G, all parts of $[m]$ must be close to a part of $[n]$ *within a range determined by the product property of the f's.* With respect to the distance between the sets $[m]$ and $[n]$, $T([m], [n])$ is short ranged with a range which is independent of the thermodynamic state. This is in contrast to the direct-correlation function which seems to become long ranged at the critical point

$$T([m], [n]) \to 0, \quad (65)$$

when the distance between $[m]$ and $[n]$ is larger than the range of intermolecular forces.

Thirdly, when the sets $[m]$ and $[n]$ are each divided into two widely separated parts $[m']$, $[m'']$, and $[n']$, $[n'']$, respectively, $T([m], [n])$ breaks up into a product

$$T([m'] \oplus [m''], [n'] \oplus [n''\]) \to T([m'], [n'])T([m''], [n'']). \quad (66)$$

This property is also proved in Appendix C. The distance between the sets $[m']$ and $[m'']$, $[n']$ and $[n'']$ for Eq. (66) to be valid is the range of correlations not the range of forces.

X. PERCUS–YEVICK EQUATION

At this point it is worth while to note that if the expression for $T([m], [n])$, given by Eq. (64), is inserted into the generalized Ornstein–Zernike equation, Eq. (63), the result is a self-contained and explicit matrix relationship between the distribution functions and the physical potentials. This is in contrast to the classical Ornstein–Zernike integral equation, which is basically a relationship between two unknown functions.

Several approximate integral equations for the correlation functions have been derived by inserting appropriate approximate formulas for the direct-correlation function in the Ornstein–Zernike theory. It is an interesting fact that the most successful of these, the Percus–Yevick equation,[14] can be "derived" very naturally in the context of the expression, Eq. (64), for the direct-correlation matrix. Indeed, if we specialize Eq. (64) to the case $m = n = 1$, we obtain

$$T(1, 1') = f(1)\delta(1, 1') + f(1, 1')(\exp -\varphi(1, 1') - 1). \quad (67)$$

This is just the approximation to the direct-correlation function which gives the Percus–Yevick equation. From Eqs. (22) and (32) we find that

$$G(1, 1') = \frac{1}{f(1)}\delta(11') + \left(\frac{f(11')}{f(1)f(1')} - 1\right). \quad (68)$$

Equation (63) then becomes the Percus–Yevick equation, if all other elements of $T([m], [n])$ are set equal to zero, or, less drastically, if all elements $T(1, [m])$ are set equal to zero.

XI. PHASE TRANSITIONS

In the previous paragraphs we have discussed situations in which $\log \Xi$ has a unique maximum in the neighborhood of which it may be expanded in a functional-power series in the $\delta\psi$'s. The equilibrium state in this case is a single phase. It may happen that two local maxima are equal, i.e., that for the same set of φ's, two different sets of ψ's correspond to equal local maximum values of $\log \Xi$. For this case a small change in the φ's, that is to say a small change of external conditions, will only slightly change the ψ's

[14] J. K. Percus, Phys, Rev. Letters, 462 (1962); *The Equilibrium Theory of Classical Fluids*, H. L. Frisch and J. L. Lebowitz, Eds. (W. A. Benjamin, Inc., New York, 1964), Vol. II, pp. 33, *et seq.*

corresponding to each of the local maxima, but will make one or the other set the one which corresponds to the absolute maximum of $\log \Xi$. Thus a small change of φ's will cause a finite change in the equilibrium ψ's. This is a first-order phase transition.

Let us consider the entropy $S(\psi_L)$, $S(\psi_G)$ for the two phases in thermodynamic equilibrium. If we denote by $\log \tilde{\Xi}$ the common value of $\log \Xi$ for the two phases, say liquid and gas, the two functionals of the ψ's, $S(\psi)$ and

$$\log \tilde{\Xi} - \sum_{n=1}^{\infty} \frac{1}{n!} \int \varphi([n]) f([n])\, d[n]$$

are equal, both for $\psi = \psi_L$ and $\psi = \psi_G$. This is a consequence of the equality of the local maxima and of a rearrangement of Eq. (43). Equation (45) in its turn states that the functional derivatives of these two functionals are equal for all equilibrium states and in particular for ψ_L and ψ_G. Thus the surface

$$S = S(\psi) \tag{69}$$

and the surface

$$S = \log \tilde{\Xi} - \sum_{n=1}^{\infty} \frac{1}{n!} \int \varphi([n]) f([n])\, d[n] \tag{70}$$

are doubly tangent, once at $\psi = \psi_L$ and once at $\psi = \psi_G$.

We see that the fundamental maximum principle, when expressed in terms of the potentials of average force, yields a picture of a first-order phase transition which is very analogous to the Gibbs picture referred to in the Introduction. In the Gibbs picture,[3] the quantity

$$S(E, V) - \frac{E + PV}{T} = G \tag{71}$$

is a maximum for fixed P and T considered as a function of E and V. A first-order phase transition corresponds to two equal maxima of this quantity. If $\tilde{G}$ is the common value of G for the two maxima, the surface

$$S = S(E, V) \tag{72}$$

and the plane

$$S = \tilde{G} + \frac{E + PV}{T} \tag{73}$$

are doubly tangent at the points (E_L, V_L) and (E_G, V_G) corresponding to liquid and gas respectively. The points on the line joining (S_L, E_L, V_L) and (S_G, E_G, V_G),

$$((XS_L + (1 - X)S_G), (XE_L + (1 - X)E_G), (XV_L + (1 - X)V_G)),$$

$0 \leq X \leq 1$, are also equilibrium points corresponding to the coexistence of the two phases in the proportion of X to $1 - X$. The states of equilibrium belong to the convex hull of the surface $S(E, V)$. Those points which lie on $S(E, V)$ itself are single phase points. Those which lie on planes multiply-tangent to $S(E, V)$ correspond to several phases in equilibrium.

It has been shown[15] that the distribution function

$$f([n]) = Xf_L([n]) + (1 - X)f_G([n]) \tag{74}$$

corresponds to the coexistence of liquid and gas. We may assume that equilibrium states correspond to the convex hull of the surface $S(\psi)$ in the space of molecular-distribution functions. Distribution functions of the form Eq. (74), corresponding to the coexistence of several phases, do not have the product property.

XII. THE CRITICAL POINT AND THE CRITICAL EIGENVECTOR

In the Gibbs picture it may happen that as the doubly tangent plane rolls on the entropy surface the points of contact approach each other until they coalesce. This means that the characteristics of the liquid and gas phases approach each other and become identical at the critical point. This picture of the critical point has a very close analogy in the present formalism. It may happen that as the external parameters contained in the φ's are varied, the set of distribution functions and the potentials of average force corresponding to the liquid f_L, ψ_L and the set corresponding to the gas f_G, ψ_G become identical. Such a set of values is the critical point.

In the neighborhood of the critical point, the behavior of the functional $\log \Xi$ is given by the quadratic form, Eqs. (47) and (48). It is easy to make plausible that at the critical point the matrix $T([m], [n])$ will have an eigenvector with eigenvalue zero and the "surface" described by $\log \Xi$ as functional of the ψ's is cylindrical at the critical point. Let us consider, with fixed intermolecular potentials, the change in the potentials of average force which take place when the density of the system is changed, keeping the temperature constant and equal to T_c. The compressibility is infinite at the critical point which implies that

$$\left(\frac{\partial \mu}{\partial \rho}\right)_T = \left(\frac{\partial \varphi(1)}{\partial \rho}\right)_T = 0 \tag{75}$$

at T_c, i.e., $\delta\varphi(1) = 0$ in such a displacement. At the same time, since the temperature and intermolecular potentials are constant, we have

$$\delta\varphi([n]) = 0, \tag{76}$$

[15] M. E. Fisher, J. Math. Phys. 6, 1643 (1965).

$n > 1$. In such a change the $\delta\psi([n])$, $n > 1$, will also change. The two sets of changes will be related by the derivative matrix

$$\delta\varphi([n]) = \sum_{m=1}^{\infty} \frac{1}{m!} \int \frac{\delta\varphi([n])}{\delta\psi[m]} \delta\psi([m])\, d[m]. \quad (77)$$

For $n = 1$ the left-hand side of Eq. (77) will be zero because the inverse compressibility is zero. For $n > 1$, the left-hand side of Eq. (77) is zero because then the $\varphi[n]$ are constant along an isotherm. If we divide the isothermal changes $\delta\psi([n])$, $n > 1$, by $\delta\psi(1) = \delta\rho/\rho$, assuming the limits exist, the vector $\delta([m])$ with components

$$\begin{aligned} \delta(1) &= 1, \\ \delta(12) &= \rho \frac{\delta\psi(12)}{\delta\rho}\bigg|_{T=T_c}, \\ \delta(123) &= \rho \frac{\delta\psi(123)}{\delta\rho}\bigg|_{T=T_c} \end{aligned} \quad (78)$$

is an eigenvector of the matrix $\delta\psi([m])/\delta\varphi([n])$, with eigenvalue zero

$$0 = \sum_{n=1}^{\infty} \frac{1}{m!} \int \frac{\delta\varphi([n])}{\delta\psi([m])} \delta([m])\, d[m]. \quad (79)$$

If we replace $[n]$ by $([m] \oplus [h])$ in Eq. (79), multiply by $f([n] \oplus [h])$, integrate, and sum, we obtain

$$0 = \sum_{h=0}^{\infty} \sum_{m=1}^{\infty} \frac{1}{m!} \frac{1}{h!} \int f([n] \oplus [h]) \frac{\delta\varphi([n] \oplus [h])}{\delta\psi([m])} \times \delta([m])\, d[m]. \quad (80)$$

Noting Eq. (54) and the symmetry of $T([m], [n])$, we have

$$0 = \sum_{m=1}^{\infty} \frac{1}{m!} \int T([n], [m])\delta([m]\, d[m]); \quad (81)$$

i.e., $\delta([m])$ is an eigenvector of $T([m], [n])$ with eigenvalue zero.

In the Gibbs picture of phase transitions the boundary of the two-phase region does not represent a locus of singularities of the single-phase E, S, V surface, but simply the locus of the points of tangency of a double tangent. Gibbs also assumed the existence of a locus of incipient mechanical instability, i.e., a locus of infinite compressibility on the *single-phase* E, S, V, surface which lies in the interior of the equilibrium E, S, V surface except at the critical point. By the above argument, the matrix $T([m], [n])$ will have an eigenvector with zero eigenvalue at every point of this locus. Without committing ourselves either favorably or unfavorably to this aspect of the Gibbs picture, we may make a further characterization of the critical point and of the eigenvector with zero eigenvalue which corresponds to it.

Let us consider the differences

$$\Delta\psi([m]) = \psi_L([m]) - \psi_G([m]) \quad (82)$$

on the phase boundary. These must approach zero as the critical point is approached since the properties of the liquid and the gas become identical at the critical point. As we have pointed out above, the φ's corresponding to the liquid and gas sides of the coexistence boundary are identical. The $\Delta\psi$'s thus represent a displacement of the ψ's corresponding to no change in the φ's. It is not unreasonable to suppose, therefore, that as the critical point is approached, the $\Delta\psi([m])$'s become proportional to the eigenvector $\delta([m])$ of T and that

$$\lim_{T \to T_c} \frac{\Delta\psi([m])}{\Delta\psi([1])} = \frac{\delta\psi([m])}{\delta\psi(1)}\bigg|_{T=T_c} = \delta([m]). \quad (83)$$

The critical point is thus characterized by the existence of a eigenvector with eigenvalue zero of $T([m], [n])$, which we might call the critical eigenvector. The critical eigenvector $\delta([m])$ plays the following three roles in the theory: (1) Displacements in the potentials of average force proportional to $\delta([m])$ produce no change in $\log \Xi$ and, according to the heuristic principle expressed in Eq. (49), fluctuations proportional to $\delta([m])$ are exceptionally probable. (2) Displacements proportional to $\delta([m])$ imply no changes in the physical potentials or the chemical potential or the temperature. (Or the pressure since $\log \tilde{\Xi} = PV/kT$.) (3) The differences in the potentials of average force between liquid and vapor phase are proportional to $\delta([m])$.

XIII. PROSPECT

The definitions and properties of the correlation matrix and direct-correlation matrix, the generalized Ornstein–Zernike equation relating them, the generalization of the Gibbs picture of phase transitions and points, and the existence of a critical eigenvector constitute the elements and principles out of which we hope to construct a theory of critical phenomena. We do not do so in this paper. In order to give some motive, in the absence of concrete results, for the rather involved formalism of the present paper we wish to show how the above elements and principles might fit together to form a theory of critical phenomena.

First of all, we do not prove the existence of a critical eigenvector in the present paper but merely give a plausible argument. It is possible, albeit with considerable difficulty, to test the hypothesis of the

existence of a critical eigenvector for the two-dimensional Ising model. In this case it is appropriate to consider the averages

$$\langle \sigma_i \sigma_j \cdots \sigma_l \rangle_{\pm},$$

where ijl represent a set of lattice sites, $\sigma_i \cdots \sigma_l$ the corresponding spin variables, and $\pm$ refer to values on the two sides of the coexistence line. The differences

$$\langle \sigma_i \cdots \sigma_l \rangle_{+} - \langle \sigma_i \cdots \sigma_l \rangle_{-}$$

will vanish for an even number of spins, and will be equal to

$$2\langle \sigma_i \cdots \sigma_l \rangle_{+}$$

for an odd number of spins. The question to be decided is: do the limits

$$\frac{\langle \sigma_i \cdots \sigma_l \rangle_{+}}{\langle \sigma \rangle}$$

exist as the critical point is approached from below.[16] Since the relevant averages can be evaluated in principle from the Onsager solution, the question should be answerable at least for averages of a few spins. To verify the existence of a fundamental eigenvector and to determine its properties, especially asymptotic properties when groups of molecules are widely separated, would make an important contribution to the present program.

Secondly, just as in the Ornstein–Zernike theory, significant fluctuations are determined by the small eigenvalues of the quadratic form, Eq. (5), i.e., by small values of $1 - \rho t(k)$, it is reasonable to expect that in the generalized theory significant fluctuations will be determined by small eigenvalues of the quadratic form, Eq. (43), i.e., by small eigenvalues of the direct-correlation matrix $T([m], [n])$. It appears, although we will not expand further on this point here, that the general properties of $T([m], [n])$ proved above, especially those expressed in Eqs. (65) and (66), give a rich structure to these low-lying eigenvalues and eigenfunctions and therefore, to the significant critical fluctuations in the potentials of average force. Once the nature of these significant critical fluctuations are understood, the critical behavior of such thermodynamic properties as specific heats and compressibilities can be determined through equations such as Eq. (41).

Finally, the generalized Ornstein–Zernike relation, Eq. (63), when the expression (64) is inserted for T, is a nonlinear relationship for the elements of the correlation matrix $G([m], [n])$. By Eqs. (67) and (68), this is a generalization of the Percus–Yevick equation for the pair-correlation function. As the critical point is approached, $G([m], [n])$ will become long-ranged when the groups of molecules $[m]$ and $[n]$ are separated. By Eqs. (31) and (22), the distance for the product property to become valid also becomes long. Thus, by Eq. (64), the distance for the product property of T Eq. (66) to become valid is long. Presumably not every asymptotic behavior of $G([m], [n])$ is compatible with the generalized Ornstein–Zernike equation. We may, therefore, hope to determine the asymptotic behavior of $G([m], [n])$, and therefore, of all relevant quantities by an appropriate analysis of the generalized Ornstein–Zernike equation when T has a small eigenvalue. The structure of eigenvalues and eigenvectors of T will clearly be an important part of the analysis.

[16] R. B. Griffiths has shown (private communication) on the basis of inequalities proved by himself, Kelly and Sherman that for a ferromagnetic Ising system the ratios $\langle \sigma_1 \cdots \sigma_l \rangle / \langle \sigma_1 \rangle$ are bounded above and below; R. B. Griffiths, J. Math. Phys. **8**, 478 and 484 (1967). O. G. Kelly and S. Sherman (unpublished).

APPENDIX A

It is the purpose of this appendix to derive Eqs. (22), (56), and (57) of the main text. Consider then a phase function

$$A([m]) = \sum_{[\mu] \subseteq [m]} a([\mu]), \tag{A1}$$

$$B([n]) = \sum_{[\nu] \subseteq [n]} b([\nu]), \tag{A2}$$

where the $a([\mu])$ have the cluster property. The product $A([m])B([n])$ may be written

$$\begin{aligned} A([m])B([n]) &= \sum_{[\nu] \subseteq [n]} \sum_{[\mu] \subseteq [m]} a([\mu])b([\nu]) \\ &= \sum_{[n] \subseteq [m] \cup [m]} \sum_{[\mu] \cup [\nu] = [n']} a([\mu])b([\nu]), \end{aligned} \tag{A3}$$

where the second equation is obtained from the first by summing over all subsets $[\nu]$, $[\mu]$ whose union equals a fixed set $[n']$ and then summing over $[n']$. Using Eq. (19), we have

$$\langle AB \rangle = \sum \frac{1}{n!} \int f([n]) \left[\sum_{[\mu] \cup [\nu] = [n]} a([\mu]b([\nu]) \right] d[n]. \tag{A4}$$

Since we integrate over all particles of the set $[n]$, there will be many terms which will give the same contribution to the integral. If we write

$$\begin{aligned} [\lambda] &= [\mu] \cap [\nu], \\ [\bar{\mu}] &= [\mu] \ominus [\lambda], \\ [\bar{\nu}] &= [\nu] \ominus [\lambda], \end{aligned} \tag{A5}$$

there will be $(n!/\bar{\mu}!\,\bar{\nu}!\,\lambda!)$ terms in the sum in Eq. (A4) corresponding to each choice of the numbers $\bar{\mu}$, $\bar{\nu}$, λ.

We have, therefore,

$$\langle AB\rangle = \sum_{\bar{\mu},\bar{\nu},\lambda=0}^{\infty} \frac{1}{\bar{\mu}!}\frac{1}{\bar{\nu}!}\frac{1}{\lambda!}\int f([\bar{\mu}]\oplus[\bar{\nu}]\oplus[\lambda]) \times a([\bar{\mu}]\oplus[\lambda])b([\nu]\oplus[\lambda])\,d[\bar{\mu}]\,d[\bar{\nu}]\,d[\lambda]. \quad \text{(A6)}$$

If we introduce the delta function defined by Eq. (23), we may rewrite this as

$$\langle AB\rangle = \sum_{\bar{\mu},\bar{\nu},\lambda,\kappa=0}^{\infty} \frac{1}{\bar{\mu}!}\frac{1}{\bar{\nu}!}\frac{1}{\lambda!}\frac{1}{\kappa!}\int f([\bar{\mu}]\oplus[\bar{\nu}]\oplus[\lambda]) \times \delta([\lambda],[\kappa])a([\bar{\mu}]\oplus[\lambda])b([\bar{\nu}]\oplus[\kappa]) \times d[\bar{\mu}]\,d[\bar{\nu}]\,d[\lambda]\,d[\kappa]. \quad \text{(A7)}$$

This in its turn may be rewritten as

$$\langle AB\rangle \sum_{\mu,\nu=1}^{\infty} \frac{1}{\mu!}\frac{1}{\nu!}\int f([\mu]\cup[\nu]) \times \sum_{\substack{[\lambda]\subseteq[\mu]\\ [\kappa]\subseteq[\nu]}} \delta([\lambda],[\kappa])a([\mu])b([\nu])\,d[\mu]\,d[\nu], \quad \text{(A8)}$$

since the

$$\frac{\mu!}{\lambda!\,\bar{\mu}!} = \frac{\nu!}{\kappa!\,\bar{\nu}!}$$

terms in the sum in Eq. (A8) are equal in value and correspond to one term in Eq. (A7), and since $[\bar{\mu}]\oplus[\bar{\nu}]\oplus[\lambda]$ is the same as $[\mu]\cup[\nu]$ when the set $[\lambda]$ coincides with the set $[\kappa]$ as required by the delta function. We note finally that

$$\begin{aligned}\langle (A-\langle A\rangle)(B-\langle B\rangle)\rangle &= \langle AB\rangle - \langle A\rangle\langle B\rangle \\ &= \sum_{\mu,\nu=1}^{\infty}\frac{1}{\mu!}\frac{1}{\nu!}\int\Big[f([\mu]\cup[\nu])\sum_{\substack{[\lambda]\subseteq[\mu]\\ [\kappa]\subseteq[\nu]}}\delta([\lambda],[\kappa]) \\ &\quad - f([\mu],[\nu])\Big]a([\mu])a([\nu])\,d[\mu]\,d[\nu]. \quad \text{(A9)}\end{aligned}$$

Equation (9) is equivalent to Eqs. (21) and (22).

To derive Eqs. (56) and (57), we note that $\tilde{\Xi}$ is the grand-partition function and, therefore, may be written as

$$\tilde{\Xi} = \sum_{n=0}^{\infty}\frac{1}{n!}\int \exp\sum_{[\nu]\subseteq[n]}\varphi([\nu])\,d[n]. \quad \text{(A10)}$$

If we vary the φ's, the linear terms in the variation of $\log\tilde{\Xi}$ are

$$\delta\log\tilde{\Xi} = \frac{\delta\tilde{\Xi}}{\tilde{\Xi}} = \sum_{n=0}^{\infty}\frac{1}{n!}\int\sum_{[\nu]\subseteq[n]}\delta\varphi([\nu]) \times \exp\sum_{[\bar{\nu}]\subseteq[n]}\varphi([\bar{\nu}])\,d[n]. \quad \text{(A11)}$$

If we write $[n]\equiv[\nu]\oplus[\nu']$ and note that for every choice of ν, $0<\nu<n$, there are $n!/\nu!\,\nu'!$ equal terms, we have

$$\delta\log\tilde{\Xi} = \sum_{\nu=1}^{\infty}\frac{1}{\nu!}\int\delta\varphi([\nu])f([\nu])\,d[\nu], \quad \text{(A12)}$$

where

$$f([\nu]) = \frac{1}{\tilde{\Xi}}\sum_{\nu=0}^{\infty}\int\frac{1}{\nu'!}\exp\sum_{[k]\subseteq[\nu]\oplus[\nu']}\varphi([k])\,d[\nu'] \quad \text{(A13)}$$

is the grand-canonical expression for the molecular-distribution functions. Eq. (56) then follows immediately from Eq. (A12).

The second-order terms in the variation of $\log\tilde{\Xi}$ are

$$\begin{aligned}\delta^2\log\tilde{\Xi} = \frac{\delta^2\tilde{\Xi}}{\tilde{\Xi}} - \left(\frac{\delta\tilde{\Xi}}{\tilde{\Xi}}\right)^2 &= \frac{1}{\tilde{\Xi}}\sum_{n=1}^{\infty}\frac{1}{n!} \\ &\times\int\Big(\sum_{[\nu']\subseteq[n]}\delta\varphi([\nu'])\Big)^2\exp\sum_{[\nu]\subseteq[n]}\varphi([\nu])\,d[n] \\ &-\frac{1}{\tilde{\Xi}}\sum_{n=1}^{\infty}\frac{1}{n!}\int\sum_{[\nu']\subseteq[n]}\delta\varphi([\nu'])\exp\sum_{[\nu]\subseteq[n]}\varphi([\nu])\,d[(n)]. \quad \text{(A14)}\end{aligned}$$

If we define a sum function

$$\delta U[n] = \sum_{[\nu]\subseteq[n]}\delta\varphi(\nu), \quad \text{(A15)}$$

we may write

$$\delta^2\log\tilde{\Xi} = \langle\delta U([\nu])^2\rangle - \langle\delta U[n]\rangle^2, \quad \text{(A16)}$$

where the brackets mean average in a grand-canonical ensemble. By analogous transformations to those leading to Eq. (A9), we may write

$$\delta^2\log\tilde{\Xi} = \sum_{n=1,m=1}^{\infty}\frac{1}{n!}\frac{1}{m!}\int\delta\varphi([m])\Gamma([m],[n]) \times \delta\varphi[n]\,d[m]\,d[n], \quad \text{(A17)}$$

where $\Gamma([m],[n])$ is defined by Eq. (22) and everywhere grand canonical distribution functions are meant. Equation (A17) implies Eq. (57).

We close this appendix by giving the following representation for $\delta([\alpha][\alpha'])$:

$$\begin{aligned}\delta([\alpha'][\alpha]) &= \sum\delta(i_1,1)\cdots\delta(i_\alpha,\alpha), \quad \text{if } \alpha=\alpha', \\ &= 0 \text{ otherwise} \quad \text{(A18)}\end{aligned}$$

where $\delta(i,j)$ is the single-particle delta function and $i_1,\cdots,i_\alpha$ is a permutation of the symbols $1'\cdots\alpha'$, and the sum is over all permutations. This expression may be used to demonstrate a formula needed to derive Eq. (B9):

$$\sum_{\beta=0}^{\infty}\frac{1}{\beta!}\int h([\alpha]\oplus[\beta])\delta([\alpha]\oplus[\beta],[\gamma])\,d[\beta] = \sum_{[\gamma']\oplus[\gamma'']=[\gamma]}h([\alpha]\oplus[\gamma'])\delta([\alpha],[\gamma'']), \quad \text{(A19)}$$

where $h([\nu])$ is an arbitrary sequence of functions.

APPENDIX B

The purpose of this appendix is to derive the formula for $T([m], [n])$, given in Eq. (64), from the expression for $T([m], [n])$, given in Eq. (54). In doing so we make use of a reciprocal relationship between the molecular distribution functions $f([n])$ and the functions $\eta([n])$ defined by

$$\eta([n]) = \exp \sum_{[\nu]\subseteq[n]} \varphi([\nu]) \tag{B1}$$

Referring to Eqs. (27), we see that the η's are related to the φ's in the same way that the f's are related to the ψ's. Equation 31 may be inverted to give

$$\varphi([n]) = \sum_{[\nu]\subseteq[n]} (-1)^{n-\nu} \log \eta([\nu]). \tag{B2}$$

The reciprocity between the η's and the f's is expressed in the following two formulas:

$$\tilde{\Xi}^{-1} = \sum_{n=0}^{\infty} \frac{1}{n!} (-1)^n \int f([n])\, d[n], \tag{B3}$$

which is to be compared to Eq. (A10), and

$$\Xi^{-1}\eta([\nu]) = \sum_{\nu'=0}^{\infty} \frac{1}{\nu'!} (-1)^{\nu'} \int f([\nu] \oplus [\nu'])\, d[\nu'], \tag{B4}$$

which is to be compared to Eq. (A12). The derivation and background of these relationships are given, for instance, in Ref. 17.

If in Eqs. (B4) and (B3) we express the f's in terms of the ψ's through Eqs. (27), we may compute the functional derivative of the $\eta([m])$ with respect to $\psi([n])$. In analogy to Eqs. (22) and (A16), we obtain

$$(-1)^m \frac{\delta\eta([m])}{\delta\psi([n])} = \sum_{\substack{0\subseteq[\kappa]\subseteq[m] \\ 0\subseteq[\kappa']\subseteq[n]}} (-1)^{[m]\cup[n]} \eta([m] \cup [n]) \times \delta([\kappa], [\kappa']) - (-1)^{m+m}\eta([m])\eta([n]). \tag{B5}$$

Taking the functional derivative of Eq. (B2) with respect to $\psi([m])$ and substituting Eq. (B2) into Eq. (54), we have

$$T([\mu], [m]) = \sum_{\kappa=0}^{\infty} \frac{1}{\kappa!} \int d[\kappa] f([\mu] \oplus [\kappa]) \times \sum_{[\lambda]\subseteq[\mu]\oplus[\kappa]} (-1)^{\mu+\kappa-\lambda} \frac{1}{\eta([\lambda])} \frac{\delta\eta([\lambda])}{\delta\psi([m])}. \tag{B6}$$

The summation is over all sets $[\lambda]$ included in $[\mu] \oplus [\kappa]$. The set $[\lambda]$ will thus be the sum of two sets $[\nu] \subseteq [\mu]$ and $[\bar{\kappa}] \subseteq [\kappa]$. If we call $[\bar{\kappa}']$ the complimentary set to $[\bar{\kappa}]$ in $[\kappa]$, the variables of $[\bar{\kappa}']$ are contained only among the arguments of f and not among the arguments of $\delta\eta/\delta\psi$. Thus, we may carry out the integration and summation over $[\bar{\kappa}']$ keeping $[\kappa]$ and all other variables fixed. Since the variables of $[\kappa]$ are dummy the value of the integral will depend only on the number of variables in $[\bar{\kappa}]$ and $[\bar{\kappa}']$. There will be $\kappa!/(\bar{\kappa}!\,\bar{\kappa}'!)$ such terms which have the same value. Noting Eq. (B4), we obtain

$$T([\mu], [m]) = \tilde{\Xi}^{-1} \sum_{0}^{\infty} \frac{1}{\bar{\kappa}!} \int d[\bar{\kappa}] \times \sum_{0\subseteq[\nu]\subseteq[\mu]} (-1)^{\mu-\nu} \frac{\eta([\mu] \oplus [\bar{\kappa}])}{\eta([\nu] \oplus [\bar{\kappa}])} \frac{\delta\eta([\nu] \oplus [\bar{\kappa}])}{\delta\psi([m])} \tag{B7}$$

Like Eq. (54), the representation of $T([\mu], [m])$ given by Eq. (B7) is not manifestly symmetric. We will obtain a manifestly symmetric representation of T by eliminating the $\delta\eta/\delta\psi$ through Eq. (B5). We obtain

$$T([\mu], [m]) = \tilde{\Xi}^{-1} \sum_{0}^{\infty} \frac{1}{\kappa!} \int d[\kappa] \sum_{0\subseteq[\nu]\subseteq[\mu]} (-1)^{\mu-\nu} \times \frac{\eta([\mu] \oplus [\kappa])}{\eta([\nu] \oplus [\kappa])} \Bigg\{ \sum_{\substack{0\subseteq[\alpha]\subseteq[\nu]\oplus[\kappa] \\ 0\subseteq[\alpha']\subseteq[m]}} (-1)^{[\kappa]\cup[m]-[\kappa]} \times \delta([\alpha], [\alpha'])\eta(([\nu] \oplus [\kappa]) \cup [m]) - (-1)^n \eta([\nu] \oplus [\kappa])\eta([m]) \Bigg\}. \tag{B8}$$

We note that, owing to its independence of the set $[\nu]$, the sum over the second term in the brace is zero. We consider then a particular term in the first summation; i.e., a particular choice of the sets $[\nu]$, $[\alpha]$, $[\alpha']$, $[\kappa]$ which satisfies the conditions of the summation. The set $[\alpha]$ will have a part $[\beta] \subseteq [\nu]$ and a part $[\gamma] \subseteq [\kappa]$. Since they are involved in the delta function, the integration over the variables $[\gamma]$ can be carried out immediately, leaving the integration over $[\bar{\kappa}]$, the complimentary set of $[\gamma]$ in $[\kappa]$. This integration will identify $[\beta]$ with a subset $[\alpha']$ of $[m]$; i.e., we must replace the set $[\kappa]$ by $[\bar{\kappa}] \oplus [\gamma']$ where $[\gamma']$ is a subset of $[m]$. Since the variables of $[\kappa]$ are dummy, there are $\kappa!/(\gamma!\,\bar{\kappa}!)$ terms in the sum which will have the same value. Using Eq. (A18), we obtain

$$T([\mu], [m]) = \tilde{\Xi}^{-1} \sum_{0}^{\infty} \frac{1}{\kappa!} \int d[\kappa] \sum_{\substack{0\subseteq[\beta]\subseteq[\nu]\subseteq[\mu] \\ 0\subseteq[\beta']\oplus[\gamma']\subseteq[m]}} \times (-1)^{\mu-\nu+[([\nu]\oplus[\nu']\oplus[\kappa])\cup[m]]-([\nu]\oplus[\gamma']\oplus[\kappa])} \times \delta([\beta], [\beta']) \times \frac{\eta([\mu] \oplus [\gamma'] \oplus [\kappa])\eta(([\gamma] \oplus [\gamma'] \oplus [\kappa]) \cup [m])}{\eta([\nu] \oplus [\alpha'] \oplus [\kappa])}. \tag{B9}$$

We may express Eq. (B9) in a symmetric form by noting that the delta function implies that a particular

[17] M. S. Green, *Lectures in Theoretical Physics, Vol. III*, W. E. Britten, B. W. Downs, and J. Downs, Eds. (Interscience Publishers Inc., New York, 1961), p. 195.

term contributes only when the set of variables $[\beta]$ is coincident with the set of variables $[\beta']$. Since $[\beta]$ is a part of $[\mu]$ and $[\beta']$ is a part of $[m]$, this means that the sets of variables $[\mu]$ and $[m]$ must be partially coincident. We may assume that the set $[\beta] \equiv [\beta']$ is also identical to the intersection $[\mu] \cap [m]$, since the coincidence of any of the variables of $[\mu]$ and $[m]$ not involved in the delta function will not give rise to a contribution. If we write $[\gamma]$ for the complement of $[\beta]$ in $[\nu]$, we may express the set $([\nu] \oplus [\beta'] \oplus [\kappa]) \cup [m]$ as

$$([\beta] \oplus [\gamma] \oplus [\gamma'] \oplus [\kappa]) \cup [m] \equiv [\gamma] \oplus [m] \oplus [\kappa],$$

since $[\beta]$ and $[\gamma']$ are included in $[m]$, and $[\gamma]$ and $[\kappa]$ are disjoint from $[m]$. Thus we have

$$T([\mu], [m]) = \tilde{\Xi}^{-1} \sum_{\substack{[\beta]\subseteq[m] \\ [\beta']\subseteq[\mu]}} \sum_{0}^{\infty} \frac{1}{\kappa!} \times \int d[\kappa] \sum_{\substack{0\subseteq[\gamma']\subseteq[m]\smallsetminus[m]\cap[\mu] \\ 0\subseteq[\gamma]\subseteq[\mu]\smallsetminus[m]\cap[\mu]}} (-1)^{\mu+m-\gamma-\gamma'} \delta([\beta], [\beta']) \times \frac{\eta([\mu] \oplus [\gamma'] \oplus [\kappa])\eta([m] \oplus [\gamma] \oplus [\kappa])}{\eta([m] \cap [\mu] \oplus [\gamma'] \oplus [\gamma] \oplus [\kappa])}. \quad \text{(B10)}$$

Equation (B10) is the manifestly symmetric form of T which we have sought.

We will make one further transformation, which in contrast to Eq. (B10) will express each matrix element of T as a finite sum of terms. From this form it will be possible to make positive statements about the behavior of $T([m], [\mu])$ when subsets of the particles $[m]$ or $[\mu]$ are separated from the rest. This will be done in Appendix C. At a certain stage in the transformation, the calculations will be much simplified, if the $\phi([\nu])$ are assumed to vanish if the set $[\nu]$ contains more than two members. This is an inessential restriction which merely simplifies the calculations.

We note first that in Eq. (B10) we have to do repeatedly with expressions of the form

$$\frac{\eta([\lambda])\eta([\zeta])}{\eta([\lambda] \cap [\xi])}, \quad \text{(B11)}$$

where $[\lambda]$ and $[\xi]$ are sets which have common elements. In Eq. (B10), $[\lambda]$ corresponds to $[\mu] \oplus [\gamma'] \oplus [\kappa]$ and $[\xi]$ to $[m] \oplus [\gamma] \oplus [\kappa]$, while $[\lambda] \cap [\xi]$ corresponds to $[m] \cap [\mu] \oplus [\gamma] \oplus [\gamma'] \oplus [\kappa]$. Let us write

$$\begin{aligned} \eta([\lambda]) &= \sum_{G_{[\lambda]}} \Pi\Phi([\nu]), \\ \eta([\xi]) &= \sum_{G_{[\xi]}} \Pi\Phi([\nu]), \\ \eta([\lambda] \cap [\xi]) &= \sum_{G_{[\lambda]\cap[\xi]}} \Pi\Phi([\nu]), \end{aligned} \quad \text{(B12)}$$

where $1 + \Phi([\nu]) = \exp \varphi([\nu])$ and in each case the product is over a collection $G_{[\lambda]}$ of subsets $[\nu]$ of $[\lambda]$ and the sum is over all distinct collections of subsets $G_{[\lambda]}$ such that no subset is repeated. The product $\eta([\lambda])\eta([\xi])$ will be a sum of products over collections of subsets of the union $[\lambda] \cup [\xi]$ in which each subset belongs entirely to $[\lambda]$ or entirely to $[\mu]$, but some subsets of a collection may be repeated. Those subsets which are repeated belong entirely to $[\lambda] \cap [\xi]$. If we now consider the sum over products of collections of subsets of $[\lambda] \cup [\xi]$, *none of which are repeated*, and each of which belong entirely either to $[\lambda]$ or to $[\xi]$, we obtain a product belonging to $\eta[\lambda]\eta[\xi]$ after multiplying a term in the sum over product by a product over a collection of unrepeated subsets of $[\lambda] \cap [\xi]$. Thus we have

$$\eta([\lambda])\eta([\xi]) = \eta([\lambda] \cap [\xi]) \sum_{G} \Pi\Phi([\nu]), \quad \text{(B13)}$$

where G is a collection of subsets of $[\lambda] \cup [\xi]$ such that each subset is contained wholly either in $[\lambda]$ or in $[\xi]$ and no set is repeated. Therefore,

$$\frac{\eta([\lambda])\eta([\xi])}{\eta([\lambda] \cap [\xi])} = \sum_{G} \Pi\Phi([\nu]), \quad \text{(B14)}$$

where G is as specified above.

Since G is a collection of subsets of $[\lambda] \cup [\xi]$ it is possible to write another formula for

$$n([\lambda])\eta([\xi]/\eta([\lambda] \cap [\xi]).$$

By similar argument to the one by which Eq. (B14) was derived, we may write

$$\frac{\eta([\lambda])\eta([\xi])}{\eta([\lambda] \cap [\xi]} = \frac{\eta([\lambda] \cup [\xi])}{1 + \sum_{G'} \Pi\Phi([\nu])}, \quad \text{(B15)}$$

where G' is the collection of subsets of $[\lambda] \cup [\xi]$ such that no subsets are contained in entirety in either $[\lambda]$ or in $[\xi]$. Every such subset will contain members of $[\lambda]$ not in $[\xi]$ and members of $[\xi]$ not in $[\lambda]$. Now in the special case of pairwise forces, i.e., when $\varphi([\nu]) = 0$, if ν is greater than two, the collection of sets G' is simply the collection of all pairs of points: one belonging to $[\bar{\lambda}] \equiv ([\lambda] \ominus [\lambda] \cap [\xi]$ and one belonging to $[\bar{\xi}] \equiv [\xi] \ominus [\lambda] \cap [\xi]$. Still, under the assumption of pairwise forces, the denominator in Eq. (B15) can be written as

$$1 + \sum_{G'} \Pi\Phi([\nu]) = \frac{\eta([\bar{\lambda}] \cup [\bar{\xi}])}{\eta([\bar{\lambda}])\eta([\bar{\xi}])}; \quad \text{(B16)}$$

i.e., in the case of pairwise forces we have

$$\frac{\eta([\lambda])\eta[\xi])}{\eta([\lambda] \cap [\xi])} = \frac{\eta([\bar{\lambda}]\eta([\bar{\xi}])\eta([\lambda] \cup [\xi])}{\eta([\bar{\lambda}] \cup [\bar{\xi}])}. \quad \text{(B17)}$$

Returning now to Eq. (B10) we define

$$[\bar{m}] = [m] \ominus [m] \cap [\mu],$$
$$[\bar{\mu}] = [\mu] \ominus [m] \cap [\mu]. \tag{B18}$$

By Eq. (B17), we may write the products of the η's on the left in Eq. (B10) as

$$\frac{\eta([\bar{\mu}] \ominus [\gamma])\eta([\bar{m}] \ominus [\gamma'])}{\eta([\bar{\mu}] \oplus [\gamma'] \oplus [\bar{m}] \oplus [\gamma'])} \times \eta([m] \cup [\mu] \oplus [\kappa]). \tag{B19}$$

Substituting this expression into Eq. (B10) and noting that the appearance of $[\kappa]$ in only one of the η's in Eq. (B19) permits the summation over κ to be carried out and that summation over $[\gamma]$ and $[\gamma']$ can be replaced by summation over

$$[\bar{\mu}] - [\gamma] = [\alpha], \quad [\bar{m}] - [\gamma'] = [\alpha'],$$

we obtain

$$T([\mu], [m]) = \sum_{\substack{0 \subseteq [\beta] \subseteq [m] \\ 0 \subseteq [\beta'] \subseteq [\mu]}} \delta([\beta], [\beta']) \times \Theta([\bar{m}], [\bar{\mu}] f([\mu] \cup [m]), \tag{B20}$$

where

$$\Theta([\mu], [m]) = \sum_{\substack{0 \subseteq [\alpha] \subseteq [m] \\ 0 \subseteq [\alpha'] \subseteq [\mu]}} (-1)^{\alpha+\alpha'} \frac{\eta([\alpha])\eta([\alpha'])}{\eta([\alpha] \oplus [\alpha'])}. \tag{B21}$$

Since we are assuming pairwise forces

$$\frac{\eta([\alpha])\eta([\alpha'])}{\eta([\alpha] \oplus [\alpha'])} = \exp - \sum_{\substack{i \in [\alpha] \\ j \in [\alpha']}} \varphi(ij).$$

[Pairs i, j, which are contained wholly in $[\alpha]$ or wholly in $[\alpha']$, are removed by the factors $\eta([\alpha])$, $\eta([\alpha'])$.] We have

$$\Theta([m], [\mu]) = \sum_{\substack{0 \subseteq [\alpha] \subseteq [m] \\ 0 \subseteq [\alpha'] \subseteq [\mu]}} (-1)^{\alpha+\alpha'} \exp - \sum_{\substack{i \in [\alpha] \\ j \in [\alpha']}} \varphi(ij).$$

Equations (B20) and (B21) together yield Eq. (64).

APPENDIX C

It is the purpose of this appendix to demonstrate the two asymptotic properties of $T([m], [n])$ expressed in Eqs. (65) and (66). We do this by first demonstrating analogous properties of $\Theta([m], [n])$. Suppose, referring to the definition of $\Theta([m], [n])$, Eq. (B21), that there is a point in $[m]$ which is distant from all points of $[n]$, i.e., for which all factors $\varphi(i, j)$, $j \in [n]$, i fixed, are zero. The sets $[\alpha]$ can then be divided into two classes, those which contain i and those which do not. These two classes of sets can be put into one-to-one correspondence by pairing a set $[\alpha]$ not containing i to the set $[\alpha] \oplus i$. The empty set corresponds to set containing i alone. For a given choice of $[\alpha']$ each of the paired sets will have the same value of the exponential but opposite signs of the factor $(-1)^{\alpha+\alpha'}$. Thus $\Theta([m], [n])$ will be zero in this case.

$$\Theta([m], [n]) = 0, \tag{C1}$$

if any point of $[m]$ is distant from all points of $[m]$; i.e., Θ will be different from zero only if every point of $[m]$ is close to some point of $[n]$ and vice versa.

Let us suppose, on the other hand, that all points of $[m]$ are close to some point of $[n]$ and vice versa but $[m] \equiv [\mu] \oplus [\nu]$, and $[n] \equiv [\mu'] \oplus [\nu']$ with $[\mu]$ far from $[\mu']$, and $[\nu]$ far from $[\nu']$. Then, in Eq. (B21) every set $[\alpha]$, $[\alpha']$ will be broken into two parts, $[\alpha] \equiv [\beta] \oplus [\gamma]$, $[\alpha'] \equiv [\beta'] \oplus [\gamma']$, one of which may be empty, such that

$$[\beta] \subseteq [\mu], \quad [\beta'] \subseteq [\mu'],$$
$$[\gamma] \subseteq [\nu], \quad [\gamma'] \subseteq [\nu']. \tag{C2}$$

The factors $\exp - \varphi(ij)$ where i lies in $[\beta]$ and j lies in $[\gamma']$ or i lies in $[\gamma]$ and j lies in $[\beta']$ will be unity. The remaining factors can be grouped into two products of the same form. We have

$$\Theta([m], [n]) = \Theta([\mu], [\mu'])\Theta([\nu], [\nu']). \tag{C3}$$

It is clear that the factor $\delta([m], [n])$ has properties analogous to Eq. (C1) and Eq. (C3); i.e.,

$$\delta([m], [n]) = 0, \tag{C4}$$

unless each point of $[m]$ coincides with some point of $[n]$ and

$$\delta([\mu] \oplus [\nu], [\mu'] \oplus [\nu']) = \delta([\mu], [\mu'])\delta([\nu], [\nu']), \tag{C5}$$

if $[\mu]$ is far from $[\nu]$ and $[\mu']$ is far from $[\nu']$. Moreover, from the product property of the f's we have

$$f([m] \cup [n]) = f([\mu] \cup [\nu])f([\mu'] \cup [\nu']). \tag{C6}$$

For a configuration in which $[m]$ and $[n]$ overlap Eq. (66) follows from Eqs. (B20) and (C3). In case one or both of the intersections $[\mu] \cap [\nu]$ and $[\mu'] \cap [\nu']$ are not empty, the only term in Eq. (20) which need be considered is the one for which

$$[\beta] = [\beta'] = [m] \cap [n].$$

Then Eq. (66) follows from Eq. (B20) together with Eq. (C3) and (C5). Equation (65) follows from Eqs. (C1), (C4), and (B20).

Convergence of the Born Series

JOHN V. CORBETT*
Crocker Nuclear Laboratory, University of California, Davis, California

(Received 6 September 1967)

Three types of Born series which can be associated with a transition amplitude are discussed, and the criteria for the convergence of the different series are compared. With regard to the divergence of the Born series, the ordering, matrix-element series diverges implies vector series diverges implies operator series diverges, is obtained for the natural vector and operator Born series that can be abstracted from the expression for the transition amplitude. The conclusion that the divergence of the operator Born series does not ensure that the Born series of physical matrix-elements divergences is applied to an example of three-body rearrangement scattering.

CONVERGENCE OF THE BORN SERIES

Any transition amplitude can be expanded in an infinite series in three ways: (1) as an operator series in which each term of the series is an operator which acts upon a space of functions, (2) as a vector series in which each term is a member of a function space, (3) as a series of matrix elements that are complex numbers constructed from the operators on, and the vectors of, the function space.

Although this decomposition can be made for two-body scattering amplitudes, it becomes more important when many-body amplitudes are expanded in power series. This importance stems from the well-known fact that three-body systems, for example, can have additional asymptotic states, which provide additional right-hand cuts and therefore make the usual perturbation theory, developed for two-body operators, ineffectual.[1]

By using this classification of Born series we are able to remove much of the confusion that has arisen in the problem of the convergence of the Born series for the three-body amplitudes of rearrangement collisions.[2] The way in which this is done is best seen through an example.

Consider the rearrangement process: $(1, 2) + 3 \rightarrow (1, 3) + 2$, for which the transition amplitude is

$$T_{fi} = (\phi_f, V_f \psi_i^{(+)}). \tag{1}$$

$\psi_i^{(+)} \equiv \psi_i(w) = (1 + G(w)V_i)\phi_i$, $w = E + i\eta$, $\eta > 0$, and E is the total three-body energy. ϕ_f and ϕ_i are the final- and initial-state vectors, and $V_f = (V_1 + V_3)$ and $V_i = (V_1 + V_2)$ are the interactions which are not involved in forming the final and initial configurations. V_α ($\alpha = 1, 2, 3$) are the two-particle interactions of particles not labeled by α. Finally, $G(w) = (w - H)^{-1}$, $G_0(w) = (w - H_0)^{-1}$ and $H = H_0 + V_1 + V_2 + V_3$. Then

$$T_{fi} = (\phi_f, V_f\phi_i) + (\phi_f, V_f G(w) V_i \phi_i). \tag{2}$$

The operator series follows when the total Green's function, or resolvent operator, $G(w)$ is extracted from the second matrix element and is expanded in powers of the operator $K(w) = G_0(w)(V_1 + V_2 + V_3)$:

$$G(w) = \sum_{n=0}^{\infty} K^n(w) G_0(w). \tag{3}$$

If the vector $G(w)V_i\phi_i$ is written as the series,

$$G(w)V_i\phi_i = \sum_{n=0}^{\infty} \chi_n(w), \tag{4}$$

where the $\chi_n(w) = K^n(w)G_0(w)V_i\phi_i \equiv K^n(w)\psi_0$ are a sequence of vectors, the vector series is defined. Furthermore the Born series may be taken to be the series of matrix elements:

$$T_{fi} = \sum_{n=0}^{\infty} a_n(w) + (\phi_f, V_f\phi_i),$$

where $a_n(w) = (\phi_f, V_f\chi_n(w))$.

It is now obvious that the three series, although they are concerned in the description of the same physical process, will have quite different mathematical properties. In the following, we show that the vector Born series may converge when the operator Born series diverges, and the Born series of matrix elements may converge when the vector series diverges.

In order to obtain these results, it is necessary to generalize the problem by considering the resolvent operator as the function of a complex parameter λ:

$$G(w, \lambda) = [w - H_0 - \lambda(V_1 + V_2 + V_3)]^{-1}.$$

Now the radii of convergence of the operator, vector, and ordinary power series of the complex variable λ can be computed and are shown to have the ordering

$$R_O(w) \leq R_V(w) \leq R_M(w),$$

where $R_O(w)$, $R_V(w)$, and $R_M(w)$ are, respectively, the

* Supported by U.S. Atomic Energy Commission.

[1] S. Weinberg, Phys. Rev. **133**, B232 (1964); C. Lovelace, Phys. Rev. **135**, B1225 (1964); and L. V. Faddeev, Sov. Phys. JETP **12**, 1014 (1961).

[2a] R. Aaron, R. D. Amado, and B. W. Lee, Phys. Rev. **121**, 319 1961).

[2b] K. Dettmann and G. Leibfried, Phys. Rev. **148**, 1271 (1966).

radius of convergence for the operator, vector, and matrix-element series at the complex energy w. Unfortunately, in this approach λ has not the immediate interpretation of the strength of any single interaction, but is instead an over-all strength parameter. In certain special cases, however, it can be given a more physical meaning.

The important mathematical point to be deduced from this work is that the divergence of the operator Born series does not imply that the Born series for the matrix elements diverge. This also has a physical content because the series of matrix elements is the object calculated and compared with the observed data. On the other hand the operator and vector series have the advantage, when they converge, of implying the convergence of the series of matrix elements for all initial and final states and all final (or initial, depending on which vector series is chosen) states, respectively.

These results also provide the reminder that, although operator equations can be easily manipulated in a formal manner, they contain more structure than is required for particular processes when they must be taken to act upon specific states.

In this work, we do not consider the problem of how fast the various series converge, nor how they can be made to converge more quickly. The discrepancy between the conclusions of the papers by Aaron, Amado, and Lee and Dettmann and Leibfried can be understood from our point of view, even though in neither of these papers are the techniques of functional analysis used. The latter paper analyzes the convergence properties of a series of matrix elements, while the former first discusses the corresponding operator series in a particular representation, and then on the basis of the divergence of this series argues that the series of physical matrix elements must also diverge, since each term is only a weighted mean of the corresponding term in the operator series. The divergence of the operator series may be due the divergence of the matrix elements of the operator series between certain vectors which are not associated with the physical states under discussion.

In the first part of this paper, the three types of series are discussed and the expressions for their radii of convergence obtained. The three radii of convergence are then compared and the result $R_O(w) \leq R_V(w) \leq R_M(w)$ deduced. Finally the vector-series approach is used on a Hilbert-space model of the problem discussed by Dettmann and Leibfried. We are able to obtain results compatible with those of Dettmann and Leibfried. The theorems for the vector series are proved in the Appendix.

OPERATOR-VALUE BORN SERIES

The results of the theory of operator-valued functions are well known[1] and can be presented in terms of the spectral properties of these operators. The operator $(\alpha I - K)^{-1}$ is the resolvent operator of K and exists as a bounded operator for α in the resolvent set of K.

If we choose K to be bounded, then its spectrum lies in a bounded region of the complex α plane. After defining the spectral radius of K to be $r(K) = \sup_{\alpha \in \sigma(K)} |\alpha| = \lim_{n \to \infty} \|K^n\|^{1/n} \leq \|K\|$, it can be shown that $(\alpha I - K)^{-1}$ is analytic in α for all α such that $|\alpha| > r(K)$. Since $(\alpha I - K)^{-1}$ is analytic in this region, it can be expanded in a Taylor series: $(\alpha I - K)^{-1} = \sum_{n=0}^{\infty} K^n/\alpha^{n-1}$, which is uniformly and absolutely convergent for $|\alpha| > r(K)$. The uniform convergence has two meanings, both of which are valid here: the convergence is uniform with respect to the parameter α in $|\alpha| > r(K)$, and it is also convergent in the uniform operator topology (convergence in the mean).

Now, if we put $\alpha = \lambda^{-1}$, $(I - \lambda K)^{-1} = \sum_{n=0}^{\infty} \lambda^n K^n$, where the series is uniformly and absolutely convergent for $|\lambda| < [r(K)]^{-1}$. In other words, the radius of convergence of this series is $[r(K)]^{-1} = \left[\lim_{n \to \infty} \|K^n\|^{1/n}\right]^{-1}$. Sometimes the symbol $\rho(0)$ is used to describe the radius of convergence of the series; this is the distance from the origin in the λ plane to the nearest point of the λ spectrum, that is, the distance to the point in the λ spectrum for which $|\lambda|$ is smallest. In terms of the spectrum of K, this latter point will be the point α in the spectrum, for which $|\alpha|$ is largest. Then $\rho(0) = [r(K)]^{-1}$.[3] Then the operator Born series is convergent at a fixed energy E, if $K(E)$ has no spectrum outside the unit circle in the α plane, for then $\rho(0) > 1$.

In this presentation, the emphasis is on the relationship between the analytic structure of $(\alpha I - K)^{-1}$ as a function of the complex variable α and the location of the spectrum of K. This is not the only approach but it is more attractive, mathematically, than the others which lead to the same conclusion for the Born series. The advantage of using the operator approach is that, once the convergence for the Born series has been obtained, then the convergence of Born series in either of the other two forms mentioned earlier follows. That is to say, that $(\phi_b, (1 - K)^{-1}\phi_a) = \sum_{n=0}^{\infty} (\phi_b, K^n\phi_a)$ is convergent for any choice of ϕ_a and ϕ_b in the underlying Hilbert space. From another point of view, this generality provides a disadvantage for, as we shall see, if the

[3] It can be shown that if α lies in spectrum of K, α^{-1} is in the λ spectrum and belongs to the point or continuous spectrum when α does. We have assumed K has no residual spectrum.

operator series diverges, we are still not sure whether the particular matrix element that is to be calculated can be expanded in a convergent series or not. Certain choices of the elements ϕ_a and ϕ_b may lead to a convergent Born series.

VECTOR-VALUED BORN SERIES

The mathematical idea behind the adoption of the vector approach is the following: The operator $[I - \lambda K(\omega)]^{-1}$ becomes undefined as λ approaches some point in that spectrum of $K(\omega)$, and thus any power series in λ for this operator must diverge at any point in the spectrum. However, the vector $\psi(\lambda) = [I - \lambda K(\omega)]^{-1}\phi$ may be well behaved at a point in the spectrum, if ϕ is chosen in such a way that it is not associated with that part of the spectrum. For example, if λ approaches the eigenvalue λ_n of $K(\omega)$ and ϕ is an eigenvector of $K(\omega)$ with eigenvalue $\lambda_m \neq \lambda_n$, then $[I - \lambda K(\omega)]^{-1}\phi$ is well defined as $\lambda \to \lambda_n$. We may thus expect that the approach through a vector series can be made to be independent of the singularities of $[I - \lambda K(\omega)]^{-1}$ and hence the restriction to operators with only isolated singularities, i.e., compact operators, can be bypassed. Furthermore, we may be able to extend the region of convergence of the power series past the location of the singularities in the λ plane. In the approach using vector-valued functions, we begin by considering the series

$$(I - \lambda K)^{-1}\phi_a = \sum_{n=0}^{\infty} \lambda^n K^n \phi_a = \sum_{n=0}^{\infty} \lambda^n \chi_n, \qquad (5)$$

where $\chi_n = K^n\phi_a$ is, for each integer n, a vector belonging to the Hilbert space; these vectors differ from the other vectors we will use in the following in that we will not normalize them to unity. The question now is whether this series of vectors converges. Usually, we would have to consider separately weak and strong convergence of the series, but here we will take the vectors to be functions of the complex variable λ and this will enable us to decide both of these questions at the same time.

We consider a sequence $\{\psi_n(\lambda)\}$, where $\psi_n(\lambda) = \sum_{p=0}^{n} \lambda^p \chi_p$, of vector-valued functions of the complex variable λ, and we wish to discuss the convergence and analyticity properties of this sequence. A vector-valued function $\xi(\lambda)$ is said to be an analytic function in some open domain U of the complex λ plane if the ordinary complex function $(\phi, \xi(\lambda))$ is analytic for all λ in U and all ϕ in $\mathcal{H}$. ($\mathcal{H}$ is the Hilbert space of the vectors considered; thus we always assume that $\|\xi(\lambda)\| = |(\xi(\lambda), \xi(\lambda))|^{\frac{1}{2}} < \infty$.) Now the analyticity properties are derived from the corresponding properties of the ordinary complex-valued functions $(\phi, \xi(\lambda))$ and from the concept of a linear functional.

The following two theorems are pertinent to our work. The proofs of these theorems is given in the Appendix.

Theorem 1: If $\psi(\lambda)$ is an analytic vector for λ in some finite open domain U, then it possesses bounded analytic derivatives $\psi^{(n)}(\lambda_0)$ for $\lambda_0 \in U$, where

$$(\phi, \psi^{(n)}(\lambda_0)) = \frac{d^n}{d\lambda^n}(\phi, \psi(\lambda))\big|_{\lambda=\lambda_0},$$

for any ϕ in the Hilbert space. Furthermore, if the circle $|\lambda - \lambda| \leq r$ is contained in U, then $\psi(\lambda)$ is given by the Taylor series

$$\psi(\lambda) = \sum_{n=0}^{\infty} \frac{(\lambda - \lambda_0)^n}{n!} \psi^{(n)}(\lambda_0),$$

which, within this circle, converges uniformly and absolutely.

Theorem 2: Any power series $\sum_{p=0}^{\infty} \psi_p(\lambda - \lambda_0)^p$ defines an analytic function $\psi(\lambda)$ in an open set $|\lambda - \lambda_0| < s$ of the complex λ plane, where $s = \left(\overline{\lim_{p\to\infty}} \|\psi_p\|^{1/p}\right)^{-1}$. The series converges absolutely and uniformly on any set $|\lambda - \lambda_0| \leq d$, where $d \leq s$. Furthermore, the series is defined uniquely by its sum

$$\chi(\lambda) = \sum_{p=0}^{\infty} \psi_p(\lambda - \lambda_0)^p,$$

since

$$\psi_p = \frac{\chi^{(p)}(\lambda)}{p!},$$

for all integer p. Absolute convergence means that the sum of the norms converges;

$$\lim_{p\to\infty} \sum_{n=0}^{p} \left\| \psi_n(\lambda - \lambda_0)^n \right\| < \infty.$$

This implies that we have strong convergence, and the proof exhibits the fact that weak analyticity is the same as strong analyticity. The convergence of the series is uniform for all λ in the circle $|\lambda - \lambda_0| < s$.

When we apply these theorems to the Born series, we must examine the series $\sum_{n=0}^{\infty} \lambda^n \psi_n$, where $\psi_n = K^n\psi_a$. Suppose this converges, in $|\lambda| < s(\phi_a)$, to the vector $\chi(\lambda)$; we still have to show that $\chi(\lambda)$ is a solution of the original Lippmann–Schwinger equation, $\Psi(\lambda) = \psi_a + \lambda K(\omega)\Psi(\lambda)$. That this is the case is shown in the Appendix under the assumption either that K is bounded or $(I - \lambda K)^{\dagger}$ has dense range in the basic Hilbert space. In all the scattering problems that we will consider, K is bounded.

Now we can state the result for the vector Born series. The vector Born series $\Psi = \sum_{n=0}^{\infty} \psi_n = \sum_{n=0}^{\infty} K^n(\omega)\psi_a$ converges absolutely if $\overline{\lim_{n\to\infty}} \|\psi_n\|^{1/n} < 1$. It is immediately obvious that this condition is less restrictive than the corresponding criterion for the Born series for an operator. We do not have to bother with compactness; the fact that K is bounded is sufficient to ensure that the question of convergence is not trivial. The comparison between the two methods will be drawn in greater detail later.

Although it may appear that we have only discussed the convergence of the Born series for the vector and still have to prove that the series of matrix elements converges when the vector series does, this last step is trivial. The vector series convergences in the strong topology and as this implies weak convergence we have immediately that

$$\lim_{p\to\infty} \left| \sum_{n=0}^{p} (\phi_b, VK^n\phi_a) - (\phi_b, V\psi) \right| = 0.$$

BORN SERIES FOR MATRIX ELEMENTS

The last point suggests a further and final specialization for the meaning of the Born series, for the approach using vector-valued functions requires more than may be demanded by a physicist who wants to decide if the Born series for a particular matrix element converges. In essence, we have asked that $\sum_{n=0}^{\infty} (\phi_f, V\psi_n)$ converges to $(\phi_f, V\Psi)$ for all vectors $\phi_f \in \mathcal{H}$, while the experimentally determined matrix element only requires convergence for one particular final state ϕ_f. That is, we want to discuss the convergence of $\sum_{n=0}^{\infty} (\phi_f, V\psi_n)$ for fixed ϕ_f. This takes the question out of the domain of functional analysis because we have a simple series of ordinary complex-valued functions to examine. This problem is easily solved theoretically, especially as we are only dealing with one value of the energy.

Consider the series $\sum_{n=0}^{\infty} \lambda^n a_n$, where $a_n = (\phi_f, V\psi_n)$; in order to more easily compare the result of this section with the others we will only consider the coefficients $a_n = (\xi, K^n\phi_a)$, where $\xi = V^\dagger\phi_b/\|V^\dagger\phi_b\|$. This redefinition will not alter the convergence properties of the series but will alter the sum of the series if it does converge, thus the final result must be multiplied by $\|V^\dagger\phi_b\|$ to get a numerically correct result.

The series $\sum_{n=0}^{\infty} \lambda^n a_n$ converges absolutely and uniformly for all λ inside the circle $|\lambda| < \left(\overline{\lim_{n\to\infty}} |a_n|^{1/n} \right)^{-1}$. The adjective absolute just means that the sum of the moduli of the individual terms converges. Hence the Born series for the matrix element converges if $\overline{\lim_{n\to\infty}} |a_n|^{1/n} < 1$.

COMPARISON OF RESULTS

In this section we show that the radius of convergence of the power series gets no smaller as the method is specialized from a series of operators to a series of vectors to a series of matrix elements. We will also demonstrate an application of the vector series to the situation in which the operator K is noncompact.

The comparison between the radii of convergence for the series of operators and the vector series is considered first. The radius of convergence for the operators series is the inverse of

$$r(K) = \overline{\lim_{n\to\infty}} \sup_{\psi\in\mathcal{H}} \|K^n\psi\|^{1/n}, \quad \|\psi\| = 1, \quad \text{if} \quad \psi \in \mathcal{H}, \tag{6}$$

while

$$s(K, \phi_a) = \overline{\lim_{n\to\infty}} \|K^n\phi_a\|^{1/n}, \quad \text{where} \quad \|\phi_a\| = 1, \quad \phi_a \in \mathcal{H}, \tag{7}$$

is the inverse of the radius of convergence for the vector series. From the definitions of these terms,

$$\|K^n\| = \sup_{\psi\in\mathcal{H}} \|K^n\psi\| \geq \|K^n\phi_a\|,$$

thus

$$\|K^n\|^{1/n} \geq \|K^n\phi_a\|^{1/n}$$

and

$$r(K) \geq s(K, \phi_a).$$

The radius of convergence of the vector series is greater than or equal to that of the operator series, or in terms of the Born series: since $s(K, \phi_a) \leq r(K)$, and the criterion for convergence of the Born series is either $r(K) < 1$ or $s(K, \phi_a) < 1$ in the two cases, the vector Born series may converge when the operator Born series does not. Similarly we can show the Born series for the matrix elements may converge even though its counterpart for the vector series does not. The inverse of the radius of convergence for the former series is

$$t(K; \phi_b, \phi_a) = \overline{\lim_{n\to\infty}} |(\xi, K^n\phi_a)|^{1/n}, \quad \xi = \frac{V^\dagger\phi_b}{\|V^\dagger\phi_b\|},$$

$$|(\xi, K^n\phi_a)| \leq \|\xi\| \, \|K^n\phi_a\|$$

by the Schwartz inequality.[4] Therefore

$$|(\xi, K^n\phi_a)|^{1/n} \leq \|K^n\phi_a\|^{1/n}.$$

These results can be written in order of implication of convergence: since $t(K; \phi_b, \phi_a) \leq s(K, \phi_a) \leq r(K)$, we

[4] We have assumed that the matrix element is composed of objects which are well defined in terms of Hilbert spaces and linear operations on them.

have that the convergence of the operator series implies convergence of vector series implies convergence of matrix-element series. In symbols: operator $\rightarrow$ vector $\rightarrow$ matrix-element series converges. On the other hand, we have for divergence of the various series: Matrix-element $\rightarrow$ vector $\rightarrow$ operator series diverges. The important point is that the divergence of the operator series does not ensure the divergence of the vector or matrix-element series, the results for the latter two cases depend upon the choice of elements ϕ_a and ϕ_b in the Hilbert space.

The methods used in the vector and matrix-element series in no way depend upon the compactness of the operator K. If K is a bounded operator then we are certain that the radius of convergence in these two cases is finite, and we have a method of deciding if the Born series converges in these cases. In the operator approached discussed by Weinberg and Lovelace,[1] the operator K was interpreted as the kernel of an integral equation, and the methods used were basically ways to reduce the kernel to a compact form. In the case of the three-body scattering, the kernel for the Born series could not be reduced to a compact form and therefore the power series could not be adequately discussed. It is, therefore, interesting that the method of a vector series presents an algorithm with which we can decide the convergence or divergence of the vector Born series for noncompact operators.

The example of a rearrangement process, which was discussed by Dettmann and Leibfried,[2b] is compatible with our result that the radius of convergence for the series of matrix elements is greater than or equal to that for the series of operators, although the details of their paper do not fit into the scheme discussed here. For example, they use functions which are not members of a Hilbert space and they also let the energy vary so that they have a function of two variables.

If K is assumed to be compact, as in the two-body scattering process, we can discuss in more detail the differences between the operator and vector series. In this situation the radius of convergence of the operator series is $\rho(0) = \left(\sup_{\alpha\in\sigma(K)} |\alpha|\right)^{-1} = |\alpha_a|^{-1}$, where ϕ_a is the eigenfunction of $K(w)$ corresponding to the eigenvalue α_a. If we had chosen ϕ_a as the initial vector in the sequence $K^n\phi_a$, $n = 0, 1, 2, \cdots$, $s(K, \phi_a) = \overline{\lim_{n\to\infty}} \|K^n\phi_a\|^{n/1} = |\alpha_a|$ and hence $s(K, \phi_a) = r(K)$. But if we had chosen the eigenfunction ϕ_b, with eigenvalue $\alpha_b \neq \alpha_a$, as our initial vector then the inverse of the radius of convergence $s(K, \phi_b) = |\alpha_b| < |\alpha_a| = r(K)$ and the radius of convergence of the series of vectors is greater than that for the operator series. The vector Born series may converge, for $|\alpha_b| < 1$, even though the operator Born series, with $|\alpha_a| > 1$, does not.

In scattering theory, the initial vector is always given and in general is not an eigenfunction of the operator K; therefore the analysis is not as direct as we have presented in the previous paragraph. For example, in the two-body scattering problem in which $K(w) = G_0(w)V = (wI - H_0)^{-1}V$, $K(w)$ is compact if V is square-integrable, the initial vector ϕ_a is always an approximate eigenfunction associated with the continuous spectrum of the unperturbed Hamiltonian H_0; i.e., there exists an $\epsilon > 0$ upon which Φ_a depends, such that for some energy E, $\|(EI - H_0)\Phi_a\| < \epsilon$. Even if we relax the strictures of the Hilbert-space approach and allow ϕ_a to be eigenfunction of H_0, $H_0\phi_a = E\phi_a$, ϕ_a is not an eigenfunction of K, for, if $\lambda K\chi = \chi$, then $G_0(w)(WI - H_0 - \lambda V)\chi = 0$, and since $G_0(w)$ is bounded for W in the resolvent set of H_0, χ is an eigenfunction of $H_0 + \lambda V$, with eigenvalue $W = E + i\epsilon$, not an eigenfunction of H_0. (The case $\lambda = 0$ tells us that $\chi = 0$.)

VECTOR SERIES AND NONCOMPACT OPERATORS

The operator $K(\omega) = G_0(\omega)\sum_{i<j} V_{ij}$, which is the generator of the Born series in the scattering processes involving three or more particles, is not a compact operator. This means that the operator methods, which worked for two-body processes, where the pertinent operator was compact or reducible to a compact form for well-behaved potentials, are inapplicable. On the other hand, the vector series, whose efficacy does not depend upon the compactness of the generating operator, can be used in the same way for two-body or for many-body calculations.

We will demonstrate this property of the vector series by first displaying a simple mathematical example.

$K = -d^2/dx^2$ acts upon the space of real functions which are square-integrable over the interval $[0, \infty]$. Its domain is the set of functions which have piecewise continuous second derivative and for which $u(0) = 1$ and both $u(x)$ and $u''(x)$ are square-integrable. It is easy to show that K is not self-adjoint, it is semi-bounded from below and has a continuous spectrum on the positive real axis of the λ plane. This last property implies that K is not compact. Nevertheless, if we consider the series of vectors $\sum_{n=0}^{\infty} \lambda^n\chi_n$, where $\chi_n = K^n\phi$, and take $\phi = e^{-x/2}$, we find, by computing the radius of convergence of the series, that it converges in the circle $|\lambda| < 2$. Therefore the vector Born series

($\lambda = 1$) converges, and converges to $\frac{4}{5}e^{-x/2} = (I - K)^{-1}\phi$. The vector series does not converge for all choices of ϕ, for example, if $\phi = e^{-x}$, the Born series diverges. Thus we see the importance of the choice of initial vector ϕ, which choice we expect to remain important in more complicated situations. This problem is not the best analogy to the situation that arises in scattering theory because, in order to simplify the calculation, we took ϕ to be an eigenvector of K corresponding to a discrete eigenvalue.

We would like, finally, to show the way in which the one-dimensional example considered by Dettmann and Leibfried is compatible with our theoretical considerations. To this end we compute the radius of convergence of the vector series for the initial state in this problem. These authors do not consider the full Born series but only the subseries $\sum_{n=0}^{\infty}(G_0v)^n\psi_0$, where $\psi_0 = G_0v'\phi$.[5] In this exercise, one of the three particles is infinitely heavy and is bound to the origin and the coordinates of the other two particles with masses m and M are designated respectively by x and X. In the initial channel, the particle of mass m is bound to the origin by a potential $v(x) = -\alpha\delta(x)$, and the initial-state vector is

$$\phi(X, x) = \left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}} e^{-\kappa|X|}e^{-iKx}.$$

Unfortunately, functions of this type are not L^2 and thus cannot be included in the spaces which were used in our theoretical discussion. Even though it is not necessary to use Hilbert spaces to frame the properties of vector-valued functions of a complex variable, all we need is a Banach space; it is simpler to modify the function ϕ so that it does belong to L^2 than to search for a Banach space to which ϕ belongs, and which furnishes the desired convergence properties for the linear functionals. In our calculation, ϕ was multiplied by a factor $(2\pi\beta)^{\frac{1}{2}}e^{-\beta|x|}$, so that ϕ had unit norm in the space $L^2(E_2)$. β may be thought of as representing a spread in the incident momentum of the particle with mass M. At the end of the calculation, after the limits have been taken, β is put equal to zero. On account of this modification, we cannot be certain that our results can be directly related to those of Dettmann and Leibfried even though the answers are numerically similar.

The unperturbed Green's function G_0 should be evaluated at the complex three-body energy $E + i\eta$, and the limit $\eta \to +0$ is taken at the end, after the limit $n \to \infty$. The only other important aspect of the calculation is that the initial state is evaluated at the total three-body energy

$$E = -\frac{\hbar^2\kappa^2}{2m} + \frac{\hbar^2K^2}{2M}.$$

Since the total energy is constant, this expression is used at each step in the calculation to eliminate the variable K.

$$\psi_0(X, x) = G_0v'\phi,$$

where

$$v' = -\alpha'\delta(x') = -\alpha'\delta(X - x),$$

is the potential which binds the particles in the final state.

$$\kappa' = \frac{mM\alpha'}{\hbar^2(m + M)},$$

$$\psi_0(X, x) = -\kappa'\frac{(\kappa\beta)^{\frac{1}{2}}}{ip}\exp iK\left(\frac{MX + mx}{M + m}\right) \times \exp -(\kappa + \beta)\left|\frac{MX + mx}{M + m}\right| \exp ip\,|X - x|,$$

where

$$p^2 = \frac{2mM}{\hbar^2(m + M)} \times \left[E + i\eta + \frac{\hbar^2}{2(M + m)}(-K^2 + \kappa^2 + \beta^2)\right]$$

and

$$(G_0v)^n\psi_0 = \frac{\kappa'}{ip}(\kappa\beta)^{\frac{1}{2}}\left(\frac{\kappa}{iq}\right)^n \exp iq\,|X| \exp ip\,|x| \times \exp -(\kappa + \beta)\frac{M\,|x|}{M + m} \exp iK\frac{MX}{M + m}.$$

$(G_0v)^n\psi_0$ is $L^2(E_2)$, provided that η is nonzero, so that we can form $\|(G_0v)^n\psi_0\|^{1/n}$ and take the limit as $n \to \infty$:

$$\overline{\lim_{n\to\infty}}\,\|(G_0v)^n\psi_0\|^{1/n} = \left|\frac{\kappa}{q}\right|,$$

where

$$q^2 = \frac{2m}{\hbar^2}\left[E + i\eta + \frac{\hbar^2}{2M} \times \left(-p^2 - \frac{K^2M^2}{(m + M)^2} + \frac{(\kappa^2 + \beta^2)M^2}{(M + m)^2}\right)\right].$$

Now we take the limits $\beta \to 0$, $\eta \to +0$, to obtain the result that the Born series $\sum_{n=0}^{\infty}(G_0v)^n\psi_0$ converges for the total three-body energy $E > E_B(m^2 + M^2)/mM$, where $E_B = \hbar^2\kappa^2/2m$ is the magnitude of the binding energy of the initial two-body bound state.

This result appears to agree rather well with the result obtained by Dettmann and Liebfried that, for $E_B < E_B'(m'/m)(M + m)^2/M^2$ (where E_B' is the binding energy in the final two-particle subsystem), the Born series for the matrix elements converges for $E > E_B$.

[5] Throughout this calculation we use the notation of the paper of Dettmann and Leibfried, and any terms undefined here are taken from that paper.

The fact that $[(m + M)^2/mM]\, E_B$ is always greater than E_B fits the theoretical conclusion that the series for the matrix elements is convergent over a larger region than the series for the corresponding vectors. Nevertheless, for the reasons we have referred to previously, this agreement should not be taken to be more than an indication of why the counter example works.

If we had chosen to expand the final state vector, it is apparent that the energies at which the series converged would depend upon the binding energy in the two particle subsystem of the final configuration. Furthermore, in this example, the region of the total energy plane, where the series for the matrix elements convergences, would depend upon both the final and initial binding energies.

This result does not prove that the actual vector Born series for this example converges in this energy range. We have considered only a subseries of the total Born series and from the form of the total vector Born series, we can see that the radius of convergence, if it is nontrivial, will depend in some complicated manner on both the coupling constants κ and κ'.

ACKNOWLEDGMENTS

I would like to express my thanks to Professor E. H. Thiel of the Mathematics Department and U. C. Davis for many helpful discussions.

APPENDIX: VECTOR-VALUED FUNCTIONS OF A COMPLEX VARIABLE

The theorems to be proved are standard in the mathematical literature,[6] nevertheless, it seems worthwhile to present them here. The methods are analogous to those used in ordinary complex analysis.

Theorem 1: If $\psi(\lambda)$ is an analytic vector for λ in some finite open domain U, then it possesses bounded analytic derivatives $\psi^{(n)}(\lambda_0)$ for $\lambda_0 \in U$, where

$$(\phi, \psi^{(n)}(\lambda)) = \frac{d^n}{d\lambda^n}(\phi, \psi(\lambda))|_{\lambda=\lambda_0},$$

for any ϕ in the Hilbert space. Furthermore, if the circle $|\lambda - \lambda_0| \leq r$ is contained in U then $\psi(\lambda)$ is given by the Taylor series

$$\psi(\lambda) = \sum_{n=0}^{\infty} \frac{(\lambda - \lambda_0)^n}{n!} \psi^{(n)}(\lambda_0),$$

which, within this circle, converges uniformly and absolutely.

In the proof we will use the notation $\overline{\lim}$ for the limit superior.

Proof. For $\lambda_0 \in U$, $(\phi, \psi(\lambda))$ is an analytic function, for all ϕ, and hence by Cauchy's integral formula

$$\frac{d^n}{d\lambda^n}(\phi, \psi(\lambda_0)) = \frac{n!}{2\pi i} \oint_C \frac{(\phi, \psi(z))}{(z - \lambda_0)^{n+1}}\, dz,$$

where C is the circle $|z - \lambda_0| = r$. Now

$$\left| \frac{d^n}{d\lambda^n}(\phi, \psi(\lambda_0)) \right| \leq \frac{n!}{r^n} \|\phi\| \, \|\psi(C)\|,$$

where $\|\psi(C)\|$ is sup of the value of $\|\psi(z)\|$ for z on the circle C. (We know they are uniformly bounded because $\psi(\lambda)$ is analytic on C.) Thus we have a bounded linear functional and there must exist a bounded vector $\psi^{(n)}(\lambda_0)$ such that

$$(\phi, \psi^{(n)}(\lambda_0)) = \frac{d^n}{d\lambda^n}(\phi, \psi(\lambda_0))$$

by a theorem on linear functionals. Since the proof is independent of the choice of $\lambda_0 \in U$, we have proven the first result; the analyticity follows directly from that of $d^n/d\lambda^n(\phi, \psi(\lambda))$ for $\lambda \in U$. The bound of

$$\psi^{(n)}(\lambda_0) \quad \text{is} \quad \|\psi^{(n)}(\lambda_0)\| \leq \frac{n!}{r^{n+1}} \|\psi(C)\|.$$

The validity of the Taylor series also follows from the Cauchy formula. Consider the remainder vector

$$\xi_p(\lambda) = \psi(\lambda) - \sum_{n=0}^{p} \frac{(\lambda - \lambda_0)^n}{n!} \psi^{(n)}(\lambda_0),$$

and form the product

$$(\phi, \xi_p(\lambda)) = \frac{(\lambda - \lambda_0)^{p+1}}{2\pi i} \oint_C \frac{(\phi, \psi(z))\, dz}{(z - \lambda_0)^{p+1}(z - \lambda)},$$

where C is the circle $|z - \lambda_0| = r$ contained in U and $|\lambda - \lambda_0| < r$. Then

$$|(\phi, \xi_p(\lambda))| \leq \frac{|\lambda - \lambda_0|^{p+1}}{r^{p+1}} \frac{r}{\rho} \|\phi\| \, \|\psi(C)\|,$$

where ρ is the shortest distance between λ and the circle C. Therefore $\xi_p(\lambda)$ is bounded, as this result holds for all ϕ, and

$$\|\xi_p(\lambda)\| \leq \frac{|\lambda - \lambda_0|^{p+1}}{r^{p+1}} \frac{r}{\rho} \|\psi(C)\|.$$

The right-hand side of this expression tends to zero as $p \to \infty$ because $|\lambda - \lambda_0| < r$. This proves the uniform convergence.

$$\lim_{p \to \infty} \left\| \psi(\lambda) - \sum_{n=0}^{p} \frac{(\lambda - \lambda_0)^n}{n!} \psi^{(n)}(\lambda_0) \right\| = 0.$$

The absolute convergence follows just as easily. Let

$$\chi_p(\lambda) = \sum_{n=0}^{p} \frac{(\lambda - \lambda_0)^n}{n!} \psi^{(n)}(\lambda_0).$$

[6] N. Dunford and J. T. Schwartz, *Linear Operators Part I.* (Interscience Publ., Inc., New York, 1958), sec. III, 14; E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups* (American Mathematical Society Colloquium Publ., Providence, R.I. 1957), Vol. 31.

Then

$$(\phi, \chi_p(\lambda)) = \sum_{n=0}^{p} \frac{(\lambda - \lambda_0)^n}{n!} (\phi, \psi^{(n)}(\lambda_0))$$
$$= \sum_{n=0}^{p} \frac{(\lambda - \lambda_0)^n}{2\pi i} \oint_C \frac{(\phi, \psi(z))}{(z - \lambda_0)^{n+1}} dz,$$

by Cauchy. Hence

$$|(\phi, \chi_p(\lambda))| \leq \|\phi\| \|\psi(C)\| \frac{\left(1 - \frac{|\lambda - \lambda_0|^p}{r^p}\right)}{\left(1 - \frac{|\lambda - \lambda_0|}{r}\right)},$$

and

$$\lim_{p \to \infty} |(\phi, \chi_p(\lambda))| \leq \|\phi\| \|\psi(C)\| \frac{1}{\left(1 - \frac{|\lambda - \lambda_0|}{r}\right)} < \infty.$$

But this is true for all $\phi \in \mathcal{H}$; therefore,

$$\lim_{p \to \infty} \|\chi_p(\lambda)\| \leq \frac{1}{\left(1 - \frac{|\lambda - \lambda_0|}{r}\right)} < \infty,$$

which is the condition for absolute convergence. Q.E.D.

Theorem 2: Any power series $\sum_{p=0}^{\infty} \psi_p(\lambda - \lambda_0)^p$ defines an analytic function $\psi(\lambda)$ in an open set $|\lambda - \lambda_0| < s$ of the complex λ plane, where

$$s = \left(\overline{\lim_{p \to \infty}} \|\psi_p\|^{1/p}\right)^{-1}.$$

Proof. Consider the partial-sum vectors $\chi_p(\lambda) = \sum_{n=0}^{p} (\lambda - \lambda_0)^n \psi_n$. For $p > q$, $|(\phi, \chi_p(\lambda) - \chi_q(\lambda))| \leq \sum_{n=q+1}^{p} |\lambda - \lambda_0|^n \|\phi\| \|\psi_n\|$, by the Schwartz inequality. The right-hand side will tend to zero for sufficiently large p and q, if λ is chosen so that

$$|\lambda - \lambda_0| < (\|\psi_n\|^{1/n})^{-1},$$

for n large enough. This will be true in the limit as $n \to \infty$, if

$$|\lambda - \lambda_0| < \left(\overline{\lim_{n \to \infty}} \|\psi_n\|^{1/n}\right)^{-1}.$$

That is to say, by the Cauchy convergence criterion, $(\phi, \chi_p(\lambda))$ will converge to some bounded linear functional $a(\phi)$. Then by a theorem on linear functionals, there exists a unique bounded vector-valued function $\chi(\lambda)$ such that

$$a(\phi) = (\phi, \chi(\lambda)) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n (\phi, \psi_n(\lambda_0)).$$

The vector $\chi(\lambda)$ is analytic within the circle

$$|\lambda - \lambda_0| < \left(\overline{\lim_{n \to \infty}} \|\psi_n\|^{1/n}\right)^{-1},$$

from the analyticity of the ordinary functions

$$(\phi, \chi(\lambda)) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n (\phi, \psi_n(\lambda_0)).$$

Q.E.D.

From the previous theorem (Theorem 1), the series is uniquely determined by its vector sum $\chi(\lambda)$ and the uniqueness of the power series within its circle of convergence,

$$\psi_n(\lambda_0) = \frac{\chi^{(n)}(\lambda_0)}{n!},$$

and, therefore, $\chi(\lambda) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n \psi_n(\lambda_0)$ converges uniformly and absolutely within the circle of convergence.

We also have to show that the vector we have constructed from the power series is a solution of the Lippmann–Schwinger equation.

Theorem 3: If $\chi(\lambda) = \sum_{n=0}^{\infty} \lambda^n \psi_n$, where $\psi_n = K^n \psi_a$, then, for λ in the circle of convergence of the power series, $\chi(\lambda)$ satisfies the Lippmann–Schwinger equation $\Psi(\lambda) = \psi_a + \lambda K \Psi(\lambda)$.

Proof. Consider the partial sum $\chi_p = \sum_{n=0}^{p} \lambda^n \psi_n$, for λ in the circle of convergence. The following equation is identically true $\chi_p(\lambda) - \psi_a = \lambda K \chi_p(\lambda) - \lambda^{p+1} \psi_{p+1}$. Therefore

$$|(\phi, \chi_p(\lambda) - \psi_a - \lambda K \chi_p(\lambda))| \leq \|\phi\| \, |\lambda|^{p+1} \, \|\psi_{p+1}\|,$$

and

$$|(\phi, \chi(\lambda) - \psi_a - \lambda K \chi(\lambda))| \leq \|\phi\| \{|\lambda|^{p+1} \|\psi_{p+1}\| + \|I - \lambda K\| \|\chi(\lambda) - \chi_p(\lambda)\|\};$$

hence, if $(I - \lambda K)$ is bounded, in the limit as $p \to \infty$, the right-hand side vanishes. The first term vanishes because within the circle of convergence $|\lambda| < \left(\overline{\lim_{p \to \infty}} \|\psi_n\|^{1/p}\right)^{-1}$ and the last by the uniform convergence for λ within the circle. From this we can conclude that $\chi(\lambda) = \psi_a + \lambda K \chi(\lambda)$, which was to be proved. If $(I - \lambda K)$ is not bounded, but the range of $(I - K)^{\dagger}$ is dense in the Hilbert space, the same result follows. Q.E.D.

Are Bloch Bands at Finite Field Adiabatically Connected to Those at Zero Field?*

G. H. WANNIER AND J. P. VAN DYKE†
Physics Department, University of Oregon, Eugene, Oregon

(Received 6 November 1967)

It has been shown previously that if a potential consists of a superposition of a periodic part and a uniform electric field, then for a particle moving in this field there are Bloch bands closed in time. The present paper addresses itself to the question whether these bands may be identified with the field-free bands. The most natural thing is to expect that the bands are slightly field dependent, but converge toward the field-free bands as E goes to zero. Bands for which this is true are said to be adiabatically connected to corresponding bands at zero field. In Sec. 2 of the paper, two model cases are given for which this adiabatic connection pertains. Section 3 is the central part of the paper and provides the conclusion that the answer to the question in the title is almost always negative. In this proof the positive cases serve an essential function. It is shown that the parameters of the periodic potential must obey at least one supplementary condition to allow adiabatic connection, and that the collected cases precisely obey this condition. Adiabatic connection is thus generally not possible. Section 4 provides an explicitly soluble case which does not allow adiabatic connection. An infinite number of field values E converging toward zero are found at each of which the two bands under consideration switch identity (hyperbolic rather than linear connection at energy crossings). The connection postulated in the effective-mass approximation must therefore be of a nonadiabatic nature. It probably involves the "sudden" approximation of quantum theory.

1. INTRODUCTION

A number of papers[1-3] involving one of the authors have appeared, investigating the compatibility of the concept of Bloch bands with the presence of electromagnetic fields. As one would expect on physical grounds, such compatibility, if existing, is entirely limited to static uniform electric and magnetic fields. Within this domain the situation is as follows. If only a uniform electric field is present, then one can prove rigorously that Bloch bands exist which are closed in time. If we look at the motion of the particle within such a band we find that it moves according to the law

$$\mathbf{k} = \mathbf{k}_0 + e\mathbf{E}t/\hbar. \tag{1}$$

The proof of these facts does not depend on power-series expansion in the field E. It is in fact possible to derive directly the basic equation used for power-series expansion with the help of the proof. Such a proof is not available in the case of a magnetic or of mixed fields. The existence of closed bands is still linked in these cases with the notion of power-series expansion. It is therefore possible in these latter cases that the existence of bands is only valid asymptotically for small fields.

One might very well think that once the notion of Bloch bands is proved for the case of a uniform electric field, this case is entirely disposed of. This is not so. There remains the question of the identity of these Bloch bands which is by no means trivial. Although Eq. (2) of Ref. 1, which is most conveniently used to expand the Bloch functions in powers of the field, is proved as a valid equation for the true Bloch functions, the expansion itself is not proved thereby as valid. Another way to see the same thing is to observe the operator which defines these bands for finite field. It is written out in Eq. (22) of Ref. 2. The operator is of such a structure that going to the limit of zero field is not possible for it. One must remember here that the question of the feasibility of this limiting operation is an extremely serious one. The entire effective-mass approximation which is used in thousands of papers depends on it. It must be possible to take this limit in some sort of way. However, the way in which to take it is not now known. We shall show in this paper that the limit is not simply an adiabatic limit. In other words the power series derived from Eq. (2) of Ref. 1 is almost always (in the mathematical sense) divergent. In fact, the bands at any finite field are separated from the bands at zero field by an infinite number of singular-field values at which they switch identity. The proof for this divergence is given below by the method of perturbed parameters. In other words we start out from a model case in which the power series is known to be valid. We embed this case in a wider class of model cases of variable parameter. We then show that there is at least one singular field value in this wider class unless this parameter β has exactly the value appropriate for the

* This work was supported by the U.S. Office of Naval Research.
† Present address: Lockheed Palo Alto Research Laboratory, Palo Alto, California.
[1] G. H. Wannier, Phys. Rev. **117**, 432 (1960).
[2] G. H. Wannier and D. R. Fredkin, Phys. Rev. **125**, 1910 (1962).
[3] G. H. Wannier, Rev. Mod. Phys. **34**, 645 (1962).

special model. By this means the power series is proved as almost always invalid.

2. PRESENTATION OF THE CONVERGENT CASES

As stated in the introduction, the proof of divergence is based on a class of model cases for which the power series is known to converge. These cases must now be presented first. Let the periodic Hamiltonian be denoted by $\mathcal{H}(\mathbf{p}, \mathbf{x})$. It obeys

$$\mathcal{H}(\mathbf{p}, \mathbf{x} + \boldsymbol{\rho}) = \mathcal{H}(\mathbf{p}, \mathbf{x}), \tag{2a}$$

where

$$\boldsymbol{\rho} = l\mathbf{a} + m\mathbf{b} + n\mathbf{c}, \tag{2b}$$

with

$$l, m, n \text{ integer}. \tag{2c}$$

Let its Bloch-type eigenfunctions be denoted by $b_q(\mathbf{x}; \mathbf{k})$ and its eigenvalues by $W_q(\mathbf{k})$. The Hamiltonian $K(\mathbf{p}, \mathbf{x})$ of our problem then takes the form

$$K(\mathbf{p}, \mathbf{x}) = \mathcal{H}(\mathbf{p}, \mathbf{x}) - \mathbf{E} \cdot \mathbf{x}, \tag{3}$$

where the previously defined dimensionless units have been employed. We look for a solution of (3) of the form

$$\psi(\mathbf{x}, t) = \sum_q \chi_q(t) b_q(\mathbf{x}, \mathbf{k}_0 + \mathbf{E}t). \tag{4}$$

The summation in (4) does not include a summation over $\mathbf{k}$ because $\mathbf{k}$ is conserved by the Hamiltonian (3) if it is made time-dependent in accordance with Eq. (1). Substitution of (4) and (3) into the time-dependent Schrödinger equation yields the equation system

$$i(\partial\chi_n/\partial t) = W_n(\mathbf{k})\chi_n - \mathbf{E} \cdot \sum_q \mathbf{X}_{nq}(\mathbf{k})\chi_q. \tag{5a}$$

Here, the quantities $\mathbf{X}_{nq}$ are the matrix elements of the Adams operator,[4] namely,

$$\mathbf{X}_{nq} = \int b_n^*(\mathbf{x}; \mathbf{k})[\mathbf{x} + i(\partial/\partial\mathbf{k})]b_q(\mathbf{x}; \mathbf{k})\, d\tau. \tag{5b}$$

The special models mentioned above result from the equation system (5) in the following way. First we make the model one-dimensional, suppressing the nonessential axes at right angles to $\mathbf{E}$. Thereupon we think of the mixing as taking place between two nonoverlapping bands 1 and 2 which we consider exclusively. We suppress therefore all matrix elements X_{1q} and X_{2q} leading out of this pair. In addition we suppress X_{11} and X_{22} as nonessential. We are then left with the equation pair

$$i(\partial\chi_1/\partial t) = W_1\chi_1 - EX\chi_2,$$
$$i(\partial\chi_2/\partial t) = W_2\chi_2 - EX^*\chi_1.$$

It is convenient in the following to eliminate the two types of variables t and k, and to make χ_1 and χ_2 also functions of the variable (1). This is no restriction because all coefficients depend only on that combination. The equations become then

$$iE(d\chi_1/dk) = W_1(k)\chi_1(k) - EX(k)\chi_2(k), \tag{6a}$$
$$iE(d\chi_2/dk) = W_2(k)\chi_2(k) - EX^*(k)\chi_1(k). \tag{6b}$$

$W_1(k)$, $W_2(k)$, and $X(k)$ are periodic functions of k of period $2\pi/d$. The system (6) is a pair of ordinary differential equations of the first order. It has two solution pairs, and the general solution is a linear combination of these two pairs. All solutions obey the law of probability conservation; in other words, Eqs. (6) have the constant integral

$$\chi_1^*\chi_1 + \chi_2^*\chi_2 = 1, \tag{7}$$

which we set equal to unity as is customary.

When we contemplate the system (6), we observe that we are dealing with two oscillators coupled by the term EX. If this term is small and if it does not upset the "labeling" of the states in a fundamental way, then this term provides only a slight modification, and the exact solutions can still be associated with one or the other of the two bands. If this is true we shall have two solution pairs behaving approximately as

$$\chi_1^{(1)} \approx \exp\left[-\frac{i}{E}\int^k W_1(k)\, dk\right], \tag{8a}$$

$$\chi_2^{(1)} \approx \frac{EX^*}{W_2 - W_1}\exp\left[-\frac{i}{E}\int^k W_1(k)\, dk\right], \tag{8b}$$

and

$$\chi_1^{(2)} \approx -\frac{EX}{W_2 - W_1}\exp\left[-\frac{i}{E}\int^k W_2(\kappa)\, d\kappa\right], \tag{9a}$$

$$\chi_2^{(2)} \approx \exp\left[-\frac{i}{E}\int^k W_2(\kappa)\, d\kappa\right]. \tag{9b}$$

The solutions (8) and (9) have an analytic singularity at $E = 0$ which was first worked out by Houston.[5] As long as the singularity affects only the phase in this predetermined way we may consider it as nonessential.

A second preliminary observation deals with the variable X. At first sight one might think of its modulation as contributing something essential to the problem. This is not so, for we can modify the independent variable k by the substitution

$$y = \int^k |X|\, dk, \tag{10}$$

and express $W_1/|X|$ and $W_2/|X|$ as new periodic functions of y. It is therefore no restriction if we occasionally treat XX^* as a constant, since it will become so after this transformation.

[4] E. N. Adams, J. Chem. Phys. **21**, 2013 (1953).

[5] W. V. Houston, Phys. Rev. **57**, 184 (1940).

After these observations we come to the main purpose of this section, which is to find real solutions, valid at finite field, which behave asymptotically as (8) for small values of E. To facilitate this we remove first the singular phase factor in (8) by the substitution

$$\chi_1^{(1)} = F \exp\left[-\frac{i}{E}\int^k W_1(\kappa)\,d\kappa\right], \tag{11a}$$

$$\chi_2^{(1)} = G \exp\left[-\frac{i}{E}\int^k W_1(\kappa)\,d\kappa\right], \tag{11b}$$

and insert into (6). We get

$$i[dF(k)/dk] = -X(k)G(k), \tag{12a}$$

$$i[dG(k)/dk] = [W(k)G(k)/E] - X^*(k)F(k), \tag{12b}$$

where

$$W(k) = W_2(k) - W_1(k) \tag{13}$$

and, as before,

$$F^*(k)F(k) + G^*(k)G(k) = 1. \tag{7}$$

We should now expect that if a solution of the type (8) exists, it means that F and G approach 1 and 0 continuously as E diminishes. Such a situation prevails *formally* if we write down the series-expansion solution in E of (12) (there is only one such). However, we are not interested in series expansions which are carried out in a much more general framework elsewhere[1] but in *actual functions* F and G for which an eventual series might be a convergent expansion.

The first case in which these conditions are satisfied is essentially trivial. It is the case in which both $W(k)$ and $X(k)$ are constants independent of k. The solutions of (12) are then simple exponentials. One finds with (7)

$$F(k) = \left(\frac{R+1}{2R}\right)^{\frac{1}{2}} \exp i\,\frac{2EXX^*k}{W(R+1)}, \tag{14a}$$

$$G(k) = \left(\frac{R-1}{2R}\right)^{\frac{1}{2}}\left(\frac{X^*}{X}\right)^{\frac{1}{2}} \exp i\,\frac{2EXX^*k}{W(R+1)}, \tag{14b}$$

where

$$R = [1 + (4E^2XX^*/W^2)]^{\frac{1}{2}}. \tag{15}$$

The solutions (14) are clearly of the required form. They are expandable in a convergent power series in E and reduce to the limit

$$F = 1, \quad G = 0$$

as E goes to zero. The solution pair is also free of singularities for all E.

The other case in which a regular solution was found for (12) arises in the following special circumstances:

$$X(k) = \tfrac{1}{2}d, \tag{16a}$$

$$W(k) = w/(1 + \epsilon\cos kd)^{\frac{1}{2}}. \tag{16b}$$

The way in which a solution can be derived need not preoccupy us here. One can verify by direct substitution into (12) and (7) that the following pair is correct

$$F(k) = (2R)^{-\frac{1}{2}}[R + 1 - \tfrac{1}{2}(\epsilon E^2d^2/w^2)\cos kd]^{\frac{1}{2}} \times \exp i\int_0^{kd} \frac{\tfrac{1}{2}(Ed/w)(1+\epsilon\cos x)^{\frac{1}{2}}}{R + 1 - \tfrac{1}{2}(\epsilon E^2d^2/w^2)\cos x}\,dx, \tag{17a}$$

$$G(k) = (2R)^{-\frac{1}{2}}[R - 1 + \tfrac{1}{2}(\epsilon E^2d^2/w^2)\cos kd]^{\frac{1}{2}} \times \exp i\left[-\arctan\frac{\epsilon Ed}{2w}\frac{\sin kd}{(1+\epsilon\cos kd)^{\frac{1}{2}}} + \int_0^{kd}\frac{\tfrac{1}{2}(Ed/w)(1+\epsilon\cos x)^{\frac{1}{2}}}{R + 1 - \tfrac{1}{2}(\epsilon E^2d^2/w^2)\cos x}\,dx\right], \tag{17b}$$

where we defined, consistently with (15),

$$R = [1 + (E^2d^2/w^2) + (\epsilon^2E^4d^2/4w^4)]^{\frac{1}{2}}. \tag{18}$$

Just as in the preceding case the solution (17) is defined and analytic for all real values of E. It is expandable in a power series in E and approaches $F = 1$ and $G = 0$ for E going to zero. In distinction from the preceding case, the solution is not trivial. It is generally supposed that the oscillations of the energy $W(k)$ with k induce transitions between bands. We have such oscillations here, and nevertheless the states for finite E are analytically connected with the states for $E = 0$. In particular, if the system was originally in the band 1, then an adiabatic increase of E will produce the state (17). Inversely, an adiabatic reduction of E will reduce the solution (17) into $F = 1$, $G = 0$. In other words, the bands at finite field are adiabatically connected to the bands at zero field.

3. PROOF THAT ADIABATIC CONNECTION IS EXCEPTIONAL

We shall now use the two solutions (14) and (17) as a base to prove the very opposite of what they seem to teach us, namely, that adiabatic connection is almost never present. For this purpose we observe that the two solutions agree in their overlapping region $X = \frac{1}{2}d$, $\epsilon = 0$. We can then think of the second solution as arising from the first by the growth of the parameter ϵ which gives the bands a width and variability in k. However, we can introduce these features in many other ways than by the formula (16b). One obvious way to do it is to change the case $W = \text{const}$ to $W(k) = W_0 + \epsilon W_1\cos kd$. If we find the solution in powers of ϵ, we know it must converge to first order since it can be, to that order, a Taylor expansion of (16b). We shall discuss a more general case by

setting

$$W(k) = w\left\{1 - \tfrac{1}{2}\epsilon \cos kd + \sum_{m=2}^{\infty} \beta_m(\epsilon \cos kd)^m\right\}.$$

To consider this in a more transparent way, write F in the form

$$F(k) = F_0(E, \epsilon) \exp\left[i \frac{2E|X|^2}{w(1+R)} \int^k f(x)\, dx\right], \quad (19a)$$

with R again defined by (15). Then from (12a),

$$G(k) = \frac{2EX^*}{w(1+R)} f(k) F_0(E, \epsilon) \times \exp\left[i \frac{2E|X|^2}{w(1+R)} \int^k f(x)\, dx\right]. \quad (19b)$$

Since $F_0(E, \epsilon)$ can be chosen to satisfy condition (7), we have reduced our considerations to the single functions $f(k)$ which because of (8b) should approach $w/W(k)$ continuously as E goes to zero.

After dropping the common factor we get from (12b) the following equation for $f(k)$:

$$i\,2E \frac{df}{dk} - \frac{4E^2|X|^2}{w(1+R)} f^2 - 2W(k) f + w(1+R) = 0. \quad (20)$$

To proceed in powers of ϵ, let

$$f(k) = \sum_{m=0}^{\infty} f_m(k) \epsilon^m.$$

This results in the equation for f_0:

$$i\,2E \frac{df_0}{dk} - \frac{4E^2|X|^2}{w(1+R)} f_0^2 - 2w f_0 + w(1+R) = 0. \quad (21)$$

Of its several solutions only $f_0 = 1$ behaves in the desired manner, and this solution makes (19a) the same as (14a) to the zeroth order—as it must be.

The equation for the $f_n (n \neq 0)$ is

$$i\,2E \frac{df_n}{dk} - \frac{4E^2|X|^2}{w(1+R)} \sum_{m=0}^{n} f_m f_{n-m} - 2w \sum_{m=0}^{n} f_{n-m} \beta_m (\cos kd)^m = 0, \quad (22a)$$

or

$$i\,2E \frac{df_n}{dk} - 2wRf_n = w(R-1) \sum_{m=1}^{n-1} f_m f_{n-m} + 2w \sum_{m=1}^{n} f_{n-m} \beta_m (\cos kd)^m, \quad (22b)$$

which can be written as:

$$iE \frac{df_n}{dk} - wRf_n = \sum_{m=0}^{n} K^c_{nm} \cos mkd + \sum_{m=1}^{n} K^s_{nm} \sin mkd. \quad (23)$$

That (23) follows from (22) can be proved by induction if one ignores solutions to the homogeneous equation $iE(df_n/dk) - wRf_n = 0$. The solution to Eq. (23) will be

$$f_n = \sum_{m=0}^{n} f^c_{nm} \cos mkd + \sum_{m=1}^{n} f^s_{nm} \sin mkd, \quad (24a)$$

where

$$f^c_{nm} = (-wRK^c_{nm} - im\,E\,dK^s_{nm})/(w^2R^2 - m^2E^2d^2), \quad (24b)$$

$$f^s_{nm} = (-wRK^s_{nm} + im\,E\,dK^c_{nm})/(w^2R^2 - m^2E^2d^2). \quad (24c)$$

Since $\beta_0 = 1$, $\beta_1 = -\frac{1}{2}$, $K^c_{11} = -\frac{1}{2}w$, and $K^c_{10} = K^c_{11} = 0$,

$$f_1 = \frac{w^2R}{2[w^2R^2 - E^2d^2]} \cos kd - \frac{iE\,dw}{2[w^2R^2 - E^2d^2]} \sin kd, \quad (25)$$

which is well-behaved for all values of the field as long as $|X| \geq d/2$.

However, if we proceed to $m > 1$, the denominator in (24b) and (24c) will vanish unless the inequality for X is made more stringent. Finally, there will be, for any fixed X, an integer m_0 such that

$$m_0 d > 2|X|.$$

For all integers from m_0 on there exists a value E_m of the field which annuls the mth denominator. The value is given by

$$E_m = w/(m^2d^2 - 4X^2)^{\frac{1}{2}}. \quad (26)$$

These field values form an infinite sequence having $E = 0$ as a limit. This situation precludes adiabatic connection of the bands for finite field with those for zero field unless the numerators also vanish for those field values.

What we ask for in the above is precisely that the system of *inhomogeneous* equations

$$im\,Edf^s_{nm} - wRf^c_{nm} = K^c_{nm}, \quad (27a)$$

$$-wRf^s_{nm} - im\,Edf^c_{nm} = K^s_{nm}, \quad (27b)$$

have a solution when the determinant of the coefficients *does* vanish. This requires that

$$wR/im\,Ed = K^c_{nm}/K^s_{nm} \quad (28)$$

simultaneously with

$$w^2R^2 = m^2E^2d^2. \quad (26)$$

Let us look at the case $m = 2$ to see under what conditions (28) and (26) are simultaneously obeyed.

A little manipulation of (22b) reveals that

$$K^c_{22} = w(R-1)\tfrac{1}{4}[f^{c^2}_{11} - f^{s^2}_{11}] - \tfrac{1}{4}wf^c_{11} + \tfrac{1}{2}w\beta_2, \tag{29a}$$

$$K^s_{22} = \tfrac{1}{4}wf^s_{11}\{2(R-1)f^c_{11} - 1\}, \tag{29b}$$

when $w^2R^2 = 4E^2d^2$, $f^c_{11} = 2/3R$ and $f^s_{11} = -iw/6Ed$.

Without too much difficulty, Eq. (28) can be solved for β_2 showing that Eqs. (26) and (28) can be simultaneously solved only if $\beta_2 = \frac{1}{2}[1 - (|X|^2/d^2)]$ which is $\frac{3}{8}$ when $X = \frac{1}{2}d$, making $W(k)$ a Taylor expansion of (16b) to second order in ϵ.

We may infer from this the following. If we analyze the fairly general two-band case

$$W(k) = w\sum_{m=0}^{\infty}\beta_m\epsilon^m\cos^m kd$$

and look for a solution pair having the asymptotic form (8), then there is an infinite number of values E_m of the electric field obeying (26) for which we fail in this objective. If we start out by imposing the asymptotic phase behavior through the substitution (11), and thereupon circumscribe the solution by the substitution (19), we find that $f(k)$ becomes infinite at those field values. It is thus generally not possible to distinguish solution (8) from solution (9) across such a singular field value. Since the sequence of field values E_m forms an infinite set converging toward zero, the difficulty persists for any electric field, however small. The connection of Bloch functions at finite field with those at zero field is thus generally not possible. Exceptions to this rule exist, but they arise from a numerical relationship between the parameters which does not seem to have any physical basis.

4. EXAMPLE OF UNCONNECTED BANDS

It is relatively difficult to find examples of the "normal" situation even in the two-band case because our proof precludes simple analytic behavior of the solutions of (5) or (6) in the neighborhood of $E = 0$. We do, however, have a sample case of the solutions of (6) having "normal" behavior. It is found that the solutions (8) and (9) switch identity at every singular field value E_m. We shall exhibit this case in the following example which is somewhat reminiscent of the Kronig–Penney model.

Let X be constant [which is no real restriction as pointed out in (10)] and let $W(k)$ have the appearance shown in Fig. 1: generally a constant W with high-wall spikes at a distance a from each other. These spikes will be taken as δ-function spikes whose integral equals Δ. These spikes are of course the ones that

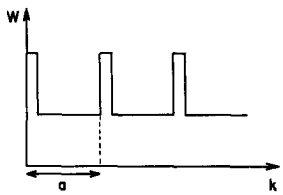


FIG. 1. Energy difference vs k in the model treated in Sec. 4.

disturb the solution which would be otherwise connectable as in (14).

For the present case we wish to discuss both solution pairs. The unsymmetric phase (11) was chosen for the specific purpose of showing power series expansion and will now be discarded. We make instead the symmetric ansatz

$$\chi_n(k) = \varphi_n(k)\exp\left[-\frac{i}{2E}\int^k (W_1 + W_2)\,d\kappa\right] \tag{30}$$

and take $X = 1$ as suggested by the transformation (10). The solution pair (14) is then still one of the valid solutions outside the spikes. We call it φ^+_i. With (30) it takes the form

$$\varphi^+_1 = \left(\frac{R+1}{2R}\right)^{\frac{1}{2}}\exp\left[i\frac{W}{2E}Rk\right], \tag{31a}$$

$$\varphi^+_2 = \left(\frac{R-1}{2R}\right)^{\frac{1}{2}}\exp\left[i\frac{W}{2E}Rk\right], \tag{31b}$$

where R is given by (15). The other solution pair of the equation system (6), which with (30) reads

$$i(d\varphi_1/dk) = -[W(k)/2E]\varphi_1 - \varphi_2, \tag{32a}$$

$$i(d\varphi_2/dk) = [W(k)/2E]\varphi_2 - \varphi_1, \tag{32b}$$

has the form

$$\varphi^-_1 = -\left(\frac{R-1}{2R}\right)^{\frac{1}{2}}\exp\left[-i\frac{W}{2E}Rk\right], \tag{33a}$$

$$\varphi^-_2 = \left(\frac{R+1}{2R}\right)^{\frac{1}{2}}\exp\left[-i\frac{W}{2E}Rk\right]. \tag{33b}$$

The solutions (31) and (33) do not, however, solve the equation system (32) entirely because of the spike of strength Δ in $W(k)$. Inserting this spike into (32) means that

$$\varphi_1(\text{high}) = \varphi_1(\text{low})\exp[i\Delta/2E], \tag{34a}$$

$$\varphi_2(\text{high}) = \varphi_2(\text{low})\exp[-i\Delta/2E]. \tag{34b}$$

The total phase shift of each solution over a period is the sum of the phase in (31) or (33), setting $k = a$ plus the phase (34). Because of this latter the solution pairs (31) and (33) fail to obey Floquet's theorem. We

have instead

$$\varphi_1^+(k+a) = \varphi_1^+(k)\exp[i(WRa+\Delta)/2E], \quad (35a)$$

$$\varphi_2^+(k+a) = \varphi_2^+(k)\exp[i(WRa-\Delta)/2E], \quad (35b)$$

$$\varphi_1^-(k+a) = \varphi_1^-(k)\exp[-i(WRa-\Delta)/2E], \quad (35c)$$

$$\varphi_2^-(k+a) = \varphi_2^-(k)\exp[-i(WRa+\Delta)/2E]. \quad (35d)$$

However, (31) and (33) are still solution pairs from which the Floquet solutions may be obtained by linear superposition. Denote by S^+ the solution pair (31) and by S^- the pair (33). A solution S obeying Floquet's theorem is then obtained by setting

$$S = p_+S^+ + p_-S^-. \quad (36)$$

Let the unknown Floquet factor be λ. Substitution of (31), (33), (35), and (36) into the Floquet condition for S yields

$$\begin{vmatrix} \left(\frac{R+1}{2R}\right)^{\frac{1}{2}}\left(\exp i\frac{WRa+\Delta}{2E}-\lambda\right) & -\left(\frac{R-1}{2R}\right)^{\frac{1}{2}}\left(\exp i\frac{-WRa+\Delta}{2E}-\lambda\right) \\ \left(\frac{R-1}{2R}\right)^{\frac{1}{2}}\left(\exp i\frac{WRa-\Delta}{2E}-\lambda\right) & \left(\frac{R+1}{2R}\right)^{\frac{1}{2}}\left(\exp i\frac{-WRa-\Delta}{2E}-\lambda\right) \end{vmatrix} = 0. \quad (37)$$

With

$$\lambda = e^{i\gamma}, \quad (38)$$

this becomes

$$\cos\gamma = \frac{1}{2}\left(1+\frac{1}{R}\right)\cos\frac{WRa+\Delta}{2E} + \frac{1}{2}\left(1-\frac{1}{R}\right)\cos\frac{WRa-\Delta}{2E} \quad (39)$$

or

$$\cos\gamma = \cos\frac{WRa}{2E}\cos\frac{\Delta}{2E} - \frac{1}{R}\sin\frac{WRa}{2E}\sin\frac{\Delta}{2E}. \quad (40)$$

Both (39) and (40) indicate that for small field R in Eq. (15) is close to 1 and $\cos\gamma$ differs little from $\cos[(WRa+\Delta)/2E]$. Equation (39) shows particularly clearly in which way the solution does depart from this value. $\cos\gamma$ is the superposition of two waves whose amplitudes sum to one and whose phases are not related. This means that $\cos\gamma$ cannot reach ±1, and $\sin\gamma$ cannot change sign. The angle γ thus stays locked in two quadrants while the phases on the right are many times 2π. To see the effect on the solution S we return the result into (36) and (37). We find after considerable manipulation

$$\frac{p_+^2}{p_-^2} = \exp\left[-i\frac{WRa}{E}\right]\frac{\sin\gamma + \frac{1}{2}(1+R^{-1})\sin[(WRa+\Delta)/2E] + \frac{1}{2}(1-R^{-1})\sin[(WRa-\Delta)/2E]}{\sin\gamma - \frac{1}{2}(1+R^{-1})\sin[(WRa+\Delta)/2E] - \frac{1}{2}(1-R^{-1})\sin[(WRa-\Delta)/2E]}. \quad (41)$$

It is seen that the ratio is either very large or very small, according to whether

$$\pm\sin\gamma \approx \sin[(WRa+\Delta)/2E]. \quad (42)$$

As the field E changes, the sign of $\sin\gamma$ cannot change, but the sign of the right-hand side in (42) switches whenever $WRa+\Delta = 2m\pi E$, which are the previously encountered singular field values (26). From (41) and (42), it is recognized that at these values the identity of the two bands is switched each time.

Clearly the result of this model case cannot describe entirely what happens when many bands are present. But the result obtained here, a switching of identity when the phases ϕ_a cross, is in line with what generally happens in adiabatic connection. It follows from this that the effective-mass approximation cannot get its validity from an adiabatic argument.

Angular Momentum and the Kerr Metric

JEFFREY M. COHEN
Institute for Space Studies, Goddard Space Flight Center, NASA, New York

(Received 14 September 1967)

For Kerr's rotating metric, it is shown that $-ma$ is the angular momentum of the body where m is the mass and a is the rotation parameter. This is true even for *large m and a.*

I. INTRODUCTION

In a previous paper[1] it was shown that when a is sufficiently small so that terms of higher power than the first are negligible, $-ma$ is the angular momentum of a slowly rotating mass shell. It was also pointed out that a slowly rotating mass shell is not the only source for the Kerr metric to first order in a; other sources are, e.g., a slowly rotating solid sphere of perfect fluid, many concentric shells, etc. The purpose of this paper is to find the expression for the *angular momentum generated by any body which has the Kerr metric exterior to it.* This is accomplished by integrating the conservation law over all space–time and applying a generalized form to the divergence theorem to this integral.

II. CONSERVATION LAWS AND CONSERVED QUANTITIES

Some of the results of this section are well known, but to facilitate a comparison between the results of Trautman and those of Komar, we give a short review here. The conservation law of general relativity

$$T^{\mu\nu}{}_{;\nu} = 0 \tag{1}$$

yields conserved quantities in spaces with symmetries. The isometry group associated with these symmetries is generated by a Killing vector ξ^μ which satisfies the equation $\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0$. Contraction of Eq. (1) with the Killing vector and integration over all space–time σ yields[2]

$$0 = \int_\sigma (\xi_\mu T^{\mu\nu}{}_{;\nu})\, dV_4. \tag{2}$$

By using the n-dimensional form of the divergence theorem,[3] we can transform Eq. (2) into an integral over the boundary of σ,

$$0 = \int_{\partial\sigma} \xi_\mu T^{\mu\nu}\, d\sigma_\nu. \tag{3}$$

If the source is bounded in space or falls off sufficiently rapidly at spatial infinity, the integral (3) reduces to the difference of the value of a integral over two different spacelike surfaces. Thus the integral is independent of the spacelike surface and consequently is a conserved quantity[4]:

$$J = \int_\Sigma \xi_\mu T^{\mu\nu}\, d\sigma_\nu. \tag{4}$$

Here Σ denotes a three-dimensional spacelike surface and $d\sigma_\nu$ its surface element. Since the stress-energy tensor is related to the Einstein tensor via Einstein's equations, the conserved quantity J in Eq. (4) can be expressed in terms of geometrical quantities[5]:

$$8\pi J = \int_\Sigma \xi_\mu G^{\mu\nu}\, d\sigma_\nu. \tag{5}$$

If the Killing vector ξ^μ is tangent to the spacelike surface Σ, then the Einstein tensor $G^{\mu\nu}$ in Eq. (5) can be expressed in terms of the second fundamental form[6] P_{ij} of the spacelike surface. When this is done, the integrand of Eq. (5) can be expressed as a divergence since ξ^μ is a Killing vector. Finally, application of the divergence theorem[3] yields

$$8\pi J = \int_{\partial\Sigma} \xi_\mu (P^{\mu\nu} - \eta^{\mu\nu} P)\, d\sigma^2_\nu, \tag{6}$$

where $\partial\Sigma$ denotes the two-dimensional boundary of Σ and $d\sigma^2_\nu$ is an area element of $\partial\Sigma$.

---

[1] J. M. Cohen, J. Math. Phys. **8**, 1477 (1967).

[2] The integrand is transformable into a divergence since ξ is a Killing vector, i.e., $\xi_\mu T^{\mu\nu}{}_{;\nu} = (\xi_\mu T^{\mu\nu})_{;\nu} - \xi_{\mu;\nu} T^{\mu\nu} = (\xi_\nu T^{\mu\nu})_{;\nu}$.

[3] A convenient form of this divergence theorem is

$$\int_\sigma b^\mu{}_{;\mu}\, dV = \int_{\partial\sigma} b^\mu\, d\sigma^\mu,$$

where σ denotes an n-dimensional surface with surface elements dV and $\partial\sigma$ denotes its boundary with surface element $d\sigma_\mu$. For a derivation of this formula see, e.g., J. L. Synge, *Relativity, the General Theory* (North-Holland Publishing Company, Amsterdam, 1960).

[4] This is a well-known result which can be found, e.g., in A. Trautman in *Gravitation*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1963), Eq. (5–38).

[5] This conserved quantity (5) differs in general from that of Komar [A. Komar, Phys. Rev. **127**, 1411 (1962)], which takes the form $8\pi J = \int_\Sigma \xi_\mu R^{\mu\nu}\, d\sigma_\nu$. (Komar's conserved quantity is obtained by applying the procedure of Sec. II to $d^2 * d\xi = 0$.) The two expressions agree if R vanishes or if $\eta_{\mu\nu}$ vanishes for the particular components fixed via ξ and Σ.

[6] Y. Fourès-Bruhat in Ref. 4.

III. APPLICATION TO THE KERR METRIC

If the Killing vector ξ_μ generates a rotation, then the conserved quantity J is an angular momentum.[1] Using Eq. (6), we can find the angular momentum of any body which generates the Kerr metric[7] by considering only the asymptotic metric at large distances from the source and the Killing vector $\partial/\partial\varphi$. By integrating Eq. (6), we find that *any source which generates the Kerr metric has the angular momentum*[8]

$$J = -ma. \tag{7}$$

This result (7) is *valid for large a as well as large m.*[9-11]

ACKNOWLEDGMENTS

For helpful discussions, I am indebted to P. B. Bergmann, D. R. Brill, W. J. Cocke, J. B. Harlte, and J. Winiscour.

This work was supported in part by the U.S. Atomic Energy Commission and the National Aeronautics and Space Administration and was completed while the author held a NAS-NRC Research Associateship at the Institute for Space Studies, Goddard Space Flight Center, NASA.

[7] R. Kerr, Phys. Rev. Letters **11**, 525 (1965).

[8] The minus sign was first noticed by R. H. Boyer and T. G. Price, [Proc. Cambridge Phil. Soc. **61**, 531 (1965)] who considered the case of both m and a small.

[9] In terms of an integral over $\partial\Sigma$, Komar's conserved quantity takes the coordinate-free form

$$16\pi J = \int_{\partial\Sigma} {}^*\, d\xi.$$

Application of this formula to the Kerr metric also yields the result (7).

[10] For a discussion of a application of Komar's formula to an axially symmetric metric see R. A. Matzner and C. W. Misner, Phys. Rev. **154**, 1229 (1967). Unfortunately their results are difficult to apply to the Kerr metric since their form of the metric becomes very complicated for Kerr's solution.

[11] By a coordinate transformation similar to that of Eq. (9) of Ref. 1, the charged metric of E. T. Newman, E. Couch, K. Chinnapared, A. Exton, P. Prakash, R. Torrence [J. Math. Phys. **6**, 918 (1965)] can be put in the form

$$ds^2 = -dt^2 + f\Sigma^{-1}(dt + a\sin^2\theta\, d\varphi^2 + \Sigma(\Delta^{-1}\, dR^2 + d\theta^2) + (R^2 + a^2)\sin^2\theta\, d\varphi^2,$$

where

$$f = 2mr - e^2,$$
$$\Sigma = R^2 + a^2\cos^2\theta,$$
$$\Delta = R^2 + a^2 - f.$$

Substitution of this metric into Eq. (16) yields the angular momentum $J = -ma$, the same as that for the Kerr metric. Also, as for the Kerr metric, this result is valid for large a as well as large m. The above form of the rotating charged metric was found independently by Brandon Carter (private communication).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

New Derivation of the Integro-Differential Equations for Chandrasekhar's X and Y Functions*

R. BELLMAN,
University of Southern California, Los Angeles, California

H. KAGIWADA AND R. KALABA
The Rand Corporation, Santa Monica, California

S. UENO
University of Kyoto, Japan

(Received 15 May 1968)

The X and Y functions of radiative transfer satisfy a system of integro-differential equations which form the basis of an effective numerical treatment. These integro-differential equations are derived from the integral equation for the source function and the differential equation for the resolvent.

I. INTRODUCTION

The X and Y functions are important functions in radiative-transfer theory.[1] Their properties are best determined analytically through their nonlinear integral equations. From the computational view, though, they are best treated as solutions of a system of integro-differential equations with initial conditions. The aim of this note is to derive the integro-differential equations from the integral equation for the source function, making use of properties of the resolvent.

This note is self-contained.

* This research is supported and monitored by the Advanced Research Projects Agency under Contract No. SD-79. Any views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of ARPA.

[1] I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge University Press, London, 1960).

III. APPLICATION TO THE KERR METRIC

If the Killing vector ξ_μ generates a rotation, then the conserved quantity J is an angular momentum.[1] Using Eq. (6), we can find the angular momentum of any body which generates the Kerr metric[7] by considering only the asymptotic metric at large distances from the source and the Killing vector $\partial/\partial\varphi$. By integrating Eq. (6), we find that *any source which generates the Kerr metric has the angular momentum*[8]

$$J = -ma. \tag{7}$$

This result (7) is *valid for large a as well as large m.*[9-11]

ACKNOWLEDGMENTS

For helpful discussions, I am indebted to P. B. Bergmann, D. R. Brill, W. J. Cocke, J. B. Harlte, and J. Winiscour.

This work was supported in part by the U.S. Atomic Energy Commission and the National Aeronautics and Space Administration and was completed while the author held a NAS-NRC Research Associateship at the Institute for Space Studies, Goddard Space Flight Center, NASA.

[7] R. Kerr, Phys. Rev. Letters **11**, 525 (1965).

[8] The minus sign was first noticed by R. H. Boyer and T. G. Price, [Proc. Cambridge Phil. Soc. **61**, 531 (1965)] who considered the case of both m and a small.

[9] In terms of an integral over $\partial\Sigma$, Komar's conserved quantity takes the coordinate-free form

$$16\pi J = \int_{\partial\Sigma} {}^*\, d\xi.$$

Application of this formula to the Kerr metric also yields the result (7).

[10] For a discussion of a application of Komar's formula to an axially symmetric metric see R. A. Matzner and C. W. Misner, Phys. Rev. **154**, 1229 (1967). Unfortunately their results are difficult to apply to the Kerr metric since their form of the metric becomes very complicated for Kerr's solution.

[11] By a coordinate transformation similar to that of Eq. (9) of Ref. 1, the charged metric of E. T. Newman, E. Couch, K. Chinnapared, A. Exton, P. Prakash, R. Torrence [J. Math. Phys. **6**, 918 (1965)] can be put in the form

$$ds^2 = -dt^2 + f\Sigma^{-1}(dt + a\sin^2\theta\, d\varphi^2 + \Sigma(\Delta^{-1}\, dR^2 + d\theta^2) + (R^2 + a^2)\sin^2\theta\, d\varphi^2,$$

where

$$f = 2mr - e^2,$$
$$\Sigma = R^2 + a^2\cos^2\theta,$$
$$\Delta = R^2 + a^2 - f.$$

Substitution of this metric into Eq. (16) yields the angular momentum $J = -ma$, the same as that for the Kerr metric. Also, as for the Kerr metric, this result is valid for large a as well as large m. The above form of the rotating charged metric was found independently by Brandon Carter (private communication).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

New Derivation of the Integro-Differential Equations for Chandrasekhar's X and Y Functions*

R. BELLMAN,
University of Southern California, Los Angeles, California

H. KAGIWADA AND R. KALABA
The Rand Corporation, Santa Monica, California

S. UENO
University of Kyoto, Japan

(Received 15 May 1968)

The X and Y functions of radiative transfer satisfy a system of integro-differential equations which form the basis of an effective numerical treatment. These integro-differential equations are derived from the integral equation for the source function and the differential equation for the resolvent.

I. INTRODUCTION

The X and Y functions are important functions in radiative-transfer theory.[1] Their properties are best determined analytically through their nonlinear integral equations. From the computational view, though, they are best treated as solutions of a system of integro-differential equations with initial conditions. The aim of this note is to derive the integro-differential equations from the integral equation for the source function, making use of properties of the resolvent.

This note is self-contained.

* This research is supported and monitored by the Advanced Research Projects Agency under Contract No. SD-79. Any views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of ARPA.

[1] I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge University Press, London, 1960).

II. DEFINITIONS

The X and Y functions of radiative transfer[1] are defined by the relations:

$$X(x, \mu) = (4/\lambda)J(x, x, \mu), \tag{1}$$

$$Y(x, \mu) = (4/\lambda)J(0, x, \mu), \tag{2}$$

where the source function $J = J(t, x, \mu)$ is defined as the solution of the Fredholm integral equation

$$J(t, x, \mu) = (\lambda/4)\exp[-(x - t)/\mu] + (\lambda/2)\int_0^x E_1(|t - y|)J(y, x, \mu)\,dy, \quad 0 \le t \le x, \quad 0 \le \mu \le 1. \tag{3}$$

As usual, E_1 is the first exponential integral function given by

$$E_1(r) = \int_0^1 e^{-r/z}\,dz/z, \quad r > 0. \tag{4}$$

Our aim is to provide a straightforward derivation of the differential-integral equations

$$X_x(x, \mu) = (\lambda/2)Y(x, \mu)\int_0^1 Y(x, z)\,dz/z, \tag{5}$$

$$Y_x(x, \mu) = -\mu^{-1}Y(x, \mu) + (\lambda/2)X(x, \mu)\int_0^1 Y(x, z)\,dz/z, \quad x > 0. \tag{6}$$

and initial conditions:

$$X(0, \mu) = 1, \tag{7}$$

$$Y(0, \mu) = 1, \quad \mu > 0, \tag{8}$$

which play such an important role in the numerical treatment of these functions.[2]

III. DERIVATION

Consider the general integral equation

$$w(t, x) = g(t) + (\lambda/2)\int_0^x E_1(|t - y|)w(y, x)\,dy. \tag{9}$$

In terms of the Fredholm resolvent K, the solution may be represented in the form

$$w(t, x) = g(t) + \int_0^x K(t, y, x)g(y)\,dy. \tag{10}$$

First, several important relations for the resolvent will be obtained. Upon differentiation with respect to x, Eq. (9) becomes

$$w_x(t, x) = (\lambda/2)E_1(x - t)w(x, x) + (\lambda/2)\int_0^x E_1(|t - y|)w_x(y, x)\,dy. \tag{11}$$

Upon introduction of the function Φ, obtained as the solution of the integral equation

$$\Phi(t, x) = (\lambda/2)E_1(x - t) + (\lambda/2)\int_0^x E_1(|t - y|)\Phi(y, x)\,dy, \tag{12}$$

it is seen that

$$w_x(t, x) = \Phi(t, x)w(x, x). \tag{13}$$

From Eq. (10) it follows that

$$w(x, x) = g(x) + \int_0^x K(x, y, x)g(y)\,dy; \tag{14}$$

Eq. (13) becomes

$$w_x(t, x) = \Phi(t, x)\left[g(x) + \int_0^x K(x, y, x)g(y)\,dy\right]. \tag{15}$$

On the other hand, Eq (10) may be differentiated with respect to x to obtain

$$w_x(t, x) = K(t, x, x)g(x) + \int_0^x K_x(t, y, x)g(y)\,dy. \tag{16}$$

From this it follows that

$$\Phi(t, x) = K(t, x, x), \tag{17}$$

$$K_x(t, y, x) = \Phi(t, x)K(x, y, x), \tag{18}$$

or

$$K_x(t, y, x) = K(t, x, x)K(x, y, x). \tag{19}$$

Equations (17) and (19) are the desired relations. A numerical procedure for determining the resolvent K using (19) is given in Ref. 3. Equation (19) has been obtained earlier by Bellman, Ref. 4, and Krein, Ref. 5.

Keeping in mind Eqs. (3) and (10), it is seen that

$$J(t, x, \mu) = (\lambda/4)\exp[-(x - t)/\mu] + \int_0^x K(t, y, x)(\lambda/4)\exp[-(x - y)/\mu]\,dy. \tag{20}$$

Upon differentiation with respect to x, this equation becomes

$$J_x(t, x, \mu) = -\mu^{-1}(\lambda/4)\exp[-(x - t)/\mu] + (\lambda/4)K(t, x, x) + (\lambda/4)\int_0^x K_x(t, y, x)\exp[-(x - y)/\mu]\,dy - (\lambda/4)\mu^{-1}\int_0^x K(t, y, x)\exp[-(x - y)/\mu]\,dy. \tag{21}$$

Upon referring to Eq. (19) for K_x, this becomes

$$J_x(t, x, \mu) = -\mu^{-1}(\lambda/4)\exp[-(x - t)/\mu] + (\lambda/4)K(t, x, x) + (\lambda/4)K(t, x, x)\int_0^x K(x, y, x)\exp[-(x - y)/\mu]\,dy - (\lambda/4)\mu^{-1}\int_0^x K(t, y, x)\exp[-(x - y)/\mu]\,dy. \tag{22}$$

[1] R. Bellman, H. Kagiwada, R. Kalaba, and S. Ueno, J. Quant. Spectry. & Radiative Transfer **6**, 479 (1966).

[2] H. Kagiwada, and R. Kalaba, J. Math. Phys. Sci. **1**, 109 (1967).

[3] R. E. Bellman, Proc. Am. Math. Soc. **8**, 435 (1957).

[4] M. G. Krein, "On a New Method of Solving Linear Integral Equations of the First and Second Kind," Dokl. Akad. Nauk. SSSR **100**, 413 (1955).

It is then seen that

$$J_x(t, x, u) = -u^{-1}J(t, x, \mu) + K(t, x, x)\Big\{(\lambda/4) + (\lambda/4)\int_0^x K(x, y, x) \times \exp[-(x - y)/\mu]\, dy\Big\}, \quad (23)$$

$$J_x(t, x, u) = -\mu^{-1}J(t, x, \mu) + K(t, x, x)J(x, x, \mu). \quad (24)$$

This equation can be utilized to compute J as in Ref. 6. In particular for $t = 0$,

$$J_x(0, x, \mu) = -\mu^{-1}J(0, x, \mu) + \Phi(0, x)J(x, x, \mu). \quad (25)$$

Note, though, that

$$\Phi(t, x) = 2\int_0^1 J(t, x, z)\, dz/z, \quad (26)$$

which follows from Eqs. (3), (4), and (12), by using the principle of superposition. In view of the definitions of X and Y in Eqs. (1) and (2), the last two relations imply that the basic equation (6) is true.

To establish Eq. (5), Eq. (3) is rewritten in the form

$$\begin{aligned} J(x - t, x, \mu) &= (\lambda/4)\exp(-t/\mu) + (\lambda/2)\int_0^x E_1(|x - t - y|)J(y, x, \mu)\, dy \\ &= (\lambda/4)\exp(-t/\mu) + (\lambda/2)\int_0^x E_1(|y - t|)J(x - y, x, \mu)\, dy. \end{aligned} \quad (27)$$

Differentiation with respect to x yields

$$\begin{aligned} &(d/dx)J(x - t, x, \mu) \\ &= (\lambda/2)E_1(|x - t|)J(0, x, \mu) + (\lambda/2)\int_0^x E_1(|y - t|)[d/dx)J(x - y, x, \mu)]\, dy. \end{aligned} \quad (28)$$

According to Eq. (12) for the function Φ, the solution of the above equation is

$$(d/dx)J(x - t, x, \mu) = J(0, x, \mu)\Phi(t, x). \quad (29)$$

In particular for $t = 0$, there results

$$(d/dx)J(x, x, \mu) = J(0, x, \mu)\Phi(0, x). \quad (30)$$

In view of Eqs. (1), (2), and (26), the last relation implies the desired differential-integral equation (5).

The initial conditions in Eqs. (7) and (8) follow directly from the definitions of X and Y and the integral equation (3).

[6] R. Bellman, H. Kagiwada, and R. Kalaba, J. Quant. Spectry. & Radiative Transfer 6, 291 (1966).

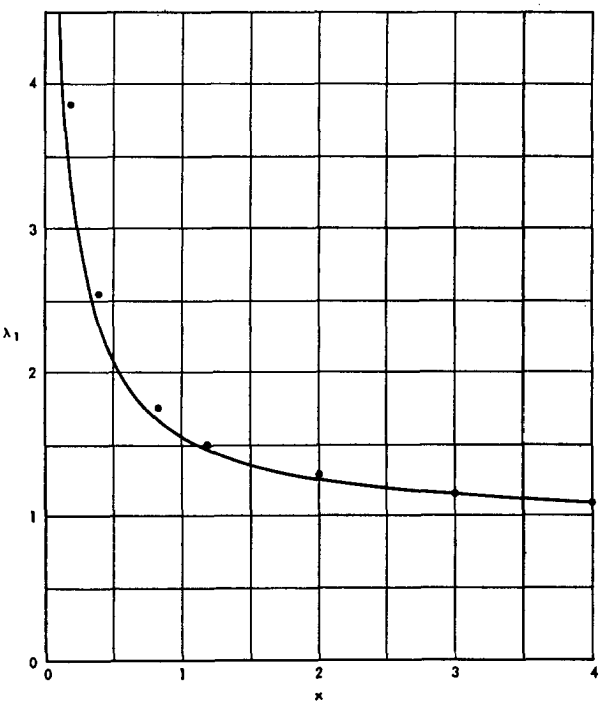


FIG. 1. Approximate curve of lowest eigenvalue as a function of length for the kernel $\frac{1}{2}E_1(|t - y|)$.

IV. DISCUSSION

As noted above, the differential-integral equations (5) and (6) together with the initial conditions of Eqs. (7) and (8) form the basis of an effective computational procedure for X and Y.[2] The integrals are approximated by sums and the resultant system of ordinary differential equations is integrated numerically from $x = 0$ to $x =$ the maximum desired thickness. For a given value of λ and for a set of μ, the X and Y functions are produced for all thicknesses between 0 and the maximum thickness. This procedure is useful for $0 < \lambda \leq 1$.

For $\lambda > 1$, the phenomenon of criticality occurs. In the course of the integration, the X and Y functions become unbounded. The value of x when X and Y become very large, say $\sim 10^3$, is a very close lower bound on the critical length for that value of λ. Such a survey of critical lengths for various $\lambda > 1$ leads to an approximate curve of the lowest eigenvalue of Eq. (3) as a function of x. Figure 1 shows such a curve and the comparison shows points plotted from a table of Wing.[7]

These results may be extended to other kernels of displacement type.[8]

[7] G. Milton Wing, J. Math. Anal. Appl. **11**, 160 (1965).

[8] H. Kagiwada and R. Kalaba, Rand Corp., Santa Monica, Calif. Report RM-5186-PR, 1966 (unpublished); Intern. J. Comp. Math. (to appear).

Diffuse Transmission of Light from a Central Source through an Inhomogeneous Spherical Shell with Isotropic Scattering

R. E. BELLMAN, H. H. KAGIWADA, AND R. E. KALABA
Rand Corporation, Santa Monica, California

AND

S. UENO
University of Kyoto, Kyoto, Japan

(Received 18 October 1966)

A partial-differential-integral equation is derived in this paper for the angular distribution of the radiation which is diffusely transmitted through an inhomogeneous, isotropically scattering, spherical shell when there is a constant net flux of radiation normally incident on the inner surface. An equation is also derived for the strength of the diffusely reflected radiation when the shell is illuminated at each point on the outer surface by constant isotropic incident radiation.

The equations obtained appear to lend themselves well to numerical solution. Astrophysically, the situation corresponds to determining the brightness of a spherical planetary nebula. As far as is known, the equations are new and exact.

I. INTRODUCTION

In recent years the scattering of light in a finite slab has been studied extensively, both analytically[1-3] and computationally.[4,5] However, it is important to consider the effects of curvature in various problems of radiative transfer in planetary and stellar atmospheres, nova envelopes, gaseous nebulas, and of neutron transport in spherical reactors. This has led to the study of radiative transfer in spherical media, an investigation which is in its infancy from the analytical and computational viewpoints.[6-19] It is the aim of this paper to derive an equation for the diffuse transmission coefficient of a shell with an absorbing core, radiation being normally incident on the inner surface of the shell. Within the shell, absorption and isotropic scattering take place.

Our plan is first to derive an equation for the intensity of the radiation which is diffusely reflected from a spherical-shell atmosphere which is illuminated isotropically at each point on its surface. Within the shell, multiple scattering and absorption processes take place. Then having derived this equation for the reflection function we can proceed to discuss the desired transmission process.

In the astrophysical context, this problem corresponds to the radiative equilibrium of a spherical planetary nebula in the region of UV radiation, whose boundary conditions, due to Milne, are as follows: There is no incident radiation on the outer boundary and the diffuse flux across the inner surface vanishes.

II. A REFLECTION FUNCTION

Consider a spherical shell with an inner radius of x_0 and an outer radius of x. These are geometrical and not optical distances. The shell is composed of an inhomogeneous material which absorbs radiation and scatters it isotropically. Assume that when energy passes through a small geometric distance Δ at a distance y from the center, the fraction $a(y)\Delta + o(\Delta)$ is absorbed. The fraction $\lambda(y)$ is reradiated isotropically. For simplicity, we shall assume that $a(y) \equiv 1$, though in the discussion in Sec. V we present the equations for the general case. We suppose that the diffuse flux across the surface of the core is zero (Milne's boundary condition).

The shell is illuminated on the outer boundary with one unit of energy per unit area per unit of time. At

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).
[2] I. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge University Press, London, 1960).
[3] V. V. Sobolev, *A Treatise on Radiative Transfer* (D. Van Nostrand Company, Inc., Princeton, N. J., 1963).
[4] R. Bellman, R. Kalaba, and M. Prestrud, *Invariant Imbedding and Radiative Transfer in Slabs of Finite Thickness* (American Elsevier Pub. Co., New York, 1963).
[5] R. Bellman, H. Kagiwada, R. Kalaba, and M. Prestrud, *Invariant Imbedding and Time-Dependent Transport Processes* (American Elsevier Publishing Company, New York, 1964).
[6] P. B. Bailey, J. Math. Anal. Appl., **8**, 144 (1964).
[7] P. B. Bailey and G. M. Wing, J. Math. Anal. Appl. **8**, 170 (1964).
[8] D. Barbier and G. W. Curtis, Ann. Astrophys. **19**, 129 (1956).
[9] V. I. Barkov, Opt. Spektrosk. **14**, 537 (1963) [Opt. Spectrosk. **14**, 285 (1963)].
[10] R. Bellman and R. Kalaba, Proc. Natl. Acad. Sci. U.S. **43**, 514 1957.
[11] M. A. Heaslet and R. F. Warming, J. Quant. Spectry. Radiative Transfer **5**, 669 (1965).
[12] J. Lenoble and Z. Sekera, Proc. Natl. Acad. Sci. U.S. **47**, 372 (1961).
[13] I. N. Minin, Astron. Zh. **41**, 662 (1964) [Sov. Astron.—AJ **8**, 528 (1965)].
[14] I. N. Minin and V. V. Sobolev, Astron. Zh. **40**, 496 (1963) [Sov. Astron.—AJ **7**, 379 (1963)].
[15] V. V. Sobolev and I. N. Minin, Iskusstv. Sputniki Zemli **14**, 7 (1962) [Planet. Space Sci. **11**, 657 (1963)].
[16] I. N. Minin and V. V. Sobolev, Kosmich. Issled. **1**, 227 (1963) [Cosmic Res. **1**, 190 (1963)].
[17] I. N. Minin and V. V. Sobolev, Kosmich. Issled. **2**, 610 (1964) [Cosmic Res. **2**, 529 (1964)].
[18] V. V. Sobolev, Astron. Zh. **37**, 3 (1960) [Sov. Astron.—AJ **4**, 1 (1960)].
[19] H. K. Sen, Astrophys. J. **110**, 276 (1949).

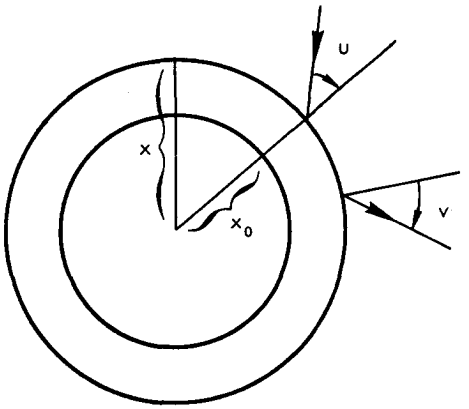


FIG. 1. The physical situation.

each point on the outer boundary the incident radiation is conical in form, and the direction cosine of an incident ray with respect to the local inward normal is u (see Fig. 1). Consider the reflection coefficient $\rho(v, u, x)$:

$$\rho(v, u, x)dv = \text{total reflected energy emanating from the outer surface } y = x \text{ per unit of area on the outer surface per unit of time, the direction cosine being between } v \text{ and } v + dv \text{ with the local outward-drawn normal and the incident radiation as described above.} \quad (1)$$

In the usual manner of invariant imbedding, the reflection coefficients for shells of outer radius $x + \Delta$ and x can now be related.

If radiation makes an angle θ with the local normal at a distance x, it makes an angle $\theta - \alpha$ with the local normal at $x + \Delta$ (see Fig. 2). Here

$$\cos(\theta - \alpha) - \cos\theta = (-\alpha)(-\sin\theta) + \cdots, \quad (2)$$

where powers of α greater than the first are neglected. In addition

$$w = \alpha x \quad (3)$$

and

$$w(1 + \Delta/x) = \Delta \tan\theta + o(\Delta), \quad (4)$$

so that

$$\alpha = (\Delta/x)\tan\theta + o(\Delta). \quad (5)$$

In view of Eq. (2), this yields

$$\cos(\theta - \alpha) - \cos\theta = (\Delta/x)\sin^2\theta/\cos\theta + o(\Delta). \quad (6)$$

In addition, from Eq. (6) it is noted that if radiation falls in the direction-cosine interval $(v, v + dv)$ at the radius x, at the radius $x + \Delta$ it will fall in the interval

$$(v + [(1 - v^2)/vx]\Delta, v + dv + \{[1 - (v + dv)^2]/x(v + dv)\}\Delta).$$

The length of this cosine interval dv^* is

$$dv^* = dv\left[1 + \frac{d}{dv}\left(\frac{1 - v^2}{v}\right)\frac{\Delta}{x}\right] + o(\Delta)$$

$$= dv\left[1 + \frac{v(-2v) - (1 - v^2)}{v^2}\frac{\Delta}{x}\right], \quad (7)$$

or

$$dv^* = dv\left[1 - \frac{1 + v^2}{v^2}\frac{\Delta}{x}\right], \quad (8)$$

where terms involving powers of Δ higher than the first are omitted.

Now consider the uniform conical radiation incident on the shell of outer radius $x + \Delta$. Write

$$\left(\frac{x}{x + \Delta}\right)^2 \rho\left(v + \frac{1 - v^2}{xv}\Delta, u + \frac{1 - u^2}{xu}\Delta, x + \Delta\right) dv^*$$

$$= \left(1 - \frac{\Delta}{u}\right)\rho(v, u, x)\, dv\left(1 - \frac{\Delta}{v}\right)\left(\frac{x}{x + \Delta}\right)^2$$

$$+ \left[\frac{\Delta}{u} + \int_0^1 \rho(v', u, x)\frac{\Delta}{v'}\, dv'\right]$$

$$\times \lambda(x)\left[\frac{dv}{2} + \int_0^1 \rho(v, u', x)\frac{du'}{2}\, dv\right], \quad (9)$$

which is correct to terms involving Δ^2. The radiation incident on the surface $y = x + \Delta$ is considered to have a direction cosine such that the ray will have direction cosine v when incident on the shell of outer radius x. This is also done for the emerging radiation.

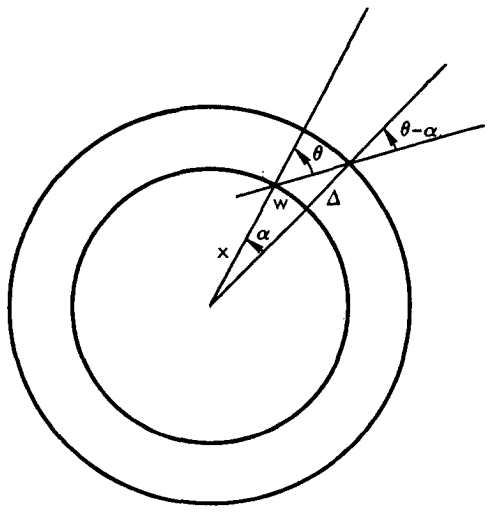


FIG. 2. Changes in angles.

The energy per unit area on $y = x + \Delta$ per unit time is taken to be $[x/(x + \Delta)]^2$ rather than unity, so that the energy per unit area on $y = x$ per unit time will be unity. In addition, the interval of direction cosines is selected to be dv^* for the surface $y = x + \Delta$, the corresponding interval for $y = x$ then being dv.

Now consider the terms on the right-hand side. The first term accounts for the energy emerging through the surface $y = x + \Delta$ where the diminutions due to absorption in the shell from $y = x$ to $y = x + \Delta$ and the differing surface areas have been taken into account. The last term is the product of three factors. The first factor is the rate of production of absorbed radiation in a partial shell of unit area on $y = x$ and extending to $y = x + \Delta$. The fraction of the energy reemitted is $\lambda(x)$, and the probability that such energy will ultimately emerge from the top surface with direction cosine in the interval $(v, v + dv)$ is given by the last factor.

Observe that the equation can be multiplied by $((x + \Delta)/x)^2$, and the factors due to geometric convergence will disappear. (The last term has Δ as a factor.) The desired relation is found by expanding in powers of Δ and equating the coefficients of Δ,

$$\rho_x + \frac{1-u^2}{xu}\rho_u + \frac{1-v^2}{xv}\rho_v - \frac{1+v^2}{xv^2}\rho = -\left(\frac{1}{u}+\frac{1}{v}\right)\rho + \frac{\lambda}{2}\left[\frac{1}{u} + \int_0^1 \rho(v', u, x)\frac{dv'}{v'}\right] \times \left[1 + \int_0^1 \rho(v, u', x)\,du'\right], \quad x \geq x_0. \tag{10}$$

As an initial condition

$$\rho(v, u, x_0) = 0, \tag{11}$$

because there is no scattering medium in the spherical shell of thickness zero.

III. A TRANSMISSION FUNCTION

This section considers the desired transmission process. Let unit energy per unit area per unit time be normally incident on the inner surface $y = x_0$ of the shell. It is desired here to determine the angular dependence of the diffusely transmitted radiation. Let

$t(v, x)dv$ = energy diffusely transmitted per unit area on the outer surface $y = x$ per unit time, the direction cosine being between v and $v + dv$, the incident radiation as described above, and the shell and the core as described earlier. (12)

This function satisfies the relation

$$t\left(v + \frac{1-v^2}{xv}\Delta, x + \Delta\right) dv^* = t(v, x)\,dv(1 - \Delta/v)(x/(x+\Delta))^2 + \left[\left(\frac{x_0}{x}\right)^2 e^{-(x-x_0)} + \int_0^1 t(v', x)\frac{dv'}{v'}\right]\Delta\lambda(x) \times \left[\frac{dv}{2} + \int_0^1 \rho(v, u', x)\frac{du'}{2}\,dv\right], \tag{13}$$

correct to terms involving Δ^2, where dv^* and $\rho(v, u, x)$ retain their earlier meanings. This leads to the equation

$$t_x + \frac{1-v^2}{xv}t_v - \frac{1+v^2}{xv^2}t = -\frac{1}{v}t - \frac{2}{x}t + \left[\left(\frac{x_0}{x}\right)^2 e^{-(x-x_0)} + \int_0^1 t(v', x)\frac{dv'}{v'}\right]\lambda(x) \times \left[\frac{1}{2} + \int_0^1 \rho(v, u', x)\frac{du'}{2}\right], \tag{14}$$

or

$$t_x + \frac{1-v^2}{xv}t_v - \frac{1-v^2}{xv^2}t = -\frac{1}{v}t + \left[\left(\frac{x_0}{x}\right)^2 e^{-(x-x_0)} + \int_0^1 t(v', x)\frac{dv'}{v'}\right] \times \frac{\lambda(x)}{2}\left[1 + \int_0^1 \rho(v, u', x)\,du'\right], \tag{15}$$

for an initial condition

$$t(v, x_0) = 0. \tag{16}$$

IV. NORMALIZATION OF EQUATION

It is convenient to introduce certain normalizations. For the reflection problem involving conical inputs on the outer surface of the shell, let the net flux incident be π. Then the intensity of the diffusely reflected radiation $r(v, u, x)$ is

$$r(v, u, x) = \pi\rho(v, u, x)(2\pi)^{-1}v^{-1}u. \tag{17}$$

Next, the function R is introduced,

$$r(v, u, x) = R(v, u, x)(4v)^{-1}. \tag{18}$$

This implies

$$\rho(v, u, x) = R(v, u, x)/(2u). \tag{19}$$

The equation for R is

$$R_x + \frac{1-v^2}{xv}R_v + \frac{1-u^2}{xu}R_u + \left(\frac{1}{u}+\frac{1}{v}\right)R - \frac{u^2+v^2}{xu^2v^2}R = \lambda(x)\left[1 + \frac{1}{2}\int_0^1 R(v', u, x)\,dv'/v'\right] \times \left[1 + \frac{1}{2}\int_0^1 R(v, u', x)\,du'/u'\right], \quad x > x_0, \tag{20}$$

and the initial condition is

$$R(v, u, x_0) = 0. \tag{21}$$

For the diffuse transmission process, the net incident flux is also normalized to be π rather than unity, and the intensity of the diffusely transmitted radiation is denoted by $T(v, x)/(4v)$. This means

$$\frac{T(v, x)}{4v} = \pi t(v, x) \frac{1}{2\pi v} \tag{22}$$

or

$$T(v, x) = 2t(v, x). \tag{23}$$

The equation for T is

$$T_x + \frac{1 - v^2}{xv} T_v - \frac{1 - v^2}{xv^2} T$$
$$= -\frac{1}{v} T + \left[\left(\frac{x_0}{x}\right)^2 e^{-(x-x_0)} + \frac{1}{2}\int_0^1 T(v', x) \frac{dv'}{v'}\right]$$
$$\times \lambda(x)\left[1 + \frac{1}{2}\int_0^1 R(v, u', x) \frac{du'}{u'}\right], \quad x > x_0, \tag{24}$$

along with the initial condition

$$T(v, x_0) = 0. \tag{25}$$

V. DISCUSSION

The equations for R and T would seem to be quite useful for computational purposes. Some results for the reflection coefficient are available in Refs. 20 and 21, and the authors plan to undertake numerical experiments on the system for R and T.

These considerations readily generalize to the case of anisotropic scattering, though, of course, the computational load is vastly increased.

For the case in which $a = a(y)$, the equations for R and T become

$$R_x + \frac{1 - v^2}{xv} R_v + \frac{1 - u^2}{xu} R_u$$
$$+ a(x)\left(\frac{1}{u} + \frac{1}{v}\right) R - \frac{u^2 + v^2}{xu^2v^2} R$$
$$= a(x)\lambda(x)\left[1 + \frac{1}{2}\int_0^1 R(v', u, x)\, dv'/v'\right]$$
$$\times \left[1 + \frac{1}{2}\int_0^1 R(v, u', x)\, du'/u'\right] \tag{26}$$

and

$$T_x + \frac{1 - v^2}{xv} T_v - \frac{1 - v^2}{xv^2} T$$
$$= -\frac{a(x)}{v} T + \left[\left(\frac{x_0}{x}\right)^2 \exp\left(-\int_{x_0}^x a(y)\, dy\right)\right.$$
$$\left. + \frac{1}{2}\int_0^1 T(v', x) \frac{dv'}{v'}\right]$$
$$\times a(x)\lambda(x)\left[1 + \frac{1}{2}\int_0^1 R(v, u', x) \frac{du'}{u'}\right],$$
$$x > x_0. \tag{27}$$

ACKNOWLEDGMENT

The authors are indebted to Paul Bailey of Sandia Corporation for a careful reading of the manuscript.

[20] R. Bellman and R. Kalaba, Proc. Natl. Acad. Sci. U.S. **54**, 1293 (1965).

[21] R. Bellman, H. Kagiwada, and R. Kalaba, J. Computational Phys. **1**, 245 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Electromagnetic Two-Body Problem for Particles with Spin

A. SCHILD
The University of Texas, Austin, Texas
AND
J. A. SCHLOSSER
Louisiana State University, Baton Rouge, Louisiana

(Received 8 October 1967)

A two-body system is considered for classical particles with charge, spin, and magnetic moment. The particles move in circular orbits, with spins orthogonal to the orbital plane, and they interact through time-symmetric electromagnetic fields. Rigorous relativistic equations of motion and rigorous expressions for the total energy and angular momentum of the system, including contributions from the field, are obtained.

1. INTRODUCTION

The first paper[1] of this series outlined a program for the study of the relativistic motion and the Bohr quantization of a system of two classical particles in electromagnetic interaction. The principal device in carrying out such a program consists in starting with a Fokker action principle. This results in time-symmetric (half-retarded plus half-advanced) interactions which permit periodic motions to which Bohr quantization can be applied. A Fokker action principle also gives, in a natural manner, finite expressions for the energy, linear momentum, and angular momentum of a system, which automatically include contributions of the electromagnetic field.

In the second paper[2] of this series a Fokker action principle was obtained for particles, each of which have charge, spin, and a magnetic moment proportional to the spin. In order to have a classical model of elementary particles with constant magnitude of the spin, a limiting procedure was adopted where the moment of inertia of each particle tends to zero while the spin remains finite. This required a renormalization of the mass. Renormalized equations of motion were obtained and renormalized expressions for the conserved quantities of the system.

In this paper the general results are applied to a two-body system with a geometry which is simple enough to permit the explicit calculation of the conserved quantities. The two particles move in concentric circles and have spins perpendicular to the plane of the orbits.

2. THE GENERAL TWO-BODY SYSTEM

The motion of a spinning particle is described by its world line $x^\mu(s)$ and by its spin $\sigma^{\mu\nu}(s) = -\sigma^{\nu\mu}$. Here s is the arc length given by

$$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu, \quad \eta_{\mu\nu} = \text{diag}\,(- - - +). \tag{2.1}$$

The spin is orthogonal to the world line,

$$\dot{x}^\mu \sigma_{\mu\nu} = 0, \tag{2.2}$$

where the dot denotes differentiation with respect to the arc length $\dot{x}^\mu = dx^\mu/ds$. The particle is characterized by its mass m, its charge e, a gyromagnetic ratio G which determines its magnetic moment $\mu^{\mu\nu} = G\sigma^{\mu\nu}$, and by the magnitude S of its spin, which is given by

$$S^2 = \tfrac{1}{2}\sigma_{\mu\nu}\sigma^{\mu\nu}. \tag{2.3}$$

It is a consequence of the equations of motion that S is a constant.

The two particles of the system will be distinguished by placing a bar over the quantities which describe the second particle: $\bar{x}^\mu(\bar{s})$, $\bar{\sigma}^{\mu\nu}(\bar{s})$, $\bar{m}$, $\bar{e}$, $\bar{G}$, $\bar{S}$, $\dot{\bar{x}}^\mu = d\bar{x}^\mu/d\bar{s}$, etc.

The time-symmetric electromagnetic field at x^μ due to the charge and magnetic moment of the particle $\bar{m}$ is given by

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu, \tag{2.4}$$

$$A^\mu = \bar{e}\int_{-\infty}^{\infty} \dot{\bar{x}}^\mu \delta\, d\bar{s} + \bar{G}\int_{-\infty}^{\infty} \bar{\sigma}^{\mu\nu}\partial_\nu \delta\, d\bar{s}. \tag{2.5}$$

Here δ is the Dirac delta function

$$\delta = \delta(\xi^2), \quad \xi^2 = \xi^\mu \xi_\mu, \quad \xi^\mu = x^\mu - \bar{x}^\mu, \tag{2.6}$$

$\partial_\mu = \partial/\partial x^\mu$ and, similarly, $\bar{\partial}_\mu = \partial/\partial \bar{x}^\mu$. At the position of particle m, the part of $F_{\mu\nu}$ which is orthogonal to the world line of m is

$$H_{\mu\nu} = F_{\mu\nu} - \dot{x}_\mu \dot{x}^\alpha F_{\alpha\nu} - \dot{x}_\nu \dot{x}^\alpha F_{\mu\alpha}, \quad \dot{x}^\mu H_{\mu\nu} = 0. \tag{2.7}$$

In the instantaneous rest frame of m, $H_{\mu\nu}$ describes the magnetic field.

[1] A. Schild, Phys. Rev. 131, 2762 (1963).
[2] A. Schild and J. A. Schlosser, J. Math. Phys. 6, 1299 (1965).

The orbital equation of motion of particle m is

$$m\ddot{x}_\mu = e\dot{x}^\nu F_{\nu\mu} - \tfrac{1}{2}G\sigma^{\alpha\beta}F_{\alpha\beta,\mu} + \frac{d}{ds}(\sigma_{\mu\nu}\ddot{x}^\nu + \tfrac{1}{2}G\sigma^{\alpha\beta}H_{\alpha\beta}\dot{x}_\mu + GF_{\alpha\beta}\dot{x}^\alpha\sigma^\beta{}_\mu), \quad (2.8)$$

and the equation of motion for its spin is

$$\dot{\sigma}_{\mu\nu} - \dot{x}_\mu\dot{x}^\alpha\sigma_{\alpha\nu} - \dot{x}_\nu\dot{x}^\alpha\sigma_{\mu\alpha} = 2GH_{\alpha[\mu}\sigma^\alpha{}_{\nu]}, \quad (2.9)$$

where the square brackets denote skew symmetrization, e.g., $A_{[\mu\nu]} = \frac{1}{2}(A_{\mu\nu} - A_{\nu\mu})$. The equations of motion have the first integrals $\dot{x}_\mu\dot{x}^\mu = 1$ and $S^2 =$ constant.

The equations of motion of particle $\bar{m}$ are obtained from the above by interchanging barred and unbarred quantities in Eqs. (2.4)–(2.9).

The total 4-momentum of the system is

$$\begin{aligned} P_\mu &= [(m - \tfrac{1}{2}G\sigma^{\alpha\beta}H_{\alpha\beta})\dot{x}_\mu - \sigma_{\mu\nu}\ddot{x}^\nu + eA_\mu - GF_{\alpha\beta}\dot{x}^\alpha\sigma^\beta{}_\mu]_s \\ &+ [(\bar{m} - \tfrac{1}{2}\bar{G}\bar{\sigma}^{\alpha\beta}\bar{H}_{\alpha\beta})\dot{\bar{x}}_\mu - \bar{\sigma}_{\mu\nu}\ddot{\bar{x}}^\nu + \bar{e}\bar{A}_\mu - \bar{G}\bar{F}_{\alpha\beta}\dot{\bar{x}}^\alpha\bar{\sigma}^\beta{}_\mu]_{\bar{s}} \\ &+ \left(\int_s^\infty\int_{-\infty}^{\bar{s}} - \int_{-\infty}^{s}\int_{\bar{s}}^{\infty}\right) ds\, d\bar{s}\,\frac{\partial\Lambda}{\partial\xi^\mu}, \end{aligned} \quad (2.10)$$

where

$$\Lambda = e\bar{e}\dot{x}_\mu\dot{\bar{x}}^\mu\delta + e\bar{G}\dot{x}_\mu\bar{\sigma}^{\mu\nu}\partial_\nu\delta - \bar{e}G\dot{\bar{x}}_\mu\sigma^{\mu\nu}\partial_\nu\delta - G\bar{G}\sigma^{\mu\alpha}\bar{\sigma}^\beta{}_\mu\partial_\alpha\partial_\beta\delta. \quad (2.11)$$

Since $\partial_\nu\delta = -\bar{\partial}_\nu\delta$, Λ is symmetric between the two particles.

The total angular momentum of the system about the origin is

$$\begin{aligned} L_{\mu\nu} &= [2(m - \tfrac{1}{2}G\sigma^{\alpha\beta}H_{\alpha\beta})x_{[\mu}\dot{x}_{\nu]} - 2x_{[\mu}\sigma_{\nu]\alpha}\ddot{x}^\alpha + \sigma_{\mu\nu} \\ &+ 2ex_{[\mu}A_{\nu]} - 2GF_{\alpha\beta}\dot{x}^\alpha x_{[\mu}\sigma^\beta{}_{\nu]}]_s \\ &+ [2(\bar{m} - \tfrac{1}{2}\bar{G}\bar{\sigma}^{\alpha\beta}\bar{H}_{\alpha\beta})\bar{x}_{[\mu}\dot{\bar{x}}_{\nu]} - 2\bar{x}_{[\mu}\bar{\sigma}_{\nu]\alpha}\ddot{\bar{x}}^\alpha + \bar{\sigma}_{\mu\nu} \\ &+ 2\bar{e}\bar{x}_{[\mu}\bar{A}_{\nu]} - 2\bar{G}\bar{F}_{\alpha\beta}\dot{\bar{x}}^\alpha\bar{x}_{[\mu}\bar{\sigma}^\beta{}_{\nu]}]_{\bar{s}} \\ &+ 2\left(\int_s^\infty\int_{-\infty}^{\bar{s}} - \int_{-\infty}^{s}\int_{\bar{s}}^{\infty}\right) ds\, d\bar{s} \\ &\times \left[x_{[\mu}\frac{\partial\Lambda}{\partial\xi^{\nu]}} + \dot{x}_{[\mu}\frac{\partial\Lambda}{\partial\dot{x}^{\nu]}} + 2\frac{\partial\Lambda}{\partial\sigma^{\alpha[\nu}}\sigma^\alpha{}_{\mu]}\right]. \end{aligned} \quad (2.12)$$

It is understood that the term $\partial\Lambda/\partial\sigma^{\alpha\nu}$ has been skew symmetrized with respect to α and ν.

It follows from the equations of motion that the quantities P_μ and $L_{\mu\nu}$ are independent of the choice of the point s on the world line of particle m and of the point $\bar{s}$ on the world line of particle $\bar{m}$. This is the conservation of the linear and angular momentum of the system.

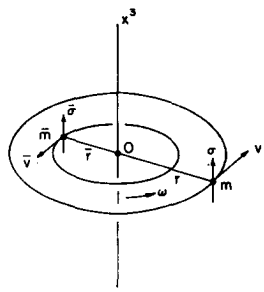


FIG. 1. Spinning particles in circular motion.

3. THE SPECIAL TWO-BODY SYSTEM

The results of a rigorous relativistic calculation will be given now for two particles in circular motion with spins parallel (or antiparallel) to the axis of rotation. The two-body system is shown in Fig. 1.

The two particles are characterized by the constants m, e, G, σ and $\bar{m}$, $\bar{e}$, $\bar{G}$, $\bar{\sigma}$. The quantity σ is positive if the spin of particle m is parallel to the axis x^3 of rotation of the system, it is negative if the spin is antiparallel; similarly $\bar{\sigma}$ may be positive or negative. The magnitudes of the spins are $S = |\sigma|$ and $\bar{S} = |\bar{\sigma}|$.

We choose as independent variables describing the motion the speeds v and $\bar{v}$ of the two particles and the angular velocity ω of the system. The radii of the orbits are then derived variables given by

$$r = v/\omega, \quad \bar{r} = \bar{v}/\omega. \quad (3.1)$$

Another derived variable is the retardation angle θ, defined as the angle from O through which one particle turns during the time $\tau = \theta/\omega$ which it takes light from the other particle to reach it. The retardation angle θ is the positive root of the retardation relation

$$v^2 + \bar{v}^2 + 2v\bar{v}\cos\theta - \theta^2 = 0, \quad (3.2)$$

which can be read from Fig. 2.

The orbital motion of the particles as a function of the time coordinate $x^4 = t$ is described by

$$x^\mu = \left(\frac{v}{\omega}\cos\omega t, \frac{v}{\omega}\sin\omega t, 0, t\right), \quad (3.3)$$

$$\bar{x}^\mu = \left(-\frac{\bar{v}}{\omega}\cos\omega t, -\frac{\bar{v}}{\omega}\sin\omega t, 0, t\right). \quad (3.4)$$

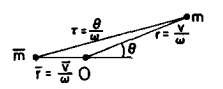


FIG. 2. The retardation relation.

The spins of the particles are

$$\sigma_{\mu\nu} = \beta\sigma \begin{pmatrix} 0 & 1 & 0 & -v\cos\omega t \\ -1 & 0 & 0 & -v\sin\omega t \\ 0 & 0 & 0 & 0 \\ v\cos\omega t & v\sin\omega t & 0 & 0 \end{pmatrix}, \tag{3.5}$$

$$\bar{\sigma}_{\mu\nu} = \bar{\beta}\bar{\sigma} \begin{pmatrix} 0 & 1 & 0 & \bar{v}\cos\omega t \\ -1 & 0 & 0 & \bar{v}\sin\omega t \\ 0 & 0 & 0 & 0 \\ -\bar{v}\cos\omega t & -\bar{v}\sin\omega t & 0 & 0 \end{pmatrix}, \tag{3.6}$$

where

$$\beta = (1 - v^2)^{-\frac{1}{2}}, \quad \bar{\beta} = (1 - \bar{v}^2)^{-\frac{1}{2}}. \tag{3.7}$$

These spins satisfy the requirements $\dot{x}^\mu\sigma_{\mu\nu} = \dot{x}^\mu\bar{\sigma}_{\mu\nu} = 0$, $S^2 = \sigma^2$, $\bar{S}^2 = \bar{\sigma}^2$.

The calculations of the equations of motion of the particles and of the conserved quantities for the system are long but straightforward. Eq. (2.9) for the spin $\sigma_{\mu\nu}$ and the corresponding equation for $\bar{\sigma}_{\mu\nu}$ are satisfied identically, Of the orbital equations of motion, only the radial components given independent conditions on the variables v, $\bar{v}$, ω. They are

$$\begin{aligned} &m\beta v + \sigma\beta^3 v\omega - e\bar{e}\omega\gamma\frac{d}{d\theta} \\ &\quad\times\{\gamma(1 + v\bar{v}\cos\theta)(v + \bar{v}\cos\theta) + \bar{v}\sin\theta\} \\ &\quad + (e\bar{G}\bar{\sigma} + \bar{e}G\sigma)\omega^2\gamma\frac{d}{d\theta}\gamma(v + \bar{v}\cos\theta) \\ &\quad + G\sigma\bar{G}\bar{\sigma}\omega^3\gamma\frac{d}{d\theta}\gamma\frac{d}{d\theta} \\ &\quad\times\{\gamma(1 + v\bar{v}\cos\theta)(v + \bar{v}\cos\theta) + \bar{v}\sin\theta\} = 0, \end{aligned} \tag{3.8}$$

$$\begin{aligned} &\bar{m}\bar{\beta}\bar{v} + \bar{\sigma}\bar{\beta}^3\bar{v}\omega - e\bar{e}\omega\gamma\frac{d}{d\theta} \\ &\quad\times\{\gamma(1 + v\bar{v}\cos\theta)(\bar{v} + v\cos\theta) + v\sin\theta\} \\ &\quad + (e\bar{G}\bar{\sigma} + \bar{e}G\sigma)\omega^2\gamma\frac{d}{d\theta}\gamma(\bar{v} + v\cos\theta) \\ &\quad + G\sigma\bar{G}\bar{\sigma}\omega^3\gamma\frac{d}{d\theta}\gamma\frac{d}{d\theta} \\ &\quad\times\{\gamma(1 + v\bar{v}\cos\theta)(\bar{v} + v\cos\theta) + v\sin\theta\} = 0. \end{aligned} \tag{3.9}$$

In these equations,

$$\gamma = (\theta + v\bar{v}\sin\theta)^{-1}, \tag{3.10}$$

and the differentiations with respect to θ are to be carried out before the retardation relation (3.2) is used to determine θ. This remark also applies to the equations below, which give the conserved quantities.

The symmetry of our two-body problem implies that the space origin O is the relativistic center of mass of the system. It follows that the only nonzero components of P^μ and $L_{\mu\nu}$ are the total energy $E = P^4$ of the system and the third component $L = L_{12}$ of the angular momentum about O. These quantities are given by

$$\begin{aligned} E = {} & m(1 - v^2)^{\frac{1}{2}} + \bar{m}(1 - \bar{v}^2)^{\frac{1}{2}} + (e\bar{G} + \bar{e}G\sigma)\omega^2\gamma \\ & + 2G\sigma\bar{G}\bar{\sigma}\omega^3\gamma\frac{d}{d\theta}\gamma(1 + v\bar{v}\cos\theta), \end{aligned} \tag{3.11}$$

$$\begin{aligned} L = {} & \beta\sigma + \bar{\beta}\bar{\sigma} - e\bar{e}\gamma(1 + v\bar{v}\cos\theta) + 2(e\bar{G}\bar{\sigma} + \bar{e}G\sigma)\omega\gamma \\ & + 3G\sigma\bar{G}\bar{\sigma}\omega^2\gamma\frac{d}{d\theta}\gamma(1 + v\bar{v}\cos\theta). \end{aligned} \tag{3.12}$$

ACKNOWLEDGMENTS

This research has been sponsored by the Aerospace Research Laboratory, Office of Aerospace Research, U.S. Air Force. Much of the work was done while one author (A. S.) was a National Science Foundation Senior Postdoctoral Fellow and a Visiting Professor at the Nordic Institute for Theoretical Atomic Physics in Copenhagen. He wishes to thank these institutions and Professor Christian Møller for their kind support and hospitality.

This research (J. A. S.) has been supported by a grant from the National Science Foundation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Possible Interpretation of Quantum Mechanics

L. DE LA PEÑA-AUERBACH*
Instituto de Física, Universidad de México, México
AND
LEOPOLDO S. GARCÍA-COLÍN
Instituto Mexicano de Petróleo, México and Facultad de Ciencias, Universidad de México, México

(Received 14 July 1967)

It is shown that to Schrödinger's equation, one may associate a Markoff process in the Smoluchowski approximation with an external force acting on the system which is a measure of its interaction with the vacuum. This interaction is, in this scheme, responsible for the stochastic character of the motion. The physical interpretation of the usual momentum and energy operators emerges in a natural way from the theory. Thus, we are immediately led to Heisenberg's uncertainty relations. However, this interpretation of quantum mechanics is valid only within the limits of validity of Smoluchowski's equation. As a simple example, we treat the one-dimensional harmonic oscillator.

INTRODUCTION

In recent papers,[1-3] we have shown that it is possible to treat the problem of the Brownian particle, formally at least, by means of a Schrödinger-like equation for the probability amplitude. We have demonstrated that in this description the mean value of the momentum of the particle is proportional to $\langle -i \operatorname{grad} \rangle$ and, due to the stochastic nature of the problem, the "uncertainty relations" appear between momentum and position coordinates. In this treatment the "diffusion equation" is used; i.e., we restrict ourselves to time intervals very large compared with the relaxation time $t \gg \beta^{-1}$, where β is the friction coefficient.

The results obtained strongly suggest an approach to the problem in the opposite direction, namely, to try to establish the connection between the motion of a quantum-mechanical particle and that of a corresponding classical particle, subjected to an additional stochastic force. Since this must hold also for the "free particle," we must postulate that the stochastic force has its origin in the vacuum with which the particle interacts continuously. Then the friction coefficient β of the Brownian problem corresponds, in the case of the quantum mechanical particle, to a new parameter which measures the mean frequency of the interaction with the vacuum (which in turn corresponds to the thermal bath of the classical particle).

Strictly speaking, this idea is not new. In fact, several authors[4] have contemplated it in different forms for a long time. Our approach, however, differs from others' in some aspects, several of which are discussed in the text.

I. PROPOSED INTERPRETATION

The starting point of this paper will be Schrödinger's equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V\psi. \tag{1}$$

As a first step in our procedure, we show that Eq. (1) may be transformed into a Smoluchowski equation with a probability density ρ which is equal to the norm of the complex probability amplitude ψ. To achieve this we shall write ψ in the usual textbook form,

$$\psi = e^{R+iS}, \tag{2}$$

where R and S are real, dimensionless functions of the coordinates and the time. From Eqs. (1) and (2) we obtain immediately the system of equations:

$$\dot{R} = -\tfrac{1}{2}\alpha \nabla^2 S - \alpha \operatorname{grad} R \cdot \operatorname{grad} S, \tag{3}$$

$$-\dot{S} = -\tfrac{1}{2}\alpha \nabla^2 R - \tfrac{1}{2}\alpha[(\operatorname{grad} R)^2 - (\operatorname{grad} S)^2] + V/\hbar, \tag{4}$$

where we have set $\alpha = \hbar/m$.

It is well known that Eq. (3) may be written in the form of a continuity equation.[5] Indeed, introducing in Eq. (3) the integrating factor e^{2R} and writing the resulting expression in the form:

$$\partial e^{2R}/\partial t = -\alpha \operatorname{div} [e^{2R} \operatorname{grad} S],$$

we have that

$$\partial \rho/\partial t + \operatorname{div} [\alpha \rho \operatorname{grad} S] = 0, \tag{5}$$

where

$$\rho = e^{2R} = |\psi|^2. \tag{6}$$

* Technical consultant, Comisión Nacional de Energía Nuclear (México).

[1] L. de la Peña-Auerbach, Phys. Letters **24A**, 603 (1967).

[2] L. de la Peña-Auerbach, E. Braun, and L. S. García-Colín, J. Math. Phys. **9**, 668 (1968).

[3] Hereafter, we shall refer to these papers as I and II, respectively.

[4] See references in Papers I and II.

[5] D. Bohm, Phys. Rev. **85**, 166 (1952).

Equation (5) is the continuity equation of quantum mechanics, which is the basis for the interpretation of ρ as the probability density, and ψ the wavefunction as the probability amplitude. The remaining equation may be rewritten as follows:

$$\frac{\partial(\hbar S)}{\partial t} + \frac{(\text{grad } \hbar S)^2}{2m} + V - \frac{\hbar^2}{2m}[\nabla^2 R + (\text{grad } R)^2] = 0,$$

which is interpreted by Bohm[5] as the Hamilton–Jacobi equation for the action $\hbar S$ for a particle acted on simultaneously by the classical potential V and the "quantum-mechanical potential" $\phi_B = -(\hbar^2/2m) \times [\nabla^2 R + (\text{grad } R)^2]$. This interpretation does not seem very convincing to us, but we shall postpone its discussion to a later stage of this paper.

For the time being we shall make use of Eq. (5) only and demonstrate that it is possible to write it in a form which corresponds to the Smoluchowski equation of a classical particle subjected to a stochastic force. This is easily achieved by introducing the new function Q defined by

$$Q = R + S. \tag{7}$$

In terms of Q and R, Eq. (3) takes the form

$$\dot{R} = -\tfrac{1}{2}\alpha[\nabla^2 Q - \nabla^2 R] - \alpha \text{ grad } R \cdot (\text{grad } Q - \text{grad } R),$$

which, with the aid of Eq. (6), may be written as

$$\partial\rho/\partial t + \text{div } [(\alpha \text{ grad } Q)\rho - \tfrac{1}{2}\alpha \text{ grad } \rho] = 0. \tag{8}$$

This equation has the sought-after form, namely that of a Smoluchowski equation of a Brownian particle acted on by an external force $\mathbf{K}$ per unit mass[6] given by

$$\mathbf{K} = \alpha\beta \text{ grad } Q \tag{9}$$

and with a diffusion coefficient D given by

$$D = \alpha/2 = \hbar/2m. \tag{10}$$

If we agree to give this physical interpretation to Eq. (8), then β is a parameter, which in view of the discussion set forth in the introduction is a measure of the interaction between the particle and the vacuum. Thus, β will have a meaning analogous to that of the corresponding parameter in the theory of the random flight.[6,7] Strictly speaking, we are therefore studying not a Brownian particle, but a "random-flight" particle subjected to an external force. It is evident that this interpretation is not in contraposition with quantum mechanics, at least from the point of view of the final results.

The postulation of Eq. (8) as the fundamental one, which implies that we are considering a description of the motion of a classical Brownian particle valid for time intervals such that $\Delta t \gg \beta^{-1}$, allows us to understand many facts of quantum mechanics with classical concepts and in a very direct way. In the first place, due to the stochastic nature of the problem, there is a mean-square deviation of the position and velocity coordinates, which gives rise to the momentum–position uncertainty relations. Secondly, there is no such thing as *the* trajectory of the particle between two given points, but every particle follows its own trajectory, giving rise, in the mean, to a current density which lends itself only to a statistical description.

This result is somewhat connected with the physical interpretation of Eq. (4) as a Hamilton–Jacobi equation for the action $\hbar S$, as was pointed out earlier. In fact, recall that the existence of such an equation implies that the particle obeys a least-action principle, by means of which one can define, at least in principle, the trajectory for the particle. But we have seen that the quantum-mechanical particle has a manifold of possible paths between the two given points and not only one, which in turn implies that such an interpretation is not legitimate. Furthermore, from Eq. (5) we conclude that ρm^{-1} grad $(\hbar S)$ represents the current density of the particle, where m^{-1} grad $(\hbar S)$ stands for the flow velocity of the particle, but not for its instantaneous velocity.

It is important to notice that Eq. (1) is considered to be valid for all time intervals. On the other hand, Eq. (8), which is a direct consequence of Eq. (1), is valid, according to our interpretation, for time intervals such that $\beta\Delta t \gg 1$. This contradiction leads us to one of the two following possibilities:

(a) It is not possible to interpret Eq. (8) as a Smoluchowski equation due to the additional restrictions not contained in Schrödinger's equation.
(b) It is possible to interpret Eq. (8) as a Smoluchowski equation provided that we assume that the restriction $\beta\Delta t \gg 1$ applies also to Schrödinger's equation.

In this paper we accept the second possibility as the correct one and thus interpret the condition $\beta\Delta t \gg 1$ as an oversimplified form of the time–energy uncertainty relation. In fact, due to the stochastic nature of the problem, the energy of an isolated particle is a fluctuating quantity and we can roughly estimate its

[6] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[7] Ming Chan Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).

fluctuation with the aid of Eq. (9) as $\Delta E \sim \alpha\beta m \Delta Q$. Eliminating the unknown parameter β with the aid of the condition $\beta\Delta t \gg 1$, we have $|\Delta E|\,\Delta t \geq \hbar\Delta Q \sim \hbar$. In our description, this classical relation has a very clear meaning: As the time interval used to measure the mean value of the energy of the particles grows, the dispersion in the measured value is reduced, because more and more fluctuations are taken into account.

The force **K** acting on the equivalent classical particle contains the information about the boundary conditions of the problem through Q according to Eq. (9) and thus depends on the whole system. This feature allows us to understand the origin of some typical quantum effects. Indeed, suppose that the configuration of the system is modified by changing the geometry of the boundaries, e.g., the opening or closing of a slit. Clearly, the system will readjust itself to the new conditions through the change of the probability amplitude from its old value to a new one. The transient is not described by Schrödinger's equation, but after it has died out the resulting Q will differ, in general, from the original one. Hence, the force **K** will also be different and thus the motion of the particle affected without having exerted a direct action on the particle itself.

This qualitative discussion has some very interesting consequences. We are postulating that the fundamental equation of quantum mechanics, namely, Schrödinger's equation, is not generally valid, but is restricted to time intervals such that $\beta\Delta t \gg 1$. This in turn means that, at least in principle, it is possible to construct a more general machinery for the description of a quantum-mechanical problem. To achieve this, we need only use as our starting point the Fokker–Planck equation, in order to eliminate the above restriction. This treatment of the problem gives us a more precise description of the motion of the particle, valid for all time intervals. Some initial results along these lines will be published elsewhere.

II. SOME BASIC RELATIONS

The aim of this section is a two-fold one. Firstly, we want to show that in our interpretation, we can measure the average value of the momentum of a particle by the expectation value of the operator $-i\hbar$ grad, and secondly, to give a physical meaning to Eq. (4). For these purposes, let us define the velocity operator $\hat{\mathbf{v}}$:

$$\hat{\mathbf{v}} = (\hbar/m)\,\text{grad}\,Q - (\hbar/2m)\,\text{grad} \tag{11}$$

so that the density current is given, according to (8), by

$$\mathbf{j} = \hat{\mathbf{v}}\rho = \mathbf{v}\rho, \tag{12}$$

$\mathbf{v}$ being an eigenvalue of the operator $\hat{\mathbf{v}}$.

Introduce also the Hermitian operator $\hat{\mathbf{p}}$ defined as

$$\hat{\mathbf{p}} = -im\alpha\,\text{grad} = -i\hbar\,\text{grad}. \tag{13}$$

As shown in Paper II, the mean value of this operator $\langle\hat{\mathbf{p}}\rangle_{\text{av}}$ is zero and also, because of its Hermiticity, both the mean and expectation values[8] of grad R are also equal to zero. These two results immediately imply that

$$\langle\hat{\mathbf{p}}\rangle = \hbar\langle\text{grad}\,S\rangle_{\text{av}}, \tag{14}$$

and furthermore, using Eq. (7), that

$$\langle m\mathbf{v}\rangle_{\text{av}} = \hbar\langle\text{grad}\,Q\rangle_{\text{av}} = \hbar\langle\text{grad}\,S\rangle_{\text{av}}. \tag{15}$$

Comparison of Eqs. (14) and (15) yield, finally, that

$$\langle\hat{\mathbf{p}}\rangle = \langle m\mathbf{v}\rangle_{\text{av}}. \tag{16}$$

Equation (16) is therefore the proof of our first assertion, namely that in our picture, we can identify the operator $\hat{\mathbf{p}} = -i\hbar$ grad as the momentum operator, in the sense that its eigenvalues are equal to the mean value of the momentum of the representative particle. This result also guarantees the validity of the usual $(\mathbf{v}, \mathbf{p})$ uncertainty relations, through the commutator $[x, \hat{\mathbf{p}}_x] = i\hbar$.

In an entirely analogous way, we can introduce the Hermitian operator

$$\hat{E} = i\hbar(\partial/\partial t), \tag{17}$$

and show that

$$\langle\hat{E}\rangle = -\hbar\langle\dot{S}\rangle_{\text{av}} = -\hbar\langle\dot{S}\rangle, \tag{18}$$

making use of the result that $\langle\dot{R}\rangle = \langle\dot{R}\rangle_{\text{av}} = 0$, already obtained in Paper II.

Now let us rewrite Eq. (4) and take its expectation value to obtain

$$-\hbar\langle\dot{S}\rangle = (\hbar^2/2m)\langle(\text{grad}\,S)^2 - (\text{grad}\,R)^2 - \nabla^2 R\rangle + \langle V\rangle. \tag{19}$$

We can write Eq. (19) in several different, but clearly equivalent, forms. In the first place, using the relation

$$\langle\hat{\mathbf{p}}^2\rangle = -\hbar^2\langle\nabla^2 R + (\text{grad}\,R)^2 - (\text{grad}\,S)^2\rangle,$$

together with Eqs. (17) and (18), we have that

$$\left\langle i\hbar\frac{\partial}{\partial t}\right\rangle = \langle\hat{E}\rangle = \left\langle\frac{\hat{\mathbf{p}}^2}{2m} + V\right\rangle = \langle\hat{H}\rangle, \tag{20}$$

where $\hat{H} \equiv \hat{\mathbf{p}}^2/2m + V$ is the Hamiltonian operator. This is an almost trivial result, but it allows us to

[8] Recall that if f is an operator, $\langle f\rangle \equiv \int\psi^* f\psi\,d\mathbf{v}$ and $\langle f\rangle_{\text{av}} \equiv \int f\rho\,d\mathbf{r}$.

interpret $\hat{E}$ as the energy operator in the usual quantum-mechanical sense. In the second place, we have from Eqs. (11) and (12) that the flow momentum is given by

$$m\mathbf{v} = \hbar \operatorname{grad} Q - \tfrac{1}{2}\hbar(\operatorname{grad} \rho/\rho) = \hbar \operatorname{grad} S. \quad (21)$$

In view of this relation, we can transform Eq. (19) into the form

$$\langle \hat{E} \rangle = \langle \tfrac{1}{2}m\mathbf{v}^2 + V + \phi_B \rangle, \quad (22)$$

where $\phi_B = -(\hbar^2/2m)[(\operatorname{grad} R)^2 + \nabla^2 R]$ is Bohm's potential.[5]

We thus see that Eq. (4) leads to two alternative interpretations for the expectation value of the energy operator $\hat{E}$. On the one hand, Eq. (20) has the usual meaning, namely that the expectation value of the energy operator is given by the sum of two terms, the total kinetic energy of the particle $\langle \hat{\mathbf{p}}^2/2m \rangle$ plus the expectation value of the external potential V. On the other hand, Eq. (22) may be interpreted as the sum of two terms, the average kinetic energy of flow $\langle \frac{1}{2}m\mathbf{v}^2 \rangle$ plus an effective potential $V + \phi_B$. This, however, is not the case. Indeed the so-called Bohm's potential ϕ_B has as its expectation value the following one, namely

$$\langle \phi_B \rangle = (1/2m)\langle \hat{\mathbf{p}}^2 - (m\mathbf{v})^2 \rangle,$$

thus showing that the two interpretations are mathematically identical, and moreover, that this quantity $\langle \phi_B \rangle$ has the true physical meaning of the difference between the expectation values of the total kinetic energy and the kinetic energy of flow, i.e., it is the mean stochastic kinetic energy.

We can write Eq. (4) in still another form, explicitly using our interpretation of the motion of the particle in terms of the Smoluchowski equation. This immediately implies that we are working in the static approximation,[9] in which the term $m(d^2\mathbf{r}/dt^2)$ is considered negligible compared to the "viscosity" term $m\beta(d\mathbf{r}/dt)$. This implies the possibility of referring to "applied velocities" $\beta^{-1}\mathbf{K}$ instead of "applied forces" $\mathbf{K}$. Consequently, in this approximation we must calculate the energy associated to the force $\mathbf{K}$ by $\frac{1}{2}m(\beta^{-1}\mathbf{K})^2$ instead of $-\int \mathbf{K} \cdot d\mathbf{r}$. Taking into account Eq. (9), the "impressed velocity" is thus equal to:

$$\mathbf{u} = \beta^{-1}\mathbf{K} = \alpha \operatorname{grad} Q = \alpha \operatorname{grad} R + \mathbf{v},$$

where we have used Eq. (21). Notice that $\langle \mathbf{u} \rangle_{\text{av}} = \langle \mathbf{v} \rangle_{\text{av}}$. From this equation it follows that

$$\operatorname{grad} R = -(m/\hbar)(\mathbf{v} - \mathbf{u}), \quad (23)$$

so that Bohm's "potential" may be rewritten as follows:

$$\phi_B = -\frac{\hbar^2}{2m}\left[\frac{m^2}{\hbar^2}(\mathbf{v} - \mathbf{u})^2 - \frac{m}{\hbar}\operatorname{div}(\mathbf{v} - \mathbf{u})\right]$$
$$= -\tfrac{1}{2}m(\mathbf{v} - \mathbf{u})^2 + \frac{\hbar}{2}\operatorname{div}(\mathbf{v} - \mathbf{u}).$$

With this result, we can write for the total energy, according to Eq. (19),

$$\hat{E} = \tfrac{1}{2}m\mathbf{v}^2 - \tfrac{1}{2}m(\mathbf{v} - \mathbf{u})^2 + (\hbar/2)\operatorname{div}(\mathbf{v} - \mathbf{u}) + V, \quad (24)$$

which is the sought for alternative form for the energy.

With the aid of Eq. (24) we can establish a formal analogy between the quantum-mechanical equations and those of fluid dynamics. Indeed, if we take the gradient of Eq. (24) and notice that

$$\operatorname{grad} E = -\hbar(\partial/\partial t) \operatorname{grad} S = -m\partial\mathbf{v}/\partial t,$$

we obtain the following result:

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \operatorname{grad} \mathbf{v} + \operatorname{grad}\left[\frac{\alpha}{2}\operatorname{div}(\mathbf{v} - \mathbf{u}) - \tfrac{1}{2}(\mathbf{v} - \mathbf{u})^2\right]$$
$$= -\operatorname{grad} V. \quad (25)$$

Equation (25) is of the form of Euler's equation for an ideal fluid, on which an external potential V is acting and supporting a hydrostatic pressure. This formal analogy has been used by some authors in a more or less explicit form,[10,11] with the purpose of giving a different interpretation of quantum mechanics. We must remark, however, that from our point of view, this analogy is purely formal and does not have any deep physical meaning. It comes out as a direct consequence of the fact that our two basic equations [e.g., (8) and (25)] are exactly of the same nature as the basic equations of hydrodynamics, namely the continuity equation and the energy conservation equation.

III. HARMONIC OSCILLATOR

It is clear from the above considerations that our model not only does not contradict usual quantum mechanics, but allows us to understand it in a different and very simple way. Nevertheless, in order to get a better understanding of the proposed interpretation, we consider it instructive to give a simple example, namely that of a one-dimensional harmonic oscillator.

Let ψ_n be the wavefunction corresponding to a stationary state of the oscillator with energy E_n. Then

$$\psi_n = C_n e^{i\hbar^{-1}Et - \frac{1}{2}\xi^2} H_n(\xi) = e^{R+iS}, \quad (26)$$

where $\xi = x/x_0$ and $x_0^2 = (m\omega)^{-1}\hbar$.

[9] J. P. Terletskiy, *Statistical Physics* (Izd. Vis. Shk., Moscow, 1966) (in Russian).

[10] E. Madelung, Z. Physik **40**, 332 (1926).

[11] D. Bohm and J. P. Vigier, Phys. Rev. **96**, 208 (1954).

According to Eqs. (9) and (21), we get that

$$\mathbf{v} = m^{-1}\hbar \operatorname{grad} S = 0, \tag{27}$$

and

$$\beta^{-1}\mathbf{K} = \mathbf{u} = m^{-1}\hbar \operatorname{grad} Q = -\frac{\hbar}{mx_0}\left(\xi - \frac{H_n'}{H_n}\right), \tag{28}$$

where $H_n' = dH_n/d\xi$. Therefore, the flow velocity is equal to zero in this case. Also, according to Eq. (28) the velocity $\mathbf{u}$ produced by the external force $\mathbf{K}$ contains two terms. The first one corresponds in this case to the velocity arising from the elastic force, whereas the second one may be considered as a macroscopic measure of the interaction between the particle and the vacuum. It is interesting to note that it is precisely this term which is responsible for the quantization of the system, because it may acquire only a well-defined discrete structure, when excited by an external elastic force. Which one of the infinitely many possible responses will be developed depends on the available energy $E_n = \hbar\omega(n + \frac{1}{2})$.

We can visualize the problem from another slightly different point of view. Suppose we consider a particle which goes from a certain point (say x_0 at $t = 0$) to another point (say x at $t > 0$). The transition amplitude (for $t > 0$) is given by Feynman's kernel for the harmonic oscillator[12]

$$K(x_0 \mid x, t) = \left(\frac{C}{\sin \omega t}\right)^{\frac{1}{2}} \exp \frac{im\omega}{2\hbar \sin \omega t} \times [(x^2 + x_0^2)\cos \omega t - 2x_0 x], \tag{29}$$

where C is some constant. Notice that this kernel is such that grad $R = 0$ and Eq. (4) simply reduces to

$$-\hbar \frac{\partial S}{\partial t} = \frac{\hbar^2}{2m}(\operatorname{grad} S)^2 + V. \tag{30}$$

Upon substitution of grad S calculated from Eq. (29), and introduction of the corresponding potential V, Eq. (30) gives the identity

$$-\hbar \frac{\partial S}{\partial t} = \frac{m\omega^2}{2} \frac{x^2 + x_0^2 - 2x_0 x \cos \omega t}{\sin^2 \omega t}. \tag{31}$$

Let us assume that the variable x may be written as the sum of a systematic part which we shall denote by $\{x\}$, and a stochastic part $x - \{x\}$. Then we can write for the purely systematic part of Eq. (31):

$$-\hbar \frac{\partial \{S\}}{\partial t} = \frac{\hbar^2}{2m}\{\operatorname{grad} S\}^2 + V(\{x\}) = \frac{m\omega^2}{2} \frac{\{x\}^2 + x_0^2 - 2x_0\{x\} \cos \omega t}{\sin^2 \omega t}. \tag{32}$$

[12] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Company, New York, 1965).

Since Eq. (32) is the Hamilton–Jacobi equation for the systematic part of the motion, we may write

$$\{E\} = -\hbar(\partial\{S\}/\partial t). \tag{33}$$

For $\{E\}$ to be a constant of the motion, it is necessary that

$$\{x\} = (v_0/\omega)\sin \omega t + x_0 \cos \omega t, \tag{34}$$

in view of Eqs. (32) and (33), where v_0 is an arbitrary constant. Defining $\{v\} \equiv \hbar\{\operatorname{grad} S\}$, one immediately obtains

$$\{v\} = d\{x\}/dt = v_0 \cos \omega t - \omega x_0 \sin \omega t. \tag{35}$$

We see that the systematic parts of the variables satisfy the classical laws for the harmonic oscillator. If further we assume that the energy $\{E\} = \frac{1}{2}m\{v\}^2 + \frac{1}{2}m\omega^2\{x\}^2$ of a representative particle coincides with $E_n = \hbar\omega(n + \frac{1}{2})$, we obtain $v_0^2 = 2n\omega^2 x_0^2$; i.e., the amplitude and energy of the oscillation in each state n are related one to each other through the classical equation

$$\{x\}_{\max}^2 = x_0^2 + \frac{v_0^2}{\omega^2} = x_0^2(1 + 2n) = \frac{2E_n}{m\omega^2}.$$

Returning to Eq. (28), we see that the knowledge of $\mathbf{u}$ allows us to obtain some information about the distribution ρ. In particular, ρ will have maxima at the points for which $u = 0$, that is, for the solutions of $\xi H_n - H_n' = 0$ (this is easily verified differentiating $\rho \sim e^{-\xi^2}H_n^2$); furthermore, for $H_n = 0$, we see from Eq. (28) that u becomes infinite, meaning that the probability of finding the particle at such points is equal to zero. Hence we see that one can describe the behavior of the particle either by specifying the values of ρ or alternatively by giving the values of the force through the velocity u.

IV. CONCLUSIONS

We have shown in this paper that by means of a simple and well-known transformation, Schrödinger's equation may be decomposed into a set of two equivalent equations. One of them corresponds to a continuity equation which may be interpreted in terms of a stochastic process described by Smoluchowski's equation. This process corresponds to that of a random-flight particle upon which an external force $\mathbf{K}$ is acting. Also, due to the limitations of this equation, the description of the motion is valid for time intervals long compared with its relaxation time. This relaxation time is, for a process of the type here contemplated, equal to the inverse of a certain parameter β which in a true Brownian motion is proportional to the viscosity of the medium. In the case of a

quantum particle, however, β must be interpreted as a new parameter which measures its interaction with the surroundings, namely, the vacuum. Also, it is shown that the diffusion coefficient appearing in Smoluchowski's equation is precisely equal to $\hbar/2m$.

The second equation of the set mentioned before cannot be interpreted as the Hamilton–Jacobi equation for the quantum particle, although, as is well known, reduces to it in the classical limit ($\hbar \to 0$). The reason for this stems from the stochastic nature that we are assigning to Schrödinger's equation which allows us to speak only of a current density for the particle and not of its instantaneous velocity at a given point in space. This means that the particle will have a manifold of possible paths between two given points, a result which is consistent with quantum mechanics.

It is convenient to emphasize that all our scheme of thoughts is valid only if we assume that the restriction imposed on the time intervals, for which the above description holds true, also holds for Schrödinger's equation.

On the basis of our previous ideas, we then proceed to show that the usual operators of quantum mechanics, namely the momentum and energy operators, acquire a physical significance which is inherent in the stochastic nature of the motion. Thus the usual expectation value of the operator $-i\hbar$ grad is indeed equal to the mean value of the momentum associated with the particle and that the expectation value of the total energy operator $i\hbar(\partial/\partial t)$ is the sum of the expectation values of the kinetic and external potential energies. Hence, the commutation relations hold true and through them we obtain, as a consequence of our procedure, the usual uncertainty relations.

Concerning the expectation value of the total energy operator, an important result is that the interpretation advanced by Bohm whereby the expectation value of this operator involves a term containing a "quantum-mechanical potential," is obtained, and furthermore, it is shown to be mathematically identical to the conventional one. Moreover, the expectation value of this "quantum potential" is seen to have the physical meaning of an average stochastic kinetic energy. Finally, it is possible to establish the formal analogy between the quantum-mechanical equations and those of hydrodynamics.

When our results are applied to a one-dimensional harmonic oscillator, we obtain explicit forms for the relevant quantities. It is then clear what the meaning of **K** is, at least for this case. Indeed, the external force contains, in addition to the usual elastic force, a term which is a direct measure of the interaction between the particle and the vacuum. This term is the one responsible for the quantization of the system. Finally it is shown, with the aid of Feynman's kernel for the oscillator, that the systematic or non-stochastic part of the dynamical variables satisfies the laws of classical mechanics.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to E. Braun for valuable discussions and Ana Maria Cetto for her assistance.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Simple Generalization of Schrödinger's Equation

L. DE LA PEÑA-AUERBACH*
Instituto de Física, Universidad de México, México 20, D.F.
AND
LEOPOLDO S. GARCÍA-COLÍN
Facultad de Ciencias, Universidad de México, México 20, D.F. and Instituto Mexicano del Petróleo, México 14, D.F.

(Received 7 August 1967)

A possible generalization of quantum mechanics is examined by showing that the motion in phase space of a classical Brownian particle may be described by a complex probability amplitude depending on the phase-space coordinates and the time, and obeying a Schrödinger-like equation. However formal this result may seem, the usual dynamical operators may be defined whose physical meaning stems directly from the theory. An outstanding feature of the formalism is that ordinary quantum mechanics in configuration space may be recovered in a limiting process whereby the velocity variable, defined now through a statistical distribution, is eliminated. Therefore, it plays the role of a hidden variable. This result supports recent reinterpretations of von Neumann's theorem on the nonexistence of such variables in quantum mechanics and serves as a counterexample of the usual interpretation of his theorem.

I. INTRODUCTION

In previous papers,[1] the first three of a series in which we have engaged ourselves to study a possible relationship between nonrelativistic quantum mechanics and the theory of stochastic processes, we have discussed the Schrödinger and Smoluchowski equations, one from the standpoint of the other. If the latter one is used to describe the motion of a Brownian particle, then, as was shown in Paper II, such a motion can be understood via a Schrödinger-like equation defining a certain complex probability amplitude in configuration space. Also, to the ordinary quantum-dynamical operators, a physical meaning was given which stems from the stochastic nature of the process. The results so obtained suggested the opposite approach to the problem, namely, to describe the motion of a quantum-mechanical particle through that of a classical particle moving under the action of a stochastic force. This problem was dealt with in Paper III, where Smoluchowski's equation was derived from Schrödinger's equation and again, the usual quantum-mechanical operators could be interpreted physically using only the stochastic nature of the classical motion. In both cases the usual commutation relations were obtained, and through them, the uncertainty principle was shown to be valid. However, the two descriptions suffer from a severe limitation; namely, that in the Smoluchowski approximation, we know that the motion is correctly described only for time intervals which are long compared with the relaxation time of the particle. Recall that this relaxation time is essentially inversely proportional to a certain parameter β which in the case of a classical Brownian particle is a measure of the viscosity of the medium but for a quantum particle it is interpreted as a measure of its interaction with the vacuum.

As was advanced in Paper III, the obvious way to remove this limitation is to consider the quantum particle as described by a motion of a classical Brownian particle in phase space, i.e., by Fokker–Planck's equation.[2] However, the step of going into phase space has a stronger motivation which may be stated as follows: We want to obtain a more precise description of the dynamics of an ordinary quantum-mechanical particle by a generalized Schrödinger equation which gives the time rate of change of a complex probability amplitude in the phase space of the particle. This equation is derived from a Fokker–Planck equation using a method which is analogous to that used in II. Also, this description is valid for all times of the particle's motion. Next, we want to recover the ordinary quantum mechanics in configuration space by taking an asymptotic limit corresponding to times long compared to the particle's relaxation time. This is achieved by eliminating one of the variables appearing in the exact description and to which a statistical distribution may be attached. This variable, the stochastic velocity $\mathbf{u}$ in our case, thus plays the role of a "hidden variable." The result so obtained is in agreement with the recent interpretation

* Technical Consultant. Comisión Nacional de Energía Nuclear (Mexico).

[1] L. de la Peña-Auerbach, Phys. Letters **24** A603 (1967); L. de la Peña-Auerbach, E. Braun, and L. S. García-Colín, J. Math. Phys. **9**, 668 (1968); L. de la Peña-Auerbach, and L. S. García-Colín, J. Math. Phys. **9**, 916 (1968). These papers will be hereafter referred to as I, II, and III, respectively. See also L. de la Peña-Auerbach and L. S. García-Colín, Rev. Mexicana Fis. **16**, 221 (1967).

[2] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943). Reprinted in *Noise and Stochastic Processes*, N. Wax, Ed. (Dover Publications, Inc., New York, 1954).

that several authors,[3] especially Feyerabend,[4] have given to von Neumann's theorem.[5] In fact, the claim is that hidden variables introduced in order to provide a more precise description of a statistical process are given by a nonfree dispersion distribution, which is precisely our result.

The idea of using the phase space of a mechanical system to discuss ordinary quantum mechanics has also been explored recently by Della Riccia and Wiener.[6] These authors indeed derive Schrödinger's equation from an extended form of Liouville's theorem, thus exhibiting the existence of hidden variables, and furthermore showing that such a description does not violate the uncertainty relations, although they cannot establish them. This is the major difference between their approach and ours. We are considering the transition to phase space to obtain a generalized Schrödinger's equation from which we shall recover, in the asymptotic sense described above, ordinary quantum mechanics together with the commutation rules for the operators and hence the uncertainty relationships. It is not our intention to go very deep into a discussion of the two methods, although some remarks are given in Sec. IV. However, let it be stated that the procedure here presented to generalize quantum mechanics is just a particular case of a more general one in which context the comparison between the two formalisms shows up in its full splendor, but we leave the details for a forthcoming paper. (See last reference in Ref. 1).

In Sec. II of this paper we show how a generalized Schrödinger equation may be obtained from Fokker–Planck's equation using methods similar to those already outlined in Papers I and II. Section III is devoted to a short discussion of a proposed physical interpretation of the results derived in the previous section, and finally, in Sec. IV we consider the special case in which the configuration and velocity spaces are independent. In this case it is shown that the extended theory reduces to the conventional one where the particle is described by Schrödinger's equation. This reduction implies, among other things, a normal distribution of velocities. As should be expected, this reduction is obtained for times which are long compared with the relaxation time of the particle. The concluding remarks in connection with the physical nature implied by this limit are given at the end of the section.

II. GENERALIZED SCHRÖDINGER'S EQUATION

The starting point of our derivation is the Fokker–Planck equation describing the motion in phase space of a Brownian particle under the influence of a non-velocity-dependent external force. Thus[2]

$$Dw/Dt = \mathrm{div}_{\mathbf{u}} \cdot (\beta \mathbf{u} w + q \,\mathrm{grad}_{\mathbf{u}} w), \tag{1}$$

where $w(\mathbf{r}, \mathbf{u}, t)$ stands for the probability density obeying a Markoff process. Indeed, Eq. (1) is valid for any process to which a Markoffian probability density in phase space may be assigned. In this equation, D/Dt stands for the substantial hydrodynamic, or macroscopic time derivative[7]

$$D/Dt = \partial/\partial t + \mathbf{u} \cdot \mathrm{grad}_{\mathbf{r}} + \mathbf{K} \cdot \mathrm{grad}_{\mathbf{u}}, \tag{2}$$

where $\mathbf{K}$ is the external force per unit mass. Furthermore, β and q, which we shall assume to be constants, are a measure of the coupling between the system with its surroundings, and of the diffusivity of the system, respectively. Thus, they are in general parameters which characterize the physical condition of the system and its surroundings. Finally, $\mathrm{grad}_{\mathbf{u}}$ and $\mathrm{grad}_{\mathbf{r}}$ imply that the operator grad acts on the velocity and space coordinates, respectively.

Since w is a real positive function, we can rewrite it as follows:

$$w = \exp 2\mathcal{R}, \tag{3}$$

where $\mathcal{R}$ is clearly a real function of $\mathbf{r}$, $\mathbf{u}$, and t. Furthermore let us define a vector $\mathbf{F}$ such that

$$\mathbf{F} = -\beta \mathbf{u} - q w^{-1} \,\mathrm{grad}_{\mathbf{u}} w. \tag{4}$$

Substitution of Eqs. (3) and (4) back into Eq. (2) yields the result

$$D\mathcal{R}/Dt = -\tfrac{1}{2} \,\mathrm{div}_{\mathbf{u}} \mathbf{F} - \mathbf{F} \cdot \mathrm{grad}_{\mathbf{u}} \mathcal{R}. \tag{5}$$

Notice should be made of the fact that the vector $\mathbf{F}$ defined by Eq. (4) has the meaning of a mean force per unit mass which is developed on the particle as it moves through its surroundings. Furthermore, $\mathbf{F}$ is velocity-dependent and we shall assume that it may be derived from a velocity-dependent potential S which may be thought of as a generalized macroscopic dissipative function. Thus,

$$\mathbf{F} = 2q \,\mathrm{grad}_{\mathbf{u}} S, \tag{6}$$

where the proportionality constant has been chosen equal to $2q$ only for the sake of convenience. Of

[3] S. John Bell, Rev. Mod. Phys. **38**, 447 (1966); D. Bohm and J. Bub, *ibid.* **38**, 453 (1966); and references in these papers.

[4] P. K. Feyerabend, Z. Physik **145**, 421 (1956).

[5] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1955); see also Ref. 3.

[6] G. Della Riccia and N. Wiener, J. Math. Phys. **7**, 1372 (1966).

[7] We shall refer to it simply as Stokes' operator.

course, q has the required dimensions, so that $\mathbf{F}$ is a force per unit mass, $\mathcal{S}$ being dimensionless. By substituting Eq. (6) into Eq. (5) we find that

$$D\mathcal{R}/Dt = -q\nabla_{\mathbf{u}}^2\mathcal{S} - 2q\,\mathrm{grad}_{\mathbf{u}}\,\mathcal{R}\cdot\mathrm{grad}_{\mathbf{u}}\,\mathcal{S}. \quad (7)$$

If we now define a probability amplitude Ψ, where

$$\Psi \equiv \exp(\mathcal{R} + i\mathcal{S}), \quad (8)$$

which is such that

$$w = \Psi^*\Psi, \quad (9)$$

we find, using Eq. (8) to express the left-hand side and the first term in the right-hand side of Eq. (7) in terms of Ψ, the expression

$$iD\Psi/Dt = -q\nabla_{\mathbf{u}}^2\Psi + \Omega\Psi, \quad (10)$$

where

$$\Omega = -D\mathcal{S}/Dt + q[\nabla_{\mathbf{u}}^2\mathcal{R} + (\mathrm{grad}_{\mathbf{u}}\,\mathcal{R})^2 - (\mathrm{grad}_{\mathbf{u}}\,\mathcal{S})^2]. \quad (11)$$

Equation (10) is the desired extended form of Schrödinger's equation defining the probability amplitude Ψ in phase space, with the function Ω given in Eq. (11) playing the role of the potential energy. Notice that the analogy is purely formal because the left-hand side contains the Stokes' operator acting on Ψ, instead of just $\partial/\partial t$, and, in the right-hand side, the Laplacian acts on the velocity coordinates plus the fact that Ω is a space-velocity-dependent function.

It is important to notice at this stage the consistency in the procedure here followed: If we use Eq. (8) to eliminate $\mathcal{R}$ and $\mathcal{S}$ from Eq. (7), we get the conservation equation for the "probability current" in phase space, namely

$$D\Psi^*\Psi/Dt = iq[\Psi^*\nabla_{\mathbf{u}}^2\Psi - (\nabla_{\mathbf{u}}^2\Psi^*)\Psi]. \quad (12)$$

This result thus supports the interpretation of Eq. (10) as a generalized Schrödinger equation in phase space.

III. PROPOSED PHYSICAL INTERPRETATION

The main objective of this section is to propose a physical interpretation to the formal results derived in Sec. II for the extended form of Schrödinger's equation. To achieve this we shall introduce an adequate language very much along the same lines of those used, for similar purposes, in Paper II. Since w is a probability distribution, we define the mean value $\langle\hat{g}\rangle_{\mathrm{av}}$ of an operator $\hat{g}$ by

$$\langle\hat{g}\rangle_{\mathrm{av}} = \int \hat{g}w(\mathbf{r}, \mathbf{u}, t)\,d\mathbf{r}\,d\mathbf{u}, \quad (13)$$

the integration being over the whole phase space. We also define its expectation value $\langle\hat{g}\rangle$ by

$$\langle\hat{g}\rangle = \int \Psi^*\hat{g}\Psi\,d\mathbf{r}\,d\mathbf{u}. \quad (14)$$

Clearly, if $\hat{g}$ is a c function, $\langle\hat{g}\rangle_{\mathrm{av}} = \langle\hat{g}\rangle$.

Using Eqs. (4) and (6) we immediately find that

$$\langle\mathbf{F}\rangle_{\mathrm{av}} = -\beta\langle\mathbf{u}\rangle_{\mathrm{av}} = 2q\langle\mathrm{grad}_{\mathbf{u}}\,\mathcal{S}\rangle_{\mathrm{av}}, \quad (15)$$

provided, as it is usually the case, that w vanishes when $\mathbf{u}\to\infty$. Also, since $-i\,\mathrm{grad}_{\mathbf{u}}$ is an Hermitian operator, the computation of its expectation value leads to the relations

$$\langle\mathrm{grad}_{\mathbf{u}}\,\mathcal{R}\rangle_{\mathrm{av}} = 0 \quad (16a)$$

and

$$\langle -i\,\mathrm{grad}_{\mathbf{u}}\rangle = \langle\mathrm{grad}_{\mathbf{u}}\,\mathcal{S}\rangle_{\mathrm{av}} = -(\beta/2q)\langle\mathbf{u}\rangle_{\mathrm{av}}, \quad (16b)$$

where use has been made of Eq. (15). This equation shows that the mean value of the dissipative force is proportional to the mean velocity of the particle with a proportionality constant equal to $-\beta$. This constant, in the case of a Brownian particle, depends on the characteristics of the particle (mass and geometry) and those of the medium (viscosity).[2] In the case of a quantum particle, it is, however, an undetermined parameter.

Let us now define an operator $\hat{\mathbf{p}}$ as[8]:

$$\hat{\mathbf{p}} = m\hat{\mathbf{u}} = i(2mq/\beta)\,\mathrm{grad}_{\mathbf{u}}. \quad (17)$$

Taking the expectation value of $\hat{\mathbf{p}}$ and using Eq. (16b), we are easily lead to the following result, namely

$$\langle\hat{\mathbf{p}}\rangle = m\langle\mathbf{u}\rangle_{\mathrm{av}}, \quad (18)$$

or in words, the expectation value of $\hat{\mathbf{p}}$ is equal to the average momentum of the particle. Thus, we may interpret this operator as its "momentum operator."

Now consider the mean deviation of the operator $\hat{\mathbf{p}}$ defined, as usual, by

$$\Delta\hat{\mathbf{p}} = \hat{\mathbf{p}} - \langle\hat{\mathbf{p}}\rangle = \hat{\mathbf{p}} - m\bar{\mathbf{u}}, \quad (19)$$

where $\bar{\mathbf{u}} = \langle\mathbf{u}\rangle_{\mathrm{av}}$. Then, one gets that

$$\langle(\Delta\hat{\mathbf{p}})^2\rangle = \langle\hat{\mathbf{p}}^2\rangle - m^2\bar{\mathbf{u}}^2, \quad (20)$$

and therefore,

$$\langle\hat{\mathbf{p}}^2/2m\rangle = -(2mq^2/\beta^2)\langle\mathrm{grad}_{\mathbf{u}}^2\rangle = \tfrac{1}{2}m\langle\mathbf{u}\rangle_{\mathrm{av}}^2 + (2m)^{-1}\langle(\Delta\hat{\mathbf{p}})^2\rangle. \quad (21)$$

This result allows us to interpret the left-hand side as the expectation value of the total kinetic-energy operator of the particle. Indeed, this total kinetic energy is equal to the mean kinetic energy due to the flow of the particle plus a term associated with the

[8] As will be shown later, $q = \hbar\beta^2/2m$ so that Eq. (17) reduces to $i\hbar\beta\,\mathrm{grad}_{\mathbf{u}}$.

kinetic energy arising from the fluctuations of its momentum.

Alternatively, we may express Eq. (21) in a different way. Consider the expectation value of the operator $\hat{\mathbf{p}}^2$. Then, if this is calculated directly using Eqs. (8) and (10) we get that

$$\langle D\mathcal{R}/Dt\rangle_{\mathrm{av}} = \langle \nabla_{\mathbf{u}}^2 \mathcal{S} + 2\,\mathrm{grad}_{\mathbf{u}}\,\mathcal{R}\cdot\mathrm{grad}_{\mathbf{u}}\,\mathcal{S}\rangle_{\mathrm{av}} = 0, \tag{22a}$$

where use has been made of Eq. (7) and the Hermitian nature of the operator. Furthermore,

$$\langle \nabla_{\mathbf{u}}^2\rangle = \langle \nabla_{\mathbf{u}}^2\mathcal{R} + (\mathrm{grad}_{\mathbf{u}}\,\mathcal{R})^2 - (\mathrm{grad}_{\mathbf{u}}\,\mathcal{S})^2\rangle. \tag{22b}$$

Thus, using Eqs. (17) and (22b), we get that

$$\langle \hat{\mathbf{p}}^2/2m\rangle = -(2mq^2/\beta^2) \times \langle \nabla_{\mathbf{u}}^2\mathcal{R} + (\mathrm{grad}_{\mathbf{u}}\,\mathcal{R})^2 - (\mathrm{grad}_{\mathbf{u}}\,\mathcal{S})^2\rangle \tag{23}$$

or

$$\langle \hat{\mathbf{p}}^2/2m\rangle = -(2mq/\beta^2)\langle D\mathcal{S}/Dt + \Omega\rangle, \tag{24}$$

using Eq. (11).

On the other hand, evaluation of $\langle i\,D/Dt\rangle$, together with Eq. (10), yields

$$\langle i\,D/Dt\rangle = (\beta^2/2mq)\langle \hat{\mathbf{p}}^2/2m\rangle + \langle\Omega\rangle, \tag{25}$$

where Eqs. (22b) and (23) have been used. Combining this result with Eq. (24) we get that

$$\langle i\,D/Dt\rangle = -\langle D\mathcal{S}/Dt\rangle. \tag{26}$$

Equation (25) establishes the energy-conservation theorem in terms of the expectation values of the total energy operator $i\,D/Dt$, the total kinetic energy operator, and Ω, which is seen to play the role of the potential energy.

It is also worthwhile to point out that in this context, the contribution to the fluctuations of the total kinetic energy arise not only from the fluctuations in velocity space but also from a term containing cross effects of the velocity and configuration spaces. This may be seen if we calculate the expectation value of $\nabla_{\mathbf{u}}^2$ by first performing an integration by parts. We get that

$$\langle \nabla_{\mathbf{u}}^2\rangle = -\int \mathrm{grad}_{\mathbf{u}}\,\Psi^*\cdot\mathrm{grad}_{\mathbf{u}}\,\Psi\,d\mathbf{r}\,d\mathbf{u} = -[\langle(\mathrm{grad}_{\mathbf{u}}\,\mathcal{R})^2\rangle_{\mathrm{av}} + \langle(\mathrm{grad}_{\mathbf{u}}\,\mathcal{S})^2\rangle_{\mathrm{av}}],$$

so when combined with Eq. (22b), yields the relation

$$\langle \nabla_{\mathbf{u}}^2\mathcal{R}\rangle_{\mathrm{av}} + 2\langle(\mathrm{grad}_{\mathbf{u}}\,\mathcal{R})_{\mathbf{u}}^2\rangle_{\mathrm{av}} = 0.$$

Substitution of this equation into Eq. (22b) and noticing that from Eqs. (3), (4), and (6),

$$\mathrm{grad}_{\mathbf{u}}(\mathcal{R}+\mathcal{S}) = -(2q)^{-1}\beta\mathbf{u},$$

we get, after some straightforward manipulations, that

$$\langle \hat{\mathbf{p}}^2/2m\rangle = \tfrac{1}{2}m\langle \mathbf{u}^2\rangle_{\mathrm{av}} - (4mq/\beta^2)\,\langle \mathrm{grad}_{\mathbf{u}}\,\mathcal{R}\cdot\mathrm{grad}_{\mathbf{u}}\,\mathcal{S}\rangle_{\mathrm{av}}, \tag{23a}$$

or, by substracting $\frac{1}{2}m\langle\mathbf{u}\rangle_{\mathrm{av}}^2$ from both sides, the equivalent expression

$$\langle (2m)^{-1}(\Delta\hat{\mathbf{p}})^2\rangle = \tfrac{1}{2}m\langle(\Delta\mathbf{u})^2\rangle_{\mathrm{av}} - (4mq/\beta^2)\langle \mathrm{grad}_{\mathbf{u}}\,\mathcal{R}\cdot\mathrm{grad}_{\mathbf{u}}\,\mathcal{S}\rangle_{\mathrm{av}},$$

where the second term in the right-hand side is a measure of such cross effects, since the $\mathbf{u}$ gradients of $\mathcal{R}$ and $\mathcal{S}$ will, in general, be function of $\mathbf{r}$ and $\mathbf{u}$.

Notice should be made of the fact that all the results here derived are equally well suited for the treatment of a true Brownian particle of mass m moving in a heat bath whose temperature is T and viscosity η. In this case, the force per unit mass $\mathbf{K}$ is known and so are the parameters β and q as functions of m, η, T, and the size of the particle. However, in our extended version of Schrödinger's equation these parameters are so far unknown and at most we can hope that when we set the conditions from which we may recover ordinary quantum mechanics, Eq. (25) will go to its usual form, and this will imply that $(2mq)^{-1}\beta^2 \to \hbar^{-1}$, $\hbar$ being Planck's constant divided by 2π. Indeed, we show in Sec. IV that this is the case. However, in the present stage of our theory we cannot fix β. Either it is determined from experiment or else additional information must be fed into the theory.

IV. TRANSITION TO ORDINARY QUANTUM MECHANICS

The purpose of this section is to study a very simple case of the extended theory advanced in the previous section in order to obtain, firstly, a better understanding of its own intrinsic structure and, secondly, to provide an example in which ordinary quantum mechanics is recovered, via a limiting process whereby the variable $\mathbf{u}$ is eliminated. In the extended theory, this variable is defined by a statistical distribution which will turn out to be a normal distribution. Thus, the velocity of the particle appears to play the role of a "hidden variable." This means that, within the framework of our theory, a hidden variable which is used to give a more precise description of the quantum-mechanical motion of a single particle, a process which is *per se* of statistical nature, is defined through a nonfree dispersion distribution. Thus, the recent interpretation[4] of von Neumann's theorem on hidden variables[5] is confirmed.

The starting point of our model is an assumption whereby the variable $\mathbf{u}$ is considered to be separated

from the configuration variable $\mathbf{r}$ and the time t. Suppose then that it is possible to write that

$$w(\mathbf{r}, \mathbf{u}, t) = g(\mathbf{u})\rho(\mathbf{r}, t), \tag{27}$$

and, furthermore, that the external force $\mathbf{K}$ does not depend on time. Substitution of Eq. (27) back into Eq. (1) yields:

$$\partial\rho/\partial t + \mathbf{u} \cdot \operatorname{grad}_{\mathbf{r}} \rho + \rho\mathbf{K} \cdot g^{-1} \operatorname{grad}_{\mathbf{u}} g = n\beta\rho + \beta\rho\mathbf{u} \cdot g^{-1} \operatorname{grad}_{\mathbf{u}} g + (\rho q/g)\nabla_{\mathbf{u}}^2 g, \tag{28}$$

where $n = \operatorname{div}_{\mathbf{u}} \mathbf{u}$ is the number of dimensions of the velocity space. Notice that Eq. (28) is satisfied only if the hybrid term $\mathbf{u} \cdot \rho^{-1} \operatorname{grad}_{\mathbf{r}} \rho + \mathbf{K} \cdot g^{-1} \operatorname{grad}_{\mathbf{u}} g$ vanishes identically, for then we get a set of differential equations: one in $\mathbf{u}$, and another one in $\mathbf{r}$ and t consistent with Eq. (27). But the above condition immediately implies that

$$g(\mathbf{u}) = \exp(-\nu\mathbf{u}^2), \tag{29}$$

where ν is a constant and that ρ must satisfy the equation

$$\operatorname{grad}_{\mathbf{r}} \rho = 2\nu\mathbf{K}\rho. \tag{30}$$

Under these conditions, Eq. (28) is then seen to be satisfied identically if we set

$$\nu = \beta/2q, \tag{31}$$

and consequently

$$\rho^{-1}(\partial\rho/\partial t) = 0, \tag{32}$$

meaning that the space distribution function is stationary in time. Summarizing, separability of the velocity and configuration spaces is possible if we have a normal distribution in the velocities and a stationary distribution in configuration space.

From Eq. (30) we also have that

$$\operatorname{div}_{\mathbf{r}} [D \operatorname{grad}_{\mathbf{r}} \rho - (\beta D/q)\mathbf{K}\rho] = 0, \tag{33}$$

where use has been made of Eq. (31) and D is the "diffusion coefficient." On the other hand, $\partial\rho/\partial t = 0$ so that Eq. (33) may be thought of as the Smoluchowski equation for the stationary distribution ρ. Comparing with the standard equation of Brownian motion,[2] we also find that $\mathbf{K}/\beta$ is the force acting on the particle per unit mass if we put $D = q/\beta^2$, which is precisely the value for the diffusion coefficient in this approximation.

From our preceding discussion we see that one can now write for the probability amplitude Ψ the expression

$$\Psi = \exp(-\beta\mathbf{u}^2/4q)\psi(\mathbf{r}, t), \tag{34}$$

with

$$\psi(\mathbf{r}, t) = \exp[R(\mathbf{r}) + iS(\mathbf{r}, t)], \tag{35}$$

R and S are real dimensionless functions, and $\rho = \exp 2R = \psi^*\psi$. Notice also that Eq. (34) satisfies the requirement imposed by Eq. (9). Furthermore, ρ has been shown to satisfy Smoluchowski's equation so that following the reasoning of Paper II we may associate to the probability amplitude ψ the conventional Schrödinger equation, namely

$$i\,\partial\psi/\partial t = -(q/\beta^2)\nabla^2\psi + V\psi, \tag{36}$$

where V is the time-independent potential also discussed in Paper II. Multiplying Eq. (36) by $\hbar$, we find that

$$q/\beta^2 = \hbar/2m, \tag{37}$$

which, of course, could have also been obtained by noticing that the diffusion coefficient D is, for ordinary quantum mechanics, equal to $(2m)^{-1}\hbar$.[9]

Finally, from Eqs. (8), (34), and (35), we get that

$$\mathcal{R} = -(\beta/4q)\mathbf{u}^2 + R(\mathbf{r}), \tag{38a}$$

and

$$\mathcal{S} = S(\mathbf{r}, t). \tag{38b}$$

Since $\mathcal{S}$ is $\mathbf{u}$-independent, Eqs. (5) and (6) immediately imply that $DR/Dt = 0$ and that $\mathbf{F} = 0$. This last result is quite interesting: $\mathbf{F}$ is the force exerted by the medium on the particle as a consequence of its motion through it. According to Eq. (4), this force has two contributions both depending, in general, on the velocity $\mathbf{u}$. However, from the orthogonality of the configuration and velocity spaces, we have just shown that both terms cancel each other. The physical meaning of this fact is quite clear: As the particle moves through its surroundings and the motion is described in phase space, the fluctuations of the stochastic variable $\mathbf{u}$ exert a force on it. After a certain time, however, these fluctuations become smoother and smoother so that the particle no longer "feels" their presence. Indeed what happens is that an equilibrium state is reached by the particle in $\mathbf{u}$ space. Its subsequent motion in configuration space is now described by Schrödinger's equation.

It is convenient to emphasize at this stage that Smoluchowski's equation (33) implies not only Schrödinger's equation (36), but also that $\hat{\mathbf{p}} = i\hbar \operatorname{grad}_{\mathbf{r}}$ may be interpreted as the momentum operator of the particle as was shown in Paper III, and through it we guarantee the validity of the momentum-space uncertainty relationships.

Since in general the force acting on the particle is velocity dependent, we cannot resort to a canonical distribution in phase space to describe its motion. This is possible, as it is shown above, only in the particular case where the configuration and velocity spaces are orthogonal. This is the reason why Della

[9] See Eq. (10) of Paper III.

Riccia and Wiener[6] necessarily obtain Schrödinger's equation, because from the beginning they assume both a velocity-independent force (through the canonical equations) and a Gibbsian canonical distribution.

Returning to our main discussion and remembering that we now have a stationary distribution in configuration space as indicated by Eq. (32) leads to the result that $S = S(t)$ as it is usual in a quantum-mechanical stationary process. Then from Eqs. (11) and (23) we get that

$$\langle \hbar\Omega \rangle = \langle -\hbar\, \partial S/\partial t \rangle - \tfrac{1}{2}m\langle \mathbf{u}^2 \rangle, \tag{39}$$

where use has been made Eq. of (37). Also, the quantity $\langle -\hbar\, \partial S/\partial t \rangle$ has been shown in Paper III to give the total energy of the particle and hence

$$\langle \hbar\Omega \rangle = E - \tfrac{1}{2}m\langle \mathbf{u}^2 \rangle. \tag{40}$$

Nevertheless, this equation has to be interpreted exercising a great deal of care. Indeed, from Eq. (23a), taking into account that in this particular case $\mathrm{grad}_{\mathbf{u}}\, S = 0$ according to Eq. (38b), we have that

$$\langle \hat{\mathbf{p}}^2/2m \rangle = \tfrac{1}{2}m\bar{\mathbf{u}}^2, \tag{41}$$

which means that all fluctuations of the kinetic energy arise only from fluctuations in the kinetic energy of flow [c.f. Eq. (23a)].

This picture is quite similar to that occurring in the approach to local equilibrium of, say, a dilute gas. Indeed, for times short compared with the time of duration of a collision, the process is quite complicated and the state of the gas described through a probability distribution which is defined in phase space. However, after a time has elapsed of the order of the mean free time, particles have undergone several collisions so that one "feels" the medium consisting of the remaining others. This process gives rise quite rapidly to a Maxwellian distribution of the velocities, whereas the approach to an equilibrium distribution in configuration space is much slower. Hence for times of order of a few mean free times, the variables describing the state of the gas no longer depend on $\mathbf{u}$, which obeys a local Maxwell distribution function, but depend only on $\mathbf{r}$ and t, and furthermore, these variables no longer satisfy a complicated Boltzmann-type equation but their time evolution is given by the ordinary equations of macroscopic hydrodynamics. In this stage the state of the gas is given through a few local thermodynamic quantities which no longer depend on $\mathbf{u}$, this latter quantity being a "hidden variable."

Summarizing our results, we have shown that to a classical particle undergoing a Brownian motion described by a Markoffian probability density, which satisfies the Fokker–Planck equation, a generalized Schrödinger equation may be associated which defines a complex probability amplitude Ψ in the phase space of the particle. The advantage of this procedure, which, as was pointed out before, is not the most general one, is that no limitations are involved insofar as the times for which the description is valid. Furthermore, ordinary quantum mechanics follows by taking an asymptotic limit which corresponds to considering times longer than the relaxation time of the particle; i.e., those times for which the description of the system by Smoluchowski's equation is practically equivalent to the description in terms of Fokker–Planck equation. Mathematically this is achieved by a separability of the velocity and space–time variables in the probability density $w = \Psi^*\Psi$. This in turn implies a normal distribution in the velocities and a spatial distribution function which is stationary in time. These characteristics lead immediately to the ordinary Schrödinger equation in configuration space together with the commutation relations for the conventional dynamical observables as shown in Papers II and III. Therefore, the uncertainty relations are obtained directly from the theory and do not have to be introduced as an additional assumption. The velocity $\mathbf{u}$ is eliminated in this limit and hence plays the role of a hidden variable, which in the extended theory is defined by a statistical distribution. This result is in agreement with the recent interpretation of von Neumann's theorem on hidden variables as presented by Feyerabend.[4]

ACKNOWLEDGMENTS

The authors wish to thank T. A. Brody for bringing to their attention Feyerabend's paper, to Ana Maria Cetto for her assistance, and to E. Braun for valuable discussions.

Local Field Theory and Isospin Invariance. I. Free-Field Theory of Spinless Bosons*

PETER CARRUTHERS†
Laboratory of Nuclear Studies, Cornell University, Ithaca, New York

(Received 18 July 1967)

The properties of free-field theories of spinless bosons are investigated. The self-conjugate boson field theories of isospin $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$ are shown to be nonlocal: the energy, isospin, and number densities fail to commute for spacelike separations. The difficulty is traced to the necessary occurrence of non-independent field variables. For the anomalous case the dependent variables are nonlocally related to the independent ones; elimination of the former makes superficially local quantities depend nonlocally on the independent fields. The connection with canonical field theory is investigated. The p and q coordinates have anomalous commutation relations in the nonlocal case. Although the Hamiltonian appears normal in every case, the anomalous theories are characterized by zero (integrated) Lagrangian. Antiparticle conjugation is investigated in detail with attention to phase questions. For anomalous theories two types of conjugation are found, one of which nonlocally relates the field to a superposition of its adjoint. A unitary transformation is constructed which converts one type to the other. Finally, *CPT* transformation properties of normal and abnormal theories are derived and compared.

1. INTRODUCTION

It was recently shown[1] that the requirements of local relativistic field theory preclude the existence of half-integral isospin, self-conjugate bosons of zero spin. Generalizations[2–7] of this result have shown, among other things, that the result is true for arbitrary-spin particles. Such anomalous particles would violate the Gell-Mann–Nishijima formula relating charge, hypercharge, and isospin component T_3, provided one requires integral charge and identifies the $SU(2)$ group in question with the isospin variable of the strong interactions. Thus one obtains partial insight into the validity of this relation.

In the present paper we investigate in detail the free-field theory of zero-spin bosons of arbitrary isospin. Pair-conjugate and self-conjugate types are developed in parallel to bring out interesting nonlocal features of the anomalous boson field theories. Although our attitude in Ref. 1 was essentially to "explain" why no such particles are seen, we wish to be prepared for the occurrence of violations of locality that would accompany their existence. Properties of interacting fields involving anomalous bosons will be described elsewhere.

In the most popular nonlocal field theories the nonlocality occurs in the interaction; here the free-field theory itself is nonlocal. The origin of the nonlocality is related to the following fact: The existence of the symmetry group requires the existence of all $2T+1$ field components $\varphi_\alpha(x)$. The manifestly isospin-invariant operators (Hamiltonian, Lagrangian, etc.) constructed from quadratics such as $\sum_\alpha \varphi_\alpha^*(x)\varphi_\alpha(x)$ appear to be local operators. However, for self-conjugate fields expressions of this type involve redundant field components. For example, $\varphi_{-\alpha}$ is related to φ_α^*; for integral isospin these operators are simply proportional. However, for half-integral isospin the relation is nonlocal [see Eq. (3.5)]. Thus the energy density and other quantities, depend nonlocally on the dynamically-independent fields. Bad commutators (i.e., commutators failing to vanish at spacelike separations) appear throughout the theory (see Sec. 3). It is interesting that a similar difficulty[8,9] occurs in massless spin $\frac{3}{2}$ and spin-2 theories: the energy density $\mathcal{H}(x)$ is noncausal. For zero mass all but two of the components of the field tensor become dependent.

In Sec. 2 we construct appropriate particle and field operators describing the general isospin particles.[10] In Sec. 3 the basic commutation rules are worked out. Here one finds that the number, energy, and isospin densities fail to commute for spacelike separations in the anomalous case. Also the local commutation rules for the isospin densities are anomalous, although, of course, the once-integrated form has to be standard. In Sec. 4 the relation to "canonical" field theory is investigated for self-conjugate bosons. The charged scalar theory is compared with the self-conjugate isospin-$\frac{1}{2}$ theory.

* Supported in part by The Office of Naval Research.
† Present Address: Instituto di Fisica G.Marconi, Universita di Roma, Rome, Italy.
[1] P. Carruthers, Phys. Rev. Letters **18**, 353 (1967).
[2] Y. S. Jin, Phys. Letters **24B**, 411 (1967). Parts of this paper are superseded by Refs. 3–7.
[3] Martin Einhorn (unpublished).
[4] G. N. Fleming and E. Kazes, Phys. Rev. Letters **18**, 764 (1967).
[5] H. Lee, Phys. Rev. Letters **18**, 1098 (1967).
[6] B. Zumino and D. Zwanziger, Phys. Rev. **164**, 1959 (1967).
[7] P. Kantor, Phys. Rev. Letters **19**, 394 (1967)
[8] J. Schwinger, Phys. Rev. **132**, 1317 (1963).
[9] C. M. Bender and B. M. McCoy, Phys. Rev. **148**, 1375 (1966).
[10] P. Carruthers and J. P. Krisch, Ann. Phys. **33**, 1 (1965).

q and p coordinates are introduced by standard Fourier-series methods. In a normal ($T = 0, 1, 2, \cdots$) theory the dependent q's are related to other q's, while in the anomalous ($T = \frac{1}{2}, \frac{3}{2}, \cdots$) theory the dependent q's are related to p's. As a consequence the commutator algebra of the q and p variables is very different in the two cases. However, the Hamiltonian looks normal in both cases. A basic distinction is that the Lagrange function vanishes identically in the anomalous cases. Finally we indicate how the method of Lagrange multipliers may be used to derive the field equations while treating all components of the field as independent variables.

Section 5 is devoted to a detailed study of antiparticle conjugation C, with special attention to phase questions. Requiring that C map causal fields into causal fields further restricts the phase factors occurring in the transformation of state vectors. G parity is discussed, and a standard phase choice recommended. For self-conjugate bosons we can construct two types of antiparticle conjugation. One maps φ_α into φ_α^* in the standard way, while the second maps φ_α into $\varphi_{-\alpha}$. For integral-isospin particles, the two are entirely equivalent, while for half-integral-isospin particles the latter operation gives a nonlocal connection between $C\varphi_\alpha C^{-1}$ and φ_α^*. However, we exhibit a (nonlocal) unitary operator which converts one operation to the other. The G parity of an arbitrary-isospin self-conjugate boson is derived.

In Sec. 6 the CPT transformation (Θ) is investigated. For anomalous bosons we find the following: If we choose the coefficients of the operators in φ_α to be real, then an imaginary factor i^{2T} appears in the CPT transformation, while the product of phases $\eta_C\eta_P\eta_T$ has tob e ± 1. The other natural choice, with $i^{2\alpha}$ appearing in one factor of φ_α, leads to

$$\eta_C\eta_P\eta_T = \pm i.$$

In either case Θ^2 is -1, in contrast to the natural choice for normal theories.

Note added in proof: A more general treatment, valid for any spin, has been given in P. Carruthers, Phys. Letters **26B**, 158 (1968).

The appendices derive explicit unitary operators necessary for the discussion of antiparticle conjugation.

2. BOSON ISOSPIN MULTIPLETS: PARTICLE AND FIELD OPERATORS

To introduce our ideas we follow a phenomenological, particle-oriented approach. Only zero-spin bosons are considered, in order to expose the structure of the theory in its simplest form. We consider two types of isospin multiplets: (1) *pair-conjugate multiplets*, in which the antiparticles of the members of an isospin multiplet constitute a distinct isospin multiplet; and (2) *self-conjugate multiplets*, in which the antiparticles belong to the same isospin multiplet. We shall abbreviate these types by PCM and SCM, respectively. The K mesons (K^+, K^0), ($\bar{K}^0$, K^-) are the most familiar case of a PCM, while the pions (π^+, π^0, π^-) are a good example of an SCM.

A. Particle Aspects

To describe a PCM we introduce $2(2T+1)$ boson operators $a_\alpha^*(k)$ and $a_\alpha(k)$ to create and annihilate the "particles" and $2(2T+1)$ independent boson operators $b_\alpha^*(k)$ and $b_\alpha(k)$ to create and annihilate the "antiparticles." Here α denotes the T_3 eigenvalue (a_α^*, b_α^* increases T_3 by α) and hence runs from $-T$ to $+T$ in integral steps. We shall omit the four-vector k, whenever convenient.

$$\begin{aligned}
&[a_\alpha(k), a_\beta^*(k')] = [b_\alpha(k), b_\beta^*(k')] = \delta_{\alpha\beta}\delta_{kk'},\\
&[a_\alpha(k), a_\beta(k')] = [b_\alpha(k), b_\beta(k')] = 0,\\
&[a_\alpha(k), b_\beta(k')] = [a_\alpha(k), b_\beta^*(k')] = 0.
\end{aligned} \tag{2.1}$$

We employ box normalization with unit volume. It is understood (unless otherwise noted) that k is a physical four-vector with $k_0 \geq m$, m being the meson mass. From our point of view it is quite arbitrary which multiplet is called "particles" and which "antiparticles." A full understanding of this distinction requires a detailed study of the charge conjugation (preferably antiparticle conjugation) operation. This theory is given in Sec. 5.

Thus for each k we have particle states $|T\alpha\rangle$ and antiparticle states $|\bar{T}\alpha\rangle$ defined by

$$\begin{aligned}
|T\alpha\rangle &= a_\alpha^*|0\rangle,\\
|\bar{T}\alpha\rangle &= b_\alpha^*|0\rangle,
\end{aligned} \tag{2.2}$$

where $|0\rangle$ denotes the vacuum state. In order to discuss conveniently the group properties of the theory we require the various $|T\alpha\rangle$ to be related to each other by isospin operators chosen to obey the standard (Condon–Shortley[11]) phase relation. (For a given α, the phase of a_α or b_α can be arbitrarily adjusted.) With the standard definition $T_\pm = T_1 \pm iT_2$, we have

$$\begin{aligned}
T_3|T\alpha\rangle &= \alpha|T\alpha\rangle,\\
T_3|\bar{T}\alpha\rangle &= \alpha|\bar{T}\alpha\rangle,\\
T_\pm|T\alpha\rangle &= \Gamma_\pm(\alpha)|T\alpha \pm 1\rangle,\\
T_\pm|\bar{T}\alpha\rangle &= \Gamma_\pm(\alpha)|\bar{T}\alpha \pm 1\rangle;\\
\Gamma_\pm(\alpha) &= [(T \mp \alpha)(T \pm \alpha + 1)]^{\frac{1}{2}},\\
\Gamma_\pm(\alpha \mp 1) &= \Gamma_\mp(\alpha).
\end{aligned} \tag{2.3}$$

[11] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, London, 1957), Chap. 3.

An explicit construction of these operators is easily given:

$$T_{\pm} = \sum_{k\alpha} \Gamma_{\mp}(\alpha)[a_{\alpha}^{*}(k)a_{\alpha\mp1}(k) + b_{\alpha}^{*}(k)b_{\alpha\mp1}(k)],$$
$$T_3 = \sum_{k\alpha} \alpha[a_{\alpha}^{*}(k)a_{\alpha}(k) + b_{\alpha}^{*}(k)b_{\alpha}(k)]. \tag{2.4}$$

Using the last of relations (2.3) one easily verifies that $(T_{\pm})^{*} = T_{\mp}$. A simpler form of the isospin operators is given as follows. Define $(2T+1)$ component vectors

$$\mathbf{a}(k) = \text{col}\,(a_{\alpha}(k)),$$
$$\mathbf{b}(k) = \text{col}\,(b_{\alpha}(k)). \tag{2.5}$$

Then, if t_i is the usual $(2T+1) \times (2T+1)$ isospin matrix, we have

$$T_i = \sum_{k} (\mathbf{a}^{\dagger}(k)t_i\mathbf{a}(k) + \mathbf{b}^{\dagger}(k)t_i\mathbf{b}(k)). \tag{2.6}$$

We cite the obvious commutation relations among the T_i, a_{α}, and b_{α}:

$$[T_i, T_j] = i\epsilon_{ijk}T_k,$$
$$[a_{\alpha}(k), T_3] = \alpha a_{\alpha}(k); \quad [b_{\alpha}(k), T_3] = \alpha b_{\alpha}(k),$$
$$[a_{\alpha}(k), T_{\pm}] = \Gamma_{\mp}(\alpha)a_{\alpha\mp1}(k); \tag{2.7}$$
$$[b_{\alpha}(k), T_{\pm}] = \Gamma_{\mp}(\alpha)b_{\alpha\mp1}(k).$$

These equations show that the a_{α}^{*} or the states $|T\alpha\rangle$ constitute a standard basis for the irreducible representation $D^{(T)}$ of the isospin group $SU(2)$. Introducing the unitary operator

$$O(\boldsymbol{\lambda}) = \exp(i\boldsymbol{\lambda}\cdot\mathbf{T}),$$
$$O(\boldsymbol{\lambda})\,|T\alpha\rangle = \sum_{\beta} |T\beta\rangle\, D_{\beta\alpha}(\boldsymbol{\lambda}), \tag{2.8}$$

where the D's are representation matrices $\exp(i\boldsymbol{\lambda}\cdot\mathbf{t})$ in the notation of Edmonds,[12] the antiparticle states $|\bar{T}\alpha\rangle$ transform exactly as $|T\alpha\rangle$ under $O(\boldsymbol{\lambda})$.

The operators a_{α}^{*}, a_{α} must transform as

$$O(\boldsymbol{\lambda})a_{\alpha}^{*}O(\boldsymbol{\lambda})^{-1} = \sum_{\beta} a_{\beta}^{*}D_{\beta\alpha}^{(T)}(\boldsymbol{\lambda}),$$
$$O^{-1}(\boldsymbol{\lambda})a_{\alpha}O(\boldsymbol{\lambda}) = \sum_{\beta} D_{\alpha\beta}^{(T)}(\boldsymbol{\lambda})a_{\beta}. \tag{2.9}$$

The antiparticle operators b_{α}^{*}, b_{α} transform exactly as a_{α}^{*}, a_{α}.

For an SCM the antiparticle multiplet is not distinct, and so there are only $2(2T+1)$ operators a_{α}^{*}, a_{α}. The description of the SCM is obtained from that of the PCM by systematically setting b equal to zero in Eqs. (2.1)–(2.9).

[12] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1960).

The 4-momentum operators are *defined* to be

$$P_{\mu} = \sum_{k\alpha} k_{\mu}(a_{\alpha}^{*}(k)a_{\alpha}(k) + b_{\alpha}^{*}(k)b_{\alpha}(k)), \quad \text{(PC)},$$
$$P_{\mu} = \sum_{k\alpha} k_{\mu}a_{\alpha}^{*}(k)a_{\alpha}(k), \quad \text{(SC)}. \tag{2.10}$$

It follows from (2.7) that P_{μ} commutes with the isospin operator. This was tacitly assumed in the labeling of the state vectors by k, T, and α. The states (or creation operators) are assumed to have the standard behavior under transformations of the Poincaré group.

B. Construction of Field Operators

In order to investigate the local behavior of a quantum-mechanical system in space and time, we introduce field operators. We require that the field operators transform according to irreducible representations of the various invariance groups under consideration [$SU(2)$, Poincaré group, etc.]. From such fields we can systematically construct all observables. Although the field itself is generally taken to be the primary concept, we prefer to regard it as the (essentially) unique construction possessing the desired group transformation properties simultaneously in the space–time coordinates x and the space of the internal symmetries (isospin, in this case). Having constructed such fields, we can investigate the consistency of various assumptions common to relativistic field theories, such as locality, causality, etc.

The positive-energy wavefunctions are

$$f_k(x) = (2\omega)^{-\frac{1}{2}}e^{-ik\cdot x}, \tag{2.11}$$

where $k\cdot x = \omega x^0 - \mathbf{k}\cdot\mathbf{x}$ defines our metric. The f_k are orthogonal within the usual Klein–Gordon inner product

$$(f_{k'}, f_k) \equiv i\int f_{k'}^{*}(x)\overleftrightarrow{\partial}_0 f_k(x)\, d^3x = \delta_{kk'}, \tag{2.12}$$

the integration running over a box of unit volume.

There are four independent operators transforming appropriately under Lorentz transformations and according to isospin rotations:

$$\chi_{1\alpha} = \sum_{k} a_{\alpha}(k)f_k(x), \quad \chi_{2\alpha}(x) = \sum_{k} b_{\alpha}(k)f_k(x),$$
$$\chi_{1\alpha}^{*} = \sum_{k} a_{\alpha}^{*}(k)f_k^{*}(x), \quad \chi_{2\alpha}^{*}(x) = \sum_{k} b_{\alpha}^{*}(k)f_k^{*}(x). \tag{2.13}$$

(Weinberg has discussed the behavior of the creation and annihilation operators under Lorentz transformations.[13]) One might also have expected objects

[13] S. Weinberg, Phys. Rev. **133B**, 1318 (1964).

like $\sum_k a_\alpha(k)f_k^*(x)$, but these do not behave properly under translations as do $\chi_{i\alpha}(x)$, $\chi_{i\alpha}^*(x)$:

$$[\chi_{i\alpha}(x), P_\mu] = i\partial_\mu\chi_{i\alpha}(x). \tag{2.14}$$

$\chi_{i\alpha}(x)$ transforms as $D^{(T)*}$, $\chi_{i\alpha}^*$ as $D^{(T)}$, under isospin rotations. This property follows trivially from (2.9).

For reasons of causality, theories are not generally described separately in terms of the $\chi_{i\alpha}$ but in terms of linear combinations of them. Clearly a superposition $\alpha\chi_1 + \beta\chi_2$ will be noncausal. Thus judicious superposition of pairs of the type (χ_1, χ_1^*), (χ_2, χ_2^*), (χ_1, χ_2^*), (χ_2, χ_1^*) with coefficients of equal magnitude is called for.

Since $\chi_{i\alpha}$ and $\chi_{i\alpha}^*$ transform only equivalently and not identically under $SU(2)$, we must transform the basis vectors in a well-known way.[14] $\chi_{i\alpha}$ transforms as $D^{(T)*}$, so $\chi'_{i\alpha}$, defined by

$$\chi'_{i\alpha} = \sum_\beta C_{\alpha\beta}\chi_{i\beta} = \bar{\eta}_\alpha\chi_{i-\alpha}, \tag{2.15}$$

$$C_{\alpha\beta} = \bar{\eta}_\alpha\delta_{\alpha,-\beta}, \quad \eta_\alpha = \xi(-1)^{T-\alpha}\,|\xi| = 1, \tag{2.16}$$

transforms as $D^{(T)}$. The phase factor ξ is arbitrary, but independent of α. We shall make repeated use of the formulas

$$\begin{aligned} \bar{\eta}_\alpha\eta_\alpha &= 1, \\ \bar{\eta}_\alpha\eta_{-\alpha} &= (-1)^{2T}, \\ \eta_{-\alpha} &= (-1)^{2T}\eta_\alpha. \end{aligned} \tag{2.17}$$

We now construct four fields transforming as $D^{(T)}$

$$\begin{aligned} \varphi_{1\alpha}^*(x) &= \chi_{1\alpha}^* + \chi'_{1\alpha} = \chi_{1\alpha}^* + \bar{\eta}_\alpha\chi_{1-\alpha}, \\ \varphi_{2\alpha}^*(x) &= \chi_{2\alpha}^* + \chi'_{2\alpha} = \chi_{2\alpha}^* + \bar{\eta}_\alpha\chi_{2-\alpha}, \\ \psi_{1\alpha}^*(x) &= \chi_{1\alpha}^* + \chi'_{2\alpha} = \chi_{1\alpha}^* + \bar{\eta}_\alpha\chi_{2-\alpha}, \\ \psi_{2\alpha}^*(x) &= \chi_{2\alpha}^* + \chi'_{1\alpha} = \chi_{2\alpha}^* + \bar{\eta}_\alpha\chi_{1-\alpha}. \end{aligned} \tag{2.18}$$

φ_1 and φ_2 are completely independent because of Eqs. (2.1), so we need only consider one of them and drop the identifying index. We thus obtain the field operator for SC multiplets:

$$\begin{aligned} \varphi_\alpha(x) &= \sum_k (a_\alpha(k)f_k(x) + \eta_\alpha a_{-\alpha}^*(k)f_k^*(x)), \\ O^{-1}(\boldsymbol{\lambda})\varphi_\alpha(x)O(\boldsymbol{\lambda}) &= \sum_\beta D_{\alpha\beta}(\boldsymbol{\lambda})\varphi_\beta(x). \end{aligned} \tag{2.19}$$

Either set, $\{\psi_{1\alpha}, \psi_{1\alpha}^*\}$ or $\{\psi_{2\alpha}, \psi_{2\alpha}^*\}$, may be used to describe PC particles. Arbitrarily choosing [as in Eq. (2.2)] the operators $\{a, a^*\}$ to correspond to the particles, we take $\psi_\alpha \equiv \psi_{1\alpha}$ to destroy "particles":

$$\psi_\alpha(x) = \sum_k (a_\alpha(k)f_k(x) + \eta_\alpha b_{-\alpha}^*(k)f_k^*(x)). \tag{2.20}$$

$\psi_\alpha(x)$ transforms identically to $\varphi_\alpha(x)$ in Eq. (2.19).

The set $\{\psi_{2\alpha}, \psi_{2\alpha}^*\}$ is not independent of the set $\{\psi_{1\alpha}, \psi_{1\alpha}^*\}$. This is shown by the relation

$$\psi_{2-\alpha}(x) = \eta_{-\alpha}(\chi_{1\alpha}^*(x) + (-1)^{2T}\bar{\eta}_\alpha\chi_{2-\alpha}(x)) \tag{2.21}$$

For integral T, the right-hand side of (2.21) is just $\eta_{-\alpha}\psi_{1\alpha}^*(x)$. For half-integral T, we can obtain equivalence by changing the sign of the b operators in the ψ_2 fields. Note that the independence of the two pieces of the field allows us to change the phase of the operators in this way. Hence the theories described by $\{\psi_{1\alpha}, \psi_{1\alpha}^*\}$ or $\{\psi_{2\alpha}, \psi_{2\alpha}^*\}$ are identical.

Thus PC bosons are described by the $2(2T + 1)$ independent operators $\{\psi_\alpha, \psi_\alpha^*\}$. It therefore seems strange that the SC bosons, which entail half as many degrees of freedom, involve the same number of field variables $\{\varphi_\alpha, \varphi_\alpha^*\}$. The latter set is not independent, however, since $\varphi_{-\alpha}$ is *not* independent of φ_α^*:

$$\varphi_{-\alpha}(x) = \eta_{-\alpha}(\chi_{1\alpha}^*(x) + (-1)^{2T}\bar{\eta}_\alpha\chi_{1-\alpha}(x)). \tag{2.22}$$

For integral T we have the expected proportionality

$$\varphi_{-\alpha}(x) = \eta_{-\alpha}\varphi_\alpha^*(x), \quad T = 0, 1, 2, \cdots. \tag{2.23}$$

For half-integral T, however, the relation is not local because of the minus sign.

The isotopic spin operators are expressed by means of the conventional isospin matrices t_i as

$$T_i = i\int d^3x \sum_{\alpha,\beta} \psi_\alpha^*(x)\overleftrightarrow{\partial}_0(t_i)_{\alpha\beta}\psi_\beta(x), \quad \text{(PC)}, \tag{2.24}$$

$$T_i = \tfrac{1}{2}i\int d^3x \sum_{\alpha,\beta} \varphi_\alpha^*(x)\overleftrightarrow{\partial}_0(t_i)_{\alpha\beta}\varphi_\beta(x), \quad \text{(SC)}. \tag{2.25}$$

One can easily check the commutation relations of these operators from those of the fields given in the next section.

3. CONNECTION BETWEEN LOCALITY AND ISOSPIN

The requirement that the field operators transform irreducibly under $SU(2)$ inevitably brings all components of the field into the theory. In particular, various observables, transforming as some tensor, need all components of the fields for their construction. In the case of SC fields the requirement of $SU(2)$ invariance thus introduces redundant fields into the theory as shown by Eq. (2.22). For integral isospin these dependent fields can be eliminated using the local

[14] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959).

relation (2.23). However, as we now show, this process is nonlocal for half-integral isospins. Thus isospin invariants like $\sum_\alpha \varphi_\alpha^*(x)\varphi_\alpha(x)$, which appear local, are nonlocal when expressed in terms of independent fields (e.g., with $\alpha > 0$).

A. Relation of $\varphi_{-\alpha}$ to φ_α^* for SC Fields

We recall the definitions of the well-known functions

$$\begin{aligned} i\Delta_+(x-y) &= \sum_k f_k(x) f_k^*(y), \\ i\Delta_-(x-y) &= -\sum_k f_k^*(x) f_k(y), \\ \Delta(x) &= \Delta_+(x) + \Delta_-(x), \\ \Delta^{(1)}(x) &= i(\Delta_+(x) - \Delta_-(x)). \end{aligned} \tag{3.1}$$

$\Delta(x)$ vanishes for spacelike x, in contrast to $\Delta^{(1)}(x)$, which only decays exponentially outside the light cone. Recall that Δ is odd, $\Delta^{(1)}$ even, and that $\dot{\Delta}\,(t=0) = -\delta(\mathbf{x})$, $\Delta(t=0) = 0$, $\dot{\Delta}^{(1)}\,(t=0) = 0$. From (3.1) and the orthogonality relations (2.12), we find that

$$\begin{aligned} i\int f_k^*(x') \overleftrightarrow{\partial}_0' \Delta(x'-x) d^3x' &= -i f_k^*(x), \\ i\int f_k^*(x') \overleftrightarrow{\partial}_0' \Delta^{(1)}(x'-x) d^3x' &= f_k^*(x). \end{aligned} \tag{3.2}$$

We thus find [cf. (2.13)] the Klein–Gordon inner products

$$\begin{aligned} (\chi_{1\alpha}(x'), \Delta(x'-x)) &= -i\chi_{1\alpha}^*(x), \\ (\chi_{1\alpha}(x'), \Delta^{(1)}(x'-x)) &= \chi_{1\alpha}^*(x). \end{aligned} \tag{3.3}$$

Thus, for integral T, (2.22) becomes

$$\begin{aligned} \varphi_{-\alpha}(x) &= \eta_{-\alpha} i([\chi_{1\alpha}(x') + \bar{\eta}_\alpha \chi_{1-\alpha}^*(x')], \Delta(x'-x)), \\ \varphi_{-\alpha}(x) &= -\eta_{-\alpha}\int \varphi_\alpha^*(x') \overleftrightarrow{\partial}_0' \Delta(x'-x) d^3x'; \\ & \qquad T = 0, 1, 2, \cdots. \end{aligned} \tag{3.4}$$

Evaluating the integral at $t' = t$ recovers Eq. (2.23).

For half-integral T the extra minus sign is compensated for by using $\Delta^{(1)}$:

$$\begin{aligned} \varphi_{-\alpha}(x) &= \eta_{-\alpha}([\chi_{1\alpha}(x') + \bar{\eta}_\alpha \chi_{1-\alpha}^*(x')], \Delta^{(1)}(x'-x)), \\ \varphi_{-\alpha}(x) &= i\eta_{-\alpha}\int \varphi_\alpha^*(x') \overleftrightarrow{\partial}_0' \Delta^{(1)}(x'-x) d^3x'; \\ & \qquad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \end{aligned} \tag{3.5}$$

Thus $\varphi_{-\alpha}(x)$ is a nonlocal superposition of $\varphi_\alpha^*(x)$. Moreover, contributions come from spacelike regions $(x-x')^2 < 0$. Setting $t = t'$ in (3.5) gives

$$\begin{aligned} \varphi_{-\alpha}(x) &= -i\eta_{-\alpha}\int \dot{\varphi}_\alpha^*(\mathbf{x}', t) \Delta^{(1)}(\mathbf{x}'-\mathbf{x}, 0) d^3x'; \\ & \qquad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \end{aligned} \tag{3.6}$$

(It is clear that the time derivative supplies the requisite minus sign.) In Eq. (3.6) the entire integral is in a spacelike region relative to x!

B. Commutation Relations

For PC fields the particle-operator commutators (2.1) and definitions (3.1) give

$$\begin{aligned} [\psi_\alpha(x), \psi_\beta(x')] &= 0, \\ [\psi_\alpha(x), \psi_\beta^*(x')] &= i\delta_{\alpha\beta}\Delta(x-x'). \end{aligned} \tag{3.7}$$

The momentum density conjugate to $\psi_\alpha(x)$ is $\dot{\psi}_\alpha^*(x)$ and the theory is entirely standard.

For SC fields the fact that the positive and negative frequency components of the field involve dependent operators brings about an entanglement for the case of half-integral isospin. We find the results

$$\begin{aligned} &[\varphi_\alpha(x), \varphi_\beta^*(x')] = i\delta_{\alpha\beta}\Delta(x-x'), \\ &[\varphi_\alpha(x), \varphi_\beta(x')] \\ &= \delta_{\alpha,-\beta}\eta_\beta \sum_k [f_k(x) f_k^*(x') - (-1)^{2T} f_k^*(x) f_k(x')]. \end{aligned} \tag{3.8}$$

Thus the second commutator is noncausal when $2T$ is an even integer:

$$\begin{aligned} &[\varphi_\alpha(x), \varphi_\beta(x')] = i\delta_{\alpha,-\beta}\eta_\beta \Delta(x-x'); \quad T = 0, 1, 2, \cdots, \\ &[\varphi_\alpha(x), \varphi_\beta(x')] = \delta_{\alpha,-\beta}\eta_\beta \Delta^{(1)}(x-x); \\ & \qquad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \end{aligned} \tag{3.9}$$

The noncommutativity of $\varphi_{-\alpha}$ with φ_α can be regarded as a consequence of the dependence of $\varphi_{-\alpha}$ on φ_α^*. [The second line of (3.9) is easily derived from (3.5) and (3.8a).]

C. Charge Density, Energy Density, and Isospin Density (Half-Integral Isospin)

The local densities of charge, isospin, and energy are expressed in terms of quadratic functions of the field variables. For these quantities to be local observables, they must all commute at spacelike separations. This is automatic for PC bosons and for integral-spin SC bosons, because the commutation function Δ vanishes outside the light cone. However, even in local theories, the charge density is not a local observable. In this section all expressions are valid for half-integral isospin.

First we verify that the energy density $\mathcal{H}(x)$ may be taken to have the usual form. From the expressions

$$\begin{aligned}
&\sum_\alpha \int \varphi_\alpha^*(x)\varphi_\alpha(x)d^3x \\
&\quad = \sum_{k\alpha} \frac{1}{2\omega_k} [2a_\alpha^*(k)a_\alpha(k) + \eta_\alpha a_\alpha^*(k)a_{-\alpha}^*(-k) \\
&\qquad + \bar{\eta}_\alpha a_{-\alpha}(k)a_\alpha(-k)], \\
&\sum_\alpha \int \dot{\varphi}_\alpha^*(x)\dot{\varphi}_\alpha(x)d^3x \\
&\quad = \sum_{k\alpha} \frac{\omega_k^2}{2\omega_k} [2a_\alpha^*(k)a_\alpha(k) - \eta_\alpha a_\alpha^*(k)a_{-\alpha}^*(-k) \\
&\qquad - \bar{\eta}_\alpha a_{-\alpha}(k)a_\alpha(-k)], \\
&\sum_\alpha \int \nabla\varphi_\alpha^*(x)\cdot\nabla\varphi_\alpha(x)d^3x \\
&\quad = \sum_{k\alpha} \frac{\mathbf{k}^2}{2\omega_k} [2a_\alpha^*(k)a_\alpha(k) + \eta_\alpha a_\alpha^*(k)a_{-\alpha}^*(-k) \\
&\qquad + \bar{\eta}_\alpha a_{-\alpha}(k)a_\alpha(-k)],
\end{aligned} \tag{3.10}$$

we find that

$$\begin{aligned}
H &= \sum_{k\alpha} \omega_k a_\alpha^*(k)a_\alpha(k) = \int \mathcal{H}(x)d^3x, \\
\mathcal{H}(x) &= \frac{1}{2}\sum_\alpha :\dot{\varphi}_\alpha^*(x)\dot{\varphi}_\alpha(x) + \nabla\varphi_\alpha^*(x) \\
&\qquad \cdot \nabla\varphi_\alpha(x) + m^2\varphi_\alpha^*(x)\varphi_\alpha(x):.
\end{aligned} \tag{3.11}$$

The charge density $\sigma(x)$ and the isospin density $\mathbf{t}(x)$:

$$\begin{aligned}
\sigma(x) &= \tfrac{1}{2}i\sum_\alpha \varphi_\alpha^*(x)\overleftrightarrow{\partial_0}\varphi_\alpha(x), \\
\mathbf{t}(x) &= \tfrac{1}{2}i\sum_{\alpha\beta} \varphi_\alpha^*(x)\overleftrightarrow{\partial_0}\mathbf{t}_{\alpha\beta}\varphi_\beta(x).
\end{aligned} \tag{3.12}$$

(For PC bosons the corresponding densities are found by removing the factor $\frac{1}{2}$ and substituting ψ_α for φ_α.) For half-integral isospins, we have the results

$$\left.\begin{aligned}
&[\mathcal{H}(x), \mathcal{H}(x')] \neq 0, \\
&[\sigma(x), \sigma(x')] \neq 0, \\
&[t_i(x), t_j(x')] \neq 0,
\end{aligned}\right\} \quad (x-x')^2 < 0, \qquad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \tag{3.13}$$

Nor does $\mathcal{H}(x)$ commute with $\sigma(x')$ or $\mathbf{t}(x')$ for spacelike separations.

These results can be verified by computing specific matrix elements. The complicated but straightforward computation is omitted here.

Although the local commutation rule

$$[t_i(x), t_j(x')]_{x_0=x_0'} \neq i\epsilon_{ijk}t_k(x)\delta(\mathbf{x}-\mathbf{x}') \tag{3.14}$$

fails in this case, the once-integrated form

$$[T_i, t_j(x)] = i\epsilon_{ijk}t_k(x) \tag{3.15}$$

is valid, as is obvious from the construction of $t_i(x)$.

Similar calculations show that $\mathcal{H}(x)$ is noncausal. We quote the additional results:

$$\begin{aligned}
[\mathbf{t}(x), \varphi_\alpha(x')] &= \tfrac{1}{2}\mathbf{t}_{\alpha\beta}\varphi_\beta(x)\overleftrightarrow{\partial_0}\Delta(x-x') \\
&\quad + \tfrac{1}{2}\eta_\alpha\mathbf{t}_{\alpha,-\beta}\varphi_\beta^*(x)\overleftrightarrow{\partial_0}\Delta^{(1)}(x-x').
\end{aligned} \tag{3.16}$$

Despite appearances, the integrated isospin operator has the correct commutator with φ_α, as is easily verified

$$\begin{aligned}
&[\mathcal{H}(x), \varphi_\alpha(x')] \\
&\quad = \frac{i}{2}\{\varphi_\alpha(x)\dot{\Delta}(x-x') + \dot{\varphi}_\alpha(x)\Delta(x-x') \\
&\qquad + \nabla\varphi_\alpha(x)\cdot\nabla\Delta(x-x')\} \\
&\qquad + \tfrac{1}{2}\eta_{-\alpha}\{\dot{\Delta}^{(1)}(x-x')\varphi_{-\alpha}^*(x) + \Delta^{(1)}(x-x')\dot{\varphi}_{-\alpha}^*(x) \\
&\qquad + \nabla\Delta^{(1)}(x-x')\cdot\nabla\varphi_{-\alpha}^*(x)\}.
\end{aligned} \tag{3.17}$$

Here the dot denotes differentiation with respect to x_0. Setting $x_0 = x_0'$ and integrating over x, we recover the equation of motion for φ_α:

$$[\varphi_\alpha(x), H] = i\dot{\varphi}_\alpha(x). \tag{3.18}$$

All these examples show the unfamiliar and even pathological features of theories of the type considered. If one postulates that relativistic quantum field theory must be local, then one can summarize our result by saying that only real representations (integral isospin) of the isospin group $SU(2)$ are permitted for self-conjugate bosons. The most serious difficulty is the failure of the energy density to commute at spacelike separations. This fact leads to difficulties with relativistic invariance.

4. RELATION TO CANONICAL FIELD THEORY

Anomalous theories of the type introduced in Ref. 1 were not discovered in conventional Lagrangian field theory. In the latter an SC multiplet of isospin T would be described by $2T+1$ Hermitian fields carrying the same mass. As such theories are really local, there was no chance of running across the anomalous SC bosons. However, the relation between the two theories is very instructive. In terms of the (redundant) set of fields $\{\varphi_\alpha, \varphi_\alpha^*\}$, all physical quantities refer to a single space–time point, for example, the energy density

$$\begin{aligned}
\mathcal{H}(x) &= \tfrac{1}{2}\sum_\alpha :\dot{\varphi}_\alpha^*(x)\dot{\varphi}_\alpha(x) + \nabla\varphi_\alpha^*(x) \\
&\qquad \cdot \nabla\varphi_\alpha(x) + m^2\varphi_\alpha^*(x)\varphi_\alpha(x):
\end{aligned} \tag{4.1}$$

and the Lagrangian density, here *defined* to be

$$\mathfrak{L}(x) = \tfrac{1}{2}\sum_\alpha :\partial_\mu\varphi_\alpha^*(x)\partial^\mu\varphi_\alpha(x) - m^2\varphi_\alpha^*(x)\varphi_\alpha(x):. \quad (4.2)$$

Only when we eliminate the dependent fields, or test commutativity at spacelike separations, do we discover the difficulties in the theory.

Before proceeding to the general case, we give a detailed comparison of the charged scalar theory and the SC isospin-$\frac{1}{2}$ theory. Both involve the same number of degrees of freedom.

A. Comparison of Charged Scalar Theory and SC Isospin-$\frac{1}{2}$ Theory

The charged scalar theory is described by a complex field:

$$\phi(x) = \sum_k (a_k f_k + b_k^* f_k^*),$$
$$\phi^*(x) = \sum_k (a_k^* f_k^* + b_k f_k). \quad (4.3)$$

Using the same a's and b's, we can construct the field φ_α ($T = \frac{1}{2}$) (choosing $\xi = 1$):

$$\varphi_{\frac{1}{2}}(x) = \sum_k (a_k f_k + b_k^* f_k^*), \quad a_k \equiv a_{\frac{1}{2}}(k),$$
$$\varphi_{-\frac{1}{2}}(x) = \sum_k (b_k f_k - a_k^* f_k^*), \quad b_k \equiv a_{-\frac{1}{2}}(k). \quad (4.4)$$

In each case the Hamiltonian is

$$H = \sum_k \omega_k(a_k^* a_k + b_k^* b_k). \quad (4.5)$$

The isospin operators are defined by

$$T_+ = \sum_k a_k^* b_k,$$
$$T_- = \sum_k b_k^* a_k, \quad (4.6)$$
$$T_3 = \tfrac{1}{2}\sum_k (a_k^* a_k - b_k^* b_k).$$

These operators commute with H, Eq. (4.5), for *either* theory. However, differences occur in local quantities. One can check in detail the result

$$[\varphi_\alpha(x), T_j] = \sum_\beta (\tfrac{1}{2}\tau_j)_{\alpha\beta}\varphi_\beta(x). \quad (4.7)$$

Thus, although $\varphi_{\frac{1}{2}} = \phi$, the operation $[\phi, T_+] = \varphi_{-\frac{1}{2}}$ produces an operator independent of ϕ and ϕ^*. Thus the *local* densities constructed from ϕ and ϕ^* will not, in general, be invariant under unitary transformations $\exp(i\boldsymbol{\lambda}\cdot\mathbf{T})$. An exception is the one-parameter subgroup $\exp(i\omega T_3)$, which is the generator of phase transformations

$$e^{i\omega T_3}\phi(x)e^{-i\omega T_3} = e^{-i\omega}\phi(x). \quad (4.8)$$

As everybody knows, this gauge transformation leaves the charged scalar theory invariant, and allows one to identify charge (Q) with T_3.

It is instructive to examine the Lagrangian density

$$\mathfrak{L}_{\text{ch sc}} = \partial_\mu\varphi^*\partial^\mu\phi - m^2\phi^*\phi,$$
$$\mathfrak{L}_{\text{SC}} = \tfrac{1}{2}\sum_\alpha(\partial_\mu\varphi_\alpha^*\partial^\mu\varphi_\alpha - m^2\varphi_\alpha^*\varphi_\alpha). \quad (4.9)$$

Introducing the operators

$$\chi_1 = \sum_k a_k f_k(x), \quad \chi_{1\mu} \equiv \partial_\mu\chi_1,$$
$$\chi_2 = \sum_k b_k f_k(x), \quad \chi_{2\mu} \equiv \partial_\mu\chi_2, \quad (4.10)$$

we find that (4.9) reduce to

$$\mathfrak{L}_{\text{ch sc}} = \mathfrak{L}_{\text{SC}} + \chi_{2\mu}\chi_1^\mu + \chi_{1\mu}^*\chi_2^{\mu *} - m^2(\chi_1\chi_2 + \chi_2^*\chi_1^*),$$
$$\mathfrak{L}_{\text{SC}} = \chi_{1\mu}^*\chi_1^\mu + \chi_{2\mu}^*\chi_2^\mu - m^2(\chi_1^*\chi_1 + \chi_2^*\chi_2). \quad (4.11)$$

$\mathfrak{L}_{\text{SC}}$ is of simpler structure and is invariant under a larger internal-symmetry group.

It is easy to show that the Lagrangian vanishes for the anomalous case

$$L_{\text{SC}} = \int \mathfrak{L}_{\text{SC}}(x)d^3x = 0. \quad (4.12)$$

The result (4.12) is true for all half-integral-isospin SC theories.

The pertinent field commutators are

$$[\phi(x), \phi^*(x')] = [\varphi_\alpha(x), \varphi_\alpha^*(x')] = i\Delta(x - x'),$$
$$[\varphi_\alpha(x), \varphi_{-\alpha}(x')] = (-1)^{2\alpha}\Delta^{(1)}(x - x'). \quad (4.13)$$

The previous analysis shows how one can proceed systematically to modify a given theory in the direction of increasing symmetry. The following sequence can be used to convert the charged scalar theory to the SC-($T = \frac{1}{2}$) theory. Begin with the fields ϕ, ϕ^*, and their operators a_k and b_k. From Schwinger[15] we know that the a's and b's can be used to construct an algebra ($SU(2)$) leaving the energy invariant:

$$H = \sum_k \xi^\dagger(k)\xi(k)\omega_k, \quad \xi(k) = \begin{pmatrix} a_k \\ b_k \end{pmatrix},$$
$$[T_i, H] = 0, \quad T_i = \sum_k \xi^\dagger(k)\tfrac{1}{2}\tau_i\xi(k). \quad (4.14)$$

These T_i coincide with those given in Eq. (4.6). Now return to the original fields and subject them to transformations of the new group. By construction, the field ϕ transforms as a representation of the group,

[15] J. Schwinger, *On Angular Momentum*, U.S. Atomic Energy Commission NYO3071, reprinted in *Quantum Theory of Angular Momentum*, L. C. Biedenharn and H. Van Dam, Eds. (Academic Press Inc., New York, 1965).

and so new fields are generated. These fields (here $\varphi_{-\frac{1}{2}}$ and $\varphi^*_{-\frac{1}{2}}$) are clearly not dynamically independent, but are needed if we modify the local densities so that they are invariant under transformations of the enlarged group. The increased symmetry is obtained, in the present case, at the expense of locality.

B. Canonical Formalism for Self-Conjugate Bosons

We have previously noted that "manifestly isospin-invariant" densities such as (4.1)–(4.2) contain dependent fields. Because of this the standard Lagrangian formalism becomes complicated. For integral isospins the problem is circumvented by eliminating the redundant fields, at the cost of manifest covariance. For half-integral isospins it is difficult (or at least ugly) to eliminate the dependent fields. Using the method of Lagrangian multipliers, one can treat the field components as independent. Before discussing this technique, we describe the formulation of the free-field theory in terms of the standard "p–q" variables.[16]

We specify the values of the operators

$$\varphi_\alpha(x), \quad \pi_\alpha(x) \equiv \dot{\varphi}^*_\alpha(x);$$
$$[\varphi_\alpha(x), \pi_\beta(x')]_{x_0=x_0'} = i\delta(\mathbf{x}-\mathbf{x}')\delta_{\alpha\beta} \tag{4.15}$$

by the discrete Fourier coefficients $q_\alpha(k)$ and $p_\alpha(k)$, defined in the usual way by

$$\varphi_\alpha(x) = \sum_k q_\alpha(k)e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{4.16a}$$

$$\pi_\alpha(x) = \sum_k p_\alpha(k)e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{4.16b}$$

$$q^*_\alpha(k) = (a_\alpha(k) + \eta_\alpha a^*_{-\alpha}(-k))/(2\omega)^{\frac{1}{2}}, \tag{4.17a}$$

$$p_\alpha(k) = \dot{q}^*_\alpha(k) = i\omega(a^*_\alpha(k) - \bar{\eta}_\alpha a_{-\alpha}(-k))/(2\omega)^{\frac{1}{2}}. \tag{4.17b}$$

The p_α and q_α are normal-mode coordinates.

From the commutation rules for a, a^*, we find

$$\begin{aligned}
&[q_\alpha(k), p_\beta(k')] = i\delta_{\alpha\beta}\delta_{kk'},\\
&[q_\alpha(k), q_\beta(k')] = -\frac{1}{2\omega}(1-(-1)^{2T})\eta_\alpha\delta_{\alpha,-\beta}\delta_{k,-k'},\\
&[p_\alpha(k), p_\beta(k')] = \tfrac{1}{2}\omega(1-(-1)^{2T})\bar{\eta}_\alpha\delta_{\alpha,-\beta}\delta_{k,-k'},\\
&[q_\alpha(k), q^*_\beta(k')] = 0,\\
&[p_\alpha(k), p^*_\beta(k')] = 0,\\
&[q^*_\alpha(k), p_\beta(k')] = \tfrac{1}{2}i\bar{\eta}_\alpha(1+(-1)^{2T})\delta_{\alpha,-\beta}\delta_{k,-k'}.
\end{aligned} \tag{4.18}$$

The various dependencies revealed by (4.18) can be understood in terms of the following relations. From the equation

$$\eta_\alpha q^*_{-\alpha}(-k) = (\eta_\alpha a^*_{-\alpha}(k) + (-1)^{2T}a_\alpha(k))/(2\omega)^{\frac{1}{2}}, \tag{4.19}$$

we find two classes of relations

$$\eta_\alpha q^*_{-\alpha}(-k) = q_\alpha(k); \qquad T = 0, 1, 2, \cdots, \tag{4.20}$$

$$\eta_\alpha q^*_{-\alpha}(-k) = \dot{q}_\alpha(k)/(i\omega); \quad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \tag{4.21}$$

Eq. (4.21) expresses the content of Eq. (3.6), as can be seen by substituting (4.16a) in the latter.

Notice that (4.7) can be put in the form

$$q_{-\alpha}(-k) = i\eta_\alpha p_\alpha(k)/\omega. \tag{4.22}$$

The difference in the commutation rules (4.18) is simply a consequence of the basic distinction revealed in Eqs. (4.20)–(4.21). For integral T, $q_{-\alpha}(-k)$ is proportional to another *coordinate* $q^*_\alpha(k)$ of opposite quantum numbers; while for half-integral T, $q_{-\alpha}(-k)$ is proportional to a *momentum* $p_\alpha(k)$.

From this point of view there is no fundamental virtue of the integral-isospin case, for which all the information in Eqs. (4.18) reduces to

$$\begin{aligned}
&[q_\alpha(k), q_\beta(k')] = [p_\alpha(k), p_\beta(k')] = 0,\\
&[q_\alpha(k), p_\beta(k')] = i\delta_{\alpha\beta}\delta_{kk'},\\
&\eta_\alpha q^*_{-\alpha}(-k) = q_\alpha(k); \quad T = 0, 1, 2, \cdots.
\end{aligned} \tag{4.23}$$

For half-integral isospin we have, instead of (4.23),

$$\begin{aligned}
&[q_\alpha(k), p_\beta(k')] = i\delta_{\alpha\beta}\delta_{kk'},\\
&[q^*_\alpha(k), p_\beta(k')] = [q_\alpha(k), p^*_\beta(k')] = 0,\\
&q_{-\alpha}(-k) = i\eta_\alpha p_\alpha(k)/\omega; \quad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots.
\end{aligned} \tag{4.24}$$

Using (4.17), we see that the Hamiltonian has the usual form

$$\begin{aligned}
H &= \sum_{k\alpha} \tfrac{1}{2}(p^*_\alpha(k)p_\alpha(k) + \omega_k^2 q^*_\alpha(k)q_\alpha(k))\\
&= \sum_{k\alpha} \omega_k a^*_\alpha(k)a_\alpha(k).
\end{aligned} \tag{4.25}$$

(As elsewhere, we normal order the a_α.) Either expression in (4.25) is equal to the space integral of $\mathcal{H}(x)$ in Eq. (4.1) [cf. Eq. (3.11)].

We can now express H in terms of independent p and q variables. From (4.20) (and a similar expression for p_α) we learn that for integral isospin

$$\begin{aligned}
&\sum_{k,\alpha=-1}^{-T} p^*_\alpha(k)p_\alpha(k) = \sum_{k,\alpha=1}^{T} p^*_\alpha(k)p_\alpha(k),\\
&\sum_{k,\alpha=-1}^{-T} \omega_k^2 q^*_\alpha(k)q_\alpha(k) = \sum_{k,\alpha=1}^{T} \omega_k^2 q^*_\alpha(k)q_\alpha(k),\\
&H = \sum_k \Big\{ \tfrac{1}{2}(p^*_0(k)p_0(k) + \omega_k^2 q^*_0(k)q_0(k))\\
&\qquad + \sum_{\alpha=1}^{T}(p^*_\alpha(k)p_\alpha(k) + \omega_k^2 q^*_\alpha(k)q_\alpha(k)) \Big\}.
\end{aligned} \tag{4.26}$$

[16] See, for example, G. Wentzel, *Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1949).

For half-integral isospin we find from (4.22)

$$\sum_{k,\alpha=-\frac{1}{2}}^{-T} p_\alpha^*(k)p_\alpha(k) = \sum_{k,\alpha=\frac{1}{2}}^{T} \omega_k^2 q_\alpha^*(k)q_\alpha(k),$$
$$\sum_{k,\alpha=-\frac{1}{2}}^{-T} \omega_k^2 q_\alpha^*(k)q_\alpha(k) = \sum_{k,\alpha=\frac{1}{2}}^{T} p_\alpha^*(k)p_\alpha(k), \quad (4.27)$$
$$H = \sum_{k,\alpha=\frac{1}{2}}^{T} (p_\alpha^*(k)p_\alpha(k) + \omega_k^2 q_\alpha^*(k)q_\alpha(k)).$$

Eq. (4.26) has the usual form. (4.27) looks so innocuous that one would never suspect the unusual features the theory has when supplied with the commutation rules (4.24) and the field $\varphi_\alpha(x)$ defined by (4.16).

Next consider the Lagrangian. We define it in a standard way by

$$L = \sum_{k,\alpha} p_\alpha(k)\dot{q}_\alpha(k) - H$$
$$= \tfrac{1}{2}\sum_{k\alpha}(p_\alpha^*(k)p_\alpha(k) - \omega_k^2 q_\alpha^*(k)q_\alpha(k)). \quad (4.28)$$

In terms of creation and annihilation operators, (4.28) is

$$L = -\tfrac{1}{2}\sum_{k\alpha} \omega_k(\eta_\alpha a_\alpha^*(k)a_{-\alpha}(-k) + \bar{\eta}_\alpha a_\alpha(k)a_{-\alpha}(-k))$$
$$= \int \mathfrak{L}(x)d^3x, \quad (4.29)$$

where $\mathfrak{L}(x)$ is given by (4.2).

Now we find a basic distinction between integral and half-integral isospin. From Eqs. (4.26) we see that for integral isospin

$$L = \sum_k \Big\{\tfrac{1}{2}(p_0^*(k)p_0(k) - \omega_k^2 q_0^*(k)q_0(k))$$
$$+ \sum_{\alpha=1}^{T}(p_\alpha^*(k)p_\alpha(k) - \omega_k^2 q_\alpha^*(k)q_\alpha(k))\Big\}, \quad (4.30)$$

while Eqs. (4.27) show that

$$L = 0; \quad T = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \quad (4.31)$$

The result (4.31) generalizes the special case (4.12) to all half-integral isospins.

Finally, we ask whether the equations of motion can be derived from the variational method. This would be trivially true if the fields entering into $\mathfrak{L}(x)$, Eq. (4.2), were independent. The fields can be treated independently if we use the method of Lagrange multipliers. Define the modified Lagrange function

$$\mathfrak{L}'(x) = \mathfrak{L}(x) + \lambda\Phi_1(x) + \mu\Phi_2(x), \quad (4.32)$$

where λ, μ are constants and $\Phi_i(x)$ are (isospin scalar) positive-definite functions:

$$\Phi_1 = \sum_\alpha \Phi_\alpha^*(x)\Phi_\alpha(x),$$
$$\Phi_2 = \sum_\alpha \Phi_{\alpha\mu}^*(x)\Phi_\alpha^\mu(x), \quad \Phi_{\alpha\mu}(x) \equiv \partial_{\mu\alpha}\Phi_\alpha(x), \quad (4.33)$$
$$\Phi_\alpha(x) \equiv \varphi_{-\alpha}(x) + i\eta_{-\alpha}\int \dot{\varphi}_\alpha^*(\mathbf{x}', t)\Delta^{(1)}(\mathbf{x}' - \mathbf{x}, 0)d^3x'.$$

Since each term in Φ_i is positive-definite, the subsidiary conditions

$$\int \Phi_i(x)d^3x = 0 \quad (4.34)$$

imply the vanishing of Φ_α, $\Phi_{\alpha\mu}$, which, in turn, enforces the dependency relation (3.6). Hence the variation of $\mathfrak{L}'$ due to $\delta\varphi_\alpha$, say, is

$$\delta\mathfrak{L}'(x) = \delta\mathfrak{L}(x) + \Big[\sum_\alpha (\lambda\Phi_\alpha^*\delta\Phi_\alpha + \mu\Phi_{\alpha\mu}^*\delta\Phi_{\alpha\mu}) + \text{H.c.}\Big]$$
$$= \delta\mathfrak{L}(x), \quad (4.35)$$

by virtue of (4.34). Hence the equation of motion

$$(\Box + m^2)\varphi_\alpha(x) = 0 \quad (4.36)$$

follows as usual.

5. ANTIPARTICLE CONJUGATION FOR ISOSPIN MULTIPLETS

Experience shows that many laws of nature are invariant under antiparticle conjugation. This seems to be the case for the strong and electromagnetic interactions. Even when this operation is not a symmetry, a proper formulation of the transformation is essential to study the asymmetry itself. Our definition is certainly based on the type of transformation which leads to a symmetry, namely, that to the behavior of a particle state with four-momentum k_μ, $(k^2 = m^2)$, additive quantum numbers $\{\nu\}$ (such as T_3, charge, hypercharge, $\cdots$) corresponds an identical evolution of an antiparticle state with four momentum k_μ and opposite additive quantum numbers $\{-\nu\}$. Here we only consider isospin variables, so that the set of variables describing a particle state is (k, T, α). The general approach has already been sketched in the appendix to Ref. 10.

As in Sec. 2, we begin with the particle approach, defining a unitary operator C which maps particle operators into antiparticle operators, preserving the commutator algebra. Our discussion of the PC case is standard, except that a careful treatment is given of the phase factors which occur. In addition, we exhibit explicitly[17] a continuous transformation

[17] See Appendix A.

(connected to the identity) of which C is a special case. By requiring that C map causal fields into causal fields (all having definite isospin-transformation properties), we can restrict the phase factors entering the theory. Virtues of different phase choices are discussed. G conjugation is also discussed.

For SC bosons two possible field transformations are possible. One, C_1, maps φ_α into φ_α^* in a conventional way; the second, C_2, maps φ_α into $\varphi_{-\alpha}$. The latter transformation is therefore nonlocal for half-integral isospin. However, we exhibit a unitary operator (nonlocal) which converts C_1 to C_2, so that the two cases can, in a sense, be regarded as equivalent. A general proof is given that the G parity of a self-conjugate, integral-isospin boson is $\eta_C(-1)^T$, where η_C is the charge parity of the $T_3 = 0$ member of the multiplet.

A. Antiparticle Conjugation for PC Isospin Multiplets

The antiparticle transformation C maps $|T\alpha\rangle$ onto $|\bar{T} - \alpha\rangle$ and vice versa. To be useful, we also require that the algebra of the field operators ψ_α be preserved under the transformation.

The most general possibility is therefore

$$\begin{aligned} C\,|T\alpha\rangle &= \lambda_\alpha\,|\bar{T} - \alpha\rangle, \\ C\,|\bar{T}\alpha\rangle &= \mu_\alpha\,|T - \alpha\rangle, \end{aligned} \tag{5.1}$$

where λ_α and μ_α depend on α. The a's and b's have to transform as

$$\begin{aligned} Ca_\alpha^* C^{-1} &= \lambda_\alpha b_{-\alpha}^*, \\ Cb_\alpha^* C^{-1} &= \mu_\alpha a_{-\alpha}^*. \end{aligned} \tag{5.2}$$

We require that C be unitary, which preserves the operator structure and requires that

$$\begin{aligned} &\bar{\lambda}_\alpha\lambda_\alpha = \bar{\mu}_\alpha\mu_\alpha = 1, \\ &[a_\alpha, a_\beta^*] \underset{C}{\to} \bar{\lambda}_\alpha\lambda_\beta[b_{-\alpha}^*, b_{-\beta}] = \delta_{\alpha\beta}. \end{aligned} \tag{5.3}$$

Two applications of C yields

$$\begin{aligned} C^2 a_\alpha^* C^{-2} &= \lambda_\alpha\mu_{-\alpha} a_\alpha^*, \\ C^2 b_\alpha^* C^{-2} &= \mu_\alpha\lambda_{-\alpha} b_\alpha^*. \end{aligned} \tag{5.4}$$

From this we cannot yet conclude that $C^2 = \lambda I$, since $\lambda_\alpha\mu_{-\alpha}$ is not 1, *a priori*. Consider the effect of C on ψ_α:

$$C\psi_\alpha(x)C^{-1} = \sum_k \bar{\lambda}_\alpha b_{-\alpha}(k) f_k(x) + \eta_\alpha\mu_{-\alpha} a_\alpha^*(k) f_k(x). \tag{5.5}$$

We wish to formulate our theory in terms of fields with definite transformation properties under $SU(2)$. There are only four which will be local: ψ_α, ψ_α^*, $\psi_{-\alpha}$, $\psi_{-\alpha}^*$. Inspection shows that only ψ_α^* has the right structure to qualify:

$$\psi_\alpha^*(x) = \sum_k a_\alpha^*(k) f_k^*(x) + \bar{\eta}_\alpha b_{-\alpha}(k) f_k(x). \tag{5.6}$$

Requiring $C\psi_\alpha C^{-1} = \zeta_\alpha\psi_\alpha^*$ gives the constraints

$$\begin{aligned} \zeta_\alpha &= \eta_\alpha\mu_{-\alpha}, \\ \bar{\eta}_\alpha\zeta_\alpha &= \bar{\lambda}_\alpha. \end{aligned} \tag{5.7}$$

Elimination of ζ_α gives (remember $\bar{\eta}_\alpha\eta_\alpha = 1$)

$$\mu_{-\alpha} = \bar{\lambda}_\alpha, \quad \zeta_\alpha = \eta_\alpha\bar{\lambda}_\alpha. \tag{5.8}$$

Therefore (5.4) simplifies to the desired form ($\lambda_\alpha\mu_{-\alpha} = 1$):

$$\begin{aligned} C^2 a_\alpha^* C^{-2} &= a_\alpha^*, \\ C^2 b_\alpha^* C^{-2} &= b_\alpha^*, \\ C^2 &= \lambda I, \end{aligned} \tag{5.9}$$

the last step following from Schur's lemma. λ can now be shown to be unity. From Eqs. (5.8) and (5.1)

$$\begin{aligned} C\,|T\alpha\rangle &= \lambda_\alpha\,|\bar{T} - \alpha\rangle, \\ C\,|\bar{T}\alpha\rangle &= \bar{\lambda}_{-\alpha}\,|T - \alpha\rangle, \\ C^2\,|T\alpha\rangle &= \lambda_\alpha C\,|\bar{T} - \alpha\rangle = \lambda_\alpha\bar{\lambda}_\alpha\,|T\alpha\rangle \\ &= |T\alpha\rangle. \end{aligned} \tag{5.10}$$

To summarize, the preservation of the operator algebra and state vector norms require $C^{-1} = C^\dagger$. Further requiring that C map ψ_α linearly to ψ_α^* implied $C^2 = 1$, so $C = C^{-1}$:

$$\begin{aligned} C &= C^{-1} = C^\dagger, \\ Ca_\alpha^* C^{-1} &= \lambda_\alpha b_{-\alpha}^*, \\ Cb_\alpha^* C^{-1} &= \bar{\lambda}_{-\alpha} a_{-\alpha}^*. \end{aligned} \tag{5.11}$$

An explicit form for such an operator C is given in Appendix A. ψ_α transforms as

$$\begin{aligned} \psi_\alpha^C &= C\psi_\alpha C^{-1} = \zeta_\alpha\psi_\alpha^*, \\ [\psi_\alpha^C(x), \psi_\beta^C(x')] &= \zeta_\alpha\bar{\zeta}_\beta[\psi_\alpha^*(x), \psi_\beta(x')] \\ &= -i\delta_{\alpha\beta}\Delta(x' - x) = i\delta_{\alpha\beta}\Delta(x - x'). \end{aligned} \tag{5.12}$$

Eq. (5.12) verifies the preservation of the equal-time commutation relations under C.

A convenient choice of λ_α is simply $\eta_C\eta_\alpha$, where η_C is arbitrary. (Thus we may take it to be unity.) This makes ψ_α transform nicely, but the particle operators inherit a phase.

Case 1: $\lambda_\alpha = \eta_\alpha\bar{\eta}_C$

$$\begin{aligned} C\psi_\alpha C^{-1} &= \eta_C\psi_\alpha^*, \\ Ca_\alpha^* C^{-1} &= \bar{\eta}_C\eta_\alpha b_{-\alpha}^*, \\ Cb_\alpha^* C^{-1} &= \eta_C\bar{\eta}_{-\alpha} a_{-\alpha}^*. \end{aligned} \tag{5.13}$$

The isospin operator transforms as

$$\begin{aligned} CT_kC^{-1} &= (\psi^*, t_k\psi^*) = -(\psi, \tilde{t}_k\psi), \\ CT_1C^{-1} &= -T_1, \\ CT_2C^{-1} &= +T_2, \\ CT_3C^{-1} &= -T_3. \end{aligned} \tag{5.14}$$

This result is also easily obtained from Eqs. (1.7). $\tilde{t}_k$ denotes the transpose of t_k. C is *not* the same as rotation by π about the 2 axis, although its effect on $\mathbf{T}$ is the same.

This brings us to G parity. Clearly, $R_2 = \exp(i\pi T_2)$ preceding C will give $\mathbf{T} \to \mathbf{T}$ under $G = CR_2$:

$$\begin{aligned} G\mathbf{T}G^{-1} &= \mathbf{T}, \\ G &\equiv CR_2. \end{aligned} \tag{5.15}$$

Thus G maps isospin multiplets into antiparticle isospin multiplets (here we choose $\eta_C = 1$):

$$R_2\begin{Bmatrix} a_\alpha^* \\ b_\alpha^* \end{Bmatrix}R_2^{-1} = (-1)^{T+\alpha}\begin{Bmatrix} a_{-\alpha}^* \\ b_{-\alpha}^* \end{Bmatrix}, \tag{5.16}$$

$$\begin{aligned} G\begin{Bmatrix} a_\alpha^* \\ b_\alpha^* \end{Bmatrix}G^{-1} &= \eta_{-\alpha}(-1)^{T+\alpha}\begin{Bmatrix} b_\alpha^* \\ (-1)^{2T}a_\alpha^* \end{Bmatrix} \\ &= \xi\begin{Bmatrix} b_\alpha^* \\ (-1)^{2T}a_\alpha^* \end{Bmatrix}. \end{aligned} \tag{5.17}$$

Thus if we choose $\xi = 1$ in η_α [Eq. (2.16)] the G conjugation simply interchanges the a and b operators:

$$\begin{aligned} Ga_\alpha G^{-1} &= b_\alpha, \\ Gb_\alpha G^{-1} &= a_\alpha(-1)^{2T}, \\ G\psi_\alpha G^{-1} &= (-1)^{T+\alpha}\psi_{-\alpha}^*. \end{aligned} \tag{5.18}$$

With these phase conventions, C has a simple effect on the field $\psi_\alpha(x)$, while G has a simple effect on the particle and isospin operators.

Case 2: $\lambda_\alpha = 1$. Choosing λ_α to be unity makes the particle states transform simply under C':

$$\begin{aligned} C'a_\alpha^*C'^{-1} = b_{-\alpha}^*, \quad C'b_\alpha^*C'^{-1} = a_{-\alpha}^*, \\ C'\psi_\alpha C'^{-1} = \eta_\alpha\psi_\alpha^*. \end{aligned} \tag{5.19}$$

Now the particle operators transform simply, while the field carries the phase factor η_α. The effect on the isospin operator is now distinct from Eq. (5.14):

$$\begin{aligned} C'T_1C'^{-1} &= T_1, \\ C'T_2C'^{-1} &= -T_2, \\ C'T_3C'^{-1} &= -T_3. \end{aligned} \tag{5.20}$$

In terms of C' we have to define a different G' by

$$G' = C'R_1, \quad R_1 = \exp(i\pi T_1), \tag{5.21}$$

so that G' commutes with $\mathbf{T}$:

$$G'\mathbf{T}G'^{-1} = \mathbf{T}. \tag{5.22}$$

The matrix elements of R_1 can be found from the expressions

$$\begin{aligned} e^{i\pi T_1} &= e^{i(\pi/2)T_3}e^{i\pi T_2}e^{-i(\pi/2)T_3}, \\ \langle T\alpha|\exp(i\pi T_2)|T\alpha'\rangle &= \delta_{\alpha,-\alpha'}(-1)^{T-\alpha}, \\ \langle T\alpha|\exp(i\pi T_1)|T\alpha'\rangle &= i^{2T}\delta_{\alpha,-\alpha'}. \end{aligned} \tag{5.23}$$

Some elementary manipulations then give the effect of G' on the important operators in the theory:

$$\begin{aligned} G'a_\alpha^*G'^{-1} &= i^{2T}b_{-\alpha}^*, \\ G'b_\alpha^*G'^{-1} &= i^{2T}a_{-\alpha}^*, \\ G'\psi_\alpha^*G'^{-1} &= i^{2\alpha}\psi_{-\alpha}^*. \end{aligned} \tag{5.24}$$

Using equations from Appendix B, we can change the phase at will. We recommend the "standard" choice summarized in Eqs. (5.11)–(5.13).

B. Antiparticle Conjugation for SC Isospin Multiplets

For SC multiplets the antiparticle transformation must satisfy

$$\begin{aligned} C|T\alpha\rangle &= \lambda_\alpha|T-\alpha\rangle, \\ Ca_\alpha^*C^{-1} &= \lambda_\alpha a_{-\alpha}^*. \end{aligned} \tag{5.25}$$

Again, λ_α obeys

$$\bar{\lambda}_\alpha\lambda_\alpha = 1, \tag{5.26}$$

and the particle-operator algebra is preserved by (5.25).

Two applications of C yields

$$C^2a_\alpha^*C^{-2} = \lambda_\alpha\lambda_{-\alpha}a_\alpha^*. \tag{5.27}$$

In order to determine the allowed values of λ_α, we require that C map φ_α into a field having definite isospin transformation properties under $SU(2)$. The independent fields which could emerge from $C\varphi_\alpha C^{-1}$ are (for fixed α) φ_α, $\varphi_{-\alpha}$, φ_α^*, $\varphi_{-\alpha}^*$.

For convenience we introduce two operators (independent for fixed α):

$$\begin{aligned} \chi_1 &= \sum_k a_\alpha(k)f_k(x) = \chi_{1\alpha}, \\ \chi_2 &= \sum_k a_{-\alpha}(k)f_k(x) = \chi_{1-\alpha}. \end{aligned} \tag{5.28}$$

The functions $\chi_{1\alpha}$ were previously defined in Eq. (2.13). Under C, the χ_i and their conjugates transform as

$$\begin{aligned} C\chi_1C^{-1} &= \bar{\lambda}_\alpha\chi_2, \\ C\chi_2C^{-1} &= \bar{\lambda}_{-\alpha}\chi_1, \\ C\chi_1^*C^{-1} &= \lambda_\alpha\chi_2^*, \\ C\chi_2^*C^{-1} &= \lambda_{-\alpha}\chi_1^*. \end{aligned} \tag{5.29}$$

The various φ_α are given in terms of the χ_i by

$$\begin{aligned}\varphi_\alpha &= \chi_1 + \eta_\alpha \chi_2^*,\\ \varphi_\alpha^* &= \chi_1^* + \bar{\eta}_\alpha \chi_2,\\ \varphi_{-\alpha} &= \chi_2 + \eta_{-\alpha}\chi_1^*,\\ \varphi_{-\alpha}^* &= \chi_2^* + \bar{\eta}_{-\alpha}\chi_1.\end{aligned} \tag{5.30}$$

From (5.29)–(5.30) we find the result

$$C\varphi_\alpha C^{-1} = \bar{\lambda}_\alpha \chi_2 + \eta_\alpha \lambda_{-\alpha}\chi_1^*. \tag{5.31}$$

Clearly, the right-hand side of (5.31) is independent of φ_α and $\varphi_{-\alpha}^*$. We next see whether λ_α can be adjusted so that (5.31) is proportional to φ_α^* or $\varphi_{-\alpha}$.

Thus we see whether we can obtain

$$C_1\varphi_\alpha C_1^{-1} = \sigma_\alpha \varphi_\alpha^* \quad \text{(Case 1)} \tag{5.32}$$

or

$$C_2\varphi_\alpha C_2^{-1} = \rho_\alpha \varphi_{-\alpha}, \quad \text{(Case 2)} \tag{5.33}$$

where $|\sigma_\alpha| = |\rho_\alpha| = 1$.

$$\mathrm{B}_1\colon\ C_1\varphi_\alpha C_1^{-1} = \sigma_\alpha \varphi_\alpha^*,\ C^2 = 1$$

In Case 1 we have to satisfy

$$\begin{aligned}\sigma_\alpha &= \eta_\alpha \lambda_{-\alpha}, \quad \text{(Case 1)}\\ \bar{\eta}_\alpha \sigma_\alpha &= \bar{\lambda}_\alpha,\end{aligned} \tag{5.34}$$

in direct analogy to Eq. (5.7). Elimination of σ_α gives the constraint

$$\lambda_{-\alpha} = \bar{\lambda}_\alpha, \quad \text{(Case 1)} \tag{5.35}$$

which then implies $\lambda_\alpha \lambda_{-\alpha} = 1$, so that from Schur's lemma Eq. (5.27) gives

$$C_1^2 = \lambda I \quad \text{(Case 1)} \tag{5.36}$$

($|\lambda| = 1$). That $\lambda = 1$ follows from Eqs. (5.25) and (5.35):

$$\begin{aligned}C_1^2 |T\alpha\rangle &= \lambda_\alpha C_1 |T - \alpha\rangle = \lambda_\alpha \lambda_{-\alpha} |T\alpha\rangle\\ &= |T\alpha\rangle.\end{aligned} \tag{5.37}$$

Because of Eq. (5.35) λ_α has to satisfy

$$\lambda_\alpha \lambda_{-\alpha} = 1. \tag{5.38}$$

As it is awkward to consider integral and half-integral multiplets simultaneously, we consider these cases separately.

Integral Isospin. In this case there is always a member with $T_3 = 0$. Thus from (5.38) we see that the charge parity η_c is ± 1:

$$\begin{aligned}\eta_c &= \lambda_0 = \pm 1,\\ C|T0\rangle &= \eta_c |T0\rangle.\end{aligned} \tag{5.39}$$

There are many solutions of (5.38). Whole families can be generated using the unitary transformations of Appendix B. As yet we have no physical interpretation of this freedom. One simple choice is to take $\lambda_\alpha = \eta_c$, independent of α. This leads to a simple transformation for the particle operators, but then the fields transform in an ugly way [see Eq. (5.19)]. Instead, a useful choice is

$$\lambda_\alpha = \eta_c(-1)^\alpha. \tag{5.40}$$

Choosing η_α real and equal to $(-1)^\alpha$ [Eq. (5.34)] gives $\sigma_\alpha = \eta_c$ independent of α, so that

$$\begin{aligned}C_1 a_\alpha^* C_1^{-1} &= (-1)^\alpha \eta_c a_{-\alpha}^*,\\ C_1\varphi_\alpha C_1^{-1} &= \eta_c \varphi_\alpha^*.\end{aligned} \tag{5.41}$$

The components of isospin transform under C just as they did in Eq. (5.13). Thus Eq. (5.15) is an appropriate definition for G conjugation. A calculation similar to that of Eqs. (5.16)–(5.17) gives

$$Ga_\alpha^* G^{-1} = \eta_c(-1)^T a_\alpha^*. \tag{5.42}$$

Thus the states $|T\alpha\rangle$ are eigenstates of G with G parity $\eta_C(-1)^T$, an oft-quoted and little-proved result:

$$\begin{aligned}G|T\alpha\rangle &= \eta_G |T\alpha\rangle,\\ \eta_G &= \eta_C(-1)^T.\end{aligned} \tag{5.43}$$

The field φ_α transforms under G as

$$G\varphi_\alpha G^{-1} = \eta_G \varphi_\alpha \tag{5.44}$$

on using (2.23) and recalling our choice $\eta_\alpha = (-1)^\alpha$. The use of η_G as a nontrivial multiplicative quantum number is well known.

Half-Integral Isospin. We now turn to the anomalous case. As in the PC case, an α-independent solution of Eq. (5.38) makes φ_α transform with an α-dependent phase. Therefore we satisfy (5.38) by generalizing (5.40) to $\eta_c(-i)^{2\alpha}$ for half-integral α. (Again η_c is ± 1.) We then find the results

$$\begin{aligned}\lambda_\alpha &= \eta_c(-i)^{2\alpha} = \eta_c e^{-i\pi\alpha},\\ Ca_\alpha C^{-1} &= i^{2\alpha}\eta_c a_{-\alpha},\\ C\varphi_\alpha C^{-1} &= i^{2T}\eta_c \varphi_\alpha^*.\end{aligned} \tag{5.45}$$

Here we have chosen η_α to be $(-1)^{T-\alpha}$. The isospin operator transforms under C in the usual way [Eq. (5.13)], and so G is defined by Eq. (5.15). Under G, we have

$$\begin{aligned}Ga_\alpha G^{-1} &= i^{2T}\eta_c a_\alpha,\\ G\varphi_\alpha G^{-1} &= i^{2\alpha}\eta_c \varphi_{-\alpha}^*.\end{aligned} \tag{5.46}$$

$$B_2: C_2\varphi_\alpha C_2^{-1} = \rho_\alpha\varphi_{-\alpha}$$

We now turn to alternative two of Eq. (5.32). In this case the coefficients are subject to the conditions

$$\eta_\alpha\lambda_{-\alpha} = \eta_{-\alpha}\rho_\alpha,$$
$$\rho_\alpha = \bar{\lambda}_\alpha. \tag{5.47}$$

Elimination of ρ_α leads to the condition

$$\lambda_\alpha\lambda_{-\alpha} = (-1)^{2T}. \tag{5.48}$$

An immediate consequence of this is that C_2^2 is not a multiple of the identity [cf. Eq. (5.37)]

$$C_2^2 = (-1)^{2T}. \tag{5.49}$$

These results suggest that this C can be represented by a rotation $R_2 = \exp(i\pi T_2)$ and a phase adjustment. However, first we dispose of the integral-isospin case, which leads to nothing of interest.

Integral Isospin. In this case (5.48) reduces to (5.38). Choosing $\lambda_\alpha = \eta_c(-1)^\alpha$ as before, we find

$$C_2 a_\alpha^* C_2^{-1} = \eta_c(-1)^\alpha a_{-\alpha}^*,$$
$$C_2\varphi_\alpha C_2^{-1} = \eta_c(-1)^\alpha\varphi_{-\alpha} = \eta_c\varphi_\alpha^*(x), \tag{5.50}$$

the last step following from the convenient choice and Eq. (5.30). There is now no distinction from Case 1.

Half-Integral Isospin. The anomalous fields are more interesting. We solve Eq. (5.48) by the choice

$$\lambda_\alpha = \eta_c(-1)^{T+\alpha}, \tag{5.51}$$

where $\eta_c = \pm 1$. The various quantities transform as

$$C_2 a_\alpha^* C_2^{-1} = \eta_c(-1)^{T+\alpha} a_{-\alpha}^*,$$
$$C_2\varphi_\alpha C_2^{-1} = \eta_c(-1)^{T+\alpha}\varphi_{-\alpha}, \tag{5.52}$$
$$C_2^2 = -1.$$

An explicit formula for C_2 is

$$C_2 = C_0 W, \tag{5.53}$$

where C_0 and W are unitary operators defined in Eqs. (B10) and (B17). [$\omega = 0$ or π, depending on whether η_c is $+1$ or -1; ξ_α is $(-1)^{T+\alpha}$ in the formula for W.] From Eqs. (5.52) it is clear that an (equivalent) expression is simply a rotation by π about the 2 axis followed by a phase transformation:

$$C_2 = \exp\left[\frac{i\pi N}{2}(1 - \eta_c)\right]\exp(i\pi T_2). \tag{5.54}$$

According to a general theorem, Eqs. (5.53) and (5.54) are equal to a phase.

Using the results of Appendix B, we can construct a connection between the operators C_1 of Eq. (5.45) and C_2 of Eq. (5.52):

$$C_2 = VC_1,$$
$$V = \exp(-i\pi NT). \tag{5.55}$$

Thus we cannot claim that C_1 and C_2 are essentially different. However, the nonlocal peculiarities of the theory are brought out once again if we inspect Eq. (5.52) more closely. Employing Eqs. (3.5) to express $\varphi_{-\alpha}$ in terms of φ_α^*, we find that

$$C_2\varphi_\alpha(x)C_2^{-1} = \eta_c i\int\varphi_\alpha^*(x')\overleftrightarrow{\partial}_0'\Delta^{(1)}(x' - x)\,d^3x', \tag{5.56a}$$

$$C_1\varphi_\alpha(x)C_1^{-1} = \eta_c i^{2T}\varphi_\alpha^*(x), \tag{5.56b}$$

where we have chosen η_α to be $(-1)^{T-\alpha}$ in arriving at the first line of (5.56). Thus C_2 is a nonlocal operation, and the operator V, which converts (5.56a) to (5.56b) is a nonlocal transformation. In configuration space, C_2 produces a smear in the region within a pion Compton wavelength of the light cone, in contrast to the local character of C_1.

6. *CPT* TRANSFORMATION FOR ISOSPIN MULTIPLETS

The difference between abnormal and normal particles also shows up in the CPT transformation, called Θ. The abnormal particles behave differently under Θ than normal particles, for which locality plays a role[18] in establishing the "normal" Θ operation. As pointed out by Kantor[7] and by Zumino and Zwanziger,[6] this property may be used to give a simple (though physically indirect) demonstration of the incompatibility of the abnormal field theories with the existence of a normal *CPT* transformation.

A. *CPT* for Pair-Conjugate Bosons

We choose η_α real. The parity transformation is

$$Pa_\alpha(k)P^{-1} = \eta_P a_\alpha(-k),$$
$$Pb_\alpha(k)P^{-1} = \bar{\eta}_P b_\alpha(-k), \tag{6.1}$$
$$P\psi_\alpha(\mathbf{x}, t)P^{-1} = \eta_P\psi_\alpha(-\mathbf{x}, t),$$

where $\bar{\eta}_P\eta_P = 1$. [In Eqs. (6.1) the symbol $-k$ means $(\omega, -\mathbf{k})$.] Time inversion is represented by an anti-unitary operator:

$$Ta_\alpha(k)T^{-1} = \eta_T a_\alpha(-k),$$
$$Tb_\alpha(k)T^{-1} = \bar{\eta}_T b_\alpha(-k), \tag{6.2}$$
$$T\psi_\alpha(\mathbf{x}, t)T^{-1} = \eta_T\psi_\alpha(\mathbf{x}_1 - t).$$

[18] See R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964).

TABLE I. The behavior of various pair-conjugate field operators under C, P, T, $G = C\exp(i\pi T_2)$, $\Theta = CPT$, and $\tilde{\Theta} = GPT$, is listed. The phase η_α is chosen as $(-1)^{T-\alpha}$. The $\pm$ signs indicate whether the indicated component of isospin T_i changes under the transformation in question.

	$a_\alpha(k)$	$b_\alpha(k)$	$\psi_\alpha(x)$	T_1	T_2	T_3
C	$\eta_\alpha\eta_C b_{-\alpha}(k)$	$\eta_{-\alpha}\bar{\eta}_C a^*_{-\alpha}(k)$	$\eta_C\psi^*_\alpha(x)$	$-$	$+$	$-$
P	$\eta_P a_\alpha(-k)$	$\bar{\eta}_P b_\alpha(-k)$	$\eta_P\psi_\alpha(-\mathbf{x}, t)$	$+$	$+$	$+$
T	$\eta_T a_\alpha(-k)$	$\bar{\eta}_T b_\alpha(-k)$	$\eta_T\psi_\alpha(\mathbf{x}, -t)$	$+$	$-$	$+$
Θ	$\eta_\alpha\eta_{CPT} b_{-\alpha}(k)$	$\eta_\alpha\bar{\eta}_{CPT} a_{-\alpha}(k)$	$\eta_{CPT}\psi^*_\alpha(-x)$	$-$	$-$	$-$
G	$\eta_C b_\alpha(k)$	$(-1)^{2T}\bar{\eta}_C a_\alpha(k)$	$(-1)^{T+\alpha}\eta_C\psi^*_{-\alpha}(x)$	$+$	$+$	$+$
$\tilde{\Theta}$	$\eta_{GPT} b_\alpha(k)$	$(-1)^{2T}\bar{\eta}_{GPT} a_\alpha(k)$	$(-1)^{T+\alpha}\eta_{GPT}\psi_{-\alpha}(x)$	$+$	$-$	$+$

Combining these results with the antiparticle conjugation results of Sec. 5 gives

$$\begin{aligned}\Theta a_\alpha(k)\Theta^{-1} &= \eta_C\eta_P\eta_T\eta_\alpha b_{-\alpha}(k),\\ \Theta b_\alpha(k)\Theta^{-1} &= \bar{\eta}_C\bar{\eta}_P\bar{\eta}_T\eta_{-\alpha}a_{-\alpha}(k),\\ \Theta\psi_\alpha(x)\Theta^{-1} &= \eta_C\eta_P\eta_T\psi^*_\alpha(-x).\end{aligned} \tag{6.3}$$

Thus it is convenient to choose phases so that

$$\eta_C\eta_P\eta_T = 1. \tag{6.4}$$

In our (standard) basis T_2 is odd under time inversion. Hence Θ anticommutes with $\mathbf{T}$ and commutes with the rotation operator:

$$\begin{aligned}\Theta\mathbf{T} + \mathbf{T}\Theta &= 0,\\ \Theta\mathcal{O}(\boldsymbol{\lambda}) &= O(\boldsymbol{\lambda})\Theta.\end{aligned} \tag{6.5}$$

Under $GPT \equiv \tilde{\Theta}$, the state vectors or particle operators transform simply, but the field transformation law is more complicated:

$$\begin{aligned}\tilde{\Theta}a_\alpha(k)\tilde{\Theta}^{-1} &= b_\alpha(k),\\ \tilde{\Theta}b_\alpha(k)\tilde{\Theta}^{-1} &= (-1)^{2T}a_\alpha(k),\\ \tilde{\Theta}\psi_\alpha(x)\tilde{\Theta}^{-1} &= (-1)^{T+\alpha}\psi^*_{-\alpha}(-x).\end{aligned} \tag{6.6}$$

The antilinear character of $\tilde{\Theta}$ makes the GPT transformation of T more complicated than for G alone.

These and other pertinent results are collected in Table I. From (6.3) we see that Θ^2 commutes with all operators and is a multiple of the identity. Defining the vacuum to be invariant under Θ gives

$$\Theta^2 = I. \tag{6.7}$$

However, $\tilde{\Theta}^2$ is $\exp(2i\pi T_2)$.

It is clear that our definition (2.2) of the antiparticle states $|\bar{T}\alpha\rangle$ coincides exactly with the *definition*

$$|\bar{T}\alpha\rangle \equiv \tilde{\Theta}\,|T\alpha\rangle, \tag{6.8}$$

except for the phase $(-1)^{2T}$ of (6.6), which occurs for half-integral isospin.

One can easily check the compatibility of (6.8), (6.5), and the constructive proof of Sec. 2, in which the states $|T\alpha\rangle$ and $|\bar{T}\alpha\rangle$ were shown to transform identically under $O(\boldsymbol{\lambda})$.

It is of interest that the antiparticle states defined by CPT transform as the c.c. representation. This follows by applying Θ to Eq. (2.8), using Eq. (6.5):

$$\begin{aligned}\Theta O(\boldsymbol{\lambda})\,|T\alpha\rangle &= \sum_\beta \Theta(|T\beta\rangle D_{\beta\alpha}(\boldsymbol{\lambda})),\\ O(\boldsymbol{\lambda})(\Theta\,|T\alpha\rangle) &= \sum_\beta (\Theta\,|T\beta\rangle)D^*_{\beta\alpha}(\boldsymbol{\lambda}).\end{aligned} \tag{6.9}$$

The operator algebra is left invariant by Θ or $\tilde{\Theta}$. The energy density transforms as

$$\Theta\mathcal{H}(x)\Theta^{-1} = \mathcal{H}(-x), \tag{6.10}$$

and the theory is Θ invariant. Note that Θ commutes with the isospin commutation relations, Eqs. (2.7).

B. *CPT* for Self-Conjugate, Integral-Isospin Bosons.

This case is not very different from the foregoing. Letting $\eta_\alpha = (-1)^\alpha$, we have

$$Pa_\alpha(k)P^{-1} = \eta_P a_\alpha(-k), \tag{6.11a}$$

$$P\varphi_\alpha(\mathbf{x}, t)P^{-1} = \eta_P\varphi_\alpha(-\mathbf{x}, t), \tag{6.11b}$$

$$Ta_\alpha(k)T^{-1} = \eta_T a_\alpha(-k), \tag{6.11c}$$

$$T\varphi_\alpha(\mathbf{x}, t)T^{-1} = \eta_T\varphi_\alpha(\mathbf{x}, -t). \tag{6.11d}$$

(Here η_T is ± 1 because of the choice of real η_α.) Again choosing phases as in (6.4), we find that

$$\begin{aligned}\Theta a_\alpha(k)\Theta^{-1} &= (-1)^\alpha a_{-\alpha}(k),\\ \Theta\varphi_\alpha(x)\Theta^{-1} &= \varphi^*_\alpha(-x).\end{aligned} \tag{6.12}$$

Actually, the requirement that Eq. (6.11a) give (6.11b) requires that η_P be real, and hence $\eta_P = \pm 1$.

Similarly, the transition from (6.11c) to (6.11d) requires that η_T be related to η_α by

$$\bar{\eta}_T\bar{\eta}_\alpha = \eta_T\eta_\alpha. \tag{6.13}$$

Thus η_T^2 is $\bar{\eta}_\alpha^2$. Finally, η_c^2 is unity. Hence the choice of real η_α forces η^2_{CPT} to unity. Hence Θ^2 is unity:

$$\Theta^2\varphi_\alpha(x)\Theta^{-2} = (\bar{\eta}_{CPT})^2\varphi_\alpha(x) = \varphi_\alpha(x). \tag{6.14}$$

TABLE II. The behavior of various self-conjugate field operators under C, P, T, $G = C\exp(i\pi T_2)$, $\Theta = CPT$, and $\tilde{\Theta} = GPT$, is listed for the phase choice $\eta_\alpha = (-i)^{2\alpha}$, $\lambda_\alpha = \eta_C\eta_\alpha$. η^2_{CPT} is $(-1)^{2T}$ for this choice.

	$a_\alpha(k)$	$\varphi_\alpha(x)$	T_1	T_2	T_3
C	$i^{2\alpha}\eta_C a_{-\alpha}(k)$	$\eta_C\varphi^*_\alpha(x)$	−	+	−
P	$\eta_P a_\alpha(-k)$	$\eta_P\varphi_\alpha(-\mathbf{x}, t)$	+	+	+
T	$\eta_T a_\alpha(-k)$	$\eta_T\varphi_\alpha(\mathbf{x}, -t)$	+	−	+
Θ	$i^{2\alpha}\eta_{CPT}a_{-\alpha}(k)$	$\eta_{CPT}\varphi^*_\alpha(-x)$	−	−	−
G	$i^{2T}\eta_C a_\alpha(k)$	$i^{2T+2\alpha}\eta_C\varphi^*_{-\alpha}(x)$	+	+	+
$\tilde{\Theta}$	$i^{2T}\eta_{CPT}a_\alpha(k)$	$i^{2T+2\alpha}\eta_{CPT}\varphi^*_{-\alpha}(-x)$	+	−	+

Eq. (6.14) and its conjugate therefore lead again to Eq. (6.7).

The transformation properties are listed in Table II. The phases are written with factors of i so that they are also valid for half-integral isospin (with the second phase choice given below).

C. *CPT* for Self-Conjugate, Half-Integral Isospin Bosons

P and T are defined as in Eq. (6.11). Again, η_P^2 is unity and η_T is related to η_α by Eq. (6.13). In Sec. 5B we discussed charge conjugation with the phase choices $\eta_\alpha = (-1)^{T-\alpha}$, $\lambda_\alpha = \eta_C(-i)^{2\alpha}$, where $\eta_C^2 = 1$. Note that this (real) η_α requires $\eta_T^2 = 1$, so that

$$\eta^2_{CPT} = 1. \tag{6.15}$$

Combining Eq. (5.45) with (6.11) gives for *CPT*

$$\Theta a_\alpha(k)\Theta^{-1} = i^{2\alpha}a_{-\alpha}(k),$$
$$\Theta\varphi_\alpha(x)\Theta^{-1} = i^{2T}\varphi^*_\alpha(-x). \tag{6.16}$$

Here we have chosen η_{CPT} to be $+1$. Table III summarizes the transformations for this phase choice.

The interesting new feature of (6.16) is the appearance of the purely imaginary factor i^{2T} in the second equation. As a consequence of this, Θ^2 no longer is equivalent to the identity. We find

$$\Theta^2 a_\alpha + a_\alpha\Theta^2 = 0,$$
$$\Theta^2\varphi_\alpha(x) + \varphi_\alpha(x)\Theta^2 = 0. \tag{6.17}$$

We solve these equations by

$$\Theta^2 = (-1)^N \equiv \exp(i\pi N), \tag{6.18}$$

where N is the number operator.

The factor i can be removed from the second of Eqs. (6.16) if we make η_α imaginary. The occurrence of i in the field operator for self-conjugate bosons may seem offensive. However, we are accustomed to the equivalence of self-conjugate theories to a collection of degenerate Hermitian fields. Such equivalence does not hold here. Although we find it convenient to have η_α real, the phase choice now presented permits a unified treatment of all self-conjugate bosons. We generalize (5.40) to

$$\lambda_\alpha = (-i)^{2\alpha}\eta_c = \eta_c e^{-i\pi\alpha}, \quad \eta_c^2 = 1. \tag{6.19}$$

Again, Eq. (5.38) is satisfied. We make σ_α in Eq. (5.34) η_c by the choice

$$\eta_\alpha = (-i)^{2\alpha} = e^{-i\pi\alpha}. \tag{6.20}$$

For integral α this reduces to the choice $\eta_\alpha = (-1)^\alpha$ used for the integral isospin SCM. From (6.20) we

TABLE III. The behavior of half-integral isospin self-conjugate field operators under C, P, T, $G = C\exp(i\pi T_2)$, $\Theta = CPT$, and $\tilde{\Theta} = GPT$, is listed for (real) $\eta_\alpha = (-1)^{T-\alpha}$, $\lambda_\alpha = (-i)^{2\alpha}\eta$. $\eta_c^2{}_{CPT}$ is $+1$ for this phase choice.

	$a_\alpha(k)$	$\varphi_\alpha(x)$	T_1	T_2	T_3
C	$i^{2\alpha}\eta_c a_{-\alpha}(k)$	$i^{2T}\eta_c\varphi^*_\alpha(x)$	−	+	−
P	$\eta_P a_\alpha(-k)$	$\eta_P\varphi_\alpha(-\mathbf{x}, t)$	+	+	+
T	$\eta_T a_\alpha(-k)$	$\eta_T\varphi_\alpha(\mathbf{x}, -t)$	+	−	+
Θ	$i^{2\alpha}\eta_{CPT}a_{-\alpha}(k)$	$i^{2T}\eta_{CPT}\varphi^*_\alpha(-x)$	−	−	−
G	$i^{2T}\eta_c a_\alpha(k)$	$i^{2\alpha}\eta_c\varphi^*_{-\alpha}(x)$	+	+	+
$\tilde{\Theta}$	$i^{2T}\eta_{CPT}a_\alpha(k)$	$i^{2\alpha}\eta_{CPT}\varphi^*_{-\alpha}(-x)$	+	−	+

learn that

$$\eta_T^2 = \eta_\alpha^2 = \eta_{CPT}^2 = -1 \ . \tag{6.21}$$

Under C and Θ we find

$$\begin{aligned} Ca_\alpha C^{-1} &= i^{2\alpha}\eta_c a_{-\alpha}, \\ C\varphi_\alpha C^{-1} &= \eta_c \varphi_\alpha^*, \\ \Theta a_\alpha \Theta^{-1} &= i^{2\alpha}\eta_{CPT} a_{-\alpha}, \\ \Theta\varphi_\alpha(x)\Theta^{-1} &= \eta_{CPT}\varphi_\alpha^*(-x). \end{aligned} \tag{6.22}$$

Since η_{CPT} is purely imaginary, we are again led to Eqs. (6.17)–(6.18). The transformation laws associated with this phase convention are given in Table II.

The proof given by Kantor[7] depends critically on the possibility of choosing Θ^2 in the form (6.7). Repeating his argument for $\Theta^2 = -1$ (for one-particle states) yields exactly the criterion for half-integral isospin representations (see Ref. 14, p. 286).

ACKNOWLEDGMENTS

The author wishes to thank Dr. Y. S. Jin, Dr. M. Einhorn, Dr. E. Kazes, Dr. H. Lee, Dr. B. Zumino, and Dr. P. Kantor for preprints and instructive correspondence.

APPENDIX A: EXPLICIT FORM FOR THE ANTIPARTICLE OPERATOR C FOR PAIR-CONJUGATE BOSONS

To discover a unitary C such that

$$\begin{aligned} Ca_\alpha^* C^{-1} &= \eta_\alpha b_{-\alpha}^*, \\ Cb_\alpha^* C^{-1} &= \bar{\eta}_{-\alpha} a_{-\alpha}^*, \end{aligned} \tag{A1}$$

we employ the identity

$$e^{\lambda A} B e^{-\lambda A} = B + \lambda[A, B] + \frac{\lambda^2}{2}[A, [A, B]] + \cdots . \tag{A2}$$

We expect C to be of the form $\exp(i\theta\Lambda)$, Λ Hermitian. Since $C^2 = 1$, $2\theta\Lambda$ has eigenvalues $2\pi n$ (n any integer).

From the commutator series (A2) we are led to consider the following Λ_1:

$$\Lambda_1 = \sum_{k\alpha} (\eta_\alpha a_\alpha(k) b_{-\alpha}^*(k) + \bar{\eta}_\alpha a_\alpha^*(k) b_{-\alpha}(k)). \tag{A3}$$

We thus obtain the following commutators:

$$\begin{aligned} &[\Lambda_1, a_\alpha^*] = \eta_\alpha b_{-\alpha}^*, \qquad [\Lambda_1, b_\alpha^*] = \bar{\eta}_{-\alpha} a_{-\alpha}^*, \\ &[\Lambda_1, [\Lambda_1, a_\alpha^*]] = a_\alpha^*, \quad [\Lambda_1, [\Lambda_1, b_\alpha^*]] = b_\alpha^*. \end{aligned} \tag{A4}$$

The closure exhibited in (A4) leads to a very simple summation of the series (A2)

$$U(\theta) \equiv \exp(i\theta\Lambda_1),$$

$$\begin{aligned} U(\theta) a_\alpha^* U(\theta)^{-1} = a_\alpha^* \sum_{n=0,2,4,\cdots,} \frac{(i\theta)^n}{n!} \\ + \eta_\alpha b_{-\alpha}^* \sum_{n=1,3,5,\cdots,} \frac{(i\theta)^n}{n!} . \end{aligned} \tag{A5}$$

The same calculation works for b_α^*. We find

$$\begin{aligned} U(\theta) a_\alpha^* U(\theta)^{-1} &= \cos\theta a_\alpha^* + i \sin\theta \eta_\alpha b_{-\alpha}^*, \\ U(\theta) b_\alpha^* U(\theta)^{-1} &= i \sin\theta \bar{\eta}_{-\alpha} a_{-\alpha}^* + \cos\theta b_\alpha^*. \end{aligned} \tag{A6}$$

Defining the two-component object v by $v_1 = a_\alpha^*$, $v_2 = i\eta_\alpha b_{-\alpha}^*$, we see that (A6) has the form of a two-dimensional rotation

$$\begin{aligned} U(\theta) v U(\theta)^{-1} &= D(\theta) v; \\ D(\theta) &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \end{aligned} \tag{A7}$$

For $\theta = \pi/2$, we have

$$\begin{aligned} U(\pi/2) a_\alpha^* U(\pi/2)^{-1} &= i\eta_\alpha b_{-\alpha}^*, \\ U(\pi/2) b_\alpha^* U(\pi/2)^{-1} &= i\bar{\eta}_{-\alpha} a_{-\alpha}^*. \end{aligned} \tag{A8}$$

This is almost C except for the appearance of the factor of i. Such a phase is easy to transform away, however.

Define the total number operator N by

$$\begin{aligned} N &= \sum_{k\alpha} (a_\alpha^*(k) a_\alpha(k) + b_\alpha^*(k) b_\alpha(k)), \\ &[N, a_\alpha^*] = a_\alpha^*, \quad [N, b_\alpha^*] = b_\alpha^*. \end{aligned} \tag{A9}$$

The transformation $\mathcal{O}(\theta)$

$$\mathcal{O}(\theta) \equiv \exp(i\theta N) \tag{A10}$$

advances the phase of a_α^* or b_α^* by θ, as follows from (A2) and (A10):

$$\mathcal{O}(\theta) \begin{pmatrix} a_\alpha^* \\ b_\alpha^* \end{pmatrix} \mathcal{O}^{-1}(\theta) = e^{i\theta} \begin{pmatrix} a_\alpha^* \\ b_\alpha^* \end{pmatrix}. \tag{A11}$$

Thus we may define C by

$$C = O\left(-\frac{\pi}{2}\right) U\left(\frac{\pi}{2}\right) = \exp\left(-i\frac{\pi}{2}N\right) \exp\left(i\frac{\pi}{2}\Lambda_1\right). \tag{A12}$$

A short calculation verifies that $[N, \Lambda_1]$ vanishes, as may be surmised from the definition of Λ, Eq. (A3). Thus we may combine exponentials in (A12), giving

$$\begin{aligned} C &= \exp\left[i\frac{\pi}{2}(\Lambda_1 - N)\right] \\ &= \exp\left[-i\frac{\pi}{2}\sum_{k\alpha}(a_\alpha^* - \eta_\alpha b_{-\alpha}^*)(a_\alpha - \bar{\eta}_\alpha b_{-\alpha})\right]. \end{aligned} \tag{A13}$$

C can be regarded as a finite transformation of the continuous Lie group C_θ:

$$C_\theta = \exp(i\theta\Xi), \quad \Xi = \Lambda_1 - N. \tag{A14}$$

There is an interesting correspondence between the finite transformation C and the infinitesimal generators Λ_1, (or Ξ). This is only the usual one which obtains when we have a continuous group, but the nature of the algebra is interesting.

Consider the transformations

$$\begin{aligned} Ca_\alpha^* C^{-1} &= \eta_\alpha b_{-\alpha}^*, \quad [\Lambda, a_\alpha^*] = \eta_\alpha b_{-\alpha}^*, \\ C\eta_\alpha b_{-\alpha}^* C^{-1} &= a_\alpha^*, \quad [\Lambda, \eta_\alpha b_{-\alpha}^*] = a_\alpha^*, \\ C\Lambda_1 C^{-1} &= \Lambda_1, \qquad [\Lambda, \Lambda] = 0. \end{aligned} \tag{A15}$$

Of course it is $i\Lambda$ that is the infinitesimal generator of $U(\theta)$. Making the identifications

$$\begin{aligned} m &= \Lambda, \\ n &= a_\alpha^*, \\ l &= \eta_\alpha b_{-\alpha}^*, \end{aligned} \tag{A16}$$

we find the algebra

$$\begin{aligned} [m, n] &= l, \\ [m, l] &= n, \\ [l, n] &= 0. \end{aligned} \tag{A17}$$

In contrast, the algebra of the Ξ, a_α^*, $\eta_\alpha b_{-\alpha}^*$ is somewhat more intricate. With $\Xi = m'$, $a_\alpha^* = n'$, $\eta_\alpha b_{-\alpha}^* = l'$, we find

$$\begin{aligned} [m', n'] &= l' - n', \\ [m', l'] &= n' - l', \\ [l', n'] &= 0. \end{aligned} \tag{A18}$$

As a consequence of (A18), $l' + n'$ commutes with m'. The quantity $l' - n'$ is exactly what occurs in the exponential (A13); with

$$C_\alpha = a_\alpha^* - \eta_\alpha b_{-\alpha}^*,$$

we have

$$C_\theta = \exp\left(-i\theta \sum_{k\alpha} C_\alpha^*(k) C_\alpha(k)\right). \tag{A19}$$

Note the identity

$$\Xi = [\Xi, a_\alpha^*]^*[\Xi, \alpha_\alpha^*]. \tag{A20}$$

APPENDIX B: OPERATOR TRANSFORMATIONS OF THE SC PARTICLE OPERATORS

We give some expressions useful for the discussion of antiparticle conjugation given in Sec. 5. We consider a single momentum state k, suppressing the label k in our equations.

I. Phase Change of a_α

Consider the operator

$$\begin{aligned} O(\theta) &= \exp(i\theta N), \\ N &= \sum_\alpha a_\alpha^* a_\alpha, \end{aligned} \tag{B1}$$

where θ is real. Using the commutator expansion (A2), we find

$$e^{i\theta N} a_\alpha^* e^{-i\theta N} = e^{i\theta} a_\alpha^*. \tag{B2}$$

II. Change of Phase which Depends on α

Let ξ_α be real and of unit magnitude [for example the function $(-1)^{T-\alpha}$],

$$\xi_\alpha^2 = 1. \tag{B3}$$

Define the Hermitian operator Λ_1 and the unitary generator U_1 by

$$\begin{aligned} U_1(\theta) &= \exp(i\theta\Lambda_1), \\ \Lambda_1 &= \sum_\alpha \xi_\alpha a_\alpha^* a_\alpha. \end{aligned} \tag{B4}$$

From the commutation rules

$$\begin{aligned} [\Lambda_1, a_\alpha^*] &= \xi_\alpha a_\alpha^*, \\ [\Lambda_1, [\Lambda_1, a_\alpha^*]] &= a_\alpha^*, \end{aligned} \tag{B5}$$

etc., we find the result

$$\begin{aligned} e^{i\theta\Lambda_1} a_\alpha^* e^{-i\theta\Lambda_1} &= \cos\theta a_\alpha^* + i\xi_\alpha \sin\theta a_\alpha^* \\ &= e^{i\xi_\alpha\theta} a_\alpha^*. \end{aligned} \tag{B6}$$

This includes (B2) as a special case.

For $\theta = \pi/2$ (B6) reduces to

$$U_1(\pi/2) a_\alpha^* U_1(\pi/2)^{-1} = i\xi_\alpha a_\alpha^*. \tag{B7}$$

We may remove the phase i by applying (B2), with $\theta = -\pi/2$. Since Λ_1 commutes with N,

$$\begin{aligned} V a_\alpha^* V^{-1} &= \xi_\alpha a_\alpha^*, \\ V = O\left(-\frac{\pi}{2}\right) U_1\left(\frac{\pi}{2}\right) &= \exp\left[\frac{i\pi}{2}(\Lambda_1 - N)\right]. \end{aligned} \tag{B8}$$

Finally, the most general transformation needed in this paper,

$$W a_\alpha^* W^{-1} = \eta_\alpha a_\alpha^*, \quad \eta_\alpha = \xi\xi_\alpha, \quad \xi = e^{i\omega}, \tag{B9}$$

is given by the operator

$$W = e^{i\omega N} V = \exp\left[i\left(\frac{\pi}{2}\Lambda + \left(\omega - \frac{\pi}{2}\right)N\right)\right]. \tag{B10}$$

III. Inversion of Sign of α

The general transformation

$$C a_\alpha^* C^{-1} = \eta_\alpha a_{-\alpha}^* \tag{B11}$$

can be built from W of (B10) and an operator C_0:

$$C_0 a_\alpha^* C_0^{-1} = a_{-\alpha}^*,$$
$$C = C_0 W. \tag{B12}$$

The Hermitian operator Λ_2,

$$\Lambda_2 = \sum_\alpha a_\alpha^* a_{-\alpha}, \tag{B13}$$

obeys the commutation rules

$$[\Lambda_2, a_\alpha^*] = a_{-\alpha}^*,$$
$$[\Lambda_2, [\Lambda_2, a_\alpha]] = a_\alpha, \tag{B14}$$

etc., so that the continuous transformation

$$U_2(\theta) = \exp(i\theta\Lambda_2) \tag{B15}$$

gives the following transformation:

$$U_2(\theta) a_\alpha^* U_2(\theta)^{-1} = a_\alpha^* \cos\theta + i \sin\theta a_{-\alpha}^*. \tag{B16}$$

Setting $\theta = \pi/2$ and removing the factor i, as in Eq. (B8), gives for C_0

$$C_0 = \exp\left[\frac{i\pi}{2}(\Lambda_2 - N)\right]$$
$$= \exp\left[\frac{i\pi}{2}\sum_\alpha a_\alpha^*(a_{-\alpha} - a_\alpha)\right]. \tag{B17}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Nonequilibrium Statistical Mechanics of Open Systems

GERARD G. EMCH AND GEOFFREY L. SEWELL*
Department of Physics and Astronomy, University of Rochester, Rochester, New York

(Received 7 June 1967)

A theoretical framework for the nonequilibrium statistical mechanics of open systems is constructed. This is concerned with a formulation of a generalized master equation governing the evolution of an arbitrary system S in interaction with a "large" reservoir R. The dynamics of S are analyzed on the basis of a precise quantum-mechanical treatment of the microscopic equations of motion for the combined system $S + R$. On proceeding to the thermodynamical limit for R we obtain a generalized master equation for S, subject to specified conditions on the many-particle structure of R, its initial state, and its coupling to S. This master equation corresponds to a self-contained law of motion for S, in which the R variables appear only in the forms of certain thermal averages, taken over the initial state. This dynamical law is a generalization of the quantum-mechanical Liouville equation to a form appropriate to open systems.

1. INTRODUCTION

A considerable body of research in nonequilibrium statistical mechanics has been founded on treatments (see, for instance, Refs. 1–4 and papers quoted there) of the microscopic equations of motion for closed systems, without recourse to statistical assumptions concerning the actual dynamics of the systems—the most familiar assumption of this type is Boltzmann's stosszahlansatz. In these aforementioned treatments, the only statistical assumptions made are ones that concern the initial states of the systems. The treatments, therefore, have the merit of being manifestly consistent with the dynamical laws of microphysics, whether classical or quantal.

The object of the present paper is to provide a formal framework for the nonequilibrium statistical mechanics of *open* systems, based likewise on a systematic treatment of their microscopic equations of motion, together with a statistical assumption concerning their initial states. Here we refer to a system S as being "open" if it is coupled to an appropriately "large" system R, which we shall term a "reservoir," and whose properties will be further specified in Secs. 2 and 3. Our definition of an open system is designed to be sufficiently general to cover not only the cases where S is a system placed in some thermostat, say, but also the cases where S is part of a much larger system, the remainder of which is identified with the reservoir R. An example of the latter cases is one where S is the assembly of conduction electrons and R is the phonon system in a semiconductor. It is evident that most systems of interest to physicists are "open" in the sense we have specified: In some cases one might expect R to exert a crucial influence; in others, a trivial one, on the behavior of S.

Our approach to the theory of open systems will be to extract the dynamical laws governing S from the microscopic equations of motion for the closed systems comprising R and S. This kind of approach has been used by a number of authors in more limited contexts. For example, some[5] have extracted dynamical laws for S in cases where the microscopic equations of motion for the compound system $R + S$ were exactly soluble, while others[6] have dóne likewise in cases where R was postulated to have especially simple microscopic properties that justified a stosszahlansatz. The present theory, on the other hand, is designed to be both general and free from any stosszahlansatz or equivalent assumption.

We shall be concerned with a description of the dynamics of S in the following situation. The reservoir R and the system S are initially prepared independently of one another, the preparation of R being effected by measurement of a set Σ_R of macroscopic variables, which are constants of the motion when R is isolated. The two systems are then coupled together so that the evolution of S is partially governed by its interaction with R.

---

* Permanent Address: Department of Physics, Queen Mary College, London, England.

[1] L. van Hove, Physica **23**, 441 (1957); E. W. Montroll, *Fundamental Problems in Statistical Mechanics*, E. G. D. Cohen, Ed. (North-Holland Publishing Company, Amsterdam, 1962), pp. 230-249; I. Prigogine, *Non-Equilibrium Statistical Mechanics* (John Wiley & Sons, New York, 1962); A. Janner, Helv. Phys. Acta **35**, 47 (1962); **36**, 155 (1963); G. L. Sewell, Physica **31**, 1520 (1965).

[2] R. Zwanzig, Physica **30**, 1109 (1964).

[3] G. Emch, Helv. Phys. Acta **37**, 532 (1964); and in *Lectures in Theoretical Physics, Summer Institute for Theoretical Physics, 1965*, W. E. Brittin, Ed. (University of Colorado Press, 1966), pp. 65–99.

[4] G. L. Sewell, Physica **34**, 493 (1967).

[5] P. C. Hemmer, Ph.D. thesis, Trondheim, 1959; R. J. Rubin J. Math. Phys. **1**, 309 (1960); R. E. Turner, Physica **26**, 269 (1960); P. Mazur and E. Braun, Physica **30**, 1973 (1964); E. Braun, Physica **33**, 528 (1967); P. Ullersma, Physica **32**, 27 (1966); **32**, 56 (1966); **32**, 74 (1966); **32**, 90 (1966).

[6] P. G. Bergmann and J. L. Lebowitz, Phys. Rev. **99**, 578 (1955); J. L. Lebowitz and P. G. Bergmann, Ann. Phys. **1**, 1 (1957); C. R. Willis and P. G. Bergmann, Phys. Rev. **128**, 391 (1962).

Our investigation of the dynamics of S is based on an extension of methods we[3,4] have previously developed for closed systems. This investigation starts from the quantum-mechanical Liouville equation for the full system comprising S and R; and proceeds from this starting point in two principal steps. The first step is to derive an exact generalized master equation for the evolution of the reduced statistical operator governing all the S observables and the macroscopic variables Σ_R. This is achieved by means of an extension of the mathematically rigorous projective techniques[2,3] which were first used to derive exact generalized master equations for macroscopic observables of closed systems. In both this earlier case and in the present one, the derivation of the master equation rests only on general principles of quantum mechanics, together with statistical assumptions on the initial state.

The second step is to reduce this master equation to one governing the evolution of S only. This reduction is based on assumptions of a general character concerning the many-particle structure of R and the form of its coupling to S. Further, the reduction is effected in the thermodynamical limit for R; i.e., in the limit where the number of particles N and the volume of R both tend to infinity, the particle density of the reservoir remaining finite.

The resultant master equation constitutes a self-contained dynamical law for S in which the R variables appear only in the forms of averages taken over the initial state. This equation represents a generalization of the quantum-mechanical Liouville equation to a form that is appropriate to open systems. In general, the equation is non-Markovian. Questions concerning its reduction to a Markovian form are not considered in this paper.

The essential factor which permits the reduction of the master equation to a self-contained law for S is that, in consequence of our assumptions, it is found that the reservoir exerts its influence on S only via its *intensive* properties. Now, although the extensive variables Σ_R may change by a finite amount as a result of the R–S coupling, the intensive variables Σ_R/N become constants of the motion in the thermodynamical limit. Consequently the law governing the evolution of S contains nothing of the time dependence of Σ_R, which means that S follows a self-contained dynamical law. It may be seen that the factors governing the existence of such a law for S are very similar to those we invoked in an earlier paper[4] for the purpose of deriving self-contained macroscopic laws for a closed system.

We set out the theory as follows: In Sec. 2 we formulate our mathematical representation of the states and observables for the combined system $S + R$. This representation is chosen so as to enable us to rigorously derive a master equation governing S and the macro-observables Σ_R. (The necessity for rigor here is brought out in discussions in Secs. 4 and 5.) Having derived this equation, we anticipate results that are obtained in Sec. 3 and employ them to reduce the master equation to one that governs the evolution of S only.

In Sec. 3 we derive the above-mentioned results, in the thermodynamic limit, subject to specified assumptions concerning the many-particle structure of R, the form of its coupling to S, and its initial state.

In Sec. 4 we make further use of these results in order to simplify the master equation to a form that should be more tractable for future use.

In Sec. 5 we summarize our conclusions.

2. GENERALIZED MASTER EQUATION FOR A SYSTEM INTERACTING WITH A RESERVOIR

We consider the evolution of the states of a quantum system, referred to as the *system of interest* (S), interacting with a "large" quantum system, referred to as the *reservoir* (R). For notational convenience the latter is essentially an energy reservoir, but the formalism is built in such a way as to allow R to be a reservoir for all other quantities which might be needed to complete its macroscopic description. As will become clear in the sequel, the reservoir will be devised in such a way that only its intensive variables will affect the evolution of the system S. The compound formed by the system S interacting with the reservoir R is referred to as the *total system* ($S + R$).

Throughout this paper we intend to use the projector techniques in Liouville space which by now have been used by numerous authors, for example Refs. 2, 3, 7, 8. For the sake of self-completeness of this paper, we first recall briefly the main facts of this mathematical formalism. We then present with somewhat more details the generalization required for our problem and the specific properties pertaining to the case under consideration.

When dealing with a quantum system, it is customary to represent pure states and observables, respectively, as vectors of a Hilbert space $\mathfrak{H}$, and as Hermitian operators acting on $\mathfrak{H}$. Mixed states (as

[7] B. Robertson, Phys. Rev. **144**, 151 (1966); **153**, 391 (1967); P. N. Argyres, e.g., in *Lectures in Theoretical Physics, Summer Institute for Theoretical Physics, 1965*, W. E. Brittin, Ed. (University of Colorado Press, 1966), pp. 183–238.

[8] G. Ludwig, in *Ergodic Theories*, P. Caldirola, Ed. (Academic Press Inc., New York, 1961), pp. 57–132; U. Fano, in *The Many-Body Problem*, E. R. Caianiello, Ed. (Academic Press Inc., New York, 1964), Vol. 2, pp. 217–239.

"mixtures") are then represented as certain operators, referred to as density matrices. The latter are the states with which statistical mechanics is concerned. In dynamical problems, for instance, one has to consider operations which map the manifold of the density matrices onto itself. Coarse-graining is also formulated in terms of operations on mixed states. In the primitive stage of the theory, this led to the introduction of a somewhat cumbersome "tetradics" notation[2] for the purpose of representing such things as the quantum analog of the classical Liouville operator. It was considered, however, that it would be useful for quantum-statistical purposes to have a formalism in which every state, whether pure or mixed, would be represented as a vector in some appropriate Hilbert space on which the Liouville operator and the coarse-graining operator would act. The Liouville space formalism precisely provides such a framework. One associates with $\mathfrak{H}$ the set $\mathfrak{L}$ of all Hilbert–Schmidt operators mapping $\mathfrak{H}$ onto itself:

$$\mathfrak{L} \equiv \{A \in B(\mathfrak{H}) \mid \mathrm{Tr}\,(A^*A) < \infty\} \equiv \mathfrak{L}(\mathfrak{H}), \quad (2.1)$$

equipped with the scalar product

$$(A, B) \equiv \mathrm{Tr}\,(A^*B) \quad \text{defined for all} \quad A, B \in \mathfrak{L}. \quad (2.2)$$

This Hilbert space $\mathfrak{L}$ is the so-called *Liouville space.* Let $\{U^t\}$ be the unitary group describing in $\mathfrak{H}$ the evolution of the pure states

$$\psi(t) = U^t\psi(0). \quad (2.3)$$

One can associate with U^t the operator $\mathfrak{U}^t$, defined by

$$\mathfrak{U}^tA = U^tAU^{-t}. \quad (2.4)$$

Thus, the unitary group $\{\mathfrak{U}^t\}$ governs the evolution of the mixed states

$$\rho(t) = \mathfrak{U}^t\rho(0). \quad (2.5)$$

In the same way as the *Hamiltonian* H is the generator of $\{U^t\}$, i.e.,

$$U^t = \exp(-iHt), \quad (2.6)$$

the Liouville operator $\mathfrak{L}$ is the generator of $\{\mathfrak{U}^t\}$, i.e.,

$$\mathfrak{U}^t = \exp(-i\mathfrak{L}t). \quad (2.7)$$

Thus, by (2.4)–(2.7),

$$\mathfrak{L}\rho = [H, \rho], \quad (2.8)$$

so that the Schrödinger equation for pure states

$$i\frac{d}{dt}\psi(t) = H\psi(t) \quad (2.9)$$

is replaced for mixtures by an equation of the same form, but in Liouville space. This is simply the von Neuman equation

$$i\frac{d}{dt}\rho(t) = \mathfrak{L}\rho(t) \equiv [H, \rho(t)]_-, \quad (2.10)$$

which then appears as the exact quantum analog of the classical Liouville equation with the commutator replaced by the Poisson bracket (P.B.)

$$i\frac{d}{dt}f(t) = \mathfrak{L}f(t) = i\{H, f(t)\}_{\text{P.B.}}, \quad (2.11)$$

where f is the classical distribution function.

Let S and R be two quantum mechanical systems, as described above. Let $\mathfrak{H}_S$ and $\mathfrak{H}_R$ be the Hilbert spaces of their respective pure states, and let us denote the corresponding Liouville spaces by $\mathfrak{L}_S$ and $\mathfrak{L}_R$. The quantum-mechanical description of the total system $S + R$ is usually given in terms of operators acting in the Hilbert space $\mathfrak{H}$, defined as the direct product of $\mathfrak{H}_R$ and $\mathfrak{H}_S$:

$$\mathfrak{H} = \mathfrak{H}_R \otimes \mathfrak{H}_S. \quad (2.12)$$

For our statistical purposes, it is more convenient to formulate the properties of the total system $S + R$ in terms of the Liouville space

$$\mathfrak{L} = \mathfrak{L}_R \otimes \mathfrak{L}_S. \quad (2.13)$$

Since there are several almost equivalent ways to define the direct product of Hilbert spaces, it is perhaps not useless to prevent possible misunderstandings by stating here what we actually mean. Let $\mathfrak{H}^{(1)}$ and $\mathfrak{H}^{(2)}$ be two Hilbert spaces. To each pair of vectors $\psi^{(1)}$ in $\mathfrak{H}^{(1)}$ and $\psi^{(2)}$ in $\mathfrak{H}^{(2)}$ we associate the operator

$$\psi^{(1)} \otimes \psi^{(2)} : \mathfrak{H}^{(2)} \to \mathfrak{H}^{(1)}, \quad (2.14)$$

defined by

$$(\psi^{(1)} \otimes \psi^{(2)})\varphi^{(2)} = (\psi^{(2)}, \varphi^{(2)})_2\psi^{(1)}. \quad (2.15)$$

We then consider the vector space $\mathfrak{H}^{(0)}$ spanned by all the operators so obtained when $\psi^{(1)}$ and $\psi^{(2)}$, respectively, run over $\mathfrak{H}^{(1)}$ and $\mathfrak{H}^{(2)}$. We equip this vector space $\mathfrak{H}^{(0)}$ with the scalar product

$$(\psi^{(1)} \otimes \psi^{(2)}, \varphi^{(1)} \otimes \varphi^{(2)}) = (\psi^{(1)}, \varphi^{(1)})_1(\psi^{(2)}, \varphi^{(2)})_2. \quad (2.16)$$

We then add to $\mathfrak{H}^{(0)}$ its limit points to get $\mathfrak{H} = \mathfrak{H}^{(1)} \otimes \mathfrak{H}^{(2)}$.

We further notice, for consistency, that, with this definition of direct product,

$$\mathfrak{L}(\mathfrak{H}_R \otimes \mathfrak{H}_S) = \mathfrak{L}(\mathfrak{H}_R) \otimes \mathfrak{L}(\mathfrak{H}_S). \quad (2.17)$$

We now present the so-called "*projector technique in Liouville space*" in the form most suited to the aims of the present paper. In its broadest sense, this

technique is devised so as to systematically extract the information pertinent to the variables under consideration from the complicated microscopic dynamics of the complete system. In the present paper the variables under consideration are, on one hand, the complete set of observables for S and, on the other hand, a set of macroscopic observables for R. The latter observables are idealized[3,9] as forming a set Σ_R of intercommuting self-adjoint operators $\{A_R\}$, acting on the Hilbert space $\mathfrak{H}_R$ pertinent to the proper microscopic description of R. Each set $\{A(E)\}$ of simultaneous eigenvalues of these observables determines a subspace $\mathfrak{H}_E$ of $\mathfrak{H}_R$; the projector from $\mathfrak{H}_R$ to this subspace $\mathfrak{H}_E$ will be denoted by D_E. We notice in passing that $\mathfrak{H}_E$ plays in quantum statistics exactly the role of the familiar *macroscopic* cell of classical statistics. We will, therefore, designate $\mathfrak{H}_E$ by the same name. In this idealization we have then that each macroscopic observable A_R takes the form

$$A_R = \sum_E A(E) D_E. \tag{2.18}$$

We denote by W_E the dimension ($<\infty$) of $\mathfrak{H}_E$. We have then:

$$\left.\begin{aligned} &\mathrm{Tr}_R\, D_E = W_E, \\ &D_E D_{E'} = D_E \delta_{EE'}, \\ &D_E^* = D_E, \\ &\sum_E D_E = I_R \text{ (the identity operator on } \mathfrak{H}_R). \end{aligned}\right\} \tag{2.19}$$

We now define a coarse-graining operator $\mathfrak{T}_R$, acting on $\mathfrak{L}_R$, as follows:

$$\mathfrak{T}_R B_R = \sum_E \langle B_R \rangle_E D_E, \quad \text{for all } B_R \text{ in } \mathfrak{L}_R, \tag{2.20}$$

with

$$\langle B_R \rangle_E = \mathrm{Tr}_R (B_R F_E), \tag{2.21}$$

where F_E, defined as

$$F_E = D_E / W_E, \tag{2.22}$$

is the density matrix which corresponds to the uniform distribution throughout the cell $\mathfrak{H}_E$. The physical meaning of $\mathfrak{T}_R$ is that it coarse-grains (see Ref. 3) the elements of $\mathfrak{L}_R$. Since, moreover, it satisfies the equation

$$\mathfrak{T}_R = \mathfrak{T}_R^* = \mathfrak{T}_R^2, \tag{2.23}$$

we refer to it as the *coarse-graining projector*. We have, in particular, for every density ρ_R in $\mathfrak{L}_R$,

$$\mathfrak{T}_R \rho_R = \sum_E p_E F_E, \tag{2.24}$$

where

$$p_E = \mathrm{Tr}_R (\rho_R D_E). \tag{2.25}$$

[9] N. G. van Kampen, Physica **20**, 603 (1954).

We can interpret p_E as the probability that the cell $\mathfrak{H}_E$ is occupied when the system R is in state ρ_R. We have indeed, for any macroscopic observable A_R and any state ρ_R,

$$\langle A_R \rangle_{\rho_r} = \sum_E p_E A(E), \tag{2.26}$$

so that we can rephrase our statement concerning the meaning of p_E by saying that it is the probability that the macroscopic observable A_R takes the value $A(E)$ when the system R is in the state ρ_R.

Let us now come back to our total system $(S + R)$. We suppose that we follow the evolution of the total system $(S + R)$ via two sets of observables, respectively, attached to S and to R. We do not want, for the moment, to impose any restriction whatsoever on the observables relative to S, which we choose to measure. The observables on R will be chosen to be the above-described set Σ_R. The conditions of observation just described lead to the definition of a new *coarse-graining projector* $\mathfrak{T}$ *for the total system* $(S + R)$, namely

$$\mathfrak{T} = \mathfrak{T}_R \otimes \mathfrak{I}_S, \quad \text{(where } \mathfrak{I}_S \text{ is the identity operator acting on } \mathfrak{L}_S). \tag{2.27}$$

The reader will convince himself immediately that $\mathfrak{T}$ is indeed a projector acting on the Hilbert space $\mathfrak{L} = \mathfrak{L}_R \otimes \mathfrak{L}_S$.

We have, in particular, for any B_R in $\mathfrak{L}_R$ and B_S in $\mathfrak{L}_S$:

$$\mathfrak{T}(B_R \otimes B_S) = \sum_E \langle B_R \rangle_E B_S^E, \tag{2.28}$$

with

$$\langle B_R \rangle_E = \mathrm{Tr}_R (B_R F_E), \tag{2.29}$$

and

$$B_S^E = D_E \otimes B_S, \tag{2.30}$$

(We notice, in passing, that this formula, extended by linearity and continuity over the whole of $\mathfrak{L}$, precisely defines $\mathfrak{T}$.)

For notational convenience in the sequel, we now want to introduce the following operators:

$$B_R^\times = B_R \otimes I_S, \tag{2.31}$$

$$B_S^\times = I_R \otimes B_S. \tag{2.32}$$

(Notice, however, that neither $B_R^\times$ nor $B_S^\times$ belong to $\mathfrak{L}$ in the general case where, respectively, $\mathfrak{H}_R$ or $\mathfrak{H}_S$ are infinite-dimensional.) We further define the operator

$$\mathrm{Tr}_R\, B: \quad \mathfrak{H}_S \to \mathfrak{H}_S \tag{2.33}$$

by

$$(\psi^{(S)}, \mathrm{Tr}_R\, B \varphi^{(S)}) = \sum_i (\psi_i^{(R)} \otimes \psi^{(S)}, B \psi_i^{(R)} \otimes \varphi^{(S)}), \tag{2.34}$$

for any operator B acting on $\mathfrak{H}_R \otimes \mathfrak{H}_S$, such that the right-hand side exists (and is finite) for all $\psi^{(S)}$ and $\varphi^{(S)}$ in $\mathfrak{H}_S$, and is independent of the chosen basis $\{\psi_i^{(R)}\}$ in $\mathfrak{H}_R$. We notice in particular that $\mathrm{Tr}_R(B_R \otimes B_S)$ is defined in the above sense whenever B_S is bounded on $\mathfrak{H}_S$ *and* B_R has a finite trace. We have then the property

$$\mathrm{Tr}_R(B_R \otimes B_S) = (\mathrm{Tr}_R B_R)B_S. \tag{2.35}$$

Using the notation just introduced, we can rewrite explicitly the action of $\mathcal{F}$ on any element of $\mathfrak{L}$ as

$$\mathcal{F}A = \sum_E D_E \otimes \mathrm{Tr}_R(F_E^{\times}A) \tag{2.36}$$

or, for the states we consider,

$$\mathcal{F}\rho = \sum_E F_E \otimes \mathrm{Tr}_R(D_E^{\times}\rho). \tag{2.37}$$

We notice at this point that this formula allows us to extend $\mathcal{F}$ somewhat outside of $\mathfrak{L}$. In particular we have, for any observable A_S on S and macroscopic observable $A_R = \sum_E A(E)D_E$ on R:

$$\begin{aligned} \mathcal{F}A_S^{\times} &= A_S^{\times}, \\ \mathcal{F}A_R^{\times} &= A_R^{\times}, \\ \mathcal{F}(A_R \otimes A_S) &= A_R \otimes A_S. \end{aligned} \tag{2.38}$$

There properties justify our use of $\mathcal{F}$ in the sense that, when computing the expectation values of "relevant" observables (namely, any observable A_S on S, any macroscopic observable A_R on R, and their combinations), we can replace ρ by $\mathcal{F}\rho$. In other words, $\mathcal{F}$ extracts from any state ρ of the total system $(S + R)$ the information relevant for our purpose, namely $\mathcal{F}\rho$.

Using now the projector technique pioneered by Zwanzig[2] and exploited subsequently by several authors, we can derive a master equation for $\mathcal{F}\rho(t)$, our quantity of interest.

To obtain this equation we just have to generalize slightly the procedure used in Ref. 3. This generalization takes account of the fact that our present projector $\mathcal{F}$ differs from that used earlier in that now $\mathcal{F}\mathfrak{L}\mathcal{F} \neq 0$. We first Laplace-transform the trivial operator equation

$$\frac{d}{dt}\mathfrak{U}(t) = -i\mathfrak{L}\mathfrak{U}(t) \tag{2.39}$$

into

$$z\mathcal{R}(z) - I = -i\mathfrak{L}\mathcal{R}(z), \tag{2.40}$$

where $\mathcal{R}(z)$ is the resolvant $(z + i\mathfrak{L})^{-1}$ of $\mathfrak{U}(t) = \exp(-i\mathfrak{L}t)$. We multiply Eq. (2.40) from the left by $\mathcal{F}$ to get

$$\mathcal{F}[z\mathcal{R}(z) - I] = -i\mathcal{F}\mathfrak{L}\mathcal{R}(z),$$

i.e.,

$$\mathcal{F}[z\mathcal{R}(z) - I] = -i\mathcal{F}\mathfrak{L}\mathcal{F}\mathcal{R}(z) - i\mathcal{F}\mathfrak{L}(I - \mathcal{F})\mathcal{R}(z). \tag{2.41}$$

If we multiply Eq. (2.40) from the left by $(I - \mathcal{F})$ we get

$$(I - \mathcal{F})[z\mathcal{R}(z) - I] = -i(I - \mathcal{F})\mathfrak{L}\mathcal{R}(z),$$

which we can rewrite as

$$z(I - \mathcal{F})\mathcal{R}(z) = (I - \mathcal{F}) - i(I - \mathcal{F})\mathfrak{L}\mathcal{R}(z). \tag{2.42}$$

On adding $i(I - \mathcal{F})\mathfrak{L}(I - \mathcal{F})\mathcal{R}(z)$ to both sides, we obtain

$$\begin{aligned} &[z + i(I - \mathcal{F})\mathfrak{L}(I - \mathcal{F})](I - \mathcal{F})\mathcal{R}(z) \\ &\qquad = (I - \mathcal{F}) - i(I - \mathcal{F})\mathfrak{L}\mathcal{F}\mathcal{R}(z), \end{aligned} \tag{2.43}$$

so that

$$\begin{aligned} (I - \mathcal{F})\mathcal{R}(z) = {} & (I - \mathcal{F})\mathcal{S}(z) \\ & \times [(I - \mathcal{F}) - i(I - \mathcal{F})\mathfrak{L}\mathcal{F}\mathcal{R}(z)], \end{aligned} \tag{2.44}$$

where $\mathcal{S}(z)$ is the resolvant of

$$\mathcal{U}'(t) = \exp[-i(I - \mathcal{F})\mathfrak{L}(I - \mathcal{F})t]. \tag{2.45}$$

We can now replace $(I - \mathcal{F})\mathcal{R}(z)$ in the second term of the right-hand side of Eq. (2.41) by Eq. (2.44) and thereby obtain

$$\begin{aligned} \mathcal{F}[z\mathcal{R}(z) - I] &+ i\mathcal{F}\mathfrak{L}\mathcal{F}\mathcal{R}(z) \\ &+ i\mathcal{F}\mathfrak{L}\mathcal{S}(z)(I - \mathcal{F}) \\ &+ \mathcal{F}\mathfrak{L}(I - \mathcal{F})\mathcal{S}(z)(I - \mathcal{F})\mathfrak{L}\mathcal{F}\mathcal{R}(z) = 0. \end{aligned} \tag{2.46}$$

We now operate on $\rho(0)$ with the inverse Laplace transform of Eq. (2.46). Thus

$$\begin{aligned} \frac{d}{dt}\mathcal{F}\rho(t) &+ i\mathcal{F}\mathfrak{L}\mathcal{F}\rho(t) \\ &+ i\mathcal{F}\mathfrak{L}\mathcal{U}'(t)(I - \mathcal{F})\rho(0) \\ &+ \int_0^t dt'\mathcal{F}\mathfrak{L}(I - \mathcal{F})\mathcal{U}'(t - t')(I - \mathcal{F})\mathfrak{L}\mathcal{F}\rho(t') = 0, \end{aligned} \tag{2.47}$$

where $\mathcal{U}'(t)$ has been defined in Eq. (2.45).

It is worth remarking here that the master equation (2.47) follows as an exact consequence of the microscopic dynamics of our total system $(S + R)$. It does not contain any statistical assumption. We will now introduce two assumptions, of statistical nature, concerning the initial state of the total system. We first assume that the systems R and S are initially independent of one another. Thus

$$\rho(0) = \rho_R(0) \otimes \sigma_S(0). \tag{2.48}$$

Secondly, we assume that the initial state of the system R is obtained by measurement of the set of macroscopic observables $\{A_R\}$. Hence,[3]

$$\rho_R(0) = \sum_E p_E(0)F_E, \tag{2.49}$$

and, consequently,

$$\rho(0) = \sum_E p_E(0) F_E \otimes \sigma_S(0). \tag{2.50}$$

We have then

$$\mathcal{P}\rho(0) = \rho(0), \quad \text{i.e.,} \quad (I - \mathcal{P})\rho(0) = 0, \tag{2.51}$$

which implies that the third term of our master equation vanishes.

The resultant master equation is then *formally* similar to the coarse-grained master equation obtained previously[3] for a closed system. There is, however, one significant difference. This is that, whereas the term $\mathcal{P}\mathcal{L}\mathcal{P}\rho(t)$ vanished in the latter case, it is generally nonzero for the case under consideration here. We discuss later the physical meaning of the new term. Let it suffice for the moment to trace back its mathematical origin to the new definition we adopted in this paper for the coarse-graining operator appropriate to the discussion of a system interacting with a reservoir. We want to emphasize that, in spite of the fact that this change in the definition of the projector $\mathcal{P}$ does not affect the *form* of the kernel of the master equation, the new equation describes a very different physical situation and hence, therefore, a significantly different explicit structure.

Some simplifications occur in the above master equation when the structure of the Hamiltonian is specified further. We now suppose that the Hamiltonian of the total system $(S + R)$ splits into three parts:

$$H = H_R^{\times} + H_S^{\times} + H_I, \tag{2.52}$$

where

$$H_R^{\times} = H_R \otimes I_S, \tag{2.53}$$

$$H_S^{\times} = I_R \otimes H_S, \tag{2.54}$$

and

$$H_I = \int_{\Omega_R} dx \int_{\Omega_S} dy V(x, y) J_R(x) \otimes J_S(y). \tag{2.55}$$

Here, H_S and H_R are the Hamiltonians for S and R, respectively, when these two systems are decoupled from one another; H_I is the interaction between these two systems; x and y are, respectively, configuration coordinates for R and S; Ω_R and Ω_S are the volumes occupied, respectively, by R and S; $J_R(x)$ and $J_S(y)$ are operators acting, respectively, on $\mathfrak{H}_R$ and $\mathfrak{H}_S$. These operators represent intensive variables such as, for instance, particle number or current densities in second quantized formalism. The function $V(x, y)$ is a c number and is suitably short-ranged in the difference of its arguments. Thus, our prescription for H_I includes the usual case of static two-body interaction between particles of R and S. More generally, J_R and J_S might be functions of the creation and annihilation operators for the particles in the reservoir and in the system. The arguments carried over in this paper apply trivially to the simpler cases, where H_I reduces to

$$H_I = V_R \otimes V_S.$$

(We do not make this latter simplifying assumption in the sequel.)

Let $\mathcal{L}_R$, $\mathcal{L}_S$, and $\mathcal{L}_I$ be the Liouville operators (acting on $\mathfrak{L}$) corresponding to $H_R^{\times}$, $H_S^{\times}$, and H_I. Then it follows from Eqs. (2.53) and (2.54) that

$$\mathcal{L}_R = \mathcal{L}_R^{(R)} \otimes \mathfrak{I}_S$$

and

$$\mathcal{L}_S = \mathfrak{I}_R \otimes \mathcal{L}_S^{(S)}, \tag{2.56}$$

where $\mathcal{L}_R^{(R)}$, $\mathcal{L}_S^{(S)}$ are the Liouville operators in $\mathfrak{L}_R$, $\mathfrak{L}_S$, respectively, that correspond to the Hamiltonians H_R, H_S; and $\mathfrak{I}_R$, $\mathfrak{I}_S$ are the identity operators for $\mathfrak{L}_R$, $\mathfrak{L}_S$. It will be henceforth assumed that the macroscopic observables $\{A_R\}$, and therefore the projection operators D_E, are constants of the motion for R when it is decoupled from S. Hence, by Eqs. (2.28)–(2.31) and Eq. (2.56)

$$\mathcal{L}_R D_E^{\times} = 0,$$

and

$$\mathcal{L}_R \mathcal{P} = \mathcal{P}\mathcal{L}_R = 0. \tag{2.57}$$

It also follows from Eqs. (2.27) and (2.56) that

$$\mathcal{P}\mathcal{L}_S = \mathcal{L}_S \mathcal{P}. \tag{2.58}$$

Hence, writing $\mathcal{P}\rho(t)$ as

$$\mathcal{P}\rho(t) = \rho_1(t), \tag{2.59}$$

we see that, in view of Eqs. (2.57)–(2.59), the master equation (2.47) reduces to

$$\frac{d}{dt}\rho_1(t) + i(\mathcal{L}_S + \mathcal{P}\mathcal{L}_I\mathcal{P})\rho_1(t) + \int_0^t dt' \mathcal{P}\mathcal{L}_I(I - \mathcal{P})\mathcal{U}'(t - t')(I - \mathcal{P})\mathcal{L}_I\mathcal{P}\rho_1(t') = 0, \tag{2.60}$$

where $\mathcal{U}'(t)$ takes now the simpler form:

$$\mathcal{U}'(t) = \exp\{-i(\mathcal{L}_0' + \mathcal{L}_I')t\}, \tag{2.61}$$

with

$$\mathcal{L}_0' = \mathcal{L}_R + (I - \mathcal{P})\mathcal{L}_S(I - \mathcal{P}) \tag{2.62}$$

and

$$\mathcal{L}_I' = (I - \mathcal{P})\mathcal{L}_I(I - \mathcal{P}). \tag{2.63}$$

$\mathcal{L}_0'$ and $\mathcal{L}_I'$ are obviously Hermitian. We can then develop Eq. (2.61) in the usual perturbation theory[10]

$$\mathcal{U}'(t) = \sum_{n=0}^{\infty} \mathcal{U}_n'(t), \tag{2.64}$$

[10] M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964), Sec. 2.5.

with

$$\mathcal{U}_0'(t) = \exp\{-i\mathcal{L}_0' t\} \tag{2.65}$$

and

$$\mathcal{U}_n'(t) = -i\int_0^t dt' \mathcal{U}_0'(t-t')\mathcal{L}_I'\mathcal{U}_{n-1}'(t'), \quad \text{for} \quad n > 0. \tag{2.66}$$

As it is well known, this series can be rewritten as a time-ordered product

$$\mathcal{U}'(t) = \mathcal{U}_0'(t) T \exp\left\{-i\int_0^t dt' \mathcal{L}_I'(t')\right\}, \tag{2.67}$$

where

$$\mathcal{L}_I'(t) = \mathcal{U}_0'(-t)\mathcal{L}_I'\mathcal{U}_0'(t). \tag{2.68}$$

We can express $\mathcal{U}'(t)$ and $\mathcal{L}_I'(t)$ in more suitable forms by using the properties already observed for our Liouville operators. We first notice that since the three operators $\mathcal{L}_R$, $\mathcal{L}_S$, and $\mathcal{P}$ intercommute, the two terms in $\mathcal{L}_0'$ also commute with one another. Consequently, by Eqs. (2.62) and (2.65),

$$\mathcal{U}_0'(t) = \exp\{-i\mathcal{L}_R t\}\exp\{-i\mathcal{L}_S' t\}, \tag{2.69}$$

with

$$\mathcal{L}_S' = (I-\mathcal{P})\mathcal{L}_S(I-\mathcal{P}). \tag{2.70}$$

Moreover, the second factor in (2.69) can be rewritten as

$$\exp\{-i\mathcal{L}_S' t\} = (I-\mathcal{P})(\mathcal{U}_S(t) - I) + I, \tag{2.71}$$

where

$$\mathcal{U}_S(t) = \exp\{-i\mathcal{L}_S t\} \equiv \mathfrak{I}_R \otimes \mathcal{U}_S^{(S)}(t),$$

with

$$\mathcal{U}_S^{(S)}(t) = \exp\{-i\mathcal{L}_S^{(S)} t\}. \tag{2.72}$$

Using Eqs. (2.58), (2.71), and (2.72), we can now rewrite Eq. (2.69) as

$$\mathcal{U}_0'(t) = \mathcal{P} + (I-\mathcal{P})\mathcal{U}_0(t)(I-\mathcal{P}), \tag{2.73}$$

with

$$\mathcal{U}_0(t) = \exp[-i(\mathcal{L}_R + \mathcal{L}_S)t], \tag{2.74}$$

which obviously commutes with $(I-\mathcal{P})$, a property we shall use shortly hereafter.

Using Eqs. (2.68), (2.73), and (2.74), we obtain

$$\mathcal{L}_I'(t) = (I-\mathcal{P})\mathcal{L}_I(t)(I-\mathcal{P}), \tag{2.75}$$

with

$$\mathcal{L}_I(t) = \mathcal{U}_0(-t)\mathcal{L}_I\mathcal{U}_0(t). \tag{2.76}$$

Let us now define:

$$\mathcal{U}(t) \equiv T\exp\left\{-i\int_0^t dt'(I-\mathcal{P})\mathcal{L}_I(t')(I-\mathcal{P})\right\}, \tag{2.77}$$

which also commutes with $(I-\mathcal{P})$. We have then, by Eqs. (2.67), (2.73), (2.75), and (2.77),

$$(I-\mathcal{P})\mathcal{U}'(t)(I-\mathcal{P}) = \mathcal{U}_0(t)(I-\mathcal{P})\mathcal{U}(t), \tag{2.78}$$

so that in view of Eqs. (2.57), (2.72), (2.74), and (2.76), we can rewrite our master equation (2.60) as

$$\frac{d}{dt}\rho_1(t) + i(\mathcal{L}_S + \mathcal{P}\mathcal{L}_I\mathcal{P})\rho_1(t) + \int_0^t dt' \mathcal{K}(t-t')\rho_1(t') = 0, \tag{2.79}$$

where the kernel $\mathcal{K}$ is the following operator on $\mathfrak{L}$:

$$\mathcal{K}(t) = \mathcal{P}\mathcal{U}_S(t)\mathcal{L}_I(t)(I-\mathcal{P})\mathcal{U}(t)\mathcal{L}_I\mathcal{P}. \tag{2.80}$$

We want to emphasize here that this master equation is an exact consequence of the microscopic dynamics of the total system $(S+R)$, under the assumption that the initial state of $(S+R)$ satisfies Eq. (2.51). In particular, we have not yet assumed any special properties that serve to characterize the reservoir R.

We now anticipate some results concerning the operators $\mathcal{P}\mathcal{L}_I\mathcal{P}$ and $\mathcal{K}(t)$ that will be shown in Sec. 3 to follow from assumptions of a general nature that serve to characterize the reservoir. These results will be stated here in order to enable us to present an uninterrupted derivation of a self-contained master equation for the open system S.

The results are:

(a) $$\mathcal{P}\rho(t) = \rho_R(0) \otimes \sigma_S(t),$$

with

$$\sigma_S(t) = \mathrm{Tr}_R\{\mathfrak{U}(t)\rho_R(0)\otimes\sigma_S(0)\}; \tag{2.81}$$

(b) $$\sum_E p_E(0)\langle V_R(y)\rangle_E F_E = \langle V_R(y)\rangle_0 \rho_R(0),$$

where

$$\langle V_R(y)\rangle_0 = \mathrm{Tr}_R\{\rho_R(0)V_R(y)\}, \tag{2.82}$$

with

$$V_R(y) = \int_{\Omega_R} dx V(x,y) J_R(x).$$

We now want to analyze the implications of these properties on the second term of the left-hand side of our master equation. We first notice that for any B_R in $\mathfrak{L}_R$ and any B_S in $\mathfrak{L}_S$, we have

$$\mathcal{L}_I(B_R\otimes B_S) = \int_{\Omega_R} dx \int_{\Omega_S} dy V(x,y)\{J_R(x)B_R \otimes [J_S(y), B_S]_- + [J_R(x), B_R]_- \otimes B_S J_S(y)\} \tag{2.83}$$

and thus

$$\mathcal{P}\mathcal{L}_I(B_R\otimes B_S) = \int_{\Omega_R} dx \int_{\Omega_S} dy V(x,y)\{\mathcal{P}_R(J_R(x)B_R) \otimes [J_S(y), B_S]_- + \mathcal{P}_R[J_R(x), B_R]_- \otimes B_S J_S(y)\}. \tag{2.84}$$

The second term in the latter integrand vanishes if we specify that B_R is a macroscopic observable A_R on the

reservoir. For such B_R we have, moreover,

$$\mathfrak{I}_R(J_R(x)A_R) = \sum_E A_R(E)\langle J_R(x)\rangle_E D_E. \quad (2.85)$$

Using Eqs. (2.81), (2.82), (2.84), and (2.85), we obtain

$$\mathfrak{I}\mathfrak{L}_I\mathfrak{I}\rho_1(t) = \rho_R(0) \otimes \mathfrak{L}_I^{(S)}\sigma_S(t), \quad (2.86)$$

with

$$\mathfrak{L}_I^{(S)}A_S \equiv [V_S, A_S]_-, \quad \text{for all } A_S \text{ in } \mathfrak{L}_S, \quad (2.87)$$

where

$$V_S = \int_{\Omega_S} dy \langle V_R(y)\rangle_0 J_S(y). \quad (2.88)$$

Consequently, by Eqs. (2.56), (2.81), (2.86), and (2.87), we can then write our master equation (2.79) as

$$\rho_R(0) \otimes \left(\frac{d}{dt} + i\mathfrak{L}_{\text{eff}}^{(S)}\right)\sigma_S(t) = -\int_0^t dt' \mathcal{K}(t - t')\rho_R(0) \otimes \sigma_S(t'), \quad (2.89)$$

with

$$\mathfrak{L}_{\text{eff}}^{(S)} = \mathfrak{L}_S^{(S)} + \mathfrak{L}_I^{(S)} \equiv [H_S + V_S,]_-. \quad (2.90)$$

Taking now the trace over R of both sides of Eq. (2.89), we get the following *reduced master equation for* $\sigma_s(t)$:

$$\left(\frac{d}{dt} + i\mathfrak{L}_{\text{eff}}^{(S)}\right)\sigma_s(t) = -\int_0^t dt' \mathcal{K}^{(S)}(t - t')\sigma_s(t'), \quad (2.91)$$

where the reduced kernel $K^{(s)}$, acting on $\mathfrak{L}_S$, is defined by

$$\mathcal{K}^{(S)}(t)A_S = \text{Tr}_R\{\mathcal{K}(t)\rho_R(0) \otimes A_S\} \quad (2.92)$$

and hence depends in a crucial way on the state $\rho_R(0)$ of the reservoir R. It may be seen that the term $\mathfrak{I}\mathfrak{L}\mathfrak{I}\rho$ of the full master equation (2.47) has given rise to the term $\mathfrak{L}_{\text{eff}}^{(S)}$ in the reduced master equation. In view of Eq. (2.90), this latter term corresponds to the rate of change of σ_S for a conservative system with Hamiltonian $H_S + V_S$. This means that, *if* the master equation (2.91) represents a dissipative law for σ_S, in the sense that it leads to an irreversible approach of σ_S to an equilibrium value, then the dissipation can only arise from the convolutionary term on the right-hand side of the equation. We shall not pursue here the problem of obtaining conditions under which the interaction of S with R leads to a dissipative law for σ_S.

Equation (2.89) is the simplified master equation which we aimed to derive. As will be shown in Sec. 4, this equation can still be significantly simplified if one takes into account more of the macroscopic character of our problem. We note here that the equation constitutes a generalization to open systems of the von Neuman equation for closed systems S:

$$\left(\frac{d}{dt} + i\mathfrak{L}_S^{(S)}\right)\sigma_S(t) = 0.$$

It is immediately seen that our equation differs from von Neuman's in two aspects. *First*, the free Liouville operator $\mathfrak{L}_S^{(S)}$ appearing in von Neuman's equation has been replaced in ours by an effective system-Liouville operator $\mathfrak{L}_{\text{eff}}^{(S)}$ which includes part of the effect of the interaction $\mathfrak{L}_I$. *Secondly*, the zero right-hand side of von Neuman's equation has been replaced by a complicated term, reminiscent in its form of the term appearing in the generalized master equation for a closed system. This formal analogy might, however, be misleading; in particular, our master equation gives the complete evolution of any density matrix for the system S whereas the coarse-grained generalized master equation for closed system gave only the evolution of its coarse-grained part.

3. RESERVOIR PROPERTIES

In this section we specify our assumptions concerning the characterization of the reservoir, and we then use these assumptions to derive the above properties (a) and (b). For simplicity we restrict our analysis to cases where the observables $\{A_R\}$ are taken to be the reservoir energy[11] and smooth functions thereof. Accordingly, the macrocells are chosen to correspond to energy shells for R. Thus, D_E is the projection operator for the subspace $\mathfrak{H}_E$ of $\mathfrak{H}_R$ which is spanned by those eigenstates of H_R in the range $(E, E + \epsilon)$. Consequently,

$$D_E = \int_{-\infty}^{\infty} d\tau \tilde{\chi}_\epsilon(\tau) \exp\{-i(H_R - E)\tau\}, \quad (3.1)$$

where $2\pi\tilde{\chi}_\epsilon(\tau)$ is the Fourier transform of the characteristic function $\chi_\epsilon(E)$ on the interval $(0, \epsilon)$ [i.e., $\chi_\epsilon(E) = 1$ for E in $(0, \epsilon)$ and is zero elsewhere]. It is readily verified that the application of D_E, as specified by (3.1), to an eigenstate of H_R yields 1 or 0 according to whether or not the energy lies in $(E, E + \epsilon)$. The magnitude of ϵ corresponds to the limit of precision for our idealized energy measurements. Its value is open to a certain amount of choice, though it must be extremely small by comparison with macroscopic energies, such as the thermal energy of R. We stipulate here that ϵ is equal to some characteristic energy quantum for a *microscopic* excitation—e.g., if R is a crystal, ϵ could be chosen as the energy of a Debye quantum. The essential point

[11] This simply means that we are considering a situation where the reservoir is initially prepared by measurement of its energy or, equivalently, its temperature.

here is that $\epsilon = O(N^0)$, where N is the number of particles of R.

We now introduce our assumptions that are designed to characterize the reservoir in the thermodynamical limit where its volume and the number of its particles both tend to infinity in such a way that their ratio remains finite. These assumptions, which are intended to be applicable to a large class of situations of physical interest, are centered on the Heisenberg operators

$$J_R(x, t) = \exp(iH_R t) J_R(x) \exp(-iH_R t), \quad (3.2)$$

the interaction "potential" $V(x, y)$, and the initial state $\rho_R(0)$. The assumptions are that:

(1) The microcanonical average of products $J_R(x, t) \cdots J_R(x_n, t_n)$ for states F_E, and similar states specified below, depends on E and N only through the *intensive* variable E/N. Thus

$$\mathrm{Tr}_R \{J_R(x_1, t_1) \cdots J_R(x_n, t_n) F_E\} = \Phi^{(n)}\left(\frac{E}{N} \middle| x_1, \cdots, x_n; t_1, \cdots, t_n\right), \quad (3.3)$$

where the function $\Phi^{(n)}(z \mid x_1, \cdots, x_n; t_1, \cdots, t_n)$ is independent of N. This assumption is generally satisfied for cases where J_R corresponds to a current or number density because (i) the product

$$J_R(x_1, t_1) \cdots J_R(x_n, t_n)$$

is an intensive quantity, and (ii) in the thermodynamic limit, its microcanonical average is equal to its canonical average, which depends on E, N only via the temperature, an intensive variable.

(2) The interaction "potential" $V(x, y)$ is of sufficiently short range to ensure that integrals

$$\int_{\Omega_R} dx_1 \cdots \int_{\Omega_R} dx_n \Phi^{(n)}(z \mid x_1, \cdots, x_n; t_1, \cdots, t_n) \times V(x_1, y_1) \cdots V(x_n, y_n)$$

tend to functions of $z; y_1, \cdots, y_n; t_1, \cdots, t_n$, independent of N, in the thermodynamic limit. The same is assumed if the set $(x_1, \cdots, x_n)$ is permuted to $(x_{l_1}, \cdots, x_{l_n})$ in $\Phi^{(n)}$.

(3) The initial energy distribution for R, as defined in (2.49), is of the form

$$p_E(0) = f_0\left(\frac{E - E_0}{N^{\frac{1}{2}}}\right), \quad (3.4)$$

where $f_0(\eta)$ is an N-independent function (apart from a normalizing constant) of finite dispersion. This form is appropriate, for example, in cases where R is initially in canonical equilibrium, provided that the reservoir is in one thermodynamical phase only, and that critical fluctuations can be discounted. Our deduction of statements (a) and (b) from these assumptions will be based on an analysis in which we discriminate throughout between intensive and extensive variables. We start by noting that, in view of Eqs. (2.5) and (2.37),

$$\mathfrak{I}\rho(t) = \sum_E F_E \otimes \sigma_E(t), \quad (3.5)$$

where

$$\sigma_E(t) = \mathrm{Tr}_R \{D_E^{\times} \mathfrak{U}(t) \rho(0)\}. \quad (3.6)$$

We now express Eq. (2.7), for $\mathfrak{U}(t)$, in the well-known interaction representational[10] form

$$\mathfrak{U}(t) = \mathfrak{U}_0(t) \sum_{n=0}^{\infty} (-i)^n \int_0^t dt_1 \times \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \mathfrak{L}_I(t_1) \cdots \mathfrak{L}_I(t_n), \quad (3.7)$$

where

$$\mathfrak{U}_0(t) = \exp[-i(\mathfrak{L}_R + \mathfrak{L}_S)t] \quad (3.8)$$

and

$$\mathfrak{L}_I(t) = \mathfrak{U}_0(-t) \mathfrak{L}_I \mathfrak{U}_0(t). \quad (3.9)$$

This latter equation may be seen to be equivalent to (2.76). Further, it follows from Eqs. (2.56), (2.57), and (3.8) that

$$\mathrm{Tr}_R \{D_E^{\times} \mathfrak{U}_0(t) \cdots\} \equiv e^{-i\mathfrak{L}_S^{(S)} t} \mathrm{Tr}_R \{D_E^{\times} \cdots\}. \quad (3.10)$$

Consequently, by Eqs. (3.7)–(3.10), we may re-express Eq. (3.6) in the form

$$\sigma_E(t) = \exp(-i\mathfrak{L}_S^{(S)} t) \sum_{n=0}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \times \mathrm{Tr}_R \{D_E^{\times} \mathfrak{L}_I(t_1) \cdots \mathfrak{L}_I(t_n) \rho(0)\}. \quad (3.11)$$

Moreover, it follows from Eqs. (2.55), (2.56), (3.8), and (3.9) that

$$\mathfrak{L}_I(t) = \sum_{j=\pm 1} \int_{\Omega_R} dx \int_{\Omega_S} dy V(x, y) \mathfrak{L}_I^{(R), j}(x, t) \otimes \mathfrak{L}_I^{(S), -j}(y, t), \quad (3.12)$$

where

$$\mathfrak{L}_I^{(R), \pm 1}(x, t) B_R = \frac{1}{\sqrt{2}} [J_R(x, t), B_R]_{\pm}, \quad (3.13)$$

$$\mathfrak{L}_I^{(S), \pm 1}(y, t) B_S = \frac{1}{\sqrt{2}} [J_S(y, t), B_S]_{\pm}, \quad (3.14)$$

$$J_S(y, t) = \exp(iH_s t) J_S(y) \exp(-iH_s t), \quad (3.15)$$

and $J_R(x, t)$ is defined by Eq. (3.2). On inserting Eq. (3.12) into Eq. (3.11) and using the initial condition given by Eqs. (2.48) and (2.49) we obtain

$$\sigma_E(t) = \exp(-i\mathfrak{L}_S^{(S)} t) \sum_{n=0}^{\infty} \sum_{\{j\}} \int_0^t dt_1 \cdots \int_0^{t_{n-1}} dt_n \times \int_{\Omega_S} dy_1 \cdots \int_{\Omega_S} dy_n \times \sum_{E'} \Gamma_{EE'}^{(n), \{j\}}(y_1, \cdots, y_n \mid t_1, \cdots, t_n) p_{E'}(0) \times \mathfrak{L}_I^{(S), -j}(y_1, t_1) \cdots \mathfrak{L}_I^{(S), -j_n}(y_n t_n) \sigma_S(0), \quad (3.16)$$

where

$$\Gamma_{EE'}^{(n)\{j\}}(y_1, \cdots, y_n \mid t_1, \cdots, t_n) = \int_{\Omega_R} dx_1 \cdots \int_{\Omega_R} dx_n V(x_1, y_1) \cdots V(x_n, y_n) \times \gamma_{EE'}^{(n)\{j\}}(x_1, \cdots, x_n \mid t_1, \cdots, t_n) \quad (3.17)$$

and

$$\gamma_{EE'}^{(n)\{j\}}(x_1, \cdots, x_n \mid t_1, \cdots, t_n) = \mathrm{Tr}_R \{D_E \mathfrak{L}_I^{(R),j_1}(x_1, t_1) \cdots \mathfrak{L}_I^{(R),j_n}(x_n, t_n) F_{E'}\}. \quad (3.18)$$

It may be seen from Eqs. (3.13), (3.14), and (3.18) that each $\gamma_{EE'}^{(n)\{j\}}$ is a sum of terms of the form

$$\theta_{EE'} = \mathrm{Tr}_R \{B_R^{(1)} D_E B_R^{(2)} F_{E'}\}, \quad (3.19)$$

where the B_R's are products of Heisenberg operators $J_R(x, t)$; i.e.,

$$B_R^{(1)} = J_R(x_1', t_1') \cdots J_R(x_l', t_l') \quad (3.20)$$

and

$$B_R^{(2)} = J_R(x_1'', t_1'') \cdots J_R(x_m'', t_m''), \quad \text{say.}$$

Our deduction of the statements (a) and (b) will be centered on an analysis of the dependence of $\theta_{EE'}$, and hence of $\Gamma_{EE'}$, on E, E', and N. The key property of Γ, which we seek to obtain, is that its dependence on the above variables takes the form

$$\Gamma_{EE'}^{(n)\{j\}}(y_1, \cdots, y_n \mid t_1, \cdots, t_n) = \Psi_{\{j\}}\left(\frac{E'}{N}, E' - E \mid y_1, \cdots, y_n; t_1, \cdots, t_n\right), \quad (3.21)$$

where the function $\Psi(u, v \mid \cdots)$ is independent of N.

In order to establish this property for Γ we divide the energy shells $(E, E + \epsilon)$ into subshells $(E + \nu, E + \nu + \delta)$. Thus we have

$$D_E = \sum_\nu D_{E\nu} \quad (3.22)$$

and

$$F_E = \delta \sum_\nu g_{E\nu} F_{E\nu}, \quad (3.23)$$

where $D_{E\nu}$ is the projection operator for a subshell, and

$$F_{E\nu} = D_{E\nu}/W_{E\nu}; \quad \delta g_{E\nu} = W_{E\nu}/W_E; \quad W_{E\nu} = \mathrm{Tr}_R (D_{E\nu}). \quad (3.24)$$

We shall choose δ so that $W_{E\nu}$, the number of eigenstates of R in the subshell, tends to infinity in the thermodynamic limit, while δ/ϵ tends to zero in this same limit. These two requirements for δ may be simultaneously fulfilled as $W_{E\nu}$ is equal to δe^S where S, the entropy of the system, is proportional to N. Consequently, we could satisfy the above conditions on δ by choosing $\delta = \epsilon/N$, for example.

As the quantity $g_{E\nu}$ occurs in $\gamma_{EE'}$, we now examine its N-dependence. It follows from Eq. (3.24) that $g_{E\nu}$ may be expressed in the form

$$g_{E\nu} = \exp [S(E + \nu) - S(E)], \quad (3.25)$$

where $S(E)$ is the entropy expressed as a function of energy. Thus, using the thermodynamic relation $T\,dS = dE$, we have

$$S(E + \nu) - S(E) = \int_0^\nu \frac{d\nu'}{T(E + \nu')}. \quad (3.26)$$

As the temperature $T(E)$ is an intensive variable, it may be expressed as a function of E/N only. Hence we may replace $T(E + \nu')$ by $T(E)$ in the above integral, since $\nu'/N \to 0$ in the thermodynamic limit. Consequently, it follows from Eqs. (3.25) and (3.26) that $g_{E\nu}$ is an intensive variable, given by

$$g_{E\nu} = \exp(\nu/T) \equiv g_0\left(\nu, \frac{E}{N}\right), \quad \text{say.} \quad (3.27)$$

We now insert the formulas for D_E and $F_{E'}$, given by Eqs. (3.1) and (3.23), into Eq. (3.19) for θ. Thus we obtain

$$\theta_{EE'} = \int_{-\infty}^{\infty} d\tau \tilde{\chi}_\epsilon(\tau) \cdot \delta \cdot \sum_\nu g_{E'\nu} \times \mathrm{Tr}_R \{B_R^{(1)}(\tau) B_R^{(2)} F_{E'\nu} e^{-iH_R\tau}\}, \quad (3.28)$$

where

$$B_R^{(1)}(\tau) = \exp(iH_R\tau) B_R^{(1)} \exp(-iH_R\tau). \quad (3.29)$$

As we have specified that δ, the width of a subshell, becomes infinitesimal as the thermodynamical limit is approached, we may replace $F_{E'\nu} \exp(-iH_R\tau)$ by $F_{E'\nu} \exp\{-i(E' + \nu)\tau\}$ in Eq. (3.28). Hence,

$$\theta_{EE'} = \int_{-\infty}^{\infty} d\tau \tilde{\chi}_\epsilon(\tau) \exp\{i(E' - E)\tau\} \tilde{g}_0(E', \tau) \times \mathrm{Tr}_R \{B_R^{(1)}(\tau) B_R^{(2)} F_{E'\nu}\}, \quad (3.30)$$

with

$$\tilde{g}_0(E', \tau) = \delta \sum_\nu g_{E'\nu} e^{-i\nu\tau};$$

i.e., by (3.27),

$$\tilde{g}_0(E', \tau) = \delta \sum_\nu g_0\left(\nu, \frac{E'}{N}\right) e^{-i\nu\tau}. \quad (3.31)$$

As δ, which is both the first factor in this expression and the spacing between values of ν, tends to zero in the thermodynamic limit, we may replace $\delta \sum_\nu$ by $\int d\nu$ in Eq. (3.31). Hence

$$\tilde{g}_0(E', \tau) = \int_0^\epsilon d\nu g_0\left(\nu, \frac{E'}{N}\right) e^{-i\nu\tau} \equiv \varphi\left(\frac{E'}{N}, \tau\right), \quad \text{say,} \quad (3.32)$$

which depends on E, E', N only through E'/N. On

inserting this formula for $\tilde{g}_0$ into (3.30), we obtain

$$\theta_{EE'} = \int_{-\infty}^{\infty} d\tau \tilde{\chi}_\epsilon(\tau) \exp\{i(E' - E)\tau\} \varphi\left(\frac{E'}{N}, \tau\right) \times \mathrm{Tr}_R\{B_R^{(1)}(\tau) B_R^{(2)} F_{E'\nu}\}. \quad (3.33)$$

Further, it follows from Eqs. (3.20) and (3.29) that the trace term in (3.33) is an expectation value of a product of Heisenberg operators J_R; i.e.,

$$\mathrm{Tr}_R[B_R^{(1)}(\tau) B_R^{(2)} F_{E'\nu}] = \mathrm{Tr}_R[J_R(x_{l_1}, t_{l_1}) \cdots J_R(x_{l_n}, t_{l_n}) F_{E'\nu}]. \quad (3.34)$$

We now complete our specification of assumption (1) by asserting that the "other mixed states" referred to therein are the set $\{F_{E'\nu}\}$. Thus it follows that the right-hand side of (3.34) depends on E', ν, N only through the intensive variable $(E' + \nu)/N$, and this may be replaced by E'/N in the thermodynamical limit. Hence, applying assumption (1) to Eq. (3.34),

$$\mathrm{Tr}_R\{B_R^{(1)}(\tau) B_R^{(2)} F_{E'\nu}\} = \Phi^{(n)}\left(\frac{E'}{N} \middle| x_{l_1}, \cdots, x_{ln}; t_{l_1}, \cdots, t_{l_n}\right), \quad (3.35)$$

the latter function depending on E', N only via E'/N. Further, as $\theta_{EE'}$ has been specified to be one of the terms contributing to the function $\gamma_{EE'}$, defined by Eq (3.18), it is evident from our analysis leading from (3.28) to (3.33) and (3.35) that each (x_{l_j}, t_{l_j}) appearing in the latter equation corresponds either to an (x_m, t_m) or to an $(x_m, t_m + \tau)$. Consequently we denote the right-hand side of Eq. (3.35) by

$$\tilde{\Phi}_\tau^{(n)}\left(\frac{E'}{N} \middle| x_1, \cdots, x_n; t_1, \cdots, t_n\right), \quad (3.35')$$

a function possessing all properties postulated with regard to $\Phi^{(n)}$ in assumption (1).

On replacing the right-hand side of Eq. (3.35) by Eq. (3.35') and inserting the resultant formula for $\mathrm{Tr}_R\{\cdots\}$ into Eq. (3.33), we obtain

$$\theta_{EE'} = \int_{-\infty}^{\infty} d\tau \tilde{\chi}_\epsilon(\tau) \exp\{i(E' - E)\tau\} \varphi\left(\frac{E'}{N}, \tau\right) \times \tilde{\Phi}_\tau^{(n)}(x_1, \cdots, x_n \mid t_1, \cdots, t_n). \quad (3.36)$$

Since $\gamma_{EE'}$ is a sum of terms $\theta_{EE'}$, it follows from Eqs. (3.17) and (3.36) that $\Gamma_{EE'}^{(n)\{j\}}$ is a sum of contributions

$$\int_{-\infty}^{\infty} d\tau \int_{\Omega_R} dx_1 \cdots \int_{\Omega_R} dx_n \varphi\left(\frac{E'}{N}, \tau\right) \tilde{\chi}_\epsilon(\tau) \exp\{i(E' - E)\tau\} \times V(x_1, y_1) \cdots V(x_n, y_n) \tilde{\Phi}_\tau^{(n)}(x_1, \cdots, x_n \mid t_1, \cdots, t_n). \quad (3.37)$$

As $\tilde{g}$ and $\tilde{\chi}$ are intensive variables, it is evident that the application of assumption (2) to this expression leads directly to the desired result (3.21).

We now turn our attention to the partial sum $\sum_{E'} \Gamma_{EE'} p_{E'}(0)$ which appears in (3.16). Using assumption (3) and Eq. (3.21) we see that

$$\sum_{E'} \Gamma_{EE'} p_{E'}(0) = \sum_{\Delta E} \Psi(\epsilon_0 + N^{-\frac{1}{2}}\eta - N^{-1}\Delta E; \Delta E) f_0(\eta + N^{-\frac{1}{2}}\Delta E), \quad (3.38)$$

where

$$\epsilon_0 = E_0/N; \quad \eta = \frac{E - E_0}{N^{\frac{1}{2}}}, \quad \Delta E = E' - E, \quad (3.39)$$

and the variables x, y, t, j have been suppressed. The variable η represents energy fluctuations about E_0 on a scale appropriate for "large" systems $(N \to \infty)$. Accordingly, we take the thermodynamic limit for the right-hand side of (3.38) as corresponding to $N \to \infty$, ϵ_0 and η finite. Hence, in this limit,

$$\sum_{E'} \Gamma_{EE'} p_{E'}(0) \to \sum_{\Delta E} \Psi(\epsilon_0; \Delta E) f_0(\eta) \equiv \sum_{E'} \Psi\left(\frac{E_0}{N}; E' - E_0\right) p_E(0) \equiv \sum_{E'} \Gamma_{E'E_0} p_E(0). \quad (3.40)$$

It should be noted that the passage from Eqs. (3.38) to (3.39) involves a change of the variable of summation since $\Delta E = (E' - E)$ in the former equation and $(E' - E_0)$ in the latter one. It follows from Eqs. (3.17), (3.18), and (3.40) that

$$\sum_{E'} \Gamma_{EE'}^{(n)\{j\}}(y_1, \cdots, y_n \mid t_1, \cdots, t_n) p_{E'}(0) = p_E(0) \int_{\Omega_R} dx_1 \cdots \int_{\Omega_R} dx_n V(x_1, y_1) \cdots V(x_n, y_n) \times \mathrm{Tr}_R\{\mathfrak{L}_I^{(R),j_1}(x_1, t_1) \cdots \mathfrak{L}_I^{(R),j_n}(x_n, t_n) F_{E_0}\}. \quad (3.41)$$

Further, it follows from Eq. (3.13) that the term $\mathrm{Tr}_R\{\cdots\}$ in (3.41) in a sum of contributions of the form

$$\mathrm{Tr}_R\{J_R(x_1', t_1') \cdots J_R(x_n', t_n') F_{E_0}\}, \quad (3.42)$$

where the set $(x_1', t_1') \cdots (x_n' t_n')$ is some permutation of $(x_1, t_1) \cdots (x_n, t_n)$. By assumptions (2) and (3), the statistical operator F_{E_0} may be replaced by $\rho_R(0)$ in Eq. (3.42) and, therefore, also in Eq. (3.41). Hence, the latter equation may be rewritten as

$$\sum_{E'} \Gamma_{EE'}^{(n)\{j\}}(y_1, \cdots, y_n \mid t_1, \cdots, t_n) p_{E'}(0) = p_E(0) \int_{\Omega_R} dx_1 \cdots \int_{\Omega_R} dx_n V(x_1, y_1) \cdots V(x_n, y_n) \times \mathrm{Tr}_R\{\mathfrak{L}_I^{(R),j_1}(x_1, t_1) \cdots \mathfrak{L}_I^{(R),j_n}(x_n, t_n) \rho_R(0). \quad (3.43)$$

We assume now that the thermodynamical limit of the sum of terms for σ_E, as given by Eq. (3.16), is equal to the sum of the limits of those terms. Hence, inserting Eq. (3.43) into Eq. (3.16), and then using Eq. (3.5) to obtain $\mathfrak{T}\rho(t)$, we find that

$$\begin{aligned}\mathfrak{T}\rho(t) = \sum_E p_E(0)F_E \otimes \sum_{n=0}^{\infty}\sum_{\{j\}}(-i)^n\int_0^t dt_1\cdots\int_0^{t_{n-1}}dt_n \\ \times\int_{\Omega_R}dx_1\cdots\int_{\Omega_R}dx_n\int_{\Omega_S}dy_1\cdots\int_{\Omega_S}dy_n \\ \times \mathrm{Tr}_R\{\mathfrak{L}_I^{(R),j_1}(x_1,t_1)\cdots\mathfrak{L}_I^{(R),j_n}(x_n,t_n)\} \\ \times e^{-i\mathfrak{L}_S^{(S)}t}\mathfrak{L}_I^{(S),-j_1}(x_1,t_1)\cdots\mathfrak{L}_I^{(S),-j_n}(x_n,t_n)\sigma_S(0).\end{aligned} \tag{3.44}$$

It may be seen from Eqs. (2.49), (2.56), (3.7), (3.8), and (3.12) that Eq. (3.44) is exactly equivalent to the pair of equations, namely (2.81), which constitute assertion (a). This completes our proof of that assertion.

The proof of assertion (b) is much simpler. Because if we use the method we employed to obtain Eq. (3.43) from Eq. (3.40), we may replace $\langle V_R(y)\rangle_E p_E(0)$, as defined in (2.82), by $\langle V_R(y)\rangle_{E_0}p_E(0)$ and thence by $\langle V_R(y)\rangle_0 p_E(0)$. Assertion (b) then follows immediately.

4. FURTHER SIMPLICATION OF THE MASTER EQUATION FOR $\sigma_S(t)$

In this section we make further use of the assumptions of Sec. 3 in order to reduce the kernel $\mathcal{K}^{(S)}$ of the master equation for σ_S, to a more tractable form. It will be shown that the projector $\mathfrak{T}$ which appears in the kernel of that equation, can be replaced by a simpler operator $\tilde{\mathfrak{T}}$.

This operator $\tilde{\mathfrak{T}}$ will be defined as acting on a space $\mathfrak{T}$. This space will be defined as a subset of $\mathfrak{L}$, which will be shown to be large enough to allow a complete description of our operations. In fact, $\mathfrak{T}$ is defined, in a way similar to that used in Sec. 2 for the specification of $\mathfrak{L}$, as

$$\mathfrak{T} = \mathfrak{T}_R \otimes \mathfrak{L}_S, \tag{4.1}$$

where $\mathfrak{T}_R$ is the subset of $\mathfrak{L}_R$ formed by all elements C_R of $\mathfrak{L}_R$ that have a finite trace.[12] We now define $\tilde{\mathfrak{T}}$ as

$$\tilde{\mathfrak{T}} = \tilde{\mathfrak{T}}_R \otimes \mathfrak{I}_S, \tag{4.2}$$

with

$$\tilde{\mathfrak{T}}_R C_R = \rho_R(0)\,\mathrm{Tr}_R\,C_R, \text{ for all } C_R \text{ in } \mathfrak{T}_R. \tag{4.3}$$

We now analyze the kernel $\mathcal{K}^{(S)}$ of the master equation and show that, on the basis of the assumptions of Sec. 3, we may replace $\mathfrak{T}$ by $\tilde{\mathfrak{T}}$ in that kernel. The advantage of this replacement is that it permits one to express $\mathcal{K}^{(S)}$ in terms of correlations between Heisenberg operators $J_R(x,t)$ in the single state $\rho_R(0)$, whereas the use of $\mathfrak{T}$ would involve averages and correlations over the full set of states F_E (cf. Ref. 4).

The kernel $\mathcal{K}^{(S)}$ is an operator on $\mathfrak{L}_S$. By Eqs. (2.27), (2.77), (2.80), and (2.92), we may express $\mathcal{K}^{(S)}$ in the form

$$\mathcal{K}^{(S)}(t) = \sum_{n=0}^{\infty}(-i)^n\mathcal{K}_n^{(S)}(t), \tag{4.4}$$

where the action of $\mathcal{K}_n^{(S)}$ on any element A_S of $\mathfrak{L}_S$ is given by

$$\begin{aligned}\mathcal{K}_n^{(S)}(t)A_S = \int_0^t dt_1\int_0^t dt_2\cdots\int_0^{t_{n-1}}dt_n \\ \times \mathrm{Tr}_R\{\mathfrak{T}\mathcal{U}_S(t)\mathfrak{L}_I(t)(I-\mathfrak{T})\mathfrak{L}_I(t_1)\cdots(I-\mathfrak{T}) \\ \times\mathfrak{L}_I(t_n)(I-\mathfrak{T})\mathfrak{L}_I\rho_R(0)\otimes A_S\}.\end{aligned} \tag{4.5}$$

Hence, by Eqs. (2.72), (3.12), and (4.5),

$$\begin{aligned}\mathcal{K}_n^{(S)}(t) = \int_0^t dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n \\ \times\int_{\Omega_R}dx\int_{\Omega_R}dx_1\cdots\int_{\Omega_R}dx_n\int_{\Omega_R}dx_0 \\ \times\int_{\Omega_S}dy\int_{\Omega_S}dy_1\cdots\int_{\Omega_S}dy_n\int_{\Omega_S}dy_0 \\ \times V(x,y)V(x_1,y_1)\cdots V(x_n,y_n)V(x_0,y_0) \\ \times\sum_{\{j\}}F_n^{\{j\}}(x,x_1,\cdots,x_n,x_0\,|\,t,t_1,\cdots,t_n) \\ \times\mathcal{U}_S^{(S)}(t)\mathfrak{L}_I^{(S),-j}(y,t)\mathfrak{L}_I^{(S),-j_1}(y_1,t_1)\cdots \\ \mathfrak{L}_I^{(S),-j_n}(y_n,t_n)\mathfrak{L}_I^{(S),-j_0}(y_0,0),\end{aligned} \tag{4.6}$$

where

$$\begin{aligned}F_n^{\{j\}}(x,x_1,\cdots,x_n,x_0\,|\,t,t_1,\cdots,t_n) \\ = \mathrm{Tr}_R\{\mathfrak{T}_R\mathfrak{L}_I^{(R),j}(x,t)(I-\mathfrak{T}_R) \\ \times\mathfrak{L}_I^{(R),j_1}(x_1,t_1)\cdots(I-\mathfrak{T}_R)\mathfrak{L}_I^{(R),j_n}(x_n,t_n) \\ \times(I-\mathfrak{T}_R)\mathfrak{L}_I^{(R),j_0}(x_0,0)\rho_R(0)\}.\end{aligned} \tag{4.7}$$

Thus the $F_n^{\{j\}}$'s are sums of terms of the form

$$\mathrm{Tr}_R\{\mathfrak{T}_R\mathcal{M}_R^{(1)}\mathfrak{T}_R\mathcal{M}_R^{(2)}\mathfrak{T}_R\cdots\mathfrak{T}_R\mathcal{M}_R^{(l)}\rho_R(0)\}, \tag{4.8}$$

where each $\mathcal{M}_R$ is a product of operators $\mathfrak{L}_I^{(R)}(x_m,t_m)$. On applying to $\mathfrak{T}_R\mathcal{M}_R\rho_R(0)$ the same analysis we used to obtain Eq. (3.44) from Eq. (3.40), we find that

$$\mathfrak{T}_R\mathcal{M}_R\rho_R(0) = \tilde{\mathfrak{T}}_R\mathcal{M}_R\rho_R(0), \tag{4.9}$$

where $\tilde{\mathfrak{T}}_R$ is defined by Eq. (4.3). Hence, by Eqs. (4.3) and (4.9),

$$\mathfrak{T}_R\mathcal{M}_R\tilde{\mathfrak{T}}_R = \tilde{\mathfrak{T}}_R\mathcal{M}_R\tilde{\mathfrak{T}}_R. \tag{4.10}$$

Consequently, by Eqs. (4.9) and (4.10), we may replace $\mathfrak{T}_R$ by $\tilde{\mathfrak{T}}_R$ throughout (4.8) and therefore throughout (4.7). This means that $\mathfrak{T}$ may be replaced by $\tilde{\mathfrak{T}}$ whenever it occurs in our formulas for $\mathcal{K}^{(S)}$.

[12] We might note in passing that a norm, but not a scalar product, can be defined on $\mathfrak{T}_R$ and, hence, on $\mathfrak{T}$. Thus $\mathfrak{T}$ is not a Hilbert space but a Banach space. This does not lead to any difficulty in what follows.

Consequently, by Eqs. (2.77), (2.80), (2.91), (2.92), (4.2), and (4.3), we may rewrite our master equation as

$$\left(\frac{d}{dt} + i\mathfrak{L}_{\text{eff}}^{(S)}\right)\sigma_S(t) + \int_0^t dt' \mathcal{K}^{(S)}(t-t')\sigma_S(t') = 0, \quad (4.11)$$

where

$$\tilde{\mathcal{K}}_S(t)A_S = \mathrm{Tr}_R\Big\{\mathcal{U}_S(t)\mathfrak{L}_I(t) \times T\exp\Big\{-i\int_0^t dt'(I-\tilde{\mathcal{P}})\mathfrak{L}_I(t')(I-\tilde{\mathcal{P}})\Big\} \times \mathfrak{L}_I\rho_0 \otimes A_S\Big\}. \quad (4.12)$$

We commented at the beginning of this section on the advantage of replacing $\mathcal{P}$ by $\tilde{\mathcal{P}}$. We wish to point out that our justification of this substitution rested on assumptions (1)–(3) which served to characterize the reservoir. If one overlooked the dependence of the master equation on those assumptions, one might try to derive it directly from the Liouville equation by using the operator $\tilde{\mathcal{P}}$, instead of $\mathcal{P}$, right from the beginning. We see, however, that this would be unjustified, in general, as $\tilde{\mathcal{P}}$ does not act on the whole Liouville space $\mathfrak{L}$. Thus, the erroneous *ab initio* use of $\tilde{\mathcal{P}}$ would reflect an inadequate consideration of the physical characteristics of the reservoir which are necessary for a replacement of $\mathcal{P}$ by $\tilde{\mathcal{P}}$. In the present treatment, the essential effect of these characteristics has been to confine $\mathcal{P}\rho$ to $\mathfrak{T}$.

5. CONCLUSION

We have derived a generalized master equation (4.11) for open systems. This equation has been obtained by a systematic treatment of the microscopic equations of motion governing a system of interest S and a reservoir R to which it is coupled. The derivation of the master equation has rested on the general principles of quantum theory, supplemented by the initial condition (2.48) and by the assumptions (1)–(3) concerning the reservoir and its coupling to S. This latter coupling appears in the master equation (4.11) only in the forms of thermal averages, taken over the initial state of the reservoir. The master equation, therefore, corresponds to a self-contained dynamical law for S, and may be regarded as a generalization to open systems of the quantum-mechanical Liouville equation.

Our derivation of the master equation has been based on an extension of Zwanzig's projective technique. It is worth emphasing that this technique must be based on the use of a projection operator $\mathcal{P}$ which acts in the appropriate Liouville space. The significance of this apparently trivial remark may be seen from our discussion in Sec. 4, where it was shown that the theory could not have been correctly founded on the use of the operator $\tilde{\mathcal{P}}$ $(= \tilde{\mathcal{P}}^2)$, *ab initio*, since this operator does not act on the whole of the Liouville space $\mathfrak{L}$. Thus, it is only in consequence of special properties of R, introduced in Sec. 3, that we are finally able to replace $\mathcal{P}$ by $\tilde{\mathcal{P}}$ in the master equation.

Having formulated the generalized master equation for S, we envisage that it can serve as a basis for further research along two principal lines. First, it could be used in the investigation of general problems, such as irreversibility and ergodicity, which have not been considered in this paper. Secondly, it could be used as a tool for investigation of specific non-equilibrium phenomena such as transport processes, Brownian motion, and spin-lattice relaxation, for example.

ACKNOWLEDGMENT

One of us (G. L. S.) wishes to express his gratitude to Professor E. W. Montroll for the hospitality he received as a visitor to the University of Rochester, where this research was carried out.

Matrix Elements of Operators of Class I Representations of $SL(n, C)$

C. FRONSDAL* AND HARUN AR-RASHID†
International Centre for Theoretical Physics, Trieste, Italy

(Received 13 August 1966)

Let D be an irreducible representation of the group $SL(n, C)$ by operators T_g in a Hilbert space H. We introduce a canonical basis of vectors $h \in H$, and show how to evaluate the "D-functions" $(h_1, T_g h_2)$. It has been suggested that one such function has the physical interpretation of the electric-charge form factor of the pion. This function is evaluated in detail.

1. INTRODUCTION

Let R be the group of rotations in three dimensions; let D_l be the irreducible representation of R in a $2l + 1$ dimensional space H, and let T_g be the linear operator in H that represents the rotation $g \in R$. One of the most important objects that a physicist has to calculate, to determine the consequences of rotational invariance, is the matrix element

$$(h_1, T_g h_2),$$

where h_1 and h_2 are vectors in H. If h_m, $m = -l, -l + 1, \cdots, l$ is the usual basis in H, then a particular case is the D function

$$D^l_{m,m'}(g) = (m', T_g m).$$

Although many other Lie groups have been applied to physical problems, the corresponding spherical functions or D functions have been of slight interest to physicists. However, it has recently been suggested that certain particular D functions, related to unitary irreducible representations of $SL(n, C)$ may have direct physical meaning.[1–3]

Let G denote the group $SL(n, C)$ considered as a Lie group with $2(n^2 - 1)$ real parameters. The maximal compact subgroup, which is isomorphic to $SU(n)$, will be denoted by the letter K. Let D be an irreducible representation of G by linear operators T_g, $g \in G$, in a Hilbert space H, and let D_K be the restriction of D to K. Then D_K is completely reducible to a direct sum of irreducible representations $D_K(m)$:

$$D_K = \sum_{(m)} \oplus \gamma_m D_K(m). \tag{1.1}$$

Here γ_m is a set of nonnegative integers, and $D_K(m)$ is the usual nomenclature for irreducible representations of $SU(n)$ in terms of a set $m = (m_1, m_2, \cdots, m_{n-1})$ of $n - 1$ nonnegative integers.

Corresponding to (1.1) there is, of course, a canonical reduction of H

$$H = \sum_{(m)} \oplus H_m, \tag{1.2}$$

where H_m is a closed, finite-dimensional subspace of H, that admits a γ_m-fold copy of the irreducible representation $D_K(m)$ of the maximal compact subgroup. A basis may be introduced in H by selecting an orthonormal basis in each one of the subspaces H_m; such a basis will be called a *canonical basis.* A basis is canonical if and only if every basis vector belongs to a closed subspace of H that transforms under K according to a γ_m-fold copy of $D_K(m)$.

Definition I: A function over G is called a "D-function" if it is one of the functions

$$D^{(D)}_{1,2}(g) = (h_1, T_g h_2), \tag{1.3}$$

where D is an irreducible representation of G by linear operators T_g in a Hilbert space H, and h_1, h_2 are two vectors belonging to a canonical basis in H.

Definition II: A function $\varphi_m(g)$ is called a "spherical function of height γ_m" if it is one of the functions

$$\varphi^{(D)}_m(g) = \sum_{h \in H_m} D^{(D)}_{h,h}(g) = \sum_{h \in H_m} (h, T_g h),$$

where the sum goes over a complete orthonormal basis in H_m. If $E(m)$ is the projection operator that projects H on H_m, then $\varphi_m(g)$ is the trace of $E(m)T_g$.[4]

General integral formulas for D functions (and for spherical functions) are easily obtainable, along with differential equations and functional equations.[5–8] Unfortunately, the explicit evaluation of the integrals is quite difficult and has been accomplished in special

* On leave of absence from University of California, Los Angeles.
† On leave of absence from Atomic Energy Centre, Dacca, East Pakistan.

[1] C. Fronsdal, *Proceedings of the Seminar on Elementary Particles and High-Energy Physics, Trieste* 1965 (IAEA, Vienna, 1965).
[2] G. Cocho, C. Fronsdal, Harun Ar-Rashid, and R. White, ICTP, Trieste, preprint IC/66/27 (to be published in Phys. Rev. Letters).
[3] H. Leutwyler, Phys. Rev. Letters **17**, 156 (1966).
[4] R. Godement, Am. Math. Soc. Transl. **73**, 496 (1952).
[5] I. M. Gel'fand, Dokl. Akad. Nauk SSSR **70**, 5 (1950) Sov. Phys.—Dokl.
[6] I. M. Gel'fand and M. A. Naimark, *Unitäre Darstellungen der Klassischen Gruppen* (Akademie-Verlag, Berlin, 1957).
[7] Harish-Chandra, Am. Math. Soc. Transl. **76**, 234 (1954).
[8] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962).

cases only. Gel'fand and Naimark[5,6] gave the complete result in the case when H_m is one-dimensional; in this case the D function $D^{(D)}_{h,h}(g)$, $h \in H_m$, coincides with the spherical function $\varphi^{(D)}_m(g)$. This is reviewed in Sec. 2. Harish-Chandra[7] obtained a generalization of the Gel'fand–Naimark formula to any complex semisimple Lie group. Our purpose is to show how other D functions may be related to the Gel'fand functions in a way that is suitable for explicit evaluation (Sec. 4). As an example we complete the calculation of the "pion form factor" (Sec. 5).

2. THE GEL'FAND FUNCTIONS

The group $SL(n, C)$ is (isomorphic to) the group of unimodular $n \times n$ matrices. It is, therefore, convenient to think of an element $g \in G$ as such a matrix, so that the group product $g \cdot g'$ is just the product of the corresponding matrices. In particular, g belongs to the maximal compact subgroup K if it is unitary. An element of K will usually be denoted by the letter u. When we next describe the representations of $SL(n, C)$ we deliberately leave out all the details, since it will be seen that all our results follow directly from the general definition (1, 3) of the D functions.

At first we shall limit ourselves to the nondegenerate principal series of unitary, irreducible representations. This entails no essential loss of generality, as will be seen. According to Gel'fand and Naimark,[6] every unitary irreducible representation of the nondegenerate principal series may be characterized as follows[9]:

(i) The Hilbert space H consists of all complex-valued functions $h(u)$, over the maximal compact subgroup K, that satisfies the conditions

$$h(\gamma u) = \alpha(\gamma)h(u), \tag{2.1}$$

where γ is any diagonal unitary matrix, the function $\alpha(\gamma)$ is discussed below, and

$$\|h\|^2 = (h, h) < \infty, \tag{2.2}$$

where the inner product in H is defined by

$$(h_1, h_2) = \int h_1^*(u)h_2(u)\, du. \tag{2.3}$$

The measure du is independent of the representation.

(ii) The operators T_g of the representation act on $h(u)$ in the following way:

$$\begin{aligned} T_g h(u) &= \alpha(g, u)h(u_g), \\ \alpha(g, u) &= \alpha(ug)/\alpha(u), \end{aligned} \tag{2.4}$$

where the unitary matrix u_g is determined by u and g up to left multiplication by a diagonal unitary matrix [compare (2.1)]. In particular, $u_g = ug$ if g is unitary ($g \in K$).

(iii) The multiplier function $\alpha(g)$ depends on the representation. The unitary irreducible representations of the nondegenerate principal series are identified by the values of $n - 1$ integers $m^0 = m_1^0, \cdots, m_{n-1}^0$ and $n - 1$ real numbers $\rho = \rho_2, \cdots, \rho_n$. The integers have a very simple meaning: of all the irreducible representations $D(m)$ of K that occur in the reduction (1.1) of D_K, the one with the least highest weight is $D(m^0)$. The multiplicities γ_m in (1.1) are completely determined by m^0, and $\gamma_{m^0} = 1$. The real parameters, along with m^0, determine the function $\alpha(g)$. The latter should, therefore, be denoted more completely as $\alpha(g, m^0, \rho)$.

Let $h(u)$ be the function that corresponds to the canonical basis vector h; then the definition (1.3) for the general D function becomes the formula

$$D^{(m^0,\rho)}_{1,2}(g) = \int h_1^*(u) \frac{\alpha(ug, m^0, \rho)}{\alpha(u, m^0, \rho)} h_2(ug)\, du. \tag{2.5}$$

Although everything on the right-hand side of this formula is known in principle, the actual integration is usually of forbidding complexity. However, Gel'fand[5] succeeded in carrying out the integrations in a special case.

An irreducible representation D of $SL(n, C)$ is said to be of *Class I* if the one-dimensional representation $D(0)$ of K occurs in the reduction (1.5) of D_K. [This definition applies to all types of irreducible representations, unitary or not, provided only that the reduction (1.1) exists; that is, it applies to all irreducible representations in a Hilbert space.] In other words, D is of Class I if and only if H contains a vector h_0 that is invariant with respect to the transformations T_g, $g \in K$. It is easy to prove that any two such vectors are linearly dependent. A unitary representation of the principal nondegenerate series is of Class I if and only if $m^0 = 0$. Now let h_0 be a canonical basis vector in H_0; that is, $h_0 \in H_0$ and $\|h_0\| = 1$. The corresponding function $h_0(u)$ is invariant under

$$h_0(u) \to T_{u'}h(u) = \frac{\alpha(uu')}{\alpha(u)} h_0(uu'), \tag{2.6}$$

from which it follows that $\alpha(u)h_0(u)$ is independent of u. By a suitable renormalization of the basis (which preserves $\|h_0\| = 1$) one may arrange that $\alpha(u) = h_0(u) = 1$, so that (2.5) reduces to

$$\begin{aligned} \varphi_0^{(0,\rho)}(g) &= D^{(0,\rho)}_{0,0}(g) = (h_0, T_g h_0) \\ &= \int \alpha(ug, 0, \rho)\, du. \end{aligned} \tag{2.7}$$

[9] Ref. 6, Sec. 8.

Because H_0 is one-dimensional, this particular D function is also a spherical function.

Gel'fand's result is[5,10]

$$\varphi_0^{(0,\rho)}(g) = \left(\frac{2}{i}\right)^{n(n-1)/2} 1!\,2!\cdots(n-1)! \times \left[\prod_{1\le p\le q\le n}(\rho_q-\rho_p)(\lambda_q^2-\lambda_p^2)\right]^{-1} \times \begin{vmatrix} 1 & 1 & \cdots & \cdots & 1 \\ \lambda_1^{i\rho_2} & \lambda_2^{i\rho_2} & \cdots & \cdots & \lambda_n^{i\rho_2} \\ \lambda_1^{i\rho_3} & \lambda_2^{i\rho_3} & \cdots & \cdots & \lambda_n^{i\rho_3} \\ & \cdots & \cdots & \cdots & \\ \lambda_1^{i\rho_n} & \lambda_2^{i\rho_n} & \cdots & \cdots & \lambda_n^{i\rho_n} \end{vmatrix}, \quad (2.8)$$

where $\lambda_1, \cdots, \lambda_n$ are the moduli of the eigenvalues of the matrix g, and $\lambda_1 \cdot \lambda_2 \cdots \lambda_n = 1$. The next step is to extend this result to arbitrary Class I representations. This is, in fact, trivial, for every irreducible Class I representation of $SL(n, C)$ may be described as an analytic continuation of those of the principal nondegenerate series, to complex values of the parameters $\rho_2, \cdots, \rho_n$. The precise nature of this process of analytic continuation of representations does not concern us here.[11]

To summarize: every irreducible Class I representation of $SL(n, C)$ may be labelled by a set of $n-1$ complex parameters $\rho_2, \cdots, \rho_n$. The spherical function $(h_0, T_g h_0)$ for any one of these representations is given by the expression (2.8), where $\lambda_1, \cdots, \lambda_n$ are the moduli of the eigenvalues of the matrix g.[12]

Two problems remain: (1) to use Gel'fand's result for $\varphi_0^{(0,\rho)}(g)$ to calculate other D-functions for Class I representations, and (2) to obtain a suitable generalization of Gel'fand's formula to representations with $m^0 \neq 0$. Only the first problem is tackled in this paper.[13]

3. DESCRIPTION OF THE ALGEBRA AND THE CANONICAL BASIS

From the representation D of the group, by operators T_g in a Hilbert space H, one may pass to a representation $\bar{D}$ of the corresponding Lie algebra, by operators L in a vector space $\mathbf{H}$ that is dense in H. We define $\mathbf{H}$ as the set of all vectors in H that have a finite number of nonzero components with respect to a canonical basis. From general results of Harish-Chandra[14] we know that the operators of the algebra are defined on $\mathbf{H}$, and that difficult concepts such as topological equivalence and topological reducibility of D are equivalent to purely algebraic properties of $\bar{D}$.

The algebra of $SL(n, C)$ is (isomorphic to) the Lie algebra of traceless, n-dimensional square matrices. We denote such matrices by the letter ϵ. If the unimodular matrix g is near 1, then

$$g = 1 + it\epsilon + O(t^2) \quad (3.1)$$

where t is a real parameter that tends to zero when g tends to 1. If ϵ_i^j denotes a matrix element of the matrix ϵ, then matrices λ_i^j and $\lambda_i'^j$ are defined by

$$\epsilon = (\epsilon_i^j + \epsilon_j^{i*})\lambda_j^i + i(\epsilon_i^j - \epsilon_j^{i*})\lambda_j'^i, \quad (3.2)$$

$$\sum_i \lambda_i^i = \sum \lambda_i'^i = 0, \quad (3.3)$$

and satisfy the commutation relations

$$[\lambda_i^j, \lambda_k^l] = -[\lambda_i'^j, \lambda_k'^l] = \delta_i^l\lambda_k^j - \delta_k^j\lambda_i^l, \qquad [\lambda_i^j, \lambda_k'^l] = 0. \quad (3.4)$$

These relations define an abstract Lie algebra. A set of linear operators in $\mathbf{H}$ that satisfies the commutation relations of the λ_i^j, $\lambda_i'^j$ will be denoted Λ_i^j and $\Lambda_i'^j$. The matrix g is unitary if the matrix ϵ is hermitian; hence Λ_i^j generate the transformations of the compact subgroup K. We refer to Λ_i^j as compact generators and to $\Lambda_i'^j$ as noncompact generators.

Even though each subspace H_m is finite dimensional, it is difficult to fix a canonical basis in H_m in a general and simple way. The reason is that, whenever $\gamma_m > 1$, there are sets of vectors in H_m that cannot be distinguished from each other by simultaneous eigenvalues of a complete set of operators in the enveloping algebra. There exists no solution of this "labelling problem," of sufficient simplicity and generality that one can hope to give general closed expressions for all D functions. There is, however, a very simple procedure that allows the labelling of the vectors of any finite number of H_m's, and this system of labelling is particularly well suited to the calculation of D functions.

We begin with H_0. Let h_0 be any nonzero vector in H_0. Since H_0 is a closed subspace of H, and H is a Hilbert space, the norm of h_0 is defined, and we choose h_0 so that $\|h_0\| = 1$. This gives a complete orthonormal basis in H_0. Next, consider the vectors

$$h_i^j \equiv \Lambda_i'^j h_0. \quad (3.5)$$

[10] The calculation is given in detail in Ref. 6, Sec. 9. See also Ref. 7, p. 253.

[11] See our review "Representations of $SL(n, C)$," ICTP, Trieste, preprint IC/66/51.

[12] I. M. Gel'fand and M. A. Naimark pointed out that this result holds for unitary representations. See Ref. 6, Appendix.

[13] For representations of the most degenerate series that are labelled by one real parameter and one integer k (with $k = 0$ for representations of Class I), we have noticed some simple relations between spherical functions for different values of k. See Ref. 2.

[14] See Ref. 11 and original papers quoted there for details.

These vectors span an $n^2 - 1$-dimensional closed subspace of H that is invariant and irreducible with respect to K. The vectors

$$h^{jl}_{ik} \equiv \Lambda'^{l}_{k}\Lambda'^{j}_{i}h_0 \tag{3.6}$$

span another closed subspace of H. It is invariant but not irreducible under K; it may easily be reduced by ordinary tensor reduction methods. Continuing this process one constructs an infinite number of closed subspaces of H, each invariant and irreducible under K. Clearly, each one of these subspaces is a subspace of some H_m. The final step, which we need not describe in detail, is to set up an orthonormal basis in each of the invariant, irreducible subspaces.

Let h_1 and h_2 be any two canonical basis vectors, constructed by the method just outlined. Then there exist operators Q_1 and Q_2 that are homogeneous polynomials in the Λ'^{j}_{i}, such that

$$\begin{aligned} h_1 &= Q_1 h_0, \\ h_2 &= Q_2 h_0. \end{aligned} \tag{3.7}$$

These operators are not unique; in practice they must be chosen so as to have the lowest possible degree of homogeneity in the Λ'^{j}_{i}.

4. D FUNCTIONS FOR CLASS I REPRESENTATIONS

From the definition (1.3) of the general D function and the choice (3.7) of the canonical basis, we find

$$D^{(0,\rho)}_{1,2}(g) = (h_0, Q_1^{\dagger}T_g Q_2 h_0). \tag{4.1}$$

The right-hand side can be expressed in terms of Gel'fand's function and its derivatives. The correctness of this statement, as well as the prescription for carrying out the calculation, will become quite clear after we have studied the cases when the degrees of homogeneity of Q_1 and Q_2 are zero or 1.

Let

$$g' = (1 + it\epsilon)g(1 + it'\eta), \tag{4.2}$$

where t and t' are arbitrarily small real parameters and ϵ and η are traceless antihermitian matrices. Then

$$\begin{aligned} T_{g'} &= (1 + iL)T_g(1 + iL'), \\ L &= 2it\epsilon^{j}_{i}\Lambda'^{i}_{j}, \quad L' = 2it'\eta^{l}_{k}\Lambda'^{k}_{l}. \end{aligned} \tag{4.3}$$

Since $\varphi_0(g)$ depends on the moduli λ_k of the eigenvalues of g' only, we have

$$\varphi_0(g') = \varphi_0(g) + (\Delta\lambda_k + \Delta'\lambda_k)\frac{\partial}{\partial\lambda_k}\varphi_0(g) + \left(\Delta''\lambda_k\frac{\partial}{\partial\lambda_k} + \tfrac{1}{2}\Delta\lambda_k\Delta'\lambda_l\frac{\partial^2}{\partial\lambda_k\partial\lambda_l}\right)\varphi_0(g) + \cdots, \tag{4.4}$$

where

$$\lambda'_k = \lambda_k + \Delta\lambda_k + \Delta'\lambda_k + \Delta''\lambda_k + \cdots, \tag{4.5}$$

and $\Delta\lambda_k$ is a linear function of ϵ, $\Delta'\lambda_k$ is a linear function of η, and $\Delta''\lambda_k$ is linear in both. Introducing (4.3) and (4.4) into the formula

$$\varphi_0(g') = (h_0, T_{g'}h_0)$$

we get the following:

$$\begin{aligned} (h_0, iLT_g h_0) &= \Delta\lambda_k\frac{\partial}{\partial\lambda_k}\varphi_0(g), \\ (h_0, T_g iL'h_0) &= \Delta'\lambda_k\frac{\partial}{\partial\lambda_k}\varphi_0(g), \\ (h_0, iLT_g iL'h_0) &= \left(\Delta''\lambda_k\frac{\partial}{\partial\lambda_k} + \tfrac{1}{2}\Delta\lambda_k\Delta'\lambda_l\frac{\partial^2}{\partial\lambda_k\partial\lambda_l}\right)\varphi_0(g). \end{aligned} \tag{4.6}$$

This procedure can obviously be repeated indefinitely, to calculate every D function represented by the formula (4.1).

The variations $\Delta\lambda_k$, $\Delta'\lambda_k$ and $\Delta''\lambda_k$ are calculated by straightforward second order perturbation theory, with the result

$$\begin{aligned} \Delta\lambda_k &= \mathrm{Re}\,(it\epsilon g)^k_k, \quad \Delta'\lambda_k = \mathrm{Re}\,(it'g\eta)^k_k, \\ \Delta''\lambda_k &= \mathrm{Re}\,(it\epsilon git'\eta)^k_k + \sum_{m\neq k}\frac{1}{\lambda_k - \lambda_m} \\ &\quad \times \mathrm{Re}\,[(it\epsilon g)^m_k(it'g\eta)^k_m + (igt'\eta)^m_k(it\epsilon g)^k_m]. \end{aligned} \tag{4.7}$$

If we insert this into (4.6) and take g to be diagonal with positive eigenvalues $\lambda_1, \cdots, \lambda_n$, we then find

$$\begin{aligned} (h_0, T_g\Lambda'^{k}_{j}h_0) = (h_0, \Lambda'^{k}_{j}T_g h_0) &= \frac{1}{2i}\delta^{k}_{j}\lambda_k\frac{\partial}{\partial\lambda_k}\varphi_0(\lambda), \\ 4(h_0, \Lambda'^{j}_{i}T_g\Lambda'^{l}_{k}h_0) &= -\tfrac{1}{2}\delta^{l}_{i}\delta^{j}_{k}\left(\lambda_i\frac{\partial}{\partial\lambda_k} + \lambda_k\frac{\partial}{\partial\lambda_i}\right)\varphi_0(\lambda) \\ &\quad - \tfrac{1}{2}\delta^{l}_{i}\delta^{j}_{k}\frac{1-\delta_{ik}}{\lambda_k - \lambda_i}\left(\lambda_i^2\frac{\partial}{\partial\lambda_i} - \lambda_k^2\frac{\partial}{\partial\lambda_k}\right) \\ &\quad \times \varphi_0(\lambda) - \delta^{j}_{i}\delta^{l}_{k}\lambda_i\lambda_k\frac{\partial^2}{\partial\lambda_i\partial\lambda_k}\varphi_0(\lambda). \end{aligned} \tag{4.8}$$

A very slight modification of the procedure gives

$$\begin{aligned} 4(h_0, \Lambda'^{j}_{i}\Lambda'^{l}_{k}T_g h_0) &= -\tfrac{1}{2}\delta^{l}_{i}\delta^{j}_{k}\left(\lambda_i\frac{\partial}{\partial\lambda_k} + \lambda_k\frac{\partial}{\partial\lambda_i}\right)\varphi_0(\lambda) \\ &\quad - \tfrac{1}{2}\delta^{l}_{i}\delta^{j}_{k}\frac{1-\delta_{ik}}{\lambda_k - \lambda_i}\lambda_i\lambda_k\left(\frac{\partial}{\partial\lambda_i} - \frac{\partial}{\partial\lambda_k}\right) \\ &\quad \times \varphi_0(\lambda) - \delta^{j}_{i}\delta^{l}_{k}\lambda_i\lambda_k\frac{\partial^2}{\partial\lambda_i\partial\lambda_k}\varphi_0(\lambda). \end{aligned} \tag{4.9}$$

5. THE PION FORM FACTOR

We shall complete the calculation of the D function in a special case. The specializations are the following: (i) Take $n = 6$ in $SL(n, C)$; (ii) take g to be diagonal, with

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4^{-1} = \lambda_5^{-1} = \lambda_6^{-1} = \lambda; \quad (5.1)$$

(iii) take the values of the parameters ρ_k to be those of one of the most degenerate representations, namely[15]

$$\begin{aligned} i\rho_2 &= -2(M+1), \quad i\rho_3 = -2(M+2), \\ i\rho_4 &= -2(M+3), \quad i\rho_5 = -2(M+4), \quad (5.2) \\ i\rho_6 &= -2(2M+5). \end{aligned}$$

These values define a unitary representation of the principal series if

$$M = -\tfrac{5}{2} + i\rho, \quad \rho \text{ real}, \quad (5.3)$$

and a unitary representation of the supplementary series if M is real between -3 and -2; (iv) take

$$\epsilon^+ = \eta = -\begin{pmatrix} 0 & i & 0 & & & \\ i & 0 & 0 & & \mathbf{0} & \\ 0 & 0 & 0 & & & \\ & & & 0 & i & 0 \\ & \mathbf{0} & & i & 0 & 0 \\ & & & 0 & 0 & 0 \end{pmatrix}.$$

This means that the left side of (4.4) becomes

$$(h, T_g h), \quad (5.4)$$

$$h \sim (\lambda_1'^2 + \lambda_4'^5)h_0. \quad (5.5)$$

In a model that applies unitary representations of $SL(6, C)$ to elementary particle physics, this function may be interpreted as the electric charge form factor of the pion.[16,17]

Approaching the limit (5.1) we obtain the following expression for the Gel'fand function (2.8):

$$\varphi_0(\lambda) \sim (\lambda^2 - \lambda^{-2})^{-9} D_{1,2},$$

where

$$D_{k,l} = \begin{vmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ \lambda^{i\rho_2} & \binom{i\rho_2}{k}\lambda^{i\rho_2-1} & \binom{i\rho_2}{l}\lambda^{i\rho_2-2} & \lambda^{-i\rho_2} & i\rho_2\lambda^{-i\rho_2+1} & \binom{i\rho_2}{2}\lambda^{-i\rho_2+2} \\ \lambda^{i\rho_3} & \binom{i\rho_3}{k}\lambda^{i\rho_3-1} & \binom{i\rho_3}{l}\lambda^{i\rho_3-2} & \lambda^{-i\rho_3} & i\rho_3\lambda^{-i\rho_3+1} & \binom{i\rho_3}{2}\lambda^{-i\rho_3+3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{vmatrix}. \quad (5.6)$$

Evaluating the determinant we get, setting $y = \lambda^2$,

$$\begin{aligned} \varphi_0(y) = \frac{(h_0, T_g h_0)}{(h_0, h_0)} = \frac{-\frac{1}{4}}{(y - y^{-1})^9} \\ \times [4(M+1)(M+4)(2M+5)(y^3 - y^{-3}) \\ - 12(M+2)(M+3)(2M+5)(y - y^{-1}) \\ - (M+3)^2(M+4)^2(y^{2M+1} - y^{-(2M+1)}) \\ + 4(M+2)^2(M+4)^2(y^{2M+3} - y^{-(2M+3)}) \\ - 6(M+1)(M+2)(M+3)(M+4) \\ \times (y^{2M+5} - y^{-(2M+5)}) + 4(M+1)^2 \\ \times (M+3)^2(y^{2M+7} - y^{-(2M+7)}) \\ - (M+1)^2(M+2)^2(y^{2M+9} - y^{-(2M+9)})]. \end{aligned} \quad (5.7)$$

Because of the special relationships (5.1) it is necessary to use the formula (4.8) in the sense of a limit.[18] We find

$$(h, T_g h) \sim \lim_{\lambda_2 \to \lambda_1} \frac{\lambda_1}{\lambda_2 - \lambda_1}\left(\lambda_2 \frac{\partial}{\partial \lambda_2} - \lambda_1 \frac{\partial}{\partial \lambda_1}\right) \times \varphi_0(\lambda) + (\lambda \to \lambda^{-1}). \quad (5.8)$$

(Using degenerate perturbation theory to determine $\Delta'' \lambda_k$ we obtain the same answer.) From (2.8) and (5.8) we find

$$(h, T_g h) \sim (\lambda^2 - \lambda^{-2})^{-9} \times \left[\left(\frac{12}{(\lambda^2 - \lambda^{-2})^2} - \frac{3}{4}\right) D_{1,2} + D_{1,3} - D_{1,4} - D_{2,3}\right],$$

where D_{kl} is the determinant (5.6). After evaluating

[15] See Ref. 11, Sec. VI, 4.

[16] G. Cocho, C. Fronsdal, Harun Ar-Rashid, and R. White, ICTP, Trieste, preprint IC/66/84.

[17] The vector h represents a pion at rest and T_g is a Lorentz transformation. See also Refs. 11 and 13.

[18] Actually the function (5.4) is given by (4.7), rather than by (4.8). In the special case studied here T_g commutes with the generator in (5.5), and (4.7) agrees with (4.8).

the determinants we get

$$
\begin{aligned}
F(y) = c[&M(M+1)^2(M+2)^2(M+3)^2 \\
&\times (M+4)^2(M+5)(2M+5)]^{-1} \\
&\times \left\{ -\frac{48}{(y-y^{-1})^2}\varphi_0(y) + \frac{1}{(y-y^{-1})^2} \right. \\
&\times [2(M+1)(M+2)(M+3)(M+4) \\
&\times (2M+5)(y^3-y^{-3}) - 2(M+2) \\
&\times (M+3)(2M+5)(3M^2+15M+20) \\
&\times (y-y^{-1}) - \tfrac{1}{3}(M+2)(M+3)^2(M+4)^2 \\
&\times (M+5)(y^{2M+1}-y^{-(2M+1)}) + \tfrac{4}{3}(M+2)^2 \\
&\times (M+3)^2(M+4)^2(y^{2M+3}-y^{-(2M+3)}) \\
&- 2(M+1)(M+2)(M+3)(M+4) \\
&\times (M^2+5M+7)(y^{2M+5}-y^{-(2M+5)}) \\
&+ \tfrac{4}{3}(M+1)^2(M+2)^2(M+3)^2 \\
&\times (y^{2M+7}-y^{-(2M+7)}) - \tfrac{1}{3}M(M+1)^2 \\
&\left. \times (M+2)^2(M+3)(y^{2M+9}-y^{-(2M+9)})] \right\},
\end{aligned}
\tag{5.9}
$$

where $c = -3150$.

For practical use, the above form of the result is awkward, because it is difficult to calculate numerical values in the neighborhood of $g = 1$ (i.e., $y = 1$). For special values of M, however, one may verify that the function is normalized to 1 at $y = 1$. Thus, for $M = 1$, (5.9) reduces to

$$\tfrac{1}{4}(y + y^{-1})^2,$$

and for $M = -\frac{5}{2}$ we find

$$
\begin{aligned}
F(y) = \frac{112}{9}\frac{y^5}{(y+1)^{10}}&[-10y^6 - 28y^5 + 235y^4 \\
&+ 722y^3 + 640y^2 + 134y + 35] \\
&+ \frac{2240}{3}\frac{y^5}{(y+1)^7}[y^4 - 26y^2 + 1](y-1)^{-7} \\
&\times [\ln y - \tfrac{1}{4}(y-1) + \cdots + \tfrac{1}{6}(y-1)^6].
\end{aligned}
\tag{5.10}
$$

The physical implications will be discussed elsewhere.[16]

ACKNOWLEDGMENTS

The authors are grateful to Professor Abdus Salam and the IAEA for hospitality at the International Center for Theoretical Physics, Trieste.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Path-Integral Calculation of the Quantum-Statistical Density Matrix for Attractive Coulomb Forces

R. G. STORER
School of Physical Sciences, The Flinders University of South Australia, South Australia

(Received 8 October 1967)

A method for the calculation of the two-particle statistical density matrix for attractive Coulomb forces is described. The path-integral expression for the density matrix is reduced to a modified path integral which involves summation over only one-dimensional paths. This expression is then approximated by an iteration procedure using direct numerical quadratures. The results obtained are related directly to the quantum-mechanical radial-distribution function for a plasma at small ion–electron separations.

1. INTRODUCTION

A classical calculation of the radial-distribution function for ions and electrons in a plasma leads to a divergence of the form $\exp(\beta e^2/r)$ at small separations ($\beta = 1/kT$). It has long been recognized that this divergence occurs because of the use of classical mechanics and that a full quantum-mechanical calculation would eliminate the divergence by providing a lower limit to the energy between two oppositely charged particles. Because one intuitively feels that the use of quantum mechanics should have little effect on high-temperature plasma calculations, the usual procedure to overcome this difficulty is to use some form of modified potential, usually one which is cutoff in some way at a distance corresponding to the Bohr radius.[1] This is not really satisfactory, particularly at temperatures of the order of 10^4 °K, so an exact calculation of the quantum-mechanical radial-distribution function at small separations would be most useful. The two-particle contribution to the radial-distribution function for a hydrogen plasma is the dominant contribution for small separations, low densities, and temperatures which are high enough to dissociate the hydrogen molecule. For large separations, many-body effects give rise to Debye screening. At these temperatures, this is an effect that can be described adequately using classical mechanics; so

[1] H. S. Green, Nucl. Fusion 1, 69 (1961).

the determinants we get

$$
\begin{aligned}
F(y) = c[&M(M+1)^2(M+2)^2(M+3)^2 \\
&\times (M+4)^2(M+5)(2M+5)]^{-1} \\
&\times \left\{ -\frac{48}{(y-y^{-1})^2}\varphi_0(y) + \frac{1}{(y-y^{-1})^2} \right. \\
&\times [2(M+1)(M+2)(M+3)(M+4) \\
&\times (2M+5)(y^3-y^{-3}) - 2(M+2) \\
&\times (M+3)(2M+5)(3M^2+15M+20) \\
&\times (y-y^{-1}) - \tfrac{1}{3}(M+2)(M+3)^2(M+4)^2 \\
&\times (M+5)(y^{2M+1}-y^{-(2M+1)}) + \tfrac{4}{3}(M+2)^2 \\
&\times (M+3)^2(M+4)^2(y^{2M+3}-y^{-(2M+3)}) \\
&- 2(M+1)(M+2)(M+3)(M+4) \\
&\times (M^2+5M+7)(y^{2M+5}-y^{-(2M+5)}) \\
&+ \tfrac{4}{3}(M+1)^2(M+2)^2(M+3)^2 \\
&\times (y^{2M+7}-y^{-(2M+7)}) - \tfrac{1}{3}M(M+1)^2 \\
&\left. \times (M+2)^2(M+3)(y^{2M+9}-y^{-(2M+9)})] \right\},
\end{aligned}
\tag{5.9}
$$

where $c = -3150$.

For practical use, the above form of the result is awkward, because it is difficult to calculate numerical values in the neighborhood of $g = 1$ (i.e., $y = 1$). For special values of M, however, one may verify that the function is normalized to 1 at $y = 1$. Thus, for $M = 1$, (5.9) reduces to

$$\tfrac{1}{4}(y + y^{-1})^2,$$

and for $M = -\frac{5}{2}$ we find

$$
\begin{aligned}
F(y) = \frac{112}{9}\frac{y^5}{(y+1)^{10}}&[-10y^6 - 28y^5 + 235y^4 \\
&+ 722y^3 + 640y^2 + 134y + 35] \\
&+ \frac{2240}{3}\frac{y^5}{(y+1)^7}[y^4 - 26y^2 + 1](y-1)^{-7} \\
&\times [\ln y - \tfrac{1}{4}(y-1) + \cdots + \tfrac{1}{6}(y-1)^6].
\end{aligned}
\tag{5.10}
$$

The physical implications will be discussed elsewhere.[16]

ACKNOWLEDGMENTS

The authors are grateful to Professor Abdus Salam and the IAEA for hospitality at the International Center for Theoretical Physics, Trieste.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 6 JUNE 1968

Path-Integral Calculation of the Quantum-Statistical Density Matrix for Attractive Coulomb Forces

R. G. STORER
School of Physical Sciences, The Flinders University of South Australia, South Australia

(Received 8 October 1967)

A method for the calculation of the two-particle statistical density matrix for attractive Coulomb forces is described. The path-integral expression for the density matrix is reduced to a modified path integral which involves summation over only one-dimensional paths. This expression is then approximated by an iteration procedure using direct numerical quadratures. The results obtained are related directly to the quantum-mechanical radial-distribution function for a plasma at small ion–electron separations.

1. INTRODUCTION

A classical calculation of the radial-distribution function for ions and electrons in a plasma leads to a divergence of the form $\exp(\beta e^2/r)$ at small separations ($\beta = 1/kT$). It has long been recognized that this divergence occurs because of the use of classical mechanics and that a full quantum-mechanical calculation would eliminate the divergence by providing a lower limit to the energy between two oppositely charged particles. Because one intuitively feels that the use of quantum mechanics should have little effect on high-temperature plasma calculations, the usual procedure to overcome this difficulty is to use some form of modified potential, usually one which is cutoff in some way at a distance corresponding to the Bohr radius.[1] This is not really satisfactory, particularly at temperatures of the order of 10^4 °K, so an exact calculation of the quantum-mechanical radial-distribution function at small separations would be most useful. The two-particle contribution to the radial-distribution function for a hydrogen plasma is the dominant contribution for small separations, low densities, and temperatures which are high enough to dissociate the hydrogen molecule. For large separations, many-body effects give rise to Debye screening. At these temperatures, this is an effect that can be described adequately using classical mechanics; so

[1] H. S. Green, Nucl. Fusion 1, 69 (1961).

since we are interested here primarily in quantal effects, we restrict our discussion to a pure Coulomb potential.

In this two-particle approximation, the radial-distribution function is proportional to the diagonal element of the coordinate representation of the density matrix

$$\langle \mathbf{x}_1, \mathbf{y}_1 | \exp(-\beta H) | \mathbf{x}_2, \mathbf{y}_2 \rangle,$$

where H is the two-particle Hamiltonian and $\mathbf{x}$ and $\mathbf{y}$ are the coordinates of the electron and ion, respectively. The center-of-mass component of this density matrix can be written down immediately and we can define the density matrix $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$ for relative motion by the equation

$$\langle \mathbf{x}_1, \mathbf{y}_1 | \exp(-\beta H) | \mathbf{x}_2, \mathbf{y}_2 \rangle = \left(\frac{2\pi\hbar^2\beta}{M}\right)^{-\frac{3}{2}} \exp\left(-\frac{(\mathbf{X}_1 - \mathbf{X}_2)^2}{2\pi\hbar^2\beta/M}\right) \rho(\mathbf{r}_1, \mathbf{r}_2, \beta), \quad (1.1)$$

where M is the combined mass, $\mathbf{r}_1 = \mathbf{x}_1 - \mathbf{y}_1$ and $\mathbf{r}_2 = \mathbf{x}_2 - \mathbf{y}_2$ are the relative coordinates, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are the center-of-mass coordinates. If we define $g_2(r)$ as the two-particle approximation to the radial-distribution function, then

$$g_2(r) = \left(\frac{2\pi\hbar^2\beta}{\mu}\right)^{\frac{3}{2}} \rho(\mathbf{r}, \mathbf{r}, \beta), \quad (1.2)$$

where $r = |\mathbf{r}|$ and μ is the reduced mass of the electron.

There are several ways by which one could attempt to calculate $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$. This function satisfies the Bloch equation

$$\frac{\partial \rho}{\partial \beta} = -H\rho \quad (1.3)$$

with $\rho = \delta(\mathbf{r}_1 - \mathbf{r}_2)$ at $\beta = 0$, where H is the Hamiltonian corresponding to the relative motion. One can obtain a formal solution to this equation in terms of a summation over the bound-state wavefunctions $\psi_s(\mathbf{r})$ and an integration over all the wavefunctions corresponding to the continuous spectrum $\psi(\mathbf{k}, \mathbf{r})$,

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = \sum_s \exp(-\beta E_s)\psi_s^*(\mathbf{r}_1)\psi_s(\mathbf{r}_2) + \int d\mathbf{k} \exp(-\beta E_{\mathbf{k}})\psi^*(\mathbf{k}, \mathbf{r})\psi(\mathbf{k}, \mathbf{r}). \quad (1.4)$$

In theory, the use of this equation is a straightforward procedure. However, in practice, the calculation of the wavefunctions, particularly for the scattering states for all values of $\mathbf{k}$, can be a time consuming operation.[2]

In this paper we use the expression for the density matrix in terms of sums over paths. This technique was first used in physics by Feynman and has been the subject of several reviews.[3–5] Some numerical calculations have been made of path integrals[6,7] using Monte Carlo methods; in contrast to these, an iteration procedure and numerical quadratures are used here. In the notation of Brush,[4] for a potential $V(\mathbf{r})$,

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = \int \exp\left\{-\int_0^\beta V(\mathbf{r}(t))\, dt\right\} d_{\omega(0,\mathbf{r}_1;\beta,\mathbf{r}_2)}\mathbf{r}(t). \quad (1.5)$$

This integral denotes a summation taken over all possible paths that a particle can take when it is moving by means of Brownian motion, starting at $\mathbf{r}_1$ and ending at $\mathbf{r}_2$ at "time" β. One can write this expression in terms of a limit which is equivalent to breaking the paths into N piecewise-straight pieces and letting N become infinite. Then,

$$\begin{aligned}\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) &= \lim_{N\to\infty} \int\cdots\int d\mathbf{s}_2 \cdots d\mathbf{s}_N \prod_{i=1}^{N} \rho^{(0)}(\mathbf{s}_i, \mathbf{s}_{i+1}, \beta) \\ &\quad \times \exp\left\{-\frac{\beta}{N}\left[\tfrac{1}{2}V(s_1) + \sum_{i=2}^{N} V(s_i) + \tfrac{1}{2}V(s_{N+1})\right]\right\} \\ &= \lim_{N\to\infty} \int\cdots\int d\mathbf{s}_2 \cdots d\mathbf{s}_N \prod_{i=1}^{N} \exp\left\{-\frac{\beta}{2N} V(s_i)\right\} \\ &\quad \times \rho^{(0)}\left(\mathbf{s}_i, \mathbf{s}_{i+1}, \frac{\beta}{N}\right) \exp\left\{-\frac{\beta}{2N} V(s_{i+1})\right\}, \quad (1.6)\end{aligned}$$

where

$$\mathbf{s}_1 = \mathbf{r}_1, \quad \mathbf{s}_{N+1} = \mathbf{r}_2,$$

and

$$\rho^{(0)}(\mathbf{r}_1, \mathbf{r}_2, \beta) = \left(\frac{2\pi\hbar^2\beta}{\mu}\right)^{-\frac{3}{2}} \exp\left(-\frac{(\mathbf{r}_1 - \mathbf{r}_2)^2}{2\hbar^2\beta/\mu}\right) \quad (1.7)$$

is the free-particle density matrix. When the path integral is written in this way, one can see readily that it is equivalent to writing

$$\begin{aligned}\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) &= \langle \mathbf{r}_1 | \exp(-\beta H) | \mathbf{r}_2 \rangle \\ &= \lim_{N\to\infty} \int\cdots\int d\mathbf{s}_2 \cdots d\mathbf{s}_N \langle \mathbf{r}_1 | \exp(-\beta H/N) | \mathbf{s}_2 \rangle \\ &\quad \times \langle \mathbf{s}_2 | \exp(-\beta H/N) | \mathbf{s}_3 \rangle \cdots \langle \mathbf{s}_N | \exp(-\beta H/N) | \mathbf{r}_2 \rangle \quad (1.8)\end{aligned}$$

(which is an exact equation even for finite N) and then replacing each factor by its high-temperature approximation, i.e., putting

$$\begin{aligned}\langle \mathbf{r}_1 | \exp(-\beta H/N) | \mathbf{r}_2 \rangle &\approx \exp\left(-\frac{\beta}{2N} V(r_1)\right) \rho^{(0)}\left(\mathbf{r}_1, \mathbf{r}_2, \frac{\beta}{N}\right) \\ &\quad \times \exp\left(-\frac{\beta}{2N} V(r_2)\right), \quad (1.9)\end{aligned}$$

which is a good approximation when the argument

[2] A. A. Barker, Australian J. Phys. **21**, 121 (1968).
[3] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Publ. Co., Inc., New York, 1965).
[4] S. G. Brush, Rev. Mod. Phys. **33**, 79 (1961).
[5] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
[6] L. D. Fosdick and H. F. Jordan, Phys. Rev. **143**, 58 (1966).
[7] M. D. Donsker and M. Kac, J. Res. Natl. Bur. Std. **B44**, 551 (1950).

of the exponential $(\beta/2N)V(r)$ is in some sense small for all values of r or at least for all values of r except a set whose measure approaches zero when $N \to \infty$. For the purposes of numerical computation, a value of N is chosen large enough so that the final result is insensitive to an increase in N.

Previous calculations using the path-integral approach have relied on a Monte Carlo sampling procedure to evaluate the many-dimensional integrals involved. This has produced satisfactory results for high enough temperatures but suffers from the defect that when the variance becomes large the amount of computation needed to produce reasonable results is increased enormously. The method used here is to note that the path integral can be expressed in terms of an iterative procedure. We note that

$$\rho(\mathbf{r}_1, \mathbf{r}_2, 2\beta) = \int d\mathbf{r}\rho(\mathbf{r}_1, \mathbf{r}, \beta)\rho(\mathbf{r}, \mathbf{r}_2, \beta), \quad (1.10)$$

so that if $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$ is known or can be approximated, for example, by

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = \exp\{-\beta V(r_1)/2\}\rho^{(0)}(\mathbf{r}_1, \mathbf{r}_2, \beta) \times \exp\{-\beta V(r_2)/2\} \quad (1.11)$$

for small β, we have an equation which could be used to determine the density matrix at half the temperature. Hence one could start an iteration procedure by using equation (1.11) for small β and evaluating the integral in Eq. (1.10) by numerical quadratures. After n steps, one would have evaluated the equivalent of an integral of $(2^n - 1)$ dimensions and reached a temperature corresponding to $2^n\beta$.

The case where the potential is a Coulomb interaction is particularly suited to this procedure because the symmetries which exist for the Coulomb potential enable us to carry out all the numerical integrations in one dimension. Before proceeding with the details of the numerical procedure, we show how this simplification can be made.

2. REDUCTION OF THE COMPLETE COULOMB DENSITY MATRIX IN TERMS OF THE S-WAVE CONTRIBUTION

In this section it is shown that the complete three-dimensional density matrix for a pure Coulomb system can be expressed entirely in terms of the S-wave contribution. This follows closely the technique used by Hostler and Pratt[8] in the consideration of Green's function for the Coulomb system. The results of Hostler[9,10] could be used directly by noting that our density matrix is the inverse Laplace transform of the Green's function that he obtains.

[8] L. Hostler and R. H. Pratt, Phys. Rev. Letters **10**, 469 (1963).
[9] L. Hostler, J. Math. Phys. **5**, 591 (1964).
[10] L. Hostler, J. Math. Phys. **8**, 642 (1967).

The equation satisfied by the Coulomb density matrix is

$$\left(\nabla_2^2 + \frac{\alpha}{r_2}\right)\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = \frac{\partial\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)}{\partial\beta} \quad (2.1)$$

with the initial condition $\rho(\mathbf{r}_1, \mathbf{r}_2, 0) = \delta^{(3)}(\mathbf{r}_1 - \mathbf{r}_2)$.

The unit of length has been chosen here to be the Bohr radius $a_0 = \hbar^2/\mu e^2$ and the unit of energy to be the ground-state energy $E_0 = \mu e^4/2\hbar^2$. With these units $\alpha = 2$ and for a proton–electron system the temperature in degrees Kelvin is related to β via

$$T = 1.5778 \times 10^5/\beta\ ^\circ\mathrm{K}. \quad (2.2)$$

We shall use these units in the remainder of this paper.

The density matrix for any central potential can be expanded in terms of the contributions from all partial waves:

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = \sum_{l=0}^{\infty} \frac{(2l+1)}{4\pi r_1 r_2}\rho_l(r_1, r_2, \beta)P_l(\cos\theta), \quad (2.3)$$

so that $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$ depends only on r_1, r_2 and the angle θ between $\mathbf{r}_1$ and $\mathbf{r}_2$. In view of Hostler's result, we now make the following change of variables:

$$x = (r_1 + r_2 + |\mathbf{r}_2 - \mathbf{r}_1|)/2,$$
$$y = (r_1 + r_2 - |\mathbf{r}_2 - \mathbf{r}_1|)/2, \quad (2.4)$$

and consider the function $\sigma(x, y)$ defined by

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = -\frac{1}{4\pi(x-y)}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\sigma(x, y). \quad (2.5)$$

By using these new variables in Eq. (2.1), one finds that σ satisfies the equation

$$\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\left\{x(x-r_1)\left[\frac{\partial^2}{\partial x^2} + \frac{\alpha}{x} - \frac{\partial}{\partial\beta}\right] + y(y-r_1)\left[\frac{\partial^2}{\partial x^2} + \frac{\alpha}{y} - \frac{\partial}{\partial\beta}\right]\right\}\sigma = 0, \quad (2.6)$$

with boundary and initial conditions $\sigma(x, y) = 0$ when $x = 0$ or $y = 0$ and $\sigma(x, y) = \delta(x - y)$ at $\beta = 0$. Since ρ is symmetrical with respect to $\mathbf{r}_1$ and $\mathbf{r}_2$, one solution to this equation is the solution of

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\alpha}{x} - \frac{\partial}{\partial\beta}\right)\sigma(x, y) = 0, \quad (2.7)$$

with the same boundary and initial conditions. This equation is exactly that satisfied by the S-wave contribution $\rho_0(r_1, r_2, \beta)$ to $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$. Symmetry excludes any contribution to σ which depends only on $(x + y)$ and r_1, so σ must be identical to ρ_0. Hence,

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = -\frac{1}{4\pi(x-y)}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\rho_0(x, y, \beta), \quad (2.8)$$

where $x = (r_1 + r_2 + |\mathbf{r}_2 - \mathbf{r}_1|)/2$

and $y = (r_1 + r_2 - |\mathbf{r}_2 - \mathbf{r}_1|)/2.$

It is worthwhile to note that this relation is satisfied by the free-particle density matrix $\rho^{(0)}(\mathbf{r}_1, \mathbf{r}_2, \beta)$, i.e., in the case where $\alpha = 0$.

Equation (2.8) is satisfactory for calculating the off-diagonal terms of $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$. However, in any numerical computation of the diagonal terms, the factor $(x - y)$ occurring in the denominator leads to an awkward limiting procedure. We can improve this situation by changing variables to

$$r = (x + y)/2 \quad \text{and} \quad s = (x - y)/2,$$

and expanding $\rho_0(x, y, \beta)$ in powers of s. Then

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \beta) = -\frac{1}{8\pi s}\frac{\partial}{\partial s}\rho_0(r + s, r - s, \beta), \quad \text{for } \mathbf{r}_1 \neq \mathbf{r}_2, \tag{2.9}$$

and

$$\rho(\mathbf{r}, \mathbf{r}, \beta) = -\frac{1}{8\pi}\frac{\partial^2}{\partial s^2}\rho_0(r + s, r - s, \beta)\Big|_{s=0}. \tag{2.10}$$

Equations (2.9) and (2.10) then enable us to reduce the complete three-dimensional problem to a one-dimensional problem for which we can obtain a direct numerical solution.

3. NUMERICAL PROCEDURE

In view of the results of the last section, we will be primarily interested in the S-wave analog of Eq. (1.6). We can obtain this by using the expansion in Eq. (2.3) and the corresponding expansion for the free-particle density matrix, i.e.,

$$\rho^{(0)}(\mathbf{r}_1, \mathbf{r}_2, \beta) = \sum_{l=0}^{\infty}\frac{(2l+1)}{4\pi r_1 r_2}\rho_l^{(0)}(r_1, r_2, \beta)P_l(\cos\theta)$$
$$= \sum_{l=0}^{\infty}(2l+1)(4\pi\beta)^{-\frac{3}{2}} \times \exp(-(r_1^2 + r_2^2)/4\beta)i^{-l}j_l(r_1 r_2 i/2\beta)P_l(\cos\theta), \tag{3.1}$$

where j_l is the spherical Bessel function [$j_0(iz) = \sinh z/z$]. Then, using the orthogonality properties of the Legendre functions and equating the coefficients of $P_l(\cos\theta)$ on both sides of Eq. (1.6), one obtains

$$\rho_l(r_1, r_2, \beta) = \lim_{N\to\infty}\int_0^{\infty} ds_1 \cdots \int_0^{\infty} ds_N \times \prod_{i=1}^{N}\exp(-\beta V(s_i)/2N - \beta V(s_{i+1})/2N) \times \rho_l^{(0)}(s_i, s_{i+1}, \beta), \tag{3.2}$$

with $s_1 = r_1$ and $s_{N+1} = r_2$. Hence the analog of Eq. (1.10) for the S-wave contribution is

$$\rho_0(r_1, r_2, 2\beta) = \int_0^{\infty} dr\,\rho_0(r_1, r, \beta)\rho_0(r, r_2, \beta), \tag{3.3}$$

and for sufficiently small β we can approximate ρ_0 by

$$\rho_0(r_1, r_2, \beta) = \exp(\beta/r_1)\rho_0^{(0)}(r_1, r_2, \beta)\exp(\beta/r_2), \tag{3.4}$$

with

$$\rho_0^{(0)}(r_1, r_2, \beta) = (4\pi\beta)^{-\frac{1}{2}}\{\exp(-(r_1 - r_2)^2/4\beta) - \exp(-(r_1 + r_2)^2/4\beta)\}. \tag{3.5}$$

If we write Eq. (2.13) as

$$\rho_0(r_1, r_2, 2\beta) = \int_0^{R} dr\,\rho_0(r_1, r, \beta)\rho_0(r, r_2, \beta) + \int_R^{\infty} dr\,\rho_0(r_1, r, \beta)\rho_0(r, r_2, \beta), \tag{3.6}$$

we can readily use a trapezoidal rule to evaluate numerically the first term over a square grid of points in (r_1, r_2) space. Since the function $\rho_0^{(0)}$ falls off so rapidly when r differs significantly from r_1 or r_2, the criteria for this rule to work efficiently are satisfied (see, e.g., Sag and Szekeres[11]), so it is not necessary to use a more sophisticated rule. For a grid with M divisions each of width Δr, and $R = (M + \frac{1}{2})\Delta r$, Eq. (3.6) is then equivalent to calculating the square of the matrix $\rho_{ij}(\beta) = \rho_0(i\Delta r, j\Delta r, \beta)$, $1 < i < M$, $1 < j < M$ and adding the correction term

$$\int_R^{\infty} dr\,\rho_0(i\Delta r, r, \beta)\rho_0(r, j\Delta r, \beta) \tag{3.7}$$

to each element of $\rho_{ij}(2\beta)$. This correction term involves knowing $\rho_0(r_1, r_2, \beta)$ at each stage of the iteration for r_1 or r_2 greater than R and since we do not calculate these values of ρ_0, we need to find some method for finding an approximation for this term. For sufficiently large values of r_1 and r_2, ρ_0 is nearly equal to the S-wave contribution to the free-particle density matrix, $\rho_0^{(0)}$, so we could replace each factor ρ_0 in (3.7) by $\rho_0^{(0)}$; however, a better approximation is to use the following replacement in the integrand:

$$\rho_0(i\Delta r, r, \beta) \approx \frac{\rho_0(i\Delta r, M\Delta r, \beta)}{\rho_0^{(0)}(i\Delta r, M\Delta r, \beta)}\rho_0^{(0)}(i\Delta r, r, \beta), \tag{3.8}$$

since the major contribution to the integral comes from $r \approx R = (M + \frac{1}{2}\Delta r)$ and we calculate $\rho_0(i\Delta r, M\Delta r, \beta)$ at each stage of the iteration. Using Eq. (3.8) we can now obtain an approximate expression for the integral from R to ∞ in terms of error functions.

The basic iteration equation (3.6) is thus approximated by

$$\rho_{ij}(2\beta) = \Delta r\sum_{k=1}^{M}\rho_{ik}(\beta)\rho_{kj}(\beta) + \frac{\rho_{iM}(\beta)\rho_{Mj}(\beta)}{\rho_{iM}^{(0)}(\beta)\rho_{Mj}^{(0)}(\beta)}A_{ij},$$

[11] T. W. Sag and G. Szekeres, Math. Computation **18**, 245 (1964).

where

$$A_{ij} = \int_R^\infty dr \rho_0^{(0)}(i\Delta r, r, \beta)\rho_0^{(0)}(r, j\Delta r, \beta)$$
$$= \tfrac{1}{2}(8\pi\beta)^{-\frac{1}{2}}\left\{\exp\left(-\frac{(i-j)^2\Delta r^2}{8\beta}\right)\right.$$
$$\times\left[\operatorname{erfc}\left(\frac{2R-(i+j)\Delta r}{(8\beta)^{\frac{1}{2}}}\right)\right.$$
$$\left.+\operatorname{erfc}\left(\frac{2R+(i+j)\Delta r}{(8\beta)^{\frac{1}{2}}}\right)\right]$$
$$-\exp\left(-\frac{(i+j)^2\Delta r^2}{8\beta}\right)$$
$$\times\left[\operatorname{erfc}\left(\frac{2R-(i-j)\Delta r}{(8\beta)^{\frac{1}{2}}}\right)\right.$$
$$\left.\left.+\operatorname{erfc}\left(\frac{2R+(i-j)\Delta r}{(8\beta)^{\frac{1}{2}}}\right)\right]\right\} \quad (3.9)$$

with $R = (M + \frac{1}{2})\Delta r$. The function erfc (x) is defined by erfc $(x) = (2/\pi)^{\frac{1}{2}}\int_x^\infty dt \exp(-t^2)$. A value of β is chosen to start the iteration procedure so that $\rho_0(\beta)$ differs from $\rho_0^{(0)}(\beta)$ by a factor close to unity for all points on the grid. For this value of β we then set

$$\rho_{ij}(\beta) = (4\pi\beta)^{-\frac{1}{2}}\exp(\beta/\Delta ri + \beta/\Delta rj)$$
$$\times\{\exp(-(i-j)^2\Delta r^2/4\beta) - \exp(-(i+j)^2\Delta r^2/4\beta)\}. \quad (3.10)$$

Equation (3.9) will then enable us to find a numerical approximation for $\rho_0(r_1, r_2, 2\beta)$, i.e., at half the temperature, at all points on the square grid in (r_1, r_2) space.

From this approximation for $\rho_0(r_1, r_2, \beta)$, Eqs. (2.9) and (2.10) can then be used to find the complete Coulomb density matrix on a square grid in the $(x = (r_1 + r_2 + |\mathbf{r}_2 - \mathbf{r}_1|)/2,\ y = (r_1 + r_2 - |r_2 - r_1|)/2)$ space. The differentiation in Eq. (2.10) to find the diagonal elements of $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$ can be carried out numerically using the following rule:

$$\rho(\mathbf{r}, \mathbf{r}, \beta) = -\frac{1}{8\pi}\frac{\partial^2}{\partial s^2}\rho_0(r+s, r-s, \beta)\Big|_{s=0}$$
$$= -\frac{1}{8\pi(\Delta r)^2}[a\rho_{i,i} + b(\rho_{i,i+1} + \rho_{i-1,i})$$
$$+ c\rho_{i+1,i-1} + d(\rho_{i-1,i-1} + \rho_{i+1,i+1})]$$
$$+ O(\Delta r^2), \quad (3.11)$$

where $a + 2b + 2c + d = 0$, $b + 2c = 2$, $b + 4d = 0$, and $r = i\Delta r$. In addition, when $\rho_0 = \rho_0^{(0)}$, the choice of $b + 8c = 0$ gives an error term of only $O(\Delta r^4)$ instead of $O(\Delta r^2)$. With this restriction, the constants are $a = -\frac{22}{3}$, $b = \frac{16}{3}$, $c = -\frac{2}{3}$, $d = -\frac{4}{3}$. These constants were found to give the least error for the free-particle density matrix and so were used to find the diagonal elements of the complete Coulomb density matrix from ρ_{ij}.

4. RESULTS

These calculations were performed on the CDC 6400 computer at the University of Adelaide, South Australia. The computing time depends almost entirely on the value of M; with $M = 80$, each iteration takes approximately 8 sec. Figure 1 shows the two-particle approximation to the quantum radial-distribution function $g_2(r)$ for a proton–electron plasma for a range of temperatures from 10^4 to 16×10^4 °K. For small radial distances these results were calculated using $\Delta r = 0.1$ and $M = 80$ with the classical approximation, Eq. (3.4), with a value of β corresponding to $T = 2.048 \times 10^7$ °K, used as the starting point for the iteration procedure. Note that by comparison with Eq. (3.2), this calculation of $g_2(r)$ at a temperature of 10^4 °K is equivalent to evaluating a 2048-dimensional integral. An integral of this type would be hard

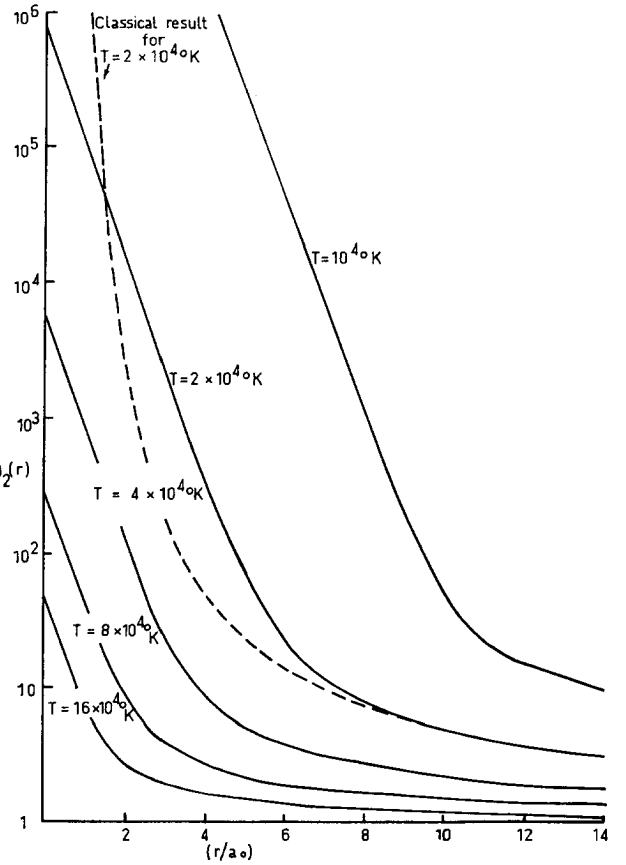


FIG. 1. The quantum radial-distribution function $g_2(r)$ for $T = 10^4$ to 16×10^4 °K. The corresponding classical radial-distribution function for $T = 2 \times 10^4$ °K is included for comparison.

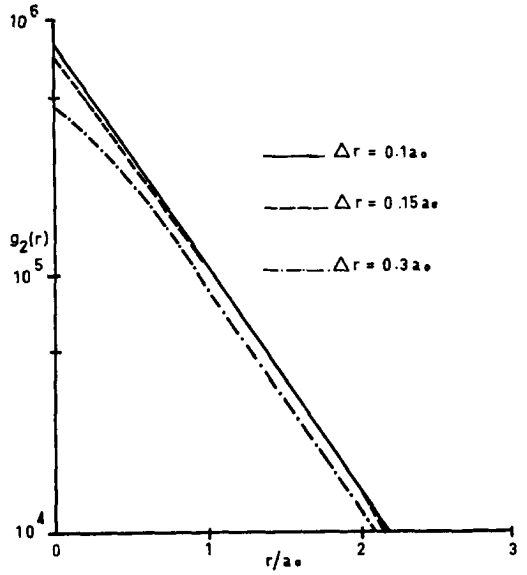


FIG. 2. An enlarged portion of Fig. 1, showing the calculated values of $g_2(r)$ using various values for the grid size Δr, for $T = 2 \times 10^4$ °K.

to evaluate by other means, e.g., Monte Carlo methods of integration, which tend to become very inefficient when the number of dimensions gets very large.

The effect of varying Δr is shown in Fig. 2. For $\Delta r < 0.1a_0$ there is little effect on the values near the origin, however, the range R of the calculation becomes rather small unless one uses prohibitively large values of M. The approximation in Eq. (3.8) causes a small error only for $r \approx R$ and this error can be detected readily by increasing the value of M. For the results shown in Fig. 1, this "edge effect" affects the values of ρ by more than 1% only for $r > R - 10\Delta r$ and one can arrange M or Δr so that this region is past the point where quantum effects on the radial-distribution function have any significant effect.

The two-particle classical approximation to the radial-distribution function is

$$\exp(-\beta V(r)) = \exp(\beta\alpha/r),$$

so to facilitate comparison between the classical and quantal results, we define an effective (temperature-dependent) potential by

$$g_2(r) = \exp[-\beta\alpha V_{\text{eff}}(r)]. \tag{4.1'}$$

Figures 3(a) and 3(b) show the effective potential for $T = 10^4$ °K to $T = 8 \times 10^4$ °K. It is interesting to note that at these temperatures, the effective potential is almost a straight line for small radial distances and joins the classical potential over quite a small range of r. Most modified potentials used previously were approximately equal to the classical potential for $r > a_0$ and had none of the marked temperature dependence shown here.

In the limit $\beta \to \infty$, the first term of Eq. (1.4), viz., $\exp(-\beta E_0)\psi_0^*(r_1)\psi_0(r_2)$, dominates $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$, so one could use the values of $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$ for large β to find the ground-state eigenvalue and eigenfunction for the Coulomb potential.

In addition, information about the higher eigenvalues could be obtained from the temperature

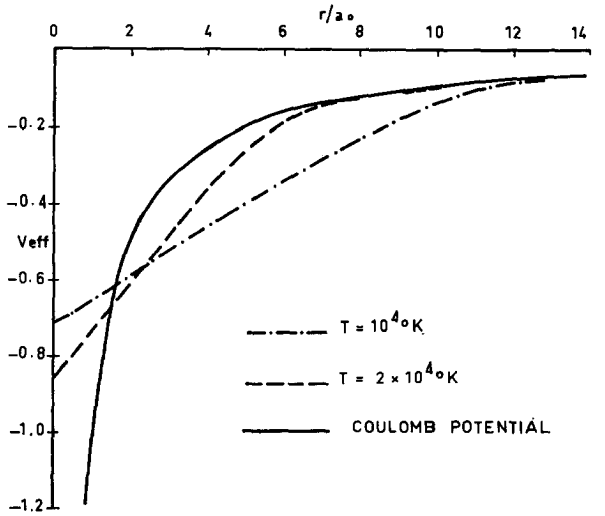


FIG. 3(a). The quantum effective potential $V_{\text{eff}}(r) = -\log g_2(r)/\beta\alpha$ compared with the attractive Coulomb potential $V_C(r) = -1/r$ for $T = 10^4$ °K and 2×10^4 °K.

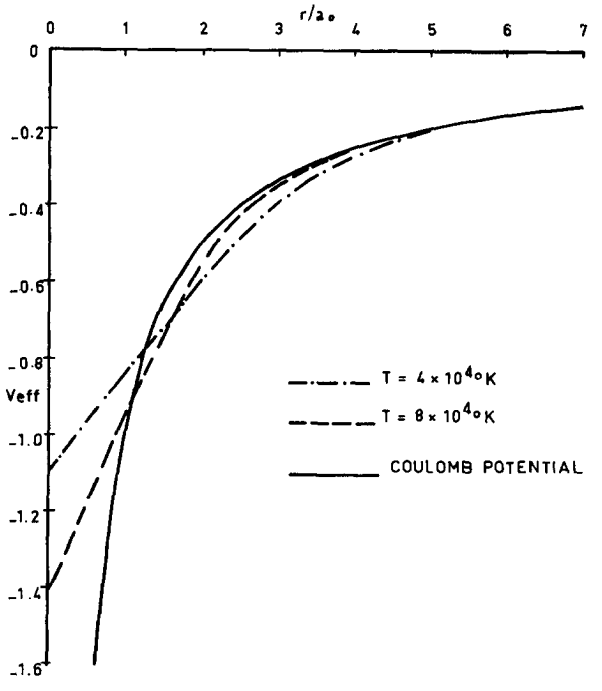


FIG. 3(b). The quantum effective potential $V_{\text{eff}}(r) = -\log g_2(r)$ $\beta\alpha$ compared with the attractive Coulomb potential $V_C(r) = -1/r$ for $T = 4 \times 10^4$ and 8×10^4 °K.

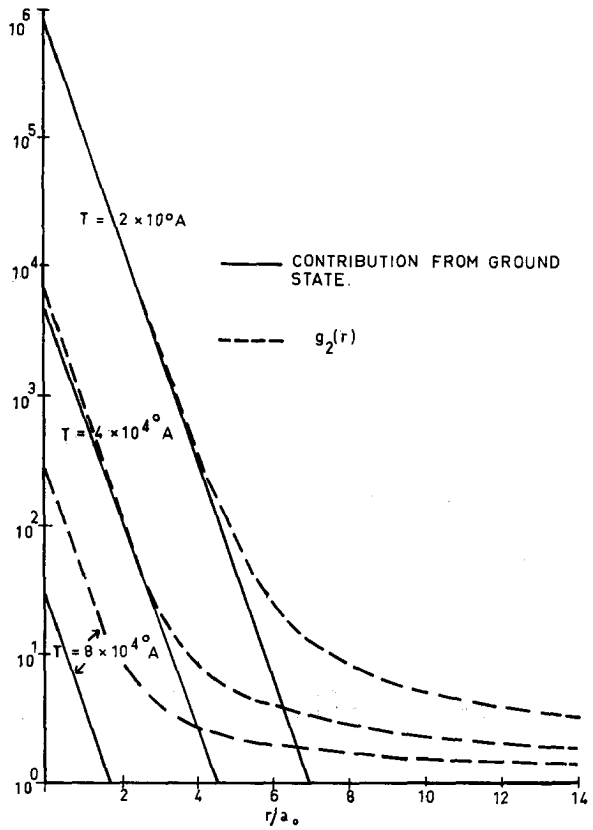


FIG. 4. A comparison between $g_2(r)$ and the contribution to $g_2(r)$ from the ground state of the two-particle system.

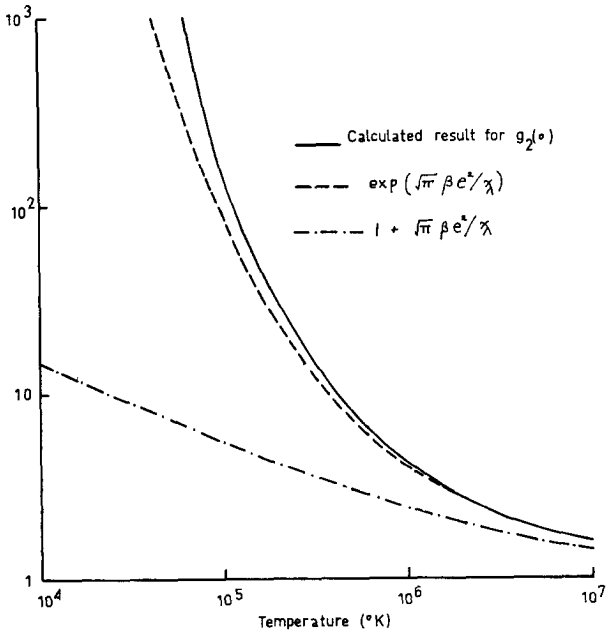


FIG. 5. A comparison between $g_2(0)$ and the approximations $\exp[(\pi)^{\frac{1}{2}}\beta e^2/\lambda\!\!\!^{-}]$ and $1 + (\pi)^{\frac{1}{2}}\beta e^2/\lambda\!\!\!^{-}$.

dependence of $\rho(\mathbf{r}_1, \mathbf{r}_2, \beta)$. Figure 4 shows a comparison between the calculated values of

$$g_2(r) = (2\pi\hbar^2\beta/\mu)^{\frac{3}{2}}\rho(\mathbf{r}, \mathbf{r}, \beta)$$

and the contribution to $g_2(r)$ from the ground state. It is the fact that one can obtain such a close correspondence between the exact and calculated values for the ground-state eigenvalue from the low-temperature results that gives one great confidence in the accuracy of the results at high temperatures. A calculation of the diagonal elements of the density matrix[2] using Eq. (1.4) and the exact Coulomb wavefunctions gives results within 0.1% of those obtained here. As is shown by these results, an iteration procedure coupled with integration, using numerical quadratures, constitutes a useful method for evaluating path integrals and hence finding the complete quantum-statistical density matrices for a range of temperatures.

There have been several analytical studies of the radial-distribution function at small separations. Trubnikov and Elesin,[12] Feix,[13] and Diesendorf and Ninham[14] have obtained the first two terms of the high-temperature expansion of $g_2(0)$,

$$g_2(0) \approx 1 + (\pi)^{\frac{1}{2}}\beta e^2/\lambda\!\!\!^{-}, \qquad (4.1)$$

where $\lambda\!\!\!^{-}$ is the thermal de Broglie wavelength $\lambda\!\!\!^{-} = \hbar(\beta/2\mu)^{\frac{1}{2}}$. The temperatures at which this expression is valid are far above those considered here; however the results obtained indicate that a conjecture made by DeWitt,[15] viz.,

$$g_2(0) \approx \exp[(\pi)^{\frac{1}{2}}\beta e^2/\lambda\!\!\!^{-}], \qquad (4.2)$$

is very close to the true result for high temperatures. Figure 5 shows a comparison between the approximations of Eqs. (4.1) and (4.2) and the calculated value of $g_2(0)$. It seems that regarding the electrons as being "smeared out" over a distance of the order of $\lambda\!\!\!^{-}$ is quite a reasonable way of picturing their behavior in this case.

ACKNOWLEDGMENTS

The author is grateful to A. A. Barker for informing him of results for a calculation of the radial-distribution functions by a direct sum over states, to Dr. Hugh E. DeWitt for suggesting Eq. (4.2) and pointing out the previous analytical results, and to the Australian Institute of Nuclear Science and Engineering for financial support.

[12] B. A. Trubnikov and V. F. Elesin, Zh. Eksp. Teor. Fiz. **47**, 1279 (1965) [Sov. Phys.—JETP **20**, 866 (1965)].

[13] M. Feix, *Proceedings of the Sixth International Conference on Ionization Phenomena in Gases, Paris, 1963* (Serma, Paris, 1964), Vol. II, pp. 185–187.

[14] M. Diesendorf and B. W. Ninham (to be published).

[15] H. E. DeWitt (private communication).