JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Spinor Structure of Space-Times in General Relativity. I

ROBERT GEROCH*
Department of Mathematics, Birkbeck College, London, England

(Received 8 January 1968)

In order to define spinor fields on a space-time M, it is necessary first to endow M with some further structure in addition to its Lorentz metric. This is the spinor structure. The definition and the elementary implications of the existence of a spinor structure are discussed. It is proved that a necessary and sufficient condition for a noncompact space-time M to admit a spinor structure is that M have a global field of orthonormal tetrads.

INTRODUCTION

It is sometimes convenient to formulate general properties which any space-time[1] should have if it is to be of physical interest. For example, one would normally reject space-times having closed timelike curves,[2] and possibly also those having incomplete geodesics.[3] It has been argued,[4] based on a recent paper of Aharonov and Susskind,[5] that another property we might require of a physically reasonable space-time M is that M admit globally defined spinor fields.[6] Now, in order that spinors may be defined on M, it is necessary that M satisfy certain global conditions.[4] In the case of space-times (but not, in general, for other signatures or dimensions), these conditions have the following consequence: *Every noncompact*[7] *space-time on which spinors may be defined carries a global field of orthonormal tetrads.* This, our main result, is proved in Sec. II.

In Sec. I we review the definition of a spinor field and the basic properties of space-times which admit such fields.

The Appendix contains a proof that every space-time is paracompact.

I. SPINOR STRUCTURE

Let P be a point of a space-time M. A spinor[8] at P consists, at least, of a rule which assigns to each[9] orthonormal tetrad w at P an array of complex numbers $\xi_{C\cdots}^{A\cdots\ B'\cdots\ D'\cdots}(w)$, this array given only up to an over-all sign. The individual numbers in each array are labeled by indices $A, \cdots, B', \cdots, C, \cdots, D', \cdots$, whose range is 1, 2. To complete the definition we must specify how such an array is to change under a change in the choice of tetrad. Let v and w be two tetrads at P, and suppose that the Lorentz transformation L which carries v to w preserves the time direction and the spatial parity, i.e., that L is in the restricted Lorentz group $\mathfrak{L}_0$. Since[8] the group $SL(2, C)$ of unimodular complex 2×2 matrices is the universal covering space of the (doubly connected) group $\mathfrak{L}_0$, there corresponds to the element L precisely two elements $\pm U^A{}_B$ of $SL(2, C)$.[10] We now require that the arrays associated with the tetrads v and w be related as follows:

$$\xi_{C\cdots\ D'\cdots}^{A\cdots\ B'\cdots}(w) = \pm U^A{}_E \cdots \bar{U}^{B'}{}_{F'} \cdots (U^{-1})^G{}_C \cdots \overline{(U^{-1})}^{H'}{}_{D'} \xi_{G\cdots\ H'\cdots}^{E\cdots\ F'\cdots}(v). \quad (1)$$

A sign ambiguity still remains in Eq. (1). It is the condition that this "two-valuedness" can be consistently eliminated over the entire space-time M that places some restrictions on the global structure of M.

Let us first of all complete our definition of a spinor at the single point P. Fix a reference tetrad w at P. We now deal, not with the set Ψ of all tetrads at P, but rather with the collection $\tilde{\Psi}$ of all pairs (v, α),

---

* U.S. Air Force Office of Scientific Research postdoctoral fellow.

[1] By a *space-time* we understand a mathematical object: a 4-manifold with a C^∞ metric of signature $(+, -, -, -)$.

[2] See, for example, H. Reichenbach, *The Philosophy of Space and Time* (Dover Publications, Inc., New York, 1958), Sec. 21.

[3] See C. W. Misner, J. Math. Phys., **4**, 924 (1963); R. Geroch, "What is a Singularity in General Relativity?," Ann. Phys. (N.Y.) (to be published).

[4] R. Penrose, "The Structure of Space-Time" in *Battelle Recontres in Mathematics and Physics: Seattle, 1967*, C. DeWitt, Ed. (W. A. Benjamin, Inc., New York, 1968).

[5] Y. Aharonov and L. Susskind, Phys. Rev. **158**, 1237 (1967).

[6] Note that we are not concerned with the existence of spinor fields having special properties, but merely with whether or not the notion of a spinor field is well defined.

[7] This assumption is quite reasonable from the physical point of view because every compact space-time is known to have closed timelike curves. "Compact" and "noncompact" refer, of course, to the entire 4-dimensional space-time manifold.

[8] See, for example, F. A. E. Pirani, in *Lectures in General Relativity, Brandeis Summer Institute in Theoretical Physics, 1964* (Prentice Hall, Inc., Englewood Cliffs, N.J., 1965). See also Ref. 4.

[9] We shall be concerned only with those tetrads having a certain preassigned temporal and spatial orientation.

[10] This ambiguity in sign cannot be removed in a continuous way. The situation is similar to the sign ambiguity in choosing a normal vector field to a Möbius band embedded in Euclidean 3-space.

Copyright © 1968 by the American Institute of Physics

where $v \in \Psi$ and α is a path in Ψ from v to w.[11] The pairs (v, α) and (u, β) are considered to represent the same point of $\tilde{\Psi}$ if $u = v$ and if the path α can be continuously distorted into the path β while keeping its endpoints fixed. In other words, $\tilde{\Psi}$ is the universal covering manifold of the six-dimensional manifold Ψ. (Intuitively, we may think of $\tilde{\Psi}$ as a collection of tetrads at P, but such that a tetrad is considered to be a different object on rotation through 360°, while it is left unchanged by a rotation through 720°.) The crucial property of $\tilde{\Psi}$ is that any two elements v and w of $\tilde{\Psi}$ are transformed into each other by a *unique* element $U^A{}_B$ of $SL(2, C)$. We may now define a spinor at P as a rule which assigns to each element v of $\tilde{\Psi}$ an array $\xi_{C\cdots D'\cdots}^{A\cdots B'\cdots}(v)$ of complex numbers such that if $v, w \in \tilde{\Psi}$ are related by $U^A{}_B \in SL(2, C)$, then

$$\xi_{C\cdots D'\cdots}^{A\cdots B'\cdots}(w) = U^A{}_E \cdots \bar{U}^{B'}{}_{F'} \cdots (U^{-1})^G{}_C \cdots \overline{(U^{-1})^{H'}{}_{D'}} \xi_{G\cdots H'\cdots}^{E\cdots F'\cdots}(v). \quad (2)$$

This definition would serve also to define spinor fields over all of Minkowski space, since a tetrad at any point can be referred to the origin by means of parallel transport.

Unfortunately, in a general space–time M, it is not possible to set up a continuous correspondence between the tangent space of each point of M and that of a single fixed point $P \in M$. What is required in order to define spinor fields on M is to do globally what we have just done locally at P. One must take the covering space of the 6-manifold of tetrads at each point of M and then piece together this 4-dimensional collection of 6-manifolds. The process of "piecing together" results in a fiber bundle.

Let the space–time M have a given time and space orientation, and let B be the principal fiber bundle[12] of oriented orthonormal tetrads on M. That is, B is a 10-manifold. Each point of B consists of a tetrad at a single point of M. The group of B is the restricted Lorentz group $\mathfrak{L}_0$, while the fiber over a point $P \in M$ is the collection Ψ of tetrads at P having the prescribed temporal and spatial orientation. A spinor field on M could now be defined (up to sign) as a mapping, with the appropriate transformation properties, from B into arrays of complex numbers. To correct the sign ambiguity, we must have a fiber bundle whose fiber is not Ψ, but rather the universal covering space $\tilde{\Psi}$ of Ψ. A *spinor structure*[13] on M is defined as a second principal fiber bundle $\tilde{B}$ [with group $SL(2, C)$] over M, along with a 2–1 mapping $\varphi: \tilde{B} \to B$, such that:

(1) φ maps each fiber of $\tilde{B}$ into a single fiber of B;

(2) φ commutes with the group operations. That is, for each $U \in SL(2, C)$, $\varphi \circ U = \Lambda(U) \circ \varphi$, where $\Lambda: SL(2, C) \to \mathfrak{L}_0$ is the covering mapping of the restricted Lorentz group.[14]

A space–time M, if it has any spinor structure at all, does not have a unique one unless M is simply connected.

Given a spinor structure $\tilde{B}$, a spinor field on M may be defined,[15] just as before, as a mapping ξ of $\tilde{B}$ into arrays of complex numbers such that ξ transforms as in Eq. (2).

The mere fact that a spinor formalism is useful in some calculations is a rather unsatisfactory reason to include the existence of a spinor structure among the "reasonable physical conditions" on space–times. However, an argument due to Penrose,[4] based on a gedanken experiment of Aharonov and Susskind,[5] makes the assumption of a spinor structure somewhat more plausible. The Aharonov–Susskind apparatus consists of a box which may be separated into two halves. If we join the two halves together, a current flows from one half to the other. If we now separate the halves, rotate one through 360° while keeping the other fixed, and then rejoin the halves, the direction of current flow is reversed. Thus the Aharonov–Susskind boxes behave somewhat like a spinor in that they are capable of keeping track of a relative rotation through 360°. This experiment demonstrates that the

[11] Our $\tilde{\Psi}$ is essentially the collection of spin frames at P. See E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[12] See, for example, N. Steenrod, *The Topology of Fibre Bundles* (Princeton University Press, Princeton, N.J., 1951).

[13] See Ref. 4. Essentially because of the way that the Lorentz group is embedded in the general linear group GL(4, R), there is a one-to-one correspondence between the spinor structures on a (space- and time-oriented) space–time M and the spin structures on the tangent bundle of the underlying manifold of M. See J. Milnor, L'enseignement math., **9**, 198 (1963); also "Remarks Concerning Spin Manifolds," in *Differential and Combinatorial Topology*, S. S. Cairns, Ed. (Princeton University Press, Princeton, N.J., 1965), p. 55.

[14] It is well known that a necessary and sufficient condition that a space- and time-oriented space–time M have spinor structure is that the second Stiefel–Whitney class of M vanish. See Ref. 13.

[15] Penrose (Ref. 4) has introduced a generalization of this definition which may be more convenient when conformal transformations are contemplated. Define an orthotetrad at P as a collection of four vectors $k^\alpha_{(i)}$, $i = 1, 2, 3, 4$, at P which satisfy

$$k^\alpha_{(i)} k^\beta_{(j)} g_{\alpha\beta} = \lambda \eta_{ij},$$

where λ is a positive number and where η_{ij} is the matrix diagonal $(+1, -1, -1, -1)$. Now define a spinor as a rule which assigns to each orthotetrad k at P an array of complex numbers $\xi_{C\cdots D'\cdots}^{A\cdots B'\cdots}(k)$ such that (a) under a Lorentz transformation the array behaves as in Eq. (2), and (b) under the transformation $k^\alpha_{(i)} \to \mu k^\alpha_{(i)}$ the array transforms as follows:

$$\xi_{C\cdots D'\cdots}^{A\cdots B'\cdots}(\mu k) = \mu^{-n/2} \xi_{C\cdots D'\cdots}^{A\cdots B'\cdots}(k).$$

Here n is the conformal weight of the spinor.

two-valuedness of spinors can be realized physically by macroscopic systems. If such Aharonov–Susskind boxes could be moved throughout space-time (i.e., along curved world lines and over regions of nonvanishing curvature) and be intercompared with one another—all without their loosing track of the "spin-orientation" information—we would, in principle, have an experimental means to measure the spinor structure of our own universe.

Two conditions, each necessary and sufficient for the existence of a spinor structure, can be obtained fairly easily from the definitions.[16] These conditions provide some insight into the global implications for a space-time of our requirement that it carry spinor fields.

Let B be the bundle of oriented frames of the space-time M. Then M has a spinor structure if and only if the fundamental groups[17] of B and M are related as follows:

$$\pi_1(B) \approx \pi_1(M) \oplus \pi_1(\Psi). \tag{3}$$

[Recall that $\pi_1(\Psi) = Z_2$.] To see the implications of this equation, assume for the moment that Eq. (3) holds. We may then take a (double) covering space of B, this space obtained by "unwrapping Ψ," i.e., by annihilating $\pi_1(\Psi)$ while leaving $\pi_1(M)$ unchanged. Such a covering space turns out to be just a spinor structure $\tilde{B}$ on M. Thus the existence of a spinor structure is equivalent to the statement that we may "unwrap each of the fibers on the bundle of frames without at the same time unwrapping any of the underlying manifold M."

Consider next a closed curve γ in B which lies entirely in the fiber over one point P of M. Suppose γ is so chosen that it cannot be contracted to a point while remaining in the fiber (i.e., γ corresponds to a rotation of a tetrad at P through 360°). We now permit γ to be distorted throughout the entire bundle B. That it still be impossible to contract γ to a point is a necessary and sufficient condition that M admit a spinor structure (assuming, once again, that M is space- and time-oriented). This is just the result we should have expected intuitively. Consider a (one-index) spinor attached[18] to a frame at $P \in M$. When the frame is rotated through 360° at P, the spinor reverses its sign. This rotation of the frame defines a curve in the fiber of B over the point P. Suppose now that this curve can be contracted, in B, to a point. We then arrive at a contradiction because a "sign change" (which occurs when the curve is in the fiber over P) has been continuously distorted to "no sign change" (which occurs when the curve has finally been contracted to a point).

We mention one further property of spinor structures: A space- and time-oriented space-time has a spinor structure if and only if each of its covering manifolds does.[19] It is of interest to compare spinor structure and other global properties of space-times with respect to behavior under taking a covering manifold. It is known,[20] for example, that (1) the experimental evidence on nonconservation of C, P, and CP in elementary-particle reactions, (2) the CPT theorem, and (3) the strong principle of equivalence[21] together imply that our universe must be orientable. However, from one point of view this result places no restrictions whatever on the underlying manifold of our universe: any space-time (whether orientable or not) has a covering manifold which, while representing exactly the same physical universe, is necessarily orientable. We see that the question of the existence of spinor fields is very different from the question of the existence of an orientation in that we cannot "create" spinor structure merely by taking a covering manifold.

We conclude Sec. I with an example of a space-time which has no spinor structure. In fact, we may just as easily display what is, in a sense, to be clarified shortly—the "generic" example. Let A_1 and A_2 each be a copy of the cross product of the closed unit 2-disc (polar coordinates θ_i and $r_i \leq 1$, $i = 1, 2$) with the 2-dimensional plane (polar coordinates φ_i and ρ_i). We now identify the boundaries (each diffeomorphically $S^1 \times R^2$) of A_1 and A_2 as follows: the point $(\theta_1, r_1, \varphi_1, \rho_1)$ of A_1 is identified with the point $(\theta_2, r_2, \varphi_2, \rho_2)$ of A_2 if

$$\theta_1 = \theta_2, \qquad \varphi_1 = \varphi_2 + m\varphi_1,$$
$$r_1 = r_2 = 1, \quad \rho_1 = \rho_2,$$

where m is a fixed nonnegative integer. That is, for each m we define a 4-manifold (without boundary) M_m. (In fact,[22] the M_m are precisely the collection of all R^2 bundles over S^2. In particular, M_2 = tangent

[16] See Refs. 4 and 13.

[17] See, for example, A. H. Wallace, *Introduction to Algebraic Topology* (Pergamon Press, Inc., New York, 1957).

[18] That is, ξ, by definition, changes as the frame w is changed, so that the components of ξ relative to w are constant.

[19] This is so essentially because the spinor structure involves the second homotopy group of M, while the operation of taking a covering space acts only on the first homotopy group.

[20] R. Geroch, "Singularities in the Spacetime of General Relativity," Ph.D. thesis, Dept. of Physics, Princeton University, 1967 (unpublished). The same type of argument has been used by Ya. B. Zeldovich and I. D. Novikov ("The Topology of the Universe: Restrictions from Elementary Particle Physics," submitted to Zh. Eksp. Teor. Fiz. Pis'ma Redaktsiya; English transl.: JETP Letters) to obtain a slightly different result.

[21] R. H. Dicke, Science **129**, 621 (1951).

[22] See Ref. 12, p. 96.

bundle of S^2.) We remark that each M_m may be given a metric of Lorentz signature.[23]

If m is odd, then, no matter what Lorentz metric is placed on M_m, M_m has no spinor structure. To verify this, choose a closed curve γ which lies entirely in the fiber over the point $\rho_1 = 0, r_1 = 0$, of M_m, and which is not homotopically zero in the fiber. Observe that if m is odd, then γ may be distorted to a point in the bundle of frames of M_m by sliding it over the 2-sphere $\rho_1 = 0, \rho_2 = 0$.

Our example is the generic one in the following sense. Let M be a space–time which cannot be given a spinor structure. Then there must be some closed curved γ lying in the fiber over a point $P \in M$ such that, while γ is not homotopically zero in the fiber, it can be contracted to a point in the entire bundle of frames. Contract γ to a point, while at the same time projecting its path into M. We thus obtain the image S in M of a 2-sphere. Suppose[24] that S can be so chosen that it is smooth and does not intersect itself. Then we may select a neighborhood V of S which is topologically a 2-dimensional vector bundle over S^2. Since it contains S as a subset, V itself does not admit a spinor structure, and must therefore be just one of the M_m for m odd. That is, the space–time M contains an open subset which is diffeomorphic to one of the M_m, m odd.

II. GLOBAL SYSTEMS OF TETRADS

Our main result reexpresses the existence of a spinor structure in more familiar terms.

Theorem: Let M be a noncompact[7] space–time.[1] Then M has a spinor structure if and only if there exists on M a global system of (orthonormal) tetrads.[25]

Proof: Suppose first of all that M has a spinor structure $\tilde{B}$. Since M is a space–time, it is paracompact (see the Appendix). Therefore[26] M may be triangulated.[27] (Since M is noncompact, there will necessarily be an infinite number of simplices.) Choose[28] a sequence $M_1, M_2, \cdots$ of subcomplexes of M such that:

(1) Each M_i is, as a topological space, a compact 4-manifold with boundary. (It follows that each M_i must consist of only a finite number of simplices.)

(2) $M_i \supset M_{i-1}$, $i = 2, 3, \cdots$; and $\bigcup_i M_i = M$.

(3) Each connected component of $M_i - M_{i-1}$ contains at least one boundary 3-simplex, $i = 2, 3, \cdots$.

The underlying group of the bundle $\tilde{B}$ is $SL(2, C)$. But $SL(2, C)$ is topologically $R^3 \times S^3$, and so its homotopy groups are easily calculated:

$$\pi_q(SL(2, C)) \approx \pi_q(R^3) \oplus \pi_q(S^3) \approx \pi_q(S^3).$$

It is convenient to list these groups here for later reference:

$$\pi_1(SL(2, C)) = 0, \quad \pi_2(SL(2, C)) = 0,$$
$$\pi_3(SL(2, C)) = Z.$$

The proof consists of constructing a cross section of the principal fiber bundle $\tilde{B}$. We proceed inductively, extending a cross section from one M_i to the next. Let us suppose, therefore, that we are given a cross section of $\tilde{B}$ over $J \equiv M_{i-1}$, and that we wish to extend this cross section to $K \equiv M_i$ (in case $i = 1$, set $J \equiv \phi$). Denote by K^j ($j = 0, 1, 2, 3, 4$) the j-skeleton of K (i.e., the union of all simplices of K of dimension less than or equal to j).

It follows[29] from the vanishing of the first two homotopy groups of $SL(2, C)$ that our given cross section over J can be extended uniquely (up to homotopy) to a cross section C over $J \cup K^2$, and further that C can be extended (though not, in general, uniquely) to a cross section over $J \cup K^3$. Since $\pi_3(SL(2, C)) \neq 0$, however, a given cross section over $J \cup K^3$ cannot, in general, be extended over $J \cup K^4 = K$. One would like to use the freedom available in selecting an extension of C over $J \cup K^3$ to find one such extension C' having the property that C' can be further extended over $J \cup K^4$. However, such a C' does not in general exist. In fact, the cross section C determines[30] an element α of the cohomology group $H^4(K, J; Z)$ (a "characteristic class"). The vanishing

[23] See L. Markus, Ann. Math. **62**, 411 (1955). In fact, this result of Markus has been strengthened slightly (Ref. 4): Every noncompact 4-manifold may be given a time-oriented Lorentz metric which has, in addition, no closed timelike curves.

[24] In fact, we need only require for this argument that there be some 2-submanifold S of M such that S is diffeomorphically an S^2, and such that S, considered as an element of $\pi_2(M)$, is not in the kernel of the homomorphism Δ in the (exact) homotopy sequence of the bundle B (Ref. 12, p. 91):

$$\cdots \to \pi_2(B) \to \pi_2(M) \xrightarrow{\Delta} \pi_1(\Phi) \to \cdots.$$

Can such an S always be found?

[25] We actually prove slightly more than this. If M has spinor structure $\tilde{B}$, then the global system of tetrads (considered now as a cross section T of B) can be so chosen that T may be lifted to a cross section of $\tilde{B}$.

[26] See J. H. C. Whitehead, Ann. Math., **41**, 809 (1940).

[27] S. S. Cairns, *Introductory Topology* (Ronald Press, New York, 1961), p. 76.

[28] Subcomplexes M_i which satisfy these conditions may be constructed in the following way. Let $\sigma_1, \sigma_2, \cdots$ denote the 4-simplices of M. Define

$$M_0 \equiv \phi,$$
$$M_i \equiv \bigcup_{k \in \Gamma_i} \bar{\sigma}_k, \quad i = 1, 2, \cdots,$$

where

$\Gamma_i = \{$positive integers k such that there exists a compact set C with $\partial C \in (M_{i-1} \cup \bar{\sigma}_i)$ and with $\sigma_k \subset C\}$.

[29] Ref. 12, p. 149.

[30] Ref. 12, p. 174.

of α is a necessary and sufficient condition for the existence of a C' which *can* be further extended over $J \cup K^4 = K$.

Our three conditions on the M_i have as a consequence, however, that $H^4(K, J; Z)$ vanishes identically.[31] We conclude that $\alpha = 0$, and therefore that a cross section over K exists.

We have shown that any cross section of $\tilde{B}$ over J can be extended to a cross section over K. Setting $K = M_1$ and $J = \phi$, we obtain a cross section over M_1. Setting $K = M_2$ and $J = M_1$, we extend this cross section from M_1 to $M_2 \supset M_1$. Continuing in this way, we obtain finally a cross section of $\tilde{B}$ over all of M. The projection of our cross section from $\tilde{B}$ to B yields the required global system of tetrads.

The converse—if M has a global system of tetrads, then it has a spinor structure—is an immediate consequence of the definition of spinor structure.

This theorem depends critically on the vanishing of homotopy groups of the spinor group $SL(2, C)$. The first homotopy group vanishes essentially because a spinor structure is defined by the property that its fiber is the universal covering space of the fiber of the bundle of frames. (Taking the universal covering space automatically annihilates the first homotopy group.) The second homotopy group vanishes for all the spin groups (in fact, for all Lie groups). The third homotopy group fails to vanish, but at this point we are sufficiently close to the dimension of the manifold that the obstruction to extending a cross section can be made to vanish. Thus the dimension of the manifold enters in an essential way. In fact, the theorem is true in four dimensions, uninteresting in lower dimensions (in this case, every orientable manifold is parallelizable[32]), and false in higher dimensions.

We remark on one application of our theorem, an application to the problem of classifying space-times according to their global structure. The homotopy classes[33] of metrics of Lorentz signature on a given manifold M are in one-to-one correspondence with homotopy classes of nowhere-vanishing vector fields on M. Let us agree that a space-time, to be of physical interest, must admit a spinor structure. Then we may, according to the theorem, take M to be parallelizable, and so its tangent bundle is the cross product $M \times R^4$. Therefore the homotopy classes of Lorentz metrics on M are in one-to-one correspondence with the collection Λ of homotopy classes of maps $\lambda: M \to S^3$. It might be interesting to see if, given M, the homotopy classes Λ may be characterized in some fairly simple way. Such a characterization would be a start toward a classification of the "homotopy classes of space-times": Which global properties of space-times are invariants of the homotopy class?

ACKNOWLEDGMENTS

I wish to thank R. Penrose for valuable discussions throughout the course of this work, as well as for reading the manuscript and suggesting several improvements.

This work was carried out under a U.S. Air Force Office of Scientific Research postdoctoral fellowship.

APPENDIX

The following theorem makes it possible to simplify the statement of many other theorems. In dealing with a manifold M which carries a metric of Lorentz signature, it is not necessary to specify that, in addition, M have a countable basis (or, equivalently, that M be paracompact).

Theorem: Let M be a (connected, Hausdorff) 4-manifold with a C^∞ metric of signature $(+, -, -, -)$. Then the topology of M has a countable basis.[34]

Proof: Fix once and for all a countable basis O_i $(i = 1, 2, \cdots)$ for the open sets of R^4 and, for each i, a point $q_i \in O_i$.

Let P be a point of M. The exponential map[35] "exp" is a continuous function from a subset of T_P, the tangent space at P, into M. Let t_P denote the union of all open subsets of T_P on which "exp" is defined and is an open map.[36] We first show that "exp" is an

[31] That is, each 4-simplex Σ of $K - J$ can be represented as the coboundary of some 3-chain $\mathcal{C}$ lying entirely in $K - J$. To see this, let γ be a curve from Σ to the boundary of $K - J$, so chosen that $\gamma \cap (K^2 \cup J) = \phi$. Then define $\mathcal{C}$ to be the (properly counted) sum of the 3-simplices which intersect γ.

[32] E. Stiefel, Comm. Math. Helv., **8**, 3 (1936).

[33] The idea of looking at homotopy classes of metrics was considered by D. Finkelstein and C. W. Misner [Ann. Phys. (N.Y.) **6**, 230 (1959)]. Their approach differs from the one contemplated here, however, in the following way. Finkelstein and Misner consider space-times which are topologically R^4. Consequently, it is necessary to impose a boundary condition (asymptotic flatness) to prevent the "twists" in the metric from being lost to infinity. We impose no boundary conditions, but as a result we must require that the manifold have non-Euclidean topology (an "O-geon" in the Finkelstein-Misner terminology) if there is to be even the possibility of more than one homotopy class of Lorentz metrics. In addition, our homotopy classes do not have a group structure, as do those of Finkelstein-Misner.

[34] The theorem actually proved here is somewhat stronger: Every n-manifold with a C^1 connection has a countable basis. Surprisingly enough, it is *not* true that every manifold with a conformal Lorentz metric (i.e., a Lorentz metric given at each point only up to an arbitrary nonzero factor) has a countable basis. A counterexample is the 2-manifold defined as the Cartesian product of two "long lines" [J. G. Hocking and G. S. Young, *Topology* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1961), p. 55]. The null directions (which define the conformal metric) are specified as lying along the coordinate axes at each point.

[35] See, for example, N. J. Hicks, *Notes on Differential Geometry* (D. Van Nostrand, Inc., Princeton, N.J., 1965), p. 131.

[36] A mapping is said to be open if the image of each open set is again an open set.

open map on t_P. Let V be an arbitrary open subset of t_P, and let V_α be a covering of t_P by open sets on each of which "exp" is an open map. For each α, $V \cap V_\alpha$ is open in V_α, and therefore $\exp(V \cap V_\alpha)$ is open in M. It follows that

$$\exp(V) = \bigcup_\alpha \exp(V \cap V_\alpha)$$

is open in M. Since V is arbitrary, "exp" is an open map on t_P. Define $S_P \equiv \exp(t_P) \subset M$.

Now let v be a tetrad at P. The tetrad v determines a natural diffeomorphism $\varphi_v: T_P \to R^4$, where, for each vector $\xi \in T_P$, $\varphi_v(\xi)$ is defined as the 4-tuple of components of ξ relative to v. We define the composition $\theta \equiv \exp \circ \varphi_v^{-1}$, a continuous map from a subset of R^4 into M.

Whenever the integer i is such that $q_i \in \theta^{-1}(S_P)$, we define

$$P_i = \theta(q_i) \in M,$$
$$U_i \equiv \theta(O_i \cap \theta^{-1}(S_P)) \subset M.$$

At each of the points P_i we define a tetrad v_i by parallel transport of the tetrad v along the geodesic[37] from P to P_i.

We now show that the U_i are a basis for the open sets of S_P. Since the O_i cover R^4, the U_i cover S_P. Since θ is an open map on $\theta^{-1}(S_P)$, each U_i is an open subset of M. Let U be any open subset of S_P. Then $\theta^{-1}(U)$ is open in R^4. Since the O_i are a basis for R^4, we may select a collection Γ of integers such that $\bigcup_{i\in\Gamma} O_i = \theta^{-1}(U)$. Therefore, $\bigcup_{i\in\Gamma} U_i = U$. Since U is arbitrary, the U_i are a basis for S_P.

To summarize, given any pair (P, v), where v is a tetrad at the point P, we have defined an open neighborhood S_P of P in M, a countable basis U_i for S_P, a countable dense set P_i in S_P, and a tetrad v_i at each P_i. We now repeat our construction for each pair (P_i, v_i). That is, we define an open neighborhood S_{P_i} of P_i, a basis U_{ij} for S_{P_i}, a dense set P_{ij} in S_{P_i}, and a tetrad v_{ij} at each P_{ij} (i fixed, $j = 1, 2, \cdots$). Applying the same construction to the pairs (P_{ij}, v_{ij}), and so on, we obtain finally a countable collection $U_\alpha \equiv \{U_i, U_{ij}, U_{ijk}, \cdots\}$ of open sets of M and a countable collection $P_\alpha \equiv \{P_i, P_{ij}, P_{ijk}, \cdots\}$ of points of M.

Since the U_α form a basis for each of the S_{P_β}, they also form a basis for $S \equiv \bigcup_\alpha S_{P_\alpha}$. We must finally show that $S = M$.

Let Q be any point in the closure of S. There is some neighborhood U of Q such that no geodesic segment in U has a pair of conjugate points in U. Since Q is in the closure of S, $U \cap S \neq \phi$. Therefore, since the P_α are dense in S, there must be some $P_\beta \in U$. But no geodesic segment has a pair of conjugate points in U, and so $S_{P_\beta} \supset U$. We have shown that $Q \in U \subset S_{P_\beta} \subset S$, i.e., that S contains each point of its closure. But S is also open in M, because it is the union of the open sets S_{P_α}. Since M is connected, it follows that $S = M$. This completes the proof.

[37] In case there are several geodesics from P to P_i, let us select, for definiteness, that one which has been used (in the exponential map) to define P_i.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

New Solutions of the Einstein-Maxwell Equations from Old*

B. KENT HARRISON†
Physics Department, Brigham Young University, Provo, Utah

(Received 14 December 1967)

Methods are discussed with which one may derive theorems which allow one to generate new solutions of the Einstein–Maxwell equations from old ones. The old solutions used to generate new ones must admit at least one nonnull Killing vector and may be required to satisfy other conditions, depending on the theorem derived. Examples of derivable theorems are shown; these theorems are used in turn to show how generation of new solutions is accomplished. Examples of the latter are shown, such as generation of Brill or electrified NUT space from the Schwarzschild solution, generation of a new twisted Melvin universe from flat space, and generation of a new generalization of the Ozsvath–Schücking metric. Possible physical interpretations, uses, and extensions of this type of theorem are discussed.

1. ASSUMED METRIC AND ITS EINSTEIN-MAXWELL EQUATIONS

We outline a method, suitable for use on a wide class of metrics, by which one can derive theorems which in turn can be used to obtain new solutions of the Einstein–Maxwell equations from old. In Sec. 1, we find the form of the Einstein–Maxwell equations for the assumed class of metrics; in Sec. 2 we show how to derive these theorems. Section 3 presents examples of theorems, including one (Theorem 2)

open map on t_P. Let V be an arbitrary open subset of t_P, and let V_α be a covering of t_P by open sets on each of which "exp" is an open map. For each α, $V \cap V_\alpha$ is open in V_α, and therefore $\exp(V \cap V_\alpha)$ is open in M. It follows that

$$\exp(V) = \bigcup_\alpha \exp(V \cap V_\alpha)$$

is open in M. Since V is arbitrary, "exp" is an open map on t_P. Define $S_P \equiv \exp(t_P) \subset M$.

Now let v be a tetrad at P. The tetrad v determines a natural diffeomorphism $\varphi_v: T_P \to R^4$, where, for each vector $\xi \in T_P$, $\varphi_v(\xi)$ is defined as the 4-tuple of components of ξ relative to v. We define the composition $\theta \equiv \exp \circ \varphi_v^{-1}$, a continuous map from a subset of R^4 into M.

Whenever the integer i is such that $q_i \in \theta^{-1}(S_P)$, we define

$$P_i = \theta(q_i) \in M,$$
$$U_i \equiv \theta(O_i \cap \theta^{-1}(S_P)) \subset M.$$

At each of the points P_i we define a tetrad v_i by parallel transport of the tetrad v along the geodesic[37] from P to P_i.

We now show that the U_i are a basis for the open sets of S_P. Since the O_i cover R^4, the U_i cover S_P. Since θ is an open map on $\theta^{-1}(S_P)$, each U_i is an open subset of M. Let U be any open subset of S_P. Then $\theta^{-1}(U)$ is open in R^4. Since the O_i are a basis for R^4, we may select a collection Γ of integers such that $\bigcup_{i\in\Gamma} O_i = \theta^{-1}(U)$. Therefore, $\bigcup_{i\in\Gamma} U_i = U$. Since U is arbitrary, the U_i are a basis for S_P.

To summarize, given any pair (P, v), where v is a tetrad at the point P, we have defined an open neighborhood S_P of P in M, a countable basis U_i for S_P, a countable dense set P_i in S_P, and a tetrad v_i at each P_i. We now repeat our construction for each pair (P_i, v_i). That is, we define an open neighborhood S_{P_i} of P_i, a basis U_{ij} for S_{P_i}, a dense set P_{ij} in S_{P_i}, and a tetrad v_{ij} at each P_{ij} (i fixed, $j = 1, 2, \cdots$). Applying the same construction to the pairs (P_{ij}, v_{ij}), and so on, we obtain finally a countable collection $U_\alpha \equiv \{U_i, U_{ij}, U_{ijk}, \cdots\}$ of open sets of M and a countable collection $P_\alpha \equiv \{P_i, P_{ij}, P_{ijk}, \cdots\}$ of points of M.

Since the U_α form a basis for each of the S_{P_β}, they also form a basis for $S \equiv \bigcup_\alpha S_{P_\alpha}$. We must finally show that $S = M$.

Let Q be any point in the closure of S. There is some neighborhood U of Q such that no geodesic segment in U has a pair of conjugate points in U. Since Q is in the closure of S, $U \cap S \neq \phi$. Therefore, since the P_α are dense in S, there must be some $P_\beta \in U$. But no geodesic segment has a pair of conjugate points in U, and so $S_{P_\beta} \supset U$. We have shown that $Q \in U \subset S_{P_\beta} \subset S$, i.e., that S contains each point of its closure. But S is also open in M, because it is the union of the open sets S_{P_α}. Since M is connected, it follows that $S = M$. This completes the proof.

[37] In case there are several geodesics from P to P_i, let us select, for definiteness, that one which has been used (in the exponential map) to define P_i.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

New Solutions of the Einstein-Maxwell Equations from Old*

B. KENT HARRISON†
Physics Department, Brigham Young University, Provo, Utah

(Received 14 December 1967)

Methods are discussed with which one may derive theorems which allow one to generate new solutions of the Einstein–Maxwell equations from old ones. The old solutions used to generate new ones must admit at least one nonnull Killing vector and may be required to satisfy other conditions, depending on the theorem derived. Examples of derivable theorems are shown; these theorems are used in turn to show how generation of new solutions is accomplished. Examples of the latter are shown, such as generation of Brill or electrified NUT space from the Schwarzschild solution, generation of a new twisted Melvin universe from flat space, and generation of a new generalization of the Ozsvath–Schücking metric. Possible physical interpretations, uses, and extensions of this type of theorem are discussed.

1. ASSUMED METRIC AND ITS EINSTEIN-MAXWELL EQUATIONS

We outline a method, suitable for use on a wide class of metrics, by which one can derive theorems which in turn can be used to obtain new solutions of the Einstein–Maxwell equations from old. In Sec. 1, we find the form of the Einstein–Maxwell equations for the assumed class of metrics; in Sec. 2 we show how to derive these theorems. Section 3 presents examples of theorems, including one (Theorem 2)

which is a generalization of and combination of previously derived theorems, one by Ehlers[1] and one by the present author,[2] Bonner,[3] Papapetrou,[4] Synge,[5] and Misra and Radhakrishna.[6] Section 4 gives examples of applications of these theorems to special metrics, and Sec. 5 discusses interpretations and uses.

We choose sign and other conventions as in Landau and Lifshitz.[7] The assumed form of the metric is as follows:

$$-ds^2 = \epsilon[e^{2U}(dx^k + af_\alpha\, dx^\alpha)^2 + a^2e^{-2U}\gamma_{\alpha\beta}\, dx^\alpha\, dx^\beta], \tag{1}$$

where

$$\epsilon = \pm 1 = \text{sgn}\,(g_{kk}); \tag{2}$$

a is an arbitrarily chosen constant scale factor; k is some particular one of 0, 1, 2, 3; Greek letters take all values of 0, 1, 2, 3, except k; and all metric coefficients are independent of x^k. Latin letters (except k) take on all values 0, 1, 2, 3. This metric thus admits a Killing vector $\xi^l = \delta^l_k$, but defines a congruence orthogonal to the hypersurfaces $x^k =$ const if $f_\alpha = 0$.

We now define $\gamma^{\alpha\beta}$ as the inverse of the 3-dimensional metric $\gamma_{\alpha\delta}$,

$$\gamma^{\alpha\beta}\gamma_{\alpha\delta} = \delta^\beta_\delta, \tag{3}$$

and $\sum^\alpha_{\beta\gamma}$ and $P_{\alpha\beta}$ as the Christoffel symbols and Ricci tensor, respectively, obtained from the $\gamma_{\alpha\beta}$. We also define

$$h_{\alpha\beta} = f_{\alpha,\beta} - f_{\beta,\alpha}, \tag{4}$$

where the comma denotes ordinary differentiation, and we define the differential parameters of first and second order,[8]

$$\Delta_1(F) = \gamma^{\alpha\beta}F_{,\alpha}F_{,\beta}, \tag{5}$$

$$\Delta_1(F, G) = \gamma^{\alpha\beta}F_{,\alpha}G_{,\beta}, \tag{6}$$

$$\Delta_2(F) = \gamma^{\alpha\beta}F_{;\alpha\beta} = \gamma^{\alpha\beta}(F_{,\alpha\beta} - \textstyle\sum^\gamma_{\alpha\beta} F_{,\gamma}), \tag{7}$$

where F and G are any functions of the coordinates. The semicolon always represents covariant differentiation with respect to the $\gamma_{\alpha\beta}$.

It is now a straightforward matter to calculate the components of the 4-dimensional Ricci tensor for metric (1). One obtains

$$R_{kk} = a^{-2}[-e^{4U}\Delta_2(U) + \tfrac{1}{4}e^{8U}\gamma^{\alpha\beta}\gamma^{\gamma\delta}h_{\alpha\gamma}h_{\beta\delta}], \tag{8}$$

$$R_{k\alpha} = af_\alpha R_{kk} + \tfrac{1}{2}a^{-1}e^{4U}\gamma^{\delta\gamma}(-h_{\alpha\gamma,\delta} + \textstyle\sum^\epsilon_{\delta\gamma} h_{\alpha\epsilon} + \sum^\epsilon_{\alpha\gamma} h_{\epsilon\delta} + 4h_{\delta\alpha}U_{,\gamma}), \tag{9}$$

$$R_{\alpha\beta} = af_\alpha R_{k\beta} + af_\beta R_{k\alpha} - a^2f_\alpha f_\beta R_{kk} + P_{\alpha\beta} - 2U_{,\alpha}U_{,\beta} + \gamma_{\alpha\beta}\Delta_2(U) - \tfrac{1}{2}e^{4U}\gamma^{\gamma\delta}h_{\alpha\gamma}h_{\beta\delta}. \tag{10}$$

We note from Eq. (4) that

$$-h_{\alpha\gamma,\delta} = h_{\delta\alpha,\gamma} + h_{\gamma\delta,\alpha}. \tag{11}$$

We thus may replace $-h_{\alpha\gamma,\delta}$ in Eq. (9) by $h_{\delta\alpha,\gamma}$ ($h_{\gamma\delta,\alpha}$ drops out because $h_{\gamma\delta}$ is antisymmetric and $\gamma^{\delta\gamma}$ is symmetric). Eq. (9) may then be written

$$R_{k\alpha} = af_\alpha R_{kk} + \tfrac{1}{2}a^{-1}e^{4U}\gamma^{\delta\gamma}(h_{\delta\alpha,\gamma} + 4h_{\delta\alpha}U_{,\gamma}). \tag{9'}$$

We choose units in the Einstein equations so that $G = c^4$. They thus may be written (for the vacuum or electromagnetic cases)

$$R_{ij} = 8\pi T_{ij}, \tag{12}$$

where T_{ij} is the energy–momentum tensor, since $T = 0$. T_{ij} is defined by

$$T_{ij} = (4\pi)^{-1}(F_{il}F_j{}^l - \tfrac{1}{4}F_{lm}F^{lm}g_{ij}). \tag{13}$$

Maxwell's equations are, for no sources,

$$(-g)^{-\frac{1}{2}}[(-g)^{\frac{1}{2}}F^{ij}]_{,j} = 0, \tag{14}$$

$$F_{ij,l} + F_{li,j} + F_{jl,i} = 0. \tag{15}$$

We assume that the F_{ij}, also, are independent of x^k. The four Eqs. (14) then yield

$$[(-g)^{\frac{1}{2}}F^{k\alpha}]_{,\alpha} = 0, \tag{16}$$

$$[(-g)^{\frac{1}{2}}F^{\beta\alpha}]_{,\alpha} = 0. \tag{17}$$

Equations (15) may be written

$$F_{\alpha k,\beta} + F_{k\beta,\alpha} = 0, \tag{18}$$

$$\epsilon^{\alpha\beta\gamma}F_{\alpha\beta,\gamma} = 0, \tag{19}$$

where $\epsilon^{\alpha\beta\gamma}$ is the alternating 3-index symbol ($\epsilon^{\alpha\beta\gamma} = +1$ if α, β, and γ are in natural order). Equations (17) and (18) may be satisfied by choosing potentials A and B:

$$F^{\alpha\beta} = (-g)^{-\frac{1}{2}}\epsilon^{\alpha\beta\gamma}A_{,\gamma}, \tag{20}$$

$$F_{k\alpha} = B_{,\alpha}. \tag{21}$$

Raising and lowering indices with the metric [Eq. (1)]

---

* Part of this work was performed during the tenure of NSF Grant No. GP-4902; another part was performed while the author was a consultant at Jet Propulsion Laboratory, Pasadena, California.

† Current (temporary) address: Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91103.

[1] J. Ehlers, "Konstruktionen and Charakterisierungen von Lösungen der Einsteinschen Gravitationsfeldgleichungen" (dissertation, Hamburg, 1957).

[2] B. Kent Harrison, Phys. Rev. **138**, B488 (1965); Report given at the London Conference on General Relativity, 1965.

[3] W. B. Bonner, Z. Physik **161**, 439 (1961).

[4] A. Papapetrou, Proc. Roy. Irish Acad. **A51**, 191 (1947).

[5] J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960), pp. 367–71.

[6] M. Misra and L. Radhakrishna, Proc. Nat. Inst. Sci. India **A28**, 632 (1962).

[7] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962), 2nd ed., Chaps. 10 and 11.

[8] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, N.J., 1949), p. 41.

now gives, after some algebra,

$$F^{k\alpha} = a^{-2}\gamma^{\alpha\beta}B_{,\beta} + a(-g)^{-\frac{1}{2}}\epsilon^{\alpha\beta\gamma}f_\beta A_{,\gamma}, \tag{22}$$

$$F_{\alpha\beta} = a(B_{,\beta}f_\alpha - B_{,\alpha}f_\beta) + a^4(-g)^{-\frac{1}{2}}e^{-4U}\epsilon^{\gamma\delta\epsilon}\gamma_{\alpha\gamma}\gamma_{\beta\delta}A_{,\epsilon}. \tag{23}$$

Substitution into Eqs. (16) and (19) yields the two remaining Maxwell equations:

$$\Delta_2(A) - 2\Delta_1(U,A) + \tfrac{1}{2}(-\gamma)^{-\frac{1}{2}}e^{2U}\epsilon^{\alpha\beta\gamma}h_{\alpha\beta}B_{,\gamma} = 0, \tag{24}$$

$$\Delta_2(B) - 2\Delta_1(U,B) - \tfrac{1}{2}(-\gamma)^{-\frac{1}{2}}e^{2U}\epsilon^{\alpha\beta\gamma}h_{\alpha\beta}A_{,\gamma} = 0, \tag{25}$$

where we have used Eq. (4) and

$$g = a^6\gamma e^{-4U}. \tag{26}$$

We can now obtain the components of T_{ij}. We get

$$T_{kk} = \epsilon(8\pi a^2)^{-1}e^{2U}[\Delta_1(A) + \Delta_1(B)], \tag{27}$$

$$T_{k\alpha} = af_\alpha T_{kk} + \epsilon(4\pi a)^{-1}(-\gamma)^{-\frac{1}{2}}\epsilon^{\beta\gamma\delta}\gamma_{\alpha\beta}A_{,\delta}B_{,\gamma}, \tag{28}$$

$$T_{\alpha\beta} = af_\alpha T_{k\beta} + af_\beta T_{k\alpha} - a^2f_\alpha f_\beta T_{kk} + \epsilon(8\pi)^{-1}e^{-2U}\{2(A_{,\alpha}A_{,\beta} + B_{,\alpha}B_{,\beta}) - \gamma_{\alpha\beta}[\Delta_1(A) + \Delta_1(B)]\}. \tag{29}$$

Equations (12) now become, with Eqs. (8)–(10) and (27)–(29) and judicious algebraic combination,

$$\Delta_2(U) - \tfrac{1}{4}e^{4U}\gamma^{\alpha\beta}\gamma^{\gamma\delta}h_{\alpha\gamma}h_{\beta\delta} = -\epsilon e^{-2U}[\Delta_1(A) + \Delta_1(B)], \tag{30}$$

$$\gamma^{\delta\gamma}(h_{\delta\alpha;\gamma} + 4h_{\delta\alpha}U_{,\gamma}) = 4\epsilon(-\gamma)^{-\frac{1}{2}}e^{-4U}\epsilon^{\beta\gamma\delta}\gamma_{\alpha\beta}A_{,\delta}B_{,\gamma}, \tag{31}$$

$$P_{\alpha\beta} - 2U_{,\alpha}U_{,\beta} + \tfrac{1}{4}e^{4U}\gamma^{\gamma\delta}(\gamma_{\alpha\beta}\gamma^{\epsilon\eta}h_{\gamma\epsilon}h_{\delta\eta} - 2h_{\alpha\gamma}h_{\beta\delta}) = 2\epsilon e^{-2U}(A_{,\alpha}A_{,\beta} + B_{,\alpha}B_{,\beta}). \tag{32}$$

The equations we wish to solve are now Eqs. (4), (24), (25), (30), (31), and (32).

The quantities $h_{\alpha\beta}$ form a 3-dimensional antisymmetric tensor and can thus be expressed in terms of an axial vector. We write

$$h_{\alpha\beta} = f_{\alpha,\beta} - f_{\beta,\alpha} = \epsilon_{\alpha\beta\gamma}\gamma^{\gamma\delta}z_\delta(-\gamma)^{\frac{1}{2}}. \tag{33}$$

The integrability condition [Eq. (11)] becomes

$$\gamma^{\alpha\beta}z_{\alpha;\beta} = \gamma^{\alpha\beta}(z_{\alpha,\beta} - z_\gamma\textstyle\sum_{\alpha\beta}^{\gamma}) = 0. \tag{34}$$

Equations (24) and (25) become

$$\Delta_2(A) - 2\Delta_1(U,A) + e^{2U}\gamma^{\alpha\beta}B_{,\alpha}z_\beta = 0, \tag{35}$$

$$\Delta_2(B) - 2\Delta_1(U,B) - e^{2U}\gamma^{\alpha\beta}A_{,\alpha}z_\beta = 0, \tag{36}$$

and Eqs. (30)–(32) become

$$\Delta_2(U) + \tfrac{1}{2}e^{4U}\gamma^{\alpha\beta}z_\alpha z_\beta = -\epsilon e^{-2U}[\Delta_1(A) + \Delta_1(B)], \tag{37}$$

$$\epsilon^{\gamma\alpha\beta}(z_{\alpha,\beta} + 4z_\alpha U_{,\beta} + 4\epsilon e^{-4U}B_{,\alpha}A_{,\beta}) = 0, \tag{38}$$

$$P_{\alpha\beta} - 2U_{,\alpha}U_{,\beta} - \tfrac{1}{2}e^{4U}z_\alpha z_\beta = 2\epsilon e^{-2U}(A_{,\alpha}A_{,\beta} + B_{,\alpha}B_{,\beta}). \tag{39}$$

We now note that Eq. (38) may be satisfied identically by choosing a "twist" potential ϕ in the following way:

$$z_\alpha = e^{-4U}[\phi_{,\alpha} + 2\epsilon(BA_{,\alpha} - AB_{,\alpha})]. \tag{40}$$

Eqs. (34), (35), (36), (38), (39) now become, respectively,

$$\Delta_2(\phi) - 4\Delta_1(U,\phi) + 2\epsilon B[\Delta_2(A) - 4\Delta_1(U,A)] - 2\epsilon A[\Delta_2(B) - 4\Delta_1(U,B)] = 0, \tag{41}$$

$$\Delta_2(A) - 2\Delta_1(U,A) + e^{-2U}[\Delta_1(\phi,B) + 2\epsilon B\Delta_1(A,B) - 2\epsilon A\Delta_1(B)] = 0, \tag{42}$$

$$\Delta_2(B) - 2\Delta_1(U,B) - e^{-2U}[\Delta_1(\phi,A) + 2\epsilon B\Delta_1(A) - 2\epsilon A\Delta_1(A,B)] = 0, \tag{43}$$

$$\Delta_2(U) + \epsilon e^{-2U}[\Delta_1(A) + \Delta_1(B)] + \tfrac{1}{2}e^{-4U}[\Delta_1(\phi) + 4\epsilon B\Delta_1(\phi,A) - 4\epsilon A\Delta_1(\phi,B) + 4B^2\Delta_1(A) - 8AB\Delta_1(A,B) + 4A^2\Delta_1(B)] = 0, \tag{44}$$

$$P_{\alpha\beta} = 2U_{,\alpha}U_{,\beta} + 2\epsilon e^{-2U}(A_{,\alpha}A_{,\beta} + B_{,\alpha}B_{,\beta}) + \tfrac{1}{2}e^{-4U}[\phi_{,\alpha} + 2\epsilon(BA_{,\alpha} - AB_{,\alpha})] \times [\phi_{,\beta} + 2\epsilon(BA_{,\beta} - AB_{,\beta})]. \tag{45}$$

The symmetrical occurrence of A and B suggests the introduction of "polar potentials" R and θ:

$$A = R\cos\theta, \tag{46}$$

$$B = R\sin\theta. \tag{47}$$

Equations (40)–(45) now become, after suitable combination of Eq. (41), (42), and (43):

$$z_\alpha = e^{-4U}(\phi_{,\alpha} - 2\epsilon R^2\theta_{,\alpha}), \tag{48}$$

$$\Delta_2(\phi) - 4\Delta_1(U,\phi) + 4\epsilon R^2\Delta_1(U,\theta) + 2Re^{-2U}[2R^2\Delta_1(R,\theta) - \epsilon\Delta_1(\phi,R)] = 0, \tag{49}$$

$$\Delta_2(R) - R\Delta_1(\theta) - 2\Delta_1(U,R) + Re^{-2U}[\Delta_1(\phi,\theta) - 2\epsilon R^2\Delta_1(\theta)] = 0, \tag{50}$$

$$R\Delta_2(\theta) + 2\Delta_1(R,\theta) - 2R\Delta_1(U,\theta) - e^{-2U}[\Delta_1(\phi,R) - 2\epsilon R^2\Delta_1(R,\theta)] = 0, \tag{51}$$

$$\Delta_2(U) + \epsilon e^{-2U}[\Delta_1(R) + R^2\Delta_1(\theta)] + \tfrac{1}{2}e^{-4U}[\Delta_1(\phi) - 4\epsilon R^2\Delta_1(\phi,\theta) + 4R^4\Delta_1(\theta)] = 0, \tag{52}$$

$$P_{\alpha\beta} = 2U_{,\alpha}U_{,\beta} + 2\epsilon e^{-2U}(R_{,\alpha}R_{,\beta} + R^2\theta_{,\alpha}\theta_{,\beta}) + \tfrac{1}{2}e^{-4U}(\phi_{,\alpha} - 2\epsilon R^2\theta_{,\alpha})(\phi_{,\beta} - 2\epsilon R^2\theta_{,\beta}). \tag{53}$$

If now we have a solution of Eqs. (49)–(53) for the $\gamma_{\alpha\beta}$ and the four functions U, ϕ, R, θ, then Eq. (48) gives z_α and Eq. (33) can then be solved for the f_α.

It will be noted that since Eq. (33) is an equation for curl **f**, there will exist an arbitrary function of integration ξ, with $\mathbf{f} = \nabla\xi +$ other terms. From the form of the metric (1) we see immediately that ξ can be absorbed into x^k by redefinition of x^k, so that we will usually drop ξ. In certain situations it may be useful to use ξ to change the form of the remaining terms in $f_\alpha\, dx^\alpha$.

2. METHOD OF DERIVATION OF THEOREMS FOR GENERATING NEW SOLUTIONS OF THE EINSTEIN-MAXWELL EQUATIONS FROM OLD

We assume first that a solution of Eqs. (49)–(53) is already known. The majority of known solutions of the Einstein–Maxwell equations fall into the class of metrics of form (1) and hence provide known solutions of the equations.

After a known solution has been selected, we identify the quantities $\gamma_{\alpha\beta}$, U, f_α, ϵ, and a (usually set = 1) by inspection. The electromagnetic potentials A and B may be identified by integrating a subset of the Eqs. (20)–(23). z_δ may be found from Eq. (33), and then ϕ may be found from Eq. (40). Equations (46) and (47) give R and θ if desired.

One sees that vacuum solutions yield $A = B = 0$, or $R = 0$; solutions admitting a hypersurface-orthogonal congruence have $z_\delta = 0$; solutions with constant g_{kk} have $U =$ const. Thus in many cases, one or more of the function U, ϕ, R, and θ are constant. We assume that at least one of these four functions is not constant.

To obtain new solutions of the Einstein–Maxwell equations, we assume the metric to have the form (1), with new U, a, and f_α—denoted by a bar—but with the same $\gamma_{\alpha\beta}$:

$$-d\bar{s}^2 = \epsilon[e^{2\bar{U}}(dx^k + \bar{a}\bar{f}^k_\alpha\, dx^\alpha)^2 + \bar{a}^2 e^{-2\bar{U}}\gamma_{\alpha\beta}\, dx^\alpha\, dx^\beta]. \tag{54}$$

We also assume new electromagnetic potentials $\bar{A}$ and $\bar{B}$—or $\bar{R}$ and $\bar{\theta}$. The $\bar{f}_\alpha$ are related to a new twist potential $\bar{\phi}$ by Eqs. (33) and (40) with bars. We now assume that $\bar{U}$, $\bar{\phi}$, $\bar{R}$, and $\bar{\theta}$ are functions of those among U, ϕ, R, and θ which are not constant and of no other quantities.

It is easily shown that, with this assumption, the differential parameters of the barred quantities are linear functions of the differential parameters of the unbarred ones. For example, if $\bar{U} = \bar{U}(U, \phi)$, then

$$\Delta_1(\bar{U}) = \bar{U}_U^2\Delta_1(U) + 2\bar{U}_U\bar{U}_\phi\Delta_1(U, \phi) + \bar{U}_\phi^2\Delta_1(\phi),$$

where subscripts denote differentiations. We now write Eqs. (49)–(52), barred, and use the assumed functional dependence to express them in terms of the differential parameters of the unbarred functions; substitute for the Δ_2's from the original Eqs. (49)–(52); then equate the coefficients of each of the Δ_1's in each equation to zero. We do this last step because we wish to avoid restricting the original solution by imposing restrictions on U, ϕ, R, and θ. In some cases, the Δ_1's of the original solution may satisfy relations among themselves; if this happens, these relations can be imposed on the Eqs. (49)–(52) before setting coefficients equal to zero, provided that the relations involve only functions of the U, ϕ, R, and θ.

When the coefficients are set equal to zero, we obtain a set of differential equations for the $\bar{U}$, etc., as functions of the U, etc. We obtain further equations by treating Eq. (53) in a similar manner. Since the $\gamma_{\alpha\beta}$ are the same in both metrics, the $P_{\alpha\beta}$ are also, and one may set the right-hand sides of the old and new Eqs. (53) equal. One then substitutes $\bar{U} = \bar{U}(U \cdots)$, etc., into one side and equates coefficients of the $U_{,\alpha}U_{,\beta}$, etc., terms.

There are four barred functions. If the number of nonconstant unbarred functions is n, one sees easily that, in general, there are $\frac{5}{2}n(n+1)$ differential equations for the four dependent variables as functions of the n independent variables. There always exists one solution—the identity. Whether or not there exist solutions which are not equivalent to the identity is not at present known in general, although such do exist in many situations. We expect some redundancy in these partial differential equations, by the nature of the general-relativity equations themselves.

If solutions are obtained which are not equivalent to the identity, then these functions $\bar{U}$, etc., provide a new solution of the Einstein–Maxwell equations. It is, of course, desirable to include in it all arbitrary constants in order to provide greatest generality.

One can approach this treatment from a slightly different direction. In this, one assumes that certain of the unbarred functions are nonconstant—without specifying the particular metric satisfying this condition—and then finds the $\bar{U}$, etc. Then, if a metric is provided which satisfies the given condition on the unbarred functions, one already has the form of the $\bar{U}$, etc., and can immediately write down the new metric. This approach is clearly a powerful way of generating *theorems* describing how to generate new *solutions* from old. There are thus 15 possible theorems (15 equals the total number of combinations of four functions in all groupings) although not all may produce new inequivalent solutions.

Another variation in approach is obtained by assuming, for the original metric, that some of the U, ϕ, R, and θ are functions of other nonconstant

ones. This clearly places a restriction on the original metric; but if it is satisfied, then the number of differential equations for the $\bar{U}$, etc., is reduced. This approach is probably equivalent to assuming that there exist relations between the Δ_1's of the original solution (possibility mentioned above). There are a large number of possible theorems derivable in this manner.

3. EXAMPLES OF THEOREMS

A simple beginning example is provided by considering a vacuum space with $g^{kk} = \epsilon$. In this case we have $U = 0$, $R = 0$. Rewrite ϕ as ψ. Examination of Eqs. (49)–(53) soon shows that

$$\Delta_2(\psi) = 0, \tag{55}$$

$$\Delta_1(\psi) = 0, \tag{56}$$

$$P_{\alpha\beta} = \tfrac{1}{2}\psi_{,\alpha}\psi_{,\beta}. \tag{57}$$

We find a new metric by writing $\bar{U}$, $\bar{\phi}$, $\bar{R}$, and $\bar{\theta}$ as functions of ψ. For simplicity we now drop the bars. Since $\Delta_1(F)$ and $\Delta_2(F)$ (where F is any of U, ϕ, R, and θ) are linear functions of $\Delta_1(\psi)$ and $\Delta_2(\psi)$, we see immediately by Eqs. (55) and (56) that the $\Delta_1(F)$ and $\Delta_2(F)$ are all zero, and that Eqs. (49)–(52) are satisfied identically. When we equate the new Eq. (53) to the old, and divide by $\psi_{,\alpha}\psi_{,\beta}$, we get

$$\tfrac{1}{2} = 2U'^2 + 2\epsilon e^{-2U}(R'^2 + R^2\theta'^2) + \tfrac{1}{2}e^{-4U}(\phi' - 2\epsilon R^2\theta')^2, \tag{58}$$

where a prime denotes $d/d\psi$. The identity transformation is the case $U = 0$, $R = 0$, and $\phi = \psi$.

Solving for ϕ' gives

$$\phi' = 2\epsilon R^2\theta' \pm e^{2U}[1 - 4U'^2 - 2\epsilon e^{-2U}(R'^2 + R^2\theta'^2)]^{\frac{1}{2}}. \tag{59}$$

Thus, using the forms of the metrics (1) and (54) and Eqs. (33) and (48), we can write the following theorem:

Theorem 1: For every vacuum metric which is a solution of the Einstein equations, and of the form (with $\partial/\partial x^k = 0$)

$$-ds^2 = \epsilon[(dx^k + f_\alpha\, dx^\alpha)^2 + \gamma_{\alpha\beta}\, dx^\alpha\, dx^\beta], \tag{60}$$

we define ψ by the equations

$$f_{\alpha,\beta} - f_{\beta,\alpha} = \epsilon_{\alpha\beta\gamma}\gamma^{\gamma\delta}\psi_{,\delta}(-\gamma)^{\frac{1}{2}} \tag{61}$$

(we assume ψ nonconstant). Then there exists a generalized solution, including electromagnetic fields, of the form

$$-ds^2 = \epsilon[e^{2U}(dx^k + af^k_\alpha\, dx^\alpha)^2 + a^2e^{-2U}\gamma_{\alpha\beta}\, dx^\alpha\, dx^\beta], \tag{62}$$

and with electromagnetic potentials R, θ such that U, R, θ are arbitrary functions of ψ, and with

$$\bar{f}_{\alpha,\beta} - \bar{f}_{\beta,\alpha} = \epsilon_{\alpha\beta\gamma}\gamma^{\gamma\delta}\psi_{,\delta}(-\gamma)^{\frac{1}{2}}G(\psi) \tag{63a}$$

$$= G(\psi)(f_{\alpha,\beta} - f_{\beta,\alpha}), \tag{63b}$$

where

$$G(\psi) = \pm e^{-2U}[1 - 4U'^2 - 2\epsilon e^{-2U}(R'^2 + R^2\theta'^2)]^{\frac{1}{2}}. \tag{64}$$

The expression $R'^2 + R^2\theta'^2$ may be replaced by $A'^2 + B'^2$.

For our second theorem, we work with the "rectangular" potentials A and B. We first assume that ϕ, A, and B are functions of U; this constitutes a restriction on the original metric chosen. (These restrictions are satisfied automatically for a vacuum metric admitting trajectories orthogonal to hypersurfaces $x^k = \text{const.}$)

We first find the form of the equations under this restriction. Equation (40) gives

$$z_\alpha = e^{-4U}[\phi' + 2\epsilon(BA' - AB')]U_{,\alpha},$$

which we write for simplicity as

$$z_\alpha = H(U)U_{,\alpha}, \tag{65}$$

and we have

$$A = A(U), \tag{66a}$$

$$B = B(U). \tag{66b}$$

It is simplest to work with Eqs. (34)–(39); they become

$$H\Delta_2(U) + H'\Delta_1(U) = 0, \tag{67}$$

$$A'\Delta_2(U) + (A'' - 2A' + HB'e^{2U})\Delta_1(U) = 0, \tag{68}$$

$$B'\Delta_2(U) + (B'' - 2B' - HA'e^{2U})\Delta_1(U) = 0, \tag{69}$$

$$\Delta_2(U) + [\tfrac{1}{2}e^{4U}H^2 + \epsilon e^{-2U}(A'^2 + B'^2)]\Delta_1(U) = 0, \tag{70}$$

$$P_{\alpha\beta} = [2 + \tfrac{1}{2}H^2e^{4U} + 2\epsilon e^{-2U}(A'^2 + B'^2)]U_{,\alpha}U_{,\beta}, \tag{71}$$

where a prime denotes d/dU. We use Eq. (70) to eliminate $\Delta_2(U)$ and assume $\Delta_1(U) \neq 0$; Eqs. (67)–(69) become

$$H' = GH, \tag{72}$$

$$A'' - (G + 2)A' + HB'e^{2U} = 0, \tag{73}$$

$$B'' - (G + 2)B' - HA'e^{2U} = 0, \tag{74}$$

where

$$G = \tfrac{1}{2}e^{4U}H^2 + \epsilon e^{-2U}(A'^2 + B'^2). \tag{75}$$

If we put

$$H = Le^{-2U}, \tag{76}$$

$$A' = Qe^{U}\cos\tau, \tag{77}$$

$$B' = Qe^{U}\sin\tau, \tag{78}$$

we find

$$Q(\tau' - L) = 0, \tag{79}$$

$$L' = L(G + 2), \tag{80}$$

$$Q' = Q(G + 1), \tag{81}$$

$$G = \tfrac{1}{2}L^2 + \epsilon Q^2. \tag{82}$$

Equation (71) becomes

$$P_{\alpha\beta} = (2 + \tfrac{1}{2}L^2 + 2\epsilon Q^2)U_{,\alpha}U_{,\beta}. \tag{83}$$

Solution of Eqs. (79)–(82) gives

$$Q^2 = \epsilon\lambda e^{2U}(1 - \lambda e^{2U} - \mu e^{4U})^{-1}, \tag{84}$$

$$L^2 = 4\mu e^{4U}(1 - \lambda e^{2U} - \mu e^{4U})^{-1}, \tag{85}$$

$$\tau = \int L\,dU + \nu \quad \text{(vacuous if } \lambda = 0\text{)}, \tag{86}$$

where λ, μ, and ν are arbitrary constants. If we consider separately the cases with $\lambda = 0$ and $\lambda \neq 0$, and those with $\mu = 0$ and $\mu \neq 0$, and if we perform the integrations indicated in Eqs. (77), (78), and (86), we find a general form of the remaining equations as follows [including Eq. (33)]:

$$f_{\alpha,\beta} - f_{\beta,\alpha} = 2\omega\epsilon_{\alpha\beta\gamma}\gamma^{\gamma\delta}[\mu(1 - \lambda e^{2U} - \mu e^{4U})^{-1}]^{\frac{1}{2}}(-\gamma)^{\frac{1}{2}}U_{,\delta}, \tag{87}$$

$$P_{\alpha\beta} = 2(1 - \lambda e^{2U} - \mu e^{4U})^{-1}U_{,\alpha}U_{,\beta}, \tag{88}$$

$$\Delta_2(U) = -(\lambda e^{2U} + 2\mu e^{4U})(1 - \lambda e^{2U} - \mu e^{4U})^{-1}\Delta_1(U), \tag{89}$$

$$A = (\xi\mu)^{\frac{1}{2}}e^{2U}\cos\alpha + \delta[\xi(1 - \lambda e^{2U} - \mu e^{4U})]^{\frac{1}{2}}\sin\alpha + A_0, \tag{90}$$

$$B = (\xi\mu)^{\frac{1}{2}}e^{2U}\sin\alpha - \delta[\xi(1 - \lambda e^{2U} - \mu e^{4U})]^{\frac{1}{2}}\cos\alpha + B_0, \tag{91}$$

where α, A_0, B_0 are constants, $\omega = \pm 1$, $\delta = \pm 1$, and $\xi = \epsilon\lambda(\lambda^2 + 4\mu)^{-1}$. If both λ and $\mu = 0$, A and B are to be taken as constants. Constants A_0 and B_0 are trivial, since they do not appear in the electromagnetic fields; the constant α signifies a duality rotation. It will be noted that Eqs. (84) and (85) require

$$\epsilon\lambda\mu \geq 0 \tag{92}$$

and

$$4\mu + \lambda^2 \geq (2\mu e^{2U} + \lambda)^2 \geq 0. \tag{93}$$

These conditions insure that the quantities under the square roots above will be positive or zero.

We now assume a second solution of the same form as the first and denote all quantities by bars, as before. We now assume $\bar{U} = \bar{U}(U)$. Equation (88), old and new, yields

$$P_{\alpha\beta} = 2(1 - \lambda e^{2U} - \mu e^{4U})^{-1}U_{,\alpha}U_{,\beta} = 2(1 - \bar{\lambda}e^{2\bar{U}} - \bar{\mu}e^{4\bar{U}})^{-1}\bar{U}_{,\alpha}\bar{U}_{,\beta},$$

which gives, since $\bar{U} = \bar{U}(U)$,

$$(d\bar{U}/dU)^2 = (1 - \bar{\lambda}e^{2\bar{U}} - \bar{\mu}e^{4\bar{U}})(1 - \lambda e^{2U} - \mu e^{4U})^{-1}. \tag{94}$$

Also

$$\Delta_1(\bar{U}) = \bar{U}'^2\Delta_1(U)$$

and

$$\Delta_2(\bar{U}) = \bar{U}''\Delta_1(U) + \bar{U}'\Delta_2(U),$$

giving

$$-(\bar{\lambda}e^{2\bar{U}} + 2\bar{\mu}e^{4\bar{U}})(1 - \bar{\lambda}e^{2\bar{U}} - \bar{\mu}e^{4\bar{U}})^{-1}\bar{U}'^2\Delta_1(U) = \bar{U}''\Delta_1(U) - \bar{U}'(\lambda e^{2U} + 2\mu e^{4U}) \times (1 - \lambda e^{2U} - \mu e^{4U})^{-1}\Delta_1(U). \tag{95}$$

We divide by $\Delta_1(U)$ and obtain a second-order differential equation for $\bar{U}(U)$. Equation (94) is a first integral of Eq. (95), so we consider only Eq. (94).

To integrate Eq. (94), we write

$$X = e^{-2U} - \tfrac{1}{2}\lambda, \tag{96a}$$

$$\bar{X} = e^{-2\bar{U}} - \tfrac{1}{2}\bar{\lambda}, \tag{96b}$$

$$\beta^2 = \mu + \tfrac{1}{4}\lambda^2, \tag{97a}$$

$$\bar{\beta}^2 = \bar{\mu} + \tfrac{1}{4}\bar{\lambda}^2. \tag{97b}$$

Equation (94) then becomes

$$(d\bar{X}/dX)^2 = (\bar{X}^2 - \bar{\beta}^2)(X^2 - \beta^2)^{-1}. \tag{98}$$

The solutions of Eq. (98) fall into a number of cases, depending on the values of β and $\bar{\beta}$. We note, by Eq. (93), that if $\beta = 0$, $2\mu e^{2U} + \lambda = 0$, yielding $\mu = \lambda = 0$.

If $\mu = 0$, we have a metric with a congruence orthogonal to the hypersurfaces $x^k = \text{const}$; such metrics will be denoted by the letter O. If $\lambda = 0$, then $\xi = 0$, and A and B reduce to constants, so that we have a vacuum space; we denote these spaces by V. If $\lambda \neq 0$, we have electromagnetic fields present—denoted by E. Thus we may classify solutions as in Table I.

TABLE I. Classification of solutions of the metric.

Type	OV	OE	V	E
λ	$= 0$	$\neq 0$	$= 0$	$\neq 0$
μ	$= 0$	$= 0$	$\neq 0$	$\neq 0$

TABLE II. Solutions of Equation (98), listing the conditions which apply, and the classifications of the old and new solutions.

Case	Conditions	$\bar{X}(X)$	Old solution	New solution
1	$\beta = \bar{\beta} = 0$	$\bar{X} = kX$	OV	OV
2	$\beta = \bar{\beta} = 0$	$\bar{X} = kX^{-1}$	OV	OV
3	$\beta \neq 0, \bar{\beta} = 0$	$\bar{X} = k\beta^{-1}(X \pm \sqrt{X^2 - \beta^2})$	OE, V, E	OV
4	$\beta = 0, \bar{\beta} \neq 0$	$\bar{X} = \bar{\beta}(2kX)^{-1}(X^2 + k^2)$	OV	OE, V, E
5	$\beta\bar{\beta} \neq 0, X^2 > \beta^2$	$\bar{X} = \bar{\beta}\beta^{-1}(X\cosh k + \sqrt{X^2 - \beta^2}\sinh k)$	OE, V, E	OE, V, E
6	$\beta\bar{\beta} \neq 0, X^2 < \beta^2$	$X = \bar{\beta}\beta^{-1}(X\cos k + \sqrt{\beta^2 - X^2}\sin k)$	OE, V, E	OE, V, E

We now (Table II) list the various kinds of solutions of Eq. (98), listing the conditions which apply and the classifications of the old and new solutions. In all cases, k is the constant of integration.

Case 1 is trivial; Case 2 very nearly so. Case 4, with $\bar{\lambda} = 0$, is the case treated by Ehlers[1]; Case 4, with $\bar{\mu} = 0$, is the one treated by the author and others.[2-6]

As noted before, OV metrics satisfy restrictions (65) and (66) automatically and thus can be used via Cases (2) and (4) above to generate new metrics. Metrics which are OE, V, or E need to be checked via Eqs. (65), (66) to see if they can be used to generate new metrics.

Thus Theorem 2 can be stated as follows: If we have a metric of form (1), satisfying Eqs. (65) and (66), then we can write down new metrics of the same form by using one or more of the cases of Table II. As a practical procedure, we identify ϵ, U, f_α, $\gamma_{\alpha\beta}$, A, B from the form (1) and the known electromagnetic fields [Eqs. (20)–(23)]. We then find z_δ from Eq. (33) and check Eqs. (65) and (66). If they are satisfied, we identify ω, δ, μ, and λ from Eqs. (87), (90), and (91). Equations (96a) and (97a) give X and β; then Table II can guide us to a selection of cases and functions $\bar{X}$. We then reverse the procedure, using Eqs. (96b), (87), (90), and (91) to get $U, \bar{f}_\alpha, \bar{A}$, and $\bar{B}$. Equations (65) and (66) are now automatically satisfied. k, $\bar{\lambda}$, $\bar{\mu}$ are to be considered as arbitrary constants of integration.

4. APPLICATIONS OF THEOREMS

Three examples of the uses of these theorems are presented here. They are not discussed in detail since that is beyond the scope of this paper.

Example 1: We generalize the Ozsvath–Schücking metric[9] by using Theorem 1. It is given by

$$-ds^2 = (dx^1)^2 - 4v\,dx^1\,dx^3 + 2\,dx^2\,dx^3 + 2v^2(dx^3)^2 + (dv)^2, \quad (99)$$

where $v = x^4$. We choose $x^k = x^1$ and then identify the following quantities: $\epsilon = 1$, $a = 1$, $U = 0$, $A = B = 0$, $f_0 = f_2 = 0$, $f_3 = -2v$, $\gamma_{00} = 1$, $\gamma_{23} = 1$, $\gamma_{33} = -2v^2$, $\gamma_{22} = \gamma_{02} = \gamma_{03} = 0$, $g = \gamma = -1$. Equation (33) gives $z_0 = z_2 = 0$, $z_3 = -2$, so that Eq. (40) (with $\phi \to \psi$) gives

$$\psi = -2x^3. \quad (100)$$

Integration of Eq. (63), dropping of arbitrary functions of integration, yields

$$\bar{f}_0 = \bar{f}_2 = 0, \quad \bar{f}_3 = -2vG(\psi). \quad (101)$$

We use Eq. (100) to express everything in terms of x^3. By Eq. (62) we get, finally, that

$$-ds^2 = e^{2U}[dx^1 - 2avF(x^3)\,dx^3]^2 + a^2e^{-2U}[2\,dx^2\,dx^3 - 2v^2(dx^3)^2 + (dv)^2] \quad (102)$$

is a solution of the Einstein–Maxwell equations, where a is an arbitrary constant, U, A, and B are arbitrary functions of x^3,

$$F(x^3) = e^{-2U}[1 - U_{,3}^2 - \tfrac{1}{2}e^{-2U}(A_{,3}^2 + B_{,3}^2)]^{\frac{1}{2}}, \quad (103)$$

and the subscript 3 indicates d/dx^3. A and B occur as electromagnetic potentials and yield the following electromagnetic fields:

$$F^{02} = a^{-3}e^{2U}A_{,3}, \quad F_{13} = B_{,3}, \quad F^{12} = a^{-2}B_{,3}, \quad F_{03} = ae^{-2U}A_{,3}. \quad (104)$$

To obtain the locally observable electromagnetic fields **E** and **H**, we write the metric in terms of differential 1 forms:

$$-ds^2 = (\omega^1)^2 + (\omega^2)^2 + (\omega^3)^2 - (\omega^0)^2, \quad (105)$$

where

$$\omega^0 = ae^{-U}[\sqrt{2}\,v\,dx^3 - (\sqrt{2}\,v)^{-1}\,dx^2], \quad (106a)$$
$$\omega^1 = e^U(dx^1 - 2avF\,dx^3), \quad (106b)$$
$$\omega^2 = ae^{-U}(\sqrt{2}\,v)^{-1}\,dx^2, \quad (106c)$$
$$\omega^3 = ae^{-U}\,dv, \quad (106d)$$

and write the electromagnetic field 2-form

$$\phi = \tfrac{1}{2}F_{ij}\,dx^i \wedge dx^j = (E_1\omega^1 + E_2\omega^2 + E_3\omega^3) \wedge \omega^0 + H_1\omega^2 \wedge \omega^3 + H_2\omega^3 \wedge \omega^1 + H_3\omega^1 \wedge \omega^2. \quad (107)$$

[9] I. Ozsvath and E. Schücking, *Recent Developments in General Relativity* (Pergamon Press, New York, 1962), p. 339.

We obtain, by using Eqs. (104), (106), and (107),

$$E_3 = -H_1 = (\sqrt{2}\, va)^{-1} A_{,3}, \tag{108a}$$

$$E_1 = H_3 = (\sqrt{2}\, va)^{-1} B_{,3}, \tag{108b}$$

$$E_2 = H_2 = 0. \tag{108c}$$

This is clearly a plane wave, since $\mathbf{E} \cdot \mathbf{H} = 0$ and $\mathbf{E}^2 - \mathbf{H}^2 = 0$ everywhere. Its direction of propagation is in the x^2 direction. This result is satisfying, since the original Oszvath–Schücking metric (and also, apparently, this new one) represents a gravitational plane wave traveling in the x^2 direction.

The above comments are somewhat faulty, since they are based on a tetrad [components of Eqs. (106)] which is singular at $v = x^4 = 0$. One can define a nonsingular tetrad for the above metric, in the manner of Oszvath–Schücking,[9] but its form is quite complicated and it will not be given here. One does find that nonsingularity is preserved only if $(Fe^{2V})^2 > \frac{1}{2}$, which yields

$$1 > 2U_{,3}^2 + e^{-2U}(A_{,3}^2 + B_{,3}^2). \tag{109}$$

This expression bears some resemblance to terms in the energy–momentum tensor and tempts one to surmise that if the energy density gets too large, a singularity develops and gravitational collapse takes place, but this point remains to be investigated elsewhere, along with more detailed studies of the metric (102).

Example 2: We begin with a flat metric in cylindrical coordinates

$$-ds^2 = -dT^2 + dr^2 + r^2\, d\theta^2 + dz^2 \tag{110}$$

and use Theorem 2, Case 4. We choose $x^k = \theta\ (= x^2)$ and note that $\epsilon = 1$, $a = 1$, $f_\alpha = 0$, $U = \ln r$, $-\gamma_{00} = \gamma_{11} = \gamma_{33} = r^2$, $\gamma_{01} = \gamma_{03} = \gamma_{13} = 0$, $\gamma = -r^6$. Thus $\lambda = \mu = 0$, $\beta = 0$. If we put $\bar{\beta} = 2kl^2$, $\bar{\lambda} = 4ckl^2$, we find

$$e^{-2\bar{U}} = l^2 r^{-2}(k^2 r^4 + 2ckr^2 + 1), \tag{111}$$

$$\bar{f}_1 = \bar{f}_3 = 0, \quad \bar{f}_0 = 4ke^2\sqrt{1 - c^2}\, z, \tag{112}$$

and

$$C = l^{-1}\sqrt{kc(1 - c^2)}\, r^2(k^2 r^4 + 2ckr^2 + 1)^{-1}, \tag{113}$$

$$D = \pm(2l)^{-1}\sqrt{ck^{-1}}\,(1 - k^2 r^4)(k^2 r^4 + 2ckr^2 + 1)^{-1}, \tag{114}$$

where

$$\bar{A} = C \cos\alpha + D \sin\alpha, \tag{115a}$$

$$\bar{B} = C \sin\alpha - D \cos\alpha \tag{115b}$$

(α represents a duality rotation).

The locally observed electromagnetic fields are both in the z direction and are given by

$$E = \zeta \cos\alpha + \gamma \sin\alpha, \tag{116a}$$

$$H = -\zeta \sin\alpha + \gamma \cos\alpha, \tag{116b}$$

with

$$\zeta = (ar)^{-1}\, dC/dr, \tag{117}$$

$$\gamma = (ar)^{-1}\, dD/dr. \tag{118}$$

Inspection shows that this is a "twisted" Melvin magnetic universe. The twist is given by $\bar{f}_0$, which plays the role of an angular velocity in a rotating coordinate system. Since $\bar{f}_0$ is proportional to z, we see that this angular velocity increases as z increases and is of opposite signs on the two sides of $z = 0$. One envisions a universe twisting upon itself (or untwisting!) as time goes on.

The case $c = \pm 1$ yields the usual Melvin universe,[10] with no twist. The case $c = 0$ gives maximum twist, but no electromagnetic field. Thus the twist and the electromagnetic field are complementary in some sense.

It is easily shown that a twisted Melvin universe, singular at $r = 0$ and reducing to the ordinary singular Melvin universe[10] for $c = \pm 1$ as above, can be obtained from the Weyl–Levi-Civita metric for a line mass.

Example 3: We use Theorem 2, Part 4 again, to generate a generalization of the Schwarzschild solution. We choose $x^k = x^0 = T$, with $a = 1$, $\epsilon = -1$, $f_\alpha = 0$, $U = \frac{1}{2}\ln(1 - 2M/R)$, $\gamma_{RR} = -1$, $\gamma_{\theta\theta} = -R^2(1 - 2M/R)$, $\gamma_{\phi\phi} = -R^2(1 - 2M/R)\sin^2\theta$, $\gamma_{\alpha\beta} = 0$ ($\alpha \neq \beta$). With $\bar{\beta} = 2kl^2$ and $\bar{\lambda} = 4ckl^2$ as before, we get

$$e^{-2\bar{U}} = l^2[(1 - 2M/R)^{-1} + k^2(1 - 2M/R) + 2ck], \tag{119}$$

$$\bar{f}_1 = \bar{f}_2 = 0, \quad \bar{f}_3 = 4kl^2(1 - c^2)^{\frac{1}{2}} M \cos\theta, \tag{120}$$

$$C = l^{-1}[-ck(1 - c^2)]^{\frac{1}{2}}[(1 - 2M/R)^{-1} + k^2(1 - 2M/R) + 2ck]^{-1}, \tag{121}$$

$$D = \pm(2kl)^{-1}[-ck]^{\frac{1}{2}}[-k^2(1 - 2M/R) + (1 - 2M/R)^{-1}] \cdot [(1 - 2M/R)^{-1} + k^2(1 - 2M/R) + 2ck]^{-1}, \tag{122}$$

where C and D are defined as before, in Eqs. (115).

If we put

$$R = at'' + b, \tag{123a}$$

$$T = h\psi, \tag{123b}$$

where a, b, and h are constants, we can show after

[10] M. A. Melvin, Phys. Letters **8**, 65 (1964).

some calculation that this is equivalent to Brill space[11] (electrified NUT space). There are just enough arbitrary constants in this generated solution to reproduce the constants of Brill space. In particular, we note that $c = -1$ yields the Reissner–Nordstrom solution and $c = 0$ yields NUT space.[12] Again we see the complementarity between twist and electromagnetic field.

Equation (123a) is both interesting and puzzling. It shows that in general the origin is shifted by this solution generation. b is usually $\neq 0$ ($b = 0$ only when $k = -c$) and is proportional to M.

5. DISCUSSION AND INTERPRETATION OR RESULTS

An obvious question is: What is the physical and/or mathematical significance of the theorems provable by this method? There is no good answer to this question. One sees in Theorem 2 that the change from old to new metrics amounts to a conformal transformation on the space of the x^α; but the x^k term experiences the inverse of that conformal transformation and may acquire nondiagonal components besides. The relationships of Petrov types, and such physical quantities as stress and shear of the old and new solutions, are also yet to be investigated. At present the method seems primarily to be a way of generalizing known solutions of the equations, while retaining some of the symmetry. It is hoped that this approach may be useful in throwing light on some fundamental properties of the Einstein–Maxwell system of equations. An ideal hope for the approach lies in the possibility of finding ways of combining solutions in arbitrary amounts to get new solutions, analogous to superposition in linear equations.

The method depends heavily on the existence of one Killing vector so that one can define a function U. (It should be noted that this Killing vector must be timelike or spacelike; no method exists at present for treating the null Killing-vector case.) One now wonders how closely such theorems as discussed in this paper are tied to the symmetries of the metric. Are there other theorems for metrics with higher symmetries?

One approach, in the case in which there exist two Killing vectors, is to apply the above theorems twice—once with respect to one variable and once with respect to another variable. Clearly, the metric must be of a rather special form to do this, and the first application of a theorem must not destroy the symmetry necessary for the second. We expect that repeated applications of the same theorem with respect to the same coordinate will merely produce another application of the same family—i.e., the generating transformations of the same type should form a group. This may be checked for the types of transformations in Theorem 2.

Another approach involves setting up a metric of higher symmetry and attempting to prove theorems in ways similar to before. So far, there is only one result available, given here:

Theorem 3: If the following metric is a vacuum solution:

$$-ds^2 = \epsilon[\eta e^{\alpha+\beta}(dx^l)^2 + e^{\alpha-\beta}(dx^k)^2 + \omega_{AB}\,dx^A\,dx^B], \tag{124}$$

and capital Latin letters take 0, 1, 2, 3, except k and l, with $\epsilon = \pm 1$, $\eta = \pm 1$, $\partial/\partial x^k = \partial/\partial x^l = 0$, then

$$-ds^2 = \epsilon[\delta\eta e^{\alpha+\beta-\mu\alpha}(dx'^l)^2 + \delta e^{\alpha-\beta+\mu\alpha}(dx'^k)^2 + a^2\exp(\tfrac{1}{2}\mu^2\alpha - \mu\beta)\omega_{AB}\,dx^A\,dx^B] \tag{125}$$

is also a solution, where $\delta = \pm 1$ (used only if $\eta = -1$), a and μ are constants, and x'^k and x'^l are linear combinations of x^k and x^l.

The proof is somewhat involved and will not be given here.

The method is not currently applicable to perfect fluids. Attempts to apply it to fluids have led to the equation of state $P = c^2\rho$, which occurs only if the matter has sound speed equal to the speed of light. This occurrence suggests that perhaps these theorems are derivable only for the presence of fields with fundamental velocity c. Indeed, one can derive similar theorems for a neutrino energy-momentum tensor. Of course, it is to be hoped—at least for solution-generation purposes!—that fluids will eventually be incorporated into the scheme. A small investigation has been made of the derivability of theorems in the Brans–Dicke theory[13]; it appears to be possible, although the mathematics is more involved.

It is anticipated that this method and similar methods will be very useful in general relativity—not only for generating new solutions of the equations, but also for examining more fundamental concepts relating the structure of the equations.

ACKNOWLEDGMENTS

The author expresses appreciation to Hugo D. Wahlquist and Frank B. Estabrook for conversations about the results of this paper, and to John H. Gardner for his continued support of this research.

[11] D. R. Brill, Phys. Rev. **133**, B845 (1964).

[12] E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).

[13] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961); R. H. Dicke, Phys. Rev. **125**, 2163 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Solution of the Equations of Radiative Transfer in a Free-Electron Atmosphere*

ROBERT L. BOWDEN AND NORMAN R. RICHARDSON†
Department of Physics, Virginia Polytechnic Institute, Blacksburg, Virginia

(Received 18 March 1968)

The radiative-transfer equations are solved for an electron-scattering stellar atmosphere as formulated by Chandrasekhar. The solution employs a transformation of the integro-differential form of the transfer equations into singular integral equations for the angular intensities of the radiation field. The Milne problem is solved to illustrate the method. In addition, the relationship is found between the above method of solution and Case's normal-mode expansion method. This leads to an alternate procedure for finding the normal-mode expansion coefficients. As an example of the method, the constant distributed source problem is solved for a half-space medium.

1. INTRODUCTION

During the past few years, several methods have been used to obtain exact results for linear transport problems in slab geometry, in which the theory of singular integral equations[1] plays a central role. Recently, Siewert and Fraley[2] have extended one of these methods, viz., Case's[3] singular normal-mode expansion method, to obtain solutions to the equations of radiative transfer for an electron-scattering atmosphere as formulated by Chandrasekhar.[4] In the normal-mode expansion method, the independent variables of the transfer equations are separated and the general solution is expressed as an expansion over the spectrum of the separation parameter. Application of boundary conditions leads to singular integral equations whose solutions are the expansion coefficients. By using orthogonality relations, the actual solution of these singular integral equations can be bypassed. In this paper we extend a method developed by Bowden, McCrosson, and Rhodes[5] for the one-velocity neutron transport equation with anisotropic scattering to the radiative-transfer equation. Our method of solution employs a transformation of the integro-differential equations of radiative transfer into singular integral equations for the angular intensity, in which the spatial variable enters only as a parameter. These equations can then be solved by standard methods. To illustrate this procedure, we solve Milne's problem. In addition, the relationship between our method and the normal-mode formulation follows readily from the orthogonality of the normal-mode eigenfunctions. This leads to an alternate method for determining the normal-mode expansion coefficients. As an example of this procedure, we solve the problem for a semi-infinite medium, containing a constant distributed source.

2. SINGULAR INTEGRAL EQUATIONS

A formulation of the transfer equation for scattering of radiation by free electrons in plane-parallel atmospheres is given by Chandrasekhar[4]:

$$\left(\mu\frac{\partial}{\partial x}+1\right)\mathbf{\Psi}(x,\mu)=\int_{-1}^{1}\mathbf{R}(\nu,\mu)\mathbf{\Psi}(x,\nu)\,d\nu\equiv\mathbf{T}(x,\mu), \tag{2.1}$$

where the intensity vector $\mathbf{\Psi}(x,\mu)$ has two components $\psi_1(x,\mu)$ and $\psi_2(x,\mu)$, corresponding to two states of polarization having electric vectors vibrating in a plane perpendicular to and along the plane of stratification of the medium, respectively. In the above equations the spatial variable x is the optical distance taken normal to the plane of stratification, in units of the Thompson scattering coefficient, μ is the direction cosine measured from the *inward* normal, and the phase matrix $\mathbf{R}(\nu,\mu)$ for azimuthally symmetric, plane-parallel geometry is given by the 2×2 matrix

$$\mathbf{R}(\nu,\mu)=\frac{3}{8}\begin{bmatrix}2(1-\nu^2)(1-\mu^2)+\nu^2\mu^2 & \mu^2\\ \nu^2 & 1\end{bmatrix}. \tag{2.2}$$

In this section we present our method of solving the integro-differential equation (2.1) by transforming it into a singular integral equation for the angular intensities in which the spatial variable enters only as a parameter. As an example of the method, we solve the classical Milne problem for a semi-infinite half-space with no incident radiation. The exact expression for the law of darkening, i.e., the emergent intensity at the vacuum interface, is found. We begin by noting the

* Supported in part by U.S. Atomic Energy Commission.
† Present address: Department of Physics, University of Florida.
[1] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).
[2] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[3] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[4] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950; also Dover Publications, Inc., New York, 1960).
[5] R. L. Bowden, F. J. McCrosson, and E. A. Rhodes, J. Math. Phys. **9**, 753 (1968).

following identity:

$$\left(\mu\frac{\partial}{\partial x}+1\right)\nu\Psi(x,\nu)$$
$$=\mu\left(\nu\frac{\partial}{\partial x}+1\right)\Psi(x,\nu)+(\nu-\mu)\Psi(x,\nu)$$
$$=\mu\mathbf{T}(x,\nu)+(\nu-\mu)\Psi(x,\nu). \tag{2.3}$$

We next define the vector $\Psi_0(x,\mu)$ as

$$\Psi_0(x,\mu)=\tfrac{1}{2}\int_{-1}^{1}\mathbf{G}(\nu,\mu)[\nu\Psi(x,\nu)-\mu\mathbf{H}(\nu,\mu)\Psi(x,\mu)]\frac{d\nu}{\nu-\mu}, \tag{2.4}$$

where

$$\mathbf{G}(\nu,\mu)=\tfrac{3}{2}\begin{bmatrix}1-\nu^2 & 0\\ \nu^2-\mu^2 & 1-\mu^2\end{bmatrix}, \tag{2.5}$$

and

$$\mathbf{H}(\nu,\mu)=\begin{bmatrix}1 & 0\\ -\dfrac{\mu(\nu-\mu)}{1-\mu^2} & 1+\dfrac{(2-3\mu^2)(\nu-\mu)}{3\mu(1-\mu^2)}\end{bmatrix}. \tag{2.6}$$

Using Eqs. (2.3), we find

$$\left(\mu\frac{\partial}{\partial x}+1\right)\Psi_0(x,\mu)$$
$$=\tfrac{1}{2}\int_{-1}^{1}\mathbf{G}(\nu,\mu)[\mu\mathbf{T}(x,\nu)+(\nu-\mu)\Psi(x,\nu)-\mu\mathbf{H}(\nu,\mu)\mathbf{T}(x,\mu)]\frac{d\nu}{\nu-\mu}. \tag{2.7}$$

If we define the moments

$$\rho_i^{(n)}(x)=\int_{-1}^{1}\nu^n\psi_i(x,\nu)\,d\nu,\quad i=1,2, \tag{2.8}$$

we can write

$$\mathbf{T}(x,\mu)=\frac{3}{8}\begin{bmatrix}2(1-\mu^2)\rho_1^{(0)}(x)-(2-3\mu^2)\rho_1^{(2)}(x)+\mu^2\rho_2^{(0)}(x)\\ \rho_1^{(2)}(x)+\rho_2^{(0)}(x)\end{bmatrix}. \tag{2.9}$$

Using Eqs. (2.5), (2.6), (2.8), and (2.9), we find that the right-hand side of Eq. (2.7) reduces to $\mathbf{T}(x,\mu)$. A comparison of this result with the equations of transfer Eq. (2.1) implies that $\Psi_0(x,\mu)$ is a *particular solution* to Eq. (2.1). To obtain the *general solution* to Eq. (2.1), we must add to $\Psi_0(x,\mu)$ a solution to the homogeneous equation

$$\left(\mu\frac{\partial}{\partial x}+1\right)\Psi_1(x,\mu)=0, \tag{2.10}$$

viz.,

$$\Psi_1(x,\mu)=-\tfrac{1}{2}\mathbf{F}(\mu)e^{-x/\mu}, \tag{2.11}$$

where $\mathbf{F}(\mu)$ is a column vector with elements $f_1(\mu)$ and $f_2(\mu)$ which are arbitrary functions of μ and whose forms are to be determined by appropriate boundary conditions. The *general solution* is then written as

$$\Psi(x,\mu)=\Psi_0(x,\mu)+\Psi_1(x,\mu), \tag{2.12}$$

i.e.,

$$\Psi(x,\mu)=\tfrac{1}{2}\int_{-1}^{1}\mathbf{G}(\nu,\mu)[\nu\Psi(x,\nu)-\mu\mathbf{H}(\nu,\mu)\Psi(x,\mu)]\times\frac{d\nu}{\nu-\mu}-\tfrac{1}{2}\mathbf{F}(\mu)e^{-x/\mu}. \tag{2.13}$$

Interpreting each part of the above integrals as Cauchy principal-valued integrals, we note that the second term on the right-hand side of Eq. (2.13) can be integrated explicitly, and we obtain the singular integral equations

$$\mathbf{\Lambda}(\mu)\Psi(x,\mu)-P\int_{-1}^{1}\mathbf{G}(\nu,\mu)\Psi(x,\nu)\frac{\nu\,d\nu}{\nu-\mu}=-\mathbf{F}(\mu)e^{-x/\mu}, \tag{2.14}$$

where we define

$$\mathbf{\Lambda}(\mu)=\begin{bmatrix}\lambda_1(\mu) & 0\\ 0 & \lambda_2(\mu)\end{bmatrix}, \tag{2.15}$$

with

$$\lambda_1(\mu)=-1+3(1-\mu^2)(1-\mu\tanh^{-1}\mu), \tag{2.16}$$

and

$$\lambda_2(\mu)=2+\lambda_1(\mu). \tag{2.17}$$

We are now concerned with the solution of the singular integral equations (2.14). If the function $\mathbf{F}(\mu)$ is known, these singular integral equations can be solved by standard methods.[1] However, before presenting this method for solving for $\Psi(x,\mu)$, we illustrate how $\mathbf{F}(\mu)$ is obtained by considering Milne's problem for a semi-infinite half-space medium. We measure x positive into the medium with $x=0$ at the vacuum boundary. The boundary conditions for the problem are that there be no radiation incident on the boundary and that the solution be of exponential order at infinity, i.e.,

$$\Psi_m(0,\mu)=0,\quad \mu>0, \tag{2.18}$$

and

$$\lim_{x\to\infty}e^{-x}\Psi_m(x,\mu)\to 0. \tag{2.19}$$

We shall also assume a constant net flux of radiation flowing through the atmosphere normal to the plane of stratification, this constant net flux being provided by the radiation coming from the "deep interior."

Applying the boundary condition (2.19), we find

$$\mathbf{F}(\mu)=0,\quad \mu<0. \tag{2.20}$$

When boundary condition (2.18) is applied, we obtain an expression for $\mathbf{F}(\mu)$:

$$\mathbf{F}(\mu) = \int_0^1 \nu \mathbf{G}(\nu, \mu)\mathbf{\Psi}_m(0, -\nu)\frac{d\nu}{\nu + \mu}, \quad \mu > 0, \quad (2.21)$$

where we have used the evenness of $\mathbf{G}(\nu, \mu)$ in both ν and μ. We now analytically continue $\mathbf{F}$ to the complex plane of z and write

$$\mathbf{F}(z) = \int_0^1 \nu \mathbf{G}(\nu, z)\mathbf{\Psi}_m(0, -\nu)\frac{d\nu}{\nu + z}, \quad \text{Re } z > 0. \quad (2.22)$$

We likewise extend $\mathbf{\Psi}$ to the complex plane and write the functional equation

$$\mathbf{\Omega}(z)\mathbf{\Psi}_m(x, z) - \int_{-1}^1 \mathbf{G}(\nu, \mu)\mathbf{\Psi}_m(x, \nu)\frac{\nu\, d\nu}{\nu - z} = \begin{cases} -\mathbf{F}(z)e^{-x/z}, & \text{Re } z > 0, \\ 0, & \text{Re } z < 0, \end{cases} \quad (2.23)$$

where

$$\mathbf{\Omega}(z) = \begin{bmatrix} \Omega_1(z) & 0 \\ 0 & \Omega_2(z) \end{bmatrix}, \quad (2.24)$$

with

$$\Omega_1(z) = -1 + 3(1 - z^2)[1 - z \tanh^{-1}(1/z)], \quad (2.25)$$

and

$$\Omega_2(z) = \Omega_1(z) + 2. \quad (2.26)$$

It is apparent from Eq. (2.21) that, if $\mathbf{\Psi}_m(0, -\mu)$, the emergent intensity, were known, then we could solve for $\mathbf{F}(\mu)$ [and hence $\mathbf{\Psi}_m(x, \mu)$], i.e., the intensity at any point in the system is given by its behavior at the boundary. A singular integral equation for $\mathbf{\Psi}_m(0, -\mu)$ can be obtained from Eq. (2.14) by setting $x = 0$ and restricting $\mu < 0$. After using the boundary condition (2.18) and the evenness of $\mathbf{\Lambda}(\mu)$ in μ, we find the singular integral equation for $\mathbf{\Psi}_m(0, -\mu)$:

$$\mathbf{\Lambda}(\mu)\mathbf{\Psi}_m(0, -\mu) - \mathrm{P}\int_0^1 \mathbf{G}(\nu, \mu)\mathbf{\Psi}_m(0, -\nu)\frac{\nu\, d\nu}{\nu - \mu} = 0, \quad \mu > 0. \quad (2.27)$$

The last equation can be solved by standard methods (as described by Muskhelishvili[1]) by considering the behavior of a certain sectionally analytic function on the complex plane of z cut along (0, 1) on the real axis

$$\mathbf{D}(z) = \frac{1}{2\pi i}\int_0^1 \mathbf{G}(\nu, \mu)\mathbf{\Psi}_m(0, -\nu)\frac{\nu\, d\nu}{\nu - z}. \quad (2.28)$$

Applying the Plemelj formulas[1] to $\mathbf{D}(z)$, we obtain

$$\mathbf{D}^+(\mu) + \mathbf{D}^-(\mu) = \frac{1}{\pi i}\mathrm{P}\int_0^1 \mathbf{G}(\nu, \mu)\mathbf{\Psi}_m(0, -\nu)\frac{\nu\, d\nu}{\nu - \mu}, \quad (2.29)$$

and

$$\mathbf{D}^+(\mu) - \mathbf{D}^-(\mu) = \tfrac{3}{2}\mu(1 - \mu^2)\mathbf{\Psi}_m(0, -\mu), \quad (2.30)$$

where $\mathbf{D}^+(\mu)$ and $\mathbf{D}^-(\mu)$ are the limits of $\mathbf{D}(z)$ as z approaches the cut (0, 1) from the upper and lower half-planes, respectively. Using the last two equations, we can write Eq. (2.27) as

$$\mathbf{\Omega}^-(\mu)\mathbf{D}^+(\mu) - \mathbf{\Omega}^+(\mu)\mathbf{D}^-(\mu) = 0, \quad (2.31)$$

where

$$\mathbf{\Omega}^\pm(\mu) = \begin{bmatrix} \Omega_1^\pm(\mu) & 0 \\ 0 & \Omega_2^\pm(\mu) \end{bmatrix}, \quad (2.32)$$

with

$$\Omega_1^\pm(\mu) = \lambda_1(\mu) \pm 3\pi i\mu(1 - \mu^2)/2 \quad (2.33)$$

and

$$\Omega_2^\pm(\mu) = \Omega_1^\pm(\mu) + 2 \quad (2.34)$$

as the limits of $\Omega_1(z)$ and $\Omega_2(z)$ as z approaches the cut $(-1, 1)$ from the upper and lower half-planes, respectively. Our singular integral equation (2.27) has been reduced to a homogeneous Hilbert problem: to find the nonvanishing sectionally analytic vector cut along (0, 1) with boundary values given by Eq. (2.31). The solution of this problem can be written

$$\mathbf{D}(z) = \mathbf{K}(z)\mathbf{X}(z), \quad (2.35)$$

where $\mathbf{X}(z)$ is a column vector with components $X_1(z)$ and $X_2(z)$ given by[2]

$$X_1(z) = \frac{1}{1 - z}\exp\left\{\frac{1}{\pi}\int_0^1 \ln\left(\frac{\Omega_1^+(\mu)}{\Omega_1^-(\mu)}\right)\frac{d\mu}{\mu - z}\right\} \quad (2.36)$$

and

$$X_2(z) = \exp\left\{\frac{1}{\pi}\int_0^1 \ln\left(\frac{\Omega_2^+(\mu)}{\Omega_2^-(\mu)}\right)\frac{d\mu}{\mu - z}\right\}, \quad (2.37)$$

and the matrix $\mathbf{K}(z)$ is given by

$$\mathbf{K}(z) = \frac{1}{2\pi i}\begin{bmatrix} a & 0 \\ 0 & b + cz \end{bmatrix}, \quad (2.38)$$

where a, b, and c are constants to be determined as follows. Since the net flux in the half-space is constant, we choose

$$2\int_0^1 \mu[\psi_{m1}(0, -\mu) + \psi_{m2}(0, -\mu)]\, d\mu = M, \quad (2.39)$$

where M is constant. This last equation can be used in obtaining the constants a, b, and c. Two additional equations for the constants are obtained by equating the right-hand sides of Eqs. (2.28) and (2.35) for $D_2(\pm 1)$:

$$(b \pm c)X_2(\pm 1) = \tfrac{3}{2}\int_0^1 \mu(\mu \pm 1)\psi_{m1}(0, -\mu)\, d\mu. \quad (2.40)$$

From Eqs. (2.30) and (2.35) we note that

$$\tfrac{3}{2}\mu(1-\mu^2)\Psi_m(0,-\mu) = \mathbf{K}(\mu)[\mathbf{X}^+(\mu)-\mathbf{X}^-(\mu)],\quad \mu>0, \tag{2.41}$$

which, when substituted into Eqs. (2.39) and (2.40), yields a set of simultaneous equations for the constants a, b, and c, which, in turn, yield

$$a = 3MA/4, \tag{2.42a}$$

$$b = 3M/4, \tag{2.42b}$$

and

$$c = -3M/4Q, \tag{2.42c}$$

where we define

$$Q = X_2(-1)X_1(+1) - X_2(+1)X_1(-1) \tag{2.43}$$

and

$$A = \frac{X_2(+1)X_1(-1) + X_2(-1)X_1(+1)}{X_2(-1)X_1(+1) - X_2(+1)X_1(-1)}. \tag{2.44}$$

Some of the integrals involved in the above determination were evaluated by noting the following identities which can be shown quickly by use of Cauchy's integral formula:

$$X_i(z) = X_i(\infty) + \frac{1}{2\pi i}\int_0^1 \frac{X_i^+(\mu)-X_i^-(\mu)}{\mu - z}\,d\mu, \quad i=1,2, \tag{2.45}$$

where $X_1(\infty)=0$ and $X_2(\infty)=1$. Now, by comparing Eqs. (2.22) and (2.28), we get

$$\mathbf{F}(z) = \frac{3M}{4}\begin{bmatrix} -X_1(-z)/Q \\ (A-z)X_2(-z)\end{bmatrix}, \tag{2.46}$$

which determines $\mathbf{F}(\mu)$ by letting z tend to $\mu\in(0,1)$. For the law of darkening from Eq. (2.41) we also obtain

$$\Psi_m(0,-\mu) = \frac{3M}{8}\begin{bmatrix} -5/QX_1(-\mu) \\ (\mu+A)/X_2(-\mu)\end{bmatrix}, \tag{2.47}$$

which is in agreement with the result found by Siewert and Fraley[2] and by Chandrasekhar.[4] In this last determination we have used the identities[2]

$$X_1(z)X_1(-z) = 5\Omega_1(z)/2 \tag{2.48}$$

and

$$X_2(z)X_2(-z) = \Omega_2(z)/2. \tag{2.49}$$

We now obtain the solution to Eq. (2.14), assuming that $\mathbf{F}(\mu)$ is known, by considering the sectionally analytic function cut along the line $(-1,1)$:

$$\mathbf{E}(x,z) = \frac{1}{2\pi i}\int_{-1}^1 \mathbf{G}(\nu,\mu)\Psi_m(x,\nu)\frac{\nu\,d\nu}{\nu - z}, \tag{2.50}$$

where x appears only as a parameter. We apply the Plemelj formula to the above equations and use the resulting boundary values to write Eq. (2.14) as

$$\frac{\Omega_i^+(\mu)}{\Omega_i^-(\mu)}E_i^-(x,\mu) - E_i^+(x,\mu) = \begin{cases} 3\mu(1-\mu^2)f_i(\mu)e^{-x/\mu}/2\Omega_i(\mu), & \mu>0,\\ 0, & \mu<0,\end{cases} \tag{2.51}$$

which are inhomogeneous Hilbert problems with cuts along $(-1,1)$. The solutions $X_{0i}(z)$, $i=1,2$, to the homogeneous Hilbert problems with boundary values along the line $(-1,1)$

$$\frac{X_{0i}^+(\mu)}{X_{0i}^-(\mu)} = \frac{\Omega_i^+(\mu)}{\Omega_i^-(\mu)} \tag{2.52}$$

are[2]

$$X_{01}(z) = \frac{1}{1-z^2}\exp\left\{\frac{1}{\pi}\int_{-1}^1 \arg\,\Omega_1^+(\mu)\frac{d\mu}{\mu - z}\right\} = X_1(z)X_1(-z) \tag{2.53}$$

and

$$X_{02}(z) = \exp\left\{\frac{1}{\pi}\int_{-1}^1 \arg\,\Omega_2^+(\mu)\frac{d\mu}{\mu - z}\right\} = X_2(z)X_2(-z). \tag{2.54}$$

With Eq. (2.52) we rewrite Eq. (2.52) as

$$\frac{E_i^+(x,\mu)}{X_{0i}^+(\mu)} - \frac{E_i^-(x,\mu)}{X_{0i}^-(\mu)} = \begin{cases}\dfrac{3\mu(1-\mu^2)f_i(\mu)e^{-x/\mu}}{2\Omega_i^-(\mu)X_{0i}^+(\mu)}, & \mu>0,\\ 0, & \mu<0.\end{cases} \tag{2.55}$$

We now consider the function

$$K_{0i}(x,z) = \frac{E_i(x,z)}{X_{0i}(z)} + \frac{1}{2\pi i}\int_0^1 \frac{3\mu(1-\mu^2)f_i(\mu)e^{-x/\mu}}{2\Omega_i^-(\mu)X_{0i}^+(\mu)}\frac{d\mu}{\mu - z}, \tag{2.56}$$

which is analytic in the finite complex plane of z, except perhaps for a cut along $(-1,1)$. Using the Plemelj formulas, we find $K_{0i}^+(x,\mu) - K_{0i}^-(x,\mu) = 0$, $\mu\in(-1,1)$. By examining the z dependence at infinity of $\mathbf{E}(x,z)$ from Eq. (2.50), and $X_{0i}(z)$ from Eqs. (2.52) and (2.53), we find

$$K_{0i}(x,z) = \frac{1}{2\pi i}[\sigma_i(x) + \omega_i(x)z], \quad i=1,2, \tag{2.57}$$

where $\sigma_i(x)$ and $\omega_i(x)$ are coefficients dependent on x to be determined below. We now solve for $E_i(x,z)$ from Eqs. (2.56) and (2.57) to obtain

$$E_i(x,z) = \frac{X_{0i}(z)}{2\pi i}\left[\sigma_i(x) + \omega_i(x)z - \int_0^1 \frac{3\nu(1-\nu^2)f_i(\nu)e^{-x/\nu}\,d\nu}{2\Omega_i^-(\nu)X_{0i}^+(\nu)(\nu - z)}\right], \quad i=1,2. \tag{2.58}$$

The condition of constant flux throughout the medium requires

$$-M = 2\int_{-1}^{1} \mu[\psi_{m1}(x, \mu) + \psi_{m2}(x, \mu)]\, d\mu, \quad (2.59)$$

where the constant M is the same as that used in Eq. (2.39). We apply the Plemelj formulas to Eq. (2.50) to obtain

$$\mu\psi_{mi}(x, \mu) = 2[E_i^+(x, \mu) - E_i^-(x, \mu)]/3(1 - \mu^2), \quad i = 1, 2. \quad (2.60)$$

With the limits $E_i^{\pm}(x, \mu)$ determined from Eq. (2.58) we substitute the last equation into our normalization condition (2.59) to obtain one equation for the coefficient

$$5\omega_1(x) + \omega_2(x) = n_1(x), \quad (2.61)$$

where

$$n_1(x) = -\tfrac{3}{2}M + \int_0^1 3\nu\left\{\frac{f_2(\nu)}{Y_2(\nu)} - \frac{f_1(\nu)}{Y_1(\nu)}\right\}e^{-x/\nu}\, d\nu, \quad (2.62)$$

with

$$Y_i(\nu) = \Omega_i^+(\nu)\Omega_i^-(\nu) = [\lambda_i(\nu)]^2 + [3\pi\nu(1 - \nu^2)/2]^2. \quad (2.63)$$

We next examine the z dependence at infinity of $E_2(x, z)$ in Eqs. (2.50) and (2.58) to find

$$\omega_2(x) = \tfrac{3}{2}\int_{-1}^{1} \nu[\psi_{m1}(x, \nu) + \psi_{m2}(x, \nu)]\, d\nu, \quad (2.64)$$

which, from Eq. (2.59), yields

$$\omega_2(x) = -3M/4. \quad (2.65)$$

The coefficient $\omega_1(x)$ is then completely determined by Eq. (2.61). Now, noting that, from Eq. (2.50), $D_2(x, 1) = -D_1(x, 1)$, from Eq. (2.59) we obtain

$$5\sigma_1(x) - \sigma_2(x) = n_2(x), \quad (2.66)$$

where

$$n_2(x) = \int_0^1 3\nu(\nu + 1)\left\{\frac{f_2(\nu)}{Y_2(\nu)} - \frac{f_1(\nu)}{Y_1(\nu)}\right\}e^{-x/\nu}\, d\nu \quad (2.67)$$

and Eqs. (2.61) and (2.65) have been used. Substituting $\psi_{m2}(x, -\mu)$, $\mu > 0$, from Eq. (2.60) into the original integro-differential equation (2.1) for $\psi_{m2}(x, \mu)$, we find the following differential equation for $\sigma_2(x)$:

$$d\sigma_2(x)/dx = 3M/4, \quad (2.68)$$

which has the solution

$$\sigma_2(x) = 3Mx/4 + \sigma_2(0). \quad (2.69)$$

The integration constant $\sigma_2(0)$ can be found by evaluating $\psi_{m2}(0, -\mu)$ from Eq. (2.60) and comparing it with Eq. (2.47). We find

$$\sigma_2(0) = \frac{3M}{4}\left\{A - \frac{1}{2\pi i}\int_0^1 [X_2^+(\mu) - X_2^-(\mu)]\, d\mu\right\}.$$

With $\sigma_2(x)$ so determined, Eq. (2.66) completely determines $\sigma_1(x)$. This completes the evaluation of the coefficients $\sigma_i(x)$ and $\omega_i(x)$.

From the functional equation (2.23) and $\mathbf{E}(x, z)$ given by Eq. (2.58) we can write

$$\psi_{mi}(x, z) = \begin{cases} \{2\pi i E_i(x, z) - f_i(z)e^{-x/z}\}/\Omega_i(z), & \operatorname{Re} z > 0, \\ 2\pi i E_i(x, z)/\Omega_i(z), & \operatorname{Re} z < 0. \end{cases}$$

The angular intensities $\psi_{mi}(x, \mu)$, $i = 1, 2$, are obtained from the last equation by letting z tend to $\mu \in (-1, 1)$.

3. COMPARISON WITH THE NORMAL-MODE EXPANSION METHOD

Although the method of Sec. II can be regarded as a completely independent method of solution, it is intimately related to the normal-mode expansion method as presented by Siewert and Fraley[2] and Smith and Siewert.[6] Following Case's method,[3] Siewert and Fraley separated the variables of Eq. (2.1) with a solution of the form

$$\mathbf{\Psi}(x, \mu) = \mathbf{\Phi}(\eta, \mu)e^{-x/\eta}, \quad (3.1)$$

where $\mathbf{\Phi}(\eta, \mu)$ is a solution to the equation

$$(\eta - \mu)\mathbf{\Phi}(\eta, \mu) = \eta\int_{-1}^{1} \mathbf{R}(\nu, \mu)\mathbf{\Phi}(\eta, \nu)\, d\nu, \quad (3.2)$$

and the spectrum of the separation parameter η is as follows:

Discrete solutions: For $\eta \notin (-1, 1)$ Siewert and Fraley found that η must be a zero of the function $\Omega_1(\eta)$ as defined by Eq. (2.25). However, Ω_1 has a double root at infinity, and they write the discrete solution as

$$\mathbf{\Phi}_+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (3.3)$$

and, because of the degeneracy due to the double root at infinity, a second linear-independent discrete eigenvector

$$\mathbf{\Psi}_-(x, \mu) = (x - \mu)\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (3.4)$$

which is a solution to Eq. (2.1).

Continuous Solutions: For $\eta \in (-1, 1)$ Siewert and Fraley write the continuous eigensolutions of Eq. (3.2):

$$\mathbf{\Phi}_1(\eta, \mu) = \begin{bmatrix} \mathrm{P}\dfrac{3\eta(1 - \mu^2)/2}{\eta - \mu} + \lambda_1(\eta)\delta(\eta - \mu) \\ 0 \end{bmatrix}, \quad (3.5a)$$

[6] O. J. Smith and C. E. Siewert, J. Math. Phys. **8**, 2467 (1967).

and

$$\mathbf{\Phi}_2(\eta, \mu) = \begin{bmatrix} -3\eta(\eta + \mu)/2 \\ \mathrm{P}\dfrac{3\eta(1-\eta^2)/2}{\eta - \mu} + \lambda_2(\eta)\delta(\eta - \mu) \end{bmatrix}, \quad (3.5b)$$

with $\lambda_i(\eta)$ defined by Eqs. (2.16) and (2.17); $\delta(\eta - \mu)$ is the Dirac delta function, and the symbol P denotes that the Cauchy principal value is to be taken in any integration involving the following term. Smith and Siewert[6] have shown that the eigensolutions $\mathbf{\Phi}_+$, $\mathbf{\Phi}_1(\eta, \mu)$, $\mathbf{\Phi}_2(\eta, \mu)$, and $\mathbf{\Psi}_-(0, \mu)$ are complete on the range $\mu \in (-1, 1)$ such that an arbitrary two-component vector $\mathbf{\Gamma}(\mu)$ defined on $-1 \leq \mu \leq 1$ can be expanded as

$$\mathbf{\Gamma}(\mu) = A_+\mathbf{\Phi}_+ + A_-\mathbf{\Psi}_-(0, \mu) + \int_{-1}^{1} \alpha(\eta)\mathbf{\Phi}_1(\eta, \mu)\, d\eta + \int_{-1}^{1} \beta(\eta)\mathbf{\Phi}_2(\eta, \mu)\, d\eta. \quad (3.6)$$

A consequence of this completeness relationship is that the general solution of Eq. (2.1) can be

$$\mathbf{\Psi}(x, \mu) = A_+\mathbf{\Phi}_+ + A_-\mathbf{\Psi}_-(x, \mu) + \int_{-1}^{1} \alpha(\eta)\mathbf{\Phi}_1(\eta, \mu)e^{-x/\eta}\, d\eta + \int_{-1}^{1} \beta(\eta)\mathbf{\Phi}_2(\eta, \mu)e^{-x/\eta}\, d\eta, \quad (3.7)$$

where A_+, A_-, $\alpha(\eta)$, and $\beta(\eta)$ are arbitrary expansion coefficients to be evaluated by boundary conditions. In addition, Siewert and Fraley have shown that $\mathbf{\Phi}_+$, $\mathbf{\Phi}_1(\eta, \mu)$, and $\mathbf{\Phi}_2(\eta, \mu)$ are complete on the range $\mu \in (0, 1)$, in the sense that an arbitrary two-component vector $\mathbf{\Theta}(\mu)$ defined on the range $0 \leq \mu \leq 1$ can be expanded as

$$\mathbf{\Theta}(\mu) = A_+\mathbf{\Phi}_+ + \int_{0}^{1} \alpha(\eta)\mathbf{\Phi}_1(\eta, \mu)\, d\eta + \int_{0}^{1} \beta(\eta)\mathbf{\Phi}_2(\eta, \mu)\, d\eta. \quad (3.8)$$

They use the completeness property (3.8) to solve the Milne problem.

Full-range orthogonality and normalization integrals are given by Smith and Siewert.[6] They showed that

$$\int_{-1}^{1} \mu\tilde{\mathbf{\Phi}}(\eta', \mu)\mathbf{\Phi}(\eta, \mu)\, d\mu = 0, \quad \eta \neq \eta', \quad (3.9)$$

where the tilde superscript indicates the transpose of the vector. They defined the full-range scalar product

$$\langle i \mid j \rangle = \int_{-1}^{1} \mu\tilde{\mathbf{\Phi}}_i(\eta', \mu)\mathbf{\Phi}_j(\eta, \mu)\, d\mu, \quad i, j = +, 1, 2, \quad (3.10)$$

and find

$$\langle i \mid j \rangle = 0, \quad i, j = +, 1, 2, \quad i \neq j, \quad (3.11)$$

and

$$\langle i \mid i \rangle = \eta\Omega_i^+(\eta)\Omega_i^-(\eta)\delta(\eta - \eta'), \quad i = 1, 2, \quad (3.12)$$

where $\Omega_i^{\pm}(\eta)$ are defined by Eqs. (2.33) and (2.34). By replacing $\mathbf{\Phi}_j(\eta, \mu)$ in Eq. (3.10) with $\mathbf{\Psi}_-(0, \mu)$, an additional orthogonality relation is

$$\langle i \mid - \rangle = 0, \quad i = -, 1, 2. \quad (3.13)$$

We can determine the relationship between our singular integral-equation method and the normal-mode expansion technique by evaluating the integrals

$$\int_{-1}^{1} \nu\tilde{\mathbf{\Phi}}_i(\mu, \nu)\mathbf{\Psi}(x, \nu)\, d\nu, \quad i = 1, 2. \quad (3.14)$$

If first we substitute the explicit forms of $\mathbf{\Phi}_i$ from Eqs. (3.5), after rearranging in matrix form we obtain

$$\begin{bmatrix} \displaystyle\int_{-1}^{1} \nu\tilde{\mathbf{\Phi}}_1(\mu, \nu)\mathbf{\Psi}(x, \nu)\, d\nu \\ \displaystyle\int_{-1}^{1} \nu\tilde{\mathbf{\Phi}}_2(\mu, \nu)\mathbf{\Psi}(x, \nu)\, d\nu \end{bmatrix} = \mu\left\{\mathbf{\Lambda}(\mu)\mathbf{\Psi}(x, \mu) - p\int_{-1}^{1} \mathbf{G}(\nu, \mu)\mathbf{\Psi}(x, \nu)\frac{\nu\, d\nu}{\nu - \mu}\right\}, \quad (3.15)$$

where $\mathbf{G}(\nu, \mu)$ and $\mathbf{\Lambda}(\mu)$ are defined by Eqs. (2.5) and (2.15). The integrals (3.14) can also be evaluated by using the general expression for $\mathbf{\Psi}(x, \nu)$ given by the normal-mode expansion (3.7). The integration involving the two discrete terms in $\mathbf{\Psi}(x, \nu)$ vanishes due to orthogonality relations (3.11) and (3.13) found by Smith and Siewert.[6] After changing the order of integration of the two continuum terms, we perform the ν integration using Eqs. (3.12) and obtain

$$\begin{bmatrix} \displaystyle\int_{-1}^{1} \nu\tilde{\mathbf{\Phi}}_1(\mu, \nu)\mathbf{\Psi}(x, \nu)\, d\nu \\ \displaystyle\int_{-1}^{1} \nu\tilde{\mathbf{\Phi}}_2(\mu, \nu)\mathbf{\Psi}(x, \nu)\, d\nu \end{bmatrix} = \begin{bmatrix} \alpha(\mu)\Omega_1^+(\mu)\Omega_1^-(\mu) \\ \beta(\mu)\Omega_2^+(\mu)\Omega_2^-(\mu) \end{bmatrix} \mu e^{-x/\mu}. \quad (3.16)$$

Equating the right-hand sides of Eqs. (3.15) and (3.16) and comparing the result with Eq. (2.14), we obtain the relationship between our method and the normal-mode expansion method:

$$\begin{bmatrix} f_1(\mu) \\ f_2(\mu) \end{bmatrix} = -\begin{bmatrix} \alpha(\mu)\Omega_1^+(\mu)\Omega_1^-(\mu) \\ \beta(\mu)\Omega_2^+(\mu)\Omega_2^-(\mu) \end{bmatrix}. \quad (3.17)$$

Equation (3.17) indicates that solving for $\mathbf{F}(\mu)$ in the singular integral equation provides an alternate means for obtaining the normal-mode expansion coefficients. To illustrate, we consider the Milne problem for a half-space with a constant, distributed source S with components s_1 and s_2. The boundary conditions, similar to those of Eqs. (2.18) and (2.19), are

$$\Psi_s(0, \mu) = 0, \quad \mu > 0, \tag{3.18}$$

and

$$\lim_{x\to\infty} e^{-x}\Psi_S(x, \mu) \to 0, \tag{3.19}$$

where the angular intensity vector $\Psi_S(x, \mu)$ represents two perpendicularly polarized intensities in the medium with sources.

The integro-differential equation for the problem with sources [cf. Eq. (2.1)] is

$$\left(\mu\frac{\partial}{\partial x} + 1\right)\Psi_S(x, \mu) = \int_{-1}^{1}\mathbf{R}(\nu, \mu)\Psi_S(x, \nu)\,d\nu + \mathbf{S} \tag{3.20}$$

and has a solution

$$\begin{aligned}\Psi_S(x, \mu) = \Psi_\infty(x, \mu) &+ A_{S+}\Phi_+ + A_{S-}\Psi_-(x, \mu)\\ &+ \int_{-1}^{1}\alpha_S(\eta)\Phi_1(\eta, \mu)e^{-x/\eta}\,d\eta\\ &+ \int_{-1}^{1}\beta_S(\eta)\Phi_2(\eta, \mu)e^{-x/\eta}\,d\eta,\end{aligned} \tag{3.21}$$

where $\Psi_\infty(x, \mu)$ is the particular solution to the constant source problem. We find the following particular solution to Eq. (3.20):

$$\Psi_\infty(x, \mu) = \begin{bmatrix} -\frac{3}{4}(s_1 + s_2)(x^2 - 2x\mu) - 5s_1\mu^2 \\ -\frac{3}{4}(s_1 + s_2)(x^2 - 2x\mu) \\ \qquad -\frac{3}{2}(s_1 + s_2)\mu^2 - \frac{9}{2}s_1 + \frac{5}{2}s_2 \end{bmatrix}. \tag{3.22}$$

We now proceed as in Eq. (3.16) and substitute $\Psi_S(x, \mu)$ into the integrals (3.14) and obtain

$$\begin{bmatrix}\int_{-1}^{1}\nu\tilde{\Phi}_1(\mu, \nu)\Psi_S(x, \nu)\,d\nu \\ \int_{-1}^{1}\nu\tilde{\Phi}_2(\mu, \nu)\Psi_S(x, \nu)\,d\nu\end{bmatrix} = \begin{bmatrix}\alpha_S(\mu)\Omega_1^+(\mu)\Omega_1^-(\mu) \\ \beta_S(\mu)\Omega_2^+(\mu)\Omega_2^-(\mu)\end{bmatrix}\mu e^{-x/\mu} + \mu\begin{bmatrix}2s_1 \\ 4s_2 - 3\mu^2(s_1 + s_2)\end{bmatrix}, \tag{3.23}$$

from which we write [cf. Eq. (2.14)]

$$\Lambda(\mu)\Psi_S(x, \mu) - p\int_{-1}^{1}\mathbf{G}(\nu, \mu)\Psi_S(x, \nu)\frac{\nu\,d\nu}{\nu - \mu} = -\mathbf{F}_S(\mu)e^{-x/\mu} + \mathbf{N}(\mu), \tag{3.24}$$

where

$$\mathbf{N}(\mu) = \begin{bmatrix}2s_1 \\ 4s_2 - 3\mu^2(s_1 + s_2)\end{bmatrix} \tag{3.25}$$

and

$$\mathbf{F}_S(\mu) = -\begin{bmatrix}\alpha_S(\mu)\Omega_1^+(\mu)\Omega_1^-(\mu) \\ \beta_S(\mu)\Omega_2^+(\mu)\Omega_2^-(\mu)\end{bmatrix}. \tag{3.26}$$

We now determine $\mathbf{F}_S(\mu)$ in a procedure similar to that used in Sec. II. We proceed to solve Eq. (3.24) subject to the boundary conditions (3.18) and (3.19). In order that our solution be of exponential order as x approaches infinity [Eq. (3.19)], we must have

$$\mathbf{F}_S(\mu) = 0, \quad \mu < 0. \tag{3.27}$$

From the condition of no incident radiation on the interface from the vacuum [Eq. (3.18)], we obtain

$$\mathbf{F}_S(\mu) = \int_0^1\mathbf{G}(\nu, \mu)\Psi_S(0, -\nu)\frac{\nu\,d\nu}{\nu + \mu} + \mathbf{N}(\mu), \quad \mu > 0, \tag{3.28}$$

where we have used the fact that $\mathbf{N}(\mu)$ is an even function of μ. When we extend $\mathbf{F}_S$ to the complex plane of z, we obtain

$$\mathbf{F}_S(z) = \int_0^1\mathbf{G}(\nu, z)\Psi_S(0, -\nu)\frac{\nu\,d\nu}{\nu + z} + \mathbf{N}(z), \quad \operatorname{Re} z > 0. \tag{3.29}$$

Using Eq. (3.18) and restricting the direction cosine to values less than zero, we obtain a singular integral equation for $\Psi_S(0, -\mu)$:

$$\Lambda(\mu)\Psi_S(0, -\mu) - \mathrm{P}\int_0^1\mathbf{G}(\nu, \mu)\Psi_S(0, -\nu)\frac{\nu\,d\nu}{\nu - \mu} = \mathbf{N}(\mu), \quad \mu > 0. \tag{3.30}$$

We define a sectionally analytic function $\mathbf{E}(z)$, cut along the line (0, 1), as

$$\mathbf{E}(z) = \frac{1}{2\pi i}\int_0^1\mathbf{G}(\nu, z)\Psi_S(0, -\nu)\frac{\nu\,d\nu}{\nu - z}. \tag{3.31}$$

Applying the Plemelj formulas,[4] we obtain, in component form,

$$E_i^+(\mu) - \frac{\Omega_i^+(\mu)}{\Omega_i^-(\mu)}E_i^-(\mu) = \tfrac{3}{2}\mu(1 - \mu^2)\frac{N_i(\mu)}{\Omega_i^-(\mu)}, \quad i = 1, 2, \tag{3.32}$$

which is an inhomogeneous vector Hilbert problem, i.e., we wish to find the sectionally analytic vector $\mathbf{E}(z)$ whose boundary values satisfy Eqs. (3.32). Using standard techniques,[4] we can write the solution as

$$E_i(z) = X_i(z)\left[R_{ii}(z) + \frac{1}{2\pi i}\int_0^1\frac{3\mu(1 - \mu^2)N_i(\mu)}{2X_i^+(\mu)\Omega_i^-(\mu)}\frac{d\mu}{\mu - z}\right], \tag{3.33}$$

where R_{ii} are the diagonal elements of the matrix

$$\mathbf{R}(z) = \frac{1}{2\pi i}\begin{bmatrix} d_1 & 0 \\ 0 & d_2 + d_3 z \end{bmatrix}, \tag{3.34}$$

with the d_i constants determined below and $X_i(z)$ are defined by Eqs. (3.36) and (3.37).

The integrals in Eqs. (3.33) can be evaluated by noting the following identities, which can be proved by use of Cauchy's integral formula

$$\frac{1}{X_1(z)} + u_1 - z = \frac{3}{2}\int_0^1 \frac{\nu(1-\nu^2)}{X_1^+(\nu)\Omega_1^-(\nu)}\frac{d\nu}{\nu - z} \tag{3.35}$$

and

$$z^2\left[1 - \frac{1}{X_2(z)}\right] + zv_0 + v_0^2 + v_1 = \frac{3}{2}\int_0^1 \frac{\nu(1-\nu^2)}{X_2^+(\nu)\Omega_2^-(\nu)}\frac{\nu^2\, d\nu}{\nu - z}, \tag{3.36}$$

where

$$u_n = \frac{1}{2\pi i}\int_0^1 [X_1^+(\mu) - X_1^-(\mu)]\mu^n\, d\mu \tag{3.37}$$

and

$$v_n = \frac{1}{2\pi i}\int_0^1 [X_2^+(\mu) - X_2^-(\mu)]\mu^n\, d\mu. \tag{3.38}$$

We finally get the components of $\mathbf{E}(z)$:

$$E_1(z) = \frac{1}{2\pi i}[d_1 + 2s_1(u_1 - z)]X_1(z) - \frac{2s_1}{2\pi i} \tag{3.39a}$$

and

$$E_2(z) = \frac{1}{2\pi i}[d_2 + d_3 z + 4s_2 - 3(s_1 + s_2)(z^2 + zv_0 + v_0^2 + v_1)]X_2(z) - \frac{1}{2\pi i}[4s_2 - 3(s_1 + s_2)z^2]. \tag{3.39b}$$

From Eqs. (3.29) and (3.31) we find

$$\mathbf{F}_S(z) = 2\pi i\mathbf{E}(-z) + \mathbf{N}(z), \tag{3.40}$$

which, from Eqs. (3.25) and (3.37), can be written as

$$f_{S1}(z) = [d_1 + 2s_1(u_1 + z)]X_1(z) \tag{3.41a}$$

and

$$f_{S2}(z) = [d_2 - d_3 z + 4s_2 - 3(s_1 + s_2)(z^2 - v_0 z + v_0^2 + v_1)]X_2(-z). \tag{3.41b}$$

Using Eqs. (3.26) and (3.41), we obtain the normal-mode expansion coefficients for the constant source problem

$$\alpha_S(\mu) = -\frac{[d_1 + 2s_1(u_1 + \mu)]X_1(-\mu)}{\Omega_1^+(\mu)\Omega_1^-(\mu)} \tag{3.42}$$

and

$$\beta_S(\mu) = -\frac{[d_2 - d_3\mu + 4s_2 - 3(s_1 + s_2)(\mu^2 - v_0\mu + v_0^2 + v_1)]X_2(-\mu)}{\Omega_2^+(\mu)\Omega_2^-(\mu)}. \tag{3.43}$$

We have found the expansion coefficients to within the three constants d_1, d_2, and d_3. The procedure for finding the constants parallels that for the source-free problem of Sec. II. As before, we normalize the exit current at the vacuum interface:

$$J = 2\int_0^1 \mu[\psi_{S1}(0, -\mu) + \psi_{S2}(0, -\mu)]\, d\mu, \tag{3.44}$$

where use has been made of boundary condition (3.18). This last equation can be used in obtaining the constants d_i. Two additional equations for the constants are obtained by equating the right-hand sides of Eqs. (3.31) and (3.39) for $E_2(\pm 1)$:

$$X_2(\pm 1)[d_2 \pm d_3 + 4s_2 - 3(s_1 + s_2)(1 \pm v_0 \pm v_0^2 + v_1)] - s_2 + 3s_1 = \tfrac{3}{2}\int_0^1 \nu[\nu \pm 1]\psi_{S1}(0, -\nu)\, d\nu. \tag{3.45}$$

We note that, by applying Plemelj's formulas to Eq. (3.31) and using Eqs. (3.39), we obtain

$$\tfrac{3}{2}\mu(1 - \mu^2)\mathbf{\Psi}_S(0, -\mu) = \mathbf{R}(\mu)[\mathbf{X}^+(\mu) - \mathbf{X}^-(\mu)] - \mathbf{N}(\mu), \tag{3.46}$$

which, when substituted into Eqs. (3.44) and (3.45), yields a set of simultaneous equations for the constants d_i. These equations then yield

$$d_1 = \{-r_1 + [1 - X_2(+1)][r_2 X_2(-1) + r_3 X_2(+1)]\}/Q, \tag{3.47a}$$

$$d_2 = \{r_1 W + \tfrac{1}{2}[r_3 X_1(+1) - r_2 X(-1)] + \tfrac{1}{2}W(r_2 - r_3)\}/Q, \tag{3.47b}$$

and

$$d_3 = [2r_1 + r_2 + r_3]/2, \tag{3.47c}$$

where Q is defined by Eq. (2.43),

$$W = X_2(+1)X_1(-1) + X_2(-1)X_1(+1), \tag{3.47d}$$

$$r_1 = \tfrac{3}{4}J - 3s_1u_1\Gamma_3 + 2s_1\Gamma_5 - 4s_2\Gamma_1 + 3(s_1 + s_2)[\Gamma_2 - \Gamma_6 + v_0\Gamma_2 + (v_0^2 + v_1)\Gamma_1], \tag{3.48a}$$

$$r_2 = 2s_1(1 - u_1)X_1(+1) + 2s_1u_0 - 4X_2(+1)s_2 + s_2 - 3s_1 + 3(s_1 + s_2)X_2(+1)(1 + v_0 + v_0^2 + v_1), \quad (3.48b)$$

and

$$r_3 = 2s_1(1 - u_1)X_1(-1) + 2s_1u_0 - 4X_2(-1)s_2 - s_2 + 3s_1 + 3(s_1 + s_2)X_2(-1)(1 - v_0 + v_0^2 + v_1), \quad (3.48c)$$

with

$$\Gamma_1 = [X_2(-1) - X_2(+1)]/2, \quad (3.49a)$$

$$\Gamma_2 = [2 - X_2(+1) - X_2(-1)]/2, \quad (3.49b)$$

$$\Gamma_3 = [X_1(-1) - X_1(+1)]/2, \quad (3.49c)$$

$$\Gamma_4 = -X_1(+1), \quad (3.49d)$$

$$\Gamma_5 = -X_1(-1), \quad (3.49e)$$

and

$$\Gamma_6 = 1 + v_0 - X_2(-1). \quad (3.49f)$$

With these d_i we have completely determined the expansion coefficients $\alpha_S(\mu)$ and $\beta_S(\mu)$. The discrete expansion coefficients A_{S-} and A_{S+} are determined from the integrals (3.14) with $\mathbf{\Phi}_+$ and $\mathbf{\Psi}_-(0, -\mu)$ given by Eqs. (3.3) and (3.4) replacing $\mathbf{\Phi}_i(x, \mu)$. Substituting for the normal-mode expansion $\mathbf{\Psi}_S(x, \mu)$ from Eq. (3.21), we find, using the orthogonality relations, that

$$\int_{-1}^{1} \mu \tilde{\mathbf{\Phi}}_+ \mathbf{\Psi}_S(x, \mu)\, d\mu = -4A_{S-}/3. \quad (3.50)$$

If instead we substitute for $\mathbf{\Phi}_+$, we obtain our normalization condition [cf. Eq. (3.44)]:

$$\frac{1}{2}\int_{-1}^{1} \mu \begin{bmatrix} \tilde{1} \\ 1 \end{bmatrix} \mathbf{\Psi}_S(x, \mu) = -J. \quad (3.51)$$

From Eqs. (3.50) and (3.51) we see that

$$A_{S-} = 3J/8. \quad (3.52)$$

In the same fashion as in Eq. (3.50), we find, when we substitute the expansion (3.21), that

$$\int_{-1}^{1} \mu \tilde{\mathbf{\Psi}}_-(0, \mu)\mathbf{\Psi}_S(x, \mu)\, d\mu = -\tfrac{4}{3}A_{S+} + \tfrac{28}{5}s_1 - \tfrac{16}{15}s_2. \quad (3.53)$$

If in Eq. (3.53) we apply the boundary condition (3.18) and use the explicit form of $\mathbf{\Psi}_-(0, \mu)$, we obtain

$$\int_{-1}^{1} \mu^2 \begin{bmatrix} \tilde{1} \\ 1 \end{bmatrix} \mathbf{\Psi}_S(0, -\mu)\, d\mu = \int_{-1}^{1} \mu^2[\psi_{S1}(0, -\mu) + \psi_{S2}(0, -\mu)]\, d\mu. \quad (3.54)$$

We now use the form of $\mathbf{\Psi}_S(0, -\mu)$ from Eq. (3.46) in (3.54) and we combine it with Eq. (3.53) to obtain

$$A_{S+} = \tfrac{1}{2}d_1(\Gamma_3 + \Gamma_5) + \tfrac{1}{2}d_2\Gamma_2 + \tfrac{1}{2}d_3(\Gamma_2 - \Gamma_6) + \tfrac{21}{5}s_1 - \tfrac{4}{5}s_2u_1(\Gamma_3 + \Gamma_5) - s_1(\Gamma_3 - \Gamma_5) + 2s_2\Gamma_2 - \tfrac{3}{2}(s_1 + s_2) \times [\Gamma_2 - v_1 + (v_0^2 + v_1)\Gamma_2 + v_0(\Gamma_2 - \Gamma_6)]. \quad (3.55)$$

4. SUMMARY

In summary, we have presented a method for obtaining exact solutions for the perpendicularly polarized intensities in an electron-scattering stellar atmosphere. The solutions are obtained by transforming the integro-differential form of the radiative transfer equations for the angular intensities. As an example of this procedure, we have solved the Milne problem and the law of darkening. The connection between our singular integral equation and the normal-mode expansion method was obtained by use of orthogonality relations. We found that this led to an alternate procedure for obtaining the normal mode expansion coefficients. We have illustrated that procedure by solving the constant distributed source problem for a semi-infinite medium for which we found the discrete and continuum coefficients. It appears for the problems considered that neither method presents an appreciable advantage over the other. However, it does seem that in the analogous multigroup neutron transport equation, where the system has absorption and the cross sections are not the same for each group, this method may present a distinct advantage. This will be the subject of a future paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Connection between Spin, Statistics, and Kinks

DAVID FINKELSTEIN*†
Belfer Graduate School of Science, Yeshiva University, New York
AND
JULIO RUBINSTEIN‡
Department of Physics, Columbia University, New York

(Received 7 August 1967)

Sufficiently nonlinear classical fields admit modes called kinks, whose number is strictly conserved in virtue of boundary conditions and continuity of the field as a function of space and time. In a quantum theory of such fields, with canonical commutation (not anticommutation) relations, kinks and their conservation still persist, and even if the intrinsic angular momentum is an integer, a rotating kink can have half-odd angular momentum, if double-valued state functionals are admitted. We formulate a natural concept of exchange appropriate for kinks. The principal result is that for fields with integer-valued intrinsic angular momentum, the observed relation between spin and (exchange) statistics follows from continuity alone, parastatistics being excluded. It is likely that in the theories with even (odd) exchange statistics, suitable creation operators will commute (anticommute). We show that, while the rotational spectrum of a kink will in general possess both integer and half-odd spin states, in fields with integer-valued intrinsic angular momentum only one of these two possibilities will ever be observed for each kind of kink, and that there is a nonzero "particle number" (strictly conserved, additive, scalar quantum number) attached to half-odd-spin kinks of each kind. It then follows that a boson and a fermion kink will always differ in at least one particle number, as well as in spin, and that, in particular, every fermion kink will have some nonzero particle number. These results are consistent with the hypothesis that the spinor fields usually employed to describe half-odd-spin quanta are not fundamental, but are useful "point-limit" approximations to operators creating or annihilating excitations in a nonlinear field of particular kinds of kinks in particular internal states.

INTRODUCTION

Quanta of integer spin are described by fields obeying canonical commutation relations, ramifications of the basic relation

$$pq - qp = \hbar/i,$$

while since the work of Jordan and Wigner,[1] quanta of half-odd spin have been described by fields obeying anticommutation relations like

$$pq + qp = \hbar/i.$$

For brevity, let us call these *canonical* and *anticanonical* quantizations, respectively. We have been exploring[2,3] a single kind of field quantization which appears to embrace both the canonical and anticanonical, and to reduce to them in appropriate circumstances. In this kind of quantization, for which we employ the term *multivalued quantization*, all the fundamental fields are supposed to obey commutation relations, as in canonical quantization, but the state-functionals on which they act are permitted to be multivalued. More succinctly, a multivalued quantization of a classical system S is a canonical quantization of the "covering system" $\tilde{S}$, and the definition of the covering system $\tilde{S}$ is formulated in close analogy to that of the (universal) covering space in topology.

It is well known that half-odd spin is related to a "quantum-mechanical double valuedness" under rotation. Here we show that Fermi–Dirac statistics can be related to a "quantum-mechanical double valuedness" under a different process which we call field exchange.

The most significant results are the following two:

1. In the framework of the usual two quantizations, it is well known that the spin and statistics of quanta are independent *per se* and further considerations (analyticity, Lorentz invariance, etc.) are necessary to account for the observed correlation. In multivalued quantized field theories, continuity arguments suffice to show that among theories with integer-valued intrinsic spin, the cases that admit half-odd spin states at all are the same as the cases that admit odd statistics. The correlation found in nature appears to follow purely from continuity considerations. Moreover, since there are only two "kinds" of spin, integer and half-odd integer, it follows that in this framework there are only two kinds of statistics and "parastatistics" is impossible.

2. In multivalued quantization, there is always a conserved additive integer attached to structures possessing half-odd spin. This coupling between the values of rotational and nonrotational quantum numbers corresponds well with the empirical fact that there exists a conserved "particle number" (strictly conserved, additive, scalar quantum number) that is nonzero for every fermion, e.g., $N_B + N_e + N_\mu$, where N_B is the baryon number, N_e the electron

* Supported by the National Science Foundation (Grant No. GP 6137).
† Young Men's Philanthropic League Professor of Physics.
‡ Supported by the U.S. Atomic Energy Commission. Present address: Dept. of Physics, University of Saskatchewan.
[1] P. Jordan and E. P. Wigner, Z. Physik **47**, 631 (1928).
[2] T. H. R. Skyrme, Nucl. Phys. **31**, 556 (1962).
[3] D. Finkelstein, J. Math. Phys. **7**, 1218 (1966).

number, and N_μ the muon number ($N_\mu + N_\mu$ is a "lepton number"). In brief, there are no truly neutral fermions. No boson and fermion can have exactly the same set of values for their conserved "particle numbers."

We may summarize our methods as follows:

Nonlinearity is crucial for our development. Loosely speaking, when a field is nonlinear enough, there are shapes it can assume which cannot be wiped out by any continuous process whatever. We have called these classically indestructible objects "kinks" and proven their possibility and conservation in quantum theory as well, by primitive considerations of continuity which ultimately reduce to homotopy calculations, recourse to detailed dynamical calculations being unnecessary.[2,3] Equally primitive is the distinction between multivalued quantized systems which can possess state functions double-valued under 2π rotation, and those which cannot: briefly, theories with half-odd spin vs integer spin theories. For example, the rigid rotator with 3 degrees of freedom is a theory with half-odd spin, but the dipole rotator with 2 degrees of freedom is not. This distinction also reduces to a homotopy calculation. Again, suitably nonlinear field theories are found to belong to the half-odd-spin type.[3] Indeed, that part of the total angular momentum of a field which is generally called the orbital angular momentum is itself found to possess half-odd-integer eigenvalues, an attribute usually supposed to be reserved for the other part of the angular momentum, the intrinsic. It is for this reason that we have refined the usual terminology, slightly reserving the words *half-odd spin* merely for the phenomenon of a sign change under 2π rotation, and calling the two parts of the total angular momentum $\mathbf{J}$ of a field *intrinsic* and *extrinsic*: $\mathbf{J} = \mathbf{J}^i + \mathbf{J}^e$.

It is possible to study the exchange of two field structures by the same methods that were used to study the 2π rotation of one structure. We ask whether a field theory admits states which are odd under exchange just as under 2π rotation, setting up an *odd/even* classification of field theories that parallels the previous *half-odd-spin/integer-spin* classification. We show that the existence of kinks is necessary (but not sufficient) for a multivalued quantized field with integer-valued intrinsic angular momentum to be of the half-odd-spin type or of the odd (Fermi–Dirac)-statistics type.

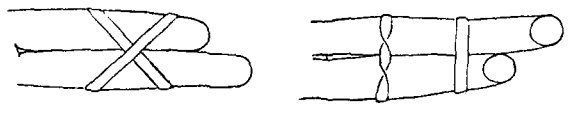


FIG. 1. The first rubber-band lemma. This relates a rotation and an exchange. See text.

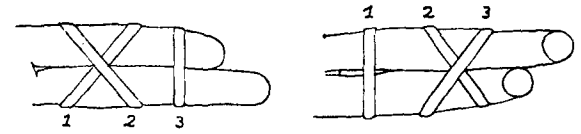


FIG. 2. The second rubber-band lemma. This relates two exchanges. See text.

Sometimes we call our work "rubber-band physics." In the first place, the simplest model of a kink is a 2π twist in a rubber band. In the second place, the topological calculations leading to the spin–statistics result are not completely transparent, and a heurism that we found indispensable is the following rubber-band lemma:

If a rubber band is wrapped twice about a rod, it exhibits at least two deformities: a self-crossing and a 2π twist.

We suggest the reader perform the experiment. If one uses a finger or two for the rod, with a little work he can get the self-crossing on the palm side, and the 2π twist concentrated in one of the two parallel strands on the back side (see Fig. 1). The crossing serves as a graph of an exchange process in which the contents of two regions of space are interchanged by a continuous deformation. (The angle around the rod is the deformation parameter.) The twist is a graph of a kind of 2π rotation of one of the two regions. The triviality of the rubber band convinced us that these two processes are homotopic, and a complete formalization and proof of this result makes up much of this paper.

The general relation proven is not between the total J and exchange, but between J^e and exchange. Just in the case of integer j^i, this coincides with the observed spin–statistics relation.

The nonexistence of parastatistics in multivalued quantized theories became clear from a second rubber-band lemma:

A rubber band can be wrapped three times around a rod with no twists and two crossings.

As graphs of exchanges among three regions 1, 2, 3 (see Fig. 2), this demonstrates that the exchange 1–2 and the exchange 2–3 are homotopic. Since parastatistics is characterized by the existence of nonequivalent exchanges, we inferred that parastatistics would be excluded. The rigorous proof given of this fact is not based on the second rubber-band lemma, however, for a simpler proof was found. But we found the second rubber-band lemma heuristically useful in another connection (see Fig. 11).

The paper is organized as follows. In Sec. I, the theory of kinks is presented. Since this is not the first exposition, the presentation is compressed and formal In Sec. II, a similar presentation is given of the idea of

multivalued quantization of field theories in general, and field theories with kinks in particular. In Sec. III, the spin concept is defined and the topological criterion for half-odd spin is formulated. Section IV deals with exchange, and the main feature is that there is a close step-for-step correspondence between the spin treatment of Sec. III and the exchange treatment of Sec. IV. Then the connection between spin, exchange, and kinks is formulated and proven in Sec. V and in the Appendices, to which some useful but noxious lemmas are relegated.

The results described here still permit us to entertain the hypothesis that the stable elementary particles are kinks in a nonlinear field, in the sense that their "particle numbers" are the numbers of kinks present in the quantum. According to this hypothesis, the linear or almost linear spinor fields usually employed to describe half-odd-spin quanta are not fundamental, but are useful "point-limit" approximations to operators creating or annihilating excitations in a nonlinear field of particular internal states. It is regrettable that the dynamical calculations required to try this hypothesis more severely are so difficult and so likely to diverge, but we are not abandoning the question yet. We suspect in particular that the present methods can be used to construct, from the underlying commutative-field creation, operators that anticommute or commute appropriately, linking our exchange statistics to the usual field-theoretic concept of statistics. When quantum mechanics was discovered, the obstinate dualism of matter and field, particle and wave, at last dissolved into the unity of the quantum, only to be reincarnated subsequently in the distinction between Bose–Einstein and Fermi–Dirac quanta with their two different quantization recipes. The present work suggests that it is finally possible to unify these two into one nonlinear field, possessing two kinds of statistics, as a direct consequence of the topology of the rotation group in three dimensions.

I. CLASSICAL FIELDS

1. A *classical field* $\varphi(x)$ on a flat (Minkowskian) space–time $\{x\}$ is a continuous mapping of $\{x\}$ into some topological space Φ. We shall assume Φ to be a connected[4] manifold, so that the value φ can be locally represented by n real numbers φ_α, and any two points φ_0, φ_1 in Φ can be joined by a path in Φ, i.e., a continuous mapping $f: I \to \Phi$ of the unit interval into Φ with $f(0) = \varphi_0, f(1) = \varphi_1$.

2. Let $X = \{\mathbf{x}\} = \{x \mid t = \text{const}\}$ be ordinary 3-dimensional space and let Φ^X be the set of all continuous mappings $\varphi: X \to \Phi$, and $\Phi^X(\varphi_0)$ the subset of Φ^X with $\varphi(\mathbf{x}) \to \varphi_0$ as $|\mathbf{x}| \to \infty$, φ_0 an arbitrary fixed point of Φ. The field $\varphi(x)|_{t=\text{const}} \equiv \varphi(\mathbf{x}, t)|$ is supposed to belong to $\Phi^X(\varphi_0)$ at any time its continuous evolution being determined by classical field equations and appropriate initial conditions.

3. We suppose that $\varphi(\mathbf{x}, t)$ has a single-valued law of transformation under the Poincaré group. This characteristic will be retained when we quantize the field, making our approach basically different from the usual ones, in which spin-$\frac{1}{2}$ particles are represented by double-valued fields.

4. Two fields $\varphi_1(\mathbf{x})$, $\varphi_2(\mathbf{x})$ in $\Phi^X(\varphi_0)$ are said to be homotopic to each other, $\varphi_1(\mathbf{x}) \sim \varphi_2(\mathbf{x})$, if there is a continuous mapping $\varphi(\mathbf{x}, u): X \otimes I \to \Phi$ such that $\lambda(\mathbf{x}, u)$ is in $\Phi^X(\varphi_0)$ for any given u in I, and

$$\varphi(\mathbf{x}, 0) = \varphi_1(\mathbf{x}),$$
$$\varphi(\mathbf{x}, 1) = \varphi_2(\mathbf{x});$$

$\varphi(\mathbf{x}, u)$ is called a homotopy between $\varphi_1(\mathbf{x})$ and $\varphi_2(\mathbf{x})$; "$\sim$" is an equivalence relation, and its equivalence classes are called the homotopy classes of $\Phi^X(\varphi_0)$.

5. Let us denote $\Phi^X(\varphi_0)$ by Q, and a general element $\varphi(\mathbf{x})$ in Q by q. We shall make of Q a topological space by introducing in it the compact-open topology (Ref. 5, p. 73), which coincides with the metric topology on Q, $d(q, q') = \max_{x \in X} d[\varphi(x) - \varphi'(x)]$, when X is compact and Φ metric (Ref. 5, p. 102). Q may or may not be connected, and its connected components Q_n are just the homotopy classes defined in Sec. I, part 4; a homotopy is a path $q(u)$ within some Q_n, and so is, by Sec. I, part 4, a time evolution $\varphi(x) = \varphi(\mathbf{x}, t)$, showing that dynamical variables that depend only on the homotopy class of $\varphi(\mathbf{x})$ are constants of motion. If there is more than one component Q_n we say that the field *admits kinks.*

6. Kinks form a group, in the following sense. The components Q_n are the elements of the group. We shall denote by $\pi_3(\Phi, \varphi_0)$ the set of all the Q_n. Starting from any $\varphi_1(\mathbf{x})$ in Q_1 and $\varphi_2(\mathbf{x})$ in Q_2 we may construct, through homotopies if necessary, representatives $\varphi_1'(\mathbf{x})$ in Q_1 and $\varphi_2'(\mathbf{x})$ in Q_2 such that

$$\varphi_1'(\mathbf{x}) = \varphi_0, \quad x_3 \geq 0,$$
$$\varphi_2'(\mathbf{x}) = \varphi_0, \quad x_3 \leq 0;$$

we define

$$Q_3 = Q_1 + Q_2 \tag{1}$$

as the homotopy class of

$$\varphi_3(\mathbf{x}) = \varphi_1'(\mathbf{x}), \quad x_3 \leq 0$$
$$= \varphi_2'(\mathbf{x}), \quad x_3 \geq 0. \tag{2}$$

[4] By "connected" we shall mean what is usually known as "arcwise connected" or "path connected"; except for this, we follow the standard terminology of topology.

[5] S. T. Hu, *Homotopy Theory* (Academic Press Inc., New York, 1959).

Under the operation (1) $\pi_3(\Phi, \varphi_0)$ is an Abelian group (Ref. 5, p. 109), the third homotopy of the pair $(\Phi, \varphi \in \Phi)$. It can be shown (Ref. 5, p. 126) that for any φ_0, φ_0' in Φ (connected), $\pi_3(\Phi, \varphi_0) \approx \pi_3(\Phi, \varphi_0')$. Accordingly, when concerned with only the group structure of $\pi_3(\Phi, \varphi_0)$, we shall denote this group by $\pi_3(\Phi)$. $\varphi_0(\mathbf{x}) = \varphi_0$, for all $\mathbf{x}$ belongs to the identity of $\pi_3(\Phi, \varphi_0)$, and if $\varphi(\mathbf{x})$ is in Q_n, then $\varphi(-\mathbf{x})$ is in Q_n^{-1} (or $-Q_n$, in the additive notation).

7. Since $\pi_3(\Phi)$ is Abelian, if it is finitely generated (Ref. 5, X, Corollary 8.3) its elements can be labeled by a set of numbers

$$(n_1, \cdots, n_k, n_{k+1}, \cdots, n_l) \equiv n, \tag{3}$$

such that $n_1, \cdots, n_k$ range over the integers and $n_{k+1}, \cdots, n_l$ range over the integers modulo

$$r_i \quad (i = k+1, \cdots, l),$$

r_i being a factor of r_{i+1}. The l components of n are conserved (Sec. I.5) and additive, in the following sense: if a field $\varphi_1(\mathbf{x})$ in $Q_n(1)$ is juxtaposed to $\varphi_2(\mathbf{x})$ in $Q_n(2)$ [see (2)], the resulting field $\varphi_3(\mathbf{x})$ belongs to the class $Q_{n(3)}$, with

$$n^{(3)} = (n_1^{(1)} + n_1^{(2)}, \cdots, n_l^{(1)} + n_l^{(2)}) \equiv n^{(1)} + n^{(2)}; \tag{4}$$

the n_i $(i = 1, \cdots, l)$ can then be considered as particle numbers or "charges." The possibility of having "modulo" conservations (the n_i's for $i > k$) has arisen also in different context,[6] but is not supported by experimental evidence so far. From now on we shall discuss only Φ's for which $l = k$ in (3), but in general our results are valid. We shall say that $\varphi(\mathbf{x})$ is one "kink of type i" if $\varphi(\mathbf{x})$ is in Q_{N_i}, where

$$N_i = (0, 0, \cdots, 1, \cdots, 0), \tag{5}$$

1 being in the ith place; then $\varphi(-\mathbf{x})$ is in Q_{-N_i} and will be called an antikink of type i. Any field $\varphi(\mathbf{x})$ in Q_n with n given by (3) can be considered as composed of a certain number of kinks or antikinks of the various types N_i.

8. In the literature the nth homotopy group of Φ is variously defined as the set of homotopy classes of mappings $(S^n, p_0) \to (\Phi, \varphi_0)$ of an n sphere S^n into Φ, with $p_0(\epsilon S^n) \to \varphi_0$, of mappings $(I^n, \partial I^n) \to (\Phi, \varphi_0)$ of an n cube I^n into Φ, with its boundary ∂I^n going into φ_0, etc. All those definitions are equivalent among themselves (this is proven by showing that after proper identifications the domains of the mappings are homeomorphic) and to ours for $n = 3$: we can, for instance, deform X into a 3-ball: $\mathbf{x} \to \mathbf{x}/1 + |\mathbf{x}|$, which is homeomorphic to a 3-sphere upon identification of its boundary $|\mathbf{x}| = 1$ into one point p_0. Notice that this is allowed by the boundary conditions introduced in Sec. I.2.

[6] See, for instance, M. Gell-Man, Phys. Letters **8**, 214 (1964).

TABLE I. Homotopy groups for the underlying manifolds of classical Lie groups.

	SO_3	SO_4	SO_5	SO_6	$SO_{n\geq 7}$	SU_2	$SU_{n\geq 3}$	$Sp_{n\geq 1}$
π_1	Z_2	Z_2	Z_2	Z_2	Z_2	0	0	0
π_2	0	0	0	0	0	0	0	0
π_3	Z	$Z+Z$	Z	Z	Z	Z	Z	Z
π_4	Z_2	Z_2+Z_2	Z_2	0	0	Z_2	0	Z_2
π_5	Z_2	Z_2+Z_2	Z_2	Z	0	Z_2	Z	Z_2

9. Any Abelian group π_3 that may be dictated by phenomenological considerations can be realized as the third homotopy group of some space Φ. Moreover, it can be shown (Ref. 5, p. 169) that given a sequence of groups

$$\pi_1, \pi_2, \cdots, \pi_n, \cdots,$$

such that all except possibly the first are Abelian and π_1 is a group of automorphisms of all the others, there exists a connected space Φ such that

$$\pi_n(\Phi) \approx \pi_n \quad \text{for all } n.$$

In particular, there are $C = 2^{\aleph_0}$ nonhomeomorphic spaces with the same π_3!

10. Some homotopy groups of spheres are tabulated in Ref. 3, Table I , and many more are given by Toda.[7] Table I shows the first five homotopy groups for the underlying manifolds of classical Lie groups; it has been compiled from several sources.[7]

The relation $\pi_n(\Phi \otimes \Psi) = \pi_n(\Phi) \oplus \pi_n(\Psi)$ is a useful tool for the construction of *ad hoc* spaces (Sec. I.9).

II. QUANTIZATION

1. When a classical field theory is quantized, the statement "the measurement of the field at time t will give the values $\varphi(\mathbf{x}, t)$" is replaced by "the probability amplitude for getting the result $\varphi(\mathbf{x})$ when measuring the field at time t is $\Psi[\varphi(\mathbf{x})](t)$," Ψ being a complex-valued time-dependent functional of $\varphi(\mathbf{x})$, continuous in t and φ. Notice that in this picture (Schrödinger) and representation [diagonal in $\varphi(\mathbf{x})$], the argument of the functional is a c-number field, an element of the function-set Φ^X of Sec. I.2.

2. We shall assume that, at any time,

$$\Psi[\varphi(\mathbf{x})](t) = 0 \quad \text{if} \quad \varphi(\mathbf{x}) \notin \Phi^X(\varphi_0) \equiv Q, \tag{6}$$

so we can use $Q = \Phi^X(\varphi_0)$ instead of Φ^X as our configuration space; this implies restrictions on the dynamics of the system, but we shall not elaborate

[7] H. Toda, *Composition Methods in Homotopy Groups of Spheres* (Princeton University Press, Princeton, N.J., 1962); A. Borel, in *Seminaine H. Cartan* (2) (Ecole Normale Superieure, Paris, 1956). L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, N.J., 1946); V. G. Boltyanskii, Transl. Am. Math. Soc. 7, 135 (1957).

this point in the present work. Since for φ in Q we can write $\Psi[\varphi](t) = \Psi(q, t)$, a continuous complex-valued function on $Q \otimes R$ (R = real number space spanned by t), another way of putting (6) is

$$\int_Q |\Psi(q, t)|^2 \, dq = 1, \tag{7}$$

provided that the concept of integral on Q can be defined; even if not, we shall use expressions of the type in (7) as symbols for expressions of the type in (6). We call Ψ the state function(al).

3. If we accept the Feynman scheme for quantization, which uses integrals over continuous histories $\varphi(\mathbf{x}, t)$ (in our case paths in Q), it can be shown (Ref. 3, Sec. III.2) that

$$\int_{Q_n} |\Psi(q, 0)|^2 \, dq = 1 \quad \text{implies} \quad \int_{Q_n} |\Psi(q, t)|^2 \, dq = 1; \tag{8}$$

then, if the "charges" defined in (Sec. I.7) are interpreted as nonlocal observables in the quantum theory, (8) tells us that they are conserved in time. In the following, whenever we talk about a homotopic conservation law it shall be understood in the quantum sense given by (8). In short, (6) and (8) allow us to extend the results of Sec. I to a quantum theory of φ in which $\varphi(\mathbf{x})$ can still be treated as having a c-number field of eigenvalues. We assume for simplicity that the particle numbers n_i obey a superselection rule,[8] i.e., that the support of any realizable physical state $\Psi(q)$ is one of the Q_n.

4. In what follows we shall be concerned with two kinds of discrete operations that leave physical systems unchanged but may multiply the state vectors by -1: a 2π rotation and the exchange of two identical subsystems. The 2π rotation can be realized continuously by a succession of infinitesimal transformations in an obvious way, and we shall give in Sec. IV an analogous realization for an operation related to the exchange.

Let $q \to F_s q$ represent a flow in Q, i.e., a one-parameter family of 1–1 mappings F_s satisfying the conditions

$$F_0 q = q, \tag{9}$$

so that $F_s q$ $(0 \leq s \leq 1)$ is a path at q. Any such flow in Q induces a continuous one-parameter family of linear transformations of the state function $\Psi(q)$:

$$\Psi(q) \to \Psi(q, s) = \Psi(F_s^{-1} q) = \Psi(q_s), \quad 0 \leq s \leq 1. \tag{10}$$

If

$$F_1 q = q, \tag{11}$$

the flow is called *closed*. Its paths are then closed paths or *loops*. (11) is the mathematical expression of the "leaving the system unchanged" used above. Since until now we have considered Ψ to be single-valued, (9) and (11) imply

$$\Psi(F_0 q) = \Psi(F_1 q). \tag{12}$$

Now we wish to consider that the state function $\Psi(q)$ might be multiple-valued; this point is elaborated upon in Sec. II.5, and (12) no longer holds.

5. Our first problem is to find whether the domain Q_n of the state function $\Psi(q)$ admits multiple-valued continuous functions. Intuitively speaking, these have more than one value at each q in Q_n (the value set changing continuously with q), such that any two values Ψ_0 and Ψ_1 at q can be connected by a continuous succession of values $\Psi[q(s)]$ by traveling some closed path $q(s)[q(0) = q(1) = q]$. It is evident that no simply connected Q_n admits such functions.[9] Moreover, since the value of a multivalued function depends on the point q and the way q is reached from some standard point, such a function on Q_n can be defined as a (single-valued) function on CQ_n, the universal covering space of Q.

We define CQ_n as follows:

We choose a base point q_{n_0} in Q_n and consider the paths $q(s)$ in Q_n $(0 \leq s \leq 1)$ such that $q(0) = q_{n_0}$; two paths $q(s)$ and $q'(s)$ are equivalent if $q(1) = q'(1) = q$ (say) and there exists a homotopy relative to $\{0, 1\}$ between $q(s)$ and $q'(s)$, i.e., a continuous mapping $q(s, u)$: $I^2 \to Q_n$ $(0 \leq s \leq 1,\ 0 \leq u \leq 1)$ such that

$$q(s, 0) = q(s), q(s, 1) = q'(s),$$
$$q(0, u) = q_{n_0}, q(1, u) = q.$$

The equivalence classes of paths $q(s)$ are the points of CQ_n, if we introduce the following topology: given $q(s)$ and an arcwise-connected open set U that contains $q(1)$, the union over U of the equivalence classes of the $q'(s)$ such that

$$q'(s) = q(2s) \quad (0 \leq s \leq \tfrac{1}{2}),$$
$$q'(s) \subset U \quad (\tfrac{1}{2} \leq s \leq 1),$$

is a set belonging to CQ_n; the collection of all such sets is taken to be a basis for the topology of CQ_n.

We now define a multivalued function on a space Q to be a single-valued function on the covering space CQ. In what follows, such functions will further be required to be continuous unless otherwise mentioned.

[8] G. C. Wick, Phys. Rev. **88**, 101 (1952).

[9] We remark that we are talking about state functions, i.e., complex-valued functions on Q. The statement is of course not true for the spinor wavefunction of ordinary quantum mechanics; indeed, its domain is simply connected, but its components do not satisfy the transformation law (10) when subject to rotations.

Q_n admits multivalued functions if and only if CQ_n is not homeomorphic to Q_n, which is equivalent to $\pi_1(Q_n) \neq 0$.

6. For any connected space Y, we shall write $\tilde{y}, \tilde{y}', \cdots$ for *different* points of CY associated with the end points $y \in Y$ and $\tilde{y}(s)$ for a representative path of the class $\tilde{y}$. We also say that $\tilde{y}, \tilde{y}', \cdots$ "cover" y, and denote by Cy the set of all the covering points of y; it is easy to see that the number of elements of Cy is constant over Y. In the following a multivalued function Ψ on Q_n shall be written as a function on CQ_n:

$$\Psi = \Psi(\tilde{q}). \tag{13}$$

7. A group that acts on Q cannot in general be defined to act on CQ continuously. For example, the action of the rotation group on spinors is not uniquely definable. However, the action of a flow on Q uniquely defines its action on CQ. If $F_s q$ is a flow on Q, then $F_s\tilde{q}$ can be defined in a natural way for each $s = s_0$ as the equivalence class of a path that consists of a path representing $\tilde{q}$ followed by the path $F_s q$, $0 \leq s \leq s_0$. Accordingly a flow on Q acts also on $\tilde{Q}$, and therefore on state-functions $\Psi(\tilde{q})$ in such a way that the value of Ψ is invariant:

$$F_s[\Psi(\tilde{q})] = \Psi(F_s^{-1}\tilde{q}) \equiv \Psi_s(\tilde{q}), \tag{14}$$

or

$$\Psi_s(\tilde{q}_s) = \Psi(\tilde{q})$$

where $\tilde{q}_s = F_s\tilde{q}$.

8. Computation of $\pi_1(Q_n)$. For Q_0, the component of $Q = \Phi^X(\varphi_0)$ that contains the field $\varphi_0(\mathbf{x}) \equiv \varphi_0$, we have (Ref. 3, Sec. V.2)

$$\pi_1(Q_0) \approx \pi_4(\Phi), \tag{15}$$

which formally solves the problem for Q_0, or at least reduces it to a standard one. Once $\pi_4(\Phi)$ is known, the problem is completely solved because it can be shown[10] that for any two connected components

$$Q_j, Q_k \text{ of } Q = \Phi^X(\varphi_0), \ \pi_1(Q_j) \approx \pi_1(Q_k),$$

so

$$\pi_1(Q_n) \approx \pi_4(\Phi) \quad \text{for all } n. \tag{16}$$

Our main interest in this result is that if $\pi_4(\Phi) \neq 0$, all Q_n admit multivalued functions; otherwise, none does.

9. Assuming that (13) is properly multivalued,

$$\Psi(\tilde{q}) \neq \Psi(\tilde{q}'), \tag{17}$$

in order to have a theory of the type described in Sec. II.4 we still have to show that the multivaluedness of $\Psi(q)$ is realized by two special flows q_s, namely, the 2π rotation of a field $\varphi(\mathbf{x})$ (Sec. III) and the "exchange" of two fields identical to $\varphi(\mathbf{x})$, a concept to be defined in Sec. IV. These are in principle two independent problems, and it is one of the main purposes of this paper to prove that their solutions are interdependent, and that, in a sense specified in Sec. V, an affirmative (negative) answer for any one of them implies an affirmative (negative) answer to the other. In this sense, the "spin" of a system determines its "statistics" and vice versa. As we shall see (Sec. V), our proofs require only continuity of the fields $\varphi(\mathbf{x})$ and of the state functionals (13) with $q = [\varphi(\mathbf{x})]$.

[10] G. W. Whitehead, Ann. Math. **47**, 460, statement 2.6 (1946). We are indebted to Professor S. T. Hu for this reference.

III. SPIN

1. The angular momentum of a field theory is defined as the generator of infinitesimal rotations

$$\delta\mathbf{x} = \delta\boldsymbol{\theta} \times \mathbf{x},$$

where $\delta\boldsymbol{\theta}$ gives the axis and angle of rotation. The change in the field resulting from this rotation may be written as

$$\delta\varphi = \frac{\partial\varphi}{\partial\mathbf{x}} \cdot \delta\mathbf{x} + \left.\frac{\partial\varphi}{\partial\boldsymbol{\theta}}\right|_{\theta=0} \cdot \delta\boldsymbol{\theta} \equiv \delta^e\varnothing + \delta^i\varnothing,$$

where the first term gives the "extrinsic" change due to the spatial variation of the φ field, and the second term is due to the "intrinsic" law of transformation

$$\boldsymbol{\theta}\colon \varphi \to \varphi' = \varphi'(\varphi, \boldsymbol{\theta})$$

of the φ field under a rotation parametrized by a vector $\boldsymbol{\theta}$. The corresponding decomposition of the generator of this rotation is written

$$\mathbf{J} = \mathbf{J}^e + \mathbf{J}^i.$$

The field theory exhibits half-odd-integer angular momentum if and only if the operator of a 2π rotation,

$$W = e^{2\pi i J_z},$$

possesses eigenvalues -1 as well as $+1$.

For brevity, half-odd-integer angular momentum will be referred to as *half-odd spin*. If a component Q_n of Q supports an eigenfunction Ψ of $W = -1$, we say Q_n (and the theory) admits half-odd spin.

2. In familiar field theories, J^e has integer eigenvalues and does not contribute to W,

$$W^e = \exp(2\pi i J_z^e) = 1.$$

Therefore half-odd angular momentum is usually attributed entirely to the transformation law of the field

$$W \approx \exp(2\pi i J_z^i).$$

In quantum-field theory, the generators $\mathbf{J}$ may be defined in terms of the transformation of the state functional of the field, rather than the field itself. If multivalued state functionals are admitted, then we have shown there exist quantum-field theories for which $\mathbf{J}^e$ contributes to W: While the field itself does not change sign under a 2π rotation, the quantum state of the field does. Heuristically, such a spin can

be thought to come from the actual rotation of very asymmetric field structures having some of the properties of a quasirigid rotator. In this kind of theory the familiar decomposition of the angular momentum of the electron,

$$\mathbf{J} = \mathbf{L} + \mathbf{S},$$

is a feature of a phenomenological model and does not correspond to the decomposition

$$\mathbf{J} = \mathbf{J}^e + \mathbf{J}^i;$$

$\mathbf{S}$ itself may result from "orbital" angular momentum $\mathbf{J}^e$ of the kink representing the electron.

3. We now specify the flow $q \to (W^e)_s q$ of (10) for a continuous 2π extrinsic rotation of a field $\varphi(\mathbf{x})$. Let $W^e(s)$ be an extrinsic rotation through an angle $2\pi s$ around some axis through $\mathbf{x} = 0$, acting in the usual way on the argument $\mathbf{x}$ and leaving invariant the value φ,

$$W^e(s)\varphi = \varphi \quad \text{(for all } s \text{ and any axis)}. \tag{18}$$

We then define the flow in Q,

$$q \to (W^e)_s q \equiv W^e(s) \cdot [\varphi(\mathbf{x})] \equiv \varphi_W(\mathbf{x}, s) = \varphi[(W^e)^{-1}(s) \cdot \mathbf{x}] \quad (\theta \leq s \leq 1), \tag{19}$$

and (1) becomes, for rotations,

$$\Psi(q) \to W^e(s) \cdot \Psi(q) \equiv \Psi[(W^e)^{-1}(s) \cdot q] = \Psi[(W^e)_{1-s} q]. \tag{20}$$

The effect of a 2π rotation on a single-valued function $\Psi(\tilde{q})$ is defined by Sec. II.7.

4. Since the $W^e(s)$ (for all s and any axis) constitute the elements of SO_3, the rotation group in 3 dimensions, W^e can be represented as a loop in the SO_3 manifold. SO_3 has a well-known graphical representation as a 3-dimensional ball of radius π with opposite points of the surface identified. A maximum cross section of this construction, including a representative path $W^e(s)$ of W^e, is shown in Fig. 3.

Any deformation of a path described by the $W^e(s)$ induces a deformation in the corresponding path described by the q's in Q, via (19); in particular, if a loop in SO_3 is homotopic relative to its end point[11] to the trivial loop, so is its induced loop $q(s)$ (but the reciprocal is not always true). A well-known example of such a loop is $(W^e)^2$, a 4π rotation, i.e.,

$$(W^e)^2 = 1. \tag{21}$$

The fact that we are dealing with a 3-dimensional theory is crucial for this. The result does not hold for the rotations in 2 dimensions, a 1-dimensional subgroup of SO_3 with the topology of S^1.

[11] By "relative to its end point" we mean that the two loops can be connected by a continuous family of *loops* with a common end point. Unless otherwise stated, whenever two loops (on X, Q, etc.) are said to be homotopic, it will mean homotopic relative to their end point.

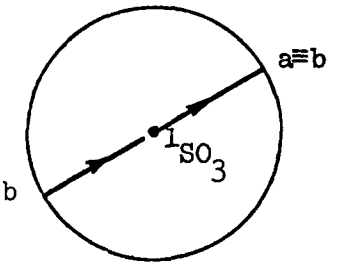


FIG. 3. Path of a 2π rotation in $SO(3, R)$. This path is not homotopic to the identity.

5. Notice that $W = \pm 1$ holds only for W as operator on state functions, whereas for the fields (or rather for their coverings), Eq. (21) only means that

$$W\tilde{q} = \tilde{q}' \Leftrightarrow W\tilde{q}' = \tilde{q}. \tag{22}$$

For Q_n to admit half-odd spin there must be a neighborhood in CQ whose points $\tilde{q}$ obey

$$W \cdot \tilde{q} \neq \tilde{q}; \tag{23}$$

this is a consequence of the following.

6. *Lemma:* Eq. (23) holds or fails to hold, simultaneously for all $\tilde{q} \in CQ_n$.

The proof of this lemma is given in Appendix A. The lemma implies that in order to see whether a given Q_n admits half-odd spin, it is enough to check if (23) holds for *any one* $\tilde{q} \in CQ_n$.

7. In what follows we give a few results helpful in finding whether or not a given Q_n admits half-odd extrinsic spin J^e.

A necessary condition for Q_n to admit half-odd J^e is that $\pi_1(Q_n)$ contain at least one element of order 2. This is an immediate consequence of (21) and (23).

A necessary condition for Q_n to admit half-odd J^e is that no $\varphi(x)$ in Q_n be axisymmetric, since for such a field, $W \cdot \tilde{q} = \tilde{q}$. In particular, Q_0 does not admit half-odd J^e, because $\varphi(\mathbf{x}) = \varphi_0$ is axisymmetric.

In other words, kinks $[\pi_3(\Phi) \neq 0]$ are necessary for half-odd extrinsic spin (Ref. 3, Sec. V.13). More strongly: kinks must be present in a state with half-odd extrinsic spin.

IV. FIELD EXCHANGE

1. In order to define an exchange operation on a field $\varphi(\mathbf{x})$, we have to identify the objects to be exchanged. Since we have neither a procedure for locating particles in a general field $\varphi(\mathbf{x})$ in Q nor the creation operators of the usual quantum-field theories, we shall at first restrict ourselves to "union" fields defined as follows. By the support $\sup \varphi$ of a field $\varphi(x)$, we mean the set of points x at which $\varphi(x) \neq \varphi_0$. $\varphi_1(x)$ and $\varphi_2(x)$ are fields of disjoint supports $\sup \varphi_1$, $\sup \varphi_2$, then their "union,"

$$\varphi = \varphi_1 \cup \varphi_2, \tag{24}$$

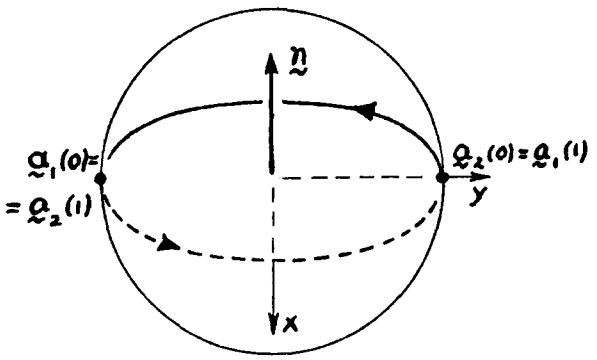


FIG. 4. Path in X of the exchange carried out on two kinks at $\mathbf{a}_1$ and $\mathbf{a}_2$.

is defined by

$$\begin{aligned}\varphi(x) &= \varphi_1(\mathbf{x}), \quad \mathbf{x} \in \sup \varphi_1, \\ &= \varphi_2(\mathbf{x}), \quad \mathbf{x} \in \sup \varphi_2, \\ &= \varphi_0 .\end{aligned} \tag{25}$$

Now let $\varphi_n(x)$ be a kink of type n with $\sup \varphi_n \subset \{|\mathbf{x}| < \epsilon\}$. We take

$$\varphi(\mathbf{x}) = \varphi_n(\mathbf{x} - \mathbf{a}_1) \cup \varphi_n(\mathbf{x} - \mathbf{a}_2), \tag{26}$$

with

$$|\mathbf{a}_1 - \mathbf{a}_2| \geq 2\epsilon. \tag{27}$$

The $\mathbf{a}_i$ are finite, and (25) and (27) guarantee that (26) is well defined, i.e., that the supports of the kinks (of type n) $\varphi_n(\mathbf{x} - \mathbf{a}_i)$ do not overlap.

We can now specify the $q(s)$ of Sec. II.4 for an exchange of the structures at $\mathbf{a}_1$ and $\mathbf{a}_2$. Let $\mathbf{n}$ be unit vector perpendicular to $\mathbf{a}_1 - \mathbf{a}_2$, and (see Fig. 4)

$$\mathbf{a}_i(s) = \tfrac{1}{2}(\mathbf{a}_1 + \mathbf{a}_2) - (-1)^i \exp s\pi\mathbf{n} \times \tfrac{1}{2}(\mathbf{a}_1 - \mathbf{a}_2); \tag{28}$$

here $\mathbf{n}\times$ is the linear operator on vectors $\mathbf{v}$ defined by $(\mathbf{n}\times)\mathbf{v} = \mathbf{n} \times \mathbf{v}$. Then

$$q_n(s) = \varphi_n[\mathbf{x} - \mathbf{a}_1(s)] \cup \varphi_n(\mathbf{x} - \mathbf{a}_2(s)), \quad 0 \leq s \leq 1 \tag{29}$$

has the desired property of being a loop starting and ending at $\varphi(\mathbf{x})$, with the structure originally at $\mathbf{a}_1$ now at $\mathbf{a}_2$, and vice versa (Fig. 4).

2. We now define a closed flow (cf. Sec. II.4) X^m in the function space Q. Physically speaking, X^m makes a clearing at infinity, creates there two kink–antikink pairs of type m, exchanges the two kinks, annihilates the kinks with their antikinks, and restores the clearing to its original form.

Symbolically, we compose X^m out of the following five flows Y_i, in the order

$$X^m = Y_5 Y_4 Y_3 Y_2 Y_1, \tag{30}$$

where Y_1 is a shrinkage of the field to a support of radius ϵ, Y_2 is a creation of two kink–antikink pairs outside the support, Y_3 is the exchange of the two created kinks (see Fig. 5), Y_4 is the annihilation of the two pairs, the inverse of Y_2, and Y_5 is an expansion of the field, the inverse of Y_1. Let

$$\varphi_{-n}(\mathbf{x}) = \varphi_n(-x, y, z), \quad \mathbf{b} = (b, 0, 0), \quad \boldsymbol{\epsilon} = (\epsilon, 0, 0), \quad b > 0.$$

$\mathbf{b}$ and $\boldsymbol{\epsilon}$ are vectors along the x axis of length b and ϵ, and let $\mathbf{n}$ now be a unit vector along the z axis. Let $r = |\mathbf{x}|$, then the Y_i are defined as follows:

$$Y_1: \varphi(\mathbf{x}, s) = \begin{cases} \varphi\left(\dfrac{\mathbf{x}}{1 - \dfrac{s}{s_1\epsilon} r}\right), & r \leq \dfrac{s_1}{s}\epsilon \\ = \varphi_0, & r \geq \dfrac{s_1}{s}\epsilon \end{cases} \quad 0 \leq s \leq s_1,$$

$$Y_2: \varphi(\mathbf{x}, s) \quad s_1 \leq s \leq s_2,$$

$$\begin{aligned} &= \varphi_m\left[\mathbf{x} - \left(\mathbf{b} + \frac{s - s_1}{s_2 - s_1}\boldsymbol{\epsilon}\right)\right], && y \geq b \\ &= \varphi_{-m}\left[\mathbf{x} - \left(\mathbf{b} - \frac{s - s_1}{s_2 - s_1}\boldsymbol{\epsilon}\right)\right], && \epsilon < y \leq b \\ &= \varphi_{-m}\left[\mathbf{x} - \left(-\mathbf{b} + \frac{s - s_1}{s_2 - s_1}\boldsymbol{\epsilon}\right)\right], && -b \leq y < -\epsilon \\ &= \varphi_m\left[\mathbf{x} - \left(-\mathbf{b} - \frac{s - s_1}{s_2 - s_1}\boldsymbol{\epsilon}\right)\right], && y \leq -b \\ &= \varphi(\mathbf{x}, s_1), && -\epsilon \leq y \leq \epsilon \end{aligned}$$

$$Y_3: \varphi(x, s) \quad s_2 \leq s \leq s_3,$$

$$\begin{aligned} &= \varphi_m\left[\mathbf{x} - \exp\left(2\pi \frac{s - s_2}{s_3 - s_2}\mathbf{n}\times\right) \times (\mathbf{b} + \boldsymbol{\epsilon})\right] \\ &\cup \varphi_m\left[\mathbf{x} + \exp\left(2\pi \frac{s - s_2}{s_3 - s_2}\mathbf{n}\times\right) \times (\mathbf{b} + \boldsymbol{\epsilon})\right], && r \geq b \\ &= \varphi(\mathbf{x}, s_2), && r \leq b \end{aligned}$$

and

$$Y_5 Y_4: \varphi(\mathbf{x}, s) = \varphi\left(\mathbf{x}, s_2 \frac{1 - s}{1 - s_3}\right) \quad s_3 \leq s \leq 1.$$

Notice that $\varphi(\mathbf{x}, s)$ is continuous at $s = s_3$ because $\varphi(\mathbf{x}, s_2) = \varphi(\mathbf{x}, s_3)$; also, during $Y_3\varphi(\mathbf{x}, s)$ is precisely the loop $q_m(s)$ of Eq. (29) performed upon the two

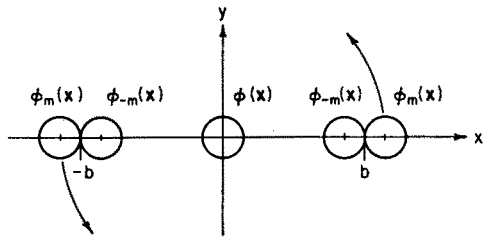


FIG. 5. Exchange operator. The field $\varphi(x)$ is shrunk into a finite sphere to leave "working space." Two kink-antikink pairs of the same kind are created in the working space. The two kinks are exchanged. The new pairs are allowed to annihilate, and the field $\varphi(x)$ expands to refill the now empty working space.

kinks φ_m created in the step Y_2, with $\mathbf{a}_2 = \mathbf{b} + \boldsymbol{\epsilon} = -\mathbf{a}_1$.

3. If the state $\varphi(\mathbf{x})$ already contains kinks of type m, as say a union field (Sec. IV.1), the exchange of two of these kinks is homotopic to $X^m\varphi(\mathbf{x})$. This is proven in Sec. V.3.

4. A double exchange of two m kinks, i.e., $(X^m)^2$, is trivial:

$$(X^m)^2 = 1. \tag{31}$$

Proof: Any deformation of a trajectory described by the $\mathbf{a}$'s induces a deformation (homotopy) in the corresponding path described by q via (29); in particular, if a set of trajectories of the $\mathbf{a}$'s is homotopic relative to its end points to the trivial trajectory $[\mathbf{a}_i(s) = \mathbf{a}_i$ for all $s]$, so is its induced loop $q(s)$.

In Fig. 6(a) we show, slightly displaced in order to distinguish them from one another, the trajectories $\mathbf{a}_1(s)$ corresponding to the loop $q_m(s)$ given by (28) and (29) with s replaced by $2s$ in the exponent, i.e., a representative loop of $(X^m)^2$.

Figure 6(b) is an intermediate stage of a deformation of the trajectories of Fig. 6(a) into the trivial trajectories of Fig. 6(c). If at every stage of the deformation $\mathbf{a}_1(s)$ and $\mathbf{a}_2(s)$ move on their respective circumferences with uniform speed, the distance between them is always $|\mathbf{a}_1 - \mathbf{a}_2|$. Since the supports of $\varphi_1[\mathbf{x} - \mathbf{a}_1(s)]$ and $\varphi_2[\mathbf{x} - \mathbf{a}_2(s)]$ do not overlap initially, they do not overlap during the process. Figure 6 constitutes a proof of (31). In symbols, let

$$\mathbf{n}(t) = [(1-t)\mathbf{n} + t(\mathbf{a}_1 - \mathbf{a}_2)]/|[\ ]| \tag{32}$$

(the denominator is the magnitude of the numerator). Then

$$\mathbf{a}_i(s,t) = \tfrac{1}{2}(\mathbf{a}_1 + \mathbf{a}_2) - (-1)^i \times \exp\,[2\pi s\mathbf{n}(t)\times]\ \tfrac{1}{2}(\mathbf{a}_1 - \mathbf{a}_2) \tag{33}$$

satisfies

$$\mathbf{a}_i(s,0) = \mathbf{a}_i(2s),\ \mathbf{a}_i(s,1) = \mathbf{a}_i. \quad \text{Q.E.D.} \tag{34}$$

Again, our proof of (31) is valid only in 3-dimensional theories! In Appendix B we show that,in general, $(X^m)^2 \neq 1$ in two dimensions. This, together with the other similarities between X and W, suggests the possibility of a relationship between W and X; the existence and nature of such a relation is the subject matter of Sec. V.

5. It is simple to prove that the homotopic triviality of the flow $X^m\varphi(\mathbf{x})$ depends only on m and not on $\varphi(\mathbf{x})$.

6. When

$$X^m\tilde{q} \neq \tilde{q} \tag{35}$$

is obtained for X^m acting as defined in Secs. II.7 and IV.2, we say that the theory, and in particular Q_m (the component of Q that contains the kink being exchanged), admit negative field exchange, or "odd statistics" for short. The reason for the term "negative" or "odd" is that (31) and (35) imply that Q admits state functions $\Psi(\tilde{q})$ with the property

$$X^m\Psi(\tilde{q}) = -\Psi(\tilde{q}). \tag{36}$$

Notice that $X^m = \pm 1$ holds only as an eigenvalue equation for X^m as an operator on state functions, whereas for fields (or rather for their coverings), Eq. (31) means only that

$$X^m\tilde{q} = \tilde{q}' \Leftrightarrow X^m\tilde{q}' = \tilde{q}. \tag{37}$$

7. As in Sec. III, we can test whether Q_m admits odd statistics by checking if

$$X^m\tilde{q} \neq \tilde{q} \tag{38}$$

is obtained for any one $\tilde{q} \in Q$, as shown by the following.

Lemma: Equation (38) holds or does not hold simultaneously for all $\tilde{q} \in Q$.

Proof: Exact analog of the proof of Lemma 3.

In other words, this lemma says the parity of a structure under field exchange depends only on the component of Q_m to which it belongs, and not on the surrounding structures.

8. A single kink of type m will admit both odd and even exchange states if it admits odd exchange states.

The conserved baryon and lepton numbers do not behave this way: for example, a single baryon can be a fermion but not a boson, in that there is a fermion but no boson with

$$N_B = 1,\ N_e = N_\mu = 0. \tag{39}$$

Why? Or does the kink model lead us to expect an

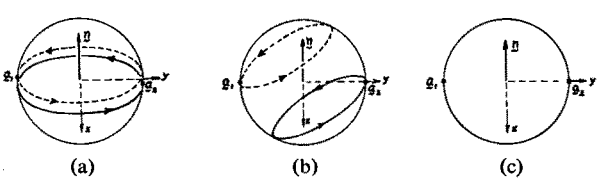


FIG. 6. (a) Path in X of an iterated exchange and (b) its deformation to (c) the identity.

excited state of the nucleon with the quantum numbers (39) but with integer spin and Bose statistics?

Our answer consists in two topological laws that combine to prevent kinks from ever exhibiting both of the statistics they may admit in principle: the invariance of statistics with respect to (a) the partition of the system into subsystems and (b) the passage of time.

In a universe of kinks of type m, the measurement of the exchange operator for any pair may give the value $+1$ or -1 in principle. However, if the result is -1, then an immediate measurement of the exchange operator for any other pair of kinks of type m will give the same result -1 by (a), and any subsequent measurement of this exchange will continue to give the result -1 by (b). Thus the statistics will appear to be a universal and permanent attribute of the type of kink. Let us now state more carefully and prove the topological laws mentioned here:

(a) Consider an eigenstate Ψ such that φ_{op} is a union field,

$$\varphi_{\text{op}}\Psi = \cup\, \varphi_x(x)\Psi,$$

with $\varphi_1, \cdots, \varphi_N$ kinks of the same type m. Then the $N(N-1)/2$ exchanges $\varphi_i \leftrightarrows \varphi_j$, as continuous processes, are all homotopic to each other, and to the effect of the exchange operator X^m (defined in Sec. IV.2). This is a consequence of Sec. IV.3.

(b) The exchange operator X^m is a constant of the motion. Indeed, the state functional $\Psi[\varphi](t)$ is continuous in (φ, t) (Sec. II.1), so when φ undergoes a continuous closed deformation $\varphi \to \varphi(s)$ (in our case an exchange of two type m kinks) we get a function $\Psi[\varphi(s)](t)$ of s and t, continuous in (s, t), with the restriction [for $\varphi \to X^m(s)\varphi$] that

$$\Psi[\varphi(0)](t) = \pm\Psi[\varphi(1)](t), \quad \text{all } t. \tag{40}$$

But $\Psi[\varphi(1)](t)$ is continuous in t, ergo the sign is $+$ for all t or $-$ for all t.

9. Parastatistics. As our terminology suggests, we intend to interpret the operator X^m as the "performer" of the exchange of certain real particles. Accepting provisionally this meaning of X^m, the relation

$$(X^m)^2 = 1 \tag{31}$$

implies that *parastatistics are forbidden within the present theory.*

Ordinary quantum field theory does not exclude the possibility of parastatistics, and some recent work has been done on the subject. However, paraparticles seem to be absent from nature,[12] and it is suggestive that they do not even arise as a possibility in theories of the kind defined here.

[12] H. S. Green, Phys. Rev. **90**, 270 (1953); A. M. L. Messiah and O. W. Greenberg, Phys. Rev. **136**, B248 (1964).

V. CONNECTIONS BETWEEN "SPIN," "STATISTICS," AND NUMBER OF KINKS

1. *Theorem:* Q_m admits half-odd extrinsic spin if and only if it admits odd statistics.

The proof involves some rather specialized topological calculations and comprises Sec. V.2 and Appendix C. The reader willing to accept the theorem can pass on to Sec. V.3.

2. *Proof:* (a) Let $O \cdot \varphi(\mathbf{x})$ stand for the trivial loop in Q starting and ending at $\varphi(\mathbf{x})$; our theorem then reads [see (19) and Sec. IV.2]

$$X^m(s) \cdot [\varphi(\mathbf{x})] \equiv \varphi_X(\mathbf{x}, s) \quad (\text{say}) \sim O \cdot \varphi(\mathbf{x}), \tag{41}$$

if and only if

$$W^e(s) \cdot [\varphi_1(\mathbf{x})] \equiv \varphi_{1W}(\mathbf{x}, s) \quad (\text{say}) \sim O \cdot \varphi_1(\mathbf{x}), \quad \varphi_1(\mathbf{x}) \in Q_m. \tag{42}$$

(b) We may express (42) more explicitly. Let X stand for the Euclidean space spanned by $\mathbf{x}$, and I for the unit interval. Then (42) means that there exists a continuous function $\varphi(\mathbf{x}, s, t)$ defined on $X \otimes I \otimes I$, such that

$$\begin{aligned} &\varphi(\boldsymbol{\infty}, s, t) = \varphi_0, \\ &\varphi(\mathbf{x}, s, 0) = \varphi_X(\mathbf{x}, s), \\ &\varphi(\mathbf{x}, s, 1) = 0 \cdot \varphi(\mathbf{x}), \\ &\varphi(\mathbf{x}, 0, t) = \varphi(\mathbf{x}, 1, t) = \varphi(\mathbf{x}). \end{aligned} \tag{43}$$

We can assume without loss of generality that the two structures exchanged by $X^m(s)$ are identical to $\varphi_1(\mathbf{x})$ and, for any s, lie completely within the 3-dimensional unit cube or 3-cube

$$I^3 = \{x, y, z \mid 0 \le x, y, z \le 1\}, \tag{44}$$

and that $\varphi(\mathbf{x})$ contains no other structures, since by Secs. IV.3 and IV.5 they are irrelevant for the statistics criterion; then (41) is equivalent to the existence of a function $\varphi(\mathbf{x}, s, t)$ on $I^3 \otimes I \otimes I = I^5$, such that ($\partial\Omega$ stands for the boundary of the region Ω):

$$\begin{aligned} &\varphi(\partial I^3, s, t) = \varphi_0, \\ &\varphi(\mathbf{x}, s, 0) = \varphi_X(\mathbf{x}, s), \\ &\varphi(\mathbf{x}, s, 1) = 0 \cdot \varphi(\mathbf{x}), \\ &\varphi(\mathbf{x}, 0, t) = \varphi(\mathbf{x}, 1, t) = \varphi(\mathbf{x}). \end{aligned} \tag{45}$$

Notice that (45) completely specifies a continuous function on the boundary ∂I^5 of $I^5 = I^3 \otimes I \otimes I$, the unit 5-cube spanned by $\mathbf{x}, s, t$; we shall denote that function by $\varphi_0[\partial I^5]$.

(c) We can also apply the above considerations to the rotation loop (42), and state that (42) means that there exists a function $\varphi_1(\mathbf{x}, s, t)$ on I^5 with boundary

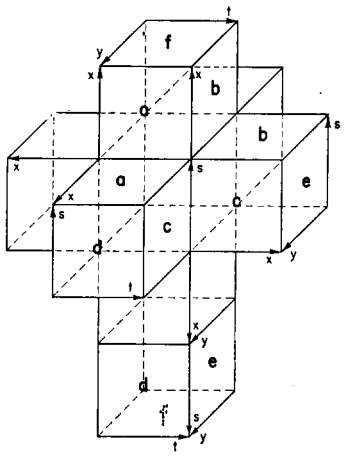


FIG. 7. Boundary of the 4-cube, or tesseract, $0 \leq (x, y, s, t) \leq 1$, where x, y are space coordinates, s is a loop parameter associated with the path of an exchange or a 2π rotation, and t is a deformation parameter associated with the homotopy of one loop into another.

values

$$\begin{aligned} &\varphi_1(\partial I^3, s, t) = \varphi_0, \\ &\varphi_1(\mathbf{x}, s, 0) = \varphi_{1W}(\mathbf{x}, s), \\ &\varphi_1(\mathbf{x}, s, 1) = 0 \cdot \varphi_1(\mathbf{x}), \\ &\varphi_1(\mathbf{x}, 0, t) = \varphi_1(\mathbf{x}, 1, t) = \varphi_1(\mathbf{x}). \end{aligned} \tag{46}$$

We shall denote the function on ∂I^5 defined by (46) as $\varphi_1[\partial I^5]$.

The existence of a continuous function $\varphi(\mathbf{x}, s, t) \equiv \varphi$ defined on I^5 and with prescribed boundary values $\varphi[\partial I^5]$, is by no means trivial, and is in fact an example of a "fundamental problem in topology, the extension problem" (Ref. 5, Sec. I.1). φ is called an extension of $\varphi[\partial I^5]$ over I^5 and (Ref. 5, Sec. I.4) there are spaces Φ such that not every $\varphi[\partial I^5]$ admits an extension over I^5.

(d) The following result is the basis for our proof[13]:

The homotopy-extension theorem: Let K be a finite simplicial complex, and L a closed subcomplex. Let $f_0: K \to Y$ and $g_u: L \to Y$ be such that $g_0 = f_0|_L$. Then f_0 admits a homotopy $f_u: K \to Y$ such that $f_u|_L = g_u$ $(0 \leq u \leq 1)$.

Since I^5 is a finite simplicial complex and ∂I^5 a closed subcomplex, the theorem applies for $K = I^5$ and $L = \partial I^5$ (note that $Y = \Phi$ is not restricted at all). Suppose that we can find a homotopy $g_u[\partial I^5]$, i.e., a family of functions on ∂I^5, such that $g_0[\partial I^5] = \varphi_0[\partial I^5]$ and $g_1[\partial I^5] = \varphi_1[\partial I^5]$; then, by the extension theorem, if $\varphi_0[\partial I^5]$ can be extended over I^5, there exists a homotopy $f_u[I^5]$ such that $f_u[I^5]|_{\partial I^5} = g_u[\partial I^5]$ so, in particular, there exists a function $f_1(I^5)$ such that

$$f_1[I^5]|_{\partial I^5} = \varphi_1[\partial I^5]. \tag{47}$$

This would prove the "only if" of the theorem in Sec. V.1, the "if" being proved by exchanging the subindices 0 and 1 in the above process.

(e) It will be convenient to represent $\varphi_0[\partial I^5]$ graphically, but this is, of course, not possible since ∂I^5 is a 4-dimensional manifold. However, for our proof we shall only need 2 dimensions of X, or I^3, so we can represent each $z = \text{const}$ (say) cross section of I^5 and ∂I^5 independently: $I^5|_{z=\text{const}}$ is a 4-cube, I^4, and $\partial I^5|_{z=\text{const}}$ is its boundary ∂I^4, a 3-dimensional manifold homomorphic to S^3, the 3 sphere. Thus, we cannot imbed ∂I^4 in X (so we cannot give a visualization of it) but we can "open" it along some "edges" (here squares) and "flatten" it, in much the same way as the surface (boundary) of a 3-cube can be brought into a flat 2-dimensional figure; at this point our ∂I^4 can be imbedded in X, so we can give a perspective visualization of it: Fig. 7. Some of the pairs of "edges" (squares) to be identified with each other are labeled with a common letter; each of the "faces" (3-cubes) is spanned by three of the variables (x, y, s, t), each going from 0 to 1 in such a way as to match, when we make the indicated identifications between the squares they span.

(f) In Fig. 8 we show only the four "faces" in which $\varphi_0[\partial I^5]|_{z=\text{const}} \equiv \varphi_0[\partial I^4] \neq \varphi_0$, two spanned by (x, y, s) and two by (x, y, t), and the "face" $(0, y, s, t)$, which has common "edges" with the former four. The support (Sec. IV.1) of $\varphi_1(\mathbf{x})|_{z=\text{const}}$ is taken to be (see Sec. V.2b) the square $|x| < \epsilon/\sqrt{2}$, $|y| < \epsilon/\sqrt{2}$. In the $(x, y, s, 0)$ 3-cube of Fig. 8 we show the exchange of two φ_1 structures, for a given z and for $\mathbf{n}$ of (28) parallel to the $-z$ axis, the exchange of two φ_1 structures, not exactly as prescribed by (29) but in a way clearly homotopic to it; the "pipes" of Fig. 8 are generated by the supports of $\varphi_1[\mathbf{x} - \mathbf{a}_1(s)]|_{z=\text{const}}$ and

$$\varphi_1[\mathbf{x} - \mathbf{a}_2(s)]|_{z=\text{const}}.$$

Analogously the $(x, y, s, 1)$ 3 cube shows the trivial

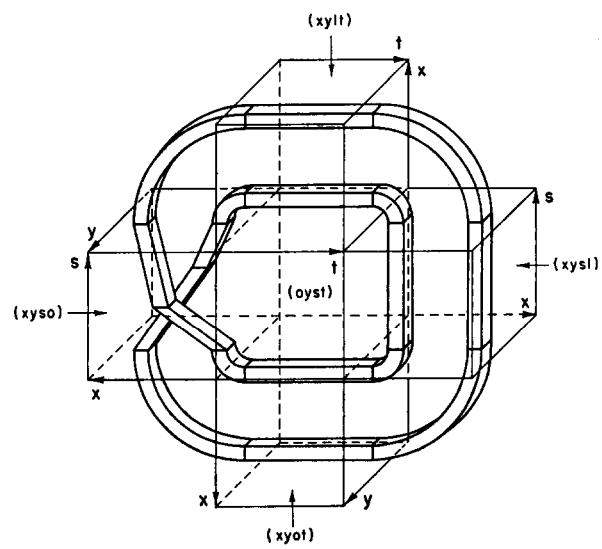


FIG. 8. Boundary values on the 4-cube of Fig. 7 whose extendibility into the interior of the 4-cube expresses the triviality of an exchange.

[13] P. J. Hilton, *An Introduction to Homotopy Theory* (Cambridge University Press, Cambridge, 1964).

loop 0. $(\varphi_1(\mathbf{x}-\mathbf{a}_1) \cup \varphi_1(\mathbf{x}-\mathbf{a}_2))$. The two remaining 3-cubes with $\varphi \not\equiv \varphi_0$ represent the last line of (45). The "curved pipes" only indicate how to identify the structures contained in identical squares; they can also be thought of as generated by the supports of the structures contained in one of the squares in which they end, if that square is "flopped over" to make it coincide with the identical one.

From now on we may speak of I^4, ∂I^4, $f[I^4]$, $g[\partial I^4]$ instead of I^5, ∂I^5, $f[I^5]$, $g[\partial I^5]$, in the understanding that we always refer to the z = const cross sections of the latter.

(g) We now construct a homotopy $g_u[\partial I^4]$ with the required properties (Sec. V.2.d). In all of ∂I^4, except the five 3-cubes shown in Fig. 8, $g_u[\partial I^4] = \varphi_0$ for all u, and $g_0[\partial I^5] = \varphi_0[\partial I^5]$ is given in Fig. 8.

In going from Fig. 8 to Fig. 9 we just "pull" $g_u[\partial I^4]$ towards the $(0yst)$ cube, in which g_u is defined in such a way as to match the boundary conditions generated when the supports of g_u in the surrounding cubes are cut by the common boundaries.

Next the "inner handle" of Fig. 9 undergoes a rigid π rotation around the α axis, in the sense shown there, with the values of the g_u rigidly carried along; the rest of g_u in the $(xy0t)$, $(xy1t)$, and $(xys1)$ cubes is unchanged. The homotopy is more complicated (not rigid) in the $(xys0)$ cube, where g_u is defined in a way that matches the boundary values for all u; then, since the shaded regions undergo a π rotation each, the two "pipes" in the $(xys0)$ cube are subject to a π twist each. By translating the content of the $(0yst)$ cube into the $(xys0)$ cube, straightening the resulting $(xys0)$ "pipe," and distributing uniformly along the s axis the two π-twists concentrated at its lower and upper ends, we

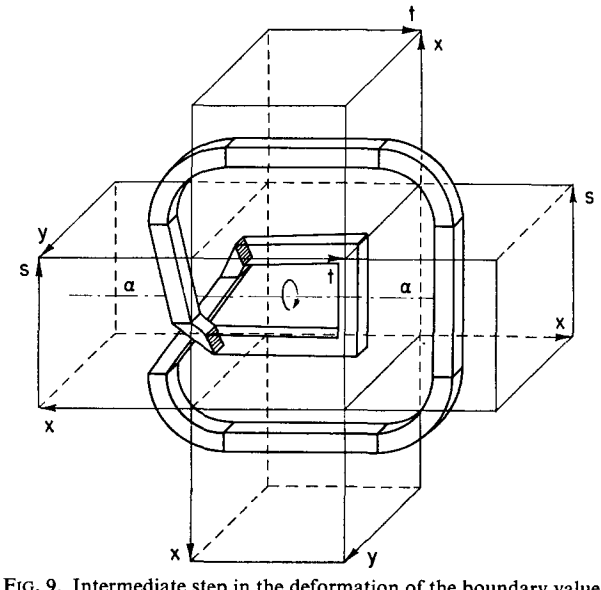


FIG. 9. Intermediate step in the deformation of the boundary values of Fig. 8 into those of Fig. 10.

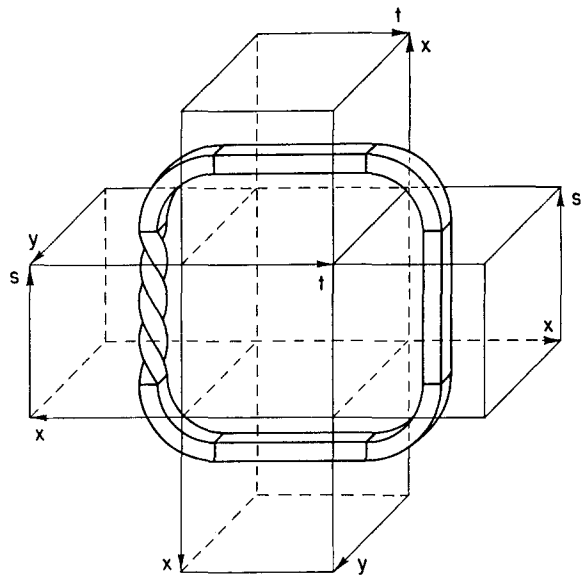


FIG. 10. Boundary values on the 4-cube of Fig. 7 whose extendability into the interior of the 4-cube expresses the triviality of a 2π rotation.

finally obtain, in Fig. 10, a graphical representation of $\varphi_1[\partial I^4]$ = (46), and thus, according to Sec. 1.d, we have proven the theorem. The detailed analytical construction of $g_u(\partial I^5)$ is given in Appendix C.

3. Consistency of the definition of X^m. We shall now prove Sec. IV.3, i.e., that if φ is the union (Sec. IV.1) of more than one pair of identical fields $\varphi_m(\mathbf{x}) \in Q_m$, the exchange of the two kinks in one pair is homotopic relative to its end point[11] to the exchange of the two kinks in any of the other pairs. The analytic proof is rather tedious, and similar to the one given in proving the theorem in Sec. V.1 (Appendix C); we shall omit it, and give only the corresponding graphic constructions (see Sec. V.2 for meaning of figures).

(a) We will first prove our statement for a particular case, i.e., when the two pairs have one kink in common. In Fig. 11 we show a field containing 3 identical kinks A, B, C. In $(x, y, s, 0)$ we exchange A, B, and in $(x, y, s, 1)$ we exchange B, C; in $(x, y, 0, t)$ and $(x, y, 1, t)$ we leave A, B, C fixed. The $g_0[\partial I^5]$ of Fig. 11 can be extended into I^5. This can be seen by treating the "inner handle" as in Sec. V.2 (Figs. 8–10): it is clear that we will obtain the $g[\partial I^5]$ of Fig. 12; again, treating the inner handle of Fig. 12 as in Sec. V.2 we are finally left with the $g[\partial I^5]$ of Fig. 13; since it can obviously be extended into I^5, our $g_0[\partial I^5]$ can too, so the exchange of A, B is homotopic to the exchange of B, C.

In case we had performed the B, C exchange in the opposite direction, we would have ended up with a (W^e) loop in the $(xys1)$ cube, instead of the trivial loop of Fig. 13. This would not invalidate our result, since $(W^e)^2 \sim 1$ (Sec. III.4).

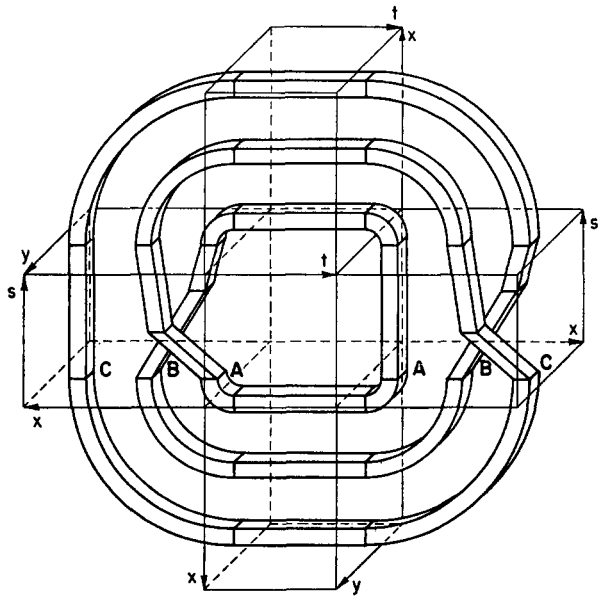


FIG. 11. Boundary values for the homotopy of one exchange (AB) into another (BC). A, B, C are identical kinks. This is a formalization of the second rubber-band lemma.

(b) In Fig. 14 we show only one space dimension (x) and two pairs of kinks, AB and CD, with

$$\varphi_B(\mathbf{x} + \mathbf{b}) = \varphi_A(\mathbf{x} + \mathbf{a})$$

and

$$\varphi_D(\mathbf{x} + \mathbf{d}) = \varphi_C(\mathbf{x} + \mathbf{c}),$$

and all of them belonging to Q_m; to distinguish them we assume the support of $\varphi_A(\mathbf{x})$ is very small and the support of $\varphi_C(\mathbf{x})$ is small, but bigger than the support of $\varphi_A(\mathbf{x})$. Our purpose will be accomplished by showing that the boundary conditions on the "cube"

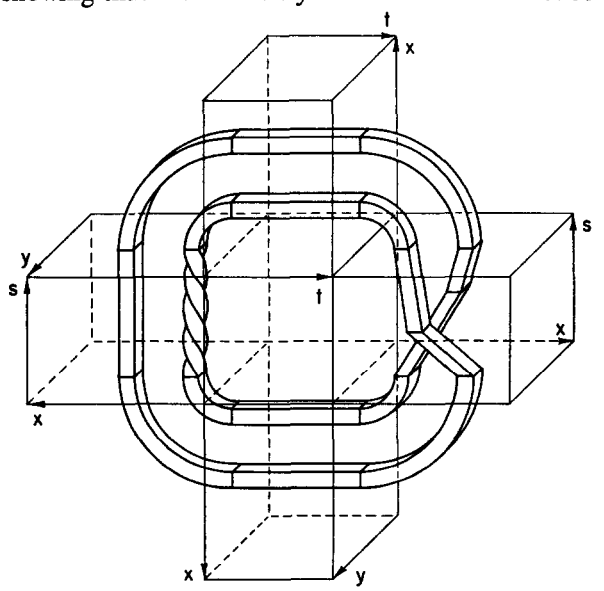


FIG. 12. Intermediate step in the deformation of Fig. 11 into Fig. 13. Note the resemblance to the first rubber-band lemma.

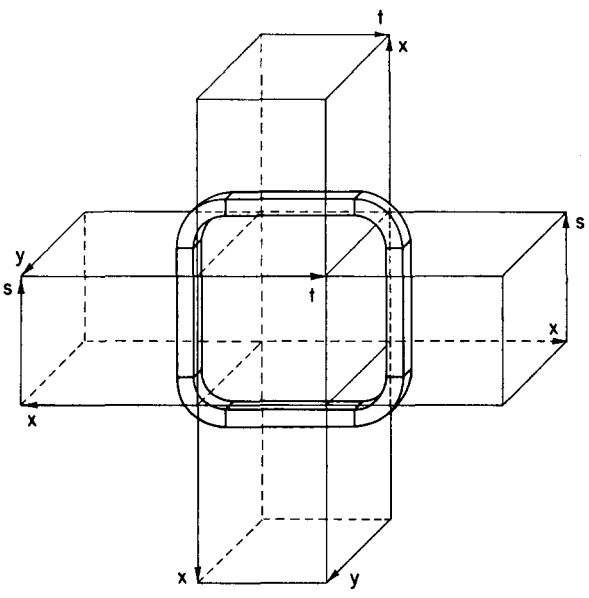


FIG. 13. Boundary values of the identity deformation on the identity loop. These values are obviously extendable into the interior.

of Fig. 14 can be extended into its interior (the t scale has been taken larger than the s, x scales to facilitate the drawing) by some function $\varphi[\partial I^5]$.

In going from $t = 0$ to $t = t_1$ we gradually transform the identity loop on C, D into the loop consisting in deforming C, D, into C', D' [with $\varphi_{D'}(\mathbf{x} + \mathbf{d}) = \varphi_{C'}(\mathbf{x} + \mathbf{c}) = \varphi_B(\mathbf{x} + \mathbf{b}) = \varphi_A(\mathbf{x} + \mathbf{a})$] for $0 \leq s \leq s_1$, C', D' remaining fixed for $s_1 \leq s \leq s_2$, and C', D' becoming C, D for $s_2 \leq s \leq 1$; also, the exchange of A, B is concentrated in the region $s_1 \leq s \leq s_2$.

In $t_1 \leq t \leq t_2$ we take $\varphi[I^5]$ identical to its value at $t = t_1$, for $0 \leq s \leq s_1$ and $s_2 \leq s \leq 1$. For $s_1 \leq s \leq s_2$ we have a situation to which we can apply the result just obtained: the loop at t_1, in which A, B are exchanged and C', D' remain fixed, is homotopic to a loop in which B, C' are exchanged and A, D' remain fixed, which in turn is homotopic to a loop in which C', D' are exchanged, whereas A, B remain fixed, at $t = t_2$.

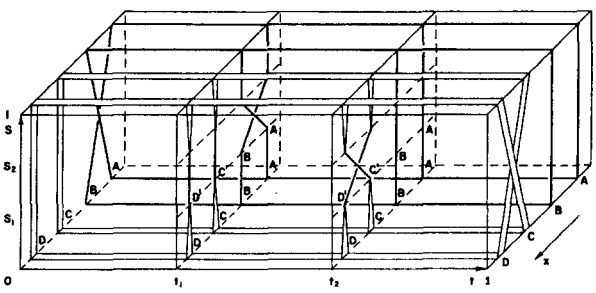


FIG. 14. Proof that two exchanges are equivalent (uniformity of statistics). Here A, B, are identical kinks of type m, and C, D are identical kinks of type m, but A and C need not be identical. At $t = 0$, the exchange AB is represented. At $t = 1$, the exchange CD is represented. The intermediate steps demonstrate their equivalence.

For $t_2 \leq t \leq 1$, $\varphi[I^5]$ consists simply in distributing uniformly along s the exchange performed at $t = t_2$, and at the same time "unmaking" the transformation made on C, D in the step $0 \leq t \leq t_1$. This completes the construction of an homotopy $\varphi[\partial I^5]$ between the exchange of A, B (at $t = 0$) and the exchange of C, D, at $t = 1$. Q.E.D.

4. In Sec. V.3a we have obtained the boundary function of Fig. 12 as an intermediate step and then proved that it can always be extended into the interior of I^5. This means that for a two-kink field, exchanging the kinks is homotopic[11] to leaving one of them fixed and rotating the other one through 2π.

Without using the homotopy extension theorem the existence of an homotopy between the above mentioned loops is not quite evident, so it is interesting to know that we can actually construct the homotopy; its detailed exposition would be too lengthy (since the interest of the problem is rather academic) so we only give, in Fig. 15, a rough sketch of the method.

5. A $2n$-kink state does not admit half-odd spin: if in $n = (n_1, \cdots, n_k)$ all the n_i are even, by Sec. III.6 we can check the statement using as a test field a union field

$$\varphi(\mathbf{x}) = \varphi_m(\mathbf{x}) \cup \varphi_m(\mathbf{x} - \mathbf{a}),$$
$$m = \tfrac{1}{2}n;$$

taking $\mathbf{a}$ and $\mathbf{n}$ along the x axis, after a trivial homotopy the above process can be symbolized by Fig. 16(a), a simplified diagram of the type used in Figs. 11–15. The homotopy (a) → (b) is obvious, and in (b) → (c) we use the result analyzed in Sec. V.4. Since (c) is simply $(X^m)^2$, (c) → (d) is proven by $(X^m)^2 = 1$, Eq. (31).

In conventional quantum-field theories, the corresponding statement would follow from the laws of composition of angular momentum: an even number of half-odd-spin particles can only have integer spin.

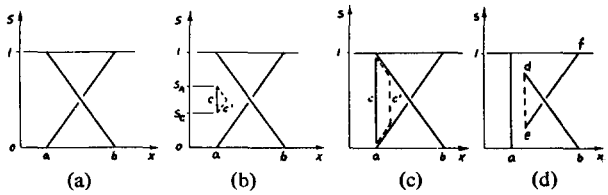


FIG. 15. Deformation of an exchange into a rotation. (a) The exchange of two kinks, originally at a and b, each represented by a dot; this is what we would see looking straight into the front face of the $(xysl)$ cube of Fig. 8 if the support of the kinks were very small. (b) We create at a (after the "a kink" has left that place) a kink c and its antikink c' (see Sec. I.7); c' moves (dotted line) towards b and then returns to a to annihilate c. In b → c s_Q and s_A, the values of s at which creation and annihilation take place approach 0 and 1 respectively. (c) Part of the path of c' overlaps the path of the "a kink" and part of the path of the "b kink"; in the corresponding regions the field is equal to φ_0: (d) The rest of the homotopy (and the most involved part of it) consists of deforming the loop $bdef$ of d into the rotation of a kink sitting at b, but at this point we would need more elaborate figures.

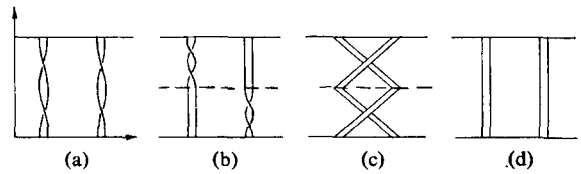


FIG. 16. Schematic of a deformation. The starting point (not shown) is a rigid rotation of two identical kinks about the line of centers. Then the axis of rotation of each kink is twined perpendicular to the line of centers, giving the deformation (a), where each kink is represented by a segment, whose endpoints traverse the intertwining helices shown. The intermediate step (b) is obtained in an obvious way from (a). In going from (b) to (c), the extendability of Fig. 12 into the 4-cube is applied to the upper and lower halves of (b) independently. Since (c) represents X^2, the deformation to the identity (d) has already been shown.

ACKNOWLEDGMENT

We thank Charles Misner for criticisms of the manuscript which helped our own understanding, as well as the presentation.

APPENDIX A: PROOF OF LEMMA 3

Let q, $q_1 \in\in Q_m$, $q(s)$ be a path between $q = q(0)$ and $q_1 = q(1)$, $W^e(s)$ a $2\pi s$ extrinsic rotation operator around some axis, and $q_1(s) = W(s) \cdot q$ a homotopically trivial loop (so $W \cdot \tilde{q}_1 = \tilde{q}_1$). Then $q(s, t) = W^e(s) \cdot q$,

$$
\begin{array}{lll}
q(s,t) = W^e(s)\cdot q, & 0 \leq s < 1, & t = 0, \\
\left.\begin{array}{ll} q, & 0 \leq s \leq t \\ W^e\left(\dfrac{s-t}{1-2t}\right)\cdot q, & t \leq s \leq 1-t \\ q, & 1-t \leq s \leq 1 \end{array}\right\} & & 0 < t < \tfrac{1}{4}, \\
\left.\begin{array}{ll} q[(4t-1)3s], & 0 \leq s \leq \frac{1}{3} \\ W^e(3s-1)\cdot q(4t-1), & \frac{1}{3} \leq s \leq \frac{2}{3} \\ q[(4t-1)(3-3s)], & \frac{2}{3} \leq s \leq 1 \end{array}\right\} & & \tfrac{1}{4} \leq t < \tfrac{2}{4}, \\
\left.\begin{array}{ll} q(3s), & 0 \leq s \leq \frac{1}{3} \\ W^e(3s-1)\cdot q(1) & \\ \quad = W(3s-1)\cdot q_1 & \\ \quad = q_1(s), & \frac{1}{3} \leq s \leq \frac{2}{3} \\ q(3-3s), & \frac{2}{3} \leq s \leq 1 \end{array}\right\} & & t = \tfrac{2}{4}, \\
\left.\begin{array}{ll} q(3s), & 0 \leq s \leq \frac{1}{3} \\ q_1(s,t) : q_1(s,\frac{2}{4}) = q_1(s), & \\ \quad q_1(s,\frac{3}{4}) = q_1 = q(1), & \frac{1}{3} \leq s \leq \frac{2}{3} \\ q(3-3s), & \frac{2}{3} \leq s \leq 1 \end{array}\right\} & & \tfrac{2}{4} \leq t \leq \tfrac{3}{4}, \\
\left.\begin{array}{ll} q[(4-4t)3s], & 0 \leq s \leq \frac{1}{3} \\ q(4-4t), & \frac{1}{3} \leq s \leq \frac{2}{3} \\ q[(4-4t)(3-3s)], & \frac{2}{3} \leq s \leq 1 \end{array}\right\} & & \tfrac{3}{4} \leq t \leq 1, \\
q(0) = q, & 0 \leq s \leq 1, & t = 1,
\end{array}
$$

is a homotopy between $W^e(s) \cdot q$ and the trivial loop, for all $q \in Q_m$.

APPENDIX B

By composition law we mean a unique and continuous prescription to define a union field $\varphi = \cup\varphi_i$ when the supports of the φ_i overlap.[14] Then

Theorem: If Q does not admit a composition law, $(X^m)^2 = 1$ is not true in 2 dimensions.

Proof: Take $a_1 = (-a, 0)$, $a_2 = (a, 0)$ and rotate them around the origin of the (xy) plane. The trajectories $\mathbf{a}_1$, $\mathbf{a}_2$ corresponding to $(X^m)^2$ [Fig. 6(a)] are

$$a_{1x}(s) = -a\cos 2\pi s, \quad a_{2x}(s) = a\cos 2\pi s,$$
$$a_{1y}(s) = -a\sin 2\pi s, \quad a_{2y}(s) = a\sin 2\pi s, \quad \text{(B1)}$$

and admit the following homotopy in the xy plane [Fig. 17(a)]:

$$\begin{aligned} a_{1x}(s,t) &= -(1-t)a\cos 2\pi s - ta, \\ a_{2x}(s,t) &= (1+t)a\cos 2\pi s - ta, \\ a_{1y}(s,t) &= -(1-t)a\sin 2\pi s, \\ a_{2y}(s,t) &= a\sin 2\pi s. \end{aligned} \quad \text{(B2)}$$

At $t = 0$ (B2) gives (B1) and, at $t = 1$, [Fig. 17(b)]

$$a_{1x}(s,1) = -a, \quad a_{2x}(s,1) = -a(1 - 2\cos 2\pi s),$$
$$a_{1y}(s,1) = 0, \quad a_{2y}(s,1) = a\sin 2\pi s;$$

at no (s, t) do $\mathbf{a}_1(s, t)$ and $\mathbf{a}_2(s, t)$ coincide. For the trajectories at $t = 1$ to be deformed into the trivial trajectory, $\mathbf{a}_1$ has to be left still and $\mathbf{a}_2(s, 1)$ shrunk into $\mathbf{a}_2(0, 1)$. This is clearly impossible to do in the x, y plane without at some stage having $\mathbf{a}_2$ and $\mathbf{a}_1$ overlap [see Fig. 17(a)].

Should (B1) be deformable into the trivial loop (in the x, y plane and without overlapping), the inverse of the homotopy (B2) followed by such a deformation would be a deformation of the trajectory of Fig. 17(b) into the trivial one; since this is not possible, the theorem is proved.

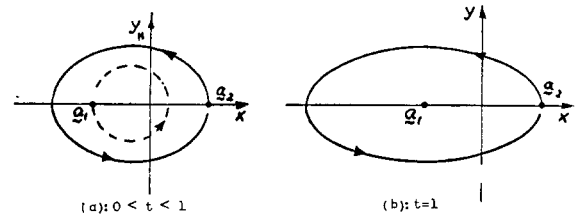


FIG. 17. A deformation of the $(X^m)^2$ loop in the plane, showing that in two dimensions, in general, $(X^m)^2 \neq 1$.

As a by-product, by means of motions out of the (x, y) plane, Fig. 17 provides an easy alternative proof of the relation $(X^m)^2 = 1$.

APPENDIX C: ANALYTIC PROOF OF THE THEOREM IN SEC. 1

The following conventions are used in the proof:

(1) (x, y) stands for $\mathbf{x} = (x, y, z)$; z remains unaffected in all transformations.

(2) x, y, s, t, u (but not z) range all between 0 and 1, if not otherwise indicated.

(3) $s \leq s_1$ stands for $0 \leq s \leq s_1$ and $s \geq s_2$ for $s_2 \leq s \leq 1$; analogously for the other variables.

(4) Figures 8–10 represent part of cross sections $z = \text{const}$ of ∂I^5, i.e., ∂I^4, by showing five of its eight sectors: $(x, y, s, 0)$, $(x, y, s, 1)$, etc. For fixed u, $\varphi(x, y, s, t, u)$ is a function on ∂I^5; it is defined as equal to φ_0 at those points of ∂I^5 at which its formal expression given below is meaningless, and is equal to φ_0 on all the sectors not mentioned explicitly.

(5) $\varphi_1(\mathbf{x}) = \varphi_0$ for $|\mathbf{x}| > \epsilon$; $a < \frac{1}{2} - 2\epsilon$ is a fixed positive number, and

$$b = 1 - a; \quad 0 < u_1 < u_2 < \cdots < u_8 < 1$$

are eight real fixed numbers.

(6) $\varphi(x, y, s, t, u)$ is the $g_u[\partial I^5]$ of the text, Sec. V.2. Let

$$\begin{aligned} f_1(s) &= (1-a) - (1-2a)s, & g_1(s) &= \tfrac{1}{2} - (1-2a)s, & s &\leq \tfrac{1}{2}, \\ & & &= a + (1-2a)(s - \tfrac{1}{2}), & s &\geq \tfrac{1}{2}, \\ f_2(s) &= a + (1-2a)s, & g_2(s) &= \tfrac{1}{2} + (1-2a)s, & s &\leq \tfrac{1}{2}, \\ & & &= (1-a) - (1-2a)(s - \tfrac{1}{2}), & s &\geq \tfrac{1}{2}. \end{aligned}$$

Then (Fig. 8)

$$\varphi(x, y, s, 0, 0) = \varphi_1(x - f_1(s), y - g_1(s)) \cup \varphi_1(x - f_2(s), y - g_2(s)),$$
$$\varphi(x, y, s, 1, 0) = \varphi(x, y, 0, t, 0) = \varphi(x, y, 1, t, 0) = \varphi_1(x - a, y - \tfrac{1}{2}) \cup \varphi_1(x - b, y - \tfrac{1}{2}).$$

Let

$$\left.\begin{aligned} f_1(s,u) &= f_1(s), & g_1(s,u) &= (1-p)g_1(s) + pa, & g_3(s,u) &= (1-p)\tfrac{1}{2} + pf_1(s) \\ f_2(s,u) &= f_2(s), & g_2(s,u) &= (1-p)g_2(s) + pb, & g_4(s,u) &= (1-p)\tfrac{1}{2} + pf_2(s) \end{aligned}\right\} \quad p = u/u_1, \quad 0 \leq u \leq u_1.$$

[14] For instance, when Φ is the manifold of a group, and φ_0 its identity element, a natural prescription is $\varphi_1(\mathbf{x}) \cup \varphi_2(\mathbf{x}) = \varphi_1(\mathbf{x}) \cdot \varphi_2(\mathbf{x})$, with "$\cdot$" indicating the product in Φ at each $\mathbf{x}$.

Then

$$\left.\begin{aligned}
&\varphi(x, y, s, 0, u) = \varphi_1(x - f_1(s, u), y - g_1(s, u)) \cup \varphi_1(x - f_2(s, u), y - g_2(s, u))\\
&\varphi(x, y, s, 1, u) = \varphi_1(x - a, y - g_3(s, u)) \cup \varphi_1(x - b, y - g_4(s, u))\\
&\varphi(x, y, 0, t, u) = \varphi_1(x - b, y - a) \cup \varphi_1(x - a, y - b)\\
&\varphi(x, y, 1, t, u) = \varphi_1(x - a, y - a) \cup \varphi_1(x - b, y - a)
\end{aligned}\right\} \quad 0 \le u \le u_1.$$

Let

$$\left.\begin{aligned}
f_1(s, u) &= (1 - g)f_1(s) + gf_1'(s), & f_1'(s) &= b - \frac{\frac{1}{2} - a - \epsilon}{\frac{1}{2} + a - \epsilon}s, && s \le \tfrac{1}{2} + a - \epsilon\\
f_2(s, u) &= (1 - g)f_2(s) + gf_2'(s), & &= \tfrac{1}{2} + \epsilon - (s - (\tfrac{1}{2} + a - \epsilon)), && \\
& & & && s \ge \tfrac{1}{2} + a - \epsilon\\
f_3(s, u) &= (1 - g)a + gf_3'(s), & f_2'(s) &= a + s, && s \le \tfrac{1}{2} - a + \epsilon\\
g_3'(s, u) &= (1 - g)f_1(s) + gg_3'(s), & &= \tfrac{1}{2} + \epsilon + \frac{\frac{1}{2} - a - \epsilon}{(\frac{1}{2} + a - \epsilon)(s - g)}, && \\
& & & && s \ge \tfrac{1}{2} - a + \epsilon\\
g_3'(s) &= b, & f_3'(s) &= a + s, && s \le \tfrac{1}{2} - a + \epsilon\\
&= b - (s - (\tfrac{1}{2} - a + \epsilon))\frac{1 - 2a}{2(a - \epsilon)}, & &= \tfrac{1}{2} + \epsilon, && \tfrac{1}{2} - a + \epsilon \le s \le \tfrac{1}{2} + a - \epsilon\\
&= a, & &= \tfrac{1}{2} - (s - (\tfrac{1}{2} + a - \epsilon)), && s \ge \tfrac{1}{2} + a - \epsilon\\
& & & && u_1 \le u \le u_2.
\end{aligned}\right\} \quad g = \frac{u - u_1}{u_2 - u_1}$$

Then

$$\left.\begin{aligned}
&\varphi(x, y, s, 0, u) = \varphi_1(x - f_1(s, u), y - a) \cup \varphi_1(x - f_2(s, u), y - b)\\
&\varphi(x, y, s, 1, u) = \varphi_1(x - f_3(s, u), y - g_3(s, u)) \cup \varphi_1(x - b, y - f_2(s))\\
&\varphi(x, y, 0, t, u) = \varphi(x, y, 0, t, u_1)\\
&\varphi(x, y, 1, t, u) = \varphi(x, y, 1, t, u_1)
\end{aligned}\right\} \quad u_1 \le u \le u_2,$$

$$\left.\begin{aligned}
&\varphi(x, y, s, 0, u) = \varphi(x + r, y, s, 0, u_2)\\
&\varphi(x, y, s, 1, u) = \varphi(x + r, y, s, 0, u_2)\\
&\varphi(x, y, 0, t, u) = \varphi(x + r, y, 0, t, u_2)\\
&\varphi(x, y, 1, t, u) = \varphi(x + r, y, 1, t, u_2)\\
&\varphi(0, y, s, t, u) = \varphi(r, y, s, 0, u_2)
\end{aligned}\right\} \quad r = \frac{1}{2}\frac{u - u_2}{u_3 - u_2} \quad u_2 \le u \le u_3.$$

Let

$$\left.\begin{aligned}
&\omega = y + is \text{ stand for } (y, s) \text{ in } \varphi(x, y, s, t, u); \quad \alpha = \tfrac{1}{2} + \tfrac{1}{2}i \text{ and}\\
&\omega_1(0) = a + i(\tfrac{1}{2} + a) - i\epsilon = \beta_1 - i\epsilon, \qquad \omega_1(u) = y_1(u) + is_1(u) = \alpha + e^{iv}(\beta_1 - \alpha) - i\epsilon\\
&\omega_2(0) = (1 - a) + i(\tfrac{1}{2} - a) + i\epsilon = \beta_2 + i\epsilon, \quad \omega_2(u) = y_2(u) + is_2(u) = \alpha + e^{iv}(\beta_2 - \alpha) + i\epsilon\\
&f_1(s, u) = \tfrac{1}{2} - a - \frac{\frac{1}{2} - a - \epsilon}{s_1(u)}s, \qquad g_1(s, u) = a - \frac{a - y_1(u)}{s_1(u)}s, \qquad s \le s_1(u)\\
&\qquad = \epsilon - (s - s_1(u)), \qquad = y_1(u), \qquad s \ge s_1(u)\\
&f_2(s, u) = \epsilon + (s - s_2(u)), \qquad g_2(s, u) = y_2(u), \qquad s \le s_2(u)\\
&\qquad = \epsilon + \frac{\frac{1}{2} - a - \epsilon}{1 - s_2(u)}(s - s_2(u)), \qquad = y_2(u) - \frac{y_2(u) - (1 - a)}{(1 - s_2(u))(s - g)}, \qquad s \ge s_2(u)\\
&\lambda_1(s, u) = s/s_1(u), \quad s \le s_1(u), \qquad \lambda_2(s, u) = 1, \qquad s \le s_2(u)\\
&\qquad = 1, \qquad s \ge s_1(u), \qquad = \frac{1 - s}{1 - s_2(u)}, \qquad s \ge s_2(u)\\
&\Omega_i = (x - f_i(s, u)) + i(y - g_i(s, u)) = (X_i + iY_i)(x, y, s, u), \qquad i = 1, 2\\
&\qquad u_3 \le u \le u_4.
\end{aligned}\right\} \quad v = \pi\frac{u - u_3}{u_4 - u_3}$$

Then

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi_1(e^{iv\lambda_1(s_1u)}\Omega_1(x, y, s, u)) \cup \varphi_1(e^{-iv\lambda_2(s_1u)}\Omega_2(x, y, s, u)) \\
\varphi(x, y, s, 1, u) &= \varphi(x, \alpha + e^{-iv}(\omega - \alpha), 1, u_3), \quad x \leq 2\epsilon \\
&= \varphi(x, y, s, 1, u_3), \quad x \geq 2\epsilon \\
\varphi(x, y, 0, t, u) &= \varphi(x, y, 0, t, u_3) \\
\varphi(x, y, 1, t, u) &= \varphi(x, y, 1, t, u_3) \\
\varphi(0, y, s, t, u) &= \varphi(0, \alpha + e^{-iv}(\omega - \alpha), t, u_3)
\end{aligned}\right\} \quad u_3 \leq u \leq u_4,$$

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi(x, y, s, 0, u_4) \\
\varphi(x, y, s, 1, u) &= \varphi(x + w, y, s, 1, u_4), \quad x \leq 2\epsilon \\
&= \varphi(x, y, s, 1, u_4), \quad x \geq 2\epsilon \\
\varphi(x, y, 0, t, u) &= \varphi(x, y, 0, t, u_4) \\
\varphi(x, y, 1, t, u) &= \varphi(x, y, 1, t, u_4) \\
\varphi(0, y, s, t, u) &= \varphi(0, y, s, t, u_4), \quad t \leq 1 - w \\
&= \varphi(t + w - 1, y, s, 1, u_4), \quad t > 1 - w
\end{aligned}\right\} \quad w = 2\epsilon\frac{u - u_4}{u_5 - u_4} \quad u_4 \leq u \leq u_5.$$

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi(x, y, s, 0, u_5) \\
\varphi(x, y, s, 1, u) &= \varphi(x, y, s, 1, u_5) \\
\varphi(x, y, 0, t, u) &= \varphi(x, y, 0, t, u_5) \\
\varphi(x, y, 1, t, u) &= \varphi(x, y, 1, t, u_5) \\
\varphi(0, y, s, t, u) &= \varphi(0, y, s, t + \mu, u_5)
\end{aligned}\right\} \quad \mu = (1 - 2\epsilon)\frac{u - u_5}{u_6 - u_5} \quad u_5 \leq u \leq u_6.$$

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi(0, y, s, \nu - x, u_6), \quad x < \nu \\
&= \varphi(x - \nu, y, s, 0, u_6), \quad x > \nu \\
\varphi(x, y, s, 1, u) &= \varphi(x - \nu, y, s, 1, u_6) \\
\varphi(x, y, 0, t, u) &= \varphi(x - \nu, y, 0, t, u_6) \\
\varphi(x, y, 1, t, u) &= \varphi(x - \nu, y, 1, t, u_6) \\
\varphi(0, y, s, t, u) &= \varphi(0, y, s, t + \nu, u_6)
\end{aligned}\right\} \quad \nu = 2\epsilon\frac{u - u_6}{u_7 - u_6} \quad u_6 \leq u \leq u_7.$$

Let

$$f(s) = \begin{cases} s, & 0 \leq s \leq \frac{1}{2} - a + \epsilon, \\ \frac{1}{2} - a + \epsilon, & \frac{1}{2} - a + \epsilon \leq s \leq \frac{1}{2} + a - \epsilon, \\ (\frac{1}{2} - a + \epsilon) - (s - (\frac{1}{2} + a - \epsilon)), & \frac{1}{2} + a - \epsilon \leq s1, \end{cases}$$

$$g(s) = \begin{cases} \dfrac{1 - 2a}{\frac{1}{2} - a - \epsilon}\, s, & 0 \leq s \leq \frac{1}{2} - a - \epsilon, \\ 1 - 2a, & \frac{1}{2} - a - \epsilon \leq s \leq \frac{1}{2} - a + \epsilon, \\ 1 - 2a - \dfrac{1 - 2a}{2(a - \epsilon)}(s - (\frac{1}{2} - a - \epsilon)), & \frac{1}{2} - a + \epsilon \leq s \leq \frac{1}{2} + a - \epsilon, \\ 0, & \frac{1}{2} + a - \epsilon \leq s \leq \frac{1}{2} + a + \epsilon, \\ \dfrac{1 - 2a}{\frac{1}{2} - a - \epsilon}(s - (\frac{1}{2} + a + \epsilon)), & \frac{1}{2} + a + \epsilon \leq s \leq 1. \end{cases}$$

Then

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi(x - \rho f(s), y + \rho g(s), s, u, u_7) \\
\varphi(x, y, s, 1, u) &= \varphi(x, y + \rho f_2(s), s, 1, u_7) \\
\varphi(x, y, 0, t, u) &= \varphi(x, y, 0, t, u_7) \\
\varphi(x, y, 1, t, u) &= \varphi(x, y + \rho(1 - 2a), 1, t, u_7)
\end{aligned}\right\} \quad \rho = \frac{u - u_7}{u_8 - u_7}, \quad u_7 \leq u \leq u_8,$$

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi(x, y, \sigma s, 0, u_8), && s \leq \tfrac{1}{2} \\
&= \varphi(x, y, \sigma(s-1)+1, 0, u_8), && s \geq \tfrac{1}{2} \\
\varphi(x, y, s, 1, u) &= \varphi(x, y, s, 1, u_8) \\
\varphi(x, y, 0, t, u) &= \varphi(x, y, 0, t, u_8) \\
\varphi(x, y, 1, t, u) &= \varphi(x, y, 1, t, u_8)
\end{aligned}\right\} \quad \sigma = 1 - \frac{u - u_8}{1 - u_8} 2(a - \epsilon), \quad u_8 \leq u \leq 1.$$

Let $(a + 2\epsilon, a, 0) = \mathbf{c}$ and $W(s)$ be a rotation operator around the z axis (Sec. III). Collecting the above expressions we get (Fig. 10)

$$\begin{aligned}
\varphi(x, y, s, 0, 1) &= \varphi_1(W(s) \cdot (\mathbf{x} - \mathbf{c})), \\
\varphi(x, y, s, 1, 1) &= \varphi_1(\mathbf{x} - \mathbf{c}), \\
\varphi(x, y, 0, t, 1) &= \varphi_1(\mathbf{x} - \mathbf{c}), \\
\varphi(x, y, 1, t, 1) &= \varphi_1(\mathbf{x} - \mathbf{c}),
\end{aligned}$$

Q.E.D.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Principle of Compensation of Dangerous Diagrams for Boson Systems. I. Maximum Overlap

DONALD H. KOBE*

Department of Physics, Northeastern University, Boston, Massachusetts

(Received 26 March 1968)

The principle of compensation of dangerous diagrams (PCDD), used by Bogoliubov to determine the coefficients in the canonical transformation to quasiparticles in boson systems, is obtained by maximizing the overlap between the true ground-state vector and the quasiparticle vacuum state. The zero-momentum state is treated exactly, which implies that the sum of all diagrams leading from the vacuum to a one-quasiparticle state must be zero, in addition to the diagrams leading from the vacuum to the two quasiparticle state. Other criteria, such as the diagonalization of the quasiparticle reaction operator up to terms cubic in the operators, and the absence of one and two quasiparticle contributions to the true ground-state wavefunction, are also shown to lead to the PCDD. A generalization of the Hartree procedure of minimizing the ground-state energy is obtained by replacing the bare quasiparticle interaction with the quasiparticle reaction operator, and is shown to be equivalent to the PCDD. Finally, a perturbation expansion of the PCDD is obtained, and the reducibility of diagrams is discussed.

1. INTRODUCTION

The canonical transformation to quasiparticles was first introduced by Bogoliubov in the theory of boson systems to obtain a model of superfluidity.[1] To choose the coefficients in the transformation, Bogoliubov diagonalized the quadratic part of the Hamiltonian. However, he later postulated the *principle of compensation of dangerous diagrams* (PCDD) as the best method to use for determining the coefficients in the transformation.[2] The PCDD states that the sum of all the diagrams leading from the vacuum to the two-quasiparticle state should be equated to zero. By so doing, divergent contributions to the perturbation expansion of the ground-state energy can be eliminated. The PCDD was taken over directly to fermion systems where it was used in the theory of superconductivity.[2,3]

The quasiparticles introduced by Bogoliubov will be called *bogolons* to avoid confusion with dressed particles which are also called quasiparticles. In the theory of superconductivity the bogolons are fermions, while in the theory of superfluidity they are bosons. In other words, the bogolons have the same statistics as the particles in the system. The canonical transformation expresses the particle operators linearly in terms of the bogolon operators. This transformation can then be made on the Hamiltonian, which is

* Present address: Department of Physics, North Texas State University, Denton, Texas, 76203.

[1] N. N. Bogoliubov, J. Phys. (USSR) **11**, 23 (1947).

[2] N. N. Bogoliubov, V. V. Tolmachev, and D. V. Shirkov, *A New Method in the Theory of Superconductivity* (Academy of Sciences of the USSR Press, Moscow, 1958), Chap. 1 (English transl.: Consultants Bureau, New York, 1959); Fortshr. Physik **6**, 605 (1958).

[3] N. N. Bogoliubov, Zh. Eksp. Teor. Fiz. **34**, 58 (1958) [Sov. Phys.—JETP **7**, 41 (1958)]; Nuovo Cimento **7**, 794 (1958); Usp. Fiz. Nauk **67**, 549 (1959) [Sov. Phys.—Usp. **2**, 236 (1959)].

$$\left.\begin{aligned}
\varphi(x, y, s, 0, u) &= \varphi(x, y, \sigma s, 0, u_8), && s \leq \tfrac{1}{2} \\
&= \varphi(x, y, \sigma(s-1)+1, 0, u_8), && s \geq \tfrac{1}{2} \\
\varphi(x, y, s, 1, u) &= \varphi(x, y, s, 1, u_8) \\
\varphi(x, y, 0, t, u) &= \varphi(x, y, 0, t, u_8) \\
\varphi(x, y, 1, t, u) &= \varphi(x, y, 1, t, u_8)
\end{aligned}\right\} \quad \sigma = 1 - \frac{u - u_8}{1 - u_8} 2(a - \epsilon), \quad u_8 \leq u \leq 1.$$

Let $(a + 2\epsilon, a, 0) = \mathbf{c}$ and $W(s)$ be a rotation operator around the z axis (Sec. III). Collecting the above expressions we get (Fig. 10)

$$\begin{aligned}
\varphi(x, y, s, 0, 1) &= \varphi_1(W(s) \cdot (\mathbf{x} - \mathbf{c})), \\
\varphi(x, y, s, 1, 1) &= \varphi_1(\mathbf{x} - \mathbf{c}), \\
\varphi(x, y, 0, t, 1) &= \varphi_1(\mathbf{x} - \mathbf{c}), \\
\varphi(x, y, 1, t, 1) &= \varphi_1(\mathbf{x} - \mathbf{c}),
\end{aligned}$$

Q.E.D.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Principle of Compensation of Dangerous Diagrams for Boson Systems. I. Maximum Overlap

DONALD H. KOBE*

Department of Physics, Northeastern University, Boston, Massachusetts

(Received 26 March 1968)

The principle of compensation of dangerous diagrams (PCDD), used by Bogoliubov to determine the coefficients in the canonical transformation to quasiparticles in boson systems, is obtained by maximizing the overlap between the true ground-state vector and the quasiparticle vacuum state. The zero-momentum state is treated exactly, which implies that the sum of all diagrams leading from the vacuum to a one-quasiparticle state must be zero, in addition to the diagrams leading from the vacuum to the two quasiparticle state. Other criteria, such as the diagonalization of the quasiparticle reaction operator up to terms cubic in the operators, and the absence of one and two quasiparticle contributions to the true ground-state wavefunction, are also shown to lead to the PCDD. A generalization of the Hartree procedure of minimizing the ground-state energy is obtained by replacing the bare quasiparticle interaction with the quasiparticle reaction operator, and is shown to be equivalent to the PCDD. Finally, a perturbation expansion of the PCDD is obtained, and the reducibility of diagrams is discussed.

1. INTRODUCTION

The canonical transformation to quasiparticles was first introduced by Bogoliubov in the theory of boson systems to obtain a model of superfluidity.[1] To choose the coefficients in the transformation, Bogoliubov diagonalized the quadratic part of the Hamiltonian. However, he later postulated the *principle of compensation of dangerous diagrams* (PCDD) as the best method to use for determining the coefficients in the transformation.[2] The PCDD states that the sum of all the diagrams leading from the vacuum to the two-quasiparticle state should be equated to zero. By so doing, divergent contributions to the perturbation expansion of the ground-state energy can be eliminated. The PCDD was taken over directly to fermion systems where it was used in the theory of superconductivity.[2,3]

The quasiparticles introduced by Bogoliubov will be called *bogolons* to avoid confusion with dressed particles which are also called quasiparticles. In the theory of superconductivity the bogolons are fermions, while in the theory of superfluidity they are bosons. In other words, the bogolons have the same statistics as the particles in the system. The canonical transformation expresses the particle operators linearly in terms of the bogolon operators. This transformation can then be made on the Hamiltonian, which is

* Present address: Department of Physics, North Texas State University, Denton, Texas, 76203.

[1] N. N. Bogoliubov, J. Phys. (USSR) **11**, 23 (1947).

[2] N. N. Bogoliubov, V. V. Tolmachev, and D. V. Shirkov, *A New Method in the Theory of Superconductivity* (Academy of Sciences of the USSR Press, Moscow, 1958), Chap. 1 (English transl.: Consultants Bureau, New York, 1959); Fortshr. Physik **6**, 605 (1958).

[3] N. N. Bogoliubov, Zh. Eksp. Teor. Fiz. **34**, 58 (1958) [Sov. Phys.—JETP **7**, 41 (1958)]; Nuovo Cimento **7**, 794 (1958); Usp. Fiz. Nauk **67**, 549 (1959) [Sov. Phys.—Usp. **2**, 236 (1959)].

normal-ordered in the bogolon operators to obtain the bogolon Hamiltonian.

The diagonalization of the bogolon Hamiltonian, neglecting terms cubic and quartic in the operators, is the same as compensating the lowest-order dangerous diagrams, i.e., the first term in the PCDD. This procedure is equivalent to the Hartree procedure of minimizing the ground-state energy using the bogolon vacuum state as the trial wavefunction.[4] Thus the compensation of the lowest-order dangerous diagrams (CLODD) is based firmly on a variational method. However, the second-order contribution to the PCDD has recently been shown to be important in the theory of the charged boson gas by Woo and Ma.[5] Thus the question arises as to whether the CLODD or the full PCDD should be used. After all, the elimination of a class of divergent diagrams from the perturbation expansion of the ground-state energy does not necessarily guarantee the convergence of the expansion.

The same question arose in the theory of fermion systems since it was shown that in nuclear matter the second-order corrections to the PCDD gave different results than the CLODD.[6] The PCDD in fermion systems was justified by obtaining it from several variation principles and other criteria.[7–9a] In this paper the PCDD for boson systems is also justified by using these criteria. The extension is not trivial, since in addition to the dangerous diagrams describing the creation of a pair of bogolons from the vacuum, new dangerous diagrams describing the creation of a single bogolon of zero momentum from the vacuum arise because the zero-momentum state is treated exactly. These additional dangerous diagrams must also be compensated by setting the sum of all diagrams leading from the vacuum to the one-bogolon state equal to zero.

In the theory of fermion systems, two slightly different forms of the PCDD were obtained, called I[7,8] and II.[9] Both of these forms satisfied the principle stated in terms of diagrams, but were obtained by different criteria and the diagrams were defined differently. The same situation arises in boson systems also, since the same general criteria can be applied. This paper will deal with the PCDD(I) and a sequel will deal with the PCDD(II).[9b]

The most physically intuitive of the criteria used in this paper is the criterion of maximum overlap between the true ground-state vector and the bogolon vacuum state.[7] This criterion is equivalent to minimizing the distance between the true ground-state vector and the bogolon vacuum state in Hilbert space. The PCDD(I) can also be obtained by eliminating the one and two bogolon excitations from the expansion of the true ground state in terms of bogolon states. The diagonalization of the reaction operator (t matrix) up to terms cubic in the bogolon operators leads to the boson form of the Brillouin–Brueckner–Bogoliubov condition[7] obtained for fermion systems and which is also equivalent to the PCDD(I). The generalization of the Hartree minimization of the ground-state energy by replacing the potential with the reaction operator, which is called the exact self-consistent-field theory,[8] also leads to the same result. The PCDD(I) can be formulated in terms of the transition density matrix and other transition amplitudes as well. The different criteria and the section of the paper where they are discussed is given in Table I.

TABLE I. Criteria for the principle of compensation of dangerous diagrams (I).

Section	Criterion
5	No divergent diagrams in the perturbation expansion of the ground-state energy.
6	Distance between the true ground state $\|0\rangle$ and the bogolon vacuum state $\|\psi_0\rangle$ is a minimum, $\| \|0\rangle - \|\psi_0\rangle \| = \text{minimum}.$
6	Maximum overlap between the true ground state $\|0\rangle$ and the bogolon vacuum state $\|\psi_0\rangle$, $\langle 0 \mid \psi_0\rangle = \text{maximum}.$
7	Transition density matrix equal to the bogolon one $\langle 0\| a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}} \|\psi_0\rangle = \langle\psi_0\| a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}} \|\psi_0\rangle,$ and $\langle 0\| a_0 \|\psi_0\rangle = \langle\psi_0\| a_0 \|\psi_0\rangle.$
8	No one- and two-bogolon contributions to the true ground-state vector, $\langle 0 \mid n \text{ bogolon}\rangle = 0$, $n = 1, 2$.
9	Bose–Brillouin–Brueckner–Bogoliubov condition, $\langle\psi_0\| t \|n \text{ bogolon}\rangle = 0,\ n = 1, 2.$
10	Exact self-consistent-field theory ($V \to t$ in the energy variation principle).

The next section defines the canonical transformation to bogolons and transforms the Hamiltonian to obtain the bogolon Hamiltonian. Section 3 gives the bogolon vacuum-state vector which is obtained in the Appendix. The unperturbed ground-state energy is minimized in Sec. 4 to obtain the compensation of the lowest-order dangerous diagrams (CLODD). After that, the next six sections are devoted to the

[4] A. Coniglio and M. Marinaro, Nuovo Cimento **48**, 249 (1967).
[5] C. W. Woo and S. K. Ma, Phys. Rev. **159**, 176 (1967).
[6] E. M. Henley and L. Wilets, Phys. Rev. **133**, B1118 (1964); Phys. Rev. Letters **11**, 326 (1963).
[7] D. H. Kobe, Phys. Rev. **140**, A825 (1965).
[8] D. H. Kobe, Ann. Phys. (N.Y.) **40**, 395 (1966).
[9a] D. H. Kobe, J. Math. Phys. **8**, 1200 (1967).
[9b] D. H. Kobe, J. Math. Phys. **9**, 1795 (1968) (following article).

various criteria for the PCDD(I) shown in Table I. Section 11 calculates the PCDD(I) to second order and discusses the reducibility of diagrams. The penultimate section is a comparison of the PCDD(I) and the PCDD(II), which will be obtained later.[9b] Finally, a comparison of the CLODD and the PCDD is made in the conclusion.

2. CANONICAL TRANSFORMATION

The Hamiltonian for a system of bosons interacting with a two-body potential is[10]

$$H = \sum_{1} (e_1 - \mu) a_1^\dagger a_1 + \tfrac{1}{2} \sum_{1234} \langle 12| v |34\rangle a_1^\dagger a_2^\dagger a_3 a_4 . \quad (2.1)$$

The kinetic energy is $e_k = k^2/2m$ where m is the mass of a particle and μ is the chemical potential. The matrix element of the two-body potential $\langle 12| v |34\rangle$ where $(1) = (\mathbf{k}_1)$, $(2) = (\mathbf{k}_2)$, etc., is assumed to be symmetric in the interchange of 1 and 2, as well as 3 and 4. The creation and annihilation operators $a_{\mathbf{k}}^\dagger$ and $a_{\mathbf{k}}$, respectively, satisfy the boson commutation relations

$$[a_{\mathbf{k}}, a_{\mathbf{l}}^\dagger] = \delta_{\mathbf{kl}},$$
$$[a_{\mathbf{k}}, a_{\mathbf{l}}] = 0 = [a_{\mathbf{l}}^\dagger, a_{\mathbf{k}}^\dagger]. \quad (2.2)$$

The Bogoliubov quasiparticles, or bogolons, have the creation and annihilation operators $\gamma_{\mathbf{k}}^\dagger$ and $\gamma_{\mathbf{k}}$, respectively. The canonical transformation which takes into account the correlations between particles of equal and opposite momentum as well as the zero-momentum condensate is[4,11]

$$a_{\mathbf{k}} = \varphi_0 \delta_{\mathbf{k}0} + u_{\mathbf{k}} \gamma_{\mathbf{k}} + v_{\mathbf{k}} \gamma_{-\mathbf{k}}^\dagger , \quad (2.3)$$

where the numbers φ_0, $u_{\mathbf{k}}$, and $v_{\mathbf{k}}$ are taken to be real. The zero-momentum state is expected to be macroscopically occupied, so the φ_0 has been added, which is approximately the square root of the number of particles in the zero-momentum state. In order for the bogolons to be bosons also, it is necessary that their creation and annihilation operators satisfy the commutation relations in Eq. (2.2). Then the coefficients in Eq. (2.3) must satisfy the following conditions:

$$u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2 = 1,$$
$$u_{\mathbf{k}} = u_{-\mathbf{k}}, \quad (2.4)$$
$$v_{\mathbf{k}} = v_{-\mathbf{k}}.$$

If the canonical transformation in Eq. (2.3) is made on Eq. (2.1) and the Hamiltonian is normal-ordered in the bogolon operators, the resulting bogolon Hamiltonian can be written as

$$H = \sum_{j,k} H_{jk}, \quad (2.5)$$

[10] See, e.g., D. H. Kobe, Am. J. Phys. **34**, 1150 (1966) for a discussion of second quantization.

[11] D. H. Kobe, Ann. Phys. (N.Y.) **47**, 15 (1968).

TABLE II. The coefficients in the Bogolon Hamiltonian.

$$H_{00} = -\mu\varphi_0^2 + \sum (e_1 - \mu) v_1^2 + \tfrac{1}{2} \langle 00| v |00\rangle \varphi_0^4 + \varphi_0^2 \sum \langle 00| v |-11\rangle u_1 v_1 + 2\varphi_0^2 \sum \langle 01| v |10\rangle v_1^2 + \tfrac{1}{2} \sum \langle 1-1| v |-22\rangle u_1 v_1 u_2 v_2 + \sum \langle 12| v |21\rangle v_1^2 v_2^2 .$$

$$h_{01} = [-\mu + \sum \langle 1-1| v |00\rangle u_1 v_1 + 2 \sum \langle 01| v |10\rangle v_1^2 + \langle 00| v |00\rangle \varphi_0^2] \varphi_0 (u_0 + v_0).$$

$$h_{11}(1,1) = U_1 (u_1^2 + v_1^2) + \Delta_1 2 u_1 v_1 .$$

$$h_{20}(1,-1) = \tfrac{1}{2}[U_1 2 u_1 v_1 + \Delta_1 (u_1^2 + v_1^2)],$$

where

$$\Delta_1 = \sum_{2} \langle 2-2| v |1-1\rangle (u_2 v_2 + \delta_{20} \varphi_0^2)$$

and

$$U_1 = e_1 - \mu + 2 \sum_{2} \langle 12| v |21\rangle (v_2^2 + \delta_{20} \varphi_0^2).$$

$$h_{30}(123) = (\varphi_0/3)[\langle 12| v |-30\rangle u_1 u_2 v_3 + (2 \leftrightarrow 3) + (1 \leftrightarrow 3)] + (u \leftrightarrow v).$$

$$h_{21}(123) = \varphi_0 \{[\langle 1-3| v |-20\rangle v_1 u_2 u_3 + (1 \leftrightarrow 2)] + \langle 12| v |30\rangle v_1 v_2 v_3\} + (u \leftrightarrow v).$$

$$h_{22}(1234) = \tfrac{1}{2}\{\langle 12| v |34\rangle u_1 u_2 u_3 u_4 + [\langle 1-3| v |-24\rangle u_1 v_2 v_3 u_4 + (3 \leftrightarrow 4)]\} + (u \leftrightarrow v).$$

$$h_{31}(1234) = \tfrac{1}{3}[\langle 12| v |-34\rangle u_1 u_2 v_3 u_4 + (2 \leftrightarrow 3) + (1 \leftrightarrow 3)] + (u \leftrightarrow v).$$

$$h_{40}(1234) = \frac{2}{4!}[\langle 12| v |-3-4\rangle u_1 u_2 v_3 v_4 + (2 \leftrightarrow 3) + (2 \leftrightarrow 4)] + (u \leftrightarrow v).$$

where $j, k = 0, 1, 2, 3, 4$ and $j + k = 0, 1, 2, 3, 4$. The term H_{jk} has j creation operators and k annihilation operators, and is of the form

$$H_{jk} = \sum_{1,2,\cdots,j+k} h_{jk}(1, 2, \cdots, j+k) \times \gamma_1^\dagger \gamma_2^\dagger \cdots \gamma_j^\dagger \gamma_{j+1} \cdots \gamma_{j+k}. \quad (2.6)$$

The coefficients h_{jk}, which are given in Table II,[12] depend on the kinetic energy, the potential energy, and the coefficients in the canonical transformation.

Since the Hamiltonian is Hermitian,

$$H_{jk} = H_{kj}^\dagger , \quad (2.7)$$

which implies that the coefficients h_{jk} satisfy

$$h_{jk}(1, 2, \cdots, j+k) = h_{kj}^*(j+k, \cdots, 2, 1). \quad (2.8)$$

Thus if one of the h_{jk} is zero, the corresponding h_{kj} is also zero. The function $h_{jk}(1, 2, 3, \cdots, j, j+1, \cdots, j+k)$ is symmetric in the first j variables, and also in the last k variables.

For the sake of later convenience Eq. (2.5) can be written as

$$H = H_0 + V, \quad (2.9)$$

[12] For fermion systems the corresponding coefficients are given in Appendix A of D. H. Kobe and W. B. Cheston, Ann. Phys. (N.Y.) **20**, 279 (1962).

where the unperturbed Hamiltonian H_0 is

$$H_0 = H_{00} + H_{11} \tag{2.10}$$

and V is the sum of the other terms in Eq. (2.5). The operator V will be called the bogolon-interaction Hamiltonian.

3. THE BOGOLON VACUUM STATE

In order to discuss the bogolon theory of boson systems further, it is important to have an explicit expression for the bogolon vacuum state $|\varphi_0\rangle$, i.e., the state of no bogolons.[13] The bogolon vacuum state contains no bogolons, so

$$\gamma_{\mathbf{k}} |\psi_0\rangle = 0 \tag{3.1}$$

for all momentum $\mathbf{k}$. It should also be normalized to unity. The bogolon vacuum state can be assumed to be generated by applying operators of the form $a_0^\dagger$ and $a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger$ to the state of no *particles*, $|\text{vac}\rangle$. As shown in the Appendix, this state has the form[14]

$$|\psi_0\rangle = C \exp\left\{\varphi_0(1 - s_0)a_0^\dagger + \tfrac{1}{2}\sum_{\mathbf{k}} s_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger\right\} |\text{vac}\rangle \tag{3.2}$$

where

$$s_{\mathbf{k}} = v_{\mathbf{k}}/u_{\mathbf{k}} \tag{3.3}$$

and the sum is over all states including the zero-momentum state. The s_0 occurs because we are treating the zero-momentum state exactly. The constant C is a normalization factor which depends on the variables $s_{\mathbf{k}}$ and φ_0, and is determined by the normalization condition

$$C^{-2} = \langle\text{vac}| \exp\left\{\varphi_0(1 - s_0)a_0 + \tfrac{1}{2}\sum_{\mathbf{k}} s_{\mathbf{k}} a_{\mathbf{k}} a_{-\mathbf{k}}\right\} \times \exp\left\{\varphi_0(1 - s_0)a_0^\dagger + \tfrac{1}{2}\sum_{\mathbf{k}} s_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger\right\} |\text{vac}\rangle. \tag{3.4}$$

In the next section we will use Eq. (3.1) to calculate the ground-state energy.

4. THE COMPENSATION OF THE LOWEST-ORDER DANGEROUS DIAGRAMS

In Bogoliubov's original theory,[1] the coefficients in the canonical transformation were determined by diagonalizing the quadratic part of the Hamiltonian. Bogoliubov only later postulated the PCDD,[2] so that his original procedure was equivalent to the compensation of the lowest-order dangerous diagrams (CLODD).

In the present theory the Hamiltonian in Eq. (2.5) can also be diagonalized up to terms cubic in the bogolon operators. The terms describing the creation or annihilation of a single bogolon from the vacuum can be set equal to zero,

$$H_{10} = 0 = H_{01}. \tag{4.1}$$

The terms describing the creation or annihilation of a pair of bogolons can also be set equal to zero,

$$H_{20} = 0 = H_{02}. \tag{4.2}$$

Bogoliubov satisfied Eq. (4.1) by choosing $\gamma_0 \equiv 0$, but in the present theory γ_0 has been retained in order to treat the zero-momentum state exactly. Thus the coefficient of γ_0 will have to be equated to zero, which will be shown to be equivalent to minimizing the unperturbed ground-state energy with respect to φ_0.[4,11] The condition in Eq. (4.2) can be satisfied by setting the coefficient of the operator $\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}$ equal to zero because of Eqs. (2.6) and (2.7). This coefficient contains terms resulting from normal-ordering the interaction, which Bogoliubov neglected. It will be shown later that Eq. (4.2) is equivalent to minimizing the unperturbed ground-state energy with respect to $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$, subject to the constraints in Eq. (2.4).

If the expectation value of the Hamiltonian in Eq. (2.9) is taken with respect to the bogolon vacuum state in Eq. (3.2), the result is

$$H_{00} = \langle\psi_0| (H_0 + V) |\psi_0\rangle \tag{4.3}$$

because of Eq. (3.1). The only term remaining in the sum of Eq. (2.5) is the one involving no creation or annihilation operators. The right side of Eq. (4.3) involves $s_{\mathbf{k}}$ and φ_0 only through the state vectors $|\psi_0\rangle$ since the total Hamiltonian is independent of them, while H_{00} involves $u_{\mathbf{k}}$, $v_{\mathbf{k}}$, and φ_0 explicitly from the normal ordering, as can be seen from Table II. The extremization of Eq. (4.3) will be shown to be a minimum in the case that bogolon interactions can be neglected.

It is convenient to use the right side of Eq. (4.3) when minimizing, since the full Hamiltonian H in Eq. (2.1) is independent of the coefficients in the canonical transformation of Eq. (2.3). The condition that the unperturbed ground-state energy be a minimum with respect to φ_0 is

$$\partial \langle\psi_0| H |\psi_0\rangle/\partial\varphi_0 = 2 \langle\psi_0| H |\partial\psi_0/\partial\varphi_0\rangle = 0, \tag{4.4}$$

since the derivative is real.[15] The differentiation of the

[13] For the fermion case, the bogolon vacuum state is just the ground-state vector used by J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957), in their theory of superconductivity.

[14] A state very similar to this one was used by E. P. Gross, Ann. Phys. (N.Y.) **9**, 292 (1960), except that he had $s_0 = 0$. The projection of this vector on the N-particle subspace would be similar to the ground-state vector used by M. Girardeau and R. Arnowitt, Phys. Rev. **113**, 755 (1959). The same state with $s_0 = 0$ was also used by F. W. Cummings and J. R. Johnston, Phys. Rev. **151**, 105 (1966).

[15] If it is complex, then the real part must be taken. In general, all of the extremum conditions obtained will also be valid for complex matrix elements of the potential if the real part of the expressions is taken.

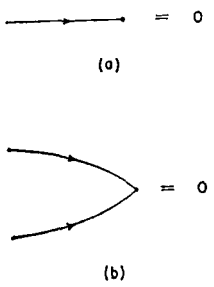


FIG. 1. Compensation of the lowest-order dangerous diagrams (CLODD). (a) Single-bogolon-state-to-vacuum; (b) two-bogolon-state-to-vacuum.

bogolon vacuum state $|\psi_0\rangle$ in Eq. (3.2) with respect to φ_0 gives

$$|\partial\psi_0/\partial\varphi_0\rangle = (1 - s_0)(a_0^\dagger - \varphi_0)\,|\psi_0\rangle, \qquad (4.5)$$

where Eq. (3.4) has been used. Thus Eqs. (4.4) and (4.5) show that the minimum in the unperturbed ground-state energy is equivalent to

$$\langle\psi_0|\, V\gamma_0^\dagger\,|\psi_0\rangle = 0 \qquad (4.6)$$

if Eqs. (2.3), (2.9), and (3.1) are used. Equation (4.6) states that the term describing the annihilation of a single bogolon is not present in the bogolon interaction V. By using Eqs. (2.5) and (2.9) it can be seen that Eq. (4.6) is equivalent to

$$h_{01} = 0 \qquad (4.7)$$

since the other terms in V will give zero. Equation (4.7) is also equivalent to Eq. (4.1) by Eqs. (2.6) and (2.7). The diagram describing the annihilation of a single bogolon is shown in the next section to be "dangerous." The compensation of this lowest-order dangerous diagram (CLODD) in Eq. (4.7) is shown graphically in Fig. 1(a).

The minimization of the unperturbed ground-state energy with respect to $s_{\mathbf{k}}$ gives the condition

$$\partial\,\langle\psi_0|\, H\,|\psi_0\rangle/\partial s_{\mathbf{k}} = 2\,\langle\psi_0|\, H\,|\partial\psi_0/\partial s_{\mathbf{k}}\rangle = 0, \qquad (4.8)$$

since it is also real.[15] The differentiation of the state vector in Eq. (3.2) with respect to $s_{\mathbf{k}}$ for $\mathbf{k} \neq 0$ gives

$$|\partial\psi_0/\partial s_{\mathbf{k}}\rangle = (a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger - u_{\mathbf{k}}v_{\mathbf{k}})\,|\psi_0\rangle \qquad (4.9)$$

if Eq. (3.4) is used. For $\mathbf{k} = 0$ the result is somewhat different,

$$|\partial\psi_0/\partial s_0\rangle = (\tfrac{1}{2}a_0^\dagger a_0^\dagger - a_0^\dagger\varphi_0 + \tfrac{1}{2}\varphi_0^2 - \tfrac{1}{2}u_0v_0)\,|\psi_0\rangle, \qquad (4.10)$$

since s_0 occurs in the coefficient of $a_0^\dagger$ in Eq. (3.2). If Eqs. (4.9) and (4.10) are used in Eq. (4.8), and the particle operators are transformed by Eq. (2.3), the minimum condition for all $\mathbf{k}$ is

$$\langle\psi_0|\, V\gamma_{\mathbf{k}}^\dagger\gamma_{-\mathbf{k}}^\dagger\,|\psi_0\rangle = 0, \qquad (4.11)$$

where V is the bogolon interaction defined in Eq. (2.9).

Equation (4.11) is the analog of the Brillouin condition[16] in Hartree–Fock theory for fermion orbitals, which states that the matrix element of the potential between the ground state and a singly-excited configuration consisting of a particle and a hole is zero. This condition was generalized to fermion systems with pairing correlations by replacing the singly-excited configuration with a pair of bogolons and the potential with the bogolon interaction. The expression obtained was the same form as Eq. (4.11), and was called the Brillouin–Bogoliubov condition.[8] Equations (4.11) and (4.6) apply to boson bogolons, and will be called the Bose–Brillouin–Bogoliubov condition (B^3 condition). In Eq. (4.6) the singly-excited configuration is replaced with a single bogolon of zero momentum, which is the lowest excited state of the system. It would be degenerate with the ground state if there were no energy gap, but in that case, a source term could be added to provide a finite energy gap. The limit as the source term goes to zero could be made at the end of the calculation. In Eq. (4.11) the singly-excited configuration is replaced with a pair of bogolons of equal but opposite momenta, which is the next class of excited states.

If Eqs. (2.5) and (2.9) are used in Eq. (4.11), the only term that will not vanish because of creation and annihilation operators will be H_{02}. Equation (4.11) is then equivalent to

$$h_{02}(\mathbf{k}, -\mathbf{k}) = 0 \qquad (4.12)$$

for all $\mathbf{k}$. This condition is completely equivalent to Eq. (4.2) because of Eqs. (2.6)–(2.8). Equation (4.12) states that the lowest-order dangerous diagram leading from the two bogolon state to the vacuum is equal to zero, which is shown graphically in Fig. 1(b). Thus it has been shown that the extremum condition on the unperturbed ground-state implies the CLODD which diagonalizes the Hamiltonian up to cubic terms in the bogolon operators.[17]

In order to show that the extremum is indeed a minimum, it is necessary to calculate the second derivatives of the ground-state energy. The ground-state energy H_{00} has its extremum value at $\{s_{\mathbf{k}}^0\}$ where we are choosing $s_{-1} = \varphi_0$ for convenience. The mean value theorem can be used at the extremum to obtain an exact expansion about the extremum

$$H_{00} = H_{00}^{(0)} + \tfrac{1}{2}\sum_{\mathbf{k},\mathbf{l}} G_{\mathbf{kl}}x_{\mathbf{k}}x_{\mathbf{l}} \qquad (4.13)$$

[16] L. Brillouin, Actualities Sci. Ind., No. 71 (1933); No. 159 (1934). See also C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[17] For fermion systems, J. G. Valatin, [Nuovo Cimento **7**, 843 (1958)] showed that minimizing the unperturbed ground-state energy resulted in the diagonalization of the Hamiltonian up to terms quartic in the bogolon operators.

TABLE III. The coefficients in the quadratic form for the ground-state energy expansion.

Element	Derivative operator on $\langle\psi_0\|H\|\psi_0\rangle$	Expression
$G_{-1,-1}$	$\partial^2/\partial\varphi_0^2$	$2(u_0 - v_0)^2 h_{11}(0, 0)$
$G_{-1,0}$	$\partial^2/\partial\varphi_0\partial s_0$	$u_0^2(u_0 - v_0)[2h_{12}(000) + 3!\, h_{03}(000)]$
[a] $G_{-1,\mathbf{k}}$	$\partial^2/\partial\varphi_0\partial s_{\mathbf{k}}$	$2u_{\mathbf{k}}^2(u_0 - v_0)[2h_{21}(\mathbf{k}, -\mathbf{k}, 0) + 3!\, h_{03}(\mathbf{k}, -\mathbf{k}, 0)]$
G_{00}	$\partial^2/\partial s_0^2$	$(u_0^4/2)[4h_{11}(00) + 4h_{22}(0000) + 4!\, h_{04}(0000)]$
[a] $G_{0\mathbf{k}}$	$\partial^2/\partial s_0\partial s_{\mathbf{k}}$	$u_0^2 u_{\mathbf{k}}^2[4h_{22}(0, 0, \mathbf{k}, -\mathbf{k}) + 4!\, h_{04}(00\mathbf{k}, -\mathbf{k})]$
[a,b] $G_{\mathbf{kl}}$	$\partial^2/\partial s_{\mathbf{k}}\partial s_{\mathbf{l}}$	$2u_{\mathbf{k}}^2 u_{\mathbf{l}}^2[4h_{22}(\mathbf{k}, -\mathbf{k}, \mathbf{l}, -\mathbf{l}) + 4!\, h_{04}(\mathbf{k}, -\mathbf{k}, \mathbf{l}, -\mathbf{l})]$
[a] $G_{\mathbf{kk}}$	$\partial^2/\partial s_{\mathbf{k}}^2$	$2u_{\mathbf{k}}^4[h_{11}(\mathbf{k}, \mathbf{k}) + h_{11}(-\mathbf{k}, -\mathbf{k}) + 4h_{22}(\mathbf{k}, -\mathbf{k}, \mathbf{k}, -\mathbf{k}) + 4!\, h_{04}(\mathbf{k}, -\mathbf{k}, \mathbf{k}, -\mathbf{k})]$

[a] $\mathbf{k} \neq 0, -1$.
[b] $\mathbf{k} \neq \mathbf{l}$.

where $H_{00}^{(0)}$ is the value of the unperturbed ground-state energy at the extremum, and the sum includes -1 as well as all momenta. The expansion variables $x_{\mathbf{k}}$ are the deviations from the extremum values

$$x_{\mathbf{k}} = s_{\mathbf{k}} - s_{\mathbf{k}}^0 \tag{4.14}$$

and the coefficients in the expansion are

$$G_{\mathbf{kl}} = \partial^2 \langle\psi_0|\, H\, |\psi_0\rangle/\partial s_{\mathbf{k}}\partial s_{\mathbf{l}}, \tag{4.15}$$

evaluated at $\{s_{\mathbf{k}}^0 + \theta x_{\mathbf{k}}\}$ where $0 \leq \theta \leq 1$.

If the sum in Eq. (4.13) is positive, then the ground-state energy is indeed a relative minimum. Thus it is necessary to show that the matrix **G** whose elements are given in Eq. (4.15) is a positive-definite matrix. The matrix elements defined in Eq. (4.15) are given in Table III. In order to show that the matrix **G** is positive-definite, it is necessary to show that all the determinants of the principal submatrices are positive. Since there are an infinite number of elements, this condition is extremely difficult to verify.

If the bogolon interaction terms in Table III can be neglected, the matrix **G** will be diagonal with positive elements. Thus the sum in Eq. (4.13) will be positive and the extremum will be a true minimum. The bogolon interaction terms would be negligible if the potential were sufficiently weak, or if the bogolon vacuum state were sufficiently close to the true ground state.

In order for the matrix element $G_{-1,-1}$ to be positive, it is necessary that

$$E_0 \equiv h_{11}(0, 0) > 0. \tag{4.16}$$

The bogolon "kinetic" energy at zero momentum must be positive! Thus in order for the extremum to be a minimum, it is necessary to have a gap in the bogolon-energy spectrum at zero momentum. Just such a gap does occur in the bogolon spectrum, however.[4,11,18] If Eq. (4.16) is multiplied by $(u_0 - v_0)^2$ and Eqs. (4.7) and (4.12) are used,[19] the condition for a minimum is that the Fourier component of the two-body potential $v(r)$ corresponding to zero momentum is positive,

$$\nu_0 = \int d\mathbf{r} v(r) > 0. \tag{4.17}$$

Therefore the integral of the interaction over all space is positive, so that it is predominantly repulsive. The condition in Eq. (4.17) is the same as for the stability of phonon modes in the original Bogoliubov theory.[1]

The surprising feature is that an energy gap is needed in the bogolon spectrum to have a minimum in the energy with respect to variation of φ_0. If the gap is eliminated, a minimum no longer exists. These conclusions can also be reached by differentiating the H_{00} in Table II which is obtained by the normal ordering of the Hamiltonian for bogolons.

5. PERTURBATION EXPANSION OF THE GROUND-STATE ENERGY

The PCDD was postulated by Bogoliubov in order to remove divergences in the perturbation expansion of the ground-state energy.[2] He stated that the sum of all the diagrams leading from the vacuum to the two-bogolon state should be set equal to zero.[20] In order to understand better why these diagrams cause difficulty, the ground-state energy will be examined. In the process it will be discovered that there are other dangerous diagrams which can cause divergences in the ground-state energy perturbation expansion. These are the diagrams that lead from the vacuum to the one-bogolon state, which can also be set equal to zero. Thus the original PCDD will be supplemented with the statement that the sum of all the diagrams leading from the vacuum to the one-bogolon state also be set equal to zero.

The true ground state $|0\rangle$ is the eigenstate of the full Hamiltonian H with the true ground-state energy $\mathcal{E}_0$ as eigenvalue

$$H|0\rangle = \mathcal{E}_0|0\rangle. \tag{5.1}$$

If the inner product is taken with the bogolon

[18] The gap was first obtained by M. Girardeau and R. Arnowitt (Ref. 14) in a similar theory.
[19] For the explicit expressions, see, e.g., Sec. III of Ref. 11.
[20] This statement is of course equivalent to the conjugate statement that the sum of all diagrams leading from the two-bogolon state to the vacuum should be set equal to zero.

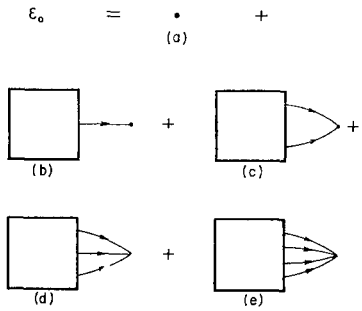


FIG. 2. Diagrammatic expansion of the ground-state energy.

vacuum, we obtain

$$\mathcal{E}_0 = \langle\psi_0|\, H\,|0\rangle \tag{5.2}$$

using intermediate normalization

$$\langle\psi_0\,|\,0\rangle = 1. \tag{5.3}$$

If Eq. (2.5) is substituted into Eq. (5.2) and Eq. (3.1) is used, the result

$$\mathcal{E}_0 = H_{00} + \langle\psi_0|\, H_{01}\,|0\rangle + \langle\psi_0|\, H_{02}\,|0\rangle + \langle\psi_0|\, H_{03}\,|0\rangle + \langle\psi_0|\, H_{04}\,|0\rangle \tag{5.4}$$

is obtained. A more explicit equation for the ground-state energy can be obtained by substituting Eq. (2.6) into Eq. (5.4). Then Eq. (5.4) can be expressed in the graphical form shown in Fig. 2, by making the one-to-one correspondences shown in Fig. 3. In order to see why divergences develop, some of these perturbation diagrams will be examined.

The original expression of the PCDD enunciated by Bogoliubov is that the sum of all the diagrams leading from the vacuum to the two-bogolon state should be set equal to zero. In order to justify this principle on the basis of the expansion of the ground-state energy, we will consider the diagram in Fig. 4. In Fig. 4 a pair of bogolons is created and scatters $n-2$ times

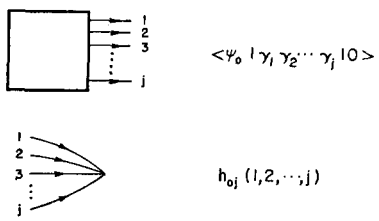


FIG. 3. Correspondence between the diagrams of Fig. 2 and mathematical quantities.

before they annihilate. The contribution to the ground-state energy from Fig. 4 is[21]

$$\sum_{\mathbf{k}} h_{02}(\mathbf{k},-\mathbf{k})h_{22}^{n-2}(\mathbf{k},-\mathbf{k},\mathbf{k},-\mathbf{k})h_{20}(\mathbf{k},-\mathbf{k})/(-2E_{\mathbf{k}})^{n-1}. \tag{5.5}$$

[21] The same argument was used in Ref. 2.

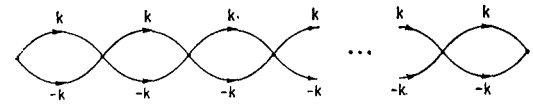


FIG. 4. An nth order contribution to the ground-state energy.

If we convert the sum to an integral and make the special choice that $v_{\mathbf{k}} = 0$ and $u_{\mathbf{k}} = 1$, then $E_{\mathbf{k}} = h_{11}(\mathbf{k},\mathbf{k}) = O(k^2)$ and the integral will diverge like k^{-2n+5} at the lower limit. In order to avoid this divergence, we can choose $h_{20}(\mathbf{k},-\mathbf{k})$ to be zero as in the CLODD of Eq. (4.12).

However, Fig. 4 is just part of a larger class of diagrams shown in Fig. 2(c). Fig. 2(c) gives the contribution of the third term on the right in Eq. (5.4),

$$\langle\psi_0|\, H_{02}\,|0\rangle = \sum_{\mathbf{k}} h_{02}(\mathbf{k},-\mathbf{k})\langle\psi_0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle, \tag{5.6}$$

which vanishes *either* if h_{02} is zero *or* the bogolon pair amplitude is zero. We will choose the condition

$$\langle\psi_0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle = 0, \tag{5.7}$$

which is equivalent to setting the box in Fig. 2(c) equal to zero. By so doing, we can also compensate

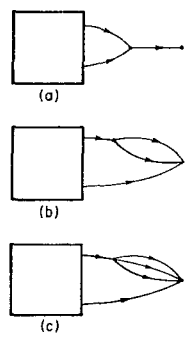


FIG. 5. Other classes of dangerous contributions to the ground-state energy.

the diagrams shown in Fig. 5(a), (b), and (c), which are subclasses of the diagrams in Fig. 2(b), (d), and (e), respectively. Equation (5.7) is just the mathematical statement of Bogoliubov's original formulation of the PCDD.

There are other diagrams in Fig. 2 that can cause divergences in the ground-state energy. One of the contributions from Fig. 2(b) is shown in Fig. 6. There is a factor h_{10} associated with the left vertex and a factor h_{01} associated with the right. The energy of a single bogolon with zero momentum is $E_0 = h_{11}(0,0)$, so the contribution of Fig. 6 to the ground-state energy is

$$-h_{01}h_{10}/E_0. \tag{5.8}$$

FIG. 6. A dangerous contribution to the ground-state energy.

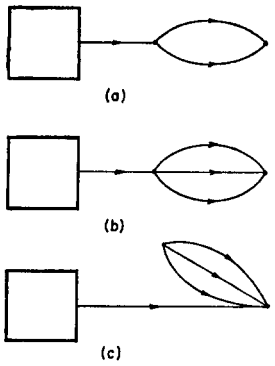


FIG. 7. Other classes of dangerous contributions to the ground-state energy.

Equation (5.8) is clearly divergent if the energy of the bogolon with zero momentum is zero. However, $E_0 = h_{11}(0, 0)$ depends on the choice of the coefficients in the transformation. If the numerator of Eq. (5.8) is taken equal to zero as in the CLODD of Eq. (4.7), the term can be eliminated altogether.

However, Fig. 6 is just part of the larger class of diagrams shown in Fig. 2(b). Fig. 2(b) gives the contribution of the second term on the right in Eq. (5.4),

$$\langle\psi_0|\, H_{01}\,|0\rangle = h_{01}\,\langle\psi_0|\,\gamma_0\,|0\rangle, \tag{5.9}$$

which vanishes if *either* h_{01} is zero *or* the bogolon amplitude is zero. We will choose the condition that the amplitude

$$\langle\psi_0|\,\gamma_0\,|0\rangle = 0, \tag{5.10}$$

which is equivalent to the vanishing of the box in Fig. 2(b). By so doing, we can compensate the larger class of dangerous diagrams shown in Fig. 7(a), (b), and (c) which are a subclass of diagrams in Fig. 2(c), (d), and (e), respectively.

Equation (5.10) is the mathematical expression for the statement that the sum of all the diagrams leading from the one bogolon state to the vacuum should be set equal to zero. This statement supplements Bogoliubov's original statement of the PCDD. He was not faced with this problem since he chose $\gamma_0 \equiv 0$.

Thus in Eqs. (5.7) and (5.10) we have the PCDD for boson systems which was obtained from the stipulation that the ground-state energy should not have any divergent diagrams. However, we are not at all certain that there are not some other divergent diagrams, or that the ground-state energy converges at all. Thus there is need to put the PCDD on a firmer foundation. In the next section we will show that the PCDD corresponds to maximizing the overlap between the true ground-state vector and the bogolon vacuum state.

6. MAXIMUM OVERLAP

A criterion which can be used to determine the coefficients in the canonical transformation of Eq. (2.3) is that the distance in Hilbert space between the true ground-state vector $|0\rangle$ and the bogolon vacuum state $|\psi\rangle$ given in Eq. (3.2) is a minimum,[7]

$$\| \,|0\rangle - |\psi_0\rangle\,\| = \text{minimum}. \tag{6.1}$$

This condition should be superior to the Hartree method used in Sec. 4 to obtain the CLODD, since it is well known that a good ground-state energy can sometimes be obtained from a poor wavefunction. However, a good wavefunction should not only give a good ground-state energy, but good expectation values to other operators as well.

On expanding Eq. (6.1) it is easily seen that the condition is the same as having maximum overlap between the true ground-state vector and the bogolon vacuum state

$$\langle 0\,|\,\psi_0\rangle = \text{maximum} \tag{6.2}$$

if the overlap is real and positive. The condition of maximum overlap was used in a previous paper to obtain the PCDD for fermion systems.[7] It was originally introduced by Brenig[22] to obtain orbitals in the independent-particle model.

The bogolon vacuum state $|\psi_0\rangle$ depends only on φ_0 and $s_{\mathbf{k}}$. The overlap is extremized with respect to variation of these two parameters, and it is shown later in this section that the extremum is indeed a maximum if bogolon interactions can be neglected. The condition that the overlap in Eq. (6.2) is an extremum with respect to the variation of φ_0 is

$$\partial\langle 0\,|\,\psi_0\rangle/\partial\varphi_0 = 0. \tag{6.3}$$

Since the true ground state $|0\rangle$ does not depend on the coefficients in the transformation of Eq. (2.3), only the state $|\psi_0\rangle$ can be varied. Equation (4.5) can thus be used, along with Eq. (2.3), to transform the particle operators to bogolon operators. If the property in Eq. (3.1) that the state $|\psi_0\rangle$ is the bogolon vacuum is used, Eq. (6.3) becomes

$$\langle 0|\,\gamma_0^\dagger\,|\psi_0\rangle = 0. \tag{6.4}$$

Equation (6.4) is just the complex conjugate of Eq. (5.10) and so is the part of the PCDD describing the compensation of the one-bogolon-to-vacuum diagrams. Equation (6.4) is shown graphically in Fig. 8(a), which is the generalization of Fig. 1(a).

The condition that the overlap in Eq. (6.2) is an extremum with respect to variation of $s_{\mathbf{k}}$ is

$$\partial\langle 0\,|\,\psi_0\rangle/\partial s_{\mathbf{k}} = 0 \tag{6.5}$$

for all $\mathbf{k}$. If Eqs. (4.9) and (4.10) are used in Eq. (6.5), along with Eqs. (2.3) and (3.1), the equation becomes

$$\langle 0|\,\gamma_{\mathbf{k}}^\dagger\gamma_{-\mathbf{k}}^\dagger\,|\psi_0\rangle = 0. \tag{6.6}$$

[22] W. Brenig, Nucl. Phys. 4, 363 (1957).

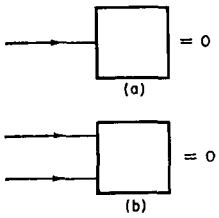


FIG. 8. The principle of compensation of dangerous diagrams (PCDD) in graphical form. (a) One-bogolon-state-to-vacuum; (b) two-bogolon-state-to-vacuum.

Equation (6.6) is just the complex conjugate of Eq. (5.7) and is the part of the PCDD describing the compensation of the two-bogolon-to-vacuum diagrams. Equation (6.6) is shown graphically in Fig. 8(b), which is the generalization of Fig. 1(b).

In order to show that the overlap is a maximum, it is necessary to examine the second derivatives of the overlap,

$$F\{s_k\} = \langle 0 \mid \psi_0 \rangle. \tag{6.7}$$

The overlap has its extremum at the values $\{s_k^0\}$ where $\varphi_0 = s_{-1}$ for convenience. The mean-value theorem can be used to obtain an exact expansion about the extremum

$$F = F^{(0)} + \tfrac{1}{2} \sum_{k,l} F_{kl} x_k x_l, \tag{6.8}$$

where $F^{(0)}$ is the value of the overlap at the extremum and the sum includes the term -1. The expansion variables x_k are the deviations in s_k from their extremum values s_k^0 as given in Eq. (4.14). The coefficients in the expansion are

$$F_{kl} = \partial^2 F / \partial s_k \partial s_l, \tag{6.9}$$

evaluated at $\{s_k^0 + \theta x_k\}$, where $0 \leq \theta \leq 1$. Since the first derivatives vanish at the extremum, they do not appear in Eq. (6.8), so it is a quadratic form. If the sum in Eq. (6.8) is always negative, the extremum is indeed a maximum. Thus it is necessary to show that the matrix **F** whose elements are given in Eq. (6.9) is negative-definite.

The matrix elements of **F** are given in Table IV. They are all extremely difficult to evaluate, since the true ground-state is not known. After the extremum values of s_k are determined from the PCDD, the matrix elements in Table IV can be evaluated by perturbation theory. The matrix **F** is thus an infinite-dimensional matrix, each matrix element of which is an infinite expansion. The determinants of all the principal submatrices of **F** must be calculated to see if the even-dimensional ones are positive, and the odd ones are negative. If they are, the matrix **F** is negative-definite, and the extremum is a true maximum. Needless to say, this program will not be carried out here.

If the bogolon interactions are negligible, which is the case when the potential is weak and/or the bogolon vacuum is close to the true ground state, all the amplitudes in Table IV describing three- or four-bogolon processes can be neglected. The remaining matrix elements are the diagonal ones which are negative if the overlap is positive. Thus in this case the overlap is a maximum. Unless the three- and four-bogolon processes are very important, this conclusion is expected to be valid in general. It is, however, conceivable that if the bogolon vacuum state is not a good approximation to the true ground state, the PCDD would not correspond to a true maximum.

TABLE IV. The coefficients in the quadratic form for the overlap expansion.

Element	Derivative operator on $\langle 0 \mid \psi_0 \rangle$	Expression
$F_{-1,-1}$	$\partial^2/\partial\varphi_0^2$	$-(u_0 - v_0)^2 \langle 0 \mid \psi_0 \rangle$
$F_{-1,0}$	$\partial^2/\partial\varphi_0 \partial s_0$	$(u_0^2/2)(u_0 - v_0) \langle 0 \mid \gamma_0^{3\dagger} \mid \psi_0 \rangle$
[a]$F_{-1,k}$	$\partial^2/\partial\varphi_0 \partial s_k$	$u_k^2(u_0 - v_0) \langle 0 \mid \gamma_0^\dagger \gamma_k^\dagger \gamma_{-k}^\dagger \mid \psi_0 \rangle$
F_{00}	$\partial^2/\partial s_0 \partial s_0$	$-(u_0^2/2)^2 \langle 0 \mid \psi_0 \rangle [2 - \langle 0 \mid \gamma_0^{4\dagger} \mid \psi_0 \rangle / \langle 0 \mid \psi_0 \rangle]$
[a]F_{0k}	$\partial^2/\partial s_0 \partial s_k$	$(u_0^2/2) u_k^2 \langle 0 \mid \gamma_0^{2\dagger} \gamma_k^\dagger \gamma_{-k}^\dagger \mid \psi_0 \rangle$
[a,b]F_{kl}	$\partial^2/\partial s_k \partial s_l$	$u_k^2 u_l^2 \langle 0 \mid \gamma_k^\dagger \gamma_{-k}^\dagger \gamma_l^\dagger \gamma_{-l}^\dagger \mid \psi_0 \rangle$
[a]F_{kk}	$\partial^2/\partial s_k^2$	$-u_k^4 \langle 0 \mid \psi_0 \rangle [1 - \langle 0 \mid (\gamma_k^\dagger \gamma_{-k}^\dagger)^2 \mid \psi_0 \rangle / \langle 0 \mid \psi_0 \rangle]$

[a] $k \neq 0, -1$.
[b] $k \neq l$.

7. TRANSITION AMPLITUDES

The criterion of maximum overlap in the fermion case was shown by Smith[23] to imply that the density matrix determined from the bogolon vacuum state be equal to the transition density matrix $\langle 0 \mid a_k^\dagger a_k \mid \psi_0 \rangle$. Maximum overlap also implies that the pair amplitude determined from the bogolon vacuum state be equal to the transition pair amplitude $\langle 0 \mid a_k a_{-k} \mid \psi_0 \rangle$. In the independent-particle model the transition density matrix is also related to maximum overlap.[24]

As another criterion for the PCDD, the density matrix calculated from the bogolon vacuum state can be equated to the transition density matrix

$$\langle \psi_0 \mid a_k^\dagger a_k \mid \psi_0 \rangle = \langle 0 \mid a_k^\dagger a_k \mid \psi_0 \rangle. \tag{7.1}$$

The expectation value of a_0 in the bogolon vacuum state can be equated to the transition one,

$$\langle \psi_0 \mid a_0 \mid \psi_0 \rangle = \langle 0 \mid a_0 \mid \psi_0 \rangle. \tag{7.2}$$

If intermediate normalization

$$\langle 0 \mid \psi_0 \rangle = 1 \tag{7.3}$$

[23] V. H. Smith, Jr., Nuovo Cimento **48**, 443 (1967).
[24] W. Kutzelnigg and V. H. Smith, Jr., J. Chem. Phys. **41**, 896 (1964).

is used, and the canonical transformation in Eq. (2.3) is substituted into Eq. (7.2), the condition

$$\langle 0|\, \gamma_0^\dagger\, |\psi_0\rangle = 0 \tag{7.4}$$

is obtained. Equation (7.1) along with Eqs. (7.3) and (7.4) implies

$$\langle 0|\, \gamma_{\mathbf{k}}^\dagger \gamma_{-\mathbf{k}}^\dagger\, |\psi_0\rangle = 0. \tag{7.5}$$

Equations (7.4) and (7.5) are just the PCDD obtained in Eqs. (6.4) and (6.6) by the maximum-overlap criterion.

A variational method could have been used here also. The sum of the squares of the differences between the density matrix calculated from the bogolon vacuum state and the transition density matrix could have been minimized. The result would have been Eq. (7.1). A similar method could have been used to obtain Eq. (7.2). It was felt that using a variational procedure to obtain Eqs. (7.1) and (7.2) would not have given any further insight, however.

An alternative criterion for the PCDD is to equate the pair amplitude calculated from the bogolon vacuum state to the transition pair amplitude

$$\langle \psi_0|\, a_{\mathbf{k}} a_{-\mathbf{k}}\, |\psi_0\rangle = \langle 0|\, a_{\mathbf{k}} a_{-\mathbf{k}}\, |\psi_0\rangle \tag{7.6}$$

and use Eq. (7.2) in addition. By substituting the canonical transformation in Eq. (2.3) into Eqs. (7.2) and (7.6), we can also obtain the PCDD in Eqs. (7.4) and (7.5).

8. ELIMINATION OF ONE- AND TWO-BOGOLON CONTRIBUTIONS TO THE TRUE GROUND STATE

One way of characterizing the PCDD in fermion systems is to eliminate the two-bogolon contributions to the true ground-state vector.[7] This condition is a generalization of the elimination of singly-excited configurations from the true ground state for the independent-particle model.[25] In the boson case, it will be shown that eliminating both one- and two-bogolon contributions to the true ground-state vector leads to the PCDD.

The true ground state can be expanded in terms of zero-, one-, two-, three-, $\cdots$, bogolon states,

$$|0\rangle = C_0\, |\psi_0\rangle + C_1 \gamma_0^\dagger\, |\psi_0\rangle + \sum_{\mathbf{k}} C_2(\mathbf{k}) \gamma_{\mathbf{k}}^\dagger \gamma_{-\mathbf{k}}^\dagger\, |\psi_0\rangle + \sum_{\mathbf{kl}} C_3(\mathbf{k}, \mathbf{l}) \gamma_{\mathbf{k}}^\dagger \gamma_{\mathbf{l}}^\dagger \gamma_{-\mathbf{k}-\mathbf{l}}^\dagger\, |\psi_0\rangle + \cdots, \tag{8.1}$$

since the set of states is assumed complete. If the one-bogolon term in Eq. (8.1) is eliminated, then the coefficient

$$C_1 = 0. \tag{8.2}$$

This coefficient can be calculated by taking the inner product of Eq. (8.1) with $\gamma_0^\dagger\, |\psi_0\rangle$. Thus Eq. (8.2) is

$$\langle \psi_0|\, \gamma_0\, |0\rangle = 0, \tag{8.3}$$

which is the same as Eq. (5.10) and the complex conjugate of Eq. (6.4). Equation (8.3) is the compensation of the dangerous diagrams describing the creation of a single bogolon from the vacuum.

The condition that there are no two-bogolon contributions to the ground-state vector in Eq. (8.1) is that

$$C_2(\mathbf{k}) = 0, \quad \text{all } \mathbf{k}. \tag{8.4}$$

This coefficient can be calculated by taking the inner product of Eq. (8.1) with $\gamma_{\mathbf{k}}^\dagger \gamma_{-\mathbf{k}}^\dagger\, |\psi_0\rangle$. Then Eq. (8.4) is

$$\langle \psi_0|\, \gamma_{\mathbf{k}} \gamma_{-\mathbf{k}}\, |0\rangle = 0, \tag{8.5}$$

which is the same as Eq. (5.7) and the complex conjugate of Eq. (6.6). Equation (8.5) is the compensation of the dangerous diagrams describing the creation of a pair of bogolons from the vacuum. The vanishing of one- and two-bogolon contributions to the true ground state is closely related to the generalization of the Bose–Brillouin–Bogoliubov condition which is discussed in the next section.

9. PARTIAL DIAGONALIZATION OF THE REACTION OPERATOR

In Sec. 4, the CLODD was obtained by a diagonalization of the Hamiltonian up to terms cubic in the bogolon operators. The PCDD can be obtained by a similar procedure, but instead of using the Hamiltonian, the reaction operator (t matrix) is diagonalized up to cubic terms. This procedure is similar to a generalization of the Brillouin condition[16] in Hartree–Fock theory by replacing the potential with the reaction operator.

The reaction operator (t matrix) is the operator such that

$$t\, |\psi_0\rangle = V\, |0\rangle, \tag{9.1}$$

where V is the bogolon interaction defined in Eq. (2.9). The reaction operator satisfies the Lippmann–Schwinger equation[26]

$$t = V + VG_0 t. \tag{9.2}$$

The propagator G_0 is

$$G_0 = \frac{1 - |\psi_0\rangle\langle\psi_0|}{\mathcal{E}_0 - H_0}, \tag{9.3}$$

where $\mathcal{E}_0$ is the exact ground-state energy, and has been defined more precisely by Löwdin.[27] From Eq. (9.1) it can be seen that the reaction operator is a

[25] R. K. Nesbet, Phys. Rev. **109**, 1632 (1958).

[26] B. A. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).

[27] P.-O. Löwdin, J. Math. Phys. **3**, 969 (1962).

many-particle operator. If the canonical transformation of Eq. (2.3) is made on it, we can write it as

$$t = \sum_{j,k} t_{jk}, \tag{9.4}$$

where t_{jk} contains j creation operators and k annihilation operators,

$$t_{jk} = \sum_{1,2,\cdots,j+k} \tau_{jk}(1, 2, \cdots, j+k) \times \gamma_1^\dagger \gamma_2^\dagger \cdots \gamma_j^\dagger \gamma_{j+1} \cdots \gamma_{j+k}. \tag{9.5}$$

If we simplify the reaction operator such that it is diagonal up to terms cubic in the bogolon operators, then we must have the operators t_{01}, t_{10}, t_{20}, and t_{02} vanish in Eq. (9.4). Thus we obtain, from Eq. (9.5),

$$\tau_{01} = 0, \tag{9.6}$$

and

$$\tau_{02}(\mathbf{k}, -\mathbf{k}) = 0. \tag{9.7}$$

Equations (9.6) and (9.7) are equivalent to

$$\langle \psi_0 |\, t\gamma_0^\dagger \,| \psi_0 \rangle = 0 \tag{9.8}$$

and

$$\langle \psi_0 |\, t\gamma_{\mathbf{k}}^\dagger \gamma_{-\mathbf{k}}^\dagger \,| \psi_0 \rangle = 0, \tag{9.9}$$

which can be seen by substituting Eqs. (9.4) and (9.5) into them.

Equations (9.8) and (9.9) are the generalization of the Bose–Brillouin–Bogoliubov condition obtained in Eqs. (4.6) and (4.11). The difference is that the bogolon interaction V has been replaced by the bogolon reaction operator t. The Brillouin condition in Hartree–Fock theory was generalized by Löwdin[28] and Nesbet[25] by replacing the potential with the reaction operator, and was called the Brillouin–Brueckner condition. In a previous paper[7] this condition was further generalized to fermion Bogoliubov quasiparticles by using a two quasiparticle state, which is a generalization of the particle-hole state, and was called the Brillouin–Brueckner–Bogoliubov condition. Equations (9.8) and (9.9) represent a further generalization to boson systems, and hence can be called the Bose–Brillouin–Brueckner–Bogoliubov condition (B^4 condition). The B^4 condition is equivalent to the usual form of the PCDD, as is shown below.

The wave operator W is defined such that

$$|0\rangle = W |\psi_0\rangle \tag{9.10}$$

and is related to the reaction operator by

$$W = 1 + G_0 t. \tag{9.11}$$

Equations (9.8) and (9.9) then imply that

$$\langle 0 |\, \gamma_0^\dagger \,| \psi_0 \rangle = 0 \tag{9.12}$$

and

$$\langle 0 |\, \gamma_{\mathbf{k}}^\dagger \gamma_{-\mathbf{k}}^\dagger \,| \psi_0 \rangle = 0, \tag{9.13}$$

which is just the PCDD obtained in Eqs. (6.4) and (6.6). Thus the partial diagonalization of the reaction operator also leads to the PCDD.

10. EXACT SELF-CONSISTENT-FIELD THEORY

In Sec. 4 it was shown that the minimization of the unperturbed ground-state energy resulted in the CLODD. Thus it appears that if the energy is the property of primary interest, the CLODD should be used instead of the PCDD. In this section it is shown that an extension of the Hartree theory which was first used by Brueckner[29] can be applied to this problem. The method was formulated more precisely by Löwdin and called the exact self-consistent-field (ESCF) theory for orbitals.[28] The application of the exact self-consistent-field theory resulted in the PCDD for the fermion case,[8] and is shown now to result in the boson PCDD.

The variational principle used here has a one-to-one correspondence with the equations of Sec. 4, but the reaction operator t defined in Eq. (9.1) is used instead of the bogolon interaction V. The exact ground-state energy $\mathcal{E}_0$ can be obtained from the reaction operator. From Eqs. (5.2), (2.9), and (9.1) the ground-state energy is

$$\mathcal{E}_0 = \langle \psi_0 |\, (H_0 + t) \,| \psi_0 \rangle, \tag{10.1}$$

where the reaction operator t is calculated from the Lippmann–Schwinger[26] equation given in Eq. (9.2). Equation (10.1) is true for any choice of the parameters in the trial-state vector $|\psi_0\rangle$, since the reaction operator adjusts itself such that the true ground-state energy is always obtained. Equation (10.1) is the analog of Eq. (4.3) for the unperturbed energy.

Since $\mathcal{E}_0$ is a constant it cannot be varied, so it is necessary to find a related function that can be. Both the trial-state vector $|\psi_0\rangle$ and the operator $H_0 + t$ depend on the choice of the parameters $u = (\varphi_0, u_{\mathbf{k}}, v_{\mathbf{k}})$ in such a way that Eq. (10.1) is a constant. However, a functional

$$\mathcal{E}(u, x) = \langle \psi_0(u) |\, [H_0(x) + t(x)] \,| \psi_0(u) \rangle \tag{10.2}$$

can be defined, where x is another independent choice of the parameters $(\varphi_0, u_{\mathbf{k}}, v_{\mathbf{k}})$. Thus, from Eq. (10.1),

$$\mathcal{E}_0 = \mathcal{E}(u, u) \tag{10.3}$$

for all values of the parameter, u.

In order to find the extremum points of the functional $\mathcal{E}(u, x)$ as a function of u, we must set the first

[28] P.-O. Löwdin, J. Math. Phys. 3, 1171 (1962).

[29] K. A. Brueckner and C. A. Levinson, Phys. Rev. 97, 1344 (1955).

derivative equal to zero,

$$\partial \mathcal{E}(u, x)/\partial u = 0, \tag{10.4}$$

which is the analog of Eqs. (4.4) and (4.8) for the unperturbed energy. In principle, this relationship can be solved to give the dependence of u on x,

$$u = u(x). \tag{10.5}$$

In order to determine the value of $\mathcal{E}$ at the minimum, the value of u in Eq. (10.5) can be used in (10.2),

$$\min \mathcal{E} = \mathcal{E}(u(x), x). \tag{10.6}$$

If this minimum value of $\mathcal{E}$ is to be the true ground-state energy $\mathcal{E}_0$, then

$$u(x) = x, \tag{10.7}$$

and the proper choice of u can be determined by setting $u = x$ in Eq. (10.4) after differentiation,

$$[\partial \mathcal{E}(u, x)/\partial u]_{u=x} = 0. \tag{10.8}$$

If the derivatives of Eq. (10.2) are taken using Eqs. (4.5), (4.9), and (4.10), the result of Eq. (10.8) is just

$$\langle \psi_0(u)| \, [H_0(u) + t(u)]\gamma_0 \, |\psi_0(u)\rangle = 0 \tag{10.9}$$

and

$$\langle \psi_0(u)| \, [H_0(u) + t(u)]\gamma_k^\dagger \gamma_{-k}^\dagger \, |\psi_0(u)\rangle = 0. \tag{10.10}$$

The bogolon operators in these equations are evaluated with the parameters u. Equations (10.9) and (10.10) just reduce to the B^4 conditions in Eqs. (9.8) and (9.9), which were shown to be equivalent to the PCDD. These conditions are the analog of the B^3 conditions in Eqs. (4.6) and (4.11). Thus this extension of Hartree theory is completely equivalent to the PCDD.

Figure 9 compares the Hartree procedure with the ESCF theory. In Sec. 4 the unperturbed ground-state energy H_{00} was minimized with respect to the parameters $u = (\varphi_0, u_k, v_k)$. The minimum condition was shown to imply the CLODD. In Fig. 9(a) the minimum of H_{00} is an upper bound to the true ground-state energy by the variation principle. Fig. 9(b) shows the ESCF theory. The condition that the functional $\mathcal{E}(u, x)$ defined in Eq. (10.2) has its minimum (or extremum) at the true ground-state energy $\mathcal{E}_0$ implies the PCDD as shown in the previous paragraphs.

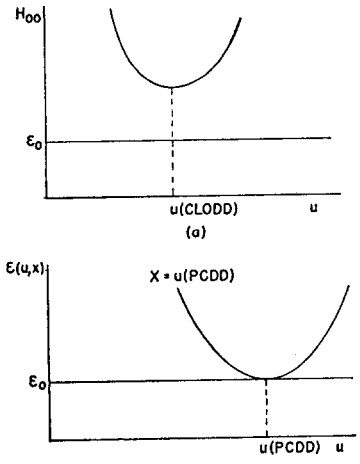


FIG. 9. A schematic comparison of the Hartree procedure and the exact self-consistent-field (ESCF) theory. (a) The minimization of the unperturbed ground-state energy H_{00} implies the compensation of the lowest-order dangerous diagrams (CLODD). (b) The condition that the minimum of the expectation value of the "effective Hamiltonian" $H_0 + t$ lie on the exact ground-state energy $\mathcal{E}_0$ implies the PCDD.

The derivative of Eq. (10.1) with respect to u is zero, so that

$$2\langle \psi_0| \, (H_0 + t) \, |\partial \psi_0/\partial u\rangle + \langle \psi_0| \, [\partial (H_0 + t)/\partial u] \, |\psi_0\rangle = 0. \tag{10.11}$$

The variational principle in Eq. (10.8) says that the first term must be zero so the second term must also vanish. Thus the variation of the ground-state wave vector holding the operator $H_0 + t$ fixed is equivalent to varying the operator holding the ground-state vector fixed.

Figure 9(b) shows that a first-order variation in the parameters of the ground-state vector holding $H_0 + t$ fixed results in a second-order variation in the energy. Equation (10.11) shows that holding the ground-state vector fixed while making a first-order change in the "effective Hamiltonian" $H_0 + t$ also makes only a second-order change in the energy. The Lippmann–Schwinger equation given in Eq. (9.2) cannot, in general, be solved for t exactly. Various approximations like, for example, including only two-body parts,[30] give some approximate reaction operator $\tilde{t}(u)$. If this approximate reaction operator can be written as the true reaction operator evaluated with a different set of parameters $u + \delta u$, i.e.,

$$\tilde{t}(u) = t(u + \delta u), \tag{10.12}$$

then using this reaction operator will result in only a second-order change in the energy from the true value $\mathcal{E}_0$. If the condition in Eq. (10.8) were not used to determine the parameters, then there would be a first-order change in the energy. Since the Lippmann–Schwinger equation is very difficult to solve exactly, this variation principle is very important in a calculation of the true ground-state energy. It is not obvious under what conditions Eq. (10.12) is satisfied, however.

The function $\mathcal{E}(u, x)$ as a function of u is bounded from below if the Hermitian "effective Hamiltonian" $H_0(x) + t(x)$ can be considered as the true Hamiltonian for some physical system. The system will then have a finite ground-state energy $\epsilon_0(x)$. If we take the

[30] This approximation is used in Refs. 25 and 29, and is the usual one.

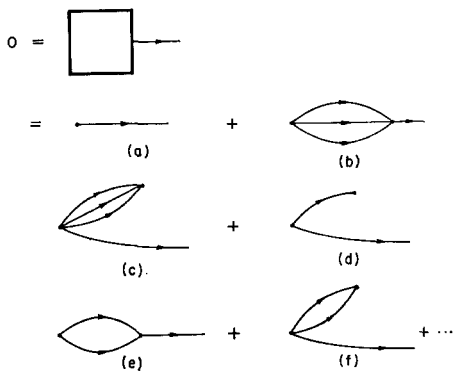


FIG. 10. The compensation of the dangerous diagrams leading from the vacuum to the one-bogolon state up to second order.

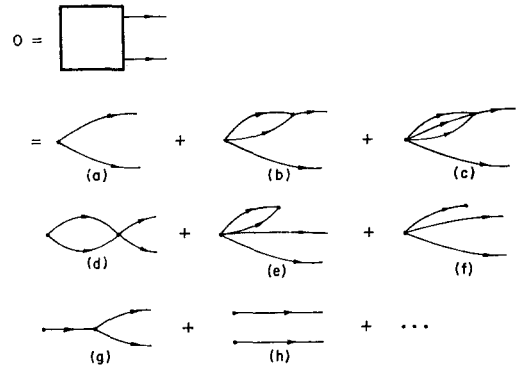


FIG. 11. The compensation of the dangerous diagrams leading from the vacuum to the two-bogolon state up to second order.

expectation value of the "effective Hamiltonian" with respect to the trial wavefunction $|\psi_0(u)\rangle$ to obtain an energy $\mathcal{E}(u, x)$ given in Eq. (10.2), we will have an upper bound on $\epsilon_0(x)$ by the energy variation principle. In other words, the true lowest eigenvalue to $H_0(x) + t(x)$ is a lower bound to $\mathcal{E}(u, x)$.

11. SECOND-ORDER PERTURBATION CALCULATIONS

The PCDD was recently used by Woo and Ma[5] in calculating corrections to the ground-state energy and excitation spectrum of a charged boson gas. They considered only one of the second-order diagrams describing the creation of a pair of bogolons from the vacuum. Thus there is some interest in obtaining the other diagrams in second order contributing to the vacuum-to-two-bogolon state and the vacuum-to-one-bogolon state. Some comments on the reducibility of the diagrams will also be made.

If the perturbation expansion for the wave operator[31] in Eqs. (9.10) and (9.11) is used in the PCDD in Eq. (5.10) for the vacuum-to-single-bogolon processes, the result to second order is

$$
\begin{aligned}
0 &= \langle\psi_0|\,\gamma_0\,|0\rangle = \langle\psi_0|\,\gamma_0 W\,|\psi_0\rangle \\
&= (-E_0)^{-1}h_{10} && \text{(a)} \\
&+ \sum (-E_0)^{-1}h_{13}(0123) \\
&\quad \times (-E_1 - E_2 - E_3)^{-1}3!\,h_{30}(123) && \text{(b)} \\
&+ \sum (-E_0)^{-1}h_{03}(123) \\
&\quad \times (-E_0 - E_1 - E_2 - E_3)^{-1}4!\,h_{40}(0123) && \text{(c)} \\
&+ \quad (-E_0)^{-1}h_{01}(-2E_0)^{-1}2!\,h_{20}(0, 0) && \text{(d)} \\
&+ \sum (-E_0)^{-1}h_{12}(012) \\
&\quad \times (-E_1 - E_2)^{-1}2!\,h_{20}(12) && \text{(e)} \\
&+ \sum (-E_0)^{-1}h_{02}(12) \\
&\quad \times (-E_0 - E_1 - E_2)^{-1}3!\,h_{30}(120) && \text{(f)} \\
&+ \cdots . && (11.1)
\end{aligned}
$$

[31] See, e.g., Ref. 27, Eq. (92).

The diagrams corresponding to the various terms in Eq. (11.1) are shown in Fig. 10. It can be seen that Fig. 10(d) involves a vertex for the annihilation of a single bogolon. This vertex can be dressed with higher-order diagrams, and thus replaced with a box describing the annihilation of a bogolon to give the vacuum. If the box describing the creation of a bogolon from the vacuum vanishes, so will the box describing the annihilation of a bogolon to give the vacuum. Figure 10(d) can be eliminated in the compensation process. It is a *reducible* diagram since it can be broken into two parts by cutting only one line.

The vertex describing the creation of two bogolons in Fig. 10(e) and the annihilation of two bogolons in Fig. 10(f) can be dressed by replacing them with boxes. If the box in Fig. 11 vanishes, then these diagrams will also, and thus do not have to be considered in the compensation process. They are also *reducible* diagrams since they can be broken into two parts by cutting two lines. Thus only the *irreducible* diagrams that cannot be broken into two parts by cutting one or two lines need be considered in the compensation process.

The propagator for a single bogolon with zero momentum is $-E_0^{-1}$, which diverges if there is no energy gap. The bare-bogolon energy does have such a gap, but the dressed bogolon presumably does not have such a gap. The energies E_1 in the denominators of Eq. (11.1) should be replaced with energies $\tilde{E}_1$ dressed in the order of the calculation. In the case that $-\tilde{E}_0^{-1}$ diverges, a source term[32] $\propto \sum_{\mathbf{k}} w_{\mathbf{k}} a_{\mathbf{k}} a_{-\mathbf{k}}$ + h.c., where $w_{\mathbf{k}}$ is arbitrary, can be added to the original Hamiltonian to insure a gap in the dressed spectrum. The coefficient of $-\tilde{E}_0^{-1}$ can be set equal to zero, and the limit $\alpha \to 0$ can be taken at the end. It is not known whether the compensation of all the diagrams of the type shown in Fig. 10 will eliminate

[32] This kind of source term is used in the theory of superconductivity, N. N. Bogoliubov, Physica **26**, S1 (1960).

the energy gap entirely,[33] but presumably a gap would not be present in an exact calculation.

A possible way of satisfying Eq. (5.8) in all orders is to use Bogoliubov's method[1] of setting $\gamma_0 \equiv 0$. This method, of course, violates the commutation relation between a_0 and $a_0^\dagger$. However, this approximation is presumably correct in the infinite volume limit for the thermodynamic functions.[34] It is not obvious that the compensation of all the dangerous diagrams leading from the vacuum to the single-bogolon state is equivalent to the Bogoliubov approximation, though. Presumably, it is an alternative but equivalent way of determining the chemical potential.

The dangerous diagrams describing the creation of two bogolons from the vacuum in Eq. (5.7) can be compensated by using the expansion for the wave operator[31] in Eq. (9.11). The result to second order is

$$
\begin{aligned}
0 &= \langle\psi_0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle = \langle\psi_0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}W\,|\psi_0\rangle \\
&= (-2E_k)^{-1}h_{20}(\mathbf{k},-\mathbf{k}) && \text{(a)}\\
&+ \sum(-2E_k)^{-1}h_{12}(\mathbf{k}12) \\
&\quad\times(-E_k-E_1-E_2)^{-1}3!\,h_{30}(-\mathbf{k}12) && \text{(b)}\\
&+ \sum(-2E_k)^{-1}h_{13}(\mathbf{k}123) \\
&\quad\times(-E_k-E_1-E_2-E_3)^{-1}4!\,h_{40}(-\mathbf{k}123) && \text{(c)}\\
&+ \sum(-2E_k)^{-1}h_{22}(\mathbf{k}-\mathbf{k}12) \\
&\quad\times(-E_1-E_2)^{-1}2!\,h_{20}(12) && \text{(d)}\\
&+ \sum\tfrac{1}{2}(-2E_k)^{-1}h_{02}(12) \\
&\quad\times(-E_1-E_2-2E_{\mathbf{k}})^{-1}4!\,h_{40}(12\mathbf{k}-\mathbf{k}) && \text{(e)}\\
&+ \tfrac{1}{2}(-2E_k)^{-1}h_{01} \\
&\quad\times(-E_0-2E_k)^{-1}3!\,h_{30}(0\mathbf{k}-\mathbf{k}) && \text{(f)}\\
&+ (-2E_k)^{-1}h_{21}(\mathbf{k},-\mathbf{k},0)(-E_0)^{-1}h_{10} && \text{(g)}\\
&+ (-2E_0)^{-1}h_{10}(-E_0)h_{10}\delta_{\mathbf{k}0} && \text{(h)}\\
&+ (\mathbf{k}\to-\mathbf{k}) + \cdots. && (11.2)
\end{aligned}
$$

Figure 11 shows the diagrams corresponding to the terms in Eq. (11.2). The diagram in Fig. 11(d) involves a vertex describing the creation of a pair of bogolons which can be dressed in higher order and replaced with a box. It vanishes if the original box describing the creation of two bogolons vanishes, and thus does not have to be considered in the compensation process. For a similar reason, the diagram in Fig. 11(e) does not have to be considered, since it is also reducible.

In the diagrams of Figs. 11(f), (g), and (h) the vertex describing the creation or annihilation of a single bogolon from the vacuum can be replaced with the box in Fig. 10, or its Hermitian conjugate. If the diagrams of Fig. 10 are compensated, then these diagrams will not contribute in Fig. 11. Thus only the irreducible diagrams need be considered in the compensation process. To second order these irreducible diagrams are Figs. 11(a), (b), and (c).[35,36]

Woo and Ma[5] considered the contribution of Figs. 11(a) and 11(b), but not Fig. 11(c). If the depletion is high, the latter diagram would be expected to be of the same order as the former, and should be investigated. The diagrams in Fig. 10 were not considered by them either, and it would be interesting to see what effect they would have.

12. RELATIONSHIP TO ANOTHER FORM OF THE PCDD

In a subsequent paper[9b] it will be shown that the criterion of minimum number of bogolons in the true ground state gives another form of the PCDD,

$$\langle 0|\,\gamma_0\,|0\rangle = 0 \qquad (12.1)$$

and

$$\langle 0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle = 0, \qquad (12.2)$$

which is called the PCDD(II). The PCDD(II) differs from the PCDD(I) in Eqs. (5.7) and (5.10) in that the bogolon vacuum state is replaced with the true ground state of the full Hamiltonian with a source term. The relationship between the PCDD(II) in Eqs. (12.1) and (12.2), and the PCDD(I) in Eqs. (5.10) and (5.7) is discussed in this section, and will also be discussed in the subsequent paper from another point of view. The diagrams used in this paper are based on the wave and reaction operator formalism, whereas the subsequent paper will use bogolon Green's function diagrams.

If the wave operator in Eq. (9.10) is substituted into the amplitude in Eq. (12.1) the result is

$$\langle 0|\,\gamma_0\,|0\rangle - \langle\psi_0|\,\gamma_0\,|0\rangle = \langle\psi_0|\,tG_0\gamma_0G_0t\,|\psi_0\rangle. \qquad (12.3)$$

The right side of Eq. (12.3) is the difference between the PCDD(II) and (I). If the perturbation expansion for the reaction operator t is substituted into the right side of Eq. (12.3), it can be seen that a bogolon with zero momentum will be annihilated before the other bogolons are annihilated. The class of diagrams that

[33] The energy gap in the bogolon spectrum was shown to vanish in the RPA by A. Coniglio and M. Marinaro [Nuovo Cimento **48**, 262 (1967)].

[34] J. Ginibre, report of work prior to publication.

[35] A similar analysis was carried out for fermion systems in Ref. 8.

[36] The above argument of reducibility does not properly take off-the-energy-shell effects into account. These effects would be present in Figs. 10(d), 10(f), and 11(h). If the gap in the spectrum is small, the effects would presumably also be small in these diagrams. On the other hand, in Figs. 11(e) and (f) the off-the-energy-shell effect would play a role, but to what extent is not known. These diagrams can be explicitly taken into account by considering them as irreducible. The concept of reducibility would then be restricted to diagrams that are on the energy shell, like Figs. 10(e), 11(d), and 11(g), and the above analysis applies exactly.

compose the right side of Eq. (12.3) is shown in Fig. 12(a). These diagrams are not dangerous in the sense that there is not necessarily any intermediate state with only one bogolon. Thus with the diagrams based on the wave and reaction operator formalism, the PCDD(I) should be used.[37]

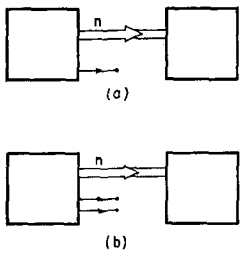


FIG. 12. Classes of diagrams which occur for the difference between the PCDD(II) and PCDD(I).

The wave operator in Eq. (9.10) can also be substituted into the amplitude in Eq. (12.2), which gives

$$\langle 0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle - \langle\psi_0|\,\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle = \langle\psi_0|\,tG_0\gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}G_0t\,|\psi_0\rangle. \quad (12.4)$$

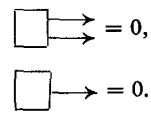
The right side of Eq. (12.4) is just the difference between the PCDD(II) and (I). It also shows that two bogolons are annihilated from a group before the others are annihilated. The class of diagrams that compose the right side of Eq. (12.4) is shown in Fig. 12(b). These diagrams are not dangerous, either, in the sense that the final state is not a two-bogolon state. Thus with the diagrams based on the wave and reaction-operator formalism, the PCDD(I) should again be used[37].

In the subsequent paper, bogolon Green's functions and their diagrams will be used. In that case, only the PCDD(II) can be expressed in terms of the Green's function diagrams. The PCDD(I) cannot be expressed in terms of the usual Green's functions for the bogolons, since they are true ground-state expectation values of time-ordered products of operators. Thus which form of the PCDD one uses will be based on the choice of perturbation formalism and the corresponding diagrams.

13. CONCLUSION

The PCDD has been shown in this paper to have many theoretical advantages over the CLODD. These advantages can be seen from Table V where the PCDD and the CLODD are compared. The CLODD was shown in Sec. 4 to be based on the minimization of the unperturbed ground-state energy. Thus it appears that from the energy point of view it is

TABLE V. Comparison between the compensation of the lowest-order dangerous diagram (CLODD) and the principle of compensation of dangerous diagrams (PCDD).

CLODD	PCDD
Lowest-order dangerous diagrams are zero: 	All dangerous diagrams are zero: 
Unperturbed ground-state energy is a minimum, $\delta\langle\psi_0\|\,(H_0+V)\,\|\psi_0\rangle = 0.$	True ground-state energy is a constant, $\delta\langle\psi_0\|\,(H_0+t)\,\|\psi_0\rangle \equiv 0.$
Hamiltonian is a constant operator, $\langle\psi_0\|\,\delta(H_0+V)\,\|\psi_0\rangle \equiv 0.$	Variation of the "effective Hamiltonian" is zero, $\langle\psi_0\|\,\delta(H_0+t)\,\|\psi_0\rangle = 0.$
Bose–Brillouin–Bogoliubov condition, $\langle\psi_0\|\,V\,\|n \text{ bogolon}\rangle = 0, \quad n = 1, 2.$	Bose–Brillouin–Brueckner–Bogoliubov condition $\langle\psi_0\|\,t\,\|n \text{ bogolon}\rangle = 0, \quad n = 1, 2.$
Distance between true ground state $\|0\rangle$ and bogolon vacuum state $\|\psi_0\rangle$ is not a minimum: $\|\,\|0\rangle - \|\psi_0\rangle\,\| \neq \min.$	Distance between the true ground state $\|0\rangle$ and the bogolon vacuum state $\|\psi_0\rangle$ is a minimum: $\|\,\|0\rangle - \|\psi_0\rangle\,\| = \min.$
Some contribution of one- and two-bogolon states to the true ground state: $\langle 0\|\,n \text{ bogolon}\rangle \neq 0, \quad n = 1, 2.$	No contribution of one- and two-bogolon states to the true ground state: $\langle 0\,\|\,n \text{ bogolon}\rangle = 0, \quad n = 1, 2.$
Lowest-order dangerous contributions to the ground-state energy vanish.	All dangerous contributions to the ground-state energy vanish.

[37] If the PCDD(I) is used, the diagrams of Fig. 12 that can be dangerous, because the intermediate state has only one or two bogolons present, vanish because of the B^4 condition in the form of Eqs. (9.6) and (9.7).

the best method and would be preferable to the PCDD obtained from the maximum-overlap criterion. However, the discussion of the exact self-consistent-field method in Sec. 10 shows that the PCDD implies that the expectation value of an "effective Hamiltonian" has the true ground-state energy as its minimum. Thus the PCDD gives the optimum way of determining the true ground-state energy, whereas the CLODD just gives an upper bound. Thus even from the energy point of view the PCDD is superior to the CLODD.

The CLODD was originally obtained by diagonalizing the Hamiltonian up to cubic terms, but the PCDD diagonalizes the reaction operator up to cubic terms. Since the reaction operator is an effective interaction describing an infinite number of bogolon-scattering processes, its diagonalization takes many of these processes into account. Thus it is more reasonable to use the PCDD.

The PCDD developed in this paper is based on maximum overlap and the related criteria given in Table I. However, another form of the PCDD, called the PCDD(II), will be developed in a subsequent paper[9b]. It is based on the criterion of minimum number of bogolons in the true ground state, and other related criteria.[9] In Sec. 12 the PCDD(II) was compared with the PCDD(I). The conclusion drawn was that the PCDD(I) is more natural to use when the diagrams based on the reaction-operator formalism are used. The PCDD(II) will be shown to be easier to formulate in terms of the diagrams based on bogolon Green's functions. Both forms of the PCDD are the same when stated in terms of diagrams, as Bogoliubov did, but the diagrams are defined differently in each case.

ACKNOWLEDGMENTS

I would like to thank Dr. A. Widom, Dr. F. Y. Wu, and Dr. E. H. Lieb for many stimulating discussions during the course of this work.

APPENDIX: THE BOGOLON VACUUM STATE

The bogolon vacuum state is assumed to be generated by application of power series in $a_k^\dagger a_{-k}^\dagger$ and $a_0^\dagger$. Thus it has the form

$$|\psi_0\rangle = C\Lambda(a_0^\dagger) \prod_1{}' \Gamma(a_1^\dagger a_{-1}^\dagger)\,|\text{vac}\rangle, \tag{A1}$$

where the prime denotes the exclusion of the zero-momentum state and only half the momentum states are included. The functions Λ and Γ are defined by the power series

$$\Lambda(z) = \sum_{n=0}^{\infty} \Lambda_n z^n \tag{A2}$$

and

$$\Gamma(z) = \sum_{n=0}^{\infty} \Gamma_n z^n, \tag{A3}$$

where z is a complex variable.

If Eq. (2.3) and its Hermitian conjugate are solved for γ_k, and Eq. (3.1) is used, the recursion relation

$$u_k\Gamma_{n+1}(n+1) - v_k\Gamma_n = 0 \tag{A4}$$

is obtained for the coefficients of Γ. If Eq. (A4) is multiplied by z^n and summed over n, the differential equation

$$\Gamma'(z) - s_k\Gamma(z) = 0, \tag{A5}$$

where s_k is given in Eq. (3.3), is obtained. The solution of Eq. (A5) with $\Gamma(0) = 1$ is

$$\Gamma(z) = \exp\{s_k z\}, \tag{A6}$$

which determines the functional form of Γ to be used in Eq. (A1).

The function Λ can be obtained in a similar way. Equation (2.3) can be solved for γ_0 and applied to Eq. (A1). The result is zero from Eq. (3.1), which gives the recursion relation

$$u_0\Lambda_{n+1}(n+1) - v_0\Lambda_{n-1} + \Lambda_n\varphi_0(v_0 - u_0) = 0 \tag{A7}$$

for the coefficients in Eq. (A2). Equation (A7) can be multiplied by z^n and converted into the differential equation

$$u_0\Lambda'(z) - [zv_0 + \varphi_0(u_0 - v_0)]\Lambda(z) = 0. \tag{A8}$$

Equation (A8) has the solution

$$\Lambda(z) = \exp\{\varphi_0(1 - s_0)z + s_0z^2/2\} \tag{A9}$$

if $\Lambda(0) = 1$. Thus the functional form of Λ in Eq. (A1) has been determined. Substituting Eqs. (A6) and (A9) into Eq. (A1) gives Eq. (3.2).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Principle of Compensation of Dangerous Diagrams for Boson Systems. II. Minimum Quasiparticle Number

DONALD H. KOBE
Department of Physics, Northeastern University, Boston, Massachusetts

(Received 6 April 1968)

The principle of compensation of dangerous diagrams (PCDD) postulated by Bogoliubov to determine the coefficients in his canonical transformation in boson systems is obtained from four criteria: (1) the number of quasiparticles in the true ground state is a minimum; (2) the "best" approximation to the true density matrix and pair amplitude is made; (3) the expectation value of an arbitrary operator is diagonalized up to terms cubic in the quasiparticle operators; (4) the most convenient starting point for dressing the quasiparticles is used. Quasiparticle Green's functions are introduced to obtain a diagrammatic expansion of the PCDD. The reducibility of the diagrams is also discussed.

1. INTRODUCTION

The *principle of compensation of dangerous diagrams* (PCDD) was enunciated by Bogoliubov[1] to determine the coefficients in his canonical transformation to quasiparticles (or bogolons) in boson systems. He argued that it is necessary to set the sum of all the diagrams leading from the vacuum to the two quasiparticle state equal to zero in order to eliminate divergences in the perturbation expansion of the ground-state energy. In a previous paper[2] it was shown that the PCDD could be obtained by maximizing the overlap between the true ground-state wavefunction and the bogolon vacuum state. Other related criteria were also discussed. These criteria lead to a form of the PCDD that could be conveniently expressed in terms of diagrams based on the wave- and reaction-operator formalism. In terms of these diagrams this form of the PCDD satisfied Bogoliubov's original statement. There are, however, diagrams leading from the vacuum to a one bogolon state that must also be compensated, in order that the zero-momentum state be treated exactly.

In this paper another form of the PCDD is obtained from other criteria.[3] The form obtained in I is called the PCDD(I), while the form obtained here is called the PCDD(II). When expressed in terms of bogolon Green's functions, the PCDD(II) also satisfies the original statement of Bogoliubov in terms of diagrams. The PCDD(II) is the natural form to use when bogolon Green's functions are to be used.

The most intuitive of the criteria used in this paper is the minimization of the number of bogolons in the true ground state. If the number of bogolons in the true ground state is small, it would be expected that the free-bogolon model in which bogolon interactions are neglected would be a good approximation to the true system. In other words, the bogolons would behave almost ideally. Mathematically, this criterion means that the expansion of the true ground state in terms of the bogolon states converges rapidly.

Other criteria for obtaining the PCDD(II) are also discussed. The "best" approximation to the true density matrix and the pair amplitude by the ones calculated using the bogolon vacuum state is another criterion. The criterion of simplifying the expectation value of an arbitrary operator is also discussed. Since the bogolon interactions should be taken into account, the most appropriate starting point for the dressing of the bogolon is used as a further criterion.

In the next section the Hamiltonian for the particles is transformed to the bogolon Hamiltonian, and the bogolon vacuum state is given. In the following four sections the criteria given in Table I are discussed. Section 7 gives the equations of motion for the bogolon Green's functions. They are used in Sec. 8 to obtain equations satisfied by the Green's functions describing the creation or annihilation of one or two bogolons. A set of four coupled integral equations is

TABLE I. Criteria for the principle of compensation of dangerous diagrams (II).

Section	Criterion
3	Expected number of quasiparticles in the true ground state is a minimum.
4	"Best" approximation to the single-particle density matrix and pair amplitude.
5	Simplification of the expectation value of an arbitrary operator.
6	Best starting point for the dressing of the quasiparticle.

[1] N. N. Bogoliubov, V. V. Tolmachev, and D. V. Shirkov, *A New Method in the Theory of Superconductivity* (Academy of Sciences of the USSR Press, Moscow, 1958), Chap. 1 (English transl.: Consultants Bureau, New York, 1959); Fortshr. Physik **6**, 605 (1958).

[2] D. H. Kobe, J. Math. Phys. **9**, 1779 (1968) (preceding paper). Henceforth this paper will be referred to as I.

[3] These criteria were used for fermion systems to obtain the PCDD by D. H. Kobe, J. Math. Phys. **8**, 1200 (1967).

obtained, from which it is shown that only the irreducible dangerous diagrams need be compensated. The PCDD(II) is compared to the PCDD(I) in Sec. 9. Finally, the results of the paper are summarized and evaluated in the conclusion.

2. CANONICAL TRANSFORMATION

The transformation originally made by Bogoliubov has the result of mixing the part of the interaction that describes the excitation of a pair of particles with equal and opposite momenta from the condensate into the kinetic energy term for the bogolons. By including terms resulting from the normal ordering of the bogolon operators, the scattering of the pair of particles an infinite number of times above the condensate can also be taken into account.[4] If the effect of triplets or quadruplets of particles scattering with each other is to be included also, it is necessary to consider the bogolon interaction processes.

The Hamiltonian for a system of bosons of mass m interacting with a two-body potential v is[5]

$$H = \sum_{\mathbf{k}} (e_k - \mu) a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \tfrac{1}{2} \sum_{1234} \langle 12| v |34\rangle a_1^\dagger a_2^\dagger a_3 a_4, \quad (2.1)$$

where $(1) = (\mathbf{k}_1)$, $(2) = (\mathbf{k}_2)$, etc. The operators $a_{\mathbf{k}}^\dagger$ and $a_{\mathbf{k}}$ are the creation and annihilation operators, respectively, for a boson with momentum $\mathbf{k}$ and satisfy the usual commutation relations. The kinetic energy is $e_k = k^2/2m$ and the chemical potential is μ. The matrix element of the potential $\langle 12| v |34\rangle$ is symmetric with respect to the interchange of 1 and 2 and also 3 and 4.

The Hamiltonian H commutes with the number operator so the number of particles is a good quantum number. However, it is invariant with respect to a gauge transformation of the first kind; that is,

$$a_{\mathbf{k}} \rightarrow a'_{\mathbf{k}} = e^{i\theta} a_{\mathbf{k}}, \quad (2.2)$$

where θ is a real constant. Thus the ground state is infinitely degenerate, corresponding to values of θ between 0 and 2π. In order to remove this degeneracy so that the usual forms of perturbation theory are applicable, it is convenient to add a source term to the Hamiltonian H to obtain[6]

$$H' = H + \alpha\sqrt{\Omega}(a_0 + a_0^\dagger). \quad (2.3)$$

The passage to the limit of infinite volume Ω is taken first and then at the end of the calculation the limit as $\alpha \rightarrow 0$ is taken.

The Hamiltonian H' no longer commutes with the number operator and is no longer invariant with respect to gauge transformations of the first kind. Thus the true ground state $|0\rangle$ of H' is not an eigenstate of the number operator. Anomalous amplitudes like $\langle 0| a_0 |0\rangle$ and $\langle 0| a_{\mathbf{k}} a_{-\mathbf{k}} |0\rangle$ can then be nonzero if the source term is present.

A generalization of the canonical transformation used by Bogoliubov[7] is

$$a_{\mathbf{k}} = \varphi_0 \delta_{\mathbf{k}0} + u_{\mathbf{k}} \gamma_{\mathbf{k}} + v_{\mathbf{k}} \gamma_{-\mathbf{k}}^\dagger, \quad (2.4)$$

where φ_0, $u_{\mathbf{k}}$, and $v_{\mathbf{k}}$ are real numbers to be determined.[4,8] In order for the bogolons to be bosons also, it is necessary for their creation and annihilation operators $\gamma_{\mathbf{k}}^\dagger$ and $\gamma_{\mathbf{k}}$, respectively, to satisfy boson commutation relations. These conditions put the following constraints on the coefficients:

$$\begin{aligned} u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2 &= 1, \\ u_{\mathbf{k}} &= u_{-\mathbf{k}}, \\ v_{\mathbf{k}} &= v_{-\mathbf{k}}. \end{aligned} \quad (2.5)$$

Since the zero-momentum state is expected to be macroscopically occupied, φ_0 has been used in Eq. (2.4), which is approximately equal to the square root of the number of particles in the zero-momentum state. The operators γ_0 were taken to be zero by Bogoliubov,[7] but are included here so that the zero-momentum state is treated exactly. In the limit of infinite volume, Bogoliubov's approximation has been shown to be exact for most potentials,[9] presumably because the γ_0 contribute only a set of measure zero. However, the γ_0 is included here to avoid any possible approximation regardless of the volume. By considering it, the condition determining the chemical potential is obtained.

If the transformation of Eq. (2.4) is made on the Hamiltonian of Eq. (2.3), the bogolon Hamiltonian can be obtained:

$$H' = \sum_{j,k} H_{jk}, \quad (2.6)$$

where $j, k = 0, 1, 2, 3, 4$, and $j + k = 0, 1, 2, 3, 4$. The term H_{jk} has j creation operators and k annihilation operators:

$$H_{jk} = \sum_{1,2,\cdots,j+k} h_{jk}(1, 2, \cdots, j+k) \times \gamma_1^\dagger \gamma_2^\dagger \cdots \gamma_j^\dagger \gamma_{j+1} \cdots \gamma_{j+k}. \quad (2.7)$$

[4] D. H. Kobe, Ann. Phys. (N.Y.) **47**, 15 (1968).
[5] For a discussion of second quantization, see, e.g., D. H. Kobe, Am. J. Phys. **34**, 1150 (1966).
[6] N. N. Bogoliubov, Physica **26**, S1 (1960).
[7] N. N. Bogoliubov, J. Phys. (USSR) **11**, 23 (1947). Reprinted in D. Pines, *The Many-Body Problem* (W. A. Benjamin, Inc., New York, 1962), p. 292.
[8] A. Coniglio and M. Marinaro, Nuovo Cimento **48**, 249 (1967); M. Girardeau, J. Math. Phys. **3**, 131 (1962); H. Ezawa, J. Math. Phys. **6**, 380 (1965).
[9] J. Ginibre (report of work prior to publication).

The coefficients h_{jk} are given in Table II of I with the value of α equal to zero. However, these coefficients can still be used if it is remembered that the anomalous amplitudes can be nonzero.

The bogolon vacuum state $|\psi_0\rangle$ is also needed. It is the state such that

$$\gamma_{\mathbf{k}} |\psi_0\rangle = 0 \tag{2.8}$$

for all $\mathbf{k}$. In I this state was determined to be

$$|\psi_0\rangle = C \exp \{a_0^\dagger \varphi_0(1 - s_0) + \tfrac{1}{2} \sum_{\mathbf{k}} s_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger\} |\text{vac}\rangle, \tag{2.9}$$

where $|\text{vac}\rangle$ is the state of no particles and

$$s_{\mathbf{k}} = v_{\mathbf{k}}/u_{\mathbf{k}}. \tag{2.10}$$

The constant C is determined from the normalization condition on $|\psi_0\rangle$. The sum in Eq. (2.9) is over *all* momentum states, including zero.

3. MINIMIZATION OF THE NUMBER OF QUASIPARTICLES

The coefficients in the canonical transformation of Eq. (2.4) have yet to be determined. They should be determined by some criterion which would make the free bogolon model a good approximation to the true system, so that bogolon interactions would not be as important.

One criterion which can be used is to choose the coefficients such that *the expected number of bogolons in the true ground state is a minimum.*[3] This condition can be formulated mathematically by taking the expectation value of the bogolon number operator in the true ground state:

$$n = \sum_{\mathbf{k}} \langle 0| \gamma_{\mathbf{k}}^\dagger \gamma_{\mathbf{k}} |0\rangle = \text{minimum}. \tag{3.1}$$

Physically, this condition means that the bogolons would behave more like an ideal gas and the bogolon interactions would not be as important. Mathematically, it means that the expansion of the true ground state in terms of bogolon states would be rapidly convergent, which is discussed in Sec. 9. Equation (3.1) leads to the PCDD(II), which is discussed from other viewpoints in the following sections.

Equation (2.4) and its Hermitian conjugate can be solved for the bogolon annihilation operator $\gamma_{\mathbf{k}}$. When it is varied with respect to $s_{\mathbf{k}}$ given in Eq. (2.10) subject to the constraints in Eq. (2.5), the annihilation operator is converted into a creation operator

$$\partial \gamma_{\mathbf{k}}/\partial s_{\mathbf{k}} = -u_{\mathbf{k}}^2 \gamma_{-\mathbf{k}}^\dagger. \tag{3.2}$$

The condition that Eq. (3.1) is an extremum is that its first derivative with respect to $s_{\mathbf{k}}$ vanishes, which gives

$$\langle 0| \gamma_{\mathbf{k}} \gamma_{-\mathbf{k}} |0\rangle = 0 \tag{3.3}$$

if the amplitude is real. (If it is complex, the real part must be taken.) Equation (3.3) is just the mathematical form of the PCDD which states that the sum of all the diagrams leading from the vacuum to the two bogolon state is zero. This condition is discussed in terms of diagrams in Sec. 8.

The expected number of bogolons n must still be minimized with respect to φ_0. The derivative of an annihilation operator with respect to φ_0 gives

$$\partial \gamma_{\mathbf{k}}/\partial \varphi_0 = -(u_0 - v_0)\delta_{0\mathbf{k}}. \tag{3.4}$$

Thus the condition obtained by differentiating n with respect to φ_0 and setting the derivative equal to zero is

$$\langle 0| \gamma_0 |0\rangle = 0 \tag{3.5}$$

if the amplitude is real, since $(u_0 - v_0) > 0$. Equation (3.5) is the mathematical form of the statement that the sum of all the diagrams leading from the vacuum to the one bogolon state is equal to zero. This condition supplements the usual statement of the PCDD.[1] Bogoliubov was not faced with these diagrams, since he chose $\gamma_0 \equiv 0$.

That the extremum is indeed a minimum can be shown by calculating the second derivatives. The mixed second derivatives can easily be shown from Eqs. (3.2) and (3.4) to be zero:

$$\partial^2 n/\partial s_{\mathbf{k}} \partial s_{\mathbf{l}} = 0 \quad \text{if} \quad \mathbf{k} \neq \mathbf{l}, \tag{3.6}$$

and

$$\partial^2 n/\partial s_{\mathbf{k}} \partial \varphi_0 = 0 \tag{3.7}$$

at the extremum. The second derivative of n with respect to $s_{\mathbf{k}}$ at the extremum is

$$\partial^2 n/\partial s_{\mathbf{k}}^2 = (2u_{\mathbf{k}})^2 \{1 + \langle 0| (\gamma_{\mathbf{k}}^\dagger \gamma_{\mathbf{k}} + \gamma_{-\mathbf{k}}^\dagger \gamma_{-\mathbf{k}}) |0\rangle\} > 0, \tag{3.8}$$

which is obviously positive. The second derivative of n with respect to φ_0 is

$$\partial^2 n/\partial \varphi_0^2 = 2(u_0 - v_0)^2 > 0, \tag{3.9}$$

which is also positive. Therefore, a Taylor series expansion of n about the extremum point shows that the value of n will increase if the variables $\{\varphi_0, s_{\mathbf{k}}\}$ deviate from their extremum values. Thus the extremum is truly a minimum.

The PCDD in Eqs. (3.3) and (3.5) is thus obtainable from a variational principle which is intuitively appealing. In the next sections the PCDD will be obtained from other equivalent criteria.

4. BEST APPROXIMATION TO THE TRUE DENSITY MATRIX AND PAIR AMPLITUDE

Another criterion that can be used to determine the coefficients in the canonical transformation is that in

some sense the theory gives the "best" approximation to the true density matrix (or occupation number) $\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle$ and the true pair amplitude $\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle$. In the independent-particle model the orbitals that give the best approximation to the true single-particle density matrix are Löwdin's natural spin orbitals.[10a]

The density matrix in the bogolon model is just

$$\langle \psi_0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|\psi_0\rangle = v_{\mathbf{k}}^2 \quad \text{if} \quad \mathbf{k} \neq 0 \tag{4.1}$$

and the pair amplitude is

$$\langle \psi_0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|\psi_0\rangle = u_{\mathbf{k}}v_{\mathbf{k}} \quad \text{if} \quad \mathbf{k} \neq 0. \tag{4.2}$$

In general, these amplitudes cannot just be equated to the true ones because $v_{\mathbf{k}}^2$ and $u_{\mathbf{k}}v_{\mathbf{k}}$ depend on only one variable $s_{\mathbf{k}}$. Different values could be obtained for $s_{\mathbf{k}}$ by equating either the true density matrix to Eq. (4.1) or the true pair amplitude to Eq. (4.2). If the numerical values of $s_{\mathbf{k}}$ obtained by equating $\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle$ and $\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle$ to Eqs. (4.1) and (4.2), respectively, are the same, then both $\langle 0|\, \gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle$ and $\langle 0|\, \gamma_{\mathbf{k}}^{\dagger}\gamma_{\mathbf{k}}\,|0\rangle$ are zero. If $\langle 0|\, \gamma_{\mathbf{k}}^{\dagger}\gamma_{\mathbf{k}}\,|0\rangle$ is zero, then $\gamma_{\mathbf{k}}\,|0\rangle = 0$ and $|0\rangle$ is the bogolon vacuum state. Thus only in the case where the free bogolon model is an exact solution can the true density matrix and pair amplitude be both equated to the model ones.[10b]

A criterion must be found for determining $s_{\mathbf{k}}$ if $\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle$ and $\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle$ are given. Several criteria are possible and in fact reasonable. However, the criterion that gives the PCDD is that $s_{\mathbf{k}}$ and φ_0 are determined to minimize the function

$$\begin{aligned}
&[\langle 0|\, a_0\,|0\rangle - \langle \psi_0|\, a_0\,|\psi_0\rangle]^2 \\
&\quad + \lim_{k\to 0}\left[\frac{\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle}{\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle + \langle 0|\, a_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}\,|0\rangle} - \frac{\langle \psi_0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|\psi_0\rangle}{\langle \psi_0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|\psi_0\rangle + \langle \psi_0|\, a_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}\,|\psi_0\rangle}\right]^2 \\
&\quad + \sum_{\mathbf{k}}{}'\left[\frac{\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle}{\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle + \langle 0|\, a_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}\,|0\rangle} - \frac{\langle \psi_0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|\psi_0\rangle}{\langle \psi_0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|\psi_0\rangle + \langle \psi_0|\, a_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}\,|\psi_0\rangle}\right]^2 \\
&\quad = \text{minimum.}
\end{aligned} \tag{4.3}$$

The prime on the summation denotes that the zero-momentum state is not included. Minimization of Eq. (4.3) with respect to $s_{\mathbf{k}}$ and use of Eqs. (4.1) and (4.2) gives

$$\frac{\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle}{\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle + \langle 0|\, a_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}\,|0\rangle} = \frac{u_{\mathbf{k}}v_{\mathbf{k}}}{v_{\mathbf{k}}^2 + u_{\mathbf{k}}^2}. \tag{4.4}$$

If Eq. (2.4) is used, Eq. (4.4) can be shown to be equivalent to

$$\langle 0|\, \gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle = 0, \tag{4.5}$$

which is the same as Eq. (3.3) for the PCDD.

The expectation value of a_0 in the bogolon vacuum state is

$$\langle \psi_0|\, a_0\,|\psi_0\rangle = \varphi_0 \tag{4.6}$$

from Eqs. (2.4) and (2.8). Minimization of Eq. (4.3) with respect to φ_0 gives

$$\varphi_0 = \langle 0|\, a_0\,|0\rangle. \tag{4.7}$$

By using Eq. (2.4), we can rewrite Eq. (4.7) as

$$\langle 0|\, \gamma_0\,|0\rangle = 0, \tag{4.8}$$

which is the same as Eq. (3.5) for the PCDD.

The extremum conditions correspond to a minimum, since the diagonal second derivatives of Eq. (4.3) are positive. The nondiagonal second derivatives of Eq. (4.3) are zero.

The momentum distribution $\langle 0|\, a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\,|0\rangle$ and the pair amplitude $\langle 0|\, a_{\mathbf{k}}a_{-\mathbf{k}}\,|0\rangle$ have been calculated by McMillan[11] for He II using a variational method with a Lennard–Jones 12–6 potential. From his values and Eq. (4.4), which is equivalent to the PCDD in Eq. (4.5), the values of $s_{\mathbf{k}}$ in Eq. (2.10) and φ_0 can be determined for He II. These values can then be used, assuming a bare *single-particle* energy $k^2/2m$, to find the energy spectrum E_k of the bogolons. The roton dip does not appear if this is done, which supports the calculation of Parry and ter Haar.[12] The bare single-particle energy should be dressed,[4] however, and the effect could be sufficiently large that the roton dip might appear.

The function in Eq. (4.3) is used in the variational principle because it results in the PCDD. The condition in Eq. (4.4) can be satisfied and yet the approximation of Eqs. (4.1) and (4.2) to the true expectation values can be poor. In this case, the validity of the theory is questionable. In order to check the theory for He II, more accurate values of the momentum distribution and pair amplitude should be calculated.

5. SIMPLIFICATION OF EXPECTATION VALUES

In his original paper, Bogoliubov[7] diagonalized the Hamiltonian up to terms cubic in the bogolon operators. He neglected the terms coming from the normal ordering of the bogolon interaction, however. In I it was shown that the maximum overlap criterion was equivalent to the diagonalization of the reaction operator up to terms cubic in the bogolon operators.

[10] (a) P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955); see also Ref. 19. (b) This discussion also applies to Sec. 4 of Ref. 3. A criterion similar to Eq. (4.3) can also be used there to circumvent this difficulty.

[11] W. L. McMillan, Phys. Rev. **138A**, 442 (1965).

[12] W. E. Parry and D. ter Haar, Ann. Phys. (N.Y.) **19**, 496 (1962).

However, too much attention can perhaps be paid to the energy operators and too little to other operators. If the expectation value of another operator is of primary interest, then attention should be focused on it. For fermion systems, it was shown that the PCDD(II) can be obtained by diagonalizing the expectation value of the quadratic part of an arbitrary operator.[3] In this section the PCDD(II) is obtained for boson systems by diagonalizing the expectation value of an arbitrary operator up to terms cubic in the bogolon operators.

An arbitrary operator Q may be a sum of one, two, three, $\cdots$ body operators. It can be easily written in terms of its matrix elements and the *particle* creation and annihilation operators.[13] If the canonical transformation of Eq. (2.4) is made on the operator and the bogolon operators are normal ordered, the result for the expectation value in the true ground state is

$$\begin{aligned}\langle 0|\, Q\,|0\rangle = {} & Q_{00} + \langle 0|\, Q_{01}\,|0\rangle + \langle 0|\, Q_{10}\,|0\rangle \\ & + \langle 0|\, Q_{11}\,|0\rangle + \langle 0|\, Q_{20}\,|0\rangle \\ & + \langle 0|\, Q_{02}\,|0\rangle + \langle 0|\, Q_{30}\,|0\rangle + \cdots . \end{aligned} \tag{5.1}$$

The operator Q_{jk} has j bogolon creation and k annihilation operators. It is given by an equation similar to Eq. (2.7), except that H_{jk} and h_{jk} are replaced by Q_{jk} and q_{jk}, respectively. The coefficient q_{jk} depends on the matrix elements of Q and the coefficients in the canonical transformation.

If the operator Q is Hermitian, then Q_{jk} will satisfy

$$Q_{jk} = Q_{kj}^{\dagger}. \tag{5.2}$$

Equation (5.2) shows that the expectation value of Q_{jk} is just the complex conjugate of the expectation value of Q_{kj}. Thus the terms with two creation or annihilation operators in Eq. (5.1) are

$$\begin{aligned}&\langle 0|\, Q_{02}\,|0\rangle + \langle 0|\, Q_{20}\,|0\rangle \\ &\qquad = 2\,\mathrm{Re}\sum_{\mathbf{k}} q_{02}(\mathbf{k}, -\mathbf{k})\,\langle 0|\, \gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle . \end{aligned} \tag{5.3}$$

These terms will vanish if

$$\langle 0|\, \gamma_{\mathbf{k}}\gamma_{-\mathbf{k}}\,|0\rangle = 0 \tag{5.4}$$

which is taken to be real. Equation (5.4) is just the PCDD obtained in Eq. (3.3).

The terms in Eq. (5.1) with only one creation or annihilation operator are

$$\langle 0|\, Q_{01}\,|0\rangle + \langle 0|\, Q_{10}\,|0\rangle = 2\,\mathrm{Re}\, q_{01}\langle 0|\, \gamma_0\,|0\rangle . \tag{5.5}$$

These terms will both vanish if

$$\langle 0|\, \gamma_0\,|0\rangle = 0, \tag{5.6}$$

which is also real. This equation is the same as Eq. (3.5) for the compensation of the new dangerous diagrams.

If Eqs. (5.3)–(5.6) are used, Eq. (5.1) becomes

$$\langle 0|\, Q\,|0\rangle = Q_{00} + \langle 0|\, Q_{11}\,|0\rangle + \langle 0|\, Q_{30}\,|0\rangle + \cdots , \tag{5.7}$$

which is considerably simplified. Although this criterion of simplicity is rather subjective and other criteria of simplicity could have been used, it is interesting that the PCDD can be obtained by using a generalization of Bogoliubov's original diagonalization argument.

6. BEST STARTING POINT FOR DRESSING THE QUASIPARTICLE

In order to dress the bogolon in higher orders in the bogolon interaction terms, it is useful to use bogolon Green's functions. By calculating the bogolon self-energy,[14] the effect of the emission and absorption of collective excitations on the energy of the bogolon can be taken into account. The bogolon interaction terms cause a shift in the bare bogolon energy, but the propagator has the same form as the bare propagator if the PCDD(II) is used.

The single-bogolon Green's function or propagator is defined as[15a]

$$\mathfrak{G}_{11}(1, 2) = i\,\langle 0|\, T\{\gamma_1\gamma_2^{\dagger}\}\,|0\rangle , \tag{6.1}$$

where the operators are in the Heisenberg picture and the operator T time orders the product with the largest (smallest) time to the left (right). If the complete set of eigenstates $|s\rangle$ of the Hamiltonian H with the eigenvalues $\mathcal{E}_s$ are inserted into Eq. (6.1), the result is

$$\begin{aligned}\mathfrak{G}_{11}(1, 2) = i\sum_s \{ & \exp\,[-i\omega_s(t_1 - t_2)] \\ & \times \langle 0|\, \gamma_1\,|s\rangle\langle s|\, \gamma_2^{\dagger}\,|0\rangle\, \theta(t_1 - t_2) \\ & + \exp\,[i\omega_s(t_1 - t_2)] \\ & \times \langle 0|\, \gamma_2^{\dagger}\,|s\rangle\langle s|\, \gamma_1\,|0\rangle\, \theta(t_2 - t_1)\}, \end{aligned} \tag{6.2}$$

where

$$\omega_s = \mathcal{E}_s - \mathcal{E}_0 \tag{6.3}$$

are the excitation energies. The Heaviside step function $\theta(t)$ is defined as

$$\theta(t) = \begin{cases} 1, & \text{if } t > 0, \\ \tfrac{1}{2}, & \text{if } t = 0, \\ 0, & \text{if } t < 0. \end{cases} \tag{6.4}$$

[13] See, e.g., D. H. Kobe, Proc. Phys. Soc.(London)**88**, 9 (1966); E. R. Pike, *ibid.* **81**, 427 (1963).

[14] D. H. Kobe, Ann. Phys. (N.Y.) **28**, 400 (1964); J. R. Schrieffer, Nucl. Phys. **35**, 363 (1962).

[15] (a) For a discussion of Green's functions and dressing see, e.g., A. A. Abrikosov, L. P. Gorkov, I. E. Dzyaloshinski, *Method of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963). (b) The normalized state $\gamma_1^{\dagger}|0\rangle/\langle 0|\, \gamma_1\gamma_1^{\dagger}\,|0\rangle^{\frac{1}{2}}$ is actually used, of course.

The first term in the sum in Eq. (6.2) is a retarded term since it vanishes if the bogolon is destroyed before it is created. The second term is the advanced part which vanishes if the bogolon is created before it is destroyed.

The single-bogolon propagator in the free-bogolon model is obtained by using the bogolon vacuum state in Eq. (6.1) and the Hamiltonian for free bogolons. The free-bogolon propagator is then just

$$G^0_{11}(1, 2) = i \exp [-iE_1(t_1 - t_2)]\delta_{12}\theta(t_1 - t_2), \quad (6.5)$$

where $E_1 = h_{11}(1,1)$ is the energy of a bare bogolon. The advanced part in Eq. (6.2) vanishes in the free-bogolon model, since an annihilation operator acts on the bogolon vacuum state.

In general, though, the bogolons are dressed by the interaction. However, if the interaction is not too strong, the approximation can be made that only the state with one bogolon excited from the true ground state $\gamma_1^\dagger |0\rangle$ with excitation energy ξ_1 contributes.[15b] If the momenta $\mathbf{k}_1$ and $\mathbf{k}_2$ are both zero, it is also necessary to include the true ground state $|0\rangle$ in the sum. In this approximation Eq. (6.2) becomes

$$\begin{aligned}\mathcal{G}_{11}(1, 2) = {} & i \langle 0| \gamma_0 |0\rangle\langle 0| \gamma_0^\dagger |0\rangle \delta_{10}\delta_{20} \\ & + i \exp [-i\xi_1(t_1 - t_2)] \\ & \times \langle 0| \gamma_1\gamma_2^\dagger |0\rangle \theta(t_1 - t_2) \\ & + i \exp [i\xi_1(t_1 - t_2)]\langle 0| \gamma_{-1}\gamma_{-1}^\dagger |0\rangle^{-1} \\ & \times \langle 0| \gamma_2^\dagger\gamma_{-1}^\dagger |0\rangle\langle 0| \gamma_{-1}\gamma_1 |0\rangle \theta(t_2 - t_1).\end{aligned} \quad (6.6)$$

The single-bogolon Green's function is coupled to the Green's functions describing the creation (annihilation) of one or two bogolons from the vacuum, as well as to higher-order bogolon Green's functions, through the equations of motion. The Green's function describing the creation of one bogolon from the vacuum is

$$\mathcal{G}_{01}(1) = i \langle 0| T\{\gamma_1^\dagger\} |0\rangle, \quad (6.7)$$

where the operator is in the Heisenberg picture. The time-ordering operator T is unnecessary for just one operator, and the propagator can be written as

$$\mathcal{G}_{01}(1) = i \langle 0| \gamma_0^\dagger |0\rangle \delta_{10}. \quad (6.8)$$

The Green's function $\mathcal{G}_{10}$ describing the annihilation of a bogolon to form the vacuum is defined in a manner similar to Eq. (6.8).

The Green's function describing the creation of a pair of bogolons from the vacuum is

$$\mathcal{G}_{02}(1, 2) = i \langle 0| T\{\gamma_1^\dagger\gamma_2^\dagger\} |0\rangle. \quad (6.9)$$

If the complete set of eigenstates $|s\rangle$ of H is inserted, Eq. (6.9) becomes

$$\begin{aligned}\mathcal{G}_{02}(1, 2) = i \sum_s \{ & \exp [-i\omega_s(t_1 - t_2)] \\ & \times \langle 0| \gamma_1^\dagger |s\rangle\langle s| \gamma_2^\dagger |0\rangle \theta(t_1 - t_2) \\ & + \exp [i\omega_s(t_1 - t_2)] \\ & \times \langle 0| \gamma_2^\dagger |s\rangle\langle s| \gamma_1^\dagger |0\rangle \theta(t_2 - t_1)\}.\end{aligned} \quad (6.10)$$

The assumption made in going from Eq. (6.2) to Eq. (6.6)—that only the state with a single bogolon excited from the true ground state $\gamma_1^\dagger |0\rangle$ with excitation energy ξ_1 contributes to the sum for $\mathbf{k}_1 \neq 0$—can also be made in Eq. (6.10).[15b] For $\mathbf{k}_1 = 0$ the vacuum state $|0\rangle$ must also be taken into account. The two-bogolon creation propagator thus becomes

$$\begin{aligned}\mathcal{G}_{02}(1, 2) = {} & i \langle 0| \gamma_0^\dagger |0\rangle\langle 0| \gamma_0^\dagger |0\rangle \delta_{10}\delta_{20} \\ & + i \exp [-i\xi_1(t_1 - t_2)] \\ & \times \langle 0| \gamma_1^\dagger\gamma_{-1}^\dagger |0\rangle \delta_{2,-1}\theta(t_1 - t_2) \\ & + i \exp [i\xi_1(t_1 - t_2)] \\ & \times \langle 0| \gamma_{-1}^\dagger\gamma_1^\dagger |0\rangle \delta_{2,-1}\theta(t_2 - t_1).\end{aligned} \quad (6.11)$$

The propagator describing the annihilation of two bogolons has a similar form.

These propagators can all be simplified by choosing the bogolon-pair amplitude

$$\langle 0| \gamma_{\mathbf{k}}\gamma_{-\mathbf{k}} |0\rangle = 0 \quad (6.12)$$

and the one-bogolon amplitude

$$\langle 0| \gamma_0 |0\rangle = 0, \quad (6.13)$$

which is just the PCDD. Both the single-bogolon creation propagator in Eq. (6.8) and the bogolon-pair creation propagator in Eq. (6.11) then vanish:

$$\begin{aligned}\mathcal{G}_{01}(1) &= 0, \\ \mathcal{G}_{02}(1, 2) &= 0,\end{aligned} \quad (6.14)$$

as do the single-bogolon and bogolon-pair annihilation propagators. The single-bogolon propagator in Eq. (6.6) becomes

$$\mathcal{G}_{11}(1, 2) = iZ_1\delta_{12} \exp [-i\xi_1(t_1 - t_2)]\theta(t_1 - t_2). \quad (6.15)$$

The renormalization factor Z_1 in Eq. (6.15) is just the coefficient of the exponential in the retarded part of Eq. (6.6):

$$Z_1 = 1 + \langle 0| \gamma_1^\dagger\gamma_1 |0\rangle. \quad (6.16)$$

Equation (6.15) for the dressed bogolon propagator is of the same form as the bare propagator in Eq. (6.5), which is expected if the bogolon interactions are not too strong. Thus the bogolon interactions do not change the form of the propagator if the PCDD in

Eqs. (6.12) and (6.13) is satisfied. The PCDD is therefore the most convenient criterion to use when considering the subsequent dressing of the bogolons.

7. BOGOLON GREEN'S FUNCTIONS

In order to discuss the expansion of the PCDD in terms of Feynman diagrams, it is necessary to derive the equations satisfied by the Green's function introduced in the last section. Then in the next section the relationship between the PCDD and the bogolon Green's functions can be investigated.

The most general bogolon Green's function which describes the annihilation of n bogolons and the creation of m bogolons is[16]

$$\mathcal{G}_{nm}(1, 2, \cdots, n, n+1, \cdots, n+m) = i \langle 0| T\{\gamma_1\gamma_2 \cdots \gamma_n\gamma^\dagger_{n+1} \cdots \gamma^\dagger_{n+m}\} |0\rangle, \quad (7.1)$$

where the expectation value is in the true ground state and the operator T time orders the bogolon operators in the Heisenberg picture. The equation satisfied by this function can be obtained essentially by differentiating it with respect to t_1 and using the Heisenberg equations of motion. A more sophisticated technique was used in a previous paper[16] to obtain the equations of motion for the fermion–bogolon Green's functions.

The equation for the Fourier transform of Eq. (7.1) is easier to represent by diagrams. The Fourier transform of Eq. (7.1) is defined as

$$G_{nm} = (2\pi)^{-n-m}\int_{-\infty}^{\infty} \cdots \int dt_1 \cdots dt_{n+m} \times \exp\left\{i\sum_{l=1}^{n+m} \omega_{-l}t_l\right\}\mathcal{G}_{nm}, \quad (7.2)$$

where the frequency ω_{-l} is defined as

$$\omega_{-l} = \begin{cases} \omega_l, & \text{if } l \leq n, \\ -\omega_l, & \text{if } l > n, \end{cases} \quad (7.3)$$

for the sake of convenience. The equation satisfied by the function G_{nm} is

$$G_{nm}(1, 2, \cdots, n+m) = \mathcal{S}_n \sum_{j,k} \int' (-2\pi j)h'_{jk}(1'2', \cdots, j+k')G^0(11') \times G_{n+k-1,m+j-1}(j+1', \cdots, j+k', 2, \cdots, n; 2', \cdots, j', n+1, \cdots, n+m) - im\mathcal{S}_n\mathcal{S}_m G^0(1, n+1) \times G_{n-1,m-1}(2, \cdots; n; n+2, \cdots, n+m), \quad (7.4)$$

where the variables $(1) = (\mathbf{k}_1, \omega_1)$, $(2) = (\mathbf{k}_2, \omega_2)$, etc., represent the momentum $\mathbf{k}$ and frequency (energy) ω of the bogolons. The integral in Eq. (7.4) is an integral (sum) over all the primed frequencies (momenta). The primes on the variable $(1')$ means (k'_1, ω'_1), for example, and $j + k' = (j + k)'$. The prime on the $h_{jk}(1, 2, \cdots, j+k)$ means that it is multiplied by a delta function for energy conservation:

$$\delta(\omega_1 + \omega_2 + \cdots + \omega_j - \omega_{j+1} - \cdots - \omega_{j+k}).$$

The sum over j and k is such that $j, k = 0, 1, 2, 3, 4$, and $j + k = 0, 1, 2, 3, 4$, but the term $j = 1, k = 1$ is specifically excluded. The operator $\mathcal{S}_m$ is the symmetrizer for m bogolons with the variables $n+1, n+2, \cdots, n+m$. The operator $\mathcal{S}_n$ is the symmetrizer for n bogolons with the variables $1, 2, 3, \cdots, n$. Thus Eq. (7.4) is symmetric in the first n variables and the last m variables, as it should be from the definition in Eq. (7.1).

The free-bogolon propagator is defined as

$$G^0(1, n+1) = (-2\pi)^{-1}(\omega_1 - E_1 + i0)^{-1}\delta_{1,n+1}\delta(\omega_1 - \omega_{n+1}), \quad (7.5)$$

where $E_1 = h_{11}(1, 1)$ and $+i0$ is an infinitesimal imaginary quantity. The Kronecker delta is for momentum conservation, and the delta function is for energy conservation.

Equation (7.4) is represented graphically in Fig. 1.

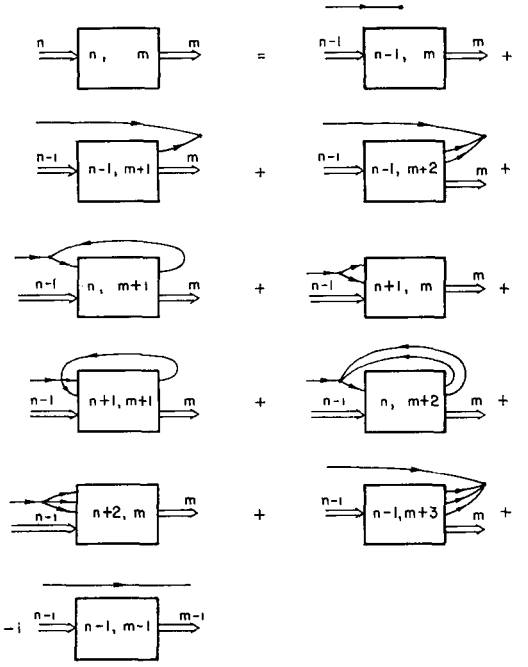


FIG. 1. The equation of motion for the Green's function G_{nm} in graphical form.

[16] For fermion systems, these Green's functions were investigated by D. H. Kobe and W. B. Cheston, Ann. Phys. (N.Y.) **20**, 279 (1962). The equations of motion for the Green's functions were obtained in this paper in a way that maintains the antisymmetrization of the function at all steps of the derivation. Here, of course, the Green's function must be symmetric in the coordinates of the bogolons destroyed and also for the bogolons created.

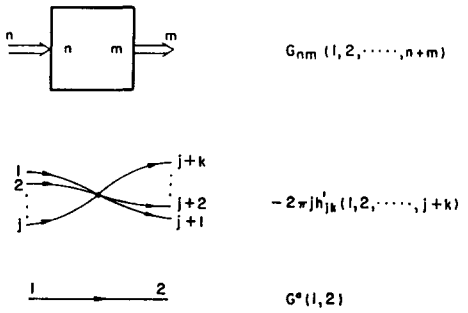


FIG. 2. The correspondence between the graphical quantities and the mathematical quantities in the Green's function equation of motion.

All the bogolon interactions have been put in specifically to show the true structure of the equation. The correspondence between the graphs and the mathematical quantities is shown in Fig. 2. Figure 1 shows that Eq. (7.4) is essentially an "expansion" in terms of the first interaction the bogolon undergoes. As the bogolon propagates, it can undergo one of the interactions given by the Hamiltonian in Eq. (2.6). Then the resulting bogolons enter a "black box" where they can interact in all possible ways. Some of the emerging bogolons interact with the incoming bogolon, but this process happens simultaneously with their entering the box, so there is no breakdown in causality.

The equation which is adjoint to Eq. (7.4) can be obtained essentially by differentiating Eq. (7.1) with respect to t_{n+1} and then taking the Fourier transform. When the equation of motion for a creation operator is used, the result is

$$\begin{aligned} G_{nm}(1,2,\cdots,n+m) &= \mathcal{S}_m \sum_{j,k}{}' \int 2\pi k h'_{jk}(1'2',\cdots,j+k')G^0(j+1',n+1) \\ &\quad \times G_{n+k-1,m+j-1}(j+k',\cdots,j+2',1,2,\cdots,n; \\ &\qquad 1',\cdots,j',n+2,\cdots,n+m) \\ &\quad - in\mathcal{S}_n\mathcal{S}_m G^0(1,n+1) \\ &\quad \times G_{n-1,m-1}(2,\cdots,n,n+2,\cdots,n+m). \quad (7.6) \end{aligned}$$

The same notation as used in Eq. (7.4) has also been used in this equation. The sum on j and k is restricted such that $j \neq 1$, $k \neq 1$, but other values such that $j, k = 0, 1, 2, 3, 4$, and $j + k = 0, 1, 2, 3, 4$ are included.

Equation (7.6) is shown graphically in Fig. 3, where the graphical quantities are defined in a way similar to Fig. 2. The exact definition can be seen by comparing Eq. (7.6) and Fig. 3. The interpretation of Fig. 3 is similar to Fig. 1, except that it represents an "expansion" in terms of the last interaction the bogolon undergoes.

Equations (7.4) and (7.6) form a hierarchy of coupled integral equations to be solved for the Green's functions. Although it is in general necessary to make some approximations to solve them, it will be seen in the next section that they can be resumed to give equations from which a reformulation of the PCDD can be obtained.

8. GREEN'S-FUNCTION FORMULATION OF THE PCDD

The PCDD as formulated in Eqs. (3.3) and (3.5) can be expressed in terms of the one-bogolon annihilation (or creation) propagator G_{10} (or G_{01}) and the two-bogolon annihilation (or creation) propagator G_{20} (or G_{02}). Equations (3.3) and (3.5) are in fact equivalent to

$$\int_{-\infty}^{\infty} \int d\omega_1 \, d\omega_2 G_{20}(1,2) = 0 \quad (8.1)$$

and

$$\int_{-\infty}^{\infty} d\omega_1 G_{10}(1) = 0, \quad (8.2)$$

which are shown graphically in Fig. 4, where the dashed lines indicate the integrations over the energy.

In order to discuss the reducibility of the diagrams it is necessary to apply the equations of motion in the last section to the functions G_{01}, G_{10}, G_{20}, and G_{02}. The higher-bogolon Green's functions can be substituted into the equations to obtain the perturbation

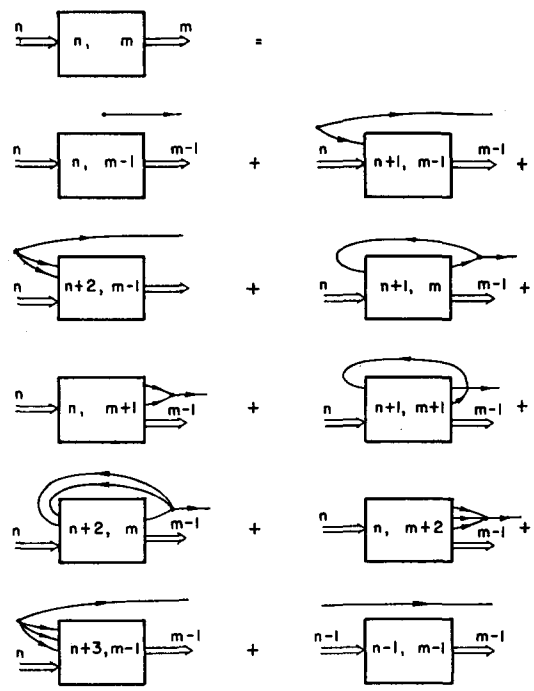


FIG. 3. The adjoint equation of motion for the Green's function G_{nm} in graphical form.

series expansion for the functions. These series can be resummed and expressed in terms of integral equations with irreducible vertex parts. The equations obtained for the four functions are shown in Figs. 5, 6, 7, and 8, respectively. The circles, which will be denoted by the symbol f with the appropriate subscripts, denote the irreducible vertex parts, which cannot be cut in two by merely cutting one or two lines. The single-bogolon propagators in Figs. 5–8 are dressed propagators. If the equations are integrated as in Eqs. (8.1) and (8.2) and set equal to zero, only the irreducible vertex parts f_{10} (f_{01}) describing the annihilation (creation) of a single bogolon and f_{20} (f_{02}) describing the annihilation (creation) of a pair of bogolons remain on the right side. They must then be zero, since the left side of the equation is zero.

However, we will assume that f_{10} (or f_{01}) and f_{20} (or f_{02}) are set equal to zero. The remaining coupled integral equations are homogeneous and, therefore, possess the trivial solution that frequency integrals of G_{10} and G_{20} are zero. Thus it is not necessary to compensate all of the dangerous diagrams, but only the irreducible ones.[17] Compensating the irreducible dangerous diagrams results in all dangerous diagrams being compensated as well. The PCDD can be reformulated in the following way: The sum of all *irreducible* diagrams leading from the vacuum to a two-bogolon state should be set equal to zero. The irreducible diagrams leading from the vacuum to a one-bogolon state are likewise compensated.[18]

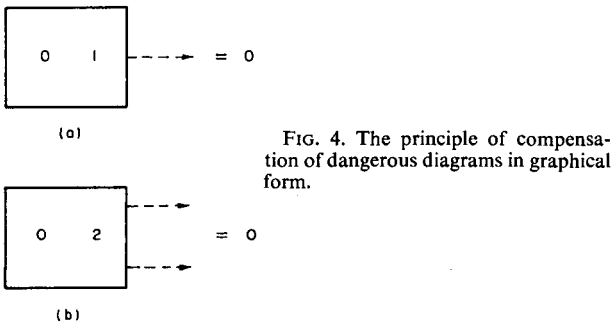


FIG. 4. The principle of compensation of dangerous diagrams in graphical form.

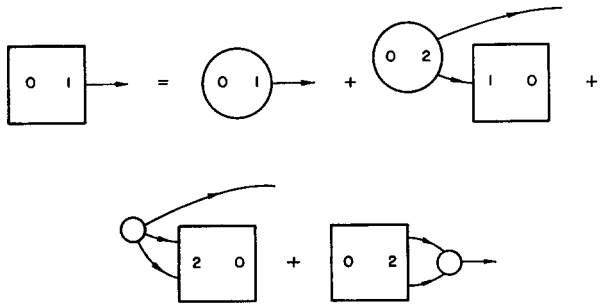


FIG. 5. The resummed integral equation for the single-bogolon creation propagator G_{01}. The outgoing line is dressed.

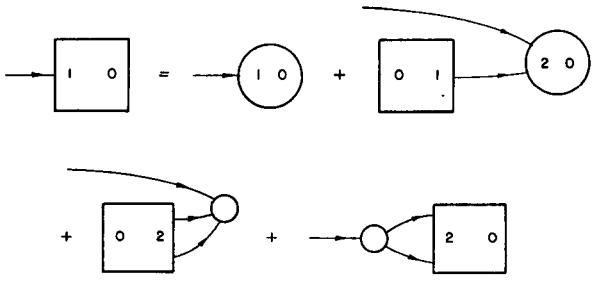


FIG. 6. The resummed integral equation for the single-bogolon annihilation propagator G_{10}. The incoming line is dressed.

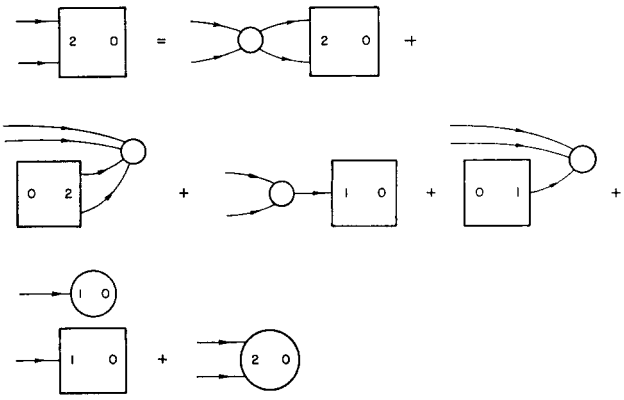


FIG. 7. The resummed integral equation for the two-bogolon creation propagator G_{02}. The outgoing lines are dressed.

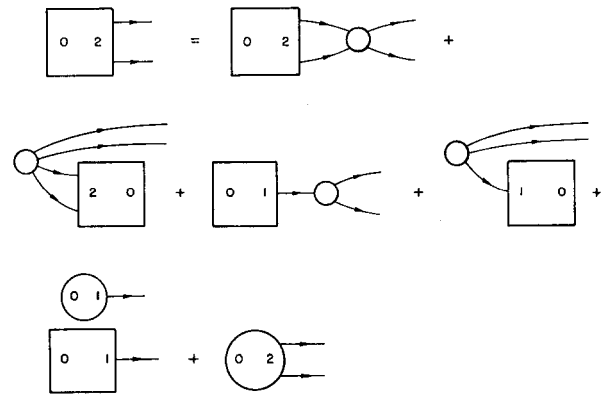


FIG. 8. The resummed integral equation for the two-bogolon annihilation propagator G_{20}. The incoming lines are dressed.

9. RELATIONSHIP BETWEEN THE TWO CONDITIONS

In a previous paper,[2] the condition for the PCDD was obtained from the maximum overlap and other criteria. These criteria give the PCDD(I)

$$\langle \psi_0 | \gamma_k \gamma_{-k} | 0 \rangle = 0 \qquad (9.1)$$

[17] In Ref. 3, the only reducible diagrams were the ladder diagrams, which were eliminated by a similar argument.

[18] In I a less rigorous argument based on ordinary perturbation theory lead to the same result.

and

$$\langle\psi_0|\,\gamma_0\,|0\rangle = 0. \tag{9.2}$$

These conditions are not in general the same as the PCDD(II) given in Eqs. (3.3) and (3.5). The relationship between the two conditions can be seen by expanding the true ground state in terms of the bogolon states

$$\begin{aligned}|0\rangle = {} & C_0\,|\psi_0\rangle + C_1\gamma_0^\dagger\,|\psi_0\rangle \\ & + \sum_{1,2} C_2(1,2)\gamma_1^\dagger\gamma_2^\dagger\,|\psi_0\rangle \\ & + \sum_{1,2,3} C_3(1,2,3)\gamma_1^\dagger\gamma_2^\dagger\gamma_3^\dagger\,|\psi_0\rangle + \cdots,\end{aligned} \tag{9.3}$$

where the sums are restricted so as to insure that the ground state has zero momentum. The PCDD(I) corresponds to

$$\begin{aligned}C_0 &= \text{maximum}, \\ C_1 &= 0, \\ C_2(1,2) &= 0,\end{aligned} \tag{9.4}$$

which was shown in I. The PCDD(II) is obtained from the criterion for the minimum number of bogolons in Eq. (3.1). By substituting Eq. (9.3) into it, we obtain

$$\begin{aligned}n = {} & 0\cdot|C_0|^2 + 1\cdot|C_1|^2 + 2\cdot\sum_{1,2} 2!\,|C_2(1,2)|^2 \\ & + 3\cdot\sum_{1,2,3} 3!\,|C_3(1,2,3)|^2 + \cdots, \\ = {} & \text{minimum}.\end{aligned} \tag{9.5}$$

Thus the series in Eq. (9.3) must converge rapidly if the number of bogolons is to be small, since the higher terms in Eq. (9.5) are more heavily weighted. In general, Eqs. (9.4) and (9.5) will not be the same. There is no reason to be concerned, however, since it is known that different criteria in the independent particle model give different orbitals.[19] Which condition to use then is a matter of taste.

However, the PCDD(II) does have the advantage if bogolon Green's functions are to be used. It is extremely difficult to express the PCDD(I) in terms of the bogolon Green's functions. For the PCDD(I) the wave- and reaction-operator formalism is most suitable. The PCDD(II) also has the advantage that to extend it to finite temperatures it is only necessary to replace the ground-state average with a thermal average. The derivation based on the minimization of the number of bogolons is unchanged in this case.

10. CONCLUSION

The four criteria presented in this paper all lead to the PCDD(II). The criterion of minimizing the expected number of bogolons is the one that is the most physically appealing, since the bogolons would be expected to behave more as an ideal gas if their number were small. The "best" approximation of the bogolon density matrix and pair amplitude to the true amplitudes as formulated in Sec. 4 focuses attention on particle amplitudes. However, there are a large number of other criteria that could be used, but do not lead to the PCDD(II). The simplification of expectation values of arbitrary operators and simplification of propagators are criteria that by themselves are somewhat subjective. However, it is interesting that these criteria also lead to the PCDD(II) as obtained by the other criteria.

Both the PCDD(I) and the PCDD(II) satisfy Bogoliubov's original statement in terms of diagrams, but the diagrams are defined differently. The PCDD(I) is most advantageous when the wave- and reaction-operator formalism is used with their associated diagrams. The exact self-consistent field approach to obtain the PCDD(I) shows that the condition also leads to the true ground-state energy. The PCDD(II), on the other hand, is most convenient when the Feynman diagrams of the Green's-function method are used. It is perhaps more convenient to discuss the self-energy of the bogolon[14] in terms of bogolon Green's functions and develop a systematic way of dressing the energy denominators. In order to deal with the problem at finite temperatures, thermal averages can be used instead of the ground-state averages. However, in the final analysis which form of the PCDD to be used is a matter of taste. In the independent particle model there are several criteria for choosing the "best" single-particle orbitals, which give different equations.[19] It is necessary to ask in which regard the orbitals are the "best." Usually, the simplest method is used, but that is only a practical consideration. In the case of bogolons the compensation of the lowest-order dangerous diagram (CLODD) is the simplest, but it was shown in I that it has theoretical drawbacks.

There is a deeper question that can be asked about the PCDD. Is the PCDD even necessary to use at all? The formalism developed by Beliaev[20] using anomalous-*particle* Green's functions is in lowest order equivalent to the CLODD (omitting pair-correlation terms coming from the normal ordering). If the particle Green's functions are extended to all orders, an exact result can be obtained if perturbation theory converges. The use of the anomalous propagators can be justified by using a source term in the Hamiltonian.[6]

[19] W. Kutzelnigg and V. H. Smith, Jr., J. Chem. Phys. **41**, 899 (1964). Equation (3) of this reference applies to the best overlap, *not* the best density orbitals.

[20] S. T. Beliaev, Zh. Eksp. Teor. Fiz. **34**, 417 (1958) [English transl.: Soviet Phys.—JETP **7**, 289 (1958)]. See also N. M. Hugenholtz and D. Pines, Phys. Rev. **116**, 489 (1959).

On the other hand, if the PCDD is used in all orders, the many-boson problem can also be solved exactly. It is necessary to use bogolon Green's functions or another type of perturbation theory when dealing with more than just the CLODD. The diagrams are also more complicated. However, if the canonical transformation to bogolons is made, it is not necessary to introduce anomalous propagators in an *ad hoc* manner. Since the bogolon interaction is "richer" than the particle interaction, it is possible perhaps to see some new and better approximations. The PCDD has a well founded physical significance and is no longer just a "recipe" for determining the coefficients. If all orders of perturbation theory are to be used, it does not make any difference what choice of coefficients is made, since the exact answer will be obtained if the series converges. In particular, the Beliaev formalism corresponds to the trivial canonical transformation with $u_k = 1$, and $v_k = 0$. However, by using the PCDD to determine the coefficients, the series will converge more rapidly.

There is an advantage to having more than one way of looking at a problem. It is even better when the different methods can be extended to show that they are essentially equivalent. Which method to use then depends on the problem to be handled and the inclination of the user.

ACKNOWLEDGMENTS

I would like to thank Dr. A. Widom, Dr. F. Y. Wu, and Dr. E. H. Lieb for valuable and stimulating discussions during the course of this work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Relativistic Stochastic Processes

RÉMI HAKIM
*Laboratoire de Physique Théorique et Hautes Energies, Orsay, France**†

(Received 24 January 1968)

Relativistic stochastic processes in μ-space are defined and studied in a completely (and manifestly) covariant manner, without particularizing the time variable. It is shown that a number of usual definitions such as "Gaussian process," etc., cannot be given a fully invariant meaning. Markovian processes are also studied. We find anew, as a particular case, results already obtained by Łopuszański [Acta Phys., Polon. **12**, 87 (1953)] in the case of Markovian processes in Minkowski space–time. Several suggestions are made to generalize these results.

1. INTRODUCTION

Beside their intrinsic interest, relativistic stochastic processes may be involved in a series of semiphenomenological theories. For instance they permit the establishment of a theory of relativistic irreversible processes and hence relativistic Onsager relations. They also permit one to give a probabilistic interpretation of various Fokker–Planck equations considered when dealing with relativistic plasmas. Furthermore, the study of relativistic Brownian motion might be of importance in some theoretical problems.[1] In two further papers, the present theory will be applied to problems of turbulence of a relativistic plasma and to an acceleration mechanism of cosmic rays of extragalactic origin.

In this paper we limit our study to relativistic stochastic processes in μ-space, i.e., in a one-particle phase space. In fact, the generalization to more complicated spaces is straightforward, although the notations might be rather sophisticated. For all practical purposes these processes are sufficient. However, the definitions we give are not the only possible ones (for more general spaces) and therefore we give an outline of other possibilities.

In Sec. 2, we give the basic definitions of stochastic processes in μ-space. However, these definitions do not correspond to the modern and precise mathematical definitions[2] used nowadays, but rather to an "old" point of view,[3,4] although it is more useful in

* Laboratoire associé au Centre Nationale de la Recherche Scientifique (CNRS).
† Postal address: Laboratoire de Physique Théorique et Hautes Energies, Bâtiment 211, Faculté des Sciences, 91-Orsay, France.

[1] E. Nelson, Phys. Rev. **150**, 1079 (1966).

[2] See, e.g., J. Neveu, *Bases mathématiques du calcul des probabilités* (Masson et Cie., Paris, 1964).

[3] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, Science Publishers, New York, 1963).

[4] Ming Chen Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).

On the other hand, if the PCDD is used in all orders, the many-boson problem can also be solved exactly. It is necessary to use bogolon Green's functions or another type of perturbation theory when dealing with more than just the CLODD. The diagrams are also more complicated. However, if the canonical transformation to bogolons is made, it is not necessary to introduce anomalous propagators in an *ad hoc* manner. Since the bogolon interaction is "richer" than the particle interaction, it is possible perhaps to see some new and better approximations. The PCDD has a well founded physical significance and is no longer just a "recipe" for determining the coefficients. If all orders of perturbation theory are to be used, it does not make any difference what choice of coefficients is made, since the exact answer will be obtained if the series converges. In particular, the Beliaev formalism corresponds to the trivial canonical transformation with $u_k = 1$, and $v_k = 0$. However, by using the PCDD to determine the coefficients, the series will converge more rapidly.

There is an advantage to having more than one way of looking at a problem. It is even better when the different methods can be extended to show that they are essentially equivalent. Which method to use then depends on the problem to be handled and the inclination of the user.

ACKNOWLEDGMENTS

I would like to thank Dr. A. Widom, Dr. F. Y. Wu, and Dr. E. H. Lieb for valuable and stimulating discussions during the course of this work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Relativistic Stochastic Processes

RÉMI HAKIM
*Laboratoire de Physique Théorique et Hautes Energies, Orsay, France**†

(Received 24 January 1968)

Relativistic stochastic processes in μ-space are defined and studied in a completely (and manifestly) covariant manner, without particularizing the time variable. It is shown that a number of usual definitions such as "Gaussian process," etc., cannot be given a fully invariant meaning. Markovian processes are also studied. We find anew, as a particular case, results already obtained by Łopuszański [Acta Phys., Polon. **12**, 87 (1953)] in the case of Markovian processes in Minkowski space–time. Several suggestions are made to generalize these results.

1. INTRODUCTION

Beside their intrinsic interest, relativistic stochastic processes may be involved in a series of semiphenomenological theories. For instance they permit the establishment of a theory of relativistic irreversible processes and hence relativistic Onsager relations. They also permit one to give a probabilistic interpretation of various Fokker–Planck equations considered when dealing with relativistic plasmas. Furthermore, the study of relativistic Brownian motion might be of importance in some theoretical problems.[1] In two further papers, the present theory will be applied to problems of turbulence of a relativistic plasma and to an acceleration mechanism of cosmic rays of extragalactic origin.

In this paper we limit our study to relativistic stochastic processes in μ-space, i.e., in a one-particle phase space. In fact, the generalization to more complicated spaces is straightforward, although the notations might be rather sophisticated. For all practical purposes these processes are sufficient. However, the definitions we give are not the only possible ones (for more general spaces) and therefore we give an outline of other possibilities.

In Sec. 2, we give the basic definitions of stochastic processes in μ-space. However, these definitions do not correspond to the modern and precise mathematical definitions[2] used nowadays, but rather to an "old" point of view,[3,4] although it is more useful in

* Laboratoire associé au Centre Nationale de la Recherche Scientifique (CNRS).
† Postal address: Laboratoire de Physique Théorique et Hautes Energies, Bâtiment 211, Faculté des Sciences, 91-Orsay, France.

[1] E. Nelson, Phys. Rev. **150**, 1079 (1966).

[2] See, e.g., J. Neveu, *Bases mathématiques du calcul des probabilités* (Masson et Cie., Paris, 1964).

[3] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, Science Publishers, New York, 1963).

[4] Ming Chen Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).

physics. However, our considerations may probably be made as rigorous as we want. In Sec. 3, we study a case of particular importance, the case of relativistic Markovian processes. Section 4 is devoted to stochastic processes in Minkowski space–time. Section 5 is mainly concerned with a brief discussion of our results. In the Appendix we give a completely different approach based on the use of proper time which we illustrate in treating the case of a particle embedded in a random force field.

In this paper we mainly use the notation of a previous article.[5] However, let us recall that

$$g^{00} = -g^{ii} = +1, \quad g^{\mu\nu} = 0 \quad \text{for} \quad \mu \neq \nu \qquad (\mu, \nu = 0, \cdots, 3;\ i = 1, \cdots, 3),$$
$$c = 1$$

(where $g^{\mu\nu}$ is the metric tensor of Minkowski space–time). The Einstein convention is also used.

Finally we want to warn our readers that the use of classical mathematical symbols throughout this paper is adopted for the sake of clarity and concision. It is not intended to give our results a mystifying appearance of mathematical rigor. Our approach is mainly a physicist's approach, although we are generally conscious of the mathematical problems we have discarded.

2. RELATIVISTIC STOCHASTIC PROCESSES IN μ-SPACE

Let us consider the problem of a particle whose motions are random. The motions of this particle are random for different reasons. For instance, the particle may be embedded in an external random force field. The "initial data" necessary to specify completely the behavior of the particle may also be random, an "observer" being unprecise while performing his "measurements." Another example of a random particle is the one of a *test particle* (that is, a generic particle of a gas), whose motions are assumed to represent more or less the main properties of the system under study.

A convenient way of describing this problem is by using what is commonly called the μ-space, i.e., the one-particle phase space. As μ-space we choose[6]

$$\mu = \mathcal{M}^4 \times U^4, \tag{1}$$

where $\mathcal{M}^4$ is the Minkowski space–time, while U^4 is the four-dimensional 4-velocity space.[6] In fact, the choice (1) for the μ-space is not completely general: we implicitly assume that we are interested only in space–time and velocity properties of physical quantities. However, in all that follows we limit ourselves to (1), the extension of our results to more general μ-spaces being straightforward (for instance, we could add the acceleration space,[7] etc.). In what follows, x^A $(A = 1, \cdots, 8)$ denotes an arbitrary coordinate system in μ space; for instance, the usual coordinates in μ-space are $(x^\mu, u^\mu) \equiv x^A$.

Physics imposes that the trajectories of a random particle (or of the stochastic process) be timelike (in Minkowski space–time). However, there are also random processes that are not related to "particles" and hence whose trajectories may be spacelike in the space under consideration. We shall come back to that question in the sequel. In 4-velocity space, trajectories are spacelike by construction. In what follows we call a trajectory in μ-space "timelike" when its projection in Minkowski space–time is timelike. "Spacelike" trajectories in μ-space are defined in the same way. Similarly, a 7-dimensional surface in μ-space is called "spacelike" when its normal is a "timelike" 8-vector.

Definition of a Relativistic Stochastic Process in μ-Space

Let now Σ^7 be a spacelike hypersurface embedded in μ-space (for instance, $\Sigma^7 = S^3 \times U^4$, with $S^3 \subset \mathcal{M}^4$ and S^3 an arbitrary spacelike hypersurface). Furthermore we impose that Σ^7 cuts all the possible timelike trajectories of the process in μ-space. As we explained at length in two preceding papers,[5,8] "physical space at a given time" is represented in Minkowski space–time by an *arbitrary* spacelike hypersurface so that we have to deal with a spacelike hypersurface Σ^7 in μ-space. Physical measures or observations always refer to such arbitrary spacelike hypersurfaces. Therefore we set the problem in the following way. Let Δ be a Lebesgue-measurable subset of Σ^7. We want to define the probabilities

$$\text{Prob}\,\{x^A \in \Delta \subset \Sigma^7 \subset \mu\}, \quad \forall\Delta. \tag{2}$$

These probabilities are completely determined by the distribution $\mathfrak{W}_1(x^A)$ through

$$\text{Prob}\,\{x^A \in \Delta \subset \Sigma^7 \subset \mu\} = \int_{\Delta \subset \Sigma^7 \subset \mu} \eta^B \mathfrak{W}_1(x^A)\, d\Sigma_B, \tag{3}$$

where $d\Sigma_B$ is the usual differential form "element of

[5] R. Hakim, J. Math. Phys. **8**, 1315 (1967).

[6] Actually, the 4-velocity space is 3-dimensional, because of the constraint

$$u^\mu u_\mu = 1, \quad u^0 > 0, \quad u^\mu \in U^4.$$

However, the use of the 4-dimensional space U^4 is more interesting (from a practical point of view), as is explained in Ref. 5.

[7] R. Hakim, J. Math. Phys. **8**, 1379 (1967).

[8] R. Hakim, J. Math. Phys. **6**, 1482 (1965).

hypersurface"

$$d\Sigma_B = \frac{1}{7!}\sqrt{G}\,\epsilon_{B,A_1,\cdots,A_7}\,dx^{A_1}\wedge,\cdots,\wedge\,dx^{A_7}, \quad (4)$$

with

$$\epsilon_{B,A_1,\cdots,A_7} = \begin{cases} +1, \text{ when } (B, A_1, \cdots, A_7) \text{ is an even permutation of } (1, \cdots, 8), \\ -1, \text{ when } (B, A_1, \cdots, A_7) \text{ is an odd permutation of } (1, \cdots, 8), \\ 0, \text{ otherwise.} \end{cases}$$

In Eq. (4), G represents the absolute value of the determinant of the metric tensor in μ-space.[9] In Eq. (3), η^B is an 8-vector which need not be specified at this stage: essentially, η^B is an "8-velocity"; i.e., η^B is "tangent" to the trajectories of the process, which property insures the invariance of definition (3) with respect to coordinate transformations. We see later what η^B actually is. Of course, $\mathfrak{W}_1(x^A)$ is a positive density and is normalized by

$$\text{Prob}\,\{x^A \in \Sigma^7 \subset \mu\} = \int_{\Sigma^7} \eta^B \mathfrak{W}_1(x^A)\,d\Sigma_B = +1, \quad \forall \Sigma^7. \quad (5)$$

In the same way we can look for the probability

$$\text{Prob}\,\{[x_1^A \in \Delta_1 \subset \Sigma_1^7 \subset \mu] \cap [x_2^A \in \Delta_2 \subset \Sigma_2^7 \subset \mu]\}, \quad (6)$$

i.e., the probability of finding x^A within $\Delta_1 \subset \Sigma_1^7$ *and* $\Delta_2 \subset \Sigma_2^7$. Exactly as above, these probabilities are entirely determined by a density $\mathfrak{W}_2(x_1^A, x_2^A)$:

$$\text{Prob}\,\{[x_1^A \in \Delta_1 \subset \Sigma_1^7] \cap [x_2^A \in \Delta_2 \subset \Sigma_2^7]\} = \int_{\Delta_1 \times \Delta_2 \subset \Sigma_1^7 \times \Sigma_2^7 \subset \mu} \xi^{B_1 B_2} \mathfrak{W}_2(x_1^A, x_2^A)\,d\Sigma_{B_1 B_2} \quad (7)$$

$(A = 1, \cdots, 8;\ B_i = 1, \cdots, 16,\ i = 1, 2)$, where $d\Sigma_{B_1 B_2}$ is the usual differential form "element of surface" corresponding to a 14-dimensional surface embedded in a 16-dimensional manifold. In Eq. (7), $\xi^{B_1 B_2}$ is defined as follows[5]:

$$\xi^{B_1 B_2} = \xi_1^{B_1} \wedge \xi_2^{B_2}, \quad (8)$$

with

$$\boldsymbol{\xi}_1 = \boldsymbol{\eta}_1 \oplus \mathbf{0}, \qquad \boldsymbol{\xi}_2 = \mathbf{0} \oplus \boldsymbol{\eta}_2. \quad (9)$$

$\mathfrak{W}_2(x_1^A, x_2^A)$ also gives the following probabilities:

$$\text{Prob}\,\{(x_1^A, x_2^A) \in \Delta \subset \Sigma^{14} \subset \mu^2\}. \quad (10)$$

$\mathfrak{W}_2(w_1^A, x_2^A)$ is a positive, not necessarily symmetric, function. In the nonrelativistic case, the density $\mathfrak{W}_n(\mathbf{x}_1, \cdots, \mathbf{x}_n; t_1, \cdots, t_n)$ is a symmetric function with respect to the permutations of the set of variables $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ and *not* with respect to the permutations of the set of variables $(\mathbf{x}_1, t_1; \cdots; \mathbf{x}_n, t_n)$. Consequently, there is no reason why the relativistic density should be symmetric with respect to the permutations of $(x_1^\mu, \cdots, x_n^\mu)$. However, in some particular cases such a symmetry property may be verified.[5] In what follows we shall see how the probabilities (7) are restricted by causality conditions.

More generally, a relativistic stochastic process in μ-space is completely determined when all the probabilities

$$\text{Prob}\left\{\bigcap_{i=1}^{i=k} [x_i^A \in \Delta_i \subset \Sigma_i^7 \subset \mu]\right\}; \quad k \in Z^+, \quad \forall \Sigma_i^7 \ \text{("spacelike")} \quad (11)$$

are given (Z^+ being the positive integers). This definition is completely similar to the customary one.[2-4] Of course, we have

$$\text{Prob}\left\{\bigcup_{i=1}^{i=k} [x_i^A \in \Sigma_i^7 \subset \mu]\right\} = +1. \quad (12)$$

Now, we *assume* (and this assumption is, as usual, quite *restrictive*, although it is useful for practical purposes) that the set of probabilities (11) is fully characterized by the densities $\mathfrak{W}_k(x_1^A, \cdots, x_k^A)$. These densities $\mathfrak{W}_k(x_1^A, \cdots, x_k^A)$ should verify the following conditions:

(a) $\mathfrak{W}_k(x_1^A, \cdots, x_k^A) \geq 0$;

(b) $\mathfrak{W}_k(x_1^A, \cdots, x_k^A)$ is not necessarily a symmetric function of $(x_1^A, \cdots, x_k^A)$;

(c)

$$\int_{\Sigma^{7k} \subset \mu^k} \xi^{B_1, \cdots, B_k} \mathfrak{W}_k(x_1^A, \cdots, x_k^A)\,d\Sigma_{B_1, \cdots, B_k} = 1, \quad \forall \Sigma^{7k} \quad (B_i = 1, \cdots, 8k,\ i = 1, \cdots, k;\ A = 1, \cdots, 8);$$

(d)

$$\mathfrak{W}_k(x_1^A, \cdots, x_k^A) = \int_{\Sigma^{7(n-k)} \subset \mu^{(n-k)}(n, \cdots, n-k)} \xi^{B_{n-k}, \cdots, B_n} \times \mathfrak{W}_n(x_1^A, \cdots, x_n^A)\,d\Sigma_{B_{n-k}, \cdots, B_n}, \quad \forall \Sigma^{7(n-k)};\ \forall n > k.$$

In Conditions (c) and (d) we have set[5]

$$\xi^{B_1, \cdots, B_k} = \bigwedge_{i=1}^{i=k} \xi_i^{B_i} \quad (13)$$

with

$$\boldsymbol{\xi}_i = \mathbf{0} \oplus \mathbf{0} \oplus \cdots \oplus \boldsymbol{\eta}_i \oplus \cdots \oplus \mathbf{0}. \quad (14)$$

[9] Since Minkowski space-time is a (pseudo-) metric space and since μ can be considered as its tangent fiber bundle (with Lorentz group as structure group), we can always endow μ with a metric G^{AB} (which need not be specified).

In Condition (d), the notation

$$\Sigma^{7(n-k)} \subset \mu^{(n-k)}(n, \cdots, n-k)$$

signifies that the surface $\Sigma^{7(n-k)}$ is embedded in a space spanned by the coordinates $(x_{nA}, \cdots, x_{n-k\,A})$. Condition (d) is simply a consistency condition: $\mathfrak{W}_k$ may be obtained from all $\mathfrak{W}_n$ $(n > k)$ and independently of the surface on which $\mathfrak{W}_n$ is integrated. Condition (c) is the normalization condition of $\mathfrak{W}_k$. Conditions (a)–(d) are the natural generalization of the similar nonrelativistic ones.[2-4]

In order that Conditions (c) and (d) be valid whatever the various surfaces Σ^{7k}, it is necessary and locally sufficient that the differential forms

$$\xi^{B_{n-k}, \cdots, B_n}\mathfrak{W}_n(x_1^A, \cdots, x_n^A)\, d\Sigma_{B_{n-k}, \cdots, B_n} \quad (15)$$

be *closed forms*,[5] i.e.,

$$d\,\{\xi^{B_1, \cdots, B_{n-k}}\mathfrak{W}_n(x_1^A, \cdots, x_n^A)\, d\Sigma_{B_1, \cdots, B_{n-k}}\} = 0, \quad k < n. \quad (16)$$

Conditions (16) imply the infinite number of integrability conditions[5]

$$\nabla_{A_i}\{\eta_i^{A_i}\mathfrak{W}_n(x_1^A, \cdots, x_n^A)\} = 0, \quad i \leq n. \quad (17)$$

Later, we shall see that, in fact, Eqs. (17) are *kinetic equations* such as Fokker–Planck equations.

Let us now try to characterize a relativistic stochastic process in μ-space in a more mathematical manner. Usually, a stochastic process in a given space E is defined as being a measurable application of the product of the "sample space" by an interval (possibly infinite) of real numbers (i.e., time) into E [2]:

$$T \times (\Omega, \mathcal{A}, \varpi) \xrightarrow{X_t(\omega)} E \quad (18)$$

[Ω = sample space $\equiv \{\omega\}$; $T \subseteq R$, real numbers (time); $\mathcal{A} = \sigma$ algebra on Ω; ϖ = probability on $\mathcal{A}$; $X_t(\omega) \in E$ = stochastic process]. In the relativistic case, instead of the diagram (18) we have

$$\{\Sigma\} \times (\Omega, \mathcal{A}, \varpi) \xrightarrow{X_\Sigma(\omega)} \mu \quad (19)$$

($\{\Sigma\}$ = set of all "regular" spacelike hypersurfaces in μ-space). In other words a relativistic stochastic process in μ-space is a random variable in μ-space indexed by a hypersurface! The fact that the set $\{\Sigma\}$ cannot be totally ordered (for a natural order, of course) can give rise to difficulties. However, $\{\Sigma\}$ may be partially ordered[8]: $\Sigma_1 \prec \Sigma_2$ if and only if Σ_2 is in the future[10] of Σ_1. Note also that if we restrict ourselves to a σ partition of μ-space (i.e., to a family of hypersurfaces indexed by one parameter and satisfying some regularity conditions[8]), namely $\{\Sigma_s\}_{s\in R}$, then definition (19) is essentially equivalent[11] to the classical one (18).

Average Values, Characteristic Functions, etc.

In another paper[5] we defined average values. If $\mathbf{\Phi}(x^A)$ is a given function on μ-space with tensorial values, then

$$\langle\mathbf{\Phi}(x^A)\rangle_{\Sigma^7} = \int_{\Sigma^7}\mathbf{\Phi}(x^A)\eta^B\mathfrak{W}_1(x^A)\, d\Sigma_B \quad (20)$$

is the average of $\mathbf{\Phi}$ *on* Σ^7. Classically we would say "average value of $\mathbf{\Phi}$ at time t." In general $\langle\mathbf{\Phi}\rangle_{\Sigma^7}$ actually depends on Σ^7 except when the integrability condition

$$\nabla_A\{\Phi(x_B)\eta^A\mathfrak{W}_1(x_B)\} = 0 \quad (21)$$

is verified. Because of Eq. (17), we can have[12]

$$[\eta^A\mathfrak{W}_1(x_B)]\cdot\nabla_A\mathbf{\Phi}(x_B) = 0, \quad (22)$$

when, for instance, $[\nabla_A, \eta^A] = 0$. More generally, average values are defined by[5]

$$\langle\mathbf{\Phi}(x_{1A}, \cdots, x_{kA})\rangle_{\Sigma^{7k}} = \int_{\Sigma^{7k}}\mathbf{\Phi}\xi^{B_1, \cdots, B_k}\mathfrak{W}_k\, d\Sigma_{B_1, \cdots, B_k} \quad (23)$$

and do not depend on Σ^{7k} when the integrability conditions

$$\nabla_{B_i}\{\mathbf{\Phi}\xi^{B_1, \cdots, B_k}\mathfrak{W}_k\} = 0, \quad i \leq k \quad (24)$$

are fulfilled.

These definitions allow us to determine the entropy[13] of a given distribution $\mathfrak{W}_k$. Indeed, we have[5]

$$S[\{\mathfrak{W}_k\}, \Sigma^{7k}] \overset{\text{DEF}}{\equiv} -\lambda\langle\log\mathfrak{W}_k\rangle_{\Sigma^{7k}}, \quad \lambda > 0. \quad (25)$$

By analogy with statistical mechanics, a relativistic stochastic process in μ-space is termed *irreversible of* k*th order* whenever

$$S[\{\mathfrak{W}_i\}, \Sigma_1^{7i}] < S[\{\mathfrak{W}_i\}, \Sigma_2^{7i}], \quad i \leq k, \quad (26)$$

with

$$\Sigma_1^{7i} \prec \Sigma_2^{7i} \quad (i \leq k).$$

It is easy to see that the necessary and sufficient conditions for a relativistic stochastic process in μ-space to be irreversible is that the following inequalities be fulfilled:

$$\nabla_{B_j}\{\log\mathfrak{W}_i\cdot\xi^{B_1, \cdots, B_i}\mathfrak{W}_i\} < 0, \quad j \leq i \leq k \quad (27)$$

[10] Σ_2 is said to be in the future of Σ_1, if its projection in Minkowski space–time is in the future of the projection of Σ_1 in Minkowski space–time.

[11] With the restriction that "admissible" measures in μ-space should have their support on Σ.

[12] We shall see later that η^A is, in general, an operator acting on densities $\mathfrak{W}_1$. Therefore we have to be careful and avoid "simplifying" Eq. (22) by $\mathfrak{W}_1$ (for instance).

[13] From the point of view of information theory.

or, equivalently,

$$\xi^{B_1,\cdots,B_j,\cdots,B_i}\mathfrak{W}_i \cdot \nabla_{B_i}\mathfrak{W}_i < 0, \tag{28}$$

where use has been made of Eq. (17) and of the implicit assumption $[\nabla, \xi] = 0$. A relativistic process in μ-space is called *reversible* when Eq. (26), Eq. (27), or Eq. (28) are equalities, with k running from 1 to ∞.

In the same way, we can determine the characteristic functions of the densities $\mathfrak{W}_k$. For instance, the characteristic function of $\mathfrak{W}_1$ *on* Σ^7 is

$$\varphi(K^\Lambda, \Sigma^7) \overset{\text{DEF}}{\equiv} \langle \exp\{iK^A x_A\}\rangle_{\Sigma^7}. \tag{29}$$

This characteristic function implies, as usual, the knowledge of all moments of $\mathfrak{W}_1$ *on* Σ^7 (and on Σ^7 only) through

$$\langle (x^{A_1})^{q_1}, \cdots, (x^{A_8})^{q_8}\rangle_{\Sigma^7} = (-i)^{\Sigma_k q_k} \frac{\partial^{\Sigma_k q_k}}{\partial (K_{A_1})^{q_1}, \cdots, \partial (K_{A_8})^{q_8}} \varphi(K^A, \Sigma^7)\Big|_{K_{A_i}=0} \tag{30}$$

(with the restriction that the moments exist, of course). Therefore we see that in the relativistic case there is not a unique characteristic function corresponding to a given density. This feature was, of course, expected, since the dependence on a surface is the relativistic analog of the usual time dependence of characteristic functions and more generally of average values.

At this point it seems to be worthwhile to make several remarks about average values. Let us return to definition (20) with the choice $\mathbf{\Phi} = \mathbf{x}$, i.e., to the definition of $\langle x^A\rangle_{\Sigma^7}$. In the classical case, $\langle X_t(\omega)\rangle$ generally represents a curve (an average trajectory) either in E or in $T \times E$ [see Eq. (18)]. What about the relativistic case? When Σ^7 belongs to a given σ partition, then it is clear that $\langle x^A\rangle_{\Sigma^7}$ also determines a curve in μ-space (μ-space is the relativistic analog of $T \times E$). However, this curve depends, in general, on the chosen σ partition of μ-space so that there is not actually a unique average trajectory of the process as in the nonrelativistic case. This point will be made clearer in the Appendix. The same features are also true for higher moments of the process. For instance, in the classical case $\langle X_{t_1}(\omega) \otimes X_{t_2}(\omega)\rangle$ determines in general a two-dimensional surface either in $E \otimes E$ or in $(T \times E) \otimes (T \times E)$ while there is in general no such unique surface generated in $\mu \otimes \mu$ by

$$\langle x^A \otimes x^B\rangle_{\Sigma^{14}},$$

etc. In some cases, of course, $\langle x^A\rangle_{\Sigma^7}$ may really determine an average trajectory of the process in μ-space. For instance, let us take

$$\mathfrak{W}_1(x_\mu, u_\mu) = \frac{n_0(x_\mu)}{4\pi K_2(1)} \exp\{-\xi^\mu u_\mu\}\, 2\theta(u^0)\delta(u^\mu u_\mu - 1), \tag{31}$$

i.e., the Jüttner–Synge density (with $\xi^\mu \xi_\mu = 1$, $\partial_\mu \xi^\nu = 0$, and where K_2 is a Kelvin function of order two). In Eq. (31), $n_0(x_\mu)$ is such that all the possibly used integrals are convergent. Since $\mathfrak{W}_1$ depends on only one 4-vector ξ^μ, then necessarily

$$\langle x^\mu\rangle_{S^3} \propto \xi^\mu, \quad \langle u^\mu\rangle_{S^3} \propto \xi^\mu. \tag{32}$$

Therefore, $\langle x^\mu\rangle_{S^3}$ determines a straight line in $\mathcal{M}^4$, while $\langle u^\mu\rangle_{S^3}$ merely yields a point in U^4 (or rather in the hyperboloid $u^\mu u_\mu = +1$, $u^0 > 0$). Finally, $\langle x^A\rangle_{\Sigma^7}$ gives rise to a curve in μ-space.

Of course, we could try to limit ourselves to those sets of $\mathfrak{W}_k$ which give rise to unique moments in μ, $\mu \otimes \mu$, $\mu \otimes \cdots \otimes \mu$, etc. (i.e., to unique surfaces: a k-dimensional surface in $\mu^{\otimes k}$). However, we have not yet been able to find the necessary and sufficient conditions the $\mathfrak{W}_k$ should satisfy in order to get these unique surfaces. Furthermore, this would be only a very particular case of stochastic processes in μ-space. However, these remarks suggest another way of defining this latter class of stochastic processes. Indeed, in the nonrelativistic case a large class of stochastic processes (at least those considered in physics) are determined by the data of all their moments

$$\langle X_t(\omega)\rangle \in E, \quad \langle X_{t_1}(\omega) \otimes X_{t_2}(\omega)\rangle \in E \otimes E, \text{ etc.,} \tag{33}$$

i.e., by a curve in $T \times E$, a 2-dimensional surface in $(T \times E) \otimes (T \times E)$, etc. This suggests that a particular class of relativistic stochastic processes in μ-space may be defined not through the $\mathfrak{W}_k$'s but rather through the data of an infinite number of merely geometrical objects, i.e., a curve in μ-space, $\cdots$, a k-dimensional surface in $\mu^{\otimes k}$, etc. These surfaces have to satisfy a number of requirements connected with "timelikeness" or "spacelikeness."

Gaussian Processes

In the nonrelativistic case, Gaussian processes may be defined in a number of equivalent manners. For instance, $X_t(\omega)$ is a Gaussian process when one of the following conditions is verified (the list below is not exhaustive):

(a) Its distribution functions are Gaussian.[14]

[14] Gaussian distributions are themselves defined in a large number of equivalent manners.

(b) It is completely determined by $\langle X_t(\omega)\rangle$ and $\langle X_{t_1}(\omega) \otimes X_{t_2}(\omega)\rangle$, where the higher moments are obtained from the first two through a definite way.[3]

In the relativistic case, definitions (a) and (b) are no longer equivalent. Suppose that $\mathfrak{W}_1$ and $\mathfrak{W}_2$ are exponentials of quadratic forms of x^A, i.e., are "Gaussian." Then it is easy to realize that these densities *are not* determined by their first two moments; the coefficients of the quadratic forms are not directly related to the first two moments. Conversely, if we adopt definition (b), i.e., if we assume that we are given the moments $\langle x^A\rangle_{\Sigma^7}$ and $\langle x^{A_1} \otimes x^{A_2}\rangle_{\Sigma^{14}}$ ($\forall\Sigma^7$, $\forall\Sigma^{14}$), then it is easily seen that these data *do not* determine a unique $\mathfrak{W}_1$ (or $\mathfrak{W}_2$): On *each* Σ^7 (or Σ^{14}) they determine a $\mathfrak{W}_1$ (or $\mathfrak{W}_2$), but these densities do depend on the particular Σ^7 (or Σ^{14}) chosen.

It results from the above considerations that it seems hardly possible to define Gaussian processes in an intrinsic manner. Of course, if we are given a particular σ partition of μ-space, then [and *only* for this σ partition) a Gaussian process can be defined in the usual way (i.e., with either definitions (a) or (b)]. There is another possibility of defining Gaussian processes for the subclass of processes considered at the end of the last paragraph. Indeed, such a process may be defined—at least in principle—by the data of an average trajectory in μ-space and by a two-dimensional surface in $\mu \otimes \mu$. However, the obtention of higher-order moments (i.e., surfaces) from these ones is far from being clear.

Causality

An important difference between usual and relativistic stochastic processes is related to causality. Indeed, usual stochastic processes—either in μ-space or in configuration space—may have arbitrary trajectories. However, in most cases a physically admissible relativistic stochastic process in μ-space *must* have *timelike trajectories* (either in Minkowski space–time or in μ-space). This causality requirement imposes conditions on the various $\mathfrak{W}_n$'s and more particularly on the various transition probabilities. For instance, if we define[15] the second-order transition density by

$$P_2(x_0^A \mid x_1^A) = \frac{\mathfrak{W}_2(x_0^A, x_1^A)}{\mathfrak{W}_1(x_0^A)}, \tag{34}$$

then we have the following obvious property:

Proposition 1: If a relativistic stochastic process in μ-space has timelike trajectories, then its second-order transition density is such that

$$P_2(x_0^\mu, u_0^\mu \mid x_1^\mu, u_1^\mu) = 0 \begin{cases} \forall x_0^\mu, x_1^\mu \in \mathcal{M}^4 & \text{such that:} \\ (x_0^\mu - x_1^\mu)(x_{\mu 0} - x_{\mu 1}) < 0 \\ \text{or} \quad x_0^0 > x_1^0. \end{cases}$$

Proof: See Ref. 8.

This proposition may be generalized in a straightforward manner for higher-order transition densities. It can also be generalized to the case of spacelike trajectories. In such a case P_2 vanishes when x_0^μ and x_1^μ satisfy

$$(x_0^\mu - x_1^\mu)(x_{\mu 0} - x_{\mu 1}) > 0. \tag{35}$$

For a relativistic Markovian process (see Sec. 3) the above proposition becomes a necessary and sufficient condition since P_2 characterizes the process.

Stationarity

The notion of stationarity in the relativistic framework is much more complicated than the Newtonian one. In fact, there are several possible generalizations of such a notion. A stationary process is commonly defined as one whose statistical characteristics are invariant under time translations. *A priori*, the first natural relativistic generalization is to call stationary those processes whose statistical characteristics are invariant under space–time translations. This kind of stationarity is referred to as the *strong stationarity*. It implies immediately that $\mathfrak{W}_1$ is a constant, $\mathfrak{W}_2$ depends on x_0^μ and x_1^μ only through the combination $(x_0^\mu - x_1^\mu)$, etc. It also implies that the process is both stationary *and* homogeneous in *each* Galilean frame of reference. In fact, we are looking for a definition of stationarity, valid for each Galilean frame of reference (Lorentz invariance), which would not also imply the spatial homogeneity of the process. To do so, we might call *stationary* a relativistic stochastic process in μ-space whose statistical characteristics are invariant under the semigroup of space–time translations $\mathfrak{G}^+$ defined by

$$\alpha^\mu \in \mathfrak{G}^+ \Leftrightarrow \alpha^\mu \alpha_\mu \geq 0, \quad \alpha^0 \geq 0. \tag{36}$$

However, although this definition seems to achieve our goals, it is not completely satisfactory in so far as (a) it involves a semigroup and not a group, (b) this semigroup is a four-parameter semigroup as opposed to the one-parameter group (or semigroup) of time translations of the usual case, and (c) it is easy to show that the invariance under this semigroup implies the strong stationarity (essentially because $\mathfrak{G}^+$ *algebraically* generates the group of space–time translations).

[15] See the next section for the details.

Therefore we are led to the notion of *stationarity with respect to a one-parameter group* of transformations in space–time. This transformation group must have *timelike orbits* in space-time. Let β^μ be the timelike unit 4-vector field tangent to the orbits of the group in space–time. The stationary property of the process then implies that all the $\mathcal{W}_n$'s are invariant under *this* group of transformation. More explicitly, we have

$$\pounds(\beta^\mu)\mathcal{W}_1 = 0, \tag{37}$$

$$\{\pounds(\beta^\mu) \otimes I + I \otimes \pounds(\beta^\mu)\}\mathcal{W}_2 = 0, \tag{38}$$

and similar expressions for higher-order distributions. In Eqs. (37) and (38), $\pounds(\beta^\mu)$ is the Lie derivative with respect to the field β^μ while I is the identity operator in μ space. When the field β^μ is a constant field, and *in* the frame of reference where it reduces to $(1, \mathbf{0})$, one can easily check that one recovers the usual definition of stationarity. Another expression of stationarity with respect to the field β^μ (or equivalently, to the group of transformations generated by β^μ) is the following:

$$\langle x_1^{A_1} \otimes \cdots \otimes x_k^{A_k}\rangle_{\Sigma^{7k}} = \langle x_1^{A_1} \otimes \cdots \otimes x_k^{A_k}\rangle_{\sigma(\Sigma^{7k})},$$
$$k = 1, \cdots, \infty; \quad \forall \Sigma^{7k} \subset \mu^k; \quad \text{and}$$
$$\forall \sigma \in (\text{transformation group generated by } \beta^\mu). \tag{39}$$

All these features (and more particularly the nonuniqueness of the definition of stationarity) occur because of the equivalence of all timelike axes (principle of relativity): in Newtonian physics all these axes are parallel and therefore there is a unique definition of stationarity. We must bear in mind that the above definitions are probably the most natural ones, but there exist other possible notions of stationarity, such as the stationarity with respect to a σ partition, etc.

Definitions of ergodism may also be obtained, but they are all very artificial and not easy to handle.

Comparison with Usual Definitions

In order to consider the connection with usual definitions let us look at the probability:

$$\text{Prob}\left\{\bigcap_{i=1}^{i=k} [x_i^A \in \Delta_i \subset \Sigma_i^7 \subset \mu]\right\}$$
$$= \int_{\Delta_1} \cdots \int_{\Delta_k} \xi^{A_1, \cdots, A_k}\mathcal{W}_k(x_1^A, \cdots, x_k^A)\, d\Sigma_{A_1, \cdots, A_k}. \tag{40}$$

Since the surfaces Σ^7 considered are of the form $\Sigma^7 = S^3 \times U^4$, we can set $\Delta = \delta \times \gamma (\delta \subset S^3, \gamma \subset U^4)$. To compare with usual definitions let us assume that the surfaces $\Sigma_1^7, \cdots, \Sigma_k^7$ belong to a given σ partition of μ-space (or equivalently let us assume that the surfaces $S_1^3, \cdots, S_k^3$ belong to a given σ partition of Minkowsky space–time). Let s be the parameter indexing the surfaces of the σ partition:

$$\Sigma_1^7 \equiv \Sigma^7(s_1), \cdots, \Sigma_k^7 \equiv \Sigma^7(s_k)$$

[or equivalently: $S_1^3 \equiv S^3(s_1), \cdots, S_k^3 \equiv S^3(s_k)$]. Then Eq. (40) represents the probability of having $x_1^A \in \Delta_1$ at "time" $s_1, \cdots,$ and $x_k^A \in \Delta_k$ at "time" s_k. Apparently this is similar to the common definitions. However, we must remember that the various Δ_k *do not* belong to the same space, *except* when the Σ^7's are all obtained from a given one through timelike translations. This latter case is of practical importance although it is most restrictive. Indeed the σ partitions constituted by spacelike 3-planes enter in this category. So far we have dealt with the lhs of Eq. (40) only. Now we consider the rhs of this equation. To achieve our goal we use the coordinate system $x^A \equiv (x^\mu, u^\mu)$, in μ-space. This coordinate system is adapted to the structure of Σ^7, i.e., takes into account the fact that $\Sigma^7 = S^3 \times U^4$. It follows that

$$d\Sigma_A = (d\Sigma_\mu, 0_\mu) \tag{41}$$

and more generally that

$$\xi^{A_1, \cdots, A_k}\mathcal{W}_k\, d\Sigma_{A_1, \cdots, A_k}$$
$$= \eta_1^{\mu_1} \cdots \eta_k^{\mu_k}\mathcal{W}_k\, d\Sigma_{\mu_1} \cdots d\Sigma_{\mu_1} d_4u_1 \cdots d_4u_k, \tag{42}$$

where η^μ designates the first four components of η^A. However, since η^A is tangent to each realization of the process,[5] it follows that $\eta^\mu \equiv u^\mu$, whatever the last four components of η^A. Hence, Eq. (42) reads

$$\xi^{A_1, \cdots, A_k}\mathcal{W}_k\, d\Sigma_{A_1, \cdots, A_k}$$
$$= \mathcal{W}_k u_1^{\mu_1} \cdots u_k^{\mu_k} d_4u_1\, d\Sigma_{\mu_1} \cdots d_4u_k\, d\Sigma_{\mu_1}. \tag{43}$$

Taking into account the constraint $u^\mu u_\mu = 1$, $u^0 > 0$ (i.e., d_4u has to be replaced by d_3u/u^0) and using locally a coordinate system with a local time axis orthogonal to the surface S^3, we can finally write Eq. (42) in the form

$$\xi^{A_1, \cdots, A_k}\mathcal{W}_k d\Sigma_{A_1, \cdots, A_k} = \mathcal{W}_k d_3x_1 d_3u_1 \cdots d_3x_k d_3u_k, \tag{44}$$

which is the usual expression. When S^3 is restricted to being a spacelike 3-plane, then the probability (40) has the same expression as usual. At the nonrelativistic limit, it reduces to the usual expression whatever S^3, since all spacelike hypersurfaces "get flat" and reduce to 3-planes, $t = \text{const}$.

3. RELATIVISTIC MARKOVIAN PROCESSES

Let us now restrict ourselves to the important class of stochastic processes constituted by Markovian

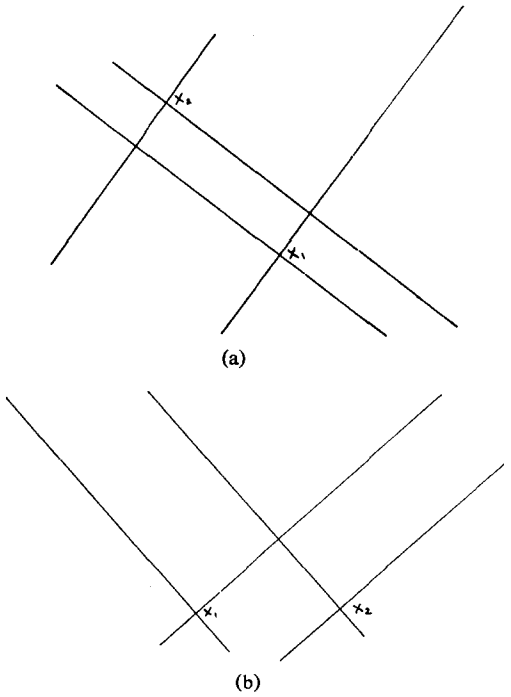


FIG. 1(a). Partial ordering of space-time: $x_1 \prec x_2$. (b) Partial ordering of space-time: x_1 and x_2 are not comparable.

processes. To this end we first define the conditional density of probability

$$P_n(x_{1B}, \cdots, x_{nB} \mid x_{0B}) = \frac{\mathfrak{W}_{n+1}(x_{0B}, x_{1B}, \cdots, x_{nB})}{\mathfrak{W}_n(x_{1B}, \cdots, x_{nB})}. \tag{45}$$

Expression (45) represents the density of probability for the process to reach the state x_{0B} knowing that it passed through the states $x_{1B}, \cdots, x_{nB}$. Of course this sequence of states is (partially) ordered as

$$x_{nB} \prec \cdots \prec x_{1B} \prec x_{0B}, \tag{46}$$

where the order $\prec$ is generated in μ-space by the order of Minkowski space-time. [In space-time we have

$$x_1^\mu \prec x_2^\mu, \tag{47}$$

if and only if

$$(x_1^\mu - x_2^\mu)(x_{1\mu} - x_{2\mu}) \geq 0, \qquad x_2^0 - x_1^0 > 0 \tag{48}$$

(see Fig. 1)].

Note that the density (45) is normalized by

$$\int_\Sigma \eta^A(x_{0B}) P_n(x_{0B}, \cdots, x_{nB} \mid x_{0B})\, d\Sigma_A(x_{0B}) = 1, \qquad \forall \Sigma \subset \mu, \tag{49}$$

and verifies the conservation relation[16]

$$\nabla_A[\eta^A(x_{0B}) P_n(x_{1B}, \cdots, x_{nB} \mid x_{0B})] = 0. \tag{50}$$

[16] Equation (50) represents the conservation of a "conditional current."

We now define a relativistic Markovian process, as usual, as one such that

$$P_n(x_{1B}, \cdots, x_{nB} \mid x_{0B}) \equiv P_2(x_{1B} \mid x_{0B}), \tag{51}$$

which property immediately implies that $\mathfrak{W}_2(x_{1B}, x_{0B})$ completely determines the process.[3,4] In particular, we have

$$\mathfrak{W}_n(x_{1B}, \cdots, x_{nB}) = P_2(x_{nB} \mid x_{n-1\,B}) \cdots P_2(x_{2B} \mid x_{1B}) \mathfrak{W}_1(x_{nB}) \qquad x_{nB} \prec x_{n-1\,B} \prec \cdots \prec x_{1B}. \tag{52}$$

Using now Eq. (52), with $n = 3$, and the compatibility relation

$$\int_\Sigma \eta^A(x_{2B}) \mathfrak{W}_3(x_{1B}, \cdots, x_{3B})\, d\Sigma_A(x_{2B}) = \mathfrak{W}_2(x_{1A}, x_{3A}), \tag{53}$$

we find

$$P_2(x_{0B} \mid x_{1B}) = \int_\Sigma \eta^A(x_{2B}) \{P_2(x_{0B} \mid x_{2B}) P_2(x_{2B} \mid x_{1B})\}\, d\Sigma_A(x_{2B}), \tag{54}$$

with

$$x_{0B} \prec x_{2B} \prec x_{1B}.$$

Equation (54), which is the relativistic Chapman–Kolmogorov equation, leads to

$$\nabla_{A2}\{\eta^A(x_{2B})[P_2(x_{0B} \mid x_{2B}) P_2(x_{2B} \mid x_{1B})]\} = 0, \tag{55}$$

which is nothing but Eq. (17) with $n = 3$. From the compatibility relation

$$\int_\Sigma \eta^A(x_{2B}) \mathfrak{W}_2(x_{1B}, x_{2B})\, d\Sigma_A(x_{2B}) = \mathfrak{W}_1(x_{1B}), \tag{56}$$

we obtain in a straightforward way

$$\int_\Sigma \eta^A(x_{2B})[P_2(x_{2B} \mid x_{1B}) \mathfrak{W}_1(x_{2B})] = \mathfrak{W}_1(x_{1B}). \tag{57}$$

In principle, and modulo well-known supplementary conditions, the Chapman–Kolmogorov equation (54) allows the derivation of the relativistic form of the Fokker–Planck equation. Using the coordinates $(t, \mathbf{x}, \mathbf{u})$, we are immediately led to this equation,[3,4] which reads

$$\nabla_A\left\{-B^A(x_B) P_2 + \frac{1}{2} \nabla_B D^{AB}(x_B) P_2\right\} = 0, \tag{58}$$

once rewritten covariantly. Equation (58) reduces to

$$u^\mu \partial_\mu P_2 + \frac{\partial}{\partial u^\mu}\left\{-B^\mu P_2 + \frac{1}{2}\frac{\partial}{\partial u^\nu} D^{\mu\nu} P_2\right\} = 0, \tag{59}$$

when using the coordinates (x_ν, u_ν) and after taking

into account the fact that, since

$$\mathbf{v} = \lim_{\Delta t \to 0} \left\langle \frac{\Delta \mathbf{x}}{\Delta t} \right\rangle, \tag{60}$$

we have

$$\lim_{\Delta t \to 0} \left\langle \frac{\Delta \mathbf{x} \otimes \Delta \mathbf{x}}{\Delta t} \right\rangle = 0. \tag{61}$$

Equation (61) explains why Eq. (59) does not involve terms in $\partial^2_{\mu\nu} P^2$. Unfortunately we have not been able to derive Eq. (58) or (59) in an arbitrary coordinate system or even with the coordinates (x_ν, u_ν). *A priori*, the Fokker–Planck tensors B^μ and $D^{\mu\nu}$ (or B^A and D^{AB}) seem to depend on an arbitrary surface, since they represent average values of increments of x^A [or (x^μ, u^μ)]. However, it is not so, since P_2 does not depend itself on an arbitrary surface [see Eqs. (54) and (55)].

Equation (59) is extremely interesting. It represents, indeed, a conservation equation[3] in μ-space. It shows that in the systems of coordinates (x_ν, u_ν), η^A has the following form:

$$\eta^A = u^\mu, \quad A = 1, \cdots, 4,$$
$$\eta^A = \frac{\partial}{\partial u^\mu} \left\{ -B^\mu + \frac{1}{2} \frac{\partial}{\partial u^\nu} D^{\mu\nu} \right\}, \quad A = 5, \cdots, 8, \tag{62}$$

and thus *is* an operator. Of course for a non-Markovian process, η^A would have a completely different form and might be an integral operator. This circumstance is by no means unusual and also occurs in the nonrelativistic framework.[3]

4. STOCHASTIC PROCESSES IN MINKOWSKI SPACE-TIME

Although stochastic processes in μ-space implicitly include processes in space–time, in this section we derive their main properties in a more direct manner because of their particular importance.

In Minkowski space–time, a stochastic process is assumed to be determined by the data of the "time-like" currents

$$J^\mu_1(x_\nu), J^{\mu_1\mu_2}_2(x_{1\nu}, x_{2\nu}), \cdots, J^{\mu_1, \cdots, \mu_n}_n(x_{1\nu}, \cdots, x_{n\nu}), \cdots, \tag{63}$$

normalized through

$$\int_{\Sigma^{3n} \subset \mathcal{M}^{4n}} J^{\mu_1, \cdots, \mu_n}_n(x_{1\nu}, \cdots, x_{n\nu})\, d\Sigma_{\mu_1, \cdots, \mu_n} = 1, \tag{64}$$

where Σ^{3n} is "spacelike." The currents (63) have to verify the following compatibility relations:

$$\int_{\Sigma^{3(n-k)} \subset \mathcal{M}^{4(n-k)}} J^{\mu_1, \cdots, \mu_n}_n(x_{1\nu}, \cdots, x_{n\nu})\, d\Sigma_{\mu_{k+1}, \cdots, \mu_n} = J^{\mu_1, \cdots, \mu_k}_k(x_{1\nu}, \cdots, x_{k\nu}), \quad \forall \Sigma^{3(n-k)}; \forall k < n. \tag{65}$$

Besides the positive and antisymmetric character of these currents, they must satisfy the conservation relations

$$\partial_{\mu i} J^{\mu_1, \cdots, \mu_i, \cdots, \mu_n}_n(x_{1\nu}, \cdots, x_{i\nu}, \cdots, x_{n\nu}) = 0 \tag{66}$$

in order that they should not depend on the various surfaces involved in Eqs. (64) and (65). [Of course, we used above Lorentzian coordinates in $\mathcal{M}^{4n}$; otherwise $\partial_{\mu i}$ in Eq. (66) should be replaced by $\nabla_{\mu i}$.]

If we restrict the surfaces Σ to be products of spacelike 3-planes of the type $t = \text{const}$, then we can easily see that the zeroth components of the currents (63) may be interpreted as ordinary densities of probability.[3,4]

The currents (63) are linked with the distributions $\mathfrak{W}_n$ of the preceding sections through

$$J^{\mu_1, \cdots, \mu_n}_n(x_{1\nu}, \cdots, x_{n\nu}) = \int \mathfrak{W}_n(x_{1\nu}, u_{1\nu}; \cdots; x_{n\nu}, u_{n\nu}) \prod_{i=1}^{i=n} u_i^{\mu_i} d_4 u_i, \tag{67}$$

and after antisymmetrizing the indices $(\mu_1, \cdots, \mu_n)$, as can be seen from Ref. 5.

Properties of the Conditional Current $J^\mu(x_{\nu 0} \mid x_{\nu 1})$

Since it will play the basic role in the theory of Markovian processes in Minkowski space–time, let us now define the current of transition probability as being

$$J^\mu(x_{\nu 0} \mid x_{\nu 1}) = \frac{J^{\mu\nu}_2(x_{\nu 0}, x_{\nu 1}) J_{1\nu}(x_{\nu 0})}{J^\lambda_1(x_{\nu 0}) J_{1\lambda}(x_{\nu 0})}, \tag{68}$$

with

$$x_{\nu 0} \prec x_{\nu 1}.$$

The conditional current (68) is obviously normalized by

$$\int_{\Sigma \subset \mathcal{M}^4} J^\mu(x_{\nu 0} \mid x_{\nu 1})\, d\Sigma_\mu(x_{\nu 1}) = 1, \quad \forall \Sigma, \tag{69}$$

which implies the following conservation relation:

$$\partial_{\mu 1} J^\mu(x_{\nu 0} \mid x_{\nu 1}) = 0, \tag{70}$$

resulting from Eq. (66) in a straightforward way $(n = 2)$.

Conversely, given the currents $J^\mu_1(x_{\nu 0})$ and

$$J^\mu(x_{\nu 0} \mid x_{\nu 1}),$$

we have

$$J^{\mu\nu}_2(x_{\nu 0} \mid x_{\nu 1}) = J^\mu_1(x_{\nu 0}) J^\nu(x_{\nu 0} \mid x_{\nu 1}), \tag{71}$$

which may be found either from the probabilistic interpretation of the various currents or from Eq. (68) considered in a local system of coordinates, where $J^\mu_1(x_{\nu 0})$ reduces to its zeroth component.

The current of transition probability (68) has to satisfy some causality requirements. First it should be

such that

$$J^\mu(x_{v0} \mid x_{v1}) = 0, \quad \text{whenever} \quad x_{v0} \not\prec x_{v1}. \tag{72}$$

Next it must be such that the restriction of the differential form

$$J^\mu(x_{v0} \mid x_{v1})\, d\Sigma_\mu(x_{v1}) \equiv \Omega \tag{73}$$

on the forward null cone $\Gamma^+(x_{v0})$:

$$(x_{v0} - x_{v1})(x_0^v - x_1^v) = 0,$$
$$x_{01} - x_{00} > 0 \tag{74}$$

vanishes identically. Indeed, if we denote by Σ_1 and Σ_2 two spacelike 3-surfaces cutting $\Gamma^+(x_{v0})$ and such that $\Sigma_2 \prec \Sigma_1$, then the conservation of the number of particles joined to the causality condition (72) implies that there should not exist any contribution from the part of $\Gamma^+(x_{v0})$ lying between Σ_1 and Σ_2, say S. We have (see Fig. 2)

$$\int_{\Sigma_1} \Omega = \int_{\Sigma_2} \Omega \quad \text{(conservation of "particles")}, \tag{75}$$

and

$$\int_{\Sigma_1 \cup S \cup \Sigma_2} \Omega = 0, \tag{76}$$

from which it follows that

$$\int_S \Omega = 0, \tag{77}$$

and since J^μ is positive, Eq. (77) leads to $\Omega \equiv 0$ on $\Gamma^+(x_{v0})$.

This property shows that *on* $\Gamma^+(x_{v0})$, the current $J^\mu(x_{v0} \mid x_{v1})$ is necessarily a null 4-vector: It should be orthogonal to the null 4-vector $d\Sigma_\mu$.

Finally let us note the important property

$$\int_\Sigma d\Sigma_\mu(x_{v0}) J_1^\mu(x_{v0}) J^\mu(x_{v0} \mid x_{v1}) = J_1^\nu(x_{v1}), \tag{78}$$

obtained by integrating Eq. (71) and using the compatibility relation (65) with $n = 2$, $k = 1$. Equation (78) immediately leads to the following conservation relation:

$$\partial_{\mu 0}\{J_1^\mu(x_{v0}) J^\nu(x_{v0} \mid x_{v1})\} = 0, \tag{79}$$

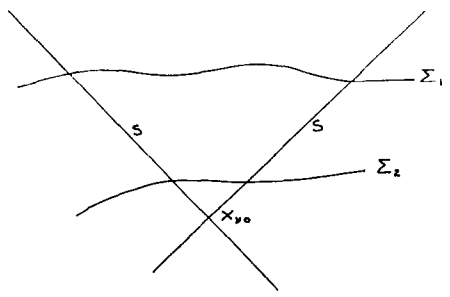


FIG. 2. Illustration of Eq. (77).

or, using Eq. (66) with $n = 1$,

$$J_1^\mu(x_{v0}) \partial_{\mu 0} J^\nu(x_{v0} \mid x_{v1}) = 0. \tag{80}$$

Markovian Processes in Minkowski Space-Time

Now in the same way as usual[3,4] we can find the covariant form of the Chapman–Kolmogorov equation. We derive it in the same way as in the preceding section except that we no longer deal with densities but rather with currents. For instance, the Markovian property is expressed on $J_3^{\mu\nu\rho}(x_{v1}, x_{v2}, x_{v3})$ as

$$J_3^{\mu\nu\rho}(x_{v1}, x_{v2}, x_{v3}) = J_1^\mu(x_{v1}) J^\nu(x_{v1} \mid x_{v2}) J^\rho(x_{v2} \mid x_{v3}), \tag{81}$$

with

$$x_{v1} \prec x_{v2} \prec x_{v3}.$$

Integrating Eq. (81) over x_{v2} and using Eq. (65) with $n = 3$, $k = 1$, we find

$$J^\mu(x_{v1} \mid x_{v3}) = \int_\Sigma d\Sigma_\nu(x_{v2}) J^\nu(x_{v1} \mid x_{v2}) J^\mu(x_{v2} \mid x_{v3}), \tag{82}$$

with

$$x_{v1} \prec x_{v2} \prec x_{v3},$$

which is nothing but the relativistic Chapman–Kolmogorov equation. It leads to the conservation relation

$$\partial_{v2}\{J^\nu(x_{v1} \mid x_{v2}) J^\mu(x_{v2} \mid x_{v3})\} = 0, \tag{83}$$

which reduces to

$$J^\nu(x_{v1} \mid x_{v2}) \partial_{v2} J^\mu(x_{v2} \mid x_{v3}) = 0, \tag{84}$$

after using Eq. (70).

Equation (82) has already been given by Łopuszański,[17] though without proof. Note that the relativistic Chapman–Kolmogorov equation *cannot* be obtained simply by integrating Eq. (54) over the 4-velocity variables. It is indeed well known that the projection of a Markovian process is not necessarily a Markovian process itself (see, e.g., Ref. 4).

Łopuszański has also shown[17] that the Fokker–Planck equations we could get from Eq. (82) reduce to the conservation relation (84) and the adjoint equation as a consequence of causality. As another consequence he also obtained the "physical" interpretation of the spatial components of the current of transition probability: They are simply the increments

$$\lim_{\Delta t \to 0} \left\langle \frac{\Delta x_i}{\Delta t} \right\rangle J_0.$$

This absence of a second-order Fokker–Planck equation shows that a theory of relativistic Brownian motion at *local equilibrium* cannot be erected on the

[17] J. Łopuszański, Acta Phys. Polon. **12**, 87 (1953).

basis of the assumption that it may be represented by a relativistic Markovian process.

Invariance Properties

(1) The strong stationarity of a relativistic process immediately implies that

$$\begin{aligned} J_1^\mu(x_\nu) &= J_1^\mu(0) = (\text{const})^\mu, \\ J_2^{\mu\nu}(x_{\nu 1}, x_{\nu 2}) &= J_2(x_{\nu 1} - x_{\nu 2}, 0_\mu), \\ &\cdots \end{aligned} \tag{85}$$

and therefore the conditional current (68) is a function of the only variable: $x_\nu = x_{\nu 1} - x_{\nu 0}$. From the continuity equation (70) and the fact that this current has necessarily the form[18]

$$J^\mu(x_\nu) = x^\mu \cdot f(\tau = \{x^\lambda x_\lambda\}^{\frac{1}{2}}), \tag{86}$$

it follows that

$$\tau \frac{d}{d\tau} f(\tau) + 4f(\tau) = 0, \tag{87}$$

or

$$f(\tau) = \frac{\text{const}}{\tau^4}. \tag{88}$$

However, the current (86), with f given by Eq. (88), is singular on the null cone $\Gamma^+(0)$ and therefore satisfies neither the normalization condition (it is too singular on the light cone) nor all the causality conditions [it does not vanish on $\Gamma^+(0)$ after multiplication by the normal vector of $\Gamma^+(0)$ at point x_ν]. Therefore we can conclude:

Proposition 2: There exists no relativistic Markovian process invariant under the inhomogeneous Lorentz group.

Therefore, it seems *a priori* difficult to generalize the very interesting results of Nelson[1] in a relativistic framework.

(2) Let us now consider the invariance of a process under timelike translations parallel to a 4-vector α^μ. In such a case we have [see Eqs. (37) and (38)]:

$$\begin{aligned} \mathfrak{L}(\alpha^\mu) J_1^\mu &= 0, \\ \{\mathfrak{L}(\alpha^\mu) \otimes I + I \otimes \mathfrak{L}(\alpha^\mu)\} J_2^\mu &= 0, \\ &\cdots \end{aligned} \tag{89}$$

which reduces to

$$\begin{aligned} J_1^\mu(x_\nu) &= J_1^\mu(\mathbf{x}, 0), \\ J_2^{\mu\nu}(x_{\nu 1}, x_{\nu 2}) &= J_2^{\mu\nu}(\mathbf{x}_1, t_1 - t_2; \mathbf{x}_2, 0), \\ &\cdots \end{aligned} \tag{90}$$

in a frame of reference where $\alpha^\mu = (1, \mathbf{0})$, etc.

Other invariance properties may be found in a similar way. However, when the invariance group is such that the field α^μ actually depends on the space–time position, then the conditions fulfilled by the various currents are more complicated.

5. DISCUSSION AND CONCLUDING REMARKS

We have now to discuss several points about the preceding results.

(1) The approaches (see Appendix for an alternative approach) that we have given have essentially been suggested by relativistic statistical mechanics[5] and therefore present the same essential characteristic features. In particular, instead of using the proper time in the Appendix, we could as well have used any other parameter and the corresponding 4-velocity $u^\mu = dx^\mu/ds$, no longer normalized to one. This last remark raises the point that such an expression as dx^μ/ds or $dx^\mu/d\tau$ does not always exist as a *mathematical object* (even $d\tau$ does not necessarily exist). However, we do not want to enter into these details since such quantities may be given a precise meaning.[19]

In fact we can always think that a stochastic process in μ-space is nothing but an idealization of an underlying dynamical problem, given assumptions about scales of time and length, etc. It is the reason why we do not have to worry too much about the mathematical problems involved, though they have an interest of their own.

(2) Let us now come to the question of Markovian processes. First, if we think that they occur as an idealization and simplification of hidden dynamical processes, then we may *wonder whether they do exist at all.* It is indeed well known that nonquantal relativistic interactions between particles are profoundly *nonlocal*; i.e., the system "keeps the memory" of its past during at least a finite interval of "time" and therefore is *not Markovian.* However, if we denote by λ an invariant parameter characterizing the spatial extension of the system and by T the correlation time of the process, and when the condition $T \gg \lambda c^{-1}$ is

[18] Where we have assumed that we have no other 4-vector than x^μ in the theory, i.e., a strict invariance under the proper homogeneous Lorentz group.

[19] See, e.g., E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton University Press, Princeton, N.J., 1967).

valid, then a Markovian process may be a consistent idealization.

(3) As we have seen, causality requirements impose strong limitations. In particular, in Minkowski space–time it limits the Fokker–Planck equation to be a first-order equation.[17] Furthermore, it demands that there should not exist stochastic processes invariant under the inhomogeneous proper Lorentz group[20] and this is a very strong demand. However, causality is required only if we consider a random particle and therefore can be relaxed if we look at our stochastic process as describing a system of particles, i.e., a gas. Obviously, two particles of a gas may be separated by a spacelike interval, whereas it is not possible for a single particle to pass through points separated by such an interval.

(4) Throughout this paper we have given a formalism of relativistic stochastic processes in μ-space. If we except the case of stochastic processes in more general spaces such as μ^N (which can be dealt with in a similar way, with the only difficulty of complex notations), we may say that we have probably treated the most complicated case, although not the most general one.

Unfortunately, we *cannot* consider an abstract relativistic random process in an arbitrary (but given) space. Indeed, the random process has to be specified further before we build up the corresponding theory. For instance, we have to answer the following questions. What is the tensorial character of the process? Is it "timelike"? "Spacelike"? How may causality be expressed? What is (are) the "time" parameter(s) that indexes (index) the process? etc. The answer to these questions depends on the physical meaning of the process under consideration.

In order to be a little more specific let us consider an "abstract" relativistic random process α of unspecified tensorial character. *A priori*, a covariant way of indexing this process might be to do so with space–time points. Therefore, such an abstract process would be a measurable mapping of $\mathcal{M}^4 \times (\Omega, \mathcal{A}, \varpi)$ into the space $\{\alpha\}$. This could be very convenient, essentially because $\mathcal{M}^4$ is partially ordered. However, this procedure does not obviously make much sense when applied to processes occurring in $\mathcal{M}^4$ itself. Of course this does not mean that it never makes sense at all. This just shows that we have to be careful and need to know the specific physical meaning of α. Another way of indexing the process would be with the help of an invariant parameter, for instance the proper time of an "observer," etc. Here again we have to specify α. More generally it seems to us that not only α has to be specified, but also the way it can be observed or measured.

(5) Finally, it is easy to realize that the preceding results can be transposed in a straightforward manner into a general relativistic framework.

[20] This invariance also leads to nonnormalizable currents. However, this property is not as important as causality.

ACKNOWLEDGMENT

The author is indebted to Professor P. Havas for his kind hospitality at Temple University, where this work has been partially carried out.

APPENDIX: ANOTHER APPROACH TO RELATIVISTIC STOCHASTIC PROCESSES

In a preceding paper[5] we gave two equivalent formalisms for relativistic statistical mechanics: a "geometrical" one and a more intuitive one based on the use of proper time. Here we give also an approach based on proper-time-dependent densities. This approach is completely equivalent to the geometrical one studied in the preceding sections insofar as equivalent statistical assumptions are used.

Generalities

Let us consider the random point $x^A \in \mu$. It may be considered as proper-time-dependent: $x^A(\tau)$. Hence, if we consider τ as a time variable, then $x^A(\tau)$ is a "true" stochastic process in μ space. However, the proper-time parametrization of the trajectories of the process is only a possible one among many others. Thus the results we obtain hereafter have no direct physical meaning *unless* they are linked to the "geometrical" approach of Sec. 2. We come back to this point below.

The stochastic process $x^A(\tau)$ is assumed, as usual, to be completely determined[3,4] by the data of the following distribution functions:

$$W_n(x_1^A, \tau_1; \cdots; x_n^A, \tau_n), \quad n = 1, 2, \cdots, \infty, \quad \text{(A1)}$$

which verifies

$$\int \cdots \int \prod_{i=1}^{i=n} d\mu_i W_n(x_1^A, \tau_1; \cdots; x_n^A, \tau_n) = 1, \qquad \forall (\tau_1, \cdots, \tau_n) \in R^n, \quad \text{(A2)}$$

$$\int \cdots \int \prod_{i=k+1}^{i=n} d\mu_i W_n(x_1^A, \tau_1; \cdots; x_n^A, \tau_n) = W_k(x_1^A, \tau_1; \cdots; x_k^A, \tau_k), \quad \forall k < n. \quad \text{(A3)}$$

Note that the densities W_n are normalized in the entire μ-space and not merely on a hypersurface as in Sec. 2. With Eqs. (1), (2), and (3) we can develop the theory of relativistic stochastic processes in the usual way. However, it is important to realize that the

stationarity of the process $x^A(\tau)$ with respect to the proper time has in general no physical meaning. Furthermore, all the discussion of Sec. 2 on stationarity or causality remains valid.

In Ref. 5, we have shown, *mutatis mutandis*, that (a) $\lim_{\tau_i \to \pm\infty} W_n = 0$, and (b)

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W_n(x_1^A, \tau_1; \cdots; x_n^A, \tau_n) = \mathfrak{W}_n(x_1^A, \cdots, x_n^A), \quad \text{(A4)}$$

which establishes the link between the two formalisms.

Markovian Processes

Once again we can apply conventional methods and definitions and find the following properties:

(a) The Chapman–Kolmogorov equation

$$\mathfrak{F}_2(x_0^A, \tau_0 \mid x_1^A, \tau_1) = \int d\mu \mathfrak{F}_2(x_0^A, \tau_0 \mid x^A, \tau)\mathfrak{F}_2(x^A, \tau \mid x_1^A, \tau_1), \quad \text{(A5)}$$

with

$$\tau_0 < \tau < \tau_1,$$

and, with appropriate assumptions[2-4];

(b) The Fokker–Planck equation

$$\frac{\partial \mathfrak{F}_2}{\partial \tau} + u^\mu \partial_\mu \mathfrak{F}_2 = -\frac{\partial}{\partial u^\mu}\{B^\mu \mathfrak{F}_2\} + \frac{1}{2}\frac{\partial^2}{\partial u^\mu \partial u^\nu}\{D^{\mu\nu}\mathfrak{F}_2\}, \quad \text{(A6)}$$

where the coefficients B^μ and $D^{\mu\nu}$ may be called, respectively, the friction and diffusion tensors which we study below.

It is easy to realize that the "proper-time" Markovian processes studied in this Appendix are not equivalent to the "true" ones studied in Sec. 3. Indeed, the integration over τ_1 of Eq. (A5) does not lead to the Chapman–Kolmogorov equation (54). Furthermore, if we consider the definitions of the Markovian property in both approaches in a closer way and use Eq. (A4), we can see that they do not mutually agree. However, in the case where the Markovian process is stationary with respect to the proper time, we do recover *the form* of the Fokker–Planck equation used in relativistic statistical mechanics,[7] after integration on the proper-time variable. Indeed, in this case, the Fokker–Planck coefficients $B^\mu(x_\nu, u_\nu)$ and $D^{\mu\nu}(x_\nu, u_\nu)$ do not really depend[3] on the proper time and therefore Eq. (A6) reduces to the following:

$$u^\mu \partial_\mu P_2 + \frac{\partial}{\partial u^\mu}\{B^\mu P_2\} - \frac{1}{2}\frac{\partial^2}{\partial u^\mu \partial u^\nu}\{D^{\mu\nu}P_2\} = 0, \quad \text{(A7)}$$

after using the above properties (a) and (b).

Let us now give some properties of the Fokker–Planck coefficients. Note that the properties given below are valid whatever the formalism used.

(1) If the process $x^A(\tau)$ is strongly stationary, then B^μ and $D^{\mu\nu}$ do not depend on x_ν:

$$B^\mu(x_\nu, u_\nu) = B^\mu(u_\nu),$$
$$D^{\mu\nu}(x_\nu, u_\nu) = D^{\mu\nu}(u_\nu). \quad \text{(A8)}$$

(2) In a large number of problems, the Jüttner–Synge[21] equilibrium distribution function is a stationary solution of Eq. (A7), i.e.,

$$P(u_\nu) \sim \exp\{-m\xi^\mu u_\mu\} \quad \text{(A9)}$$

is a solution of

$$\frac{\partial}{\partial u^\mu}\left\{-B^\mu P + \frac{1}{2}\frac{\partial}{\partial u^\nu}[D^{\mu\nu}P]\right\} = 0. \quad \text{(A10)}$$

Such a case occurs, for instance, when we think of Eq. (A7) as a kinetic equation.[5] Then, Eq. (A10) leads to

$$-\frac{\partial}{\partial u^\mu}B^\mu + \xi^\mu B_\mu + \frac{1}{2}\frac{\partial^2}{\partial u^\mu \partial u^\nu}D^{\mu\nu} - \frac{1}{2}\left\{\xi_\nu \frac{\partial}{\partial u^\mu}D^{\mu\nu} + \xi_\mu \frac{\partial}{\partial u^\nu}D^{\mu\nu}\right\} = 0. \quad \text{(A11)}$$

(3) In Ref. 5 we noticed that provided the 4-force F^μ is such that $F^\mu u_\mu = 0$, the uniform distribution $\delta[u^\mu u_\mu - 1]$ is identically a solution of Eq. (A10). This leads to the relations

$$D^{\mu\nu}u_\mu u_\nu = 0, \quad \text{(A12)}$$

$$\frac{1}{2}\frac{\partial^2}{\partial u^\mu \partial u^\nu}D^{\mu\nu} = \frac{\partial}{\partial u^\mu}B^\mu, \quad \text{(A13)}$$

$$\frac{1}{2}\left\{u_\mu \frac{\partial}{\partial u^\nu}D^{\mu\nu} + u_\nu \frac{\partial}{\partial u^\mu}D^{\mu\nu}\right\} = B^\mu u_\mu. \quad \text{(A14)}$$

Of course, Eqs. (A11)–(A14) depend on the nature and the interpretation we give to the basic stochastic process.

Illustration of the Formalism

As an application of the preceding ideas let us consider the problem of a particle of mass m embedded in a random force field.

There are two possible *equivalent ways* of dealing with this problem. Either we study *the random differential system*[3]

$$m\frac{du^\mu}{d\tau} = F^\mu,$$
$$\frac{dx^\mu}{d\tau} = u^\mu, \quad \text{(A15)}$$

[21] J. L. Synge, *The Relativistic Gas* (North-Holland Publishing Co., Amsterdam, 1957).

with given (or random) initial data, or we directly write the *equivalent random Liouville equation*[5] satisfied by

$$R(x_\nu, u_\nu; \tau) \equiv \delta(x_\nu - x_\nu(\tau)) \otimes \delta(u_\nu - u_\nu(\tau)),$$

$$\frac{\partial}{\partial \tau} R + u^\mu \partial_\mu R + \frac{F^\mu}{m} \frac{\partial}{\partial u^\mu} R = 0, \qquad \text{(A16)}$$

where we have assumed implicitly that $(\partial/\partial u^\mu)F^\mu = 0$, and next solve it. The former method is inspired by Stratonovich,[3] while the latter is from Kubo.[22]

The random force field F^μ is assumed to be completely specified (in a statistical sense) when all the moments

$$\langle F^{\mu_1} \otimes \cdots \otimes F^{\mu_k} \rangle = \Phi^{\mu_1, \cdots, \mu_k}, \quad k = 1, 2, \cdots, \infty \qquad \text{(A17)}$$

are known. Mathematically, F^μ is in fact *a random variable in a functional space.*

Let us now set

$$L_0 = -u^\mu \partial_\mu,$$

$$L_1 = \frac{F^\mu}{m} \frac{\partial}{\partial u^\mu}. \qquad \text{(A18)}$$

With these notations, Eq. (A16) can be written

$$\frac{\partial}{\partial \tau} R = (L_0 + L_1) R, \qquad \text{(A19)}$$

and, in the *interaction representation*, it reads

$$\frac{\partial}{\partial \tau} \mathcal{R} = \mathfrak{L} \mathcal{R}, \qquad \text{(A20)}$$

with

$$\mathcal{R} = \exp[-L_0 \tau] R, \qquad \text{(A21a)}$$

$$\mathfrak{L} = \exp[-L_0 \tau] L_1 \exp[L_0 \tau]. \qquad \text{(A21b)}$$

Equation (A20) is formally integrated and yields

$$\mathcal{R} = \exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right] \mathcal{R}(0) = \exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right] \cdot R(0), \qquad \text{(A22)}$$

where exp [ ] designates symbolically the series

$$\exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right]$$

$$= \sum_{n=0}^{\infty} \int_0^\tau d\tau_1 \int_0^{\tau_1} \cdots \int_0^{\tau_{n-1}} d\tau_n P[\mathfrak{L}(\tau_1), \cdots, \mathfrak{L}(\tau_n)], \qquad \text{(A23)}$$

with $\tau_1 < \tau_2 < \cdots < \tau_n$, and where P designates the chronological operator. Our purpose is to derive an equation satisfied by $\langle R \rangle = W_1$. It is easy to show that $\langle \mathcal{R} \rangle$ satisfies

$$\frac{\partial}{\partial \tau} \langle \mathcal{R} \rangle = \frac{\partial}{\partial \tau} \left\langle \exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right] \right\rangle \times \left\{ \left\langle \exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right] \right\rangle \right\}^{-1} \langle \mathcal{R} \rangle, \qquad \text{(A24)}$$

so that W_1 satisfies

$$\frac{\partial}{\partial \tau} W_1 = L_0 W_1 + \left\{ \exp[L_0 \tau] \times \frac{\partial}{\partial \tau} \left\langle \exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right] \right\rangle \times \left\langle \exp\left[\int_0^\tau \mathfrak{L}(s)\, ds\right] \right\rangle^{-1} \times \exp[-L_0 \tau] \right\} W_1. \qquad \text{(A25)}$$

Of course, the second term of the rhs of Eq. (A25) can be explicitly written with the help of definition (A21b), (A23) and the moments of F^μ. If we consider that F^μ is a term of order 1 in a squared coupling constant, then the various series of operators involved in the rhs of Eq. (A25) give a development in powers of this constant. Equation (A25) may be useful in connection with problems of kinetic equations. To the second order in the squared coupling constant, Eq. (A25) reads

$$\frac{\partial}{\partial \tau} W_1 - L_0 W_1 - \langle L_1 \rangle W_1 = \int_0^\tau ds\, K[L_1(\tau) \exp\{L_0(\tau - s)\} L_1(s) \times \exp\{-L_0(\tau - s)\}] \times W_1, \qquad \text{(A26)}$$

where $K[\ ]$ signifies "correlation function of [ ]." Next we have to solve Eq. (26) and use its solution to find the "true" density $\mathfrak{W}_1$ by using Eq. (4). The assumption that F^μ does not depend on u^μ is, of course, an oversimplification of the real situation and thus a fully correct treatment is in fact more complicated.

In another paper we shall show that these equations are useful when dealing with problems of acceleration of relativistic particles by random fields (i.e., acceleration of the charged component of extragalactic cosmic rays) and also in problems of turbulence.

Concluding Remarks

We can repeat here what has been said in Ref. 5 about the proper-time-dependent formalism. It has no specific meaning by itself and has only a heuristic value. It may be used as a useful intermediary in the calculations leading to the "true" geometrical densities. The main advantage of this approach lies in the fact that it allows the use of *standard methods* although usual definitions should be used only after a careful analysis of their "physical" meaning.

[22] R. Kubo, J. Math. Phys. **4**, 174 (1963).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Invariants in the Motion of a Charged Particle in a Spatially Modulated Magnetic Field

D. A. DUNNETT
Department of Applied Mathematics, University of Liverpool

E. W. LAING
Department of Natural Philosophy, University of Glasgow

AND

J. B. TAYLOR
U.K.A.E.A., Culham Laboratory, Abingdon, Berkshire, England

(Received 16 January 1968)

In this paper we study the effect of a spatial modulation in the magnetic field on invariants such as the magnetic moment $\mu = V_\perp^2/B$. In particular, we investigate whether an invariant still exists when the wavelength of the modulation is comparable to the gyro radius of the particle. In an axially symmetric magnetic field, with a square-wave modulation, the orbit equations reduce to algebraic relations convenient for numerical study. We find from such studies that orbits are of two types: (a) regular orbits which generate an invariant; (b) orbits which are quasi-ergodic. We have also calculated an invariant by perturbation theory with the depth of modulation of the field as a small parameter. For this we developed a modified form of perturbation theory which overcomes the difficulty of infinities arising at resonance between the perturbation and the cyclotron period. This difficulty in fact corresponds to a change in topology of the invariant curves. The invariant calculated from this theory shows very good agreement with the numerically computed orbits of type (a). The transition to quasi-ergodic behavior cannot be predicted analytically, but some indication of it may exist in the complex topology of the invariant curves in the ergodic regions.

1. INTRODUCTION

The invariants of motion play an important role both in the trapping and containment of charged particles in magnetic-confinement systems. For motion in sufficiently smooth fields, an important adiabatic invariant is the magnetic moment $\mu = V_\perp^2/B$, where $V_\perp$ is the particle velocity transverse to the magnetic field **B**. In this paper we consider the effect of a spatial modulation, superimposed on an otherwise uniform magnetic field, on the behavior of such invariants. We are particularly interested in the situation where the wavelength of the modulation may be comparable with the gyromagnetic radius of the particle. In this case μ is no longer a valid invariant and we investigate whether any alternative invariant exists.

Interest in a possible invariant in a modulated magnetic field arose out of studies of the containment properties of magnetic traps incorporating such fields,[1-3] but here we are concerned only with the invariant itself, not with any possible containment. Accordingly we consider a very simple field, with axial symmetry and without any end effects. The axial component of the magnetic field B_z is $B_0[1 + \epsilon g(z)]$ where $g(z)$ is the periodic modulation with wavelength λ; there is no radial dependence so that $\nabla \times \mathbf{B} \neq 0$, although $\nabla \cdot \mathbf{B} = 0$.

The motion of a charged particle in this field is studied both numerically (Sec. 2) and analytically (Sec. 3). In the numerical studies $g(z)$ is taken to be a step function, alternately ± 1 with discontinuities at $z = n\lambda/2$. With this form of magnetic field the orbit equations reduce to a set of transfer matrices so that the orbit can be computed over many field periods with speed and accuracy.

The results of these numerical calculations show that, depending on their initial values, orbits can be classified into two types: (a) regular orbits which correspond to the existence of an invariant of the motion, and (b) orbits which fill quasi-ergodically all the available part of phase space not mapped by orbits of type (a).

In the analytic investigation we study invariants using perturbation theory, with ϵ as a small parameter. Straightforward perturbation theory fails because the perturbation may "resonate" with the cyclotron period of the unperturbed orbit, leading to the problem of "vanishing denominators." However, a modified form of perturbation theory is introduced which overcomes this difficulty and permits us to generate an adiabatic invariant J, as a series in ϵ. This is valid throughout phase space, even in the region of the resonances. Comparison of the first few terms of the

[1] D. A. Dunnett, E. W. Laing, S. J. Roberts, and A. E. Robson, J. Nucl. Energy: Pt. C, **7**, 359 (1965).
[2] A. E. Robson and J. B. Taylor, Phys. Fluids **8**, 2026 (1965).
[3] D. A. Dunnett and E. W. Laing, J. Nucl. Energy: Pt. C, **8**, 399 (1966).

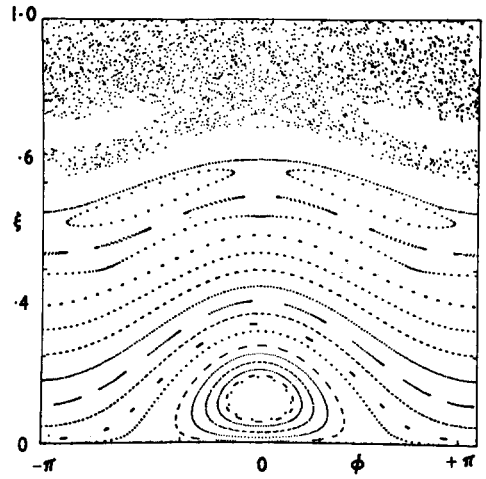


FIG. 1. Numerically computed orbits, $\epsilon = 0.025$.

invariant $J \simeq J_0 + \epsilon J_1$ with the numerical computations shows excellent agreement with the regular orbits calculated numerically.

2. NUMERICAL COMPUTATIONS

It has been shown by Laing and Robson[4] that, within a range of z for which $f(z) = 1 + \epsilon g(z)$ has a constant value f_i, the orbit equations may be put in the form

$$d^2r/d\tau^2 = r_0^4/r^3 - rf_i^2, \tag{1}$$

$$d^2z/d\tau^2 = 0. \tag{2}$$

We have used as a dimensionless time variable $\tau = \frac{1}{2}\omega t$, $\omega = eB_0/mc$, and the length r_0 is related to the constant canonical momentum p_θ conjugate to the azimuthal coordinate θ about the axis of symmetry.

The general solution of (1) is

$$r^2 = \alpha_i + \beta_i \cos(2f_i\tau + \phi_i), \tag{3}$$

where

$$\alpha_i^2 - \beta_i^2 = r_0^4/f_i^2.$$

At any time τ, the state of motion of the particle is thus described by only two parameters α_i, ϕ_i. At a discontinuity in f, both r and $dr/d\tau$ are continuous, so that at the boundary between region i and region $(i+1)$,

$$\alpha_i + \beta_i \cos\phi_i = \alpha_{i+1} + \beta_{i+1} \cos\psi_{i+1}$$

and

$$f_i\beta_i \sin\phi_i = f_{i+1}\beta_{i+1} \sin\psi_{i+1}. \tag{4}$$

Here we have set $\tau = 0$ at the boundary between the ith and $(i+1)$th regions and ψ_i, ϕ_i denote the phase at the beginning and end, respectively, of the ith region. Thus

$$\phi_i = \psi_i + \lambda f_i/u_i \tag{5}$$

where $u_i = (dz/d\tau)_i$ and can be determined from (3) and the energy equation. This takes the form

$$\left(\frac{dr}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 + \left(\frac{r_0^2}{r} - rf_i\right)^2 = v^2, \tag{6}$$

where we have written $v = 2V/\omega$. A more detailed account of the procedure for computing the trajectory of a particle is given by Dunnett *et al.*[1]

Instead of using ϕ and α to represent the orbit, the results are presented in terms of the phase ϕ and the normalized magnetic moment

$$\xi = \frac{v_\perp^2}{v^2} = 1 - \frac{1}{v^2}\left(\frac{dz}{d\tau}\right)^2.$$

The value of ξ is constant with a region of uniform magnetic field, and is related to α_i by

$$\xi_i = 2(f_i/v^2)(\alpha_i f_i - r_0^2).$$

In the calculation, then, values of (ξ, ϕ) are computed in successive periods of the magnetic field and the successive points (ξ_i, ϕ_i) are plotted to give a representation of the "orbit." Some typical results are shown in Figs. 1–3. These all refer to orbits with $v = 2$, $\lambda = 2\pi$, $r_0 = 2$, which were of particular interest in the containment problem studied by Dunnett *et al.*[1] They represent a particle injected parallel to the axis at a distance $r_0 = 2$ from the axis with a velocity satisfying the resonance criterion $V = \omega\lambda/2\pi$, i.e., $v = 2$ for $\lambda = 2\pi$. The three figures are computed with amplitudes of the modulating field which are 0.025, 0.05, and 0.10, respectively, of the main field.

From these diagrams it can be seen that orbits are of two distinct types. In some regions of the (ξ, ϕ) plane the successive values of (ξ_i, ϕ_i), along an orbit,

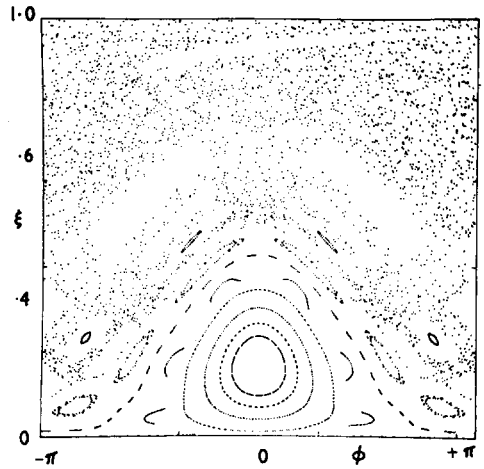


FIG. 2. Numerically computed orbits, $\epsilon = 0.05$.

[4] E. W. Laing and A. E. Robson, J. Nucl. Energy: Pt. C, 3, 146 (1961).

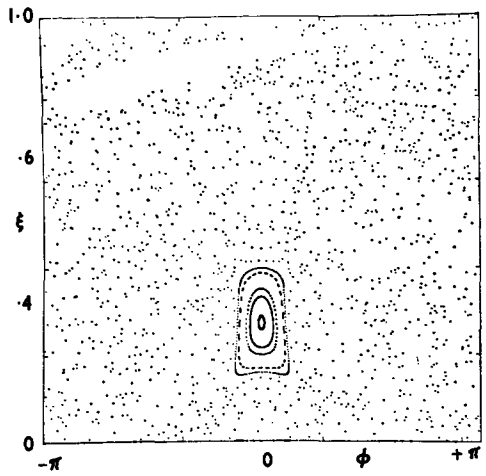


FIG. 3. Numerically computed orbits, $\epsilon = 0.10$.

lie on a smooth curve. These [type (a)] orbits correspond to the existence of an invariant of motion. However, for orbits in other regions of (ξ, ϕ) the successive values of (ξ_i, ϕ_i) do not lie on a regular curve but fill, quasi-ergodically, part of the (ξ, ϕ) plane. In fact, one of these [type (b)] orbits eventually fills all the (ξ, ϕ) plane except for regions already mapped by type (a) orbits and some regions of large ξ. (These excluded large ξ regions correspond to particle reflection at the field discontinuities.) It should be stressed that with type (a) orbits each initial value (ξ_0, ϕ_0) generates only *one* invariant curve (except that it may generate a chain of "islands," as shown). However, with the type (b) quasi-ergodic orbits, *any single* initial state (ξ, ϕ_0) will ultimately map the whole area. It is in this sense that we claim that the motion is quasi-ergodic.

The region occupied by type (a) (regular) orbits decreases, and that occupied by type (b) (quasi-ergodic) orbits increases, as the strength of the field modulation ϵ is increased. In Fig. 1, where $\epsilon = 0.025$, almost all possible orbits are regular and possess a valid invariant. In Fig. 3, where $\epsilon = 0.1$, almost the whole plane is mapped by a quasi-ergodic orbit and only a small class of orbits possess a valid invariant. In all cases there is a sharp transition between regions where an invariant exists and those where quasi-ergodic behavior pertains.

This kind of behavior is similar to that found in several other problems, notably in the quest for a third invariant of galactic motion.[5] It is worth noting, however, that in our present problem the motion for $\epsilon = 0$ is unbounded, whereas interest usually centers around motion which is periodic when $\epsilon = 0$.

[5] M. Henon and C. Heiles, Astron. J. **69**, 73 (1964).

We note in passing that, if μ were an invariant, the orbits (ξ_i, ϕ_i) would be on horizontal straight lines $\xi = \text{const.}$

3. THE INVARIANT

The results of the numerical calculations indicate that, at least for small ϵ, an invariant may exist even when the usual magnetic-moment invariant is no longer valid. In this section we show how this invariant may be calculated by a perturbation expansion in ϵ. The essential problem is to devise a perturbation expansion which remains valid even when the modulated magnetic field gives a perturbation in resonance with the unperturbed cyclotron orbit, i.e., when its effect is no longer "small." In conventional perturbation calculations this resonance gives rise to vanishing denominators in the coefficients of the ϵ power series. We shall show how this difficulty may be overcome, in principle, to any finite order in ϵ.

Perturbation Theory

The Hamiltonian describing the motion of a charged particle in the magnetic vector potential $\mathbf{A}$ is given by

$$H = (\mathbf{p} - e\mathbf{A}/c)^2/2m.$$

In our case, using cylindrical coordinates, the vector potential has only one nonzero component, $A_\theta = \frac{1}{2}rB_z$. Using C for r_0^2, but otherwise the same notation as in Sec. 2, we may then write

$$H = \tfrac{1}{2}(p_r^2 + p_z^2 + [C/r - r(1 + \epsilon g(z))]^2). \quad (7)$$

As we evaluate the invariant only to first order in ϵ, we express H approximately in the form $H_0 + \epsilon H_1$:

$$H_0 = \tfrac{1}{2}[p_r^2 + p_z^2 + (C/r - r)^2],$$
$$H_1 = (r^2 - C)g(z). \quad (8)$$

H_0 is thus the Hamiltonian for motion in a uniform magnetic field.

The object is to generate a new constant of the motion J which is to replace the constant P_z or, equivalently, the constant ξ, which exists when $\epsilon = 0$. The system is still conservative and $H = H_0 + \epsilon H_1$ is a constant of the motion so that any other constant of motion J has the property $[J, H] = 0$, where $[\ ,\ ]$ is the Poisson bracket. Setting $J = J_0 + \epsilon J + \cdots$, $H = H_0 + \epsilon H_1$, and expanding the Poisson bracket, one obtains a set of recurrence equations, of which the first two are

$$[J_0, H_0] = 0,$$
$$[J_1, H_0] + [J_0, H_1] = 0. \quad (9)$$

We first perform a simplifying canonical transformation[6] $(r, p_r) \rightarrow (Q, P)$, so that $\frac{1}{2}[p_r^2 + (C/r - r)^2] \rightarrow P$.

[6] B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).

This is effected by the generating function

$$W = \int^r dr[2P - (C/r - r)^2]^{\frac{1}{2}}, \tag{10}$$

which gives

$$r^2 = P + C + A \sin 2Q, \tag{11}$$

where

$$A = [(P + C)^2 - C^2]^{\frac{1}{2}}.$$

The transformed Hamiltonian then becomes

$$H \to P + \tfrac{1}{2}p_z^2 + \epsilon\Omega, \tag{12}$$

where

$$\Omega = (P + A \sin 2Q)g(z).$$

The recurrence equations (9) become, writing $p_z = p$ for brevity,

$$\frac{\partial J_0}{\partial Q} + p\frac{\partial J_0}{\partial z} = 0, \tag{13}$$

$$\frac{\partial J_1}{\partial Q} + p\frac{\partial J_1}{\partial z} + [J_0, \Omega] = 0. \tag{14}$$

Equation (13) indicates that J_0 is an arbitrary function of P, p, $pQ - z$. However, dependence on the last of these is ruled out by requiring that J_0 be periodic both in Q and in z for all p. In fact, we need regard J_0 only as an arbitrary function of p. Then, after a change of variables to $\alpha = pQ - z$, $\beta = z$, Eq. (14) may be written

$$\frac{\partial J_1}{\partial \beta} = \frac{1}{p}\frac{\partial J_0}{\partial p}\frac{\partial \Omega}{\partial z},$$

so that

$$J_1 = \frac{1}{p}\frac{\partial J_0}{\partial p}\int^\beta d\beta'\left[P + A \sin\frac{2}{p}(\alpha + \beta')\right]\frac{dg(\beta')}{d\beta'}. \tag{15}$$

Now $g(\beta)$ is a periodic function of period λ, so that the integral is unbounded in β whenever $p = \lambda/\pi n$. In the neighborhood of such points our expansion in ϵ must apparently break down—the well-known vanishing-denominator problem.

This failure of perturbation theory near a resonance has a simple interpretation. It is a manifestation of the fact that at such a point the topology of the true invariant curves, $J = \text{const}$, differs from that of the curves $J_0 = \text{const}$. Generally, if ϵ is small, the contours of $(J_0 + \epsilon J_1) = \text{const}$ can be topologically different than those of $J_0 = \text{const}$ only if J_1 is large. Thus the appearance of a large J_1 is simply the response of perturbation theory to the change in topology. Consequently, a valid perturbation expansion *can* be obtained if J_0 is chosen so that a *small* J_1 *can* make the topology of the $(J_0 + \epsilon J_1)$ curves differ from that of the $J_0 = \text{const}$ curves. This is the case if $\partial J_0/\partial p$ vanishes at the points concerned.

We illustrate this by evaluating J_1 for the square-wave modulation used in the numerical calculation, $g(\beta') = \pm 1$. Then

$$\frac{dg}{d\beta} = 2\sum_{-\infty}^{\infty}(-1)^n\delta(\beta' - n\lambda/2),$$

and so

$$J_1 = \frac{2}{p}\left(\frac{\partial J_0}{\partial p}\right) \times \left\{\frac{1 + (-1)^N}{2} + \frac{A}{2\cos(\lambda/2p)}\left[\sin\frac{(2\alpha - \lambda/2)}{p} \times (-1)^N \sin\frac{(2\alpha + (N + \tfrac{1}{2})\lambda)}{p}\right]\right\} + C(\alpha, p, P). \tag{16}$$

In (16), β has been chosen in the interval $[N\lambda/2, (N + 1)\lambda/2]$ so that $\beta = N\lambda/2 + \gamma$, $0 < \gamma < \lambda/2$. The integration "constant" C may conveniently be chosen so that terms independent of N disappear; then, if J_1 is evaluated at the midpoint of an interval,

$$J_1 = (-1)^N\frac{1}{p}\frac{\partial J_0}{\partial p}\left[P + \frac{A}{\cos(\lambda/2p)}\sin 2Q\right]. \tag{17}$$

We now observe that, as expected, J_1 is unbounded at $p = 0$ and at the zeros of $\cos(\lambda/2p)$, unless $\partial J_0/\partial p$ is chosen so that it vanishes at these points. A suitable choice for J_0 is

$$J_0 = p^3(\sin(\lambda/2p) - 6p/\lambda\cos(\lambda/2p)),$$

giving finally

$$J \simeq J_0 + \epsilon J_1 = p^3(\sin(\lambda/2p) - 6p/\lambda\cos(\lambda/2p)) \mp \tfrac{1}{2}\epsilon\lambda[1 + 48(p/\lambda)^2][P\cos(\lambda/2p) + A\sin 2Q]. \tag{18}$$

Of course, other functions J_0 may be chosen, provided $\partial J_0/\partial p$ vanishes at the appropriate points, and these would apparently lead to a different function for $J_0 + \epsilon J_1$ $(= \tilde{J}$, say). However, it can be shown that there would then be a functional relationship between J and $\tilde{J}$, so that the curves $\tilde{J} = \text{const}$ would be identical with the curves $J = \text{const}$.

For comparison with the numerical computations, J must be expressed in terms of the variables ξ, ϕ. The only dynamical variable in J_0 is p, and this is exactly $v(1 - \xi)^{\frac{1}{2}}$. In J_1, we need only zero-order (in ϵ) relations between P, Q and ξ, ϕ. These are

$$P = \tfrac{1}{2}v^2\xi, \quad Q = \frac{\pi}{4} - \tfrac{1}{2}\phi.$$

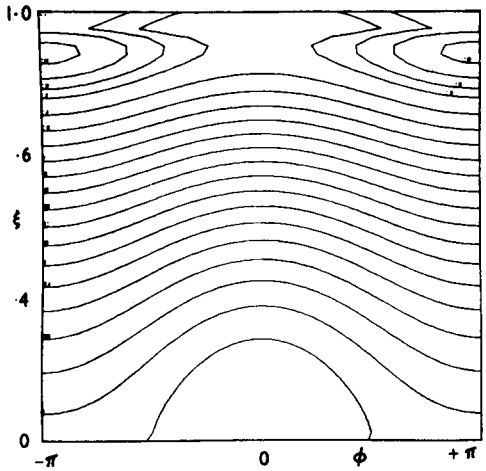


FIG. 4. Invariant curves $J_0 + \epsilon J_1 = \text{const}$, $\epsilon = 0.025$.

Consequently J can be written

$$J = v^3(1-\xi)^{\frac{3}{2}} \sin \frac{\lambda}{2v(1-\xi)^{\frac{1}{2}}} - \frac{6v^4}{\lambda}(1-\xi)^2 \cos \frac{\lambda}{2v(1-\xi)^{\frac{1}{2}}} \mp \tfrac{1}{2}\epsilon\lambda[1 + 48v^2(1-\xi)/\lambda^2] \times \left\{ \tfrac{1}{2}v^2\xi \cos \frac{\lambda}{2v(1-\xi)^{\frac{1}{2}}} + [(\tfrac{1}{2}v^2\xi)^2 + v^2\xi C]^{\frac{1}{2}} \cos\phi \right\}. \quad (19)$$

Curves $J = \text{const}$ are shown in Figs. 4–6 for the same parameters ($v = 2$, $C = 4$, $\lambda = 2\pi$, and $\epsilon = 0.025$, 0.05, 0.10) as were used in the numerical computations of Figs. 1–3. It is seen that there is close agreement between curves derived from the invariant (19) and the computed orbits of type (a).

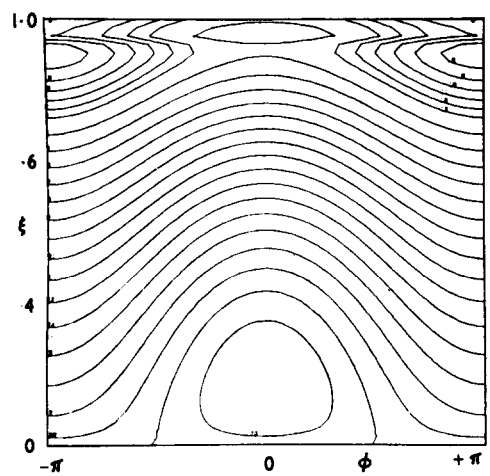


FIG. 5. Invariant curves $J_0 + \epsilon J_1 = \text{const}$, $\epsilon = 0.05$.

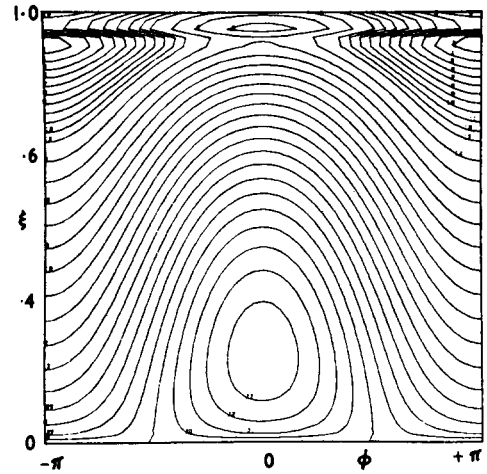


FIG. 6. Invariant curves $J_0 + \epsilon J_1 = \text{const}$, $\epsilon = 0.10$.

SUMMARY AND DISCUSSION

We have investigated the motion of charged particles in a spatially modulated magnetic field, in particular, to see whether any invariant exists when the modulations give resonance with the cyclotron period and destroy the magnetic-moment invariant. A study of numerically computed orbits in a modulated field shows that they fall into two classes: (a) regular orbits, corresponding to the existence of an invariant; and (b) quasi-ergodic orbits. Provided the modulation ϵ of the field is not too great, the regular orbits predominate.

For small ϵ, then, these numerical studies indicate that an invariant exists, even in the resonant situation. We have, therefore, computed this invariant as a power series in ϵ. For this we developed a form of perturbation theory which avoids the usual problem of vanishing denominators near resonance and shows that this phenomena really indicates a change in topology of the invariant curves. The invariant calculated in this way is given by Eq. (18) and shows very good agreement with the numerically computed orbits of type (a). This agreement confirms that, *when an invariant exists*, it can be calculated by our modified form of perturbation theory.

However, perturbation theory can give no direct indication of the transition to quasi-ergodic behavior. An indirect indication may lie in the chain of "islands" which the numerical studies show near the boundary between regular and ergodic behavior. These islands represent further changes in topology which would be reproduced if our perturbation theory were carried to higher order, and we may speculate that the increasingly complex behavior of the analytic curves represent the onset of quasi-ergodic behavior.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Kinematic Singularities of Helicity and Transversity Amplitudes and Asymptotic Regge-Pole Contributions

A. McKERRELL*

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England

(Received 24 January 1968)

The crossing relation for transversity amplitudes is used in a simple proof of the kinematic-singularity structure of helicity amplitudes, both in the general-mass and in various particular-mass cases. A comparison with earlier results is made. The kinematic properties of helicity and transversity amplitudes are employed in the derivation of an asymptotic expression for the contribution of Regge poles to the amplitude.

1. INTRODUCTION

The kinematic-singularity structure of helicity amplitudes has been given by Hara[1] and by Wang.[2] Both the proofs and the final results are, however, rather complicated. At thresholds and pseudothresholds the behavior of helicity amplitudes may be related to that of partial-wave amplitudes, and Frautschi and Jones[3] have shown how, at least in certain cases, this relationship may be used to obtain the powers of the singularities very simply. We show that this result is quite general.

Cohen-Tannoudji *et al.*[4] also derived the kinematic-singularity structure of helicity amplitudes, this time from the properties of invariant amplitudes, without using crossing, and Stapp[5] obtained similar results from basic analyticity properties. Reference 4 also employs the transversity amplitudes of Kotański[6] in a derivation of kinematic constraints on the helicity amplitudes. Section 2 of the present paper is devoted largely to the presentation of a very simple proof of the kinematic-singularity structure of helicity amplitudes. The method employs the same physical assumptions as does Wang,[2] but by using transversity amplitudes[6] and their kinematic properties[4] great simplification is obtained.

In Sec. 3 some results on the high-energy limit of a Regge-pole contribution[7–10] to the helicity amplitude are derived (for general mass), and questions about daughter contributions, conspiracy, and evasion are considered. In Sec. 4 it is shown how recourse to transversity amplitudes may make for a simpler phenomenological analysis, in spite of the various helicity-dependent kinematic factors. Since this part of the paper is for application to inelastic processes, where data are at present rather inadequate, no apology is made for the somewhat crude approximations suggested.

The Appendix is devoted to showing the equivalence of the results of Sec. 2 with those of Wang[2,9] and to the derivation of the kinematic-singularity structure of the helicity amplitudes in various particular-mass cases.

2. KINEMATIC-SINGULARITY STRUCTURE OF HELICITY AMPLITUDES

We review briefly the properties of helicity states in order to establish some notations and conventions. Single-particle rest states of a particle of mass m and spin s transform under rotations according to

$$R\,|\tilde{p}\lambda\rangle = \sum_{\mu} D^{s}_{\mu\lambda}(R)\,|\tilde{p}\mu\rangle, \quad \tilde{p} = (m, \mathbf{0}). \tag{1}$$

We use the conventions of Edmonds[11] for the rotation matrices.

States of momentum p with $p^2 = m^2$ are defined by

$$|p\lambda\rangle = H(p)\,|\tilde{p}\lambda\rangle, \tag{2}$$

where

$$H(p) = R(p)Z(p), \tag{3}$$

and, for

$$p = [m\cosh\zeta,\ m\sinh\zeta(\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta)], \tag{4}$$

we have

$$Z(p) = \exp(-i\zeta K_z), \tag{5}$$

$$R(p) = \exp(-i\varphi J_z)\exp(-i\theta J_y)\exp(-i\varphi J_z). \tag{6}$$

In particular, for

$$q_1 = (m_1, 0, 0, q), \tag{7}$$

$$q_2 = (m_2, 0, 0, -q), \quad q \geq 0, \tag{8}$$

* Present address: Department of Applied Mathematics, The University, Liverpool 3, England.

[1] Y. Hara, Phys. Rev. **136**, B507 (1964).

[2] L.-L. Wang, Phys. Rev. **142**, 1187 (1966).

[3] S. Frautschi and L. Jones, Phys. Rev. **164**, 1918 (1967).

[4] G. Cohen-Tannoudji, A. Morel, and H. Navelet, Ann. Phys. (N.Y.) **46**, 239 (1968).

[5] H. P. Stapp, Phys. Rev. **160**, 1251 (1967).

[6] A. Kotański, Acta Phys. Polon. **29**, 699 (1966).

[7] M. Gell-Mann, M. L. Goldberger, F. E. Low, E. Marx, and F. Zachariasen, Phys. Rev. **133**, B145 (1964).

[8] L.-L. Wang, Phys. Rev. Letters, **16**, 756 (1966).

[9] L.-L. Wang, Phys. Rev. **153**, 1664 (1967).

[10] A. McKerrell and L. Sertorio, Nuovo Cimento **52A**, 1223 (1967).

[11] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1960), 2nd ed.

we define

$$R(q_1) = 1, \tag{9}$$

$$R(q_2) = e^{-i\pi J_y}. \tag{10}$$

The generators **J**, **K** of the homogeneous Lorentz group satisfy the usual commutation relations:

$$[J^i, J^j] = i\epsilon^{ijk}J^k, \tag{11}$$

$$[K^i, J^j] = i\epsilon^{ijk}K^k, \tag{12}$$

$$[K^i, K^j] = -i\epsilon^{ijk}J^k. \tag{13}$$

Two-particle states in the over-all center-of-mass frame are defined in terms of direct products of single-particle states by

$$|p_1\lambda_1 p_2\lambda_2\rangle = R(p_1)(|q_1\lambda_1\rangle \otimes |q_2\lambda_2\rangle), \tag{14}$$

where q_1 and q_2 are defined in (7) and (8), with $q = |\mathbf{p}_1| = |\mathbf{p}_2|$. We have omitted the Jacob and Wick[12] factor $(-1)^{s_2+\lambda_2}$ for the second particle. For momenta p_1, p_2 in the xz plane, the parity operator acts according to

$$P\,|p_1\lambda_1 p_2\lambda_2\rangle = \eta_1\eta_2(-1)^{s_1+\lambda_1+s_2+\lambda_2}e^{-i\pi J_y}\,|p_1 -\lambda_1 p_2 -\lambda_2\rangle, \tag{15}$$

η_1 and η_2 being the intrinsic parities.

We define the scattering amplitude f_{cdab} for the reaction

$$a + b \to c + d \tag{16}$$

in the usual way, and normalize it by the requirement that the differential cross section be

$$\frac{d\sigma}{dt} = [\pi \mathcal{S}_{ab}^2(2s_a + 1)(2s_b + 1)]^{-1}\sum_{abcd}|f_{cdab}|^2, \tag{17}$$

where

$$\mathcal{S}_{ab}^2 = [s - (m_a + m_b)^2][s - (m_a - m_b)^2]. \tag{18}$$

Parity conservation is expressed [from (15)] as

$$f_{cdab} = \eta(-1)^{s_a+s_b+s_c+s_d+a+b+c+d}f_{-c\,-d\,-a\,-b} \tag{19}$$

in agreement with Ref. 4. Here

$$\eta = \eta_a\eta_b\eta_c\eta_d. \tag{20}$$

(We assume that all intrinsic parities are real.)

Transversity amplitudes[6] are defined in terms of helicity amplitudes by

$$T_{cdab} = \sum_{c'd'a'b'} u^*(s_c)_{c'c}u^*(s_d)_{d'd}u(s_a)_{a'a}u(s_b)_{b'b}f_{c'd'a'b'}, \tag{21}$$

where

$$u(s) = D^s(e^{-\frac{1}{2}i\pi J_z}e^{-\frac{1}{2}i\pi J_y}e^{+\frac{1}{2}i\pi J_z}), \tag{22}$$

$$u(s)_{ab} = e^{-\frac{1}{2}\pi(a-b)}d^s_{ab}(-\tfrac{1}{2}\pi). \tag{23}$$

[12] M. Jacob and G.-C. Wick, Ann. Phys. (N.Y.) 7, 404 (1959).

The matrices $u(s)$ are unitary and symmetric and

$$u(s)_{-a-b} = u(s)_{ab}, \tag{24}$$

$$u^*(s)_{ab} = (-1)^{a-b}u(s)_{ab}, \tag{25}$$

$$(-1)^{s+a}u(s)_{ab} = e^{i\pi b}u(s)_{-ab}. \tag{26}$$

These matrices diagonalize the rotation matrices $d^s(\theta)$, so that the crossing matrix for transversity amplitudes takes a particularly simple form[4]

$$T^t_{-c\,-a\,-d\,-b} = \epsilon_0(-1)^{s_a+s_b+s_c+s_d}e^{-i\pi(b-c)} \times e^{i(\alpha\chi_a-b\chi_b+c\chi_c-d\chi_d)}T^s_{cdab}, \tag{27}$$

where $\epsilon_0 = -1$ if a and d are fermions and is $+1$ in all other cases.

Each χ_i is the angle between the s and t rest frames as seen from the corresponding particle. Their sines and cosines are as follows:

$$\cos\chi_a = -\frac{(s + m_a^2 - m_b^2)(t + m_a^2 - m_c^2) + 2m_a^2\Delta}{\mathcal{S}_{ab}\mathcal{C}_{ca}},$$

$$\sin\chi_a = \frac{2m_a\varphi^{\frac{1}{2}}}{\mathcal{S}_{ab}\mathcal{C}_{ca}}, \tag{28}$$

$$\cos\chi_b = \frac{(s + m_b^2 - m_a^2)(t + m_b^2 - m_d^2) - 2m_b^2\Delta}{\mathcal{S}_{ab}\mathcal{C}_{db}},$$

$$\sin\chi_b = \frac{2m_b\varphi^{\frac{1}{2}}}{\mathcal{S}_{ab}\mathcal{C}_{db}}, \tag{29}$$

$$\cos\chi_c = \frac{(s + m_c^2 - m_a^2)(t + m_c^2 - m_a^2) - 2m_c^2\Delta}{\mathcal{S}_{cd}\mathcal{C}_{ca}},$$

$$\sin\chi_c = \frac{2m_c\varphi^{\frac{1}{2}}}{\mathcal{S}_{cd}\mathcal{C}_{ca}}, \tag{30}$$

$$\cos\chi_d = -\frac{(s + m_d^2 - m_c^2)(t + m_d^2 - m_b^2) + 2m_d^2\Delta}{\mathcal{S}_{cd}\mathcal{C}_{db}},$$

$$\sin\chi_d = -\frac{2m_d\varphi^{\frac{1}{2}}}{\mathcal{S}_{cd}\mathcal{C}_{db}}, \tag{31}$$

where

$$\Delta = m_b^2 + m_c^2 - m_d^2 - m_a^2. \tag{32}$$

The threshold factor $\mathcal{S}_{ab}$ was defined in (18); $\mathcal{S}_{cd}$ is defined similarly, and also $\mathcal{C}_{ca}$, $\mathcal{C}_{db}$ with t replacing s. The function φ is zero on the boundary of the physical region:

$$\varphi = st(\Sigma m^2 - s - t) - s(m_b^2 - m_d^2)(m_a^2 - m_c^2) - t(m_a^2 - m_b^2)(m_c^2 - m_d^2) - (m_a^2m_d^2 - m_c^2m_b^2)(m_a^2 + m_d^2 - m_c^2 - m_b^2). \tag{33}$$

We deduce the kinematic singularities of the helicity amplitudes from the result, basic to the argument of Ref. 2 and proved in Ref. 4, that the only

s-kinematic singularities of the t-channel amplitude f^t are those appearing explicitly in

$$f^t_{c'a'd'b'} = (\cos \tfrac{1}{2}\theta_t)^{|\lambda'+\mu'|}(\sin \tfrac{1}{2}\theta_t)^{|\lambda'-\mu'|}\bar{f}^t_{c'a'd'b'}, \quad (34)$$

where

$$\lambda' = d' - b', \quad \mu' = c' - a'. \quad (35)$$

The formulas for the sines and cosines of the s- and t-channel scattering angles follow:

$$\cos\theta_s = \frac{2st + s^2 - s\Sigma m^2 + (m_a^2 - m_b^2)(m_c^2 - m_d^2)}{\mathcal{S}_{ab}\mathcal{S}_{cd}},$$

$$\sin\theta_s = \frac{2[s\varphi]^{\frac{1}{2}}}{\mathcal{S}_{ab}\mathcal{S}_{cd}}, \quad (36)$$

$$\cos\theta_t = \frac{2st + t^2 - t\Sigma m^2 + (m_d^2 - m_b^2)(m_c^2 - m_a^2)}{\mathcal{C}_{ca}\mathcal{C}_{db}},$$

$$\sin\theta_t = \frac{2[t\varphi]^{\frac{1}{2}}}{\mathcal{C}_{ca}\mathcal{C}_{db}}. \quad (37)$$

The s-channel scattering amplitude is written in the form

$$f^s_{cdab} = (\cos \tfrac{1}{2}\theta_s)^{|\lambda+\mu|}(\sin \tfrac{1}{2}\theta_s)^{|\lambda-\mu|}\bar{f}^s_{cdab}, \quad (38)$$

where

$$\lambda = a - b, \quad \mu = c - d, \quad (39)$$

and $\bar{f}^s$ has no t-kinematic singularities.

It is convenient at this point to introduce the following linear combinations of $\bar{f}_{cdab}$ and $\bar{f}_{-c-dab}$, although the necessity for doing so is not yet apparent:

$$f^{\pm}_{cdab} = \bar{f}_{cdab} \pm \eta_c\eta_d(-1)^{s_c - s_d - v}(-1)^{\lambda+\lambda_m}\bar{f}_{-c-dab} \quad (40)$$

$$= \bar{f}_{cdab} \pm \eta_a\eta_b(-1)^{s_a - s_b - v}(-1)^{-\mu+\lambda_m}\bar{f}_{cd-a-b}, \quad \text{by (19).} \quad (41)$$

Here,

$$v = 0 \text{ or } \tfrac{1}{2} \text{ so that } s_c - s_d - v \text{ is integral,} \quad (42)$$

and

$$\lambda_m = \max(|\lambda|, |\mu|) = \tfrac{1}{2}(|\lambda - \mu| + |\lambda + \mu|). \quad (43)$$

The choice of phase in (40) is governed by the application to Regge-pole theory,[7,10] but we postpone discussion of this point since the results on kinematic singularities are quite independent of this model.

We now derive the s-kinematic singularities of the $f^{\pm}$. Since $\bar{f}^t$, $\cos\frac{1}{2}\theta_t$, $\sin\frac{1}{2}\theta_t$, and the crossing matrix are regular at $s = 0$, so also is $\bar{f}^s$. However, the factor $\cos\frac{1}{2}\theta_s$ or $\sin\frac{1}{2}\theta_s$ introduces $s^{\frac{1}{2}}$ singularities into $\bar{f}$, and we have

$$f^{\pm}_{cdab} \sim s^{-\frac{1}{2}\max(|\lambda+\mu|, |\lambda-\mu|)} = s^{-\frac{1}{2}(|\lambda|+|\mu|)}, \quad s \to 0. \quad (44)$$

Our treatment of the singularities at thresholds and pseudothresholds will use the behavior of the transversity amplitudes at these points. Since T^t is kinematically regular at $\mathcal{S}_{ab} = 0$, $\mathcal{S}_{cd} = 0$, it follows from the crossing relation (27), which of course contains no summation, that T^s behaves near these points like

$$\exp[-i(a\chi_a - b\chi_b + c\chi_c - d\chi_d)] = \exp[-\tfrac{1}{2}i[(\chi_a + \chi_b)(a - b) + (\chi_a - \chi_b)(a + b) + (\chi_c + \chi_d)(c - d) + (\chi_c - \chi_d)(c + d)]]. \quad (45)$$

We write

$$\varphi_{ab} = [s - (m_a + m_b)^2]^{\frac{1}{2}}, \quad (46)$$

$$\psi_{ab} = [s - (m_a - m_b)^2]^{\frac{1}{2}}, \quad (47)$$

and similarly for φ_{cd}, ψ_{cd}, and note, following Ref. 4, that the vanishing of the denominators of $\cos\chi_i$ and $\sin\chi_i$ leads to poles or zeros of $\exp[i(\chi_a \pm \chi_b)]$, $\exp[i(\chi_c \pm \chi_d)]$. It is clear that whether a pole or zero occurs depends on the determination of $\varphi^{\frac{1}{2}}$ in the $\sin\chi_i$, and hence on whether ϵ is $+1$ or -1 in

$$\cos\theta_s \sim -i\epsilon\sin\theta_s, \quad (48)$$

at the corresponding pole of $\cos\theta_s$ and $\sin\theta_s$. (Here we use $\sim$ to mean "is asymptotically equal to" rather than "is asymptotically proportional to"; which sense is intended will always be clear.)

The results for the behavior of T^s at thresholds and pseudothresholds are as follows[4]:

$$T^s_{cdab} \sim \varphi_{ab}^{\epsilon_1(a+b)}, \quad \varphi_{ab} \to 0, \quad (49)$$

$$T^s_{cdab} \sim \psi_{ab}^{\epsilon_{ab}\epsilon_2(a-b)}, \quad \psi_{ab} \to 0, \quad (50)$$

$$T^s_{cdab} \sim \varphi_{cd}^{\epsilon_3(c+d)}, \quad \varphi_{cd} \to 0, \quad (51)$$

$$T^s_{cdab} \sim \psi_{cd}^{\epsilon_{cd}\epsilon_4(c-d)}, \quad \psi_{cd} \to 0. \quad (52)$$

Here,

$$\epsilon_{ij} = \frac{m_i - m_j}{|m_i - m_j|}. \quad (53)$$

Since the φ_{ij} and ψ_{ij} are themselves square roots, any kinematically regular amplitude must contain only even powers of each. This is the reason for defining f^+ and f^- as will appear. Consider first the threshold $\varphi_{ab} = 0$. We have

$$\cos\theta_s \sim -i\epsilon_1\sin\theta_s \sim \varphi_{ab}^{-1}, \quad (54)$$

$$\cos\tfrac{1}{2}\theta_s = \frac{1 + \cos\theta_s}{\sin\theta_s}\sin\tfrac{1}{2}\theta_s \sim -i\epsilon_1\sin\tfrac{1}{2}\theta_s \sim \varphi_{ab}^{-\frac{1}{2}}. \quad (55)$$

Thus, from (41),

$$f^{\pm}_{cdab} \sim (\sin \tfrac{1}{2}\theta_s)^{-|\lambda+\mu|-|\lambda-\mu|}(-i\epsilon_1)^{-|\lambda+\mu|} f^{s}_{cdab} \pm \eta_a\eta_b(-1)^{s_a-s_b-v}(-1)^{-\mu+\lambda_m}(\sin \tfrac{1}{2}\theta_s)^{-|\lambda+\mu|-|\lambda-\mu|} \times (-i\epsilon_1)^{-|\lambda-\mu|} f^{s}_{cd\,-a\,-b} \quad (56)$$

$$\sim (\sin \tfrac{1}{2}\theta_s)^{-2\lambda_m}[f^{s}_{cdab} \pm \eta_a\eta_b(-1)^{s_a-s_b-v} \times e^{-i\pi\lambda\epsilon_1} f^{s}_{cd-a-b}]. \quad (57)$$

We write f^s in terms of T^s, using the inverse of (21); because of (26) we can express $f^{\pm}$ as a single sum over the T^s:

$$f^{\pm}_{cdab} \sim \varphi_{ab}^{\lambda_m} \sum_{c'd'a'b'} u(s_c)_{c'c} u(s_d)_{d'd} u^*(s_a)_{a'a} u^*(s_b)_{b'b} \times (1 \pm \eta_a\eta_b(-1)^{v+\epsilon_1(a'+b')}) T^{s}_{c'd'a'b'}. \quad (58)$$

It is now clear that the sum is over even or odd values of $a' + b'$ and therefore that multiplication by a suitable power of φ_{ab} will make $f^{\pm}$ kinematically regular. From (49) the most singular terms on the right of (58) are those with $a' = -\epsilon_1 s_a$, $b' = -\epsilon_1 s_b$, except that the factor in front of T^s may vanish. We allow for this by writing

$$f^{\pm}_{cdab} \sim \varphi_{ab}^{\lambda_m - s_a - s_b + \theta_1}, \quad (59)$$

where

$$\theta_1 = 0 \quad \text{or} \quad 1, \quad (60)$$

and

$$\pm\eta_a\eta_b(-1)^{v-s_a-s_b+\theta_1} = +1. \quad (61)$$

For total and initial orbital angular momentum j and l_1, respectively, parity is $\eta_a\eta_b(-1)^{l_1}$ and j-parity is $(-1)^{j-v}$. Thus (60) and (61) define θ_1 such that $l_1 = j - s_a - s_b + \theta_1$ is the lowest orbital angular momentum for parity $\times$ j-parity $= \pm 1$. This will acquire greater significance when we come to consider Regge poles.

Consider, now, the pseudothreshold $\psi_{ab} = 0$. From (50) we find in an exactly similar manner that

$$f^{\pm}_{cdab} \sim \psi_{ab}^{\lambda_m - s_a - s_b + \theta_2}, \quad (62)$$

where

$$\theta_2 = 0 \quad \text{or} \quad 1, \quad \pm\eta_a\eta_b(-1)^{v-\epsilon_{ab}(s_a-s_b-\theta_2)} = +1. \quad (63)$$

Equation (63) for θ_2 differs from Eq. (61) for θ_1 only in that the parity of the lighter particle is replaced by the parity of the corresponding antiparticle. The same arguments apply to the final-particle threshold and pseudothreshold.

It may be convenient to have the definitions of the θ_i and l_i displayed explicitly, though it should be emphasized that in practice one does not need to refer to formulas since the interpretation in terms of initial and final orbital angular momentum and threshold factors can be used as a mnemonic:

$$\theta_i = 0 \quad \text{or} \quad 1, \quad (64)$$

$$\pm\eta_a\eta_b(-1)^{v-s_a-s_b+\theta_1} = +1, \quad l_1 = j - s_a - s_b + \theta_1, \quad (65)$$

$$\pm[\eta_a\eta_b](-1)^{v-s_a-s_b+\theta_2} = +1, \quad l_2 = j - s_a - s_b + \theta_2, \quad (66)$$

$$\pm\eta_c\eta_d(-1)^{v-s_c-s_d+\theta_3} = +1, \quad l_3 = j - s_c - s_d + \theta_3, \quad (67)$$

$$\pm[\eta_c\eta_d](-1)^{v-s_c-s_d+\theta_4} = +1, \quad l_4 = j - s_c - s_d + \theta_4. \quad (68)$$

Here we have

$$[\eta_i\eta_j] = \begin{cases} \eta_i\bar{\eta}_j = \eta_i\eta_j(-1)^{2s_j}, & m_j < m_i, \\ \bar{\eta}_i\eta_j = \eta_i\eta_j(-1)^{2s_i}, & m_i < m_j. \end{cases} \quad (69)$$

In particular, for BB reactions $[\eta_i\eta_j] = \eta_i\eta_j$, for FF (or $F\bar{F}$) $[\eta_i\eta_j] = -\eta_i\eta_j$, for FB (or BF) $[\eta_i\eta_j] = \pm\eta_i\eta_j$, depending on whether the fermion or boson is more massive.

The final result, then, for the general-mass case is

$$f^{\pm}_{cdab} \sim s^{-\frac{1}{2}(|\lambda|+|\mu|)} \varphi_{ab}^{\lambda_m - s_a - s_b + \theta_1} \psi_{ab}^{\lambda_m - s_a - s_b + \theta_2} \times \varphi_{cd}^{\lambda_m - s_c - s_d + \theta_3} \psi_{cd}^{\lambda_m - s_c - s_d + \theta_4}. \quad (70)$$

Frautschi and Jones[3] brought out the connection between kinematic behavior at thresholds and pseudothresholds and orbital angular momentum, and remarked that this method agreed with the results of Wang[2] in all cases they checked. We show in the Appendix that the agreement is quite general.

3. ASYMPTOTIC EXPRESSIONS FOR REGGE-POLE CONTRIBUTIONS

We take from Ref. 10 the following expression for the contribution to the t-channel amplitude of a single Regge pole of signature σ, parity $\times$ j-parity $= \pm 1$, trajectory $\alpha = \alpha^{\pm}$, and residue $\beta^{\pm}$:

$$f^{t}_{cdab} \sim -\frac{\pi}{2}(1 + \sigma e^{-i\pi(\alpha-v)}) \times \frac{2\alpha+1}{\sin \pi(\alpha - v)} (-1)^{\mu-v} \partial^{\alpha}_{\mu-\lambda}(\pi - \theta_t) \beta^{\pm}_{cdab}, \quad (71)$$

where $a + b \to c + d$ is now the t channel and

$$\partial^{\alpha}_{\mu-\lambda}(\pi - \theta_t) = \frac{\Gamma(2\alpha+1)(\sin \frac{1}{2}\theta_t)^{2\alpha-\mu-\lambda}(\cos \frac{1}{2}\theta_t)^{\mu+\lambda}}{[\Gamma(\alpha+\mu+1)\Gamma(\alpha-\mu+1)\Gamma(\alpha-\lambda+1)\Gamma(\alpha+\lambda+1)]^{\frac{1}{2}}} \times F(\mu - \alpha, \lambda + \alpha; -2\alpha; \operatorname{cosec}^2 \tfrac{1}{2}\theta_t). \quad (72)$$

Changing (2.35) and (2.36) of Ref. 10 to take account of the omission in the present paper of the Jacob and Wick[12] factor $(-1)^{s_2-\lambda_2}$, we have

$$\beta^{\pm}_{cdab} = \pm\eta_a\eta_b(-1)^{s_a-s_b-v}\beta^{\pm}_{cd-a-b}$$
$$= \pm\eta_c\eta_d(-1)^{s_c-s_d-v}\beta_{-c-dab} \quad (73)$$
$$= \eta(-1)^{s_a-s_b-s_c+s_d}\beta^{\pm}_{-c-d-a-b}. \quad (74)$$

Let us write

$$m = \max(\lambda, \mu), \quad m' = \min(\lambda, \mu). \quad (75)$$

Because of (19) we may assume that

$$m \geq |m'|. \quad (76)$$

This saves some trouble with phases.

We write the denominator of (72) as

$$\Gamma(\alpha+1)^2/K_1(\alpha), \quad (77)$$

and use the result[13]

$$\Gamma(2\alpha+1)/\Gamma(\alpha+1) = 2^{2\alpha}\pi^{-\frac{1}{2}}\Gamma(\alpha+\tfrac{1}{2}) \quad (78)$$

to obtain

$$\partial^{\alpha}_{\mu-\lambda}(\pi-\theta_t)$$
$$= \frac{\pi^{-\frac{1}{2}}\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha+1)}K_1(\alpha)(2\sin\tfrac{1}{2}\theta_t)^{2\alpha-m-m'}$$
$$\times(2\cos\tfrac{1}{2}\theta_t)^{m+m'}F(m-\alpha, m'-\alpha; -2\alpha; \operatorname{cosec}^2\tfrac{1}{2}\theta_t) \quad (79)$$
$$= e^{\pm i\pi(\alpha-m)}\frac{\pi^{-\frac{1}{2}}\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha+1)}K_1(\alpha)(2\cos\tfrac{1}{2}\theta_t)^{2\alpha-m+m'}$$
$$\times(2\sin\tfrac{1}{2}\theta_t)^{m-m'}F(m-\alpha, -m'-\alpha; -2\alpha; \sec^2\tfrac{1}{2}\theta_t). \quad (80)$$

The last line follows from the properties of the hypergeometric functions [13] and is a familiar one for the ordinary rotation matrices d^j (for which, however, it is true only for physical values of j). In the asymptotic limit $s \to \infty$ we have, for $t \neq 0$,

$$\cos\tfrac{1}{2}\theta_t \sim e^{\mp i\frac{1}{2}\pi}\sin\tfrac{1}{2}\theta_t; \quad (81)$$

the alternative signs in (80) correspond with those in (81), as is easily seen from the fact that both hypergeometric functions become unity in that limit.

For $t \to 0$ we have $\sin\theta_t \sim t^{\frac{1}{2}}$ and either $\sin\frac{1}{2}\theta_t$ or $\cos\frac{1}{2}\theta_t$ behaves like $t^{\frac{1}{2}}$. The asymptotic limit $F = 1$ no longer applies, but the correct behavior of the scattering amplitude is nevertheless obtained by assuming $F = 1$ in (79) and (80) for $\sin\frac{1}{2}\theta_t \sim t^{\frac{1}{2}}$, $\cos\frac{1}{2}\theta_t \to t^{\frac{1}{2}}$, respectively. The daughter hypothesis is one way of arranging additional contributions to the scattering amplitude to make this true.[14–16] With this modification to the asymptotic limit, (71) becomes

$$f^t_{cdab} \sim -\pi^{\frac{1}{2}}(1+\sigma e^{-i\pi(\alpha-v)})$$
$$\times\frac{\Gamma(\alpha+\frac{3}{2})(-1)^{\mu-v}}{\Gamma(\alpha+1)\sin\pi(\alpha-v)}K_1(\alpha)(2\sin\tfrac{1}{2}\theta_t)^{m-m'}$$
$$\times(2\cos\tfrac{1}{2}\theta_t)^{m+m'}\left(\frac{-4st}{\mathcal{C}_{ab}\mathcal{C}_{cd}}\right)^{\alpha-m}\beta^{\pm}_{cdab},$$
$$s \to \infty, \quad \text{all } t. \quad (82)$$

Now from (19), (73), and (76) we find

$$f^t_{cdab} \sim \pm\eta_c\eta_d(-1)^{\lambda+\lambda_m}(-1)^{s_c-s_d-v}f^t_{-c-dab} \quad (83)$$
$$= \pm\eta_a\eta_b(-1)^{-\mu+\lambda_m}(-1)^{s_a-s_b-v}f^t_{cd-a-b}, \quad (84)$$

valid without the restriction (76). This shows that the choice of phase in (40) and (41) is such that a pole with parity $\times$ j-parity $= \pm 1$, residue $\beta^{\pm}$ contributes only to $f^{\pm}$ in the asymptotic limit, $f^{\mp}$ being of lower order in s.[7,10] The t-kinematic behavior of $f^{t\pm}$ is given by (70) (on replacing s by t), but β must contain an additional factor $(p_{ab}p_{cd})^{\alpha-m} = (\mathcal{C}_{ab}\mathcal{C}_{cd}/4t)^{\alpha-m}$ to cancel that appearing in (82). Thus

$$\beta^{\pm}_{cdab} \propto t^{\lambda_m-\frac{1}{2}(|\lambda|+|\mu|)-\alpha}\varphi^{l_1}_{ab}\psi^{l_2}_{ab}\varphi^{l_3}_{cd}\psi^{l_4}_{cd}, \quad (85)$$

where the l_i are defined in (65)–(68). The interpretation given earlier of the behavior of the amplitude at thresholds and pseudothresholds in terms of initial and final orbital angular momentum should now be more transparent.

Since f^t is regular at $t = 0$, we see from (38) (for f^t) that if f^t is nonzero at $t = 0$, either f^t_{cdab} or f^t_{-c-dab} will be more singular than the other there, and therefor that f^{t+} and f^{t-} differ, if at all, only by a sign. However, we have found that a single pole with residue $\beta^{\pm}$ contributes only to $f^{\pm}$ in the asymptotic limit, and therefore, either there is another trajectory of opposite parity $\times$ j-parity passing through the same value as α at $t = 0$ and contributing equally to $f^{\mp}$ (conspiracy) or evasion takes place, the residues having additional factors t beyond those in (85).[17–19] Working back from (85) we see that f_{cdab} and f_{-c-dab} behave like 1 and $t^{-|m'|}$ (not necessarily respectively) at $t = 0$, and therefore that an extra factor $t^{|m'|}$ must appear in (85) in order to make both f^t_{cdab} and f^t_{-c-dab} regular at $t = 0$. Factorization of the residues also may introduce factors t,[9] and these may be sufficient to satisfy the condition for evasion.[17,18] We write

[13] A. Erdélyi, *et al.*, *Bateman Manuscript Project, Higher Transcendental Functions* (McGraw-Hill Book Co., New York, 1953).

[14] D. Z. Freedman and J.-M. Wang, Phys. Rev. **153**, 1596 (1967).

[15] L. Sertorio (private communication).

[16] D. Z. Freedman and J.-M. Wang, Berkeley Preprint, 1967 (unpublished).

[17] E. Leader, Phys. Rev. **166**, 1599 (1968).

[18] S. Frautschi and L. Jones, Phys. Rev. **167**, 1335 (1968).

[19] H. Högaasen and Ph. Salin, CERN preprint TH 788, 1967 (unpublished).

$K_3(t)$ for any additional factor in β coming from evasion or factorization.

We require also the "sense–nonsense" factors[7–10] such as $[\alpha(\alpha+1)]^{\frac{1}{2}}$ if zero is a sense–nonsense value of α, and $\alpha(\alpha+1)$ if zero is a sense–sense value and the trajectory chooses nonsense, or a nonsense–nonsense value and the trajectory chooses sense. We write the factor coming from this source as $K_2(\alpha)$, and note that $K_1(\alpha)K_2(\alpha)$ is a polynomial in α. We now define reduced residues γ_{cdab} which are free from kinematic singularities and zeros, but are still subject to kinematic constraints at thresholds and pseudothresholds[4]:

$$\beta^{\pm}_{cdab} = K_2(\alpha)t^{-\frac{1}{2}(|\lambda|+|\mu|)}4^{\alpha}(-ts_0)^{\lambda_m-\alpha} \times K_3(t)\varphi_{ab}^{l_1}\psi_{ab}^{l_2}\varphi_{cd}^{l_3}\psi_{cd}^{l_4}\gamma^{\pm}_{cdab}, \quad (86)$$

where s_0 is a scaling constant which may be taken to be the geometric mean of the squares of the masses, or set conveniently at 1 GeV.

We write

$$K(t) = K_1(\alpha)K_2(\alpha)K_3(t)t^{-\frac{1}{2}(|\lambda|+|\mu|)}\varphi_{ab}^{\lambda_m-s_a-s_b+\theta_1} \times \psi_{ab}^{\lambda_m-s_a-s_b+\theta_2}\varphi_{cd}^{\lambda_m-s_c-s_d+\theta_3}\psi_{cd}^{\lambda_m-s_c-s_d+\theta_4}, \quad (87)$$

then, from (82),

$$f^t_{cdab} \sim -\pi^{\frac{1}{2}}(1+\sigma e^{-i\pi(\alpha-v)}) \times \frac{\Gamma(\alpha+\frac{3}{2})(-1)^{\mu-v}K(t)}{\Gamma(\alpha+1)\sin\pi(\alpha-v)}\gamma^{\pm}_{cdab}\left(\frac{s}{s_0}\right)^{\alpha-\lambda_m}. \quad (88)$$

The restriction $\max(\lambda,\mu) > 0$ still holds: other residues are obtained by using (19). The reduced residues $\gamma^{\pm}$ satisfy the same symmetries (73) and (74) as the $\beta^{\pm}$, and from (88), as from (82), we see that in the asymptotic limit $\gamma^{\pm}$ contributes only to $f^{t\pm}$, i.e., to the amplitude with the same index of parity $\times$ j-parity.

4. PHENOMENOLOGY USING TRANSVERSITY AMPLITUDES

In this section we assume that $-z = -\cos\theta_t$, as well as s, is large and positive. Then

$$z \sim \frac{2st}{\mathcal{C}_{ab}\mathcal{C}_{cd}}, \quad t<0, \quad (89)$$

$$\sin\tfrac{1}{2}\theta_t \sim \left(\frac{-st}{\mathcal{C}_{ab}\mathcal{C}_{cd}}\right)^{\frac{1}{2}}, \quad \cos\tfrac{1}{2}\theta_t \sim -i\epsilon\left(\frac{-st}{\mathcal{C}_{ab}\mathcal{C}_{cd}}\right)^{\frac{1}{2}}, \quad \epsilon = \pm 1. \quad (90)$$

From (82), for the contribution of a Regge trajectory α, we obtain

$$f^t_{cdab} \sim -\pi^{\frac{1}{2}}(1+\sigma e^{-i\pi(\alpha-v)}) \times \frac{\Gamma(\alpha+\frac{3}{2})(-1)^{\lambda+v}}{\Gamma(\alpha+1)\sin\pi(\alpha-v)}K_1(\alpha)(i\epsilon)^{\lambda+\mu} \times \left(\frac{-4st}{\mathcal{C}_{ab}\mathcal{C}_{cd}}\right)^{\alpha}\beta^{\pm}_{cdab}. \quad (91)$$

By analogy with Eq. (21), we define transversity residues β^T by

$$\beta^{T\pm}_{cdab} = \sum_{c'd'a'b'} u^*(s_c)_{c'c}u^*(s_d)_{d'd}u(s_a)_{a'a}u(s_b)_{b'b}(-1)^{\lambda+v} \times K_1(\alpha)(i\epsilon)^{\lambda+\mu}\beta^{\pm}_{cdab}, \quad (92)$$

so that the transversity amplitude (in the t channel) is given by

$$T^t_{cdab} \sim -\pi^{\frac{1}{2}}(1+\sigma e^{-i\pi(\alpha-v)}) \times \frac{\Gamma(\alpha+\frac{3}{2})}{\Gamma(\alpha+1)\sin\pi(\alpha-v)}\left(\frac{-4st}{\mathcal{C}_{ab}\mathcal{C}_{cd}}\right)^{\alpha}\beta^{T\pm}_{cdab}. \quad (93)$$

As we have seen, T^t is regular at $t=0$, while at thresholds and pseudothresholds the behavior is given in Eqs. (49)–(52). Let us define reduced residues γ^T by

$$\beta^{T\pm}_{cdab} = (-4s_0t)^{-\alpha}K^T_{12}(t)\varphi_{ab}^{\alpha+\epsilon_1(a+b)}\psi_{ab}^{\alpha+\epsilon_{ab}\epsilon_2(a+b)} \times \varphi_{cd}^{\alpha+\epsilon_3(c+d)}\psi_{cd}^{\alpha+\epsilon_{cd}\epsilon_4(c-d)}\gamma^{T\pm}_{cdab}, \quad (94)$$

and put

$$K^T(t) = K^T_{123}(t)\varphi_{ab}^{\epsilon_1(a+b)}\psi_{ab}^{\epsilon_{ab}\epsilon_2(a-b)}\varphi_{cd}^{\epsilon_3(c+d)}\psi_{cd}^{\epsilon_{cd}\epsilon_4(c-d)}. \quad (95)$$

Then

$$T^t_{cdab} \sim -\pi^{\frac{1}{2}}\{1+\sigma\exp[-i\pi(\alpha-v)]\} \times \frac{\Gamma(\alpha+\frac{3}{2})}{\Gamma(\alpha+1)\sin\pi(\alpha-v)}K^T(t)\gamma^{T\pm}_{cdab}\left(\frac{s}{s_0}\right)^{\alpha}. \quad (96)$$

The function $K^T_{123}(t)$ represents the contribution to the behavior from $K_1(\alpha)$, $K_2(\alpha)$, and $K_3(t)$. Any factor common to all helicity amplitudes will appear in $K^T_{123}(t)$, the others should be treated as constraints, but an "average" function may be defined if there is not too much variation between different amplitudes. The reduced residues are regular at $t=0$ and at thresholds and pseudothresholds. The $\varphi^{\frac{1}{2}}$ cut remains, but can be ignored in the region considered. The variation in φ_{ab} and φ_{cd} will also be small in most cases. For the purposes of phenomenological fitting of $d\sigma/dt$, the various ϵ_i in (94) are clearly irrelevant, since a change in their determination amounts only to a permutation of the amplitudes.

By employing transversity amplitudes we have obtained directly the wide variation in kinematic

behavior of the different amplitudes at the pseudothresholds, instead of having to take this into account by means of the kinematic constraints of Cohen-Tannoudji *et al.*[4] A further advantage is that all the nonzero amplitudes are independent. From an argument close to that leading to Eq. (58), we see that for a trajectory of parity $\times$ j-parity $= \pm 1$, which therefore contributes only to $f^{\pm}$ in the asymptotic limit, we have

$$T^t_{cdab} \sim \pm\eta_a\eta_b(-1)^{v+\epsilon(a+b)}T^t_{cdab}, \tag{97}$$

where ϵ is now that given by (90), and, similarly,

$$T^t_{cdab} \sim \pm\eta_c\eta_d(-1)^{v+\epsilon(c+d)}T^t_{cdab}. \tag{98}$$

Taking these two together we get the equation for parity conservation:

$$T^t_{cdab} = \eta(-1)^{a+b-c-d}T^t_{cdab}. \tag{99}$$

Thus about three quarters of the amplitudes vanish. Since K^T is clearly invariant under change of sign of λ or μ, Eqs. (97)–(99) apply also to $\gamma^{T\pm}$.

The theory outlined in this section will be used in a forthcoming paper to improve the fit to $d\sigma/dt$ for $\pi + N \to \omega + \Delta$ found in Ref. 10.

Since the matrix relating f and T is unitary [that relating $K_2(\alpha)\beta$ and β^T in (92) is also real or pure imaginary], f may be replaced by T^t in (17), the expression for the differential cross section. The situation with measurements involving polarization would of course be much less simple.

Note added in proof: The work of Jackson and Hite[20] on helicity amplitudes shows that the results of this section have to be interpreted with care. The poles in some of the transversity amplitudes at thresholds and pseudothresholds in t do not appear in $d\sigma/dt$. The powers of φ_{ab}, etc., to be used in fitting are the nonnegative ones, including zero whether or not any nonzero transversity amplitude has this behavior. This point will be discussed more fully in a later paper.[21]

ACKNOWLEDGMENTS

I am indebted to Dr. I. T. Drummond for reading the manuscript and for helpful remarks. I wish to thank Dr. L. Sertorio and Dr. D. Branson for useful discussions. I am grateful to Dr. J. C. Polkinghorne for hospitality at Cambridge and for his stimulating interest. I wish to acknowledge a Research Fellowship from Science Research Council, London.

APPENDIX

First we show the equivalence of our results on kinematic singularities with those of Wang.[2] Results for the singularities at $\varphi = 0$ and $s = 0$ are clearly identical. We have to show that our formulas (64)–(70) agree with equations (IV3)f and (III12)f of Ref. 2. Wang's[2] formulas for $\alpha_g(\pm)$, $\beta_g(\pm)$ are clearly of the form

$$J_a + J_b - \tfrac{1}{2}(|\lambda - \mu| + |\lambda + \mu|) - \theta \tag{A1}$$

and

$$J_c + J_d - \tfrac{1}{2}(|\lambda - \mu| + |\lambda + \mu|) - \theta, \tag{A2}$$

respectively, where each θ is 0 or 1. We need only show that the value of θ in every case agrees with that given by (65)–(69). For this we use the result

$$\max\eta \text{ of } n = n - \theta, \quad \theta = 0 \text{ or } 1, \quad \eta(-1)^{n-\theta} = +1, \tag{A3}$$

which is equivalent to Wang's[2] definition of "max η of n."

First consider the cases $BB \to BB$, $BB \to F\bar{F}$, $F\bar{F} \to BB$, $FF \to FF$, where $v = 0$. The sign between f_{cdab} and f_{-c-dab} which Wang writes $\pm$ is, for our amplitudes $f^{\pm}$ in (40), $\pm\eta_c\eta_d(-1)^{s_c+s_d+\lambda+\lambda_m}$, where we change back to Jacob and Wick[12] amplitudes for the purposes of this comparison. Thus for the θ corresponding to α, we have, from (A3) and Wang's[2] formula,

$$\pm\eta_c\eta_d(-1)^{s_c+s_d+\lambda+\lambda_m} \times \eta_{ab}(-1)^{J_a+J_b-\frac{1}{2}(v_a+v_b)+\frac{1}{2}(|\lambda-\mu|-|\lambda+\mu|)+\theta} = +1, \tag{A4}$$

which simplifies to

$$\pm\eta_a\eta_b(-1)^{-s_a-s_b+\theta} = +1 \tag{A5}$$

in agreement with Eq. (65) for θ_1.

The proof for β_1 is similar, if slightly simpler. Agreement for α_2 and β_2 may be shown in the same way, or we may note that α_1 and α_2 are the same or different in Wang[2] exactly when $[\eta_a\eta_b]$ is plus or minus $\eta_a\eta_b$, and similarly for β_1 and β_2.

For $FB \to FB$ interactions, the θ corresponding to Wang's[2] α'_g is defined by

$$+1 = \mp\eta_c\eta_d(-1)^{s_c+s_d-\frac{1}{2}+\lambda+\lambda_m} \times \eta_{ab}(-1)^{s_a+s_b+\frac{1}{2}(|\lambda-\mu|-|\lambda+\mu|)+\theta} \tag{A6}$$

$$= \pm\eta_a\eta_b(-1)^{\frac{1}{2}-s_a-s_b+\theta}, \tag{A7}$$

which agrees again with (65). The case of β_1 is similar. To obtain agreement for α_2 and β_2 we must assume that each fermion is more massive than the corresponding boson (so that $[\eta_i\eta_j] = +\eta_i\eta_j$).

Our formulas agree exactly with Table IV of Ref. 4 for the general-mass case.

We turn now to a consideration of particular-mass cases. At particular values of the masses, two or more singularities may coincide. A special treatment is then required at that point. For singularities with no such

[20] J. D. Jackson and G. E. Hite, Phys. Rev. **169**, 1248 (1968).
[21] A. McKerrell, Nuovo Cimento **56A**, 249 (1968).

coincidence the treatment is the same as in the general-mass case. The same cases I–V of Ref. 4 will be considered.

(i) *Elastic Scattering*: $a \equiv c$, $b \equiv d$, $m_a \neq m_b$, $\eta = +1$. Here both thresholds and both pseudothresholds coincide. Equations (49) and (51) may be combined to give

$$T^s_{cdab} \sim \varphi_{ab}^{\epsilon_1(a+b+c+d)}. \tag{A8}$$

Since by parity invariance (which we always assume),

$$T^s_{cdab} = \eta(-1)^{a+b-c-d}T^s_{cdab}, \tag{A9}$$

only even (or only odd) powers of φ_{ab} ($= \varphi_{cd}$) appear (the other components of T^s vanish). In this case, therefore, we need not define $f^{\pm}$, and it follows that each f^s behaves like the most singular T^s, that with $a = c = -\epsilon_1 s_a$, $b = d = -\epsilon_1 s_b$. [This is consistent with (A9) since $\eta = +1$.] The pseudothreshold result is similar. Thus,

$$f^s_{cdab} \sim (\mathfrak{S}^2_{ab})^{-s_a-s_b}. \tag{A10}$$

At $s = 0$, $\cos\theta_s = +1$, $\sin\theta_s \sim s^{\frac{1}{2}}$. Therefore

$$\cos\tfrac{1}{2}\theta_s \sim \pm 1, \quad \sin\tfrac{1}{2}\theta_s \sim s^{\frac{1}{2}},$$

and so

$$f^s_{cdab} \sim s^{-\frac{1}{2}|\lambda-\mu|}(\mathfrak{S}^2_{ab})^{\lambda_m-s_a-s_b}. \tag{A11}$$

(ii) $m_a = m_b$, $m_c \neq m_d$, $v = 0$. Here $\psi_{ab} = s^{\frac{1}{2}}$ so that the point $s = 0$ requires special consideration. Nevertheless, (50) still applies and

$$T^s_{cdab} \sim s^{\frac{1}{2}\epsilon_{ab}\epsilon_2(a-b)}, \quad s \to 0. \tag{A12}$$

It is true that ϵ_{ab} is no longer defined, but its value will turn out to be irrelevant. At $s = 0$, $\sin\theta_s = \pm 1$, $\cos\theta_s = 0$, $\tan\tfrac{1}{2}\theta_s = \pm 1$. Thus

$$f^{\pm}_{cdab} \sim f^s_{cdab} \pm \eta_a\eta_b(-1)^{s_a-s_b-\mu+\lambda_m}f^s_{cd-a-b} \tag{A13}$$

$$\sim \sum_{c'd'a'b'} [u]_{c'd'a'b'cdab} \times (1 \pm \eta_a\eta_b(-1)^{\lambda-\mu+\lambda_m+a'+b'})T^s_{c'd'a'b'} \tag{A14}$$

$$\sim s^{-\frac{1}{2}(s_a+s_b-\theta_2')}, \tag{A15}$$

where

$$\pm\eta_a\eta_b(-1)^{\lambda-\mu+\lambda_m+\theta_2'+s_a-s_b} = +1. \tag{A16}$$

Combining this with the results from the general-mass case for the other thresholds and pseudothreshold, we find

$$f^{\pm}_{cdab} \sim s^{-\frac{1}{2}(s_a+s_b-\theta_2')}(s - 4m_a^2)^{\frac{1}{2}(\lambda_m-s_a-s_b+\theta_1)} \times \varphi_{cd}^{\lambda_m-s_c-s_d+\theta_3}\psi_{cd}^{\lambda_m-s_c-s_d+\theta_4}. \tag{A17}$$

This is equivalent to the result of Appendix I of Ref. 9 [when due account is taken of the Jacob and Wick[12] factor $(-1)^{s_2-\lambda_2}$] and to Table VII of Ref. 4.

(iii) $m_c = m_d$, $m_a \neq m_b$, $v = 0$. This case is exactly similar to the last, and we have

$$f^{\pm}_{cdab} \sim s^{-\frac{1}{2}(s_c+s_d-\theta_4')}(s - 4m_c^2)^{\frac{1}{2}(\lambda_m-s_c-s_d+\theta_3)} \times \varphi_{ab}^{\lambda_m-s_a-s_b+\theta_1}\psi_{ab}^{\lambda_m-s_a-s_b+\theta_2}, \tag{A18}$$

where

$$\pm\eta_c\eta_d(-1)^{\lambda+\mu+\lambda_m+\theta_4'+s_c-s_d} = +1. \tag{A19}$$

(iv) $m_a = m_b$, $m_c = m_d$, $m_a \neq m_c$, $v = 0$. Again, only the point $s = 0$ (which now coincides with $\psi_{ab} = 0$, $\psi_{cd} = 0$) requires special treatment, but now $\varphi \sim s$, $\sin\chi_i \sim \pm 1$, $\cos\chi_i \sim s^{\frac{1}{2}}$, and so (50) and (52) no longer apply. Now, however, the crossing matrix is no longer singular, and we work directly with helicity amplitudes. We have $\sin\tfrac{1}{2}\theta_t \sim s^{\frac{1}{2}}$, $\cos\tfrac{1}{2}\theta_t = \pm 1$ at $s = 0$. Thus, from (34),

$$f^t_{c'a'd'b'} \sim s^{\frac{1}{2}|\lambda'-\mu'|}. \tag{A20}$$

The crossing relation for helicity amplitudes given by Cohen-Tannoudji *et al.*[4] is as follows, where we use the notation of (27) and the conventions of Edmonds[11] for the rotation matrices:

$$f^s_{cdab} = \epsilon_0(-1)^{2s_b+2s_d}e^{i\pi(b-c)} \times \sum_{a'b'c'd'} d^{s_a}_{aa'}(\chi_a)d^{s_b}_{bb'}(\chi_b)d^{s_c}_{cc'}(\chi_c)d^{s_d}_{dd'}(\chi_d)f^t_{c'a'd'b'}.$$

At $s = 0$ we have, from (28)–(31),

$$\sin\chi_a = \sin\chi_b = \pm 1, \quad \sin\chi_c = \sin\chi_d = \pm 1. \tag{A21}$$

It follows that

$$d^{s_a}_{aa'}(\chi_a)d^{s_b}_{bb'}(\chi_b) = (-1)^{s_a+s_b\mp(a+b)}d^{s_a}_{a-a'}(\chi_a)d^{s_b}_{b-b'}(\chi_b), \tag{A22}$$

and similarly for c, d. Since $a + b$ is an integer, the uncertain sign is irrelevant. Parity conservation implies that

$$f^t_{c'a'd'b'} = \eta_c\eta_{\bar{a}}\eta_{\bar{d}}\eta_b(-1)^{s_a+s_b+s_c+s_d+a'+b'+c'+d'}f^t_{-c'-a'-d'-b'}, \tag{A23}$$

and therefore we may write

$$f^s_{cdab} \sim \sum_{a'b'c'd'} d^{s_a}_{aa'}(\chi_a)d^{s_b}_{bb'}(\chi_b)d^{s_c}_{cc'}(\chi_c)d^{s_d}_{dd'}(\chi_d) \times \tfrac{1}{2}(1 + \eta(-1)^{a+b+c+d-a'+b'+c'-d'})f^t_{c'a'd'b'} \tag{A24}$$

for $s \to 0$. Thus we see from (A20) that f^s is a sum over even or odd nonnegative powers of $s^{\frac{1}{2}}$ at $s = 0$, accordingly as $\eta(-1)^{a+b+c+d}$ is even or odd. Using the

familiar θ notation, we may write

$$f^{s\pm}_{cdab} \sim s^{\frac{1}{2}\theta}(s-4m_a^2)^{\frac{1}{2}(\lambda_m-s_a-s_b+\theta_1)}(s-4m_c^2)^{\lambda_m-s_c-s_d+\theta_3}, \tag{A25}$$

where

$$\theta = 0 \quad \text{or} \quad 1, \quad \eta(-1)^{a+b+c+d+\theta} = +1. \tag{A26}$$

(v) $m_a = m_b = m_c = m_d$, $a \equiv c$, $b \equiv d$, $\eta = +1$.

At $s = 0$ the discussion of (iv) applies; at $s = 4m_a^2$ the argument given for general elastic scattering (i) goes through unchanged. The result is

$$\bar{f}_{cdab} \sim s^{\frac{1}{2}\theta}(s-4m_a^2)^{\lambda_m-s_a-s_b}, \tag{A27}$$

where

$$\theta = 0 \quad \text{or} \quad 1, \quad (-1)^{a+b+c+d+\theta} = +1. \tag{A28}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Spin–Orbit Interactions of the Configuration $l^{n-1}l'l''$

CHARLES ROTH
Department of Mathematics, McGill University, Montreal, Canada

(Received 19 February 1968)

Matrix elements for the spin–orbit interactions of the configuration $l^{n-1}l'l''$ in L–S coupling are expressed as linear combinations of radial integrals.

1. INTRODUCTION

In a previous paper[1] the techniques of the Racah algebra[2] were used to obtain the matrix elements of the Coulomb interactions for the configuration $d^{n-1}sp$ in L–S coupling. The purpose of the present paper is to obtain the spin–orbit interaction-matrix elements for the more general configuration $l^{n-1}l'l''$, again in L–S coupling.

The spin–orbit interactions for the configuration $l^{n-1}l'l''$, denoted by $\mathrm{SO}_{\mathrm{total}}$, can be written as[3,4]

$$\begin{aligned}\mathrm{SO}_{\mathrm{total}} &= \zeta_l \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_1 \cdot l_1 + \cdots + s_{n-1} \cdot l_{n-1} \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &\quad + \zeta_{l'} \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_n' \cdot l_n' \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &\quad + \zeta_{l''} \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_{n+1}'' \cdot l_{n+1}'' \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &= \mathrm{SO}(l) + \mathrm{SO}(l') + \mathrm{SO}(l'').\end{aligned} \tag{1}$$

The parameters ζ_l, $\zeta_{l'}$, and $\zeta_{l''}$ include the gyromagnetic ratios of the spin and orbital motions and an integral over radial eigenfunctions.

2. THE SPIN–ORBIT INTERACTIONS OF THE l ELECTRONS

From (1) the spin–orbit interactions of the l electrons is given by

$$\mathrm{SO}(l) = (n-1)\zeta_l \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_{n-1} \cdot l_{n-1} \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle. \tag{2}$$

Expanding (2) by means of coefficients of fractional parentage[3] yields

$$\begin{aligned}\mathrm{SO}(l) = (n-1)\zeta_l \sum_{\nu_3 S_3 L_3} &[\langle l^{n-2}(\nu_3 S_3 L_3) l_{n-1}(S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_{n-1} \cdot l_{n-1} \,|l^{n-2}(\nu_3 S_3 L_3) l_{n-1}(S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &\times \langle l^{n-1}(\nu_2 S_2 L_2)\{|l^{n-2}(\nu_3 S_3 L_3) l S_2 L_2\rangle \langle l^{n-2}(\nu_3 S_3 L_3) l S_2' L_2'|\} l^{n-1}(\nu_2' S_2' L_2')\rangle].\end{aligned} \tag{3}$$

By ITS Eq. (15.6), the matrix element of (3) [ME(l)] is given by

$$\mathrm{ME}(l) = [\exp \pi i(S' + L + J + 1)] \langle S \| s_{n-1}^{[1]} \| S' \rangle \langle L \| l_{n-1}^{[1]} \| L' \rangle \overline{W} \begin{pmatrix} S & S' & 1 \\ L' & L & J \end{pmatrix} \delta(J, J') \delta(M, M'). \tag{4}$$

---

[1] C. Roth, J. Math. Phys., **9**, 686 (1968).
[2] G. Racah and U. Fano, *Irreducible Tensorial Sets* (Academic Press Inc., New York, 1958), henceforth referred to as ITS.
[3] G. Racah, Phys. Rev. **63**, 367 (1943).
[4] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge University Press, Cambridge, England, 1935), Art. 6[6].

familiar θ notation, we may write

$$f^{s\pm}_{cdab} \sim s^{\frac{1}{2}\theta}(s-4m_a^2)^{\frac{1}{2}(\lambda_m-s_a-s_b+\theta_1)}(s-4m_c^2)^{\lambda_m-s_c-s_d+\theta_3}, \tag{A25}$$

where

$$\theta = 0 \quad \text{or} \quad 1, \quad \eta(-1)^{a+b+c+d+\theta} = +1. \tag{A26}$$

(v) $m_a = m_b = m_c = m_d$, $a \equiv c$, $b \equiv d$, $\eta = +1$.

At $s = 0$ the discussion of (iv) applies; at $s = 4m_a^2$ the argument given for general elastic scattering (i) goes through unchanged. The result is

$$\bar{f}_{cdab} \sim s^{\frac{1}{2}\theta}(s-4m_a^2)^{\lambda_m-s_a-s_b}, \tag{A27}$$

where

$$\theta = 0 \quad \text{or} \quad 1, \quad (-1)^{a+b+c+d+\theta} = +1. \tag{A28}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Spin–Orbit Interactions of the Configuration $l^{n-1}l'l''$

CHARLES ROTH
Department of Mathematics, McGill University, Montreal, Canada

(Received 19 February 1968)

Matrix elements for the spin–orbit interactions of the configuration $l^{n-1}l'l''$ in L–S coupling are expressed as linear combinations of radial integrals.

1. INTRODUCTION

In a previous paper[1] the techniques of the Racah algebra[2] were used to obtain the matrix elements of the Coulomb interactions for the configuration $d^{n-1}sp$ in L–S coupling. The purpose of the present paper is to obtain the spin–orbit interaction-matrix elements for the more general configuration $l^{n-1}l'l''$, again in L–S coupling.

The spin–orbit interactions for the configuration $l^{n-1}l'l''$, denoted by $\mathrm{SO}_{\mathrm{total}}$, can be written as[3,4]

$$\begin{aligned}\mathrm{SO}_{\mathrm{total}} &= \zeta_l \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_1 \cdot l_1 + \cdots + s_{n-1} \cdot l_{n-1} \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &\quad + \zeta_{l'} \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_n' \cdot l_n' \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &\quad + \zeta_{l''} \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_{n+1}'' \cdot l_{n+1}'' \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &= \mathrm{SO}(l) + \mathrm{SO}(l') + \mathrm{SO}(l'').\end{aligned} \tag{1}$$

The parameters ζ_l, $\zeta_{l'}$, and $\zeta_{l''}$ include the gyromagnetic ratios of the spin and orbital motions and an integral over radial eigenfunctions.

2. THE SPIN–ORBIT INTERACTIONS OF THE l ELECTRONS

From (1) the spin–orbit interactions of the l electrons is given by

$$\mathrm{SO}(l) = (n-1)\zeta_l \langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_{n-1} \cdot l_{n-1} \,|l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle. \tag{2}$$

Expanding (2) by means of coefficients of fractional parentage[3] yields

$$\begin{aligned}\mathrm{SO}(l) = (n-1)\zeta_l \sum_{\nu_3 S_3 L_3} &[\langle l^{n-2}(\nu_3 S_3 L_3) l_{n-1}(S_2 L_2) l'(S_1 L_1) l'' SLJM|\, s_{n-1} \cdot l_{n-1} \,|l^{n-2}(\nu_3 S_3 L_3) l_{n-1}(S_2' L_2') l'(S_1' L_1') l'' S'L'J'M'\rangle \\ &\times \langle l^{n-1}(\nu_2 S_2 L_2)\{|l^{n-2}(\nu_3 S_3 L_3) l S_2 L_2\rangle \langle l^{n-2}(\nu_3 S_3 L_3) l S_2' L_2'|\} l^{n-1}(\nu_2' S_2' L_2')\rangle].\end{aligned} \tag{3}$$

By ITS Eq. (15.6), the matrix element of (3) [ME(l)] is given by

$$\mathrm{ME}(l) = [\exp \pi i(S' + L + J + 1)] \langle S \| s_{n-1}^{[1]} \| S' \rangle \langle L \| l_{n-1}^{[1]} \| L' \rangle \overline{W} \begin{pmatrix} S & S' & 1 \\ L' & L & J \end{pmatrix} \delta(J, J') \delta(M, M'). \tag{4}$$

---

[1] C. Roth, J. Math. Phys., **9**, 686 (1968).
[2] G. Racah and U. Fano, *Irreducible Tensorial Sets* (Academic Press Inc., New York, 1958), henceforth referred to as ITS.
[3] G. Racah, Phys. Rev. **63**, 367 (1943).
[4] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge University Press, Cambridge, England, 1935), Art. 6[6].

Successive applications of ITS Eq. (15.7) yield

$$\begin{aligned}\langle S\| s^{[1]}_{n-1}\|S'\rangle &= \langle S_1 \ \tfrac{1}{2} \ S\| s^{[1]}_{n-1}\|S'_1 \ \tfrac{1}{2} \ S'\rangle \\ &= [\exp \pi i(S_2+S_1+S'_1+S'+1)][(2S_1+1)(2S'_1+1)(2S+1)(2S'+1)]^{\frac{1}{2}} \\ &\quad \times \overline{W}\begin{pmatrix} S_1 & S'_1 & 1 \\ S'_2 & S_2 & \frac{1}{2}\end{pmatrix}\overline{W}\begin{pmatrix} S & S' & 1 \\ S'_1 & S_1 & \frac{1}{2}\end{pmatrix}\langle S_2\| s^{[1]}_{n-1}\|S'_2\rangle. \end{aligned} \tag{5}$$

By ITS Eqs. (15.7′) and (14.9),

$$\begin{aligned}\langle S_2\| s^{[1]}_{n-1}\|S'_2\rangle &= \langle S_3 \ \tfrac{1}{2} \ S_2\| s^{[1]}_{n-1}\|S_3 \ \tfrac{1}{2} \ S'_2\rangle \\ &= [\exp \pi i(S_3+S_2)][\tfrac{3}{2}(2S_2+1)(2S'_2+1)]^{\frac{1}{2}}\overline{W}\begin{pmatrix} S_2 & S'_2 & 1 \\ \frac{1}{2} & \frac{1}{2} & S_3\end{pmatrix}. \end{aligned} \tag{6}$$

Similarly, by repeated application of ITS Eq. (15.7), we have

$$\begin{aligned}\langle L\| l^{[1]}_{n-1}\|L'\rangle &= \langle L_1 \ l'' \ L\| l^{[1]}_{n-1}\|L'_1 \ l'' \ L'\rangle \\ &= [\exp \pi i(L_2+L_1+L'_1+L'+l'+l'')][(2L_1+1)(2L'_1+1)(2L+1)(2L'+1)]^{\frac{1}{2}} \\ &\quad \times \overline{W}\begin{pmatrix} L_1 & L'_1 & 1 \\ L'_2 & L_2 & l'\end{pmatrix}\overline{W}\begin{pmatrix} L & L' & 1 \\ L'_1 & L_1 & l''\end{pmatrix}\langle L_2\| l^{[1]}_{n-1}\|L'_2\rangle. \end{aligned} \tag{7}$$

By ITS Eqs. (15.7′) and (14.9),

$$\begin{aligned}\langle L_2\| l^{[1]}_{n-1}\|L'_2\rangle &= \langle L_3 \ l \ L_2\| l^{[1]}_{n-1}\|L_3 \ l \ L'_2\rangle \\ &= [\exp \pi i(L_3+L_2+l+\tfrac{3}{2})][(2L_2+1)(2L'_2+1)l(l+1)(2l+1)]^{\frac{1}{2}}\overline{W}\begin{pmatrix} L_2 & L'_2 & 1 \\ l & l & L_3\end{pmatrix}. \end{aligned} \tag{8}$$

Inserting (4), (5), (6), (7), and (8) into (3) yields, for the spin–orbit interaction for the electrons l,

$$\begin{aligned}\mathrm{SO}(l) &= (n-1)\zeta_l\delta(J,J')\delta(M,M')[\exp \pi i(L_1+L'_1+L+L'+l+l'+l''+2S_2+S_1+S'_1+2S'+J+\tfrac{3}{2})] \\ &\quad \times [\tfrac{3}{2}l(l+1)(2l+1)(2L_2+1)(2L'_2+1)(2L_1+1)(2L'_1+1)(2L+1)(2L'+1) \\ &\quad \times (2S_2+1)(2S'_2+1)(2S_1+1)(2S'_1+1)(2S+1)(2S'+1)]^{\frac{1}{2}} \\ &\quad \times \overline{W}\begin{pmatrix} L_1 & L'_1 & 1 \\ L'_2 & L_2 & l'\end{pmatrix}\overline{W}\begin{pmatrix} L & L' & 1 \\ L'_1 & L_1 & l''\end{pmatrix}\overline{W}\begin{pmatrix} S_1 & S'_1 & 1 \\ S'_2 & S_2 & \frac{1}{2}\end{pmatrix}\overline{W}\begin{pmatrix} S & S' & 1 \\ S'_1 & S_1 & \frac{1}{2}\end{pmatrix}\overline{W}\begin{pmatrix} S & S' & 1 \\ L' & L & J\end{pmatrix} \\ &\quad \times \sum_{\nu_3 S_3 L_3}\Big\{[\exp \pi i(L_3+S_3)]\overline{W}\begin{pmatrix} L_2 & L'_2 & 1 \\ l & l & L_3\end{pmatrix}\overline{W}\begin{pmatrix} S_2 & S'_2 & 1 \\ \frac{1}{2} & \frac{1}{2} & S_3\end{pmatrix}\langle l^{n-1}(\nu_2 S_2 L_2)| \\ &\quad \times \{|l^{n-2}(\nu_3 S_3 L_3)lS_2L_2\rangle\langle l^{n-2}(\nu_3 S_3 L_3)lS'_2L'_2|\}|l^{n-1}(\nu'_2 S'_2 L'_2)\rangle\Big\}. \end{aligned} \tag{9}$$

3. THE SPIN–ORBIT INTERACTION OF THE l' ELECTRON

From (1) the spin–orbit interaction of the l' electron is given by

$$\mathrm{SO}(l') = \zeta_{l'}\langle l^{n-1}(\nu_2 S_2 L_2)l'(S_1L_1)l''SLJM|\, s'_n\cdot l'_n\,|l^{n-1}(\nu'_2 S'_2 L'_2)l'(S'_1L'_1)l''S'L'J'M'\rangle.$$

Since the interaction pertains to the nth electron, the quantum numbers of the first $(n-1)$ electrons must be the same on both sides of the matrix element. Hence we must have

$$\nu'_2 S'_2 L'_2 \equiv \nu_2 S_2 L_2.$$

Then, by ITS Eq. (15.6), we have

$$\begin{aligned}\mathrm{SO}(l') &= \zeta_{l'}[\exp \pi i(S'+L+J+1)]\langle S\| s'^{[1]}_n\|S'\rangle\langle L\| l'^{[1]}_n\|L'\rangle \\ &\quad \times \overline{W}\begin{pmatrix} S & S' & 1 \\ L' & L & J\end{pmatrix}\delta(J,J')\delta(M,M')\delta(\nu_2 S_2 L_2, \nu'_2 S'_2 L'_2). \end{aligned} \tag{10}$$

Now, by ITS Eqs. (15.7), (15.7′), and (14.9), we have

$$\langle S\| s_n'^{[1]} \|S'\rangle = \langle S_1 \ \tfrac{1}{2} \ S\| s_n'^{[1]} \|S_1' \ \tfrac{1}{2} \ S'\rangle$$
$$= [\exp \pi i(S_2 + 2S_1 + S' + \tfrac{3}{2})][\tfrac{3}{2}(2S_1 + 1)(2S_1' + 1)(2S + 1)(2S' + 1)]^{\frac{1}{2}}$$
$$\times \overline{W}\begin{pmatrix} S_1 & S_1' & 1 \\ \frac{1}{2} & \frac{1}{2} & S_2 \end{pmatrix} \overline{W}\begin{pmatrix} S & S' & 1 \\ S_1' & S_1 & \frac{1}{2} \end{pmatrix} \tag{11}$$

and also

$$\langle L\| l_n' \|L'\rangle = \langle L_1 \ l'' \ L\| l_n'^{[1]} \|L_1' \ l'' \ L'\rangle$$
$$= [\exp \pi i(L_2 + L' + l' + l'' + \tfrac{1}{2})][l'(l' + 1)(2l' + 1)(2L_1 + 1)(2L_1' + 1)(2L + 1)(2L' + 1)]^{\frac{1}{2}}$$
$$\times \overline{W}\begin{pmatrix} L_1 & L_1' & 1 \\ l' & l' & L_2 \end{pmatrix} \overline{W}\begin{pmatrix} L & L' & 1 \\ L_1' & L_1 & l'' \end{pmatrix}. \tag{12}$$

Hence substituting (11) and (12) into (10) yields, for the spin–orbit interaction of the l' electron,

$$\mathrm{SO}(l') = \zeta_{l'}[\exp \pi i(L_2 + L + L' + l' + l'' + S_2 + 2S_1 + 2S' + J + 1)]$$
$$\times [\tfrac{3}{2}l'(l' + 1)(2l' + 1)(2L_1 + 1)(2L_1' + 1)(2L + 1)(2L' + 1)(2S_1 + 1)(2S_1' + 1)(2S + 1)(2S' + 1)]^{\frac{1}{2}}$$
$$\times \overline{W}\begin{pmatrix} L_1 & L_1' & 1 \\ l' & l' & L_2 \end{pmatrix} \overline{W}\begin{pmatrix} L & L' & 1 \\ L_1' & L_1 & l'' \end{pmatrix} \overline{W}\begin{pmatrix} S_1 & S_1' & 1 \\ \frac{1}{2} & \frac{1}{2} & S_2 \end{pmatrix} \overline{W}\begin{pmatrix} S & S' & 1 \\ S_1' & S_1 & \frac{1}{2} \end{pmatrix} \overline{W}\begin{pmatrix} S & S' & 1 \\ L' & L & J \end{pmatrix}$$
$$\times \delta(J, J')\delta(M, M')\delta(\nu_2 S_2 L_2, \nu_2' S_2' L_2'). \tag{13}$$

4. THE SPIN–ORBIT INTERACTION OF THE l'' ELECTRON

From (1) the spin–orbit interaction of the l'' electron is given by

$$\mathrm{SO}(l'') = \zeta_{l''}\langle l^{n-1}(\nu_2 S_2 L_2) l'(S_1 L_1) l'' SLJM|\ s_{n+1}'' \cdot l_{n+1}''\ | l^{n-1}(\nu_2' S_2' L_2') l'(S_1' L_1') l'' S' L' J' M'\rangle.$$

Since the interaction pertains to the $(n + 1)$th electron, the quantum numbers of the first n electrons must be the same on both sides.

Hence we must have

$$\nu_2' S_2' L_2' \equiv \nu_2 S_2 L_2 \quad \text{and} \quad S_1' L_1' \equiv S_1 L_1.$$

Then, by ITS Eq. (15.6), we have

$$\mathrm{SO}(l'') = \zeta_{l''}[\exp \pi i(S' + L + J + 1)]\langle S\| s_{n+1}''^{[1]} \|S'\rangle$$
$$\times \langle L\| l_{n+1}''^{[1]} \|L'\rangle \overline{W}\begin{pmatrix} S & S' & 1 \\ L' & L & J \end{pmatrix} \delta(J, J')\delta(M, M')\delta(\nu_2 S_2 L_2, \nu_2' S_2' L_2')\delta(S_1 L_1, S_1' L_1'). \tag{14}$$

By ITS Eqs. (15.7′) and (14.9), we have

$$\langle S\| s_{n+1}''^{[1]} \|S'\rangle = \langle S_1 \ \tfrac{1}{2} \ S\| s_{n+1}''^{[1]} \|S_1' \ \tfrac{1}{2} \ S'\rangle$$
$$= [\exp \pi i(S_1 + S)][\tfrac{3}{2}(2S + 1)(2S' + 1)]^{\frac{1}{2}} \overline{W}\begin{pmatrix} S & S' & 1 \\ \frac{1}{2} & \frac{1}{2} & S_1 \end{pmatrix} \tag{15}$$

and also

$$\langle L\| l_{n+1}''^{[1]} \|L'\rangle = \langle L_1 \ l'' \ L\| l_{n+1}''^{[1]} \|L_1' \ l'' \ L'\rangle$$
$$= [\exp \pi i(L_1 + L + l'' + \tfrac{3}{2})][l''(l'' + 1)(2l'' + 1)(2L + 1)(2L' + 1)]^{\frac{1}{2}} \overline{W}\begin{pmatrix} L & L' & 1 \\ l'' & l'' & L_1 \end{pmatrix}. \tag{16}$$

Substituting (15) and (16) into (14) yields, for the spin–orbit interaction of the l'' electron,

$$\mathrm{SO}(l'') = \zeta_{l''}[\exp \pi i(L_1 + l'' + S_1 + S + S' + J + \tfrac{1}{2})]$$
$$\times [\tfrac{3}{2}l''(l'' + 1)(2l'' + 1)(2L + 1)(2L' + 1)(2S + 1)(2S' + 1)]^{\frac{1}{2}} \overline{W}\begin{pmatrix} L & L' & 1 \\ l'' & l'' & L_1 \end{pmatrix} \overline{W}\begin{pmatrix} S & S' & 1 \\ \frac{1}{2} & \frac{1}{2} & S_1 \end{pmatrix}$$
$$\times \overline{W}\begin{pmatrix} S & S' & 1 \\ L' & L & J \end{pmatrix} \delta(J, J')\delta(M, M')\delta(\nu_2 S_2 L_2, \nu_2' S_2' L_2')\delta(S_1 L_1, S_1' L_1'). \tag{17}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Local Field Theory and Isospin Invariance. II. Free Field Theory of Arbitrary Spin Particles*

P. CARRUTHERS†
Istituto di Fisica "G. Marconi," Università di Roma, Roma, Italia
and
Laboratory of Nuclear Studies, Cornell University, Ithaca, New York

(Received 18 December 1967)

The properties of free field theories of arbitrary spin and isospin particles are investigated. Self-conjugate isofermion ($I = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$) field theories of arbitrary spin S are shown to be nonlocal, with commutators (or anticommutators) failing to vanish outside the light cone. Special attention is given to discrete transformations $\mathcal{C}$, $\mathcal{T}$, and $\mathcal{G}$. For self-conjugate multiplets the parity, charge conjugation, and time-reversal phase factors η_C, η_P, η_T are not arbitrary but obey $\eta_P^2 = (-1)^{2S}$, $\eta_C^2 = (-1)^{2I}\eta_\alpha^2$, $\eta_T^2 = (\eta_\alpha^*)^2$, where the phase $\eta_\alpha = \xi(-1)^{I+\alpha}$ ($|\xi| = 1$, $\alpha = -I, \cdots, +I$) arises when the complex-conjugate representations of $SU(2)$ are transformed to the standard basis. Composite products of $\mathcal{C}$, $\mathcal{T}$, and $\mathcal{G}$ are discussed, with general phases and attention to the dependence on order of the operators. The six possible $\mathcal{CTG}$ operations Θ_i are analyzed. For pair-conjugate multiplets, the operator Θ_i^2 is, in general, a gauge transformation, with phase $(\omega_i^*)^2$ depending on the $\mathcal{C}$, $\mathcal{T}$, $\mathcal{G}$ phases and the order in which $\mathcal{C}$, $\mathcal{T}$, and $\mathcal{G}$ enter into Θ_i. By postulating that $\Theta_i^2\psi\Theta_i^{-2}$ be independent of the order of $\mathcal{C}$, $\mathcal{T}$, and $\mathcal{G}$, we derive the Yang–Tiomno parity factors (± 1, $\pm i$) for pair-conjugate fermions. For self-conjugate multiplets, however, the phase restrictions previously stated lead to a unique order-independent result $\Theta^2\psi\Theta^{-2} = (-1)^{2I+2S}\psi$. The usual local field theory result lacks the factor $(-1)^{2I}$. The latter differs from unity only for the case $2I$ = odd integer, in which case we are in fact dealing with a nonlocal field theory. The technical details of the theory involve construction of local fields from helicity wavefunctions and helicity particle operators. Fields of spin greater than one are described by the Rarita–Schwinger formalism. An appendix treats in detail the form and properties of the high-spin wavefunctions in the helicity basis.

1. INTRODUCTION

It was recently discovered that the usually accepted principles of local quantum field theory preclude the existence of self-conjugate isofermions, i.e., half-odd integral isospin multiplets in which the antiparticles lie in the same multiplet as the particles.[1–8] The result is true for any spin. In Refs. 1–5, it was shown that the field theory of selfconjugate isofermions is nonlocal; in particular, the commutators of many physical densities (energy, isospin, etc.) fail to vanish for spacelike separations, as shown in Paper I.[9] References 6–8 instead assume the existence of a "normal" *TCP* operation Θ, and show that contradictions with the latter arise for the anomalous theories in question. The latter approach depends on locality only to the extent that the usual *TCP* theorem depends on weak local commutativity.[10,11]

In the present work we explore further the relations between locality, internal symmetry, and discrete transformations. Some interesting work along these lines has already appeared.[12–14] Here we are especially interested in restrictions placed on phase factors involved in the discrete transformations.[15] For self-conjugate particles[16] these phases are greatly restricted. We use Rarita–Schwinger fields[17] to describe high spin. The detailed account of the associated helicity wavefunctions given in Appendix A may be of independent interest.

The theoretical reasons for rejecting field theories of self-conjugate isofermions are very similar to those used to arrive at the spin–statistics connection. The similarity is quite striking in the analysis given by Weinberg,[18] who showed that quantization of bosons with anticommutators (or fermions with commutators) led to noncausal commutators, a result quite similar

* Research supported in part by the U.S. Office of Naval Research and the National Science Foundation.

† National Science Foundation Senior Post-doctoral Fellow on leave from Cornell University, Ithaca, New York.

[1] P. Carruthers, Phys. Rev. Letters **18**, 353 (1967).
[2] Huan Lee, Phys. Rev. Letters **18**, 1098 (1967).
[3] Martin Einhorn, Princeton University (unpublished).
[4] G. N. Fleming and E. Kazes, Phys. Rev. Letters **18**, 764 (1967).
[5] Y. S. Jin, Physics Letters **24B**, 411 (1967).
[6] P. Kantor, Phys. Rev. Letters **19**, 394 (1967).
[7] O. Steinmann, Physics Letters **25B**, 234 (1967).
[8] B. Zumnio and D. Zwanziger, Phys. Rev. **164**, 1959 (1967).
[9] P. Carruthers, J. Math. Phys. **9**, 928 (1968). This paper is referred to as I in the text.
[10] R. F. Streater and A. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964).
[11] R. Jost, *General Theory of Quantized Fields* (American Mathematical Society, Providence, Rhode Island, 1965).
[12] L. Michel, in *Group Theoretical Concepts and Methods in Elementary Particles*, F. Gürsey, Ed. (Gordon and Breach, Science Publishers, Inc., New York, 1964), p. 135.
[13] L. C. Biedenharn, J. Nuyts, and H. Ruegg, Commun. Math. Phys. **2**, 231 (1966).
[14] T. D. Lee and G. C. Wick, Phys. Rev. **148**, 1385 (1966).
[15] P. Carruthers, Phys. Letters **26B**, 158 (1968).
[16] In this paper the concept of selfconjugacy refers to the particular group $SU(2)_I \otimes U(1)_Y$ [or better, $U(2)$; see Michel, Ref. 12]. Multiplets which are pair-conjugate with respect to this group may become self-conjugate in a larger group containing operations changing the quantum numbers distinguishing particle from antiparticle. For example, K and $\bar{K}$ belong to the self-conjugate meson multiplet **8** in $SU(3)/Z_3$.
[17] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).
[18] S. Weinberg, Phys. Rev. **133B**, 1318 (1964).

to the results of Sec. 2. An alternative expression of the spin–statistics connection is thus to remark that its violation would lead to nonlocal physical quantities. This aspect is usually ignored because of the more serious difficulties associated with negative energies or indefinite probabilities.

An even more compelling difficulty of field theories of self-conjugate isofermions is the inevitable breakdown of Lorentz covariance when interactions occur. This aspect is analyzed in a sequel to the present work.[19]

2. FIELD OPERATORS AND COMMUTATION RULES

Except for technical details associated with spin, the analysis parallels that given in detail for spin zero in Ref. 9. As there, we distinguish between pair-conjugate (PC) and self-conjugate (SC) isospin multiplets, although for some purposes the latter is a special case of the former. Particle states with isospin I, spin S, momentum p,[20] and helicity λ[21] are created by the operators $a_\alpha^*(p, \lambda)$. For antiparticle states the same task is accomplished by the $2I + 1$ operators $b_\alpha^*(p, \lambda)$. We only consider particles with finite rest mass, and employ the standard spin–statistics assumption. We require that the a_α^* and b_α^* create isospin multiplets whose members are connected by the Condon–Shortley phase convention, in order that the standard apparatus of the theory of the rotation group can be used intact.

The Rarita–Schwinger field may be expanded in the form

$$\psi_\mu^\alpha(x) = \sum_{p\lambda} (a_\alpha(p, \lambda)\chi_\mu(p, \lambda)f_p(x) + \eta_\alpha b_{-\alpha}^*(p, \lambda)\chi_\mu^c(p, \lambda)f_p^*(x)). \quad (2.1)$$

The details are given below. Here ψ_μ^α is a Rarita–Schwinger field—a four-component spinor symmetric in its four-vector indices $\mu = \mu_1, \cdots, \mu_k$, where $k = S$ for bosons and $S - \frac{1}{2}$ for fermions. ψ_μ^α obeys appropriate wave equations and subsidiary conditions. The spin S-wavefunctions are discussed in detail in Appendix A. χ^c denotes the conjugate wavefunction of the antiparticle, and $f_p(x)$ is $N_p e^{-ip\cdot x}$, where N_p is $(2E)^{-\frac{1}{2}}$ or $(M/E)^{\frac{1}{2}}$ for bosons and fermions, respectively. The phase factor η_α is introduced so that the particle and antiparticle constituents of (2.1) transform identically under isospin transformations.[9] η_α can be represented as $\xi(-1)^{I+\alpha}$, where $\xi(|\xi| = 1)$ accounts for the arbitrariness in the phase of χ^c or the operators b_α describing the antiparticle multiplet. We are especially interested in learning what freedom remains to this phase as various restrictions are put on the theory. For SC multiplets we set $b_\alpha = a_\alpha$ in Eq. (2.1).

The field (2.1) transforms as suggested by the notation under Lorentz transformations and as D^* under isospin transformations,[22] where D is an irreducible representation matrix for isospin I. For clarity we approach the generality of Eq. (2.1) in easy steps: first spin 0, $\frac{1}{2}$, 1, followed by bosons of spin $S \geq 2$ and fermions of spin $S \geq \frac{3}{2}$.

A. Spin 0

In this case, the functions χ and χ^c are unity; for PC multiplets

$$\phi^\alpha(x) = \sum_p (a_\alpha(p)f_p(x) + \eta_\alpha b_{-\alpha}^*(p)f_p^*(x)). \quad (2.2)$$

The pertinent field commutators are

$$[\phi^\alpha(x), \phi^\beta{}^*(x')] = i\delta_{\alpha\beta}\Delta(x - x'), \quad \text{(PC, SC)} \quad (2.3)$$

$$[\phi^\alpha(x), \phi^\beta(x')] = 0. \quad \text{(PC)} \quad (2.4)$$

However, for SC multiplets [$b = a$ in Eq. (2.2)] in place of Eq. (2.4) we find

$$[\phi^\alpha(x), \phi^\beta(x')] = \delta_{\alpha,-\beta}\eta_\beta \begin{cases} i\Delta(x - x'), & 2I = \text{even}, \\ \Delta^{(1)}(x - x'), & 2I = \text{odd}. \end{cases} \quad (2.5)$$

In the case $2I =$ even integer, the relation (2.5) is completely equivalent to Eq. (2.3a). However, in the isospinor case this is not so; in particular, the noncausal commutator $\Delta^{(1)}$ contaminates the whole theory. The rest of this section shows that the results (2.3)–(2.5) extend to any spin, provided that we take account of the "statistics" and supply appropriate spin factors in the commutators (or anticommutators).

B. Spin $\frac{1}{2}$

This case has been treated in the usual formalism by Einhorn.[3] The only difference here is the purely formal one of replacing ordinary spin states (referred to a fixed axis) by helicity states. Letting i designate the components of the Dirac field $\psi_i^\alpha(x)$,

$$\psi_i^\alpha(x) = \sum_{p\lambda} (a_\alpha(p, \lambda)u_i(p, \lambda)f_p(x) + \eta_\alpha b_{-\alpha}^*(p, \lambda)v_i(p, \lambda)f_p^*(x)), \quad (2.6)$$

[19] P. Carruthers, Phys. Rev. **172**, 1406 (1968).

[20] For simplicity of writing we write the four-momentum $p = (p_0, \mathbf{p})$, even though the quantity in question only depends on the three-vector $\mathbf{p}$. Later, when space and time inversions are considered, the symbol $-p$ is $(p_0, -\mathbf{p})$; in each case $p_0 = +(\mathbf{p}^2 + m^2)^{\frac{1}{2}}$. Kronecker deltas $\delta pp'$ refer to box normalization and enforce $\mathbf{p} = \mathbf{p}'$.

[21] M. Jacob and G. C. Wick, Ann. Phys. (N.Y.) **7**, 404 (1959).

[22] P. Carruthers and J. P. Krisch, Ann. Phys. (N.Y.) **33**, 1 (1965). In the present work we use I for isospin to avoid confusion of the previous symbol (T) with the time-reversal matrix.

where the wavefunctions u and v are defined in Appendix A, one easily computes the anticommutators:

$$\{\psi_j^\alpha(x), \bar{\psi}_k^\beta(x')\} = i(i\gamma\cdot\partial + m)_{jk}\delta_{\alpha\beta}\Delta(x-x'), \quad \text{(PC, SC)} \tag{2.7}$$

$$\{\psi_j^\alpha(x), \psi_k^\beta(x')\} = 0, \quad \text{(PC)} \tag{2.8}$$

$$\begin{aligned}&\{\psi_j^\alpha(x), \psi_k^\beta(x')\}\\ &= \eta_\beta\delta_{\alpha,-\beta}\sum_{p,\lambda}[u_j(p,\lambda)v_k(p,\lambda)e^{-ip\cdot(x-x')}\\ &\quad + (-1)^{2I}v_j(p,\lambda)u_k(p,\lambda)e^{ip\cdot(x-x')}]/2E\\ &= \eta_\beta\delta_{\alpha,-\beta}C_{1kl}(i\gamma\cdot\partial+m)_{jl}\begin{cases} i\Delta(x-x'), & 2I=\text{even},\\ \Delta^{(1)}(x-x'), & 2I=\text{odd}.\end{cases}\end{aligned} \tag{2.9}$$

In Eq. (2.9) the matrix C_1 is the usual charge-conjugation matrix. Comparison of Eqs. (2.3)–(2.5) with (2.7)–(2.9) reveals a one-to-one correspondence. The self-conjugate isospinor–spinor fields suffer from the same troubles as the zero-spin case. (The usual Majorana field[23] corresponds to the special case $I = 0$.)

C. Spin 1

For the PC vector-meson field we have

$$V_\mu^\alpha(x) = \sum_{p\lambda}(a_\alpha(p,\lambda)e_\mu(p,\lambda)f_p(x) + \eta_\alpha b_{-\alpha}^*(p,\lambda)e_\mu^*(p,\lambda)f_p^*(x)), \tag{2.10}$$

where $e_\mu(p,\lambda)$ is the vector wavefunction for helicity λ, transverse to p^μ, thus guaranteeing $\partial^\mu V_\mu^\alpha = 0$. The commutators are

$$[V_\mu^\alpha(x), V_\nu^{\beta *}(x')] = i\delta_{\alpha\beta}M_{\mu\nu}(-i\partial)\Delta(x-x'), \quad \text{(PC, SC)} \tag{2.11}$$

$$[V_\mu^\alpha(x), V_\nu^\beta(x')] = 0, \quad \text{(PC)} \tag{2.12}$$

$$\begin{aligned}&[V_\mu^\alpha(x), V_\nu^\beta(x')]\\ &= \eta_\beta\delta_{\alpha,-\beta}\sum_{p\lambda}[e_\mu(p,\lambda)e_\nu^*(p,\lambda)e^{-ip\cdot(x-x')}\\ &\quad - (-1)^{2I}e_\mu^*(p,\lambda)e_\nu(p,\lambda)e^{ip\cdot(x-x')}]/2E\\ &= -\eta_\beta\delta_{\alpha\beta}M_{\mu\nu}(-i\partial)\begin{cases} i\Delta(x-x'), & 2I=\text{even},\\ \Delta^{(1)}(x-x'), & 2I=\text{odd}.\end{cases}\end{aligned} \quad \text{(SC)} \tag{2.13}$$

The operator $M_{\mu\nu}$ is $(g_{\mu\nu} + \partial_\mu\partial_\nu/m^2)$, obtained by replacing p_μ in Eq. (A26) by $-i\partial_\mu$.

The extension to arbitrary spin is now quite elementary using the wavefunctions and projection operators of Appendix A.

D. Spin-S Mesons

The only change from the vector field is to replace μ by a set of μ_i $(i = 1, \cdots, S)$ and the polarization vector e_μ by the spin-S wavefunction $e_\mu^S(p,\lambda)$. Equations (2.11)–(2.13) are changed only by the replacements $(\mu,\nu)\to(\boldsymbol{\mu},\boldsymbol{\nu})$, $e_\mu\to e_{\boldsymbol{\mu}}^S$, $-M_{\mu\nu}\to(-1)^S M_{\boldsymbol{\mu}\boldsymbol{\nu}}$.

E. Spin-S Fermions

This case is quite similar to that of spin $\frac{1}{2}$, but we give a few details. The field $\psi_{\boldsymbol{\mu}}^\alpha(x)$ has the expansion

$$\psi_{\boldsymbol{\mu}}^\alpha(x) = \sum_{p\lambda}(a_\alpha(p,\lambda)u_{\boldsymbol{\mu}}^S(p,\lambda)f_p(x) + \eta_\alpha b_{-\alpha}^*(p,\lambda)v_{\boldsymbol{\mu}}^S(p,\lambda)f_p^*(x)), \tag{2.14}$$

where $v_{\boldsymbol{\mu}}^S(p,\lambda) = C\bar{u}_{\boldsymbol{\mu}}^{ST}(p,\lambda)$ (cf. Appendix A).

For either PC or SC fields we find

$$\begin{aligned}&\{\psi_{\boldsymbol{\mu}j}^\alpha(x), \bar{\psi}_{\boldsymbol{\nu}k}^\beta(x')\}\\ &= \sum_{p\lambda}\frac{M}{E}(u_{\boldsymbol{\mu}j}^S(p,\lambda)\bar{u}_{\boldsymbol{\nu}k}^S(p,\lambda)e^{-ip\cdot(x-x')}\\ &\quad + v_{\boldsymbol{\mu}j}^S(p,\lambda)\bar{v}_{\boldsymbol{\nu}k}^S(p,\lambda)e^{ip\cdot(x-x')})\\ &= (-1)^{S-\frac{1}{2}}\sum_{p\lambda}[(\mathfrak{I}_{+\boldsymbol{\mu}\boldsymbol{\nu}}^S)_{jk}e^{-ip\cdot(x-x')} - (\mathfrak{I}_{-\boldsymbol{\mu}\boldsymbol{\nu}}^S)_{jk}e^{ip\cdot(x-x')}]\\ &= i(-1)^{S-\frac{1}{2}}[\Lambda_{\boldsymbol{\mu}\boldsymbol{\nu}}(-i\partial)(i\gamma\cdot\partial+m)]_{jk}\Delta(x-x')\end{aligned} \tag{2.15}$$

on making use of Eqs. (A30) and (A34).

For PC fields we find

$$\{\psi_{\boldsymbol{\mu}j}^\alpha(x), \psi_{\boldsymbol{\nu}k}^\beta(x')\} = 0, \tag{2.16}$$

while for SC fields a calculation similar to that of Eq. (2.15) gives

$$\begin{aligned}&\{\psi_{\boldsymbol{\mu}j}^\alpha(x), \psi_{\boldsymbol{\nu}k}^\beta(x')\}\\ &= \eta_\beta\delta_{\alpha,-\beta}(-1)^{S-\frac{1}{2}}C_{1kl}[\Lambda_{\boldsymbol{\mu}\boldsymbol{\nu}}(-i\partial)(i\gamma\cdot\partial+m)]_{jl}\\ &\quad\times\begin{cases} i\Delta(x-x'), & 2I=\text{even},\\ \Delta^{(1)}(x-x'), & 2I=\text{odd}.\end{cases}\end{aligned} \tag{2.17}$$

In all these cases we expect that for SC fields $\psi^{-\alpha}$ is dependent on $\psi^{\alpha *}$. For $2I$ = even, this dependence is a simple proportionality, in which case the quantities $[\psi,\psi]_\pm$ are not independent of $[\psi,\psi^*]_\pm$. For $2I$ = odd, the dependence is more intricate. The relation between $\psi^{-\alpha}$ and $\psi^{\alpha *}$ is essentially the same as derived in I for zero spin. We do not repeat details of the calculation. Let $\psi_{\boldsymbol{\mu}}^{\alpha C}(x)$ be $\psi_{\boldsymbol{\mu}}^{\alpha *}(x)$ for mesons and $C_1\bar{\psi}_{\boldsymbol{\mu}}^{\alpha T}(x)$ for fermions. An elementary calculation gives the SC conditions:

$$\begin{aligned}&\psi_{\boldsymbol{\mu}}^{-\alpha}(x) = \eta_{-\alpha}\psi_{\boldsymbol{\mu}}^{\alpha C}(x), && 2I=\text{even},\\ &\psi_{\boldsymbol{\mu}}^{-\alpha}(x) = i\eta_{-\alpha}\int d^3x'\,\psi_{\boldsymbol{\mu}}^{\alpha C}(x')\overleftrightarrow{\partial}_0'\Delta^{(1)}(x'-x), && 2I=\text{odd}.\end{aligned} \tag{2.18}$$

Although (5.1) is nonlocal in configuration space, the general expression has a space–time translation-invariant structure. Thus, in momentum space, one

[23] E. Majorana, Nuovo Cimento **14**, 171 (1937).

can write a linear connection between $\psi_\mu^{-\alpha}(p)$ and $\psi_\mu^{\alpha C}(p)$. In the case of interactions this allows for a seemingly more general definition of self-conjugacy, as has been noticed by the authors of Refs. 4, 6, and 7.

We have shown that self-conjugate fermions of any spin give rise to noncausal commutators. It should be amply clear from I that the associated field theory is nonlocal.

3. C, $\mathcal{T}$, AND $\mathcal{C}$ TRANSFORMATIONS

Our discussion of C, $\mathcal{T}$, and $\mathcal{C}$ transformations is a straightforward generalization of the results of I for zero-spin particles. We omit details of arguments essentially the same as given there. The idea is first to define the discrete operations by particle mappings and then to require that the associated field have a sensible transformation law. Results are first given for PC multiplets; equating particle and antiparticle operators yields consistency conditions among the phases which must be satisfied in the SC case. Repeated use is made of results derived in Appendix A.

Requiring that antiparticle conjugation C maps $\psi_\mu^\alpha(x)$ into $\psi_\mu^{\alpha *}(x)$ with an α-independent phase, we obtain for the particle transformations[24]

$$\begin{aligned} \mathrm{C}a_\alpha(p,\lambda)\mathrm{C}^{-1} &= \eta_C \eta_\alpha^* b_{-\alpha}(p,\lambda), \\ \mathrm{C}b_\alpha(p,\lambda)\mathrm{C}^{-1} &= \eta_C^* \eta_{-\alpha} a_{-\alpha}(p,\lambda). \end{aligned} \tag{3.1}$$

For SC multiplets the phases on the right-hand side of Eq. (3.1) are equal, giving the constraint

$$\eta_C^2 = \eta_\alpha \eta_{-\alpha} = (-1)^{2I}\eta_\alpha^2. \tag{3.2}$$

The field transformation law follows easily, using Eqs. (2.2) and the last of Eqs. (A24):

$$\mathrm{C}\psi_\mu^\alpha(x)\mathrm{C}^{-1} = \eta_C C_1 \bar{\psi}_\mu^{\alpha T}(x). \tag{3.3}$$

The matrix C_1 is $i\gamma_0\gamma_2$ for fermions and unity for mesons. The bar-transpose operation is the usual one for fermions; for bosons it simply means taking the Hermitian adjoint. We do not consider here the other possible mapping for C discussed in I, which was nonlocal in the independent fields but unitarily equivalent to (3.2) for the anomalous case.

The parity transformation is somewhat more intricate because of the more complicated behavior of the wavefunctions under parity. For spin 0 we have

$$\begin{aligned} \mathcal{T}a_\alpha(p)\mathcal{T}^{-1} &= \eta_P a_\alpha(-p), \\ \mathcal{T}b_\alpha(p)\mathcal{T}^{-1} &= \eta_P^* b_\alpha(-p), \\ \mathcal{T}\psi^\alpha(\mathbf{x},t)\mathcal{T}^{-1} &= \eta_P \psi^\alpha(-\mathbf{x},t). \end{aligned} \tag{3.4}$$

[24] We always assume that the vacuum state $|0\rangle$ is an eigenvector of C, $\mathcal{T}$, and $\mathcal{C}$ with unit eigenvalue.

In the usual discussions of the Dirac field, one uses wavefunctions and operators referring to a fixed axis. We review this before going over to the helicity formalism. From the relations (A4) we note the well-known sign difference of the particle and antiparticle transformations:

$$\begin{aligned} \mathcal{T}a_\alpha(p,s)\mathcal{T}^{-1} &= \eta_P a_\alpha(-p,s), \\ \mathcal{T}b_\alpha(p,s)\mathcal{T}^{-1} &= -\eta_P^* b_\alpha(-p,s), \\ \mathcal{T}\psi^\alpha(\mathbf{x},t)\mathcal{T}^{-1} &= \eta_P P\psi^\alpha(-\mathbf{x},t), \end{aligned} \tag{3.5}$$

where $P = \gamma_0$.

If we use the helicity basis, the transformation laws for the operators $a_\alpha(p,\lambda)$ and $b_\alpha(p,\lambda)$ follow from (3.5) and (B6):

$$\begin{aligned} \mathcal{T}a_\alpha(p,\lambda)\mathcal{T}^{-1} &= \eta_P(-1)^{\frac{1}{2}-\lambda}e^{-2i\lambda\phi}a_\alpha(-p,-\lambda), \\ \mathcal{T}b_\alpha(p,\lambda)\mathcal{T}^{-1} &= -\eta_P^*(-1)^{\frac{1}{2}-\lambda}e^{-2i\lambda\phi}b_\alpha(-p,-\lambda). \end{aligned} \tag{3.6}$$

We note that the helicity changes sign, although s does not, as expected. The detailed structure of (3.6) could have been surmised from the identities (A11), which guarantee that the spinor-field parity-transformation law is the same as shown in (3.5).

For spin 1, inspection of (A16) leads us to the definition

$$\begin{aligned} \mathcal{T}a_\alpha(p,\lambda)\mathcal{T}^{-1} &= \eta_P(-1)^{1-\lambda}e^{-2i\lambda\phi}a_\alpha(-p,-\lambda), \\ \mathcal{T}b_\alpha(p,\lambda)\mathcal{T}^{-1} &= \eta_P^*(-1)^{1-\lambda}e^{-2i\lambda\phi}b_\alpha(-p,-\lambda). \end{aligned} \tag{3.7}$$

This gives the usual field-transformation law

$$\mathcal{T}V_\mu^\alpha(\mathbf{x},t)\mathcal{T}^{-1} = \eta_P(-g_{\mu\mu})V_\mu^\alpha(-\mathbf{x},t) \tag{3.8}$$

on using Eqs. (2.10), (3.7), and (A16).

The symmetry relations for higher-spin wavefunctions summarized in Eqs. (A21)–(A23) lead the definition for arbitrary spin (including the previous cases):

$$\begin{aligned} \mathcal{T}a_\alpha(p,\lambda)\mathcal{T}^{-1} &= \eta_P(-1)^{s-\lambda}e^{-2i\lambda\phi}a_\alpha(-p,-\lambda), \\ \mathcal{T}b_\alpha(p,\lambda)\mathcal{T}^{-1} &= (-1)^{2S}\eta_P^*(-1)^{s-\lambda}e^{-2i\lambda\phi}b_\alpha(-p,-\lambda), \end{aligned} \tag{3.9}$$

where the factor $(-1)^{2S}$ is especially to be noted. With these conventions the spin-S field transforms as

$$\mathcal{T}\psi_\mu^\alpha(\mathbf{x},t)\mathcal{T}^{-1} = \eta_P\Pi(-g_{\mu\mu})P\psi_\mu^\alpha(-\mathbf{x},t). \tag{3.10}$$

The over-all sign has been chosen to give η_P its conventional value. P is again γ_0 for fermions and unity for mesons.

In the special case of SC multiplets the consistency of Eqs. (3.9) requires that

$$\eta_P^2 = (-1)^{2S}. \tag{3.11}$$

For SC mesons the parity factor is thus ± 1, while for Majorana particles it is $\pm i$.

The discussion of time reversal proceeds in a similar way. Recalling that $\mathcal{T}$ is antilinear, for spin 0 we have

$$\mathcal{T}a_\alpha(p)\mathcal{T}^{-1} = \eta_T a_\alpha(-p),$$
$$\mathcal{T}b_\alpha(p)\mathcal{T}^{-1} = (\eta_T^*\eta_\alpha^*/\eta_\alpha)b_\alpha(-p). \quad (3.12)$$

It is important to note that the $\mathcal{T}$-transformation phase depends on the isospin variable, in contrast to the parity operation, which commutes with the isospin. The field $\psi^\alpha(x)$ transforms as

$$\mathcal{T}\psi^\alpha(\mathbf{x}, t)\mathcal{T}^{-1} = \eta_T\psi^\alpha(\mathbf{x}, -t). \quad (3.13)$$

For spin $\frac{1}{2}$ we first work in the usual basis, and then give results for the helicity basis. The result (A6) indicates the necessity for a spin-dependent phase factor:

$$\mathcal{T}a_\alpha(p, s)\mathcal{T}^{-1} = \eta_T(-1)^{\frac{1}{2}+s}a_\alpha(-p, -s),$$
$$\mathcal{T}b_\alpha(p, s)\mathcal{T}^{-1} = (\eta_T^*\eta_\alpha^*/\eta_\alpha)(-1)^{\frac{1}{2}+s}b_\alpha(-p, -s). \quad (3.14)$$

The Dirac field transforms as

$$\mathcal{T}\psi^\alpha(\mathbf{x}, t)\mathcal{T}^{-1} = \eta_T T\psi^\alpha(\mathbf{x}, -t), \quad (3.15)$$

where the matrix $T = \gamma_3\gamma_1$ is real.

The transformation of the helicity operators follows directly from Eqs. (B6) and (3.14):

$$\mathcal{T}a_\alpha(p, \lambda)\mathcal{T}^{-1} = \eta_T e^{2i\lambda\phi}a_\alpha(-p, \lambda),$$
$$\mathcal{T}b_\alpha(p, \lambda)\mathcal{T}^{-1} = (\eta_T^*\eta_\alpha^*/\eta_\alpha)e^{2i\lambda\phi}b_\alpha(-p, \lambda). \quad (3.16)$$

These results could have been derived using Eqs. (A12), requiring the transformation (3.15).

For spin 1, Eqs. (A16) suggest that the transformations of a_α and b_α are formally identical to (3.16). The field transformation law is

$$\mathcal{T}V_\mu^\alpha(\mathbf{x}, t)\mathcal{T}^{-1} = \eta_T g_{\mu\mu}V_\mu^\alpha(\mathbf{x}, -t). \quad (3.17)$$

The general case is now clear from the preceding examples and the second of Eqs. (A24); Eq. (3.16) summarizes the form of the $\mathcal{T}$ transformation for any spin. The field undergoes the transformation

$$\mathcal{T}\psi_\mu^\alpha(\mathbf{x}, t)\mathcal{T}^{-1} = \eta_T\Pi(g_{\mu\mu})T\psi_\mu^\alpha(\mathbf{x}, -t). \quad (3.18)$$

For the special case of SC fields the consistency of (3.16) implies

$$\eta_T^2 = (\eta_\alpha^*)^2. \quad (3.19)$$

Combining Eqs. (3.2), (3.11), and (3.19) leads to the SC consistency conditions

$$\eta_P^2 = (-1)^{2S},$$
$$(\eta_C\eta_T)^2 = (-1)^{2I}. \quad (3.20)$$

We use these relations extensively in the next section.

4. COMPOSITE DISCRETE TRANSFORMATIONS; *TCP*

It is of some interest to consider, instead of the separate transformations $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{T}$, various combinations of them.[25] A particularly important combination is the $\mathcal{CPT}$ operation, since every *local* relativistic quantum field theory is invariant under this transformation. In this section we study the composite transformations with emphasis on their dependence on operator order. The phases are kept absolutely general (except for SC multiplets, for which consistency restrictions arise) to exhibit an aspect of the theory usually obscured by a remark of the type "clearly we can choose phases so that $\cdots$."

In order to handle all spins at the same time, we rewrite the $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{T}$ transformations of Sec. 3 in a more convenient order. Let i ($i = 1, 2, 3, 4$) be the spinor index, when appropriate. Then the transformations in question are

$$\mathcal{C}\psi_{\mu i}^\alpha(x)\mathcal{C}^{-1} = \eta_C\psi_{\mu j}^{\alpha *}(x)C_{ji},$$
$$\mathcal{P}\psi_{\mu i}^\alpha(\mathbf{x}, t)\mathcal{P}^{-1} = \eta_P\Pi(-g_{\mu\mu})\psi_{\mu j}^\alpha(-\mathbf{x}, t)P_{ji}, \quad (4.1)$$
$$\mathcal{T}\psi_{\mu i}^\alpha(\mathbf{x}, t)\mathcal{T}^{-1} = \eta_T\Pi(g_{\mu\mu})\psi_{\mu j}^\alpha(\mathbf{x}, -t)T_{ji}'.$$

Here summation over repeated indices is understood. For fermions the real matrices $C = C_1\gamma_0$, $C = -i\gamma_2$, $P = \gamma_0$, $T = \gamma_3\gamma_1$, while for bosons $C = P = T = I$ and no spinor index occurs. T' is the transpose of T: $T' = -T$.

The matrices C, P, and T have been put to the right of the fields in (4.1) so that the matrix products occur in the same order as the operator products which induce the transformation. Since C, P, and T are real, the conjugate fields also transform as (4.1) except that each phase η_i is replaced by η_i^*. Thus far the various phases: η_C, η_P, η_T, η_α are completely general, subject to the unimodular condition and also Eq. (2.3). Hence four arbitrary phases occur in the behavior of the PC field; for SC fields η_P and one of η_C or η_T are restricted by Eqs. (3.20).

We begin by discussing bilinear products. We find easily the relations for the squares:

$$\mathcal{C}^2\psi\mathcal{C}^2 = \psi,$$
$$\mathcal{T}^2\psi\mathcal{T}^{-2} = (-1)^{2S}\psi, \quad (4.2)$$
$$\mathcal{P}^2\psi\mathcal{P}^{-2} = \eta_P^2\psi.$$

The various field labels (α, μ, x) are the same on both sides of Eqs. (4.2). The phases η_C and η_T drop out since $\mathcal{C}$ and $\mathcal{T}$ involve conjugation in one form or another. In the PC case η_P^2 is an arbitrary number of the form $\exp(2i\varphi_P)$, $\eta_P = \exp(i\varphi_P)$, so that $\mathcal{P}^2$ is a gauge transformation generated by the difference in number of particles and antiparticles. For SC multiplets this operator vanishes, so no such gauge transformation is possible. However, in this case η_P^2 is

[25] G. Feinberg and S. Weinberg [Nuovo Cimento **14**, 571 (1959)] have discussed composite transformations for spin 0, $\frac{1}{2}$, and 1 without considering internal symmetries. However, essentially nothing new results from the extension to general spin and isospin.

fixed to be $(-1)^{2S}$. Thus introducing the Fermi counting operator F_s, which is $+1$ for states with an odd number of fermions and 0 for states with an even number of Fermions or any number of bosons, we have

$$\begin{aligned} \mathcal{C}^2 &= I, \\ \mathcal{T}^2 &= (-1)^{F_s}, \\ \mathcal{P}^2 &= \begin{cases} (-1)^{F_s}, & \text{SC multiplets,} \\ \exp\,[2i\varphi_P(N-\bar{N})], & \text{PC multiplets.} \end{cases} \end{aligned} \tag{4.3}$$

The products $\mathcal{CP}$ and $\mathcal{PC}$ differ in general:

$$\begin{aligned} \mathcal{CP}\psi^{\alpha}_{\mu i}(\mathbf{x}, t)(\mathcal{CP})^{-1} &= \eta_P\eta_C\Pi(-g_{\mu\mu})\psi^{\alpha *}_{\mu j}(-\mathbf{x}, t)(CP)_{ji}, \\ (\mathcal{PC})\psi^{\alpha}_{\mu i}(\mathbf{x}, t)(\mathcal{PC})^{-1} &= \eta_C\eta_P^{*}\Pi(-g_{\mu\mu})\psi^{\alpha *}_{\mu j}(-\mathbf{x}, t)(PC)_{ji}. \end{aligned} \tag{4.4}$$

The product CP differs from PC by $(-1)^{2S}$, so the second transformation in (4.4) has phase $\eta_C\eta_P^{*}(-1)^{2S}$ as compared to $\eta_C\eta_P$ in the first. Thus, in general, $\mathcal{CP} \neq \mathcal{PC}$. However, for SC particles these phases coincide, giving

$$\mathcal{CP} = \mathcal{PC} \quad \text{(SC)}. \tag{4.5}$$

The squared operators $(\mathcal{CP})^2 = (\mathcal{PC})^2$ are independent of η_C, η_P, however:

$$\begin{aligned} (\mathcal{CP})^2\psi(\mathcal{CP})^{-2} &= (-1)^{2S}\psi, \\ (\mathcal{CP})^2 = (\mathcal{PC})^2 &= (-1)^{F_s}. \end{aligned} \tag{4.6}$$

For $\mathcal{PT}$ the situation is similar:

$$\begin{aligned} \mathcal{PT}\psi^{\alpha}_{\mu i}(x)(\mathcal{PT})^{-1} &= \eta_P\eta_T(-1)^{\eta_S}\psi^{\alpha}_{\mu j}(-x)(PT')_{ji}, \\ \mathcal{TP}\psi^{\alpha}_{\mu i}(x)(\mathcal{TP})^{-1} &= \eta_P^{*}\eta_T(-1)^{\eta_S}\psi^{\alpha}_{\mu j}(-x)(T'P)_{ji}, \end{aligned} \tag{4.7}$$

so that, in general, $\mathcal{PT} \neq \mathcal{TP}$, although for SC multiplets $\mathcal{PT} = (-1)^{F_s}\mathcal{TP}$. η_S is the greatest integer contained in S: $\eta_S = S$ for mesons and $S - \frac{1}{2}$ for fermions. However, $(\mathcal{PT})^2 = (\mathcal{TP})^2$ is independent of phases

$$\begin{aligned} (\mathcal{PT})^2\psi(\mathcal{PT})^{-2} &= (-1)^{2S}\psi, \\ (\mathcal{PT})^2 = (\mathcal{TP})^2 &= (-1)^{F_s}. \end{aligned} \tag{4.8}$$

The operations $\mathcal{CT}$ and $\mathcal{TC}$ give

$$\begin{aligned} \mathcal{CT}\psi^{\alpha}_{\mu i}(\mathbf{x}, t)(\mathcal{CT})^{-1} &= \eta_C\eta_T\psi^{\alpha *}_{\mu j}(\mathbf{x}, -t)(CT')_{ji}, \\ \mathcal{TC}\psi^{\alpha}_{\mu i}(\mathbf{x}, t)(\mathcal{TC})^{-1} &= \eta_C^{*}\eta_T^{*}\psi^{\alpha *}_{\mu j}(\mathbf{x}, -t)(T'C)_{ji}, \end{aligned} \tag{4.9}$$

so that $\mathcal{CT} = \mathcal{TC}$ only when $\eta_C\eta_T$ is real. When SC multiplets are under consideration, $\eta_C^{*}\eta_T^{*}$ is equal to $(-1)^{2I}\eta_C\eta_T$, so that $\mathcal{CT} = (-1)^{F_I}\mathcal{TC}$, F_I being the *isofermion number*.

The corresponding squared operators give

$$\begin{aligned} (\mathcal{CT})^2\psi(\mathcal{CT})^{-2} &= (-1)^{2S}(\eta_C^{*}\eta_T^{*})^2\psi, \\ (\mathcal{TC})^2\psi(\mathcal{TC})^{-2} &= (-1)^{2S}(\eta_C\eta_T)^2\psi. \end{aligned} \tag{4.10}$$

For SC multiplets $(\eta_C\eta_T)^2$ is $(-1)^{2I}$, giving

$$(\mathcal{CT})^2 = (\mathcal{TC})^2 = (-1)^{F_s+F_I}. \tag{4.11}$$

It may be helpful to classify the previous results in order of decreasing generality.

1. Relations true for arbitrary phases:

$$\begin{aligned} \mathcal{C}^2 &= I, \\ \mathcal{P}^2 &= \begin{cases} (-1)^{F_s} & \text{(SC)}, \\ \exp\,[2i\varphi_P(N-\bar{N})], & \text{(PC)} \end{cases} \\ \mathcal{T}^2 &= (-1)^{F_s}, \\ (\mathcal{CP})^2 = (\mathcal{PC})^2 &= (-1)^{F_s}, \\ (\mathcal{PT})^2 = (\mathcal{TP})^2 &= (-1)^{F_s}. \end{aligned} \tag{4.12}$$

2. SC multiplets; extra relations:

$$\begin{aligned} \mathcal{CP} &= \mathcal{PC}, \\ \mathcal{PT} &= (-1)^{F_s}\mathcal{TP}, \\ \mathcal{CT} &= (-1)^{F_I}\mathcal{TC}, \\ (\mathcal{TC})^2 = (\mathcal{CT})^2 &= (-1)^{F_s+F_I}. \end{aligned} \tag{4.13}$$

3. PC multiplets; simple useful phase choices:

$$\eta_P^2 = 1 \Rightarrow \begin{cases} \mathcal{P}^2 = I, \\ \mathcal{CP} = (-1)^{F_s}\mathcal{PC}, \\ \mathcal{PT} = \mathcal{TP}, \end{cases}$$

$$(\eta_C\eta_T)^2 = 1 \Rightarrow \begin{cases} \mathcal{CT} = \mathcal{TC}, \\ (\mathcal{CT})^2 = (\mathcal{TC})^2 = (-1)^{F_s}. \end{cases} \tag{4.14}$$

Now we discuss the $\mathcal{CPT}$ transformation Θ. It is important to consider the six distinct operations Θ_i corresponding to the permutations of the order of $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{T}$. The six operators are

$$\begin{aligned} \Theta_1 &= \mathcal{CPT}, \quad \Theta_4 = \mathcal{TPC}, \\ \Theta_2 &= \mathcal{TCP}, \quad \Theta_5 = \mathcal{PCT}, \\ \Theta_3 &= \mathcal{PTC}, \quad \Theta_6 = \mathcal{CTP}. \end{aligned} \tag{4.15}$$

Under Θ_1, the general field transforms as

$$\begin{aligned} \Theta_1\psi^{\alpha}_{\mu i}(x)\Theta_1^{-1} &= (-1)^{\eta_S}\psi^{\alpha *}_{\mu j}(-x)(PCT')_{ji}\eta_C\eta_P\eta_T, \\ PCT' &= -i\gamma_0\gamma_1\gamma_2\gamma_3 = i\gamma_5, \quad 2S \text{ odd}, \\ &= 1, \quad 2S \text{ even}. \end{aligned} \tag{4.16}$$

$\gamma_5 \equiv \gamma^0\gamma^1\gamma^2\gamma^3$ is purely imaginary, symmetric, and anti-Hermitian; $\gamma_5^2 = -I$.

Clearly, every Θ transformation can be written as

$$\Theta_i\psi^{\alpha}_{\mu j}(x)\Theta_i^{-1} = (-1)^{\eta_S}\omega_i\psi^{\alpha *}_{\mu k}(-x)(\theta_i)_{kj}, \tag{4.17}$$

where ω_i is a product of $\mathcal{CPT}$ phases and θ_i is the matrix product of C, P, and T' in the same order as

$\mathcal{C}$, $\mathcal{S}$, and $\mathcal{T}$ occur in Θ_i:

$$\theta_i = 1, \quad 2S = \text{even},$$
$$\theta_1 = \theta_2 = \theta_6 = -\theta_3 = -\theta_4 = -\theta_5 = i\gamma_5, \quad 2S = \text{odd}. \tag{4.18}$$

The phase factors ω_i are

$$\omega_1 = \omega_2^* = \omega_3^* = \eta_P\eta_C\eta_T,$$
$$\omega_4 = \omega_5^* = \omega_6^* = \eta_P\eta_C^*\eta_T^*. \tag{4.19}$$

If we write $\eta_P = \exp(i\varphi_P)$, etc., and $\omega_i = \exp(i\varphi_i)$, then two numbers (and their negatives) describe the general case

$$\varphi_A = \varphi_P + \varphi_C + \varphi_T,$$
$$\varphi_B = \varphi_P - \varphi_C - \varphi_T. \tag{4.20}$$

The square of $\mathcal{C}\mathcal{S}\mathcal{T}$ results in

$$\Theta_i^2\psi\Theta_i^{-2} = (\omega_i^*)^2\psi. \tag{4.21}$$

For SC multiplets the ω_i^2 are completely fixed by Eqs. (3.20) to be $(-1)^{2I+2S}$ *independent* of i:

$$\Theta_i^2\psi\Theta_i^{-2} = (-1)^{2I+2S}\psi, \quad \text{SC}$$
$$\Theta_i^2 = (-1)^{F_s+F_I}. \tag{4.22}$$

The lack of any dependence on order can be surmised by inspection of Eq. (4.13).

Equation (4.22) generalizes the usual result (see Ref. 12, for example) that Θ^2 yields $(-1)^{2S}$ when applied to sets of Hermitian fields which define a *local* field theory. The case of selfconjugate isofermions ($2I$ = odd) is not equivalent to a local linear combination of Hermitian fields (see I, or the next section). We see that Θ^2 provides a simple way to test a self-conjugate field theory for locality. The authors of Refs. 6–8 dispose of selfconjugate isofermions by showing incompatibility with a "normal" Θ transformation. On the surface it appears that the "global" test of the *field* using Θ^2 might be more general than the local-causal properties of the field *theory*. However, it must be recognized that locality (rather, weak local commutativity) enters the discussion of Θ in a critical way. Both aspects of the problem are interesting and should be studied. Besides providing a new type of nonlocal field theory for study, the SC isofermions might even exist. Although the latter possibility seems remote, one should not forget the history of "dubious" objects such as the two-component neutrino.

Now we turn to the PC field. The phase ω_i is quite general, and Eq. (4.21) indicates that the operators Θ_i^2 are gauge transformations

$$\Theta_j^2 = \exp[2i\varphi_j(N - \bar{N})], \tag{4.23}$$

where

$$\varphi_1 = -\varphi_2 = -\varphi_3 = \varphi_A,$$
$$\varphi_4 = -\varphi_5 = -\varphi_6 = \varphi_B$$

[cf. Eqs. (4.19)–(4.20)]. It is of some interest that the product of two discrete Θ operations is equivalent to a continuous gauge transformation whose phase depends on the discrete phases φ_C, φ_P, and φ_T. Only by making a special phase choice, or a physical assumption, can we make this phase definite. We stress this point because it is not always realized that the common choice[14] $\Theta^2 = (-1)^{F_s}$ for *all* particles relies upon a special phase convention.

The following remarks may clarify somewhat the nature of this convention and its relation to physical criteria. In particular, we have in mind the discussion of fermion parity by Yang and Tiomno.[26] These authors required that two reflections be equivalent to a rotation by 0 *or* 2π. If this geometrical requirement is imposed on the parity operator, we obtain the additional constraint that η_P^4 be unity. We want to show the following: If one ω_k is chosen to be real, then $\omega_i = \omega_k$ for all i, provided $\eta_P^4 = 1$. With these assumptions, the transformation $\Theta_i^2\psi\Theta_i^{-2} = (\omega_i^*)^2\psi$ becomes independent of i, that is, of the permutation of the order of $\mathcal{C}$, $\mathcal{S}$, and $\mathcal{T}$. For each spin ω_i^2 can be only ± 1, and the conventional choice $(-1)^{2S}$ is especially useful since it mirrors the (phase-independent) constituent transformation

$$\mathcal{T}^2\psi\mathcal{T}^{-2} = (-1)^{2S}\psi.$$

The proof is easy; we give one specimen by assuming that $\omega_1^2 = (-1)^{2S}$. Equation (4.19) then shows that $\omega_2^2 = \omega_3^2 = \omega_1^2$ and further that $(\eta_C^*\eta_T^*)^2 = (-1)^{2S}\eta_P^2$. The latter gives $\omega_4^2 = (-1)^{2S}\eta_P^4$. When η_P^4 is unity, $\omega_i^2 = (-1)^{2S}$ for all i. Inspection shows that the result is independent of which initial ω_κ^2 is chosen to be $(-1)^{2S}$.

It is very interesting that a converse theorem holds. If we require invariance of $\Theta_i^2\psi\Theta_i^{-2}$ under $\mathcal{C}$, $\mathcal{S}$, $\mathcal{T}$ permutations, then we can deduce the Yang–Tiomno equation $\eta_P^4 = 1$. If ω_i^2 is independent of i, Eqs. (4.19) show that all ω_i^2 are real. Secondly, $\omega_1^2 = (\omega_4^*)^2$ implies the results

$$\eta_P^4 = 1,$$
$$(\eta_C\eta_T)^4 = 1, \tag{4.24}$$
$$\omega_i^2 = \pm 1.$$

Again $\omega_i = (-1)^{2S}$ is a useful though special choice, and one can take Θ^2 to be $(-1)^{F_s}$. Since the parity of

[26] C. N. Yang and J. Tiomno, Phys. Rev. **79**, 495 (1950).

a *single* fermion is generally regarded as unmeasurable, the permutation invariance may be without physical significance.

ACKNOWLEDGMENT

The author is indebted to the Istituto di Fisica "Guglielmo Marconi" for its hospitality.

APPENDIX A: RARITA-SCHWINGER WAVEFUNCTIONS FOR HIGH SPIN

In order to give a complete discussion of the discrete symmetries we have to be able to relate the wavefunctions $\chi_\mu(p, \lambda)$ to other wavefunctions with spin or momentum reversed. We use a real metric with $g_{00} = 1 = -g_{ii}$ $(i = 1, 2, 3)$ and γ matrices

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}, \quad \gamma_0\gamma_\mu^\dagger\gamma_0 = \gamma_0,$$
$$\gamma_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma^k = \begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix}, \tag{A1}$$

where σ_k are the usual Pauli spin matrices.

The spin state of a wavefunction usually refers to a fixed direction (the famous z axis) or to the direction of momentum (helicity basis). Although we generally use the latter description, it is useful to review both for the spin-$\frac{1}{2}$ case. Let χ_s $(s = \pm\frac{1}{2})$ be the usual real two-component spinor quantized along the z axis $(\sigma_z\chi_s = 2S\chi_s)$. [We write $\chi_s(\hat{z})$ whenever a chance for confusion with helicity spinors occurs.] Then the Dirac wavefunction $u(p, s)$ for a particle of momentum p and "spin" s is

$$u(p, s) = \begin{pmatrix} \cosh \frac{1}{2}\omega\chi_s \\ \sigma \cdot \hat{p} \sinh \frac{1}{2}\omega\chi_s \end{pmatrix}, \tag{A2}$$

where the parameter ω is related to energy and momentum by $E = m \cosh \omega$, $p = m \sinh \omega$. The antiparticle wavefunction is defined by

$$v(p, s) = C_1\bar{u}^T(p, s) = (-1)^{\frac{1}{2}-s}\begin{pmatrix} \sigma \cdot \hat{p} \sinh \frac{1}{2}\omega\chi_{-s} \\ \cosh \frac{1}{2}\omega\chi_{-s} \end{pmatrix}, \tag{A3}$$

where the phase of the charge conjugation matrix $C_1 = i\gamma_0\gamma_2$ has been chosen in a natural way [$v(p, \frac{1}{2}) \to$ col. $(0, \chi_{-\frac{1}{2}})$ when $\mathbf{p} \to 0$].

Inspection reveals the inversion formulas[20]

$$u(p, s) = Pu(-p, s),$$
$$v(p, s) = -Pv(-p, s), \tag{A4}$$

where P is the matrix γ_0. To invert both p and s, we use the time-reversal matrix $T = \gamma_3\gamma_1$ (chosen real) and the formula

$$\chi_{-s} = (-1)^{\frac{1}{2}+s} i\sigma_2\chi_s \tag{A5}$$

to obtain

$$w^*(p, s) = (-1)^{\frac{1}{2}+s} Tw(-p, -s), \tag{A6}$$

where w stands for either u or v.

To discuss the analogous properties of Dirac helicity spinors, it is useful to summarize the corresponding properties of the two-component helicity spinors. The latter are defined by a rotation in the usual way, by rotation of a state with helicity λ in the $\hat{z}$ direction. Let $\hat{p}$ and the rotation axis $\hat{n}$ be given by

$$\hat{p} = (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta),$$
$$\hat{n} = (-\sin\phi, \cos\phi, 0). \tag{A7}$$

The state $\chi_\lambda(\hat{p})$ is then given by

$$\chi_\lambda(\hat{p}) = \exp(-i\hat{n} \cdot \boldsymbol{\sigma}\theta/2)\chi_\lambda(\hat{z}). \tag{A8}$$

The state $\chi_\lambda(-\hat{p})$ is conventionally reached by the rotation described by the angles $(\pi - \theta, \phi + \pi)$ to avoid any confusion due to the double-valuedness of $SU(2)$ transformations.

The following three formulas are useful (only two are independent):

$$\chi_\lambda(\hat{p}) = (-1)^{\frac{1}{2}-\lambda} e^{2i\lambda\phi}\chi_{-\lambda}(-\hat{p}),$$
$$\chi_\lambda(\hat{p}) = (-1)^{\frac{1}{2}-\lambda} i\sigma_2\chi^*_{-\lambda}(\hat{p}), \tag{A9}$$
$$\chi_\lambda(\hat{p}) = -e^{2i\lambda\phi} i\sigma_2\chi^*_\lambda(-\hat{p}).$$

The helicity spinor $u(p, \lambda)$ is given by

$$u(p, \lambda) = \exp(-i\hat{n} \cdot \boldsymbol{\sigma}\theta/2)u(p = p_z, s = \lambda)$$
$$= \begin{pmatrix} \cosh \frac{1}{2}\omega\chi_\lambda(\hat{p}) \\ 2\lambda \sinh \frac{1}{2}\omega\chi_\lambda(\hat{p}) \end{pmatrix}. \tag{A10}$$

The antiparticle helicity spinor is

$$v(p, \lambda) = C_1\bar{u}^T(p, \lambda) = (-1)^{\frac{1}{2}-\lambda}\begin{pmatrix} -2\lambda \sinh \frac{1}{2}\omega\chi_{-\lambda}(\hat{p}) \\ \cosh \frac{1}{2}\omega\chi_{-\lambda}(\hat{p}) \end{pmatrix}. \tag{A11}$$

(2λ is the same as $(-1)^{\frac{1}{2}-\lambda}$ for spin $\frac{1}{2}$.)

The formulas corresponding to Eqs. (A4) and (A6) are

$$u(p, \lambda) = (-1)^{\frac{1}{2}-\lambda} e^{2i\lambda\phi} Pu(-p, -\lambda),$$
$$v(p, \lambda) = -(-1)^{\frac{1}{2}-\lambda} e^{-2i\lambda\phi} Pv(-p, -\lambda),$$
$$u^*(p, \lambda) = e^{-2i\lambda\phi} Tu(-p, \lambda),$$
$$v^*(p, \lambda) = e^{2i\lambda\phi} Tv(-p, \lambda). \tag{A12}$$

These results are somewhat more complicated than the "old-fashioned" ones, (A4) and (A6), but are more useful because of the physical significance of the helicity states.

The vector helicity wavefunctions $e^\mu(p, \lambda)$ are obtained by rotation of the wavefunctions polarized

along the z axis:

$$e^\mu(p_z, \pm 1) = \mp(0, 1, \pm i, 0)/2^{\frac{1}{2}},$$
$$e^\mu(p_z, 0) = (p, 0, 0, E)/m. \tag{A13}$$

For arbitrary direction, one obtains by rotation

$$e^\mu(p, \pm 1) = \mp(0, 1 - (1 - \cos\theta)e^{\pm i\phi}\cos\phi, \pm i(1 \pm i(1 - \cos\theta)e^{\pm i\phi}\sin\phi), -\sin\theta e^{\pm i\phi}),$$
$$e^\mu(p, 0) = (p, E\sin\theta\cos\phi, E\sin\theta\sin\phi, E\cos\theta)/m. \tag{A14}$$

The polarization vectors e_μ obey the orthogonality–completeness relations

$$p \cdot e(p, \lambda) = 0,$$
$$e^*_\mu(p, \lambda)e^\mu(p, \lambda') = -\delta_{\lambda\lambda'}, \tag{A15}$$
$$\sum_\lambda e_\mu(p, \lambda)e^*_\nu(p, \lambda) = -g_{\mu\nu} + p_\mu p_\nu/m^2.$$

A short calculation leads to the relations

$$e_\mu(p, \lambda) = (-1)^{1-\lambda}e^{2i\lambda\phi}(-g_{\mu\mu})e_\mu(-p, -\lambda),$$
$$e^*_\mu(p, \lambda) = e^{-2i\lambda\phi}g_{\mu\mu}e_\mu(-p, \lambda),$$
$$e^*_\mu(p, \lambda) = (-1)^\lambda e_\mu(-p, -\lambda). \tag{A16}$$

[A quicker way to derive these results is to identify $e_\mu(2p, 2\lambda)$ with $\bar{v}(p, \lambda)\gamma_\mu u(p, \lambda)$ and then to use Eqs. (A11) to reach (A16).]

The Rarita–Schwinger helicity wavefunctions for higher spin are easily found by combining spin-$\frac{1}{2}$ and spin-1 helicity wavefunctions using ordinary Clebsch–Gordan coefficients.[27–29] For meson wavefunctions we have the recursion formula relating spin S and spin $S + 1$:

$$e^{s+1}_{\mu_1\mu_2\cdots\mu_{s+1}}(p, \lambda) = \sum_{\lambda_1\lambda_2} C(s1s+1; \lambda_1\lambda_2\lambda)e^s_{\mu_1\cdots\mu_s}(p, \lambda_1)e_{\mu_{s+1}}(p, \lambda_2). \tag{A17}$$

For fermions, we have the spinor-tensor ($S = k + \frac{1}{2}$)

$$u^{s+1}_{\mu_1\cdots\mu_{k+1}}(p, \lambda) = \sum_{\lambda_1\lambda_2} C(s1s+1; \lambda_1\lambda_2\lambda)u^s_{\mu_1\cdots\mu_k}(p, \lambda_1)e_{\mu_{k+1}}(p, \lambda_2), \tag{A18}$$

$$v^{s+1}_{\mu_1\cdots\mu_{k+1}}(p, \lambda) = C_1\bar{u}^{s+1\ T}_{\mu_1\cdots\mu_{k+1}}(p, \lambda) = \sum_{\lambda_1\lambda_2} C(s1s+1; \lambda_1\lambda_2\lambda)v^s_{\mu_1\cdots\mu_k}(p, \lambda_1)e^*_{\mu_{k+1}}(p, \lambda_2). \tag{A19}$$

For spin $\frac{3}{2}$, the ordinary Dirac spinors $u(p, \lambda)$ or $v(p, \lambda)$ appear on the right-hand side of Eq. (A18) or (A19).

For notational reasons we write $\chi_\mu(p, \lambda)$ for either the meson wavefunction $e^s_{\mu_1\cdots\mu_s}(p, \lambda)$ or the fermion wavefunction $u^s_{\mu_1\cdots\mu_k}(p, \lambda)$ and χ^c_μ for the generalized conjugate wavefunction $e^{s*}_{\mu_1\cdots\mu_s}$ for mesons, and $C_1\bar{u}^{sT}_{\mu_1\cdots\mu_k} = v^s_{\mu_1\cdots\mu_k}$ for fermions. The vector μ stands for the set of 4-vector indices μ_i, in which χ_μ is totally symmetric and divergenceless. [$p^{\mu_i}\chi_\mu(p, \lambda)$ vanishes; for fermions $\gamma^{\mu_i}\chi_\mu(p, \lambda)$ also vanishes.] More details on the subsidiary conditions can be found in the work of Fronsdal and Behrends.[30,31]

The orthogonality conditions are easily found by use of the recursion formulas:

$$e^{s*}_\mu(p, \lambda)e^{s\mu}(p, \lambda') = (-1)^s\delta_{\lambda\lambda'},$$
$$\bar{u}^s_\mu(p, \lambda)u^{s\mu}(p, \lambda') = (-1)^{s-\frac{1}{2}}\delta_{\lambda\lambda'}, \tag{A20}$$
$$\bar{v}^s_\mu(p, \lambda)v^{s\mu}(p, \lambda') = -(-1)^{s-\frac{1}{2}}\delta_{\lambda\lambda'}.$$

Symmetry relations involving the change of sign of p or λ follow easily from the recursion relations, using induction and Eqs. (A12) and (A16) for spin $\frac{1}{2}$:

$$e^s_\mu(p, \lambda) = (-1)^{s-\lambda}e^{2i\lambda\phi}\Pi(-g_{\mu\mu})e^s_\mu(-p, -\lambda),$$
$$e^{s*}_\mu(p, \lambda) = e^{-2i\lambda\phi}\Pi(g_{\mu\mu})e^s_\mu(-p, \lambda), \tag{A21}$$

$$u^s_\mu(p, \lambda) = (-1)^{s-\lambda}e^{2i\lambda\phi}\Pi(-g_{\mu\mu})Pu^s_\mu(-p, -\lambda),$$
$$u^{s*}_\mu(p, \lambda) = e^{-2i\lambda\phi}\Pi(g_{\mu\mu})Tu^s_\mu(-p, \lambda), \tag{A22}$$

$$v^s_\mu(p, \lambda) = -(-1)^{s-\lambda}e^{-2i\lambda\phi}\Pi(-g_{\mu\mu})Pv^s_\mu(-p, -\lambda),$$
$$v^{s*}_\mu(p, \lambda) = e^{2i\lambda\phi}\Pi(g_{\mu\mu})Tv^s_\mu(-p, \lambda). \tag{A23}$$

Again P and T are the real matrices γ_0 and $\gamma_3\gamma_1$, respectively. The products $\Pi(\pm g_{\mu\mu})$ run over S or $S - \frac{1}{2}$ factors $g_{\mu_i\mu_i}$.

All the preceding inversion formulas are summarized by

$$\chi^s_\mu(p, \lambda) = (-1)^{s-\lambda}e^{2i\lambda\phi}\Pi(-g_{\mu\mu})P\chi_\mu(-p, -\lambda),$$
$$\chi^{s*}_\mu(p, \lambda) = e^{-2i\lambda\phi}\Pi(g_{\mu\mu})T\chi_\mu(-p, \lambda),$$
$$\chi^{sc}_\mu(p, \lambda) = C_1\bar{\chi}^{sT}_\mu(p, \lambda), \tag{A24}$$

provided we replace all matrices by unity for mesons.

In order to write the field commutation relations in a neat form, it is useful to have projection operators, expressed in terms of the helicity wavefunctions. Here we mainly emphasize points not covered or stressed by Fronsdal and Behrends.[30,31] The reader is referred to the latter papers for more information, and some explicit formulas.

[27] P. Auvil and J. J. Brehm, Phys. Rev. **140**, B135 (1965); **145**, 1152 (1966).
[28] D. Brudnoy, Phys. Rev. **145**, 1229 (1966).
[29] P. Carruthers, Phys. Rev. **152**, 1345 (1966).
[30] C. Fronsdal [Nuovo Cimento Suppl. **9**, 416 (1958)] discusses these points in detail.
[31] R. E. Behrends and C. Fronsdal, Phys. Rev. **106**, 345 (1957).

For spin $\frac{1}{2}$ we have the well-known projection operators

$$\Lambda_+(p) = \sum_\lambda u(p,\lambda)\bar{u}(p,\lambda) = \frac{m + p\cdot\gamma}{2m},$$
$$\Lambda_-(p) = -\sum_\lambda v(p,\lambda)\bar{v}(p,\lambda) = \frac{m - p\cdot\gamma}{2m}. \tag{A25}$$

For spin 1, we define a second-rank tensor projection operator:

$$M_{\mu\nu}(p) = -\sum_\lambda e_\mu(p,\lambda)e_\nu^*(p,\lambda)$$
$$= g_{\mu\nu} - p_\mu p_\nu/m^2, \tag{A26}$$
$$M_{\mu\nu}(p)M_\lambda^\nu(p) = M_{\mu\lambda}(p).$$

For spin-S mesons, Eqs. (A26) generalize to

$$M^s_{\mu\nu}(p) = (-1)^s \sum_\lambda e^s_\mu(p,\lambda)e^s_\nu{}^*(p,\lambda),$$
$$M^s_{\mu\nu}(p)M^{s\nu}_\lambda(p) = M^s_{\mu\nu}. \tag{A27}$$

Fronsdal and Behrends have given an explicit formula[31] for $M^S_{\mu\nu}(p)$. For our purposes we only need to note that it is a real tensor constructed from an appropriate number of $g_{\mu\nu}$ and p_μ factors.

From Eqs. (A21) we find the symmetry relations

$$M^s_{\mu\nu}(\mathbf{p}) = \Pi(-g_{\mu\mu})\Pi(-g_{\nu\nu})M^s_{\mu\nu}(-\mathbf{p}),$$
$$M^{s*}_{\mu\nu}(\mathbf{p}) = \Pi(g_{\mu\mu})\Pi(g_{\nu\nu})M^s_{\mu\nu}(-\mathbf{p}). \tag{A28}$$

Since the number of μ_i equals the number of ν_i, the minus signs cancel in the first of (A28), proving the reality of $M^s_{\mu\nu}$:

$$M^s_{\mu\nu}(p) = M^{s*}_{\mu\nu}(p). \tag{A29}$$

In addition, the first equation indicates that, when $\mathbf{p}$ changes sign, $M^S_{\mu\nu}$ transforms as an ordinary vector in each of its indices, thereby excluding pseudotensors from the construction of $M^S_{\mu\nu}(p)$.

For baryons of spin S, we write the projection operators as

$$\mathfrak{P}^s_{+\mu\nu}(p) = (-1)^{s-\frac{1}{2}}\sum_\lambda u^s_\mu(p,\lambda)\bar{u}^s_\nu(p,\lambda),$$
$$\mathfrak{P}^s_{-\mu\nu}(p) = -(-1)^{s-\frac{1}{2}}\sum_\lambda v^s_\mu(p,\lambda)\bar{v}^s_\nu(p,\lambda). \tag{A30}$$

In addition to the obvious relations

$$\mathfrak{P}^s_{\pm\mu\nu}(p)\mathfrak{P}^{s\nu}_{\pm\lambda}(p) = \mathfrak{P}^s_{\pm\mu\lambda}(p),$$
$$\mathfrak{P}^s_{\pm\mu\nu}(p)\mathfrak{P}^{s\nu}_{\mp\lambda}(p) = 0, \tag{A31}$$
$$(p\cdot\gamma \mp m)\mathfrak{P}_\pm(p) = \mathfrak{P}_\pm(p)(p\cdot\gamma \mp m),$$

and the subsidiary conditions deriving from the wavefunctions, we have the symmetry relations ($P = \gamma_0$, $T = \gamma_3\gamma_1$)

$$P\mathfrak{P}^s_{\pm\mu\nu}(\mathbf{p})P^{-1} = \Pi(-g_{\mu\mu})\Pi(-g_{\nu\nu})\mathfrak{P}^s_{\pm\mu\nu}(-\mathbf{p}),$$
$$T\mathfrak{P}^{s*}_{\pm\mu\nu}(\mathbf{p})T^{-1} = \Pi(g_{\mu\mu})\Pi(g_{\nu\nu})\mathfrak{P}^s_{\pm\mu\nu}(-\mathbf{p}). \tag{A32}$$

Again, the left-hand sides of (A32) are equal. We also note the relations ($\gamma_5 = \gamma^0\gamma^1\gamma^2\gamma^3$)

$$i\gamma_5 u(p,\lambda) = (-1)^{\frac{1}{2}+\lambda}v(p,-\lambda),$$
$$i\gamma_5 v(p,\lambda) = (-1)^{\frac{1}{2}-\lambda}u(p,-\lambda), \tag{A33}$$
$$i\gamma_5\mathfrak{P}^s_{\pm\mu\nu}(p)(i\gamma_5)^{-1} = \mathfrak{P}^s_{\mp\mu\nu}(p).$$

The operators $\mathfrak{P}_{\pm\mu\nu}$ are to be constructed from the metric tensor γ_{μ_i} and p_{μ_i}. The first equation of (A32) excludes pseudoquantities like γ_5, and the second makes the coefficients of the independent tensors real. One can then write the projection operators in the form

$$\mathfrak{P}^s_{\pm\mu\nu}(p) = \Lambda_{\mu\nu}(p)\Lambda_\pm(p) = \Lambda_\pm(p)\Lambda_{\mu\nu}(p), \tag{A34}$$

where the $\Lambda_\pm(p)$ are given in Eq. (A25). The quantities $\Lambda_{\mu\nu}(p)$ are then determined by the subsidiary conditions.

APPENDIX B. SPIN-$\frac{1}{2}$ PARTICLE OPERATORS

The Dirac field can be expanded in terms of wavefunctions and operators referring to a fixed axis or in terms of helicity wavefunctions and operators. We have the relations

$$\sum_s a(p,s)u(p,s) = \sum_\lambda a(p,\lambda)u(p,\lambda),$$
$$\sum_s b^*(p,s)v(p,s) = \sum_\lambda b^*(p,\lambda)v(p,\lambda), \tag{B1}$$

where the wavefunctions were given in Eqs. (A2) and (A3), and (A10) and (A11). From the orthogonality relations we find

$$a(p,\lambda) = \sum_s M_{\lambda s}a(p,s),$$
$$b^*(p,\lambda) = \sum_s R^*_{\lambda s}b^*(p,s), \tag{B2}$$

where a little calculation gives

$$M_{\lambda s} \equiv \bar{u}(p,\lambda)u(p,s) = \chi^\dagger_\lambda(\hat{p})\chi_s(\hat{z}),$$
$$M_{\lambda s} = \chi^\dagger_\lambda(\hat{z})e^{i\hat{n}\cdot\sigma\theta/2}\chi_s(\hat{z}), \tag{B3}$$
$$R^*_{\lambda s} \equiv (-1)^{\lambda-s}\bar{v}(p,\lambda)v(p,s) = (-1)^{\lambda-s}\chi^\dagger_{-\lambda}(\hat{p})\chi_{-s}(\hat{z}),$$
$$R^*_{\lambda s} = (\chi^\dagger_\lambda(\hat{p}),\chi_s(\hat{z}))^* = M^*_{\lambda s}. \tag{B4}$$

We have used the second of Eqs. (A9) to establish that $R = M$, i.e., that the particle and antiparticle helicity operators are related in exactly the same way to the fixed axis operators. The matrix $M_{\lambda s}$ is given by

$$M = e^{i\hat{n}\cdot\sigma\theta/2} = \begin{pmatrix} \cos\frac{\theta}{2} & e^{-i\phi}\sin\frac{\theta}{2} \\ -e^{i\phi}\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}. \tag{B5}$$

Thus the transformations (B2) are, not surprisingly, unitary and unimodular.

Using the notation $+$, $-$ for $(\lambda, s) = +\frac{1}{2}$ or $-\frac{1}{2}$ and introducing subscripts h and s to distinguish helicity from z-axis operators, we have

$$a_h(p, +) = \cos\frac{\theta}{2}\, a_s(p, +) + e^{-i\phi}\sin\frac{\theta}{2}\, a_s(p, -),$$

$$a_h(p, -) = -e^{i\phi}\sin\frac{\theta}{2}\, a_s(p, +) + \cos\frac{\theta}{2}\, a_s(p, -). \quad \text{(B6)}$$

Corresponding expressions hold for the antiparticle operators. For $-p$, we obtain the appropriate relations by the substitutions $\theta \to \pi - \theta$, $\phi \to \pi + \phi$. These are useful in discussing the parity transformation. Using (B6), we can easily derive the $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{T}$ transformations of the helicity operators from those of the z axis operators $a(p, s)$ and $b(p, s)$. Of course, an equivalent result can be obtained by requiring that the field has a well-defined transformation under these operations, once the inversion properties (A12) are known.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Variational Approach to Multiple Scattering

A. D. LEVINE AND C. F. HAYES
Physics Department, West Virginia University, Morgantown, West Virginia

(Received 9 January 1968)

Using a variation technique, two coupled equations are derived whose transition matrix is shown to describe multiple scattering. The resulting index-of-refraction equation differs from that obtained in previous theories by our inclusion of the effect of the scatterer on the target and certain constants of approximation.

1. INTRODUCTION

When a particle or wave comes upon a target involving many obstacles and is scattered by the obstacles one after the other, we say multiple scattering has occurred. The importance of a satisfactory way of dealing with multiple scattering is obvious since experimentally one encounters scattering which is multiple, not single. This is the case whether the target is composed of atoms in a crystal or nucleons in a nucleus.

Theories have been presented for multiple scattering by Bothe,[1] Williams,[2] Goudsmit and Saunderson,[3] Moliere,[4] Synder and Scott,[5] Lax,[6] Watson,[7] Fano,[8] Ter-Mikayelian,[9] Muhlschegel and Koppe,[10] Weinberg,[11] Gnedin and Dolginov,[12] and Kalashnikov and Ryazanov,[13] to name only a few. The general procedure has been to use the Boltzmann transport equation or a geometrical-optics approach for high energies, perturbation techniques for low energies, and a self-consistent field method for energies of neither extreme. Some of the more recent work in Russia makes use of the kinetic equation of Migdal.[14]

The approach we propose starts from a variation principle. Assuming high energies and neglecting exchange terms, we proceed until results are obtained which agree closely with those of Lax and Watson.

2. STATEMENT OF THE PROBLEM

We expect the Hamiltonian for a particle incident on a group of N scatterers to be

$$H = \iiint d^3\mathbf{r}\, d^3\mathbf{R}\, d^3\mathbf{R}'\, \psi^+(\mathbf{r})\phi^+(\mathbf{R})\phi^+(\mathbf{R}') \times [(\mathbf{p}^2/2mN^2) + (\mathbf{P}^2/2MN) + \{G(\mathbf{R}-\mathbf{R}')/2\} + \{V(\mathbf{r}-\mathbf{R})/N\}]\phi(\mathbf{R}')\phi(\mathbf{R})\psi(\mathbf{r}), \quad (1)$$

where ψ, $\mathbf{p}$, $\mathbf{m}$, and $\mathbf{r}$ refer to the incident particle, and ϕ, $\mathbf{P}$, $\mathbf{M}$, and $\mathbf{R}$ refer to the scatterers. $G(\mathbf{R}-\mathbf{R}')$ is the interaction between scatterers, and $V(\mathbf{r}-\mathbf{R})$ is the interaction between the incident particle and the

[1] W. Bothe, Z. Physik **54**, 161 (1929).
[2] E. J. Williams, Phys. Rev. **58**, 292 (1940).
[3] S. Goudsmit and J. L. Saunderson, Phys. Rev. **58**, 36 (1940).
[4] G. Moliere, Z. Naturforsch. **3a**, 78 (1948).
[5] H. S. Snyder and W. T. Scott, Phys. Rev. **76**, 220 (1949).
[6] M. Lax, Rev. Mod. Phys. **23**, 287 (1951).
[7] K. M. Watson, Phys. Rev. **89**, 575; **92**, 291 (1953).
[8] U. Fano, Phys. Rev. **93**, 117 (1954).
[9] M. L. Ter-Mikayelian, Nucl. Phys. **9**, 679 (1959).
[10] B. Muhlschegel and H. Koppe, Z. Physik **150**, 474 (1958).
[11] S. Weinberg, Phys. Rev. **133**, B232 (1964).
[12] Yu. N. Gnedin and A. Z. Dolginov, Zh. Eksp. Teor. Fiz. **45**, 1136 (1963) [Sov. Phys.—JETP **18**, 784 (1964)].
[13] N. P. Kalashnikov and M. I. Ryazanov, Zh. Eksp. Teor. Fiz. **50**, 117 (1966) [Sov. Phys.—JETP **23**, 79 (1966)].
[14] A. B. Migdal and N. M. Polievktov-Nikoladze, Dokl. Akad. Nauk SSSR **96**, 49 (1954).

Thus the transformations (B2) are, not surprisingly, unitary and unimodular.

Using the notation $+$, $-$ for $(\lambda, s) = +\frac{1}{2}$ or $-\frac{1}{2}$ and introducing subscripts h and s to distinguish helicity from z-axis operators, we have

$$a_h(p, +) = \cos\frac{\theta}{2}\, a_s(p, +) + e^{-i\phi}\sin\frac{\theta}{2}\, a_s(p, -),$$

$$a_h(p, -) = -e^{i\phi}\sin\frac{\theta}{2}\, a_s(p, +) + \cos\frac{\theta}{2}\, a_s(p, -). \quad \text{(B6)}$$

Corresponding expressions hold for the antiparticle operators. For $-p$, we obtain the appropriate relations by the substitutions $\theta \to \pi - \theta$, $\phi \to \pi + \phi$. These are useful in discussing the parity transformation. Using (B6), we can easily derive the $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{T}$ transformations of the helicity operators from those of the z axis operators $a(p, s)$ and $b(p, s)$. Of course, an equivalent result can be obtained by requiring that the field has a well-defined transformation under these operations, once the inversion properties (A12) are known.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Variational Approach to Multiple Scattering

A. D. LEVINE AND C. F. HAYES
Physics Department, West Virginia University, Morgantown, West Virginia

(Received 9 January 1968)

Using a variation technique, two coupled equations are derived whose transition matrix is shown to describe multiple scattering. The resulting index-of-refraction equation differs from that obtained in previous theories by our inclusion of the effect of the scatterer on the target and certain constants of approximation.

1. INTRODUCTION

When a particle or wave comes upon a target involving many obstacles and is scattered by the obstacles one after the other, we say multiple scattering has occurred. The importance of a satisfactory way of dealing with multiple scattering is obvious since experimentally one encounters scattering which is multiple, not single. This is the case whether the target is composed of atoms in a crystal or nucleons in a nucleus.

Theories have been presented for multiple scattering by Bothe,[1] Williams,[2] Goudsmit and Saunderson,[3] Moliere,[4] Synder and Scott,[5] Lax,[6] Watson,[7] Fano,[8] Ter-Mikayelian,[9] Muhlschegel and Koppe,[10] Weinberg,[11] Gnedin and Dolginov,[12] and Kalashnikov and Ryazanov,[13] to name only a few. The general procedure has been to use the Boltzmann transport equation or a geometrical-optics approach for high energies, perturbation techniques for low energies, and a self-consistent field method for energies of neither extreme. Some of the more recent work in Russia makes use of the kinetic equation of Migdal.[14]

The approach we propose starts from a variation principle. Assuming high energies and neglecting exchange terms, we proceed until results are obtained which agree closely with those of Lax and Watson.

2. STATEMENT OF THE PROBLEM

We expect the Hamiltonian for a particle incident on a group of N scatterers to be

$$H = \iiint d^3\mathbf{r}\, d^3\mathbf{R}\, d^3\mathbf{R}'\, \psi^+(\mathbf{r})\phi^+(\mathbf{R})\phi^+(\mathbf{R}') \times [(\mathbf{p}^2/2mN^2) + (\mathbf{P}^2/2MN) + \{G(\mathbf{R}-\mathbf{R}')/2\} + \{V(\mathbf{r}-\mathbf{R})/N\}]\phi(\mathbf{R}')\phi(\mathbf{R})\psi(\mathbf{r}), \quad (1)$$

where ψ, $\mathbf{p}$, $\mathbf{m}$, and $\mathbf{r}$ refer to the incident particle, and ϕ, $\mathbf{P}$, $\mathbf{M}$, and $\mathbf{R}$ refer to the scatterers. $G(\mathbf{R}-\mathbf{R}')$ is the interaction between scatterers, and $V(\mathbf{r}-\mathbf{R})$ is the interaction between the incident particle and the

[1] W. Bothe, Z. Physik **54**, 161 (1929).
[2] E. J. Williams, Phys. Rev. **58**, 292 (1940).
[3] S. Goudsmit and J. L. Saunderson, Phys. Rev. **58**, 36 (1940).
[4] G. Moliere, Z. Naturforsch. **3a**, 78 (1948).
[5] H. S. Snyder and W. T. Scott, Phys. Rev. **76**, 220 (1949).
[6] M. Lax, Rev. Mod. Phys. **23**, 287 (1951).
[7] K. M. Watson, Phys. Rev. **89**, 575; **92**, 291 (1953).
[8] U. Fano, Phys. Rev. **93**, 117 (1954).
[9] M. L. Ter-Mikayelian, Nucl. Phys. **9**, 679 (1959).
[10] B. Muhlschegel and H. Koppe, Z. Physik **150**, 474 (1958).
[11] S. Weinberg, Phys. Rev. **133**, B232 (1964).
[12] Yu. N. Gnedin and A. Z. Dolginov, Zh. Eksp. Teor. Fiz. **45**, 1136 (1963) [Sov. Phys.—JETP **18**, 784 (1964)].
[13] N. P. Kalashnikov and M. I. Ryazanov, Zh. Eksp. Teor. Fiz. **50**, 117 (1966) [Sov. Phys.—JETP **23**, 79 (1966)].
[14] A. B. Migdal and N. M. Polievktov-Nikoladze, Dokl. Akad. Nauk SSSR **96**, 49 (1954).

target. We take

$$\psi = \sum_{\mathbf{k}} b_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}), \tag{2}$$

$$\phi = \sum_{\mathbf{k}} a_{\mathbf{k}}\phi_{\mathbf{k}}(\mathbf{R}), \tag{3}$$

where a_k^+ and a_k are the creation and annihilation operators for the scatters, assumed to be fermions, and b_k^+ and b_k are for the incident particle, which may be a fermion or a boson. We have taken

$$\int \psi^+\psi\, d^3\mathbf{r} = 1, \quad \int \phi^+\phi\, d^3\mathbf{R} = N. \tag{4}$$

If we assume the variation

$$\delta(H - E)\boldsymbol{\Phi} = 0, \tag{5}$$

where $\boldsymbol{\Phi}$ is a vector spanning the Hilbert space of both the incident particle and target, we obtain six equations, one from the variation of each functional. Each equation, however, may be reduced to one of the following two equations:

$$\left[(\mathbf{p}^2/2m) + H_L + \sum_{\mathbf{k}\mathbf{k}'} a_{\mathbf{k}'}^+ a_{\mathbf{k}} V_{\mathbf{k}'\mathbf{k}}\right]\psi\boldsymbol{\Phi} = E\psi\boldsymbol{\Phi}, \tag{6}$$

$$\Big\{H_p + (N\mathbf{P}^2/2M) + \int d^3\mathbf{R}'\,\phi^+(\mathbf{R})G(\mathbf{R} - \mathbf{R}')\phi(\mathbf{R}') + N\sum_{\mathbf{k}\mathbf{k}'} b_{\mathbf{k}'}^+ b_{\mathbf{k}} V_{\mathbf{k}'\mathbf{k}}\Big\}\boldsymbol{\phi}\boldsymbol{\Phi} = \phi\boldsymbol{\Phi}, \tag{7}$$

where

$$H_L = \int d^3\mathbf{R}\phi^+(\mathbf{R})(\mathbf{P}^2/2M)\phi(\mathbf{R}) + \int d^3\mathbf{R}\, d^3\mathbf{R}'\,\phi^+(\mathbf{R})\phi^+(\mathbf{R}') \times (\{G(\mathbf{R} - \mathbf{R}')\}/2)\phi(\mathbf{R}')\phi(\mathbf{R}), \tag{8}$$

$$H_p = \int d^3\mathbf{r}\psi^+(\mathbf{r})(\mathbf{p}^2/2m)\psi(\mathbf{r}), \tag{9}$$

$$V_{\mathbf{k}'\mathbf{k}} = \int d^3\mathbf{R}\phi_{\mathbf{k}'}^+(\mathbf{R})V(\mathbf{r} - \mathbf{R})\phi_{\mathbf{k}}(\mathbf{R}), \tag{10}$$

$$V_{\mathbf{k}'\mathbf{k}} = \int d^3\mathbf{r}\psi_{\mathbf{k}}^+(\mathbf{r})V(\mathbf{r} - \mathbf{R})\psi_{\mathbf{k}}(\mathbf{r}). \tag{11}$$

The coupled equations (6) and (7) form the basis of this approach to multiple scattering. If the initial wavefunction of the scatterers is known, H_L and $V_{\mathbf{K}'\mathbf{K}}$ may be calculated (and hence $\psi_{\mathbf{k}}$) from the solution of (6). With this value of $\psi_{\mathbf{k}}$ one calculates H_p and $V_{\mathbf{k}'\mathbf{k}}$ and hence $\phi_{\mathbf{k}}$ from the solution of (7). This procedure is continued until self-consistent solutions are obtained.

Letting

$$v = \sum_{\mathbf{k}'\mathbf{k}} a_{\mathbf{k}'}^+ a_{\mathbf{k}} V_{\mathbf{k}'\mathbf{k}} \tag{12}$$

and

$$H_0 = (\mathbf{p}^2/2m) + H_L, \tag{13}$$

from Eq. (6) we have

$$H_0\psi_0\boldsymbol{\Phi} = E\psi_0\boldsymbol{\Phi},$$

where the total energy is equal to the sum of the initial energy of the particle E_p and the initial energy of the target E_L:

$$E = E_L + E_p. \tag{14}$$

We define k from

$$(\hbar^2k^2/2m) = E_p, \tag{15}$$

and since, in general, we know

$$H_L\boldsymbol{\Phi} = E_L\boldsymbol{\Phi}, \tag{16}$$

we further define q from

$$H_L\boldsymbol{\Phi} = \{E - (\hbar^2q^2/2m)\}\boldsymbol{\Phi}. \tag{17}$$

Note that k and q are equal when the target and scattered particle are infinitely removed from each other. We choose the same direction for k and q.

Consequently, we have the integral equation

$$\psi^{\pm}(\mathbf{r})\boldsymbol{\Phi} = \psi_0(\mathbf{r})\boldsymbol{\Phi} + \int d^3q'(q^2 - q'^2 \pm i\epsilon)^{-1}\psi_{0\mathbf{q}}(\mathbf{r}) \times \int d^3\mathbf{r}'\psi_{0\mathbf{q}'}^*(\mathbf{r}')\{2mv(\mathbf{r}')/\hbar^2\}\psi^{\pm}(\mathbf{r}')\boldsymbol{\Phi}. \tag{18}$$

Adding and subtracting H_L to the denominator, we represent Eq. (18) by

$$\psi^{(\pm)}(\mathbf{r})\boldsymbol{\Phi} = \psi_0(\mathbf{r})\boldsymbol{\Phi} + (E - H_0 \pm i\epsilon)^{-1}v(\mathbf{r})\psi^{(\pm)}(\mathbf{r})\boldsymbol{\Phi}. \tag{19}$$

3. THE TRANSITION MATRIX AND INDEX OF REFRACTION

The transition matrix is

$$T_{\mathbf{k}'\mathbf{k}} = (\boldsymbol{\Phi}\psi_{0\mathbf{k}'}; v\psi_{\mathbf{k}}\boldsymbol{\Phi}). \tag{20}$$

We choose diagonal elements

$$T_{\mathbf{k}\mathbf{k}} = \sum_{\mathbf{k}} t'_{\mathbf{k}\mathbf{k}\mathbf{K}}, \tag{21}$$

where

$$t'_{\mathbf{k}\mathbf{k}\mathbf{K}} = (\boldsymbol{\Phi}\psi_{0\mathbf{k}}\ a_{\mathbf{K}}^+ a_{\mathbf{K}} V_{\mathbf{K}\mathbf{K}}\psi_{\mathbf{k}}\boldsymbol{\Phi}). \tag{22}$$

We denote $\psi_{\mathbf{k}}$ as the solution of (6) when only one energy level of scatterers are present:

$$[(\mathbf{p}^2/2m) + H_L + a_{\mathbf{K}}^+ a_{\mathbf{K}} V_{\mathbf{K}\mathbf{K}}]\psi_{\mathbf{K}}\boldsymbol{\Phi} = E\psi_{\mathbf{k}}\boldsymbol{\Phi}. \tag{23}$$

We now define a transition matrix

$$t_{\mathbf{k}\mathbf{K}\mathbf{K}} = (\boldsymbol{\Phi}\psi_{0\mathbf{K}}; a_{\mathbf{K}}^+ a_{\mathbf{K}} V_{\mathbf{K}\mathbf{K}}\psi_{\mathbf{k}\mathbf{K}}\boldsymbol{\Phi}). \tag{24}$$

Using the technique of Watson,[7] it is easy to show

that

$$T_{\mathbf{kk}} = \sum_{\mathbf{K}_1} t_{\mathbf{kkK}_1} + \sum_{\mathbf{K}_1 \neq \mathbf{K}_2} t_{\mathbf{kkK}_1}\alpha^{-1}t_{\mathbf{kkK}_2} + \cdots + \sum_{\mathbf{K}_1 \neq \mathbf{K}_2 \cdots \neq \mathbf{K}_n} t_{\mathbf{kkK}_1}\alpha^{-1}t_{\mathbf{kkK}_2}\cdots\alpha^{-1}t_{\mathbf{kkK}_n}, \quad (25)$$

where

$$\alpha = E - H_0 + i\epsilon. \quad (26)$$

Therefore, we see T is indeed the transition matrix for multiple scattering. The first sum in Eq. (25) is obviously scattering from each level by itself. The second term depicts the scattering-off of one level followed by a second scattering from another level. The next term depicts triple scattering. This continues until all possible combinations are accounted for. Since $\mathbf{K}_1 \neq \mathbf{K}_2 \neq \mathbf{K}_3 \cdots$, we see no level can scatter the particle again without being scattered in the meantime by another level. It was for this reason we stipulated in the beginning that the scatterers be fermions and, hence, that there only be one scatterer per energy level.

The Møller operator may be written:

$$\Omega = 1 + \alpha^{-1}\sum_{\mathbf{K}_1} t_{\mathbf{K}_1}\Omega(\mathbf{K}_1), \quad (27)$$

$$\Omega(\mathbf{K}_1) = 1 + \alpha^{-1}\sum_{\mathbf{K}_2 \neq \mathbf{K}_1} t_{\mathbf{K}_2}\Omega(\mathbf{K}_2), \quad (28)$$

where $\Omega(\mathbf{K}_1)$ is the Møller operator for a group of scatterers with the K_1 level removed. Therefore we define C_1 by

$$\Omega(\mathbf{K}_1) = C_1\Omega \quad (29)$$

and note that the more scatterers there are, the closer C_1 approaches one. If we impose the impulse approximation

$$t_{\mathbf{kkK}} = C_2 t'_{\mathbf{kkK}}, \quad (30)$$

Eq. (28) becomes

$$\Omega = 1 + \alpha^{-1}C_1C_2T_{\mathbf{kk}}\Omega. \quad (31)$$

Clearly, C_1 and C_2 may be estimated through the usual Neumann expansion. Further, if we renormalize for a possible shift $\Delta_{\mathbf{k}}$,

$$\Delta_{\mathbf{k}} = E - E', \quad (32)$$

where

$$H_0\psi_0\phi = E'\psi_0\phi, \quad (33)$$

$$(H_0 + v)\psi\phi = E\psi\phi, \quad (34)$$

we have

$$\psi_{\mathbf{k}}^R = Z_p^{-(\frac{1}{2})}\psi_{\mathbf{k}}, \quad (35)$$

$$\phi_{\mathbf{K}}^R = Z_L^{-(\frac{1}{2})}\phi_{\mathbf{K}}. \quad (36)$$

As an estimate, we calculate that

$$Z_p = |\{1 + [E - (H_0 + \Delta) + i\epsilon]^{-1}(v - \Delta) + [E - (H_0 + \Delta) + i\epsilon]^{-1} \times (v - \Delta)[E - (H_0 + \Delta) + i\epsilon]^{-1} + \cdots\}\psi_0\phi|^2, \quad (37)$$

where

$$\Delta = \sum_{\mathbf{k}'} |\psi_{0\mathbf{k}'}\rangle \Delta_{\mathbf{k}'} \langle\psi_{0\mathbf{k}'}|. \quad (38)$$

A similar expression holds for Z_L. Thus, Eq. (32) becomes

$$Z_p^{(\frac{1}{2})}\Omega^R = 1 + [E - H + i\epsilon]^{-1}C_1C_2Z_pZ_LT^RD^R, \quad (39)$$

where

$$\psi^R = \Omega^R\psi$$

and

$$T^R = v^R\Omega^R.$$

From Eq. (39) Schrödinger's equation is

$$[(\mathbf{p}^2/2m) + H_L - E + C_1C_2Z_p^{(\frac{1}{2})}Z_LT^R]\psi^R\phi = 0. \quad (40)$$

Since

$$q^2 = k^2 + (2m/\hbar^2)(E_L - H_L), \quad (41)$$

we have the index of refraction from

$$k'^2 = k^2 + (2m/\hbar^2)(EL - H_L) - 2mC_1C_2Z_p^{(\frac{1}{2})}Z_LT^R. \quad (42)$$

Comparing this result with Lax's,

$$k'^2 = k^2 - (2mC/\hbar^2)\bar{T}_{\alpha\alpha}, \quad (43)$$

we find his C is analogous to our C_1; of course, since he does not consider the effect of the scattered particle on the target, he does not have the term $(2m/\hbar^2)(E_L - H_L)$. Comparing with Watson, we see that he has assumed the equivalent of our "C_1 and C_2 equal to one" for high energies.

CONCLUSION

Starting with a system of N identical fermions acting as scatterers, we have examined a variational approach to multiple scattering. Neglecting spin effects because of an assumed high energy of the incident particle, two coupled equations were found which give a transition matrix that describes multiple scattering. The resulting index-of-refraction equation parallels those of Lax and Watson. We assume that the theory may be applied to high-energy scattering of any system of fermions where the initial wavefunction of the scatterers is known [as was discussed in connection with Eqs. (6) and (7)]. Clearly the next step in the extension of this approach is to incorporate spin and exchange effects.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Narrow Resonance Saturation of Sum Rules: A Nonuniqueness Theorem

D. ATKINSON*
Istituto di Fisica dell'Università, Roma

(Received 30 January 1968)

It is shown that, if the superconvergence relations appertaining to a two-particle scattering process are saturated by an infinite tower of resonances of mass m_J, where J is the spin, then m_J must increase less quickly than J in all cases. With this observation, it is shown that the complex of all the superconvergence relations for all the processes $a(J_1) + a(J_2) \to a(J_3) + a(J_4)$, where $a(J)$ is the spin-J member of the same tower of particles, possesses an infinity of solutions for the coupling constants. It is concluded that the superconvergence relations are incomplete, and need to be supplemented by some new physical principle, if they are to be of any practical use. A possible exception to this rule is when all but a finite number of residues are constrained to be positive, with no negative coefficient from the isospin crossing matrices. In this case, a given solution may be unique if m_J increases not more quickly than $J^{\frac{1}{2}}$.

1. INTRODUCTION

The use of superconvergent sum rules for two-particle scattering amplitudes has been attended by a certain number of practical successes.[1] The superconvergence relations are generally evaluated by inserting only the resonance contributions, and these often in the "δ function approximation." The advantage is that simple relations between coupling constants can be obtained and these sometimes agree with experiment.[2] However, there are cases where disagreement has been found[3] and the reason for this may be sought at different levels. The simplest possibility is that an insufficient number of states has been added to the sum rule to "saturate" it; the hope may be entertained that the disagreement will disappear if a few more states are included. Indeed, a given superconvergence relation can always be saved by adding a few high-mass states whose couplings to the external particles are unknown experimentally.

On the other hand, for a given superconvergent system, there exists in general an infinite number of sum rules; and the addition of higher states should be such as to be at least approximately consistent with this infinity of equations. Some interest has been evinced lately in such infinite systems with, correspondingly, an infinite number of saturating states.[4] Interest may center on the algebraic structure of any solution,[5] if such exists, or may concentrate rather upon finding a particular solution.[4] In this latter case, the question of uniqueness is important. It has recently been shown[6] that, if the saturating states have a mass spectrum that increases less rapidly than the spin, then the superconvergence relations for spinless external particles or for pseudoscalar-vector scattering yield equations between the couplings that have an infinite number of solutions. This suggests that the superconvergence relations, at least in the one-particle δ-function approximation, may be empty, since the class of solutions is very wide.

The hope remained that, when all the superconvergence relations corresponding to all possible external particles are taken into account, there may be a unique solution, or at least a small number of solutions. The purpose of this paper is to investigate this complete system of superconvergence relations; the result is to dash the hope: the infinite set of superconvergence relations still has an infinity of solutions for the couplings in most cases.

In greater detail, the model examined in this paper is that of an infinite tower of particles characterized by a mass spectrum m_J that increases indefinitely with the spin J. Let a_1, a_2, a_3, a_4 be the spin s_1, s_2, s_3, s_4 members of this tower. Then the superconvergence relations for the scattering $a_1 + a_2 \to a_3 + a_4$ are considered in the approximation of saturation by the same infinite tower of resonances. All the relations, corresponding to all possible external particles a_1, a_2, a_3, a_4, are considered together as one complicated set of simultaneous quadratic equations for the couplings.

In Sec. 2, the detailed structure of the superconvergence relations for general external spins is given. These relations are saturated by the tower of

* Present address: Rutherford High Energy Laboratory, Didcot, Berks., England.

[1] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962).

[2] S. Fubini, G. Furlan, and C. Rossetti, Nuovo Cimento **40A**, 1171 (1965).

[3] M. B. Halpern, Phys. Rev. **160**, 1441 (1967).

[4] K. Bardakci, M. B. Halpern, and G. Segré, "Perturbative Approach to Superconvergence Relations and Current Algebra," Berkeley Preprint, 1967; M. Gell-Mann, D. Horn, and J. Weyers, "Representation of Current Algebra at Infinite Momentum: Solution of the Case in which Isotopic Spin Factors Out," California Institute of Technology Preprint, 1967.

[5] K. Bardakci and G. Segré, Phys. Rev. **159**, 1263 (1967).

[6] D. Atkinson and M. B. Halpern, "Construction of Solutions to Superconvergence Relations," Berkeley Preprint, 1967.

resonances—and it is shown that this leads to a set of equations for the coupling constants, each equation involving an infinite series which must sum to zero identically for a continuum of values of the momentum transfer. However, it is a physical requirement that the superconvergence relation should not hold for arbitrarily large (spacelike) momentum transfers, and this is interpreted as meaning that the infinite series over coupling constants must diverge for sufficiently large momentum transfers. However, this can only happen if the mass spectrum m_J increases less quickly than J, as is shown in Appendix E. Under these conditions, it is shown in Sec. 3 that for given external spins, an infinite number of solutions can be found for the couplings, in general. The proof relies on a theorem of Pólya, which is given in Appendix A. In Sec. 4, it is shown that, because of the extreme nonuniqueness of the solution for given external spins, it is possible to order the infinite set of sets of equations, corresponding to all the different external spins possible, in such a way that one can find solutions of each set of equations which are consistent with all the other set of equations.

The proof up to this point takes no account of any constraint of positivity which may occur, since, in some equations, some couplings occur as their squares, which must be positive or zero. However, there are cases where these squares of coupling constants are multiplied by positive isospin crossing-matrix elements, and where they always occur in conjunction with other squares of coupling constants, which are multiplied by negative crossing-matrix elements, in such a way that the whole term may be positive or negative. To these cases the proof applies without modification. In Sec. 5, it is shown that if there *are* sign restrictions on some terms, but in such a way that there should be an infinite number of positive and an infinite number of negative terms, then if there exists one solution of the system, there exists an infinite number of solutions. In Sec. 6, it is shown that this weakened theorem can even be extended to the case where *every* term is required to be positive, but in this case only if m_J increases more quickly than $J^{\frac{1}{2}}$.

If the members of the tower are imagined to lie on a Regge trajectory $\alpha(t)$, then the results may be rephrased: In order that the superconvergence relation should not hold for arbitrary momentum transfer, $\alpha(t)$ must increase more rapidly than $t^{\frac{1}{2}}$. If $\alpha(t)$ increases more rapidly than $t^{\frac{1}{2}}$ but less rapidly than t, then it has been shown in all cases that any solution of the complex of superconvergence relations cannot be unique. Except in the case of all positive terms, this restriction to a trajectory that increases less quickly than linearly need not be made.

It would appear, then, that attempts at finding global solutions to superconvergence relations must be abandoned, unless these can be supplemented by new physical requirements that pick out one from the infinity of solutions. However, in this latter case, it would seem that the brunt of the restriction must fall on these new requirements, since the superconvergence relations are permissive to the point of vacuity. Indeed, one can adopt the view that, in passing from a crossing-symmetric dispersion relation to a superconvergence relation, one has completely lost contact with the physical solution.

A possible exception to these negative considerations is provided if one takes seriously the failure of the proof in the case that $\alpha(t)$ increases linearly (or faster) in t, and all the residues in some superconvergence relation are constrained to be positive. In this domain the possibility of a unique solution remains open. Indeed, the fact that a linear Regge trajectory is picked out as the slowest rate of increase that is compatible with the possibility of a unique solution is, perhaps, of some interest.

2. SUPERCONVERGENCE RELATION FOR GENERAL SPINS

It is the purpose of this section to set out the superconvergence relations that exist for a general scattering $a_1 + a_2 \to a_3 + a_4$, where the particle a_p has spin s_p, $p = 1, 2, 3, 4$.[7] For definiteness, the particles may be supposed to lie on some Regge trajectory, although this is not necessary to the proof. It will be convenient to consider the ensemble of all possible two-particles to two-particle scattering within this Regge family. Let

$$f(s, t; \lambda_3, \lambda_4; \lambda_1, \lambda_2) \equiv f(s, t; \lambda) \tag{2.1}$$

be a helicity amplitude for the reaction

$$a_1 + a_2 \to a_3 + a_4, \tag{2.2}$$

where s is the invariant square of the energy, t is the invariant square of the momentum transfer, and λ_p, $p = 1, 2, 3, 4$ are the helicities. For a fixed value of s, the amplitude (2.1) has, in general, certain zeros and singularities, in the complex t plane, of a purely kinematical nature, which may be removed by dividing by the factor

$$[\cos(\theta_s/2)]^{|\lambda_i+\lambda_f|} \cdot [\sin(\theta_s/2)]^{|\lambda_i-\lambda_f|}, \tag{2.3}$$

[7] T. L. Trueman and G. C. Wick, Ann Phys. (N.Y.) **26**, 322 (1964); T. L. Trueman, Phys. Rev. Letters **17**, 1198 (1966); L. L. Wang, Phys. Rev. **142**, 1187 (1966).

where

$$\lambda_i = \lambda_1 - \lambda_2 \quad \text{and} \quad \lambda_f = \lambda_3 - \lambda_4, \tag{2.4}$$

and where

$$\cos\theta_s = \frac{2st + s^2 - sM^2 + (m_1^2 - m_2^2)(m_3^2 - m_4^2)}{\{[s-(m_1+m_2)^2][s-(m_1-m_2)^2][s-(m_3+m_4)^2][s-(m_3-m_4)^2]\}^{\frac{1}{2}}}. \tag{2.5}$$

Here m_p is the mass of particle a_p, $p = 1, 2, 3, 4$, and

$$M^2 = m_1^2 + m_2^2 + m_3^2 + m_4^2. \tag{2.6}$$

In the unphysical region of reaction (2.2) $s \approx 0$, $t \to \infty$, one expects the asymptotic behavior

$$f(s, t; \lambda) \sim t^{\alpha(s)}, \tag{2.7}$$

where $\alpha(s)$ is identified as the location of the right-most singularity in the angular momentum plane associated with the partial-wave helicity amplitude. Now the kinematical factor (2.3) has the asymptotic behavior $(t \to \infty)$

$$t^{|\lambda_i+\lambda_f|/2} \cdot t^{|\lambda_i-\lambda_f|/2}. \tag{2.8}$$

This may be written t^σ, where

$$\sigma \equiv \max\,[|\lambda_1 - \lambda_2|, |\lambda_3 - \lambda_4|]. \tag{2.9}$$

Collecting these results, one sees that the following amplitude is free from kinematical singularities:

$$[\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|} \cdot [\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|} \cdot f(s, t; \lambda), \tag{2.10}$$

and that it has the asymptotic behavior $t^{\alpha(s)-\sigma}$ for fixed s and $t \to \infty$. Particular interest attaches to the cases where

$$\sigma > \alpha(s) + 1 \tag{2.11}$$

for the same range of s values including $s = 0$. By the Froissart theorem, one has $\alpha(0) \leq 1$, and in some cases, where no particle with the quantum numbers of the channel (2.2) exists, it may be the case that $\alpha(0)$ is much less than unity. However, in such a case the leading singularity may not be a pole, but a many-particle angular-momentum branch point. Nevertheless, in a reaction involving high spins, either in the initial or in the final states (or both), it will generally be possible to find some helicity amplitudes for which the maximum helicity flip σ is so large that (2.11) is satisfied. For such a choice of helicities, one may apply the Cauchy theorem to the amplitude (2.10) in the complex t plane at fixed s:

$$\int_\Gamma dt[\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|} \cdot [\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|} \cdot f(s, t; \lambda), \tag{2.12}$$

where the contour Γ is shown in Fig. 1. The right-hand cut comes from the t channel,

$$\bar{a}_4 + a_2 \to a_3 + \bar{a}_1, \tag{2.13}$$

and the branch point t_0 is given by

$$t_0 = [\text{lowest two-particle threshold in } t \text{ channel}]. \tag{2.14}$$

The left-hand cut comes from the u channel,

$$\bar{a}_3 + a_2 \to \bar{a}_1 + a_4, \tag{2.15}$$

and the branch point u_0 is

$$u_0 = M^2 - s - [\text{lowest two-particle threshold in } u \text{ channel}]. \tag{2.16}$$

In some cases it is possible that u_0 is larger than t_0, so that the cuts overlap, but this alters none of the arguments. In writing Eq. (2.12), it was assumed that there were no bound states. If these do contribute to a particular amplitude, the right-hand side of Eq. (2.12) will be the sum of the residues of all these bound-state poles. The subsequent proof is scarcely altered, and all the results apply.

The condition (2.11) implies that the integrand of (2.12) decreases more rapidly than $t^{-1-\epsilon}$ for some $\epsilon > 0$, so that the contour Γ may be deformed to infinity, and the only remaining contributions are the integrals of absorptive parts over the t and u cuts. It will be convenient to concentrate attention first upon the contribution from the t cut, viz.,

$$\int_{t_0}^{\infty} dt\ \mathrm{Im}\,\{[\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|} \times [\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|} \cdot f(s, t; \lambda)\}. \tag{2.17}$$

With $s \approx 0$, the integrand is in an unphysical region for reaction (2.2), but a physical (or mainly physical) region for reaction (2.13). Thus it is necessary to use the helicity crossing matrices to express the s-channel amplitudes in terms of the t-channel amplitudes. Let

$$F(s, t; \mu_3, \mu_1; \mu_4, \mu_2) \equiv F(s, t; \mu) \tag{2.18}$$

be a helicity amplitude for the reaction (2.13). Here t is the invariant energy square, s is the square of the

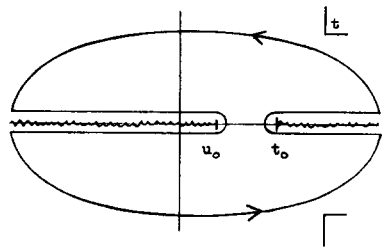


FIG. 1. Contour of integration for super-convergence relation.

momentum transfer, and μ_p, $p = 1, 2, 3, 4$, are the helicities. The connection between (2.18) and (2.1) is

$$f(s, t; \lambda) = \beta \sum_{\mu} K(s, t; \lambda; \mu) F(s, t; \mu). \quad (2.19)$$

Here β is the isospin crossing matrix, the sum is over all possible t-channel helicities, μ_1, μ_2, μ_3, μ_4, and K is the helicity crossing matrix, viz.,

$$K(s, t; \lambda; \mu) = \prod_{p=1}^{4} d^{s_p}_{\mu_p, \lambda_p}(\chi_p), \quad (2.20)$$

where

$$\cos \chi_p = \frac{(-1)^p(s + m_p^2 - \bar{m}_p^2)(t + m_p^2 - \bar{\bar{m}}_p^2) + 2m_p^2(m_1^2 - m_2^2 - m_3^2 + m_4^2)}{\{[s - (m_p + \bar{m}_p)^2][s - (m_p - \bar{m}_p)^2][t - (m_p + \bar{\bar{m}}_p)^2][t - (m_p - \bar{\bar{m}}_p)^2]\}^{\frac{1}{2}}} \quad (2.21)$$

with

$$\begin{bmatrix} \bar{m}_1 \\ \bar{m}_2 \\ \bar{m}_3 \\ \bar{m}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix},$$

$$\begin{bmatrix} \bar{\bar{m}}_1 \\ \bar{\bar{m}}_2 \\ \bar{\bar{m}}_3 \\ \bar{\bar{m}}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix},$$

and where $d^s_{\mu,\lambda}$ is the standard representation of the rotation group, which can be written

$$d^s_{\mu,\lambda}(\chi) = \left[\frac{(s + \sigma)!\,(s - \sigma)!}{(s + \tau)!\,(s - \tau)!}\right]^{\frac{1}{2}} [\cos(\chi/2)]^{|\lambda+\mu|} \times [\sin(\chi/2)]^{|\lambda-\mu|} \cdot P_{s-\sigma}^{|\lambda-\mu|,|\lambda+\mu|}(\cos \chi), \quad (2.22)$$

where

$$\sigma = \max(|\lambda|, |\mu|),$$
$$\tau = \min(|\lambda|, |\mu|),$$

and $P_J^{\alpha,\beta}(x)$ is the Jacobi polynomial defined by

$$P_J^{\alpha,\beta}(x) = 2^{-J} \sum_{m=0}^{J} \binom{J + \alpha}{m} \binom{J + \beta}{J - m} \times (x - 1)^{J-m}(x + 1)^m. \quad (2.23)$$

The t channel amplitude $F(s, t; \mu)$ can be expanded in partial waves:

$$F(s, t; \mu) = \sum_J (2J + 1) d^J_{\mu_i,\mu_f}(\theta_t) F_J(t; \mu), \quad (2.24)$$

where

$$\mu_i = \mu_2 - \mu_4, \quad \mu_f = \mu_3 - \mu_1, \quad (2.25)$$

where $F_J(t; \mu)$ is a partial-wave amplitude, and where

$$\cos \theta_t = \frac{2st + t^2 - tM^2 + (m_4^2 - m_2^2)(m_3^2 - m_1^2)}{\{[t - (m_1 + m_3)^2][t - (m_1 - m_3)^2][t - (m_2 + m_4)^2][t - (m_2 - m_4)^2]\}^{\frac{1}{2}}}. \quad (2.26)$$

For the sake of simplicity, it will be supposed that there is just one resonance in each partial-wave amplitude $F_J(t; \mu)$, namely, the spin-J member of the same Regge family to which the external particles belong. (For simplicity, no account will be taken of the fact that a given Regge family only contributes to alternate partial waves. This point can be allowed for very easily.) However, it is easy to extend the treatment to the case of more than one resonance in each wave. Moreover, any particular resonance may be suppressed by letting its coupling to the external particles tend to zero. The contribution of the graph of Fig. 2 to the partial-wave helicity amplitude may be written

$$\gamma(s_2, s_4, J)\gamma(s_1, s_3, J)C(t; \mu; J)/(m_J^2 - t), \quad (2.27)$$

where $\gamma(a, b, c)$ is the coupling at the vertex formed by the particles of spin a, b, c and where $C(t; \mu; J)$ is a kinematical factor that depends on the helicities, $\mu_1, \mu_2, \mu_3, \mu_4$. Lastly, m_J is the mass of the resonance, i.e., the mass of the spin-J particle. This mass is, strictly speaking, complex, but the "δ-function approximation" consists in writing the imaginary part of (2.27) as

$$\pi\gamma(s_2, s_4, J)\gamma(s_1, s_3, J)C(\mu; J)\delta(t - m_J^2), \quad (2.28)$$

where m_J is now the real mass, and where $C(\mu; J) \equiv C(m_J^2; \mu; J)$. This is the only contribution to the superconvergence relation that is retained.

In general, of course, there can be more than one coupling constant at a vertex (a, b, c), say, and these

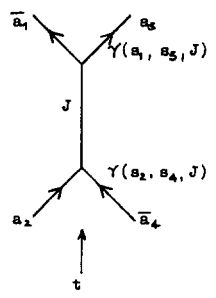


FIG. 2. Exchange of spin-J particle in t channel.

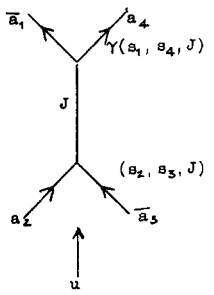


FIG. 3. Exchange of spin-J particle in u channel.

couplings will contribute with different kinematical factors to the different helicity amplitudes. However, it is not the case that there are as many couplings as there are helicity states of the external particles. Indeed, in some cases there will be no possible couplings, if only zeroth and first derivatives of the field operators are allowed in the Lagrangian. In this paper, it is proposed to show that there exist an infinite number of solutions to the superconvergence relations for the nonzero couplings. For this purpose it will be sufficient to select a vertex (a, b, c) for which at least one coupling is possible, and to set all the coupling constants except one to zero. It will be shown that an infinite number of solutions exists even with this severe constraint. It is to be expected that there are even more solutions when this artificial requirement is removed.

The t contribution to the superconvergence relation, Eq. (2.17), reads

$$\beta \sum_J (2J+1) \sum_\mu [\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|}[\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|} \times K(s,t;\lambda;\mu)\cdot d^J_{\mu_i,\mu_f}(\theta_t)\gamma(s_2,s_4,J) \times \gamma(s_1,s_3,J)C(\mu;J)\big|_{t=m_J^2}. \quad (2.29)$$

There is a similar contribution from the u channel (i.e., the left-hand cut in the complex t plane, at fixed s), viz.,

$$\gamma \sum_J (2J+1) \sum_\nu [\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|}[\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|} \times K(s,u;\lambda;\nu)d^J_{\nu_i,\nu_f}(\theta_u)\gamma(s_2,s_3,J) \times \gamma(s_1,s_4,J)C(\nu;J)\big|_{u=m_J^2}, \quad (2.30)$$

where $\nu_1, \nu_2, \nu_3, \nu_4$ are the helicities in the u channel and θ_u is the same function of s and $u \equiv M^2 - s - t$ as θ_t is of s and t [Eq. (2.26)]. Here γ is the isospin crossing matrix from the s to the u channel.

For those combinations of $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ for which the superconvergence condition (2.11) holds, the sum of the two contributions (2.29) and (2.30) is zero. Notice first that the relation may only hold for some of the possible isospin states in the s channel, since $\alpha(s)$ certainly depends on the isospin. Secondly, some of the relations may be trivial, due to cancellation of the t and u contributions. This occurs, for instance, if all four external particles are the same scalar, isoscalar boson, and in many other cases also. In other cases, the t and u channels are identical, but add constructively, as in the s-channel isospin-zero superconvergence relation for the elastic scattering of identical scalar, isospinor bosons.

These detailed considerations are not of immediate concern. It is the present purpose to consider all the nontrivial superconvergence relations for given external particles a_1, a_2, a_3, a_4. In general, there will be a set of relations between the coupling constants

$$\gamma(s_2,s_4,J),\ \gamma(s_1,s_3,J),\ \gamma(s_2,s_3,J),\ \gamma(s_1,s_4,J),$$

for all physical J, corresponding to the various superconvergent combinations of s-channel helicities and isospin. Some of these couplings may be the same, if some of the external particles are identical; and some couplings will certainly be the same, if the internal particles belong to the same family as the external particles, for example,

$$\gamma(s_2,s_4,J) = \gamma(J,s_4,s_2),$$

and these equalities must be taken into account.

The superconvergence relation, for a given s-channel isospin I and given helicities λ may be written

$$\sum_J \left\{ \sum_\mu A_J(s;\lambda;\mu)g^I_J + \sum_\nu B_J(s;\lambda;\nu)h^I_J \right\} = 0, \quad (2.31)$$

where

$$A_J(s;\lambda;\mu) \equiv [\cos\theta_s/2]^{-|\lambda_i+\lambda_f|}[\sin\theta_s/2]^{-|\lambda_i-\lambda_f|} \times K(s,t;\lambda;\mu)C(\mu,J)d^J_{\mu_i,\mu_f}(\theta_t)\big|_{t=m_J^2}, \quad (2.32)$$

$$B_J(s;\lambda;\nu) \equiv [\cos\theta_s/2]^{-|\lambda_i+\lambda_f|}[\sin\theta_s/2]^{-|\lambda_i-\lambda_f|} \times K(s,u;\lambda;\nu)C(\nu,J)d^J_{\nu_i,\nu_f}(\theta_u)\big|_{u=m_J^2}, \quad (2.33)$$

$$g^I_J \equiv \sum_{I'} \beta_{I,I'}(2J+1)\gamma^{I'}(s_2,s_4,J)\gamma^{I'}(s_1,s_3,J), \quad (2.34)$$

$$h^I_J \equiv \sum_{I'} \gamma_{I,I'}(2J+1)\gamma^{I'}(s_2,s_3,J)\gamma^{I'}(s_1,s_4,J). \quad (2.35)$$

The superconvergence relation is assumed to hold in some neighborhood of $s = 0$, so that Eq. (2.31) is equivalent to an infinite number of equations for the quantities g^I_J and h^I_J.

It is convenient to set to zero separately the t-channel and u-channel contributions, when these are distinct. Thus

$$\sum_J g^I_J \sum_\mu A_J(s;\lambda;\mu) = 0, \quad (2.36)$$

$$\sum_J h^I_J \sum_\nu B_J(s;\lambda;\nu) = 0. \quad (2.37)$$

This is not necessary, but it will simplify the ensuing algebra. The fact that it is possible to find solutions of Eqs. (2.36) and (2.37) indicates how very far the equation (2.31) is from possessing a unique solution. It will in fact only be necessary to consider Eq. (2.36) for all possible external spins s_1, s_2, s_3, s_4, since this will then entail all Eqs. (2.37).

It is necessary now to convert Eq. (2.36), which is valid for a continuum of values of s, into a discrete infinity of equations. This can be done by expanding $\sum_\mu A_J(s; \lambda; \mu)$ about a point $s = s_0$, at which it is analytic, in a Taylor series, say

$$\sum_\mu A_J(s; \lambda; \mu) = \sum_{n=0}^{\infty} A_{n,J}(\lambda)(s - s_0)^n. \quad (2.38)$$

Thus Eq. (2.36) may be written

$$\sum_{n=0}^{\infty} (s - s_0)^n \sum_J g_J^I A_{n,J}(\lambda) = 0. \quad (2.39)$$

Since this series converges uniformly for $|s - s_0|$ small enough, it follows that each coefficient of $(s - s_0)$ must vanish independently, i.e.,

$$\sum_J g_J^I A_{n,J}(\lambda) = 0, \quad (2.40)$$

for $n = 0, 1, 2, \cdots$, and all permissible I, λ.

It is the purpose of this paper to study the solubility of Eq. (2.40). This depends upon the asymptotic behavior of $A_{n,J}(\lambda)$ as $J \to \infty$, as will be shown in the next section, and so it is of interest to calculate this explicitly. If it is assumed that the mass spectrum increases indefinitely, i.e.,

$$m_J^2 \xrightarrow[J\to\infty]{} \infty, \quad (2.41)$$

then, from Eq. (2.32), this limit may be calculated explicitly to be

$$A_{n,J}(\lambda) \underset{J\to\infty}{\sim} E_n(\lambda)[J/m_J]^{2n} F_J(\lambda), \quad (2.42)$$

where E and F are two functions that are given in Appendix C. The important point is that $E_n(\lambda)$ does not depend on J and $F_J(\lambda)$ does not depend on n. Hence

$$\lim_{J\to\infty} [A_{n,J}(\lambda)/A_{n_0,J}(\lambda)] = [E_n(\lambda)/E_{n_0}(\lambda)] \lim_{J\to\infty} [m_J/J]^{2(n_0-n)}. \quad (2.43)$$

If $n_0 > n$, this limit is zero only if

$$\lim_{J\to\infty} [m_J/J] = 0. \quad (2.44)$$

This is a key equation, the derivation of which was the goal of this rather long section.

3. SOLUTION FOR GIVEN EXTERNAL SPINS

In this section it will be shown how Eq. (2.40) can be solved for the g_J^I, for any given mass spectrum m_J that satisfies (2.41) and (2.44). It will then be shown that, once the g's are known, the coupling constants $\gamma(a, b, c)$ can be obtained from Eq. (2.34). All this applies to the superconvergence relations appropriate to given external spins s_1, s_2, s_3, s_4. In Sec. 4, it is shown that simultaneous solutions for all external spins can be constructed.

For a given s-channel isospin I and helicity state λ, Eq. (2.40) reads

$$\sum_{J=0}^{\infty} a(n, J)g(J) = 0, \quad n = 0, 1, 2, \cdots, \quad (3.1)$$

where

$$a(n, J) \equiv A_{n,J}(\lambda) \quad (3.2)$$

and

$$g(J) \equiv g_J^I, \quad (3.3)$$

and where, according to Eqs. (2.43) and (2.44),

$$\lim_{J\to\infty} [a(n, J)/a(n_0, J)] = 0, \quad \text{all} \quad n < n_0. \quad (3.4)$$

In Appendix A, it is shown that, under these conditions supplemented by the observation that $a(0, J) \neq 0$ for an infinite number of values of J, an infinite number of linearly independent, absolutely convergent solutions $g(J)$ exists. The proof was originally invented by Pólya[8] thirty years ago.

It is the case, then, that one can find a solution for given s channel helicities. The question now is whether solutions $g(J)$ can be found which simultaneously satisfy Eq. (2.40) for all permissible values of the helicities λ. The fact that this is possible hinges upon the observation that, for given spins s_1, s_2, s_3, s_4, there is only a finite number of possible helicity combinations λ. Suppose that these combinations are ordered in some way, and set

$$a(n, J; q) \equiv A_{n,J}(\lambda_q), \quad (3.5)$$

where q takes on the values $q = 1, 2, \cdots, Q$, say, thus exhausting all the permissible ways of combining the s channel helicities. For any value of q, the corresponding condition (3.4) holds. This means that the equations

$$\sum_{J=0}^{\infty} a(n, J; q)g(J) = 0, \quad \begin{matrix} \text{for} & n = 0, 1, 2, \cdots, \\ \text{and} & q = 1, 2, \cdots, Q, \end{matrix} \quad (3.6)$$

can be reordered as

$$\sum_{J=0}^{\infty} A(n', J)g(J) = 0, \quad n' = 0, 1, 2, \cdots, \quad (3.7)$$

[8] G. Pólya, Commentarii Math. Helvetici **11**, 234 (1938–9).

where $A(n', J)$ is equal to one of the $a(n, J, q)$, in such a way that each of the distinct equations (3.6) occurs once, and only once, in (3.7), and such that

$$\lim_{J\to\infty} [A(n, J)/A(n_0, J)] = 0, \tag{3.8}$$

for all $n < n_0$. It is, strictly speaking, not clear that one can preserve the Pólya condition (3.8) when interleaving the different equations (3.6), since one might find two of Eqs. (3.6) with the same asymptotic behavior, say

$$a(n_1, J; q_1)/a(n_2, J; q_2) \xrightarrow[J\to\infty]{} \text{const} \tag{3.9}$$

or perhaps this ratio might oscillate, with no definite limit. This point is taken up in Appendix D, where it is shown that, even if this happens, one can find an equivalent set of equations (3.7) for which this "coincidental" asymptotic behavior does not obtain. It follows then that one can construct an infinite set of absolutely convergent solutions $g(J)$ to Eq. (3.7), that is to say, to the complete set (3.6).

It remains to be shown that, given a set of solutions g_J^I, according to Eq. (2.34), viz., one can always solve for the individual γ's:

$$g_J^I = (2J + 1) \sum_{I'} \beta_{I,I'} \gamma^{I'}(s_2, s_4, J)\gamma^{I'}(s_1, s_3, J). \tag{3.10}$$

However, this is very straightforward, for the crossing matrix satisfies

$$\beta^2 = 1 \tag{3.11}$$

so that (3.10) can be inverted to give

$$(2J + 1)\gamma^{I'}(s_2, s_4, J)\gamma^{I'}(s_1, s_3, J) = \sum_I \beta_{I',I} g_J^I. \tag{3.12}$$

It may be that not all values of I yield superconvergence relations (this depends on the detailed dynamical assumptions), but in that case there are more γ's than g's, so that Eq. (3.10) can again be solved for the γ's, this time nonuniquely.

4. SOLUTION COMBINING ALL EXTERNAL SPINS

It has been shown that, for given external spins s_1, s_2, s_3, s_4, solutions can be found to the superconvergence conditions (2.40), where g_J^I is given by (2.34). These equations may be written

$$\sum_J \sum_{I'} A_{n,J}(I, I'; \lambda)\gamma^{I'}(s_2, s_4, J)\gamma^{I'}(s_1, s_3, J) = 0, \tag{4.1}$$

where

$$A_{n,J}(I, I'; \lambda) \equiv (2J + 1)\beta_{I,I'} A_{n,J}(\lambda). \tag{4.2}$$

The problem now is to show that solutions of the whole set of Eqs. (4.1), for all possible external spins, can be found which are consistent with one another.

Suppose first that the complete set of sets of equations be ordered as follows:

(a)

$s_1=s_2=s_3=s_4=0$, so that (4.1) involves only $[\gamma^{I'}(0, 0, J)]^2$;

(b)

$s_1=1, s_3=0$: $s_2=0=s_4$ involving $\gamma^{I'}(1, 0, J)\gamma^{I'}(0, 0, J)$,

$s_2=1,\ s_4=0$ involving $[\gamma^{I'}(1, 0, J)]^2$;

(c)

$s_1=2, s_3=0$: $s_2=0=s_4$ involving $\gamma^{I'}(2, 0, J)\gamma^{I'}(0, 0, J)$,

$s_2=1,\ s_4=0$ involving $\gamma^{I'}(2, 0, J)\gamma^{I'}(1, 0, J)$,

$s_2=2,\ s_4=0$ involving $[\gamma^{I'}(2, 0, J)]^2$;

(d)

$s_1=1, s_3=1$: $s_2=0=s_4$ involving $\gamma^{I'}(1, 1, J)\gamma^{I'}(0, 0, J)$,

$s_2=1,\ s_4=0$ involving $\gamma^{I'}(1, 1, J)\gamma^{I'}(1, 0, J)$,

$s_2=2,\ s_4=0$ involving $\gamma^{I'}(1, 1, J)\gamma^{I'}(2, 0, J)$,

$s_2=1,\ s_4=1$ involving $[\gamma^{I'}(1, 1, J)]^2$;

and so on.

By this method, all sets of equations are enumerated. The difficulty is that, for example, once $\gamma^{I'}(1, 0, J)$ has been determined by the first equation (b) above, it must then be shown to be consistent with the last equation (b), the second equation (c), and so on, and similarly for all the other couplings, each of which occurs an infinite number of times. Moreover, even in determining $\gamma^{I'}(1, 0, J)$ from (b), not every term is free, since $\gamma^{I'}(1, 0, 0)$ must agree with $\gamma^{I'}(0, 0, 1)$, determined at stage (a), in the solution for $\gamma^{I'}(0, 0, J)$.

Nevertheless, it can be shown that these stringent cross-consistency conditions can be more than matched by the extreme indeterminacy of any one equation, in such a way that in fact there exist an infinity of solutions for the entire system. To show this, one should recall that a given set of equations can, in general, be solved even if an infinite set of the couplings are set identically equal to zero, so long as an infinite set are left free.

Let S_q represent the infinite set of integers

$$S_q \equiv \{q, 2q, 2^2q, 2^3q, \cdots\}, \tag{4.3}$$

where q is a prime number. Evidently there exists an infinite set of disjoint sets S_q corresponding to all the prime numbers q. Suppose now that a particular solution be constructed for $\gamma^{I'}(0, 0, J)$, from stage (a) above, in which every $\gamma^{I'}(0, 0, J)$ is set equal to zero, except for those values of J belonging to S_3. This is possible because S_3 is an infinite set. Next, $\gamma^{I'}(1, 0, J)$ is determined from the first equation (b), where

$\gamma^{I'}(0, 0, J)$ is already known, and thus is part of the coefficient. Moreover, for a given nonzero $\gamma^{I'}(0, 0, J)$, the Pólya condition for the system reads

$$\lim_{J\to\infty} [A_{n,J}(I, I'; \lambda)\gamma^{I'}(0, 0, J)]/ [A_{n_0,J}(I, I'; \lambda)\gamma^{I'}(0, 0, J)] = 0 \quad (4.4)$$

for $n < n_0$, and the coupling $\gamma^{I'}(0, 0, J)$ cancels out, giving

$$\lim_{J\to\infty} [A_{n,J}(I, I'; \lambda)/A_{n_0,J}(I, I'; \lambda)] = 0, \quad (4.5)$$

which is implied by Eq. (2.44), as has already been indicated. Hence, an absolutely convergent solution of (b) can be obtained, for those $\gamma^{I'}(1, 0, J)$ which multiply nonzero $\gamma^{I'}(0, 0, J)$. All the other $\gamma^{I'}(1, 0, J)$ are left arbitrary, and this is the crucial point, for now the last equation (b) can be solved for $\gamma^{I'}(1, 0, J)$ by first injecting the known values for J belonging to S_3. This gives a known inhomogeneous term in the equations. It is convenient also to set all $\gamma^{I'}(1, 0, J)$ to zero, except when J belongs to S_3 and S_5. Then the last equation (b) can be solved for the unknown $\gamma^{I'}(1, 0, J)$, for J in S_5. The presence of the inhomogeneous term does not alter Pólya's proof at all; the presentation in Appendix A is given for a completely arbitrary inhomogeneous term.

The next steps should be clear. The first equation (c) serves to determine $\gamma^{I'}(2, 0, J)$ for J belonging to S_3, and the second for J belonging to S_5. The last equation is satisfied by setting $\gamma^{I'}(2, 0, J)$ equal to zero, except for J belonging to S_3, S_5, S_7, and thus determining the coupling for J in S_7. Then $\gamma^{I'}(1, 1, J)$ is determined for J in S_3, S_5, S_7, successively, from the first three equations (d), and then for J in S_{11} from the last equation, with $\gamma^{I'}(1, 1, J) = 0$ for all other values of J. This process can be continued indefinitely.

Three comments may be added. It is the case that, when one solves a given equation for a coupling constant $\gamma^{I'}(s_1, s_3, J)$, for J belonging to a given set S_p, not all these couplings are free, since $\gamma^{I'}(s_1, s_3, J) = \gamma^{I'}(s_1, J, s_3)$, for example, and, for small enough J, the latter quantity may already have been determined at an earlier stage. However, it is easy to see that, at each stage, there is only a finite number of such terms. These then can be added to the inhomogeneous term, leaving an infinite number of couplings with which to solve the equation.

The second point concerns the convergence of the infinite series constituting the inhomogeneous terms that are present at the last steps of each of the stages (b), (c), (d), $\cdots$. In stage (c), for instance, it is necessary to find a solution $\gamma^{I'}(2, 0, J)$ at the first step that converges so quickly that the corresponding inhomogeneous term at the second and third steps also converge. However, since series solutions can be constructed that converge arbitrarily quickly, there is no difficulty in ensuring this. No further convergence requirements are made on $\gamma^{I'}(2, 0, J)$ after stage (c), since, at later stages, this coupling is always multiplied by some new coupling, which can be constructed to satisfy the later convergence requirements.

The third point is that the purist might object that an enumeration of the infinite set of infinite sets of equations has not been given, since it seems that one must first solve an infinite set of equations at stage (a) before starting stage (b), and so on. However, this objection can be overcome by solving first a finite number of the equations at stage (a), à la Pólya, then a finite number at stage (b), using only those values of $\gamma^{I'}(0, 0, J)$ already determined, then some more equations (a), some more (b), and a first set (c), and so. This technique involves a slight generalization of the Pólya technique, since the inhomogeneous terms at the later steps of stages (b), (c), and (d) are never completely known. This paper will not be burdened by the details of this generalization, which is straightforward and is left to the reader.

5. SIGN RESTRICTIONS ON RESIDUES

Up to this point, no account has been taken of any possible requirement of positivity of squares of couplings. For instance, if $s_1 = s_2$, and $s_3 = s_4$, so that the t channel scattering is elastic, then Eq. (2.34) becomes

$$g_J^I = (2J + 1) \sum_{I'} \beta_{I,I'}[\gamma^{I'}(s_1, s_3, J)]^2. \quad (5.1)$$

If the particles are isoscalars, $\beta_{0,0} = 1$, and there is the strong requirement that all the g_J^0 must be positive (or zero). This would generally not be observed by the solutions constructed by the Pólya method. On the other hand, if the particles were isospinors and a superconvergence relation existed only for the $I = 0$ channel, then one would have

$$g_J^0 = (2J + 1)\{-\tfrac{1}{2}[\gamma^0(s_1, s_3, J)]^2 + \tfrac{3}{2}[\gamma'(s_1, s_3, J)]^2\} \quad (5.2)$$

and there would be no sign restriction for the g_J^0, so that the analysis of the preceding sections would apply directly. Similarly, if the particles were isovectors, and the superconvergence relation arose for the $I = 2$ channel, one would have

$$g_J^2 = (2J + 1)\{\tfrac{1}{3}[\gamma^0(s_1, s_3, J)]^2 - \tfrac{1}{2}[\gamma^1(s_1, s_3, J)]^2 + \tfrac{1}{6}[\gamma^2(s_1, s_3, J)]^2\}, \quad (5.3)$$

and again there is no sign restriction.

If, further, $s_1 = s_3$ and the initial and final particles in the t channel are identical, then the Pauli principle is operative. Thus, in the isospinor example, only odd partial waves contribute to the state $I = 0$ and only even waves to $I = 1$, if the particles are bosons (the reverse obtaining for fermions). In this case one would have

$$g_J^0 = \begin{cases} \frac{3}{2}(2J+1)[\gamma^1(s_1, s_1, J)]^2 & \text{for } J \text{ even}, \\ -\frac{1}{2}(2J+1)[\gamma^0(s_1, s_1, J)]^2 & \text{for } J \text{ odd}, \end{cases} \tag{5.4}$$

and here the constraint is that g_J^0 must be positive/negative for J even/odd. Similarly, in the isovector case (again for bosons), one has

$$g_J^2 = \begin{cases} (2J+1)\{\frac{1}{3}[\gamma^0(s_1, s_1, J)]^2 \\ \qquad + \frac{1}{6}[\gamma^2(s_1, s_1, J)]^2\} & \text{for } J \text{ even}, \\ -(2J+1)\frac{1}{2}[\gamma^1(s_1, s_1, J)]^2 & \text{for } J \text{ odd}, \end{cases} \tag{5.5}$$

and so the same sign restrictions apply.

It can be shown in some of these cases that solutions exist that satisfy the sign restrictions.[6] However, what will be shown in this section is slightly weaker than an existence theorem, but for practical purposes it is equivalent. It will be shown that, if an infinite number of the g_J^I are constrained to be positive (or zero), and an infinite number negative, as in the examples (5.4) and (5.5) above, then, if one solution exists, so do an infinite number of solutions. Thus there is an infinite number of solutions or there are none at all. The more difficult case that all the g_J^I are constrained to be positive, as in the isoscalar example, will be taken up in the next section.

If one solution G_J^I of the system

$$\sum_{J=0}^{\infty} A_{n,J}(\lambda) G_J^I = 0 \tag{5.6}$$

exists, that satisfies the positivity constraints, then it is shown in Appendix B that an infinite number of solutions exists. This proof depends on the assumption that $A_{n,J}(\lambda)$ becomes positive-semidefinite for J sufficiently large. In the present case, this is true, since the analysis of Appendix C shows that $A_{n,J}(\lambda)$ has a definite limiting behavior. Should this limit have the wrong sign for a given n, the equation (5.6) for that n can be multiplied throughout by -1. The proof of Appendix B then applies. It proceeds by showing first that an intermediate solution g_J^I exists, satisfying

$$\sum_{J=0}^{\infty} A_{n,J}(\lambda) g_J^I = 0, \tag{5.7}$$

and such that, for every J, either g_J^I has the required sign or, if not, then $|g_J^I| < |G_J^I|$. Under these circumstances, it is easy to see that

$$G_J'^I \equiv G_J^I + g_J^I \tag{5.8}$$

is a new solution that satisfies the sign requirements. There exists an infinite number of such solutions.

Next, it must be shown that the analysis of Sec. 4 carries through, so that if one solution exists when all the different superconvergence relations are taken into account, then one can find an infinite number of such solutions. Let $\Gamma(a, b, c)$ be the given coupling constant, where the isospin superscript I' has been dropped, and consider the enumeration of equations given in Sec. 4. Instead of finding a solution $[\gamma(0, 0, J)]^2$ at stage (a) in which every term is zero unless J is in class S_3, one first sets $[\gamma(0, 0, J)]^2 = -[\Gamma(0, 0, J)]^2$ for all J not in S_3. This constitutes an inhomogeneous term in the stage (a) equation, which can then be solved for the remaining $[\gamma(0, 0, J)]^2$, i.e., for J in S_3, by the method of Appendix B, such that, within S_3, $[\gamma(0, 0, J)]^2$ either is positive or it is less than the corresponding $[\Gamma(0, 0, J)]^2$. Then

$$[\Gamma'(0, 0, J)]^2 = [\Gamma(0, 0, J)]^2 + [\gamma(0, 0, J)]^2 \tag{5.9}$$

is a new solution of the stage (a) equation; moreover,

$$\Gamma'(0, 0, J) = 0 \quad \text{for } J \text{ not in } S_3.$$

The first equation of stage (b) is solved for the product $\Gamma'(1, 0, J)\Gamma'(0, 0, J)$, where $\Gamma'(0, 0, J)$ is known, thus determining $\Gamma'(1, 0, J)$ for J in S_3. A solution $\gamma(1, 0, J)$ of stage (b), step 2, is obtained by setting

$$[\gamma(1, 0, J)]^2 = \begin{cases} [\Gamma'(1, 0, J)]^2 - [\Gamma(1, 0, J)]^2 & \text{for } J \text{ in } S_3, \\ -[\Gamma(1, 0, J)]^2 & \text{for } J \textit{ not} \text{ in } S_3, S_5 \end{cases} \tag{5.10}$$

as the inhomogeneous term, and solving for

$$[\gamma(1, 0, J)]^2,$$

J in S_5, as in Appendix B, such that this quantity either is positive or is smaller than the corresponding $[\Gamma(1, 0, J)]^2$. Then the following solution satisfies the positivity requirements. [Since the particles are not identical in this case, possessing different spins, there would be no Pauli principle, and so no sign restriction, since $[\gamma(1, 0, J)]^2$ would always occur in isospin combinations like Eqs. (5.2), (5.3). However, there are cases in which the special construction of this section would be necessary—for example, stage (d)]:

$$[\Gamma'(1, 0, J)]^2 = [\Gamma(1, 0, J)]^2 + [\gamma(1, 0, J)]^2. \tag{5.11}$$

Note that this definition agrees with the solution of stage (b), step 1, for J in S_3, and that $\Gamma'(1, 0, J) = 0$ for J not in S_3, S_5. Hence stages (a) and (b) have been completely satisfied, and one can then treat stage (c) in a similar manner, and so on. This completes the proof that if one solution of the complete system

exists, then one can construct an infinite number of solutions.

6. POSITIVE-DEFINITE RESIDUES

In this section the purpose is to study a system where *all* the residues g_J^I are constrained to be positive. Examples are any of the diagonal terms for isoscalar scattering, the $I = 1$ component of isospinor scattering, the $I = 0$ component of isovector scattering. In any given situation, not all of these cases might yield superconvergence relations, but there are certainly some high spin amplitudes that should do so.

Suppose, then, that g_J is to be positive or zero, and is to satisfy

$$\sum_J A_{n,J}(\lambda) g_J = 0. \tag{6.1}$$

It is easy to see that no nontrivial solution is possible if all the $A_{n,J}(\lambda)$ are positive. However, when the external particles have spin, a finite number of the coefficients are negative.

A proof of nonuniqueness is often possible along the lines of the previous section. Thus, if one solution exists

$$\sum_J A_{n,J}(\lambda) G_J = 0, \tag{6.2}$$

one can attempt to find another g_J, such that $|g_J| \leq |G_J|$, and then

$$G_J' = G_J + g_J \tag{6.3}$$

is a new solution that satisfies the positivity requirement. The difference between this case and that of Sec. 5 is that there is no guarantee that such a solution g_J can be found. The reason lies in a rather subtle shortcoming of the Pólya proof. Although solutions g_J can be constructed that cause the series

$$\sum_J |A_{n,J}(\lambda) g_J| \tag{6.4}$$

to converge arbitrarily quickly, it is not the case that g_J, considered as a function of J, can be made to decrease arbitrarily quickly with J. In fact, there is a limiting behavior for this function. If the *given* solution G_J tends to zero less quickly than this "Pólya limit," then an infinite number of solutions may be constructed, as in the preceding section. If, on the other hand, the given solution does not tend to zero less quickly than this, then it may be unique. At any rate, the Pólya construction cannot be used to find new solutions.

What is shown in this section is that, if

$$\lim_{J\to\infty} [J^{\frac{1}{2}}/m_J] = 0, \tag{6.5}$$

then, if one solution exists, so do an infinite number of solutions. If one thinks of the states as lying on a Regge trajectory, then condition (6.5) is equivalent to asserting that the solutions are nonunique if the trajectory increases less rapidly than linearly in t. If the trajectory increases more rapidly than this and a solution exists, then it may be unique.

In Appendix E it is shown that, if the mass spectrum has a power-law dependence

$$m_J^2 \underset{J\to\infty}{\sim} J^a, \tag{6.6}$$

then, in order that the superconvergence relation should not hold for arbitrarily large s (which constitutes a physical requirement), one has

$$a < 2, \tag{6.7}$$

that is, the Regge trajectory $\alpha(t)$ must increase more quickly than $t^{\frac{1}{2}}$. (This fact was first pointed out by Grodsky *et al.*[9]) If condition (6.7) is satisfied, then a solution G_J of (6.2) must satisfy

$$G_J \underset{J\to\infty}{\sim} \exp[-Kj^{1-a/2}], \tag{6.8}$$

where K is a positive constant. This is also shown in Appendix E.

In Appendix F it is shown that the fastest rate of decrease of a Pólya solution G_J, when (6.7) is satisfied, is

$$g_J \underset{J\to\infty}{\sim} \exp[-BJ^{\frac{1}{2}}], \tag{6.9}$$

where B is a positive constant. Evidently (6.9) is a faster rate than (6.8) whenever

$$a > 1, \tag{6.10}$$

i.e., whenever the Regge trajectory increases less quickly than linearly. Under these circumstances, any given solution cannot be unique.

ACKNOWLEDGMENTS

I would like to thank Dr. F. Calogero and Dr. M. B. Halpern for useful discussions, and Professor N. Cabibbo for the hospitality that has been extended to me at the Istituto di Fisica "G. Marconi."

APPENDIX A: PÓLYA'S THEOREM

In this appendix, it will be shown that the system of equations

$$\sum_{J=0}^{\infty} a(n, J) g(J) = b(n), \quad n = 0, 1, 2, \cdots, \tag{A1}$$

has an infinity of linearly independent solutions $g(J)$ if the coefficients $a(n, J)$ satisfy the condition

$$\lim_{J\to\infty} |a(n, J)/a(n_0, J)| = 0, \quad \text{for} \quad n_0 = 1, 2, 3, \cdots \text{ and all } n < n_0, \tag{A2}$$

[9] I. T. Grodsky, M. Martinis, and M. Święcki, Phys. Rev. Letters **19**, 332 (1967).

and if the first row $a(0, J)$ contains an infinity of nonzero elements. Moreover, all the infinite sums in Eq. (A1) can be made absolutely convergent. That is,

$$\sum_{J=0}^{\infty} |a(n, J)g(J)| \tag{A3}$$

exists for $n = 0, 1, 2, \cdots$. The sequence of numbers $b(n)$, $n = 0, 1, \cdots$, can be quite arbitrary.

The proof follows closely that given by Cooke[10]; but it is hoped that the presentation given here will be more readily comprehensible. The method of proof is by construction.

A particular solution $g(J)$, $J = 0, 1, 2, \cdots$, is built up successively by partially satisfying more and more of Eqs. (A1), until eventually all the equations are completely satisfied. It is a most important observation that any number, even an infinite number, of the $g(J)$ can be set equal to zero, on condition that an infinite number are nonzero.

The zeroth stage in the construction of a typical solution consists in picking two integers c and J_0 such that $c < J_0$, and $a(0, c) \neq 0$, $a(0, J_0) \neq 0$. Then $g(c)$ is given any arbitrary value and $g(J_0)$ is determined by the equation

$$a(0, c)g(c) + a(0, J_0)g(J_0) = b(0), \tag{A4}$$

i.e.,

$$g(J_0) = [b(0) - a(0, c)g(c)]/a(0, J_0). \tag{A5}$$

To avoid triviality, it is required that the term in square brackets be nonvanishing. Thus, it is convenient to pick $g(c) = 0$ if $b(0) \neq 0$, and $g(c) \neq 0$ if $b(0) = 0$. Lastly, one defines $g(J) = 0$ for all J that satisfy $0 \leq J \leq J_0$, except for $J = c$ and $J = J_0$. Then it is clear that the $n = 0$ equation has been partially satisfied, viz.,

$$\sum_{J=0}^{J_0} a(0, J)g(J) = b(0). \tag{A6}$$

The first stage consists in determining a number J_1 with $J_1 > J_0$, and the values of $g(J)$ for $J = J_0 + 1$, $J_0 + 2, \cdots, J_1$, such that Eq. (A1) is partially satisfied up to $J = J_1$ for $n = 0$ and $n = 1$, i.e.,

$$\sum_{J=0}^{J_1} a(n, J)g(J) = b(n), \quad \text{for} \quad n = 0, 1. \tag{A7}$$

This is done without altering the $g(J), J = 0, 1, 2, \cdots, J_0$, that were determined at the zeroth stage. In the second stage, the first three equations, those for $n = 0, 1, 2$, are satisfied up to some $J = J_2$ $(J_2 > J_1)$, and so on. Eventually any given equation is satisfied up to an arbitrarily high order.

The Nth stage will be described in detail. This then constitutes an inductive construction for the entire sequence $\{g(J)\}$. At the $(N-1)$th state, a number J_{N-1} has been determined such that

$$\sum_{J=0}^{J_{n-1}} a(n, J)g(J) = b(n), \quad \text{for} \quad n = 0, 1, \cdots, N-1. \tag{A8}$$

The Nth stage consists in determining J_N and $g(J)$, for $J = J_{N-1} + 1, J_{N-1} + 2, \cdots, J_N$, such that

$$\sum_{J=0}^{J_N} a(n, J)g(J) = b(n), \quad \text{for} \quad n = 0, 1, \cdots, N, \tag{A9}$$

i.e., such that one more equation is partially satisfied. It follows, by subtraction of Eq. (A8) from Eq. (A9), that

$$\sum_{J=J_{N-1}+1}^{J_N} a(n, J)g(J) = 0, \quad \text{for} \quad n = 0, 1, \cdots, N-1, \tag{A10}$$

and

$$\sum_{J=0}^{J_N} a(N, J)g(J) = b(N). \tag{A11}$$

There are here $(N + 1)$ equations and, in general, it would be possible to satisfy them by introducing $(N + 1)$ new $g(J)$, which would then be determined by the equations. However, it is convenient to introduce $(N + 2)$ new $g(J)$—one more than necessary—in order to expedite the proof of absolute convergence, as will appear.

It will be shown how to choose $(N + 1)$ numbers, $k_0, k_1, k_2, \cdots, k_N$, such that

$$J_{N-1} < k_0 < k_1 < \cdots < k_N < J_N, \tag{A12}$$

and such that

$$\sum_{p=0}^{N} a(n, k_p)g(k_p) = -a(n, J_N)g(J_N), \quad \text{for} \quad n = 0, 1, 2, \cdots, N-1, \tag{A13}$$

and

$$\sum_{p=0}^{N} a(N, k_p)g(k_p) = b(N) - \sum_{J=0}^{J_{N-1}} a(N, J)g(J) - a(N, J_N)g(J_N). \tag{A14}$$

It can be seen that Eq. (A13) and Eq. (A14) are equivalent to Eq. (A10) and Eq. (A11), if one defines $g(J) = 0$ for $J_{N-1} < J < J_N$, except for $J = k_p$, $p = 0, 1, 2, \cdots, N$.

The crucial step in the whole proof is to define $g(J_N)$ so that the right-hand side of Eq. (A14) vanishes. That is, one sets

$$g(J_N) = [a(N, J_N)]^{-1}\left\{b(N) - \sum_{J=0}^{J_{N-1}} a(N, J)g(J)\right\}. \tag{A15}$$

Thus $g(J_N)$ is defined in terms of the $g(J)$, $J = 0, 1,$

[10] R. G. Cooke, *Infinite Matrices and Sequence Spaces* (MacMillan and Co., Ltd., London, 1950).

$2, \cdots, J_{N-1}$, that have already been determined at earlier stages.

With this definition, Eq. (A13) and Eq. (A14) take the form

$$\sum_{p=0}^{N} a(n, k_p)g(k_p) = -[a(n, J_N)/a(N, J_N)]\left\{b(N) - \sum_{J=0}^{J_{N-1}} a(N, J)g(J)\right\}, \quad \text{for} \quad n = 0, 1, \cdots, N-1, \quad \text{(A16)}$$

and

$$\sum_{p=0}^{N} a(N, k_p)g(k_p) = 0. \quad \text{(A17)}$$

This is a system of $(N + 1)$ inhomogeneous equations, which can be solved for the unknown $g(k_p)$, $p = 0, 1, \cdots, N$, on condition that the matrix

$$A \equiv \begin{bmatrix} a(0, k_0) & a(0, k_1) & \cdots & a(0, k_N) \\ a(1, k_0) & a(1, k_1) & \cdots & a(1, k_N) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a(N, k_0) & & \cdots & a(N, k_N) \end{bmatrix} \quad \text{(A18)}$$

has a nonvanishing determinant. The first task is to show that $k_0, k_1, \cdots, k_N$ can always be chosen so that this is indeed the case.

The determinant would be zero if one row were a constant multiple of another. However, this is only a sufficient, not a necessary, condition for the vanishing of the determinant. A necessary and sufficient condition is that a linear relation exist between the rows, that is,

$$c_0 a(0, k_p) + c_1 a(1, k_p) + \cdots + c_N a(N, k_p) = 0, \quad \text{for all} \quad p = 0, 1, \cdots, N, \quad \text{(A19)}$$

where not all of the c_n, $n = 0, 1, \cdots, N$, vanish. It will be shown that the k_p can be chosen so that no such relation (A19) exists, so that in fact the determinant of A[Eq. (A18)] can be made nonzero.

The only way in which a relation of the form (A19) could be inescapable would be if a relation of the form

$$c_0 a(0, J) + c_1 a(1, J) + \cdots + c_N a(N, J) = 0 \quad \text{(A20)}$$

were to exist for *all* $J > J_{N-1}$, for otherwise one could choose, say, k_N such that (A19) were violated for $p = N$. It will be shown that (A20) cannot hold for all $J > J_{N-1}$, unless $c_0 = c_1 = \cdots = c_N = 0$. This constitutes the proof that the k_p can be chosen such that $\det A \neq 0$.

With any J for which $a(N, J) \neq 0$, one can write Eq. (A20) as

$$c_N = c_0 a(0, J)/a(N, J) + \cdots + C_{N-1} a(N-1, J)/a(N, J). \quad \text{(A21)}$$

Because of condition (A2), it follows that the right-hand side of this equation tends to zero as $J \to \infty$; therefore, if (A21) is to hold for all $J > J_{N-1}$, it is clear that the only possibility is $c_N = 0$. Hence Eq. (A20) reduces to

$$c_0 a(0, J) + c_1 a(1, J) + \cdots + C_{N-1} a(N-1, J) = 0. \quad \text{(A22)}$$

In the same way, it may be shown successively that

$$c_{N-1} = 0, c_{N-2} = 0, \cdots, c_1 = 0. \quad \text{(A23)}$$

Finally, since it was originally assumed also that the first row $a(0, J)$ contains an infinity of nonzero members, and therefore certainly a nonzero member for $J > J_{N-1}$, it follows that $c_0 = 0$. This completes the proof that the k_p can be chosen so that the system [(A16) and (A17)] is nonsingular (i.e., $\det A \neq 0$).

With a suitable choice of k_p, the solution of Eq. (A16) and Eq. (A17) can accordingly be written

$$\begin{bmatrix} g(k_0) \\ g(k_1) \\ \cdot \\ \cdot \\ \cdot \\ g(k_{N-1}) \\ g(k_N) \end{bmatrix} = -A^{-1} \begin{bmatrix} a(0, J_N)/a(N, J_N) \\ a(1, J_N)/a(N, J_N) \\ \cdot \\ \cdot \\ \cdot \\ a(N-1, J_N)/a(N, J_N) \\ 0 \end{bmatrix} \times \left\{b(N) - \sum_{J=0}^{J_{N-1}} a(N, J)g(J)\right\}, \quad \text{(A24)}$$

where the inverse matrix A^{-1} exists and is independent of the value of J_N, which has not yet been chosen. Because of the condition (A2), it is possible to make the column vector in Eq. (A24), and hence the $g(k_p)$, arbitrarily small; and in particular one can, by making J_N large enough, cause the sum

$$\sum_{p=0}^{N} |a(n, k_p)g(k_p)|$$

to be as small as one pleases for any given n. Lastly, $g(J_N)$ was defined by Eq. (A15), so that one has

$$|a(n, J_N)g(J_N)| = |a(n, J_N)/a(N, J_N)| \cdot \left|b(N) - \sum_{J=0}^{J_{N-1}} a(N, J)g(J)\right| \quad \text{(A25)}$$

and this also can be made arbitrarily small by making J_N large enough, but only for $n = 0, 1, \cdots, N-1$.

In fine, it has been shown that, given any $\epsilon_N > 0$, numbers $k_0, k_1, \cdots, k_N, J_N$, exist, such that, if Eq. (A8) holds, then Eq. (A9) is true; moreover,

$$\sum_{J=J_{N-1}+1}^{J_N} |a(n, J)g(J)| < \epsilon_N, \quad n = 0, 1, \cdots, N-1, \tag{A26}$$

where it should be recalled that $g(J)$ has been set to zero for $J_{N-1} < J < J_N, J \neq k_p, p = 0, 1, \cdots, N$.

This completes the description of the Nth state in the construction of a solution. It is only necessary to choose the sequence $\{\epsilon_N\}$, $N = 0, 1, 2, \cdots$, to be such that the series

$$\sum_{N=0}^{\infty} \epsilon_N \tag{A27}$$

converges, in order to be sure that the sum (A3) exists for any n, for then

$$\sum_{J=J_N+1}^{\infty} |a(N, J)g(J)| < \sum_{m=N+1}^{\infty} \epsilon_m \quad \text{for} \quad N = 0, 1, 2, \cdots. \tag{A28}$$

It should not be necessary to labor the point that the construction is nonunique, that there exist, in fact, an infinity of linearly independent, absolutely convergent solutions.

APPENDIX B: POSITIVITY CONSTRAINTS

In this section it will be shown that the system of equations (A1) for which condition (A2) holds, possesses either no solution or an infinite number of solutions when the following positivity condition is satisfied by the coefficients:

(i) For any n, there exists a J_n, such that

$$a(n, J) \geq 0, \quad \text{for all} \quad J > J_n;$$

and when the following constraint is imposed on the solutions:

(ii) $g(\sigma_p) \geq 0$, $g(\tau_p) \leq 0$, for $p = 0, 1, 2, \cdots$, where $\{\sigma_p\}$. $\{\tau_p\}$ are two infinite disjoint sets that exhaust the set of nonnegative integers.

For suppose that at least one solution $\{G(J)\}$ satisfies constraint (ii), so that

$$\sum_{J=0}^{\infty} a(n, J)G(J) = b(n), \quad n = 0, 1, 2, \cdots. \tag{B1}$$

Then one may proceed to construct a solution of the system

$$\sum_{J=0}^{\infty} a(n, J)g(J) = 0, \quad n = 0, 1, 2, \cdots, \tag{B2}$$

after the manner of Appendix A, but with $b(n) = 0$. At the zeroth stage, c is chosen to be σ_0 and $g(c)$ is given an arbitrary positive value. If $a(0, \sigma_0) > 0$, then J_0 is chosen to be τ_b; but if $a(0, \sigma_0) < 0$, J_0 is chosen to be σ_b, where b is chosen to be so large that $a(0, J_0) > 0$. Then $g(J_0)$ is given by

$$a(0, c)g(c) + a(0, J_0)g(J_0) = 0 \tag{B3}$$

so that $g(c)$ and $g(J_0)$ satisfy constraint (ii). As in Appendix A, one requires

$$g(J) = 0, \quad \text{for} \quad 0 \leq J < J_0, \quad J \neq c. \tag{B4}$$

As the Nth stage of the construction, as in Eq. (A15), one has

$$g(J_n) = -[a(N, J_N)]^{-1} \sum_{J=0}^{J_{N-1}} a(N, J)g(J). \tag{B5}$$

Now J_N can be chosen to be τ_d or σ_d, depending on whether

$$\sum_{J=0}^{J_{N-1}} a(N, J)g(J) \tag{B6}$$

is positive or negative, where d is so large that

$$a(N, J_N) > 0, \tag{B7}$$

which is always possible [condition (i)]. Then $g(J_N)$ satisfies constraint (ii). [Note that, if expression (B6) is zero, J_N may be chosen to be τ_d or σ_d.]

It can easily be shown that the infinite matrix $a(n, J)$ does not have finite rank, so that an infinite number of the $G(J)$ in Eq. (B1) must be nonzero. Hence $k_p, p = 0, 1, \cdots, N$, in Eqs. (A13) and (A14), with $b(N) = 0$, may be chosen so that

$$G(k_p) \neq 0, \quad \text{for} \quad p = 0, 1, 2, \cdots, N. \tag{B8}$$

Then Eq. (A24) shows that J_N may be made so large that the $g(k_p)$ are arbitrarily small, in particular, so that

$$|g(k_p)| < |G(k_p)|, \quad p = 0, 1, \cdots, N. \tag{B9}$$

Thus an absolutely convergent solution of Eq. (B2), $g(J)$, can be constructed, each term of which satisfies one of the following three conditions:

Either

(a) $g(J) = 0$, or
(b) $g(J)$ satisfies constraint (ii), or (B10)
(c) $|g(J)| < |G(J)|$.

Then, since Eqs. (B1) and (B2) imply

$$\sum_{J=0}^{\infty} a(n, J)[G(J) + g(J)] = b(n), \quad n = 0, 1, 2, \cdots, \tag{B11}$$

it follows that

$$G'(J) = G(J) + g(J) \tag{B12}$$

is another solution of Eq. (B1) that satisfies the constraint (ii).

It is obvious, then, that if one solution of (B1) exists, then so do an infinite number of solutions.

APPENDIX C: ASYMPTOTIC LIMITS

The problem in this appendix is to convert Eq. (2.36), viz.,

$$\sum_{J=0}^{\infty} g_J \sum_{\mu} A_J(s;\lambda;\mu) = 0, \tag{C1}$$

where

$$A_J(s;\lambda;\mu) = [\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|}[\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|} \times C(\mu,J)\cdot K(s,t;\lambda;\mu)d^J_{\mu_i,\mu_f}(\theta_t)\big|_{t=m_J^2} \tag{C2}$$

and which holds for a continuum of s values in a neighborhood of $s = 0$, into an infinite set of discrete equations of the form

$$\sum_J A_{n,J}(\lambda) g_J = 0, \tag{C3}$$

and to study the behavior of $A_{n,J}(\lambda)$ as $J \to \infty$.

Basically, the problem is reduced to an expansion of $A_J(s;\lambda;\mu)$ about some regular point in the complex s plane; and $s = 0$ is the obvious candidate for such a point. The difficulty is that there are some "kinematical" singularities at $s = 0$. While these can be factored out for finite J, it turns out that there are other singularities in the limit $J \to \infty$, which cannot be so factored. In view of these difficulties, which are of a purely formal nature, and have no fundamental significance, the easiest procedure is to expand about another fixed point $s = s_0$, close to $s = 0$, at which $A_J(s;\lambda;\mu)$ is analytic. A series expansion

$$\sum_{\mu} A_J(s;\lambda;\mu) = \sum_{n=0}^{\infty} A_{n,J}(\lambda)\cdot(s-s_0)^n \tag{C4}$$

can be made which will converge for $|s - s_0|$ sufficiently small. Then Eqs. (C1) and (C4) imply

$$\sum_n (s-s_0)^n \sum_J A_{n,J}(\lambda) g_J = 0, \tag{C5}$$

and since this must hold identically in s, for $|s - s_0|$ small enough, it follows that Eq. (C3) must hold.

The asymptotic behavior $\lim_{J\to\infty} A_{n,J}(\lambda)$ may be most easily studied by considering the limit $J \to \infty$ of the left-hand side of Eq. (C4), and then by expanding in this limit, since the convergence is uniform with respect to J.

In the limit $J \to \infty$, one has $m_J^2 \to \infty$, and the first two terms in (C2) become

$$[\cos(\theta_s/2)]^{-|\lambda_i+\lambda_f|} \times [\sin(\theta_s/2)]^{-|\lambda_i-\lambda_f|}\big|_{t=m_J^2} \underset{J\to\infty}{\sim} p(s;\lambda)(m_J^2)^{-\sigma}, \tag{C6}$$

where

$$p(s;\lambda) = (-1)^{|\lambda_i-\lambda_f|}\{[s-(m_1+m_2)^2][s-(m_1-m_2)^2] \times [s-(m_3+m_4)^2][s-(m_3-m_4)^2]/s^2\}^{\sigma}, \tag{C7}$$

with

$$\sigma = \max\{\lambda_i, \lambda_f\}. \tag{C8}$$

The third term in (C2), the helicity crossing-matrix $K(s,t;\lambda;\mu)$, which is given explicitly in Eq. (2.20), will be studied next. Now

$$\cos\chi_p\big|_{t=m_J^2} \xrightarrow[J\to\infty]{} \frac{(-1)^p(s+m_p^2-\bar{m}_p^2)}{\{[s-(m_p+\bar{m}_p)^2][s-(m_p-\bar{m}_p)^2]\}^{\frac{1}{2}}}, \quad p = 1,2,3,4. \tag{C9}$$

Thus $\chi_p|_{t=m_J^2}$ has a definite limit. This must be true also for the whole crossing matrix, so that one may write

$$K(s,m_J^2;\lambda;\mu) \xrightarrow[J\to\infty]{} K(s;\lambda;\mu), \tag{C10}$$

where the J-independent quantity on the right is defined by this relation.

The fourth term in (C2) may be written

$$d^J_{\mu_i,\mu_f}(\theta_t)\big|_{t=m_J^2} = \left[\frac{(J+M)!\,(J-M)!}{(J+N)!\,(J-N)!}\right]^{\frac{1}{2}}[\cos(\theta_t/2)]^{|\mu_i+\mu_f|} \times [\sin(\theta_t/2)]^{|\mu_i-\mu_f|}P^{|\mu_i-\mu_f|,|\mu_i+\mu_f|}_{J-M}(\cos\theta_t)\big|_{t=m_J^2}, \tag{C11}$$

where

$$M = \max(|\mu_i|,|\mu_f|),$$
$$N = \min(|\mu_i|,|\mu_f|).$$

Each of these four terms will be considered in turn. The limiting behavior of the first term may be evaluated by using Stirling's formula. This gives

$$\left[\frac{(J+M)!\,(J-M)!}{(J+N)!\,(J-N)!}\right]^{\frac{1}{2}} \xrightarrow[J\to\infty]{} 1. \tag{C12}$$

The second and third terms give

$$[\cos(\theta_t/2)]^{|\mu_i+\mu_f|}[\sin(\theta_t/2)]^{|\mu_i-\mu_f|}\big|_{t=m_J^2} \underset{J\to\infty}{\sim} (-s)^{|\mu_i-\mu_f|/2}(m_J^2)^{-|\mu_i-\mu_f|/2}. \tag{C13}$$

The Jacobi polynomial has the expansion

$$P^{|\mu_i-\mu_f|,|\mu_i+\mu_f|}_{J-M}(\cos\theta_t)\big|_{t=m_J^2} \underset{J\to\infty}{\sim} \sum_{n=0}^{J-M} B_{n,J}(\mu)(-s)^n(m_J^2)^{-n}, \tag{C14}$$

where

$$B_{n,J}(\mu) = \binom{J-M+|\mu_i-\mu_f|}{J-M-n} \times \binom{J-M+|\mu_i+\mu_f|}{n}. \tag{C15}$$

It is necessary to re-expand (C14) about $s = s_0$, which can be done, using

$$(-s)^n = (-1)^n \sum_{n'=0}^{n} \binom{n}{n'} s_0^{n-n'}(s - s_0)^{n'} \quad \text{(C16)}$$

to give

$$P_{J-M}^{|\mu_i-\mu_f|,|\mu_i+\mu_f|}(\cos\theta_t)\big|_{t=m_J^2} \underset{J\to\infty}{\sim} \sum_{n=0}^{J-M} B_{n,J}(\mu; s_0)(s - s_0)^n, \quad \text{(C17)}$$

where

$$B_{n,J}(\mu; s_0) = \sum_{n'=n}^{J-M} \binom{n'}{n} B_{n',J}(\mu)(-m_J^2)^{-n'} s_0^{n'-n}. \quad \text{(C18)}$$

Thus, in the limit $J \to \infty$, one has

$$A_J(s; \lambda; \mu) \underset{J\to\infty}{\sim} p(s, \lambda)(-s)^{|\mu_i-\mu_f|/2} K(s; \lambda; \mu) \cdot (m_J^2)^{-\sigma-|\mu_i-\mu_f|/2} \times \sum_{n=0}^{J-M} B_{n,J}(\mu; s_0)(s - s_0)^n. \quad \text{(C19)}$$

For sufficiently small $|s - s_0|$, there is a convergent series expansion

$$p(s, \lambda)(-s)^{|\mu_i-\mu_f|/2} K(s; \lambda; \mu) = \sum_{r=0}^{\infty} D_r(\lambda; \mu; s_0)(s - s_0)^r, \quad \text{(C20)}$$

where it is not necessary to calculate the coefficients D_r; the essential point is that they are independent of J. Hence Eq. (C19) becomes

$$A_J(s; \lambda; \mu) \underset{J\to\infty}{\sim} (m_J^2)^{-\sigma-|\mu_i-\mu_f|/2} C(\mu, J) \sum_{r=0}^{\infty} (s - s_0)^r \times \sum_{n=r}^{J-M} B_{n,J}(\mu; s_0) D_{n-r}(\lambda; \mu; s_0). \quad \text{(C21)}$$

This limiting form can be inserted into (C1) to give the limiting form of (C3), with the result

$$A_{n,J}(\lambda) \underset{J\to\infty}{\sim} \sum_{\mu} (m_J^2)^{-\sigma-|\mu_i-\mu_f|/2} C(\mu, J) \times \sum_{r=n}^{J-M} B_{r,J}(\mu; s_0) D_{r-n}(\lambda; \mu; s_0). \quad \text{(C22)}$$

The object now is to show that the expression (C22) satisfies the Pólya condition, namely,

$$\lim_{J\to\infty} |A_{n,J}(\lambda)/A_{n_0,J}(\lambda)| = 0, \quad \text{(C23)}$$

for all $n < n_0$. This will be shown to hold if

$$\lim_{J\to\infty} [m_J/J] = 0. \quad \text{(C24)}$$

From (C15) one finds that

$$B_{n,J}(\mu) \underset{J\to\infty}{\sim} J^{2n+|\mu_i-\mu_f|}[n!\,(n + |\mu_i - \mu_f|)!]^{-1}. \quad \text{(C25)}$$

Hence (C18) yields

$$B_{n,J}(\mu; s_0) \underset{J\to\infty}{\sim} \sum_{n'=n}^{J-M} \binom{n'}{n} (-1)^{n'} s_0^{n'-n} J^{|\mu_i-\mu_f|} \times [n'!\,(n' + |\mu_i - \mu_f|)!]^{-1}[J/m_J]^{2n'}. \quad \text{(C26)}$$

Because of the inverse factorials, the first term in the series dominates (since J/m_J cannot increase faster than J), i.e.,

$$B_{n,J}(\mu; s_0) \underset{J\to\infty}{\sim} (-1)^n J^{|\mu_i-\mu_f|} \times [n!\,(n + |\mu_i - \mu_f|)!]^{-1}[J/m_J]^{2n}. \quad \text{(C27)}$$

Again, in (C22) the first term, corresponding to $r = n$, dominates, giving for the asymptotic value of the expansion coefficients

$$A_{n,J}(\lambda) \underset{J\to\infty}{\sim} (m_J^2)^{-\sigma}[J/m_J]^{2n-\mu'}(-1)^n[n!\,(n + \mu')!]^{-1} \times \sum_{\mu}' C(\mu, J) D_0(\lambda; \mu; s_0), \quad \text{(C28)}$$

where $\mu' = \min |\mu_i - \mu_f|$ for all possible t-channel helicities, and where $\sum_{\mu}'$ means a summation over all t-channel helicities consistent with $|\mu_i - \mu_f| = \mu'$.

The final result follows:

$$|A_{n,J}(\lambda)/A_{n_0,J}(\lambda)| \underset{J\to\infty}{\sim} [m_J/J]^{2(n-n_0)} \frac{n_0!\,(n_0 + \mu')!}{n!\,(n + \mu')!} \quad \text{(C29)}$$

and, by (C24), this tends to zero as $J \to \infty$, for any $n < n_0$.

APPENDIX D: REORDERING THE EQUATIONS

In Sec. 3, the set of sets of equations

$$\sum_{J=0}^{\infty} A_{n,J}(\lambda_q) g(J) = 0, \quad n = 0, 1, 2, \cdots, \quad \text{(D1)}$$

where $q = 1, 2, \cdots, Q$, corresponding to the different permissible helicity states in the s channel, is reordered as a single set of equations

$$\sum_{J=0}^{\infty} A(n', J) g(J) = 0, \quad n' = 0, 1, 2, \cdots. \quad \text{(D2)}$$

With a mass spectrum

$$m_J^2 \underset{J\to\infty}{\sim} J^a, \quad 0 < a < 2, \quad \text{(D3)}$$

and fixed helicities λ, formula (C28) shows that

$$\lim_{J\to\infty} |A_{n,J}(\lambda_q)/A_{n_0,J}(\lambda_q)| = 0, \quad \text{(D4)}$$

for any $n < n_0$. In reordering (D1) in the form (D2), it is clear that the corresponding condition

$$\lim_{J\to\infty} |A(n', J)/A(n_0', J)| = 0 \quad \text{(D5)}$$

can be maintained for any $n' < n_0'$, aside from possible coincidences in asymptotic behavior of $A_{n,J}(\lambda_q)$ for different values of q. The maximum order of coincidence is Q, the number of distinct, permissible helicity states. The purpose of this appendix is to show that an equivalent system of equations (D2) can be defined for which all the coincidences have been removed, so that (D5) holds without exception.

Consider the asymptotic expression (C28) for $A_{n,J}(\lambda)$. There will be coincident asymptotic behaviors, e.g.,

$$A_{n_1,J}(\lambda_1) \underset{J\to\infty}{\sim} A_{n_2,J}(\lambda_2) \tag{D6}$$

for two *different* values of the maximum helicity flip σ [defined in Eq. (2.9)], say for $\sigma = \sigma_1$ and $\sigma = \sigma_2$, only if

$$n_1 - n_2 = \frac{a}{2-a}(\sigma_1 - \sigma_2). \tag{D7}$$

But this will be possible only if the right-hand side of (D7) is an integer. For a general value of a, this will not be so, and so there will not be any coincidences between different values of σ. To be quite certain about this, it is enough to require a to be irrational.

However, a glance at Eq. (C28) shows that all the coefficients corresponding to the same value of σ will have degenerate asymptotic behaviors. Suppose that there are Q' such coefficients which have the behavior

$$A_{n,J}(\lambda_q) \underset{J\to\infty}{\sim} A^0_{n,J}(\sigma)\left[1 + \sum_{r=1}^{\infty} K_{n,r}(\lambda_q) m_J^{-2r}\right], \tag{D8}$$

$q = 1, 2, \cdots, Q'$, where $A^0_{n,J}(\sigma)$ is the leading term, common to all the λ_q which have the same value of σ, and which is given in (C28). The higher terms are obtained by expanding $A_J(s;\lambda;\mu)$ [Eq. (2.32)] in inverse powers of m_J^2. Since this involves expanding the helicity crossing matrix $K(s,t;\lambda;\mu)$ in powers of $1/t$ with $t = m_J^2 \to \infty$, it is clear from Eqs. (2.20)–(2.23) that coefficients $K_{n,r}(\lambda_q)$ in Eq. (D8) will be different for each of the four helicities $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. This means that the Q' sets of equations

$$\sum_J A_{n,J}(\lambda_q) g_J = 0, \tag{D9}$$

$q = 1, 2, \cdots, Q'$, which have the degenerate behavior

$$A_{n,J}(\lambda_q) \underset{J\to\infty}{\sim} J^{(2-a)(n-\mu'/2)-a\sigma} \sum_\mu{}' C(\mu, J) D_0(\lambda;\mu;s_0), \tag{D10}$$

can be replaced by Q' linear combinations, one of which has the asymptotic behavior (D10), i.e., that of $A^0_{n,J}(\sigma)$, one of which behaves like $m_J^{-2}A^0_{n,J}(\sigma)$, one like $m^{-4}A^0_{n,J}(\sigma)$, and so on, until the last one behaves like $m_J^{-2(Q'-1)}A^0_{n,J}(\sigma)$. Let these linear combinations be labeled $A_{n,J}(\lambda_q;\sigma)$, $q = 1, 2, \cdots, Q'$, starting with the one that tends to infinity least quickly. Then the system

$$\sum_J A_{n,J}(\lambda_q;\sigma) g_J = 0, \tag{D11}$$

$q = 1, 2, \cdots, Q'$, is equivalent to (D9), and it satisfies

$$|A_{n,J}(\lambda_q;\sigma)/A_{n_0,J}(\lambda_q;\sigma)| \underset{J\to\infty}{\sim} J^{-|q_0-q|a}, \tag{D12}$$

which tends to zero as $J \to \infty$ for any $q < q_0$. Moreover, if a is irrational, it is guaranteed that there is no degeneracy of asymptotic behavior between different values of n or different values of σ.

Hence a complete ordering of the system (D1) in the form (D2), with strict observance of condition (D5), is possible. Notice that this is so because there is only a finite number of infinite sets of equations (D1), so that the recursive redefinitions of degenerate coefficients involve only a finite number of steps.

It is clear that the restriction that "a," the exponent of the mass-spectrum power law, be irrational need not be taken seriously. This was only necessary to avoid the possibility of accidental coincidences between different n and σ; since any rational can be approximated arbitrarily well by an irrational, for practical purposes this restriction may be neglected.

APPENDIX E: PHYSICAL CONVERGENCE LIMITATIONS

A typical superconvergence relation, Eq. (2.12), is supposed to hold only for a limited range of s values. This means that the infinite sum (2.36), viz.,

$$\sum_{J=0}^{\infty} g_J \sum_\mu A_J(s;\lambda;\mu) = 0, \tag{E1}$$

must converge for only a limited range also. There must be an s_0 such that the infinite J sum in (E1) converges for $s < s_0$ and diverges for $s > s_0$. In this appendix, it will be shown that this simple requirement implies a powerful restriction on both the mass spectrum and upon the asymptotic dependence of the g_J.

From Appendix C, one has the following asymptotic expression for the coefficients:

$$\sum_\mu A_J(s;\lambda;\mu) \underset{J\to\infty}{\sim} p(s,\lambda)(-s)^{\mu'/2}(m_J^2)^{-\sigma-\mu'/2} \times \sum_\mu{}' K(s;\lambda;\mu) C(\mu;J) P^{|\mu_i-\mu_f|,|\mu_i+\mu_f|}_{J-M}(\cos\theta_t)\big|_{t+m_J^2}. \tag{E2}$$

In the large-J limit[11]

$$P_J^{\alpha,\beta}(\cosh u) \underset{J\to\infty}{\sim} \tfrac{1}{2}e^{Ju}\exp[(\alpha+\beta+1)u/2]\times(\pi J)^{-\frac{1}{2}} \times[\sinh(u/2)]^{-(\alpha+\frac{1}{2})}[\cosh(u/2)]^{-(\beta+\frac{1}{2})}, \quad \text{(E3)}$$

so that (E2) reduces to

$$\sum_\mu A_J(s;\lambda;\mu) \underset{J\to\infty}{\sim} \tfrac{1}{2}s^{-\frac{1}{2}}p(s,\lambda)[m_J^2]^{-\sigma+\frac{1}{2}}(\pi J)^{-\frac{1}{2}} \times\exp[2s^{\frac{1}{2}}J/m_J]. \quad \text{(E4)}$$

Suppose now that the mass spectrum m_J increases faster than J:

$$\lim_{J\to\infty}[J/m_J]=0. \quad \text{(E5)}$$

Then the exponential term in (E4) reduces to unity, and the series (E1) becomes, asymptotically,

$$\tfrac{1}{2}s^{-\frac{1}{2}}p(s,\lambda)\sum_J g_J[m_J^2]^{-\sigma+\frac{1}{2}}(\pi J)^{-\frac{1}{2}}. \quad \text{(E6)}$$

Since the s and the J dependences have factored, it is clear that if this series converges for any s, it converges for all s, and so there cannot be an s_0 such that (E1) converges for $s<s_0$ and diverges for $s>s_0$. Hence the class of mass spectra (E5) must be rejected on physical grounds. The same reasoning applies if J/m_J tends to a finite limit, for then the exponential factor in (E4) tends to a function only of s, and there is again factorization between J and s.

Hence, in place of (E5) one must have

$$\lim_{J\to\infty}[m_J/J]=0, \quad \text{(E7)}$$

which means that the Regge trajectory $\alpha(t)$ must rise more rapidly than $t^{\frac{1}{2}}$. Suppose that the mass spectrum obeys the power law

$$m_J^2 \underset{J\to\infty}{\sim} J^a, \quad \text{(E8)}$$

with $a<2$. Then (E1) has the asymptotic form

$$\tfrac{1}{2}\pi^{-\frac{1}{2}}s^{-\frac{1}{4}}p(s,\lambda)\sum_J g_J J^{-a(\sigma-\frac{1}{4})-\frac{1}{2}}\exp[2s^{\frac{1}{2}}J^{1-a/2}]. \quad \text{(E9)}$$

If this series is to converge for $s<s_0$ and diverge for $s>s_0$, then g_J must have the asymptotic behavior

$$g_J \underset{J\to\infty}{\sim} \exp[-2s_0^{\frac{1}{2}}\cdot J^{1-a/2}]. \quad \text{(E10)}$$

This formula is of use in Sec. 6.

APPENDIX F: LIMITING BEHAVIOR OF PÓLYA SOLUTIONS

It may not be immediately obvious, from the Pólya proof (Appendix A), that there is a limiting asymptotic behavior for the solutions $g(J)$ of Eq. (A1). At the Nth stage in the construction, $g(k_0)$, $g(k_1),\cdots,g(k_N)$, can certainly be made as small as one pleases for fixed $k_0,k_1,\cdots,k_N$ simply by making J_N large enough. Moreover, if $a(n,J)\to\infty$ as $J\to\infty$, as in the cases of interest in this paper, then $g(J_N)$ defined by (A15) can be made indefinitely small simply by making J_N large enough. However, this does not mean necessarily that $g(J_N)$ decreases arbitrarily quickly *as a function of* J_N.

Consider Eq. (A15) in the homogeneous case $b(n)\equiv 0$:

$$g(J_N)=-[a(N,J_N)]^{-1}\sum_{J=0}^{J_{N-1}}a(N,J)g(J). \quad \text{(F1)}$$

The asymptotic behavior will be governed by $g(J_N)$, $N\to\infty$. From Eq. (2.42), and the power-law mass spectrum

$$m_J^2 \underset{J\to\infty}{\sim} J^a, \quad 0<a<2, \quad \text{(F2)}$$

one has

$$a(N,J) \underset{J\to\infty}{\sim} J^{(2-a)N}. \quad \text{(F3)}$$

In order to solve (F1) asymptotically, it is now necessary to know how J_N depends on N. It is clear that the more quickly J_N increases with N, the less quickly N increases with J_N, and so the less quickly $g(J_N)$ decreases as a function of J_N. Hence, to obtain the most rapid possible decrease of $g(J_N)$ with J_N, one must choose J_N to increase as slowly as possible with N. Now the smallest possible J_N, consistent with the formal Pólya construction, is $\frac{1}{2}N^2+3N+2$, so the following class of J_N dependences will yield the minimal $g(J_N)$:

$$J_N=K^2N^2, \quad \text{(F4)}$$

where K is a constant satisfying $K^2>\frac{1}{2}$. With this expression the term in (F1) corresponding to $J=J_n=K^2n^2$ has the asymptotic behavior

$$(N/n)^{-(2-a)N}g(j_n). \quad \text{(F5)}$$

The ansatz

$$g(J) \underset{J\to\infty}{\sim} \exp\left[-\frac{2-a}{K}J^{\frac{1}{2}}\right] \quad \text{(F6)}$$

satisfies Eq. (F1) asymptotically, for then the largest term of the form (F5) occurs for $n=N-1$, when one has

$$\exp\left\{(2-a)N\log\left(1-\frac{1}{N}\right)-\frac{2-a}{K}K(N-1)\right\} \underset{N\to\infty}{\sim}\exp\left\{-\frac{2-a}{K}J_N^{\frac{1}{2}}\right\}, \quad \text{(F7)}$$

which is consistent. Thus the most rapidly decreasing solution that one can construct by the Pólya method has the behavior (F6). It is not difficult to show that the full proof of Sec. 4 remains valid for a J_N dependence as in Eq. (F4), with K sufficiently large, although the disjoint sets S_q need to be chosen more carefully than those defined by Eq. (4.3).

[11] *Bateman Manuscript Project, Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill Book Co., New York, 1953).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

One- and Two-Dimensional Green's Functions for Electromagnetic Waves in Moving Simple Media

R. T. COMPTON, JR.
Technische Hochschule, Munich, Germany*

(Received 19 January 1968)

This paper considers the one- and two-dimensional Green's functions associated with electromagnetic radiation in a moving medium. The medium is assumed to be lossless, to have constant permittivity and permeability, and to move with constant velocity with respect to a given inertial coordinate system. For each source (line source or plane source) both the time-dependent and the harmonic Green's functions are found. The results are obtained by means of special transformations which convert the equations to be solved into standard forms. The significance of the results, in terms of one- and two-dimensional Cerenkov radiation, is discussed.

1. INTRODUCTION

Electromagnetic radiation from a point source in a simple moving medium has been discussed in several recent papers. Tai[1], Lee and Papas,[2] in independent works, discussed the fields radiated by a harmonic point source in a moving medium, and Compton[3] has given the time-dependent Green's function for such a medium. The Fourier transform relation between the time-dependent and harmonic solutions has also been discussed in a short note by Tai.[4]

In this paper, we solve for the one- and two-dimensional Green's functions for electromagnetic waves in a moving simple medium. That is, we will find the Green's functions associated with radiation from an infinite line source and from an infinite plane source. For each source, both the time-dependent and time-harmonic Green's functions will be found. A knowledge of these functions will allow us to discuss various interesting problems (e.g., two-dimensional Cerenkov radiation) and will complete our catalog of Green's functions for simple moving media.

In the previous paper by the author,[3] a Fourier integral method was used to construct the solution for the three-dimensional Green's function. As was remarked in that paper [see Eq. (37) of Ref. 3], however, the solution can also be found by an alternate method making use of an affine transformation of coordinates that converts the differential equation to be solved into the standard time-dependent wave equation. It turns out that this method is the simpler of the two, and a second purpose of this paper will be to illustrate this procedure in detail. Tai,[1] in solving for the three-dimensional harmonic Green's function, makes use of a special transformation with which he is able to cast his differential equation into a two-dimensional Klein–Gordon equation, whose solution is known. As will be apparent below, the transformation used by Tai and the affine transformation suitable for solving the time-dependent equation are closely related in the frequency and time domains. For the one- and two-dimensional Green's functions studied here, the use of an affine transformation in the time domain and a transformation similar to Tai's in the frequency domain will allow us to find the solutions quickly, with no integrals to evaluate.

2. FORMULATION OF THE PROBLEM

We consider a lossless medium with constant permittivity ϵ and permeability μ (both measured in a reference frame attached to the medium) that moves with constant velocity $\mathbf{v} = v\hat{\mathbf{z}}$ with respect to a given inertial reference frame. [For a detailed derivation of Eqs. (1)–(10) of this section, the reader is referred to Eqs. (1)–(37) of Ref. 3.] The electromagnetic fields, as measured in the given frame, satisfy the Maxwell–Minkowski equations

$$\mathbf{D}_0 \times \mathbf{E} = -\frac{\partial}{\partial t}(\mu\boldsymbol{\alpha} \cdot \mathbf{H}), \tag{1a}$$

$$\mathbf{D}_0 \times \mathbf{H} = \frac{\partial}{\partial t}(\epsilon\boldsymbol{\alpha} \cdot \mathbf{E}) + \mathbf{J}, \tag{1b}$$

$$\mathbf{D}_0 \cdot (\epsilon\boldsymbol{\alpha} \cdot \mathbf{E}) = \rho + \boldsymbol{\Omega} \cdot \mathbf{J}, \tag{1c}$$

$$\mathbf{D}_0 \cdot (\mu\boldsymbol{\alpha} \cdot \mathbf{H}) = 0, \tag{1d}$$

where $\mathbf{D}_0$ is the differential operator

$$\mathbf{D}_0 = \nabla - \boldsymbol{\Omega}\frac{\partial}{\partial t}, \tag{2}$$

* Present address: Department of Electrical Engineering, Ohio State University, 1320 Kinnear Road, Columbus, Ohio 43212. The work reported in this paper was performed while the author was in tenure as a National Science Foundation Postdoctoral Fellow.

[1] C. T. Tai, IEEE Trans., AP-13, 322 (1965).
[2] K. S. H. Lee and C. H. Papas, J. Math. Phys. **5**, 1668 (1964).
[3] R. T. Compton, Jr., J. Math. Phys. **7**, 2145 (1966).
[4] C. T. Tai, J. Math. Phys. **8**, 646 (1967).

$\boldsymbol{\alpha}$ is the dyadic with components,

$$\boldsymbol{\alpha} = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3a}$$

and where

$$a = (1 - \beta^2)/(1 - n^2\beta^2), \tag{3b}$$

$$\boldsymbol{\Omega} = [(n^2 - 1)\beta/c(1 - n^2\beta^2)]\,\hat{\mathbf{z}}, \tag{3c}$$

$$\beta = v/c, \tag{3d}$$

and

$$n = (\mu\epsilon/\mu_0\epsilon_0)^{\frac{1}{2}}. \tag{3e}$$

ϵ_0 and μ_0 are the permittivity and permeability of free space, and $c = (\epsilon_0\mu_0)^{-\frac{1}{2}}$ is the velocity of light in free space. **E**, **H**, **B**, **D**, ρ, and **J** are the Maxwell fields and sources in the usual notation, and all quantities are measured in rationalized MKS units. We assume also that $n > 1$.

Equations (1) can be solved by defining a vector potential function $\mathbf{A}_0$ and a scalar potential function Ψ such that

$$\mathbf{H} = \frac{1}{\mu}\boldsymbol{\alpha}^{-1} \cdot \mathbf{D}_0 \times (\boldsymbol{\alpha}^{-1} \cdot \mathbf{A}_0), \tag{4a}$$

$$\mathbf{E} = -\boldsymbol{\alpha}^{-1} \cdot \frac{\partial \mathbf{A}_0}{\partial t} - \mathbf{D}_0\Psi. \tag{4b}$$

($\boldsymbol{\alpha}^{-1}$ denotes the inverse of $\boldsymbol{\alpha}$.) If the potentials are chosen to satisfy the gauge relation

$$\mathbf{D}_0 \cdot \mathbf{A}_0 = -\epsilon\mu a^2 \frac{\partial \Psi}{\partial t}, \tag{5}$$

then it is found that they satisfy the following wave equations:

$$(\mathbf{D}_a \cdot \mathbf{D}_0)\mathbf{A}_0 - \epsilon\mu a \frac{\partial^2 \mathbf{A}_0}{\partial t^2} = -\mu a \mathbf{J}, \tag{6a}$$

$$(\mathbf{D}_a \cdot \mathbf{D}_0)\Psi - \epsilon\mu a \frac{\partial^2 \Psi}{\partial t^2} = -\frac{\rho + \boldsymbol{\Omega} \cdot \mathbf{J}}{a\epsilon}, \tag{6b}$$

where

$$\mathbf{D}_a = \frac{1}{a}\boldsymbol{\alpha} \cdot \mathbf{D}_0 = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\left(\frac{1}{a}\frac{\partial}{\partial z} - \frac{\Omega}{a}\frac{\partial}{\partial t}\right). \tag{7}$$

In the previous work,[3] we solved Eqs. (6) by defining the time-dependent Green's function G as the solution to

$$(\mathbf{D}_a \cdot \mathbf{D}_0)G - \epsilon\mu a \frac{\partial^2 G}{\partial t^2} = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'), \tag{8}$$

where $\delta(\mathbf{r} - \mathbf{r}')\delta(t - t')$ is a delta-function source occurring at $\mathbf{r} = \mathbf{r}'$, $t = t'$. [G is the solution to Eq. (8) that satisfies the causality condition.] The solutions for $\mathbf{A}_0$ and Ψ are then given by

$$\mathbf{A}_0(\mathbf{r}, t) = -\mu a \iiiint \mathbf{J}(\mathbf{r}', t')G(\mathbf{r}, t \mid \mathbf{r}', t')\,d\mathbf{r}'\,dt', \tag{9a}$$

$$\Psi(\mathbf{r}, t) = \frac{-1}{a\epsilon}\iiiint [\rho(\mathbf{r}', t') + \boldsymbol{\Omega} \cdot \mathbf{J}(\mathbf{r}', t')] \times G(\mathbf{r}, t \mid \mathbf{r}', t')\,d\mathbf{r}'\,dt', \tag{9b}$$

and the fields may be found from Eqs. (4) plus the constitutive relations

$$\mathbf{D} = \epsilon\boldsymbol{\alpha} \cdot \mathbf{E} + \boldsymbol{\Omega} \times \mathbf{H}, \tag{10a}$$

$$\mathbf{B} = \mu\boldsymbol{\alpha} \cdot \mathbf{H} - \boldsymbol{\Omega} \times \mathbf{E}. \tag{10b}$$

In this paper we will study the Green's functions appropriate for solving Eqs. (6) when the source terms are one- and two-dimensional. That is, we study the equation

$$(\mathbf{D}_a \cdot \mathbf{D}_0)G - \epsilon\mu a \frac{\partial^2 G}{\partial t^2} = \text{source term}, \tag{8'}$$

where the "source term" is a one- or two-dimensional delta function, having either impulsive or harmonic time dependence.

It will simplify matters if, before we begin, we rotate the coordinate system so the velocity **v** is at an arbitrary direction with respect to the coordinate axes. We may then let the line source or plane source lie on the coordinate axes. Eq. (8'), when written out in scalar form, is

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{1}{a}\frac{\partial^2}{\partial z^2} - 2\frac{\Omega}{a}\frac{\partial^2}{\partial t \partial z} + \frac{\Omega^2}{a}\frac{\partial^2}{\partial t^2} - \epsilon\mu a \frac{\partial^2}{\partial t^2}\right)G = \text{source term}. \tag{11}$$

We first relabel the coordinate axes so that

$$y = w \quad z = u. \tag{12}$$

Eq. (11) for G is then

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial w^2} + \frac{1}{a}\frac{\partial^2}{\partial u^2} - 2\frac{\Omega}{a}\frac{\partial^2}{\partial t \partial u} + \frac{\Omega^2}{a}\frac{\partial^2}{\partial t^2} - \epsilon\mu a \frac{\partial^2}{\partial t^2}\right)G = \text{source term}. \tag{13}$$

Now we consider a second coordinate system xyz, where the y and z axes are rotated by an angle ϕ with respect to the w and u axes, as shown in Fig. 1. From the chain rule of differentiation, it is readily found that

$$\frac{\partial G}{\partial u} = \cos\phi \frac{\partial G}{\partial z} + \sin\phi \frac{\partial G}{\partial y}, \tag{14a}$$

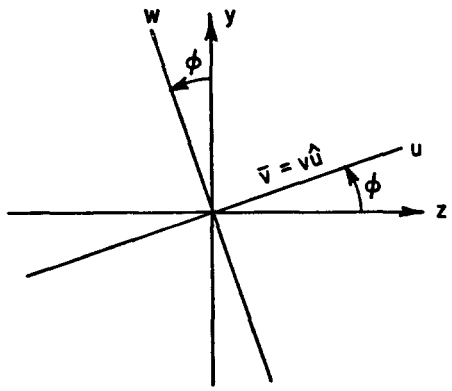


FIG. 1. *xyz* and *xwu* coordinate systems.

$$\frac{\partial^2 G}{\partial u^2} = \cos^2\phi \frac{\partial^2 G}{\partial z^2} + 2\sin\phi\cos\phi \frac{\partial^2 G}{\partial z \partial y} + \sin^2\phi \frac{\partial^2 G}{\partial y^2}, \tag{14b}$$

$$\frac{\partial^2 G}{\partial w^2} = \cos^2\phi \frac{\partial^2 G}{\partial y^2} - 2\sin\phi\cos\phi \frac{\partial^2 G}{\partial y \partial z} + \sin^2\phi \frac{\partial^2 G}{\partial z^2}, \tag{14c}$$

and hence the differential equation for G in terms of the xyz axes is

$$\begin{aligned}\frac{\partial^2 G}{\partial x^2} &+ \left(\cos^2\phi + \frac{1}{a}\sin^2\phi\right)\frac{\partial^2 G}{\partial y^2} \\ &- 2\left(1 - \frac{1}{a}\right)\sin\phi\cos\phi \frac{\partial^2 G}{\partial y \partial z} \\ &+ \left(\sin^2\phi + \frac{1}{a}\cos^2\phi\right)\frac{\partial^2 G}{\partial z^2} \\ &- 2\frac{\Omega}{a}\sin\phi \frac{\partial^2 G}{\partial t \partial y} - 2\frac{\Omega}{a}\cos\phi \frac{\partial^2 G}{\partial t \partial z} \\ &+ \frac{\Omega^2}{a}\frac{\partial^2 G}{\partial t^2} - \epsilon\mu a \frac{\partial^2 G}{\partial t^2} = \text{source term}. \end{aligned} \tag{15}$$

This equation will simplify considerably for the types of sources discussed below.

3. TWO-DIMENSIONAL TIME-DEPENDENT GREEN'S FUNCTION

For this Green's function we assume an infinite line source parallel to the z axis, with an impulsive time dependence. That is,

$$\text{source term} = \delta(x - x')\delta(y - y')\delta(t - t'). \tag{16}$$

The solution for G then has no z dependence, and Eq. (15) simplifies to the following:

$$\begin{aligned}\frac{\partial^2 G}{\partial x^2} &+ b\frac{\partial^2 G}{\partial y^2} - c\frac{\partial^2 G}{\partial y \partial t} - d\frac{\partial^2 G}{\partial t^2} \\ &= \delta(x - x')\delta(y - y')\delta(t - t'), \end{aligned} \tag{17}$$

where

$$b = \cos^2\phi + (1/a)\sin^2\phi, \tag{18a}$$

$$c = 2(\Omega\sin\phi)/a, \tag{18b}$$

$$d = \epsilon\mu a - (\Omega^2/a). \tag{18c}$$

As mentioned above, we may solve this equation by making an affine transformation of variables to eliminate the mixed derivative and then comparing the resulting differential equation with the standard equation for the two-dimensional Green's function. Specifically, suppose we define new variables t_1 and y_1 by

$$\left.\begin{aligned} y &= y_1 - At_1 \\ t &= t_1 \end{aligned}\right\} \quad \text{or} \quad \left\{\begin{aligned} y_1 &= y + At \\ t_1 &= t, \end{aligned}\right. \tag{19}$$

where A is a constant, as of yet unspecified. Then, by the chain rule, we find that

$$\frac{\partial^2 G}{\partial y^2} = \frac{\partial^2 G}{\partial y_1^2}, \tag{20a}$$

$$\frac{\partial^2 G}{\partial t^2} = A^2\frac{\partial^2 G}{\partial y_1^2} + 2A\frac{\partial^2 G}{\partial y_1 \partial t_1} + \frac{\partial^2 G}{\partial t_1^2}, \tag{20b}$$

$$\frac{\partial^2 G}{\partial t \partial y} = A\frac{\partial^2 G}{\partial y_1^2} + \frac{\partial^2 G}{\partial y_1 \partial t_1}. \tag{20c}$$

In terms of x, y_1, and t_1, the differential equation (17) becomes

$$\begin{aligned}\frac{\partial^2 G}{\partial x^2} &+ (b - cA - dA^2)\frac{\partial^2 G}{\partial y_1^2} \\ &- (c + 2Ad)\frac{\partial^2 G}{\partial y_1 \partial t_1} - d\frac{\partial^2 G}{\partial t_1^2} \\ &= \delta(x - x')\delta(y_1 - At_1 - y')\delta(t - t') \\ &= \delta(x - x')\delta(y_1 - At' - y')\delta(t - t'). \end{aligned} \tag{21}$$

We see that we may eliminate the mixed derivative by choosing the constant A such that

$$A = -\frac{c}{2d} = \frac{\Omega}{\Omega^2 - \epsilon\mu a^2}\sin\phi = \frac{(n^2 - 1)\beta c}{\beta^2 - n^2}\sin\phi. \tag{22}$$

The coefficient of $\partial^2 G/\partial y_1^2$ then becomes

$$\begin{aligned} b - cA - dA^2 &= b + \frac{c^2}{4d} \\ &= \cos^2\phi + \frac{1}{a}\left[1 + \frac{\Omega^2\sin^2\phi}{\epsilon\mu a^2 - \Omega^2}\right]\sin^2\phi \\ &= \cos^2\phi + \left[\frac{n^2(1 - \beta^2)}{n^2 - \beta^2}\right]\sin^2\phi. \end{aligned} \tag{23}$$

Also, from Eq. (18c), we find

$$d = [(n^2 - \beta^2)/c^2(1 - \beta^2)], \tag{24}$$

and hence Eq. (21) becomes

$$\frac{\partial^2 G}{\partial x^2}+\frac{1}{\kappa^2}\frac{\partial^2 G}{\partial y_1^2}-\frac{1}{c^2}\frac{1}{\gamma^2}\frac{\partial^2 G}{\partial t_1^2}$$
$$=\delta(x-x')\delta(y_1-At'-y')\delta(t_1-t'), \quad (25)$$

where we let

$$\frac{1}{\kappa^2}=\cos^2\phi+\left[\frac{n^2(1-\beta^2)}{n^2-\beta^2}\right]\sin^2\phi$$
$$=\frac{n^2(1-\beta^2)+(n^2-1)\beta^2\cos^2\phi}{n^2-\beta^2}, \quad (26a)$$

$$\frac{1}{\gamma^2}=\frac{n^2-\beta^2}{1-\beta^2}. \quad (26b)$$

Because we assume $n>1$ and because $\beta<1$, κ^2, and γ^2 are always positive, and we may define

$$y_0=\kappa y_1 \quad (\kappa>0), \quad (27a)$$
$$t_0=\gamma t_1 \quad (\gamma>0), \quad (27b)$$

and then Eq. (25) becomes

$$\frac{\partial^2 G}{\partial x^2}+\frac{\partial^2 G}{\partial y_0^2}-\frac{1}{c^2}\frac{\partial^2 G}{\partial t_0^2}$$
$$=\kappa\gamma\delta(x-x')\delta(y_0-\kappa At'-\kappa y')\delta(t_0-\gamma t'). \quad (28)$$

We want the solution to Eq. (17) which satisfies the causality condition, viz., it is zero before $t=t'$. It is not difficult to check that this boundary condition carries over to the same causality condition for Eq. (28), namely, G must be zero for $t_0<\gamma t'$. Hence Eq. (28) for G and the boundary condition satisfied by G are formally identical with the problem of the two-dimensional free-space Green's function[5] and we may write down the solution to Eq. (28) immediately:

$$G=\begin{cases}-\dfrac{\kappa\gamma c}{2\pi\sqrt{c^2\tau^2-P^2}}, & P<c\tau,\\ 0, & P>c\tau,\end{cases} \quad (29)$$

where

$$\tau=t_0-\gamma t', \quad (30a)$$
$$P=[(x-x')^2+(y_0-\kappa At'-\kappa y')^2]^{\frac{1}{2}}. \quad (30b)$$

By substituting back for t_0, y_0, κ, γ, and A by means of Eqs. (19), (22), (26), and (27), this result is found to be

$$G=\begin{cases}-\dfrac{\kappa}{2\pi}\left\{(t-t')^2-\left[\dfrac{1-(\beta/n)^2}{1-\beta^2}\right]\left(\dfrac{R_0}{v_0}\right)^2\right\}^{-\frac{1}{2}}, & \\ \qquad (t-t')>\left[\dfrac{1-(\beta/n)^2}{1-\beta^2}\right]\dfrac{R_0}{v_0}, & (31)\\ 0, \qquad (t-t')<\left[\dfrac{1-(\beta/n)^2}{1-\beta^2}\right]\dfrac{R_0}{v_0}, & \end{cases}$$

where

$$R_0=\left\{(x-x')^2+\kappa^2\left[y-y'-\frac{(n^2-1)\beta c}{n^2-\beta^2}\sin\phi(t-t')\right]^2\right\}^{\frac{1}{2}}, \quad (32a)$$

$$v_0=(\mu\epsilon)^{-\frac{1}{2}}=(c/n). \quad (32b)$$

v_0 is the phase velocity of a wave in the medium, as measured from a reference frame attached to the medium.

The interpretation of Eq. (31) is similar to that of the three-dimensional time-dependent Green's function.[3] The field is nonzero only inside an elliptical cylinder parallel to the z axis, with center at the point

$$x=x', \quad y=y'+\frac{(n^2-1)\beta c}{n^2-\beta^2}\sin\phi(t-t'). \quad (33)$$

The outer limit of the elliptical cylinder in the xy plane is defined by

$$R_0=\left[\frac{1-\beta^2}{1-(\beta/n)^2}\right]v_0(t-t'). \quad (34)$$

By manipulating Eqs. (32) and (34), one can show that, if the velocity of the medium is small, the elliptical cylinder propagates away from the source point in all directions in the xy plane, as shown in Fig. 2. If the velocity is very high, the cylinder propagates away from the source on one side only, remaining at all times inside and tangent to a wedge-shaped region defined by half-angle θ_0, where

$$\cos\theta_0=\left[1-\frac{1-\beta^2}{(n^2-1)\beta^2\sin^2\phi}\right]^{\frac{1}{2}} \quad (35)$$

as shown in Fig. 3. In this case the source point (x', y') is *outside* the cylinder, and we note that a source at (x', y') can never produce any fields outside the wedge-shaped region.

The critical velocity between these two cases is given by

$$\beta_c=\frac{v_c}{c}=\left[\frac{1}{(n^2-1)\sin^2\phi+1}\right]^{\frac{1}{2}}, \quad (36)$$

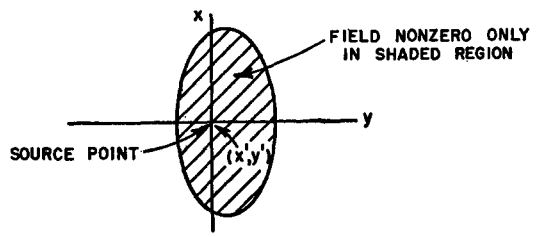


FIG. 2. Low velocity case: $\beta<\beta_c$.

[5] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., Inc., New York, 1953), p. 842.

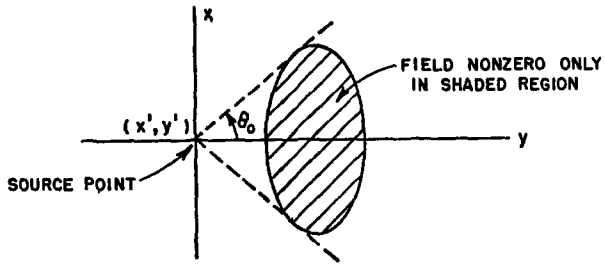


FIG. 3. High velocity case: $\beta > \beta_c$.

and when $v > v_c$ we have the case of two-dimensional Cerenkov radiation.

4. TWO-DIMENSIONAL HARMONIC GREEN'S FUNCTION

For this Green's function we assume an infinite line source parallel to the z axis, the same as above, but now with harmonic time dependence. In Eq. (15), we put

$$\text{source term} = \delta(x - x')\delta(y - y')e^{j\omega t}, \tag{37}$$

where ω is the radian frequency. The solution for G again has no z dependence, and by writing

$$G = ge^{j\omega t}, \tag{38}$$

we find from Eq. (15) that g satisfies

$$\frac{\partial^2 g}{\partial x^2} + b\frac{\partial^2 g}{\partial y^2} - j\omega c\frac{\partial g}{\partial y} + \omega^2 dg = \delta(x - x')\delta(y - y'). \tag{39}$$

To solve for g, we make the transformation

$$g = g_0 e^{\alpha y}, \tag{40a}$$

where α is a constant to be specified. [This α has nothing to do with the dyadic $\boldsymbol{\alpha}$ defined in Eq. (3a).] Then

$$\frac{\partial g}{\partial y} = \frac{\partial g_0}{\partial y}e^{\alpha y} + \alpha g_0 e^{\alpha y}, \tag{40b}$$

$$\frac{\partial^2 g}{\partial y^2} = \frac{\partial^2 g_0}{\partial y^2}e^{\alpha y} + 2\alpha\frac{\partial g_0}{\partial y}e^{\alpha y} + \alpha^2 g_0 e^{\alpha y}, \tag{40c}$$

and Eq. (39) becomes

$$\begin{aligned}\frac{\partial^2 g_0}{\partial x^2} + b\frac{\partial^2 g_0}{\partial y^2} + (2b\alpha - j\omega c)\frac{\partial g_0}{\partial y} &\\ + (b\alpha^2 - j\omega c\alpha + \omega^2 d)g_0 &\\ = e^{-\alpha y}\delta(x - x')\delta(y - y') &\\ = e^{-\alpha y'}\delta(x - x')\delta(y - y'). &\end{aligned} \tag{41}$$

Hence, we may eliminate the first derivative by choosing

$$\begin{aligned}\alpha &= \frac{j\omega c}{2b} \\ &= \frac{j\omega\Omega\sin\phi}{a\cos^2\phi + \sin^2\phi} \\ &= \frac{j\omega(n^2 - 1)\beta\sin\phi}{c(1 - \beta^2)\cos^2\phi + c(1 - n^2\beta^2)\sin^2\phi}.\end{aligned} \tag{42}$$

The coefficient of g_0 in Eq. (41) becomes

$$\begin{aligned}b\alpha^2 - j\omega c\alpha + \omega^2 d &\\ = \omega^2 d - \frac{(j\omega c)^2}{4b} &\\ = \omega^2\epsilon\mu a - \frac{\omega^2\Omega^2}{a} + \frac{\omega^2\Omega^2\sin^2\phi}{a^2[\cos^2\phi + (1/a)\sin^2\phi]} &\\ = \frac{\omega^2}{c^2}\left[\frac{n^2(1 - \beta^2) + \beta^2(n^2 - 1)\cos^2\phi}{1 - n^2\beta^2 + \beta^2(n^2 - 1)\cos^2\phi}\right] \equiv k_1^3, &\end{aligned} \tag{43}$$

and Eq. (41) is

$$\frac{\partial^2 g_0}{\partial x^2} + b\frac{\partial^2 g_0}{\partial y^2} + k_1^2 g_0 = e^{-\alpha y'}\delta(x - x')\delta(y - y'). \tag{44}$$

The above transformation is analogous to the one used by Tai[1] in the three-dimensional case, and is closely related to the affine transformation in Eq. (19) for the time-dependent case.

Before solving this equation, we must examine the signs of the coefficients b and k_1^2. From Eqs. (3b) and (18a) we find

$$b = \frac{1 - \beta^2[(n^2 - 1)\sin^2\phi + 1]}{1 - \beta^2}, \tag{45}$$

so b is positive for

$$\beta < \beta_c = [(n^2 - 1)\sin^2\phi + 1]^{-\frac{1}{2}} \tag{46}$$

and negative for $\beta > \beta_c$. Similarly, from Eq. (43), we have

$$k_1^2 = \frac{\omega^2}{c^2}\left[\frac{n^2(1 - \beta^2) + \beta^2(n^2 - 1)\cos^2\phi}{1 - \beta^2[(n^2 - 1)\sin^2\phi + 1]}\right], \tag{47}$$

so k_1^2 also changes sign at $\beta = \beta_c$, being positive for $\beta < \beta_c$ and negative for $\beta > \beta_c$. It is obvious from the discussion of the time-dependent Green's function above that we would expect the behavior of the harmonic solution to change completely at the critical velocity $v_c = c\beta_c$. Hence it is not surprising that the coefficients b and k_1^2 change sign at this velocity.

Knowing the signs of b and k_1^2, we can now discuss the solution of Eq. (44). We have:

Case I: $\beta < \beta_c$. Then $b > 0$, $k_1^2 > 0$, and if we let

$$y_0 = b^{-\frac{1}{2}}y \quad (b^{\frac{1}{2}} > 0), \tag{48}$$

we have

$$\frac{\partial^2 g_0}{\partial x^2} + \frac{\partial^2 g_0}{\partial y^2} + k_1^2 g_0 = \frac{1}{b^{\frac{1}{2}}} e^{-\alpha y'} \delta(x - x') \delta\left(y_0 - \frac{y'}{b^{\frac{1}{2}}}\right). \tag{49}$$

From the form of the time-dependent Green's function for $\beta < \beta_c$, it is clear that waves propagate away from the source in all directions, and hence the boundary condition for Eq. (49) is that g_0 must represent outward-going waves at infinity. Therefore the problem of solving Eq. (49) is identical with the problem of the ordinary two-dimensional harmonic Green's function; from the known solution[6] we have immediately

$$g_0 = \frac{j}{4(b^{\frac{1}{2}})} e^{-\alpha y'} H_0^{(2)}(k_1 \rho), \tag{50a}$$

where

$$\rho = [(x - x')^2 + (y_0 - y/b^{\frac{1}{2}})^2]^{\frac{1}{2}}, \tag{50b}$$

and where $H_0^{(2)}$ is the Hankel function of the second kind. Finally, substituting back for y_0 and b, and using Eq. (40a) to find g, we have

$$g = \frac{j e^{\alpha(y - y')}}{4[\cos^2 \phi + (1/a) \sin^2 \phi]^{\frac{1}{2}}} H_0^{(2)}(k_1 \rho), \tag{51a}$$

where

$$\rho = \left[(x - x')^2 + \frac{a}{a \cos^2 \phi + \sin^2 \phi} (y - y')^2\right]^{\frac{1}{2}}, \tag{51b}$$

where α is given in Eq. (42) and k_1 in Eq. (43) ($k_1 > 0$).

Case II: $\beta > \beta_c$. Then $b < 0$, $k_1^2 < 0$. If we define

$$k_2^2 = -k_1^2 \qquad (k_2 > 0) \tag{52a}$$

and

$$k_3^2 = -(1/b) \quad (k_3 > 0), \tag{52b}$$

then Eq. (44) is

$$\frac{\partial^2 g_0}{\partial x^2} - \frac{1}{k_3^2} \frac{\partial^2 g_0}{\partial y^2} - k_2^2 g_0 = e^{-\alpha y'} \delta(x - x') \delta(y - y'), \tag{53}$$

which is a one-dimensional Klein–Gordon equation. The boundary condition for g_0 may be obtained from our previous remarks about the time-dependent Green's function. It is clear from Fig. 3 that the fields produced by a harmonic source at (x', y') are nonzero only inside the wedge-shaped region of half-angle θ_0.

[6] Reference 5, p. 811. We have the opposite time convention from that of Morse and Feshbach.

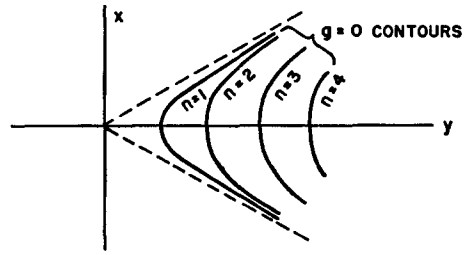


FIG. 4. $g = 0$ contours in xy plane.

Since

$$\tan \theta_0 = \frac{(1 - \cos^2 \theta_0)^{\frac{1}{2}}}{\cos \theta_0} = \left[\frac{1 - \beta^2}{n^2 \beta^2 - 1 - \beta^2 (n^2 - 1) \cos^2 \phi}\right]^{\frac{1}{2}} = k_3, \tag{54}$$

g_0 is zero everywhere except in the region $|x - x'| \leq k_3(y - y')$. The solution of Eq. (53) subject to this boundary condition is mathematically identical to the problem of solving an ordinary one-dimensional Klein–Gordon equation

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \eta^2 \psi = \delta(x - x') \delta(t - t'), \tag{55}$$

subject to the causality condition, namely,

$$\psi \equiv 0 \quad \text{for} \quad |x - x'| > c(t - t'). \tag{56}$$

Since Eq. (55) has the solution[7]

$$\psi = -(c/2) J_0\{\eta[c^2(t - t')^2 - (x - x')^2]^{\frac{1}{2}}\} \times u[(t - t') - |x - x'|/c], \tag{57}$$

where J_0 is the Bessel function of zero order and $u(\xi)$ is a unit step function at $\xi = 0$, we find

$$g_0 = -e^{-\alpha y'} (k_3/2) J_0\{k_2[k_3^2 (y - y')^2 - (x - x')^2]^{\frac{1}{2}}\} \times u[y - y' - |x - x'|/k_3], \tag{58}$$

or finally, from Eq. (40a),

$$g = -\tfrac{1}{2} e^{\alpha(y - y')} k_3 J_0\{k_2[k_3^2 (y - y')^2 - (x - x')^2]^{\frac{1}{2}}\} \times u[y - y' - |x - x'|/k_3]. \tag{59}$$

It is interesting to note that this Green's function, unlike the harmonic Green's function for the three-dimensional case,[1] is finite everywhere in the xy plane, even on the wedge surface at angle θ_0. It is, of course, discontinuous at the wedge, so the fields, which are obtained from derivatives of g, have singularities there.

[7] This result is easily derived using the method shown in Ref. 5, pp. 854–857. The integral which must be evaluated to find ψ in Eq. (57) is equal to one-half of the integral for $h(R, \tau)$ given in Eq. (7.3.37) of Morse and Feshbach.

Furthermore, g is zero whenever

$$k_2[k_3^2(y-y')^2-(x-x')^2]^{\frac{1}{2}}=x_n, \quad n=1,2,\cdots, \tag{60}$$

where x_n, $n=1,2,\cdots$, are the zeros of the Bessel function

$$J_0(x_n)=0. \tag{61}$$

That is, g is zero on a family of hyperbolas in the xy plane, as shown in Fig. 4.

5. ONE-DIMENSIONAL TIME-DEPENDENT GREEN'S FUNCTION

For this case we assume a source term in Eq. (15) of the form

$$\text{source term}=\delta(y-y')\delta(t-t'), \tag{62}$$

i.e., an infinite plane source parallel to the xz plane. (There is no need to rotate the velocity vector of the medium into a more arbitrary direction than that shown in Fig. 1. For any velocity of the medium and orientation of the source plane, coordinate axes may always be chosen so the source plane is parallel to the xz plane and the velocity vector lies in the yz plane.) For such a source term, the solution for G has no x or z dependence, so Eq. (15) simplifies to

$$b\frac{\partial^2 G}{\partial y^2}-c\frac{\partial^2 G}{\partial y\partial t}-d\frac{\partial^2 G}{\partial t^2}=\delta(y-y')\delta(t-t'), \tag{63}$$

where b, c, and d are defined in Eqs. (18). To solve Eq. (63), we may use the same affine transformation as was used above to solve Eq. (17). Repeating the steps of Eqs. (19)–(26), we find that the Green's function must satisfy

$$\frac{1}{\kappa^2}\frac{\partial^2 G}{\partial y_1^2}-\frac{1}{c^2}\frac{1}{\gamma^2}\frac{\partial^2 G}{\partial t_1^2}=\delta(y_1-At'-y')\delta(t_1-t'). \tag{64}$$

Since, as before, κ^2 and γ^2 are always positive, we may solve this equation by comparison with the standard one-dimensional time-dependent Green's function equation. Again making the substitution of Eq. (27), we find

$$\frac{\partial^2 G}{\partial y_0^2}-\frac{1}{c^2}\frac{\partial^2 G}{\partial t_0^2}=\kappa\gamma\delta(y_0-\kappa At'-\kappa y')\delta(t_0-\gamma t'), \tag{65}$$

subject to the causality boundary condition, and hence[8]

$$G=-\frac{\kappa\gamma c}{2}u[c(t_0-\gamma t')-|y_0-\kappa At'-\kappa y'|], \tag{66}$$

[8] Reference 5, p. 843.

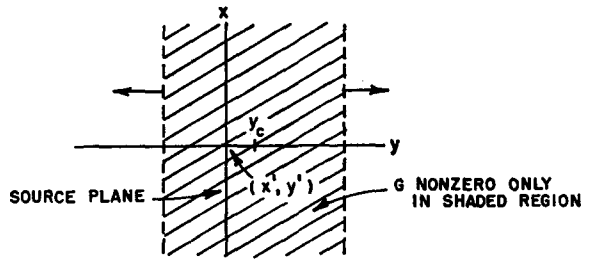


FIG. 5. Low-velocity case: $\beta<\beta_c$.

where $u(\xi)$ is a unit step function. Substituting back for y_0, t_0, and A, finally gives

$$G=-\frac{\kappa\gamma c}{2}u\left[\frac{c\gamma}{\kappa}(t-t')-\left|y-y'-\frac{(n^2-1)\beta c}{n^2-\beta^2}\sin\phi(t-t')\right|\right]. \tag{67}$$

The meaning of this result is easy to see. The Green's function consists of a square pulse of amplitude $-(\kappa\gamma c/2)$ centered at the point

$$y_c=y'+\frac{(n^2-1)\beta c}{n^2-\beta^2}\sin\phi(t-t'), \tag{68}$$

and extending over the region

$$\frac{(n^2-1)\beta c}{n^2-\beta^2}\sin\phi(t-t')-\frac{c\gamma}{\kappa}(t-t') \le |y-y'| \le \frac{(n^2-1)\beta c}{n^2-\beta^2}\sin\phi(t-t')+\frac{c\gamma}{\kappa}(t-t'). \tag{69}$$

Outside this region, G is zero. As time progresses, the edges of the pulse propagate away from the center of the pulse. If the velocity of the medium is low, the pulse edges travel away from the source plane on both sides, as shown in Fig. 5. On the other hand, if the velocity of the medium is high enough, the pulse lies entirely on one side of the source plane, as shown in Fig. 6. The transition velocity between these two cases is found by setting the left-hand side of Eq. (69)

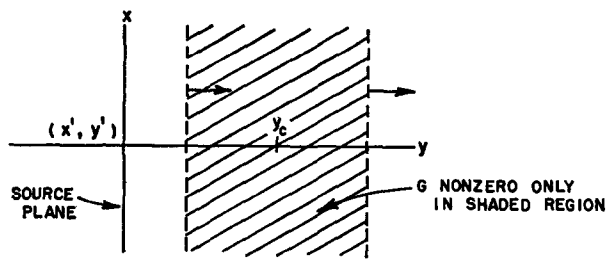


FIG. 6. High-velocity case: $\beta>\beta_c$.

to zero. This gives

$$\beta_c = [(n^2 - 1)\sin^2\phi + 1]^{-\frac{1}{2}}, \tag{70}$$

which is the same as in Eq. (46). For $\beta < \beta_c$ we have the result shown in Fig. 5, and for $\beta > \beta_c$ we have the case in Fig. 6. The case $\beta > \beta_c$ might be called one-dimensional Cerenkov radiation.

6. ONE-DIMENSIONAL HARMONIC GREEN'S FUNCTION

For this Green's function we choose an infinite plane source the same as in the previous section except with harmonic time dependence:

$$\text{source term} = \delta(y - y')e^{j\omega t}. \tag{71}$$

G again has no x or z dependence, and making the substitution of Eq. (38), we find from Eq. (15) that g satisfies

$$b\frac{\partial^2 g}{\partial y^2} - j\omega c\frac{\partial g}{\partial y} + \omega^2 dg = \delta(y - y'). \tag{72}$$

We solve this equation with the same transformation as in Eqs. (40)–(43). Thus we find g_0 satisfies

$$b\frac{\partial^2 g_0}{\partial y^2} + k_1^2 g_0 = e^{-\alpha y'}\delta(y - y'). \tag{73}$$

Since both b and k_1^2 change sign at $\beta = \beta_c$, as discussed below Eq. (44), we again have two cases:

Case I: $\beta < \beta_c$. Then $b > 0$ and $k_1^2 > 0$, and making the substitution of Eq. (48) gives

$$\frac{\partial^2 g_0}{\partial y_0^2} + k_1^2 g_0 = \frac{1}{b^{\frac{1}{2}}} e^{-\alpha y'}\delta\left(y_0 - \frac{y'}{b^{\frac{1}{2}}}\right). \tag{74}$$

From the time-dependent Green's function in Eq. (67) and the fact that for the case $\beta < \beta_c$ the wave fronts propagate away from the source plane in both directions, it is clear that the boundary condition for Eq. (74) is that g_0 represent outward-going waves at $y_0 \to \pm\infty$. Hence the solution is

$$g_0 = -\frac{e^{-\alpha y'}}{2jk_1(b^{\frac{1}{2}})}\exp\left(-jk_1\left|y_0 - \frac{y'}{b^{\frac{1}{2}}}\right|\right), \tag{75}$$

or, after substituting for y_0 and using Eq. (40a) for g, we have

$$g = -\frac{e^{\alpha(y-y')}}{2jk_1(b^{\frac{1}{2}})}\exp\left(-j\frac{k_1}{b^{\frac{1}{2}}}|y - y'|\right). \tag{76}$$

Case II: $\beta > \beta_c$. Then $b < 0$, $k_1^2 < 0$, and we may again make the substitutions of Eq. (52). Equation (73) then becomes ($k_3 > 0$, $k_2 > 0$):

$$\frac{1}{k_3^2}\frac{\partial^2 g_0}{\partial y^2} + k_2^2 g_0 = -e^{-\alpha y'}\delta(y - y') \tag{77}$$

or, with

$$y_2 = k_3 y, \tag{78}$$

$$\frac{\partial^2 g_0}{\partial y_2^2} + k_2^2 g_0 = -k_3 e^{-\alpha y'}\delta(y_2 - k_3 y'), \tag{79}$$

the same type of equation obtained in Case I. For this equation, however, we must be careful to note that the boundary condition is different than for Case I. From the behavior of the time-dependent Green's function for $\beta > \beta_c$, it is clear that the boundary condition for Eq. (79) is

$$g_0 \equiv 0, \quad \text{for} \quad y_2 < k_3 y' \quad (\text{for } y > y'), \tag{80a}$$

since a source at $y = y'$ cannot excite any fields to the left of the source plane (see Fig. 6). Since $\partial^2 g_0/\partial y_2^2$ has a delta-function singularity at $y_2 = k_3 y'$, g_0 itself is continuous at $y_2 = k_3 y'$, and hence it is also required that

$$g_0 = 0, \quad \text{at} \quad y_2 = k_3 y'. \tag{80b}$$

The solution for Eq. (79) subject to Eqs. (80) is

$$g_0 = \begin{cases} -\dfrac{k_3 e^{-\alpha y'}}{k_2}\sin k_2(y_2 - k_3 y'), & y_2 > k_3 y', \\ 0, & y_2 < k_3 y', \end{cases} \tag{81}$$

or, from Eqs. (40a), (52), and (78),

$$g = \begin{cases} -\dfrac{e^{\alpha(y-y')}}{(bk_1^2)^{\frac{1}{2}}}\sin\sqrt{\dfrac{k_1^2}{b}}\,(y - y'), & y > y', \\ 0, & y < y'. \end{cases} \tag{82}$$

Here we again find that g is finite everywhere, and also has zeros at

$$(k_1^2/b)^{\frac{1}{2}}(y - y') = n\pi, \quad n = 0, 1, 2, \cdots, \tag{83}$$

i.e., on planes spaced equidistant from the source plane.

CONCLUSION

The one- and two-dimensional Green's functions have been found and are given in Eqs. (31), (51), (59), (67), (76), and (82). As in the three-dimensional case, it is found that the transition between ordinary radiation and Cerenkov radiation occurs when the velocity is high enough that the time-dependent Green's function lies entirely on one side of the source. In that case, a harmonic source produces wavefronts which interfere, thus causing the harmonic Green's function to have a standing-wave behavior. Below the Cerenkov velocity, the harmonic Green's function exhibits the usual traveling-wave behavior.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Diffusion Length and Criticality Problems in Two- and Three-Dimensional, One-Speed Neutron Transport Theory. I. Rectangular Coordinates

M. M. R. WILLIAMS
Nuclear Engineering Department, Queen Mary College, University of London

(Received 9 September 1967)

Several problems in linear transport theory have been solved involving more than one dimension. We have considered the diffusion-length problem in a slab system and have studied how the flux in the transverse direction varies as a function of distance from the source. The solution is found to consist of an asymptotic term, composed of a finite number of harmonics, plus a transient which is nonseparable in the x and z coordinates. When the slab is sufficiently thin only the transient term survives and the concept of a diffusion length loses its value. A method for overcoming this difficulty is presented. In another problem, the thickness of the slab is allowed to become semi-infinite and the effect of decay in the z direction on the emergent angular distribution and surface flux are assessed. By approximating the flux in the y and z directions by a function of the form $\exp\{iB_y y + iB_z z\}$ and solving the resulting one-dimensional problem in the x direction exactly, it has been possible to obtain a statement of the critical conditions in a bare rectangular parallelepiped system. The application of this method to the diffusion length problem in a system with a rectangular cross section is discussed. Finally, by comparison with the exact solution for the slab, we estimate the accuracy of a reduced Boltzmann equation deduced by the author in a previous publication.

1. INTRODUCTION

In the field of one-dimensional, one-speed linear transport theory, nearly every conceivable problem has been or can be solved exactly. The Wiener–Hopf technique,[1,2] the oldest method for dealing with these problems, has been employed in the past, although more recently the method of singular eigenfunctions devised by Case[3] has assumed importance. We will not list the many papers dealing with applications of Case's technique, but simply refer the interested reader to the publications of the past five years appearing in *Annals of Physics (New York)* and *Journal of Mathematical Physics.*

Once we leave the one-dimensional problem, and consider simultaneously the effect of particle-density variation in another direction, the problem becomes exceedingly complex. Progress has been made, however, by attempting to reduce the three-dimensional transport equation to a pseudo-one-dimensional one.[4] This reduction may be achieved if the spatial variation in two of the three orthogonal directions is approximated by a known function; for example, by the asymptotic flux $\exp\{i\boldsymbol{\beta}\cdot\mathbf{r}\}$ or $\exp\{-\nu z\}$. The Boltzmann equation may then be reduced to one in a single space coordinate, the effect of the other directions entering via a single parameter, e.g., a transverse buckling or a spatial-decay constant.

In a previous publication,[4] we have considered this problem quite generally, including the effect of energy transfer. As an example we obtained an exact solution for the one-speed Milne problem in a finite prism, the transverse leakage being allowed for by a buckling. This problem has also recently been solved by Kaper[5] who used the equation derived in Ref. 4 but employed the techniques of the theory of generalized analytic functions. Mitsis[6] has solved critical problems in infinite cylinders and spheres, and some problems in two dimensions have also been considered by Smith[7,8] in connection with certain radiative-transfer situations.

Progress has also been made in the general energy-dependent problem (and as a special case the one-speed problem) of the diffusion-length experiment.[9] It has been demonstrated that a diffusion-length, i.e., a discrete eigenvalue, ceases to exist when the transverse dimensions of the system are sufficiently small.

Apart from one or two special problems of point sources,[10,11] the discussion given above appears to summarize the progress to date on solutions of the transport equation in more than one dimension. In this paper, we propose to extend the work described above to include criticality and diffusion-length problems in rectangular parallelepipeds. The criticality problem is solved approximately by assuming an asymptotic $\exp\{i\boldsymbol{\beta}\cdot\mathbf{r}\}$ variation in two of the orthogonal directions and treating the other one exactly

[1] G. Placzek and W. Seidel, Phys. Rev. **72**, 550 (1947).
[2] S. Frankel and S. Goldberg, Atomic Energy Commission Document 2056 (1945).
[3] K. M. Case, Ann. Phys. (N.Y.) **9**, 305 (1967).
[4] M. M. R. Williams, Nukleonik **9**, 305 (1967).
[5] H. Kaper, "Elementary Solutions of the Reduced Three-Dimensional Transport Equation." Submitted to J. Math. Phys. (to be published).
[6] G. Mitsis, ANL 6787, 1963.
[7] M. G. Smith, Proc. Cambridge Phil. Soc. **60**, 909 (1964).
[8] M. G. Smith and G. E. Hunt, Proc. Cambridge Phil. Soc. **63**, 209 (1967).
[9] M. M. R. Williams, "Symposium on Neutron Thermalization and Reactor Spectra," Ann Arbor, Mich. (1967), paper SM 96/3.
[10] J. P. Elliott, Proc. Roy. Soc. (London) **A228**, 424 (1955).
[11] R. C. Erdmann, J. Math. Phys. **8**, 1040 (1967).

(this will also include the pulsed-neutron problem). The diffusion-length problem will contain no approximations for the case of a slab with exponential decay parallel to its surfaces, but to obtain tractable expressions we shall have to make one or two reasonable assumptions. Our method of attack is based upon a replication property discussed by Case[12] of the integral form of the neutron-transport equation which reduces the problem to the solution of certain singular integral equations.

2. THE BASIC EQUATION AND PROBLEMS UNDER DISCUSSION

The neutron density $\rho(\mathbf{r})$ in a non-re-entrant body is described by the following integral equation[13]:

$$\rho(\mathbf{r}) = \int d\mathbf{r}' \frac{\exp\{-|\mathbf{r}-\mathbf{r}'|\}}{4\pi|\mathbf{r}-\mathbf{r}'|^2}\{c\rho(\mathbf{r}') + S(\mathbf{r}')\}, \quad (1)$$

where distance is measured in units of total mean free path l and $c = l/l_s$, l_s being the scattering mean free path. $S(\mathbf{r})$ is a source.

Consider now Eq. (1) when the body is infinite, in, say, the z direction, and of arbitrary shape in the x and y directions. Let the source $S(\mathbf{r})$ be of the form

$$S(\mathbf{r}) = Q(x, y)\delta(z)$$

and take the infinite-medium Fourier transform of (1). Then,[6]

$$\Psi(\xi, k) = \frac{1}{2\pi}\int_A d\xi' \int_0^1 K_0\left[\frac{|\xi-\xi'|}{\nu}(1+\nu^2k^2)^{\frac{1}{2}}\right]\frac{d\nu}{\nu^2} \times [c\Psi(\xi', k) + Q(\xi')], \quad (2)$$

where ξ is a vector in the x–y plane and

$$\Psi(\xi, k) = \int_{-\infty}^{\infty} e^{-ikz}\rho(\mathbf{r})\, dz. \quad (3)$$

$K_0(x)$ is a modified Bessel function.

We consider Eq. (2) in two specific geometries: (A) for a slab of width $2a$, infinite in the y direction (i.e., no spatial variation in the y direction); (B) a rectangular parellepiped of width $2a$ in the x direction, and in which the flux shape in the other two directions is approximated by the form $\exp\{iB_y y + iB_z z\}$.

In Case (A), we can integrate Eq. (2) over $y(-\infty, \infty)$ to obtain

$$\Psi(x, k) = \frac{1}{2}\int_{-a}^{a} dx' \int_0^1 \frac{\exp\{-(|x-x'|/\nu)(1+\nu^2k^2)^{\frac{1}{2}}\}}{(1+\nu^2k^2)^{\frac{1}{2}}}\frac{d\nu}{\nu} \times [c\Psi(x', k) + Q(x')]. \quad (4)$$

[12] K. M. Case, *Developments in Transport Theory*, E. Inönu and P. Zweifel, Eds. (Academic Press Inc., New York 1967).

[13] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).

In Case (B), which we consider only from the point of view of criticality, the equation becomes homogeneous but otherwise it is identical to Eq. (4) with $k^2 = B_y^2 + B_z^2 = \beta^2$, B_y^2 and B_z^2 being the bucklings in the y and z directions, respectively; i.e.,

$$B_y = \pi/[2(b + y_0)], \quad (5)$$

$$B_z = \pi/[2(c + z_0)], \quad (6)$$

where $2b$ and $2c$ are the widths of the system in the y and z direction, respectively. y_0 and z_0 are the corresponding extrapolated end points, about which we shall have more to say later.

We also consider briefly the diffusion-length problem in which the flux distribution is treated exactly in the x direction and approximated by the function $\exp\{iB_y y\}$ in the y direction.

3. REPLICATION PROPERTY OF EQUATION (4)

From the symmetry of the problem it would seem profitable to seek solutions of Eq. (4) as a superposition of terms of the type

$$\cosh\left\{\frac{x}{\tilde{\nu}}(1+\tilde{\nu}^2k^2)^{\frac{1}{2}}\right\}. \quad (7)$$

Operating on this function with the integral operator defined by Eq. (4), we see that

$$\frac{c}{2}\int_{-a}^{a} dx' \int_0^1 \frac{\exp\{-|x-x'|(1+\nu^2k^2)^{\frac{1}{2}}/\nu\}}{(1+\nu^2k^2)^{\frac{1}{2}}}\frac{d\nu}{\nu} \times \cosh\left\{\frac{x'}{\tilde{\nu}}(1+\tilde{\nu}^2k^2)^{\frac{1}{2}}\right\}$$

$$= \cosh\left(\frac{\tilde{\alpha}}{\tilde{\nu}}x\right)c\int_0^1 \frac{d\nu}{\nu^2(\alpha^2/\nu^2 - \tilde{\alpha}^2/\tilde{\nu}^2)} - \frac{c}{2}\int_0^1 \frac{d\nu}{\nu\alpha}\, e^{-a\alpha/\nu}\cosh\left(\frac{\alpha x}{\nu}\right) \times \left[\frac{e^{-a\tilde{\alpha}/\tilde{\nu}}}{\alpha/\nu + \tilde{\alpha}/\tilde{\nu}} + \frac{e^{a\tilde{\alpha}/\tilde{\nu}}}{\alpha/\nu - \tilde{\alpha}/\tilde{\nu}}\right], \quad (8)$$

where $\alpha = (1+\nu^2k^2)^{\frac{1}{2}}$ and $\tilde{\alpha} = (1+\tilde{\nu}^2k^2)^{\frac{1}{2}}$.

Thus, the action of the operator reproduces the original function, plus a distribution of it over ν between (0, 1).

If we take the spatial variation of the source in the x direction to be a constant, i.e., $Q(x) = Q_0$, then we can look for a general solution to Eq. (4) in the form

$$\Psi(x, k) = S + a_0\cosh\left\{\frac{x}{\nu_0}(1+\nu_0^2k^2)^{\frac{1}{2}}\right\} + \int_0^1 \phi(\nu)\cosh\left\{\frac{x}{\nu}(1+\nu^2k^2)^{\frac{1}{2}}\right\} d\nu. \quad (9)$$

Inserting this expression into (4) and collecting coefficients of $\cosh(x\alpha/\nu)$, $\cosh(x\alpha_0/\nu_0)$, and the constant term, we find

$$S = \frac{Q_0(1/k)\tan^{-1}k}{1-(c/k)\tan^{-1}k}, \tag{10}$$

$$\Lambda(\nu_0) = 1 - c\int_0^1 \frac{d\nu}{\nu^2(\alpha^2/\nu^2 - \alpha_0^2/\nu_0^2)} = 0$$

$$= 1 - c\nu_0^2\int_0^1 \frac{d\nu}{\nu_0^2 - \nu^2}, \tag{11}$$

which defines ν_0, and finally the integral equation for $\phi(\nu)$:

$$\Lambda(\nu)\psi(\nu) + \frac{c}{2}\int_0^1 \psi(\nu')\frac{1}{\nu\alpha}\left(\frac{\alpha}{\nu} - \frac{\alpha'}{\nu'}\right)^{-1} d\nu'$$
$$+ \frac{c}{2}\int_0^1 \psi(\nu')\frac{1}{\nu\alpha}\left(\frac{\alpha}{\nu} + \frac{\alpha'}{\nu'}\right)^{-1} e^{-2a\alpha'/\nu'}\, d\nu'$$
$$+ \left[ca_0\Big/\nu^2\left(\frac{\alpha^2}{\nu^2} - \frac{\alpha_0^2}{\nu_0^2}\right)\right]\left[\cosh\left(\frac{a\alpha_0}{\nu_0}\right) + \frac{\nu\alpha_0}{\nu_0\alpha}\sinh\left(\frac{a\alpha_0}{\nu_0}\right)\right]$$
$$+ \left[Q_0\Big/(1+\nu^2k^2)\left(1 - \frac{c}{k}\tan^{-1}k\right)\right] = 0, \tag{12}$$

where $\phi(\nu) = e^{-a\alpha/\nu}\psi(\nu)$.

Let us now define a new variable

$$\omega = \frac{\nu(1+k^2)^{\frac{1}{2}}}{(1+\nu^2k^2)^{\frac{1}{2}}} \quad \text{or} \quad \nu = \omega g(\omega) \tag{13}$$

in terms of which Eq. (12) becomes

$$\Lambda(\omega g)\Phi(\omega) + \frac{c}{2}\int_0^1 \frac{\omega' g(\omega')\Phi(\omega')\, d\omega'}{\omega' - \omega}$$
$$= -\frac{c}{2}\int_0^1 \frac{\omega' g(\omega')\Phi(\omega')\exp\{-2a(1+k^2)^{\frac{1}{2}}/\omega'\}}{\omega' + \omega}\, d\omega'$$
$$- \frac{ca_0\omega_0}{2(1+k^2)}\left[\frac{\exp\{a(1+k^2)^{\frac{1}{2}}/\omega_0\}}{\omega_0 - \omega} + \frac{\exp\{-a(1+k^2)^{\frac{1}{2}}/\omega_0\}}{\omega_0 + \omega}\right]$$
$$- \left[Q_0\Big/(1+k^2)\left(1 - \frac{c}{k}\tan^{-1}k\right)\right]. \tag{14}$$

In Eq. (14) we have defined

$$\Lambda(\omega g) = 1 - \frac{c}{2}\int_{-1}^1 \frac{\omega g(\omega')\, d\omega'}{\omega - \omega'}, \tag{15}$$

$$g(\omega) = \{1 + k^2(1-\omega^2)\}^{-\frac{1}{2}},$$
$$\Phi(\omega) = g(\omega)\psi[\omega g(\omega)], \tag{16}$$

and

$$d\nu = (1+k^2)g^3(\omega)\, d\omega.$$

We also note that k (the Fourier-transform variable) is, in general, a complex number, thus ω also is complex. In the case under consideration, $\Psi(x, k)$ is analytic in the k plane when it is cut from i to $i\infty$ and $-i$ to $-i\infty$, apart from some poles at $k = \pm ik_n$ $(n = 0, 1, \cdots, N)$, where $k_n < 1$. For $|k| < 1$, therefore, ω is entirely real and varies from zero to ν as k_n varies from unity to zero. In our analysis we assume that ω is real and deal with the analytic continuation to complex ω (or k) later.

Equation (14) is a standard singular integral equation which we write as

$$\Lambda_0(\omega)\Phi_0(\omega) + \frac{c}{2}\int_0^1 \frac{\omega'\Phi_0(\omega')\, d\omega'}{\omega' - \omega} = \psi'(\omega), \tag{17}$$

where $\Phi_0 = g\Phi$ and $\Lambda_0 = \Lambda/g$. ψ' represents the inhomogeneous term on the right-hand side of (14).

4. SOLUTION OF THE INTEGRAL EQUATION

Following the method used by Case,[3] the solution of (17) may be written down directly as

$$\Phi_0(\omega) = \Lambda_0(\omega)g_0(c, \omega)\psi'(\omega)$$
$$- \frac{1}{X^-(\omega)[\Lambda_0(\omega) + \frac{1}{2}c\pi i\omega]}\int_0^1 \frac{\gamma(\omega')\psi'(\omega')\, d\omega'}{\omega' - \omega}, \tag{18}$$

together with the side condition

$$\int_0^1 \gamma(\psi)\psi'(\omega)\, d\omega = 0. \tag{19}$$

The following notation is employed:

$$g_0(c, \omega) = \{\Lambda_0^2(\omega) + \tfrac{1}{4}c^2\pi^2\omega^2\}^{-1}, \tag{20}$$

$$\gamma(\omega) = \frac{\frac{1}{2}c\omega X^-(\omega)}{\Lambda_0(\omega) - \frac{1}{2}c\pi i\omega}, \tag{21}$$

$$X(\omega) = (1-\omega)^{-1}\exp\{\Gamma(\omega)\}, \tag{22}$$

$$\Gamma(\omega) = \frac{1}{2\pi i}\int_0^1 \frac{\ln G(\omega')\, d\omega'}{\omega' - \omega}, \tag{23}$$

$$G(\omega) = \frac{\Lambda_0(\omega) + \frac{1}{2}c\pi i\omega}{\Lambda_0(\omega) - \frac{1}{2}c\pi i\omega}. \tag{24}$$

$X^-(\omega)$ is the value of $X(\omega)$ as we approach the cut (0, 1) from below.

It may readily be shown that

$$\frac{1}{X^-(\omega)[\Lambda_0(\omega) + \frac{1}{2}c\pi i\omega]} = \frac{\Lambda_0(\omega)g_0(c, \omega)}{X(\omega)} \tag{25}$$

and hence that Eq. (18) reduces to

$$(1+k^2)\Phi_0(\omega)g(\omega) = -\tfrac{1}{2}cg_0(c,\omega)X(-\omega)[1-(c/k)\tan^{-1}k](\omega_0^2-\omega^2) \times\left[a_0\omega_0\left\{\frac{X(\omega_0)}{\omega_0-\omega}\exp\left(a(1+k^2)^{\frac{1}{2}}/\omega_0\right)+\frac{X(-\omega_0)}{\omega_0+\omega}\exp\left(-a(1+k^2)^{\frac{1}{2}}/\omega_0\right)\right\}+(1+k^2)\int_0^1\frac{\omega' X(-\omega')\Phi(\omega')}{\omega'+\omega}\times\exp\left\{-2a(1+k^2)^{\frac{1}{2}}/\omega'\right\}d\omega'\right], \quad (26)$$

with $\omega = \nu(1+k^2)^{\frac{1}{2}}/(1+\nu^2k^2)^{\frac{1}{2}}$. Equation (19) simplifies to

$$\frac{Q_0}{1-(c/k)\tan^{-1}k} + \tfrac{1}{2}ca_0\omega_0[-X(\omega_0)\exp\{a(1+k^2)^{\frac{1}{2}}/\omega_0\} + X(-\omega_0)\exp\{-a(1+k^2)^{\frac{1}{2}}/\omega_0\}] + \tfrac{1}{2}c(1+k^2)\int_0^1\Phi_0(\omega')\omega' X(-\omega')\times\exp\left\{\frac{-2a(1+k^2)^{\frac{1}{2}}}{\omega'}\right\}d\omega'. \quad (27)$$

Substituting for ω, (27) becomes

$$\frac{Q_0}{1-(c/k)\tan^{-1}k} + \frac{ca_0(1+\nu_0^2k^2)^{\frac{1}{2}}}{[1-(c/k)\tan^{-1}k]^{\frac{1}{2}}}\left[\frac{1-\nu_0^2(1-c)}{2(\nu_0^2-1)}\right]^{\frac{1}{2}}\times\cosh\{(a+x_0)(1+\nu_0^2k^2)^{\frac{1}{2}}/\nu_0\} + \frac{c}{2}\int_0^1\psi(\nu')\frac{\nu'(1+k^2)}{(1+\nu'^2k^2)}X(-\omega')\times\exp\{-2a(1+\nu'^2k^2)^{\frac{1}{2}}/\nu'\}d\nu', \quad (27')$$

where we have defined x_0 as

$$-X(\omega_0)/X(-\omega_0) = \exp\{2x_0(1+k^2)^{\frac{1}{2}}/\omega_0\}. \quad (28)$$

In arriving at Eq. (27′) we have used the relation

$$X(\omega)X(-\omega) = \left[1-\tfrac{1}{2}c\omega g(\omega)\log\left\{\frac{\omega g(\omega)+1}{\omega g(\omega)-1}\right\}\right]\times\left[\left(1-\frac{c}{k}\tan^{-1}k\right)(\omega_0^2-\omega^2)\right]^{-1}$$

and the limit of this as $\omega\to\omega_0$, i.e.,

$$X(\omega_0)X(-\omega_0) = -\frac{(1+\nu_0^2k^2)^2}{[1-(c/k)\tan^{-1}k]\nu_0^2(1+k^2)}\left[\frac{1-\nu_0^2(1-c)}{2(\nu_0^2-1)}\right]. \quad (29)$$

Equation (9) together with Eqs. (26) and (27) constitute the Fourier transform of the neutron density in the z direction.

5. INVERSION OF THE TRANSFORM

The complete spatial distribution $\rho(x, z)$ is given by

$$\rho(x,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{ikz}\Psi(x,k)\,dk. \quad (30)$$

Inserting Eq. (9), we note the appearance of poles at the roots of $1-(c/k)\tan^{-1}k = 0$, i.e., $k = \pm i/\nu_0$, where

$$1-\tfrac{1}{2}c\nu_0\log\left(\frac{\nu_0+1}{\nu_0-1}\right) = 0.$$

It is readily verified, however, that $\psi(\nu_0)\equiv 0$, and also that

$$\lim_{k\to\pm i/\nu_0}\{S + a_0(k)\cosh(x(1+\nu_0^2k^2)^{\frac{1}{2}}/\nu_0)\} = 0,$$

hence the residue is zero, as we expect, since ν_0 is the diffusion length of the infinite-medium problem and not of the finite medium one considered here.

Direct inversion of $\Psi(x, k)$ is difficult because it is not possible to obtain explicit expressions for a_0 or $\psi(\nu)$. We resort, therefore, to an approximate method and neglect the terms involving $\psi(\nu)$, which are small because of the term $\exp\{-2a(1+\nu^2k^2)^{\frac{1}{2}}/\nu\}$. Such a procedure is not as accurate as in the one-dimensional critical problem. This is due to the fact that k is imaginary, i.e., it is a Fourier-transform variable that runs along the imaginary axis; thus $k = \pm i|k|$. We see that $\Psi(x, k)$ has a finite number of poles $k = \pm ik_n$ ($n = 0, 1, \cdots, N$), where $k_n < 1$. Thus, as $k_n\to 1$, the approximation becomes less accurate, but still valid because the small ν values in the integrand remain effective. For $|k| > 1$, the integral, which becomes multivalued, must be split as follows:

$$\int_0^{|k|^{-1}}\cdots\exp\{-2a(1-\nu^2|k|^2)^{\frac{1}{2}}/\nu\}\,d\nu + \int_{|k|^{-1}}^1\cdots\exp\{-2ai(\nu^2|k|^2-1)^{\frac{1}{2}}/\nu\}\,d\nu.$$

For $|k|$ near unity, the second term is expected to be small and both may still be neglected. For larger values of $|k|$, where the second term may become appreciable, the contribution to the transient part of $\rho(x, z)$ is small anyway. On these nonrigorous, but plausible, arguments we neglect $\psi(\nu)$ entirely. This means, of course, that we cannot obtain information near the surfaces $x = \pm a$, but this defect will be remedied in due course.

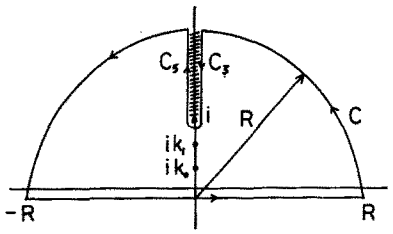


FIG. 1. Singularities in the k plane and inversion contour.

We find now that a_0 is given by (27) with $\psi(\nu) = 0$, and hence

$$\rho(x, z) = \frac{Q_0}{2\pi}\int_{-\infty}^{\infty} e^{ikz}\left\{\frac{(1/k)\tan^{-1}k}{1-(c/k)\tan^{-1}k} - \frac{1}{c(1+\nu_0^2k^2)}\right.$$
$$\times\left[\frac{1+\nu_0^2k^2}{1-(c/k)\tan^{-1}k}\right]^{\frac{1}{2}}\left[\frac{2(\nu_0^2-1)}{1-\nu_0^2(1-c)}\right]^{\frac{1}{2}}$$
$$\left.\times\frac{\cosh\{(x/\nu_0)(1+\nu_0^2k^2)^{\frac{1}{2}}\}}{\cosh\{[(a+x_0/\nu_0)](1+\nu_0^2k^2)^{\frac{1}{2}}\}}\right\}dk. \quad (31)$$

We are interested in $\rho(x, z)$ for $z > 0$, hence we look for singularities in the upper half of the k plane. The integrand in Eq. (31) is single-valued in the upper half-plane if a cut is introduced from i to $i\infty$. In addition, there are some simple poles given by the roots of the equation

$$\cosh\{[(a+x_0)/\nu_0](1+\nu_0^2k^2)^{\frac{1}{2}}\} = 0. \quad (32)$$

The general picture of these singularities is shown in Fig. 1. The roots of Eq. (32) are given by

$$[(a+x_0)/\nu_0](1+\nu_0^2k^2)^{\frac{1}{2}} = i(n+\tfrac{1}{2})\pi. \quad (33)$$

Setting $k = ik_n$, we find

$$k_n^2 = 1/\nu_0^2 + B_n^2, \quad (34)$$

where

$$B_n = \tfrac{1}{2}(2n+1)\pi[a + x_0(ik_n)]^{-1}. \quad (35)$$

$x_0(ik_n)$ is the extrapolated end point associated with the transverse buckling B_n^2 of the nth mode; it is defined by Eq. (28) and may be written out explicitly as

$$x_0(k) = \frac{\nu_0}{2(1+\nu_0^2k^2)^{\frac{1}{2}}}\log\left[\frac{\nu_0(1+k^2)^{\frac{1}{2}} + (1+\nu_0^2k^2)^{\frac{1}{2}}}{\nu_0(1+k^2)^{\frac{1}{2}} - (1+\nu_0^2k^2)^{\frac{1}{2}}}\right]$$
$$-\frac{\nu_0^2}{\pi}\int_0^1\frac{\vartheta(\nu)\,d\nu}{(1+\nu^2k^2)^{\frac{1}{2}}(\nu_0^2-\nu^2)}, \quad (36a)$$

with

$$\vartheta(\nu) = \tan^{-1}\left\{\frac{c\pi\nu/2}{1-\frac{1}{2}c\nu\log[(1+\nu)/(1-\nu)]}\right\}.$$

From the numerical point of view it is more convenient to integrate (36a) by parts to get

$$x_0(k) = \frac{\nu_0}{4(1+\nu_0^2k^2)^{\frac{1}{2}}}\log\left[\frac{\nu_0(1+k^2)^{\frac{1}{2}} + (1+\nu_0^2k^2)^{\frac{1}{2}}}{\nu_0(1+k^2)^{\frac{1}{2}} - (1+\nu_0^2k^2)^{\frac{1}{2}}}\right]$$
$$+\frac{\nu_0^2}{\pi}\int_0^1\frac{\vartheta_0(\nu)\,d\nu}{(1+\nu^2k^2)^{\frac{1}{2}}(\nu_0^2-\nu^2)}, \quad (36b)$$

where

$$\vartheta_0(\nu) = \tan^{-1}\left\{\frac{2}{\pi c\nu} - \frac{1}{\pi}\log\left(\frac{1+\nu}{1-\nu}\right)\right\}.$$

Hence

$$x_0(ik_n) = \frac{\nu_0}{2(\nu_0^2k_n^2-1)^{\frac{1}{2}}}\tan^{-1}\left\{\frac{(\nu_0^2k_n^2-1)^{\frac{1}{2}}}{\nu_0(1-k_n^2)^{\frac{1}{2}}}\right\}$$
$$+\frac{\nu_0^2}{\pi}\int_0^1\frac{\vartheta_0(\nu)\,d\nu}{(1-\nu^2k_n^2)^{\frac{1}{2}}(\nu_0^2-\nu^2)}. \quad (37)$$

It has been shown[9] that discrete decay constants of the diffusion length problem must be less than unity. Furthermore, Eqs. (34) and (37) constitute a relationship between k_n and $2a$, and therefore it may be confirmed from these equations that, for a given a, we find a finite number (N) of roots k_n less than unity. As a decreases, the number N decreases, until finally only a single discrete root k_0 remains. On decreasing a still further, even this root disappears, which is in agreement with the theorem proved in Ref. 9.

The exact value a_N^* at which the Nth decay constant disappears is given by (34) and (37) with $k_n = 1$. Thus

$$a_N^* = (N+\tfrac{1}{2})\pi\nu_0(\nu_0^2-1)^{-\frac{1}{2}} - x_0(i), \quad (38)$$

where

$$x_0(i) = \frac{\pi\nu_0}{4(\nu_0^2-1)^{\frac{1}{2}}} + \frac{\nu_0^2}{\pi}\int_0^1\frac{\vartheta_0(\nu)\,d\nu}{(1-\nu^2)^{\frac{1}{2}}(\nu_0^2-\nu^2)}. \quad (39)$$

As an example, Fig. 2 shows $2a_0^*$ as a function of $\kappa = \nu_0^{-1}$. It is clear that at small values of absorption, the "critical thickness" for the existence of a diffusion length is about one mean free path; with increasing absorption, a_0^* increases accordingly.

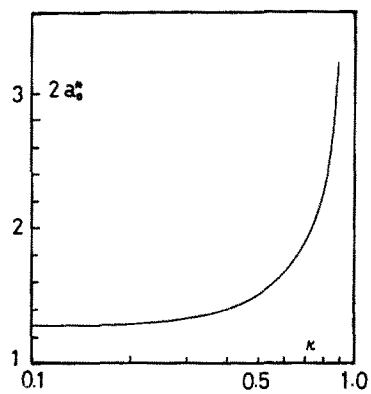


FIG. 2. Slab width at which the fundamental mode disappears.

It should be remembered that Eq. (34) is only approximate since we have neglected $\psi(\nu)$. The question arises then: What is the effect on a_0^* of this approximation? To assess this error we have found values of a_0^* directly from the homogeneous part of the integral equation (4). An eigenvalue problem for k_n results and is solved by a variational technique. It is found that Eq. (38) is extremely accurate and therefore we can have some confidence in our results.

Associated with the poles k_n are the residues. These may be obtained directly from (31); thus denoting the contribution to $\rho(x, z)$ from the residues by $\rho_{\text{asy}}(x, z)$ we find

$$\rho_{\text{asy}}(x, z) = \frac{Q_0}{c}\left[\frac{2(\nu_0^2 - 1)}{1 - \nu_0^2(1 - c)}\right]^{\frac{1}{2}} \sum_{n=0}^{N} \frac{\exp(-k_n z)}{B_n D_n (1 + B_n^2 \nu_0^2)^{\frac{1}{2}}} \times \frac{\cos(B_n x)}{\left\{\frac{c}{2k_n}\log\left(\frac{1 + k_n}{1 - k_n}\right) - 1\right\}^{\frac{1}{2}}}, \quad (40)$$

where

$$D_n = [B_n \chi_n + (n + \tfrac{1}{2})\pi B_n^{-2}]$$

and

$$\chi_n = \frac{i}{k_n}\frac{dx_0(k)}{dk}\bigg|_{ik_n}.$$

From Eq. (36) we find

$$\chi_n = \frac{\nu_0^2}{2(1 - k_n^2)^{\frac{1}{2}}(\nu_0^2 k_n^2 - 1)} - \frac{\nu_0^3}{2(\nu_0^2 k_n^2 - 1)^{\frac{1}{2}}}\tan^{-1}\left\{\frac{(\nu_0^2 k_n^2 - 1)^{\frac{1}{2}}}{\nu_0(1 - k_n^2)^{\frac{1}{2}}}\right\} + \frac{\nu_0^2}{\pi}\int_0^1 \frac{\mu^2 \vartheta_0(\mu)\, d\mu}{(1 - k_n^2\mu^2)^{\frac{1}{2}}(\nu_0^2 - \mu^2)}. \quad (41)$$

We note that for $a < a_0^*$, $\rho_{\text{asy}}(x, z) \equiv 0$.

The contribution to $\rho(x, z)$ from the integration around the contour is more difficult. However, by considering the closed contour C in Fig. 1, and allowing $R \to \infty$, we find that

$$\rho(x, z) = \rho_{\text{asy}}(x, z) + \rho_{\text{trans}}(x, z), \quad (42)$$

where

$$\rho_{\text{trans}}(x, z) = \frac{1}{2\pi}\int_i^{i\infty} e^{ikz}(C_3 - C_5)\, dk. \quad (43)$$

C_3 and C_5 represent symbolically the values of the integrand in Eq. (31) along C_3 and C_5. Several of the terms in Eq. (31) take on different values at each side of the cut. For example, setting $k = is$, we find along C_3

$$1 - \frac{c}{k}\tan^{-1} k = 1 - \frac{c}{s}\tanh\left(\frac{1}{s}\right) + \frac{i\pi}{2s} = 1 - c\mathfrak{X} + \frac{i\pi}{2s}, \quad (44)$$

$$(1 + \nu_0^2 k^2)^{\frac{1}{2}} = i(\nu_0^2 s^2 - 1)^{\frac{1}{2}}, \quad (45)$$

$$x_0(k) = x_1(s) - ix_2(s), \quad (46)$$

$$\cosh\left\{\frac{(a + x_0)}{\nu_0}(1 + \nu_0^2 k^2)^{\frac{1}{2}}\right\} = \cos\left\{\frac{(\nu_0^2 s^2 - 1)^{\frac{1}{2}}}{\nu_0}(a + x_1 - ix_2)\right\} = P + iQ, \quad (47)$$

where

$$P = \cos A \cosh B,$$
$$Q = \sin A \sinh B,$$
$$A = \frac{1}{\nu_0}(\nu_0^2 s^2 - 1)^{\frac{1}{2}}(a + x_1),$$
$$B = \frac{1}{\nu_0}(\nu_0^2 s^2 - 1)^{\frac{1}{2}} x_2.$$

Along C_5 these functions take on their complex conjugate values.

Inserting (44)–(47) into (43), and using considerable algebra, we obtain

$$\rho_{\text{trans}}(x, z) = \frac{1}{2}\int_1^{\infty}\frac{ds}{s}\frac{e^{-sz}}{(1 - c\mathfrak{X})^2 + (\pi^2 c^2/4s^2)} \times [1 + I(s)\cos\{(\nu_0^2 s^2 - 1)^{\frac{1}{2}} x/\nu_0\}], \quad (48)$$

where

$$I(s) = \frac{2s}{c\pi}\left[\frac{\nu_0^2 - 1}{1 - \nu_0^2(1 - c)}\right]^{\frac{1}{2}} \times \frac{\{Q(s)[T(s)^{\frac{1}{2}} - 1 + c\mathfrak{X}]^{\frac{1}{2}} - P(s)[T(s)^{\frac{1}{2}} + 1 - c\mathfrak{X}]^{\frac{1}{2}}\}}{(\nu_0^2 s^2 - 1)^{\frac{1}{2}}(P^2(s) + Q^2(s))} \quad (49)$$

and

$$T(s) = (1 - c\mathfrak{X})^2 + \pi^2 c^2/4s^2.$$

More explicitly

$$P(s) = \cos\left\{\frac{1}{\nu_0}(\nu_0^2 s^2 - 1)^{\frac{1}{2}}(a + x_1)\right\} \times \cosh\left\{(\nu_0^2 s^2 - 1)^{\frac{1}{2}}\frac{x_2}{\nu_0}\right\},$$

$$Q(s) = \sin\left\{\frac{1}{\nu_0}(\nu_0^2 s^2 - 1)^{\frac{1}{2}}(a + x_1)\right\} \times \sinh\left\{(\nu_0^2 s^2 - 1)^{\frac{1}{2}}\frac{x_2}{\nu_0}\right\},$$

$$P^2 + Q^2 = \cos^2\left\{\frac{1}{\nu_0}(\nu_0^2 s^2 - 1)^{\frac{1}{2}}(a + x_1)\right\} + \sinh\left\{(\nu_0^2 s^2 - 1)^{\frac{1}{2}}\frac{x_2}{\nu_0}\right\}. \quad (50)$$

We note that ρ_{trans} reduces to the well-known infinite-medium result when $a \to \infty$.

The rather complicated expressions arising in Eq. (49) are due mainly to the square-root term in Eq. (31), and also because $x_0(k)$ is discontinuous across the cut. Let us now demonstrate how $x_1(s)$ and $x_2(s)$ are to be obtained in terms of real integrals.

Along C_3 we find that

$$x_0(k) = \frac{\nu_0}{2i(\nu_0^2 s^2 - 1)^{\frac{1}{2}}} \times \log\left[\frac{e^{i\pi/2}[\nu_0(s^2-1)^{\frac{1}{2}} + (\nu_0^2 s^2 - 1)^{\frac{1}{2}}]}{(\nu_0^2 - 1)^{\frac{1}{2}}}\right] + \frac{\nu_0^2}{\pi}\int_0^1 \vartheta_0(\mu)\, d\mu \times [(\nu_0^2 - \mu^2)(1/\mu + s)^{\frac{1}{2}}(s - 1/\mu)^{\frac{1}{2}} e^{i\pi/2}]^{-1}. \quad (51)$$

Simplifying the first term, and splitting up the integral term into $\mu \gtrless 1/s$, we obtain

$$x_0(k) = \frac{\pi\nu_0}{4(s^2\nu_0^2 - 1)^{\frac{1}{2}}} + \frac{\nu_0^2}{\pi}\int_0^{1/s} \frac{\vartheta_0(\mu)\, d\mu}{(\nu_0^2 - \mu^2)(1 - s^2\mu^2)^{\frac{1}{2}}} - i\left\{\frac{\nu_0}{4(\nu_0^2 s^2 - 1)^{\frac{1}{2}}} \log\left[\frac{(\nu_0^2 s^2 - 1)^{\frac{1}{2}} + \nu_0(s^2-1)^{\frac{1}{2}}}{(\nu_0^2 s^2 - 1)^{\frac{1}{2}} - \nu_0(s^2-1)^{\frac{1}{2}}}\right] + \frac{\nu_0^2}{\pi}\int_{1/s}^1 \frac{\vartheta_0(\mu)\, d\mu}{(\nu_0^2 - \mu^2)(s^2\mu^2 - 1)^{\frac{1}{2}}}\right\} = x_1(s) - i x_2(s). \quad (52)$$

We note that $x_2(s)$ increases uniformly from zero at $s = 1$, goes through a maximum, and eventually tends to zero as $s \to \infty$. $x_1(s)$ on the other hand is equal to $x_0(i)$ at $s = 1$, then it goes through a maximum in the neighborhood of $s = 1$, eventually tending to zero as $s \to \infty$. Table I lists some values of $x_2(s)$ for two values of κ. In Fig. 3 we have plotted $x_0(ik_0)$ from $k_0 = 0.91$ to 1.0. $x_0(ik_0)$ appears to be an extremely weak function of k_0 below 0.9. In the same figure we also give $x_1(s)$ for $s = 1.0$ to 2.0.

TABLE I. Values of $x_2(s)$ for $\kappa = 0.1, 0.5$.

s	$\kappa = 0.1$	$\kappa = 0.5$
1.0	0	0
1.01	0.0310	0.0423
1.03	0.0658	0.0893
1.05	0.0932	0.1259
1.08	0.1278	0.1709
1.1	0.1478	0.1964
1.2	0.2256	0.2909
1.5	0.3477	0.4218
1.8	0.3975	0.4653
2.0	0.4126	0.4753

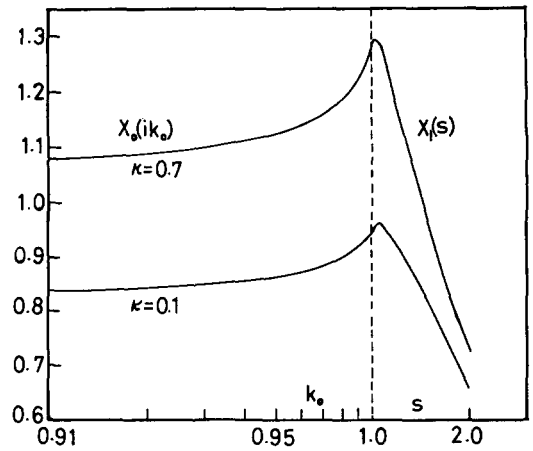


FIG. 3. Extrapolated endpoint as a function of buckling and absorption. $x_1(s)$ is the continuation for $k_0 > 1$.

6. SURFACE DISTRIBUTIONS

The expression for $\rho(x, z)$ given by Eq. (31) is an asymptotic one, in the sense that it neglects the effects of transients induced by the surfaces $x = \pm a$. For example, $\rho(\pm a, z)$ is not given at all accurately by Eq. (31). We can, however, remedy this and obtain an expression for $\rho(\pm a, z)$, which is accurate to $O\{\exp(-2a)\}$, directly from the integral equation (26).

Let us consider Eq. (9) rewritten in terms of the ω variable

$$\Psi(x, k) = \frac{Q_0(1/k)\tan^{-1}k}{1 - (c/k)\tan^{-1}k} + a_0(k)\cosh\{x(1+k^2)^{\frac{1}{2}}/\omega_0\} + (1+k^2) \times \int_0^1 \exp\{-a(1+k^2)^{\frac{1}{2}}/\omega\}\Phi_0(\omega)g(\omega) \times \cosh\left\{\frac{x}{\omega}(1+k^2)^{\frac{1}{2}}\right\} d\omega. \quad (53)$$

Thus the surface distribution at $x = -a$ is given by

$$\Psi(-a, k) = \frac{Q_0(1/k)\tan^{-1}k}{1 - (c/k)\tan^{-1}k} + a_0(k)\cosh\{a(1+k^2)^{\frac{1}{2}}/\omega_0\} + \tfrac{1}{2}(1+k^2)\int_0^1 \Phi_0(\omega)g(\omega) \times [1 + \exp(-2a(1+k^2)^{\frac{1}{2}}/\omega)]\, d\omega. \quad (54)$$

Neglecting the terms $O\{\exp(-2a\cdots)\}$ in (54) and (26) and substituting $\Phi_0 g$ from (26) into (54), we obtain

$$\Psi(-a, k) = \frac{Q_0(1/k)\tan^{-1}k}{1 - (c/k)\tan^{-1}k} + a_0(k)\cosh\{a(1+k^2)^{\frac{1}{2}}/\omega_0\} - \tfrac{1}{2}a_0[T(\omega_0)\exp\{a(1+k^2)^{\frac{1}{2}}/\omega_0\} + T(-\omega_0)\exp\{-a(1+k^2)^{\frac{1}{2}}/\omega_0\}], \quad (55)$$

where

$$T(\omega_0) = X(\omega_0)\frac{c}{2}\int_0^1 \frac{\omega_0\, d\omega}{(\omega_0 - \omega)X^-(\omega)[\Lambda_0(\omega) + \frac{1}{2}ic\pi\omega]}. \tag{56}$$

Using the integral,[14]

$$\frac{c}{2}X(-\omega)\int_0^1 \frac{t\, dt}{(t + \omega)X^-(t)[\Lambda_0(t) + \frac{1}{2}i\pi ct]} = X(-\omega)(1 + \omega) - 1, \tag{57}$$

we readily find that

$$T(\omega_0) = 1 - X(\omega_0)\left[\frac{1}{X(0)} - \omega_0\right]. \tag{58}$$

Then, with a_0 from (27), we obtain the very simple result

$$\Psi(-a, k) = \frac{Q_0}{c}\left[\frac{\tanh\{(1 + \nu_0^2k^2)^{\frac{1}{2}}(a + x_0)/\nu_0\}}{\{1 - (c/k)\tan^{-1}k\}^{\frac{1}{2}}} - 1\right]. \tag{59}$$

Observe that the zeros of the denominator at $\pm i/\nu_0$ are exactly cancelled by those in the numerator.

We do not invert the transform (59) completely but simply give the asymptotic part, viz.,

$$\rho_{\text{asy}}(-a, z) = \frac{Q_0}{c}\sum_{n=0}^{N}\frac{(-)^n}{D_n(1 + \nu_0^2B_n^2)^{\frac{1}{2}}} \times \exp(-k_nz)\left\{\frac{c}{2k_n}\log\left(\frac{1 + k_n}{1 - k_n}\right) - 1\right\}^{-\frac{1}{2}}. \tag{60}$$

We also note that for $k = 0$ and $a \to \infty$, Eq. (59) reduces to the well-known half-space constant-source value of

$$\Psi(-\infty) = (Q_0/c)[(1 - c)^{-\frac{1}{2}} - 1].$$

This completes our mathematical analysis of the diffusion-length problem in a slab.

7. PHYSICAL CONSIDERATIONS

It is of some interest to consider Eq. (42) from a physical point of view. If, for example, our problem represented some experimental situation to measure the diffusion length ν_0, we would make measurements at distances from the source such that ρ_{trans} was negligible. Then, allowing the higher spatial harmonics in ρ_{asy} to decay, our final density distribution would be of the form

$$\rho(x, z) \sim \frac{Q_0}{c}\left[\frac{2(\nu_0^2 - 1)}{1 - \nu_0^2(1 - c)}\right]^{\frac{1}{2}} \times \frac{\exp(-k_0z)\cos(B_0x)}{B_0(1 + \nu_0^2B_0^2)^{\frac{1}{2}}D_0\left\{\frac{c}{2k_0}\log\left(\frac{1 + k_0}{1 - k_0}\right) - 1\right\}^{\frac{1}{2}}}, \tag{61}$$

where

$$k_0^2 = 1/\nu_0^2 + B_0^2$$

and

$$B_0 = \pi/[2(a + x_0(ik_0))]. \tag{62}$$

Experimentally, we would measure k_0, guess a value of $x_0(ik_0)$ and hence obtain ν_0. It should be noted, however, that $x_0(ik_0)$ is itself a function of k_0 and so a little iterating is necessary. This technique is a fairly standard one for diffusion-length measurements. What happens, however, when $a < a_0^*$? In this case $\rho_{\text{asy}} = 0$ and $\rho(x, z) = \rho_{\text{trans}}(x, z)$. The question remains as to how ν_0 is to be extracted from this nonexponential function.

A possible way out of this dilemma stems from the observation that, for $a < a_0^*$, $I(s) \gg 1$ and also, for a not very much less than a_0^*, $I(s)$ has a *sharp* peak near $s = 1$. This sharp peak arises from the fact that we can find a root $s = s_0$ of $P(s) = 0$, and at the same time $Q(s_0)$ is small. Thus, in the neighborhood of s_0, $I(s)$ can be approximated by the following expression:

$$I(s) \simeq \vartheta^*(s_0)\,\gamma(s_0)/[(s - s_0)^2 + \gamma^2(s_0)], \tag{63}$$

where $\vartheta^*(s_0)$ is a constant and we have expanded

$$P^2(s) \simeq \alpha^2(s_0)(s - s_0)^2$$

and set

$$\gamma(s_0) = Q(s_0)/\alpha(s_0).$$

s_0 is the root of

$$\cos\{[(\nu_0^2s^2 - 1)^{\frac{1}{2}}/\nu_0](a + x_1(s))\} = 0$$

or

$$s_0^2 = 1/\nu_0^2 + \beta_N^2, \tag{64}$$

where

$$\beta_N = \frac{(2N + 1)\pi}{2(a + x_1(s_0))}. \tag{65}$$

We are interested in the smallest β_N, viz., β_0. It is clear, therefore, that for $s_0 \simeq 1$, but slightly greater than it, we can write

$$\rho_{\text{trans}}(x, z) \simeq \vartheta_1^*(s_0)\gamma(s_0) \times \int_1^\infty \frac{\exp(-sz)\cos\{(\nu_0^2s^2 - 1)x/\nu_0\}\, ds}{(s - s_0)^2 + \gamma^2(s_0)}, \tag{66}$$

where $\vartheta_1^* = \vartheta^*/2s_0 \cdot I(s_0)$.

If γ is sufficiently small, the Lorentzian function can be approximated by a delta function $\delta(s - s_0)$ and

$$\rho_{\text{trans}}(x, z) \simeq \text{const}\, e^{-s_0z}\cos(\beta_0x). \tag{67}$$

Equation (67) is similar in form to the asymptotic solution given by Eq. (61). The main difference is that

[14] A. Leonard and T. W. Mulliken, "Solutions to the Criticality problem for Spheres and Slabs," Rand Corp. Report RM-3256-PR, 1962.

β_0 has replaced the conventional buckling B_0. Again, the only difference between β_0 and B_0 is $x_1(s_0)$ and $x_0(ik_0)$, respectively. From Fig. 3 we note that $x_1(s_0)$ is the apparent continuation of $x_0(ik_0)$ for values of $k_0 > 1$ (whether it is the analytic continuation is not yet clear). It would appear then that, even if the dimensions of the system are too small to sustain an asymptotic solution, the neutron density maintains a pseudo-asymptotic character but with a modified transverse buckling. Of course, this approximation cannot be carried too far and breaks down if γ is not sufficiently small. In that case, the flux distribution does not fall off exponentially from the source, but progressively changes in shape in the x direction as z increases.

It is worth noting that these remarks about the pseudo-asymptotic nature of the transient term also apply for energy-dependent problems where the value of $2a_0^*$ may be significant, e.g., of the order of one meter.[9] The concept of a pseudo-asymptotic distribution is also discussed by Corngold and Durgun[15] in connection with the pulsed-neutron experiment.

8. ANGULAR DISTRIBUTIONS

The angular distribution in our slab system may be obtained from the original integro-differential form of the Boltzmann equation, i.e.,

$$\left[\eta\frac{\partial}{\partial x}+(1-\eta^2)^{\frac{1}{2}}\left\{\cos\Theta\frac{\partial}{\partial y}+\sin\Theta\frac{\partial}{\partial z}\right\}+1\right]f(\mathbf{r},\boldsymbol{\Omega})=(1/4\pi)[c\rho(\mathbf{r})+S(\mathbf{r})],\quad(68)$$

where $\rho(\mathbf{r})=\int d\boldsymbol{\Omega}f(\mathbf{r},\boldsymbol{\Omega})$, $f(\mathbf{r},\boldsymbol{\Omega})$ being the angular flux defined more explicitly as

$$f(\mathbf{r},\boldsymbol{\Omega})=f(x,y,z,\eta,\Theta).$$

The angular coordinates Θ and $\eta=\cos\vartheta$ are shown in Fig. 4.

Remembering that there is no variation in the y direction and taking the Fourier transform in the z direction, Eq. (68) becomes

$$\left[\eta\frac{\partial}{\partial x}+1+ik(1-\eta^2)^{\frac{1}{2}}\sin\Theta\right]\tilde{f}(x,k,\eta,\Theta)=(1/4\pi)(c\Psi(x,k)+Q_0),\quad(69)$$

where

$$\tilde{f}(x,k,\cdots)=\int_{-\infty}^{\infty}e^{-ikz}f(x,z,\cdots)\,dz.\quad(70)$$

Equation (69) may be integrated with respect to x,

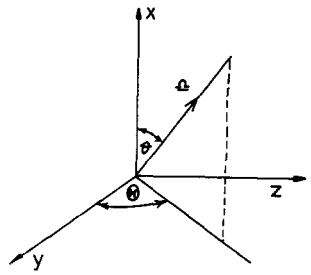


FIG. 4. Angular coordinates of the neutron motion.

when we find

$$\tilde{f}(x,k,\eta,\Theta)=\frac{1}{4\pi\eta}\int_{-a}^{x}dx'\times\exp\left\{-\frac{(x-x')}{\eta}[1+ik(1-\eta^2)^{\frac{1}{2}}\sin\Theta]\right\}\times[c\Psi(x',k)+Q_0],\quad\eta>0,\quad\text{all }\Theta,\quad(71)$$

and

$$\tilde{f}(x,k,\eta,\Theta)=-\frac{1}{4\pi\eta}\int_{x}^{a}dx'\exp\left\{\frac{(x'-x)}{\eta}[1+ik(1-\eta^2)^{\frac{1}{2}}\sin\Theta]\right\}\times[c\Psi(x',k)+Q_0],\quad\eta<0,\quad\text{all }\Theta.\quad(72)$$

Setting $c\Psi(x,k)+Q_0=\tilde{\Phi}(x,k)$, we see from the definition of the Fourier transform that

$$f(x,z,\eta,\Theta)=\frac{1}{4\pi\eta}\int_{-a}^{x}dx'\times e^{-(x-x')/\eta}\Phi\left[x',z-\frac{(x-x')}{\eta}(1-\eta^2)^{\frac{1}{2}}\sin\Theta\right],\quad\eta>0,\quad\text{all }\Theta,\quad(73)$$

and a similar expression for $\eta<0$. $\Phi(x,z)$ is the Fourier inverse of $\tilde{\Phi}(x,k)$. We see therefore that the angular distribution may be expressed directly in terms of $\rho(x,z)$, which we have already obtained.

The angular distribution at the surface $x=a$ comes from Eq. (73) and is

$$f(a,z,\eta,\Theta)=\frac{1}{4\pi\eta}\int_{-a}^{a}e^{-(a-x)/\eta}\Phi\left[x,z-\frac{(a-x)}{\eta}(1-\eta^2)^{\frac{1}{2}}\sin\Theta\right]dx,\quad\eta>0,\quad\text{all }\Theta.\quad(74)$$

The angular distribution within the prism may be obtained if Eq. (42) is substituted into (73). Evaluating the integral involving Q_0, we find

$$f(x,z,\eta,\Theta)=f_s(x,z,\eta,\Theta)+f_M(x,z,\eta,\Theta),\quad(75)$$

[15] N. Corngold and K. Durgun, Nucl. Sci. Eng. **29**, 354 (1967).

where f_s is the contribution from the uncollided neutrons coming directly from the source, viz.,

$$f_s(x, z, \eta, \Theta) = \frac{Q_0}{4\pi(1-\eta^2)^{\frac{1}{2}} \sin\Theta} \exp\left\{-\frac{z}{(1-\eta^2)^{\frac{1}{2}} \sin\Theta}\right\},$$
$$0 < z < \frac{(a+x)}{\eta}(1-\eta^2)^{\frac{1}{2}} \sin\Theta,$$
$$= 0, \quad \text{otherwise.} \tag{76}$$

The limit on f_s is a mathematical statement of the physical fact that uncollided neutrons, leaving the source plane at position x and angular coordinates (ϑ, Θ), will not be found for

$$z > (a+x)(1-\eta^2)^{\frac{1}{2}} \sin\Theta/\eta.$$

This value of z is just the distance from the source plane at which the neutron emerges from the body.

f_M is the contribution from multiply-scattered neutrons. Its form is rather complicated but a concise expression can be obtained if we write

$$\rho(x, z) = \sum_{n=0}^{N} A_n e^{-k_n z} \cos B_n x + \int_1^\infty A(s) e^{-sz}[1 + I(s) \cos\beta(s)x]\, ds \tag{77}$$

in an obvious notation. Then we find, for $\eta > 0$ and all Θ, that

$$f_M(x, z, \eta, \Theta) = \frac{c}{4\pi} \sum_{n=0}^{N} \frac{A_n F_n(\eta, \Theta, x) \exp(-k_n z)}{[1 - k_n(1-\eta^2)^{\frac{1}{2}} \sin\Theta]^2 + \eta^2 B_n^2}$$
$$+ \frac{c}{4\pi} \int_1^\infty \frac{A(s) e^{-sz}}{[1 - s(1-\eta^2)^{\frac{1}{2}} \sin\Theta]} \times \left[1 - \exp\left(-\frac{(a+x)}{\eta}\{1 - s(1-\eta^2)^{\frac{1}{2}} \sin\Theta\}\right)\right] ds$$
$$+ \frac{c}{4\pi} \int_1^\infty \frac{A(s) I(s) F_s(\eta, \Theta, x) \exp(-sz)\, ds}{[1 - s(1-\eta^2)^{\frac{1}{2}} \sin\Theta]^2 + \eta^2\beta^2}, \tag{78}$$

where

$$F_n(\eta, \Theta, x) = [1 - k_n(1-\eta^2)^{\frac{1}{2}} \sin\Theta] \cos B_n x + \eta B_n \sin B_n x$$
$$- [(1 - k_n(1-\eta^2)^{\frac{1}{2}} \sin\Theta) \cos B_n a - \eta B_n \sin B_n a]$$
$$\times \exp\left\{-\frac{(a+x)}{\eta}[1 - k_n(1-\eta^2)^{\frac{1}{2}} \sin\Theta]\right\}. \tag{79}$$

For F_s we simply replace k_n by s and B_n by $\beta(s)$ in (79).

When there are no discrete roots the term involving the summation is absent. It may then be verified that the angular distribution, as described by the integral term in (78), exhibits a marked anisotropy along the z axis.

When the fundamental term $n = 0$ exists, and is well established, the angular distribution will be of the following asymptotic form:

$$f_M(x, z, \eta, \Theta) \sim \frac{c}{4\pi} \frac{A_0 e^{-k_0 z} F_0(\eta, \Theta, x)}{[1 - k_0(1-\eta^2)^{\frac{1}{2}} \sin\Theta]^2 + \eta^2 B_0^2}. \tag{80}$$

9. THE HALF-SPACE PROBLEM

If the origin of coordinates is shifted so that $x = 0$ and $x = 2a$ are now the surfaces of our slab, we may obtain the half-space results very easily. The situation is then of a half-space with spatial decay parallel to its free surface.

The form of the flux in this problem is as follows:

$$\Psi(x, k) = S + b_0 \exp(-x\alpha_0/\nu_0) + \int_0^1 \phi(\nu) \exp(-x\alpha/\nu)\, d\nu. \tag{81}$$

Use of the replication property then leads to the following singular integral equation for $\phi(\nu)$:

$$\Phi_0(\omega)\Lambda_0(\omega) + \frac{c}{2}\int_0^1 \frac{\omega'\Phi_0(\omega')\, d\omega'}{\omega' - \omega} = -\frac{Q_0}{(1+k^2)[1 - (c/k)\tan^{-1} k]} - \frac{b_0 c\omega_0}{(1+k^2)2(\omega_0 - \omega)}, \tag{82}$$

where $g^3\phi(\omega g) = \Phi_0$.

This equation may be solved exactly, the result being

$$b_0 = \frac{Q_0}{1 - (c/k)\tan^{-1} k} \frac{1}{c\omega_0 X(\omega_0)} \tag{83}$$

and

$$(1+k^2)\Phi_0(\omega) = -\tfrac{1}{2} Q_0 g_0(c, \omega) X(-\omega)(\omega_0 + \omega). \tag{84}$$

Hence we have

$$\Psi(x, k) = \frac{Q_0 (1/k)\tan^{-1} k}{[1 - (c/k)\tan^{-1} k]} \times \left\{1 + \frac{k}{\tan^{-1} k} \frac{\exp\{-x(1+k^2)^{\frac{1}{2}}/\omega_0\}}{c\omega_0 X(\omega_0)}\right\} - \tfrac{1}{2} Q_0 \int_0^1 g_0(c, \omega) X(-\omega)(\omega_0 + \omega) \times \exp\{-x(1+k^2)^{\frac{1}{2}}/\omega\}\, d\omega. \tag{85}$$

Similarly we can obtain the surface flux

$$\Psi(0, k) = \frac{Q_0}{c}\left[\left\{1 - \frac{c}{k}\tan^{-1} k\right\}^{-\frac{1}{2}} - 1\right], \quad (86)$$

and the emergent angular distribution

$$\tilde{f}(0, k, \eta, \Theta) = \frac{Q_0}{4\pi}\frac{-1}{[1 - (c/k)\tan^{-1} k]} \times \frac{-1}{[\xi\omega_0 + \eta(1 + k^2)^{\frac{1}{2}}]}\frac{1}{X[-(\eta/\xi)(1 + k^2)^{\frac{1}{2}}]}, \quad (87)$$

where $\xi = 1 + ik(1 - \eta^2)^{\frac{1}{2}} \sin\Theta$.

Clearly, as $x \to \infty$, $\Psi(x, k)$ goes over to the well-known infinite-medium result[13] given by the first term in Eq. (85). For x near the surface, this simple expression is modified by leakage. We demonstrate the effect by performing the inversion of $\Psi(0, k)$ which we write in the form

$$\rho(0, z) = \frac{Q_0}{2\pi c}\int_{-\infty}^{\infty} e^{ikz}\left\{\left[\frac{k^2 + \nu_0^{-2}}{1 - (c/k)\tan^{-1} k}\right]^{\frac{1}{2}} \times \frac{1}{(k^2 + \nu_0^{-2})^{\frac{1}{2}}} - 1\right\} dk, \quad (88)$$

where $k = \pm i/\nu_0$ are the roots of

$$1 - (c/k)\tan^{-1} k.$$

ν_0 is, of course, the conventional diffusion length.

The quantity in the square brackets in the integrand of Eq. (88) has no poles or zeros for $-1 < \text{Im}\,(k) < 1$, but, due to the arc tangent, it has branch points at $k = \pm i$. The function may be made single-valued, however, by introducing two cuts in the k plane extending from i to $i\infty$ and $-i$ to $-i\infty$. Let us denote the function by $F(k)$ and its inverse by $f(z)$. The other factor in (88), i.e., $(k^2 + \nu_0^{-2})^{-\frac{1}{2}}$ has the Fourier inverse $K_0(z/\nu_0)$, where $K_0(x)$ is a modified Bessel function. From the convolution theorem, therefore, we may write (88) for $z \neq 0$, as

$$\rho(0, z) = \frac{Q_0}{\pi c}\int_0^{\infty} K_0\left(\frac{z'}{\nu_0}\right)[f(|z + z'|) + f(|z - z'|)]\, dz', \quad (89)$$

where

$$f(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikz}F(k)\, dk. \quad (90)$$

Performing the integration leads to

$$f(z) = \int_1^{\infty} I_0(s)e^{-sz}\, ds, \quad (91)$$

where

$$I_0(s) = \frac{\sqrt{2}}{\pi}\left(s^2 - \frac{1}{\nu_0^2}\right)^{\frac{1}{2}}(1 - c\mathfrak{T} + T^{\frac{1}{2}})^{\frac{1}{2}}T^{-\frac{1}{2}}. \quad (92)$$

Insertion of (91) into (89) leads to

$$\rho(0, z) = \frac{1}{\pi}\int_1^{\infty} I_0(s)\frac{e^{-sz}}{(s^2 - \nu_0^{-2})^{\frac{1}{2}}} \times \log\left\{\frac{s\nu_0 + (s^2\nu_0^2 - 1)^{\frac{1}{2}}}{s\nu_0 - (s^2\nu_0^2 - 1)^{\frac{1}{2}}}\right\} + \frac{2}{\pi}\int_1^{\infty}\frac{\exp(-yz/\nu_0)}{(y^2 - 1)^{\frac{1}{2}}}\int_1^{\infty}\frac{I_0(s)\, ds}{s(s^2 - y^2\nu_0^{-2})}. \quad (93)$$

No simple expression can be obtained for (93). We note, however, that the second term dominates after about one mean free path from the source and decays at least as fast as $\exp(-z/\nu_0)$. It is interesting to note also that, although the total flux far from the free surface (i.e., large x) behaves asymptotically for large z as $\exp(-z/\nu_0)$, the corresponding surface flux does not. Thus the ratio $\rho(x, z)/\rho(0, z)$ does not become independent of z as $z \to \infty$. It must be concluded, therefore, that the nature of the source affects the surface flux (and angular distribution) of the half-space for all z. This is not the case in the finite slab for which the ratio discussed above becomes constant provided k_0 exists.

10. CRITICAL PROBLEM

As we discussed in Sec. 2, the present method can be used to obtain approximate criticality conditions in rectangular systems. The critical problem for a rectangular parallelepiped of sides a, b, c and in which the distribution in the y and z directions can be written $\exp(iB_y y + iB_z z)$ is described by the following eigenvalue problem:

$$\Psi(x, \beta) = \frac{c}{2}\int_{-a}^{a} dx' \times \int_0^1 \frac{\exp\{-|x - x'|(1 + \nu^2\beta^2)/\nu\}}{(1 + \nu^2\beta^2)^{\frac{1}{2}}}\frac{d\nu}{\nu}\Psi(x', \beta), \quad (94)$$

where $\beta^2 = B_y^2 + B_z^2$.

Equation (94) may be solved directly by the method described in Sec. 4. Indeed, the solution of Eq. (14) is simply carried over with $k^2 = \beta^2$ and $Q_0 = 0$. Then the flux distribution is given by

$$\Psi(x, \beta) = a_0 \cos\left\{x\left(\frac{1}{\nu_0^2} - \beta^2\right)^{\frac{1}{2}}\right\} + \int_0^1 \phi(\nu)\cosh\{x(1 + \nu^2\beta^2)^{\frac{1}{2}}/\nu\}\, d\nu, \quad (95)$$

where we have taken $\nu_0 = i\,|\nu_0|$ because $c > 1$ for criticality.

At the same time, the criticality condition is given by

$$\frac{ca_0(1-\nu_0^2\beta^2)^{\frac{1}{2}}}{[1-(c/\beta)\tan^{-1}\beta]^{\frac{1}{2}}}\left[\frac{\nu_0^2(c-1)-1}{2(1+\nu_0^2)}\right]^{\frac{1}{2}} \times \cos\left\{(a+x_0)\left(\frac{1}{\nu_0^2}-\beta^2\right)^{\frac{1}{2}}\right\} + \frac{c}{2}\int_0^1 \psi(\nu')\frac{\nu'(1+\beta^2)}{(1+\nu'^2\beta^2)}X(-\omega') \times \exp\{-2a(1+\nu'^2\beta^2)^{\frac{1}{2}}/\nu'\}\,d\nu' = 0, \quad (96)$$

with an analogous expression for $\psi(\nu)$ obtained from Eq. (26).

In the end-point approximation, the criticality condition is simply

$$\cos\left\{(a+x_0)\left(\frac{1}{\nu_0^2}-\beta^2\right)^{\frac{1}{2}}\right\} = 0$$

or

$$2a = |\nu_0|\,\pi(1-\nu_0^2\beta^2)^{-\frac{1}{2}} - 2x_0(i\beta). \quad (97)$$

Recalling the definitions of B_y and B_z, we see that Eq. (97) constitutes a relationship between the dimensions of the system and its nuclear properties. The greatest uncertainty in (97) lies in the extrapolated end points y_0 and z_0. Even this can be removed, however, if we assume a cubic system $a = b = c$, for then $x_0 = y_0 = z_0$. It is also interesting to note that the correct extrapolated end-point to use in the definition of B_x is that corresponding to $B_y^2 + B_z^2$, i.e., $x_0[i(B_y^2+B_z^2)^{\frac{1}{2}}]$. Similarly, y_0 is defined as $x_0[i(B_x^2+B_z^2)^{\frac{1}{2}}]$ and z_0 as $x_0[i(B_x^2+B_y^2)^{\frac{1}{2}}]$. Thus Eq. (97), together with the definitions of B_x, B_y, and B_z constitute four equations for ν_0, B_x, B_y, and B_z, given a, b, and c.

For systems greater than about 2 mean free paths, Eq. (97) gives a very accurate description of the criticality condition. Corrections to this formula may be obtained from (96).

The angular distribution both within the body and on its surface may be obtained very easily from the methods described in Ref. 4.

11. EXTENSION OF DIFFUSION-LENGTH PROBLEM TO RECTANGULAR CROSS SECTION

Earlier we considered the diffusion-length problem in a slab, infinite in the y direction. Now, although it is not yet possible to solve the complete two-dimensional x–y problem, we can approach this by Fourier transforming Eq. (2) in the y direction and associating the square of the Fourier-transform variable with the buckling B_y^2. This gives a first approximation of the effect of leakage in the y direction, whilst the distribution in the x direction is treated exactly. We shall not carry through the analysis of this problem but simply state the equation to be solved. In fact this is just identical to Eq. (4) with the exception that k^2 is replaced by $k^2 + B_y^2$.

12. LIMITATIONS OF THE REDUCED BOLTZMANN EQUATION

Our analysis for the slab problem, which may be taken as essentially exact, enables us to assess the accuracy of the reduced Boltzmann equation suggested by the author in Ref. 4. Making the $\exp(iB_x x)$ assumption in Eq. (4) leads then to

$$\Psi(k) = \frac{Q_0[1/2(k^2+B_x^2)^{\frac{1}{2}}]\tan^{-1}\{(k^2+B_x^2)^{\frac{1}{2}}\}}{1-[c/2(k^2+B_x^2)^{\frac{1}{2}}]\tan^{-1}\{(k^2+B_x^2)^{\frac{1}{2}}\}}, \quad (98)$$

which is to be compared with the exact expression (31).

The singularities of $\Psi(k)$ consist of branch cuts extending from $i(1+B_x^2)^{\frac{1}{2}}$ to $i\infty$ and $-i(1+B_x^2)^{\frac{1}{2}}$ to $-i\infty$, and also two poles defined by the roots of

$$1 - \frac{c}{2(k^2+B_x^2)^{\frac{1}{2}}}\tan^{-1}\{(k^2+B_x^2)^{\frac{1}{2}}\} = 0. \quad (99)$$

Calling these poles $\pm ik_0$, we see that

$$k_0^2 = \frac{1}{\nu_0^2} + B_x^2. \quad (100)$$

Evaluating the Fourier inverse leads to

$$\rho(x,z) = \frac{Q_0(\nu_0^2-1)e^{-k_0 z}\cos B_x x}{c\nu_0(1+\nu_0^2B_x^2)^{\frac{1}{2}}[1-\nu_0^2(1-c)]} + \int_{(1+B_x^2)^{\frac{1}{2}}}^{\infty} e^{-sz}I(s,B_x)\,ds\cos B_x x, \quad (101)$$

where $I(s, B_x)$ is rather lengthy and will not be given.

Comparing Eq. (101) with Eqs. (40) and (48) highlights the approximations involved in the asymptotic assumption; these are: (1) no harmonics are predicted; (2) transient term has lower limit $(1+B_x^2)^{\frac{1}{2}}$, not unity; (3) transient term is separable in z and x; (4) the amplitude of the pole contribution is incorrect; (5) no limit on the transverse dimensions for the existence of k_0 are available, although such a bound may be obtained by imposing the condition of non-negativeness of the angular distribution, and leads to $k_0 \not> 1$.

These failures would seem to limit the use of the reduced Boltzmann equation considerably. However, such a view would be pessimistic for it is found that, provided $\nu_0^2B^2$ is not too near unity, the expression (101) gives an adequate representation of the neutron density and, in particular, the angular distribution

within the body. Moreover, the diffusion-length problem is a severe test of the reduced equation which may be shown to be much more acceptable for criticality and half-space problems, and certainly better than the infinite-medium approximations used hitherto.

13. SUMMARY

In this paper we have developed methods for dealing with the transport equation in more than one dimension. It has been demonstrated that important effects can arise due to a coupling between the different space dimensions. These effects would be missed in any asymptotic approximation.

An accurate formulation of the critical conditions in a rectangular parallelepiped has been derived which should provide a useful benchmark problem against which to compare approximate techniques.

The methods discussed above can be used in other branches of physics in which a linear transport equation arises, and, in particular, radiative-transfer and rarefied-gas dynamic problems.

Cylindrical geometries may also be treated and these are the subject of a subsequent paper.

Finally, it should be mentioned that the problems discussed in this report may all be derived via the Wiener–Hopf technique. We have used the present method since it is more readily generalizable to more complicated geometries.

Note added in proof: Since submitting this paper for publication, the author has become aware of the work of Bareiss and Abu-Shumays (Argonne National Laboratory Report 7328). Whilst this work does not deal explicitly with the problems discussed above, it does throw considerable light on the structure of the eigenvalue spectrum of the three-dimensional Boltzmann equation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Diffusion Length and Criticality Problems in Two- and Three-Dimensional, One-Speed Neutron Transport Theory. II. Circular Cylindrical Coordinates

M. M. R. WILLIAMS
Nuclear Engineering Department, Queen Mary College, University of London

(Received 10 September 1967)

An integral equation is derived which describes the diffusion of neutrons from a plane source perpendicular to the axis of an infinite circular cylinder. The equation is solved by a Fourier transform in the axial direction and by the subsequent solution of a singular integral equation for the radial distribution; this is accomplished as a result of a certain replication property possessed by the original integral equation. We have inverted the Fourier transform to obtain the simultaneous $r - z$ variation of the neutron density in the cylinder. In addition, a formula is given from which the angular distribution of neutrons may be obtained. By approximating the density along the axis of the prism with the asymptotic distribution, it has been possible to assess the effect of axial leakage on the cylindrical critical problem. A critical equation is derived and the appropriate extrapolated endpoints to be used in this are defined.

1. INTRODUCTION

In a previous publication[1] (hereafter referred to as I) we have studied the diffusion of neutrons from a plane source in an infinite slab, the source plane being perpendicular to the faces of the slab. In addition, we studied the critical problem in a rectangular parallelepiped when the density in two of the orthogonal directions was approximated by the asymptotic distribution.

The present paper extends the conclusions reached in I to circular cylindrical geometry. We shall therefore calculate the neutron density and angular distribution arising from a plane source situated in an infinite circular cylinder, the source plane being perpendicular to the axis of the cylinder. We shall also consider the critical problem of a finite cylinder in which the density in the axial direction is approximated by the asymptotic distribution.

2. THE INTEGRAL EQUATION

In I, an integral equation describing the neutron density in a system of arbitrary, but non-re-entrant,

[1] M. M. R. Williams, preceding paper, J. Math. Phys. **9**, 1873 (1968).

within the body. Moreover, the diffusion-length problem is a severe test of the reduced equation which may be shown to be much more acceptable for criticality and half-space problems, and certainly better than the infinite-medium approximations used hitherto.

13. SUMMARY

In this paper we have developed methods for dealing with the transport equation in more than one dimension. It has been demonstrated that important effects can arise due to a coupling between the different space dimensions. These effects would be missed in any asymptotic approximation.

An accurate formulation of the critical conditions in a rectangular parallelepiped has been derived which should provide a useful benchmark problem against which to compare approximate techniques.

The methods discussed above can be used in other branches of physics in which a linear transport equation arises, and, in particular, radiative-transfer and rarefied-gas dynamic problems.

Cylindrical geometries may also be treated and these are the subject of a subsequent paper.

Finally, it should be mentioned that the problems discussed in this report may all be derived via the Wiener–Hopf technique. We have used the present method since it is more readily generalizable to more complicated geometries.

Note added in proof: Since submitting this paper for publication, the author has become aware of the work of Bareiss and Abu-Shumays (Argonne National Laboratory Report 7328). Whilst this work does not deal explicitly with the problems discussed above, it does throw considerable light on the structure of the eigenvalue spectrum of the three-dimensional Boltzmann equation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Diffusion Length and Criticality Problems in Two- and Three-Dimensional, One-Speed Neutron Transport Theory. II. Circular Cylindrical Coordinates

M. M. R. WILLIAMS
Nuclear Engineering Department, Queen Mary College, University of London

(Received 10 September 1967)

An integral equation is derived which describes the diffusion of neutrons from a plane source perpendicular to the axis of an infinite circular cylinder. The equation is solved by a Fourier transform in the axial direction and by the subsequent solution of a singular integral equation for the radial distribution; this is accomplished as a result of a certain replication property possessed by the original integral equation. We have inverted the Fourier transform to obtain the simultaneous $r - z$ variation of the neutron density in the cylinder. In addition, a formula is given from which the angular distribution of neutrons may be obtained. By approximating the density along the axis of the prism with the asymptotic distribution, it has been possible to assess the effect of axial leakage on the cylindrical critical problem. A critical equation is derived and the appropriate extrapolated endpoints to be used in this are defined.

1. INTRODUCTION

In a previous publication[1] (hereafter referred to as I) we have studied the diffusion of neutrons from a plane source in an infinite slab, the source plane being perpendicular to the faces of the slab. In addition, we studied the critical problem in a rectangular parallelepiped when the density in two of the orthogonal directions was approximated by the asymptotic distribution.

The present paper extends the conclusions reached in I to circular cylindrical geometry. We shall therefore calculate the neutron density and angular distribution arising from a plane source situated in an infinite circular cylinder, the source plane being perpendicular to the axis of the cylinder. We shall also consider the critical problem of a finite cylinder in which the density in the axial direction is approximated by the asymptotic distribution.

2. THE INTEGRAL EQUATION

In I, an integral equation describing the neutron density in a system of arbitrary, but non-re-entrant,

[1] M. M. R. Williams, preceding paper, J. Math. Phys. **9**, 1873 (1968).

shape was given. When applied to cylindrical geometry, this equation reduces to the following form:

$$\Psi(r,k)=\int_0^R r'dr'\int_0^1 G_\nu(r,r';k)\frac{d\nu}{\nu^2}\,[c\Psi(r',k)+Q(r')], \tag{1}$$

where

$$G_\nu(r,r';k)=\begin{cases} K_0\left(\frac{r}{\nu}(1+k^2\nu^2)^{\frac{1}{2}}\right)I_0\left(\frac{r'}{\nu}(1+k^2\nu^2)^{\frac{1}{2}}\right), & r'\le r,\\ K_0\left(\frac{r'}{\nu}(1+k^2\nu^2)^{\frac{1}{2}}\right)I_0\left(\frac{r}{\nu}(1+k^2\nu^2)^{\frac{1}{2}}\right), & r'\ge r. \end{cases} \tag{2}$$

In Eq. (1), $\Psi(r,k)$ is the infinite-medium Fourier transform of the density $\rho(r,z)$, i.e.,

$$\Psi(r,k)=\int_{-\infty}^{\infty}e^{-ikz}\rho(r,z)\,dz. \tag{3}$$

$K_0(x)$ and $I_0(x)$ are the familiar Bessel functions.

The main body of this paper is concerned with solving Eq. (1) and then inverting $\Psi(r,k)$ to find $\rho(r,z)$.

3. REPLICATION PROPERTY OF EQ. (1)

From the symmetry of the problem, it would seem profitable to seek solutions to Eq. (1) in the following form:

$$\Psi(r,k)=S+A_0I_0\left(\frac{r}{\tilde\nu}(1+\tilde\nu^2k^2)^{\frac{1}{2}}\right)+\int_0^1\phi(\nu)I_0\left(\frac{r}{\nu}(1+\nu^2k^2)^{\frac{1}{2}}\right)d\nu. \tag{4}$$

Therefore let us consider the action of the integral operator on the right-hand side of Eq. (1) on a function of the type

$$I_0\left(\frac{r}{\tilde\nu}(1+k^2\tilde\nu^2)^{\frac{1}{2}}\right).$$

We find that

$$\int_0^R r'\,dr'\int_0^1 G_\nu(r,r';k)\frac{d\nu}{\nu^2}I_0\left(\frac{r'}{\tilde\nu}(1+\nu^2k^2)^{\frac{1}{2}}\right)=I_0\left(\frac{r}{\tilde\nu}\tilde\alpha\right)\int_0^1\frac{d\nu}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\tilde\alpha^2}{\tilde\nu^2}\right]^{-1}-\int_0^1\frac{d\nu}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\tilde\alpha^2}{\tilde\nu^2}\right]^{-1}I_0\left(\frac{r}{\nu}\alpha\right)\Phi(R,\nu,\tilde\nu), \tag{5}$$

where $\alpha=(1+k^2\nu^2)^{\frac{1}{2}}$ and $\tilde\alpha=(1+k^2\tilde\nu^2)^{\frac{1}{2}}$. Also

$$\Phi(R,\nu,\tilde\nu)=\frac{R\tilde\alpha}{\tilde\nu}K_0\left(\frac{R}{\nu}\alpha\right)I_1\left(\frac{R}{\tilde\nu}\tilde\alpha\right)+\frac{R\alpha}{\nu}I_0\left(\frac{R\tilde\alpha}{\tilde\nu}\right)K_1\left(\frac{R\alpha}{\nu}\right). \tag{6}$$

Therefore we see that the action of the operator on $I_0(\cdots)$ reproduces the function, and adds a distribution of it over ν between (0, 1). This behavior confirms the correctness of our assumed solution (4).

Inserting Eq. (4) into Eq. (1) with $Q(r)=Q_0$, and using (5), leads, after equating coefficients of terms in $I_0(r\tilde\alpha/\tilde\nu)$, $I_0(r\alpha/\nu)$, and the constant, to

$$S=Q_0\frac{1}{k}\tan^{-1}k\Big/\left(1-\frac{c}{k}\tan^{-1}k\right) \tag{7}$$

and

$$\Lambda(\tilde\nu)=1-c\tilde\nu^2\int_0^1\frac{d\nu}{\tilde\nu^2-\nu^2}=0$$

$$=1-c\int_0^1\frac{d\nu}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\tilde\alpha^2}{\tilde\nu^2}\right]^{-1}, \tag{8}$$

which defines $\tilde\nu=\nu_0$, the infinite-medium diffusion length.

Finally, the integral equation for $\phi(\nu)$ becomes

$$\Lambda(\nu)\phi(\nu)+c\int_0^1\frac{d\nu'}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\alpha'^2}{\nu'^2}\right]^{-1}\phi(\nu')\Phi(R,\nu,\nu')=-\frac{c}{\nu^2}A_0\left[\frac{\alpha^2}{\nu^2}-\frac{\alpha_0^2}{\nu_0^2}\right]^{-1}\Phi(R,\nu,\nu_0)-\frac{Q_0}{1-\frac{c}{k}\tan^{-1}k}\cdot\frac{R/\nu}{(1+\nu^2k^2)^{\frac{1}{2}}}K_1(R\alpha/\nu). \tag{9}$$

We now define a new function $\psi(\nu)$ by

$$\psi(\nu)\frac{R}{\nu}(1+k^2\nu^2)^{\frac{1}{2}}K_1\left(\frac{R}{\nu}(1+k^2\nu^2)^{\frac{1}{2}}\right)=\phi(\nu) \tag{10}$$

and rearrange Eq. (9) so that we may write

$$\Lambda(\nu)\psi(\nu)+c\int_0^1\frac{d\nu'}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\alpha'^2}{\nu'^2}\right]^{-1}\psi(\nu')\times\left[1-p\left(\frac{R\alpha'}{\nu'}\right)\left\{q\left(\frac{R\alpha'}{\nu'}\right)-\frac{\nu\alpha'}{\nu'\alpha}q\left(\frac{R\alpha}{\nu}\right)\right\}\right]$$
$$=-\frac{cA_0}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\alpha_0^2}{\nu_0^2}\right]^{-1}\left\{I_0\left(\frac{R\alpha_0}{\nu_0}\right)+\frac{\nu\alpha_0}{\nu_0\alpha}q\left(\frac{R\alpha}{\nu}\right)I_1\left(\frac{R\alpha_0}{\nu_0}\right)\right\}-\frac{Q_0}{1-\frac{c}{k}\tan^{-1}k}\cdot\frac{1}{(1+\nu^2k^2)}. \tag{11}$$

In arriving at Eq. (11) we have used the relation

$$x[I_0(x)K_1(x)+K_0(x)I_1(x)]=1$$

and have defined

$$\left.\begin{aligned} q(x)&=K_0(x)/K_1(x)\sim 1+O(x^{-1})\\ \text{and}\quad p(x)&=xI_1(x)K_1(x)\sim\tfrac{1}{2}+O(x^{-1}) \end{aligned}\right\},\quad \text{for } x\to\infty.$$

If we rearrange the square bracket under the integral sign in Eq. (11) to the form

$$\frac{1}{2}\left(1+\frac{\nu\alpha'}{\nu'\alpha}\right)+\left[\tfrac{1}{2}-p'q'+\frac{\nu\alpha'}{\nu'\alpha}(p'q-\tfrac{1}{2})\right]$$
$$\equiv\frac{1}{2}\left(1+\frac{\nu\alpha'}{\nu'\alpha}\right)+W(\nu,\nu';R),$$

where p' and q' are in an obvious notation, then Eq. (11) may be written as

$$\Lambda(\nu)\psi(\nu)+\frac{c}{2}\int_0^1\frac{d\nu'}{\nu\alpha}\left[\frac{\alpha}{\nu}-\frac{\alpha'}{\nu'}\right]^{-1}\psi(\nu')=-\eta', \quad (12)$$

where

$$\eta'=c\int_0^1\frac{d\nu'}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\alpha'^2}{\nu'^2}\right]^{-1}W(\nu,\nu';R)\psi(\nu')$$
$$+\frac{cA_0}{\nu^2}\left[\frac{\alpha^2}{\nu^2}-\frac{\alpha_0^2}{\nu_0^2}\right]^{-1}$$
$$\times\left\{I_0\left(\frac{R\alpha_0}{\nu_0}\right)+\frac{\nu\alpha_0}{\nu_0\alpha}q\left(\frac{R\alpha}{\nu}\right)I_1\left(\frac{R\alpha_0}{\nu_0}\right)\right\}$$
$$+\frac{Q_0}{1-\frac{c}{k}\tan^{-1}k}\cdot\frac{1}{(1+\nu^2k^2)}. \quad (13)$$

Equation (12) is now identical in form to one already solved in I, with the exception of η', which constitutes a new, nonsingular source term.

It should be noted that the function $W(\nu,\nu';R)$ has the following property:

$$W(\nu,\nu';R)\sim O(R^{-1})$$

as $R\to\infty$. Thus, although Eq. (12) goes over to the correct infinite medium equation as $R\to\infty$, it does so more slowly than in the slab problem, and consequently the "end-point approximation" will be less accurate. This behavior has also been noted by Kiesewetter in connection with the simpler problem of criticality in an infinite cylinder.[2]

4. SOLUTION OF EQ. (12)

The solution of Eq. (12) follows the same lines as in I. Thus the change to the new variable

$$\omega=\nu(1+k^2)^{\frac{1}{2}}/(1+\nu^2k^2)^{\frac{1}{2}}$$

reduces Eq. (12) to

$$\Lambda_0(\omega)\psi_0(\omega)+\frac{c}{2}\int_0^1\frac{\omega'\psi_0(\omega')\,d\omega'}{\omega'-\omega}=-\eta_0', \quad (14)$$

with

$$\eta_0'=c\int_0^1\frac{W_0(\omega,\omega';R)\psi_0(\omega')\omega'^2\,d\omega'}{\omega'^2-\omega^2}$$
$$+\frac{cA_0\omega_0^2}{(1+k^2)(\omega_0^2-\omega^2)}\left\{I_0\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega_0}\right)\right.$$
$$\left.+\frac{\omega}{\omega_0}q\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega}\right)I_1\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega_0}\right)\right\}$$
$$+\frac{Q_0}{(1+k^2)\left(1-\frac{c}{k}\tan^{-1}k\right)}, \quad (15)$$

where

$$g^3(\omega)\psi[\omega g(\omega)]=\psi_0(\omega),$$
$$\Lambda_0(\omega)=\Lambda[\omega g(\omega)]/g(\omega),$$
$$g(\omega)=\{1+k^2(1-\omega^2)\}^{-\frac{1}{2}},$$

and

$$W_0(\omega,\omega';R)=W[\omega g(\omega),\omega' g(\omega');R].$$

Neglecting the term containing W_0, the solution of Eq. (14) may be written down immediately[3] as

$$(1+k^2)\psi_0(\omega)$$
$$=-cA_0\omega_0X(-\omega)\left(1-\frac{c}{k}\tan^{-1}k\right)(\omega_0^2-\omega^2)$$
$$\times\left[I_0\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega_0}\right)\frac{1}{2}\left\{\frac{X(\omega_0)}{\omega_0-\omega}+\frac{X(-\omega_0)}{\omega_0+\omega}\right\}\right.$$
$$\left.+I_1\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega_0}\right)\int_0^1\frac{\gamma(\omega')\omega' q(R(1+k^2)^{\frac{1}{2}}/\omega')\,d\omega'}{(\omega'-\omega)(\omega_0^2-\omega'^2)}\right], \quad (16)$$

with

$$A_0=-\frac{Q_0}{1-\frac{c}{k}\tan^{-1}k}\cdot\frac{1}{c\omega_0}$$
$$\times\left\{I_0\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega_0}\right)\int_0^1\frac{\omega_0\gamma(\omega)\,d\omega}{\omega_0^2-\omega^2}\right.$$
$$+I_1\left(\frac{R(1+k^2)^{\frac{1}{2}}}{\omega_0}\right)$$
$$\left.\times\int_0^1\frac{\gamma(\omega)\omega q(R(1+k^2)^{\frac{1}{2}}/\omega)\,d\omega}{\omega_0^2-\omega^2}\right\}^{-1}. \quad (17)$$

Unfortunately, the integrals involving q in Eqs. (16) and (17) do not seem to be reducible any further. It is clear, however, that in the spirit of our previous approximations, we may consistently set $q=1$.

[2] Helmut Kiesewetter, "Solutions of the Neutron Transport Equation for a Slab, a Sphere and a Cylinder in the Sense of Wiener and Hopf," private communication, 1966.

[3] G. Mitsis, Argonne National Laboratory preprint 6787, 1963.

Equations (16) and (17) then simplify considerably to

$$(1 + k^2)\psi_0(\omega) = -\tfrac{1}{2}cA_0\omega_0 X(-\omega)\left(1 - \frac{c}{k}\tan^{-1} k\right)(\omega_0^2 - \omega^2) \times \left[\frac{X(\omega_0)}{\omega_0 - \omega}\{I_0(t_0) + I_1(t_0)\} + \frac{X(-\omega_0)}{\omega_0 + \omega}\{I_0(t_0) - I_1(t_0)\}\right], \quad (18)$$

where $t_0 = R(1 + k^2)^{\frac{1}{2}}/\omega_0$ and

$$A_0 = -\frac{Q_0}{1 - \frac{c}{k}\tan^{-1} k} \cdot \frac{2}{c\omega_0}[X(-\omega_0)\{I_0(t_0) - I_1(t_0)\} - X(\omega_0)\{I_0(t_0) + I_1(t_0)\}]^{-1}. \quad (19)$$

In terms of the original variable v_0, (19) can also be written

$$A_0 = -\frac{Q_0}{1 - \frac{c}{k}\tan^{-1} k} \frac{2}{c(1 - v_0^2k^2)^{\frac{1}{2}}}\left[\frac{2(v_0^2 - 1)}{1 - v_0^2(1 - c)}\right]^{\frac{1}{2}} \times [(I_0(t_0) - I_1(t_0)) \exp\{-x_0(1 + v_0^2k^2)^{\frac{1}{2}}/v_0\} + (I_0(t_0) + I_1(t_0)) \exp\{x_0(1 + v_0^2k^2)^{\frac{1}{2}}/v_0\}]^{-1}, \quad (20)$$

with $t_0 = R(1 + v_0^2k^2)^{\frac{1}{2}}/v_0$.

It may be verified that, for large R, Eqs. (18) and (19) go over to slablike values as obtained in I.

5. INVERSION IN THE k PLANE

From Eq. (4) and after neglecting the surface transients, we may write

$$\Psi(r, k) = \frac{Q_0 \frac{1}{k}\tan^{-1} k}{1 - \frac{c}{k}\tan^{-1} k} + A_0 I_0\left(\frac{r}{v_0}(1 + v_0^2k^2)^{\frac{1}{2}}\right). \quad (21)$$

It may be verified that the apparent pole in Eq. (21) due to the zeros of $1 - (c/k)\tan^{-1} k$ does not in fact exist owing to exact cancellation from the second term. The singularities of Eq. (21) in the k plane arise therefore from the roots of the equation

$$[I_0(t_0) - I_1(t_0)] \exp\{-x_0(1 + k^2v_0^2)^{\frac{1}{2}}/v_0\} + [I_0(t_0) + I_1(t_0)] \exp\{x_0(1 + k^2v_0^2)^{\frac{1}{2}}/v_0\} = 0 \quad (22)$$

and from the branch points at $k = \pm i$.

Let us consider the pole contribution arising from the zeros of (22). If we designate the roots of (22) by $k = \pm ik_n$, and define a quantity B_n as

$$k_n^2 = B_n^2 + (1/v_0^2), \quad (23)$$

then B_n will be given by the roots of the following equation:

$$J_0(BR)\cos(Bx_0) - J_1(BR)\sin(Bx_0) = 0, \quad (24)$$

where $x_0(k)$ is evaluated at $k^2 = -k_n^2$. For large BR, where we may use the asymptotic expansions of the Bessel functions, Eq. (24) becomes

$$\cos\{B(R + x_0) - \tfrac{1}{4}\pi\} = 0$$

or

$$B_n = (n + \tfrac{3}{4})\pi/(R + x_0). \quad (25)$$

The case of $n = 0$ gives $B_0 = 2.356/(R + x_0)$, which is close to the familiar diffusion-theory result of $2.405/(R + x_0)$. As n increases, Eq. (25) becomes more accurate.

It should be added that only values of B_n such that $k_n \leq 1$ are acceptable solutions of Eq. (24). This restriction limits the number of roots to a finite value N. As R decreases, N also decreases, until for a certain critical R (R_0^* say) no roots can be found. In such a situation there will be no asymptotic, $r - z$ separable, solution and the density will decay nonexponentially from the source. The physical consequences of this behavior and a possible method for dealing with it have been discussed in I.

For the sake of example we shall write down the asymptotic part of $\rho(r, z)$ which arises from the residues of $\Psi(r, k)$ at $k = ik_n$ ($z > 0$). We find

$$\rho_{\text{asy}}(r, z) = \frac{2Q_0}{cv_0}\left[\frac{2(v_0^2 - 1)}{1 - v_0^2(1 - c)}\right]^{\frac{1}{2}} \times \sum_{n=0}^{N} \frac{J_0(B_nr)}{B_nD_n\left\{\frac{c}{2k_n}\log\left(\frac{1 + k_n}{1 - k_n}\right) - 1\right\}^{\frac{1}{2}}}, \quad (26)$$

where

$$iD_n = \frac{d}{dk}[(I_0(t_0) - I_1(t_0)) \exp\{-x_0(1 + v_0^2k^2)^{\frac{1}{2}}/v_0\} + (I_0(t_0) + I_1(t_0)) \exp\{x_0(1 + v_0^2k^2)^{\frac{1}{2}}/v_0\}]_{k=ik_n}.$$

The explicit form of D_n will not be given due to its great length. It is straightforward to obtain, provided it is recalled that $x_0(k)$ is a function of k.

$\rho_{\text{asy}}(r, z)$ is seen to be similar in form to the classic diffusion-theory expression, but now the amplitudes of the modes are given exactly and the transverse buckling B_n is uniquely defined in terms of the known extrapolated endpoint.

The transient contribution to $\rho(r, z)$ has the general form

$$\rho(r, z) = \int_1^{\infty} K(s)\{1 + B(s)J_0(r(v_0^2s^2 - 1)^{\frac{1}{2}}/v_0)\}e^{-sz}\,ds \quad (27)$$

and is not separable in r and z.

A more accurate description of the density in the neighborhood of the boundary may be obtained if Eq. (18) is inserted into the last term of Eq. (4), and the surface transient is then inverted in the k plane. It is readily observed, however, that this term becomes negligible very rapidly as we move away from the boundary.

6. THE ANGULAR DISTRIBUTION

The Fourier transform in k space of the angular distribution at position (r, z) in the cylinder may be reconstructed from the well-known integral formula[3]:

$$\hat{f}(r, k, \theta, \phi) = \frac{1}{4\pi}\int_0^{s_0(r,\theta,\phi)} e^{-s(1+ik\cos\theta)}[c\Psi(\{r^2 + s^2\sin^2\theta - 2sr\sin\theta\cos\phi\}^{\frac{1}{2}}, k) + Q_0]\,ds, \quad (28)$$

where

$$S_0 = \{r\cos\phi + [R^2 - r^2\sin^2\phi]^{\frac{1}{2}}\}/\sin\theta, \quad (29)$$

and we have accounted for the variation in the z direction by the factor $(1 + ik\cos\theta)$.

The angular coordinates (θ, ϕ) are defined in Ref. 3.

Defining $\Phi(r, z) = c\rho(r, z) + Q_0$, we may formally write the angular distribution at the point (r, z) as

$$f(r, z, \theta, \phi) = \frac{1}{4\pi}\int_0^{s_0(r,\theta,\phi)} e^{-s}\Phi(\{r^2 + s^2\sin^2\theta - 2sr\sin\theta\cos\phi\}^{\frac{1}{2}}, z - s\cos\theta)\,ds. \quad (30)$$

We shall not, however, pursue the algebraic problem that results when Eqs. (26) and (27) are inserted into (30).

7. THE CRITICAL PROBLEM

The results obtained in Sec. 4 can be taken over directly to solve the problem of a critical cylinder of finite length H. Making the assumption that the flux distribution in the z direction may be approximated by the asymptotic distribution $\exp(iB_z z)$, where $B_z = \pi/(H + 2Z_0)$, we find that the solution of the critical problem is given by Eq. (14) with $Q_0 = 0$. Then $A_0^{-1} = 0$ gives the critical condition, and, with the approximations discussed earlier, this is given explicitly by Eq. (24). The lowest root of (24), $B_0 = B_r$, corresponds to the radial geometric buckling. From Eqs. (22) and (23), therefore, and remembering that in a critical problem $\nu_0 = i\,|\nu_0|$, we may write

$$1/\nu_0^2 = B_z^2 + B_r^2. \quad (31)$$

The appropriate extrapolated end-points to use in (31) raises an interesting point. For example, from Eq. (24), B_r^2 depends upon $x_0(iB_z)$, i.e., upon the buckling in the other orthogonal direction. We conclude, therefore, that z_0 will be defined as $x_0(iB_r)$. To find ν_0, B_r, and B_z it is necessary to solve Eqs. (24) and (31), and the one relating B_z to z_0, simultaneously.

Additional evidence for the validity of Eq. (31) may be obtained from the results of I. For example, we obtained the critical equation for a slab with a distribution $\exp\{iB_y y + iB_z z\}$ in the y and z directions. Thus we obtained

$$1/\nu_0^2 = B_x^2 + \beta^2,$$

where $\beta^2 = B_y^2 + B_z^2$. We could equally well assume, however, that the slab has a cylindrical boundary and write $\beta^2 = B_r^2$, with B_r given by Eq. (24). In this way we arrive back at Eq. (31) with the same definitions of B_z and B_r.

The corresponding angular distribution may be obtained quite readily from Eq. (28) with $Q_0 = 0$.

8. SUMMARY AND DISCUSSION

The use of a replication property and an infinite-medium Fourier transform have enabled us to solve a simple two-dimensional transport-theory problem. Although we have concentrated on circular cylinders in this paper, the method may be used for any non-re-entrant, cross-sectional shape whose boundary coincides with one of the coordinate surfaces for which the Helmholtz equation is separable.[4] However, it has not yet been possible to deal exactly with rectangular cross sections, although an approximate method was discussed in I.

The present technique is also applicable to time-dependent problems, e.g., initial-value and wave-propagation problems, if the time variable is eliminated by a Laplace transform. Subsequent inversion of the solution $\tilde{\rho}(r, z; s)$ will yield $\rho(r, z; t)$; we shall consider such problems in a future publication.

[4] K. M. Case, *Developments in Transport Theory*, E. Inönu and P. Zweifel, Eds. (Academic Press Inc., New York, 1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Strong-Coupling Limit in Potential Theory. II

WILLIAM M. FRANK
Bar-Ilan University, Ramat Gan, Israel

(Received 29 March 1968)

The analytic properties of the Jost function in the coupling constant g, derived in an earlier article for a restricted class of potentials, are rederived generally for potentials $V(r)$, for which $|V(r)|$ can be bounded by a monotonically decreasing potential $\bar{V}(r)$, such that $\int dr\,|\bar{V}(r)|^{\frac{1}{2}} < \infty$. In particular, potentials with nonlinear exponential tails are studied as a function of the energy and coupling constant. Earlier results of Sartori on the energy dependence of the Jost function, derived only for special cases of such potentials in the Born approximation, are demonstrated to be true generally for large classes of such potentials.

I. INTRODUCTION

In the earlier article[1] (to be referred to as I) the analytic properties in g were studied for the Jost function based on the Schrödinger equation

$$\left[\frac{d^2}{dr^2} + \kappa^2 - \frac{l(l+1)}{r^2} + gV(r)\right]u_l(r) = 0, \quad (1)$$

describing the scattering states for $\kappa^2 > 0$ in a potential $-gV(r)$ as well as the bound states of energy κ^2 ($\hbar^2/2m \equiv 1$) for $\kappa^2 < 0$, when the potential is attractive. In particular, the large-g behavior gave a new expression for the number of bound states which was related to the large-g behavior of the scattering phase shifts. That analysis is derived from the conditions

(i) $V(r)$ is $L^{\frac{1}{2}}$, i.e., $\int_0^\infty dr\,|V(r)|^{\frac{1}{2}} < \infty$,

(ii) $|V(r)|^{\frac{1}{2}}$ is expressible as an absolutely convergent Laplace transform, i.e., one can write

$$U(r) \equiv |V(r)|^{\frac{1}{2}} = \int_0^\infty d\alpha\,\sigma(\alpha)e^{-\alpha r}, \quad (2)$$

where

$$\int_0^\infty d\alpha\,|\sigma(\alpha)| < \infty.$$

Condition (ii) limits one to potentials which damp for large r no faster than a linear exponential in the sense that

$$\overline{\lim_{r\to\infty}} -\frac{1}{r}\ln|V(r)| < \infty. \quad (3)$$

If the limit in Eq. (3) has a finite nonzero value, the potential will be said to possess a "linear exponential tail" (l.e.t.). This type of potential with the exception of the square well is most frequently considered in standard potential theory. Equation (2) could not represent potentials which have a nonlinear exponential tail (n.l.e.t.) such as a Gaussian or square well. It does include potentials with power tails (p.t.), i.e., potentials such that

$$\lim_{r\to\infty}\frac{1}{\ln r}\ln|V(r)| < \infty. \quad (4)$$

It is of interest to broaden the scope of potential theory to include n.l.e.t. potentials for the sake of mathematical generality, as well as for the sake of understanding better the transition from often used model potentials, which vanish identically beyond a certain range, and the more realistic l.e.t. types. The Jost function of an l.e.t. potential has singularities in the $k = E^{\frac{1}{2}}$ plane, which are absent in the n.l.e.t. case if

$$-\ln|V(r)| \sim r^p, \quad (5)$$

for large r with $p > 1$. The case $0 < p < 1$ is generally expressible in the form of Eq. (2), as is demonstrated subsequently. The k dependence of Jost functions for n.l.e.t. potentials with $p > 1$, has been considered in the first Born approximation by Sartori[2] for a very artificial form of n.l.e.t. potential.

In this article the results of Paper I are extended to a general $L^{\frac{1}{2}}$ potential. We find the exponential order[3] in g of the Jost function to be $\frac{1}{2}$ for all k, and the same estimate as before applies to the number of bound states. In Sec. III the analytical behavior of the Jost function in g and k is studied for n.l.e.t. potentials expressible in the general form

$$|V(r)|^{\frac{1}{2}} = \int_\beta^\infty d\alpha\,\sigma(\alpha)\exp[-(\alpha r)^p], \quad (6)$$

with $p > 1$. We shall restrict ourselves to the condition $\beta > 0$. $\beta = 0$ leads to p.t. potentials which are

* Present address: U.S. Naval Ordinance Laboratory, Silver Spring, Maryland.

[1] W. M. Frank, J. Math. Phys. **8**, 466 (1967). The following errors in this article should be noted: The right sides of Eqs. (11), (14), (51), and (53) require an additional factor of 2; the G in Eq. (51) should be barred; the sign of ikr in the expression for $h_l^{(+)}(kr)$ following Eq. (51) should be +; and the right side of Eq. (61) should have a plus sign.

[2] L. Sartori, J. Math. Phys. **4**, 1408 (1963).

[3] A function $f(z)$ analytic in the finite z plane is said to have exponential order ρ if ρ is the greatest lower bound of numbers α, such that $|f(z)| \leq \exp K_\alpha|z|^\alpha$ for some K_α. If $f(z)$ has exponential order α, its type τ is defined by $\tau = \limsup|z|^{-\alpha}|\ln f(z)|$.

generally expressible by Eq. (2). The results of Sartori are established in this general case.

In a subsequent article the WKB method is applied to establish the same large-coupling behavior of the Jost function, and to derive an asymptotic expansion about the point at infinity as well. The strong-coupling limit of scattering phase shifts is discussed there.

II. $\bar{L}^{\frac{1}{2}}$ POTENTIALS

The results derived apply to all potentials $V(r)$, for which there exists a $\bar{V}(r)$ such that

$$0 \leq \bar{V}(r+a) \leq \bar{V}(r) \quad \text{for all } a \geq 0, \tag{7a}$$

$$|V(r)| \leq \bar{V}(r), \tag{7b}$$

$$\int_0^\infty dr\, \bar{V}(r)^{\frac{1}{2}} < \infty. \tag{7c}$$

It follows from Eqs. (7b) and (7c) that

$$\int_0^\infty dr\, |V(r)|^{\frac{1}{2}} < \infty. \tag{8}$$

If Eq. (8) is obeyed we say that $V(r)$ is $L^{\frac{1}{2}}$, while if, in addition, the conditions of Eqs. (7) hold we say that $V(r)$ is $\bar{L}^{\frac{1}{2}}$.

It will be shown that the Jost function[4] $\Delta_l(g, -\kappa)$ for Eq. (1) is an entire function of g with exponential order $\frac{1}{2}$ for any fixed κ which is not a singular point of $\Delta_l(g, -\kappa)$. As derived in I [see Eqs. (13)–(18) and (52)–(54)] and rederived in Appendix A, the modulus of $\Delta_l(g, -\kappa) \equiv \Delta(g)$ can be bounded by the series (for real κ, or κ in the lower half-plane):

$$\begin{aligned}
|\Delta(g)| &\leq \sum_{n=0}^\infty |g|^n \int_0^\infty dr_1 \\
&\quad \times \int_0^{r_1} dr_2 \cdots \int_0^{r_{n-1}} dr_n (r_1 - r_2)\cdots(r_{n-1} - r_n) \\
&\quad \times r_n\, |V(r_1)\cdots V(r_n)| \\
&\equiv \sum_{n=0}^\infty |g|^n \int_0^\infty d\xi_1 \cdots \int_0^\infty d\xi_n \xi_1 \cdots \xi_n \\
&\quad \times |V(\xi_n) V(\xi_{n-1} + \xi_n)\cdots V(\xi_1 + \cdots + \xi_n)| \\
&\equiv \sum_{n=0}^\infty |g|^n L_n,
\end{aligned} \tag{9}$$

where the variables $\xi_j = r_j - r_{j+1}$ $(j = 1, \cdots, n-1)$ and $\xi_n = r_n$ have been introduced. From Eqs. (7) we find

$$\begin{aligned}
L_n &\leq \int_0^\infty d\xi_1 \cdots \int_0^\infty d\xi_n \xi_1 \cdots \xi_n \\
&\quad \times |\bar{V}(\xi_n)\bar{V}(\xi_{n-1} + \xi_n)\cdots \bar{V}(\xi_1 + \cdots + \xi_n)| \\
&= \left(\frac{2}{\pi}\right)^n \int_0^\infty d\xi_1 \cdots \int_0^\infty d\xi_n \int_0^{\pi/2} d\varphi_1 \cdots \int_0^{\pi/2} d\varphi_n \xi_1 \cdots \xi_n \\
&\quad \times |\bar{V}(\xi_n)\cdots \bar{V}(\xi_1 + \cdots + \xi_n)|.
\end{aligned} \tag{10}$$

We introduce the new variables $\zeta_1, \cdots, \zeta_{2n}$ defined by

$$\begin{aligned}
\zeta_{2k-1} &= \frac{1}{\sqrt{2}} \xi_k \cos\varphi_k, \\
\zeta_{2k} &= \frac{1}{\sqrt{2}} \xi_k \sin\varphi_k \quad (k = 1, \cdots, n).
\end{aligned} \tag{11}$$

From the inequalities

$$\zeta_{2k-1} + \zeta_{2k} \leq \sqrt{2}\,(\zeta_{2k-1}^2 + \zeta_{2k}^2)^{\frac{1}{2}} \equiv \xi_k, \tag{12a}$$

and

$$\begin{aligned}
&\bar{V}(\xi_k + \cdots + \xi_n) \\
&= \bar{V}^{\frac{1}{2}}(\xi_k + \cdots + \xi_n)\bar{V}^{\frac{1}{2}}(\xi_k + \cdots + \xi_n) \\
&\leq \bar{V}^{\frac{1}{2}}(\zeta_{2k-1} + \cdots + \zeta_{2n})\bar{V}^{\frac{1}{2}}(\zeta_{2k} + \cdots + \zeta_{2n})
\end{aligned} \tag{12b}$$

which follow from the monotonicity condition of Eq. (7a), we derive

$$\begin{aligned}
L_n &\leq \left(\frac{4}{\pi}\right)^n \int_0^\infty d\zeta_1 \cdots \int_0^\infty d\zeta_{2n} \bar{V}^{\frac{1}{2}}(\zeta_{2n}) \\
&\quad \times \bar{V}^{\frac{1}{2}}(\zeta_{2n-1} + \zeta_{2n}) \cdots \bar{V}^{\frac{1}{2}}(\zeta_1 + \cdots + \zeta_{2n}).
\end{aligned} \tag{13}$$

Introducing the variables $x_1, \cdots, x_{2n}$ defined by

$$\zeta_k = x_k - x_{k+1}, \quad \zeta_{2n} \equiv x_{2n} \quad (k = 1, \cdots, 2n-1), \tag{14}$$

we find

$$\begin{aligned}
L_n &\leq \left(\frac{4}{\pi}\right)^n \int_0^\infty dx_1 \int_0^{x_1} dx_2 \cdots \int_0^{x_{2n-1}} dx_{2n} \\
&\quad \times \bar{V}^{\frac{1}{2}}(x_1)\cdots \bar{V}^{\frac{1}{2}}(x_{2n}) \\
&= \left(\frac{4}{\pi}\right)^n \frac{1}{(2n)!} \left[\int_0^\infty dx\, \bar{V}^{\frac{1}{2}}(x)\right]^{2n} \\
&\equiv \frac{\bar{\tau}^{2n}}{(2n)!},
\end{aligned} \tag{15}$$

where

$$\bar{\tau} \equiv \frac{2}{\pi^{\frac{1}{2}}} \int_0^\infty dx\, \bar{V}^{\frac{1}{2}}(x).$$

This upper bound on the coefficients of the power-series expansion establishes the exponential order of $\Delta_l(g, \kappa)$ to be at most $\frac{1}{2}$. The argument presented at the end of Sec. I in I shows the exponential order to be no less than $\frac{1}{2}$ and therefore exactly $\frac{1}{2}$. It follows that the upper limit on the number of bound states to be of the form of Eq. (40) of I with $\bar{\tau}$ replacing τ. It was shown in I that the type[3] for a class of purely repulsive l.e.t. potentials is exactly

$$\tau = \int_0^\infty dr\, |V(r)|^{\frac{1}{2}}. \tag{16}$$

This result for the general case of purely repulsive $\bar{L}^{\frac{1}{2}}$ potentials is deduced by WKB methods in a subsequent article. A not too rigorous argument was

[4] The Jost function for radial momentum k is identical with the Fredholm determinant $\Delta(g, -k)$ for the scattering integral equation; see R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951), Appendix.

also presented in I to the effect that the attractive phase shift also has a leading $g^{\frac{1}{2}}$-type behavior for large g. This result is also spelled out more rigorously for the general case in the subsequent article.

III. N.L.E.T. POTENTIALS

We show first that a large class of n.l.e.t. potentials for which Eq. (5) is true with $0 < p < 1$ are representable by Eq. (2). Thus if $\sigma(\alpha)$ behaves like

$$\sigma(\alpha) \sim \alpha^{m-1} \exp(-c/\alpha^q), \tag{17}$$

for α near zero, one can readily establish from a steepest-descent-type estimate that, for large r,

$$|V(r)|^{\frac{1}{2}} = \int_0^\infty d\alpha\, \sigma(\alpha) e^{-\alpha r} \sim \text{const} \times r^{-(m+\frac{1}{2}q)/(q+1)} \exp[-Br^{q/(q+1)}], \tag{18}$$

so that with $q = p/(1-p)$ the large-r behavior satisfies Eq. (5). It is clear that the Jost function $\Delta(g, -k)$ will be analytic in k in a cut k plane with a cut in the upper half-plane extending to $k = 0$.

For the case $p > 1$, we deal with potentials which allow the representation

$$U(r) \equiv |V^{\frac{1}{2}}(r)| = \int_\beta^\infty d\alpha\, \sigma(\alpha) \exp[-(\alpha r)^p]. \tag{6}$$

We focus explicitly on S waves, though the same results for real energies apply to all l waves (see Appendix C). Though not completely general, Eq. (6) represents a wide class of n.l.e.t. potentials, and more general cases can also be treated by the methods to be presented. We assume $\beta > 0$. We verify that the Jost function as a function of g is entire with exponential order $\frac{1}{2}$, while as a function of k it is entire with exponential order $p/(p-1)$ as found by Sartori for the Born term.

We begin with the expression for the Jost function[5] (see I or Appendix A):

$$\Delta(g, -k) = \sum_{n=0}^\infty g^n \int_0^\infty d\xi_1 \cdots \int_0^\infty d\xi_n \bar{H}_0(\xi_1) \cdots \bar{H}_0(\xi_n) \times V(\xi_1) V(\xi_1 + \xi_2) \cdots V(\xi_1 + \cdots + \xi_n) \equiv \sum_{n=0}^\infty g^n L_n,$$

where

$$\bar{H}_0(\xi) = \frac{\exp(-ik\xi)(\sin k\xi)}{k} \tag{19}$$

for S waves, and is defined in Appendix C for general l waves. We consider complex $k = k_0 + ik_1$, then

$$|\bar{H}_k(\xi)| \leq \sqrt{2}\, \xi \exp(k_1 + |k_1|) \equiv \sqrt{2}\, \xi \exp[2k_1\theta(k_1)\xi]. \tag{20}$$

(The constant $\sqrt{2}$ is superfluous when $k_1 \leq 0$, but is retained for uniformity.) By reproducing the steps of Eqs. (9)–(14) in the same notation we find

$$|L_n| \leq \left(\frac{32^{\frac{1}{2}}}{\pi}\right)^n \int_0^\infty d\zeta_1 \cdots \int_0^\infty d\zeta_{2n} \exp[\kappa(\zeta_1 + \cdots + \zeta_{2n})] \times U(\zeta_{2n}) U(\zeta_{2n-1} + \zeta_{2n}) \cdots U(\zeta_1 + \cdots + \zeta_{2n}), \tag{21}$$

where $\kappa = 8^{\frac{1}{2}} k_1 \theta(k_1)$, $U(r) \equiv |V^{\frac{1}{2}}(r)|$ and we have used the inequality

$$\xi_k = \sqrt{2}\,(\zeta_{2n-1}^2 + \zeta_{2k}^2)^{\frac{1}{2}} \leq \sqrt{2}\,(\zeta_{2k-1} + \zeta_{2n}).$$

Introduction of the representation of Eq. (6) leads from Eq. (21) to

$$|L_n| \leq \left(\frac{32^{\frac{1}{2}}}{\pi}\right)^n \int d\alpha_1 \sigma(\alpha_1) \cdots \int d\alpha_{2n} \sigma(\alpha_{2n}) K_\kappa(\alpha_1, \cdots, \alpha_{2n}), \tag{22}$$

where

$$K_\kappa(\alpha_1, \cdots, \alpha_{2n}) \equiv \int_0^\infty d\zeta_1 \cdots \int_0^\infty d\zeta_{2n} \exp[\kappa(\zeta_1 + \cdots + \zeta_{2n}) - \alpha_{2n}^p \zeta_{2n}^p - \alpha_{2n-1}^p (\zeta_{2n-1} + \zeta_{2n})^p \cdots - \alpha_1^p (\zeta_1 + \cdots + \zeta_{2n})^p]. \tag{23}$$

At this point we utilize the following simple consequences of the Holder inequality[6] ($p \geq 1$):

$$\sum_{i=1}^n |a_i|^p \geq n^{1-p} \left[\sum_{i=1}^n |a_i|\right]^p, \tag{24}$$

to obtain

$$K_\kappa(\alpha_1, \cdots, \alpha_{2n}) \leq \int_0^\infty d\zeta_1 \cdots \int_0^\infty d\zeta_{2n} \times \exp\{\kappa(\zeta_1 + \cdots + \zeta_{2n}) - (2n)^{1-p}[\alpha_1\zeta_1 + (\alpha_1 + \alpha_2)\zeta_2 + \cdots + (\alpha_1 + \cdots + \alpha_{2n})\zeta_{2n}]^p\}, \tag{25}$$

which, after the introduction of the variables ($q \equiv 1/p$),

$$u_k = (2n)^{q-1}(\alpha_1 + \cdots + \alpha_k)\zeta_k \quad (k = 1, \cdots, 2n), \tag{26}$$

[5] We have expressed the potential as $-gV(r)$, so that positive $V(r)$ corresponds to an attractive potential. For this reason g^n appears in place of $(-g)^n$ in the Fredholm power series.

[6] The Holder inequality which appears in any book on functional analysis has the form for discrete sums ($1/p + 1/q = 1$):

$$\left|\sum_i a_i b_i\right| \leq \left[\sum_i |a_i|^p\right]^{1/p} \left[\sum_i |b_i|^q\right]^{1/q}.$$

results in

$$K_\kappa(\alpha_1, \cdots, \alpha_{2n}) \leq \frac{(2n)^{2n(1-q)}}{\alpha_1(\alpha_1+\alpha_2)\cdots(\alpha_1+\cdots+\alpha_{2n})}\int_0^\infty du_1 \cdots \int_0^\infty du_{2n} \times \exp\left[\kappa_n\left(\frac{u_1}{\alpha_1}+\cdots+\frac{u_{2n}}{\alpha_1+\cdots+\alpha_{2n}}\right) - (u_1+u_2+\cdots+u_{2n})^p\right] \equiv \frac{(2n)^{2n(1-q)}}{\alpha_1(\alpha_1+\alpha_2)\cdots(\alpha_1+\cdots+\alpha_{2n})} I_{2n}(\kappa_n), \quad (27)$$

where $\kappa_n = (2n)^{1-q}\kappa$. We first consider the case $k_1 \leq 0$ so that $\kappa_n = \kappa = 0$. In Appendix B we show that

$$I_{2n}(0) = \int_0^\infty du_1 \cdots \int_0^\infty du_{2n} \exp -(u_1+\cdots+u_{2n})^p = \frac{1}{p}\frac{\Gamma(2n/p)}{\Gamma(2n)}. \quad (28)$$

Combining Eqs. (22), (27), and (28) we find that

$$|L_n| \leq \frac{1}{p}\left(\frac{32^{\frac{1}{2}}}{\pi}\right)^n \frac{(2n)^{2n(1-q)}\Gamma(2n/p)}{\Gamma(2n)} \times \int d\alpha_1 \sigma(\alpha_1)\cdots\int d\alpha_{2n}\sigma(\alpha_{2n}) \times \frac{1}{\alpha_1(\alpha_1+\alpha_2)\cdots(\alpha_1+\cdots+\alpha_{2n})} = \frac{2n}{p}\left(\frac{32^{\frac{1}{2}}}{\pi}\right)^n \frac{(2n)^{2n(1-q)}\Gamma(2n/p)U_1^{2n}}{(2n)!^2}, \quad (29)$$

as shown in I, where

$$U_1 \equiv \int_0^\infty \frac{d\alpha\, \sigma(\alpha)}{\alpha}. \quad (30)$$

The bound on the power-series coefficients L_n determines[7] $\Delta_l(g, -k)$ to have an exponential order of at most $\frac{1}{2}$. Relying on the argument in I that the exponential order cannot be less than $\frac{1}{2}$, we conclude that $\Delta(g, -k)$ has an exponential order of exactly $\frac{1}{2}$ for an entire function of g.

The convergence of $\Delta(g, -k)$ for k in the lower half-plane also entails the analyticity in k of $\Delta(g, -k)$ in the lower half-plane. It is a familiar result[8] that the Jost function in this half k plane behaves like $1 + O(|k|^{-1})$ for large $|k|$. Such a result can be derived by using an alternative bound to Eq. (20)

$$|\bar{H}_0(\xi)| \leq \frac{1}{|k|}. \quad (31)$$

Utilizing the expression Eq. (A5) of Appendix A for $\Delta(g, -k)$ we find from Eq. (31) that

$$|\Delta(g, -k)| \leq \exp\left[\left|\frac{g}{k}\right| \int_0^\infty dx\, |V(x)|\right], \quad (32)$$

confirming the stated estimate. The bound in Eq. (32) grows like a linear exponential in g, which in view of earlier results must be an overestimate for large g. The origin of the discrepancy lies in the choice of the differing estimates of $|H_0(\xi)|$ in Eqs. (20) and (31). The underlying significance of the differing choice of estimates is discussed subsequently.

We now consider the behavior of $\Delta_l(g, -k)$ for k in the upper half-plane. Once again the large-$|g|$ behavior is correctly determined from the bound of Eq. (20) while the large-$|k|$ behavior is derived from the bound in Eq. (31). We first consider the large-g behavior, which involves a study of the integral $I_{2n}(k)$ of Eq. (27).

$$I_{2n}(\kappa_n) \equiv \int_0^\infty du_1 \cdots \int_0^\infty du_{2n} \times \exp\left[\kappa_n\left(\frac{u_1}{\alpha_1}+\cdots+\frac{u_{2n}}{\alpha_1+\cdots\alpha_{2n}}\right) - (u_1+\cdots+u_{2n})^p\right] < \int_0^\infty du_1 \cdots \int_0^\infty du_{2n} \times \exp\left[\frac{\kappa_n}{\beta}(u_1+\cdots+u_{2n}) - (u_1+\cdots+u_{2n})^p\right] \equiv J_{2n}\left(\frac{\kappa_n}{\beta}\right), \quad (33)$$

where $\alpha_j \geq \beta$ has been used. In Appendix B the integral $J_{2n}(\kappa_n/\beta)$ is estimated by the steepest-descent method and shown to behave for large n like

$$J_{2n}\left(\frac{\kappa_n}{\beta}\right) \sim K(\text{Im } k)^{-\frac{1}{2}}n^{-\frac{1}{2}(1-q)}[B(\text{Im } k)^{1/(p-1)} \times \exp(C\,|\text{Im } k|^{p/(p-1)})]^{2n}\frac{n^{2n/p}}{\Gamma(2n)}, \quad (34)$$

with B, C, and K denoting constants. In conjunction with Eqs. (22) and (27) one finally finds an upper bound for $|L_n|$ of the form

$$|L_n| \leq [B(\text{Im } k)^{1/(p-1)} \times \exp(C\,|\text{Im } k|^{p/(p-1)})]^{2n}\frac{n^{2n}}{(2n)!\,\Gamma(2n)}. \quad (35)$$

This type of large-n behavior determines[7] $\Delta(g, -k)$ to represent an entire function of g of exponential order

[7] See, e.g., E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, London, 1939), 2nd ed., Sec. 8.3.

[8] See, e.g., R. G. Newton, J. Math. Phys. **1**, 319 (1960), Eq. (4.16′).

at most $\frac{1}{2}$. The lower bound $\frac{1}{2}$ for the exponential order in g for k in the upper half-plane can be achieved through employment of the inequality

$$|\bar{H}_k(\xi)| \geq \frac{e^{2k_1\xi} - 1}{2|k|} \quad (k_1 \equiv \operatorname{Im} k), \tag{36}$$

together with a lower bound for $V(r)$ in the form of a square-well potential $S(r)$ such that $|V(r)| \geq S(r)$ everywhere. The resulting lower bound on L_n is readily related to the coefficients of the power-series expansion for the Jost function of $S(r)$ at the complex energy $k' = i \operatorname{Im} k$. The latter Jost function is explicitly verified[9] to have an exponential order of $\frac{1}{2}$.

The k dependence of the bound in Eq. (35) corresponds to an entire function of k of infinite exponential order. This is a gross overestimate originating in the crudeness of the bound in Eq. (33). A more accurate bound for the k behavior can be achieved by choosing for $|V(r)|$ a representation of the form

$$|V(r)| = \int d\alpha\, \rho(\alpha) \exp(-\alpha r^p), \tag{37}$$

as well as the bound

$$|H_k(\xi)| \leq \frac{1}{|k|} \exp(2k_1\xi). \tag{38}$$

This bound will be bad for $|k| = 0$, and we rely on other arguments[10] for the basic fact that $\Delta(g, -k)$ is analytic in the neighborhood of $k = 0$ if for some $a > 0$, $V(r)e^{ar} \to 0$ with diverging r. Substitution into the expression Eq. (A5) for L_n leads to

$$\begin{aligned} |L_n| &\leq \frac{1}{(|k|)^n} \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n e^{2k_1 r_1} |V(r_1) \cdots V(r_n)| \\ &= \frac{1}{(2|k|)^n (n-1)!} \int_0^\infty dr\, e^{2 \operatorname{Im} kr} |V(r)| [W(r)]^{n-1}, \end{aligned} \tag{39}$$

where

$$W(r) = \int_0^r dr\, |V(r)|. \tag{40}$$

Then

$$\begin{aligned} |\Delta(g, -k)| &\leq 1 + \gamma \int_0^\infty dr\, e^{2k_1 r} |V(r)|\, e^{\gamma W(r)} \\ &< 1 + \gamma e^{\gamma V_1} \int_0^\infty dr\, e^{2k_1 r} |V(r)| \\ &= 1 + \gamma e^{\gamma V_1} \int d\alpha\, \rho(\alpha) \int_0^\infty dr\, e^{2k_1 r - \alpha r^p}, \end{aligned} \tag{41}$$

where $|\gamma| = |g/k|$ and $V_1 \equiv \int_0^\infty dr\, |V(r)|$. The behavior of the integral

$$\int_0^\infty dr\, e^{2k_1 r - \alpha r^p} \tag{42}$$

for large k_1 may be determined through the use of a steepest-descent estimation technique or by expansion of a power series in k_1. Both of these methods which are explicitly presented in Sartori's article, show $\Delta(g, -k)$ to correspond to a function of k of exponential order of at most $p/(p-1)$. In the case of the power-series expansion, one must note that the coefficients of the power-series expansion determined the exponential order through standard theorems.[7] The representation, Eq. (37), can in fact be dispensed with in favor of an asymptotic statement of the form

$$|V(r)| \sim \exp - [\beta r^p (1 + O(r^p))]. \tag{43}$$

That $p/(p-1)$ is the true exponential order in k for potentials with the representation, Eq. (6), can be verified if the potential can be bounded pointwise from below by a nonnegative potential $U(r)$ of the form

$$U(r) = ar^m \exp(-\beta r^p). \tag{44}$$

Explicit calculation of the first Born term in the Jost-function expansion along the lines of the calculation by Sartori gives a contribution which has exponential order $p/(p-1)$, and the result is established.

IV. DISCUSSION

We noted that different inequalities on the modulus of the Green's function used to estimate the large-$|g|$ and large-$|k|$ behavior of the Jost function were used. These led to the general result that the best possible estimates on the behavior of the Jost function in one variable provided poor estimates in the other variable. This situation was necessitated by the need for bounds that would make for feasible estimations. One can qualitatively understand why doing one's best for large values of one variable would tend to denigrate the estimates for large values of the other variable. The modulus of the Jost function is a reciprocal measure of the penetrability of the repulsive potential. Strong coupling tends to decrease the penetrability, while high energy tends to increase it. Thus the correct large-g dependence of the Jost function requires realistic appraisal of the Green's function for small r which is expressed by the inequality (for real k):

$$|H_k(\xi)| = \left| \frac{e^{-ik} \sin k\xi}{k} \right| \leq \xi,$$

while the large-k dependence is better expressed through

$$|H_k(\xi)| \leq \frac{1}{k}.$$

[9] Reference 8, Sec. 10.
[10] Reference 8, Sec. 3.

The present results have been derived for S waves only, but we show in Appendix C that the inequalities of Eq. (20) (for $\mathrm{Im}\, k \leq 0$) and Eq. (31) remain essentially valid for higher partial waves.

Note Added in Proof: I am indebted to Dr. D. J. Land for showing that the factor $(4/\pi)^n$ on the right-hand side of Eq. (15) is superfluous. A similar type result was communicated to the author recently by Dr. K. Chadan.

A more general result for the lower half k plane can be obtained in place of Eq. (32) in terms of an expression involving

$$\exp\left[|g|\int_0^\infty dx \frac{x\,|V(x)|}{1+|\mathrm{Im}\, k|\, x}\right].$$

APPENDIX A: THE JOST-FUNCTION SERIES

The radial partial wave of Eq. (1) can be transformed into the integral equations[10]

$$\begin{aligned} u_l(r) &= krj_l(kr) + gkr\int_0^r dr' r'[j_l(kr)h_l^{(+)}(kr') \\ &\quad - j_l(kr')h_l^{(+)}(kr)]V(r')u_l(r') \\ &\equiv krj_l(kr) + g\int_0^r dr' \bar{\mathfrak{G}}_l(r, r'; k)V(r')u_l(r'), \end{aligned} \tag{A1}$$

for the scattering case with $\kappa^2 = k^2$, and

$$\begin{aligned} u_l(r) &= ig\mu r\int_0^r dr' r'[j_l(i\mu r')h_l^{(+)}(i\mu r') \\ &\quad - j_l(i\mu r')h_l^{(+)}(i\mu r)]V(r')u_l(r') \\ &\equiv g\int_0^r dr' \mathfrak{G}_l(r, r'; \mu)V(r')u_l(r'), \end{aligned} \tag{A2}$$

for the bound-problem case with $\kappa^2 = -\mu^2$. The functions $j_l(kr)$, $h_l^{(+)}(kr)$ are the spherical Bessel functions obeying

$$\begin{aligned} j_l(0) = 0, \quad j_l(kr) &\underset{r\to\infty}{\sim} \frac{\sin(kr - l\pi/2)}{kr}, \\ h_l^{(+)}(kr) &\underset{r\to\infty}{\sim} \frac{e^{+i(kr-l\pi/2)}}{kr}, \end{aligned} \tag{A3}$$

and the corresponding quantities in Eq. (A3) are the appropriate analytic continuations to imaginary k. The Jost function, as is well known, is identical with the Fredholm determinant,[4] which has a well-known power-series expansion. For Eq. (A1) this power series has the form[4]

$$\bar{\Delta}(g, k) = \sum_{n=0}^\infty (-g)^n L_n, \tag{A4}$$

where

$$\begin{aligned} L_n &= k\int_0^\infty dr_1 \int_0^{r_1} dr_2 \cdots \int_0^{r_{n-1}} dr_n r_1 h^{(+)}(kr_1) \\ &\quad \times \bar{\mathfrak{G}}(r_1, r_2; k)\cdots\bar{\mathfrak{G}}(r_{n-1}, r_n; k) \\ &\quad \times r_n j_l(kr_n)V(r_1)\cdots V(r_n) \\ &= \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n \bar{H}_l(r_1, r_2)\cdots\bar{H}_l(r_n, 0) \\ &\quad \times V(r_1)\cdots V(r_n), \end{aligned} \tag{A5}$$

where

$$\bar{H}_l(rs) = kr^2 \frac{h_l^{(+)}(kr)}{h_l^{(+)}(ks)}[j_l(kr)h_l^{(+)}(ks) - j_l(ks)h_l^{(+)}(kr)]. \tag{A6}$$

[The suppressed energy variable is k, while in Eqs. (20), (31), and (36) it is $-k$.] From the inequality

$$|H_l(rs)| < |r - s| \tag{A7}$$

proven in Appendix C we obtain the bound

$$\begin{aligned} |L_n| &\leq \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n (r_1 - r_2)\cdots(r_{n-1} - r_n) \\ &\quad \times r_n\,|V(r_1)\cdots V(r_n)|. \end{aligned} \tag{A8}$$

Through the introduction of the variables

$$\xi_n = r_n, \quad \xi_j = r_j - r_{j+1} \quad (j = 1, \cdots, n-1)$$

we find an alternative expression for the right-hand side of Eq. (A8)

$$\begin{aligned} |L_n| &\leq \int_0^\infty d\xi_1 \cdots \int_0^\infty d\xi_n \xi_1 \cdots \xi_n \\ &\quad \times |V(\xi_n)V(\xi_{n-1} + \xi_n)\cdots V(\xi_1 + \cdots + \xi_n)|. \end{aligned} \tag{A9}$$

Equations (A8) and (A9) correspond to the right-hand side of Eq. (9).

For the bound-state case one finds analogously

$$\Delta(g, \mu) \equiv \sum_{n=0}^\infty g^n L_n, \tag{A10}$$

with

$$\begin{aligned} L_n &= i\mu\int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n r_1 h_l^{(+)}(i\mu r_1) \\ &\quad \times \mathfrak{G}_l(r_1, r_2; \mu)\cdots\mathfrak{G}_l(r_{n-1}, r_n; \mu) r_n j_l(i\mu r_n) \\ &= \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n \\ &\quad \times H_l(r_1, r_2)\cdots H_l(r_{n-1}, r_n)H_l(r_n, 0), \end{aligned} \tag{A11}$$

with

$$\begin{aligned} H_l(rs) &= [i\mu r^2 h_l^{(+)}(i\mu r)/h_l^{(+)}(i\mu s)] \\ &\quad \times [j_l(i\mu r)h_l^{(+)}(i\mu s) - j_l(i\mu s)h_l^{(+)}(i\mu r)], \end{aligned} \tag{A12}$$

for which one again has an inequality (see Appendix C):

$$|H_l(rs)| \leq |r - s|, \tag{A13}$$

leading again to the bound, Eq. (9).

APPENDIX B: EVALUATION OF SOME INTEGRALS

We consider first the integral I_n defined by

$$I_n \equiv \int_0^\infty du_1 \cdots \int_0^\infty du_n \exp -(u_1 + \cdots + u_n)^p. \tag{B1}$$

We set $u_j = v_j^2$ and find

$$I_n = 2^n \int_0^\infty dv_1 \cdots \int_0^\infty dv_n v_1 \cdots v_n \exp -(v_1^2 + v_n^2)^p$$
$$= \left(\frac{4}{\pi}\right)^n \int_0^{\pi/2} d\varphi_1 \cdots \int_0^{\pi/2} d\varphi_n \int_0^\infty dr_1 v_1 \cdots \int_0^\infty dr_n v_n$$
$$\times \exp -(v_1^2 + \cdots + v_n^2)^p. \tag{B2}$$

We introduce the Cartesian variables $y_1, \cdots, y_{2n}$ defined by

$$y_{2k-1} = v_k \cos \varphi_k,$$
$$y_{2k} = v_k \sin \varphi_k \quad (k = 1, \cdots, n), \tag{B3}$$

which leads to

$$I_n = \left(\frac{4}{\pi}\right)^n \int_0^\infty dy_1 \cdots \int_0^\infty dy_{2n} \exp -(y_1^2 + \cdots + y_{2n}^2)^p, \tag{B4}$$

which in $2n$-dimensional spherical coordinates can be written

$$I_n = \Omega_{2n} \left(\frac{4}{\pi}\right)^n \int_0^\infty dy\, y^{2n-1} e^{-y^{2p}} = \frac{1}{2p} \left(\frac{4}{\pi}\right)^n \Omega_{2n} \Gamma\left(\frac{n}{p}\right), \tag{B5}$$

where Ω_{2n} is the total spherical angle in $2n$ dimensions and

$$y^2 = \sum_{j=1}^{2n} y_j^2.$$

From the known expression $\Omega_{2n} = 2\pi^n/\Gamma(n)$, we find that

$$I_n = \frac{1}{p} 4^n \frac{\Gamma(n/p)}{\Gamma(n)}, \tag{B6}$$

as expressed in Eq. (28).

In Eq. (33) we encounter an integral of the form

$$J_n(\sigma) \equiv \int_0^\infty du_1 \cdots \int_0^\infty du_n$$
$$\times \exp [\sigma(u_1 + \cdots + u_n) - (u_1 + \cdots + u_n)^p], \tag{B7}$$

where σ is to be a large parameter. The transformations applied to evaluate I_n result in

$$J_n(\sigma) = \Omega_{2n} \left(\frac{4}{\pi}\right)^n \int_0^\infty dy\, y^{2n-1} \exp (\sigma y^2 - y^{2p})$$
$$\equiv \Omega_{2n} \left(\frac{4}{\pi}\right)^n V_n(\sigma). \tag{B8}$$

A steepest-descents type of estimate may be applied to $V_n(\sigma)$. Setting

$$h(y) = -y^{2p} + \sigma y^2 + (2n - 1) \ln y, \tag{B9}$$

we find that the saddle point y_0 for which $h'(y_0) = 0$, obeys the equation

$$y_0^{2p} = \frac{\sigma}{p} y_0^2 + \frac{2n - 1}{2p}. \tag{B10}$$

With

$$\sigma = \sigma_n = \frac{1}{\beta} \kappa_{n/2} = \frac{8^{\frac{1}{2}} k_1}{\beta} n^{(1-1/p)} \equiv \theta n^{(1-1/p)}$$

one finds for large n that

$$y_0 \sim A n^{\frac{1}{2}p} [1 + O(1/n)], \tag{B11}$$

where

$$A^{2p} = \frac{\theta}{p} A^2 + \frac{2}{p}, \tag{B12}$$

so that for large θ (i.e., large k_1):

$$A(\theta) \sim \left(\frac{\theta}{p}\right)^{1/(2p-2)} \tag{B13}$$

Then, for large θ,

$$h(y_0) = \sigma(1 - 1/p) y_0^2 + (2n - 1)(\ln y_0 - 1/2p)$$
$$\sim \frac{1}{p} n \ln n + \left[\frac{p(p-1)}{p^{p/(p-1)}} \theta^{p/(p-1)} + \frac{1}{p-1} \ln \theta - \frac{\ln p}{p-1} - \frac{1}{2p}\right] n, \tag{B14}$$

and

$$h''(y_0) = 4\left[(1 - p)\sigma - \frac{p(n - \frac{1}{2})}{y_0^2}\right]$$
$$\sim 4(1 - p)\theta n^{(1-1/p)} \tag{B15}$$

and the steepest-descent estimate combined with factors in Eq. (B8) leads to the result in Eq. (34).

APPENDIX C: HIGHER PARTIAL WAVES

The inequalities of Eqs. (20) and (31) will be established for general partial waves for $\operatorname{Im} k \leq 0$. The $\operatorname{Im} k > 0$ case with higher partial waves will not be dealt with presently. In I the inequality[11]

$$|\bar{H}_l(rs)| = |kr^2[h_l^{(+)}(ks) j_l(kr) - h_l^{(+)}(kr) j_l(ks)] h_l^{(+)}(kr)/h^{(+)}(ks)|$$
$$\leq |r - s|, \tag{C1}$$

$(r > s)$ was quoted but not proven. We now prove it. We define

$$u_+(r) \equiv kr h_l^{(+)}(kr). \tag{C2}$$

[11] Reference 7, line following Eq. (53).

Since $u_+(r)$ solves

$$\left[\frac{d^2}{dr^2} + k^2 - \frac{(l+1)}{r^2}\right]u_+(r) = 0, \tag{C3}$$

so does

$$u_0(r) \equiv u_+(r)\int_0^r \frac{dr'}{u_+(r)^2}. \tag{C4}$$

Clearly $u_0(0) = 0$, while, for large r,

$$u_0(r) \sim \frac{1}{2ik}\left[e^{-i(kr-l\pi/2)} + Ce^{ikr}\right] + O\left(\frac{1}{r}\right) \tag{C5}$$

with C some constant, since for large r:

$$u_+(r) \sim e^{-i(kr-l\pi/2)}\,[1 + O(1/r)]. \tag{C6}$$

This determines that

$$u_0(r) = rj_l(kr) \tag{C7}$$

whence

$$[h_l^{(+)}(ks)j_l(kr) - h_l^{(+)}(kr)j_l(ks)] = \frac{1}{krs}u_+(r)u_+(s)\int_s^r \frac{dr'}{u_+(r')^2}. \tag{C8}$$

Then

$$|\bar{H}_l(rs)| \leq |u_+(r)|^2\int_s^r \frac{dr'}{|u_+(r')|^2}. \tag{C9}$$

Since, as we shall show $|u_+(r)|^2$ is a monotonically decreasing function of the radial argument we readily conclude that

$$|\bar{H}_l(rs)| \leq |r - s|.$$

We offer a physical proof that

$$|u_+(r)|^2 \equiv k^2r^2\,|h^{(+)}(kr)|^2$$

is a monotonically decreasing function of r. Consider a particle of momentum k moving in one dimension in the direction of the positive x axis through the potential barrier

$$U(x) \equiv \theta(x - a)\frac{l(l+1)}{x^2} \tag{C10}$$

(see Fig. 1) with a some arbitrary nonnegative value of x. [$\theta(x)$ denotes the step function which is unity for positive x.] The wavefunction of the particle in the region of this barrier is of the form

$$\psi(x) \equiv B \times h_l^{(+)}(kx). \tag{C11}$$

Clearly from physical considerations the probability density $\rho = |\psi(x)^2|$ will be monotonically decreasing in the direction of increasing x, first due to reflection and second due to the increase in local velocity with increasing x, since current density is conserved. In fact if $l > l'$

$$|rh_l^{(+)}(kr)|^2 \leq |rh_{l'}^{(+)}(kr)|^2, \tag{C12}$$

since the corresponding repulsive barrier is weaker.

The parallel inequality [Eq. (16)] of I

$$|H_l(rs)| = |\mu r^2[h_l^{(+)}(i\mu s)j_l(i\mu r) - h_l^{(+)}(i\mu r)j_l(i\mu s)] \times h_l^{(+)}(\mu r)/h_l^{(+)}(\mu s)| \leq 1, \tag{C13}$$

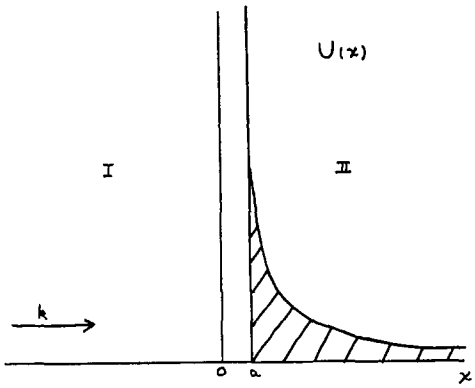


FIG. 1. A stream of monoenergetic particles with momentum to k travels from $x = -\infty$ to $x = +\infty$, encountering the repulsive potential barrier in region II, $U(x) = \theta(x - a)l(l+1)/\lambda^2$. Reflection by the barrier in region II causes the probability density to decrease with increasing x.

can be proven by similar methods. In a parallel notation, let

$$v_+(r) \equiv i\mu r h_l^{(+)}(i\mu r), \tag{C14}$$

which obeys

$$\left[\frac{d^2}{dr^2} - \mu^2 - \frac{l(l+1)}{r^2}\right]v_+(r) = 0, \tag{C15}$$

then

$$v_0(r) = v_+(r)\int_0^r \frac{dr'}{v_+(r')^2} \tag{C16}$$

is another solution of the same equation to be identified as

$$v_0(r) = irj_l(i\mu r), \tag{C17}$$

so that

$$|H_l(rs)| \leq |v_+(r)|^2\int_s^r \frac{dr'}{[v_+(r')]^2}. \tag{C18}$$

The function $v_+(r)$ is identical with

$$v_+(r) = e^{(-i\pi)/2(l+\frac{1}{2})}\left(\frac{2r}{\pi}\right)^{\frac{1}{2}}K_{l+\frac{1}{2}}(r), \tag{C19}$$

where $K_n(r)$ is the Bessel function of imaginary argument, which damps exponentially for large positive values of its argument. In terms of the integral representation[12]

$$K_n(r) = \frac{\Gamma(\frac{1}{2})}{\Gamma(n+\frac{1}{2})}\left(\frac{z}{2}\right)^n\int_1^\infty dt\, e^{-zt}(t^2-1)^{n-\frac{1}{2}} = \frac{\pi^{\frac{1}{2}}2^{-n}}{\Gamma(n+\frac{1}{2})}\frac{1}{z^{n-1}}\int_z^\infty du\, e^{-u}(u^2 - z^2)^{n-\frac{1}{2}}. \tag{C20}$$

We readily find by explicit differentiation that

$$|v_+(r)| = \frac{2^{-l}}{\Gamma(l+1)}r^{-l+1}\int_r^\infty du\, e^{-u}(u^2 - r^2)^l \tag{C21}$$

[12] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, The Macmillan Co., New York, 1944), 2nd ed., Sec. 6.3.

decreases with increasing r. This, via Eq. (C18), leads to the inequality of Eq. (C13).

An inequality of the form

$$|\bar{H}_l(rs)| \leq A/k \tag{C22}$$

can be established with A independent of l. In terms of the previous notation ($r > s$):

$$|\bar{H}_l(rs)| = \frac{1}{k}\left| u_0(r)u_+(s) - \frac{u_0(s)u_+^2(r)}{u_+(s)} \right| \leq \frac{1}{k}\left[|u_0(r)u_+(r)| + |u_0(s)u_+(r)|\right], \tag{C23}$$

in view of earlier results. From the inequalities for real positive k[13]:

$$|u_0(r)| \leq C\left(\frac{kr}{1+kr}\right)^{l+1}, \tag{C24a}$$

$$|u_+(r)| \leq C\left(\frac{kr}{1+kr}\right)^{-l}, \tag{C24b}$$

one readily concludes that

$$|\bar{H}_l(rs)| \leq 2C^2/k. \tag{C25}$$

[13] Reference 8, Eqs. (3.9).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Derivation of Nonrelativistic Sum Rules from the Causality Condition of Wigner and Van Kampen

K. CHADAN* AND A. MONTES*
Laboratoire de Physique Théorique et Hautes Energies,† Orsay, France

(Received 19 January 1968)

Nonrelativistic sum rules previously obtained for each phase shift in the framework of elastic scattering by a local central potential of finite radius are shown to follow from the causality principle of Wigner and Van Kampen, under the assumptions that the phase shift satisfies the Levinson theorem and that, at high energies, the integrability condition holds. In fact, it is shown that under these assumptions, any interaction of finite radius which is causal in the sense of Wigner and Van Kampen is equivalent to a local potential of the same radius. First we recall the sum rules and mention some of their applications. We give a brief survey of their proof in potential scattering, which is based essentially on the analytic and asymptotic properties of the Jost function (or the S matrix). Then we show that, under the assumptions mentioned above on the phase shifts, the properties of the R matrix derived by Wigner and Van Kampen from causality lead to the same analytic and asymptotic properties of the Jost function (defined now directly from the S matrix) as in potential scattering, providing, therefore, sufficient information for the direct derivation of the sum rules. Finally, using the Gel'fand–Levitan and Marchenko integral equations of the inverse-scattering problem, we show that, in fact, this information is sufficient to entail that the causal interaction of Wigner and Van Kampen is equivalent to a local potential of the same radius.

I. INTRODUCTION

In a previous paper,[1] one of the authors has derived nonrelativistic sum rules in the framework of potential scattering with central potentials of finite radius. These sum rules are integro-differential relations between each of the phase shifts and the binding energies of the possible bound states with the same angular momentum. Denoting these binding energies by $E_j = -\gamma_j^2, j = 1, 2, \cdots, n$, the sum rules read[2]

$$r + \frac{d\delta_l}{dk} = -2\sum_{j=1}^{n} \frac{\gamma_j}{k^2+\gamma_j^2} - \frac{1}{\pi}\int_{-\infty}^{\infty} dz \times \frac{\log|\cos[zr + \delta_l(k+z) - \delta_l(k) + \theta(k+z) - \theta(k)]|}{z^2}, \tag{1}$$

$$\theta(x) = -2\sum_{j=1}^{n} \tan^{-1}\frac{\gamma_j}{x}. \tag{2}$$

They are valid for all values of r greater than the radius R of the potential, so that we have in fact a continuous infinity of sum rules.

Different types of sum rules have also been obtained in the more general case of potentials which may have

* Postal address: Laboratoire de Physique Théorique et Hautes Energies, Bâtiment 211, Faculté des Sciences, 91-Orsay, France.
† Laboratoire associé au C.N.R.S.

[1] K. Chadan, Nuovo Cimento **40**, 1194 (1965).

[2] As usual, we choose the units in such a way that $\hbar = 2M = 1$, where M is the mass of the particle. The energy is then given by $E = k^2$, k being the wave number. For negative values of k, the phase shifts are defined by $\delta_l(k) = -\delta_l(-k)$.

decreases with increasing r. This, via Eq. (C18), leads to the inequality of Eq. (C13).

An inequality of the form

$$|\bar{H}_l(rs)| \leq A/k \tag{C22}$$

can be established with A independent of l. In terms of the previous notation ($r > s$):

$$|\bar{H}_l(rs)| = \frac{1}{k}\left| u_0(r)u_+(s) - \frac{u_0(s)u_+^2(r)}{u_+(s)} \right| \leq \frac{1}{k}\left[|u_0(r)u_+(r)| + |u_0(s)u_+(r)|\right], \tag{C23}$$

in view of earlier results. From the inequalities for real positive k[13]:

$$|u_0(r)| \leq C\left(\frac{kr}{1+kr}\right)^{l+1}, \tag{C24a}$$

$$|u_+(r)| \leq C\left(\frac{kr}{1+kr}\right)^{-l}, \tag{C24b}$$

one readily concludes that

$$|\bar{H}_l(rs)| \leq 2C^2/k. \tag{C25}$$

[13] Reference 8, Eqs. (3.9).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Derivation of Nonrelativistic Sum Rules from the Causality Condition of Wigner and Van Kampen

K. CHADAN* AND A. MONTES*
Laboratoire de Physique Théorique et Hautes Energies,† Orsay, France

(Received 19 January 1968)

Nonrelativistic sum rules previously obtained for each phase shift in the framework of elastic scattering by a local central potential of finite radius are shown to follow from the causality principle of Wigner and Van Kampen, under the assumptions that the phase shift satisfies the Levinson theorem and that, at high energies, the integrability condition holds. In fact, it is shown that under these assumptions, any interaction of finite radius which is causal in the sense of Wigner and Van Kampen is equivalent to a local potential of the same radius. First we recall the sum rules and mention some of their applications. We give a brief survey of their proof in potential scattering, which is based essentially on the analytic and asymptotic properties of the Jost function (or the S matrix). Then we show that, under the assumptions mentioned above on the phase shifts, the properties of the R matrix derived by Wigner and Van Kampen from causality lead to the same analytic and asymptotic properties of the Jost function (defined now directly from the S matrix) as in potential scattering, providing, therefore, sufficient information for the direct derivation of the sum rules. Finally, using the Gel'fand–Levitan and Marchenko integral equations of the inverse-scattering problem, we show that, in fact, this information is sufficient to entail that the causal interaction of Wigner and Van Kampen is equivalent to a local potential of the same radius.

I. INTRODUCTION

In a previous paper,[1] one of the authors has derived nonrelativistic sum rules in the framework of potential scattering with central potentials of finite radius. These sum rules are integro-differential relations between each of the phase shifts and the binding energies of the possible bound states with the same angular momentum. Denoting these binding energies by $E_j = -\gamma_j^2, j = 1, 2, \cdots, n$, the sum rules read[2]

$$r + \frac{d\delta_l}{dk} = -2\sum_{j=1}^{n} \frac{\gamma_j}{k^2+\gamma_j^2} - \frac{1}{\pi}\int_{-\infty}^{\infty} dz \times \frac{\log|\cos[zr + \delta_l(k+z) - \delta_l(k) + \theta(k+z) - \theta(k)]|}{z^2}, \tag{1}$$

$$\theta(x) = -2\sum_{j=1}^{n} \tan^{-1}\frac{\gamma_j}{x}. \tag{2}$$

They are valid for all values of r greater than the radius R of the potential, so that we have in fact a continuous infinity of sum rules.

Different types of sum rules have also been obtained in the more general case of potentials which may have

* Postal address: Laboratoire de Physique Théorique et Hautes Energies, Bâtiment 211, Faculté des Sciences, 91-Orsay, France.
† Laboratoire associé au C.N.R.S.

[1] K. Chadan, Nuovo Cimento **40**, 1194 (1965).

[2] As usual, we choose the units in such a way that $\hbar = 2M = 1$, where M is the mass of the particle. The energy is then given by $E = k^2$, k being the wave number. For negative values of k, the phase shifts are defined by $\delta_l(k) = -\delta_l(-k)$.

an infinite tail.[3] These new sum rules, more complicated than (1), are of course also valid inside a potential of finite radius ($r \leq R$). If the potential has a finite radius R, and if r is outside this radius, the new sum rules, for the S-wave case, read

$$r + \frac{d\delta_0}{dk} = S(k, r) - \frac{P}{\pi}\int_{-\infty}^{\infty} dk' \times \log\left|\frac{k \sin [k'r + \delta_0(k')]}{k' \sin [kr + \delta_0(k)]}\right| \Big/ (k' - k)^{-2}, \quad r \geq R, \quad (3)$$

where $S(k, r)$ is given, in general, by[4]

$$S(k, r) = 2\sum_{j=1}^{n}\left[\frac{\gamma_j(r)}{k^2 + \gamma_j^2(r)} - \frac{\gamma_j}{k^2 + \gamma_j^2}\right], \quad (4)$$

$i\gamma_j(r)$ being the imaginary zeros, in the upper-half k plane, of the regular solution $\varphi_l(k, r)$ of the radial Schrödinger equation, normalized according to[5,6]

$$\lim_{r\to 0} r^{-l-1}[(2l + 1)!!]\varphi_l(k, r) = 1. \quad (5)$$

For other waves, one simply has to replace in (3) the sine functions by appropriate combinations of the spherical Hankel functions $w_l(kr)$, and exp $(\pm i\delta_l)$.[5]

Both types of sum rules may be used in practice to calculate in a very simple and direct manner the radius of the potential and the energies of the bound states from the knowledge of the phase shift, without using the Schrödinger equation. Such calculations have been performed in the case of neutron–proton singlet and triplet interactions at low energies.[4] The results are in very good agreement with the experimental results of effective-range theory and the deuteron binding energy. As is seen from (1) and (3), no internal quantities like the potential or the internal wavefunction appear in any form in these sum rules.[7]

One may therefore ask whether it would be possible to derive the above sum rules from some general principles, without making use of the Schrödinger equation in the interior region. Everything in (1) and (3) being directly given by the phase shifts [or the R matrix of Wigner, formula (44) below], one thinks immediately of the causality principle, and its consequences on the analytic properties of the R matrix, as an "ersatz" for the Schrödinger equation in the interior region. The purpose of the present paper is to show that, indeed, the sum rules can be directly derived from the results of Wigner on the R matrix.[8] In fact, one has to make, at some point of the proof, a mild technical assumption about the asymptotic behavior of the phase shifts in the limit of infinite k, namely, that

$$\delta_l(k) \underset{k\to\infty}{\simeq} -N\pi + \eta(k), \quad (6a)$$

where N is a nonnegative integer, and that

$$\int^{\infty} k^{-1}|\eta(k)|\, dk < \infty, \quad (6b)$$

which is true in general in potential scattering if (see Appendix A):

$$\int_0^R r\,|V(r)|\, dr < \infty.$$

The reason for taking N nonnegative is as follows. As we shall see later, the phase shift is an analytic function of k in the neighborhood of the real axis, and because of the symmetry property $\delta_l(-k) = -\delta_l(k)$, we take $\delta_l(0) = 0$. On the other hand, if the Levinson theorem[5]—which is true in potential scattering—holds, we have $\delta(\infty) = -n\pi$, n being the number of bound states. Here, wishing to derive our sum rules with a finite number of bound states, we must assume something of the sort (6a). Moreover, the interaction cannot be equivalent to a local potential unless the Levinson theorem holds. Although we have some reasons to believe that condition (6a) is also a consequence of the causality condition and the finiteness of the number of bound states, and not an extra

[3] K. Chadan, Nuovo Cimento **41**, 115 (1966); **44**, 838 (1966).

[4] K. Chadan and A. Montes Lozano, Phys. Rev. **164**, 1762 (1967). It is shown there that (1) these zeros, and their complex conjugates, are the only complex zeros of $\varphi_l(k, r)$; (2) $\gamma_j(r)$ are monotonically increasing functions of r; (3) $\lim_{r\to\infty} \gamma_j(r) = \gamma_j$.

[5] For the study of the solutions of the radial Schrödinger equation we refer the reader to the review article by R. G. Newton, J. Math. Phys. **1**, 319 (1960). We shall follow the notation of this article.

[6] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Book Co., New York, 1966); V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland Publ. Co., Amsterdam, 1965).

[7] At first sight, it may seem that, even for $r \geq R$, the zeros $\{i\gamma_j(r)\}$ are somehow explicitly related to the internal region because they are the zeros of $\varphi_l(k, r)$ normalized according to (5). In the absence of any information on the internal region, the wavefunction outside is given by (see Ref. 5) $\varphi_l(k, r) = A(k)[u_l(kr) - \tan\delta_l v_l(kr)]$ with an arbitrary $A(k)$, which depends explicitly on the internal region. However, it is shown in Ref. 4 that $\{\gamma_j(r)\}$ depend explicitly on r, with strictly positive derivatives, so that they necessarily come from the expression inside the bracket, and not from $A(k)$. It follows that, for $r \geq R$, $\{\gamma_j(r)\}$ and $\{\gamma_j\}$ can be calculated directly from the phase shift. This is of course a matter of principle. In practice, one knows the phase shifts only numerically, and an analytic continuation outside the real axis of the expression inside the bracket is in general impossible. This is why, in practice, one has to use the sum rules to obtain $\{\gamma_j(r)\}$ and $\{\gamma_j\}$, as done in Ref. 4.

[8] E. P. Wigner, *Proceedings of the International School of Physics "Enrico Fermi," Varenna, Course 29* (Academic Press Inc., New York, 1964). These lectures contain complete references to earlier works. The R-matrix theorem was first proved by Wigner and coworkers from the assumption that the Hamiltonian is a self-adjoint operator bounded below and having a finite radius. It was then realized that the same conclusions on the properties of the R matrix could be obtained if one uses only the causality condition. Both approaches are described by Wigner in the above lectures. See also N. G. Van Kampen [Phys. Rev. **91**, 1267 (1953)], who studies the implications of the causality condition directly on the S matrix. His results are somewhat weaker than those of Wigner on the R matrix. However, with some weak additional assumptions, Van Kampen [Physica **20**, 115 (1954)], is able to show that the S matrix has analytic properties closely related to those of Wigner's R matrix

assumption, we admit it as such with no further comments.

In Sec. II, we give, for the sake of completeness, a brief survey of the essential mathematical tools and the analytic properties of various functions which are used to prove the sum rules in potential scattering. In Sec. III, we summarize the analytic properties of the R matrix obtained by Wigner and Van Kampen from causality, and show their bearings on the properties of the functions considered in Sec. II. It turns out that the analytic properties thus obtained are identical with those of Sec. II, and therefore that the sum rules can be derived directly from causality. In fact, in Sec. IV, we prove that causality and conditions (6a), (6b) are equivalent to a local potential. In other words, we show that any causal interaction for which the phase shift satisfies (6a), (6b) is equivalent to a local potential. Finally, in the several Appendices we work out in detail the proofs of some technical points which otherwise may obscure the progression of the reasoning.

II. MATHEMATICAL TOOLS AND A BRIEF SURVEY OF THE PROOFS IN POTENTIAL SCATTERING

The definitions and theorems we begin this section with are given in the book of Boas on entire functions,[9] to which the reader is referred for details and original references. We deal here exclusively with entire functions of order one and finite type. They are commonly called functions of exponential type (exponential type τ, say, if the type is needed). The Phragmén–Lindelöf indicator function of such a function $f(z)$, is defined by

$$h_f(\theta) = \limsup_{r\to\infty} r^{-1} \log |f(re^{i\theta})|. \tag{7}$$

The function $h_f(\theta)$, for a given value of θ, is either finite or $-\infty$.

Class P: This class has been introduced by Levin.[9] Let $f(z)$ be an entire function of exponential type τ (finite). $f(z)$ is said to belong to the class P if the two following conditions are satisfied:

(a) $f(z)$ has no zeros in $\operatorname{Im} z < 0$, (b) $h_f(-\alpha) \geq h_f(\alpha)$ for some α in the range $0 < \alpha < \pi$.

For functions of the class P, we can prove the following.

Theorem 1[9]: If $f(z) \in P$, one has

$$|f(z)| \geq |f(z^*)| \tag{8}$$

for all z in $\operatorname{Im} z < 0$. In other words, if (8) holds on one ray, even in the weak form of condition (b) above, it holds on all rays.

This theorem is essential in what follows. As we shall see, we have mostly to deal with functions of the form

$$g(z) = \exp(i\alpha) f(z) \pm \exp(-i\alpha) f(-z), \ \alpha \text{ real}, \tag{9}$$

where $f(z) \in P$, with strict inequality sign in (8),[10] and

$$f(-z) = [f(z^*)]^*. \tag{10}$$

It is then obvious that because of (8), $|f(z)| > |f(-z)|$ if $\operatorname{Im} z < 0$. Therefore, $g(z)$ cannot vanish in the lower half-plane. The same is true in $\operatorname{Im} z > 0$ because there $|f(z)| < |f(-z)|$, so that, finally, the zeros of $g(z)$ are all real. The importance of this remark stems from the fact that our sum rules (1) are nothing else but the results of the following theorems of Paley and Wiener,[9] and Pfluger,[9] applied to appropriate solutions of the Schrödinger equation outside the potential ($r > R$), which have real zeros in the k plane (a fact which is proved exactly in the above manner).

Theorem 2[11]: If $f(z)$ is an entire function of exponential type with real zeros only, and $f(0) = 1$, the three conditions

$$h_f\left(\pm\frac{\pi}{2}\right) = \lim_{y\to\infty} |y|^{-1} |f(\pm iy)| = \pi(B \mp \operatorname{Im}\alpha), \tag{11a}$$

$$\lim_{A\to\infty} \frac{n(A)}{A} = 2B, \tag{11b}$$

$$\lim_{A\to\infty} P\int_{-A}^{A} \frac{\log|f(x)|}{x^2}\, dx = -\pi^2 B \tag{11c}$$

are equivalent, $n(A)$ denotes the number of zeros of $f(z)$ in $|z| < A$, counted according to their multiplicity, and P means the principal value of the integral.

Suppose now that $f(z)$ has arbitrary zeros, real or complex, and is bounded on the real axis. We have the following theorem on the density of the zeros.

[9] R. P. Boas, *Entire Functions* (Academic Press Inc., New York, 1954).

[10] It is obvious that if $f(z) \in P$, (8) holds with strict inequality sign for $f_\epsilon(z) = e^{i\epsilon z} f(z)$, no matter how small ϵ (> 0) is. It follows that $g_\epsilon(z)$, defined by (9) with $f_\epsilon(z)$, has all its zeros real for any $f(z) \in P$. In general, by the continuity theorem, $g(z) = \lim_{\epsilon\to 0} g_\epsilon(z)$ will also have only real zeros, except in some pathological cases where the continuity theorem breaks down.

[11] This theorem, as stated here, is an obvious modification of the Theorem 8.2.1. of Ref. 9. According to Hadamard's factorization theorem of entire functions, we multiply the canonical product considered in the book by $\exp(\alpha z)$ to cover all cases. If the function is real on the real axis, we have $\operatorname{Im}\alpha = 0$. This is in fact what happens later, when we use Theorem 2 to derive (1).

Theorem 3[12]: Let $f(z)$ be an entire function of exponential type τ in each of the half-planes $\operatorname{Im} z > 0$ and $\operatorname{Im} z < 0$. Let $f(z)$ be bounded on the real axis. If we denote by $n_{\pm}(A)$ the number of zeros with modulus less than A in $\operatorname{Re} z > 0$ and $\operatorname{Re} z < 0$, respectively, then

$$\lim_{A\to\infty} \frac{n_+(A)}{A} = \lim_{A\to\infty} \frac{n_-(A)}{A} = \frac{\tau}{\pi}. \tag{12}$$

In other words, the zeros have the density τ/π in each of the above half-planes (total density $2\tau/\pi$).

For functions of exponential type with arbitrary zeros, we have the following theorem.

Theorem 4[13]: Let $f(z)$ be an entire function of exponential type τ, z_n its zeros, and $f(0) = 1$. The conditions

$$\pi C = \sum_1^\infty \frac{|\sin\theta_n|}{r_n} < \infty, \quad r_n = |z_n|, \quad \theta_n = \arg z_n, \tag{13a}$$

and

$$\lim_{A\to\infty} \frac{n(A)}{A} = 2B, \tag{13b}$$

together, are equivalent to

$$\lim_{A\to\infty} \frac{P}{\pi} \int_{-A}^{A} \frac{\log|f(x)|}{x^2}\, dx = -\pi(B - C). \tag{13c}$$

As we shall see, Theorems 3 and 4 are the basic tools for the derivation of the sum rules (3).

Having now collected all the necessary definitions and theorems, we show briefly how the sum rules may be derived in potential scattering. As is well known, if the potential satisfies

$$\int_0^\infty r\,|V(r)|\, dr < \infty \tag{14a}$$

and

$$\int_0^\infty r^2\,|V(r)|\, dr < \infty, \tag{14b}$$

the scattering process and the bound states may be entirely described in terms of the so-called Jost function $f_l(k)$ (Refs. 5 and 6). If we denote by $f_l(\pm k, r)$ the two independent solutions of the radial Schrödinger equation satisfying

$$\lim_{r\to\infty} \exp(\pm ikr) f_l(\pm k, r) = i^l, \tag{15}$$

the Jost function $f_l(k)$ is defined by

$$\varphi_l(k, r) = \tfrac{1}{2} i k^{-l-1} [f_l(-k) f_l(k, r) - (-1)^l f_l(k) f_l(-k, r)], \tag{16}$$

φ_l being the regular solution, normalized according to (5). The properties of the above functions are as follows[5]:

(A) $\varphi_l(k, r)$, for each value of r, is a real[14] even entire function of k of order 1 and type r. Its asymptotic properties in the k plane are given by

$$\varphi_l(k, r) \underset{|k|\to\infty}{=} k^{-l-1} \sin(kr - \tfrac{1}{2} l\pi) + o(|k|^{-l-1} \exp|\operatorname{Im} k|\, r). \tag{17}$$

(B) For $r \neq 0$, $f_l(k, r)$ is regular analytic in $\operatorname{Im} k < 0$, and continuous on $\operatorname{Im} k = 0$ except at $k = 0$. Also, if $r \neq 0$,

$$\lim_{k\to 0} k^l f_l(k, r) \tag{18}$$

exists if the limit is carried out in the region of regularity. As $|k| \to \infty$ in $\operatorname{Im} k \leq 0$, we have

$$f_l(k, r) = i^l e^{-ikr} + o(\exp \operatorname{Im} kr). \tag{19}$$

The "symmetry" property[15]

$$f_l(-k^*, r) = (-1)^l [f_l(k, r)]^* \tag{20}$$

follows from the radial Schrödinger equation [$V(r)$ real] and (15). Notice that both k and $-k^*$ are in the same half-plane.

If the potential has a finite radius, $k^l f_l(k, r)$ becomes an entire function of k for $r \neq 0$.

(C) The Jost function, which can also be defined by

$$f_l(k) = k^l W[f_l(k, r), \varphi_l(k, r)], \tag{21}$$

W being the Wronskian, is a regular analytic function of k in $\operatorname{Im} k < 0$, and continuous in $\operatorname{Im} k \leq 0$. From its definition and (20), it follows that, wherever it exists,

$$f_l(-k^*) = [f_l(k)]^*. \tag{22}$$

This shows that $f_l(k)$ is real on the imaginary axis.

Note that, in general, (16) makes sense only for real values of k. In the limit of large k, we have

$$\lim_{\substack{|k|\to\infty \\ \operatorname{Im} k \leq 0}} f_l(k) = 1. \tag{23}$$

On the real axis, according to (22) we have

$$f_l(\pm k) = |f_l(\pm k)| \exp[\pm i\delta_l(k)] \tag{24}$$

[12] N. Levinson, *Gap and Density Theorems* (Am. Math. Soc. Colloquium Publ., No. 26, New York, 1940), Chap. III; P. Koosis, Bull. Soc. Math. France **86**, 27 (1958). The density of the zeros is defined by $\lim_{A\to\infty} n(A)/A$, $n(A)$ being the number of zeros in $|z| < A$ counted according to their multiplicity. Similarly, one can define the density in a half-plane (left, right, upper, lower, $\cdots$, etc.).

[13] Reference 9, p. 143, Theorem 8.41.

[14] Real in the sense of analytic functions. It is real on the real axis, so that, by the Schwarz reflection principle, $\varphi_l(k^*, r) = \varphi_l(k, r)^*$. Also, because of its evenness, φ_l is real on the imaginary k axis.

[15] It follows from (20) that $i^l f_l(k, r)$ is real on the imaginary axis.

and

$$\delta_l(-k) = -\delta_l(k), \tag{25}$$

$\delta_l(k)$ being the phase of $f_l(k)$. Now, using this, together with (15) and (16), we see that, for k real, we have

$$\varphi_l(k, r) \underset{r\to\infty}{\simeq} (-1)^l k^{-l-1} |f_l(k)| \sin [kr - \tfrac{1}{2}l\pi + \delta_l(k)], \tag{26}$$

so that $\delta_l(k)$ is the phase shift. The S matrix on the real axis is therefore given by

$$S_l(k) = \exp [2i\delta_l(k)] = f_l(k)[f_l(-k)]^{-1}. \tag{27}$$

If there are bound states, they are determined by the zeros of $f_l(k)$ in $\operatorname{Im} k < 0$.[16] These zeros, $k_j = -i\chi_j$, $j = 1, 2, \cdots, n$, are simple, situated on the imaginary axis, finite in number, and they correspond to bound states with energies $E_j = -\chi_j^2$. On the real axis, $f_l(k)$ never vanishes.[16] Finally, the phase shifts satisfy the Levinson theorem

$$\delta_l(0) - \delta_l(\infty) = n\pi. \tag{28}$$

If the potential vanishes outside a radius R, $f_l(k)$ becomes an entire function of order 1 and type $2R$. More precisely,

$$[f_l(k) - 1] \exp (2i\,k\,R) = o(1), \qquad |k| \to \infty, \quad \operatorname{Im} k > 0. \tag{29}$$

This, together with (23) and Theorem 1 show that,[1] for $r \geq R$,

$$g(k, r) = \prod_{j=1}^{n} \frac{k - i\chi_j}{k + i\chi_j} \exp (i\,k\,r) f_l(k) \in P, \tag{30}$$

its type being r. Note that, if $r > R$, $g(k, r)$ satisfies (8) with strict inequality sign. Also, $g(k, r)$ satisfies (22) [or (10)] in the entire plane. As is easily seen, all of the above properties are also valid for $g[(k' - k), r]$, k being fixed and real. Combining now $\exp (i\alpha)g[(k' - k), r]$ with $\exp (-i\alpha)g[-(k' - k), r]$, with an appropriate phase α, we obtain functions similar to (9), with real zeros, to which, after normalizing them to 1 at $k' = 0$, we may apply Theorem 2. This, after some easy algebra, leads to two kinds of sum rules.[17] If we take the plus sign in the above combination, we obtain (1), whereas if we take the minus sign, we obtain formulas similar to (3) in which $S(k, r)$, the zeros now being all real, is given by (4) with all $\gamma_j(r) = 0$ [i.e., by a sum identical with the one that figures in the right-hand side of (1)], but where $\theta(k)$, Eq. (2), is present in the argument of the sine functions.

[16] We leave out the possibility of $f_l(0) = 0$, which corresponds to a resonance ($l = 0$) or a bound state ($l \neq 0$), at zero energy. There is no loss of generality in doing so. One can handle the problem in the same way (Ref. 1) as when $f_l(0) \neq 0$.

[17] For details, see Ref. 1.

The derivation of (3) is similar to that of the previous case. We consider now the wavefunction φ_l outside the potential ($r \geq R$). It is given by (16) where $f_l(k, r)$ is now the free solution[5]

$$w_l(k\,r) = -i\,k\,r\,h_l^{(2)}(k\,r), \tag{31}$$

$$w_l(z) \underset{|z|\to\infty}{=} i^l e^{-iz}\left[1 + O\left(\frac{1}{|z|}\right)\right], \tag{32}$$

$h_l^{(2)}(z)$ being the spherical Hankel function. From (17) and Theorem 3, it follows that[3] the zeros of φ_l have a total density $2\,r/\pi$. This, together with the properties of the complex zeros[4] $\pm i\gamma_j(r)$ of φ_l, shows that we can apply Theorem 4 to $\varphi_l(k' + k, r)/\varphi_l(k, r)$, k being fixed and real. The net result is the sum rule (3) and the similar ones for higher l.

From the above considerations, it is now clear that the essential ingredients in the derivation of the sum rules are the analytic and asymptotic properties of $f_l(k)$, namely that $f_l(k)$ is an entire function of order 1 and type $2R$, satisfies (22), (23), and (29), and that its zeros in the lower half-plane are pure imaginary and correspond to bound states. To derive the sum rules from causality and the analytic properties of the R matrix, we have to show that the S matrix [see formula (47) below] can be written in the form (27), where $f_l(k)$, the "Jost function," has all the essential properties, mentioned above, of the true Jost function of potential scattering. This is the main conclusion of the present paper, the proof of which we devote Sec. III.

III. "JOST FUNCTION" IN R-MATRIX THEORY

By definition, the R matrix relates the value of the wavefunction to its normal derivative on the sphere inside which the interaction takes place. Assuming the interaction to be invariant under spatial rotations, we can consider separately each angular-momentum state. The wavefunction outside the interaction region being given in general by

$$\psi_l(k, r) = A_l(k)w_l(kr) + B_l(k)w_l(-kr), \tag{33}$$

where $w_l(kr)$ is the free solution introduced before, Eq. (31), the R matrix (rather its lth element) is simply

$$R_l(E) = \psi_l(k, R)/\psi_l'(k, R). \tag{34}$$

The R theorem of Wigner[8] now states that $R_l(E)$ is a real[14] meromorphic function of E of the following form:

$$R_l(E) = \sum_\lambda \frac{\gamma_{\lambda l}^2}{E_{\lambda l} - E}, \tag{35}$$

where $\gamma_{\lambda l}$ and $E_{\lambda l}$ are real constants. The poles $E_{\lambda l}$

may approach infinity in either direction but cannot accumulate in a finite point. It is obvious from (35) that $R_l(E)$ is a Herglotz function,[18] and maps the upper and lower half-planes into themselves. In other words,

$$\operatorname{Im} R_l(E) \cdot \operatorname{Im} E > 0 \quad \text{unless} \quad \operatorname{Im} E = 0. \tag{36}$$

Also,

$$\sum_\lambda \frac{\gamma_{\lambda l}^2}{E_{\lambda l}} < \infty. \tag{37}$$

Let us now see the relation between $R_l(E)$ and the S matrix. From (33) and (32), we have

$$S_l(k) = \exp[2i\delta_l(k)] = -(-1)^l B_l(k)[A_l(k)]^{-1}. \tag{38}$$

On the other hand, using the Wronskian[19]

$$W[w_l(kr), w_l(-kr)] = (-1)^l 2ik, \tag{39}$$

$$W[f, g] \equiv fg' - f'g \tag{40}$$

in (33), we obtain

$$A_l(k) = -(-1)^l (2ik)^{-1} W[w_l(-kr), \psi_l(k, r)] \tag{41}$$

and

$$B_l(k) = (-1)^l (2ik)^{-1} W[w_l(kr), \psi_l(k, r)], \tag{42}$$

so that

$$S_l(k) = \frac{W[k^l w_l(kr), \psi_l(k, r)]}{W[(-k)^l w_l(-kr), \psi_l(k, r)]}. \tag{43}$$

From this, (33), and (34), it is easily found that

$$R_l(E) = \frac{-w_l(kR) + (-1)^l S_l(k) w_l(-kR)}{-w_l'(kR) + (-1)^l S_l(k) w_l'(-kR)}, \tag{44}$$

$$w_l'(kR) = \frac{d}{dr} w_l(kr)\Big|_{r=R}. \tag{45}$$

Conversely, we have

$$S_l(k) = g_l(k)[g_l(-k)]^{-1}, \tag{46}$$

$$g_l(k) = k^l[w_l(kR) - R_l(E) w_l'(kR)]. \tag{47}$$

Now comes the crucial point. We want to write S_l [Eq. (46)] in the form (27), with $f_l(k)$ having all the properties, mentioned at the end of Sec. II, of the true Jost function of potential scattering. The R matrix being meromorphic, it is obvious that we cannot simply take $f_l(k) \equiv g_l(k)$. To remedy this, we have to show that there exists an appropriate real[14] entire function $F_l(k^2)$, such that

$$f_l(k) = g_l(k) F_l(k^2). \tag{48}$$

That $F_l(k^2)$ must be a real function is the consequence of the fact that, because of[5]

$$w_l(-z^*) = (-1)^l [w_l(z)]^* \tag{49}$$

and the reality of $R_l(E)$,

$$R_l(E^*) = [R_l(E)]^*, \tag{50}$$

$g_l(k)$ satisfies already (22), i.e.,

$$g_l(-k^*) = [g_l(k)]^*. \tag{51}$$

Therefore, to ensure that $f_l(k)$, Eq. (48), also satisfies (22), we must choose $F_l(k^2)$ real:

$$F_l(E^*) = [F_l(E)]^*. \tag{52}$$

Before giving the explicit construction of $F_l(E)$, let us enumerate the properties of the S matrix one can derive from those of the R matrix, namely (35), (36), and (37). We consider for simplicity the S wave, and drop henceforth the subscript l. The other waves can be handled in the same way, with some minor changes, as shown in Appendix B. As has been shown by Van Kampen,[20] the above properties of the R matrix imply (and are implied by) the following properties of S:

(a) $S(k)$ is a meromorphic function of k in the entire plane;

(b) on the real axis,

$$[S(k)]^* = [S(k)]^{-1} = S(-k); \tag{53}$$

(c) the poles of S lie either on the positive imaginary axis or in the lower half-plane.[21] The poles on the positive imaginary axis are simple. It has been shown in the first paper of Van Kampen[8] that poles of higher order on the $+i$ axis would violate causality;

(d) in the first quadrant of the k plane,

$$\operatorname{Im}[\exp(2ikR)S] \leq 1; \tag{54}$$

(e) the absolute value of $S_R(k) = \exp(2ikR)S(k)$ is uniformly bounded in $0 \leq \arg k \leq \frac{1}{2}\pi - \epsilon$, $\epsilon > 0$.

Note that properties (a) and (b) are trivial consequences of the reality and meromorphy of the R matrix, and obviously hold for all l.

Let us now study the properties of the poles E_λ of the R matrix. They are needed in the construction of $F(k^2)$. Consider first the negative poles. We assume that the number of bound states is finite, i.e., there is a finite number of poles of $S(k)$ on the imaginary

[18] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (Am. Math. Soc., New York, 1963). Quite generally, a real meromorphic Herglotz function admits the Mittag–Leffler expansion

$$R(z) = \alpha z + \beta + \sum_n \left\{\frac{\gamma_n^2}{u_n - z} - \frac{\gamma_n^2}{u_n}\right\},$$

where all the constants are real, and $\alpha \geq 0$. This is equivalent to (35) if $\alpha = 0$ and the sum in (37) converges to β.

[19] All the derivatives are taken with respect to r, Eq. (45).

[20] N. G. Van Kampen, Rev. Mex. Fis. **2**, 233 (1953).

[21] W. Schützer and J. Tiomno, Phys. Rev. **83**, 249 (1951).

axis.[22] This is a reasonable assumption in general, and is dictated here by our wish to prove the sum rules. If we think in terms of potential scattering, the number of bound states is finite if $rV(r)$ is integrable at the origin.[5,6] On the other hand, as discussed by Wigner[8] and Van Kampen,[8] an attractive potential behaving like r^{-n} $(n \geq 2)$ at the origin—which has indeed an infinite number of bound states—or, more generally, a self-adjoint Hamiltonian which is not bounded below, would certainly violate the causality. In fact, in this case, because of the well-known phenomenon of collapse into the origin, the R matrix cannot be defined. The number of bound states being finite, we can see at once that the same must be true for the number of negative poles of the R matrix. Indeed, a bound state $(k = i\gamma,\ \gamma > 0)$ is determined by the vanishing of the denominator of (46)[23]:

$$w(-i\gamma R) - R(-\gamma^2)w'(-i\gamma R) = 0. \tag{55}$$

Because of (49), the functions $w(-i\gamma R)$ and $w'(-i\gamma R)$ are either real (l even) or pure imaginary (l odd), and $R(E)$ is an increasing function on the real axis because of (35). Also, if E_{j-1} and E_j $(\geq E_{j-1})$ are two consecutive poles of the R matrix, one has $R(E_{j-1} + 0) = -\infty$ and $R(E_j - 0) = +\infty$. It follows that there is one bound state between each two consecutive negative poles of $R(E)$. This implies that the number of these poles must be finite.

Consider now the positive poles of the R matrix. From (44) and (31), the R matrix is given by $(l = 0)$,

$$R(k^2) = (1/k)\tan(kR + \delta). \tag{56}$$

Using condition (6a), one sees easily,[24] that the poles and their residues in the k plane $(E_\lambda = k_\lambda^2)$ are given asymptotically by

$$\pm k_\lambda = \nu_\lambda + O(\eta_\lambda), \tag{57a}$$

$$\pm \rho_\lambda = b_\lambda + O(\eta_\lambda), \tag{57b}$$

$$\eta_\lambda \cong \eta[\lambda(\pi/R)], \tag{57c}$$

where λ is a large positive integer, and

$$\nu_\lambda = (\lambda + \tfrac{1}{2})(\pi/R), \tag{58a}$$

$$b_\lambda = -1/\nu_\lambda R\,. \tag{58b}$$

If k_λ is a pole, $-k_\lambda$ is also a pole with an opposite residue. Collecting first the contribution of these two poles together and then summing over λ, we get (35) back, with the asymptotic residues

$$\gamma_\lambda^2 = -2k_\lambda\rho_\lambda = (2/R) + O(\eta_\lambda). \tag{59}$$

The above expressions for the poles and their residues are asymptotic, and hold for large values of the index. In fact, it is clear that there exists an integer Λ such that, for $\lambda > \Lambda$, there is a one-to-one correspondence between the poles k_λ and their asymptotic form ν_λ. Note that the first few k_λ are purely imaginary $(E_\lambda < 0)$ if there are bound states.

From the above asymptotic values of the poles and the residues, it can be shown that[25]

$$\lim_{k\to\infty} -ikR(E) = 1 + o(1), \tag{60}$$

uniformly in any sector $\epsilon \leq \arg k \leq \pi - \epsilon$, $\epsilon > 0$. Using this property in (46), and remembering (32) we see that

$$\exp(2ikR)S \underset{|k|\to\infty}{=} o(1), \tag{61}$$

uniformly in any sector $0 < \epsilon \leq \arg k \leq \pi - \epsilon < \pi$. If $k_j = i\gamma_j$, $\gamma_j > 0$, $j = 1, \cdots, n$, are the poles of the S matrix on the $+i$ axis, we get from $|S(k)| = 1$ on the real axis and the maximum modulus principle, that

$$\left|\exp(2ikR)\prod_{j=1}^{n}(k - i\gamma_j)(k + i\gamma_j)^{-1}S(k)\right| \leq 1 \quad \text{in} \quad \operatorname{Im} k \geq 0. \tag{62}$$

This is the generalization of the property (e) of the S matrix to the whole upper half-plane.

Let us now consider tentatively

$$F(k^2) \equiv \prod_\lambda \left(1 - \frac{E}{E_\lambda}\right), \tag{63}$$

where E_λ are the poles of the R matrix. That this function is entire and real is obvious because of the uniform convergence of the infinite product in any compact of the E plane [consider $\log F$, and use (57a)]. With this choice of F, the function $f(k)$ defined by (48) is obviously entire. Using the general theory of meromorphic functions,[26] it is shown in Appendix

[22] That these poles correspond to bound states is clear from the definition of ψ_l, (33), and the fact that they come from the vanishing of the denominator of (46), which is proportional to $A_l(k)$.

[23] Notice that $k = 0$ is neither a pole of S, nor a pole of R.

[24] To calculate the residues, we need the first derivative of $\delta(k)$. S being holomorphic in a neighborhood of the real axis, and having no zeros on the real axis because of $|S| = 1$, the phase shift $\delta = (2i)^{-1}\log S$ is also holomorphic in the neighborhood of the real axis, and therefore has bounded derivatives of all order in that neighborhood.

[25] The E plane is cut along the positive-real axis, $k = E^{\frac{1}{2}}$ is defined with $\operatorname{Im} k > 0$ if E is not real-positive, and k real-positive if E is real positive on the upper rim of the cut. It is well-known that the asymptotic properties of entire functions and meromorphic functions are closely related to the asymptotic behavior of the zeros and poles (see Refs. 9 and 26). More precisely, there is a definite relation between the asymptotic behavior of $n(r, 0)$ and $n(r, \infty)$, where these two symbols represent, respectively, the number of zeros and poles in $|z| < r$, and the growth of the corresponding meromorphic function. From (56) it is clear that in our case the zeros and the poles of $R(k^2)$ are asymptotically those of $k^{-1}\tan(kR)$, and we expect the asymptotic behavior of $R(k^2)$ in the upper half of the k plane to be the same as $k^{-1}\tan(kR - N\pi) \sim -(ik)^{-1}$. In Appendix C we show rigorously that this is indeed the case.

[26] R. Nevanlinna, *Eindeutige Analytische Funktionen* (Springer-Verlag, Berlin, 1953), p. 163ff.

C that $R(E)$ has the factorization

$$R(E) = R(0)\frac{G(E)}{F(E)} = R(0)\prod_{\mu}\left(1-\frac{E}{E'_{\mu}}\right)\left[\prod_{\lambda}\left(1-\frac{E}{E_{\lambda}}\right)\right]^{-1}, \quad (64)$$

where E'_{μ} are the zeros of $R(E)$. These zeros are simple and interlaced between the poles E_{λ}. They have the asymptotic form

$$E'_{\mu} = \left(\mu+\frac{l}{2}\right)^2\frac{\pi^2}{R^2} + \mu\alpha(\mu), \quad \mu \text{ a large integer}, \quad (65)$$

as is clear from (56) and (B15). Also, it is shown in Appendix C that $R(E)$ has the following asymptotic behavior in the entire plane:

$$R(E) \underset{k\to\infty}{=} \frac{1}{k}\tan(kR - \tfrac{1}{2}l\pi) + \frac{o(1)}{|k|}, \quad (66)$$

and in Appendix D that in all complex directions, $0 < \arg k < \pi$, F, and G have the uniform asymptotic behavior

$$F = \prod_{\lambda}\left(1-\frac{E}{E_{\lambda}}\right) \underset{k\to\infty}{=} Ak^{-l}\cos(kR - \tfrac{1}{2}l\pi)[1+o(1)], \quad (67)$$

$$G = \prod_{\mu}\left(1-\frac{E}{E'_{\mu}}\right) \underset{k\to\infty}{=} Bk^{-l-1}\sin(kR - \tfrac{1}{2}l\pi)[1+o(1)], \quad (68)$$

where the constants A and B are related by $A = R(0)B$ (for the actual values of A and B in terms of E_{λ} and E'_{μ}, see Appendix D).

Using the above asymptotic behaviors in the definition of the "Jost function" (47) and (48), we obtain

$$f_l(k) = k^l[w_l(kR)F(E) - R(0)G(E)w'_l(kR)] \underset{k\to\infty}{=} A[1 + e^{-2ikR}o(1) + o(1)].$$

This behavior is in agreement with both (23) and (29) if $A = 1$. The "Jost function" should therefore be defined by

$$f_l(k) = (k^l/A)[w_l(kR)F(E) - R(0)G(E)w'_l(kR)]. \quad (69)$$

It is an entire function of order 1 and type $2R$.

It remains to be shown that its zeros in the lower half-plane are simple and on the imaginary axis. This has been shown for the S wave by Schützer and Tiomno,[21] as was mentioned before under property (c) of the causal S matrix [note that a pole of S in $\operatorname{Im} k > 0$ is due to a zero of its denominator $f(-k)$]. In Appendix B, we show that the above property is true in general ($l \geq 0$). It is in fact sufficient to prove only that the zeros in the lower half-plane are on the imaginary axis. That they are simple is a consequence of the causality condition,[8] as we saw before under property (c) of the S matrix. The property (22) is also automatically satisfied. We have therefore completed our proof that the causal S matrix can be written in the form (27), with $f_l(k)$ having all the properties of the Jost function of potential scattering. We can therefore derive our sum rules directly from the causality condition and the Levinson theorem.

In fact, we are going to show in Sec. IV, using the Gel'fand–Levitan equation of the inverse scattering problem, that the above properties of $f_l(k)$ entail the existence of a local potential of radius R which reproduces our causal S matrix. This of course does not mean that the interaction is itself local, but only that it is equivalent to a local potential as far as the R and the S matrix are concerned.

IV. THE EQUIVALENCE OF THE CAUSAL INTERACTION WITH A LOCAL POTENTIAL OF FINITE RADIUS

It will be shown now that a causal interaction of radius R satisfying the Levinson theorem and (6b) is in fact equivalent to a local potential of the same radius. For simplicity, we consider the S wave with one bound state. As will be seen, the generalization to higher waves and more than one bound state is straightforward.

We start from the Gel'fand–Levitan integral equation[27] of the inverse scattering problem which implies that given an arbitrary phase shift[28] (satisfying the Levinson theorem) and the bound-states energies, there exists a family of potentials (the so-called phase-equivalent family), depending on as many arbitrary real-positive parameters as there are bound states, satisfying conditions (14a), (14b), such that all the potentials of the family reproduce the phase shift and the binding energies. The arbitrary real positive parameters are in fact the norm of the bound-state wavefunctions defined by the boundary condition (5) of Sec. I. They can vary between zero and infinity. In general, different members of the family have different asymptotic behaviors at $r = \infty$. We therefore have to show that the properties of the Jost functions (or the S matrix) which were derived in Sec. III entail the existence of a potential of radius R in the

[27] Reference 5, Sec. 8.

[28] A sufficient condition for the existence of the potential would be the continuity of the phase shift on the real axis. Here, we have in fact much more than this since $\delta = (2i)^{-1}\log S$ is analytic in a neighborhood of the real axis because of $|S(k)| = 1$, which guarantees that S has neither zeros nor poles on the real axis.

family. According to a theorem by Newton,[29] this potential is unique among the phase-equivalent family. All the other members behave at infinity as a linear combination, with nonzero coefficients, of the bound-state wavefunctions, so that they do not have a finite radius. To prove the existence of a potential of finite radius in the family, we use another integral equation, derived by Marchenko,[30,31] which is equivalent to the Gel'fand–Levitan equation, and is such that the asymptotic properties of the potentials are much more transparent. This equation reads

$$F(x+y)+K(x,y)+\int_x^\infty K(x,t)F(t+y)\,dt=0, \qquad 0\le x<y, \quad (70)$$

$$F(t)=Ce^{-\gamma t}+\frac{1}{2\pi}\int_{-\infty}^{\infty}[1-S(k)]e^{ikt}\,dk, \quad (71)$$

$-\gamma^2$ being the energy of the bound state, and C (positive) the normalization constant:

$$C^{-1}=\int_0^\infty |f(-i\gamma,r)|^2\,dr. \quad (72)$$

The potential is given by

$$V(r)=-2\frac{d}{dr}K(r,r). \quad (73)$$

Necessary and sufficient conditions for the existence of a unique solution of the above integral equation are also known.[32] It can be easily seen that these conditions are satisfied in our case. In fact, we do not need to look at these conditions since, as was mentioned before, the Gel'fand–Levitan equation guarantees the existence and the uniqueness of the potential if the phase shift (satisfying the conditions stated before), the bound-state energy, and the constant C are known.

Let us consider first the case of no bound state. In this case, $F(t)$ reduces to the Fourier-integral transform of $1-S$. The S matrix is now holomorphic in the upper half-plane and satisfies there the condition (61). By closing the contour of integration in (71) in the upper half-plane, it is seen that $F(t)=0$ for $t>2R$. It follows then from Eq. (70) that $K(x,y)=0$ for $y\ge x>R$, and this, according to (73), implies that the potential vanishes beyond the radius R.

The case where there is a bound state is quite similar. The S matrix still satisfies (61), but now has a pole at $k=i\gamma$. By closing the contour of integration as before, we find

$$F(t)=\left[C+\frac{if(i\gamma)}{f'(-i\gamma)}\right]e^{-\gamma t}, \quad t>2R, \quad (74)$$

$$f'(-i\gamma)=f'(k)|_{k=-i\gamma}.$$

This, together with (70) and (73), implies that $V(r)$ behaves in general like $\exp(-2\gamma r)$ for $r>R$, except when

$$C=-if(i\gamma)/f'(-i\gamma), \quad (75)$$

in which case it has, as before, the radius R. This is equivalent to saying that the unique member of the phase-equivalent family which has the radius R is such that the bound-state wavefunction normalization is determined by (75). It remains to be shown now, that the right-hand side of Eq. (75) is indeed positive.

This is true in general, in the absence of ghosts, and can be shown here in a very simple manner. According to (46) and (47), and if we notice that $w_0(z)=\exp(-iz)$, the condition $f(-i\gamma)=0$ [which is equivalent to $g(-i\gamma)=0$] gives

$$g(-i\gamma)=e^{-\gamma R}[1+\gamma R(-\gamma^2)]=0.$$

Accordingly,

$$g(i\gamma)=e^{\gamma R}[1-\gamma R(-\gamma^2)]=2e^{\gamma R}.$$

Also,

$$g'(-i\gamma)=-ie^{-\gamma R}[R+2\gamma^2\dot{R}(-\gamma^2)-R\gamma R(-\gamma^2)]$$
$$=-ie^{-\gamma R}[2R+2\gamma^2\dot{R}(-\gamma^2)],$$

where $\dot{R}(-\gamma^2)=dR/dE$ at $E=-\gamma^2$. The R matrix being an increasing function on the real axis, we get, finally,

$$-i\frac{f(i\gamma)}{f'(-i\gamma)}=-i\frac{g(i\gamma)}{g'(-i\gamma)}=\frac{2e^{2\gamma R}}{2R+2\gamma^2\dot{R}(-\gamma^2)}>0.$$

Q.E.D.

If there is more than one bound state, the first term in the right-hand side of (71) is replaced by a sum over the bound states, and the above analysis can be carried out exactly in the same way. For each bound state γ_j, the normalization constant C_j, corresponding to the potential of finite radius, would be

$$C_j=\frac{2\exp(2\gamma_j R)}{2R+2\gamma_j^2\dot{R}(-\gamma_j^2)}>0.$$

For $l>0$, the proof can be carried out as for the S wave. The starting point is again the Marchenko equation for $l\neq 0$, given by Blažek,[33] which is quite similar to (70) and (71) except that the exponentials are replaced by appropriate combinations of Hankel functions. The calculations are elementary, and lead

[29] R. G. Newton, Phys. Rev. **101**, 1588 (1956).
[30] Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, Science Publ., Inc., New York, 1963), Chap. 5.
[31] L. D. Faddeev, J. Math. Phys. **4**, 72 (1963).
[32] Reference 30, p. 132, theorem 5.6.1.
[33] M. Blažek, Commun. Math. Phys. **3**, 282 (1966).

to the same conclusion: In the phase-equivalent family of potentials, there is one with radius R. It corresponds to the normalization constant

$$C_l = \left[\int_0^\infty |f_l(-i\gamma, r)|^2\, dr\right]^{-1} = -i\,\frac{f_l(i\gamma)}{\dot{f}_l(-i\gamma)}.$$

To prove the positivity of the right-hand side, one uses the same method as for $l = 0$, above, combined with the recursion formulas of Appendix B for the functions $w_l(kR)$. It is elementary but tedious and we shall not reproduce it here.

As is obvious now, the decomposition of the R matrix into the product (64) and the detailed properties of the "Jost function" are not necessary at all for the proof of the equivalence of causal interaction with a local potential of the same radius, since this proof is based entirely on the meromorphy properties of the S matrix (46) and (61), the latter being a direct consequence of (60). Once this equivalence is shown, it follows that the S matrix is given by (27), with $f(k)$ having all the required properties, and the sum rules become a trivial consequence of the above equivalence. The decomposition formula (64) is also obtained easily from formula (44) and the remarks of Sec. II on the zeros of functions which are combinations of functions of class P. This is the reason why the proofs of (64), (67), and (68) are given only in Appendices C and D. That they are given at all is simply because these properties, which are established without knowing beforehand the existence of the equivalent local potential, seem to hold, in a slightly modified form, under more general circumstances. One may therefore hope that our sum rules, in a generalized form, would be true for more general classes of causal interactions, especially for some restricted class of nonlocal interactions which are not equivalent to a local potential. We intend to develop this in a forthcoming paper.

Finally, it may be shown that the properties of the "Jost function" established directly from those of the R matrix, when used in the Gel'fand–Levitan integral equation, lead also to the equivalence theorem. However, the proof is much more involved than given here.

ACKNOWLEDGMENTS

One of the authors (K. C.) would like to thank F. Lurçat and J. J. Loeffel for stimulating discussions.

APPENDIX A

In general, under conditions (14a), (14b), the Jost function in $\mathrm{Im}\, k \leq 0$ is given by [Ref. 5, formula (4.4)]

$$f_l(k) = 1 + \int_0^R \varphi_l(k, r) V(r) k^l w_l(kr)\, dr. \quad \text{(A1)}$$

Here, of course, we only need condition (14a).

The phase shift on the real k axis is the phase of the Jost function, i.e.,

$$\delta_l(k) = \mathrm{Im} \log f_l(k). \quad \text{(A2)}$$

Because of[5]

$$f_l(k) \underset{k\to\infty}{=} 1 + o(1), \quad \text{(A3)}$$

we have

$$\log f_l(k) \underset{k\to\infty}{\simeq} f_l(k) - 1, \quad \text{(A4)}$$

so that, for k large enough,

$$\delta_l(k) = \mathrm{Im} \log f_l(k) \simeq \mathrm{Im}\, f_l(k). \quad \text{(A5)}$$

We have therefore to show that $k^{-1}\,|\mathrm{Im}\, f_l(k)|$ is integrable at infinity. Now[5]

$$\mathrm{Im}\, w_l(kr) = -u_l(kr) = -\left(\frac{\pi}{2} kr\right)^{\frac{1}{2}} J_{l+\frac{1}{2}}(kr). \quad \text{(A6)}$$

Using the upper bounds given in Ref. 5 (k real-positive),

$$|u_l(kr)| \leq A\left(\frac{kr}{1 + kr}\right)^{l+1} \quad \text{(A7)}$$

and

$$|\varphi_l(k, r)| \leq B\left(\frac{r}{1 + kr}\right)^{l+1}, \quad \text{(A8)}$$

in (A1), we find

$$|\mathrm{Im}\, f_l| \leq C\int_0^R dr\, r\, |V(r)|\, \frac{(kr)^{2l+1}}{(1 + kr)^{2l+2}} \leq C\int_0^R r\, |V(r)|\, \frac{kr}{(1 + kr)^2}\, dr. \quad \text{(A9)}$$

It follows that

$$\int_{K_0}^K k^{-1}\, |\mathrm{Im}\, f_l(k)|\, dk \leq C\int_0^R dr\, r\, |V(r)| \int_{K_0 r}^{Kr} \frac{du}{(1 + u)^2} \leq C\int_0^R r\, |V(r)|\, dr. \quad \text{(A10)}$$

Q.E.D.

APPENDIX B

This Appendix is devoted to a detailed study of the functions considered in Sec. III, for all values of the angular momentum l. First, let us mention some properties of the free solutions $w_l(kr)$. We have[5]

$$w_l(z) = -v_l(z) - iu_l(z) \underset{z\to\infty}{=} i^l e^{-iz}[1 + O(1/z)], \quad \text{(B1)}$$

$$u_l(z) = zj_l(z) = (-1)^{l+1} u_l(-z), \quad \text{(B2)}$$

$$v_l(z) = zn_l(z) = (-1)^l v_l(-z), \quad \text{(B3)}$$

$$u_l(z) = \begin{cases} \sin(z - \frac{1}{2}l\pi), & z \to \infty, \quad \text{(B4a)} \\ z^{l+1}[(2l+1)!!]^{-1}, & z \to 0, \quad \text{(B4b)} \end{cases}$$

$$v_l(z) = \begin{cases} -\cos(z - \frac{1}{2}l\pi), & z \to \infty, \quad \text{(B5a)} \\ -z^{-l}(2l-1)!!, & z \to 0. \quad \text{(B5b)} \end{cases}$$

Accordingly, the asymptotic form of the R matrix (44), for $k \to +\infty$, is

$$R_l(k^2) = \frac{1}{k}\tan\left(kR - \frac{l\pi}{2} + \delta_l\right) + \frac{1}{\sin\left[2(kR - \frac{1}{2}l\pi + \delta_l)\right]} O\left(\frac{1}{k^2}\right). \quad \text{(B6)}$$

From this and (6a), it follows that the poles $k_{\lambda l}$ and the residues $\rho_{\lambda l}$ are given asymptotically by

$$k_{\lambda l} = \nu_{\lambda l} + O[\eta(k_{\lambda l})] = \nu_{\lambda l} + O[\eta(\pi\lambda/R)], \quad \text{(B7)}$$

$$\rho_{\lambda l} = b_{\lambda l} + O[\eta(k_{\lambda l})] = \nu_{\lambda l} + O[\eta(\pi\lambda/R)], \quad \text{(B8)}$$

where λ is a large integer ($\gtrless 0$), and

$$\nu_{\lambda l} = \left(\lambda + \frac{l+1}{2}\right)\frac{\pi}{R}, \quad \text{(B9)}$$

$$b_{\lambda l} = \frac{-1}{R\nu_{\lambda l}}. \quad \text{(B10)}$$

There is therefore a one-to-one correspondence between the $k_{\lambda l}$ and $\nu_{\lambda l}$ for $|\lambda| > \Lambda$, Λ being a fixed large positive integer. Combining the poles $\pm k_{\lambda l}$ which have opposite residues, we get

$$R_l(k^2) = \sum_{\lambda} \frac{\gamma_{\lambda l}^2}{E_{\lambda l} - E}, \quad \text{(B11)}$$

the sum now being over the poles of the R matrix which we number also symbolically by the set $\{\lambda\}$, although the set $\{1, 2, \cdots\}$ would be more appropriate. The first few $E_{\lambda l}$ may be negative if there are bound states. For $\lambda > \Lambda$, we have

$$E_{\lambda l} = \left(\lambda + \frac{l+1}{2}\right)^2\left(\frac{\pi^2}{R^2}\right) + \lambda\alpha(\lambda), \quad \text{(B12)}$$

$$\gamma_{\lambda l}^2 = (2/R) + \alpha(\lambda). \quad \text{(B13)}$$

According to (6b),

$$\int^{\infty} \lambda^{-1}|\alpha(\lambda)|\, d\lambda < \infty. \quad \text{(B14)}$$

It is clear also that the zeros of $R_l(E)$ are given asymptotically by

$$E'_{\mu} = E^{0\prime}_{\mu} + \mu\alpha(\mu), \quad \text{(B15)}$$

$$E^{0\prime}_{\mu} = (\mu + \tfrac{1}{2}l)^2\frac{\pi^2}{R^2}, \quad \text{(B16)}$$

where μ is a large integer. Formulas (B12–B16) are the generalizations of formulas of the Sec. III given for $l = 0$.

We have now to generalize to $l \neq 0$, the properties of the S matrix, stated for $l = 0$ in Sec. III, which are needed in our analysis. As was noted before, properties (a) and (b) are trivial consequences of the reality and meromorphy of the R matrix, and (49). In fact, it is obvious that because of (46) and (51), we have in general (k complex)

$$S_l(-k) = [S_l(k)]^{-1} \quad \text{(B17)}$$

and

$$S_l(-k^*) = S_l^*(k), \quad \text{(B18)}$$

i.e., the S matrix is real on the imaginary axis.

Property (c) can be established as follows. According to (46), we have to prove that $g_l(k)$, given by formula (47), does not vanish in $\operatorname{Im} k < 0$ outside the imaginary axis. The zero due to k^l being compensated by the pole k^{-l} of $w_l(kR)$, we are left with the solution of

$$w_l(kR) - kR_l(E)\dot{w}_l(kR) = 0, \quad \text{(B19)}$$

$$\dot{w}_l(kR) = \left(\frac{d}{d\rho}\right)w_l(\rho)\Big|_{\rho=kR}. \quad \text{(B20)}$$

We note now that $w_l(kR)$ does not vanish in the lower half-plane.[34] To simplify the writing, let us put $kR = \rho$ and $R_l(\rho^2/R^2) = \tilde{R}_l(\rho^2)$.

If we divide by $w_l(\rho)$, (6.19) becomes

$$\rho\frac{w'(\rho)}{w_l(\rho)} = \frac{R}{\tilde{R}_l(\rho^2)}, \quad \text{(B21)}$$

$R_l(E)$ being a Herglotz function [$\operatorname{Im} R_l(E) \cdot \operatorname{Im} E > 0$ if $\operatorname{Im} E \neq 0$]; we have

$$\operatorname{Im}[\tilde{R}_l(\rho^2)]^{-1} = -\operatorname{Re}\rho \cdot \operatorname{Im}\rho \cdot [+],$$

where the symbol [+] means a positive function which does not vanish except perhaps for real values of ρ^2, i.e., ρ either real or pure imaginary. It follows that, in the lower half of the ρ plane,

$$\operatorname{Im}\tilde{R}_l^{-1}(\rho^2) = \operatorname{Re}\rho \cdot [+].$$

We are now going to prove that in the same domain, the left-hand side of (B21) satisfies

$$\operatorname{Im}\left[\frac{\rho w_l'(\rho)}{w_l(\rho)}\right] = -\operatorname{Re}\rho \cdot [+].$$

Clearly, this would mean that (B19) does not vanish in $\operatorname{Im}\rho < 0$ except perhaps on $\operatorname{Re}\rho = 0$. For this proof, we proceed by induction.[35] We start from the recursion formula[36]

$$\rho w_l(\rho) + \rho w'_{l-1}(\rho) - l w_{l-1}(\rho) = 0.$$

This relation gives, respectively, for l and $l+1$,

$$l - \rho\left(\frac{w'_{l-1}}{w_{l-1}}\right) = \rho\left(\frac{w_l}{w_{l-1}}\right)$$

[34] *Bateman Manuscript Project, Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. II, p. 62.

[35] We wish to thank E. Cremmer for his help in this proof.

[36] These recursion relations are easily obtained from those for the Hankel functions, Ref. 34, and our definition.

and

$$l + \rho\left(\frac{w_l'}{w_l}\right) = 2l + 1 - \rho\left(\frac{w_{l+1}}{w_l}\right).$$

Note that we always divide by a function which does not vanish in Im $\rho < 0$. Taking the product of these two formulas, and using the recursion relation[36]

$$(2l + 1)w_l = \rho(w_{l+1} + w_{l-1}),$$

we get

$$[l + \rho(w_l'/w_l)][l - \rho(w_{l-1}'/w_{l-1})] = \rho^2.$$

Separating the real and imaginary parts, and using the self-explanatory notation

$$\rho(w_l'/w_l) = G_l + iH_l, \quad \rho = x + iy,$$

we obtain

$$(l - G_{l-1})G_l + H_{l-1}H_l = (x^2 - y^2) - l(l - G_{l-1}),$$
$$-H_{l-1}G_l + (l - G_{l-1})H_l = 2xy + lH_{l-1}.$$

Solving these two equations for G_l and H_l, we find

$$G_l = -l - (\Delta_{l-1})^{-1}[(y^2 - x^2)A_{l-1} - 2xyH_{l-1}], \tag{B22}$$

$$H_l = (\Delta_{l-1})^{-1}[(x^2 - y^2)H_{l-1} + 2xyA_{l-1}], \tag{B23}$$

where

$$A_{l-1} = l - G_{l-1}$$

and

$$\Delta_{l-1} = A_{l-1}^2 + H_{l-1}^2 > 0.$$

We are now going to prove that $B_l = -x^{-1}H_l = -(\text{Re } \rho)^{-1} \text{ Im } [\rho(w_l'/w_l]$ is positive in the half-plane $y = \text{Im } \rho < 0$. Indeed, we get from (B22) and (B23) that

$$A_l = 2l + 1 + (\Delta_{l-1})^{-1}[(y^2 - x^2)A_{l-1} - 2x^2yB_{l-1}], \tag{B24}$$

$$B_l = (\Delta_{l-1})^{-1}[(x^2 - y^2)B_{l-1} - 2yA_{l-1}]. \tag{B25}$$

Writing now (B25) for $l + 1$, and eliminating A_l and B_l with the help of (B24) and (B25), we finally find

$$B_{l+1} = \Delta_l^{-1}[-2(2l + 1)y + \Delta_{l-1}^{-1} |\rho|^4 B_{l-1}].$$

It follows that, in the lower half-plane, B_{l+1} is strictly positive if B_{l-1} is so. Now, for S and P waves, we have

$$B_0 = -(\text{Re } \rho)^{-1} \text{ Im } \left[\rho\left(\frac{w_0'}{w_0}\right)\right]$$
$$= -(\text{Re } \rho)^{-1} \text{ Im } \left[\frac{-i\rho e^{-i\rho}}{e^{-i\rho}}\right] = 1$$

and

$$B_1 = -(\text{Re } \rho)^{-1} \text{ Im } \left\{\frac{\rho[(i + \rho^{-1})e^{-i\rho}]'}{[(i + \rho^{-1})e^{-i\rho}]}\right\}$$
$$= \frac{(|\rho|^2 - 2 \text{ Im } \rho)}{|1 + i\rho|^2} > 0.$$

It follows that, for all l,

$$\text{Im } [\rho(w_l'/w_l)] = -\text{Re } \rho \cdot [+]$$

in Im $\rho < 0$.

The fact that the eventual zeros on the negative-imaginary axis are simple, i.e., the poles of $S_l(k)$ in the upper half-plane are simple, is a consequence of causality, as was mentioned in Sec. III.

As for properties (d) and (e), they can also be generalized to all l, and k in the whole upper half-plane, by using a product similar to the one in the left-hand side of (62), and by replacing everywhere exp $(2ikR)$ by appropriate combinations of $w_l(\pm kR)$. However, we do not use them in our analysis, and so we do not pursue the matter further.

We are going now to establish some expansion formulas for tan $(kr - \frac{1}{2}l\pi)$ and sin $(kR - \frac{1}{2}l\pi)$ that we use in Appendices C and D. For the first function, we start from the Euler formula[37]

$$\tan (z - \tfrac{1}{2}l\pi)$$
$$= -\cot\left(z - \frac{l+1}{2}\pi\right)$$
$$= -\left(z - \frac{l+1}{2}\pi\right)^{-1} - \sum_{n=1}^{\infty}\left[\left(z - \frac{l+1}{2}\pi - n\pi\right)^{-1} + \left(z - \frac{l+1}{2}\pi + n\pi\right)^{-1}\right] \tag{B26}$$

to which we add

$$0 = \sum_1^{\infty}\left[\left(z + \frac{l+1}{2}\pi + n\pi\right)^{-1} - \left(z + \frac{l+1}{2}\pi + n\pi\right)^{-1}\right].$$

It is then easily found that

$$\tan (z - \tfrac{1}{2}l\pi)$$
$$= -\left(z - \frac{l+1}{2}\pi\right)^{-1} - \sum_1^{\infty} 2z\Big/\left[z^2 - \left(n + \frac{l+1}{2}\right)^2\pi^2\right] - (l+1)\pi\sum_1^{\infty}\left[(z + n\pi)^2 - \frac{(l+1)^2}{4}\pi^2\right]^{-1}.$$

The second series being absolutely and uniformly convergent, we finally obtain

$$\tan (z - \tfrac{1}{2}l\pi) \underset{z\to\infty}{=} -2\sum\left\{z\Big/\left[z^2 - \left(n + \frac{l+1}{2}\right)^2\pi^2\right]\right\} + o(1) \tag{B27}$$

in all complex directions.

[37] G. Valiron, *Théorie des fonctions* (Masson et Cie., Paris, 1948), p. 384.

For the second function, we begin with[38]

$$\sin(z-\tfrac{1}{2}l\pi)=(z-\tfrac{1}{2}l\pi)\prod_{\mu=1}^{\infty}\left(1-\frac{z-\frac{1}{2}l\pi}{\mu\pi}\right)e^{z/\mu\pi}\times\prod_{\nu=1}^{\infty}\left(1+\frac{z-\frac{1}{2}l\pi}{\nu\pi}\right)e^{-z/\nu\pi}. \quad \text{(B28)}$$

The second infinite product can be written

$$\prod_{\nu=1}^{\infty}=\prod_{\nu=1}^{l}\times\prod_{\nu=l+1}^{\infty}.$$

Making the change of variable $\nu=\mu+l$ in the second product above, we get

$$\prod_{\nu=1}^{\infty}=\prod_{\nu=1}^{l}\left(1+\frac{z-\frac{1}{2}l\pi}{\nu\pi}\right)\times\exp(-z/\nu\pi)\prod_{\mu=1}^{\infty}\left(\frac{\mu\pi+\frac{1}{2}l\pi+z}{(\mu+l)\pi}\right)\times\exp\left[-\left(\frac{1}{\mu}-\frac{1}{\mu}+\frac{1}{\mu+l}\right)z/\pi\right].$$

Collecting now all the exponentials in the right-hand side, except the first exponential of the second product, we obtain

$$\exp\left[\left(-\sum_{\nu=1}^{l}\frac{1}{\nu}+\sum_{\mu=1}^{\infty}\frac{l}{\mu(\mu+l)}\right)\frac{z}{\pi}\right]$$

which is equal to one by virtue of

$$\sum_{\nu=1}^{l}\frac{1}{\nu}=\sum_{\mu=1}^{\infty}\frac{l}{\mu(\mu+l)}. \quad \text{(B29)}$$

(For the proof of this, see below.) Going back now to (B28), and putting the two infinite products together, we find

$$\sin(z-\tfrac{1}{2}l\pi)=(z-\tfrac{1}{2}l\pi)\prod_{\mu=1}^{l}\left(1+\frac{z-\frac{1}{2}l\pi}{\mu\pi}\right)\prod_{\mu=1}^{\infty}\frac{(\mu+\frac{1}{2}l)^2\pi^2-z^2}{(\mu+\frac{1}{2}l)^2\pi^2}\prod_{\mu=1}^{\infty}\frac{(\mu+\frac{1}{2}l)^2}{\mu(\mu+l)}=O(z^{l+1})\prod_{\mu=1}^{\infty}\left(1-\frac{z^2}{(\mu+\frac{1}{2}l)^2\pi^2}\right). \quad \text{(B30)}$$

Note that all the operations we have performed on the infinite products are legitimate because these products are all absolutely convergent, and we have extracted from them other absolutely convergent infinite products. To prove (B29), we proceed by induction. For $l=1$, it holds trivially that

$$\sum_{\mu=1}^{\infty}\frac{1}{\mu(\mu+1)}=\sum_{\mu=1}^{\infty}\left(\frac{1}{\mu}-\frac{1}{\mu+1}\right)=\sum_{\mu=1}^{\infty}\int_{\mu}^{\mu+1}\frac{dx}{x^2}=\int_{1}^{\infty}\frac{dx}{x^2}=1.$$

[38] Reference 37, p. 433.

Suppose now that (B29) holds for $l-1$. Taking the difference of each of its sides for l and $l-1$, we get, as above,

$$\frac{1}{l}=\sum_{\mu=1}^{\infty}\frac{1}{(\mu+l-1)(\mu+l)}=\sum_{\nu=l}^{\infty}\frac{1}{\nu(\nu+1)}=\int_{l}^{\infty}\frac{dx}{x^2}=\frac{1}{l}.$$

Q.E.D.

APPENDIX C

Here we have to show that in general ($l\geq 0$)

$$R_l(k^2)=k^{-1}\tan(kR-\tfrac{1}{2}l\pi)+|k|^{-1}o(1) \quad \text{(C1)}$$

for $k\to\infty$ in all complex directions in the upper half-plane. By definition,

$$R_l(k^2)=\sum_{\lambda}\frac{\gamma_{\lambda l}^2}{E_{\lambda l}-E}, \quad \text{(C2)}$$

where, for $\lambda>\Lambda$ (a large fixed integer),

$$E_{\lambda l}=E_{\lambda l}^0+\lambda\alpha(\lambda), \quad \text{(C3)}$$

$$E_{\lambda l}^0=\left(\lambda+\frac{l+1}{2}\right)^2\frac{\pi^2}{R^2}, \quad \text{(C4)}$$

$$\gamma_{\lambda l}^2=\frac{2}{R}+\alpha(\lambda), \quad \text{(C5)}$$

and $\alpha(\lambda)$ is such that

$$\int^{\infty}\lambda^{-1}|\alpha(\lambda)|\,d\lambda<\infty. \quad \text{(C6)}$$

As is obvious, we can neglect the first Λ terms of the series in (C2) which are $O(E^{-1})$ for $E\to\infty$. For $\lambda>\Lambda$, there is a one-to-one correspondence between the poles of $R_l(E)$ and those of $k^{-1}\tan(kR-\frac{1}{2}l\pi)$. To simplify the writing, we drop henceforth the subscript l. As is clear from the above formulas, we have, uniformly in all complex directions ($0<\epsilon\leq\arg E\leq 2\pi-\epsilon$):

$$\lim_{|E|\to\infty}R(E)=0. \quad \text{(C7)}$$

Let us see first what we can get from the above properties of $R(E)$ by using the general theory of meromorphic functions.[26,39] According to (C3), the number of poles of the R matrix in $|E|\leq r$ is given, asymptotically, by

$$n(r,\infty)=(R/\pi)r^{\frac{1}{2}}+O(r^{-\frac{1}{2}}). \quad \text{(C8)}$$

Also, the origin not being a pole, we have

$$n(0,\infty)=0. \quad \text{(C9)}$$

If we denote by $n(r,0)$ the number of zeros inside the

[39] E. C. Titchmarsh, *Theory of Functions* (Oxford University Press, London, 1939), p. 284bff.

circle $|E| \leq r$, counted according to their multiplicity, we also have

$$n(r, 0) = (R/\pi)r^{\frac{1}{2}} + O(r^{-\frac{1}{2}}). \tag{C10}$$

Indeed, $R(E)$ being a Herglotz function does not vanish outside the real axis. On the other hand, there are no zeros for $E < E_1$, and those in $E > E_1$ are interlaced between the poles E_λ, and are simple because $R(E)$ is monotonically increasing between two successive poles. There is exactly one simple zero between two successive poles. In fact, according to Appendix B, and formula (6a), these zeros are asymptotically close to $(\lambda + \frac{1}{2}l)^2\pi^2/R^2$, i.e., they are situated, in the k plane, half way between two neighboring poles.

Consider now the characteristic function $T(r)$ (Refs. 26 and 39). It is given by ($a \neq \infty$):

$$\begin{aligned} T(r) &= m(r, \infty) + N(r, \infty) \\ &= m(r, a) + N(r, a) + \varphi(r, a), \end{aligned} \tag{C11}$$

where $\varphi(r, a)$ is a bounded function of r, and

$$m(r, \infty) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ |R(re^{i\varphi})|\, d\varphi, \tag{C12}$$

$$N(r, \infty) = \int_0^r \frac{n(t, \infty)}{t}\, dt = \frac{2R}{\pi} r^{\frac{1}{2}} + O(r^{-\frac{1}{2}}), \tag{C13}$$

$$m(r, a) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ \left| \frac{1}{R(re^{i\varphi}) - a} \right| d\varphi, \tag{C14}$$

$$N(r, a) = \int_0^r \frac{n(t, a)}{t}\, dt, \tag{C15}$$

$n(t, a)$ denoting the number of the roots of $R(E) - a$ in $|E| \leq t$, counted according to their multiplicity, and ($\alpha \geq 0$)

$$\log^+ \alpha = \max (0, \log \alpha). \tag{C16}$$

It follows from (C10) that

$$N(r, 0) = (2R/\pi)r^{\frac{1}{2}} + O(r^{-\frac{1}{2}}). \tag{C17}$$

Consider now the function $m(r, \infty)$. It is obvious that because of (C7), this function is bounded when $r \to \infty$ by staying a finite distance (no matter how small) away from the poles E_λ, i.e., if we exclude small intervals $|r - E_\lambda| = d$, $\lambda = 1, 2, \cdots$. Indeed, $R(re^{i\varphi})$ goes to zero at $r = \infty$, $\epsilon < \varphi < 2\pi - \epsilon$, and is bounded on the arc $-\epsilon \leq \varphi \leq \epsilon$ if we exclude the above intervals. Note that we can choose d small enough so that all the zeros of $R(E)$, E'_μ, are outside the intervals $|r - E_\lambda| = d$.

Therefore, if $r \notin [E_\lambda - d, E_\lambda + d]$, $\lambda = 1, 2, \cdots$, we have

$$T(r) = N(r, \infty) + m(r, \infty) = (2R/\pi)r^{\frac{1}{2}} + O(1). \tag{C18}$$

Now, according to (C11) with $a = 0$, the asymptotic equality of $N(r, \infty)$ and $N(r, 0)$, and the boundedness of $\varphi(r, 0)$, $m(r, 0)$, and $m(r, \infty)$ are equivalent, so that, for $r \in [E_\lambda - d, E_\lambda + d]$, and $\lambda \to \infty$, we can use $m(r, 0)$ instead of $m(r, \infty)$. Consider therefore

$$m(r, 0) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ \left| \frac{1}{R(re^{i\varphi})} \right| d\varphi. \tag{C19}$$

The intervals $[E_\lambda - d, E_\lambda + d]$ being free from zeros E'_μ, the integrand is bounded on the arc $-\epsilon < \varphi < \epsilon$. In fact, it is mostly zero because $R(E)$ is large close to the poles. On the other hand, $R(E)$ being a Herglotz function, we have, for $\varphi \neq 0, 2\pi$, the lower bound[18]

$$|R(re^{i\varphi})| \geq \frac{C}{r}, \tag{C20}$$

C independent of φ.

Therefore, if $r \to \infty$ by staying within the intervals $(E_\lambda - d, E_\lambda + d)$, we have

$$m(r, 0) \leq O(\log r). \tag{C21}$$

It follows that no matter how $r \to \infty$, we have, neglecting $O(\log r)$,

$$T(r) \simeq (2R/\pi)r^{\frac{1}{2}}. \tag{C22}$$

The order of the meromorphic function[40] $R(E)$ is, accordingly,

$$\rho = \limsup_{r\to\infty} \frac{\log T(r)}{\log r} = \tfrac{1}{2}. \tag{C23}$$

It follows from the factorization theorem of meromorphic functions that[41]

$$R(E) = R(0)\left[\prod_\mu \left(1 - \frac{E}{E'_\mu}\right)\right] \Big/ \left[\prod_\lambda \left(1 - \frac{E}{E_\lambda}\right)\right]. \tag{C24}$$

This formula will be useful in Appendix D.

That the asymptotic behavior of $R(E)$ is given by $k^{-1}\tan(kR - \frac{1}{2}l\pi)$ can be shown as follows. First of all, we have the Mittag–Leffler expansion

$$\tan(z - \tfrac{1}{2}l\pi) = \frac{-1}{z - \dfrac{l+1}{2}\pi} - \sum_{\lambda=-\infty}^{\infty}{}' \left(\frac{1}{z - \nu_\lambda} + \frac{1}{\lambda\pi}\right), \tag{C25}$$

where

$$\nu_\lambda = \left(\lambda + \frac{l+1}{2}\right)\pi,$$

[40] Reference 26, p. 217ff; Ref. 39, p. 284e.

[41] Reference 26, p. 224, formulas (7) and (8); Ref. 39, p. 284f. Here $\rho = \frac{1}{2}$, and the zeros and poles have genus 0 ($q = 0$ in Nevanlinna's terminology; $p_1 = p_2 = 0$ in Titchmarsh). We have assumed that $R(0) \neq 0$. If the origin were a zero (simple) of $R(E)$, we would have a factor $R'(0)E$ instead of $R(0)$ in front of our formula. However, the factor E could be absorbed in the numerator (function G), and nothing would change in our analysis.

and the prime on $\sum$ means that the term $\lambda = 0$ is omitted.

As shown in Appendix B, formula (B27), the above formula leads to

$$\frac{2}{R}\sum_{\lambda}\frac{1}{E_\lambda^0 - k^2}_{k\to\infty} = k^{-1}\tan(kR - \tfrac{1}{2}l\pi) + k^{-1}o(1) \quad \text{(C26)}$$

in all complex directions.

Consider now

$$\sum_{\lambda}\left(\frac{\gamma_\lambda^2}{E_\lambda - E} - \frac{2/R}{E_\lambda^0 - E}\right) = \sum_{\lambda}\left[E_\lambda^0\gamma_\lambda^2 - \frac{2}{R}E_\lambda - E\left(\gamma_\lambda^2 - \frac{2}{R}\right)\right] \times [(E_\lambda - E)(E_\lambda^0 - E)]^{-1}. \quad \text{(C27)}$$

We are going to show that this difference behaves like $|E|^{-\frac{1}{2}}o(1)$. As before, we can drop any finite number of terms we wish from (C27). They will contribute only by $O(E^{-1})$. From (C3) and (C5), it is seen that the main terms in the numerator are $E_\lambda^0\alpha(\lambda)$ and $E\alpha(\lambda)$. Once these are taken care of, the remainder $\lambda\alpha(\lambda)/R$ can be shown, with the same method, to give a smaller contribution.

Now, for E not on the positive real axis, $\varphi = \arg E \neq 0, 2\pi$, we have (draw a figure and project E on the real axis, and E_λ on the direction φ):

$$|E_\lambda^0 - E| \geq |E|\,|\sin\varphi|,$$
$$|E_\lambda^0 - E| \geq E_\lambda^0\,|\sin\varphi|,$$

or

$$|E_\lambda^0 - E| \geq \tfrac{1}{2}(E_\lambda^0 + |E|)\,|\sin\varphi|,$$

and similar formulas with E_λ instead of E_λ^0.

Therefore,

$$\left|\sum\frac{E_\lambda^0\alpha(\lambda)}{(E_\lambda - E)(E_\lambda^0 - E)}\right| \leq 4\sum\frac{E_\lambda^0\,|\alpha(\lambda)|}{(E_\lambda + |E|)(E_\lambda^0 + |E|)\sin^2\varphi} \leq \frac{4}{\sin^2\varphi}\sum\frac{|\alpha(\lambda)|}{E_\lambda + |E|} \leq \frac{4}{\sin^2\varphi}\frac{1}{|E|^{\frac{1}{2}}}\sum\frac{|\alpha(\lambda)|}{(E_\lambda + |E|)^{\frac{1}{2}}},$$

the series in the rhs being convergent by virtue of (C6). For $|E| \to \infty$, we get therefore $|E|^{-\frac{1}{2}}o(1)$. The other main term can be worked out similarly:

$$\left|\sum\frac{E\alpha(\lambda)}{(E_\lambda^0 - E)(E_\lambda - E)}\right| \leq \frac{4}{\sin^2\varphi}\sum\frac{|\alpha(\lambda)|}{E_\lambda^0 + |E|} = |E|^{-\frac{1}{2}}o(1).$$

As for the remainder, with numerator $\lambda\alpha(\lambda)$, the same method gives $O(|E|^{-1})$. We have therefore proved formula (C1) rigorously. In fact, it is now easy to see that, in all directions, including the positive real axis, we have the asymptotic behavior

$$R(E) \simeq k^{-1}\tan(kR - \tfrac{1}{2}l\pi - N\pi). \quad \text{(C28)}$$

To end this appendix, let us say a few words about the properties of the two entire functions

$$F = \prod_{\lambda} \quad \text{and} \quad G = \prod_{\mu}$$

which figure in (C24). Each one is an entire function of order $\frac{1}{2}$ in E. Indeed, the convergence exponent of the zeros in both cases is[42] $\rho_1 = \frac{1}{2}$, the genus of the zeros (or of the canonical products) is zero, and $n(r) \simeq (R/\pi)r^{\frac{1}{2}}$. According to Theorem 2.6.5,[42] both functions are of order $\rho = \rho_1 = \frac{1}{2}$. That both are also of finite type is clear from Theorem 2.9.5. According to Definition 2.7.3, the genus of F and G is zero (see also Theorem 2.12.5).[42] The asymptotic properties of F and G are studied in Appendix D.

APPENDIX D

We have to study here the asymptotic properties of

$$F = \prod_{\lambda} \quad \text{and} \quad G = \prod_{\mu}.$$

They are quite similar to each other as is intuitively clear from the asymptotic expressions of the zeros E_λ and E'_μ. We consider in detail G, because it enters directly into the definition of the "Jost function" (70). According to the estimates of canonical products at large,[43] we have, uniformly in all complex directions ($\varphi \neq 0, 2\pi$):

$$G(|E|\,e^{i\varphi}) = \prod_{\mu}\left[1 - \frac{E}{E'_\mu}\right] = \exp[(|E|^{\frac{1}{2}}Re^{i(\varphi-\pi)/2})(1 + o(1))], \quad \text{(D1)}$$

and an identical expression for $F(E)$. However, the above estimate is clearly not sufficient to ensure that the "Jost function" would satisfy (29). We have to obtain a more precise statement than merely $o(1)$. We start from formula (B30):

$$\sin(z - \tfrac{1}{2}l\pi) = O(z^l)(z - \tfrac{1}{2}l\pi)\prod_{\mu=1}^{\infty}\left(1 - \frac{z^2}{z_\mu^2}\right), \quad \text{(D2)}$$
$$z_\mu = (\mu + \tfrac{1}{2}l)\pi, \quad u = 1, 2, \cdots.$$

As we saw before, for μ larger than some integer M, there is a one-to-one correspondence between E'_μ Eq. (65) and $E_\mu^{0\prime} = (\mu + \frac{1}{2}l)^2\pi^2/R^2$.

Accordingly, we can write

$$\frac{G(E)}{\sin(kR - \tfrac{1}{2}l\pi)} = \left[\prod_{\mu>M}\left(1 - \frac{E}{E'_\mu}\right)\right]\Big/\left[\prod_{\mu>M}\left(1 - \frac{E}{E_\mu^{0\prime}}\right)\right]O(k^P), \quad \text{(D3)}$$

[42] Reference 9, Chap. 2. The number of the theorems mentioned below refers to this book.

[43] Reference 9, Lemma 10.3.6.

where P is an integer ($\gtrless 0$). Note that for $\mu < M$, there is not necessarily a one-to-one correspondence between the zeros of $R_l(E)$ and those of

$$k^{-1} \tan (kR - \tfrac{1}{2}l\pi).$$

This is the origin of the term $O(k^P)$.

To see the asymptotic behavior of the above ratio, we take its logarithm:

$$\log \frac{\prod_{\mu > M}}{\prod_{\mu > M}} = \sum_{\mu > M}^{\infty} \left[\log \left(1 - \frac{E}{E'_\mu}\right) - \log \left(1 - \frac{E}{E^{0\prime}_\mu}\right) \right]. \tag{D4}$$

Each series is absolutely and uniformly convergent in any compact of the E plane because

$$\log \left(1 - \frac{E}{E^{0\prime}_\mu}\right) \underset{\mu \to \infty}{\simeq} - \frac{E}{E^{0\prime}_\mu} = O(\mu^{-2}),$$

and a similar statement for E'_μ instead of $E^{0\prime}_\mu$.

Now (D4) can be written

$$-\sum \log \frac{E'_\mu}{E^{0\prime}_\mu} + \sum \log \frac{E - E'_\mu}{E - E^{0\prime}_\mu}. \tag{D5}$$

The first series is absolutely convergent because of

$$\frac{E'_\mu}{E^{0\prime}_\mu} \underset{\mu \to \infty}{=} \frac{E^{0\prime}_\mu + \mu\alpha(\mu)}{E^{0\prime}_\mu} \simeq 1 + \frac{\alpha(\mu)}{\mu},$$

and (C6). The second series in (D5) is, therefore, also absolutely and uniformly convergent (it is the difference of two absolutely convergent series). This can also be directly checked in the above manner. Now, for $E \to \infty$, $\arg E \neq 0, 2\pi$, each term of the second series in (D5) vanishes, and so does the series itself. It follows that, in all complex directions, we have

$$\log \frac{\prod_{\mu > M} \left(1 - \frac{E}{E'_\mu}\right)}{\prod_{\mu > M} \left(1 - \frac{E}{E^{0\prime}_\mu}\right)} \underset{|E| \to \infty}{=} O(1). \tag{D6}$$

We have therefore proved that

$$G(E) = \prod_{\mu} \left(1 - \frac{E}{E'_\mu}\right) \underset{E \to \infty}{=} \sin (kR - \tfrac{1}{2}l\pi) O(k^Q)(1 + o(1)), \tag{D7}$$

Q being an appropriate integer ($Q \gtrless 0$).

A similar analysis can be made for $F(E)$. Because of the shift $\pi/2R$ in the position of the zeros in the k plane, we have to compare $F(E)$ with the cosine function. The result is

$$F(E) \underset{E \to \infty}{\simeq} O(k^{Q'}) \cos (kR - \tfrac{1}{2}l\pi), \tag{D8}$$

Q' being an integer ($\gtrless 0$). Note that, because of (C1), we have

$$Q' = Q + 1. \tag{D9}$$

To obtain the actual value of Q (or Q') is a more delicate matter. As is clear from the analysis leading to (D7), Q is related to the difference between the total number of zeros of $R_l(k^2)$ and $k^{-1} \tan (kR - \frac{1}{2}l\pi)$, in the k plane. We will concentrate now on the calculation of Q', using the well-known formula[44]

$$n_0(f; D) - n_\infty(f; D) = \frac{1}{2i\pi} \int_\Gamma \frac{f'(z)}{f(z)}\, dz, \tag{D10}$$

where $n_\alpha(f; D)$ denotes the number of times the meromorphic function $f(z)$ takes the value α inside the domain D bounded by the simple Jordan curve Γ. The integral is taken in the counterclockwise sense along Γ, and it is assumed that there are no zeros or poles on the boundary Γ of the domain.

We have to compare the zeros of F [poles of $R_l(E)$] with zeros of $\cos (kR - \frac{1}{2}l\pi)$. From now on, we will work in the k plane where it is easier to keep track of the zeros and poles. Now, according to (44), the poles of $R_l(k^2)$ come from the zeros of the denominator:

$$D_l(k) = -w'_l(kR) + (-1)^l w'_l(-kR) S_l(k). \tag{D11}$$

The origin not being a pole of the R matrix,[8] it is easily seen from (46) and (47) that[45]

$$S_l(k) \underset{k \to 0}{=} 1 + O(k^{2l+1}). \tag{D12}$$

From this, it follows that both the numerator N_l, and the denominator D_l, of (44) behave like k^l near the origin.[45] It follows that the number of poles of $R_l(k^2)$ is given by the number of zeros of $\tilde{D}_l(k) = k^{-l} D_l(k)$. We are therefore interested in the difference

$$n_0(\tilde{D}_l; |k| < r) - n_0[\cos (kR - \tfrac{1}{2}l\pi); |k| < r], \tag{D13}$$

r large enough, so as to include all the zeros E_λ and E^0_λ which are not in a one-to-one correspondence to each other, i.e., all the zeros with $\lambda < \Lambda$. It is obvious that, making r larger and larger, because of the one-to-one correspondence between E_λ and E^0_λ for $\lambda > \Lambda$, the above difference remains fixed once we have included all the zeros corresponding to $\lambda \leq \Lambda$. We can therefore take the limit $r \to \infty$ in (D13). Let us denote by $n^{\pm}_{0,\infty}(f; r)$ the number of zeros or poles of a meromorphic function $f(z)$ inside the upper (+) and the lower (−) half of the circle $|z| = r$, respectively, and

[44] Reference 37, p. 391, formula (54).

[45] We have $w_l(z) = -v_l(z) - iu_l(z)$, where v_l has the parity l, and u_l the parity $l + 1$. In general, $w_l(z) = (G_l - H_l) \exp(-iz)$, where G_l and H_l are polynomials, with real coefficients, in z^{-1}, of degree l and $l - 1$, respectively. Their parities are l and $l - 1$.

by $n_{0,\infty}(f; r)$ their total number in $|z| < r$. We have

$$n_\infty(R_l; r) = n_0(F; r) = n_0(\tilde{D}_l = k^{-l}D_l; r) = -l + n_0(D_l; r), \tag{D14}$$

$$n_0(D_l; r) = n_0(D_l; r) - n_\infty(S_l; r) + n_\infty(S_l; r) = n_0(D_l; r) - n_\infty(D_l; r) + n_\infty(S_l; r), \tag{D15}$$

where use has been made of (D11), and[46]

$$n_\infty(S_l; r) = n_\infty^+(S_l; r) + n_\infty^-(S_l; r) = n + n_\infty^-(S_l; r), \tag{D16}$$

n being the number of bound states (poles in Im $k >$ 0). Now, from the symmetry relation

$$S_l(-k) = [S_l(k)]^{-1},^{47} \tag{D17}$$

and (D16), we deduce

$$n_\infty(S_l; r) = n + n_\infty^-(S_l; r) = 2n - [n_0^-(S_l; r) - n_\infty^-(S_l; r)]. \tag{D18}$$

Collecting the above results, we get finally

$$n_0(F, r) = -l + n_0(D_l; r) = n_0(D_l; r) - n_\infty(D_l; r) - l + 2n - [n_0^-(S_l; r) - n_\infty^-(S_l; r)]. \tag{D19}$$

To simplify the writing, let us make the change of variable $kR = \rho$. Applying (D10) to D_l and cos $(\rho - \frac{1}{2}l\pi)$, we get

$$[n_0(D_l; r) - n_0(\cos(\rho - \tfrac{1}{2}l\pi); r)]_{r\to\infty} = 2n + \frac{1}{2i\pi}\lim_{r\to\infty}\int_{\substack{|\rho|=r \\ \mathrm{Im}\,\rho\geq 0}}\left\{\frac{D_l'}{D_l} + \tan(\rho - \tfrac{1}{2}l\pi)\right\}d\rho + \frac{1}{2i\pi}\int_{-\infty}^{\infty}\frac{S_l'}{S_l}d\rho + \frac{1}{2i\pi}\lim_{r\to\infty}\int_{\substack{|\rho|=r \\ \mathrm{Im}\,\rho\leq 0}} \times\left\{\frac{D_l'}{D_l} + \tan(\rho - \tfrac{1}{2}l\pi) - \frac{S_l'}{S_l}\right\}d\rho. \tag{D20}$$

Now,

$$\frac{1}{2i\pi}\int_{-\infty}^{\infty}\frac{S_l'}{S_l}d\rho = \frac{1}{\pi}[\delta(\infty) - \delta(-\infty)] = -2N, \tag{D21}$$

and the other two integrals in the upper and the lower half-planes vanish in the limit $r \to \infty$ because of (61), (D11), and (32), as is easily verified.[48] It is understood here that, in accordance with the conditions of validity of (D10), the limit $r \to \infty$ has to be taken in such a way as to avoid the zeros and poles of D_l, S_l, and tan $(\rho - \frac{1}{2}l\pi)$. This is always possible because all these functions are meromorphic in the whole ρ plane, so that, excluding appropriate small intervals, we can make $r \to \infty$ without ever having zeros or poles on the circle $|\rho| = r$. The final result is

$$[n_0(F; r) - n_0(\cos(\rho - \tfrac{1}{2}l\pi))]_{r\to\infty} = -l + 2n - 2N. \tag{D22}$$

It follows that, for $0 < \arg k < \pi$,

$$F(k^2)_{k\to\infty} = A\,k^{-l+2n-2N}\cos(kR - \tfrac{1}{2}l\pi)(1 + o(1)), \tag{D23}$$

and

$$G(k^2)_{k\to\infty} = B\,k^{-l-1+2n-2N}\sin(kR - \tfrac{1}{2}l\pi)(1 + o(1)), \tag{D24}$$

where

$$A = \frac{\prod_{\mu\leq M}E_\mu^{0\prime}}{\prod_{\mu\leq M}E_\mu'}\prod_{\mu>M}^{\infty}\frac{E_\mu^{0\prime}}{E_\mu'}, \tag{D25}$$

$$B = \frac{\prod_{\lambda\leq\Lambda}E_\lambda^0}{\prod_{\lambda\leq\Lambda}E_\lambda}\prod_{\lambda>\Lambda}^{\infty}\frac{E_\lambda^0}{E_\lambda}. \tag{D26}$$

In fact, $A = BR(0)$ because of (C1). Note that the above asymptotic behaviors are even under $k \to -k$ irrespective of the parity of l. If the Levinson theorem holds, we have $n = N$.

[46] Remember that S_l has no poles or zeros on the real axis.

[47] In Sec. III, this property was stated for k real. However, it is obvious that it holds in general in the complex plane.

[48] It is sufficient to use the asymptotic behavior of the functions under the integrals, and the fact that $w_l'' + [1 - l(l+1)z^{-2}]w_l = 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Asymptotic Behavior of Certain Nonlinear Boundary-Value Problems

JOSÉ CANOSA
IBM Scientific Center, Palo Alto, California
AND
JULIAN COLE
California Institute of Technology, Pasadena, California

(Received 5 February 1968)

The asymptotic properties of a class of nonlinear boundary-value problems are studied. For large values of a parameter, the differential equation is of the singular-perturbation type, and its solution is constructed by means of matched asymptotic expansions. In two special cases, very simple approximate analytic solutions are obtained, and their accuracy is illustrated by showing their good agreement with the exact numerical solution of the problem.

I. INTRODUCTION

In a previous paper,[1] an expansion method for nonlinear boundary-value problems was presented. The eigenfunction-expansion basis used was defined by the associate linear boundary-value problem. As the problem became strongly nonlinear, more terms in the expansion were needed for accuracy. We present here a study of the asymptotic properties of the same class of problems in slab geometry. In two special cases, approximate asymptotic solutions are obtained whose accuracy increases as the nonlinear distribution departs more from the linear distribution. Thus, when more than two terms in the expansion method are needed for accuracy, the present asymptotic solutions give very simply a distribution which is in excellent agreement with the exact numerical solution. It is shown that in the general case an approximate solution can be obtained in terms of asymptotic expansions.

II. THE NONLINEAR PROBLEM

The problem studied arose in connection with the distribution of the energy released in a nuclear power reactor as a result of a power excursion.[2] The distribution is given by the following nonlinear boundary-value problem:

$$\frac{d^2y}{dx^2} + r^2(x)y - y^n = 0,$$
$$y(0) = y(\pi) = 0. \tag{1}$$

In (1), $r^2(x)$ is the space-dependent perturbation in the neutron multiplication of a reactor with an infinitesimally small initial neutron population; the perturbation makes the reactor supercritical, thus leading to the release of energy. The energy release through various nonlinear physical effects, such as Doppler broadening of the neutron-absorption cross section or moderator-expansion effects, reduces the neutron multiplication and the neutron population eventually becomes zero. Equation (1) gives the distribution of the energy release from the start of the perturbation till the neutron population again becomes zero. The exponent n is a positive number, and $n \geq 2$.

Due to the physical nature of the problem, we assume the existence of a solution which is continuous, positive, and bounded in the domain. The uniqueness of the solution is proved by the following:

Theorem.[3] If a positive solution of Eq. (1) exists, it is unique.

Proof. Let us assume that there are two positive solutions, y_1 and y_2, and let

$$\eta = y_1 - y_2. \tag{2}$$

Then

$$\frac{d^2\eta}{dx^2} + r^2(x)\eta = y_1^n - y_2^n,$$
$$\eta(0) = \eta(\pi) = 0. \tag{3}$$

Define the following implicit function of x:

$$g(y_1, y_2) = n\int_0^1 [ty_1 + (1-t)y_2]^{n-1}\,dt$$
$$= \frac{[ty_1 + (1-t)y_2]^n}{y_1 - y_2}\Bigg|_0^1$$
$$= \frac{y_1^n - y_2^n}{y_1 - y_2} = \frac{y_1^n - y_2^n}{\eta}. \tag{4}$$

Thus η satisfies

$$\frac{d^2\eta}{dx^2} + [r^2(x) - g(y_1, y_2)]\eta = 0, \quad \eta(0) = \eta(\pi) = 0, \tag{5}$$

[1] J. Canosa, J. Math. Phys. **8**, 2180 (1967).
[2] W. L. Ergen, Trans. Am. Nucl. Soc. **8**, 221 (1965).
[3] H. B. Keller (private communication).

and also

$$\frac{d^2y_1}{dx^2} + [r^2(x) - g(y_1, 0)]y_1 = 0, \quad y_1(0) = y_1(\pi) = 0. \tag{6}$$

But, as $n \geq 2$ and y_1 and y_2 are assumed positive, obviously

$$[ty_1 + (1-t)y_2]^{n-1} \geq [ty_1]^{n-1}, \quad 0 \leq t \leq 1, \tag{7}$$

and so from (4),

$$g(y_1, y_2) > g(y_1, 0) \tag{8}$$

in any interval of x where y_1 and y_2 are positive. Therefore,

$$r^2(x) - g(y_1, y_2) < r^2(x) - g(y_1, 0). \tag{9}$$

Sturm's fundamental comparison theorem[4] can now be directly applied to Eqs. (5) and (6), even if these are nonlinear. As $\eta(0) = \eta(\pi) = 0$ and $y_1(0) = 0$ by hypothesis, Sturm's theorem shows that, as y_1 oscillates more rapidly than η in the interval $(0, \pi)$, y_1 must vanish at least once between 0 and π. This contradicts the assumption that two positive solutions exist.

We now present some results of interest that have been obtained for the general case where $r^2(x)$ is a *positive* function in the domain, at least piecewise continuous, and $n \geq 2$.

It is of some interest to point out first that no positive solution can exist unless

$$r^2(\bar{x}) > 1, \tag{10}$$

where $r^2(\bar{x})$ is the value of $r^2(x)$ at some point $\bar{x}$ inside the domain. For r^2 constant, we have the condition

$$r^2 > 1. \tag{11}$$

The proof is quite simple and is based on the extension of an idea due to Kastenberg.[5] The fundamental eigenfunction of the Helmholtz equation obeys

$$\frac{d^2y_1}{dx^2} + \lambda_1 y_1 = 0. \tag{12}$$

If we multiply (12) by y, the solution of (1) and (1) by y_1, subtract and integrate over the domain, we get, after using Green's theorem,

$$\int_0^\pi [\lambda_1 - r^2(x)]yy_1\,dx + \int_0^\pi y_1 y^n\,dx = 0. \tag{13}$$

As y and y_1 are strictly positive, (13) can only be satisfied if

$$\int_0^\pi [\lambda_1 - r^2(x)]yy_1\,dx < 0. \tag{14}$$

Applying the mean-value theorem,

$$[\lambda_1 - r^2(\bar{x})]\int_0^\pi yy_1\,dx < 0,$$

that is,

$$\lambda_1 - r^2(\bar{x}) < 0,$$

and so

$$r^2(\bar{x}) > \lambda_1 = 1, \tag{15a}$$

which is the desired result (10) from which (11) follows when r^2 is constant. The usefulness of this result is that if

$$r^2 = \text{const} = 1, \tag{15b}$$

Eq. (1) has no solution, while the linear equation associated with it has one ($y = \sin x$). Therefore, if we confine our discussion to the case

$$r^2(x) \geq 1, \quad 0 \leq x \leq \pi, \tag{15c}$$

(as it is done in this paper), the average value of $r^2(x)$ over the domain is a measure of the nonlinearity of the problem; i.e., if

$$\overline{r^2(x)} \to 1,$$

the solution is close to that of the linear problem associated with (1) with the condition (15b); and if

$$\overline{r^2(x)} \to \infty,$$

the solution departs strongly from that of the linear problem.

III. UPPER AND LOWER BOUNDS FOR THE MAXIMUM OF THE DISTRIBUTION

One should also note that, when

$$\overline{r^2(x)} \to 1, \tag{16}$$

corresponding physically to very small perturbation increases in the neutron multiplication, the energy release is quite small and the distribution given by (1) approaches (except a multiplicative constant) the fundamental mode of the general eigenvalue equation

$$\frac{d^2y_1}{dx^2} + r^2(x)y_1 - \lambda_1 y_1 = 0, \quad y(0) = y(\pi) = 0. \tag{17}$$

The expression (16) defines an asymptotic limit of the problem. In this paper, we will study mainly the other asymptotic limit when

$$\overline{r^2(x)} \to \infty, \tag{18}$$

[4] E. L. Ince, *Ordinary Differential Equations* (Dover Publications, Inc., New York, 1956), p. 224.

[5] W. E. Kastenberg, Ph.D. dissertation, University of California, Berkeley (1966).

corresponding to large perturbations in the neutron multiplication, large energy releases, and thus strong nonlinear effects. In this case, the distribution given by (1) departs strongly from the linear distribution (17).

We now derive upper and lower bounds for the maximum value of the distribution given by (1). Because of the conjecture in Sec. II, the solution must have at least a maximum. Let the maximum value of the distribution occur at $x = a$; then

$$y(a) = M, \quad \frac{dy(a)}{dx} = 0, \quad \frac{d^2y(a)}{dx^2} < 0. \tag{19}$$

Thus, from (1) and (19),

$$\frac{d^2y(a)}{dx^2} = y^n(a) - r^2(a)y(a) < 0. \tag{20}$$

As $r^2(a)$ and $y(a)$ are strictly positive quantities, we finally have

$$y(a) < [r^2(a)]^{1/n-1}, \tag{21}$$

so that an upper bound for the solution is given by

$$y(x) < [\text{maximum of } r^2(x)]^{1/(n-1)}. \tag{22}$$

If r^2 is constant, the solution has only one maximum at the center of the domain.[1]

The lower bound for the peak of the distribution is now derived. We multiply (1) by y_1, the fundamental eigenfunction given by (17), Eq. (17) by y, the solution of (1), subtract and integrate over the domain. After applying Green's theorem and noting that the resulting surface integrals vanish because of the boundary conditions satisfied by y and y_1, we get

$$-\int_0^\pi y^n y_1\,dx + \lambda_1\int_0^\pi yy_1\,dx = 0. \tag{23}$$

Therefore

$$\lambda_1 = \int_0^\pi y^n y_1\,dx \Big/ \int_0^\pi yy_1\,dx. \tag{24}$$

As the lower bound of the distribution is zero, the application of the mean-value theorem leads to the following inequality:

$$0 \le \int_0^\pi y^n y_1\,dx \le M^{n-1}\int_0^\pi yy_1\,dx, \tag{25}$$

where M is the maximum value of the distribution y in the domain. Using now (24), Eq. (25) becomes

$$0 \le \lambda_1\int_0^\pi yy_1\,dx \le M^{n-1}\int_0^\pi yy_1\,dx \tag{26}$$

or

$$M^{n-1} \ge \lambda_1, \tag{27}$$

and finally,

$$M \ge \lambda_1^{1/(n-1)}. \tag{28}$$

Therefore, from Eqs. (22) and (28), it follows that the maximum value of the distribution is bounded by

$$\lambda_1^{1/(n-1)} \le M < [\text{maximum of } r^2(x)]^{1/(n-1)}. \tag{29}$$

To illustrate the inequality (29) explicitly, consider Eq. (1) where r^2 is constant and $n = 2$; then the fundamental eigenvalue of (17) is $\lambda_1 = r^2 - 1$, and (29) becomes

$$r^2 - 1 \le M < r^2. \tag{30}$$

In the limit case $r^2 \to \infty$, it is clear from (30) that

$$M \to r^2. \tag{31}$$

IV. GENERAL ASYMPTOTIC SOLUTION

We seek the solution of

$$\frac{d^2y}{dx^2} + r^2(x)y - y^n = 0, \quad y(0) = y(\pi) = 0, \tag{32}$$

in the asymptotic case

$$r^2(x) = \lambda\rho(x), \tag{33}$$

where

$$\lambda \to \infty \quad \text{and} \quad \rho(x) = O(1), \tag{34}$$

and both λ and $\rho(x)$ are positive. The general result (29) indicates that the solution of (32) away from the boundaries is of the same order of magnitude as $[r^2(x)]^{1/(n-1)}$. This suggests the introduction of a new variable

$$y = \lambda^{1/(n-1)}Y, \quad Y = O(1), \tag{35}$$

in terms of which (32) becomes

$$\epsilon\frac{d^2Y}{dx^2} + \rho(x)Y = Y^n \tag{36}$$

with

$$\epsilon \equiv 1/\lambda \to 0. \tag{37}$$

Equation (36) is in the form of a singular perturbation equation, the solution of which will be sought in the form of asymptotic expansions.

A. Special Cases

We will first give the formal solution of Eq. (36) in the simplest case, where

$$\rho(x) = \text{const} = 1, \tag{38}$$

and then extend the treatment to the general case in which $\rho(x)$ is not a constant.

From a representation of Eq. (36), with the property (38), in the phase plane $(Y, dY/dx)$, it can be shown that its solution has a maximum at the domain center and is symmetric about it, so that one needs only to construct the solution in $0 \le x \le \pi/2$.

For $n = 2$, Eq. (36) becomes

$$\epsilon \frac{d^2Y}{dx^2} + Y - Y^2 = 0 \tag{39}$$

and its solution proceeds as follows:

(i) *Outer Expansion.* We assume the solution in the form of an asymptotic series

$$Y(x, \epsilon) = \sum_{i=0}^{\infty} \epsilon^i f_i(x) = f_0(x) + \epsilon f_1(x) + \cdots. \tag{40}$$

Substituting into (39) and grouping like powers of ϵ, we can determine the functions $f_i(x)$ recursively. One gets

$$f_0(x) = 1 \quad \text{and} \quad f_i(x) = 0, \quad i > 0, \tag{41}$$

so that (40) degenerates into

$$Y(x, \epsilon) = 1. \tag{42}$$

Transforming back to the original variable y, Eq. (35),

$$y(x, \epsilon) = \lambda Y(x, \epsilon) = \lambda = 1/\epsilon = r^2 \tag{43}$$

[see (33) and (38)]. That is, away from the boundaries, the solution is given asymptotically by the straight line (43), a result already suggested by (30).

(ii) *Boundary-Layer (Inner) Solution.* To study the solution near the boundary $x = 0$, define a new length

$$\tilde{x} = x/(\epsilon)^{\frac{1}{2}}. \tag{44}$$

Then Eq. (39) becomes

$$\frac{d^2Y}{d\tilde{x}^2} + Y - Y^2 = 0. \tag{45}$$

Equation (45) has a first integral

$$\left(\frac{dY}{d\tilde{x}}\right)^2 + Y^2 - \tfrac{2}{3}Y^3 = \text{const.} \tag{46}$$

The boundary-layer solution near $x = 0$ is associated with the limit process $\epsilon \to 0$, $\tilde{x}$ fixed, and has the form

$$Y(\tilde{x}, \epsilon) = g_0(\tilde{x}) + \beta_1(\epsilon) g_1(\tilde{x}) + \cdots. \tag{47}$$

$\beta_1(\epsilon) \to 0$ as $\epsilon \to 0$; it is to be found in general so that consistent problems occur for the higher approximations. Only the first term is considered here and satisfies essentially (45):

$$\frac{d^2g_0}{d\tilde{x}^2} + g_0 - g_0^2 = 0. \tag{48}$$

Now the boundary conditions at ∞ for g_0 no longer involve ϵ but are determined by matching with the outer expansion (42). For this simple case the matching occurs as $\tilde{x} \to \infty$, $x \to 0$, and requires

$$g_0(\infty) = 1, \quad dg_0(\infty)/d\tilde{x} = 0. \tag{49}$$

The boundary condition (49) fixes the constant of integration in the integral corresponding to (46):

$$\left(\frac{dg_0}{d\tilde{x}}\right)^2 + g_0^2 - \tfrac{2}{3}g_0^3 = \tfrac{1}{3}. \tag{50}$$

The variables in (50) can be separated and the problem is reduced to one quadrature, where the integration constant is determined by the boundary condition at $\tilde{x} = 0$:

$$g_0(0) = 0. \tag{51}$$

After carrying the integration, we get

$$g_0(\tilde{x}) = \frac{1 - 4(2 - \sqrt{3})e^{-\tilde{x}} + (7 - 4\sqrt{3})e^{-2\tilde{x}}}{1 + 2(2 - \sqrt{3})e^{-\tilde{x}} + (7 - 4\sqrt{3})e^{-2\tilde{x}}}. \tag{52}$$

There is a corresponding boundary layer near $x = \pi$ with $\tilde{x}$ in (44) replaced by $\tilde{\tilde{x}} = (\pi - x)/(\epsilon)^{\frac{1}{2}}$. Because of the symmetry of this case, it is sufficient to consider $0 \leq x \leq \pi/2$. Therefore the boundary layer solution contains the outer solution, and thus is uniformly valid in the range $0 \leq x \leq \pi/2$. In Table I we give

TABLE I. Asymptotic approximation vs exact solution.

$x/(\pi/2)$	$r^2 = 8$		$r^2 = 10$		$r^2 = 16$		$r^2 = 20$	
	Exact	Asymptotic	Exact	Asymptotic	Exact	Asymptotic	Exact	Asymptotic
0	0	0	0	0	0	0	0	0
0.1	1.989	1.994	2.766	2.768	5.495	5.495	7.586	7.586
0.2	3.700	3.709	5.069	5.073	9.679	9.680	13.05	13.05
0.3	5.029	5.041	6.772	6.778	12.39	12.39	16.33	16.33
0.4	5.993	6.009	7.941	7.949	14.00	14.00	18.12	18.12
0.5	6.660	6.682	8.704	8.717	14.91	14.91	19.05	19.06
0.6	7.106	7.138	9.187	9.205	15.41	15.41	19.53	19.53
0.7	7.394	7.440	9.483	9.511	15.68	15.68	19.76	19.76
0.8	7.569	7.638	9.657	9.701	15.82	15.83	19.88	19.88
0.9	7.663	7.766	9.746	9.817	15.88	15.91	19.93	19.94
1.0	7.693	7.850	9.774	9.888	15.90	15.95	19.94	19.97

TABLE II. Asymptotic approximation vs exact solution

$x/(\pi/2)$	$r^2 = 2.25$		$r^2 = 4$		$r^2 = 9$		$r^2 = 16$	
	Exact	Asymptotic	Exact	Asymptotic	Exact	Asymptotic	Exact	Asymptotic
0	0	0	0	0	0	0	0	0
0.1	0.2367	0.2476	0.4351	0.4371	0.9641	0.9642	1.668	1.669
0.2	0.4608	0.4821	0.8306	0.8344	1.748	1.748	2.843	2.843
0.3	0.6620	0.6930	1.160	1.165	2.284	2.284	3.480	3.480
0.4	0.8340	0.8739	1.414	1.421	2.610	2.610	3.777	3.777
0.5	0.9743	1.023	1.599	1.609	2.793	2.793	3.907	3.907
0.6	1.083	1.142	1.728	1.740	2.891	2.892	3.961	3.961
0.7	1.164	1.235	1.813	1.829	2.943	2.944	3.984	3.984
0.8	1.219	1.305	1.865	1.889	2.969	2.971	3.993	3.993
0.9	1.250	1.357	1.894	1.930	2.981	2.985	3.997	3.997
1.0	1.260	1.396	1.903	1.953	2.985	2.992	3.998	3.999

the results obtained from (52) together with the direct numerical solution of Eq. (39) for several values of r^2 (note that $r^2 = \lambda = 1/\epsilon$). The agreement with the exact results is already quite good for $r^2 = 8$, the maximum error (2%) occurring at the domain center. The table shows that, as r^2 increases, the agreement improves rapidly and the discrepancy from the exact solution is only one or two units in the fourth significant figure. The general result (30) is also illustrated in Table I. The expansion method,[1] in return, gave good results for $r^2 \to 1$, where the distribution approaches the linear (sine) distribution defined by (17), and where only one term in the expansion was necessary; with two terms, the expansion method gave good results for $r^2 \lessapprox 8$, above which more terms were needed and so the method became quite laborious. The present asymptotic solution thus complements the expansion method, and gives the solution for strongly nonlinear problems where the distribution is markedly different from the linear distribution.

Another special case in which an explicit asymptotic solution can be obtained is when $n = 3$ in Eq. (36). The general result (29) becomes

$$(r^2 - 1)^{\frac{1}{2}} \leq M < r \equiv 1/(\epsilon)^{\frac{1}{2}}, \tag{53}$$

so that in the limit $r^2 \to \infty$, the solution at the center approaches asymptotically the upper bound

$$y(\pi/2) \to 1/(\epsilon)^{\frac{1}{2}}.$$

The procedure is now formally the same as that followed for Eq. (39). The results corresponding to Eqs. (42) and (52) are, respectively,

$$Y(x, \epsilon) = 1 \tag{54}$$

and

$$g_0(\tilde{x}) = \frac{\exp(\sqrt{2}\,\tilde{x}) - 1}{\exp(\sqrt{2}\,\tilde{x}) + 1}. \tag{55}$$

In Table II we give the results obtained from Eq. (55) together with the numerical solution of Eq. (36) with $n = 3$, for several values of r^2. It is seen that the agreement is excellent for $r^2 \gtrapprox 4$.

The ranges of validity of the asymptotic approximations are obtained readily from Eqs. (52) and (55). For if these equations are asymptotically to approach the outer solution, Eq. (42), we must have

$$Y\left(\frac{\pi}{2(\epsilon)^{\frac{1}{2}}}\right) \to 1,$$

that is,

$$2(2 - \sqrt{3})e^{-\pi/2(\epsilon)^{\frac{1}{2}}} \ll 1, \quad e^{2^{\frac{1}{2}}\pi/2(\epsilon)^{\frac{1}{2}}} \gg 1,$$

respectively, so that finally we get

$$r^2 \geq 7 \quad \text{and} \quad r^2 \geq 4$$

for the ranges of validity of (52) and (55).

These two special cases ($n = 2$ and 3) seem to be the only ones in which it is possible to obtain the indefinite integrals arising from the separation of variables in the result corresponding to (48) for a general value of n.

When $n \geq 4$, although no explicit solution can be obtained, the solution of (36) can be reduced to a simple quadrature. The general result (29) is now

$$(r^2 - 1)^{1/(n-1)} \leq M < r^{2/(r-1)} \equiv (1/\epsilon)^{1/(n-1)}, \tag{56}$$

so that for $r^2 \to \infty$, the maximum of the solution at the center of the domain approaches asymptotically

$$y(\pi/2) = (1/\epsilon)^{1/(n-1)}. \tag{57}$$

That is, as is seen by inspection of Eq. (36) together with the restriction (38), the outer solution is always

$$Y(x, \epsilon) = 1, \tag{58}$$

independently of n. Furthermore, the equation

corresponding to (48),

$$\frac{d^2g_0}{d\tilde{x}^2} + g_0 - g_0^n = 0, \tag{59}$$

always has a first integral

$$\left(\frac{dg_0}{d\tilde{x}}\right)^2 + g_0^2 - \frac{2}{n+1}\, g_0^{n+1} = \text{const}, \tag{60}$$

whose constant is obtained with the usual matching condition with (58), i.e.,

$$g_0(\infty) = 1, \quad \frac{dg_0(\infty)}{d\tilde{x}} = 0. \tag{61}$$

Thus (60) becomes

$$\left(\frac{dg_0}{d\tilde{x}}\right)^2 = \frac{n-1}{n+1} - g_0^2 + \frac{2g_0^{n+1}}{n+1} \equiv P_{n+1}(g_0). \tag{62}$$

It can be shown by elementary means that the asymptotic value of the boundary layer solution, i.e., $g_0 = 1$, is a double root of the polynomial $P_{n+1}(g_0)$ appearing in (62). This remarkable property of the problem reduces the solution of (59) to a straightforward quadrature. Because of this property, we can rewrite (62) in the form

$$\left(\frac{dg_0}{d\tilde{x}}\right)^2 = (g_0 - 1)^2 Q_{n-1}(g_0), \tag{63}$$

where Q_{n-1} means simply a certain polynomial of degree $n-1$ in g_0. Therefore the solution of (59) is given by

$$-\int_0^{g_0} \frac{dg_0}{(g_0 - 1)[Q_{n-1}(g_0)]^{\frac{1}{2}}} = \tilde{x}, \tag{64}$$

where the integration constant has been determined by the boundary condition at $\tilde{x} = 0$, and the sign by the fact that $dg_0(0)/d\tilde{x} > 0$. For $g_0 = 1$, the integral (64) becomes infinite as it should [see first equation in (61)]. In an actual calculation we are only interested in the range $0 \leq \tilde{x} \leq \pi/2(\epsilon)^{\frac{1}{2}}$. For values of g_0 not too near 1, the numerical integration of (64) can be done by elementary means (e.g., Simpson's rule) because the integrand is well behaved. For values $g_0 \to 1$, the form of the integrand (64) allows us to avoid numerical difficulties, because $Q_{n-1}(g_0)$ does not vanish at $g_0 = 1$, and can be extracted from the integrand; the form of the remaining part of the integrand, which can be integrated in closed form, shows clearly the exponential decay of the boundary layer solution into the outer solution.

B. General Case

We now briefly describe the asymptotic behavior of the solution of Eq. (36) in the general case in which $\rho(x)$ is a well-behaved function. The formal procedure is quite similar to that of the previous section:

(i) *Outer Expansion.* We assume a solution

$$Y(x, \epsilon) = f_0(x) + \epsilon f_1(x) + \cdots, \tag{65}$$

and, following the usual procedure, we get

$$f_0(x) = \rho(x)^{1/(n-1)}, \tag{66}$$

$$f_1(x) = \frac{1}{n-1}\,\frac{[\rho(x)^{1/(n-1)}]''}{\rho(x)}. \tag{67}$$

In the first-order treatment to which we restrict ourselves, only the first term in (65) is needed, and we expect it to approximate the solution of (36) away from the boundaries.

(ii) *Boundary-Layer (Inner) Expansion.* We define a new length by

$$\tilde{x} = \frac{x}{\epsilon^{\frac{1}{2}}}, \tag{68}$$

where $\tilde{x}$ is fixed, i.e., $O(1)$. As (68) shows, when $\epsilon \to 0$, the actual thickness of the boundary layer approaches zero as $\epsilon^{\frac{1}{2}}$. The outer expansion (65) is then expected to be valid in most of the domain, except in the infinitesimally thin boundary layer. With (68), Eq. (36) becomes

$$\frac{d^2Y}{d\tilde{x}^2} + \rho(\epsilon^{\frac{1}{2}}\tilde{x})Y = Y^n. \tag{69}$$

Let us expand

$$\begin{aligned} Y(\tilde{x}, \epsilon) &= g_0(\tilde{x}) + \epsilon^{\frac{1}{2}} g_1(\tilde{x}) + \cdots, \\ \rho(\epsilon^{\frac{1}{2}}\tilde{x}) &= \rho_0 + \rho_0'\epsilon^{\frac{1}{2}}\tilde{x} + \cdots, \end{aligned} \tag{70}$$

where $\rho_0 = \rho(0)$, $\rho_0' = \rho'(0)$, etc. Substituting (70) into (69) and proceeding in the usual fashion, we

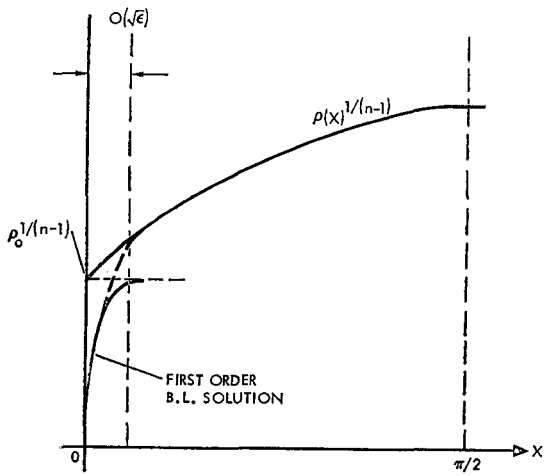


FIG. 1. Illustration of the results of a first-order treatment.

obtain the first approximation to the boundary layer:

$$\frac{d^2g_0}{d\tilde{x}^2} + \rho_0 g_0 = g_0^n, \tag{71}$$

which is of the same form as (59). Equation (71) shows then that, to first order, the boundary-layer solution has the same form independently of the coefficient $\rho(x)$ in Eq. (36). Therefore, all the discussion about the boundary layer presented in the previous section applies equally here. In this first-order treatment, the rough matching of (71) with the outer solution (65) is achieved by imposing

$$g_0(\infty) = \rho_0^{1/(n-1)},$$

i.e., the boundary-layer solution must approach asymptotically the value of the outer solution at the origin. Note that, for x very small although fixed, as $\tilde{x} = x/\epsilon^{\frac{1}{2}}$, if $\epsilon \to 0$, $\tilde{x} \to \infty$, and so the actual distance from the origin at which the boundary layer solution blends with the outer solution becomes increasingly small with ϵ. For clarity, the results of the treatment are illustrated in Fig. 1, assuming that $\rho(x)$ and thus the solution of (36) is symmetric. The matching of the boundary layer and the outer solutions will be somewhat rough in this first-order solution. However, a uniformly valid first approximation is

$$Y(x, \epsilon) = f_0(x) + g_0(\tilde{x}) - \rho_0^{1/(n-1)} + O(1)$$

and defines a smooth curve.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Microscopic Approach to Kinetic Theory: Inhomogeneous Systems

WILLARD R. CHAPPELL*

Department of Physics and Astrophysics, and Joint Institute for Laboratory Astrophysics
University of Colorado, Boulder, Colorado

(Received 1 December 1967)

The microscopic linearized Vlasov equation is solved in terms of a generalized inverse dielectric function $\epsilon^{-1}(\mathbf{r}, \mathbf{r}'; t, t')$ and the initial phase-space density fluctuation. This expression is then used to calculate the density autocorrelation function and to obtain a generalized kinetic equation for plasmas and gravitational gases. It is shown that many of the results for the inhomogeneous system have a close similarity to the corresponding results for homogeneous systems. In particular, the test-particle theory is exhibited and an expression is obtained for the phase-space density of the polarization cloud associated with a test particle.

1. INTRODUCTION

Much attention in recent years has been devoted to the problem of deriving kinetic equations for homogeneous plasmas. Several methods have been applied to this problem and many cases have been treated successfully. Various authors have succeeded in obtaining kinetic equations for systems with only the coulomb interaction,[1–4] for systems including the electromagnetic field,[5–7] and for quantum systems.[8–10] The problem of a convergent kinetic equation has also been considered by several authors.[11–13]

The methods used for these problems in general fall into three groups. One of these groups uses the BBKGY hierarchy for the reduced distribution functions as its starting point.[2–4] Another group follows the techniques introduced by Prigogine and his co-workers.[14] A third technique[15–17] which has recently become popular is the outgrowth of work by

* Supported in part by ARL, OAR, USAF 33(615)2705, through P.E.C. Research Associates, Boulder, Colorado.

[1] R. Balescu, Phys. Fluids **3**, 52 (1960).
[2] A. Lenard, Ann. Phys. (N.Y.) **10**, 390 (1960).
[3] R. L. Guernsey, Ph.D. thesis, University of Michigan (1960).
[4] N. Rostoker and M. N. Rosenbluth, Phys. Fluids **3**, 1 (1960).
[5] T. H. Dupree, Phys. Fluids **6**, 1714 (1963).
[6] N. Rostoker, R. E. Aamodt, and O. C. Eldridge, Ann. Phys. (N.Y.) **31**, 243 (1965).
[7] C. S. Wu, *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach, Science Publ., Inc., New York, 1967), Vol. IX.
[8] R. Balescu, Phys. Fluids **4**, 94 (1961).
[9] A. Ron, Phys. Rev. **134**, A70 (1964).
[10] W. R. Chappell, Phys. Rev. **152**, 113 (1966); J. Math. Phys. **7**, 1153 (1966).
[11] J. Weinstock, Phys. Rev. **133**, 673 (1964); Phys. Fluids **10**, 127 (1967).
[12] R. L. Guernsey, Phys. Fluids **7**, 1600 (1964).
[13] H. Gould, *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach Science Publ., Inc., New York, 1967), Vol. IX.
[14] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962); R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience Publishers, Inc., New York, 1963).
[15] J. Dawson and T. Nakayama, Phys. Fluids **9**, 252 (1966); J. Math. Phys. **8**, 553 (1967).
[16] W. R. Chappell, J. Math. Phys. **8**, 298 (1967); *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach Science Publ., Inc., New York, 1967), Vol. IX.
[17] J. Price, Ph.D. thesis, University of California (1966).

obtain the first approximation to the boundary layer:

$$\frac{d^2g_0}{d\tilde{x}^2} + \rho_0 g_0 = g_0^n, \tag{71}$$

which is of the same form as (59). Equation (71) shows then that, to first order, the boundary-layer solution has the same form independently of the coefficient $\rho(x)$ in Eq. (36). Therefore, all the discussion about the boundary layer presented in the previous section applies equally here. In this first-order treatment, the rough matching of (71) with the outer solution (65) is achieved by imposing

$$g_0(\infty) = \rho_0^{1/(n-1)},$$

i.e., the boundary-layer solution must approach asymptotically the value of the outer solution at the origin. Note that, for x very small although fixed, as $\tilde{x} = x/\epsilon^{\frac{1}{2}}$, if $\epsilon \to 0$, $\tilde{x} \to \infty$, and so the actual distance from the origin at which the boundary layer solution blends with the outer solution becomes increasingly small with ϵ. For clarity, the results of the treatment are illustrated in Fig. 1, assuming that $\rho(x)$ and thus the solution of (36) is symmetric. The matching of the boundary layer and the outer solutions will be somewhat rough in this first-order solution. However, a uniformly valid first approximation is

$$Y(x, \epsilon) = f_0(x) + g_0(\tilde{x}) - \rho_0^{1/(n-1)} + O(1)$$

and defines a smooth curve.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Microscopic Approach to Kinetic Theory: Inhomogeneous Systems

WILLARD R. CHAPPELL*

Department of Physics and Astrophysics, and Joint Institute for Laboratory Astrophysics
University of Colorado, Boulder, Colorado

(Received 1 December 1967)

The microscopic linearized Vlasov equation is solved in terms of a generalized inverse dielectric function $\epsilon^{-1}(\mathbf{r}, \mathbf{r}'; t, t')$ and the initial phase-space density fluctuation. This expression is then used to calculate the density autocorrelation function and to obtain a generalized kinetic equation for plasmas and gravitational gases. It is shown that many of the results for the inhomogeneous system have a close similarity to the corresponding results for homogeneous systems. In particular, the test-particle theory is exhibited and an expression is obtained for the phase-space density of the polarization cloud associated with a test particle.

1. INTRODUCTION

Much attention in recent years has been devoted to the problem of deriving kinetic equations for homogeneous plasmas. Several methods have been applied to this problem and many cases have been treated successfully. Various authors have succeeded in obtaining kinetic equations for systems with only the coulomb interaction,[1–4] for systems including the electromagnetic field,[5–7] and for quantum systems.[8–10] The problem of a convergent kinetic equation has also been considered by several authors.[11–13]

The methods used for these problems in general fall into three groups. One of these groups uses the BBKGY hierarchy for the reduced distribution functions as its starting point.[2–4] Another group follows the techniques introduced by Prigogine and his co-workers.[14] A third technique[15–17] which has recently become popular is the outgrowth of work by

* Supported in part by ARL, OAR, USAF 33(615)2705, through P.E.C. Research Associates, Boulder, Colorado.

[1] R. Balescu, Phys. Fluids **3**, 52 (1960).
[2] A. Lenard, Ann. Phys. (N.Y.) **10**, 390 (1960).
[3] R. L. Guernsey, Ph.D. thesis, University of Michigan (1960).
[4] N. Rostoker and M. N. Rosenbluth, Phys. Fluids **3**, 1 (1960).
[5] T. H. Dupree, Phys. Fluids **6**, 1714 (1963).
[6] N. Rostoker, R. E. Aamodt, and O. C. Eldridge, Ann. Phys. (N.Y.) **31**, 243 (1965).
[7] C. S. Wu, *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach, Science Publ., Inc., New York, 1967), Vol. IX.
[8] R. Balescu, Phys. Fluids **4**, 94 (1961).
[9] A. Ron, Phys. Rev. **134**, A70 (1964).
[10] W. R. Chappell, Phys. Rev. **152**, 113 (1966); J. Math. Phys. **7**, 1153 (1966).
[11] J. Weinstock, Phys. Rev. **133**, 673 (1964); Phys. Fluids **10**, 127 (1967).
[12] R. L. Guernsey, Phys. Fluids **7**, 1600 (1964).
[13] H. Gould, *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach Science Publ., Inc., New York, 1967), Vol. IX.
[14] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962); R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience Publishers, Inc., New York, 1963).
[15] J. Dawson and T. Nakayama, Phys. Fluids **9**, 252 (1966); J. Math. Phys. **8**, 553 (1967).
[16] W. R. Chappell, J. Math. Phys. **8**, 298 (1967); *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach Science Publ., Inc., New York, 1967), Vol. IX.
[17] J. Price, Ph.D. thesis, University of California (1966).

Klimontovich[18] and Dupree[5] and is based on the fact that the microscopic distribution function satisfies exactly the Vlasov equation.

In a recent article[16] we discussed this approach for the case of homogeneous systems. We showed that we could obtain an asymptotic ($t \to \infty$) expression for the microscopic distribution function by solving an approximate equation of motion. This approximate equation is simply the linearized Vlasov equation for the microscopic distribution function. It is equivalent to the equations of motion for the two-particle correlation and autocorrelation functions when the three-particle correlation function and the effects of close collisions are negligible. The use of this asymptotic expression to calculate the correlation function yields the Balescu–Lenard–Guernsey kinetic equation and other results. This asymptotic expression for the microscopic distribution function also yields the test-particle results which express the system as a collection of noninteracting quasiparticles. A quasiparticle is composed of a particle moving at constant velocity plus the associated polarization cloud induced in the medium by the particle. We also indicated that the test-particle approach could be extended to inhomogeneous systems. But we were unable to obtain an expression for the phase-space density of the polarization cloud. Finally, we obtained a set of equations describing a homogeneous, slowly varying system which contains small, inhomogeneous, quickly varying perturbations.

In this paper we discuss the generalization of the previous results to inhomogeneous systems. The inhomogeneous case has been considered by a number of authors.[19–21] But these considerations have been limited to considering the inhomogeneous system as a perturbation on a homogeneous system (generally upon the equilibrium system). The resulting equations are quite difficult to interpret. This is quite different from the homogeneous case where, generally, the kinetic equations have very simple interpretations in terms of the, golden rule, using dynamically shielded potentials or in terms of the Boltzmann equation.[13,22] Another difficulty is that these results cannot be applied to the gravitational gas because, unlike the plasma, it has no tendency toward homogeneity.

We show that the inhomogeneous system does have many of the characteristics of the homogeneous system. In particular, we can solve the microscopic linearized Vlasov equation in terms of a generalized dielectric function. In the first section we discuss some general properties of the microscopic distribution functions and autocorrelation functions. In the next section we solve the approximate equation of motion. We are then able to obtain approximate expressions for correlation functions in terms of the initial one-particle distribution function and the initial two-particle distribution function propagated along the unperturbed orbits. The effect of the medium on a test particle is again (as in the homogeneous case) to modify the field due to the test particle by the generalized dielectric function. The test-particle results extend as expected,[16,23] and we obtain an expression for the phase-space density of the polarization cloud. In particular, we can consider the system to be a collection of noninteracting quasiparticles. The quasiparticle consists of a particle traveling an orbit determined by the mean fields plus the associated polarization cloud induced by the particle. Finally, we are able to obtain a generalized kinetic equation which contains the effects of the dielectric function but not those of the three-particle correlation function and close collisions. We confine ourselves throughout to the one-component system.

2. MICROSCOPIC DISTRIBUTION FUNCTIONS

The microscopic or "exact" one-particle distribution function has been discussed by many authors recently.[15–17] It is given by

$$\hat{f}(\mathbf{r}, \mathbf{p}, t) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i(t))\delta(\mathbf{p} - \mathbf{p}_i(t)), \qquad (1)$$

where $\mathbf{r}_i(t)$ and $\mathbf{p}_i(t)$ are the position and momentum of the ith particle at time t. It is easy to show that this distribution function obeys the microscopic Vlasov equation

$$\frac{\partial \hat{f}(x, t)}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial \hat{f}}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial \hat{f}}{\partial p} = 0, \qquad (2)$$

where $x = (\mathbf{r}, \mathbf{p})$ denotes a point in phase space, $\mathbf{F}$ denotes the force which includes all external forces and the "exact" fluctuating force arising from interparticle interactions. For example, if the particles interact through a two-body potential such as the Coulomb or gravitational potentials, then this contribution to the force is given by

$$\mathbf{F}(\mathbf{r}, t) = -\frac{\partial}{\partial \mathbf{r}} \int dx' \Phi(\mathbf{r} - \mathbf{r}') \hat{f}(\mathbf{r}', \mathbf{p}', t). \qquad (3)$$

[18] Iu. L. Klimontovich. Zh. Eksp. Teor. Fiz., 33, 982 (1957) [Sov. Phys.—JETP 6, 753 (1958)].
[19] R. L. Guernsey, Phys. Fluids 5, 322 (1962); *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach, Science Publ., Inc., New York, 1967), Vol. IX.
[20] M. K. Sundaresan and T. Y. Wu, Can. J. Phys. 42, 794 (1964).
[21] R. Balescu and A. Kuszell, J. Math. Phys. 5, 1140 (1964).
[22] H. W. Wyld and D. Pines, Phys. Rev. 127, 1851 (1962).
[23] M. E. Rensink, Phys. Rev. 164, 175 (1967).

It is also easy to show that $\hat{f}$ obeys another equation of motion which involves the N-particle Liouville operator $\mathfrak{L}_N$. This equation is

$$\frac{\partial \hat{f}(x, t)}{\partial t} - i\mathfrak{L}_N \hat{f}(x, t) = 0. \tag{4}$$

In the above equation we are considering $\hat{f}$ to be a function of a point Γ in the $6N$-dimensional Γ space. The operator $\mathfrak{L}_N$ is defined by

$$i\mathfrak{L}_N = \sum_{i=1}^{N} \frac{\mathbf{p}_i}{m} \cdot \frac{\partial}{\partial \mathbf{r}_i} + \sum_{i=1}^{N} \mathbf{F}(x_i) \cdot \frac{\partial}{\partial \mathbf{p}_i}, \tag{5}$$

where $\mathbf{F}(x_i)$ is the force experienced by the ith particle. The Liouville equation for the N-particle distribution function $f_N(\Gamma, t)$ is given by

$$\frac{\partial f_N(\Gamma, t)}{\partial t} + i\mathfrak{L}_N f_N(\Gamma, t) = 0. \tag{6}$$

We can solve Eqs. (4) and (6) formally as

$$\hat{f}(x, t) = \hat{f}(x, \Gamma(t, \Gamma_0)) = e^{i\mathfrak{L}_N t}\hat{f}(x, \Gamma_0) \tag{7}$$

and

$$f_N(\Gamma_0, t) = e^{-i\mathfrak{L}_N t} f_N(\Gamma_0, 0) = f_N(\Gamma(-t, \Gamma_0), 0), \tag{8}$$

where

$$\Gamma(t, \Gamma_0) = e^{i\mathfrak{L}_N t}\Gamma_0 \tag{9}$$

defines the time-dependent, $6N$-dimensional phase point $\Gamma(t, \Gamma_0)$, whose initial value is Γ_0.

The average value of the microscopic distribution function is given by

$$\begin{aligned}\langle \hat{f}(\mathbf{r}, \mathbf{p}, t)\rangle_0 &= \int d\Gamma_0 \hat{f}(\mathbf{r}, \mathbf{p}; \Gamma(t, \Gamma_0)) f_N(\Gamma_0, 0) \\ &= \int d\Gamma_0 (e^{i\mathfrak{L}_N t}\hat{f}(\mathbf{r}, \mathbf{p}; \Gamma_0)) f_N(\Gamma_0, 0),\end{aligned} \tag{10}$$

where $d\Gamma_0$ is a volume element in the $6N$-dimensional phase space. The above equation has a close similarity to the "Heisenberg picture" in quantum mechanics because all of the time dependence is in the "operator" $\hat{f}$ and the average is taken with respect to the initial distribution function. The anti-Hermitian properties of the Liouville operator also allow us to write Eq. (10) as

$$\begin{aligned}\langle \hat{f}(\mathbf{r}, \mathbf{p}, t)\rangle_0 &= \int d\Gamma_0 \hat{f}(\mathbf{r}, \mathbf{p}; \Gamma_0) e^{-i\mathfrak{L}_N t} f_N(\Gamma_0, 0) \\ &= \langle \hat{f}(\mathbf{r}, \mathbf{p}, 0)\rangle_t.\end{aligned} \tag{11}$$

The above equation represents the "Schrödinger picture" where all of the time dependence is in the distribution function. From Eq. (11) it is easy to show that

$$\langle f(\mathbf{r}, \mathbf{p}, t)\rangle = n f_1(\mathbf{r}, \mathbf{p}, t), \tag{12}$$

where n is the average particle density and f_1 is the usual one-particle reduced distribution function. In a similar manner it is easy to show that

$$\begin{aligned}\langle \hat{f}(\mathbf{r}, \mathbf{p}, t)\hat{f}(\mathbf{r}, \mathbf{p}, t)\rangle_0 &= n^2 f_2(\mathbf{r}, \mathbf{p}; \mathbf{r}', \mathbf{p}', t) \\ &\quad + n\delta(\mathbf{r} - \mathbf{r}')\delta(\mathbf{p} - \mathbf{p}') f_1(\mathbf{r}, \mathbf{p}, t).\end{aligned} \tag{13}$$

One of the interesting aspects of the "Heisenberg picture" is that it is a very convenient way to treat multitime correlation functions. In particular, any two-time correlation function involving one-particle observables[16] (e.g., the particle density) can be written in terms of the quantity

$$\begin{aligned}&\langle \hat{f}(x, t)\hat{f}(x', t')\rangle_0 \\ &= \int d\Gamma_0 (e^{i\mathfrak{L}_N t}\hat{f}(x; \Gamma_0))(e^{i\mathfrak{L}_N t'}\hat{f}(x'; \Gamma_0)) f_N(\Gamma_0, 0)\end{aligned} \tag{14}$$

$$\begin{aligned}&= \int d\Gamma_0' \int d\Gamma_0 \hat{f}(x; \Gamma_0)\hat{f}(x'; \Gamma_0') \\ &\quad \times e^{-i\mathfrak{L}_N t} e^{-i\mathfrak{L}_N' t'} f_N(\Gamma_0, 0)\delta(\Gamma_0 - \Gamma_0').\end{aligned} \tag{15}$$

With the use of Eqs. (8) and (9) we obtain

$$\begin{aligned}&\langle \hat{f}(x, t)\hat{f}(x', t')\rangle_0 \\ &= \int d\Gamma_0 \int d\Gamma_0' \hat{f}(x; \Gamma_0)\hat{f}(x', \Gamma_0') D_2(\Gamma_0, t; \Gamma_0', t'),\end{aligned} \tag{16}$$

where D_2 is the joint probability density that the system will be at the point Γ_0 at time t and at the point Γ_0'' at t' and is given by

$$D_2(\Gamma_0, t; \Gamma_0', t') = f_N(\Gamma_0, t)\delta(\Gamma_0 - \Gamma(t - t', \Gamma_0')). \tag{17}$$

If we substitute the expression

$$\hat{f}(\mathbf{r}, \mathbf{p}; \Gamma_0) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i)\delta(\mathbf{p} - \mathbf{p}_i) \tag{18}$$

for $\hat{f}$ into Eq. (17), we obtain

$$\begin{aligned}\langle \hat{f}_1(x, t)\hat{f}_1(x', t')\rangle &= n^2 W_{12}(x, t; x', t') \\ &\quad + n W_{11}(x, t; x', t'),\end{aligned} \tag{19}$$

where W_{11} and W_{12} are the quantities introduced by Rostoker.[16,24]

We can also write the two-time correlation in another form. We have

$$\begin{aligned}&\langle \hat{f}(x, t)\hat{f}(x', t')\rangle_0 \\ &= \int d\Gamma_0 \hat{f}(x; \Gamma_0) e^{-i\mathfrak{L}_N t}\hat{f}(x'; \Gamma(t', \Gamma_0)) f_N(\Gamma_0, 0) \\ &= \int d\Gamma_0 \hat{f}(x; \Gamma_0) f_N'(x', t'; \Gamma_0, \tau),\end{aligned} \tag{20}$$

where $\tau = t - t'$ and

$$f_N'(x', t'; \Gamma_0, \tau) = e^{-i\mathfrak{L}_N \tau}\hat{f}(x', 0) f_N(\Gamma_0, t'). \tag{21}$$

[24] N. Rostoker, Phys. Fluids 7 479, 491 (1964).

We note that f'_N obeys the Liouville equation as a function of τ. The two-time correlation function is now seen to be simply the reduced one-particle distribution function corresponding to f'_N. That is,

$$\langle \hat{f}(x, t)\hat{f}(x', t')\rangle_0 = N\int d\Gamma_0 \delta(\mathbf{r} - \mathbf{r}_1)\delta(\mathbf{p} - \mathbf{p}_1) f'_N(\mathbf{r}', \mathbf{p}', t'; \Gamma_0, \tau)$$

$$= n f_1(\mathbf{r}', \mathbf{p}', t'; \mathbf{r}, \mathbf{p}, \tau), \tag{22}$$

$$f_1(\mathbf{r}', \mathbf{p}', t'; \mathbf{r}, \mathbf{p}, 0) = \langle \hat{f}(x, t)\hat{f}(x' \cdot t)\rangle_0, \tag{23}$$

which is given by Eq. (13). This result is equivalent to the theorem proved by Weinstock which relates the autocorrelation function to a one-particle distribution function.[25]

3. LINEARIZED MICROSCOPIC VLASOV EQUATION

It has been pointed out by several authors[15–17] that the linearized microscopic Vlasov equation is equivalent to the equation for the two-particle correlation function in which the three-particle correlation function and the effects of close collisions have been omitted. In the homogeneous case, the linearized microscopic Vlasov equation is easily solved and the resulting kinetic equation is the well-known Balescu–Lenard–Guernsey equation.[16] In this section, we consider the case in which inhomogeneities in space are allowed and the Bogoliubov adiabatic assumption is not used.

We write the microscopic distribution function as

$$\hat{f}(x, t) = n f_1(x, t) + \delta\hat{f}(x, t), \tag{24}$$

where the quantity $\delta\hat{f}$ represents the fluctuating phase-space density. From Eq. (2) we obtain for f_1 the equation

$$\frac{\partial f_1(x, t)}{\partial t} + iL_0 f_1(x, t) = -\frac{1}{n}\frac{\partial}{\partial \mathbf{p}} \cdot \langle \delta\mathbf{F}(x, t)\delta\hat{f}(x, t)\rangle, \tag{25}$$

where $\delta\mathbf{F}$ represents the fluctuating force arising from δf, i.e.,

$$\delta\mathbf{F}(x, t) = \mathbf{F}(x, t) - \langle \mathbf{F}(x, t)\rangle, \tag{26}$$

and we assumed that it commutes with the divergence operator in momentum space. The one-particle operator iL_0 is given by

$$iL_0 = \frac{\mathbf{p}}{m} \cdot \frac{\partial}{\partial \mathbf{r}} + \langle \mathbf{F}\rangle \cdot \frac{\partial}{\partial \mathbf{p}}. \tag{27}$$

[25] J. Weinstock, Phys. Rev. 139, A388 (1965).

We can define the time-dependent operators[26]

$$\mathbf{r}_0(t) = U_0(t, 0)\mathbf{r}, \quad \mathbf{p}_0(t) = U_0(t, 0)\mathbf{p}, \tag{28}$$

where

$$\frac{\partial U_0(t, t')}{\partial t} - iL_0(t)U_0(t, t') = 0,$$

which describe a particle moving in the average force field $\langle \mathbf{F}\rangle$ which includes all of the external forces and any forces which arise from inhomogeneities in the system.

The equation of motion for the fluctuating phase-space density is

$$\left(\frac{\partial}{\partial t} + iL_0\right)\delta\hat{f}(x, t) + n\delta\mathbf{F}(x, t) \cdot \frac{\partial}{\partial \mathbf{p}} f_1(x, t)$$

$$= -\frac{\partial}{\partial \mathbf{p}} \cdot [\delta\mathbf{F}(x, t)\delta\hat{f}(x, t) - \langle \delta\mathbf{F}(x, t)\delta\hat{f}(x, t)\rangle]. \tag{29}$$

The linearized microscopic Vlasov equation is obtained by neglecting the right-hand side of Eq. (29). Clearly, this assumes that the fluctuations are small and not very "stiff," such as those arising from soft collisions as opposed to hard collisions. This assumption is equivalent to neglecting the three-particle correlation function and the effects of close collisions.[7,16]

We can solve the resulting equation in terms of the initial value of $\delta\hat{f}$ as[27]

$$\delta\hat{f}(x, t) = U_0(-t, 0)\delta\hat{f}(x, 0) - n\int_0^t dt' U_0(t', t)\delta\mathbf{F}(x, t') \cdot \frac{\partial}{\partial \mathbf{p}} f_1(x, t'). \tag{30}$$

The way in which the fluctuating force is related to $\delta\hat{f}$ depends on the system under consideration. We now restrict ourselves to a one-component system in which the particles interact via a two-body potential $\Phi(\mathbf{r})$. We have in mind either the Coulomb or gravitational potentials. We can then write

$$\delta\mathbf{F}(x, t) = -\frac{\partial}{\partial \mathbf{r}}\int dx' \Phi(\mathbf{r} - \mathbf{r}')\delta\hat{f}(x', t). \tag{31}$$

Equation (30) can then be rewritten as

$$\delta\hat{f}(x, t) = U_0(-t, 0)\delta\hat{f}(x, 0) + n\int_0^t dt' U_0(t', t)\frac{\partial}{\partial \mathbf{r}} \times \int dx' \Phi(\mathbf{r} - \mathbf{r}') \cdot \frac{\partial}{\partial \mathbf{p}} f_1(x, t')\delta\hat{f}(x', t'). \tag{32}$$

If we then integrate both sides of the above equation

[26] The operator $U_0(t, 0)$ is the time-translation operator for the unperturbed orbits. Its initial value is the unit operator and it has the semigroup property $U_0(t', t'')U_0(t'', t) = U_0(t', t)$.

[27] A somewhat related approach was discussed at the IXth annual meeting of the Division of Plasma Phys., A.P.S. (Nov., 1967), by J. Price.

over the momenta and define the fluctuating particle density

$$\delta\hat{\rho}(\mathbf{r}, t) = \int d\mathbf{p}\delta\hat{f}(\mathbf{r}, \mathbf{p}, t), \tag{33}$$

we obtain[28]

$$\int d\mathbf{r}' \int_0^t dt' \epsilon(\mathbf{r}, \mathbf{r}'; t, t')\delta\rho(\mathbf{r}', t') = \int d\mathbf{p} U_0(-t, 0)\delta\hat{f}(x, 0), \tag{34}$$

where the generalized dielectric function is given by

$$\epsilon(\mathbf{r}, \mathbf{r}'; t, t') = \delta(t - t')\delta(\mathbf{r} - \mathbf{r}') - n\int d\mathbf{p} U_0(t', t) \times \frac{\partial\Phi(\mathbf{r} - \mathbf{r}')}{\partial\mathbf{r}} \cdot \frac{\partial f_1(\mathbf{r}, \mathbf{p}, t')}{\partial\mathbf{p}}. \tag{35}$$

It is easy to see that if f_1 is independent of space and time, the usual dielectric function is obtained by a Laplace–Fourier transform. It is not surprising that a dielectric function of this form should arise. The fact that the system is inhomogeneous in space and time means that the Fourier transform of the dielectric function is not just a function of k and ω, but is a function of two wavevectors and two frequencies. Such generalized dielectric functions have been discussed elsewhere in recent years.[29]

In order to obtain an expression for the fluctuating particle density, we need to introduce the inverse of the dielectric function which is defined by the integral equation

$$\int d\mathbf{r} \int_0^{t''} dt \epsilon^{-1}(\mathbf{r}'', \mathbf{r}; t'', t)\epsilon(\mathbf{r}, \mathbf{r}'; t, t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t'' - t'). \tag{36}$$

With the use of Eq. (35) we obtain

$$\epsilon^{-1}(\mathbf{r}'', \mathbf{r}'; t'', t') = \delta(\mathbf{r}'' - \mathbf{r}')\delta(t'' - t') + n\int dx \int_0^{t''} dt \epsilon^{-1}(\mathbf{r}, \mathbf{r}'; t, t')U_0(t', t) \times \frac{\partial\Phi(\mathbf{r} - \mathbf{r}')}{\partial\mathbf{r}} \cdot \frac{\partial f_1(\mathbf{r}, \mathbf{p}, t')}{\partial\mathbf{p}}. \tag{37}$$

If the one-particle distribution is independent of space and time, we obtain the usual expression for the dielectric constant.

The use of Eq. (36) in Eq. (34) yields

$$\delta\hat{\rho}(\mathbf{r}, t) = \int dx' \int_0^t dt' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'; t, t')U_0'(-t', 0)\delta\hat{f}(x', 0). \tag{38}$$

[28] The integral over time in Eq. (34) should be extended infinitesimally beyond t so as not to require a factor of 2 in the first term of Eq. (35).

[29] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

The force due to $\delta\hat{\rho}$ is given by

$$\delta\mathbf{F}(\mathbf{r}, t) = -\frac{\partial}{\partial\mathbf{r}} \int d\mathbf{r}' \Phi(\mathbf{r} - \mathbf{r}')\delta\hat{\rho}(\mathbf{r}', t). \tag{39}$$

We thus have the usual description in terms of a generalized dielectric function. If we replace $\epsilon^{-1}(\mathbf{r}, \mathbf{r}'; t, t')$ in Eq. (38) by $\delta(\mathbf{r} - \mathbf{r}')\delta(t - t')$, then the corresponding fluctuating force obtained from Eq. (39) is that due to noninteracting particles traveling along the unperturbed orbits given by Eq. (28). The effect of the medium is to modify the force by a dielectric function.

We can now write $\delta\hat{f}$ in terms of its initial value as

$$\delta\hat{f}(x, t) = U_0(-t, 0)\delta\hat{f}(x, 0) + n\int_0^t dt' U_0(t', t) \frac{\partial}{\partial\mathbf{r}} \times \int d\mathbf{r}' \Phi(\mathbf{r} - \mathbf{r}') \cdot \frac{\partial}{\partial\mathbf{p}} f_1(x, t') \int dx'' \times \int_0^t dt'' \epsilon^{-1}(\mathbf{r}', \mathbf{r}''; t' \; t'')U_0''(-t'', 0)\delta\hat{f}(x'', 0). \tag{40}$$

As we have mentioned previously, the effect of the operator $U_0(-t, 0)$ is to propagate backward in time along the unperturbed orbits. If we define

$$f^0(x, t) = U_0(-t, 0)f(x, 0), \tag{41}$$

where f is any function of $x = (\mathbf{r}, \mathbf{p})$, then we can write Eq. (40) as

$$\delta\hat{f}(x, t) = \delta\hat{f}^0(x, t) + \int_0^t dt' \int dx' P(x', t' \mid x, t)\delta\hat{f}^0(x', t'), \tag{42}$$

where

$$P(x't' \mid xt) = n\int_0^t dt'' \theta(t'', t')U_0(t'', t) \times \frac{\partial}{\partial\mathbf{r}} \int d\mathbf{r}'' \Phi(\mathbf{r} - \mathbf{r}'') \cdot \frac{\partial}{\partial\mathbf{p}} f_1(x, t'')\epsilon^{-1}(\mathbf{r}'', \mathbf{r}'; t'', t'), \tag{43}$$

where

$$\theta(x) = \begin{cases} 0, & x > 0 \\ 1, & x < 0 \end{cases}.$$

The quantity $P(x't' \mid xt)$ is the generalization of a similar quantity introduced by Rostoker[24] for the homogeneous case. It is the phase-space density at x of the shielding cloud induced by a test particle at x'.[16] It has been shown previously that the test-particle theory holds for inhomogeneous systems.[16,23] Equation (43) gives an explicit expression for $P(x't' \mid xt)$ in terms of the inverse dielectric function. The expression given by Eq. (42) can then be interpreted in terms of

the test-particle theory. It says that the system can be considered to be a superposition of N noninteracting quasiparticles which are composed of a bare test particle following the unperturbed orbit given by Eq. (28) plus the polarization cloud induced by the test particle.

We are now in the position to calculate various correlation and autocorrelation functions. We only need an expression for the quantity $\langle \delta\hat{f}^0(x,t)\delta\hat{f}^0(x',t')\rangle$. This is very easy to obtain. We find

$$\langle \delta\hat{f}^0(x,t)\delta\hat{f}^0(x',t')\rangle = n^2 g(x_0(-t), x_0'(-t'), 0) + n\delta(x_0(-t) - x_0'(-t'))f_1^0(x,t), \tag{44}$$

where $x_0(t)$ is the unperturbed orbit given by Eq. (28). Thus, from Eqs. (42) and (44), we can obtain an expression for the quantity $\langle \delta\hat{f}(x,t)\delta\hat{f}(x',t')\rangle$ and all related quantities in terms of the inverse dielectric function and the initial one- and two-particle distribution functions propagated along the unperturbed orbits.

In particular, we obtain for the density autocorrelation function the expression[30]

$$\begin{aligned}\langle \delta\hat{\rho}(\mathbf{r},t)\delta\hat{\rho}(\mathbf{r}',t')\rangle &= \int dx'' \int_0^t dt'' \int dx''' \int_0^t dt''' \epsilon^{-1}(\mathbf{r},\mathbf{r}'';t,t'') \\ &\times \epsilon^{-1}(\mathbf{r}',\mathbf{r}'';t',t''')[n^2 g_2(x_0''(-t''), x_0'''(-t'''), 0) \\ &+ n\delta(x_0''(-t'') - x_0'''(-t'''))f_1^0(x'',t'')]. \end{aligned} \tag{45}$$

In order to obtain a kinetic equation we need an expression for $\langle \delta\rho(x,t)\delta\hat{f}(x,t)\rangle$. This can be calculated by using Eqs. (38), (42), and (44). We then obtain a generalized kinetic equation when we substitute the result into Eq. (25). The generalized kinetic equation is

$$\begin{aligned}&\frac{\partial f_1(x,t)}{\partial t} + \frac{\mathbf{p}}{m}\cdot\frac{\partial f_1(x,t)}{\partial \mathbf{r}} + \langle \mathbf{F}(x,t)\rangle \cdot \frac{\partial f_1(x,t)}{\partial \mathbf{p}} \\ &= \frac{1}{n}\frac{\partial}{\partial \mathbf{p}}\cdot\frac{\partial}{\partial \mathbf{r}} \int d\mathbf{r}'\Phi(\mathbf{r}-\mathbf{r}') \int dx' \int_0^t dt' \epsilon^{-1}(\mathbf{r},\mathbf{r}';t,t') \\ &\times \Big\{[n\delta(x_0(-t) - x_0'(-t'))f_1^0(x,t) \\ &+ n^2 g_2(x_0(-t), x_0'(-t'), 0)] \\ &+ \int_0^t dt'' \int dx'' P(x''t'' \mid x,t) \\ &\times [n\delta(x_0''(-t'') - x_0'(-t'))f_1^0(x',t') \\ &+ n^2 g_2(x_0''(-t''), x_0'(-t'), 0)]\Big\}. \end{aligned} \tag{46}$$

The above equation is relatively complicated, but it does represent a solution of the problem when we ignore the three-particle correlation function and the close collision effects. These approximations are reasonable for the Coulomb and gravitational forces. To lowest order, we can replace $f_1^0(x,t)$ by the exact quantity $f_1(x,t)$ on the right-hand side, giving us a kinetic equation involving only f_1 and the initial correlation function. Equation (46) gives a general result to which approximation procedures, such as the adiabatic assumption, can be applied to obtain kinetic equations for particular systems. Of course, we also need to obtain approximate expressions for the inverse dielectric function. In reality, the kinetic equation consists of Eq. (46) and Eq. (37). This is really the case with the Balescu–Lenard–Guernsey equation as well, because the dielectric function there contains $f_1(t)$ as well. We note also that the problem of obtaining the kinetic equation is intimately connected with orbit theory because of the U_0 operator. This was to be expected.

Finally, we can obtain the generalization of the test-particle result for the correlation function for inhomogeneous systems. We find that

$$\begin{aligned} g_2(x,x',t) &= \int_0^t dt'' P(x_0'(t'-t''), t'' \mid x,t) f_1(x',t') \\ &+ \int_0^t dt'' P(x_0(t-t''), t'' \mid x',t') f_1(x,t) \\ &+ \int_0^t dt'' \int_0^t dt''' \int dx'' P(x'',t'' \mid x,t) \\ &\times P(x_0'''(t'''-t''), t''' \mid x',t') f_1(x'',t''). \end{aligned} \tag{47}$$

Equation (47) is subject to the usual interpretation. The correlation between two particles arises from three effects. The first two terms are the contribution due to one particle being in the polarization cloud of the particle. The last term is due to both particles being in the polarization cloud of a third particle.

ACKNOWLEDGMENT

The author would like to express his gratitude to Dr. C. S. Wu for helpful suggestions.

[30] A very similar expression for the quantum case was obtained by D. Dubois, *Lectures in Theoretical Physics*, W. E. Brittin, Ed. (Gordon and Breach, Science Publ., Inc., New York, 1967), Vol. IX.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Galilean Tensor Calculus

G. PINSKI
Physics Department, Drexel Institute of Technology, Philadelphia, Pennsylvania

(Received 7 August 1967)

Galilean transformations are expressed as transformations in a five-dimensional space, with a subsidiary condition, and a Galilean tensor calculus with a nonsingular metric is developed. It is shown that the homogeneous Galilei group is isomorphic to a subgroup of the pseudo-orthogonal group $O(4, 1)$, which leaves the difference of two components of a vector invariant. A set of scalar variables for a Galilean-invariant S matrix is selected. A Galilean-invariant phase space is defined and a recursion relation derived.

INTRODUCTION

Galilean invariance[1–5] is of interest to the physicist since it is the nonrelativistic limit of Lorentz invariance and is the invariance obeyed by classical mechanics. There has also been recent attention given to the construction of kinematically consistent nonrelativistic quantum theories.[6] However, the elegance of the manifestly covariant formulation of Lorentz invariance is lost in the usual treatment of Galilean invariance. In what follows it is shown that a manifestly covariant formulation of Galilean invariance follows in a natural way from the limit of Lorentz covariance. One can then systematically construct Galilean covariants and choose scalar variables for a Galilean invariant S matrix in an identical manner to the construction in the Lorentz-invariant case. A Galilean-invariant phase space can also be defined and a recursion relation similar to the relation for Lorentz-invariant phase space will be shown to hold.

1. GALILEAN COVARIANCE

In considering Galilean invariance it is convenient to work with Galilean quantities which are the nonrelativistic limits of the corresponding Lorentz quantities. Rather than regard the limiting process as one in which we let $c \to \infty$, we retain rest-energy terms as well as the next order term in v/c. Thus, the relativistic energy–momentum relation $E = [m^2c^4 + (\mathbf{p})^2c^2]^{\frac{1}{2}}$ becomes $E = mc^2 + (\mathbf{p})^2/2m$. Similarly, the general homogeneous Lorentz transformations (including spatial rotations R) reduce, in the Galilean limit, to

$$\begin{aligned} p' &= Rp + mV, \\ E' &= E + V^{\mathrm{T}}Rp + \tfrac{1}{2}mV^2, \\ x' &= Rx + Vt, \\ t' &= t. \end{aligned} \tag{1}$$

A homogeneous Lorentz transformation can be represented as a linear transformation on a Lorentz four-vector. On the other hand, a Galilean transformation of energy and momentum cannot be represented as a linear transformation acting on an energy–momentum four-vector because of the explicit appearance of the mass in the transformation. In order to have a linear transformation, we must introduce a fifth component to our Galilean vector. With a five-vector[7] $p^\alpha = \begin{pmatrix} \mathbf{p} \\ E \\ m \end{pmatrix}$, a homogeneous Galilean transformation can be represented by a matrix of the form

$$G = \left(\begin{array}{c|cc} \mathsf{R} & \mathbf{0} & \mathbf{V} \\ \hline \mathbf{V}^{\mathrm{T}}\mathsf{R} & 1 & V^2/2 \\ \mathbf{0} & 0 & 1 \end{array}\right). \tag{2}$$

The three-by-three "space part" of each five-by-five matrix will be abbreviated in the upper left corner.

A space–time five-vector has $\mathbf{x}$ for its first three components and t for its fifth component. Its fourth component must be related to $\mathbf{x}$ and t in the same way that E is related to $\mathbf{p}$ and m, i.e., $x^4 = t + (\mathbf{x})^2/2t$.

The metric in this five-dimensional space can be found by taking the Galilean limit of the Lorentz scalar product. The Lorentz scalar product $E_1E_2 - \mathbf{p}_1 \cdot \mathbf{p}_2$ becomes, in the Galilean limit, $m_1E_2 + m_2E_1 - m_1m_2 - \mathbf{p}_1 \cdot \mathbf{p}_2$. Our nonsingular[8] Galilean metric is therefore determined to be

$$(g)_{\alpha\beta} = \left(\begin{array}{c|cc} -1 & 0 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 1 & -1 \end{array}\right), \tag{3}$$

The group of transformations G is just the group which leaves the metric g invariant under a congruent

[1] V. Bargmann, Ann. Math. **40**, 149 (1939); **59**, 1 (1954).
[2] E. Inonu and E. P. Wigner, Nuovo Cimento **9**, 705 (1952).
[3] M. Hamermesh, Ann. Phys. **9**, 518 (1960).
[4] J. M. Levy-Leblond, J. Math. Phys. **4**, 776 (1963).
[5] J. Voisin, J. Math. Phys. **6**, 1519 (1965).
[6] J. M. Levy-Leblond, Commun. Math. Phys. **4**, 157 (1967).

[7] From now on factors of c will be understood and not written explicitly.
[8] P. Havas, Rev. Mod. Phys. **36**, 938 (1964). A four-dimensional formulation of Galilean invariance using a singular metric is developed here.

TABLE I. Contravariant and covariant Galilean five-vectors.

Space–time differentials	Velocity	Acceleration	Momentum	Force	Gradient
dx^α	v^α	a^α	p^α	F^α	∇^α
$d\mathbf{x}$	$\mathbf{v}$	$\mathbf{a}$	$\mathbf{p}$	$\mathbf{F}$	$-\nabla$
$dt(1+\mathbf{v}^2/2)$	$1+\mathbf{v}^2/2$	$\mathbf{a}\cdot\mathbf{v}$	E	$\mathbf{F}\cdot\mathbf{v}$	$\partial/\partial t$
dt	1	0	m	0	0
dx_α	v_α	a_α	p_α	F_α	∇_α
$-d\mathbf{x}$	$-\mathbf{v}$	$-\mathbf{a}$	$-\mathbf{p}$	$-\mathbf{F}$	∇
dt	1	0	m	0	0
$dt(\mathbf{v}^2/2)$	$\mathbf{v}^2/2$	$\mathbf{a}\cdot\mathbf{v}$	$E-m$	$\mathbf{F}\cdot\mathbf{v}$	$\partial/\partial t$

transformation $G^{\mathrm{T}}gG = g$, with the subsidiary condition that it leave the "submetric"

$$(g_0)_{\alpha\beta} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 1 \end{pmatrix} \tag{4}$$

invariant, $G^{\mathrm{T}}g_0G = g_0$; i.e., it must leave the fifth component of our five-vector invariant.

We may now lower indices on contravariant vectors to form covariant vectors. Some Galilean five-vectors of interest are given in Table I. We observe that for a covariant vector it is the fourth component which is invariant. The following relations may be noted:

$$p^\alpha p_\alpha = m^2, \tag{5a}$$

$$v^\alpha v_\alpha = 1, \tag{5b}$$

$$v^\alpha a_\alpha = 0 = v_\alpha a^\alpha, \tag{5c}$$

$$F^\alpha = d(p^\alpha)/dt = ma^\alpha, \tag{5d}$$

$$d\tau^2 \equiv g_{\alpha\beta}\,dx^\alpha\,dx^\beta = dx^\alpha\,dx_\alpha = dt^2, \tag{5e}$$

$$\nabla^\alpha\nabla_\alpha = -\nabla^2. \tag{5f}$$

There are two n-particle invariants which should be distinguished. The invariant center-of-mass energy or "effective mass" of n particles, which in the Lorentz case is $[(\sum E_i)^2 - (\sum \mathbf{p}_i)^2]^{\frac{1}{2}}$, becomes in the Galilean limit

$$U = \sum E_i - (\sum \mathbf{p}_i)^2/2\sum m_i. \tag{6}$$

The invariant "effective-mass squared" $(\sum p_i)_\alpha(\sum p_j)^\alpha$ becomes

$$M^2 = \sum_{i<j}[2(m_iE_j + m_jE_i - \mathbf{p}_i\cdot\mathbf{p}_j) + (m_i - m_j)^2], \tag{7}$$

and is no longer the square of the center-of-mass energy. These two invariants are not independent, but are related by

$$M^2 = 2(\sum m_i)U - (\sum m_i)^2. \tag{8}$$

We note that $U \geq M$, with the equality holding only when all the particles are at rest.

The Levi-Civita symbol $\epsilon_{\alpha\beta\gamma\eta\delta}$ is a fifth-rank covariant isotropic tensor under transformations for which $\det G = 1$, and may be used to construct Galilean covariants. In particular, the Galilean analog of any Lorentz covariant constructed with the fourth-rank symbol $\epsilon_{\mu\nu\rho\sigma}$ may be found immediately. For example, the Lorentz quadrilinear invariant

$$\epsilon_{\mu\nu\rho\sigma}p_1^\mu p_2^\nu p_3^\rho p_4^\sigma = \begin{vmatrix} p_{1x} & p_{2x} & p_{3x} & p_{4x} \\ p_{1y} & p_{2y} & p_{3y} & p_{4y} \\ p_{1z} & p_{2z} & p_{3z} & p_{4z} \\ E_1 & E_2 & E_3 & E_4 \end{vmatrix} \tag{9}$$

has as its Galilean analog

$$\epsilon_{\alpha\beta\gamma4\delta}p_1^\alpha p_2^\beta p_3^\gamma p_4^\delta = \begin{vmatrix} p_{1x} & p_{2x} & p_{3x} & p_{4x} \\ p_{1y} & p_{2y} & p_{3y} & p_{4y} \\ p_{1z} & p_{2z} & p_{3z} & p_{4z} \\ m_1 & m_2 & m_3 & m_4 \end{vmatrix}, \tag{10}$$

recalling that the fourth component of a covariant vector is an invariant.

The square of the spin for an n-particle system is given, for the Lorentz case, by

$$L^2 = -w_\mu w^\mu, \tag{11}$$

where

$$w_\mu = [2(P_\tau P^\tau)^{\frac{1}{2}}]^{-1}\epsilon_{\mu\nu\rho\sigma}P^\nu M^{\rho\sigma}, \tag{12}$$

$$P^\nu = \sum p_i^\nu, \quad M^{\rho\sigma} = \sum (x_i^\rho p_i^\sigma - x_i^\sigma p_i^\rho). \tag{13}$$

In the Galilean case we have

$$L^2 = -w_\alpha w^\alpha, \tag{14}$$

where

$$w_\alpha = [2(P_\tau P^\tau)^{\frac{1}{2}}]^{-1}\epsilon_{\alpha\beta\gamma4\delta}P^\beta M^{\gamma\delta}. \tag{15}$$

P^β and $M^{\gamma\delta}$ are again given by (13). The fourth components of the momentum and position vectors do not appear in w_α.

There are, in addition, Galilean tensors which do not correspond to any Lorentz tensor, namely those for which an index of the Levi-Civita symbol is not held fixed. The quintilinear Galilean invariant

$$Q = \epsilon_{\alpha\beta\gamma\eta\delta}p_1^\alpha p_2^\beta p_3^\gamma p_4^\eta p_5^\delta \tag{16}$$

has no Lorentz analog. Apart from constant factors,

this invariant can be written as

$$\begin{vmatrix} v_{1x} & v_{2x} & v_{3x} & v_{4x} & v_{5x} \\ v_{1y} & v_{2y} & v_{3y} & v_{4y} & v_{5y} \\ v_{1z} & v_{2z} & v_{3z} & v_{4z} & v_{5z} \\ v_1^2 & v_2^2 & v_3^2 & v_4^2 & v_5^2 \\ 1 & 1 & 1 & 1 & 1 \end{vmatrix}.$$

We now find an equivalent group $\bar{G}$ which leaves the pseudo-orthogonal $O(4, 1)$ metric

$$\bar{g} = \begin{pmatrix} -1 & & \\ & 1 & \\ & & -1 \end{pmatrix} \tag{17}$$

invariant. If we find a congruent transformation C such that $C^{\mathrm{T}}gC = \bar{g}$, then the equivalent group $\bar{G} = C^{-1}GC$ will leave the metric $\bar{g}$ invariant: $\bar{G}^{\mathrm{T}}\bar{g}\bar{G} = \bar{g}$. It will also leave the metric $\bar{g}_0$ invariant: $\bar{G}^{\mathrm{T}}\bar{g}_0\bar{G} = \bar{g}_0$, where $\bar{g}_0 = C^{\mathrm{T}}g_0C$. Under such a transformation, a Galilean vector p^α becomes $\bar{p} = C^{-1}p$. One choice[9] of C is

$$C = \begin{pmatrix} 1 & & \\ & 1 & 0 \\ & 1 & -1 \end{pmatrix}. \tag{18}$$

The matrices of the group $\bar{G}$ then have the form

$$\bar{G} = \begin{pmatrix} \mathsf{R} & \mathbf{V} & -\mathbf{V} \\ \mathbf{V}^{\mathrm{T}}\mathsf{R} & 1 + V^2/2 & -V^2/2 \\ \mathbf{V}^{\mathrm{T}}\mathsf{R} & V^2/2 & 1 - V^2/2 \end{pmatrix} \tag{19}$$

The "submetric" $\bar{g}_0$ is

$$\bar{g}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}. \tag{20}$$

The new Galilean vector is $\bar{p}^\alpha = \begin{pmatrix} \mathbf{p} \\ E \\ E - m \end{pmatrix}$, a momentum, total-energy, kinetic-energy five-vector. It is now evident that the homogeneous Galilei group is isomorphic to a subgroup of the pseudo-orthogonal group $O(4, 1)$ which leaves the difference between the fourth and fifth component of a vector invariant.

[9] More generally, the matrix C may be chosen to be

$$C = \begin{pmatrix} 1 & & \\ & k\left(\frac{1}{2} + \frac{1}{2k^2}\right) & \pm k\left(\frac{1}{2} - \frac{1}{2k^2}\right) \\ & k & \pm k \end{pmatrix},$$

k arbitrary.

2. SCALAR VARIABLE PROBLEM

The problem of selecting a set of scalar variables for a Lorentz-invariant S matrix has been discussed by Rohrlich.[10] He has shown the following:

For an n-particle reaction there are $3n - 10$ independent bilinear scalars needed to reconstruct the momentum configuration with respect to three momenta, say p_1, p_2, p_3, chosen as base vectors. The $3n - 9$ scalar variables

$$\begin{aligned} &p_1 \cdot p_k, \quad k = 2, \cdots, n-1, \\ &p_2 \cdot p_k, \quad k = 3, \cdots, n-1, \\ &p_3 \cdot p_k, \quad k = 4, \cdots, n-1, \end{aligned}$$

together with the $n - 1$ masses $m_k, k = 1, \cdots, n - 1$, determine the $n - 1$ four-vectors up to a choice of two vectors, for $k = 4, \cdots, n - 1$. The nth vector is determined by momentum conservation. If the mass of the nth particle is specified, then p_n is in fact overdetermined, placing a single constraint on the $3n - 9$ variables. Only $3n - 10$ are therefore independent. The dichotomic invariant sign $\epsilon_{\mu\nu\rho\sigma}p_k^\mu p_1^\nu p_2^\rho p_3^\sigma$ serves to distinguish between the two choices of p_k for each $k = 4, \cdots, n - 1$. Therefore, $n - 4$ of these dichotomic quadrilinear invariants in addition to the aforementioned bilinear invariants specify the configuration uniquely.

There is an entirely analogous situation for the case of Galilean scalar variables. The appropriate bilinear scalars are Galilean scalar products of the form $p_{(a)\alpha}p^\alpha_{(b)}$. Suppose that we are given three base vectors $p_i \equiv (\mathbf{p}_i, E_i, m_i)$, $i = 1, 2, 3$, and that for an arbitrary p_k, $k = 4, \cdots, n - 1$, the scalar products

$$p_i \cdot p_k = m_iE_k + m_kE_i - m_im_k - \mathbf{p}_i \cdot \mathbf{p}_k, \quad i = 1, 2, 3, \tag{21}$$

are specified. We may then solve these three equations simultaneously for $\mathbf{p}_k$ in terms of E_k in the form $\mathbf{p}_k = \boldsymbol{\alpha}_k + \boldsymbol{\beta}_kE_k$. Since $E_k - (\mathbf{p}_k)^2/2m_k = m_k$, we have a quadratic equation for each E_k; if one solution is real, the other must also be, leaving us with two physical vectors as in the Lorentz case.

The quadrilinear Galilean invariant

$$\text{sign } \epsilon_{\alpha\beta\gamma\delta}p_k^\alpha p_1^\beta p_2^\gamma p_3^\delta$$

is now the dichotomic parameter which distinguishes between the two vectors. The specification of $n - 4$ of these quadrilinear invariants, in addition to the $3n - 10$ independent bilinear invariants, therefore determines the n-particle configuration.

[10] F. Rohrlich, Nuovo Cimento **38**, 673 (1965).

3. GALILEAN-INVARIANT PHASE SPACE

Under Galilean transformations, both d^3p and $d^4p \equiv d^3p\,dE$ are invariants. Although a phase space defined by

$$\int \prod_{i=1}^{n} [d^3p_i]\delta^3\Big(\mathbf{P} - \sum_{i=1}^{n} \mathbf{p}_i\Big)$$

is Galilean-invariant, we may choose not to work with it since it does not have the same dimensions as, and is not the nonrelativistic limit of, the Lorentz invariant phase space. However a Galilean-invariant phase space defined by

$$F_n = \int \prod_{1}^{n} [d^4p_i\delta(p_i^2 - m_i^2)] \times \delta^3\Big(\mathbf{P} - \sum_{1}^{n} \mathbf{p}_i\Big)\delta\Big(E - \sum_{1}^{n} E_i\Big) \quad (22)$$

does have these properties. If we note that

$$d^4p_i\delta(p_i^2 - m_i^2) = d^3p_i\,dE_i\delta[(2m_iE_i - m_i^2 - \mathbf{p}_i^2) - m_i^2] = \frac{d^3p_i\,dE_i}{2m_i}\delta\Big[E_i - \Big(m_i + \frac{\mathbf{p}_i^2}{2m_i}\Big)\Big] \quad (23)$$

and introduce the notation

$$u_i = E_i - \mathbf{p}_i^2/2m_i, \quad U_k = \sum_{1}^{k} E_i - \Big(\sum_{1}^{k} \mathbf{p}_i\Big)^2\Big/2\sum_{1}^{k} m_i,$$

$$\delta^4\Big(P - \sum_{1}^{n} p_i\Big) = \delta^3\Big(\mathbf{P} - \sum_{1}^{n} \mathbf{p}_i\Big)\delta\Big(E - \sum_{1}^{n} E_i\Big), \quad (24)$$

we can write

$$F_n(U_n; m_1, \cdots, m_n) = \frac{1}{\prod_{1}^{n}(2m_i)} \int \prod_{1}^{n} [d^4p_i\delta(u_i - m_i)]\delta^4\Big(P - \sum_{1}^{n} p_i\Big). \quad (25)$$

A recursion relation for the quantity

$$R_n(U_n; m_1, \cdots, m_n) \equiv \int \prod_{1}^{n} [d^4p_i\delta(u_i - m_i)]\delta^4\Big(P - \sum_{1}^{n} p_i\Big), \quad (26)$$

similar to the recursion relation[11,12] for Lorentz-invariant phase space, may be easily derived. Inserting into the integral (26) the factor

$$1 = \int d^4r\delta^4\Big(r - \sum_{1}^{n-1} p_i\Big)\int dU\delta(U_{n-1} - U),$$

we have

$$R_n(U_n; m_1, \cdots, m_n) = \int \Big\{\int \prod_{1}^{n-1} [d^4p_i\delta(u_i - m_i)]\delta^4\Big(r - \sum_{1}^{n-1} p_i\Big) \times \int d^4p_n\delta(u_n - m_n)\,d^4r \times \delta(U_{n-1} - U)\delta^4(P - r - p_n)\Big\}\,dU = \int_{\sum_{1}^{n-1} m_i}^{U_n - m_n} R_{n-1}(U; m_1, \cdots, m_n)R_2^{(1)}(U_n; U, m_n)\,dU. \quad (27)$$

Here

$$R_2^{(1)}(U_n; U, m_n) \equiv \int d^4p_n\delta(u_n - m_n)d^4r\delta(U_{n-1} - U)\delta^4(P - r - p_n) \quad (28)$$

is readily evaluated to be

$$R_2^{(1)}(U_n; U, m_n) = 2\pi(U_n - U - m_n)^{\frac{1}{2}}\Big(1/2m_n + 1/2\sum_{1}^{n-1} m_i\Big)^{-\frac{3}{2}}.$$

Also

$$R_2(U; m_1, m_2) = 2\pi(U - m_1 - m_2)^{\frac{1}{2}}(1/2m_1 + 1/2m_2)^{-\frac{3}{2}}. \quad (29)$$

It should be noted that, unlike the situation for the Lorentz recursion relation, the function $R_2^{(1)}(U_n; U, m_n)$ is an explicit function of the sum of the masses of those particles which have effective mass U, and must be distinguished from the function $R_2(U; m_1, m_2)$.

The two-particle Lorentz and Galilean phase space may now be compared:

$$F_2^{\mathrm{Lor}}(U; m_1, m_2) = \pi\,|\mathbf{p}_1^{\mathrm{Lor}}|/U,$$
$$F_2^{\mathrm{Gal}}(U; m_1, m_2) = \pi\,|\mathbf{p}_1^{\mathrm{Gal}}|/(m_1 + m_2), \quad (30)$$

where the center-of-mass momentum of a particle is

$$|\mathbf{p}_1^{\mathrm{Lor}}| = \frac{U}{2}\Big\{\Big[1 - \Big(\frac{m_1 + m_2}{U}\Big)^2\Big]\Big[1 - \Big(\frac{m_1 - m_2}{U}\Big)^2\Big]\Big\}^{\frac{1}{2}},$$
$$|\mathbf{p}_1^{\mathrm{Gal}}| = \Big[\frac{U - m_1 - m_2}{(1/2m_1 + 1/2m_2)}\Big]^{\frac{1}{2}}. \quad (31)$$

Unlike the two-particle Lorentz phase space which approaches a limiting value with increasing energy, the two-particle Galilean phase space is unbounded.

Application of these relations to Galilean statistical theories will be treated in a separate publication.

ACKNOWLEDGMENTS

The author is indebted to Professor E. C. G. Sudarshan for suggesting an examination of Galilean invariance and for his continued interest in this work. The author also wishes to thank Professor G. Rosen for a valuable suggestion.

[11] P. P. Srivastava and E. C. G. Sudarshan, Phys. Rev. **110**, 765 (1958).
[12] G. Pinski, Nuovo Cimento **24**, 719 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Approach to the Three-Body Scattering Problem

J. R. JASPERSE

Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, Massachusetts

(Received 17 October 1967)

We derive some new equations for the wavefunction for the three-body scattering problem using the "partial-wavefunction approach." All scattering processes are solutions of inhomogeneous coupled integral equations for the scattered parts of the total wavefunction. We show how the Pauli principle is incorporated when some or all of the particles are identical. This leads to a reduction in the number of independent scattered partial wavefunctions when some of the particles are identical. The kernels in the integral equations contain no δ functions but each scattered partial wavefunction does have a self-coupling term. We apply the method to the scattering of three identical bosons interacting through zero-range pair potentials. The method gives the same equation for this problem as obtained by the Faddeev T-matrix approach.

I. INTRODUCTION

The advantages of writing the wavefunction as a sum of three parts for a bound-state system of three identical particles has been shown by Eyges.[1] Faddeev has applied a similar idea to the scattering problem, and written the T matrix as the sum of three parts.[2] Using the "partial-wavefunction" approach, we have also shown how solutions for a bound-state system of three *arbitrary* particles can be constructed which are eigenstates of a complete set of commuting operators.[3] The method has been applied with success to the heliumlike atom,[3] and has led to a new technique for calculating bound-state energies for three-body systems.

The purpose of this paper is to show how the three-body scattering problem can be treated in a similar way by breaking up the scattered part of the total wavefunction into a sum of three parts which we call "scattered partial wavefunctions." General operator equations for the three-body scattering problem are derived and the matrix elements of these equations for several specific situations are written out in the momentum representation. In order to show the reader how these equations are applied we consider explicitly the case of three identical particles interacting through zero-range pair potentials and demonstrate that the method, in this case, gives the same results as the Faddeev T-matrix approach.

II. THE EQUATIONS FOR THREE-PARTICLE SCATTERING

The Schrödinger equation where the center-of-mass and time motion have been separated out is

$$H\Psi_t = E\Psi_t, \tag{1}$$

where Ψ_t is the Schrödinger state-vector and H is the Hamiltonian operator written as

$$H = H_0 + V_{ij} + V_{jk} + V_{ki}. \tag{2}$$

Consider a system of three *distinguishable* particles where particle k is incident on a *bound state* of particles i and j. We may write the total state-vector as the sum of an incident part, which is a known eigenstate of $H_0 + V_{ij}$, and a scattered part which is unknown,

$$\Psi_t = \Psi^{(i)} + \Psi.$$

Substituting this expression into Eq. (1), rearranging, and writing the result in terms of the resolvent operator $G_0(s) = (H_0 - sI)^{-1}$, we obtain

$$\Psi = -G_0(s)\{(V_{jk} + V_{ki})\Psi^{(i)} + (V_{ij} + V_{jk} + V_{ki})\Psi\},$$

where s is defined as the complex energy parameter. Let this definition for $G_0(s)$ and a consistent extension of the energy of the in-state to complex values serve to define the continuation of the energy to the complex plane in our equations. It can be shown that a solution of the following form exists[4]:

$$\Psi_t = \Psi^{(i)} + \psi_{ij} + \psi_{jk} + \psi_{ki},$$

where the three scattered partial state-vectors, ψ_{ij}'s, are given by a set of coupled operator equations. In matrix form we have:

$$\begin{pmatrix}\psi_{ij}\\ \psi_{jk}\\ \psi_{ki}\end{pmatrix} = -G_0(s)\begin{pmatrix}0\\ V_{jk}\\ V_{ki}\end{pmatrix}\Psi^{(i)} - G_0(s)\begin{pmatrix}V_{ij} & V_{ij} & V_{ij}\\ V_{jk} & V_{jk} & V_{jk}\\ V_{ki} & V_{ki} & V_{ki}\end{pmatrix}\begin{pmatrix}\psi_{ij}\\ \psi_{jk}\\ \psi_{ki}\end{pmatrix}. \tag{3}$$

[1] L. Eyges, Phys. Rev. **115**, 1643 (1959).
[2] L. D. Faddeev, Zh. Eksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys.—JETP **12**, 1014 (1961)].
[3] J. R. Jasperse and M. H. Friedman, Phys. Rev. **159**, 69 (1967).
[4] This result follows using arguments similar to those developed in Ref. 3.

Rewriting these equations so that the self-coupled terms appear together gives

$$\begin{pmatrix}[1 + G_0(s)V_{ij}]\psi_{ij}\\ [1 + G_0(s)V_{jk}]\psi_{jk}\\ [1 + G_0(s)V_{ki}]\psi_{ki}\end{pmatrix} = -G_0(s)\begin{pmatrix}0\\ V_{jk}\\ V_{ki}\end{pmatrix}\Psi^{(i)} - G_0(s)\begin{pmatrix}0 & V_{ij} & V_{ij}\\ V_{jk} & 0 & V_{jk}\\ V_{ki} & V_{ki} & 0\end{pmatrix}\begin{pmatrix}\psi_{ij}\\ \psi_{jk}\\ \psi_{ki}\end{pmatrix}. \quad (4)$$

If the incident wave were a state where each particle were free, then the inhomogeneous term of the first equation in Eqs. (4) would be $-G_0(s)V_{ij}\Psi^{(i)}$ instead of 0. The momentum representation of the homogeneous equations embodied in Eqs. (4), representing the three-body bound states, is the same as that derived in Ref. 3.

Consider a system of three particles interacting through pair potentials, where particles i and k are *identical* and particle j is different. In dealing with situations where some of the particles are identical, we may either solve the Schrödinger equation without any symmetry requirement on the solution and then symmetrize (or antisymmetrize) afterward, or we symmetrize (or antisymmetrize) both the solution and the equations in the beginning. In a two-body problem with two identical particles, there is no advantage in using one method over the other. However, for the three-body problem using the partial-wavefunction approach, it is advantageous to symmetrize (or antisymmetrize) the solutions first. Since all pair potentials are even functions of their arguments, the three coupled equations which result are no longer all independent and the system of equations can be simplified. In this way the number of independent scattered partial wavefunctions can be reduced. The Hamiltonian is given by Eq. (2) where the pair potentials V_{ij} and V_{jk} have the same form. It can be shown that the incident state-vector and the scattered state vector are[4]

$$\Psi^{S,A(i)} = \psi_{ij}^{(i)} \pm \psi_{jk}^{(i)},$$
$$\Psi^{S,A} = \psi_{ij}^{+,-} \pm \psi_{jk}^{+,-} + \psi_{ki}^{e,o},$$

and that the total state-vector is given by

$$\Psi_t = \Psi^{S,A(i)} + \Psi^{S,A},$$

where the incident partial state-vector is a known eigenstate of $H_o + V_{ij}$. The + and − superscripts refer to either the symmetric or antisymmetric total solution. The *e* and *o* superscripts refer to the evenness or oddness of the scattered partial state-vector upon exchange of particles i and k. For particle k, incident on the *bound state* of particles i and j, we substitute this form for the total state-vector into the Schrödinger equation and use procedures similar to those of the previous case to obtain

$$\begin{pmatrix}[1 + G_0(s)V_{ij}]\psi_{ij}^{+,-}\\ \pm[1 + G_0(s)V_{jk}]\psi_{jk}^{+,-}\\ [1 + G_0(s)V_{ki}]\psi_{ki}^{e,o}\end{pmatrix} = -G_0(s)\begin{pmatrix}\pm V_{ij}\psi_{jk}^{(i)}\\ V_{jk}\psi_{ij}^{(i)}\\ V_{ki}[\psi_{ij}^{(i)} \pm \psi_{jk}^{(i)}]\end{pmatrix} - G_0(s)\begin{pmatrix}0 & V_{ij} & V_{ij}\\ V_{jk} & 0 & V_{jk}\\ V_{ki} & V_{ki} & 0\end{pmatrix}\begin{pmatrix}\psi_{ij}^{+,-}\\ \pm\psi_{jk}^{+,-}\\ \psi_{ki}^{e,o}\end{pmatrix}. \quad (5)$$

Note that the total state-vector and the coupled set of equations giving the partial scattered state-vectors are explicitly symmetrized (or antisymmetrized). For this system it is possible to show that there are only two independent equations by proving that[4]

$$\psi_{ij}^{(i)} = \psi_{jk}^{(i)} \quad \text{and} \quad \psi_{ij}^{+,-} = \psi_{jk}^{+,-}.$$

By this we mean that when we take the matrix elements of the operator Eqs. (5) in momentum space we can prove that the form of $\psi_{ij}^{(i)}$ is the same as that of $\psi_{jk}^{(i)}$ and the form of $\psi_{ij}^{+,-}$ is the same as that of $\psi_{jk}^{+,-}$. If each particle were initially free, then the inhomogeneous terms would be

$$-G_0(s)\begin{pmatrix}V_{ij}[\psi_{ij}^{(i)} \pm \psi_{jk}^{(i)}]\\ V_{jk}[\psi_{ij}^{(i)} \pm \psi_{jk}^{(i)}]\\ V_{ki}[\psi_{ij}^{(i)} \pm \psi_{jk}^{(i)}]\end{pmatrix}.$$

In order to make these operator equations more explicit we write out the matrix elements for the case of two identical particles interacting with a third which is different. The center-of-mass coordinates in position space and in momentum space may be defined as follows:

$$\mathbf{R} \equiv (m_i\mathbf{r}_i + m_j\mathbf{r}_j + m_k\mathbf{r}_k)M^{-1},$$
$$\mathbf{r}_{ij} \equiv (\mathbf{r}_i - \mathbf{r}_j),$$
$$\boldsymbol{\rho}_k \equiv \mathbf{r}_k - (m_i\mathbf{r}_i + m_j\mathbf{r}_j)(m_i + m_j)^{-1},$$
$$\mathbf{K} \equiv \mathbf{k}_i + \mathbf{k}_j + \mathbf{k}_k,$$
$$\mathbf{k}_{ij} \equiv (m_j\mathbf{k}_i - m_i\mathbf{k}_j)(m_i + m_j)^{-1},$$
$$\boldsymbol{\kappa}_k \equiv [(m_i + m_j)\mathbf{k}_k - m_k(\mathbf{k}_i + \mathbf{k}_j)]M^{-1},$$

where the three sets of coordinates for each representation are generated by allowing i, j, and k to assume the values 1, 2, and 3 in cyclical permutation and M is the total mass. Letting particles 1 and 3 be the identical pair and taking the matrix elements of the two independent equations of Eqs. (5) in the frame

$(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3) \equiv (\mathbf{k}, \boldsymbol{\varkappa})$ gives[5]

$$\phi_{12}^{+,-}(\mathbf{k}, \boldsymbol{\varkappa}) + [(2\pi)^3(\alpha_{12}k^2 + \beta_3\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V_{12}(\mathbf{r})\phi_{12}^{+,-}(\mathbf{k}', \boldsymbol{\varkappa}) \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k}) = I_{12}^{+,-}(\mathbf{k}, \boldsymbol{\varkappa}) - [(2\pi)^3(\alpha_{12}k^2 + \beta_3\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V_{12}(\mathbf{r})\{\pm\phi_{12}^{+,-}(b\mathbf{k}', \mathbf{k}' - b^{-1}\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[\mathbf{k}' - \mathbf{k} - (b^{-1} - b)\boldsymbol{\varkappa}] + \phi_{31}^{e,o}(\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[-\mathbf{k}' - \mathbf{k} + (2 - a)\boldsymbol{\varkappa}]\} \quad (6a)$$

$$\phi_{31}^{e,o}(\mathbf{k}, \boldsymbol{\varkappa}) + [(2\pi)^3(\alpha_{31}k^2 + \beta_2\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V_{31}(\mathbf{r})\phi_{31}^{e,o}(\mathbf{k}', \boldsymbol{\varkappa}) \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k}) = I_{31}^{+,-}(\mathbf{k}, \boldsymbol{\varkappa}) - [(2\pi)^3\alpha_{31}k^2 + \beta_2\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V_{31}(\mathbf{r})\{\phi_{12}^{+,-}(-a\mathbf{k}', \mathbf{k}' - a^{-1}\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[\mathbf{k}' - \mathbf{k} - (a^{-1} - \tfrac{1}{2})\boldsymbol{\varkappa}] \pm \phi_{12}^{+,-}(-a\mathbf{k}', \mathbf{k}' - a^{-1}\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[-\mathbf{k}' - \mathbf{k} + (a^{-1} - \tfrac{1}{2})\boldsymbol{\varkappa}]\}. \quad (6b)$$

Here we have defined

$$\langle \mathbf{k}, \boldsymbol{\varkappa} \mid \psi_{ij}^{+,-} \rangle \equiv \phi_{ij}^{+,-}(\mathbf{k}, \boldsymbol{\varkappa})$$

and the inhomogeneous terms are

$$I_{12}^{+,-}(\mathbf{k}, \boldsymbol{\varkappa}) = -(\pm)[(2\pi)^3(\alpha_{12}k^2 + \beta_3\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V_{12}(\mathbf{r})\phi_{12}^{(i)}(b\mathbf{k}', \mathbf{k}' - b^{-1}\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[\mathbf{k}' - \mathbf{k} - (b^{-1} - b)\boldsymbol{\varkappa}], \quad (7a)$$

$$I_{31}^{+,-}(\mathbf{k}, \boldsymbol{\varkappa}) = -[(2\pi)^3(\alpha_{31}k^2 + \beta_2\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V_{31}(\mathbf{r})\{\phi_{12}^{(i)}(-a\mathbf{k}', \mathbf{k}' - a^{-1}\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[\mathbf{k}' - \mathbf{k} - (a^{-1} - \tfrac{1}{2})\boldsymbol{\varkappa}] \pm \phi_{12}^{(i)}(-a\mathbf{k}', \mathbf{k}' - a^{-1}\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot[-\mathbf{k}' - \mathbf{k} + (a^{-1} - \tfrac{1}{2})\boldsymbol{\varkappa}]\}. \quad (7b)$$

We have also defined

$$\langle \mathbf{k}, \boldsymbol{\varkappa} \mid \psi_{ij}^{(i)} \rangle \equiv \phi_{ij}^{(i)}(\mathbf{k}, \boldsymbol{\varkappa}),$$

and the constants appearing in these equations for $m_1 = m_3$ are given by

$$\alpha_{12} = \tfrac{1}{2}a^{-1}, \qquad \beta_3 = \tfrac{1}{2}(1 + b),$$
$$\alpha_{31} = 1, \qquad \beta_2 = \tfrac{1}{2}(a^{-1} - \tfrac{1}{2}),$$
$$a = m_2(m_1 + m_2)^{-1}, \quad b = m_1(m_1 + m_2)^{-1}.$$

[5] Note that the homogeneous parts of these equations differ slightly from Eqs. (14a) and (14b) of Ref. 3. This comes about because we have redefined the total wavefunction for the case of two identical particles [given in this paper by Eq. (8)] and have corrected a slight error in the symmetry discussion in Ref. 3.

The space part of the total wavefunction in the frame $(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3)$ is defined as

$$\langle \mathbf{k}_{12}, \boldsymbol{\varkappa}_3 \mid \Psi_t^{S,A} \rangle \equiv \Phi_t^{S,A}(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3),$$

and is given by

$$\Phi_t^{S,A}(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3) = \phi_{12}^{(i)}(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3) \pm \phi_{12}^{(i)}(-\mathbf{k}_{23}, \boldsymbol{\varkappa}_1) + \phi_{12}^{+,-}(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3) \pm \phi_{12}^{+,-}(-\mathbf{k}_{23}, \boldsymbol{\varkappa}_1) + \phi_{31}^{e,o}(\mathbf{k}_{31}, \boldsymbol{\varkappa}_2). \quad (8)$$

Note that $\Phi_t^{S,A}(\mathbf{k}_{12}, \boldsymbol{\varkappa}_3)$ is symmetric or antisymmetric upon the exchange of particles 1 and 3.

For three *identical* particles it is also possible to show that a solution of the following form exists:

$$\Psi_t^{S,A} = \Psi^{S,A(i)} + \Psi^{S,A},$$

where

$$\Psi^{S,A(i)} = \psi_{ij}^{e,o(i)} + \psi_{jk}^{e,o(i)} + \psi_{ki}^{e,o(i)},$$
$$\Psi^{S,A} = \psi_{ij}^{e,o} + \psi_{jk}^{e,o} + \psi_{ki}^{e,o}.$$

The coupled equations, which result when the incident state is the one in which one particle scatters from the bound state of the other two, are given by

$$\begin{pmatrix} [1 + G_0(s)V_{ij}]\psi_{ij}^{e,o} \\ [1 + G_0(s)V_{jk}]\psi_{jk}^{e,o} \\ [1 + G_0(s)V_{ki}]\psi_{ki}^{e,o} \end{pmatrix} = -G_0(s)\begin{pmatrix} V_{ij}[\psi_{jk}^{e,o(i)} + \psi_{ki}^{e,o(i)}] \\ V_{jk}[\psi_{ij}^{e,o(i)} + \psi_{ki}^{e,o(i)}] \\ V_{ki}[\psi_{ij}^{e,o(i)} + \psi_{jk}^{e,o(i)}] \end{pmatrix} - G_0(s)\begin{pmatrix} 0 & V_{ij} & V_{ij} \\ V_{jk} & 0 & V_{jk} \\ V_{ki} & V_{ki} & 0 \end{pmatrix}\begin{pmatrix} \psi_{ij}^{e,o} \\ \psi_{jk}^{e,o} \\ \psi_{ki}^{e,o} \end{pmatrix}. \quad (9)$$

Again, we have explicitly symmetrized or antisymmetrized the equations and as a consequence we may show that the form of each incident partial state-vector is the same, and the form of each scattered partial state-vector is the same. This in turn implies that there is only one independent equation in Eqs. (9). Taking the matrix elements in momentum space of the single independent equation in Eqs. (9), and dropping the subscripts i, j, and k we obtain

$$\phi^{e,o}(\mathbf{k}, \boldsymbol{\varkappa}) + [(2\pi)^3(k^2 + \tfrac{3}{4}\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V(\mathbf{r})\phi^{e,o}(\mathbf{k}', \boldsymbol{\varkappa}) \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k}) = I^{e,o}(\mathbf{k}, \boldsymbol{\varkappa}) - [(2\pi)^3(k^2 + \tfrac{3}{4}\kappa^2 - s)]^{-1} \times \iint d\mathbf{r}\, d\mathbf{k}' V(\mathbf{r})\{\phi^{e,o}(\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot(-\mathbf{k}' - \mathbf{k} + \tfrac{3}{2}\boldsymbol{\varkappa}) + \phi^{e,o}(-\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\varkappa}) \times \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k} - \tfrac{3}{2}\boldsymbol{\varkappa})\}, \quad (10)$$

where we have defined $\langle \mathbf{k}, \boldsymbol{\kappa} \mid \psi_{ij}^{e,o} \rangle \equiv \phi_{ij}^{e,o}(\mathbf{k}, \boldsymbol{\kappa})$ and where the inhomogeneous term is

$$\begin{aligned} I^{e,o}(\mathbf{k}, \boldsymbol{\kappa}) = &-[(2\pi)^3(k^2 + \tfrac{3}{4}\kappa^2 - s)]^{-1} \\ &\times \iint d\mathbf{r}\, d\mathbf{k}' V(\mathbf{r})\{\phi^{e,o(i)}(\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\kappa}) \\ &\times \exp i\mathbf{r}\cdot(-\mathbf{k}' - \mathbf{k} + \tfrac{3}{2}\boldsymbol{\kappa}) \\ &+ \phi^{e,o(i)}(-\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\kappa}) \\ &\times \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k} - \tfrac{3}{2}\boldsymbol{\kappa})\}. \end{aligned} \tag{11}$$

The total wavefunction in the frame $(\mathbf{k}_{ij}, \boldsymbol{\kappa}_k)$ is

$$\begin{aligned} \langle \mathbf{k}_{ij}, \boldsymbol{\kappa}_k \mid \Psi_i^{S,A} \rangle &\equiv \Phi_i^{S,A}(\mathbf{k}_{ij}, \boldsymbol{\kappa}_k) \\ &= \phi^{e,o(i)}(\mathbf{k}_{ij}, \boldsymbol{\kappa}_k) + \phi^{e,o(i)}(\mathbf{k}_{jk}, \boldsymbol{\kappa}_i) \\ &\quad + \phi^{e,o(i)}(\mathbf{k}_{ki}, \boldsymbol{\kappa}_j) + \phi^{e,o}(\mathbf{k}_{ij}, \boldsymbol{\kappa}_k) \\ &\quad + \phi^{e,o}(\mathbf{k}_{jk}, \boldsymbol{\kappa}_i) + \phi^{e,o}(\mathbf{k}_{ki}, \boldsymbol{\kappa}_j). \end{aligned} \tag{12}$$

III. THREE IDENTICAL PARTICLES INTERACTING THROUGH ZERO-RANGE PAIR POTENTIALS

In Sec. II we have derived some general equations for the three-body scattering problem using the partial-wavefunction approach. We have also shown how the Pauli principle may be incorporated so as to symmetrize or antisymmetrize the equations when some or all of the particles become identical. In this section we consider one case explicitly, i.e., the case where three identical particles interact through zero-range pair potentials, and show that our method produces the same result as the Faddeev T-matrix approach.

The method we propose for calculating the total wavefunction consists of performing the following three steps:

(1) Take the matrix elements in momentum space of the appropriate operator equations for a given incident state;

(2) Use the knowledge of the bound states of each pair problem and the requirement that the Fourier transform exist and represent all possible outgoing states to deduce the most general expansion for each scattered partial wavefunction;

(3) Substitute these expansions into the appropriate equations and derive a set of coupled integral equations for the unknown functions appearing in each expansion.

Let us carry this out for the zero-range pair potential problem. The pair problem has one bound state given by

$$\tilde{N}^\beta \psi^\beta(\mathbf{r}) = \tilde{N}^\beta e^{-\beta r}/r; \quad \tilde{N}^\beta = (\beta/2\pi)^{\frac{1}{2}},$$

for $r > r_0$, where r_0 is the range of the potential. In momentum space we have

$$N^\beta \phi^\beta(\mathbf{k}) = N^\beta(k^2 + \beta^2)^{-1}; \quad N^\beta = (\beta/\pi^2)^{\frac{1}{2}}.$$

The incident partial wavefunction, when one particle is incident on the bound state of the other two, is then

$$\phi^{(i)}(\mathbf{k}, \boldsymbol{\kappa}) = N^\beta \phi^\beta(\mathbf{k})(2\pi)^3 \delta(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0), \tag{13}$$

where $\boldsymbol{\kappa}_0$ is the wavevector of the incident particle. We have dropped the e and o superscripts since it is understood that we are looking for the symmetric solution. Since the pair potential has zero range we see from Eqs. (10) and (11) that the following substitution is valid:

$$\phi(\mathbf{k}, \boldsymbol{\kappa}) = (k^2 + \tfrac{3}{4}\kappa^2 - s)^{-1} g(\boldsymbol{\kappa}; \boldsymbol{\kappa}_0).$$

The wavefunction in momentum space must have the appropriate singularities to produce all possible outgoing states. Since the pair problem has only one bound state we may write

$$\begin{aligned} \phi(\mathbf{k}, \boldsymbol{\kappa}) = &\, 3\pi N^\beta (k^2 + \tfrac{3}{4}\kappa^2 - s)^{-1} \\ &\times (-\beta^2 + \tfrac{3}{4}\kappa^2 - s)^{-1} f(\boldsymbol{\kappa}; \boldsymbol{\kappa}_0), \end{aligned} \tag{14}$$

where f is a bounded continuous function of κ and θ (angle between $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}_0$) with bounded continuous derivatives on κ and θ. Substituting Eq. (14) into Eq. (10), noticing that the factors $(k^2 + \frac{3}{4}\kappa^2 - s)^{-1}$ cancel out, multiplying through by $\phi^{\beta *}(\mathbf{k})$, and integrating with respect to $\mathbf{k}$, we obtain

$$\begin{aligned} [\Delta^\beta + D^\beta(\kappa)] \frac{3\pi N^\beta f(\boldsymbol{\kappa}; \boldsymbol{\kappa}_0)}{(-\beta^2 + \frac{3}{4}\kappa^2 - s)} = &-(2\pi)^{-3} \iiint d\mathbf{k}\, d\mathbf{r}\, d\mathbf{k}' V(\mathbf{r}) \phi^{\beta *}(\mathbf{k}) \Big\{ \phi^{(i)}(\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\kappa}) \\ &\times \exp i\mathbf{r}\cdot(-\mathbf{k}' - \mathbf{k} + \tfrac{3}{2}\boldsymbol{\kappa}) + \phi^{(i)}(-\tfrac{1}{2}\mathbf{k}', \mathbf{k}' - 2\boldsymbol{\kappa}) \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k} - \tfrac{3}{2}\boldsymbol{\kappa}) \\ &+ \frac{3\pi N^\beta f(\mathbf{k}' - 2\boldsymbol{\kappa}; \boldsymbol{\kappa}_0) \exp i\mathbf{r}\cdot(-\mathbf{k}' - \mathbf{k} + \frac{3}{2}\boldsymbol{\kappa})}{[(\frac{1}{2}k')^2 + \frac{3}{4}(\mathbf{k}' - 2\boldsymbol{\kappa})^2 - s][-\beta^2 + \frac{3}{4}(\mathbf{k}' - 2\boldsymbol{\kappa})^2 - s]} \\ &+ \frac{3\pi N^\beta f(\mathbf{k}' - 2\boldsymbol{\kappa}; \boldsymbol{\kappa}_0) \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k} - \frac{3}{2}\boldsymbol{\kappa})}{[(-\frac{1}{2}k')^2 + \frac{3}{4}(\mathbf{k}' - 2\boldsymbol{\kappa})^2 - s][-\beta^2 + \frac{3}{4}(\mathbf{k}' - 2\boldsymbol{\kappa})^2 - s]} \Big\}. \end{aligned} \tag{15}$$

Here we have defined the following expression:

$$\Delta^\beta + D^\beta(\kappa) \equiv \int d\mathbf{k}\, \phi^{\beta*}(\mathbf{k})\Big\{1 + (2\pi)^{-3}\iint d\mathbf{r}\, d\mathbf{k}' V(\mathbf{r}) \times [(k'^2 + \tfrac{3}{4}\kappa^2 - s)^{-1} \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k})]\Big\}.$$

The integrations over the pair potential may be eliminated by making use of the two-body Schrödinger equation in momentum space:

$$\phi^{\beta*}(\mathbf{k}') = -[(2\pi)^3(k'^2 + \beta^2)]^{-1} \times \iint d\mathbf{k}\, d\mathbf{r} V(\mathbf{r})\phi^{\beta*}(\mathbf{k}) \exp i\mathbf{r}\cdot(\mathbf{k}' - \mathbf{k}).$$

This reduces the vector integrations from three to one. Substituting Eq. (13) for the $\phi^{(i)}$ and making an appropriate transformation of variables for the one remaining vector variable appearing in Eq. (15), we obtain

$$(\tfrac{3}{4})[\Delta^\beta + D^\beta(\kappa)]\frac{f(\boldsymbol{\varkappa};\boldsymbol{\varkappa}_0)}{(-\beta^2 + \tfrac{3}{4}\kappa^2 - s)} = 4\pi^2[(\boldsymbol{\varkappa}_0 + \tfrac{1}{2}\boldsymbol{\varkappa})^2 + \beta^2]\phi^{\beta*}(\boldsymbol{\varkappa}_0 + \tfrac{1}{2}\boldsymbol{\varkappa})\phi^\beta(\tfrac{1}{2}\boldsymbol{\varkappa}_0 + \boldsymbol{\varkappa}) + (\tfrac{3}{2})\int d\mathbf{y}\frac{f(\mathbf{y};\boldsymbol{\varkappa}_0)}{(y^2 + \mathbf{y}\cdot\boldsymbol{\varkappa} + \kappa^2 - s)(-\beta^2 + \tfrac{3}{4}y^2 - s)}.$$

The expression for $\Delta^\beta + D^\beta(\kappa)$ becomes

$$\Delta^\beta + D^\beta(\kappa) = \int d\mathbf{k}\Big\{\phi^{\beta*}(\mathbf{k}) - \frac{(k^2 + \beta^2)\phi^{\beta*}(\mathbf{k})}{(k^2 + \tfrac{3}{4}\kappa^2 - s)}\Big\}.$$

For the zero-range pair potential we may calculate $\Delta^\beta + D^\beta(\kappa)$ by using the following formula:

$$\int dx \frac{x^2}{(x^2 \pm z)} = x \mp z\int dx \frac{1}{(x^2 \mp z)} + C.$$

We see that the infinite parts cancel and we are left with

$$\Delta^\beta + D^\beta(\kappa) = 2\pi^2[-\beta - i(s - \tfrac{3}{4}\kappa^2)^{\frac{1}{2}}].$$

The integral equation for f becomes[6]

$$\left(\frac{3}{8}\right)\frac{f(\boldsymbol{\varkappa};\boldsymbol{\varkappa}_0)}{[\beta - i(s - \tfrac{3}{4}\kappa^2)^{\frac{1}{2}}]} = (\kappa^2 + \boldsymbol{\varkappa}\cdot\boldsymbol{\varkappa}_0 + \kappa_0^2 - s)^{-1} + \frac{3}{8\pi^2}\int d\mathbf{y}\frac{f(\mathbf{y};\boldsymbol{\varkappa}_0)}{(y^2 + \mathbf{y}\cdot\boldsymbol{\varkappa} + \kappa^2 - s)(-\beta^2 + \tfrac{3}{4}y^2 - s)}. \tag{16}$$

Note that the self-coupling of the scattered partial wavefunction causes no difficulty at the point $\boldsymbol{\varkappa} = \boldsymbol{\varkappa}_0$, and that the kernel of the integral equation for f is a Schmidt operator for all complex s except for part of the real axis from $-\beta^2$ to ∞. It is also possible to prove[7] that the limit on the cut exists for $s \to (m/\hbar^2)E < 0$ in the Banach space C_1 (the space of bounded continuous functions with bounded continuous derivatives) and hence, for $E < 0$, the kernel is compact in C_1. The compactness proof fails in C_1 for $s \to (m/\hbar^2)E > 0$. This does not mean that the kernel cannot be made compact in some other Banach space. In fact, making the substitution

$$\phi'(\boldsymbol{\varkappa};\boldsymbol{\varkappa}_0) = (\tfrac{3}{8})[\beta - i(s - \tfrac{3}{4}\kappa^2)^{\frac{1}{2}}]^{-1}f(\boldsymbol{\varkappa};\boldsymbol{\varkappa}_0)$$

we obtain the following equation:

$$\phi'(\boldsymbol{\varkappa};\boldsymbol{\varkappa}_0) = (\kappa^2 + \boldsymbol{\varkappa}\cdot\boldsymbol{\varkappa}_0 + \kappa_0^2 - s)^{-1} - \frac{1}{\pi^2}\int d\mathbf{y}\frac{\phi'(\mathbf{y};\boldsymbol{\varkappa}_0)}{(\kappa^2 + \boldsymbol{\varkappa}\cdot\mathbf{y} + y^2 - s)(\beta + i(s - \tfrac{3}{4}y^2)^{\frac{1}{2}})}. \tag{17}$$

This equation is identical with that derived by Faddeev for this problem.[2] Faddeev has shown in general that the fifth power of the kernel is compact in a certain Banach space.[8] The space part of the total wavefunction for symmetric states is then

$$\Phi_i^S(\mathbf{k}_{ij}, \boldsymbol{\varkappa}_k) = \mathrm{SYM}\{N^\beta\phi^\beta(\mathbf{k}_{ij})(2\pi)^3\delta(\boldsymbol{\varkappa}_k - \boldsymbol{\varkappa}_k^0) + 3\pi N^\beta(k_{ij}^2 + \tfrac{3}{4}\kappa_k^2 - s)^{-1} \times (-\beta_{ij}^2 + \tfrac{3}{4}\kappa_k^2 - s)^{-1}f(\boldsymbol{\varkappa}_k;\boldsymbol{\varkappa}_k^0)\},$$

where the SYM denotes symmetry with respect to the variables $\mathbf{k}_{ij}$ and $\boldsymbol{\varkappa}_k$ as written out explicitly in Eq. (12). We note here that Eqs. (16) and (17) are not unique as they stand due to the pathological behavior of the three-body bound state for the zero-range pair potential.[9] The three-body transition amplitude for this problem can be calculated in a straightforward way. For example, in the frame $(\mathbf{k}_{ij}, \boldsymbol{\varkappa}_k)$ we get the following result for elastic scattering:

$$T_{f\leftarrow i} = [f(\boldsymbol{\varkappa};\boldsymbol{\varkappa}_0)]_{\boldsymbol{\varkappa}\to\boldsymbol{\varkappa}_0} \equiv f(\kappa_0, \theta; \kappa_0).$$

IV. DISCUSSION

We have derived some new equations for the three-body scattering problem using the partial-wavefunction approach and showed how the equations simplify when some or all of the particles are identical.

[6] The inhomogeneous term may be written this way since the equation only has meaning for $s = (m/\hbar^2)E \pm i\epsilon$ in the limit as $\epsilon \to 0$. For a different derivation of this equation, see G. V. Skornia-kov and K. A. Ter-Martirosian, Zh. Eksp. Teor. Fiz. **31**, 775 (1956) [Sov. Phys.—JETP **4**, 648 (1957)].

[7] The proof is similar to that of Lovelace, [C. Lovelace, Phys. Rev. **135**, B1225 (1964)].

[8] L. D. Faddeev, *Mathematical Aspects of the Three-body Problem in the Quantum Scattering Theory* (Steklov Mathematical Institute, Leningrad, 1963), No. 69.

[9] The nonuniqueness is due to the fact that the three-body bound state has a solution for arbitrary E. For a discussion on how a satisfactory solution can be chosen using these equations, see G. S. Danilov, Zh. Eksp. Teor. Fiz. **40**, 498 (1961) [Sov. Phys.—JETP **13**, 349 (1961)].

We have also indicated how each partial wavefunction can be expanded and applied the method to the problem of three identical particles interacting through zero-range pair potentials. It is important to note that with this method we do *not* need to know the off-shell, two-body t matrices; only a knowledge of the two-body bound states for each pair potential is necessary. Our feeling at present is that this method can be applied to more complicated pair interactions. Three-dimensional, square-well scattering and electron–hydrogen scattering are currently being considered using this approach.

ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with Professor M. Friedman and Dr. L. Eyges.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Representation Functions for the Degenerate Baryon Series in $U(6) \otimes U(6)$*

R. DELBOURGO
Imperial College, London

(Received 6 December 1967)

The most degenerate class of representation functions $d^{NB}_{W,W'}(\theta)$ pertaining to $U(6) \otimes U(6)$ have been calculated for arbitrary quark number N with the baryon number B, and W, W' weights of the $U(6)_W$ subgroup assuming the following values: $B = 0, \frac{1}{3}, \frac{2}{3}, 1$; $[W], [W'] = 1, 6, 21, 56$. These functions are the matrix elements of the rotation operator $\exp(-i\theta J_2)$ between basis vectors of the degenerate $U(6) \otimes U(6)$ Feynman series (characterized entirely by BNW labels), and are essential to the generalized Reggeization program.

1. REPRESENTATION FUNCTIONS OF THE MOST-DEGENERATE SERIES

A generalized Reggeization scheme based on the supermultiplet symmetry $U(6) \otimes U(6)$ has recently been proposed[1] where it has been assumed for definiteness that the sequence of physically observed states falls in an especially simple class of representations[2]—the Feynman series—which are characterized by only two Casimir labels: the baryon number B and the quark number N. Correspondingly, the poles of the generalized partial-wave components of the S matrix are presumed to occur just in this degenerate sequence and to provide the supermultiplet Regge poles. An analysis shows that the N-plane Regge-pole contribution to the amplitude is of the form $\beta d^{\alpha B}_{W,W'}(\theta)/\sin \pi\alpha$, where $d^{\alpha B}_{W,W'}(\theta)$ is the continuation of the rotation-matrix element

$$d^{NB}_{W,W'}(\theta) = \langle NBW| e^{-i\theta J_2} |NBW'\rangle \qquad (1)$$

to complex $N = \alpha(E)$ for the degenerate states $|BNW\rangle$. In practice, therefore, one requires a knowledge of these functions for $B = 0$ (meson exchange) and $B = 1$ (baryon exchange). The $B = 0$ functions have been stated in a previous publication[1]; in this paper we present the $B = 1$ functions, building these up from the known $B = 0$ set through $B = \frac{1}{3}$ (quark sequence) and $B = \frac{2}{3}$ (biquark sequence).

Before evaluating the $d(\theta)$, one should briefly recall the precise definition of the $U(6) \otimes U(6)$ degenerate-states set which we are proposing to adopt in order to appreciate the comparatively simple nature of the ensuing results. To do so, it is useful to state the significant properties of the fundamental spinor ("quark") representations from which all others, and in particular the degenerate series, can be constructed. Letting B_1 and B_2 refer to the "baryon numbers" of the commuting phase transformations of $U(1) \otimes U(1)$ and writing (W_1, W_2) for the representations of the commuting $SU(6)$ groups of $SU(6) \otimes SU(6)$, we have the four basic spinors:

ψ_a transforming as $(6, 1)$; $B_1 = \frac{1}{3}$, $B_2 = 0$,
ψ^a transforming as $(\bar{6}, 1)$; $B_1 = -\frac{1}{3}$, $B_2 = 0$,
$\psi_{\dot{a}}$ transforming as $(1, 6)$; $B_1 = 0$, $B_2 = \frac{1}{3}$,
$\psi^{\dot{a}}$ transforming as $(1, \bar{6})$; $B_1 = 0$, $B_2 = -\frac{1}{3}$.

By constructing multispinor products of these four

* The research reported in this document has been sponsored in part by the Air Force Office for Scientific Research (OAR) through the European Office of Aerospace Research, United States Air Force.

[1] A. Salam and J. Strathdee, Phys. Rev. Letters, **19**, 339 (1967); R. Delbourgo, M. A. Rashid, A. Salam, and J. Strathdee, Phys. Rev. Phys. Rev. **170**, 1477 (1968).

[2] Y. Dothan, M. Gell-Mann,and Y. Neeman, Phys. Letters. **17**, 148 (1965); A. Gotsman and Y. Neeman, Tel-Aviv University preprint (1966); R. Delbourgo, A. Salam, and J. Strathdee, Proc. Roy. Soc. (London) **289A**, 177 (1965).

We have also indicated how each partial wavefunction can be expanded and applied the method to the problem of three identical particles interacting through zero-range pair potentials. It is important to note that with this method we do *not* need to know the off-shell, two-body t matrices; only a knowledge of the two-body bound states for each pair potential is necessary. Our feeling at present is that this method can be applied to more complicated pair interactions. Three-dimensional, square-well scattering and electron–hydrogen scattering are currently being considered using this approach.

ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with Professor M. Friedman and Dr. L. Eyges.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Representation Functions for the Degenerate Baryon Series in $U(6) \otimes U(6)$*

R. DELBOURGO
Imperial College, London

(Received 6 December 1967)

The most degenerate class of representation functions $d^{NB}_{W,W'}(\theta)$ pertaining to $U(6) \otimes U(6)$ have been calculated for arbitrary quark number N with the baryon number B, and W, W' weights of the $U(6)_W$ subgroup assuming the following values: $B = 0, \frac{1}{3}, \frac{2}{3}, 1$; $[W], [W'] = 1, 6, 21, 56$. These functions are the matrix elements of the rotation operator $\exp(-i\theta J_2)$ between basis vectors of the degenerate $U(6) \otimes U(6)$ Feynman series (characterized entirely by BNW labels), and are essential to the generalized Reggeization program.

1. REPRESENTATION FUNCTIONS OF THE MOST-DEGENERATE SERIES

A generalized Reggeization scheme based on the supermultiplet symmetry $U(6) \otimes U(6)$ has recently been proposed[1] where it has been assumed for definiteness that the sequence of physically observed states falls in an especially simple class of representations[2]—the Feynman series—which are characterized by only two Casimir labels: the baryon number B and the quark number N. Correspondingly, the poles of the generalized partial-wave components of the S matrix are presumed to occur just in this degenerate sequence and to provide the supermultiplet Regge poles. An analysis shows that the N-plane Regge-pole contribution to the amplitude is of the form $\beta d^{\alpha B}_{W,W'}(\theta)/\sin \pi\alpha$, where $d^{\alpha B}_{W,W'}(\theta)$ is the continuation of the rotation-matrix element

$$d^{NB}_{W,W'}(\theta) = \langle NBW| e^{-i\theta J_2} |NBW'\rangle \qquad (1)$$

to complex $N = \alpha(E)$ for the degenerate states $|BNW\rangle$. In practice, therefore, one requires a knowledge of these functions for $B = 0$ (meson exchange) and $B = 1$ (baryon exchange). The $B = 0$ functions have been stated in a previous publication[1]; in this paper we present the $B = 1$ functions, building these up from the known $B = 0$ set through $B = \frac{1}{3}$ (quark sequence) and $B = \frac{2}{3}$ (biquark sequence).

Before evaluating the $d(\theta)$, one should briefly recall the precise definition of the $U(6) \otimes U(6)$ degenerate-states set which we are proposing to adopt in order to appreciate the comparatively simple nature of the ensuing results. To do so, it is useful to state the significant properties of the fundamental spinor ("quark") representations from which all others, and in particular the degenerate series, can be constructed. Letting B_1 and B_2 refer to the "baryon numbers" of the commuting phase transformations of $U(1) \otimes U(1)$ and writing (W_1, W_2) for the representations of the commuting $SU(6)$ groups of $SU(6) \otimes SU(6)$, we have the four basic spinors:

ψ_a transforming as $(6, 1)$; $B_1 = \frac{1}{3}$, $B_2 = 0$,
ψ^a transforming as $(\bar{6}, 1)$; $B_1 = -\frac{1}{3}$, $B_2 = 0$,
$\psi_{\dot{a}}$ transforming as $(1, 6)$; $B_1 = 0$, $B_2 = \frac{1}{3}$,
$\psi^{\dot{a}}$ transforming as $(1, \bar{6})$; $B_1 = 0$, $B_2 = -\frac{1}{3}$.

By constructing multispinor products of these four

* The research reported in this document has been sponsored in part by the Air Force Office for Scientific Research (OAR) through the European Office of Aerospace Research, United States Air Force.

[1] A. Salam and J. Strathdee, Phys. Rev. Letters, **19**, 339 (1967); R. Delbourgo, M. A. Rashid, A. Salam, and J. Strathdee, Phys. Rev. Phys. Rev. **170**, 1477 (1968).

[2] Y. Dothan, M. Gell-Mann,and Y. Neeman, Phys. Letters. **17**, 148 (1965); A. Gotsman and Y. Neeman, Tel-Aviv University preprint (1966); R. Delbourgo, A. Salam, and J. Strathdee, Proc. Roy. Soc. (London) **289A**, 177 (1965).

quarks we can represent any $U(6) \otimes U(6)$ state. Thus a supermultiplet of baryon number $B = B_1 + B_2$ and quark number $N \equiv \frac{3}{2}(B_1 - B_2)$ is described by the multispinor[3]

$$\Phi^{b_1 \cdots b_{n_1},\ \hat{b}_1 \cdots \hat{b}_{n_2-3B_2}}_{a_1 \cdots a_{n_1+3B_1},\ \hat{a}_1 \cdots \hat{a}_{n_2}}.$$

By imposing tracelessness and particular symmetries in indices of the same type, we may select out the desired (W_1, W_2) level.

The degenerate series referred to above is defined by the additional conditions $n_1 = n_2 = 0$, together with complete symmetry in the remaining indices, viz., by multispinors of the variety $\Phi^{(\hat{b}_1 \cdots \hat{b}_N)}_{(a_1 \cdots a_{N+3B})}$. Thus for the mesons ($B = 0$) and the baryons ($B = 0$) we encounter the Feynman series:

$$N = \quad 0, \qquad 1, \qquad 2, \qquad \cdots$$

$$(W_1, W_2) = \quad (1, 1), \quad (6, \bar{6}), \quad (21, \overline{21}), \cdots ; B = 0$$

$$(W_1, W_2) = (56, 1), (126, \bar{6}), (252, \overline{21}), \cdots ; B = 1.$$

The simplifying feature of these sequences is that, under the reduction $SU(6) \otimes SU(6)$ to $SU(6)$, the $SU(6)$ (W) representations occur singly in any (W_1, W_2) representation of $SU(6) \otimes SU(6)$. Consequently, the labeling $|BNW\rangle$ is sufficient to characterize the degenerate states. Nevertheless, it must be emphasized that the $|BNW\rangle$ vectors on their own cannot constitute a complete set since, for given B and N, they need to be supplemented by other less-degenerate vectors.

In the simplest instance[4] the Reggeization program for $U(6) \otimes U(6)$ calls for the evaluation of the special representations $d^{NB}_{W,W'}(\theta)$ of Eq. (1), for $B = 0$ and $B = 1$. In previous work[1] it was shown that arbitrary $d^N_{W,W'}(\theta)$ for given B can be built up from the lowest W, W' values by differential techniques, and that it was sufficient to consider the basic functions

$$d^N_{[1][1]}(\theta), \quad B = 0; \quad d^N_{[6][6']}(\theta), \quad B = \tfrac{1}{3};$$
$$d^N_{[21][21']}(\theta), \quad B = \tfrac{2}{3}; \quad d^N_{[56][56']}(\theta), \quad B = 1;$$

and so on. We shall demonstrate below how the $B \neq 0$ functions can themselves be built up from the $B = 0$ set also by successive differentiations through $B = \frac{1}{3}, \frac{2}{3}$, etc.

2. MULTISPINOR $d(\theta)$ FUNCTIONS

The calculation of $d(\theta)$ will be carried out in two steps: the first consists in casting the arguments in the M-function framework of the inhomogeneous $U(6, 6)$ theory to obtain $U(6, 6)$ tensor expressions for the $d(\theta)$; the second step, deferred to Sec. 3, consists in contracting over external $U(6, 6)$ wavefunctions[3] $\Phi^{(B_1 \cdots B_N)}_{(A_1 \cdots A_{N+3B})}$ to obtain the canonical expressions for $d(\theta)$.

Now the fundamental function $d^N_{[1][1]}(\theta)$ arises in the supersinglet scattering process

$$(1, 1)_{\frac{1}{2}p+q} + (1, 1)_{\frac{1}{2}p-q} \to (1, 1)_{\frac{1}{2}p+q'} + (1, 1)_{\frac{1}{2}p-q'}$$

by including an intermediate $(N, \bar{N})_p$ meson state and imposing $U(6)_W$ conservation[3] at the 3-point vertices—that is in fact the basic source of $d^N_{WW'}(\theta)$ in the Reggeization scheme. As the rules for such M-function calculations have been stated so often,[3] we shall not repeat them here. The significant point to be made is that $d^N_{[1][1]}(\theta)$ devolves upon reducing the expression

$$F_N = \prod_{j=1}^{N} \{(p - m)q'(p + m)\}^{B_j}_{A_j} \Big[\sum_B \prod_{k=1}^{N} q^{A_k}_{B_k} \Big] \Big/ N! \quad (2)$$

for $p^2 = m^2$; a straightforward, if tedious, computation shows that the right-hand side of Eq. (2) reduces, for $U(v) \otimes U(v)$, to

$$F_N = (4m^2 |q| |q'|)^N C^{\frac{1}{2}v}_N(\hat{q} \cdot \hat{q}'), \quad (3)$$

where

$$\mathbf{q} \cdot \mathbf{q}' \equiv -q \cdot q' + q \cdot pq' \cdot p/m^2 \equiv |q| |q'| \cos\theta. \quad (4)$$

Thus

$$d^N_{[1][1]}(\theta) \propto C^3_N(\cos\theta) \quad (5)$$

for $U(6) \otimes U(6)$.

Consider now the functions $d^N_{[6][6']}(\theta)$ belonging to the $B = \frac{1}{3}$ family of excitations. These arise from the scattering process

$$(6, 1) + (1, 1) \to (6', 1) + (1, 1)$$

by including an intermediate $B = \frac{1}{3}$ state of arbitrary N. The basic quantities which enter are the vertex function

$$\langle (1, 1)| \mathcal{T}^{(A_1 \cdots A_N)}_{(B_1 \cdots B_{N+1})} |(6, 1)\rangle = \sum_B \Big[u^{[6]}_{B_{N+1}} \prod_{k=1}^{N} q^{A_k}_{B_k} \Big] \Big/ (2m)^N (N + 1)!$$

and the numerator of the propagator

$$(p^2 - m^2)\langle \Phi^{(B_1 \cdots B_{N+1})}_{(A_1 \cdots A_N)} \Phi^{(A_1' \cdots A_{N'})}_{(B_1' \cdots B_{N+1}')} \rangle = \sum_{A,B} \Big[(p + m)^{B_{N+1}}_{B_{N+1}'} \prod_{k=1}^{N} (p + m)^{B_k}_{B_k'} (p - m)^{A_k'}_{A_k} \Big] \Big/ (2m)^{2N+1} N! (N + 1)!$$

which together provide the expression

$$T_{[6][6']} = \bar{u}^B_{[6']} T^A_B u^{[6]}_A / (2m)^{4N+1}, \quad (6)$$

[3] R. Delbourgo, M. A. Rashid, A. Salam and J. Strathdee, *Trieste Conference Proceedings* (I.A.E.A., Vienna, 1965).

[4] One should not categorically preclude the possibility that the physically observed supermultiplets fall into sequences of levels more complicated than the degenerate sequence considered in this paper. If that is the case, the calculations will be correspondingly more complicated; that is all.

where

$$T^A_B = (p+m)^{B_{N+1}}_B \prod_{j=1}^{N} \{(p-m)q'(p+m)\}^{B_j}_{A_j} \times \sum_B \left[\delta^A_{B_{N+1}} \prod_{k=1}^{N} q^{A_k}_{B_k}\right] \Big/ (N+1)!. \quad (6')$$

The right-hand side of (6) is not so difficult to simplify as it appears to be at first sight, if we make use of the following trick. We notice that, since

$$F_{N+1} = (4m^2|q||q'|)^{N+1} C^{\frac{1}{2}\nu}_{N+1}(\cos\theta) = \prod_{j=1}^{N+1} \{(p-m)q'(p+m)\}^{B_j}_{A_j} \times \left[\sum_B \prod_{k=1}^{N+1} q^{A_k}_{B_k}\right] \Big/ (N+1)!,$$

$$\frac{\partial F_{N+1}}{\partial q^A_B} = \frac{\{(p-m)q'(p+m)\}^{B_{N+1}}_B}{N!} \times \prod_{j=1}^{N} \{(p-m)q'(p+m)\}^{B_j}_{A_j} \times \sum_B \left[\delta^A_{B_{N+1}} \prod_{k=1}^{N} q^{A_k}_{B_k}\right].$$

Hence,

$$\left[(p+m)q'(p-m)\frac{\partial}{\partial q}\right]^A_B F_{N+1} = -\frac{8m^3q'^2}{N!}(p+m)^{B_{N+1}}_B \prod_{j=1}^{N} \{(p-m)q'(p+m)\}^{B_j}_{A_j} \times \sum_B \left[\delta^A_{B_{N+1}} \prod_{k=1}^{N} q^{A_k}_{B_k}\right] = -8m^3q'^2(N+1)T^A_B. \quad (7)$$

The advantage of this manipulation lies in the fact that the left-hand side can be readily evaluated according to techniques which have already been described elsewhere.[1] Thus

$$\frac{\partial F_{N+1}}{\partial q} = \frac{(4m^2|q||q'|)^N}{\nu}\left[\{(p-m)q'(p+m)\}C^{\frac{1}{2}\nu'}_{N+1} - \{(p-m)q(p+m)\}\frac{|q'|}{|q|}C^{\frac{1}{2}\nu'}_N\right] \quad (8)$$

and leads to

$$T^A_B = \frac{(4m^2|q||q'|)^N}{(N+1)\nu}[(\Lambda_+)^A_B C^{\frac{1}{2}\nu'}_{N+1} - (\Lambda_-)^A_B C^{\frac{1}{2}\nu'}_N], \quad (9)$$

where

$$\Lambda_+ \equiv (p+m)/2m,$$
$$\Lambda_- \equiv (p+m)q'(p-m)q(p+m)/8m^3|q||q'|. \quad (10)$$

In this way we have arrived at our multispinor formula for $d(\theta)$ for the degenerate family belonging to $B = \frac{1}{3}$:

$$d^N(\theta) \propto [\Lambda_+ C^{3'}_{N+1} - \Lambda_- C^{3'}_N]/\nu(N+1). \quad (11)$$

The same trick method can be used with success to obtain the $B = \frac{2}{3}$, 1 degenerate-series functions. Performing the same manipulations as outlined above, we get the multispinor expressions for $B = \frac{2}{3}$:

$$d^N_{[21][21']}(\theta) = \bar{u}^{(B_1B_2)}_{[21']} d^{NA_1A_2}_{B_1B_2}(\theta) u^{[21]}_{(A_1A_2)}, \quad (12)$$

with

$$d^{NA_1A_2}_{B_1B_2}(\theta) \propto [(-8m^3q'^2)^2(N+1)(N+2)]^{-1} \times \left[(p+m)q'(p-m)\frac{\partial}{\partial q}\right]^{A_1}_{B_1} \times \left[(p+m)q'(p-m)\frac{\partial}{\partial q}\right]^{A_2}_{B_2} F_{N+2} = [(\Lambda_+)^{A_1}_{B_1}(\Lambda_+)^{A_2}_{B_2}C^{\frac{1}{2}\nu''}_{N+1} - 2(\Lambda_+)^{A_1}_{B_1}(\Lambda_-)^{A_2}_{B_2}C^{\frac{1}{2}\nu''}_{N+1} + (\Lambda_-)^{A_1}_{B_1}(\Lambda_-)^{A_2}_{B_2}C^{\frac{1}{2}\nu''}_N]/(N+1)(N+2)\nu^2 \quad (12')$$

or, abbreviating to an obvious notation,

$$d^N(\theta) \propto [\Lambda_+ \otimes \Lambda_+ C''_{N+2} - 2\Lambda_+ \otimes \Lambda_- C''_{N+1} + \Lambda_- \otimes \Lambda_- C''_N]; \quad (12'')$$

$B = 1$:

$$d^N(\theta) \propto [\Lambda_+ \otimes \Lambda_+ \otimes \Lambda_+ C'''_{N+3} - 3\Lambda_+ \otimes \Lambda_+ \otimes \Lambda_- C'''_{N+2} + 3\Lambda_+ \otimes \Lambda_- \otimes \Lambda_- C'''_{N+1} - \Lambda_- \otimes \Lambda_- \otimes \Lambda_- C'''_N], \quad (13)$$

relating to

$$d^N_{[56][56']}(\theta) = \bar{u}^{(B_1B_2B_3)}_{[56']}(k') d^{NA_1A_2A_3}_{B_1B_2B_3} u^{56]}_{(A_1A_2A_3)}(k). \quad (13')$$

The reader is urged to test the veracity of the results for small N by direct explicit calculation. Note that all $d^N(\theta)$ are associated with the "threshold" factor $(4m^2|q||q'|)^N$.

3. CANONICAL $d(\theta)$ FUNCTIONS

We have only to perform the contractions $\bar{u} \wedge u$ over the Bargmann–Wigner wavefunctions u in Eqs. (6), (12), and (13) in order to arrive at the canonical forms $d^N_{W,W'}(\theta)$. These are straightforwardly done if we make use of the basic contractions

$$\bar{u}_{r',\pm}\Lambda_+ u_{r,\pm} = \bar{u}_{r',\pm}\Lambda_- u_{r,\pm} = \delta_{rr'}\cos\tfrac{1}{2}\theta,$$
$$\bar{u}_{r',\mp}\Lambda_+ u_{r\pm} = -\bar{u}_{r',\mp}\Lambda_- u_{r,\pm} = \pm\delta_{rr'}\sin\tfrac{1}{2}\theta, \quad (14)$$

which formulas are easily proved in the rest frame of p. $\pm$ are helicity indices and r, r' are $SU(3)$ indices of the quark wavefunctions. The remaining thing to be done is to impose the boundary conditions

$$d^N_{W,W'}(\theta = 0) = \delta_{WW'}$$

to find the over-all normalization factors, and this is also easily done from the well-known properties

$$C^{\frac{1}{2}\nu}_N(1) = \Gamma(\nu+N)/\Gamma(N+1)\Gamma(\nu),$$
$$C^{\frac{1}{2}\nu'}_N(x) = \nu C^{\frac{1}{2}\nu+1}_{N-1}(x). \quad (15)$$

We present below all relevant expressions for the reduced functions $d^N_{\lambda\lambda'}(\theta)$ defined by

$$d^N_{W,W'}(\theta) = d^N_{R\lambda,R'\lambda'}(\theta) = \delta_{RR'}\, d^N_{\lambda\lambda'}(\theta), \quad (16)$$

where R and R' refer to $SU(3)$ labels contained in W, W'. These are derived by inserting the appropriate wavefunctions[3] in (6), (12), and (13):

$$u_{[W=6]} = u_{r\lambda},$$
$$u_{[W=21]} = u_{(r\lambda,s\mu)} = u^{[R=6]}_{(rs)(\lambda\mu)} + u^{[R=\bar{3}]}_{[rs][\lambda\mu]},$$
$$u_{[W=56]} = u_{(r\lambda,s\mu,tv)} = u^{[R=10]}_{(rst)(\lambda\mu\nu)} + \{u^{[R=8]}_{[rs]t,[\lambda\mu]\nu} + \text{cyclic}\}.$$

The final results for $d^N_{\lambda\lambda'}(\theta)$ are ($\nu = 6$):

Case: $B = 0$; $[W] = [W'] = 1$

$$d^N_{00}(\theta) = a^0_{N\nu} C^{\frac{1}{2}\nu}_N(\cos\theta),$$
$$a^{3B}_{N\nu} \equiv \Gamma(\nu + 3B)\Gamma(N+1)/\Gamma(N+\nu+3B) \times 2(2+\nu)\cdots(6B-2+\nu); \quad (17)$$

Case: $B = \frac{1}{3}$; $[W] = [W'] = 6$

$$d^N_{\frac{1}{2}\frac{1}{2}}(\theta) = a^1_{N\nu}\cos\tfrac{1}{2}\theta[C'_{N+1} - C'_N],$$
$$d^N_{-\frac{1}{2}\frac{1}{2}}(\theta) = a^1_{N\nu}\sin\tfrac{1}{2}\theta[C'_{N+1} + C'_N]. \quad (18)$$

Note the close similarity of these expressions with the ones encountered[5] for $SU(2)$ by putting $\nu = 1$.

Case: $B = \frac{2}{3}$; $[W] = [W'] = 21$

When $[R] = [R'] = 6$:

$$d^N_{11}(\theta) = a^2_{N\nu}\cos^2\tfrac{1}{2}\theta[C''_{N+2} - 2C''_{N+1} + C''_N],$$
$$d^N_{01}(\theta) = \sqrt{2}a^2_{N\nu}\cos\tfrac{1}{2}\theta\sin\tfrac{1}{2}\theta[C''_{N+2} - C''_N],$$
$$d^N_{-11}(\theta) = a^2_{N\nu}\sin^2\tfrac{1}{2}\theta[C''_{N+2} + 2C''_{N+1} + C''_N], \quad (17')$$
$$d^N_{00}(\theta) = a^2_{N\nu}[\cos\theta C''_{N+2} - 2C''_{N+1} + \cos\theta C''_N].$$

For $[R] = [R'] = \bar{3}$ we have instead

$$d^N_{00}(\theta) = a^2_{N\nu}[C''_{N+2} - 2\cos\theta C''_{N+1} + C''_N]. \quad (18')$$

Case: $\mathbf{B} = 1$; $[W] = [W'] = 56$

When $[R] = [R'] = 10$:

$$d^N_{\frac{3}{2}\frac{3}{2}}(\theta) = a^3_{N\nu}\cos^3\tfrac{1}{2}\theta[C'''_{N+3} - 3C'''_{N+2} + 3C'''_{N+1} - C'''_N],$$
$$d^N_{\frac{1}{2}\frac{3}{2}}(\theta) = \sqrt{3}a^3_{N\nu}\cos^2\tfrac{1}{2}\theta\sin\tfrac{1}{2}\theta \times [C'''_{N+3} - C'''_{N+2} - C'''_{N+1} + C'''_N],$$
$$d^N_{-\frac{1}{2}\frac{3}{2}}(\theta) = \sqrt{3}a^3_{N\nu}\cos\tfrac{1}{2}\theta\sin\tfrac{1}{2}\theta \times [C'''_{N+3} + C'''_{N+2} - C'''_{N+1} - C'''_N],$$
$$d^N_{-\frac{3}{2}\frac{3}{2}}(\theta) = a^3_{N\nu}\sin^3\tfrac{1}{2}\theta[C'''_{N+3} + 3C'''_{N+2} + 3C'''_{N+1} + C'''_N],$$
$$d^N_{\frac{1}{2}\frac{1}{2}}(\theta) = d^N_{\frac{3}{2}\frac{3}{2}}(\theta) - \frac{2}{\sqrt{3}}d^N_{-\frac{1}{2}\frac{3}{2}}(\theta),$$
$$d^N_{-\frac{1}{2}\frac{1}{2}}(\theta) = -d^N_{-\frac{3}{2}\frac{3}{2}}(\theta) + \frac{2}{\sqrt{3}}d^N_{\frac{1}{2}\frac{3}{2}}(\theta). \quad (19)$$

[5] See the Appendix to M. Jacob and G. C. Wick, Ann. Phys. (N.Y.) 7, 404, (1959).

Finally, for $[R] = [R'] = 8$:

$$d^N_{\frac{1}{2}\frac{1}{2}}(\theta) = a^3_{\nu N}\cos\tfrac{1}{2}\theta \times [C'''_{N+3} - 3C'''_{N+2} + 3\cos\theta C'''_{N+1} - C'''_N],$$
$$d^N_{-\frac{1}{2}\frac{1}{2}}(\theta) = a^3_{\nu N}\sin\tfrac{1}{2}\theta \times [C'''_{N+3} + 3C'''_{N+2} - 3\cos\theta C'''_{N+1} + C'''_N]. \quad (20)$$

At this point it is instructive to write down the orthogonality properties of these functions. For the $B = 0$ and $B = \frac{1}{3}$ functions given above, one can definitely state the relations

$B = 0$:

$$\int d^N(\theta)\, d^M(\theta)(\sin\theta)^\nu\, d\theta = \frac{\delta_{NM}2^\nu\Gamma(N+1)[\Gamma(\frac{1}{2}\nu+\frac{1}{2})]^2}{(2N+\nu)\Gamma(N+\nu)},$$

$B = \frac{1}{3}$:

$$\int d^N_{WW'}(\theta)\, d^M_{WW'}(\theta)(\sin\theta)^\nu\, d\theta = \frac{\delta_{NM}(N+2\nu)\Gamma(N+1)[\Gamma(\frac{1}{2}\nu+\frac{1}{2})]^2}{N(2N+\nu)\Gamma(N+\nu+1)}.$$

Unfortunately, for the $B = \frac{2}{3}$ and $B = 1$ functions, such simple orthogonality properties no longer exist[6]; there one really requires a full knowledge of the $U(6) \otimes U(6)$ invariant-group element and an exhaustive study of *all* the group representations, which we have not examined in this paper. The general orthonormality conditions therefore remain a problem.

ACKNOWLEDGMENT

I wish to thank Dr. M. Scadron for many pertinent comments.

[6] For example, in $O(4)$, the basic representation functions

$$d^{jo\sigma}_{jmj'}(\theta) = \langle jm|\, e^{-i\theta k_3}|\, j'm\rangle$$

satisfy more involved orthonormality conditions:

$$\sum_{m=-j}^{j}\int d^{jo\sigma}_{jmj'}(\theta)\, d^{jo'\sigma}_{jm'j}(\theta)\sin^2\theta\, d\theta = \frac{\delta_{jojo'}\delta_{\sigma\sigma'}}{j_o^2 - \sigma^2}.$$

We may conjective that the $U(\nu) \otimes U(\nu)$ functions satisfy similar sorts of conditions with a summation over various multiplicity labels that serve to distinguish between the less degenerate sets of states.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Cell Model of a Fluid. I. Evaluation of the Partition Function and Series Expansions

RALPH G. TROSS*†‡ AND LOUIS H. LUND
University of Missouri at Rolla, Rolla, Missouri

(Received 3 May 1968)

The partition function of a one-dimensional system of particles with all interactions active is formulated in terms of Ising ferromagnetic spin states. It is shown how the partition function for two- and three-dimensional systems can be obtained from the one-dimensional one by restricting the number of bonds per particle. After writing the partition function in matrix form, the operator of interest is diagonalized and its trace expressed in the form of a convenient infinite series. The series is shown to be absolutely convergent and its analytic properties are briefly investigated. It is then applied to one-, two-, and three-dimensional systems with nearest-neighbor interactions. The validity of the model is established by summing the one-dimensional series and comparing it with the known solution obtainable by other methods. Two- and three-dimensional series are determined by algebraic techniques identical to the one-dimensional series. These are seen to agree with the low-temperature expansions obtained by other authors. The method of this article is seen to have the advantage of simplicity and uniformity regardless of the dimension of the system. A subsequent article is devoted to the thermodynamics of the model.

1. INTRODUCTION

It is well known that it has not been possible as yet to formulate a completely satisfactory molecular theory of the liquid state which correctly portrays phase transitions and remains valid in the transition region. Basically, the difficulty lies in the high densities of liquids which approach those of solids, while symmetries of the solid state are lacking. Consequently, in liquid theory one is forced to deal with a general N-body problem without recourse to the simplifications and approximations that are available in dealing with solids and gases. Nor has it proved feasible thus far to extend the cluster expansions and virial series to the transition region itself and to liquids.[1–4] While these series converge absolutely and generally quite rapidly for gases, it has not been possible to extend these analytically to other phase regions. In view of the work of Yang and Lee, such an extension may well be impossible on theoretical grounds,[5] since, in the thermodynamic limit, it seems probable that the logarithm of the partition function is a lacunary function[6] with a solid wall of singularities surrounding the origin in the complex y plane (y = activity). Not surprisingly, there have been very few successful attempts to find an exact mathematical representation of a system exhibiting a phase transition. Some typical exceptions are the two-dimensional Ising model,[7,8] the ideal Bose gas,[3] the spherical Ising model,[9,10] and the Kac–Baker model.[4,11–13] All but two of these deal with models in one and two dimensions. While these do not correspond to any real system except perhaps a surface film or filament, they have the decisive advantage of being mathematically tractable. One hopes, then, that some of the conclusions and properties of these models may generalize to three dimensions.

We address ourselves in this series of two articles to the study of a simple one-dimensional model. Van Hove has shown that one-dimensional systems with finite-range pairwise potentials with a cutoff will show no phase transition.[14] Accordingly, the interaction potential chosen for the model investigated here has an infinite range with a hard-rod repulsive core and possesses, as we shall see, variable range and depth parameters. In these respects the model resembles the Kac–Baker one, but it differs from the latter in the analysis and technique of solution. Moreover, the potential is a modified Lennard–Jones one, rather than the exponential interaction potential used by Kac and Baker. The system investigated here is a cell model of a simple fluid in which the cells have

* Based upon a dissertation submitted in partial fulfillment of the Ph.D. degree, University of Missouri at Rolla (1968).
† Work supported in part by a NASA Predoctoral Traineeship.
‡ Present address: Department of Mathematics, University of Ottawa, Ottawa, Ontario.

[1] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1940).
[2] T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Co., Inc., New York, 1956).
[3] K. Huang, *Statistical Mechanics* (John Wiley & Sons, New York, 1963).
[4] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216 (1963).
[5] C. N. Yang and T. D. Lee, Phys. Rev., **87**, 404, 410 (1952).
[6] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, 1963).
[7] L. Onsager, Phys. Rev., **65**, 117 (1944).
[8] B. Kaufman, Phys. Rev., **76**, 1232 (1949).
[9] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[10] H. A. Gersch, Phys. Fluids **6**, 599 (1963).
[11] M. Kac, Phys. Fluids **2**, 8 (1959).
[12] G. Baker, Phys. Rev. **122**, 1477 (1961).
[13] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **5**, 60 (1964).
[14] L. Van Hove, Physica, **15**, 951 (1949); **16**, 137 (1950).

the purpose of specifying the instantaneous location of the particles with an uncertainty equal to the cell parameter. It is therefore not a lattice gas in the usual sense, since the particles are not fixed at their lattice sites but are free to move throughout the system. We shall find that, by selecting the cell parameter equal to the particle's exclusion length, we are able to separate completely the repulsive-core potential from the attractive tail. This choice is equivalent to introducing an exclusion principle into the system which limits cell occupancy to one particle or none. It is therefore convenient to designate such occupancy by means of Ising spin indices. We shall see, however, that the spinor-algebraic approach which proved so successful in the two-dimensional Ising model appears not to lead to a solution in the present instance. This is because the infinite-range potential causes every particle to interact with every other one, so that in the thermodynamic limit the coordination number becomes infinite. It is possible, however, to linearize the partition function by a Fourier-type transform, thereby obtaining a solution of the problem in the form of an infinite series which converges absolutely for all finite values of the activity. By applying the solution to a nearest-neighbor potential, one can confirm the validity of the approach, since the series can then be summed and compared with the known closed-form solution. This calculation at the same time lays the ground work for obtaining low-temperature series in two and three dimensions by identical algebraic techniques. We shall prove that, in order to obtain such series, one need only modify the interaction potential in a suitable way. It will be established that by this simple stratagem two- and three-dimensional systems with more limited interactions can be generated from the basic one-dimensional model. While it is not necessary to restrict the interaction to nearest neighbors in these higher-dimensional systems, it is essential that the coordination number remain small so that the approach be mathematically tractable. The resulting series will be seen to agree completely with those obtained by other authors.

The thermodynamics of this model will be explored in the second article of this series. We shall find there that the behavior is surprisingly realistic and that the system undergoes a change of phase under certain conditions.

2. MATHEMATICAL FORMULATION OF THE PROBLEM

A. One-Dimensional System

The basic system considered is a one-dimensional fluid in which every particle interacts with every other one. Extension to higher dimensions will be considered later. Let the particles be distributed on a line of length L, bent to form a ring. All distance measurements then occur mod. N. The periodic boundary condition is introduced for mathematical convenience only; it becomes insignificant in the thermodynamic limit. We divide the line into N equal segments or cells, each of length $l = L/N$. Cells may be either occupied or empty, particles being free to move from one cell to the next. The particles are assumed to be completely symmetric and to interact with a modified Lennard–Jones potential of the form

$$V(x) = \begin{cases} +\infty & (x < l/2), \\ -\zeta/|x|^{\gamma} & (x \geq l/2), \end{cases} \tag{2.1}$$

where x is the separation between particles, ζ is the potential depth, and γ is a parameter fixing the range of interaction. The hard-core repulsive interaction is most conveniently accounted for by letting the cell-size l coincide with the exclusion length of the particle and restricting cell occupation numbers to zero and one. All other choices of l can be shown to be trivial variations of this scheme. The interaction potential in terms of the present discrete model is then

$$\phi_s = \tfrac{1}{2}(V_s + V_{N-s}),$$
$$V_s = \begin{cases} 0 & (s = 0) \\ -\zeta/s^{\gamma} & (s \neq 0) \end{cases} \quad \text{(Mod. } N\text{)}, \tag{2.2}$$

where s is the distance between the two particles of interest measured in units of the cell parameter l from cell center to cell center. The ϕ potential is introduced to symmetrize the interaction around the ring for finite systems. The hard-core repulsive potential having been built into the model by the above scheme, the potential of the zeroth cell can be chosen arbitrarily for mathematical convenience. The ϕ potential satisfies the following symmetry conditions:

$$\phi_s = \phi_{s+N} = \phi_{s-N}. \tag{2.3}$$

Letting n_k represent the occupation number of the kth cell, the energy of the system is given by

$$H = \frac{1}{2m}\sum_{r=1}^{n} p_r^2 + \sum_{r=1}^{N}\sum_{s=1}^{N} n_r n_{r+s}\phi_s, \quad \left(\sum_{r=1}^{N} n_r = n\right) \tag{2.4}$$

where it is assumed that there are n particles in the system and that the total interaction energy of the system can be obtained by summing all pair-wise interactions. "Surface" terms (i.e., interactions with the boundary) do not occur by virtue of the periodic boundary conditions.

Now let the occupation number n_k be replaced by a cell occupation index $\sigma_k = 2n_k - 1$. The configurational part of the system energy can then be written as

$$U_c = \tfrac{1}{4}\sum_{r=1}^{N}{}'\sum_{s=1}^{N}{}'(1+\sigma_r)(1+\sigma_{r+s})\phi_s$$
$$= \left(N/4 + \tfrac{1}{2}\sum_{r=1}^{N}{}'\sigma_r\right)\Phi + \tfrac{1}{4}\sum_{r=1}^{N}{}'\sum_{s=1}^{N}{}'\sigma_r\sigma_{r+s}\phi_s, \quad (2.5)$$

where $\Phi = \sum_{s=1}^{N}\phi_s$, and the primed summation implies the restriction $\sum_{r=1}^{N}\sigma_r = 2n - N$.

The partition function appropriate for a system of n particles is then

$$\mathfrak{Q}_n = 1/(n!\,h^n)\int dx_1\cdots\int dx_n\int dp_1\cdots\int dp_n$$
$$\times \exp\{-\beta H(p_1,\cdots,p_n;x_1,\cdots x_n)\}$$
$$= 1/(n!\,\lambda^n)Z_n \quad (\beta = 1/kT), \quad (2.6)$$

where $\lambda = (\beta h^2/2\pi m)^{\frac{1}{2}}$ is the thermal wave length and results from the evaluation of the momentum integrals.[3] A straightforward analysis shows that for large systems the configurational part of the partition function takes the form

$$Z_n = \int dx_1\cdots\int dx_n \exp\{-\beta U_c(x_1,\cdots,x_n)\}$$
$$= n!\,l^n\sum_{\langle\sigma\rangle}\exp\{-\beta U_c(\sigma_1,\cdots,\sigma_N)\}, \quad (2.7)$$

where

$$\sum_{\langle\sigma\rangle} = \sum_{\sigma_1=-1}^{1}{}'\cdots\sum_{\sigma_N=-1}^{1}{}'.$$

The partition function is then

$$\mathfrak{Q}_n = (l/\lambda)^n\exp\{-\beta N\Phi/4\}$$
$$\times\sum_{\langle\sigma\rangle}\exp\left\{-\beta/4\left(\sum_{r,s}\sigma_r\sigma_{r+s}\phi_s + 2\Phi\sum_r\sigma_r\right)\right\} \quad (2.8)$$

Since the present investigation is concerned with systems subject to phase transitions, i.e., with systems in which the number of particles is not constant, it is preferable to deal with a grand canonical partition function. This also eliminates the awkward restriction on the sum of σ's. One then has

$$Q_N = \sum_{n=0}^{N}\xi^n\mathfrak{Q}_n, \quad \text{where} \quad \xi = \exp\{\beta g\} \quad (2.9)$$

and g is the chemical potential per particle. Writing $\xi l/\lambda = \exp\{\beta\mu\}$ and using the identity

$$n = \tfrac{1}{2}\sum_{r=1}^{N}(1+\sigma_r),$$

we obtain

$$Q_N = A_N\sum_{\{\sigma\}}\exp\left\{\sum_{r=1}^{N}\sum_{s=1}^{N}\sigma_r\sigma_{r+s}\theta_s + \nu\sum_{r=1}^{N}\sigma_r\right\}, \quad (2.10)$$

where

$$\sum_{\{\sigma\}} = \sum_{\sigma_1=-1}^{1}\cdots\sum_{\sigma_N=-1}^{1}, \quad A_N = \exp\{N(\nu-\Theta)\},$$
$$\theta_s = -\tfrac{1}{4}\beta\phi_s, \quad \Theta = \sum_{r=1}^{N}\theta_r,$$

and $\nu = \frac{1}{2}\beta(\mu - \Phi)$. The remainder of the investigation will use this form of the partition function as its point of departure. Each term in this sum of 2^N terms represents one possible configuration or microstate of the system. Since the particles comprising the fluid are indistinguishable, the probability of finding the system in a state of n observable particles is

$$W(n) = A_N Q_N^{-1}\sum_i{}'\exp\left\{\sum_{r,s}\sigma_r^i\sigma_{r+s}^i\theta_s + \nu\sum_r\sigma_r^i\right\}, \quad (2.11)$$

where σ_r^i represents the ith configuration of the rth cell and $\sum_i'$ implies a sum over all configurations consistent with the macroscopic state of the system. It is worthwhile noting that the cell structure of the system does not imply any spurious symmetry or lattice structure. The cells are used merely as a convenient mathematical artifice to establish a discrete space grid and fix the instantaneous position of the particles.

B. Multi-Dimensional Systems

The partition function of two- and three-dimensional systems with a limited number of bonds per particles is readily obtained from the preceding analysis of a one-dimensional system with all interactions active. Focussing attention first on a square array with nearest neighbor interactions only, we find that we can construct such a configuration by winding the one-dimensional chain around a three-dimensional torus (Fig. 1) and letting all interactions vanish except those between nearest neighbors in the chain and those between nearest-neighbor cells in adjacent rows. Figure 2 is a plane projection of such a configuration.

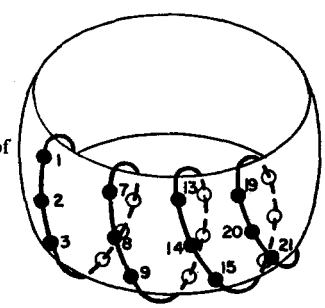


FIG. 1. Toroidal topology of 2-dimensional model.

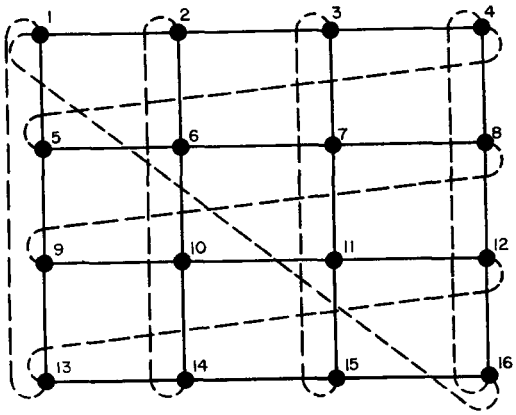


FIG. 2. Two-dimensional model in plane projection.

Thus we obtain from Eq. (2.10) the partition function for a two-dimensional square array with m cells per row simply by use of the following potential:

$$V_r = \begin{Bmatrix} -\zeta & (r = 1, m) \\ 0 & (r \neq 1, m) \end{Bmatrix} \text{(Mod. } N\text{)}, \quad (2.12)$$

with ϕ_r given by Eq. (2.3) as before. It is obvious that this potential introduces certain spurious interactions. In Fig. 2 these are shown by broken lines. While such interactions are significant in finite models, they become negligible in the thermodynamic limit. One can proceed in exactly the same way to obtain the partition function of a three-dimensional system. A typical square array of 64 cells is shown in Fig. 3. For a cubic array of N cells with m cells per row, k cells per layer, the appropriate nearest-neighbor

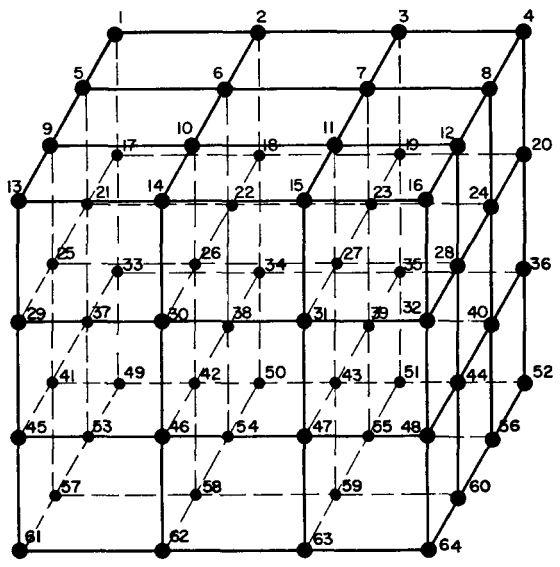


FIG. 3. Three-dimensional array.

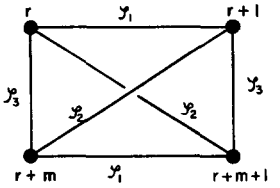


Fig. 4. Anisotropic nearest-neighbor and next-nearest-neighbor interaction.

interaction potential is

$$V_r = \begin{Bmatrix} -\zeta & (r = 1, m, k) \\ 0 & (r \neq 1, m, k) \end{Bmatrix} \text{(Mod. } N\text{)}. \quad (2.13)$$

Extension of the one-dimensional system partition function to higher dimensions is not restricted to nearest-neighbor interactions. More complicated interactions can also be accommodated, provided only that the number of bonds per cell remains finite and sufficiently small to make the analysis feasible mathematically. An element of a typical anisotropic square array with nearest-neighbor and next-nearest-neighbor interactions is shown in Fig. 4. The partition function of such a system can be obtained from Eq. (2.10) by means of the following potential:

$$V_r = \begin{Bmatrix} -\zeta_1 & (r = 1), \\ -\zeta_2 & (r = m - 1, m + 1), \\ -\zeta_3 & (r = m), \\ 0 & (r \neq 1, m - 1, m, m + 1), \end{Bmatrix} \text{(Mod. } N\text{)}. \quad (2.14)$$

In two- and three-dimensional systems, just as in one dimension, the cell structure serves merely as a space grid for the purpose of specifying the instantaneous location of the particles with an uncertainty equal to the cell dimension. To take account of the change in the momentum integrals in multidimensional systems, the parameter λ must, of course, be redefined as follows: $\lambda = (\beta h^2/2\pi m)^{d/2}$, where d is the dimensionality of the system.

C. Multiple Cell Occupancy

One can readily extend the model to take account of cell occupancies other than 0 and 1, provided that intermediate occupancies are not considered. If, for instance, one limits cell occupancy to 0 or m particles, the partition function takes the same form as Eq. (2.10) if one makes the following definitions:

ϕ_0 = average interaction potential of particles in the same cell,

$$\theta_s = -\tfrac{1}{4}\beta m^2 \phi_s; \quad \Theta' = \sum_{r=1}^{N-1} \theta_r,$$

$$\nu = \tfrac{1}{2}\beta m[\mu + \tfrac{1}{2}(1 + m)\phi_0 - m\Phi],$$

$$A_N = \exp\{N(\nu - \Theta')\}.$$

For the simple fluid with which this investigation is concerned, such a modification is undesirable since the cell size would have to be increased, thus reducing the fineness of the space grid. In addition, the hard-core repulsive potential could not be satisfied in this case.

D. Ferromagnetic Systems

The partition function in (2.10) is identical with that of an Ising ferromagnet.[3,15] Comparison of the definitions of the various parameters for the fluid and ferromagnet yields the well-known isomorphisms noted by Yang and Lee[5] and conveniently summarized by Huang.[3] The analysis can be broadened considerably, but this will not be done in the present context.

3. OPERATOR FORMULATION

Let μ_i be a state vector representing the state of the ith cell and Z be an operator defined as follows:

$$Z\mu_i = \sigma_i\mu_i. \tag{3.1}$$

We define now a vector ψ^α representing the αth configuration of the system:

$$\psi^\alpha = |\mu_1^\alpha, \mu_2^\alpha, \cdots, \mu_N^\alpha\rangle. \tag{3.2}$$

A convenient representation of ψ^α is a 2^N-dimensional one in which it is the direct product of the N two-dimensional μ-spinors. A corresponding operator Z_k can be defined in the same space as follows:

$$Z_k\psi^\alpha = \sigma_k^\alpha\psi^\alpha. \tag{3.3}$$

In the 2^N-dimensional representation the operator Z_k takes the form

$$Z_k = I \otimes I \otimes \cdots \otimes I \otimes \underset{\substack{\uparrow \\ k\text{th factor}}}{Z} \otimes I \otimes \cdots \otimes I, \tag{3.4}$$

where the I's are the two-dimensional identity. It is easy to show that in terms of these operators the partition function, Eq. (2.10), can be written as

$$Q_N = A_N \sum_{\alpha=1}^{2^N} \psi^{\alpha\dagger} S \psi^\alpha = A_N T_r(S), \tag{3.5}$$

where

$$S = S_1 S_2,$$

$$S_1 = \prod_{r=1}^{N} S_{1r} = \prod_{r=1}^{N} \exp\left\{\sum_{s=1}^{N} Z_r Z_{r+s}\theta_s\right\},$$

$$S_2 = \prod_{r=1}^{N} S_{2r} = \prod_{r=1}^{N} \exp\{\nu Z_r\}.$$

The partition functions for small systems are readily computed from Eq. (3.5). Such partition functions are tabulated in Appendix A. In attempting a closed-form solution of the problem it seems natural to follow the approach of Kaufman and Onsager,[7,8] setting $S_2 = 1$ initially. It is obvious that this will not be easy since the preceding analysis shows that the general one-dimensional system considered here contains the unsolved three-dimensional Ising model as a particular subset. The analysis is outlined in Appendix B and shows that the operator S_1 is not the spin representative of commuting rotations, as it is for a two-dimensional square lattice, but is rather part of a tensor transformation. It does not appear feasible therefore to determine the eigenvalues of S in this way.

4. EVALUATION OF THE PARTITION FUNCTION

In the partition function, Eq. (3.5), consider the exponent

$$U = \sum_{r=1}^{N}\sum_{s=1}^{N} Z_r Z_{r+s}\theta_s + \nu\sum_{r=1}^{N} Z_r.$$

This can be written in the form $U = Y^\dagger \Phi Y + \Phi_0$, where $Y = \zeta - \zeta_0$, $\zeta = |Z_1, Z_2, \cdots, Z_N\rangle$, ζ_0 is a constant vector yet to be determined, Φ_0 is a scalar, also as yet undetermined, and Φ is a doubly-cyclic matrix[16]:

$$\Phi = \begin{pmatrix} 0 & \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_3 & \theta_2 & \theta_1 \\ \theta_1 & 0 & \theta_1 & \theta_2 & \cdots & \theta_4 & \theta_3 & \theta_2 \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \theta_1 & \theta_2 & \theta_3 & \theta_4 & \cdots & \theta_2 & \theta_1 & 0 \end{pmatrix}; \tag{4.1}$$

i.e., $\Phi_{ij} = \theta_{j-i}$ (we must remember that $\theta_{N-r} = \theta_r$). One finds readily that V_k, the kth normalized eigenvector of Φ, is given by $V_k = 1/N^{\frac{1}{2}}\omega_{k(l-1)}$, where $\omega_{kl} = \exp\{2\pi kli/N\}$, while the kth eigenvalue is

$$\lambda_k = \sum_{r=1}^{N} \cos[(2\pi/N)kr]\theta_r.$$

The λ's are thus the finite Fourier cosine transforms of the θ potentials. The matrix Φ being symmetric, its eigenvectors form, of course, an orthonormal set. We now make the following identifications:

$$\zeta^\dagger\Phi\zeta = \sum_{r=1}^{N}\sum_{s=1}^{N} Z_r Z_{r+s}\theta_s, \quad \zeta^\dagger\Phi\zeta_0 + \zeta_0^\dagger\Phi\zeta = -\nu\sum_{r=1}^{N} Z_r, \tag{4.2}$$

$$\Phi_0 = -\zeta_0^\dagger\Phi\zeta_0.$$

[15] S. Fluegge, *Handbuch der Physik* (Springer-Verlag, Berlin, 1962), Vol. XIII.

[16] The symbols $\Phi, \Phi_0, \zeta, \zeta_0$, and V_k have the meaning defined here only in the present section and must not be confused with the corresponding symbols used elsewhere in this paper.

One shows easily that $\zeta_0 = aV_N$, while

$$a = -N^{\frac{1}{2}}\nu/2\lambda_N.$$

The partition function then takes the form

$$Q_N = B_N \operatorname{Tr} [\exp \{Y^\dagger \Phi Y\},$$
$$B_N = A_N \exp \{-a^2\lambda_N\}]. \tag{4.3}$$

The quadratic form $Y^\dagger \Phi Y$ can be diagonalized in the usual way by letting $Y = S\eta$, where S is an $N \times N$ matrix whose columns are the eigenvectors of Φ. Then we have that

$$Y^\dagger \Phi Y = \sum_{r=1}^{N} |\eta_r|^2 \lambda_r.$$

Remembering that the η's are operators, we determine their eigenvalues from the following equation:

$$\begin{aligned}\eta_r \psi^\alpha &= \left[1/N^{\frac{1}{2}} \sum_{s=1}^{N} \omega_{-r(s-1)} Z_s - a\delta_{r,N}\right]\psi^\alpha \\ &= \left[1/N^{\frac{1}{2}} \sum_{s=1}^{N} \omega_{-r(s-1)} \sigma_s^\alpha - a\delta_{r,N}\right]\psi^\alpha \\ &= \chi_r^\alpha \psi^\alpha. \end{aligned} \tag{4.4}$$

The partition function can then be written as follows:

$$\begin{aligned} Q_N &= B_N \operatorname{Tr} \left[\exp \left\{\sum_{r=1}^{N} |\eta_r|^2 \lambda_r\right\}\right] \\ &= B_N \sum_{\{C_i\}} g(C_i) \exp \left\{\sum_{r=1}^{N} |\chi_r(C_i)|^2 \lambda_r\right\}. \end{aligned} \tag{4.5}$$

The sum in the last equation is over all macroscopic configurations, $g(C_i)$ being the degeneracy of the ith such configuration, i.e., the number of microstates corresponding to the ith macrostates. It proves most convenient to choose as the macroscopic configurations appearing in the sum the number of particles (n) in the system. The sum can then be carried out in terms of successive deviations from the perfectly ordered state. One readily shows that $|\chi_r|^2$ is given by

$$|\chi_r(n)|^2 = \begin{cases} 4/N \sum_{i=1}^{n} \sum_{j=1}^{n} \cos [2\pi/Nr(q_i - q_j)] & (r \neq N), \\ (N + aN^{\frac{1}{2}} - 2n)^2/N & (r = N), \end{cases} \tag{4.6}$$

where it was assumed that the occupied cells are cells q_1 through q_n. One finds that this expression is invariant to the substitution $n \to N - n$; i.e., Eq. (4.6) is the same for a system of n occupied or n empty cells, $|\chi_r(n)|^2 = |\chi_r(N-n)|^2$. Before the sum in Eq. (4.5) can be carried out, a consistency condition must be satisfied. This results from the fact that

$$\psi^{\alpha\dagger} \psi^\alpha = \sum_{r=1}^{N} (\sigma_r^\alpha)^2 = N.$$

The consistency condition here is

$$\chi^\dagger \chi = N - 4an/N^{\frac{1}{2}} + 2aN^{\frac{1}{2}} + a^2.$$

One can give this condition a geometrical interpretation. The original sum in the partition function in terms of the σ's can be thought of as a sum over the corners of an N-dimensional cube having sides two units in length. The transformation used above to diagonalize the quadratic form $Y^\dagger \Phi Y$ rotates the coordinate system and translates the origin. The consistency condition then states that distances must remain invariant under this transformation. One readily verifies that every term in the expansion of Eq. (4.5) satisfies this condition. Carrying out the sum in (4.5) in terms of successive deviations from the perfectly ordered state, after some simplification and rearrangement one obtains

$$Q_N = 2C_N \sum_{n=0}^{[N/2]} T_n, \quad C_N = \begin{cases} \exp \{N\nu\} & \text{(fluid)}, \\ \exp \{N\Theta\} & \text{(ferromagnet)}, \end{cases} \tag{4.7}$$

$$T_n = \cosh [(N - 2n)\nu] X^{-n/2} t_n (1 - \tfrac{1}{2}\delta_{n,N/2}),$$
$$t_n = (1/n!) \sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' \prod_{r=1}^{n} \prod_{s=1}^{n} x_{(q_r - q_s)},$$
$$x_r = \exp \{8\theta_r\}, \quad X = \prod_{r=1}^{N} x_r.$$

The primed sums indicate that no index may be repeated in any one term, while $[m]$ means "nearest integer equal to or less than m."

5. CONVERGENCE AND ANALYTICITY

A. Convergence

As long as N is finite, the series in Eq. (4.7) obviously converges, so that the question of convergence is of interest only in the thermodynamic limit, i.e., as $N \to \infty$. Convergence of the series in that limit depends on the sign of the exponent

$$4\left[\sum_{r=1}^{n} \sum_{s=1}^{n} \theta_{(q_r - q_s)} - n\Theta\right] = 4\left[\sum_{r=1}^{n} \sum_{s=1}^{n} \theta_{(q_r - q_s)} - n \sum_{r=1}^{N} \theta_r\right]. \tag{5.1}$$

Remembering that in the limit $N \to \infty$ symmetries due to boundary conditions disappear, we find that none of the θ's in the double sum occur with a multiplicity greater than $(n - 1)$. Hence for any attractive interaction with hard core (not necessarily a Lennard-Jones potential) the expression in (5.1) is less than zero (since $\theta_r \geq 0$, $\forall r$). For example, for a potential of constant depth ϵ, expression (5.1) is

$4n(n-N)\epsilon \leq 0$, since $\epsilon \geq 0$ and $0 \leq n \leq N/2$. Thus we have

$$0 \leq \cosh(N\nu) + \sum_{n=1}^{[N/2]} 1/n! \cosh[(N-2n)\nu] \times \sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' \exp\left\{4\left[\sum_{r=1}^{n}\sum_{s=1}^{n}\theta_{q_r-q_s} - n\Theta\right]\right\}(1-\tfrac{1}{2}\delta_{n,N/2}) \leq \tfrac{1}{2}\exp\{N\nu\}\sum_{n=0}^{N}\binom{N}{n}\exp\{-2n\nu\} = 2^{N-1}\cosh^N\nu. \tag{5.2}$$

Hence

$$0 \leq Q_N(\nu,\theta) \leq 2^N C_N \cosh^N\nu = \begin{cases}(1+\exp\{2\nu\})^N & \text{(fluid)},\\ 2\exp\{N\Theta\}\cosh^N\nu & \text{(ferromagnet)}.\end{cases} \tag{5.3}$$

Consequently,

$$0 \leq \lim_{N\to\infty} Q_N^{1/N} \leq \begin{cases}(1+\exp\{2\nu\}) & \text{(fluid)},\\ 2\exp\{\Theta\}\cosh\nu & \text{(ferromagnet)}.\end{cases} \tag{5.4}$$

The series in Eq. (4.7) therefore has the following properties:

(1) $\lim_{N\to\infty} Q_N^{1/N}$ exists;

(2) $q = \lim_{N\to\infty}(1/N)\ln(Q_N)$ exists;

(3) The series representing these functions are absolutely and uniformly convergent for all finite values of the parameter ν.

B. Analyticity

The partition function of Eq. (4.7) can be written in the form

$$Q_N = 2C_N\sum_{n=0}^{[N/2]} a_n\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}) = C_N\begin{cases}y^{-N/2}P_N(\Theta,y) & (y\leq 1),\\ y^{N/2}P_N(\Theta,y^{-1}) & (y>1),\end{cases} \tag{5.5}$$

where

$$P_N(\Theta,y) = \sum_{n=0}^{N} a_n y^n, \quad y = \exp\{2\nu\},$$

$$a_n = (1/n!)X^{-n/2}\sum_{q_1=1}^{N}{}'\cdots\sum_{q_n=1}^{N}{}'\prod_{\substack{r=1\\(r<s)}}^{n}\prod_{s=1}^{n} x_{q_r-q_s}.$$

It is readily shown that the coefficients have the following properties:

$$a_n = a_n^*, \quad a_n \leq \binom{N}{n}, \quad a_n = a_{N-n}, \quad a_n \geq 0, \quad a_0 = 1. \tag{5.6}$$

Since $C_N = y^{N/2}$ for a fluid, we see that the partition function is analytic throughout the entire finite complex y plane, but has a pole of order N at the ideal point at infinity. This agrees with physical reasoning, since the point $y = 0$ corresponds to a system of zero density, while $y \to \infty$ represents a system in which every cell is filled, i.e., of density $\rho = 1$, where ρ is defined as[17]

$$\rho = \langle n\rangle/N. \tag{5.7}$$

Conversely, a ferromagnet has poles of order $N/2$ at the origin and at infinity, corresponding to a diverging free energy at these points. This behavior is again what one would expect on physical grounds since these points represent an infinitely large external magnetic field aligned with the negative and positive z axis, respectively.

In view of the connection established by Yang and Lee between phase changes and zeros of the partition function,[5] the distribution of these zeros is of particular interest. One concludes from Eq. (5.5) that the zeros of Q_N coincide with those of the polynomial P_N. It is apparent from Eq. (5.5) and the properties of the coefficients in Eq. (5.6) that the roots occur in complex conjugate pairs and inverse pairs; i.e., if y_k is such a root, then y_k^*, y_k^{-1}, and $(y_k^*)^{-1}$ are roots also. Furthermore, no roots can lie on the positive real axis for any finite N. Using Yang and Lee's result that $|y_k| \leq 1$, it follows at once that all roots must lie on the unit circle in the complex y plane. Furthermore, if N is odd, at least one of the roots must lie at $y = -1$.

The analysis can be carried considerably further; this will be done in a later publication.

6. NEAREST-NEIGHBOR APPROXIMATIONS

In this section the series form of the partition function in Eq. (4.7) will be applied to one-, two-, and three-dimensional systems with nearest-neighbor interactions. In the one-dimensional case the solution so obtained can be checked against the known closed-form solution to establish the validity of the present approach and to develop the algebraic machinery for two- and three-dimensional low-temperature series approximations.

[17] It will be shown in the second of this series [J. Math. Phys. **9**, 1957 (1968), following paper] that the density ρ is given by $\rho = (\tfrac{1}{2})(1 + T_\nu/NT)$, where

$$T_\nu = \sum_{n=0}^{[N/2]}(N-2n)\cdot\sinh[(N-2n)\nu]X^{-n/2}t_n(1-\tfrac{1}{2}\delta_{n,N/2})$$

and

$$T = \sum_{n=0}^{[N/2]}\cosh[(N-2n)\nu]\cdot X^{-n/2}t_n(1-\tfrac{1}{2}\delta_{n,N/2}).$$

Hence for $\nu \to -\infty$ or $y = 0$, $\rho = 0$; while for $\nu \to +\infty$ or $\nu \to +\infty$, $\rho = 1$.

A. One-Dimensional System

The potential to be used is

$$V_r = \begin{cases} -\zeta & (r = 1), \\ 0 & (r \neq 1). \end{cases} \tag{6.1}$$

Then

$$X_r = \exp\{4\epsilon(\delta_{r,1} + \delta_{r,N-1})\} = 1 + a\Delta_r, \tag{6.2}$$

where

$$\epsilon = 2\theta_1 = \tfrac{1}{4}\beta\zeta, \quad a = X_1 - 1,$$
$$\Delta_r = \delta_{r,1} + \delta_{r,N-1} \text{ (modulo } N).$$

We can now evaluate the partition function term by term. Thus [cf. Eq. (4.7)]

$$t_0 = 1, \quad t_1 = \sum_{q=1}^{N} \cdot 1 = N, \tag{6.3}$$
$$t_2 = \tfrac{1}{2}\Sigma'_{q_1q_2}x_{12} = \tfrac{1}{2}\{\Sigma_{q_1q_2}x_{12} - \Sigma_{\underline{q_1q_2}}x_{12}\}$$
$$= \tfrac{1}{2}\{\Sigma_{q_1q_2}(1 + a\Delta_{12}) - N\}$$
$$= \tfrac{1}{2}(N^2 + 2aN - N)$$
$$= (N/2)(N + 2x_1 - 3),$$

where the following abbreviated notation was used:

$$\Sigma_{q_1 \cdots q_n} = \sum_{q_1=1}^{N} \cdots \sum_{q_n=1}^{N}, \quad x_{ij} = x_{q_i - q_j}, \quad \Delta_{ij} = \Delta_{q_i - q_j} \tag{6.4}$$

and

$$\Sigma_{\underline{q_1q_2}}x_{12} = \Sigma_{q_1q_2}x_{12}\delta_{q_1,q_2}.$$

For the next term in the expansion we find

$$t_3 = 1/(3!)\,\Sigma'_{q_1q_2q_3}x_{12}x_{13}x_{23}$$
$$= 1/(3!)\,\Sigma_{q_1q_2q_3}(1 + a\Delta_{12})(1 + a\Delta_{13})\cdot(1 + a\Delta_{23}) \times (1 - \delta_{q_1,q_2})(1 - \delta_{q_1,q_3})(1 - \delta_{q_2,q_3})$$
$$= 1/(3!)\,\{\Sigma_{q_1q_2q_3} - 3\Sigma_{\underline{q_1q_2}q_3} + 2\Sigma_{\underline{q_1q_2q_3}}\}(1 + a\Delta_{12}) \times (1 + a\Delta_{13})(1 + a\Delta_{23})$$
$$= (N/3!)[6x_1^2 + 6(N - 4)x_1 + (N - 4)(N - 5)]. \tag{6.5}$$

In arriving at this result the following was used:

$$\Sigma_{q_1q_2q_3}\Delta_{12}\Delta_{13}\Delta_{23} = 0, \tag{6.6}$$

which follows directly if one notes that all δ-function products occurring in this sum are incompatible. If one associates with each Δ function the corresponding Boltzmann factor x_1, or, equivalently, the corresponding a factor, one can think of the function Δ_{ij} as a bond between cells q_i and q_j. From that point of view, which will be useful in evaluating higher terms, one can say that any triangular configuration of bonds makes a vanishing contribution to the partition function. This result is readily generalized: Any configuration of bonds which includes a closed multilateral with an odd number of sides makes a vanishing contribution to the partition function.

In expanding the primed sums in t_4, one obtains

$$t_4 = (1/4!)\Sigma'_{q_1q_2q_3q_4}x_{12}x_{13}x_{14}x_{23}x_{24}x_{34}$$
$$= (1/4!)\{\Sigma_{q_1q_2q_3q_4} - 6\Sigma_{\underline{q_1q_2}q_3q_4} + 8\Sigma_{\underline{q_1q_2q_3}q_4} + 3\Sigma_{\underline{q_1q_2}\,\underline{q_3q_4}} - 6\Sigma_{\underline{q_1q_2q_3q_4}}\}x_{12}x_{13}x_{14}x_{23}x_{24}$$
$$= (1/4!)\Big\{\Sigma_{q_1q_2q_3q_4}\prod_{\substack{r=1\\s=1\\(r<s)}}^{4}(1 + a\Delta_{rs}) - 6\Sigma_{q_1q_2q_3}(1 + a\Delta_{12})^2(1 + a\Delta_{13})^2\cdot(1 + a\Delta_{23}) + \Sigma_{q_1q_2}[8(1 + a\Delta_{12})^3 + 3(1 + a\Delta_{12})^4] - 6N\Big\}. \tag{6.7}$$

The sums are easily evaluated except for the first one where the algebra becomes somewhat tedious. It can be simplified considerably by determining the number of ways of distributing one, two, . . . , six bonds among four cells. In this way, for the entire term one has

$$t_4 = (N/4!)[24x_1^3 + 36(N - 5)x_1^2 + 12(N - 5)(N - 6)x_1 + (N - 5)(N - 6) \times (N - 7)]. \tag{6.8}$$

In evaluating t_5 the following result is helpful:

$$\sum_{k=1}^{m}\binom{m}{k}a^k = \sum_{k=1}^{m}\sum_{l=0}^{k}\binom{m}{k}\binom{k}{l}(-1)^{k-l}x_1^l = x_1^m - 1. \tag{6.9}$$

Expansion of the primed sums yields

$$t_5 = 1/(5!)\Sigma'_{q_1 \cdots q_5}\prod_{\substack{r=1\\s=1\\(r<s)}}^{5}x_{rs}$$
$$= 1/(5!)\Sigma_{q_1 \cdots q_5} - 10\Sigma_{\underline{q_1q_2q_3q_4q_5}} + 15\Sigma_{\underline{q_1q_2}\,\underline{q_3q_4}q_5} + 20\Sigma_{\underline{q_1q_2q_3}q_4q_5} - 20\Sigma_{\underline{q_1q_2q_3q_4}q_5} - 30\Sigma_{\underline{q_1q_2q_3q_4q_5}} + 24\Sigma_{\underline{q_1q_2q_3q_4q_5}}\Big\}\prod_{\substack{r=1\\s=1\\(r<s)}}^{5}x_{rs}$$
$$= 1/(5!)\Big\{\Sigma_{q_1 \cdots q_5}\prod_{\substack{r,s=1\\(r<s)}}^{5}(1 + a\Delta_{rs}) - 10\Sigma_{q_1q_2q_3q_4}(1 + a\Delta_{12})^2(1 + a\Delta_{13})^2 \times (1 + a\Delta_{14})^2(1 + a\Delta_{23})(1 + a\Delta_{24})(1 + a\Delta_{34}) + \Sigma_{q_1q_2q_3}[15(1 + a\Delta_{12})^4\cdot(1 + a\Delta_{13})^2(1 + a\Delta_{23})^2 + 20(1 + a\Delta_{12})^3(1 + a\Delta_{13})^3(1 + a\Delta_{23})] - \Sigma_{q_1q_2}[20(1 + a\Delta_{12})^6 + 30(1 + a\Delta_{12})^4] + 24N\Big\}. \tag{6.10}$$

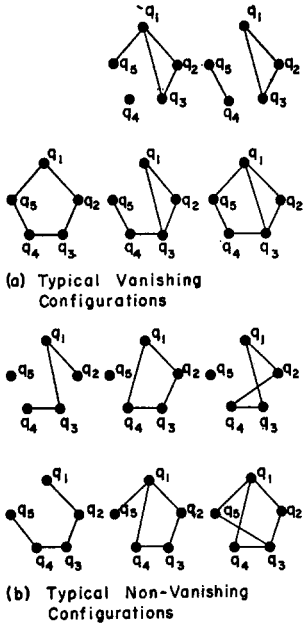


FIG. 5. Typical bond configurations for t_5.

The sums occurring here are identical with those encountered in previous terms and are evaluated in the same way except for the first one. In carrying out the first sum one observes that it can be decomposed into ten equivalent single-bond terms, 45 equivalent two-bond terms, 120 triple-bond terms of which ten include triangular bond configurations and therefore vanish, etc. Figure 5 shows some typical vanishing and nonvanishing bond configurations occurring in this sum. Reasoning in this way, one finds

$$\begin{aligned}
&\Sigma_{q_1q_2q_3q_4q_5}\prod_{\substack{r,s=1\\(r<s)}}^{5}(1+a\Delta_{rs})\\
&=\Sigma_{q_1\cdots q_5}\{1+10a\Delta_{12}+45a^2\Delta_{12}\Delta_{13}\\
&\quad+110a^3\Delta_{12}\Delta_{13}\Delta_{14}+15a^4\Delta_{12}\Delta_{23}\Delta_{34}\Delta_{14}\\
&\quad+125a^4\Delta_{12}\Delta_{23}\Delta_{34}\Delta_{45}+60a^5\Delta_{12}\Delta_{23}\Delta_{34}\\
&\quad\times\Delta_{14}\Delta_{15}+10a^6\Delta_{12}\Delta_{14}\Delta_{15}\Delta_{23}\Delta_{34}\Delta_{35}+\text{zero terms}\}\\
&=N[N^4+20N^3(x_1-1)+180N^2(x_1^2-2x_1+1)\\
&\quad+10N(9x_1^4+52x_1^3-210x_1^2+228x_1-79)\\
&\quad+20(5x_1^6+6x_1^5-5x_1^4-140x_1^3\\
&\quad+315x_1^2-250x_1+69].
\end{aligned}\tag{6.11}$$

For the entire term one finds

$$\begin{aligned}
t_5=(N/5!)[&120x_1^4+240(N-6)x_1^3\\
&+120(N-6)(N-7)x_1^2\\
&+20(N-6)(N-7)(N-8)x_1\\
&+(N-6)(N-7)(N-8)(N-9)].
\end{aligned}\tag{6.12}$$

Combining these results, for the one-dimensional partition function correct to this term one has

$$\begin{aligned}
Q_N=2C_N\{&\cosh(N\nu)+Nx_1^{-1}\cosh[(N-2)\nu]\\
&+(N/2)x_1^{-2}(2x_1+N-3)\cosh[(N-4)\nu]\\
&+(N/3!)x_1^{-3}[6x_1^2+6(N-4)x_1\\
&+(N-4)(N-5)]\cosh[(N-6)\nu]\\
&+(N/4!)x_1^{-4}[24x_1^3+36(N-5)x_1^2\\
&+12(N-5)(N-6)x_1\\
&+(N-5)(N-6)(N-7)]\cosh[(N-8)\nu]\\
&+(N/5!)x_1^{-5}[120x_1^4+240(N-6)x_1^3\\
&+120(N-6)(N-7)x_1^2\\
&+20(N-6)(N-7)(N-8)x_1\\
&+(N-6)(N-7)(N-8)(N-9)]\\
&\times\cosh[(N-10)\nu]+\cdots\}.
\end{aligned}\tag{6.13}$$

Equation (6.13) suggests strongly that one should be able to write the one-dimensional partition function in the following form:

$$Q_N=2C_N\sum_{n=0}^{[N/2]}x_1^{-n}t_n\cosh[(N-2n)\nu(1-\tfrac{1}{2}\delta_{n,N/2}),\tag{6.14}$$

where

$$t_n=\begin{cases}1 & (n=0),\\ \dfrac{N}{n}\displaystyle\sum_{k=1}^{n}\binom{n}{k}\binom{N-n-1}{k-1}x_1^{n-k} & (n\neq 0).\end{cases}$$

One can re-express t_n in the form

$$\begin{aligned}
t_n&=\frac{N}{n}\sum_{k=1}^{n}\binom{n}{k}\binom{N-n-1}{k-1}(1+a)^{n-k}\\
&=\frac{N}{n}\sum_{l=0}^{n-1}\sum_{k=1}^{n-l}\binom{n}{k}\binom{N-n-1}{k-1}\binom{n-k}{l}a^l\quad(n\neq 0).
\end{aligned}\tag{6.15}$$

Conversely, from Eqs. (4.7) and (6.2) one has that

$$t_n=(1/n!)\Sigma'_{q_1\cdots q_n}\prod_{\substack{r,s=1\\(r<s)}}^{n}(1+a\Delta_{rs})\quad(n\neq 0).\tag{6.16}$$

If Eq. (6.14) is to be correct, then these two expressions for t_n must be equal. While an explicit proof of this equality does not seem feasible, the coefficients of a were determined from these two expressions for a wide range of N and n by computer. The equality of the coefficients in all cases examined presents persuasive evidence that the formulation of the partition function in Eq. (6.14) is in fact correct.

It is possible to carry out the sum in Eq. (6.14), thus obtaining a closed-form solution for the one-dimensional nearest-neighbor model which agrees with the solution found by other methods. This analysis is carried out in Appendix C, where it is shown that

the partition function of such a system has the form

$$Q_N = C_N \cosh^N(\nu)\{(1+(1+\omega)^{\frac{1}{2}})^N + (1-(1+\omega)^{\frac{1}{2}})^N\},$$
$$C_N = \begin{cases}\exp\{N\nu\} & \text{(fluid)},\\ \exp\{N\epsilon\} & \text{(ferromagnet)},\end{cases} \qquad (6.17)$$
$$\omega = (x_1^{-1}-1)\operatorname{sech}^2\nu.$$

Agreement with the known solution of this model establishes the validity of the present approach as a preliminary to applying this method to systems of higher dimensions.[18]

B. Multi-Dimensional Systems

Considering now two- and three-dimensional systems of N cells, the appropriate potentials are those of Eqs. (2.12) and (2.13), respectively. The definitions of Eq. (6.2) then change as follows:

$$\epsilon = \tfrac{1}{4}\beta\zeta = \begin{cases}2\theta_1 = 2\theta_m = 2\theta_{N-1} = 2\theta_{N-m} & \text{(2-dim. system)},\\ 2\theta_1 = 2\theta_m = 2\theta_k = 2\theta_{N-1} = 2\theta_{N-m} = 2\theta_{N-k} & \text{(3-dim. system)},\end{cases} \qquad (6.18)$$

$$\Delta_r = \begin{cases}\delta_{r,1}+\delta_{r,m}+\delta_{r,N-1}+\delta_{r,N-m} & \text{(2-dim. system)},\\ \delta_{r,1}+\delta_{r,m}+\delta_{r,k}+\delta_{r,N-1}+\delta_{r,N-m}+\delta_{r,N-k} & \text{(3-dim. system)}.\end{cases}$$

The change in definition of the Δ functions reflects, of course, the change in coordination number of the system. One can now proceed to determine the terms in the series expansion of the partition function algebraically in exactly the same way as in the one-dimensional case. The expansions of the primed sums occurring in t_n remain unchanged, as do the coefficients (expressed in series form) of the various powers of a. Indeed, there are only two changes in this entire process. One is the changed value of X defined in Eq. (4.7):

$$X = \prod_{r=1}^{N} x_r = \prod_{r=1}^{N}(1+a\Delta_r) = \begin{cases}x_1^2 & \text{(1-dim. system)},\\ x_1^4 & \text{(2-dim. system)},\\ x_1^6 & \text{(3-dim. system)}.\end{cases} \qquad (6.19)$$

Secondly, the value of sums of Δ-function products changes because of the changed definition of Δ_r. This causes a change in the actual numerical value of the coefficients of a. Table I shows the values of typical such sums. These changes are sufficient, of course, to modify the actual terms in the series materially. As a typical example, one finds that the fourth term in the series expansion is given by

$$T_4 = \cosh[(N-8)\nu]t_4(1-\tfrac{1}{2}\delta_{4,N/2}) \times \begin{cases}x_1^{-2} & \text{(1-dim. system)},\\ x_1^{-8} & \text{(2-dim. system)},\\ x_1^{-12} & \text{(3-dim. system)},\end{cases} \qquad (6.20)$$

$$t_4 = 1/(4!)\Sigma'_{q_1q_2q_3q_4}\prod_{\substack{r,s=1\\(r<s)}}^{4}(1+a\Delta_{rs}) = \Big\{1/(4!)\Sigma_{q_1q_2q_3q_4} - 6\Sigma_{\underline{q_1q_2}q_3q_4} + 8\Sigma_{\underline{q_1q_2q_3}q_4} + 3\Sigma_{\underline{q_1q_2}\underline{q_3q_4}} - 6\Sigma_{\underline{q_1q_2q_3q_4}}\Big\}\prod_{\substack{r,s=1\\(r<s)}}^{4}(1+a\Delta_{rs})$$
$$= 1/(4!)\{\Sigma_{q_1q_2q_3q_4}[1+6a\Delta_{12}+15a^2\Delta_{12}\Delta_{13}+16a^3\Delta_{12}\Delta_{13}\Delta_{14}+4a^3\Delta_{12}\Delta_{13}\Delta_{23}+12a^4\Delta_{12}\Delta_{13}\Delta_{14}\Delta_{34}+3a^4\Delta_{12}\Delta_{23}\Delta_{34}\Delta_{14}] - 6\Sigma_{q_1q_2q_3}[1+(3+2a)a\Delta_{12}+(4+4a+a^2)a^3\Delta_{12}\Delta_{13}\Delta_{23}] + \Sigma_{q_1q_2}[11+(36+42a+20a^2+3a^3)\times a\Delta_{12}]-6N\},$$
$$= \frac{N}{4!}\begin{cases}[24x_1^3+36(N-5)x_1^2+12(N-5)(N-6)x_1+(N-5)(N-6)(N-7)] & \text{(1-dim. system)},\\ [24x_1^4+432x_1^3+24(8N-85)x_1^2+24(N^2-21N+18)x_1+N^3-30N^2+323N-1254] & \text{(2-dim. system)},\\ [72x_1^4+1992x_1^3+36(13N-219)x_1^2+36(N^2-31N+270)x_1+N^3-42N^2+659N-3906] & \text{(3-dim. system)}.\end{cases}$$

Structuring the other terms of the partition function from the one-dimensional one in the same way, one obtains for the two-dimensional series

$$Q_N = 2C_N\{\cosh(N\nu)+Nx_1^{-2}\cosh[(N-2)\nu] + (N/2!)x_1^{-4}[4x_1+(N-5)]\cosh[(N-4)\nu] + (N/3!)x_1^{-6}[36x_1^2+12(N-8)x_1+N^2-15N+62]\cosh[(N-6)\nu] + (N/4!)x_1^{-8}[24x_1^4+432x_1^3+24(8N-85)x_1^2+24(N^2-21N+18)x_1+N^3-30N^2+323N-1254]\cosh[(N-8)\nu] + (N/5!)x_1^{-10}[960x_1^5+120(N+43)x_1^4+1200(3N-40)x_1^3+120(5N^2-132N+926)x_1^2+40(N^3-39N^2+536N-2616)x_1+N^4-50N^3+995N^2-9370N+35424]\times\cosh[(N-10)\nu]+\cdots\}. \qquad (6.21)$$

[18] Further proof of the validity of the model appears in the second article, where it is shown that for a hard-rod potential the present solution reduces correctly to the Tonks equation of state.

TABLE I. Sums of Δ-function products for systems of various dimensions.

Reference figure	Sums of Δ functions	Value of sums		
		1-Dim. system	2-Dim. system	3-Dim. system
	$\Sigma_{q_1q_2}\Delta_{12}$	$2N$	$4N$	$6N$
	$\Sigma_{q_1q_2q_3}\Delta_{12}\Delta_{13}$	$4N$	$16N$	$36N$
	$\Sigma_{q_1q_1q_3}\Delta_{12}\Delta_{13}\Delta_{23}$	0	0	0
	$\Sigma_{q_1q_2q_3q_4}\Delta_{12}\Delta_{13}\Delta_{14}$	$8N$	$64N$	$216N$
	$\Sigma_{q_1q_2q_3q_4}\Delta_{12}\Delta_{23}\Delta_{34}\Delta_{14}$	$6N$	$36N$	$90N$
	$\Sigma_{q_1q_2q_3q_4q_5}\Delta_{12}\Delta_{23}\Delta_{34}\Delta_{45}\Delta_{15}$	0	0	0
	$\Sigma_{q_1q_2q_3q_4q_5}\Delta_{12}\Delta_{14}\Delta_{15}\Delta_{23}\Delta_{34}\Delta_{35}$	$10N$	$100N$	$318N$

For the corresponding three-dimensional series one has

$$\begin{aligned}Q_N = {} & 2C_N\{\cosh(N\nu) + Nx_1^{-3}\cosh[(N-2)\nu] \\ & + (N/2!)x_1^{-6}[6x_1 + (N-7)]\cosh[(N-4)\nu] \\ & + (N/3!)x_1^{-9}[90x_1^2 + 18(N-12)x_1 \\ & + N^2 - 21N + 128]\cosh[(N-6)\nu] \\ & + (N/4!)x_1^{-12}[72x_1^4 + 1992x_1^3 \\ & + 36(13N-219)x_1^2 + 36(N^2 - 31N + 270)x_1 \\ & + N^3 - 42N^2 + 659N - 3906]\cosh[(N-8)\nu] \\ & + (N/5!)x_1^{-15}[5760x_1^5 + 120(3N+426)x_1^4 \\ & + 480(32N-701)x_1^3 \\ & + 360(20N^2 - 163N + 1844)x_1^2 \\ & + 60(N^3 - 57N^2 + 1190N - 9216)x_1 + N^4 \\ & - 70N^3 + 2015N^2 - 28490N + 169744] \\ & \times \cosh[(N-10)\nu] + \cdots\}. \end{aligned} \tag{6.22}$$

Since one wishes eventually to go to the thermodynamic limit, it is more useful to work with a series expansion of $e^q = Q_N^{1/N}$ where q is Kramer's grand potential.[19] Using a method due to Domb,[20] one obtains the following series for a two-dimensional system from Eq. (6.21):

$$\begin{aligned}Q_N^{1/N} = {} & C_N^{1/N}y^{-\frac{1}{2}}\{1 + x_1^{-2}y + 2(x_1^{-3} - x_1^{-4})y^2 \\ & + 2(3x_1^{-4} - 7x_1^{-5} + 4x_1^{-6})y^3 \\ & + (x_1^{-4} + 18x_1^{-5} - 77x_1^{-6} + 98x_1^{-7} - 40x_1^{-8})y_4 \\ & + (8x_1^{-5} + 44x_1^{-6} - 370x_1^{-7} + 799x_1^{-8} \\ & - 706x_1^{-9} + 225x_1^{-10})y^5 + \cdots\}. \end{aligned} \tag{6.23}$$

For a three-dimensional system from Eq. (6.22) in the same way one has

$$\begin{aligned}Q_N^{1/N} = {} & C_N^{1/N}y^{-\frac{1}{2}}\{1 + x_1^{-3}y + 3(x_1^{-5} - x_1^{-6})y^2 \\ & + 3(5x_1^{-7} - 11x_1^{-8} + 6x_1^{-9})y^3 \\ & + (3x_1^{-8} + 83x_1^{-9} - 309x_1^{-10} \\ & + 360x_1^{-11} - 137x_1^{-12})y^4 \\ & + (48x_1^{-10} + 429x_1^{-11} - 2676x_1^{-12} + 5055x_1^{-13} \\ & - 4041x_1^{-14} + 1185x_1^{-15})y^5 + \cdots\}. \end{aligned} \tag{6.24}$$

These series are valid for $y \leq 1$. Corresponding series for $y > 1$ can be obtained by the simple transformation of Eq. (5.5). It is also possible to obtain from (6.21) or (6.22) a series for the grand potential. Thus, for the three-dimensional system one has

$$\begin{aligned}q = {} & \beta\rho\ell \\ = {} & \lim_{N\to\infty}(1/N)\ln Q_N \\ = {} & 1 + x_1^{-3}y + (3x_1^{-5} - \tfrac{7}{2}x_1^{-6})y^2 \\ & + (15x_1^{-7} - 36x_1^{-8} + 21\tfrac{1}{8}x_1^{-9})y^3 \\ & + (3x_1^{-8} + 83x_1^{-9} - 328\tfrac{1}{2}x_1^{-10} \\ & + 405x_1^{-11} - 162\tfrac{3}{4}x_1^{-12})y^4 \\ & + (48x_1^{-10} + 426x_1^{-11} - 2804x_1^{-12} + 5532x_1^{-13} \\ & + 4608x_1^{-14} + 1406\tfrac{1}{5}x_1^{-15})y^5 + \cdots. \end{aligned} \tag{6.25}$$

As mentioned in Sec. 4, these series represent successive deviations from the perfectly ordered state and hence are in fact low-temperature expansions. Comparison with similar series computed by others using entirely different techniques shows complete agreement insofar as the terms overlap and after making allowance for differences in notation.[20–22] This serves to confirm the validity of the transformation discussed

[19] D. ter Haar, *Elements of Statistical Mechanics* (Holt, Rinehart and Winston, New York, 1964).
[20] C. Domb, Adv. Phys. 9, 149 (1960).
[21] A. J. Wakefield, Proc. Cambridge Phil. Soc. 47, 419, 799 (1951).
[22] S. G. Brush, "History of the Lenz–Ising Model," University of California, Lawrence Radiation Laboratory Report UCRL-7940, 1964.

in Sec. 2, whereby a one-dimensional system with all interactions active can be converted to a system of higher dimensionality with a more limited number of interparticle bonds. It is not necessary to restrict the bonds to nearest-neighbor ones; additional bonds can readily be considered as long as their number remains small enough so that the sums of Δ functions are manageable. The two- and three-dimensional series, it will be observed, were obtained by the same algebraic technique as the one-dimensional one. Consequently, once a term in the one-dimensional series is known, the corresponding term for series in higher dimensions can be found without much additional effort. Since the technique is basically algebraic, it lends itself to programming by computer. It was pointed out that coefficients for the one-dimensional series were verified in this way. It may also prove feasible to determine expansion terms for series in higher dimensions by this method.

9. SUMMARY AND CONCLUSIONS

It was seen that the cell model underlying this study describes a simple one-dimensional fluid in which all particles interact with one another. The cells, as pointed out, serve only to specify the instantaneous position of the particles with an uncertainty equivalent to the particle dimension. Artificialities introduced by the model, in addition to this uncertainty in position, are the periodic boundary conditions and the hard-core repulsive potential that was assumed. The latter restricts cell occupancy to zero or one and hence causes the system to behave as an assembly of fermions. This permits an isomorphism to be established between cell occupation numbers and Ising spin states so that the Ising formalism can be applied to the model. In addition, the fermionlike nature of the system results in characteristic behavior, as evidenced by the primed sums and their expansion in the series formulation of the partition function and the commutation rules of the operators in Appendix B. This fermionlike behavior also underlies, of course, the success of the various operator methods and field-theoretical techniques that have been applied to the Ising model in recent years.[23-26]

We saw that in the present model the spinor-algebraic approach, which proved so successful in solving the two-dimensional Ising model, will not lead to a solution. The difficulty lies in the very much larger number of bonds per particle in our model. This results in an infinite coordination number in the thermodynamic limit. The operators of interest are therefore not the spin representatives of plane rotations but turn out to be parts of tensor transformations.

It proved possible, however, to diagonalize the operator, thus permitting the solution to be written as an infinite series. We saw that the series converges absolutely for all densities and offers a convenient method of examining the analyticity of the partition function. The results of this analysis agree with those of Yang and Lee.

In applying the solution to a one-dimensional system in which only nearest neighbors interact, we were able to verify the validity of the approach and develop a straightforward algebraic technique which proved very convenient for obtaining low-temperature series for two- and three-dimensional systems. These series were seen to agree with those obtained by other authors using entirely different methods. It is rather remarkable that the simple scheme of applying the potentials in Eqs. (2.12)–(2.14) will convert a one-dimensional system into one of higher dimension but with a reduced number of bonds per particle. Looked at from that point of view, the unsolved three-dimensional Ising model is really a subset of the more general one-dimensional system that forms the central problem of the present series of papers. If we can find a closed-form solution to the latter, then we have automatically solved the three-dimensional Ising model as well. The present analysis also sheds some interesting light on the reason why a two-dimensional array with nearest-neighbor and next-nearest-neighbor bonds should pose the same level of difficulty as the three-dimensional Ising problem. From Eqs. (2.13) and (2.14), we see that these two systems are mathematically completely isomorphic.

In the second of this series of two papers the solution will be applied to certain potentials for which the series can be summed in closed form. We shall see that for a system of point particles with no interaction potential the model correctly reproduces the equation of state of an ideal gas. If, on the other hand, we retain the repulsive core and consider a system of hard rods, we obtain the well-known Tonks equation of state. This solution will be seen to hold in one, two, and three dimensions. A one-dimensional fluid with only nearest-neighbor interactions is then examined briefly and is seen to have interesting thermodynamic properties but exhibits no phase transition, of course. If we let the range parameter in the

[23] H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience Publishers, New York, 1964).
[24] I. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).
[25] R. W. Gifford and R. A. Hurst, J. Math. Phys. **7**, 305 (1966); **8**, 1427 (1967).
[26] L. P. Kadanoff, Nuovo Cimento **44**, 276 (1966).

Lennard–Jones potential go to zero, we shall find that the system does experience a phase transition. The thermodynamic properties of this system are examined by computer for various volumes and are then examined analytically. In the thermodynamic limit the transition temperature goes to infinity so that a change of phase can occur at any finite temperature. Finally, finite systems of different sizes interacting with the unmodified potential are examined by computer. They are seen to exhibit very realistic thermodynamic behavior and give a number of indications of an incipient phase transitions. One of the most cogent of these is the behavior of the pair-correlation functions. There is obvious long-range order up to a certain temperature which coincides with the temperature at which the specific-heat curve has a maximum. For higher temperatures long-range order changes very rapidly to short-range order. Two- and three-dimensional systems with nearest-neighbor interactions are also examined numerically and are seen to have an incipient change of phase at temperatures that agree with those of the corresponding Ising model within quite close bounds.

APPENDIX A: PARTITION FUNCTIONS OF SMALL SYSTEMS

The following partition functions were computed directly from Eq. (3.5) of the text. All are written for a fluid, the symbols having the meaning defined in Sec. 2. To obtain the corresponding ferromagnetic partition function, it is only necessary to premultiply the expression given by the factor $\exp\{N(\Theta - \nu)\}$ and reinterpret the symbols as follows:

$$x_r = \exp\{8\theta_r\}, \qquad \theta_r = -\beta\phi_r,$$
$$\phi_r = \tfrac{1}{2}(J_r + J_{N-r}), \qquad \nu = \beta mB, \tag{A1}$$

where J_r is the exchange energy of two spins separated by r lattice spacings, m is the magnetic moment per spin, while B is the external magnetic field. Then we have

$$Q_2 = 2\exp(2\nu)\{\cosh(2\nu) + x_1^{-\frac{1}{2}}\}, \tag{A2}$$

$$Q_3 = 2\exp(3\nu)\{\cosh(3\nu) + 3x_1^{-1}\cosh(\nu)\}, \tag{A3}$$

$$Q_4 = 2\exp(4\nu)\{\cosh(4\nu) + 4x_1^{-1}x_2^{-\frac{1}{2}}\cosh(2\nu) + 2x_1^{-1}x_2^{-1} + x_1^{-2}\}, \tag{A4}$$

$$Q_5 = 2\exp(5\nu)\{\cosh(5\nu) + 5(x_1x_2)^{-1}\cosh(3\nu) + 5(x_1x_2)^{-2}(x_1 + x_2)\cosh(\nu)\}, \tag{A5}$$

$$Q_6 = 2\exp(6\nu)\{\cosh(6\nu) + 6(x_1x_2x_3^{\frac{1}{2}})^{-1}\cosh(4\nu) + 3(x_1x_2x_3^{\frac{1}{2}})^{-2}(2x_1 + 2x_2 + x_3)\cosh(2\nu) + (x_1x_2x_3^{\frac{1}{2}})^{-3}(6x_1x_2x_3 + 3x_1^2x_2 + x_2^3)\}, \tag{A6}$$

$$Q_7 = 2\exp(7\nu)\{\cosh(7\nu) + 7(x_1x_2x_3)^{-1}\cosh(5\nu) + 7(x_1x_2x_3)^{-2}(x_1 + x_2 + x_3)\cosh(3\nu) + 7(x_1x_2x_3)^{-3}(x_1^2x_2 + 2x_1x_2x_3 + x_1x_3^2 + x_2^2x_3)\cosh(\nu)\}, \tag{A7}$$

$$Q_8 = 2\exp(8\nu)\{\cosh(8\nu) + 8X^{-1}\cosh(6\nu) + 4X^{-2}(2x_1 + 2x_2 + 2x_3 + x_4)\cosh(4\nu) + 8X^{-3}(x_1^2x_2 + x_2^2x_4 + x_2x_3^2 + 2x_1x_2x_3 + 2x_1x_3x_4)\cosh(2\nu) + X^{-4}(4x_1^3x_2^2x_3 + 8x_1^2x_2^2x_3x_4 + 8x_1^2x_2x_3^2x_4 + 8x_1x_2^2x_3^2x_4 + 4x_1x_2^2x_3^3 + 2x_1^2x_3^2x_4^2 + x_2^4x_4^2)\}, \tag{A8}$$

where $X = x_1x_2x_3x_4^{\frac{1}{2}}$, and

$$Q_9 = 2\exp(9\nu)\{\cosh(9\nu) + 9X^{-1}\cosh(7\nu) + 9X^{-2}(x_1 + x_2 + x_3 + x_4)\cosh(5\nu) + 9X^{-3}(x_1^2x_2 + 2x_1x_2x_3 + 2x_1x_3x_4 + x_2^2x_4 + x_1x_4^2 + 2x_2x_3x_4 + \tfrac{1}{3}x_3^3)\cosh(3\nu) + 9X^{-4}(x_1^3x_2^2x_3 + 2x_1^2x_2^2x_3x_4 + x_1^2x_2x_3^2x_4 + 2x_1^2x_2x_3x_4^2 + 2x_1x_2^2x_3x_4^2 + x_1^2x_3x_4^3 + 2x_1x_2x_3^3x_4 + x_1x_2x_3^2x_4^2 + x_1x_2^2x_3^2x_4 + x_2^3x_3x_4^2)\cosh(\nu)\}, \tag{A9}$$

where $X = x_1x_2x_3x_4$.

APPENDIX B: SPINOR-ALGEBRAIC FORMULATION

Starting from Eq. (3.5) in the text, this appendix shows that the operator S is not the spin representative of a set of commuting plane rotations, but is part of a tensor transformation. Hence it does not appear feasible to determine its eigenvalues by the spinor-algebraic approach. In order to reduce the problem to its simplest terms, we shall consider only the case $\nu = 0$, i.e., $S_2 = 1$. We are then concerned with the operator

$$S_1 = \prod_{r=1}^{N}\prod_{s=1}^{N} D_{rs}, \quad \text{where} \quad D_{rs} = \exp\{Z_rZ_{r+s}\theta_s\}.$$

To proceed we define a set of matrices that are the generators of a complete set of linearly independent operators that span the space of the operator S and obey the following anticommutation rule:

$$[\Gamma_\alpha, \Gamma_\beta]_+ = 2\delta_{\alpha,\beta}. \tag{B1}$$

In the 2^N-dimensional representation that has been adopted for the Z_k [cf. Eq. (3.4)], the base matrices are the Γ matrices and all possible products of Γ matrices. It is readily shown that there are $2N$ independent

matrices of this type. We choose the following representation for the Γ's:

$$\Gamma_{2\alpha-1} = Z_1 \cdots Z_{\alpha-1} X_\alpha;\ \Gamma_{2\alpha} = Z_1 \cdots Z_{\alpha-1} Y_\alpha. \quad \text{(B2)}$$

We find then that the operator Z is given by

$$Z_\alpha = i\Gamma_{2\alpha}\Gamma_{2\alpha-1}, \quad \text{(B3)}$$

so that

$$\begin{aligned} D_{rs} &= \exp\{Z_r Z_{r+s}\theta_s\} \\ &= \exp\{-\theta_s \Gamma_{2r}\Gamma_{2r-1}\Gamma_{2(r+s)}\Gamma_{2(r+s)-1}\}. \end{aligned} \quad \text{(B4)}$$

Consider now the operator

$$D_{\mu\nu\rho\sigma} = \exp\{(\theta/2)\Gamma_\mu\Gamma_\nu\Gamma_\rho\Gamma_\sigma\}. \quad \text{(B5)}$$

One proves readily that

$$D_{\mu\nu\rho\sigma} = \cosh(\theta/2) + \Gamma_\mu\Gamma_\nu\Gamma_\rho\Gamma_\sigma \sinh(\theta/2), \quad \text{(B6)}$$

where the fact was used that $(\Gamma_\mu\Gamma_\nu\Gamma_\rho\Gamma_\sigma)^2 = 1$, provided that $\mu \neq \nu \neq \rho \neq \sigma$, which will be assumed in what follows. Let us first examine the properties of a related operator

$$G_{\mu\nu\rho\sigma} = \exp\{(\theta/2)(\Gamma_\mu\Gamma_\nu + \Gamma_\rho\Gamma_\sigma)\}. \quad \text{(B7)}$$

Observing that

$$(\Gamma_\mu\Gamma_\nu + \Gamma_\rho\Gamma_\sigma)^2 = 2(\Gamma_\mu\Gamma_\nu\Gamma_\rho\Gamma_\sigma - 1), \quad \text{(B8)}$$

we have

$$\begin{aligned} G_{\mu\nu\rho\sigma} &= \cos^2(\theta/2) + (\Gamma_\mu\Gamma_\nu + \Gamma_\rho\Gamma_\sigma)\sin(\theta/2) \\ &\quad \times \cos(\theta/2) + \Gamma_\mu\Gamma_\nu\Gamma_\rho\Gamma_\sigma \sin^2(\theta/2) \\ &= \tfrac{1}{2}\{S_{\mu\nu} + S_{\rho\sigma} + aD_{\mu\nu\rho\sigma} + b\}. \end{aligned} \quad \text{(B9)}$$

Here $\tanh\phi = 1 - \cos\theta$, $a = \operatorname{sech}\phi$, and b is related to a and ϕ by the equation

$$a^2 + b(b+2) = 0. \quad \text{(B10)}$$

The operator S in Eq. (B9) is defined as follows:

$$S_{\mu\nu}(2\theta) = \cos\theta + \Gamma_\mu\Gamma_\nu \sin\theta = \exp\{\theta\Gamma_\mu\Gamma_\nu\}. \quad \text{(B11)}$$

It is not difficult to show that this operator is the spin representative of a plane rotation in the $\mu - \nu$ plane through an angle 2θ.[3,8] We must now investigate the properties of the operator G. The analysis is simplified if we note that

$$[\Gamma_\mu\Gamma_\nu, \Gamma_\rho\Gamma_\sigma]_- = 0 \quad (\mu, \nu, \rho, \sigma \text{ distinct}), \quad \text{(B12)}$$

so that

$$\begin{aligned} G_{\mu\nu\rho\sigma} &= \exp\{(\theta/2)(\Gamma_\mu\Gamma_\nu + \Gamma_\rho\Gamma_\sigma) \\ &= \exp\{(\theta/2)\Gamma_\mu\Gamma_\nu\} \exp\{(\theta/2)\Gamma_\rho\Gamma_\sigma\} \\ &= S_{\mu\nu}(\theta)S_{\rho\sigma}(\theta) = S_{\rho\sigma}(\theta)S_{\mu\nu}(\theta). \end{aligned} \quad \text{(B13)}$$

It follows that $G^{-1}G = 1$ so that G is a unitary operator. One finds then that

$$\begin{aligned} \bar{\Gamma}_\mu\bar{\Gamma}_\nu &= G^{-1}\Gamma_\mu\Gamma_\nu G \\ &= \exp\{-(\theta/2)(\Gamma_\mu\Gamma_\nu + \Gamma_\rho\Gamma_\sigma)\}\Gamma_\mu\Gamma_\nu \\ &\quad \times \{\exp(\theta/2)(\Gamma_\mu\Gamma_\nu + \Gamma_\rho\Gamma_\sigma)\} \\ &= \Gamma_\mu\Gamma_\nu. \end{aligned} \quad \text{(B14)}$$

Similarly,

$$\begin{aligned} \bar{\Gamma}_\rho\bar{\Gamma}_\sigma &= G^{-1}\Gamma_\rho\Gamma_\sigma G = \Gamma_\rho\Gamma_\sigma, \\ \bar{\Gamma}_\alpha\bar{\Gamma}_\beta &= G^{-1}\Gamma_\alpha\Gamma_\beta G = \Gamma_\alpha\Gamma_\beta \quad (\alpha, \beta \neq \mu, \nu, \rho, \sigma). \end{aligned} \quad \text{(B15)}$$

But

$$\begin{aligned} \bar{\Gamma}_\mu\bar{\Gamma}_\rho &= G^{-1}\Gamma_\mu\Gamma_\rho G \\ &= \exp\{-(\theta/2)\Gamma_\mu\Gamma_\nu\}\Gamma_\mu \exp\{(\theta/2)\Gamma_\mu\Gamma_\nu\} \\ &\quad \times \exp\{-(\theta/2)\Gamma_\rho\Gamma_\sigma\}\Gamma_\rho \exp\{(\theta/2)\Gamma_\rho\Gamma_\sigma\} \\ &= (\Gamma_\mu\cos\theta + \Gamma_\nu\sin\theta)(\Gamma_\rho\cos\theta + \Gamma_\sigma\sin\theta) \\ &= \Gamma_\mu\Gamma_\rho\cos^2\theta + (\Gamma_\nu\Gamma_\rho + \Gamma_\mu\Gamma_\sigma)\sin\theta\cos\theta \\ &\quad + \Gamma_\nu\Gamma_\sigma\sin^2\theta. \end{aligned} \quad \text{(B16)}$$

And, similarly,

$$\begin{aligned} \bar{\Gamma}_\mu\bar{\Gamma}_\sigma = G^{-1}\Gamma_\mu\Gamma_\sigma G &= \Gamma_\mu\Gamma_\sigma\cos^2\theta + (\Gamma_\nu\Gamma_\sigma - \Gamma_\mu\Gamma_\rho) \\ &\quad \times \sin\theta\cos\theta - \Gamma_\nu\Gamma_\rho\sin^2\theta. \end{aligned} \quad \text{(B17)}$$

Other transformations follow the same pattern. Let us define a new vector X, whose $4N^2$ components are the second rank tensors $\Gamma_{\alpha\beta} = \Gamma_\alpha\Gamma_\beta$ ($\alpha, \beta = 1, 2, \cdots, 2N$). We can write these transformations symbolically as follows:

$$\bar{X} = \Omega X. \quad \text{(B18)}$$

The nonidentity elements of this transformation are

$$\begin{vmatrix} \bar{\Gamma}_\mu\bar{\Gamma}_\rho \\ \bar{\Gamma}_\mu\bar{\Gamma}_\sigma \\ \bar{\Gamma}_\nu\bar{\Gamma}_\rho \\ \bar{\Gamma}_\nu\bar{\Gamma}_\sigma \end{vmatrix} = \begin{vmatrix} \cos^2\theta & \sin\theta\cos\theta & \sin\theta\cos\theta & \sin^2\theta \\ -\sin\theta\cos\theta & \cos^2\theta & -\sin^2\theta & \sin\theta\cos\theta \\ -\sin\theta\cos\theta & -\sin^2\theta & \cos^2\theta & \sin\theta\cos\theta \\ \sin^2\theta & -\sin\theta\cos\theta & -\sin\theta\cos\theta & \cos^2\theta \end{vmatrix} \begin{vmatrix} \Gamma_\mu\Gamma_\rho \\ \Gamma_\mu\Gamma_\sigma \\ \Gamma_\nu\Gamma_\rho \\ \Gamma_\nu\Gamma_\sigma \end{vmatrix}. \quad \text{(B19)}$$

We note that

$$\Omega = \omega \otimes \omega \quad \text{(B20)}$$

and

$$X = \Gamma \otimes \Gamma,$$

where ω is a plane rotation with components

$$\begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix} \quad \text{(B21)}$$

and Γ is a vector having the $2N$ Γ matrices as its components. Thus Ω is a tensor transformation transforming Γ dyadics into themselves. Using summation covention and covariant–contravariant notation, the transformation appears as follows:

$$\Gamma'^{\mu\nu} = \Omega^{\mu\nu}_{\alpha\beta}\Gamma^{\alpha\beta}. \tag{B22}$$

It follows then that the operator G is the spin representative of a tensor transformation. D, on the other hand, is associated with the operator G through Eq. (B9) but is not itself a spin representative. Accordingly, it does not seem feasible to determine the eigenvalues of D by spinor-algebraic analysis.

APPENDIX C: CLOSED FORM SOLUTION FOR THE ONE-DIMENSIONAL MODEL WITH NEAREST-NEIGHBOR INTERACTIONS

We consider first the case $\nu = 0$. In that case Eq. (6.14) becomes

$$\begin{aligned} &Q_N(\nu = 0) \\ &= 2C_N\Big\{1 + N\sum_{n=1}^{[N/2]}\sum_{k=1}^{n}\frac{1}{k}\binom{n-1}{k-1} \\ &\quad\times\binom{N-n-1}{k-1}x_1^{-k}(1 - \tfrac{1}{2}\delta_{n,N/2})\Big\} \\ &= 2C_N\Big\{1 + N\sum_{k=1}^{[N/2]}\left(\frac{1}{2k}\right)x_1^{-k}\sum_{n=k}^{N-k}\binom{n-1}{k-1}\binom{N-n-1}{k-1}, \end{aligned} \tag{C1}$$

where the order of summation was changed and the δ-function term was eliminated by taking advantage of the symmetry of the summand with respect to the substitution $n \to N - n$. If we now use the following relationship (which follows directly from the definition of the combinatorial symbol),

$$\binom{r-s}{t} = (-1)^t\binom{s-r+t-1}{t}, \tag{C2}$$

where for $a < 0$ we define

$$\binom{a}{b} = \frac{1}{b!}a(a-1)\cdots(a-b+1),$$

we find after some algebra that

$$\begin{aligned} &\sum_{n=k}^{N-k}\binom{n-1}{k-1}\binom{N-n-1}{k-1} \\ &\qquad= \sum_{n=k}^{N-k}(-1)^{N-2k}\binom{-k}{N-k-n}\binom{-k}{n-k} \\ &\qquad= \binom{N-1}{N-2k}. \end{aligned} \tag{C3}$$

The following identity was also used:

$$\sum_{r=0}^{k}\binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}. \tag{C4}$$

Thus for the partition function we have

$$\begin{aligned} Q_N(\nu = 0) &= 2C_N(\nu = 0)\Big\{1 + \sum_{k=1}^{[N/2]}\binom{N}{2k}x_1^{-k}\Big\} \\ &= C_N\sum_{k=0}^{N}\binom{N}{k}[1 + (-1)^k]x_1^{-k/2} \\ &= C_N\{(1 + x_1^{-\frac{1}{2}})^N + (1 - x_1^{-\frac{1}{2}})^N\} \\ &= 2^N C_N\exp(-N\epsilon)(\cosh^N\epsilon + \sinh^N\epsilon). \end{aligned} \tag{C5}$$

This agrees with the well-known solution of the one-dimensional Ising model in the absence of an external magnetic field.[3,15]

When $\nu \neq 0$, the analysis becomes considerably more complex. The closed-form solution of the one-dimensional Ising model is in that case (cf. Appendix D):

$$\begin{aligned} Q_N = C_N\cosh^N(\nu)\{&[1 + (1+\omega)^{\frac{1}{2}}]^N \\ &+ [1 - (1+\omega)^{\frac{1}{2}}]^N\}, \end{aligned} \tag{C6}$$

where $\omega = (x_1^{-1} - 1)\operatorname{sech}^2(\nu)$. For the work to follow it is desirable to rewrite this by means of the binomial theorem in the form

$$\begin{aligned} Q_N = 2C_N\cosh^N(\nu)&\sum_{n=0}^{[N/2]}\sum_{k=0}^{n}\sum_{m=0}^{k}\binom{N}{2n}\binom{n}{k}\binom{k}{m} \\ &\times(-1)^{k-m}x_1^{-m}\cosh^{-2k}(\nu). \end{aligned} \tag{C7}$$

In order to show that Eq. (6.14) of the text leads to this expression, we must then establish the following equivalence:

$$\begin{aligned} &\cosh^N(\nu)\sum_{n=0}^{[N/2]}\sum_{k=0}^{n}\sum_{m=0}^{k}\binom{N}{2n}\binom{n}{k}\binom{k}{m} \\ &\qquad\times(-1)^{k-m}x_1^{-m}\cosh^{-2k}(\nu) \\ &\overset{?}{=} \cosh(N\nu) + \sum_{n=1}^{[N/2]}\sum_{k=1}^{n}\frac{N}{k}\binom{n-1}{k-1}\binom{N-n-1}{k-1}x_1^{-k} \\ &\qquad\times\cosh[(N-2n)\nu](1 - \tfrac{1}{2}\delta_{n,N/2}). \end{aligned} \tag{C8}$$

For $m = 0$ we must show that

$$\begin{aligned} &\cosh^N(\nu)\sum_{n=0}^{[N/2]}\sum_{k=0}^{n}\binom{N}{2n}\binom{n}{k}(-1)^k\cosh^{-2k}(\nu) \\ &\qquad= \cosh(N\nu). \end{aligned} \tag{C9}$$

By retracing the steps that led from Eq. (C6) to (C7), we find that this proof is trivial. Suppose now that

Eq. (C8) holds for $m = m$; i.e., suppose that

$$\sum_{n=m}^{[N/2]}\sum_{k=m}^{n}\binom{N}{2n}\binom{n}{k}\binom{k}{m}(-1)^{k-m}\cosh^{-2k}(\nu)$$
$$=\binom{N}{m}\cosh^{-N}(\nu)\sum_{n=m}^{[N/2]}\binom{n-1}{m-1}\binom{N-n-1}{m-1}$$
$$\times\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}). \quad \text{(C10)}$$

We wish to show that it holds for $m = m + 1$, so that the equivalence of the following expression is to be proved:

$$\sum_{n=m}^{[N/2]}\sum_{k=m+1}^{n}\binom{N}{2n}\binom{n}{k}\binom{k}{m+1}(-1)^{k-m-1}\cosh^{-2k}(\nu)$$
$$\stackrel{?}{=}\binom{N}{m+1}\cosh^{-N}(\nu)\sum_{n=m+1}^{[N/2]}\binom{n-1}{m}\binom{N-n-1}{m}$$
$$\times\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}). \quad \text{(C11)}$$

This can be recast in the following more convenient way:

$$\cosh^{N}(\nu)\sum_{n=m}^{[N/2]}\sum_{k=m}^{n}\binom{N}{2n}\binom{n}{k}\binom{k}{m}$$
$$\times(k-m)(-1)^{k-m-1}\cosh^{-2k}(\nu)$$
$$=\frac{N}{m^2}\sum_{n=m}^{[N/2]}\binom{n-1}{m-1}\binom{N-n-1}{m-1}(n-m)(N-n-m)$$
$$\times\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}). \quad \text{(C12)}$$

We now let

$$F(N,\nu,m)$$
$$=\sum_{n=m}^{[N/2]}\sum_{k=m}^{n}\binom{N}{2n}\binom{n}{k}\binom{k}{m}(-1)^{k-m}\cosh^{-2k}(\nu). \quad \text{(C13)}$$

Then, using the assumed identity, Eq. (C10), the equivalence to be established is

$$\sum_{n=m}^{[N/2]}\sum_{k=m}^{n}\binom{N}{2n}\binom{n}{k}\binom{k}{m}(k-m)(-1)^{k-m-1}\cosh^{-2k}(\nu)$$
$$=\tfrac{1}{2}(\coth\nu\,\partial/\partial\nu+2m)\Big[\frac{N}{m}\operatorname{sech}^{N}(\nu)$$
$$\times\sum_{n=m}^{[N/2]}\binom{n-1}{m-1}\binom{N-n-1}{m-1}$$
$$\times\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2})\Big]$$
$$=(N/2m)\operatorname{sech}^{N}(\nu)\sum_{n=m}^{[N/2]}\binom{n-1}{m-1}\binom{N-n-1}{m-1}$$
$$\times\{(N-2n)\coth(\nu)\sinh[(N-2n)\nu]$$
$$+(2m-N)\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}). \quad \text{(C14)}$$

This last equation can be re-expressed as follows:

$$\sum_{n=m}^{[N/2]}\sum_{k=m}^{n}\binom{N}{2n}\binom{n}{k}\binom{k}{m}(k-m)(-1)^{k-m-1}\cosh^{-2k}(\nu)$$
$$=(N/m)\operatorname{sech}^{N}(\nu)$$
$$\times\sum_{n=m+1}^{[N/2]}\Big\{\sum_{k=m+1}^{n}\binom{k-2}{m-1}\binom{N-k}{m-1}(N-2k+2)$$
$$-\binom{n-1}{m-1}\binom{N-n-1}{m-1}(n-m)\Big\}$$
$$\times\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}), \quad \text{(C15)}$$

where the following identities were used (which are readily established):

$$\coth(\nu)\sinh[(N-2n)\nu]$$
$$=\begin{cases}0 & (n=N/2),\\ \cosh[(N-2n)\nu] & \\ \quad+2\sum_{k=1}^{[N/2]-n}\cosh[(N-2n-2k)\nu] & \\ \quad\times(1-\tfrac{1}{2}\delta_{k,[N/2]-n}) & (n\neq N/2),\end{cases} \quad \text{(C16)}$$

and

$$\sum_{n=m}^{[N/2]}\sum_{k=1}^{[N/2]-n}\binom{n-1}{m-1}\binom{N-n-1}{m-1}(N-2n)$$
$$\times\cosh[(N-2n-2k)\nu]$$
$$\times(1-\tfrac{1}{2}\delta_{k,[N/2]-n})(1-\delta_{n,N/2})$$
$$=\sum_{n=m+1}^{[N/2]}\sum_{k=m+1}^{m}\binom{k-2}{m-1}\binom{N-k}{m-1}(N-2k-2)$$
$$\times\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}). \quad \text{(C17)}$$

To complete the proof by induction, one must show [cf. Eq. (C12)] that

$$(N/m)\sum_{n=m+1}^{[N/2]}\sum_{k=m+1}^{n}\Big\{\binom{k-2}{m-1}\binom{N-k}{m-1}(N-2k+2)$$
$$-\binom{n-1}{m-1}\binom{N-n-1}{m-1}(n-m)\Big\}$$
$$\times\cosh(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2})$$
$$=(N/m^2)\sum_{n=m}^{[N/2]}\binom{n-1}{m-1}\binom{N-n-1}{m-1}(n-m)$$
$$\times(N-n-m)\cosh[(N-2n)\nu](1-\tfrac{1}{2}\delta_{n,N/2}). \quad \text{(C18)}$$

Recognizing that ν is arbitrary, this is equivalent to

establishing the following identity:

$$\sum_{k=m+1}^{n}\left\{\binom{k-2}{m-1}\binom{N-k+1}{m}-\binom{k-1}{m}\binom{N-k}{m-1}\right\}=\binom{n-1}{m}\binom{N-n}{m}. \quad \text{(C19)}$$

For $m = 1$ the proof is trivial. The identity is readily established in general if one expands the left-hand side of (C19) as

$$\begin{aligned}&\binom{m-1}{m-1}\binom{N-m}{m}-\binom{m}{m}\binom{N-m-1}{m-1}\\&\quad+\binom{m}{m-1}\binom{N-m-1}{m}\\&\quad-\binom{m+1}{m}\binom{N-m-2}{m-1}+\cdots\\&\quad+\binom{n-2}{m-1}\binom{N-n+1}{m}-\binom{n-1}{m}\binom{N-n}{m-1}\end{aligned} \quad \text{(C20)}$$

and uses the following two relations:

$$\binom{r}{s}-\binom{r-1}{s-1}=\binom{r-1}{s}, \quad \text{(C21)}$$

$$\binom{r}{s}+\binom{r}{s-1}=\binom{r+1}{s}. \quad \text{(C22)}$$

This completes the inductive proof of the equivalence of Eq. (6.14) and (C6).

APPENDIX D: CLOSED FORM SOLUTION OF THE ONE-DIMENSIONAL SYSTEM WITH NEAREST-NEIGHBOR INTERACTIONS

From Eq. (2.10) using the potential of Eq. (6.1), we obtain

$$\begin{aligned}Q_N&=A_N\sum_{\{\sigma\}}\exp\left\{\epsilon\sum_{r=1}^{N}\sigma_r\sigma_{r+1}+\nu\sum_{r=1}^{N}\sigma_r\right\}\\&=A_N\sum_{\{\sigma\}}\langle\sigma_1|\,P\,|\sigma_2\rangle\langle\sigma_2|\,P\,|\sigma_3\rangle\cdots\langle\sigma_N|\,P\,|\sigma_1\rangle\\&=A_N T_r\{P^N\}=A_N(\lambda_+^N+\lambda_-^N),\end{aligned} \quad \text{(D1)}$$

where

$$P=\begin{vmatrix}e^{\epsilon+\nu} & e^{-\epsilon}\\ e^{-\epsilon} & e^{\epsilon-\nu}\end{vmatrix} \quad \text{(D2)}$$

and λ_+ and λ_- are the eigenvalues of P. We find readily that

$$\lambda_\pm=\exp(\epsilon)\{\cosh\nu\pm[\cosh^2\nu-2\exp(-2\epsilon)\sinh(2\epsilon)]^{\frac{1}{2}}\}. \quad \text{(D3)}$$

Hence

$$Q_N=A_N\exp(N\epsilon)\cosh^N(\nu)\{[1+(1+\omega)^{\frac{1}{2}}]^N+[1-(1+\omega)^{\frac{1}{2}}]^N\}, \quad \text{(D4)}$$

where $\omega=(x_1^{-1}-1)\operatorname{sech}^2(\nu)$ and $x_1=\exp(4\epsilon)$. And since

$$A_N=\begin{cases}N(\nu-\epsilon) & \text{(fluid)},\\ 1 & \text{(ferromagnet)},\end{cases} \quad \text{(D5)}$$

we obtain

$$Q_N=C_N\cosh^N(\nu)\{[1+(1+\omega)^{\frac{1}{2}}]^N+[1-(1+\omega)^{\frac{1}{2}}]^N\}, \quad \text{(D6)}$$

where

$$C_N=\begin{cases}\exp(N\nu) & \text{(fluid)},\\ \exp(N\epsilon) & \text{(ferromagnet)}.\end{cases} \quad \text{(D7)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Cell Model of a Fluid. II. Thermodynamic Properties of the System

RALPH G. TROSS*† AND LOUIS H. LUND
University of Missouri at Rolla, Rolla, Missouri

(Received 3 May 1968)

In Paper I of this series (preceding paper) the model was formulated mathematically and the solution derived in the form of an infinite series. The convergence and analyticity of the solution were investigated. It was then applied to one-, two-, and three-dimensional systems with nearest-neighbor interactions, and low-temperature series for these systems were obtained. In the present article the thermodynamics of the model are investigated. The series solution of the partition function derived in Paper I is applied to systems with interaction potentials of increasing complexity. The validity of the model is established by showing that in the appropriate limits it leads correctly to the ideal gas and the Tonks equation of state. It is shown that the model is capable of portraying phase transitions and gives realistic results thermodynamically. Finally various finite one-, two-, and three-dimensional systems are analyzed numerically by high-speed computer and their thermodynamic properties and pair-correlation functions are examined. Interesting conclusions emerge concerning the range of order in such systems and the probable critical temperatures.

I. INTRODUCTION

In the first[1] of this series of two articles, a one-dimensional fluid was examined, in which every particle interacts with every other one with a modified Lennard–Jones-type potential. The system was cast in the mold of a cell model which could be treated by means of the Ising formalism. We derived the partition function for such a system, showed how the partition function for two- and three-dimensional systems with a limited number of bonds per particle could be derived from the one-dimensional one and found a solution to the problem in the form of an infinite series. This series was shown to be absolutely convergent and proved convenient for investigating the analyticity of the system. The analysis confirmed the results previously obtained by Yang and Lee.[2] It was seen in I that the series could be summed for a one-dimensional system with nearest-neighbor interactions and that the resultant closed-form solution agreed with that obtainable by matrix methods. This served to establish the validity of the model in that limit. In addition, the corresponding low-temperature series for two- and three-dimensional systems with nearest-neighbor interactions were readily obtained from the one-dimensional series. Thus it proved possible to calculate series expansions in all dimensions by identical algebraic techniques.

In this article we shall investigate the thermodynamic behavior of this fluid model. First, certain potentials will be considered for which the system can be analyzed in the thermodynamic limit. If there is no interaction of any kind, we shall find that the model leads correctly to the equation of state of an ideal gas. If the hard repulsive core is retained, the fluid obeys the well-known Tonks equation of state.[3] Next a system with hard-rod repulsive and nearest-neighbor attractive interactions is examined. It has the advantage that one can write the equation of state in closed form. Aside from the absence of a phase transition the system exhibits interesting and quite realistic thermodynamic properties and has the advantage of being very tractable mathematically. Another potential which allows one to proceed to the thermodynamic limit is one in which the range parameter γ is allowed to go to zero, resulting in an infinitely long-range interaction. While this does not correspond to any real fluid, it is an interesting limit mathematically. One can show rigorously, as we shall see, that such a system exhibits phase transitions at all finite temperatures.

Although it has not been possible thus far to go to the thermodynamic limit in systems interacting with the general Lennard–Jones potential,[1] the series solution of the partition function, Eq. (4.7) of Paper I, is well suited for numerical analysis by high-speed computer. Accordingly one-, two-, and three-dimensional systems of finite size are analyzed and their thermodynamic properties determined. The results are surprisingly realistic. Although no true phase transition is encountered, as one would expect for systems of finite size, there are a number of indications that such a transition may occur in the thermodynamic limit. The most convincing of these are the conclusions on long-range order that emerge from a study of the pair-correlation functions of the system.

* Based upon a dissertation submitted to the Graduate Faculty of the University of Missouri at Rolla in partial fulfillment of the requirements for the Ph.D. degree. Work supported in part by a NASA Predoctoral Traineeship and an NSF post-doctoral grant.

† Present address: Dept. of Math., University of Ottawa, Ottawa, Canada.

[1] R. G. Tross and L. H. Lund, J. Math. Phys. **9**, 1940 (1968), hereafter referred to simply as I or Paper I.

[2] C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404, 410 (1952).

[3] L. Tonks, Phys. Rev. **50**, 955 (1936).

The point of departure for the analysis in this article is the model developed in I and the series form of the partition function in Eq. (4.7)

$$
\begin{aligned}
&Q_N = 2C_N \sum_{n=0}^{[N/2]} T_n; \quad C_N = \exp(N\nu) \\
&T_n = \cosh[(N-2n)\nu] X^{-n/2} t_n (1 - \tfrac{1}{2}\delta_{n,N/2}) \\
&t_n = \left(\frac{1}{n!}\right) \sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' \prod_{\substack{r=1 \\ (r<s)}}^{n} \prod_{s=1}^{n} x_{q_r - q_s} \\
&x_r = \exp\{8\theta_r\}; \quad X = \prod_{r=1}^{N} x_r \\
&\theta_r = -(\tfrac{1}{4})\beta\phi_r; \quad \phi_r = (\tfrac{1}{2})(V_r + V_{N-r}) \\
&V_r = -\zeta / r^{\gamma} \\
&\beta = 1/kT; \quad e^{2\nu} = (l/\lambda) \exp\left\{\beta\left(g - \sum_{r=1}^{N} \phi_r\right)\right\},
\end{aligned}
\tag{1.1}
$$

where ζ defines the depth of the potential cell and γ is a range parameter. (For further definitions see Paper I.)

II. THERMODYNAMICS OF INFINITE SYSTEMS

In this section we investigate the thermodynamic behavior of certain infinite systems, i.e., systems which are infinite in the sense that the solution holds in the thermodynamic limit, $N \to \infty$, where N is the total number of cells in the system. This limit implies that either the cell parameter $l = L/N$ goes to zero (L being the length of the one-dimensional system) or the length of the system becomes infinite. These two limits are not equivalent as we shall see.

A. Ideal Gas

We wish first to check the validity of the model in the ideal-gas limit and assume the interaction potential to be $V_r = 0$. This potential does not, however, eliminate the hard-core repulsive interaction which was built into the model by our choice of cell parameter and the introduction of an exclusion principle. We neglect this point for the moment and come back to it later. The series, Eq. (1.1), now takes the form

$$
\begin{aligned}
Q_N &= 2\exp(N\nu) \sum_{n=0}^{[N/2]} \binom{N}{n} \\
&\quad \times \cosh[(N-2n)\nu](1 - \tfrac{1}{2}\delta_{n,N/2}) \\
&= \sum_{n=0}^{N} \binom{N}{n} \exp(2n\nu) = \left(1 + \frac{V\xi}{\lambda N}\right)^N,
\end{aligned}
\tag{2.1}
$$

where we have used the obvious identity

$$
t_n = \left(\frac{1}{n!}\right) \sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' 1 = \binom{N}{n}.
$$

In Eq. (2.1), V is the "volume" of the system. For the basic one-dimensional fluid the volume is, of course, the length of the line L. However, since there is no attractive interaction among the particles, we may assume the cell to be a tiny square of dimensions $l \times l$ or cube of dimensions $l \times l \times l$. The string of squares or cubes can then be wrapped around a torus as in Sec. 2B of I to generate a two- or three-dimensional system. For higher-dimensional systems we must, of course, redefine λ in an appropriate way. [We must then use the following definition: $\lambda = (\beta h^2/2\pi m)^{d/2}$, where d is the dimension of the system. This follows directly from integrating the kinetic part of the partition function.]

The hard-core repulsive potential can now be eliminated in going to the thermodynamic limit. To do so, we let the cell parameter l shrink to zero while $N \to \infty$, i.e., we keep the volume V fixed. Then

$$
Q = \lim_{N\to\infty} Q_N = \exp(V\xi/\lambda) = \exp(\beta p V). \tag{2.2}
$$

We have then that $\ln Q = \beta p V = V\xi/\lambda$. From the well-known relation $\langle n \rangle = \xi[\partial/\partial\xi(\ln Q)]_\beta$ we then obtain the equation of state of an ideal gas[4,5]

$$
\beta p V = \langle n \rangle, \tag{2.3}
$$

where $\langle\ \rangle$ designates an ensemble average or expectation value. It should be observed once more that in obtaining this equation of state from our model we were obliged to let the particles shrink to ideal points in order to eliminate the built-in repulsive potential.

B. A Fluid of Hard Rods

We next consider a fluid of hard rods, i.e., a system of particles with a hard repulsive core and no attractive interaction. The partition function, before we go to the thermodynamic limit, is then the same as that in Eq. (2.1), i.e., $Q_N = (1 + l\xi/\lambda)^N$. The hard repulsive core has, of course, not been eliminated. In order to retain the repulsive potential, we must go to the limit in a different way: we keep l, the cell parameter, fixed while we let $N \to \infty$. This is tantamount to letting the volume go to infinity. For the density of the system we obtain

$$
\rho = \xi[\partial/\partial\xi(1/N)\ln(Q_N)]_\beta = (l\xi/\lambda)/(1 + l\xi/\lambda). \tag{2.4}
$$

Here $\rho = \langle n \rangle / N$ is a dimensionless density. [This density is related to the usual density as follows: $\rho = \langle n \rangle / N = l/v$. If m_0 is the mass per particle, then $\tilde{\rho} = m_0(1/v) = m_0\rho/l$ is the density as usually

[4] K. Huang, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1963).

[5] D. ter Haar, *Elements of Statistical Mechanics* (Holt, Rinehart, and Winston, Inc., New York, 1964).

defined.] One can also think of ρ as the proportion of the line occupied by particles, for we can write $\rho = \langle n\rangle l/Nl = L_0/L$, where L_0 is the length which the particles would occupy if they were tightly packed. Using (2.4) and keeping the cell parameter l fixed, one can now go to the thermodynamic limit and obtain:

$$\beta pL = \lim_{N\to\infty} \ln (Q_N) = \langle n\rangle/(1 - \rho) \qquad (2.5)$$

or

$$pL(1 - \rho) = pL_{\text{eff}} = \langle n\rangle KT$$

which is the well-known Tonks equation of state,[3] with $L_{\text{eff}} = L(1 - \rho)$ denoting the net length of the line not occupied by particles, i.e., the free "volume" of the system. (We use the symbol p to represent the "pressure" of the system in the appropriate units. Thus, in a one-dimensional system pressure has the units of force.)

By identical arguments to those given in the preceding section one sees that the analysis holds for systems of all dimensions and that consequently L and L_{eff} can be replaced by V and V_{eff}, respectively. Thus the model yields the correct equations of state for a fluid of hard rods, hard disks, or hard spheres.

One can show in this connection the effect of the hard-core potential on the series solution of Eq. (1.1) and on the choice of the cell parameter l. This is done in Appendix A.

C. One-Dimensional Fluid with Nearest-Neighbor Interactions

We examine in this section the thermodynamic properties of a one-dimensional fluid comprised of particles with hard-core repulsive potential and nearest-neighbor attractive interactions. While such a fluid is of very limited practical interest, it has the advantage of being mathematically tractable since the equation of state can be written in closed form. Furthermore, there are many striking parallels with real fluids, although such a system cannot have a phase transition, as we know from Van Hove's work.[6]

It was shown in I (Sec. 6A) that the partition function of this fluid is given by

$$Q_N = C_N \cosh^N(\nu) \times \{[1 + (1 + \omega)^{\frac{1}{2}}]^N + [1 - (1 + \omega)^{\frac{1}{2}}]^N\}, \qquad (2.6)$$

where $C_N = \exp(N\nu)$ and $\omega = (x_1^{-1} - 1)\,\text{sech}^2\,\nu$. In the thermodynamic limit we obtain for the grand potential[5]

$$q = \beta pl = \lim_{N\to\infty} 1/N \ln (Q_N) = \nu + \ln(\cosh\nu) + \ln[1 + (1 + \omega)^{\frac{1}{2}}]. \qquad (2.7)$$

[6] L. Van Hove, Physica 16, 137 (1950).

TABLE I. Relationship between density ρ, Gibbs free energy g, and parameter ν.

ρ	0	$\frac{1}{2}$	1
g	$-\infty$	$1/\beta \ln(\lambda x_1^{-1}/l)$	$+\infty$
ν	$-\infty$	0	$+\infty$

The density is then

$$\rho = \langle n\rangle/N = \xi(\partial q/\partial\xi)_\beta = (\tfrac{1}{2})(\partial q/\partial\nu)_\beta = (\tfrac{1}{2})[1 + \tanh\nu/(1 + \omega)^{\frac{1}{2}}]. \qquad (2.8)$$

While ν is a function of the temperature, interaction potential, and Gibbs free energy [cf. Eq. (1.1)], we can in fact consider it a free parameter in Eqs. (2.7) and (2.8) so that these two equations give the equation of state parametrically. Table I gives the relationship between ρ, g, and ν. Pressure–density data computed from these two equations are plotted in Figs. 1 and 2. (Actually $q = \beta pl$ is plotted vs ρ. For a given isotherm this is, of course, proportional to p.) The graphs show clearly that the pressure of the system is reduced by a decrease in the temperature or by an increase in the strength of the interaction potential. We wish next to compute the isothermal compressibility. This is given by

$$K_T = -(1/v)(\partial v/\partial p)_\beta = (1/\rho)(\partial\rho/\partial p)_\beta = (\beta l/2\rho^2)(\partial\rho/\partial\nu)_\beta = (\beta l x_1^{-1}\,\text{sech}^2\,\nu)/[\rho^2(1 + \omega)^{\frac{3}{2}}]. \qquad (2.9)$$

Figures 3 and 4 are plots of compressibility vs density corresponding to the p–ρ plots in Figs. 1 and 2, respectively. The behavior is physically reasonable: compressibility increases with decreasing temperature and increasing strength of the interaction. This parameter is, of course, of interest because of its relation to the slope of the p–ρ curve [cf. Eq. (2.9)] and its well-known connection with fluctuations in the number of particles of the system. The compressibility is therefore a valuable indicator of a change in phase and one expects the compressibility curve to have a discontinuity at the critical density. No such behavior is apparent in these two figures in consonance with Van Hove's theorem.[6]

The internal energy density of the system is determined as follows:

$$u = \langle H\rangle/N = -(\partial q/\partial\beta)_\xi = -[(\partial q/\partial\nu)_\beta(\partial\nu/\partial\beta)_\xi + (\partial q/\partial\beta)_\nu] = -2\rho(\partial\nu/\partial\beta)_\xi + (\partial q/\partial\beta)_\nu = (\tfrac{1}{2})\rho kT - (\zeta/2) \times \{2\rho - [x_1^{-1}\,\text{sech}^2\,\nu]/[1 + \omega + (1 + \omega)^{\frac{1}{2}}]\}. \qquad (2.10)$$

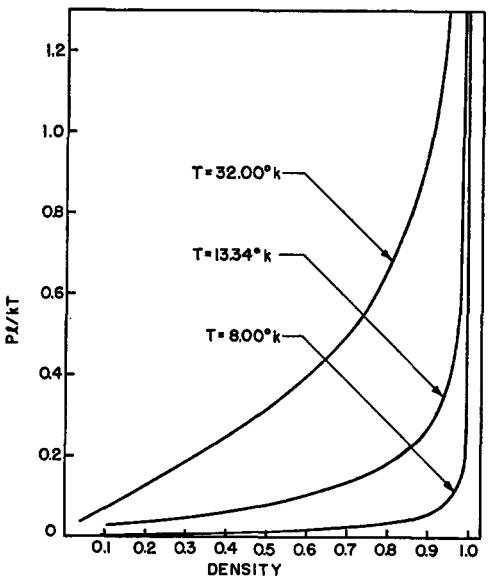


FIG. 1. Pressure-vs-density isotherms for a one-dimensional system with nearest-neighbor interactions; $\zeta = 64k$.

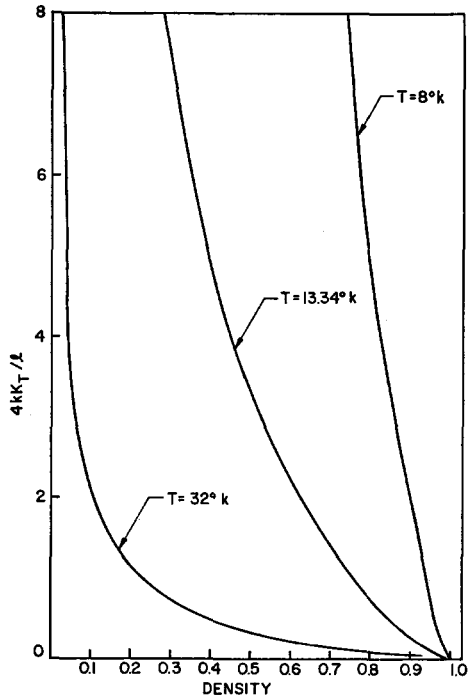


FIG. 3. Compressibility vs density for a one-dimensional system with nearest-neighbor interactions; $\zeta = 64k$, various temperatures.

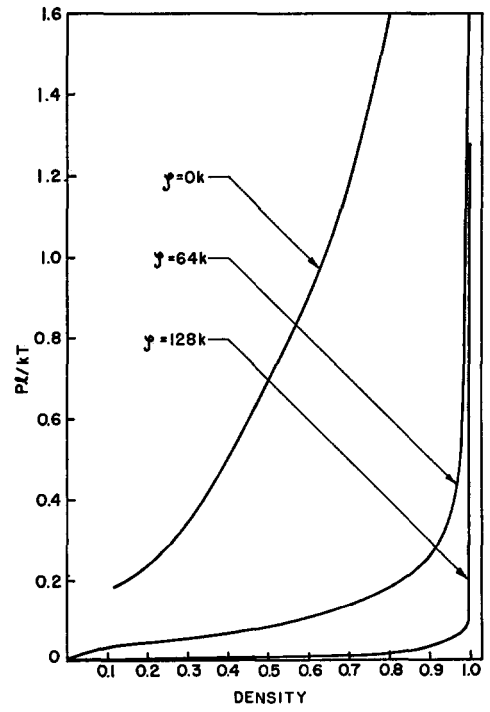


FIG. 2. Pressure-vs-density curves for various potential-well depths for a one-dimensional system with nearest-neighbor interactions; $T = 13.34°K$.

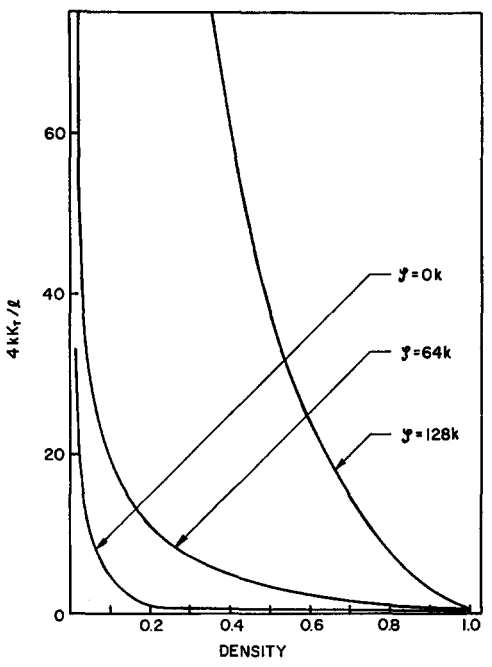


FIG. 4. Compressibility vs density for a one-dimensional system with nearest-neighbor interactions; $T = 13.34°K$; various potential-well depths.

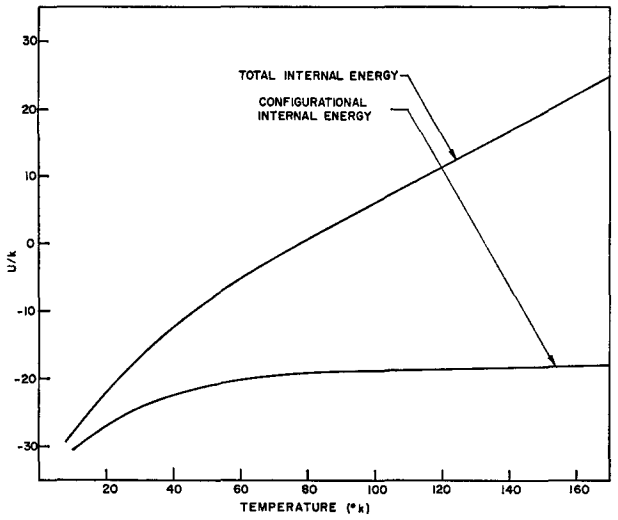


FIG. 5. Internal energy density vs temperature for a one-dimensional system with nearest-neighbor interactions.

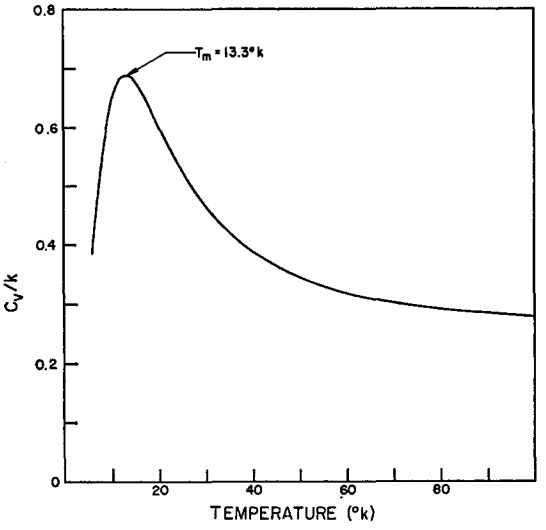


FIG. 6. Specific heat vs temperature for a one-dimensional system with nearest-neighbor interactions; $\rho = \frac{1}{2}$.

Energy density is plotted vs temperature in Fig. 5 for $\rho = \frac{1}{2}$. One can readily check the validity and meaning of the various terms in (2.10) by proceeding directly from the partition function. This analysis is carried out in Appendix B. It is apparent from (2.10) that the zero-point energy density of the system is $u = -\zeta\rho$, which corresponds to a zero-point energy per particle of $-\zeta$ and thus accords with physical intuition.

From (2.10) one can compute the specific heat of the system as follows:

$$C_v = (\partial u/\partial T)_v = -k\beta^2(\partial u/\partial\beta)_\rho$$
$$= (\tfrac{1}{2})k\rho + (\tfrac{1}{4})k(\beta\zeta)^2 x_1^{-1}$$
$$\times \operatorname{sech}^4 \nu/[(1+\omega)^{\frac{1}{2}}(1+(1+\omega)^{\frac{1}{2}})^2]. \quad (2.11)$$

As is well-known, the interest in the specific heat arises from the obvious connection between it and the fluctuations in energy of the system. One expects these to be very large in the critical region so that the specific heat is a valuable indicator of phase changes. In the critical region, (2.11) reduces to

$$C_v(\rho = \tfrac{1}{2}) = k/4 + k\epsilon^2 \operatorname{sech}^2 \epsilon, \quad (2.12)$$

where $\epsilon = (\frac{1}{4})\beta\zeta$. Figure 6 is a plot of this relation. While the maximum at $T_m = 0.208\zeta/k$ obviously represents large fluctuations in density, it does not correspond to a phase transition. If we let $\eta = 2kT/\zeta$, we can recast (2.12) in the form

$$(C_v - k/4) = k\eta^{-2}\exp(-1/\eta)$$
$$\times \sum_{r=0}^{\infty}(-1)^{r+1} r \exp[-(r-1)/\eta]$$
$$\doteq k\eta^{-2}\exp[-1/\eta] \quad (\eta \ll 1), \quad (2.13)$$

which shows that the configurational specific heat approaches zero exponentially as $T \to 0$, i.e., faster than the T^3 law of Debye.[7]

It is possible to eliminate the parameter ν in the equation of state and to write the pressure as an explicit function of the density. One obtains then from Eqs. (2.7) and (2.8):

$$pl(1-\rho) = kTx_1^{-1}$$
$$\times \{(\rho - \tfrac{1}{2}) + [(\rho - \tfrac{1}{2})^2 + x_1\rho(1-\rho)]^{\frac{1}{2}}\}. \quad (2.14)$$

One may notice in passing that for $x_1 \to 1$ this merges into the correct equation of state for the Tonks fluid. One can rewrite (2.14) in a somewhat more suggestive way:

$$pl_{\text{eff}} = kT\rho_{\text{eff}}, \quad (2.15)$$

where $l_{\text{eff}} = l(1-\rho) = (1/N)L(1-\rho) = (1/N)L_{\text{eff}}$ [cf. Eq. (2.5)] and $\rho_{\text{eff}} = (\rho - \frac{1}{2}) + [(\rho - \frac{1}{2})^2 + x_1\rho(1-\rho)]^{\frac{1}{2}}$. l_{eff} thus represents an effective cell parameter related to the effective or free "volume" by the relation $l_{\text{eff}} = \lim_{N\to\infty}\{(L - L_0)/N\}$. Similarly ρ_{eff} represents an effective density which, by virtue of the interaction potential, is lower than the true density. The pressure of the system is thus reduced by the interaction as we saw in Fig. 2. One can also cast the equation of state in the form of a van der Waals equation

$$(p + a)(L - b) = \langle n\rangle kT \quad (2.16)$$

where

$$a = p\{x_1/[(1 - L/2L_0) + (1 - L/2L_0)^2$$
$$+ x_1(L/L_0 - 1)] - 1\}$$

[7] J. M. Ziman, *Principles of the Theory of Solids* (Cambridge University Press, Cambridge, England, 1964).

and

$$b = L\rho = \langle n \rangle l = L_0.$$

The effect of the hard-core repulsion is contained in b, while a reflects the effect of the attractive potential. It is readily shown that $a \geq 0$. This form of the equation of state lends itself to a virial expansion which yields

$$p = \rho kT[1 + (2 - x_1)\rho + (4 - 5x_1 + 2x_1^2)\rho^2 + (8 - 18x_1 + 16x_1^2 - 5x_1^3)\rho^3 + (16 - 56x_1 + 82x_1^2 - 55x_1^3 + 14x_1^4)\rho^4 + (32 - 160x_1 + 340x_1^2 - 365x_1^3 + 196x_1^4 - 42x_1^5)\rho^5 + \cdots]. \quad (2.17)$$

Figure 7 is a plot of the second virial coefficient obtained from this expansion and of the true virial coefficient for a three-dimensional fluid. The similarity of these curves is rather surprising; their limit as $T \to \infty$, however, is different. From (2.14) it is possible to compute all thermodynamic functions in nonparametric form. Only a few results will be given here. For the internal energy density one finds, for example,

$$u = (\tfrac{1}{2})\rho kT - \rho\zeta\{1 - 2(1 - \rho)/[1 + 2((\rho - \tfrac{1}{2})^2 + x_1\rho(1 - \rho))^{\frac{1}{2}}]\}, \quad (2.18)$$

while the specific heat is given by

$$C_v = (\tfrac{1}{2})k\rho + 2k(\beta\zeta)^2\{\rho(1 - \rho)/[1 + 2((\rho - \tfrac{1}{2})^2 + x_1\rho(1 - \rho))^{\frac{1}{2}}]\}^2\{x_1/[(\rho - \tfrac{1}{2})^2 + x_1\rho(1 - \rho)]^{\frac{1}{2}}\}, \quad (2.19)$$

which for $\rho = \frac{1}{2}$ reduces again to expression (2.12). The chemical potential per particle is

$$g = kT \ln(\lambda/l) + 2kT \ln\{\rho_{\text{eff}}/[x_1(\rho(1 - \rho))^{\frac{1}{2}}]\}, \quad (2.20)$$

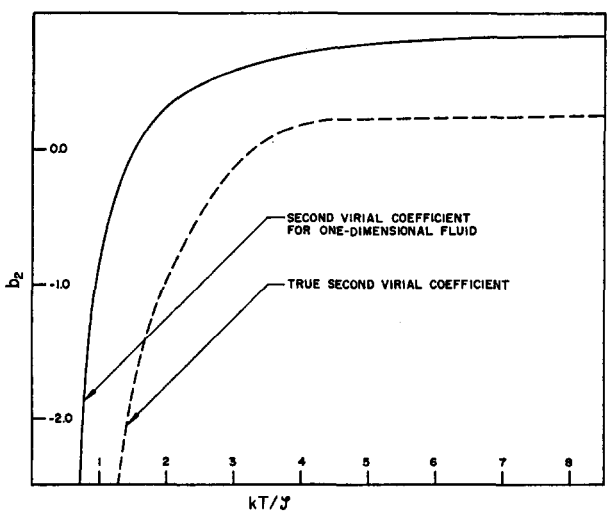


FIG. 7. Second virial coefficient vs temperature for a one-dimensional fluid with nearest-neighbor interactions.

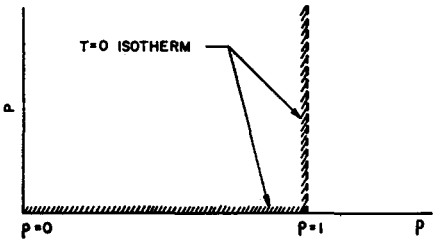


FIG. 8. Degenerate phase transition of the one-dimensional fluid with nearest-neighbor interactions.

where ρ_{eff} is given by (2.15). Other functions are readily determined from (2.14) but are not included here.

Before leaving this system we observe from (2.14) that $p = 0$ at $T = 0$ for all values of the density except $\rho = 1$. For the latter density we find

$$\lim_{\substack{T\to 0 \\ \rho\to 1}} p = \lim_{\substack{T\to 0 \\ g\to\infty}} (kT/l)[\tfrac{1}{2} \ln T + \zeta/kT + g/kT] + \text{finite terms} = +\infty. \quad (2.21)$$

Thus at $T = 0$ this system experiences a degenerate phase transition as shown in Fig. 8. One can also arrive at this conclusion by considering the analyticity of the partition function. One finds then that the zeros of the partition function of this fluid, just like those of the corresponding one-dimensional Ising ferromagnet, are distributed on the unit circle in the complex y plane.[8] These zeros cannot approach the real axis closer than $\alpha_{\min} = 2\cos^{-1}[(1 - x_1^{-1})^{\frac{1}{2}}]$. Hence[2] a phase transition can occur only as $T \to 0$ or $x_1 \to \infty$, and these transitions are degenerate.

D. A Fluid with Infinite-Range Interactions

We shall consider next a fluid with an attractive potential of infinitely long range, i.e., a potential where the range parameter γ goes to zero [cf. Eq. (1.1)]. A comparable magnetic potential in which all exchange interactions are identical has been applied by Kittel and Shore to a Heisenberg ferromagnet.[9] For a fluid such a potential has no physical counterpart; however, it constitutes an interesting limiting case to systems with interaction potentials of increasingly longer range. An alternative point of view is also possible. It was shown in I that, in considering systems with nearest-neighbor interactions, the coordination number increases as the dimension of the system is increased. The system considered here could therefore be thought of as a limiting case to a sequence of

[8] T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Co., Inc., New York, 1956).

[9] C. Kittel and H. Shore, Phys. Rev. **138**, 1165 (1965).

nearest-neighbor systems of increasingly higher dimensionality. Specifically, then, the potential we wish to examine is

$$V_s = \begin{Bmatrix} 0 & (s=0) \\ -\zeta & (s \neq 0) \end{Bmatrix} \quad \text{(Mod. } N\text{)}. \qquad (2.22)$$

If we substitute this potential in the partition function, Eq. (1.1), we find that

$$t_n = (1/n!) \sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' \exp\left(2 \sum_{\substack{r=1 \\ (r<s)}}^{n} \sum_{s=1}^{n} \beta\zeta\right)$$

$$= \binom{N}{n} \exp\left[n(n-1)\beta\zeta\right]. \qquad (2.23)$$

Consequently the partition function takes the form

$$Q_N = C_N \sum_{n=0}^{N} \binom{N}{n} x^{-n(N-n)} \cosh\left[(N-2n)v\right] \qquad (2.24)$$

or

$$Q_N = \sum_{n=0}^{N} \binom{N}{n} y^n x^{-n(N-n)},$$

where $y = \exp(2v)$, $C_N = \exp(Nv)$, and $x = \exp(\beta\zeta)$. The sum in (2.24), although suggestive, is not readily evaluated. It is suited, however, to numerical evaluation and theoretical analysis. We present some numerical results first. For the density one obtains

$$\rho = (\tfrac{1}{2})(\partial q/\partial v)_\beta = (\tfrac{1}{2})(1 + T_v/NT), \qquad (2.25)$$

where

$$T(v, x) = \sum_{n=0}^{N} \binom{N}{n} x^{-n(N-n)} \cosh\left[(N-2n)v\right]$$

and

$$T_v = (\partial T/\partial v)_x.$$

For the compressibility we find:

$$K_T = (\beta l/4\rho^2)[T_{vv}/NT - (2\rho-1)^2 N], \qquad (2.26)$$

where $T_{vv} = (\partial^2 T/\partial v^2)_x$. The internal energy density is

$$u = (\tfrac{1}{2})\rho kT - \zeta[(N-1)\rho - (T_x/NT)], \qquad (2.27)$$

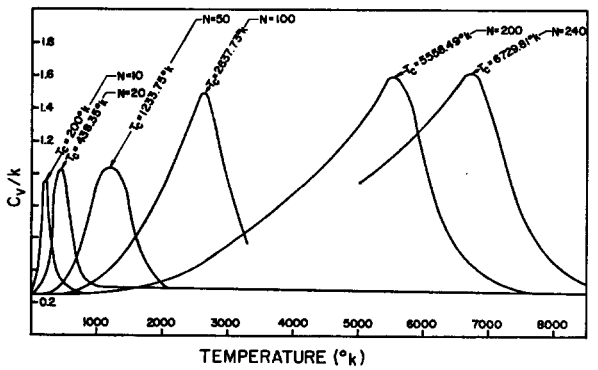


FIG. 9. Specific heat vs temperature for a one-dimensional system with infinite-range interactions; $\rho = \frac{1}{2}$.

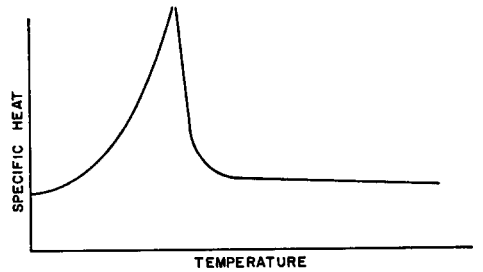


FIG. 10. Typical experimental specific-heat curve. (See Refs. 10 and 11.)

where $T_x = (\partial T/\partial x)_v$. Finally, the specific heat is

$$C_v = (\tfrac{1}{2})k\rho + k(\beta\zeta)^2 \times \{(1/NT)[(2\rho-1)NT_x - T_{xv}]^2/[(2\rho-1) \times NT_v - T_{vv}] + (1/NT^2)(TT_{xx} - T_x^2)\}, \qquad (2.28)$$

where $T_{xv} = (\partial^2 T/\partial x \partial v)$ and $T_{vv} = (\partial^2 T/\partial v^2)_x$.

Figure 9 is a plot of specific heat vs temperature for systems ranging in size from 10 to 240 cells and for a density of $\rho = \frac{1}{2}$. It appears from the graphs that the maxima tend to some sort of a limit, although it is not clear what this limit is. However, inspection of the data in Table II suggests that the quantity $N\zeta\beta_m$

TABLE II. Tabulated data on the maxima of the specific heat curves in Fig. 9.

System size (N)	Temperature of specific-heat maxima (T_m)	$N\zeta\beta_m$
10	193.95	3.30
20	438.35	2.92
50	1233.76	2.69
100	2637.53	2.43
200	5558.49	2.30
240	6729.81	2.28

approaches a limit 2.0. The analysis which follows at the end of this section shows that this is in fact the correct limit. One also observes from Fig. 9 that the specific heat curves assume a characteristic shape. A typical experimental specific-heat curve is shown in Fig. 10. It is based on the curves obtained by Nix and Shockley for alloys exhibiting typical order–disorder phase transitions.[10] A similar specific heat curve for xenon was published recently by Edwards, Lipa, and Buckingham.[11] It is quite apparent that the curves in Fig. 9 approach the characteristic shape of the experimental curves. Pressure–density curves are

[10] F. C. Nix and W. Shockley, Rev. Mod. Phys. **10**, 1 (1938).
[11] C. Edwards, J. A. Lipa, and M. J. Buckingham, Phys. Rev. Letters **20**, 496 (1968).

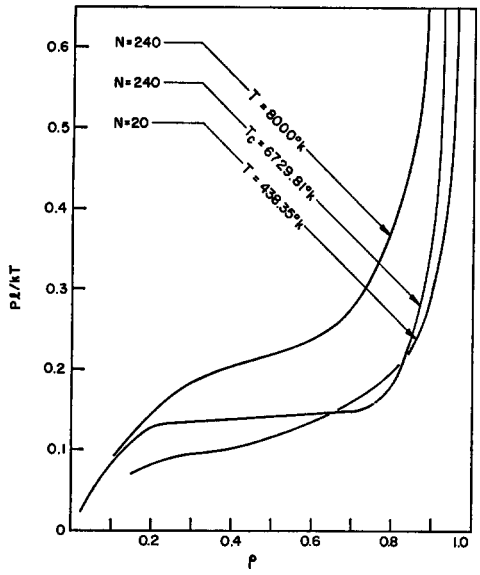


FIG. 11. Pressure-vs-density isotherms of one-dimensional systems with infinite-range interactions.

presented in Fig. 11. It is obvious that the isotherm of the 240-cell system at its "critical" temperature (i.e., the temperature where the specific heat curve has a maximum) is very nearly that of a system undergoing a phase transition. Conversely, the 8000°K isotherm, which is well above the "critical" temperature for this system, shows no such behavior. For comparison the

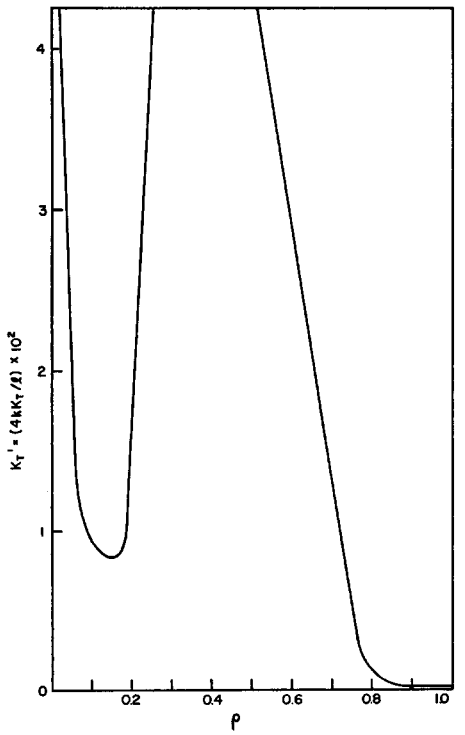


FIG. 12. Compressibility-vs-density curve for a one-dimensional system of 240 cells with infinite-range interactions.

"critical" isotherm for a very much smaller system also gives no indication of any phase change. "Critical" isotherms for systems of intermediate size (not shown to avoid confusion) gradually approach a phase transition with increasing system size.

A compressibility-vs-density plot for the 240-cell system at its "critical" temperature appears in Fig. 12. The compressibility curve very clearly approaches a discontinuity which is in keeping with the nearly horizontal isotherm in Fig. 11. Since the system is finite there is, of course, no true discontinuity, but the maximum value of the K_T curve cannot be determined from the computer data.

Figures 11 and 12 show quite convincingly that the system approaches a change of phase. It is possible to generalize and confirm the numerical results by analyzing Eq. (2.24). Suppose that T_m is the maximum term in the partition function, then clearly

$$q = \lim_{N\to\infty} (1/N) \ln (Q_N) = \nu + \lim_{N\to\infty} (1/N) \ln (T_m) + R. \tag{2.29}$$

Here

$$R = \lim_{N\to\infty} (1/N) \ln \left(1 + \sum_n T_n/T_m\right) = 0,$$

since there are N terms appearing in the argument of the logarithm, and $|T_n/T_m| < 1$ so that

$$|R| < \lim_{N\to\infty} (1/N) \ln N = 0.$$

In the thermodynamic limit we can thus determine all thermodynamic functions from the maximum term, T_m. Since we are interested in phase transitions, furthermore, it suffices to examine the case $\rho = \frac{1}{2}$, i.e., $\nu = 0$. To find the maximum term in the sum

$$T_n(x, 0) = \sum_{n=0}^{N} \binom{N}{n} x^{-n(N-n)}$$

we observe that the summand can be written as an exponential by means of the Stirling approximation. Noting that the summand attains a maximum as the exponent does, we obtain the following condition:

$$\ln (N/n - 1) + (2n - N)\beta\zeta = 0. \tag{2.30}$$

Letting $n/N = \rho$ as before [where we have used (2.29)] this leads to the following transcendental equation:

$$z = \tanh (az), \tag{2.31}$$

where $z = (2\rho - 1)$ and $a = N\beta\zeta/2$.

One obvious solution is $\rho = \frac{1}{2}$, which does not correspond to a phase transition. However, if we now plot the curves $f_1(z) = z$ and $f_2(z) = \tanh (az)$ as in Fig. 13, we see that two other solutions are possible corresponding to $z = 2\rho - 1 = \pm z_1$, or

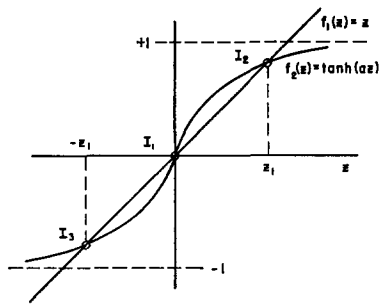


FIG. 13. Graphic solution of Eq. (2.31).

$\rho = \frac{1}{2} \pm x_1$. It is obvious that these additional solutions can exist if and only if the curve $f_2(z)$ has a slope at the origin greater than 1, i.e., if $a > 1$. Hence a phase transition can occur only if $N\beta\zeta > 2$; then ρ is truly a multiple-valued function of the pressure, having the values $\rho = \frac{1}{2} + z_1, \frac{1}{2}$, and $\frac{1}{2} - z_1$, all corresponding to the pressure $p = (kT/l) \lim_{N\to\infty} (1/N) \ln (T_m)$. The critical temperature of the system must therefore be given by the condition $N\beta\zeta = 2$. This, then, confirms the limit of the quantity $N\beta\zeta$ discussed in connection with Table II. We note that in the thermodynamic limit the critical temperature is infinite. For finite systems the preceding conclusions hold only approximately, since in that case T_m is no longer the only significant term. There is then also a contribution from the remainder term. Table III allows a comparison of the "critical" temperatures of finite systems obtained numerically and from the relation $a = 1$ (which, as pointed out, does not hold exactly for such systems). For infinite systems one notes that $\lim_{N\to\infty} a = \infty$ so that from Eq. (2.31) z can only take on the values ± 1; i.e., ρ is restricted to the values zero and one. Consequently the phase transition of this system in the thermodynamic limit is a degenerate one resembling that in Fig. 8.

III. ANALYSIS OF FINITE SYSTEMS

In Sec. II we examined various fluids with interaction potentials that allowed us to go to the thermodynamic limit. This section is devoted to an investigation of systems in which every particle interacts with every other one, i.e., which interact with the full Lennard–Jones potential. In that case we cannot go to the limit $N \to \infty$. However, the series solution of the partition function, Eq. (1.1), is well suited to numerical analysis. Hence we consider systems of finite and in fact quite modest size for which the thermodynamic properties were determined by means of an IBM 360 model-40 computer. Computer time increased very rapidly with system size so that it was not feasible to investigate systems with more than 15 cells. Some caution is in order in speaking of the thermodynamic functions of such systems since we are dealing here with finite assemblies and cannot go to the thermodynamic limit. We shall denote these functions by their usual symbols but interpret them in a common-sense way.

TABLE III. Comparison of critical temperatures for a one-dimensional system with infinite-range interactions.

N	T_c (Numerical)	T_c' (Analytic)[a]	T_c/T_c'
10	193.95	320	0.60
20	438.35	640	0.68
50	1233.76	1600	0.78
100	2637.53	3200	0.82
200	5558.49	6400	0.87
240	6729.81	7680	0.88

[a] The analytically determined temperature T_c' holds exactly only in the thermodynamic limit. For the finite systems considered here it is an approximation.

A. Formulation of the Problem for Numerical Analysis

We define a function T as follows [cf. Eq. (1.1)]:

$$T(\nu, \beta, N) = \sum_{n=0}^{[N/2]} T_n. \tag{3.1}$$

The grand potential is then given by

$$q = \beta p l = \nu + (1/N) \ln (2T). \tag{3.2}$$

The density of the system is

$$\rho = (\tfrac{1}{2})(1 + T_\nu/NT), \tag{3.3}$$

where $T_\nu = (\partial T/\partial \nu)_\beta$. For the compressibility one obtains:

$$K_T = (\beta l/4\rho^2)[T_{\nu\nu}/NT - (2\rho - 1)^2 N], \tag{3.4}$$

where $T_{\nu\nu} = (\partial^2 T/\partial \nu^2)_\beta$. The internal energy density is

$$u = (\tfrac{1}{2})\rho kT + \rho \sum_{r=1}^{N} V_r - T_\beta/NT, \tag{3.5}$$

where $T_\beta = (\partial T/\partial \beta)_\nu$. For the specific heat one finds

$$C_\nu = \tfrac{1}{2}k\rho + k\beta^2[(T_{\beta\beta}/NT) - (1/N)(T_\beta/T)^2] + k\beta^2[(T_{\nu\beta}) - (2\rho - 1)NT_\beta]^2/[NT((2\rho - 1)^2 N^2 T - T_{\nu\nu})], \tag{3.6}$$

where $T_{\beta\beta}$, $T_{\nu\beta}$, etc., have the obvious interpretation.

We next wish to find a pair-correlation function that is amenable to numerical analysis. For this purpose we define an average correlation parameter as follows:

$$S_i = (1/N)\left\langle \sum_{r=1}^{N} \sigma_r \sigma_{r+i} \right\rangle. \tag{3.7}$$

One observes that $S_i = 1$ if cells in the system separated by i cell parameters are either all occupied or all empty. If one cell is occupied while the other is empty for every pair of cells separated by i cell parameters, then perfect anticorrelation exists and $S_i = -1$. Finally, $S_i = 0$ corresponds obviously to a state of no correlation. In order to find the relationship between the correlation parameter defined above and the usual pair correlation functions, we restrict ourselves to the case $\rho = \frac{1}{2}$ since we are principally interested in the order of the system in the transition region. It is shown in Appendix C that this relationship is as follows[8]:

$$S_i = (1/N)\sum_{r=1}^{N} C^{(2)}(\sigma_r, \sigma_{r+i}) - 1, \qquad (3.8)$$

or in terms of the Kirkwood correlation functions, which have found more frequent use in dealing with fluids:

$$S_i = 1/(N-1)\sum_{r=1}^{N} g^{(2)}(\sigma_r, \sigma_{r+i}) - 1. \qquad (3.9)$$

For convenience we shall refer to the correlation parameter S_i simply as correlation function in what follows.

The correlation functions are readily computed from (1.1) by noting that

$$S_i = (\tfrac{1}{2}N)(\partial/\partial\theta_i) \times \ln\left\{\sum_{\{\sigma\}} \exp\left(\sum_{r=1}^{N}\sum_{s=1}^{N}\sigma_r\sigma_{r+s}\theta_s + \nu\sum_{r=1}^{N}\sigma_r\right)\right\} \qquad (3.10)$$

It follows then that:

$$S_i = 1 + (T_{\theta_i}/2NT) \qquad (3.11)$$

where

$$T_{\theta_i} = (\partial T/\partial\theta_i)_v = 8(\partial T/\partial x_i)_v = \sum_{n=0}^{[N/2]} \cosh[(N-2n)\nu]X^{-n/2}\{(t_n)_{\theta_i} - 8n\} \qquad (3.12)$$

and

$$(t_n)_{\theta_i} = (\partial t_n/\partial\theta_i) = (8/n!)\sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' \prod_{r=1}^{n}\prod_{\substack{s=1\\(r<s)}}^{n} x_{(q_r-q_s)}\delta_{(q_r-q_s),i}.$$

(We assume for convenience that N is odd.)

B. Results for One-Dimensional Systems

Specific-heat curves for systems ranging in size from two to fourteen cells are shown in Fig. 14. The curves are for a potential-well depth of $\zeta = 64k$ which corresponds roughly to that of Neon.[12] One observes

[12] W. Band, *An Introduction to Quantum Statistics* (D. Van Nostrand Co., Inc., New York, 1955).

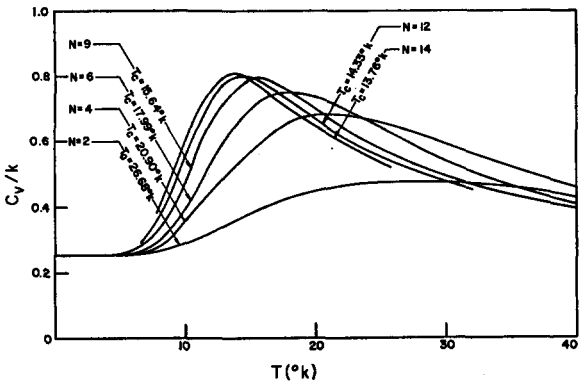


FIG. 14. Specific heat vs temperature for a one-dimensional system; $\zeta = 64k$; $\gamma = 6$; $\rho = \frac{1}{2}$.

that the maxima of these curves becomes higher and steeper with increasing system size and moves toward lower temperatures. Table IV summarizes this information. The term "critical" temperature is used loosely in this table and in what follows to mean the temperature at which the specific-heat curve has a maximum. It is apparent that the specific-heat maxima approach a limit; however, from the information available one cannot say with any certainty what this limit is. If one expands ζ/kT_c in inverse powers of N and fits the expansion parameters to the numerical data, one obtains

$$\zeta/kT_c \simeq 6.43 - 31.16/N + 87.36/N^2. \qquad (3.13)$$

Table IV also shows the values of ζ/kT_c obtained from this equation. The fit is seen to be excellent for systems of six cells or more. If Eq. (3.13) were to remain valid for larger systems one would conclude that $\lim_{N\to\infty} \zeta/kT_c = 6.43$ which, for the neonlike potential used in Fig. 14 would correspond to $T_c \simeq 10°K$.

Specific-heat curves for different potential-well depths are plotted in Fig. 15 for a 12-cell system. The

TABLE IV. Summary of critical data for one-dimensional specific heat curves.

System size (N)	Critical temperature (T_c)	ζ/kT_c True	ζ/kT_c Eq. (3.13)
2	26.68	2.39	12.69
4	20.90	3.06	4.10
6	17.99	3.56	3.66
9	15.64	4.09	4.04
10	15.14	4.22	4.18
11	14.74	4.34	4.32
12	14.33	4.44	4.44
13	14.08	4.55	4.55
14	13.76	4.65	4.65

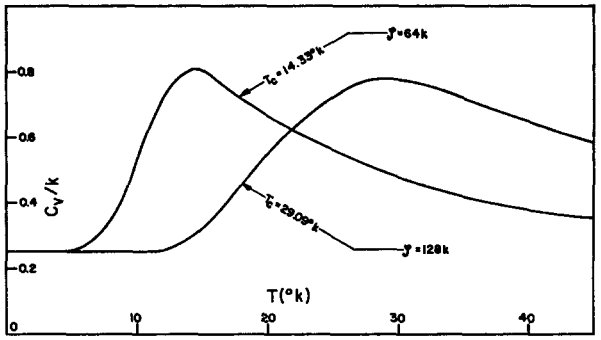


FIG. 15. Specific-heat vs temperature for a one-dimensional system; $N = 12$; various potential-well depths; $\gamma = 6$; $\rho = \frac{1}{2}$.

maxima of these curves occur at higher temperatures as the potential-well depth is increased. Comparable curves for a fixed depth of the potential well but for various range parameters are shown in Fig. 16. It appears from these data that the effect of increasing the effective range of interaction, i.e., of decreasing the parameter γ, is similar to that of increasing the potential-well depth. Both shift the maxima to higher temperatures. One can incorporate the γ and ζ dependence into an equation similar to (3.13)

$$\zeta/kT_c = \alpha_\gamma \exp\{(m-1)/100\} \times [1.53 - 7.42/N + 20.80/N^2], \quad (3.14)$$

where m is given by $\zeta = 2^m k$ and

$$\alpha_\gamma = \gamma - (\gamma-1)(\gamma-2)(\gamma-3)/(18+2\gamma).$$

This equation fits the numerical data with quite good accuracy.

Pressure-vs-density data for systems of various sizes are plotted in Fig. 17. As the size of the system increases, the interaction becomes more effective and the pressure for a given density is reduced. Figure 18 shows the change of the p–ρ isotherms with temperature for a 12-cell system. Pressure is seen to decrease with decreasing temperature as one would expect. There is,

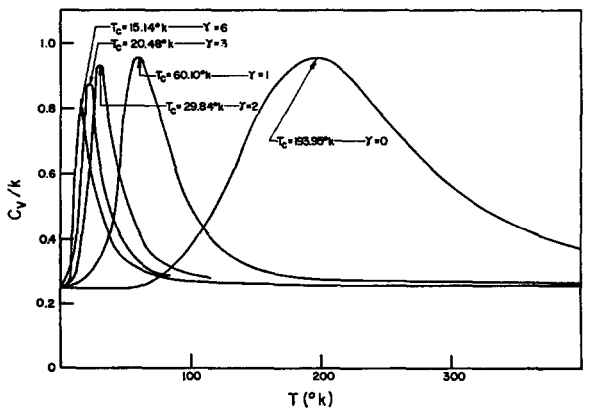


FIG. 16. Specific-heat vs temperature for a one-dimensional system with different potential range parameters; $N = 10$; $\zeta = 64k$; $\rho = \frac{1}{2}$.

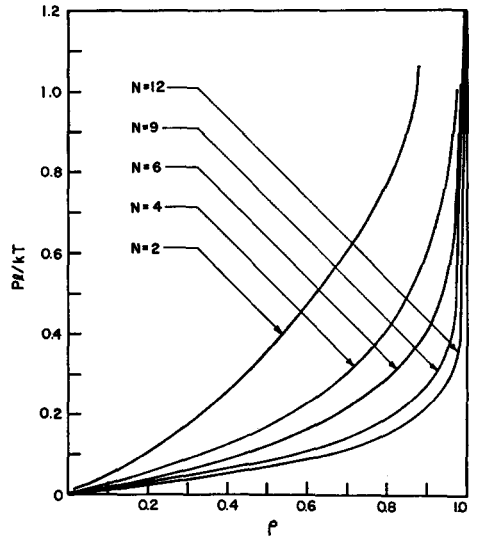


FIG. 17. Pressure-vs-density isotherms for one-dimensional systems of various sizes; $T = 10°K$; $\zeta = 64k$; $\gamma = 6$.

however, no indication of a phase transition. However, in Fig. 19, which depicts the "critical" isotherms for a 10-cell system with various interaction potentials, we observe a noticeable inflection and flattening of these curves. This is rather reminiscent of the curves in Fig. 11 and could presage a change of phase. Interestingly enough it is found that all critical isotherms fall within the narrow band bordered by the curves marked A and B. This seems to suggest that a law of corresponding states may apply approximately. At the critical density one finds from the upper curve, which coincides with all but two of the isotherms, that $pl/kT_c = \beta_c p_c l = 0.135$. Hence, $\beta_c p_c V_c = \beta_c p_c l/\rho_c = 0.270$. Hill shows in Table II, p. 232, that the mean

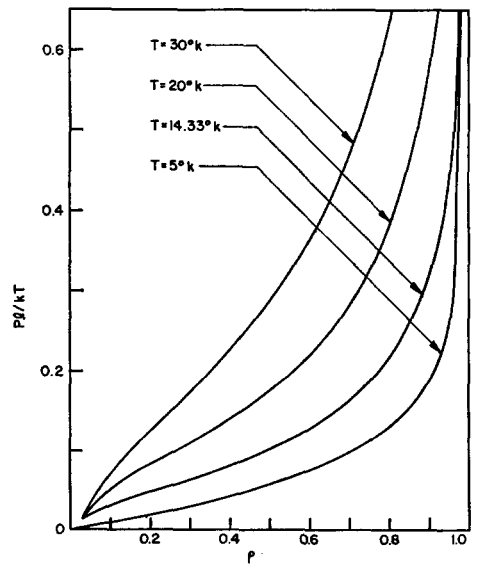


FIG. 18. Pressure-vs-density isotherms for a one-dimensional system; $N = 12$; $\zeta = 64k$; various temperatures.

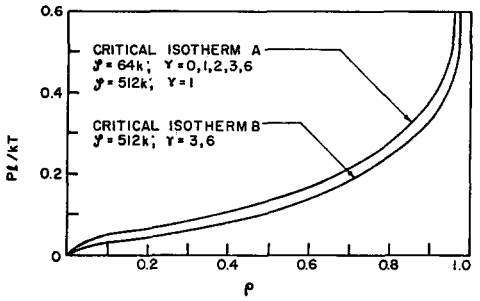


FIG. 19. "Critical" pressure-vs-density isotherms for a one-dimensional system with different interaction potentials; $N = 10$.

value of this parameter for Ne, N_2, A, and CH_4 is 0.292.[8] The potentials plotted in Fig. 19 are approximately those for Ne and A. Consequently the critical value obtained from the present one-dimensional model lies within 7.5% of the experimental value for real three-dimensional fluids.

Compressibility curves for systems of various sizes are plotted in Fig. 20. These are seen to exhibit normal fluid behavior, without any indication of a phase transition. Figure 21, on the other hand, shows compressibility evidencing some rather interesting anomalous behavior. These curves are the analogs of those in Fig. 19 for a potential-well depth of $64k$. The anomalies clearly correspond to the inflections in

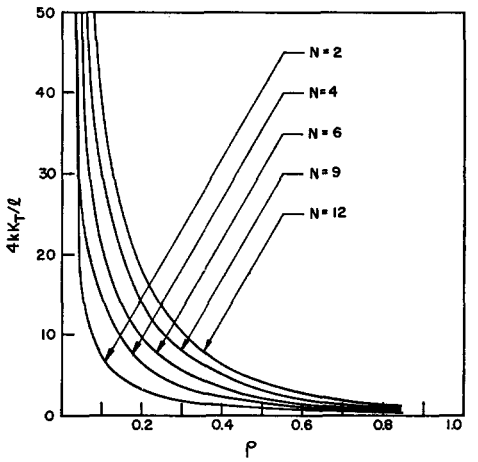


FIG. 20. Compressibility vs density for one-dimensional systems of various sizes; $T = 10°K$; $\zeta = 64k$; $\gamma = 6$.

ranging in size from 5 to 15 cells are plotted in Fig. 23 for a temperature of 32°K and a density of $\rho = \frac{1}{2}$. These are seen to agree quite well qualitatively with experimental curves and curves for radial distribution functions computed by other methods.[8,13,14] Correlation is largest for nearest neighbors, then drops off and approaches zero. However, one observes an absence of the local maxima and minima that are usually so characteristic of these curves. The difference in behav-

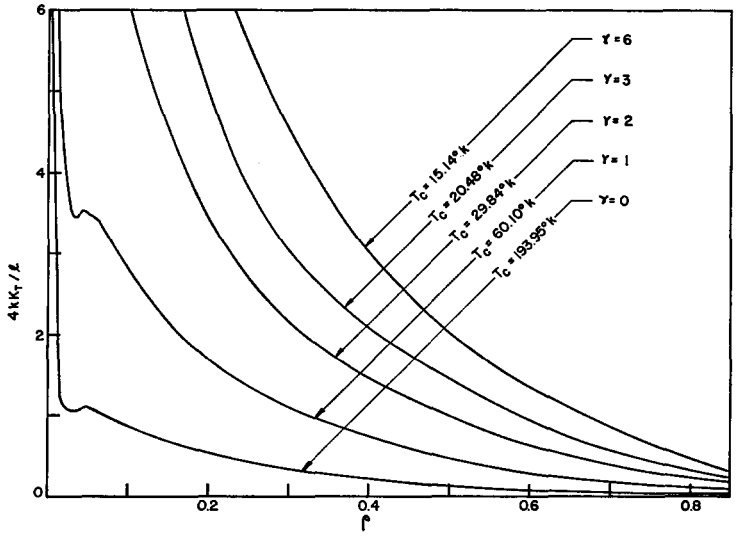


FIG. 21. Compressibility vs density for a one-dimensional system with various potential ranges; $N = 10$; $\zeta = 64k$.

slope in Fig. 19 and may be associated with an incipient phase transition. This behavior appears even more clearly in Fig. 22, where pressure and compressibility isotherms are plotted on the same graph. Both are for a 12-cell system at its "critical" temperature. The inflection in the p–ρ curve and the corresponding anomaly in the compressibility curve are plainly visible.

The pair-correlation functions for selected systems

ior appears to be due to the discrete nature of the cellular model used here. De Boer has shown that the undulations are due to contributions to the potential of average force from the screening effect of particles

[13] S. Fluegge, *Handbuch der Physik, Vol. XIII*, (Springer Verlag, Berlin, 1962).

[14] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954).

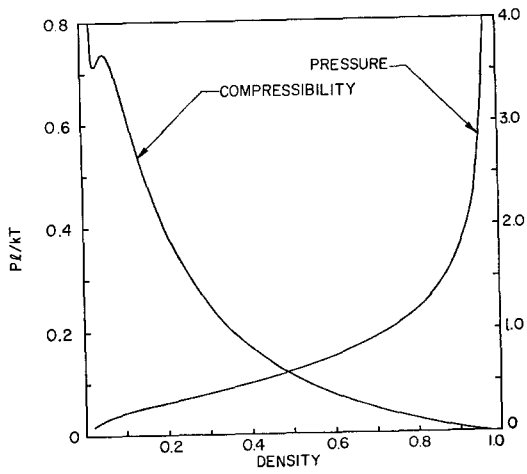


FIG. 22. Pressure and compressibility vs density for a 12-cell one-dimensional system; $T = 14.33°\text{K}$; $\zeta = 64k$.

surrounding the particle of interest.[15] The effect, according to de Boer, is to superpose an extra repulsive force at distances smaller than $1.5l$, an extra attractive force between $1.5l$ and $2.2l$, and a small repulsive force at larger distances. In the present model all potentials are averaged over the whole cell. Thus the variations discussed by de Boer are smeared out over the entire cell and are more than compensated for by the attractive Lennard–Jones potential. In Fig. 24 we see the effect of temperature on correlation for a 13-cell system. At temperatures well below the critical point one sees that correlation is perfect, or very nearly so,

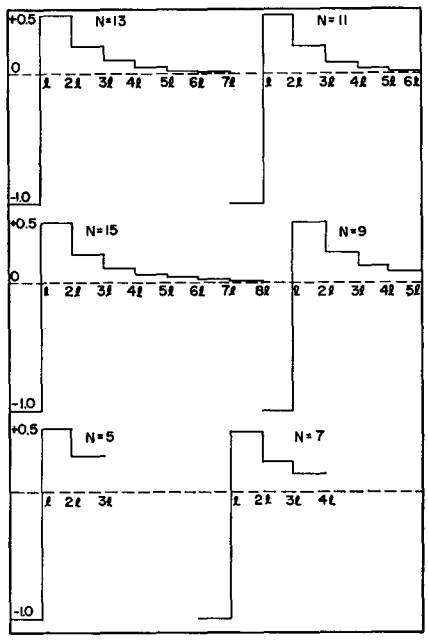


FIG. 23. Pair-correlation functions for one-dimensional systems of various sizes; $T = 32°\text{K}$; $\zeta = 64k$; $\gamma = 6$; $\rho = \frac{1}{2}$.

[15] J. de Boer, Rept. Progr. Theoret. Phys. (Kyoto) **12**, 305 (1949).

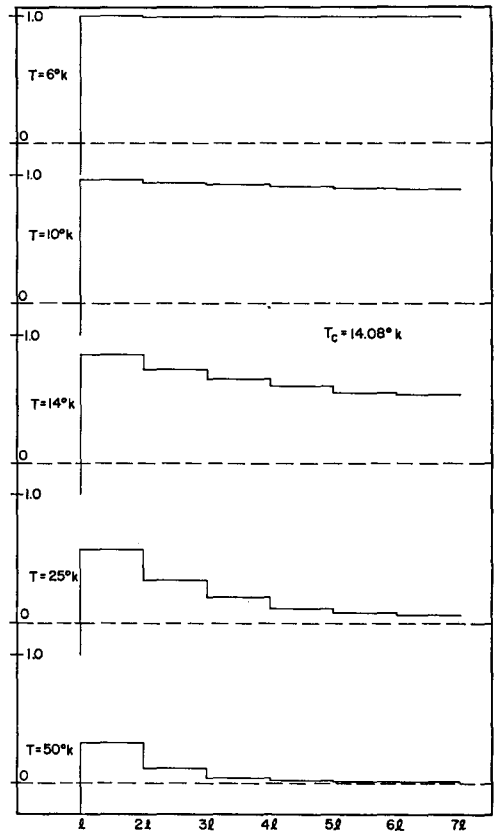


FIG. 24. Pair-correlation functions for a 13-cell system at various temperatures; $\zeta = 64k$; $\gamma = 6$; $\rho = \frac{1}{2}$.

throughout the entire system. Long-range order exists. As the temperature increases, correlation is gradually reduced, less so for nearest-neighbor cells and more for cells far removed from one another. Long-range order changes to short-range order. The drop in correlation is particularly marked for temperatures higher than the critical one. This is shown more

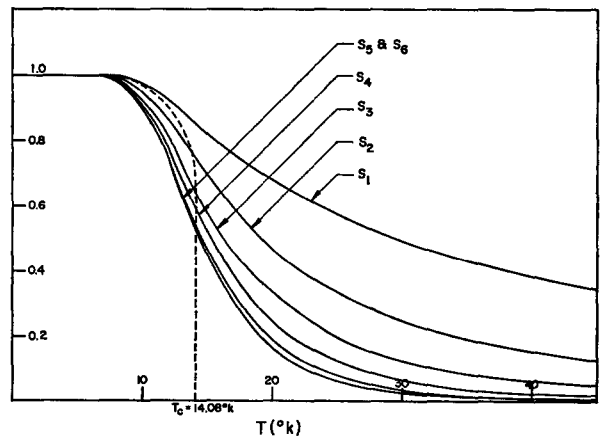


FIG. 25. Pair-correlation functions vs temperature for a 13-cell system; $\zeta = 64k$; $\gamma = 6$; $\rho = \frac{1}{2}$.

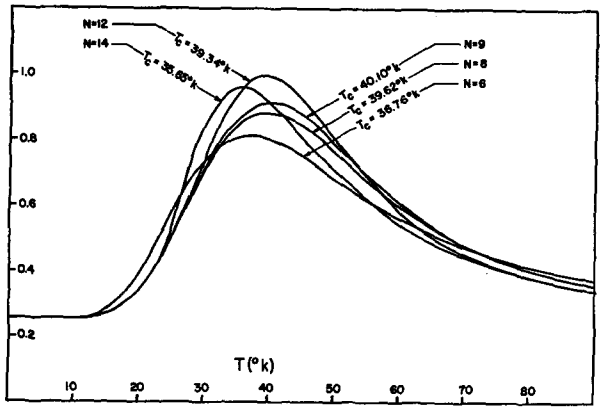


FIG. 26. Specific heat vs temperature for typical two-dimensional systems; $\zeta = 64k$; $\rho = \frac{1}{2}$.

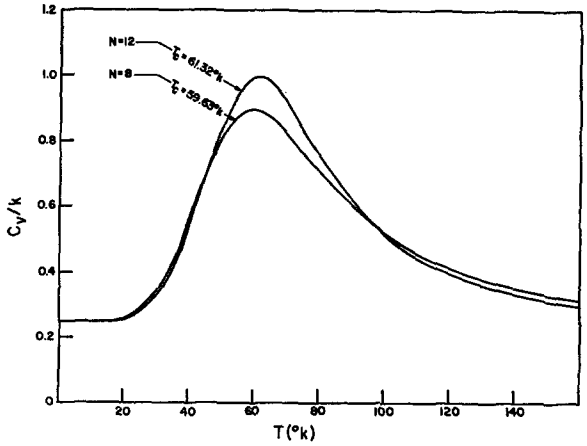


FIG. 27. Specific heat vs temperature for two typical three-dimensional systems; $\zeta = 64k$; $\rho = \frac{1}{2}$.

clearly in Fig. 25 where the correlation functions for this same system are plotted against temperature. We see that nearest-neighbor correlation decreases only slowly; however, the longer-range correlations drop off markedly as the critical temperature is approached and passed. The graph shows very clearly how long-range order changes to short-range order at the "critical" temperature. The behavior is completely analogous to that found by Kaufman and Onsager in their investigation of the two-dimensional Ising ferromagnet.[16] Since in the two-dimensional Ising model this behavior is associated with a phase transition, one surmises that in the present instance it might be indicative of an incipient change of phase. The surmise is reinforced by the behavior discussed in connection with Fig. 22.

C. Results for Two- and Three-Dimensional Systems

Section 2B and 6B of Paper I dealt with application of the present model to multi-dimensional systems. In this section we present some results for small two- and three-dimensional systems with nearest-neighbor interactions only that were obtained by computer. The potential used was:

$$V_s = \begin{cases} -64k & (s = 1, m) \quad (s = 1, m, k) \\ 0 & (s \neq 1, m) \quad (s \neq 1, m, k) \end{cases} \text{Mod. } N. \tag{3.15}$$

two-dimensional system *three-dimensional system*

Two-dimensional systems accessible within the restriction $N \leq 15$ are those having 4, 6, 8, 9, 10, 12, 14, and 15 cells, respectively, while three-dimensional systems are confined to $N = 8$ and 12.

Specific-heat curves for typical two- and three-dimensional fluids are shown in Figs. 26 and 27, respectively. As in the one-dimensional case, the curves exhibit a characteristic peak which is shifted to higher temperatures with increasing dimensionality of the system. Table V summarizes the "critical" temperature data available for these systems. For the two-dimensional system the exact value for the ratio ζ/kT_c is 1.76.[17] It is seen that the values in Table V

TABLE V. "Critical" temperature data for two- and three-dimensional systems for $\zeta = 64k$, $\rho = \frac{1}{2}$.

System size (N)	"Critical" temperature (T_c)		ζ/kT_c	
	2-Dim. system	3-Dim. system	2-Dim. system	3-Dim. system
6	36.76	—	1.74	—
8	39.62	59.63	1.61	1.07
9	40.10	—	1.59	—
12	39.34	61.32	1.62	1.04
14	35.65	—	1.79	—

fluctuate about the exact ratio within approximately $\pm 10\%$. These fluctuations are attributable to the small size of the systems investigated, the differences in symmetry of the systems, and the consequent strong influence of boundary effects (spurious interactions). (The 14-cell system, for example, consists of two rows of seven cells each, while the 12-cell one has three rows of four cells. There are thus obvious differences in symmetry.) For the three-dimensional system, Wakefield's value of the corresponding ratio is generally considered the most accurate[13]:

$$\exp\{-2J/kT_c\} = 0.641.$$

[16] B. Kaufman and L. Onsager, Phys. Rev. 76, 1244 (1949).

[17] L. Onsager, Phys. Rev. 65, 117 (1944).

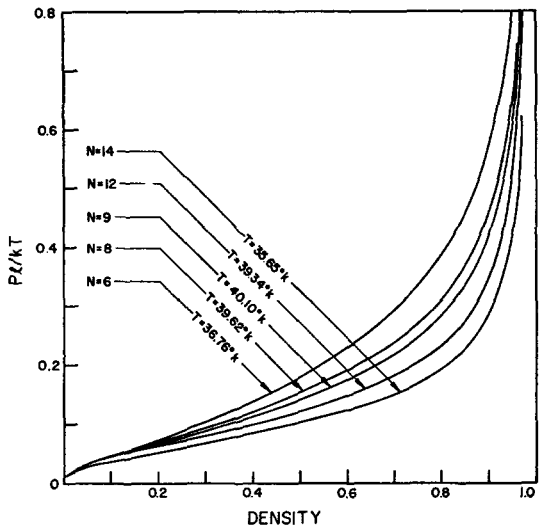


FIG. 28. Pressure vs density, two-dimensional systems; $\zeta = 64k$; $T = T_c$.

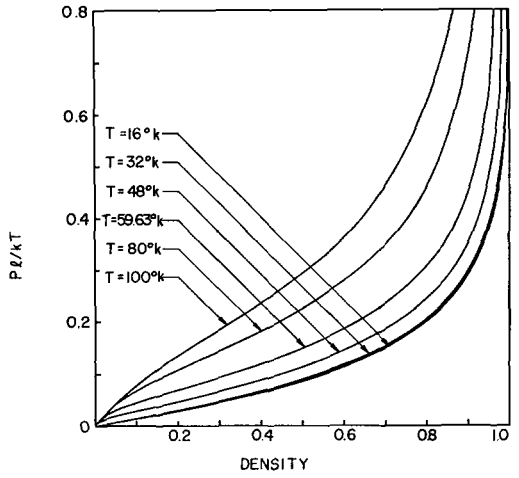


FIG. 30. Pressure–density curves for an eight-cell three-dimensional system; $\zeta = 64k$.

In terms of the parameters of the present model this is equivalent to $\zeta/kT_c = 1.05$. (ξ for the nearest-neighbor fluid corresponds to $4J$ for the ferromagnet.) Other estimates of this ratio range from 0.98 to 1.09.[13] The values obtained from the model considered here fall well within this range and differ from the best value by less than 2%. This accuracy is most likely due to the perfect and almost perfect cubic symmetry of these two systems. It is nonetheless remarkable that such close agreement should be obtained for assemblies of so small a size.

Pressure-vs-density graphs for typical two- and three-dimensional systems are shown in Figs. 28 through 30. They exhibit normal fluid behavior except that there is, of course, no phase transition. Lowering the temperature and increasing the system size are both seen to lower the pressure for a given density. In addition, one observes again the rather characteristic inflection of the isotherms. The compressibility curves for the 9-cell two-dimensional system in Fig. 31 correspond to the p–ρ curves in Fig. 29. One clearly discerns anomalous behavior in these curves at temperatures below the "critical" one. This same behavior and the correlation between the inflection of the p–ρ curve on the one hand and the corresponding inversion of the K_T–ρ curve on the other is shown even more clearly in Fig. 32, where these two curves for a 12-cell three-dimensional fluid have been super-

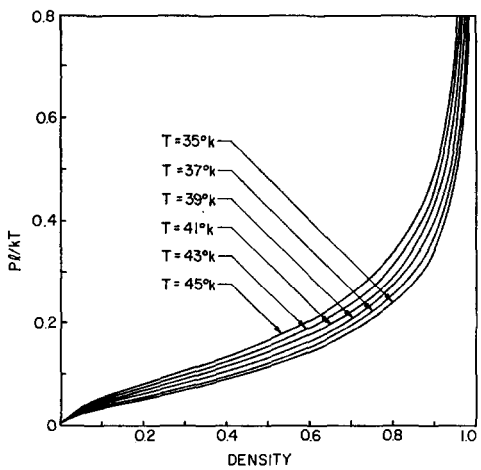


FIG. 29. Pressure–density curves for a nine-cell two-dimensional system, various temperatures; $\zeta = 64k$.

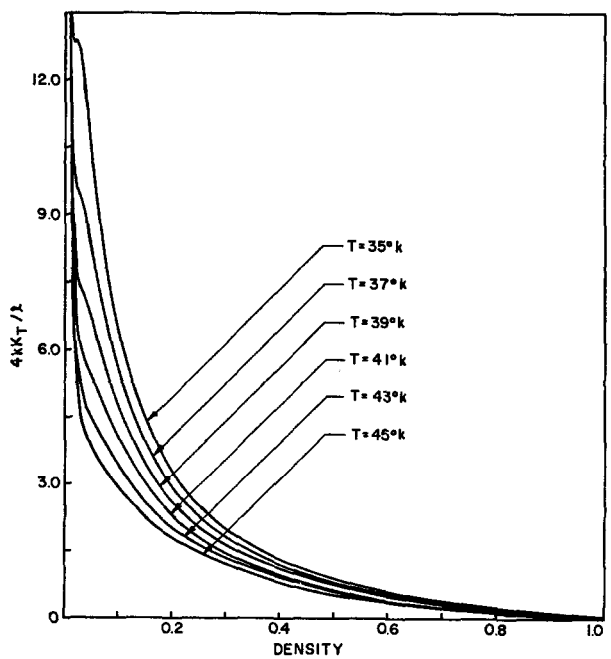


FIG. 31. Compressibility-vs-density isotherms for a nine-cell two-dimensional system; $\zeta = 64k$.

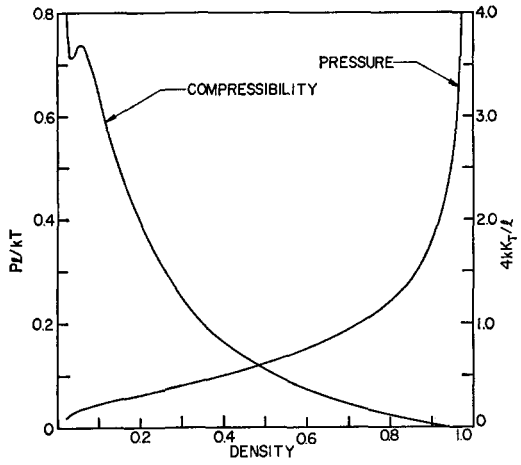


FIG. 32. Pressure-and-compressibility-vs-density isotherms for a 12-cell three-dimensional system.

posed on the same graph. Since we know that two- and three-dimensional systems in the thermodynamic limit are subject to phase transitions, one surmises that this behavior in the pressure-and-compressibility isotherms is indicative of an incipient transition. The analysis in Sec.II.Dand in particular Figs. 11 and 12 lend further credence to this surmise.

IV. CONCLUSIONS

The fluid model examined in this series of two articles was seen to combine mathematical convenience with considerable flexibility. Despite its simplicity it proved capable of yielding realistic and consistent thermodynamic results and of portraying phase transitions correctly. In I it was found that the introduction of a cellular structure made an otherwise difficult problem mathematically much more tractable. The price paid for this simplification was an uncertainty in the instantaneous locations of the particles comprising the system. One could only say that a given particle is located within a certain cell with an uncertainty proportional to the cell parameter. By keeping the cell size small and in fact letting it coincide with the particle's exclusion sphere it was possible to keep this uncertainty within reasonable bounds. The model then represents an approximation to real physical systems but is not exactly equivalent to any such system. This is not the only approximation incorporated in the model. Others include the assumption that the system can be treated classically, the restriction of the configurational energy to pairwise interactions, the adoption of the Lennard–Jones potential with hard-core cutoff for the interaction energy, etc. These latter approximations are fairly standard, however, in theoretical investigations of this type and appear to be quite reasonable for simple nonquantum fluids such as the inert gases. The discrete cellular structure also does not seem to affect the thermodynamic behavior of the model too adversely. In any event, qualitative agreement with the behavior of real fluids is seen to be quite good and even phase transitions can be reproduced in a qualitatively correct manner. Since the basic model investigated here is a one-dimensional one, qualitative agreement is perhaps all that can be expected. Nonetheless, on occasions we found quantitative agreement as well. The critical ratio $\beta_c p_c v_c$ mentioned in Sec. II is a case in point. Such agreement may, of course, be fortuitous.

Despite the relative mathematical simplicity of the model it is surprisingly versatile. We noted this in I already, where we saw how the basic one-dimensional system in which all particles interact with one another can be converted into a two- or three-dimensional one with more limited interactions. It proved possible by virtue of this transformation to obtain series expansions in all dimensions by an identical straight-forward algebraic method. Also, we noted there that the model offers a convenient method for investigating the zeros of the partition function, thereby providing further insight into the analyticity of that function and therefore into the nature of the phase transitions to which it may give rise. Current research is being focused on this area.

In the present article we have explored this versatility of the model further by applying it to various types of systems in order to investigate their thermodynamic behavior. Thus we saw that the model correctly and almost trivially yields the equation of state of an ideal gas and of a Tonks gas in the limit of no interaction and of only a hard-core interaction, respectively. This provided further evidence of the validity of the model, at least in that limit, over and above the analysis in I. The one-dimensional fluid with nearest-neighbor interactions, which was investigated next, offered the unique advantage that the equation of state and all thermodynamic functions can be written in closed form. This system consequently offers a mathematically very tractable model for investigating thermodynamic behavior in a fluid having attractive interactions of limited range combined with a hard repulsive core. Obvious limitations of the system are its one-dimensional nature and the consequent (by virtue of Van Hove's theorem) absence of a phase transition. Nonetheless the behavior of this system turns out to be surprisingly realistic in many ways.

If one is primarily concerned with phase transitions, the one-dimensional fluid with attractive forces of infinite range (i.e., where $\gamma \to 0$) provides an interest-

ing test case. While physically this obviously does not correspond to any real system, it represents an interesting limit exhibiting typical transition behavior. Investigation of finite systems of this type showed how a phase transition is approached gradually with increasing system size without need for recourse to a Maxwell-type construction. In a 240-cell system the pressure–density isotherm is practically indistinguishable from that of a fluid undergoing a true phase transition, while the compressibility curve shows a very realistic approach to a true discontinuity. Exact analysis confirms the numerical results and indicates that in the thermodynamic limit this fluid has an infinite critical temperature.

In investigating finite one-, two-, and three-dimensional systems of quite modest size the series solution of the partition function used in this study proved quite convenient for computer analysis. It was found that small one-dimensional fluid systems interacting with the full Lennard–Jones potential exhibit many of the characteristics of real fluids. There is, however, no phase transition for systems of so small a size. Nevertheless several indicators suggest that if one could investigate larger systems one would see an approach to a phase transition just as one did in the case of the infinite-range fluid:

(1) The maxima of the specific-heat curves seem to approach a limiting critical temperature, other than absolute zero, and also approach a typical shape.

(2) The pressure–density isotherms begin to show a typical inflection that is usually associated with systems approaching a phase transition. The compressibility isotherms display a corresponding inversion which could well presage an approach to a discontinuity.

(3) Analysis of pair-correlation functions with respect to temperature shows quite conclusively that long-range order prevails at low temperatures and changes quite abruptly to short-range order as the "critical" temperature of the system is passed.

Analysis of small two- and three-dimensional systems with nearest-neighbor attractive interactions indicates that the "critical" temperatures obtained from their specific-heat maxima agree within a few percent with those obtained by other methods in the thermodynamic limit. The excellent agreement achieved is rather surprising in view of the very small size of the systems investigated here and the consequently quite pronounced effect of spurious interactions. These systems exhibit the same behavior in the pressure-and-compressibility isotherms as that mentioned in (2) above. Since we know from other evidence that two- and three-dimensional systems in the thermodynamic limit are subject to phase transitions, one can by analogy construe this analogous behavior of the pressure-and-compressibility isotherms as further indication of a possible approach to a phase transition in the one-dimensional case.

The model investigated in this study appears quite promising and worthy of further examination. It would be of particular interest to investigate the general one-dimensional system with all interactions active in the thermodynamic limit. To that end one must either aim at finding a closed-form solution or a valid approximation. Alternatively one should try to extend computer analysis to very much larger systems. These approaches are currently being investigated. Other present research is concerned with critical exponents and it is hoped that some results concerning these will be available in the near future.

ACKNOWLEDGMENTS

One of the authors (R. G. T.) is greatly indebted to the Physics Department, University of Missouri at Rolla, for partial support of this research and to Professor R. E. Lee and his staff for valuable assistance and for so generously making available computer time and the facilities of the UMR Computer Center.

APPENDIX A: THE HARD-CORE POTENTIAL AND CELL PARAMETER

We first wish to show the connection between the hard-core potential and the primed sums that occur in the series solution, Eq. (1.1) of this article or Eq. (4.7) of I. Let us temporarily revoke the convention adopted in I for mathematical convenience that $V_0 = \phi_0 = 0$ (Mod. N). Instead we write the potential explicitly as follows:

$$\eta_s = \begin{Bmatrix} \zeta_{hc} & (s = 0) \\ -\zeta/s^\nu & (s \neq 0) \end{Bmatrix} \quad \text{(Mod. } N\text{)}. \qquad \text{(A1)}$$

where $\zeta_{hc} = +\infty$ is the hard-core potential. The corresponding Boltzmann factor is

$$z_s = \exp\{-\beta\eta_s\} = \begin{Bmatrix} \exp(-\beta\zeta_{hc}) & (s = 0) \\ \exp(\beta\zeta/s^\nu) & (s \neq 0) \end{Bmatrix} \quad \text{(Mod. } N\text{)}. \qquad \text{(A2)}$$

The hard-core part of z_s can be written as $(1 - \delta_{s,0})$ so that the entire Boltzmann factor becomes:

$$z_s = (1 - \delta_{s,0}) \exp(\beta V_s)$$
$$V_s = \begin{cases} 0 & (s = 0) \\ -\zeta/s^\nu & (s \neq 0). \end{cases} \qquad \text{(A3)}$$

It will be observed that V_s is the potential of Paper I with the convention $V_s = \phi_s = 0$, which is seen to arise naturally in this way. Consider now the primed

sums in Eq. (1.1)

$$t_n = (1/n!)\sum_{q_1=1}^{N}{}' \cdots \sum_{q_n=1}^{N}{}' \prod_{\substack{r=1\\(r<s)}}^{n}\prod_{s=1}^{n} x_{(q_r-q_s)}. \quad \text{(A4)}$$

The primes imply that no index may be repeated in any given term of the sum. Hence one can also write

$$t_n = (1/n!)\sum_{q_1=1}^{N} \cdots \sum_{q_n=1}^{N} \prod_{\substack{r=1\\(r<s)}}^{n}\prod_{s=1}^{n}(1-\delta_{(q_r-q_s)})x_{(q_r-q_s)}$$
$$= (1/n!)\sum_{q_1=1}^{N} \cdots \sum_{q_n=1}^{N} \prod_{\substack{r=1\\(r<s)}}^{n}\prod_{s=1}^{n} z_{(q_r-q_s)}, \quad \text{(A5)}$$

where the sums in (A5) are now unrestricted and where Eq. (A3) was used. Thus the primed summations arise from the hard-core repulsive potential and incorporate the exclusion principle that has been built into the model. The expansion of these primed sums is given, for the first few terms, in Sec. 6 of I.

Next, we wish to clarify the relationship that exists between the cell parameter l and the hard-core repulsive potential. Let d be the diameter of the particle's exclusion sphere (the particle's exclusion length in one dimension). Suppose we choose a cell parameter such that $l_0 < d$. Then each particle occupies $-[-d/l_0] = m'$ cells, where [ ] means "nearest integer equal to or less than." The entire system has $L/l_0 = k$ cells where $k > N$, N being given by L/l and l is the usual cell parameter used in this study ($l = d$). Assuming for convenience that m is odd, we define $m = (\frac{1}{2})(m' - 1)$. The appropriate true potential is now given by (A2). In view of the change in cell dimension we must rewrite Eq. (A3) as follows:

$$z_s = (1 - \delta_{s,0} - \delta_{s,1} - \cdots - \delta_{s,m} - \delta_{s,N-1} - \cdots - \delta_{s,N-m}) \exp(\beta v_s). \quad \text{(A6)}$$

Upon introducing this Boltzmann factor into the expression for t_n one finds that one can write

$$t_n = (1/n!)\sum_{q_1=1}^{k} \cdots \sum_{q_n=1}^{k} \prod_{\substack{r=1\\(r<s)}}^{n}\prod_{s=1}^{n}(1-m'\delta_{(q_r-q_s)})x_{(q_r-q_s)}, \quad \text{(A7)}$$

for a fluid of hard rods $x_s = 1$, so that one has:

$$t_n(x_s = 1, \forall s) = (1/n!)k(k-m')(k-2m')\cdots(k-(n-1)m'). \quad \text{(A8)}$$

Letting now $k = Nm'$ (we shall choose l_0 such that N is an integer), we find $t_n = \binom{N}{n}(m')^n$ and the partition function is:

$$Q_k = \sum_{n=0}^{k}\binom{N}{n}(m')^n y^n = \sum_{n=0}^{m'N}\binom{N}{n}\eta^n \quad \text{(A9)}$$

where $y = \exp(2\nu)$ and $\eta = m'y$. Observing that $\binom{N}{n} = 0$ for $n > N$, we obtain:

$$Q_k = \sum_{n=0}^{m'N}\binom{N}{n}\eta^n = (1+\eta)^N = [1 + (m'l_0/\lambda)\xi]^N$$
$$= (1 + l\xi/\lambda)^N = Q_N. \quad \text{(A10)}$$

This is the same as Eq. (2.1) and shows that choosing the cell parameter smaller than the particle's exclusion length is a trivial variation of the choice adopted in this study. It was pointed out in I that a choice of $l > d$ would not be useful in the present context.

APPENDIX B: COMPUTATION OF THE INTERNAL ENERGY DENSITY BY A DIFFERENT METHOD FOR THE ONE-DIMENSIONAL FLUID WITH NEAREST-NEIGHBOR INTERACTIONS

The following analysis serves to check the validity and meaning of Eq. (2.10). Obviously the first term in this equation is the kinetic contribution and reflects the one degree of freedom of the system. It requires no further analysis. To proceed with the other terms we define the following quantities:

N_+ = Expected number of occupied cells = $\langle n \rangle$.
N_- = Expected number of empty cells = $N - \langle n \rangle$.
N_{++} = Expected number of nearest-neighbor cells with both cells occupied.
N_{--} = Expected number of nearest-neighbor cells with both cells empty.
N_{+-} = Expected number of nearest-neighbor cells with one cell occupied and one empty.

The following relations are easily seen to hold between these quantities:

$$2N_{++} + N_{+-} = 2N_+$$
$$2N_{--} + N_{+-} = 2N_- \quad \text{(B1)}$$
$$N_{++} + N_{--} + N_{+-} = N.$$

Let us define a parameter (short-range order parameter) σ as follows:

$$\sigma = (1/N)\left\langle \sum_{r=1}^{N}\sigma_r\sigma_{r+1}\right\rangle = (1/N)(N_{++} + N_{--} - N_{+-})$$
$$= 1 - N_{+-}/N = 1 + 4\eta - 4\rho, \quad \text{(B2)}$$

where $\eta = N_{++}/N$. One can determine σ from the partition function as follows [see Eq. (2.10) of I]:

$$\sigma = A_N/N \sum_{\{\sigma\}}\left\{\sum_{r=1}^{N}\sigma_r\sigma_{r+1} \times \exp\left[\epsilon\sum_{s=1}^{N}\sigma_s\sigma_{s+1} + \nu\sum_{s=1}^{N}\sigma_s\right]\right\}\Big/ Q_N$$
$$= (1/N)[(\partial/\partial\epsilon)\ln \Xi_N]_\nu, \quad \text{(B3)}$$

where $\epsilon = \theta_1 = (\frac{1}{4})\beta\zeta$ and $\Xi_N = A_N^{-1}Q_N$. We have

then, using (2.7),

$$\sigma = (\partial/\partial\epsilon)\{\epsilon + \ln(\cosh\nu) + \ln[1 + (1+\omega)^{\frac{1}{2}}]\}_\nu$$
$$= 1 - [2x_1^{-1}\operatorname{sech}^2\nu]/[1 + \omega + (1+\omega)^{\frac{1}{2}}]. \quad \text{(B4)}$$

For the configurational energy of the system we have:

$$U_c = (\tfrac{1}{4})\left\langle \sum_{r=1}^{N}\sum_{s=1}^{N}(1+\sigma_r)(1+\sigma_{r+s})\right\rangle\phi_s, \quad \text{(B5)}$$

where $\phi_s = -(\zeta/2)(\delta_{s,1} + \delta_{s,N-1})$. Carrying out the summations indicated one obtains

$$U_c = -\zeta[(N/4) + (N/2)(2\rho - 1) + (N/4)\sigma], \quad \text{(B6)}$$

so that, using (B4), we have

$$u_c = U_c/N$$
$$= -\zeta\{\rho - [x_1^{-1}\operatorname{sech}^2\nu]/2[1 + \omega + (1+\omega)^{\frac{1}{2}}]\}. \quad \text{(B7)}$$

This is exactly the configurational part of (2.10). By means of (B1) and (B2) one notes that (B6) can be rewritten

$$U_c = -N(\rho - \tfrac{1}{4} + \tfrac{1}{4}\sigma)\zeta = -\zeta N\eta = -\zeta N_{++}. \quad \text{(B8)}$$

This represents just the interaction of all nearest-neighbor occupied cells, as one would expect on physical grounds.

APPENDIX C: RELATIONSHIP BETWEEN VARIOUS CORRELATION FUNCTIONS

We wish to establish the relationship between the correlation parameter used in this study and defined in (3.7) on the one hand and the usual pair-correlation functions. Attention is confined to the case $\rho = \frac{1}{2}$ for reasons indicated following Eq. (3.7). The *a priori* probability that a cell is occupied or empty is $\frac{1}{2}$. The joint probability that cells i and j have occupation indices σ_i and σ_j, respectively is

$$W(\sigma_i, \sigma_j) = W(\sigma_j)W(\sigma_i/\sigma_j) = (\tfrac{1}{2})W(\sigma_i/\sigma_j), \quad \text{(C1)}$$

where $W(\sigma_i/\sigma_j)$ is the conditional probability that cell i has occupation index σ_i given that cell j has occupation index σ_j. Let us suppose that we can write this conditional probability in the form

$$W(\sigma_i/\sigma_j) = W(\sigma_i)(1 + M_{ij}) = (\tfrac{1}{2})(1 + M_{ij}). \quad \text{(C2)}$$

Then

$$W(\sigma_i, \sigma_j) = (\tfrac{1}{4})(1 + M_{ij}). \quad \text{(C3)}$$

Now, given that cell j has occupation index σ_j, the probability that cell i has the same occupation index and the probability that it has the opposite index are mutually exclusive events. Consequently,

$\langle\sigma_i\sigma_j\rangle$ = (probability that cells i and j have the same occupation index) $(+1)$ + (probability that cells i and j have opposite indices) (-1)

$$= M_{ij},$$

and so

$$W(\sigma_i, \sigma_j) = (\tfrac{1}{4})(1 + \langle\sigma_i\sigma_j\rangle). \quad \text{(C5)}$$

We can therefore write the correlation parameter in the following form, with the help of Eqs. (3.7) and (C5)

$$S_i = (4/N)\sum_{r=1}^{N} W(\sigma_r, \sigma_{r+i}) - 1. \quad \text{(C6)}$$

One readily verifies that this equation yields the correct limiting values of $+1$, -1, and 0 for perfect correlation, perfect anticorrelation, and no correlation respectively.

In terms of the usual pair-correlation function, defined as follows:

$$W(\sigma_i, \sigma_j) = W(\sigma_i)W(\sigma_j)C^{(2)}(\sigma_i, \sigma_j)$$
$$= (\tfrac{1}{4})C^{(2)}(\sigma_i, \sigma_j); \quad \text{(C7)}$$

we can write the correlation parameter in the following form[8]:

$$S_i = (1/N)\sum_{r=1}^{N} C^{(2)}(\sigma_r, \sigma_{r+i}) - 1. \quad \text{(C8)}$$

Or, in terms of the Kirkwood correlation functions[8]:

$$S_i = 1/(N-1)\sum_{r=1}^{N} g^{(2)}(\sigma_r, \sigma_{r+i}) - 1. \quad \text{(C9)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Class of Exact Invariants for Classical and Quantum Time-Dependent Harmonic Oscillators*

H. R. LEWIS, JR.
Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico

(Received 21 September 1967)

A class of exact invariants for oscillator systems whose Hamiltonians are

$$H = (1/2\epsilon)\,[p^2 + \Omega^2(t)q^2]$$

is given in closed form in terms of a function $\rho(t)$ which satisfies

$$\epsilon^2 d^2\rho/dt^2 + \Omega^2(t)\rho - \rho^{-3} = 0.$$

Each particular solution of the equation for ρ determines an invariant. The invariants are derived by applying an asymptotic theory due to Kruskal to the oscillator system in closed form. As a consequence, the results are more general than the asymptotic treatment, and are even applicable with complex $\Omega(t)$ and quantum systems. A generating function is given for a classical canonical transformation to a class of new canonical variables which are so chosen that the new momentum is any particular member of the class of invariants. The new coordinate is, of course, a cyclic variable. The meaning of the invariants is discussed, and the general solution for $\rho(t)$ is given in terms of linearly independent solutions of the equations of motion for the classical oscillator. The general solution for $\rho(t)$ is evaluated for some special cases. Finally, some aspects of the application of the invariants to quantum systems are discussed.

I. INTRODUCTION

The systems to be discussed are those for which there is a Hamiltonian of the form

$$H = \frac{1}{2\epsilon}\,[p^2 + \Omega^2(t)q^2], \tag{1}$$

where q is a canonical coordinate, p is its conjugate momentum, $\Omega(t)$ is an arbitrary complex function of t, and ϵ is a positive real parameter. If the system is a quantum system, then q and p satisfy the usual commutation relation

$$[q, p] = i\hbar. \tag{2}$$

For brevity, we refer to a system whose Hamiltonian is given by Eq. (1) as a time-dependent harmonic oscillator.

For a time-dependent harmonic oscillator there is a class of *exact* invariants I of the form[1,2]

$$I = \tfrac{1}{2}[\rho^{-2}q^2 + (\rho p - \epsilon\rho' q)^2], \tag{3}$$

where ρ is any function of t satisfying

$$\epsilon^2\rho'' + \Omega^2(t)\rho - \rho^{-3} = 0, \tag{4}$$

and the prime denotes differentiation with respect to t. Equations (3) and (4) define a class of invariants because ρ may be *any* particular solution of Eq. (4). The quantity I is an invariant for quantum as well as classical systems.

If Eq. (4) be solved recursively to give ρ as a series in positive powers of ϵ, then that ρ can be substituted into Eq. (3) to give I as a series in positive powers of ϵ. For classical systems with real Ω, that series for I is the usual adiabatic-invariant series whose leading term is proportional to $(\epsilon H)/\Omega$. The study of the adiabatic invariant has received considerable attention in the literature,[3–11] often in connection with the motion of a charged particle in a particular electromagnetic field.[12] Recently, Kruskal[13] has developed a general

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] H. R. Lewis, Jr., Phys. Rev. Letters **18**, 510 (1967); also an *erratum*: Phys. Rev. Letters **18**, 636 (1967).

[2] Professor K. R. Symon has brought to my attention that these invariants are related to quantities used in the analysis of beam oscillations in an alternating-gradient synchrotron. See, for example, E. D. Courant and H. S. Snyder, Ann. Phys. (N.Y.) **3**, 1 (1958).

[3] A. Einstein, Inst. intern. phys. Solvay, Conseil phys., Rapports et discussions, **1**, 450 (1911).

[4] R. M. Kulsrud, Phys. Rev. **106**, 205 (1957).

[5] F. Hertweck and A. Schlüter, Z. Naturforsch. **12a**, 844 (1957).

[6] G. Backus, A. Lenard, and R. Kulsrud, Z. Naturforsch. **15a**, 1007 (1960).

[7] S. Chandrasekhar, *The Plasma in a Magnetic Field*, R. K. M. Landshoff, Ed. (Stanford University Press, Stanford, 1958), p. 3.

[8] A. Lenard, Ann. Phys. (N.Y.) **6**, 261 (1959).

[9] L. M. Garrido, Progr. Theoret. Phys. (Kyoto) **26**, 577 (1961).

[10] J. E. Littlewood, Ann. Phys. (N.Y.) **21**, 233 (1963); also, Ann. Phys. (N.Y.) **29**, 13 (1964).

[11] G. Knorr and D. Pfirsch, Max-Planck-Institut für Physik und Astrophysik (Munich, Germany), Report No. MPI-PAE/Pl.15/65, 1965.

[12] The equations of motion for a classical time-dependent harmonic oscillator with complex q and p are exactly equivalent to the equations of motion for a charged particle moving nonrelativistically in the electromagnetic field for which the scalar potential is zero and the vector potential is $\mathbf{A} = \frac{1}{2}h(t)\mathbf{B}_0 \times \mathbf{r}$, where $h(t)$ is a function of time, $\mathbf{B}_0$ is a constant vector, and $\mathbf{r}$ is the position vector. (See the reference given in Footnote 5.) With these electromagnetic potentials, the magnetic field $\mathbf{B}$ is given by $h(t)\mathbf{B}_0$, and the electric field $\mathbf{E}$ is given by $-(1/2c)(dh/dt)\mathbf{B}_0 \times \mathbf{r}$. The quantities $\Omega(t)$ and ϵ in Eq. (1) are, respectively, $\frac{1}{2}B(t)$ and the ratio of mass to charge of the particle.

There has been some confusion in the literature about the generality with which the equations of motion for a time-dependent

asymptotic theory of a class of nearly periodic classical systems, among which is included the classical time-dependent harmonic oscillator with real Ω. The class of exact invariants defined by Eqs. (3) and (4) was derived as a result of applying Kruskal's asymptotic theory to the classical oscillator with real Ω.[14] It is possible to apply the theory to that system in closed form, in terms of q, p, and the function $\rho(t)$, without demanding the adiabatic results in the limit of small ϵ. As a consequence, from the asymptotic theory we obtain exact, nonasymptotic results that are even applicable with complex Ω and quantum systems. This is apparently the first example of an application of Kruskal's theory in which nonasymptotic results are obtained, or in which the structure of the functions defined in the theory can be examined in detail. Possibly this example can serve as a model for application of the theory to more complicated problems to obtain similar results.

In Sec. II we derive the invariant defined by Eqs. (3) and (4) starting from Kruskal's asymptotic theory. In Sec. III we give a generating function for a classical canonical transformation to new canonical variables which are so chosen that the new momentum is I. The new coordinate is a cyclic variable, as it must be, since I is an invariant. In that section we also discuss the interpretation of I and its relation to the adiabatic invariant series, and we give the general solution for ρ in terms of linearly independent solutions of the equations of motion for a classical oscillator. The function ρ is given explicitly in Sec. IV for some special cases. In Sec. V we show that I is a constant of the motion for a quantum time-dependent oscillator, derive its eigenvalues and eigenstates, and calculate the expectation values of H in those eigenstates. A complete discussion of the quantum system in terms of I will be given later.[15]

II. DERIVATION OF THE INVARIANT STARTING FROM KRUSKAL'S THEORY

A. The Equations in Standard Form

We begin by considering a classical time-dependent harmonic oscillator for which Ω is a real function. In that case Hamilton's equations of motion for the system can be written in the standard form to which Kruskal's asymptotic theory[13] is applicable. The usual form of Hamilton's equations of motion is

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} = \frac{1}{\epsilon} p,$$
$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} = -\frac{1}{\epsilon} \Omega^2(t) q. \tag{5}$$

In order to apply Kruskal's theory, these equations must be replaced by an equivalent first-order autonomous system, all of whose solutions are periodic in the independent variable in the limit $\epsilon = 0$. This can be achieved by introducing a new independent variable s, defined by $s = t/\epsilon$, and treating t as an additional dependent variable. The system of equations so obtained can be represented compactly as

$$d\mathbf{x}/ds = \mathbf{F}(\mathbf{x}, \epsilon), \tag{6}$$

where $\mathbf{x}$ and $\mathbf{F}(\mathbf{x}, \epsilon)$ are vectors defined by

$$\mathbf{x} = (q, p, t),$$
$$\mathbf{F}(\mathbf{x}, \epsilon) = (p, -\Omega^2(t)q, \epsilon). \tag{7}$$

Because t is now a dependent variable, this system is autonomous. The $\epsilon = 0$ system, obtained by replacing $\mathbf{F}(\mathbf{x}, \epsilon)$ in Eq. (6) by $\mathbf{F}(\mathbf{x}, 0)$, is

$$dq/ds = p,$$
$$dp/ds = -\Omega^2(t)q, \tag{8}$$
$$dt/ds = 0.$$

The solution of the last equation is $t = \text{const}$ and, therefore, the other two equations are just the

harmonic oscillator are equivalent to the equations of motion for a particle moving in a uniform but time-dependent magnetic field. (See the references given in Footnotes 5, 7, 9, and 11.) The point that is sometimes overlooked is that there are many possible electric fields that are consistent with the uniform magnetic field. The electric field given above is the only axially symmetric electric field, for which the charge density is everywhere zero, that is consistent with the uniform but time-dependent magnetic field $h(t)\mathbf{B}_0$. However, if the symmetry and charge-free restrictions are removed, then there are other electric fields possible. We can change $\mathbf{E}$ without changing $\mathbf{B}$ by introducing any space- and time-dependent scalar potential φ and by adding to the vector potential the gradient of any space- and time-dependent scalar function χ. The charge density corresponding to the new electric field is proportional to

$$\nabla^2[\varphi + (1/c)\partial\chi/\partial t].$$

Even if this expression vanishes, the new electric field can still differ from the one given above. The equations of motion for a particle moving in the new electromagnetic field are generally not equivalent to the equations of motion for a time-dependent harmonic oscillator.

[13] M. Kruskal, J. Math. Phys. 3, 806 (1962).

[14] See also R. M. Kulsrud, Princeton University Plasma Physics Laboratory Semiannual Report No. MATT-Q-20 (July 1, 1962–Dec. 31, 1962). The existence of this unpublished reference was kindly made known to me by Dr. Clifford Gardner after the present work was completed. The last term of Eq. (5) of the reference should be multiplied by $\frac{1}{4}$, with corresponding changes in subsequent equations. Except for that error, an invariant equivalent to I was derived by Kulsrud in this reference by a different method. His objective was to obtain the asymptotic series representation of the adiabatic invariant of the system in positive powers of ϵ. It was apparently not realized that the result obtained was an exact, nonasymptotic result; that a class of exact invariants had been defined, rather than just the one adiabatic invariant; and that the result is applicable to classical systems with complex Ω and to quantum systems.

[15] H. R. Lewis, Jr., and W. B. Riesenfeld, "An Exact Quantum Theory of the Time-Dependent Harmonic Oscillator and of a Charged Particle in a Time-Dependent Electromagnetic Field" (to be published).

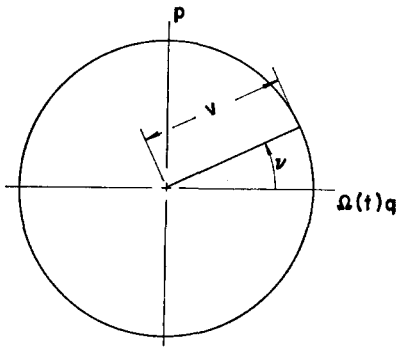


FIG. 1. Definition of the variables v and ν.

harmonic-oscillator equations with a constant frequency. Since Ω is real, this means that the dependent variables of the $\epsilon = 0$ system are all periodic in s with period $2\pi/\Omega(t)$. Therefore, Eq. (6) is indeed of the standard form required by Kruskal's theory.

B. More Appropriate Variables

The next step in applying Kruskal's theory is to transform variables from $\mathbf{x}$ to so-called "more appropriate variables" $\mathbf{y} = \mathbf{Y}(\mathbf{x})$ and $\nu = N(\mathbf{x})$. These variables are to be defined such that $\mathbf{y}$, a two-component vector, is constant on the integral curves of the $\epsilon = 0$ system, and such that ν, an angle variable, changes monotonically by an amount 2π around each of those closed curves. The definition of the more appropriate variables that we use is

$$\mathbf{y} = \mathbf{Y}(\mathbf{x}) = (v, t),$$

and

$$\nu = N(\mathbf{x}) = \tan^{-1}\left[\frac{p}{\Omega(t)q}\right], \tag{9}$$

where

$$v = [\Omega^2(t)q^2 + p^2]^{\frac{1}{2}}.$$

The relation between the variables (v, ν) and the variables (q, p) is shown in Fig. 1. The inverse transformation is given by

$$\mathbf{x} = \left[\frac{v}{\Omega(t)}\cos\nu, v\sin\nu, t\right]. \tag{10}$$

The equations satisfied by $\mathbf{y}$ and ν are[16]

$$\begin{aligned} d\mathbf{y}/ds &= \epsilon\mathbf{g}(\mathbf{y}, \nu), \\ d\nu/ds &= \psi(\mathbf{y}, \nu), \end{aligned} \tag{11}$$

where

$$\epsilon\mathbf{g}(\mathbf{y}, \nu) = \mathbf{f}\cdot\nabla_{\mathbf{x}}\mathbf{Y} = \epsilon\left(v\frac{d\ln\Omega}{dt}\cos^2\nu, 1\right),$$

$$\psi(\mathbf{y}, \nu) = \mathbf{f}\cdot\nabla_{\mathbf{x}}N = -\Omega - \epsilon\frac{d\ln\Omega}{dt}\sin\nu\cos\nu. \tag{12}$$

[16] The quantities ν, N, and ψ correspond, respectively, to the quantities $-2\pi v$, $-2\pi\Upsilon$, and $-2\pi\psi$ defined by Kruskal.

C. Nice Variables

The reason for defining the more appropriate variables is to provide a convenient means for defining one further transformation of variables—from the more appropriate variables to so-called "nice variables,"

$$\mathbf{z} = \mathbf{Z}(\mathbf{y}, \nu),$$

and

$$\varphi = \Phi(\mathbf{y}, \nu). \tag{13}$$

The variable $\mathbf{z}$ is a two-component vector and φ is an angle variable. The functions $\mathbf{Z}$ and Φ are to be chosen such that the equations of motion for the oscillator, written in terms of $\mathbf{z}$ and φ, are

$$\begin{aligned} d\mathbf{z}/ds &= \epsilon\mathbf{h}(\mathbf{z}), \\ d\varphi/ds &= \omega(\mathbf{z}), \end{aligned} \tag{14}$$

where $\mathbf{h}(\mathbf{z})$ and $\omega(\mathbf{z})$ are functions *independent of* φ that are to be determined along with $\mathbf{Z}(\mathbf{y}, \nu)$ and $\Phi(\mathbf{y}, \nu)$.

The motivation for defining nice variables is that we can use them to specify an important set of closed curves known as *rings*. The rings are those curves on which $\mathbf{z}$ is constant, and they are closed because φ is an angle variable. The definition of the rings does not depend on the variable s because, by virtue of the fact that the function $\mathbf{h}(\mathbf{z})$ appearing in the first of Eqs. (14) is independent of φ, the family of rings is continuously mapped into itself if each ring is allowed to change with s according to Eqs. (14). In our oscillator problem it is possible to choose the rings such that they lie in the planes given by $t = \text{const}$, and we shall choose them in that way. As a result, the rings can be used to define an invariant I in terms of an action integral:

$$\begin{aligned} I &= \frac{1}{2\pi}\oint_{\text{ring}} p\,dq \\ &= -\frac{1}{2\pi}\int_0^{2\pi} X_2\frac{\partial X_1}{\partial\varphi}\,d\varphi, \end{aligned} \tag{15a}$$

where

$$\begin{aligned} q &= X_1(\mathbf{z}, \varphi), \\ p &= X_2(\mathbf{z}, \varphi). \end{aligned} \tag{15b}$$

This invariant is the one given by Eqs. (3) and (4).

The equations that must be satisfied by $\mathbf{Z}$, Φ, $\mathbf{h}$, and ω are easily determined. The periodicity conditions on $\mathbf{Z}$ and Φ that result from taking φ and ν to be angle variables are

$$\begin{aligned} \mathbf{Z}(\mathbf{y}, \nu + 2\pi) &= \mathbf{Z}(\mathbf{y}, \nu), \\ \Phi(\mathbf{y}, \nu + 2\pi) &= \Phi(\mathbf{y}, \nu) + 2\pi. \end{aligned} \tag{16}$$

The partial differential equations that must be satisfied subject to these periodicity conditions can be derived by substituting Eqs. (13) into Eqs. (14), performing the differentiations, and using Eqs. (11) to eliminate dy/ds and dv/ds. The equations are

$$\epsilon \mathbf{g} \cdot \nabla_y \mathbf{Z} + \psi \frac{\partial \mathbf{Z}}{\partial v} = \epsilon \mathbf{h}[\mathbf{Z}(\mathbf{y}, v)],$$

$$\epsilon \mathbf{g} \cdot \nabla_y \Phi + \psi \frac{\partial \Phi}{\partial v} = \omega[\mathbf{Z}(\mathbf{y}, v)]. \tag{17}$$

Thus, to calculate the invariant I, we must first solve Eqs. (17) for $\mathbf{Z}$ and Φ subject to the periodicity conditions. Then we must invert the transformation to nice variables to obtain q and p in terms of $\mathbf{z}$ and φ. Finally, we must evaluate the action integral given in Eq. (15a).

Kruskal showed that nice variables can be found which are asymptotic series in positive powers of ϵ, and he gave a prescription for finding the terms of the series recursively. He also showed that the inverse of the asymptotic transformation to nice variables exists and can be constructed. For our oscillator problem it is possible to find a class of transformations to nice variables expressed in closed form. The transformations depend on the function ρ which is defined by Eq. (4), and each particular solution of that equation determines a transformation. The recursive solution of Eq. (4) in positive powers of ϵ gives the asymptotic transformation considered by Kruskal.

Probably the major complication of Eqs. (17) is that the terms to the right of the equality signs are unknown functions of unknown functions, and the first step in solving the equations is to remove that complication. The clue to doing so is obtained by actually carrying out the recursive solutions of Eqs. (17) for two or three terms of the series, and then to notice that the dependences of these terms on their arguments can be represented as follows:

$$\begin{aligned}
\mathbf{Z}(\mathbf{y}, v) &= \{Z_1(\mathbf{y}, v), Z_2(\mathbf{y}, v)\} \\
&= \{v f_1(t, v), t\}, \\
\Phi(\mathbf{y}, v) &= f_2(t, v), \\
\mathbf{h}[\mathbf{Z}(\mathbf{y}, v)] &= \{Z_1(\mathbf{y}, v) k_1(t), 1\} \\
&= \{v f_1(t, v) k_1(t), 1\}, \\
\omega[\mathbf{Z}(\mathbf{y}, v)] &= k_2(t).
\end{aligned} \tag{18}$$

The next step is to realize that these special forms of the unknown functions are completely consistent with Eqs. (17) and with the periodicity conditions. We now take Eqs. (18) as a special ansatz for finding a class of particular solutions of Eqs. (17). By substituting Eqs. (18) into Eqs. (17) we obtain the following equations for $f_1(t, v), f_2(t, v)$, $k_1(t)$, and $k_2(t)$:

$$\begin{aligned}
\epsilon \frac{\partial f_1}{\partial t} + \psi \frac{\partial f_1}{\partial v} &= \epsilon f_1 \left[k_1(t) - \frac{d \ln \Omega}{dt} \cos^2 v \right], \\
\epsilon \frac{\partial f_2}{\partial t} + \psi \frac{\partial f_2}{\partial v} &= k_2(t),
\end{aligned} \tag{19}$$

where, as before,

$$\psi = -\Omega - \epsilon \frac{d \ln \Omega}{dt} \sin v \cos v.$$

The periodicity conditions on $\mathbf{Z}$ and Φ require

$$\begin{aligned}
f_1(t, v + 2\pi) &= f_1(t, v), \\
f_2(t, v + 2\pi) &= f_2(t, v) + 2\pi.
\end{aligned} \tag{20}$$

We have now replaced the problem of solving Eqs. (17) subject to the periodicity conditions on $\mathbf{Z}$ and Φ by that of solving Eqs. (19) subject to the periodicity conditions on f_1 and f_2. By so doing we have achieved an enormous simplification. Firstly, Eqs. (19) do not involve unknown functions of unknown functions. Secondly, the equations for f_1 and f_2 are not coupled. Thirdly, f_1, f_2, k_1, and k_2 do not depend on the variable v.

It is convenient to approach the problem of solving Eqs. (19) by the method of characteristics. The two sets of characteristic equations are

$$\begin{aligned}
\epsilon \frac{dv}{dt} &= \psi, \\
\frac{df_1}{dt} &= f_1 \left[k_1 - \frac{d \ln \Omega}{dt} \cos^2 v \right],
\end{aligned} \tag{21}$$

and

$$\begin{aligned}
\epsilon (dv/dt) &= \psi, \\
\epsilon (df_2/dt) &= k_2.
\end{aligned} \tag{22}$$

We shall solve Eqs. (19) via their characteristic equations by first solving them for a special choice of $\Omega(t)$, and then generalizing that method of solution to include arbitrary choices of Ω. In determining the solutions, for a general Ω as well as for the special choice, we shall always take f_1, f_2, k_1, and k_2 to be $O(1)$ in ϵ, so that they agree with the asymptotic treatment given by Kruskal. This will mean that the function $\rho(t)$ will be taken as a solution of Eq. (4) that is $O(1)$ in ϵ. However, after the problem has been solved with those restrictions, we shall notice that we still have valid solutions for f_1, f_2, k_1, and k_2 even if $\rho(t)$ is taken to be *any* particular solution of Eq. (4), without regard to its order in ϵ.

D. A Special Case

We can learn how to solve the characteristic equations [Eqs. (21) and (22)] with an arbitrary $\Omega(t)$ by first solving them with an $\Omega(t)$ for which the $\epsilon(dv/dt) = \psi$ equation is separable. We choose

$$\Omega = 1/(B - At), \tag{23}$$

where A and B are constants, so that

$$d \ln \Omega/dt = A\Omega.$$

Then Eqs. (22) take the form

$$\epsilon \frac{dv}{dt} = \frac{1}{At - B}(1 + \epsilon A \sin v \cos v),$$
$$\epsilon(df_2/dt) = k_2. \tag{22'}$$

The general solutions of these equations are given by

$$\phi(t, v) = \tan^{-1}\left\{\frac{\tan v + \epsilon(A/2)}{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}\right\} - \frac{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}{\epsilon A} \ln (At - B) = C,$$
$$f_2 - \frac{1}{\epsilon}\int^t k_2(t')\, dt' = \text{const}, \tag{24}$$

where $C = \text{const}$. Thus, the general solution of the second of Eqs. (19) is

$$f_2(t, v) = \frac{1}{\epsilon}\int^t k_2(t')\, dt' + \chi[\phi(t, v)], \tag{25}$$

where χ is an arbitrary function. In order to satisfy $f_2(t, v + 2\pi) = f_2(t, v) + 2\pi$, we take χ such that $\chi(w) = w$. (A constant could be added to this, but it would be redundant, because of the indefinite integral in the expression for f_2.) Then, in order that f_2 be $O(1)$ in ϵ, we take k_2 such that

$$\int^t k_2(t')\, dt' = \frac{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}{A} \ln (At - B) + \epsilon F_2(t),$$

where F_2 is an arbitrary function that is $O(1)$ in ϵ. The expressions for $f_2(t, v)$ and $k_2(t)$ are then

$$f_2(t, v) = \tan^{-1}\left\{\frac{\tan v + \epsilon(A/2)}{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}\right\} + F_2(t),$$
$$k_2(t) = -\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}} \frac{1}{B - At} + \epsilon F_2'(t). \tag{26}$$

In order to find f_1 and k_1, we must integrate Eqs. (21). For the Ω given by Eq. (23), those equations can be written as

$$\epsilon \frac{dv}{dt} = \frac{1}{At - B}(1 + \epsilon A \sin v \cos v),$$
$$df_1/dt = f_1[k_1(t) - A\Omega \cos^2 v]. \tag{21'}$$

The first of these equations has already been solved [Eqs. (24)]. Using that solution, we rewrite the second equation as

$$d \ln f_1/dt = k_1(t) - G(t, C),$$

where

$$G(t, C) = \frac{A\Omega}{1 + \left\{-\epsilon(A/2) + \{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}} \tan\left[C - \frac{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}{\epsilon}\int^t \Omega(t')\, dt'\right]\right\}^2}.$$

The general solution of this equation is given by

$$\ln f_1 - \int^t \{k_1(t') - G[t', \phi(t, v)]\}\, dt' = \text{const}.$$

The integral involving G can be brought into a standard form[17] by changing the variable of integration from t' to

$$\omega = \phi(t, v) - \frac{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}{\epsilon}\int^{t'} \Omega(t'')\, dt''.$$

The result is

$$\int^t G[t', \phi(t, v)]\, dt' = \int^W \frac{\cos^2 \omega}{\{1 + [\epsilon(A/2)]^2\} \cos^2 \omega + \{1 - [\epsilon(A/2)]^2\} \sin^2 \omega - 2[\epsilon(A/2)]\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}} \sin \omega \cos \omega}\, d\omega,$$

where

$$W = \phi(t, v) - \frac{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}{\epsilon}\int^t \Omega(t')\, dt' = \tan^{-1}\left\{\frac{\tan v + \epsilon(A/2)}{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}\right\}.$$

Remembering that any function of $\phi(t, v)$ may be added to the integral, we can write the value of the

[17] W. Gröbner and N. Hofreiter, *Integraltafel, Erster Teil: Unbestimmte Integrale* (Springer-Verlag, Vienna, 1961), 3rd ed., p. 124.

integral as

$$\int^t G[t', \phi(t, \nu)]\, dt'$$
$$= \frac{A}{2}\int^t \Omega(t')\, dt' + \tfrac{1}{2}\ln(\cos^2 W + \{-\epsilon(A/2)\cos W + \{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}\sin W\}^2)$$
$$= \frac{A}{2}\int^t \Omega(t')\, dt' + \tfrac{1}{2}\ln\left[\frac{1+\tan^2\nu}{1+\tan^2 W}\right]. \quad (27)$$

Thus the general solution of the first of Eqs. (19) is

$$\ln f_1 = \int^t \{k_1(t')\, dt' - G[t', \phi(t, \nu)]\}\, dt' + \chi[\phi(t, \nu)]$$
$$= \int^t k_1(t')\, dt' - \frac{A}{2}\int^t \Omega(t')\, dt' - \tfrac{1}{2}\ln\left[\frac{1+\tan^2\nu}{1+\tan^2 W}\right] + \chi[\phi(t, \nu)],$$

where χ is an arbitrary function. In order to satisfy $f_1(t, \nu + 2\pi) = f_1(t, \nu)$, and simultaneously to make f_1 be $O(1)$ in ϵ, we must take χ to be a constant function:

$$\ln f_1 = \int^t k_1(t')\, dt' - \frac{A}{2}\int^t \Omega(t')\, dt' - \tfrac{1}{2}\ln\left[\frac{1+\tan^2\nu}{1+\tan^2 W}\right].$$

We now introduce another function, $F_1(t)$, such that

$$\int^t k_1(t')\, dt' = \frac{A}{2}\int^t \Omega(t')\, dt' + \ln F_1(t),$$

where $F_1(t)$ is an arbitrary function that is $O(1)$ in ϵ. The expressions for $f_1(t, \nu)$ and $k_1(t)$ are then

$$f_1(t, \nu) = F_1(t)\left[\frac{1+\tan^2 W}{1+\tan^2\nu}\right]^{\frac{1}{2}}$$
$$= F_1(t)\left(\frac{1 + \left\{\dfrac{\tan\nu + \epsilon(A/2)}{\{1 - [\epsilon(A/2)]^2\}^{\frac{1}{2}}}\right\}^2}{1+\tan^2\nu}\right)^{\frac{1}{2}},$$
$$k_1(t) = \frac{1}{2}\frac{A}{B - At} + \frac{d\ln F_1}{dt}. \quad (28)$$

Thus, with Eqs. (26) and (28), we have solutions of Eqs. (19) and (20) for our special Ω. We shall generalize these solutions to a wider class of solutions that are also valid for an arbitrary Ω.

E. f_1, f_2, k_1, and k_2 for Arbitrary Ω

We now turn to the solution of Eqs. (19) for arbitrary Ω and, as with the special case, work with the characteristic equations, Eqs. (21) and (22). The initial step is to generalize the first of Eqs. (24) to find a function $\phi(t, \nu)$ such that $\phi(t, \nu) = C = \text{const}$ is an integral of $\epsilon(d\nu/dt) = \psi$. This is equivalent to requiring $\phi(t, \nu)$ to satisfy

$$\epsilon\frac{\partial\phi}{\partial t} + \psi\frac{\partial\phi}{\partial\nu} = 0. \quad (29)$$

It is convenient to replace the variable ν by $u = \tan\nu$, so that Eq. (29) becomes

$$\epsilon\frac{\partial\phi}{\partial t} - \left[\Omega(1+u^2) + \epsilon u\frac{d\ln\Omega}{dt}\right]\frac{\partial\phi}{\partial u} = 0. \quad (30)$$

As a generalization of the first of Eqs. (24), we let $\phi(t, \nu)$ be of the form

$$\phi(t, \nu) = \tan^{-1}\left(\frac{u + \epsilon g_1(t)}{\{1 - [\epsilon g_2(t)]^2\}^{\frac{1}{2}}}\right) + \frac{1}{\epsilon}\int^t \{1 - [\epsilon g_3(t')]^2\}^{\frac{1}{2}}\Omega(t')\, dt',$$

where g_1, g_2, and g_3 are functions to be determined such that ϕ satisfies Eq. (30). If we substitute ϕ into Eq. (30) and clear of fractions, we obtain a quadratic expression in u that must vanish. Setting the coefficients of 1, u, and u^2 equal to zero separately, we obtain three equations for g_1, g_2, and g_3. After some manipulation, we find that these functions can be replaced by the single function $\rho(t)$ defined by Eq. (4). The final expression for $\phi(t, \nu)$ in terms of $\rho(t)$ is

$$\phi(t, \nu) = \tan^{-1}\left[\rho^2\left(\Omega\tan\nu - \epsilon\frac{d\ln\rho}{dt}\right)\right] + \frac{1}{\epsilon}\int^t \rho^{-2}(t')\, dt', \quad (31)$$

where ρ is *any* particular solution of

$$\epsilon^2\rho'' + \Omega^2(t)\rho - \rho^{-3} = 0. \quad (4)$$

The particular solution of Eq. (4) that appears in the first of Eqs. (24) is

$$\rho = \{1 - [\epsilon(A/2)]^2\}^{-\frac{1}{4}}(B - At)^{\frac{1}{2}}.$$

The determination of f_2 and k_2 now proceeds exactly as in the special case. The general expression for f_2 is given by Eq. (25), and we again take χ such that $\chi(w) = w$. Assuming for the moment that $\rho(t)$ is $O(1)$ in ϵ, we choose k_2 such that f_2 is $O(1)$ in ϵ also:

$$\int^t k_2(t')\, dt' = -\int^t \rho^{-2}(t')\, dt' + \epsilon F_2(t),$$

where $F_2(t)$ is an arbitrary function that is $O(1)$ in ϵ. Then the final expressions for $f_2(t, \nu)$ and $k_2(t)$ are

$$f_2(t, \nu) = \tan^{-1}\left[\rho^2\left(\Omega\tan\nu - \epsilon\frac{d\ln\rho}{dt}\right)\right] + F_2(t),$$
$$k_2(t) = -\rho^{-2}(t) + \epsilon F_2'(t). \quad (32)$$

Even if we drop the requirement that $\rho(t)$ and $F_2(t)$ be $O(1)$ in ϵ, Eqs. (32) represent a valid solution of the second of Eqs. (19), with the correct periodicity

condition on $f_2(t, v)$. We do drop that requirement and thereby obtain a more general solution.

We now determine f_1 and k_1. Using the solution of the $\epsilon(dv/dt) = \psi$ equation, we write the second of Eqs. (21) as

$$\frac{d \ln f_1}{dt} = k_1(t) - G(t, C),$$

where

$$G(t, C) = \frac{d \ln \Omega}{dt} \frac{1}{1 + \Omega^{-2}\left\{\epsilon \frac{d \ln \rho}{dt} + \rho^{-2} \tan\left[C - \frac{1}{\epsilon}\int^t \rho^{-2}(t')\, dt'\right]\right\}^2}.$$

As in the special case, the solution of this equation is given by

$$\ln f_1 - \int^t \{k_1(t') - G[t', \phi(t, v)]\}\, dt' = \text{const.}$$

The integral involving $G[t', \phi(t, v)]$ can be represented by a formula similar to Eq. (27), which holds for the special case. In particular, it can be represented by

$$\int^t G[t', \phi(t, v)]\, dt' = \int^t r_1(t')\, dt' + \tfrac{1}{2} \ln \{\cos^2 W + [-r_2(t) \cos W + r_3(t) \sin W]^2\},$$

where

$$W = \phi(t, v) - \frac{1}{\epsilon}\int^t \rho^{-2}(t')\, dt', \tag{33}$$

and where r_1, r_2, and r_3 are functions to be determined. By differentiating this formula and clearing of fractions, we obtain a homogeneous quadratic equation in $\sin W$ and $\cos W$. Satisfying the equation separately for the coefficients of $\sin^2 W$, $\sin W \cos W$, and $\cos^2 W$, we obtain equations for r_1, r_2, and r_3. After some manipulation, we find that the integral can be represented in terms of $\rho(t)$, with no further need of r_1, r_2, and r_3. The result is

$$\int^t G[t', \phi(t, v)]\, dt' = \ln \Omega - \tfrac{1}{2} \ln \rho^{-2} + \tfrac{1}{2} \ln \left\{\cos^2 W + \Omega^{-2}\left[\epsilon \frac{d \ln \rho}{dt} \cos W + \rho^{-2} \sin W\right]^2\right\}. \tag{34}$$

Using this formula, we can now write the general solution of the first of Eqs. (19) as

$$\begin{aligned} \ln f_1 &= \int^t \{k_1(t') - G[t', \phi(t, v)]\}\, dt' + \chi[\phi(t, v)] \\ &= \int^t k_1(t')\, dt' - \ln \Omega + \tfrac{1}{2} \ln \rho^{-2} \\ &\quad - \tfrac{1}{2} \ln \left\{\cos^2 W + \Omega^{-2}\left[\epsilon \frac{d \ln \rho}{dt} \cos W + \rho^{-2} \sin W\right]^2\right\} + \chi[\phi(t, v)] \\ &= \int^t k_1(t')\, dt' - \ln \Omega + \tfrac{1}{2} \ln \rho^{-2} \\ &\quad - \tfrac{1}{2} \ln \left[\frac{1 + \tan^2 v}{1 + \tan^2 W}\right] + \chi[\phi(t, v)], \end{aligned} \tag{35}$$

where χ is an arbitrary function. As in the special case, we take χ to be a constant function, and we choose k_1 such that

$$\int^t k_1(t')\, dt' = \ln \Omega - \tfrac{1}{2} \ln \rho^{-2} + \ln F_1(t),$$

where F_1 is an arbitrary function. We no longer require ρ, F_1, f_1, or k_1 to be $O(1)$ in ϵ. The final expressions for $f_1(t, v)$ and $k_1(t)$ are

$$\begin{aligned} f_1(t, v) &= F_1(t)\left\{\frac{1 + \tan^2 W}{1 + \tan^2 v}\right\}^{\frac{1}{2}} \\ &= F_1(t)\left\{\frac{1 + \left[\rho^2\left(\Omega \tan v - \epsilon \frac{d \ln \rho}{dt}\right)\right]^2}{1 + \tan^2 v}\right\}^{\frac{1}{2}}, \\ k_1(t) &= \frac{d}{dt} \ln [F_1 \Omega \rho]. \end{aligned} \tag{36}$$

F. The Invariant

The transformation from $\mathbf{x} = (q, p, t)$ to nice variables $\mathbf{z} = (z_1, z_2)$ and φ, which is defined by Eqs. (9), (13), (18), (32), and (36), can now be written as

$$\begin{aligned} z_1 &= F_1(t)\Omega(t)\left\{q^2 + \rho^4\left[p - \epsilon q \frac{d \ln \rho}{dt}\right]^2\right\}^{\frac{1}{2}}, \\ z_2 &= t, \\ \varphi &= \tan^{-1}\left[\rho^2\left(\frac{p}{q} - \epsilon \frac{d \ln \rho}{dt}\right)\right] + F_2(t). \end{aligned} \tag{37}$$

The rings, defined by $\mathbf{z} = \text{const}$, are simply ellipses in the planes defined by $t = \text{const}$. The inverse transformation, from the nice variables to $\mathbf{x}$, is

$$\begin{aligned} q = X_1(\mathbf{z}, \varphi) &= \pm \frac{z_1}{F_1 \Omega [1 + \tan^2(\varphi - F_2)]^{\frac{1}{2}}}, \\ p = X_2(\mathbf{z}, \varphi) &= \pm \frac{z_1\left[\epsilon \frac{d \ln \rho}{dt} + \rho^{-2} \tan(\varphi - F_2)\right]}{F_1 \Omega [1 + \tan^2(\varphi - F_2)]^{\frac{1}{2}}}, \\ t &= z_2. \end{aligned} \tag{38}$$

At this point it is easy to calculate the invariant I by means of Eq. (15a). The result is

$$I = \frac{1}{2}\left(\frac{z_1}{F_1 \Omega \rho}\right)^2,$$

or, expressed in terms of q and p,

$$I = \tfrac{1}{2}[\rho^{-2}q^2 + (\rho p - \epsilon\rho' q)^2]. \tag{3}$$

Although, in order to satisfy the requirements of Kruskal's asymptotic theory, we originally assumed Ω to be real, it is clear that the derivation of the invariant in no way depends on that. As a matter of fact, it is easy to verify $dI/dt = 0$ for the general case of Ω, q, and p complex, with *any* particular solution of Eq. (4), by differentiating Eq. (3), using Eqs. (5) to eliminate dq/dt and dp/dt, and using Eq. (4) to eliminate $d^2\rho/dt^2$. We shall see later that I is also an invariant of the quantum time-dependent harmonic oscillator even if Ω is complex. Throughout the remainder of this paper, we shall allow Ω to be complex, except where the contrary is explicitly stated.

III. GENERAL DISCUSSION OF I AND ρ

A. A Generating Function

Having found the explicit form of the invariant I, it is possible to define a transformation to new canonical variables in which the new momentum is the invariant itself. A generating function of the canonical transformation can also be found.[18] Kruskal[13] showed that the new canonical coordinate Q and its conjugate momentum P can be taken as

$$\begin{aligned} Q &= -\varphi = -\tan^{-1}\left[\rho^2\left(\frac{p}{q} - \epsilon\frac{d\ln\rho}{dt}\right)\right], \\ P &= I = \tfrac{1}{2}[\rho^{-2}q^2 + (\rho p - \epsilon\rho' q)^2]. \end{aligned} \tag{39}$$

[The arbitrary function $F_2(t)$ has been arbitrarily set equal to zero.] It is simple to check that the Poisson bracket of Q with P satisfies

$$[Q, P] = 1$$

for any $\rho(t)$, even if $\rho(t)$ does not satisfy Eq. (4).

The transformation to the variables Q and P can be generated by a function $F(q, Q, t)$ by the following scheme[19]:

$$\begin{aligned} p &= \partial F/\partial q, \\ P &= -(\partial F/\partial Q), \\ K &= H + (\partial F/\partial t), \end{aligned} \tag{40}$$

where K is the Hamiltonian appropriate to the new canonical variables. The first two of these equations may be integrated to give $F(q, Q, t)$ by expressing p and P in terms of q, Q, and t:

$$\begin{aligned} p &= \rho^{-2}q(\epsilon\rho\rho' - \tan Q), \\ P &= \tfrac{1}{2}\rho^{-2}q^2(1 + \tan^2 Q). \end{aligned}$$

When that is done, we find that a suitable generating function and new Hamiltonian are

$$\begin{aligned} F(q, Q, t) &= -\tfrac{1}{2}\rho^{-2}q^2(\tan Q - \epsilon\rho\rho'), \\ K &= (1/\epsilon)\,\rho^{-2}P. \end{aligned} \tag{41}$$

Equation (4) is necessary in deriving this expression for K. Since K does not involve Q, we see again that P is an invariant provided that ρ satisfies Eq. (4). Hamilton's equations of motion written in terms of Q and P are

$$\begin{aligned} \frac{dQ}{dt} &= \frac{\partial K}{\partial P} = \frac{1}{\epsilon}\rho^{-2}, \\ \frac{dP}{dt} &= -\frac{\partial K}{\partial Q} = 0. \end{aligned} \tag{42}$$

It is interesting that the first of these equations can be written as

$$\epsilon[P^{\frac{1}{2}}\rho]^2\frac{dQ}{dt} = P.$$

This is of the same form as the equation expressing conservation of angular momentum for a particle of mass ϵ moving in an axially symmetric force field, if we interpret $(P^{\frac{1}{2}}\rho)$ as a generalized radius, Q as a generalized angle, and P as a generalized angular momentum.

B. Interpretation of I

To within a constant factor, the invariant I is the most general invariant of a time-dependent harmonic oscillator that is a homogeneous quadratic form in q and p. This is because there is a two-parameter family of functions $\rho(t)$ that satisfy Eq. (4). By writing the most general such invariant in terms of two linearly independent solutions, $f(t)$ and $g(t)$, of

$$\epsilon^2\frac{d^2q}{dt^2} + \Omega^2(t)q = 0, \tag{43}$$

and then using the equivalence between I and the general homogeneous quadratic invariant, it is possible to write the general solution of the nonlinear equation for $\rho(t)$ in terms of $f(t)$ and $g(t)$.

We can write the general solution of Eqs. (5) in terms of f and g as

$$\begin{aligned} q &= D_1 f + D_2 g, \\ p &= \epsilon(D_1 f' + D_2 g'), \end{aligned}$$

[18] More than one generating function can be found, one of which is given in Ref. 1. A much simpler generating function is given in the present paper.

[19] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publ. Co., Inc., Cambridge, Mass., 1950), p. 240.

where D_1 and D_2 are arbitrary complex constants. Solving these equations for D_1 and D_2 in terms of q, p, f, and g, we obtain

$$D_1 = (\epsilon\alpha)^{-1}(\epsilon g'q - gp),$$
$$D_2 = (\epsilon\alpha)^{-1}(fp - \epsilon f'q),$$

where α is a constant (the Wronskian) defined by

$$\alpha = fg' - gf'. \tag{44}$$

Using these expressions for the constants D_1 and D_2, we can write the most general invariant that is a homogeneous quadratic form in q and p as

$$A^2D_1^2 + B^2D_2^2 + C^2D_1D_2,$$

where A, B, and C are arbitrary complex constants. The allegation is that this is proportional to I. That is,

$$A^2D_1^2 + B^2D_2^2 + C^2D_1D_2 = 2E^2I,$$

where E is another arbitrary complex constant. By expanding both sides of this equation and comparing terms, it is seen that the equation is indeed correct, provided that A, B, C, and E are related by

$$\left(\frac{C}{E}\right)^4 = 4\left[\left(\frac{A}{E}\right)^2\left(\frac{B}{E}\right)^2 - (\epsilon\alpha)^2\right],$$

and provided that ρ is given by

$$\rho = \gamma_1(\epsilon\alpha)^{-1}\left\{\left(\frac{A}{E}\right)^2 g^2 + \left(\frac{B}{E}\right)^2 f^2 + 2\gamma_2\left[\left(\frac{A}{E}\right)^2\left(\frac{B}{E}\right)^2 - (\epsilon\alpha)^2\right]^{\frac{1}{2}} fg\right\}^{\frac{1}{2}},$$

where

$$\gamma_1 = \pm 1$$

and

$$\gamma_2 = \pm 1. \tag{45}$$

This is the general solution of Eq. (4), written in terms of $f(t)$, $g(t)$, and the constants (A/E), (B/E), γ_1, and γ_2. In Sec. IV we shall use Eq. (45) to work some examples explicitly.

Before leaving the question of the interpretation of I, we make note of an interesting connection between the *classical* time-dependent harmonic oscillator and the Schrödinger equation. Equation (43) is of the same form as the time-independent, one-dimensional Schrödinger equation if we let t represent the spatial coordinate and q represent the wavefunction. For bound states Ω is imaginary, and for continuum states Ω is real. Thus, the invariant I specifies a relation between the wavefunction and its first derivative.

C. The Adiabatic Invariant Series

It is possible to solve Eq. (4) recursively to give ρ as a series in positive powers of ϵ^2 or $(1/\epsilon^2)$. If we take that solution in positive powers of ϵ^2 whose leading term is $\Omega^{-\frac{1}{2}}$, substitute it into Eq. (3), and expand the result in powers of ϵ, then we obtain the usual adiabatic invariant series. As we shall see in Sec. IV, there are examples in which the solutions of Eq. (4) that are $O(1)$ in ϵ can also be obtained in analytic, closed form. In some examples at least, the analytic solutions have series expansions in positive powers of ϵ^2 which are convergent. Sometimes the radius of convergence is even infinite, as is the case when the analytic solution is a polynomial in ϵ^2. When there is a convergent series expansion of that ρ which approaches $\Omega^{-\frac{1}{2}}$ as ϵ approaches zero, then the adiabatic invariant series is a *convergent* expansion of an *exact* invariant, usually with a different radius of convergence. It would be interesting to know how general these results are. In any event they give new insight into the nature of the adiabatic invariant series.

The recursive solution of Eq. (4) in positive powers of ϵ^2 whose leading term is $\Omega^{-\frac{1}{2}}$ is given in the Appendix to order ϵ^6.

IV. EXAMPLES

In this section we use Eq. (45) to calculate ρ explicitly for some soluble cases.

A. $\Omega = \text{const}$

We take as the linearly independent solutions of Eq. (43):

$$f = \sin\left(\frac{\Omega}{\epsilon}t\right), \quad g = \cos\left(\frac{\Omega}{\epsilon}t\right).$$

The Wronskian is

$$\alpha = -(\Omega/\epsilon),$$

and the general solution for ρ is

$$\rho = \gamma_1\Omega^{-1}\left\{\left(\frac{A}{E}\right)^2\cos^2\left(\frac{\Omega}{\epsilon}t\right) + \left(\frac{B}{E}\right)^2\sin^2\left(\frac{\Omega}{\epsilon}t\right) + 2\gamma_2\left[\left(\frac{A}{E}\right)^2\left(\frac{B}{E}\right)^2 - \Omega^2\right]^{\frac{1}{2}}\sin\left(\frac{\Omega}{\epsilon}t\right)\cos\left(\frac{\Omega}{\epsilon}t\right)\right\}^{\frac{1}{2}}.$$

If we want ρ to be $O(1)$ in ϵ, then we must take

$$(A/E)^2 = (B/E)^2 = \pm\Omega.$$

In that case ρ becomes

$$\rho = \gamma_1(\gamma_2\Omega)^{-\frac{1}{2}}.$$

B. $\Omega = bt^{-1}$

In this case we take the linearly independent solutions of Eq. (43) to be

$$f = t^{\frac{1}{2}+\lambda}, \quad g = t^{\frac{1}{2}-\lambda},$$

where

$$\lambda = \tfrac{1}{2}[1 - (2b/\epsilon)^2]^{\frac{1}{2}}.$$

The Wronskian is

$$\alpha = -2\lambda,$$

and the general solution for ρ is

$$\rho = \gamma_1(2\epsilon\lambda)^{-1}\left\{\left(\frac{A}{E}\right)^2 t^{1-2\lambda} + \left(\frac{B}{E}\right)^2 t^{1+2\lambda} + 2\gamma_2\left[\left(\frac{A}{E}\right)^2\left(\frac{B}{E}\right)^2 - (2\epsilon\lambda)^2\right]^{\frac{1}{2}} t\right\}^{\frac{1}{2}}.$$

For $\epsilon < 2b$, λ is imaginary, so that

$$t^{2\lambda} = t^{i2|\lambda|} = e^{i2|\lambda| \ln t}.$$

If we want ρ to be $O(1)$ in ϵ, then we must take

$$A/E = B/E = 0,$$

in which case ρ becomes

$$\rho = \gamma_1 i^{\frac{3}{2}+\frac{1}{2}\gamma_2} b^{-\frac{1}{2}}[1 - (\epsilon/2b)^2]^{-\frac{1}{4}} t^{\frac{1}{2}}.$$

C. $\Omega = bt^{m/2}$, $m \neq -2$

The solution of this example is in terms of Bessel functions.[20] We take the f and g to be

$$f = t^{\frac{1}{2}} J_\beta(y), \quad g = t^{\frac{1}{2}} N_\beta(y),$$

where

$$y = \frac{2b}{\epsilon(m+2)} t^{(m+2)/2}$$

and

$$\beta = 1/(m+2).$$

The Wronskian is

$$\alpha = (m+2)/\pi,$$

and the general solution for ρ is

$$\rho = \gamma_1 \frac{\pi}{\epsilon(m+2)} t^{\frac{1}{2}}\left\{\left(\frac{A}{E}\right)^2 N_\beta^2(y) + \left(\frac{B}{E}\right)^2 J_\beta^2(y) + 2\gamma_2\left[\left(\frac{A}{E}\right)^2\left(\frac{B}{E}\right)^2 - \frac{\epsilon^2(m+2)^2}{\pi^2}\right]^{\frac{1}{2}} J_\beta(y) N_\beta(y)\right\}^{\frac{1}{2}}.$$

As ϵ approaches zero, y approaches infinity. By using the asymptotic forms of J_β and N_β, we find that we must take

$$(A/E)^2 = (B/E)^2 = \pm[\epsilon(m+2)]/\pi$$

if we want ρ to be $O(1)$ in ϵ. In that case ρ becomes

$$\rho = \gamma_1\left[\frac{\gamma_2\pi}{\epsilon(m+2)}\right]^{\frac{1}{2}} t^{\frac{1}{2}}[H_\beta^{(1)}(y) H_\beta^{(2)}(y)]^{\frac{1}{2}}, \quad (46)$$

where

$$H_\beta^{(1)} = J_\beta + iN_\beta \quad \text{and} \quad H_\beta^{(2)} = J_\beta - iN_\beta.$$

We obtain an interesting specialization of Eq. (46) if we choose

$$\frac{m}{2} = \frac{1}{2n+1} - 1, \quad n = 0, \pm 1, \pm 2, \cdots,$$

so that Ω, β, and y are given by

$$\Omega = bt^{-2n/(2n+1)}, \quad \beta = n + \tfrac{1}{2},$$

$$y = \frac{b}{\epsilon}(2n+1) t^{1/(2n+1)}.$$

Then Eq. (46) becomes

$$\rho = \gamma_1 \gamma_2^{\frac{1}{2}} b^{-\frac{1}{2}} t^{n/(2n+1)} |G(t, \epsilon)|^2,$$

where

$$G(t, \epsilon) = \left[\sum_{k=0}^{n} (-1)^k \frac{(n+k)!}{k!\,(n-k)!} \times \left(\frac{\epsilon}{2ib(2n+1)}\right)^k t^{-k/(2n+1)}\right]^{\frac{1}{2}}.$$

This ρ is a *polynomial* in ϵ^2 and, therefore, its expansion in powers of ϵ^2 has an infinite radius of convergence.

V. QUANTUM SYSTEMS

In this section we still consider systems whose Hamiltonians are given by Eq. (1), but we now require that q and p satisfy the commutation relation

$$[q, p] = i\hbar. \quad (2)$$

By using this commutation relation, it is straightforward to show that the invariant I, now considered as an operator, is also a constant of the motion for the quantum system for any function ρ that satisfies Eq. (4). That is, I satisfies

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + \frac{1}{i\hbar}[I, H] = 0. \quad (47)$$

Therefore, I has eigenstates whose eigenvalues are time-independent.

We now restrict attention to the special situation in which $\rho(t)$ is a real function, as is always possible if Ω^2 is real. In that case the eigenstates and eigenvalues of I can be found by a method that is completely analogous to the method introduced by Dirac[21] for finding the eigenstates and eigenvalues of the Hamiltonian for a harmonic oscillator (Ω real and time-independent). We first introduce "raising" and

[20] E. Jahnke *et al.*, *Tables of Higher Functions* (McGraw-Hill Book Company, New York, 1960), 6th ed.

[21] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1947), 3rd ed.

"lowering" operators, $a^\dagger$ and a, that are defined by

$$a^\dagger = \frac{1}{(2\hbar)^{\frac{1}{2}}}[\rho^{-1}q - i(\rho p - \epsilon\rho' q)],$$

$$a = \frac{1}{(2\hbar)^{\frac{1}{2}}}[\rho^{-1}q + i(\rho p - \epsilon\rho' q)]. \quad (48)$$

These operators satisfy the relations

$$[a, a^\dagger] = 1,$$
$$\hbar a a^\dagger = I + \tfrac{1}{2}\hbar. \quad (49)$$

The operator a operating on an eigenstate of I produces an eigenstate of I whose eigenvalue is $\hbar$ lower than that of the original eigenstate. Similarly, $a^\dagger$ acting on an eigenstate of I increases the eigenvalue by $\hbar$. Once these properties are established, the normalizability of the eigenstates of I can be used to demonstrate that the eigenvalues of I are $(n + \frac{1}{2})\hbar$, where n is zero or a positive integer. Letting $|n\rangle$ denote the normalized eigenstate of I whose eigenvalue is $(n + \frac{1}{2})\hbar$, we can express the relation between $|n + 1\rangle$ and $|n\rangle$, to within an arbitrary time-dependent phase factor, as

$$|n + 1\rangle = (n + 1)^{-\frac{1}{2}} a^\dagger |n\rangle. \quad (50)$$

The condition which determines the state whose eigenvalue is $\frac{1}{2}\hbar$ is

$$a|0\rangle = 0. \quad (51)$$

The wavefunction in the q representation associated with the state whose eigenvalue is $\frac{1}{2}\hbar$ can be calculated easily. Denote that wavefunction by Ψ_0. It is determined by the differential equation that is obtained by writing Eq. (51) in the q representation:

$$\{\rho^{-1}q + i[-i\hbar\rho(\partial/\partial q) - \epsilon\rho' q]\}\Psi_0 = 0.$$

The normalized solution of this equation is

$$\Psi_0(q, t) = e^{i\delta_0}(\pi\hbar\rho^2)^{-\frac{1}{4}} \exp\left[-\frac{q^2}{2\hbar\rho^2}(1 - i\epsilon\rho\rho')\right], \quad (52)$$

where δ_0 is an arbitrary real quantity that may depend on t but is independent of q.

By means of Eqs. (49), we can use operator techniques to calculate the expectation value of the Hamiltonian in a state $|n\rangle$. The result is

$$\langle n| H |n\rangle = \frac{1}{2\epsilon}(\rho^{-2} + \Omega^2\rho^2 + \epsilon^2\rho'^2)(n + \tfrac{1}{2})\hbar. \quad (53)$$

It is interesting to note that the expectation values of H are equally spaced at every instant and that the lowest value is always obtained with $n = 0$, just as with the harmonic oscillator.

All of the quantum-mechanical results reduce to the usual ones for a harmonic oscillator if we take Ω to be real, positive, and constant, and take $\rho = \Omega^{-\frac{1}{2}}$, so that $I = \epsilon H/\Omega$. A further and more complete discussion of the application of the invariant I to quantum systems will be given in a later paper.[15]

APPENDIX: SERIES FOR ρ IN POWERS OF ϵ^2

The recursive solution of Eq. (4) in positive powers of ϵ^2 was carried out and checked to order ϵ^6 with the IBM 7094 computer, using the IBM FORMAC computing system. The leading term was chosen to be $\Omega^{-\frac{1}{2}}$. The result is

$$\rho = \rho_0 + \rho_2\epsilon^2 + \rho_4\epsilon^4 + \rho_6\epsilon^6 + \cdots.$$

In the following expressions for the coefficients, derivatives with respect to t are indicated by the notation

$$\Omega^{(n)} = d^n\Omega/dt^n.$$

The coefficients are

$$\rho_0 = \Omega^{-\frac{1}{2}},$$
$$\rho_2 = \tfrac{1}{8}\Omega^{-\frac{7}{2}}\Omega^{(2)} - \tfrac{3}{16}\Omega^{-\frac{9}{2}}\Omega^{(1)^2},$$
$$\rho_4 = -\tfrac{1}{32}\Omega^{-\frac{11}{2}}\Omega^{(4)} + \tfrac{5}{16}\Omega^{-\frac{13}{2}}\Omega^{(3)}\Omega^{(1)} - \tfrac{207}{128}\Omega^{-\frac{15}{2}}\Omega^{(2)}\Omega^{(1)^2} + \tfrac{29}{128}\Omega^{-\frac{13}{2}}\Omega^{(2)^2} + \tfrac{621}{512}\Omega^{-\frac{17}{2}}\Omega^{(1)^4},$$
$$\rho_6 = \tfrac{1}{128}\Omega^{-\frac{15}{2}}\Omega^{(6)} - \tfrac{21}{128}\Omega^{-\frac{17}{2}}\Omega^{(5)}\Omega^{(1)} - \tfrac{83}{256}\Omega^{-\frac{17}{2}}\Omega^{(4)}\Omega^{(2)} + \tfrac{879}{512}\Omega^{-\frac{19}{2}}\Omega^{(4)}\Omega^{(1)^2} + \tfrac{709}{128}\Omega^{-\frac{19}{2}}\Omega^{(3)}\Omega^{(2)}\Omega^{(1)} - \tfrac{2883}{256}\Omega^{-\frac{21}{2}}\Omega^{(3)}\Omega^{(1)^3} - \tfrac{49}{256}\Omega^{-\frac{17}{2}}\Omega^{(3)^2} + \tfrac{2060637}{40096}\Omega^{-\frac{23}{2}}\Omega^{(2)}\Omega^{(1)^4} - \tfrac{456615}{2048}\Omega^{-\frac{21}{2}}\Omega^{(2)^2}\Omega^{(1)^2} + \tfrac{12887}{1024}\Omega^{-\frac{19}{2}}\Omega^{(2)^3} - \tfrac{2060637}{8192}\Omega^{-\frac{25}{2}}\Omega^{(1)^6}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 11 NOVEMBER 1968

Representations of the Three-Dimensional Rotation Group in Terms of Direction and Angle of Rotation

MOSHE CARMELI*
General Physics Research Laboratory, Aerospace Research Laboratories
Wright-Patterson Air Force Base, Ohio

(Received 1 March 1968)

We find the irreducible representations of the three-dimensional pure rotation group by Weyl's method, which makes use of the homomorphism of the special unitary group of order two onto the whole three-dimensional rotation group. The representations are realized, however, in terms of the angle of rotation in a specified direction and the spherical angles of the direction of the rotation rather than in terms of the familiar Euler angles. The results are then compared with those obtained by different methods and the advantages of the present technique are pointed out. We also derive the differential operators corresponding to infinitesimal rotations about the coordinate axis in terms of the new variables.

1. INTRODUCTION

Recently, the irreducible representations of the rotation group O_3 were given in terms of a parametrization of the rotation through the finite angle and direction of the rotation.[1-3] The expression obtained for the matrix elements of the operator $\exp(i\psi\mathbf{n}\cdot\mathbf{J})$,

$$(m|\exp(i\psi\mathbf{n}\cdot\mathbf{J})|n), \tag{1.1}$$

where $\mathbf{J}=(J_1,J_2,J_3)$ are the usual infinitesimal generators of O_3, was shown to be rather simple. In Eq. (1.1), ψ is the angle of rotation in a direction specified by the unit vector $\mathbf{n}$, the latter being expressed by its spherical angles

$$\mathbf{n}=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta). \tag{1.2}$$

Such a parametrization is possible since every rotation can be realized as a rotation about some axis through a suitable angle ψ. In all previous discussions on the representations of O_3 the parameters used to be taken as the Euler angles.

The new representation was also shown to be useful in discussing the behavior of wavefunctions under rotation in a prescribed direction. Presumably, for many applications one would prefer a parametrization in terms of the new variables instead of the Euler angles.[2]

Consequently, the orthogonality relations between the matrices (1.1) were demonstrated. Also, two sets of Hermitian operators with the usual commutation relations were introduced and their action on the matrices (1.1) was found.[3]

In this paper we deduce the new form by an alternative method which was originally suggested by Weyl and has been widely adopted as a method for finding the representations of O_3 when Euler's angles are used.[4] The method is based on the fact that O_3 is homomorphic to the unimodular unitary group of order two (SU_2), such that to every rotation $g\in O_3$ there correspond two matrices $\mp u\in SU_2$ and, conversely, to every $u\in SU_2$ there corresponds some rotation $g\in O_3$. It thus follows that the description of the representation of O_3 is equivalent to that of SU_2; a representation $g\to T_g$ of O_3 is single- or double-valued according to whether or not T_u is equal to T_{-u}.[5]

We point out that, by using Weyl's method, one can obtain a general invariant result that is a function of $u\in SU_2$, valid for any parametrization one uses to describe the rotation [see Eq. (3.9) below]. To find the representations of O_3 in terms of the new parameters, one has merely to express u in terms of these parameters, as is the case when the Euler angles are employed. This, indeed, can easily be done (see Sec. 2 below). Thus the present method is much easier for the interested physicist to follow than the previous works[1-3] since it is most closely related to the usual derivation of the representation, such as Wigner's,[4] when the Euler angles are used; the two forms become two particular realizations of the same representation.

In addition, by having the results as functions over the group SU_2, certain relations will be obtained which are then invariant under change of the parameters. As an example, the orthogonality relations between the matrix elements of the irreducible representation can be written as an invariant integral[6] over SU_2 [see Eq.

* This work was initiated while the author was at the Department of Physics and Astronomy, University of Maryland, College Park, Maryland.

[1] H. E. Moses, Ann. Phys. (N.Y.) **37**, 224 (1966).

[2] H. E. Moses, Nuovo Cimento **40A**, 1120 (1965).

[3] H. E. Moses, Ann. Phys. (N.Y.) **42**, 343 (1967).

[4] E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).

[5] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, Inc., New York, 1964).

[6] A. Weil, Actualites Sci. Ind., No. 869 (1938). [This article was also published as a book: *L'integration dans les groupes topologiques et ces applications* (Hermann & Cie., Paris, 1940).]

(3.15)]. Hence the relations are valid for any parametrization, and one does not have to prove them separately for the new variables, as is done in Ref. 3.

After mentioning some elementary facts about the rotation group in Sec. 2, we find the explicit form of the unitary matrix $u \in SU_2$ corresponding to a rotation in a specified direction in terms of ψ, θ, and ϕ. Typically, the unitary matrix is expressed in terms of Euler's angles.[4,5]

In Sec. 3 we derive the matrix elements of the irreducible representation in terms of ψ, θ, and ϕ. For completeness, we briefly review the method for deriving these matrix elements as functions of $u \in SU_2$, but after that we substitute for u its expression as a function of ψ, θ, and ϕ. The matrices obtained are then compared with those obtained in Refs. 1 and 2.

Finally, in Sec. 4 we derive the differential operators corresponding to infinitesimal rotations about the Ox_1, Ox_2, and Ox_3 in terms of the new variables. The corresponding operators when Euler's angles are used are well known in the literature. We here use a method similar to that of Gel'fand and Shapiro.[7,8]

Throughout this paper we will adopt the notation and terminology of Naimark.[5]

2. PRELIMINARIES

An orthogonal matrix describing a rotation with angle ψ about some direction $\mathbf{n}$ is given by[2,9]

$$g_{rs} = \delta_{rs}\cos\psi + n_r n_s(1 - \cos\psi) - \epsilon_{rst}n^t \sin\psi, \tag{2.1}$$

where $\mathbf{n}$ is given by (1.2) and r, s, and t run from 1 to 3. Rotations $g_1(\psi)$, $g_2(\psi)$, and $g_3(\psi)$ around Ox_1, Ox_2, and Ox_3 are obtained from (2.1) by putting the proper values of θ and ϕ.[10] The infinitesimal matrices g_r corresponding to rotations about the axis Ox_r are defined by[11]

$$g_r = \left[\frac{d}{d\psi}\, g_r(\psi)\right]_{\psi=0} \tag{2.2}$$

and satisfy the commutation relations

$$[g_r, g_s] = \epsilon_{rst}g_t. \tag{2.3}$$

We denote a representation of O_3 in an n-dimensional Euclidean space R by $g \to T_g$ and for convenience[12] we put

$$A_r(\psi) = T_{g_r(\psi)}. \tag{2.4}$$

The basic infinitesimal operators of the representation are then obtained by

$$A_r = \left[\frac{dA_r(\psi)}{d\psi}\right]_{\psi=0}. \tag{2.5}$$

A representation of O_3 is uniquely determined by its basic infinitesimal operators A_r. The determination of all the finite-dimensional representations of O_3 is based on the fact that the operators A_r satisfy the same commutation relation that exists among the infinitesimal matrices g_r:

$$[A_r, A_s] = \epsilon_{rst}A_t. \tag{2.6}$$

The A_r are skew-Hermitians, $A_r^\dagger = -A_r$, since, without loss of generality, every finite-dimensional representation of O_3 can be considered to be unitary.

Defining

$$H_\mp = iA_1 \pm A_2, \quad H_3 = iA_3, \tag{2.7}$$

one finds that

$$[H_\mp, H_3] = \pm H_\mp, \quad [H_+, H_-] = 2H_3,$$
$$H_+^\dagger = H_-, \quad H_3^\dagger = H_3. \tag{2.8}$$

The problem then reduces to the determination of $H_\mp$, H_3 satisfying conditions (2.8). This is answered by the following[5]: Every finite-dimensional representation of O_3 is uniquely determined by a nonnegative integer or half-integer j, the weight of the representation. The space of the representation corresponding to such a number j has the dimension $2j + 1$; the operators $H_\mp$, H_3 of this representation are given relative to its canonical basis $f_{-j}, f_{-j+1}, \cdots, f_j$ by

$$H_+f_n = a_{n+1}f_{n+1}, \quad H_-f_n = a_nf_{n-1}, \quad H_3f_n = nf_n,$$
$$a_n = [(j + n)(j - n + 1)]^{\frac{1}{2}}, \tag{2.9}$$

where $n = -j, -j + 1, \cdots, j$.[13]

[7] I. M. Gel'fand and Z. Ya. Shapiro, Usp. Mat. Nauk 7, 3 (1952), Am. Math. Soc. Transl. Ser. 2., 2, 207 (1956).

[8] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon Press, Inc., New York, 1963).

[9] The direction of rotation here is taken opposite to that of Refs. 1–3 and is in accordance with Ref. 5.

[10] These matrices are given by

$$g_1(\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix}, \cdots.$$

[11] The g_r are related to $g_r(\psi)$ by $g_r(\psi) = \exp(\psi g_r)$, and are given by

$$g_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \cdots.$$

[12] $A_r(\psi)$ are called the basic one-parameter groups of the given representation and define one-parameter groups of operators that satisfy $A_r(\psi_1)A_r(\psi_2) = A_r(\psi_1 + \psi_2)$; they are differentiable functions of ψ and may be expanded as $A_r(\psi) = \exp(\psi A_r)$, where A_r is defined by Eq. (2.5).

[13] It also follows that for each j there corresponds an irreducible representation of O_3. If the operators $H_\mp$ and H_3 of a representation of O_3 in a $(2j + 1)$-dimesional space are given relative to some basis $f_{-j}, f_{-j+1}, \cdots, f_j$, then by Eqs. (2.9), that representation is irreducible.

We now find the unitary matrix u corresponding to the rotation g, Eq. (2.1). They are related by[14]

$$g_{rs} = \tfrac{1}{2}\,\mathrm{Tr}\,(\sigma^r u \sigma^s u^\dagger), \tag{2.10}$$

$$u = \mp(1 + \sigma^r\sigma^s g_{rs})/2(1 + \mathrm{Tr}\,g)^{\frac{1}{2}}, \tag{2.11}$$

where σ^r are the Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.12}$$

Using Eq. (2.1), a direct calculation gives

$$u = \mp \begin{pmatrix} \cos\frac{\psi}{2} + i\sin\frac{\psi}{2}\cos\theta & i\sin\frac{\psi}{2}\sin\theta e^{i\phi} \\ i\sin\frac{\psi}{2}\sin\theta e^{-i\phi} & \cos\frac{\psi}{2} - i\sin\frac{\psi}{2}\cos\theta \end{pmatrix}. \tag{2.13}$$

This is the unitary matrix $u \in SU_2$ corresponding to a rotation with angle ψ around the direction $\mathbf{n}$ specified by θ and ϕ. The corresponding matrix when Euler's angles are employed is well known in the literature.[5] It will be noted that $u(-\psi, \theta, \phi) = u^{-1}(\psi, \theta, \phi)$.

The unitary matrices $u_1(\psi)$, $u_2(\psi)$, and $u_3(\psi)$, corresponding to the rotations $g_1(\psi)$, $g_2(\psi)$, and $g_3(\psi)$ around the axis of coordinates, are obtained from (2.13)[15]:

$$u_1(\psi) = \mp \begin{pmatrix} \cos\frac{\psi}{2} & i\sin\frac{\psi}{2} \\ i\sin\frac{\psi}{2} & \cos\frac{\psi}{2} \end{pmatrix}, \tag{2.14a}$$

$$u_2(\psi) = \mp \begin{pmatrix} \cos\frac{\psi}{2} & -\sin\frac{\psi}{2} \\ \sin\frac{\psi}{2} & \cos\frac{\psi}{2} \end{pmatrix}, \tag{2.14b}$$

$$u_3(\psi) = \mp \begin{pmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2} \end{pmatrix}. \tag{2.14c}$$

Using these matrices, the operators $A_r(\psi)$ of SU_2 will be determined in the next section.

3. MATRIX ELEMENTS OF AN IRREDUCIBLE REPRESENTATION

A matrix u of SU_2 can be considered as that of a linear transformation of the space of all pairs of complex numbers (ξ^1, ξ^2):

$$\xi'^p = \sum_{q=1}^{2} u_{pq}\xi^q \quad (p = 1, 2). \tag{3.1}$$

A representation of SU_2 can be obtained if one considers several pairs $(\xi_1^1, \xi_1^2), \cdots, (\xi_k^1, \xi_k^2)$ and forms all products $\xi_1^{p_1}\cdots\xi_k^{p_k}$, letting $p_1, \cdots, p_k$ take the values 1, 2, independently.[5] Under the transformation (3.1), this product transforms like

$$\xi_1'^{p_1}\cdots\xi_k'^{p_k} = \sum_{q_1,\cdots,q_k=1}^{2} u_{p_1q_1}\cdots u_{p_kq_k}\xi_1^{q_1}\cdots\xi_k^{q_k}. \tag{3.2}$$

The products $\xi_1^{p_1}\cdots\xi_k^{p_k}$ may be considered as a vector in the linear space R_k of all 2^k complex numbers $\xi^{p_1\cdots p_k}$. The linear transformation $T_u^{(k)}$ of the space R_k is then given by

$$\xi'^{p_1\cdots p_k} = \sum_{q_1,\cdots,q_k=1}^{2} u_{p_1q_1}\cdots u_{p_kq_k}\xi^{q_1\cdots q_k}. \tag{3.3}$$

The correspondence $u \to T_u^{(k)}$ is a representation of SU_2, not irreducible in general, since the subspace S_k of R_k of all symmetrical vectors ξ is invariant with respect to all the operators $T_u^{(k)}$. The correspondence $u \to T_u^{(k)}$ is irreducible, however, in the space S_k. We denote this representation by Z_k.[16]

An equivalent realization of the representation Z_k is obtained if one identifies the space S_k with the $(k+1)$-dimensional space of homogeneous polynomials $p(z_1, z_2)$ of degree k in the two complex variables z_1 and z_2 and sets up a one-to-one correspondence between ξ of S_k and $p(z_1, z_2)$ in the form

$$p(z_1, z_2) = \sum_{p_1,\cdots,p_k=1}^{2} \xi^{p_1\cdots p_k} z_{p_1}\cdots z_{p_k}. \tag{3.4}$$

The operator $T_u^{(k)}$ for this new realization of the space S_k is then given by

$$T_u^{(k)}p(z_1, z_2) = p(z_1', z_2'),$$

$$z_q' = \sum_{p=1}^{2} u_{pq}z_p \quad (q = 1, 2). \tag{3.5}$$

Introducing a new variable $z = z_1/z_2$, the polynomial $p(z_1, z_2)$ can be written as $z_2^k p(z)$, where $p(z)$ is a polynomial in z of degree not exceeding k. The operators $T_u^{(k)}$ of the representation Z_k are then given by

$$T_u^{(k)}p(z) = (u_{12}z + u_{22})^k p\left(\frac{u_{11}z + u_{21}}{u_{12}z + u_{22}}\right). \tag{3.6}$$

This relation gives, in particular, the operators

[14] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).

[15] The infinitesimal matrices u_r corresponding to rotations around Ox_r,

$$u_r = [du_r(\psi)/d\psi]_{\psi=0},$$

are related to the Pauli matrices by $u_r = \mp(i/2)\sigma_r$.

[16] It will be noted that Z_k is the spinor representation of weight $k/2$. For more details see Ref. 5.

$A_r(\psi) = T_{u_r(\psi)}$ when the matrices $u_r(\psi)$, Eqs. (2.14),[17] are used.[18]

It follows that every irreducible finite-dimensional representation of SU_2 is uniquely determined by some nonnegative integer or half-integer $j = k/2$, the weight of the representation.[19] The functions

$$f_n(z) = (-1)^{j-n} \frac{z^{j-n}}{[(j-n)!\,(j+n)!]^{\frac{1}{2}}}, \tag{3.7}$$

where $n = -j, -j+1, \cdots, j$, form a canonical basis for the representation Z_k in the space S_k.[20] Using Eq. (3.6), one finds

$$T_u^{(k)} f_n(z) = \sum_{m=-j}^{j} T_{mn}^j(u) f_m(z), \tag{3.8}$$

where $T_{mn}^j(u)$ are the matrix elements of the operator T_u of the irreducible representation of weight j relative to the canonical basis, which corresponds to an arbitrary rotation g. Its explicit expression is[21]

$$T_{mn}^j(u) = (-1)^{2j-m-n}\left[\frac{(j-m)!\,(j+m)!}{(j-n)!\,(j+n)!}\right]^{\frac{1}{2}} \times \sum C_a^{j-n} C_{j-m-a}^{j+n} u_{11}^a u_{12}^{j-m-a} u_{21}^{j-n-a} u_{22}^{m+n+a}, \tag{3.9}$$

where the summation runs from $a = \max(0, -m-n)$ to $\min(j-m, j-n)$, and C_m^n denotes the number of combinations of m elements from n:

$$C_m^n = n!/(n-m)!\,m!.$$

In Eq. (3.9) the indices m and n take the values $-j, -j+1, \cdots, j; j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots$.

To find the matrix elements (3.9) in terms of the variables ψ, θ, and ϕ, we simply substitute for u_{pq} their expressions as functions of these variables. These expressions were calculated in Sec. 2 and are given by Eq. (2.13). We obtain

$$T_{mn}^j(u) = (-1)^{2j-m-n}\left[\frac{(j-m)!\,(j+m)!}{(j-n)!\,(j+n)!}\right]^{\frac{1}{2}} \times \left(\cos\frac{\psi}{2} - i\sin\frac{\psi}{2}\cos\theta\right)^{m+n} e^{i\phi(n-m)} \times \sum C_a^{j-n} C_{j-m-a}^{j+n} \left(1 - \sin^2\frac{\psi}{2}\sin^2\theta\right)^a \times \left(i\sin\frac{\psi}{2}\sin\theta\right)^{2j-2a-m-n} \tag{3.10}$$

The matrix $T_{mn}^j(u)$ can also be written in the form

$$T_{mn}^j(u) = (-1)^{2j-m-n}\left[\frac{(j-m)!\,(j+m)!}{(j-n)!\,(j+n)!}\right]^{\frac{1}{2}} \times \left(i\sin\frac{\psi}{2}\sin\theta e^{-i\phi}\right)^{m-n} \times \left(\cos\frac{\psi}{2} - i\sin\frac{\psi}{2}\cos\theta\right)^{m+n} S(j,m,n,x), \tag{3.11}$$

where[22]

$$S(j,m,n,x) = 2^{m-j}(j-n)!\,(j+n)! \times \sum \frac{(x+1)^a (x-1)^{j-m-a}}{a!\,(j-n-a)!\,(j-m-a)!\,(a+m+n)!} \tag{3.12}$$

and x is defined by

$$x = 1 - 2\sin^2(\psi/2)\sin^2\theta. \tag{3.13}$$

If the direction of the rotation is reversed, ψ will then have to be replaced by $-\psi$ in the above formulas since $u^{-1}(\psi, \theta, \phi) = u(-\psi, \theta, \phi)$. The matrix elements are then given by

$$T_{mn}^j(u^{-1}) = \left[\frac{(j-m)!\,(j+m)!}{(j-n)!\,(j+n)!}\right]^{\frac{1}{2}} \times \left(i\sin\frac{\psi}{2}\sin\theta e^{-i\phi}\right)^{m-n} \times \left(\cos\frac{\psi}{2} + i\sin\frac{\psi}{2}\cos\theta\right)^{m+n} S(j,m,n,x). \tag{3.14}$$

This form of the matrix elements was first obtained by Moses through different methods.

[17] For the determination of the operators $A_r(\psi)$, one needs $u_r(\psi)$ only for small values of ψ. The signs in (2.14) are determined by the conditions $\lim u_r(\psi) = 1$ when $\psi \to 0$. Hence the + sign must be used.

[18] For example, $A_3(\psi)$ is given by

$$A_3(\psi)p(z) = e^{-ik\psi/2} p(e^{i\psi}z).$$

The basic infinitesimal operators A_r are obtained by differentiating such relations with respect to ψ and putting $\psi = 0$. For A_3, for example, one obtains

$$A_3 p(z) = i\left(z\frac{\partial}{\partial z} - \tfrac{1}{2}k\right)p(z).$$

Consequently, one obtains

$$H_+ p = -\partial p/\partial z,$$
$$H_- p = z^2(\partial p/\partial z) - kzp,$$
$$H_3 p = -z(\partial p/\partial z) + \tfrac{1}{2}kp.$$

[19] Conversely, for any nonnegative integer or half-integer j, there exists an irreducible representation of SU_2 of weight j. A representation of weight j can be realized as the representation Z_k, where $k = 2j$; and every finite-dimensional irreducible representation of SU_2 is equivalent to one of the representations Z_k.

[20] This can easily be seen by checking that $H_\pm f_n$ and $H_3 f_n$ are indeed given by Eqs. (2.9) when f_n is put for p in the corresponding equations of Footnote 18.

[21] It will be noted that $T_{mn}^j(-u) = (-1)^{2j} T_{mn}^j(u)$. Thus the representation is single-valued for integer j and double-valued for half-integer j. In the sequel the matrix u of Eq. (2.13) will be taken with the + sign.

[22] It will be noted that $S(j,m,n,x)$ is equal to the Jacobi polynomial $P_s^{\alpha\beta}(x)$ when $s = j - \frac{1}{2}(|m+n| + |m-n|)$, $\alpha = |m-n|$ and $\beta = |m+n|$.

We conclude this section by pointing out that the matrix T^j satisfies the invariant orthogonality relation[5]

$$\int T^j_{mn}(u)T^{*j_1}_{m_1n_1}(u)\,du = \frac{1}{2j+1}\,\delta_{jj_1}\delta_{mm_1}\delta_{nn_1}. \quad (3.15)$$

The integral in (3.15) is an invariant integral over SU_2, independent of the parametric representation of the rotation group. In fact, relations similar to (3.15) are valid for any compact group.[23]

4. DIFFERENTIAL OPERATORS CORRESPONDING TO INFINITESIMAL ROTATIONS

We are now in a position to find the differential operators corresponding to infinitesimal rotations about the coordinate axis, namely, the operators A_1, A_2, and A_3 (see Sec. 2) and, consequently, the operators $H_\mp$, H_3. These operators are well known in the literature when the Euler angles are employed. We here apply Gel'fand and Shapiro's method[7] to derive these operators in terms of the new variables.

Let $g \to T_g$ be an irreducible representation of weight j of the group O_3 and let $T_{mn} = T^j_{mn}$ be its matrix elements. We consider these elements as functions of the rotation g, $T_{mn} = T_{mn}(g)$. Since $g \to T_g$ is a representation, we have $T_{gg'} = T_gT_{g'}$. In terms of matrix elements, the last relation is

$$T_{mn}(gg') = \sum_{l=-j}^{j} T_{ml}(g)T_{ln}(g'), \quad (4.1)$$

where $T_{mn}(gg')$ are the matrix elements of the operator $T_{gg'}$. Define a transformation U such that

$$U_{g'}T_{mn}(g) = T_{mn}(gg'). \quad (4.2)$$

Comparing Eqs. (4.1) and (4.2), we obtain

$$U_{g'}T_{mn}(g) = \sum_{l=-j}^{j} T_{ln}(g')T_{ml}(g). \quad (4.3)$$

Furthermore, one can show that

$$U_{g'}U_{g''} = U_{g'g''}. \quad (4.4)$$

It thus follows that the transformation $U_{g'}$ realizes a representation of O_3 in the space of $2j+1$ functions of the mth row of the matrix T_g [compare Eq. (3.8)], and that the matrix elements of $U_{g'}$ are $T_{ln}(g')$.[24]

To find the operators A_r we take g' as the rotation through some angle α around the axis Ox_r and expand the relation (4.2) in powers of α.[7,12] Expansion of $T_{mn}(gg')$, which we denote by $T_{mn}(\bar\psi, \bar\theta, \bar\phi)$, gives

$$T_{mn}(\bar\psi, \bar\theta, \bar\phi) = T_{mn}(\psi, \theta, \phi) + \alpha\left[\frac{\partial T_{mn}}{\partial\psi}\frac{d\bar\psi}{d\alpha} + \frac{\partial T_{mn}}{\partial\theta}\frac{d\bar\theta}{d\alpha} + \frac{\partial T_{mn}}{\partial\phi}\frac{d\bar\phi}{d\alpha}\right]_{\alpha=0} + \cdots. \quad (4.5)$$

To obtain A_r we have to determine

$$\left.\frac{d\bar\psi}{d\alpha}\right|_{\alpha=0}, \quad \left.\frac{d\bar\theta}{d\alpha}\right|_{\alpha=0}, \quad \text{and} \quad \left.\frac{d\bar\phi}{d\alpha}\right|_{\alpha=0}$$

for each case.

Now the matrix of the rotation g is a function of the angles ψ, θ, and ϕ, which, by Eq. (2.1), has the form

$$g(\psi, \theta, \phi) = \begin{pmatrix} \cos\psi + \sin^2\theta\cos^2\phi(1-\cos\psi) & \sin^2\theta\cos\phi\sin\phi(1-\cos\psi) - \cos\theta\sin\psi & \sin\theta\cos\theta\cos\phi(1-\cos\psi) + \sin\theta\sin\phi\sin\psi \\ \sin^2\theta\sin\phi\cos\phi(1-\cos\psi) + \cos\theta\sin\psi & \cos\psi + \sin^2\theta\sin^2\phi(1-\cos\psi) & \sin\theta\cos\theta\sin\phi(1-\cos\psi) - \sin\theta\cos\phi\sin\psi \\ \sin\theta\cos\theta\cos\phi(1-\cos\psi) - \sin\theta\sin\phi\sin\psi & \sin\theta\cos\theta\sin\phi(1-\cos\psi) + \sin\theta\cos\phi\sin\psi & \cos\psi + \cos^2\theta(1-\cos\psi) \end{pmatrix}. \quad (4.6)$$

The matrix of rotation gg' is given by some angles $\bar\psi$, $\bar\theta$, and $\bar\phi$ which depend on the rotation angle α and which are equal to ψ, θ, and ϕ when $\alpha = 0$. Expansion of the matrix gg' in a power series in α gives

$$gg' = g(\psi, \theta, \phi) + \alpha\left\{\left.\frac{\partial g}{\partial\psi}\frac{d\bar\psi}{d\alpha}\right|_{\alpha=0} + \left.\frac{\partial g}{\partial\theta}\frac{d\bar\theta}{d\alpha}\right|_{\alpha=0} + \left.\frac{\partial g}{\partial\phi}\frac{d\bar\phi}{d\alpha}\right|_{\alpha=0}\right\} + \cdots. \quad (4.7)$$

To find A_1 we identify g' with the rotation with angle α around Ox_1 given by[10]

$$g_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \alpha\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} + \cdots, \quad (4.8)$$

[23] See, for example, L. S. Pontrjagin, *Topological Groups* (Princeton Univ. Press, Princeton, N.J., 1946); M. A. Naimark, *Normed Rings* (P. Noordhoff Ltd., Groningen, The Netherlands, 1959).

[24] The representation $g' \to U_{g'}$ in the space of functions $T_{ml}(g)$, $l = -j, -j+1, \cdots, j$, is irreducible, and the $T_{ml}(g)$ form a canonical basis in this space. Hence the operators $H_\pm$ and H_3 of this representation satisfy the relation (2.9). See Ref. 7.

Therefore

$$gg_1 = g(\psi, \theta, \phi) + \alpha \begin{pmatrix} 0 & g_{13} & -g_{12} \\ 0 & g_{23} & -g_{22} \\ 0 & g_{33} & -g_{32} \end{pmatrix} + \cdots . \quad (4.9)$$

On the other hand, gg_1 is given by (4.7) when $g_1 = g'$. Comparing these two expressions for gg_1, we obtain equations from which

$$\left.\frac{d\bar{\psi}}{d\alpha}\right|_{\alpha=0}, \quad \left.\frac{d\bar{\theta}}{d\alpha}\right|_{\alpha=0}, \quad \text{and} \quad \left.\frac{d\bar{\phi}}{d\alpha}\right|_{\alpha=0}$$

can be determined for the case of rotation about Ox_1. We obtain[25]

$$2 \sin\theta \cos\phi \sin\frac{\psi}{2}\left(-\sin\theta \sin\phi \left.\frac{d\bar{\phi}}{d\alpha}\right|_{\alpha=0} + \cos\theta \cos\phi \left.\frac{d\bar{\theta}}{d\alpha}\right|_{\alpha=0}\right) + \left(-\cos\frac{\psi}{2} + \sin^2\theta \cos^2\phi \cos\frac{\psi}{2}\right)\left.\frac{d\bar{\psi}}{d\alpha}\right|_{\alpha=0} = 0, \quad (4.10a)$$

$$2 \sin\theta \sin\phi \sin\frac{\psi}{2}\left(\sin\theta \cos\phi \left.\frac{d\bar{\phi}}{d\alpha}\right|_{\alpha=0} + \cos\theta \sin\phi \left.\frac{d\bar{\theta}}{d\alpha}\right|_{\alpha=0}\right) + \left(-\cos\frac{\psi}{2} + \sin^2\theta \sin^2\phi \cos\frac{\psi}{2}\right)\left.\frac{d\bar{\psi}}{d\alpha}\right|_{\alpha=0} = \sin\theta\left(\cos\theta \sin\phi \sin\frac{\psi}{2} - \cos\phi \cos\frac{\psi}{2}\right), \quad (4.10b)$$

$$2 \cos\theta \sin\frac{\psi}{2}\left.\frac{d\bar{\theta}}{d\alpha}\right|_{\alpha=0} + \cos\frac{\psi}{2} \sin\theta \left.\frac{d\bar{\psi}}{d\alpha}\right|_{\alpha=0} = \cos\theta \sin\phi \sin\frac{\psi}{2} + \cos\phi \cos\frac{\psi}{2}. \quad (4.10c)$$

[25] One obtains nine equations; only three of them are independent. Our equations (4.10) are obtained by equating the diagonal elements of the matrices (4.7) and (4.9).

The solution of Eqs. (4.10) is

$$\left.\frac{d\bar{\phi}}{d\alpha}\right|_{\alpha=0} = \tfrac{1}{2} \operatorname{cosec}\theta\left(\cos\theta \cos\phi - \cot\frac{\psi}{2} \sin\phi\right),$$
$$\left.\frac{d\bar{\theta}}{d\alpha}\right|_{\alpha=0} = \frac{1}{2}\left(\sin\phi + \cot\frac{\psi}{2} \cos\theta \cos\phi\right),$$
$$\left.\frac{d\bar{\psi}}{d\alpha}\right|_{\alpha=0} = \cos\phi \sin\theta. \quad (4.11)$$

Using Eq. (4.5), we find

$$A_1 = \cos\phi \sin\theta \frac{\partial}{\partial\psi} + \frac{1}{2}\left(\sin\phi + \cot\frac{\psi}{2} \cos\theta \cos\phi\right)\frac{\partial}{\partial\theta} + \tfrac{1}{2} \operatorname{cosec}\theta\left(\cos\theta \cos\phi - \cot\frac{\psi}{2} \sin\phi\right)\frac{\partial}{\partial\phi}. \quad (4.12a)$$

The operators A_2 and A_3 are found in a similar way:

$$A_2 = \sin\phi \sin\theta \frac{\partial}{\partial\psi} - \frac{1}{2}\left(\cos\phi - \cot\frac{\psi}{2} \cos\theta \sin\phi\right)\frac{\partial}{\partial\theta} + \tfrac{1}{2} \operatorname{cosec}\theta\left(\cos\theta \sin\phi + \cot\frac{\psi}{2} \cos\phi\right)\frac{\partial}{\partial\phi}, \quad (4.12b)$$

$$A_3 = \cos\theta \frac{\partial}{\partial\psi} - \frac{1}{2} \cot\frac{\psi}{2} \sin\theta \frac{\partial}{\partial\theta} - \frac{1}{2}\frac{\partial}{\partial\phi}. \quad (4.12c)$$

Using these results and Eq. (2.7), for H_+, H_-, and H_3 we obtain[26]

$$H_\pm = ie^{\pm i\phi}\left\{\sin\theta \frac{\partial}{\partial\psi} + \frac{1}{2}\left(\mp i + \cot\frac{\psi}{2} \cos\theta\right)\frac{\partial}{\partial\theta} + \tfrac{1}{2} \operatorname{cosec}\theta\left(\cos\theta \pm i \cot\frac{\psi}{2}\right)\frac{\partial}{\partial\phi}\right\}, \quad (4.13a)$$

$$H_3 = i\left(\cos\theta \frac{\partial}{\partial\psi} - \frac{1}{2} \cot\frac{\psi}{2} \sin\theta \frac{\partial}{\partial\theta} - \frac{1}{2}\frac{\partial}{\partial\phi}\right). \quad (4.13b)$$

[26] An operator with similar expressions was also introduced in Ref. 3, but with no explanation of the way by which they came to it. Of course, the operators $H_\pm$ and H_3, using Eqs. (2.9) and (4.13), satisfy the following relations with respect to the canonical basis $T^j_{m,-j}, T^j_{m,-j+1}, \cdots, T^j_{m,j}$:

$$H_\pm T^j_{mn}(g) = [(j \pm n + 1)(j \mp n)]^{\frac{1}{2}} T^j_{m,n\pm1}(g),$$
$$H_3 T^j_{mn}(g) = n T^j_{mn}(g).$$