JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 4, NUMBER 12 DECEMBER 1963

On the Representations of the Semisimple Lie Groups. II*

G. E. BAIRD AND L. C. BIEDENHARN
Department of Physics, Duke University, Durham, North Carolina
(Received 5 April 1963; revised manuscript received 3 August 1963)

The explicit determination of the matrices of the generators of the unitary groups, SU_n, is carried out and discussed in two alternative treatments: (a) by purely algebraic infinitesimal methods, and (b) by Young-pattern techniques employing the Schwinger–Bargmann boson operator methods. The implication of this result for a tableau calculus is discussed and a determination of the $[\lambda] \times [1]$ Wigner coefficient for all SU_n is indicated.

INTRODUCTION

IT is the purpose of the present series of papers to discuss the explicit constructive determinations of the representations of the semisimple Lie groups by an extention of the Racah–Wigner angular-momentum calculus developed for the group SU_2. The emphasis throughout is intended to be physical and the work is motivated by problems in nuclear structure.

The program to be followed has been laid out in detail by the work of Wigner[1] and Racah[2] and consists of essentially three problems: (a) the determination of invariant operators ("Casimir invariants") that uniquely specify the irreducible representations, (b) the determination of sufficient "labeling operators" to uniquely specify the states of an irreducible representation, and (c) the determination of Wigner coefficients by an explicit solution of the problem of simple reducibility.[3] A discussion of these problems was given in I,[4] and a solution for the general Casimir operator (I_n) on the unitary groups was constructed.[5-7] We shall continue in the present paper to limit our attention to the unitary groups, and for these an explicit solution to the state-labeling problem has long been available: this is the famous Weyl branching formula[8,9] for the unitary groups detailing the canonical subgroup decomposition $SU_n \supset U_1 \times SU_{n-1}$. The solution of these problems for the unitary groups contains, in principle, the solution for all semisimple Lie groups (by embedding), but more useful special results are available for the rotation and symplectic groups, as will be discussed subsequently.

The first part of the present paper applies these results for the two "labeling problems" to a purely algebraic constructive determination of the matrices of the generators of the unitary groups. This program is a direct elaboration of the classic researches

* Supported in part by the U. S. Army Research Office (Durham), and the National Science Foundation.

[1] E. P. Wigner, Am. J. Math. **63**, 57 (1941); lecture notes, Princeton University, Princeton, New Jersey, 1955 (unpublished).

[2] G. Racah, lecture notes, The Institute for Advanced Study, Princeton, New Jersey, 1951 (unpublished).

[3] W. T. Sharp, "Racah Algebra and the Contraction of Groups," CRT-935; AECL-1098, Atomic Energy of Canada Ltd., Chalk River, Ontario, 1960 (unpublished).

[4] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963), which will be referred to as I, throughout.

[5] The work of Bargmann and Moshinsky[6] also contains equivalent results for the SU_3 invariants. See also the work of Umezawa.[7] *Note added in proof*: It has recently come to our attention that the basic idea used in I was given earlier by M. Gell-Mann for SU_3 [M. Gell-Mann, California Institute of Technology Synchrotron Laboratory Rept. CTSL28-1961 (unpublished) and Phys. Rev. **125**, 1067 (1963)].

[6] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).

[7] M. Umezawa (to be published).

[8] The application of this result[9] to state labeling has been referred to in various papers by Wigner and Racah.

[9] H. Weyl, *The Theory of Groups and Quantum Mechanics* translated by H. P. Robertson (Methuen and Company Ltd., London, 1931), p. 390.

of Casimir,[10] van der Waerden[11] and Racah.[12] It was Casimir who first noted the possibility of a completely algebraic infinitesimal discussion of the rotation group, assuming nothing whatsoever beyond the generators and their commutators. The partly algebraic extension by van der Waerden and Casimir to arbitrary semisimple Lie groups was completely algebrized in principle by Racah. The actual elaboration for the unitary group has never been carried out, and the present paper explicitly considers this problem in complete detail for SU_3, obtaining the generator matrices,[13] the range of quantum numbers,[14] the dimension formula, etc. The techniques are entirely algebraic and patterned precisely according to the paradigm familiar to physicists as the quantum theory of angular momentum. SU_3 is the typical group from this point of view and the extension to all unitary groups can then be sketched.[15,16]

The algebraic approach is quite rigorous, as is clearly shown by the angular-momentum paradigm, and the method furnishes a thorough familiarity with the properties of the group in question. Yet the algebraic method—aside from its intrinsic interest as an alternative methodology—suffers from an unavoidable cumbersomeness. To comprehend the structure of the unitary groups as a whole, there is no real alternative to the classic group-theoretic (integral) methods of Weyl.[17-22] These methods have attained a striking elegance by Schwinger's[23] application of boson operator techniques in conjunction with Weyl's classic methods. Subsequent important developments of this boson operator calculus have been made by Bargmann,[24] Friedrichs,[25] Moshinsky,[26] and Helmers,[27] among others.[28]

The second part of this paper *explicitly* obtains the matrices of the generators of the general unitary group by the group-theoretical methods of the Young tableaux and the Schwinger–Bargmann boson operator techniques. This part of the paper was initiated by Moshinsky who called our attention (subsequent to the work of part I) to the paper of Gelfand and Zetlin,[29] wherein the generator matrices of the unitary groups had been explicitly determined. This paper is extremely brief (three pages) and does not appear to have been translated in either the usual Journal translations or the translations on group-theoretical subjects of the American Mathematical Society, or even referred to in the review articles on group theory by Gelfand himself.[30] Moreover, the results are presented without the slightest hint as to the methods employed and contain not a single reference or citation of other work. In an effort to understand the meaning of this very impressive work, we were led to develop the proofs presented in part II. We have been able to rederive all of the results of this reference (correcting incidentally a few errors). In doing so we have been led to a further development of the Young tableau as an *operator calculus* wherein matrix elements may be calculated directly from the Young patterns using the Nakayama concept of the 'hook length' of a Young-pattern node. The beginnings of such a tableau calculus are discussed in part II.

We should like to acknowledge our indebtedness to the prior work of Gelfand and Zetlin, and hope that our proof of their results in terms of a tableau calculus will be of general interest, particularly as

[10] H. Casimir, Proc. Koninkl. Akad. Amsterdam **34**, 844 (1931).

[11] H. Casimir and B. L. van der Waerden, Math. Ann. **111**, 1 (1935). W. Pauli, "Continuous Groups in Quantum Mechanics," CERN 56-31, Geneva, 1956; in particular, p. 10.

[12] G. Racah, Rend. Atti Accad. Naz. Lincei. **8**, 108 (1950).

[13] K. T. Hecht, Bull. Am. Phys. Soc. **8**, 57 (1963). Hecht has done explicit calculations of SU_3 matrices. *Note added in proof*: See also D. L. Pursey, Proc. Roy. Soc. (London) **A275**, 284 (1963).

[14] See also the recent work of H. Goldberg, *Trieste Conference on Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963), p. 221.

[15] Most of the results of Part I were presented earlier in *Lectures in Theoretical Physics*, edited by Wesley E. Brittin, B. W. Downs, and Joanne Downs (Interscience Publishers, Inc., New York, 1963), Vol. V, and as noted in reference 16.

[16] L. C. Biedenharn, Phys. Letters **3**, 69 (1962).

[17] Application of these methods[18] to nuclear physics has been given by Yamanouchi,[19] Jahn,[20] Elliott,[21] Flowers,[22] and of course Racah[2] and Wigner.[1]

[18] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, New Jersey, 1939); lecture notes, The Institute for Advanced Study, Princeton, New Jersey, 1934 (unpublished).

[19] T. Yamanouchi, Proc. Phys. Math. Soc. Japan **18**, 623 (1936); **19**, 436 (1937); **20**, 547 (1938).

[20] H. A. Jahn, Proc. Roy. Soc. (London) **A201**, 516 (1950); **A205**, 192 (1951).

[21] J. P. Elliott, Proc. Roy Soc. (London) **A218**, 345 (1953).

[22] B. H. Flowers, Proc. Roy. Soc. (London) **A210**, 497 (1952); **A212**, 248 (1952).

[23] J. Schwinger, *On Angular Momentum* [Technical Information Services, Oak Ridge, Tennessee (Rept. W-23091, No. NYO-3071), 1952]. This classic paper is shortly to be reprinted in a volume devoted to angular momentum theory (to be published by Academic Press Inc., New York).

[24] V. Bargmann, Commun. Pure Appl. Math. **14**, 198 (1961).

[25] K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1953).

[26] M. Moshinsky, Nucl. Phys. **31**, 384 (1962); Rev. Mod. Phys. **34**, 813 (1962).

[27] K. Helmers, Nucl. Phys. **12**, 647 (1959); **23**, 594 (1961).

[28] Recently Moshinsky has applied such methods to determining unitary bases for the irreducible representations of the general unitary group (preprint, May, 1963).

[29] I. M. Gelfand and M. L. Zetlin, Doklady Akad. Nauk SSSR **71**, 825 (1950).

[30] We should like to thank Louis Wright for translating this paper.

the Gelfand and Zetlin paper is relatively inaccessible. We should like to express our appreciation to Professor Moshinsky for his calling the Gelfand–Zetlin paper to our attention.

The real motivation underlying the present series of papers is the extension of the Racah–Wigner calculus to include a definition of "Wigner and Racah coefficients" on the unitary groups. From this point of view, the generator matrices are but special cases of the Wigner coefficients.[31] A complete solution to the Wigner coefficient problem was sketched previously[32] and involves as the essential additional concept the symmetric "vector coupling" coefficient defined in I. From this point of view, the present paper is ancillary to the main results to be presented systematically in a subsequent paper devoted to the Wigner coefficients. The results of part II already include, however, a complete determination of the $[\lambda] \times [1]$ Wigner coefficient for all SU_n.

I. CONSTRUCTION OF THE GENERATOR MATRICES BY PURELY ALGEBRAIC INFINITESIMAL TECHNIQUES

A. The State-Labeling Problem and its Solution for SU_n

The irreducible representations of a semisimple Lie group are uniquely labeled by the highest weights that occur in the representation; this is the content of Cartan's second main theorem.[18] The rank of the group (l) is equal to the number of diagonal group generators H_i—and hence the number of components of the weight $\mathbf{\Lambda}$. We have shown by construction in I that there exist for the SU_n group $l = n - 1$ independent invariants $I_2, \cdots I_n$; we conclude that the inequivalent irreducible representations may be uniquely labeled by the eigenvalues of the I_n. Moreover (from the second main theorem), the state of highest weight is also unique, and hence labeled uniquely by the eigenvalues of the I_n also.

The remaining states belonging to a given inequivalent, irreducible representation (labeled by the I_n eigenvalues) may or may not be uniquely labeled by the weights. In order to give explicitly the matrices of the representation—or equivalently the matrices of the generators—it is necessary that *each state* belonging to a given representation must be labeled in one-to-one fashion. This is the "state-labeling problem" whose importance has been emphasized by Wigner[1] and by Racah,[2] as discussed in I.

It is the purpose of the present section to detail an explicit solution to the state-labeling problem applicable to all SU_n groups. While the intrinsic content of the present section cannot be claimed as completely new, such explicit results as we shall give do not appear to be in the literature. Such results would require repetition in any event in order to carry out our program of explicit construction of the SU_n representations. Let us note, however, that the solution of the state-labeling problem for SU_n has been given earlier by Wigner in an unpublished manuscript. [The Racah lecture notes (reference 2) refer indirectly to this work.] In addition, the branching law of Weyl[9] implicitly contains the solution here proposed.

Let us examine first what the state-labeling problem entails. (This argument is that of reference 2.) The problem is to label uniquely the elements of a matrix of $n^2 - 1$ parameters. We have already $l = n - 1$ invariant operators I_n which lead to $n - 1$ labels. Of the remaining $(n^2 - 1) - (n - 1) = (n - 1)(n)$ labels, we know that the operators H_i furnish $2 \times (n - 1)$ labels; the factor 2 enters since the ijth matrix element has H_i labels for both the ith state and the jth state. Hence we still need $\frac{1}{2}[n(n - 1) - 2(n - 1)] = \frac{1}{2}[(n - 1)(n - 2)]$ *independent operators, commuting with each other and with all the* H_i, in order to label the states uniquely.

For example, in SU_2 there are three generators, one invariant operator (J^2), and one $H_i(J_z)$. No further labeling operators are required, since $\frac{1}{2}[(n - 1)(n - 2)]$ vanishes for $n = 2$. This is the familiar angular-momentum example which is uniquely labeled by J^2 and J_z, as is well-known.

Consider next the group SU_3. Here there are eight generators, two invariants (I_2 = Casimir's invariant and I_3), and two additive ("magnetic") operators H_1 and H_2. We need $\frac{1}{2}[(n - 1)(n - 2)] = 1$ further operator to complete the labeling. If one examines the vector diagram of SU_3 it is clear at once that E_α and $E_{-\alpha}$ both commute with H_2 (since $\boldsymbol{\alpha}$ is perpendicular to the H_2 axis). Moreover, E_α and $E_{-\alpha}$ by themselves *constitute the vector diagram of* SU_2. Thus, directly from the vector diagram, we may assert that the Casimir invariant for SU_2 (given here by $\frac{1}{3}\Lambda^2 \equiv H_1^2 + E_\alpha E_{-\alpha} + E_{-\alpha}E_\alpha$) is a suitable operator to complete the designation of the states in SU_3. (By construction Λ^2 commutes with H_1; by choice of E_α and $E_{-\alpha}$, Λ^2 commutes with H_2.)

[31] It should be noted that the Wigner coefficients determined by Moshinsky,[26] for SU_3 are not general enough to include the generator matrices as special cases.

[32] L. C. Biedenharn, Phys. Letters **3**, 254 (1963).

This important result was found by Elliott. [Actually an explicit construction for just this case also appears in Weyl, (reference 18).] The two important papers in which Elliott discusses the SU_3 group constituted something of a "breakthrough" for nuclear physics.

The general result for the SU_n labeling problem is quite immediate, and more or less obvious now from what has been said. *For the general SU_n case, the labeling problem is solved by the canonical factorization*: $SU_n \supset U_1 \times SU_{n-1}$. Here U_1 is the one-parameter Abelian subgroup generated by *a linear combination of the $n-1$ H_i operators*; the SU_{n-1} group is a subgroup of SU_n. Each of the generators of the particular SU_{n-1} subgroup must commute with the U_1 generator in order to define the direct product. (Note that it is this requirement that distinguishes among the many isomorphic SU_{n-1} subgroups of SU_n.)

Before proving that this decomposition is always possible, let us see first if it is satisfactory. In effect, this labeling scheme assigns to every state vector of a representation of SU_n the labeling $|I_2^{(1)} I_3^{(1)} \cdots I_n^{(1)}; I_2^{(2)} \cdots I_{n-1}^{(2)}; I_2^{(3)}, \cdots I_{n-2}^{(3)}; \cdots ; I_2^{(n-1)}; H_i'\rangle$, that is: $I_n^{(1)}$ are the eigenvalues of the I_n for SU_n; $I_n^{(2)}$ are the eigenvalues of SU_{n-1}, $\cdots$, $I_2^{(n-1)}$ is the eigenvalue of SU_2; while the H_i' are the eigenvalues of the H_i of SU_n.

How many labels are involved? Aside from the $3(n-1)$ labels furnished by the H_i and I_n of SU_n, there are $\sum_{i=1}^{n-2} (i) = \frac{1}{2}[(n-2)(n-1)]$ labels furnished by the chain of subgroups SU_i. This is precisely the required number of additional labels, and demonstrates (by induction) that if the chain is possible then it possesses at least one feature of the complete solution.

To prove that it is always possible in SU_n to find subgroups U_1 and SU_{n-1} which commute, let us consider the defining $n \times n$ representation of SU_n. The canonical generators $e_{(ij)}$ and h_i must be replaced by a slightly different set. First divide the e's into two sets: $\{e_{*(1n)}\}$ and $\{e_{(ij)}'\}$, where

$$e'_{(ij)} \equiv \begin{bmatrix} 0 & 0 \\ 0 & e_{(ij)}{}^{(n-1)} \end{bmatrix},$$

with $e_{(ij)}{}^{(n-1)}$ being the generators e_{ij} for SU_{n-1}. Secondly, let the diagonal generators be: $h_n \equiv \text{diag}\,(n-1, -1, -1, \cdots, -1)$ and

$$h'_i \equiv \begin{bmatrix} 0 & 0 \\ 0 & h_i{}^{(n-1)} \end{bmatrix},$$

where $h_i{}^{(n-1)}$ are the canonical h_i for SU_{n-1}. These diagonal generators, $\{h_n, h'_i\}$ are $n-1$, independent, traceless, commuting matrices.

By construction, h_n commutes with the set $\{h'_i\}$ and the set $\{e'_{(ij)}\}$. Since h_n is the generator of U_1, and since the $\{h'_i, e'_{(ij)}\}$ are isomorphic to the generators of SU_{n-1}, the possibility of obtaining a subgroup $U_1 \times SU_{n-1}$ has been demonstrated, for the general representation of SU_n, $\{e, h\} \to \{E, H\}$.

It would be of some interest to exhibit this subgroup in terms of the canonical generators of both SU_n and SU_{n-1}; this can be carried out by means of the Racah–Wigner algebra (for SU_2), but will not be given here. While the SU_3 case is straightforward (the SU_2 group was already explicit in the canonical generators of SU_3 as could be seen from the vector diagram), the SU_4 case shows that *this decomposition expresses a symmetry property of the simplex that is the fundamental region of the group SU_n*. In SU_4, one sees that there exists a plane passing through the 3-dimensional simplex (of cubo-octohedral symmetry) which contains six vectors of the SU_4 group in precisely the form of the vector diagram of SU_3. A diagram will convince one of the truth of this statement, although it is a bit complicated to reproduce here.

Although we have demonstrated that the desired (canonical) decomposition $SU_n \supset U_1 \times SU_{n-1}$ is always possible and has the proper number of labels, it is yet to be shown that this subgroup decomposition furnishes a unique and distinct set of labels for each and every state of a given representation of SU_n. Before completing the proof, let us note that there are two ways in which one may proceed, depending on the purpose behind the demonstration. If it is desired simply to prove the uniqueness of the labeling furnished by the canonical decomposition, one may presuppose the full machinery of group theory, including the many results of Weyl and Cartan. The demonstration is quite easy in this case, for we know that a finite basis exists for every irreducible representation (corollary to Peter–Weyl theorem), and the invariant operators for the SU_{n-1} subgroup, as well as the operator H_i for the U_1 subgroup, then possess eigenstates that are orthogonal for distinct sets of eigenvalues. By simply counting the number of such orthogonal states, one may then show that the number of such states agrees with the Weyl dimension formula, and hence the theorem is demonstrated.

Such a proof, however, is not really consonant with our underlying purpose, for we wish to obtain *all* results as a direct consequence of purely infinitesimal considerations—including even the fact that the irreducible representations are all finite-dimensional. The invariance of the operators I_n has been demonstrated in I by infinitesimal means, and the subgroup decomposition $SU_n \supset U_1 \times SU_{n-1}$, similarly used such concepts only. To complete our

demonstration we propose to show, *by construction*, that these two results suffice to determine the matrices of the generators of all irreducible representations. In so doing, the fact that the canonical decomposition $SU_n \supset U_1 \times SU_{n-1}$ provides a unique labeling will be obtained as a corollary. It is to this task that we next turn.

B. The Unimodular Unitary Group SU_3 as a Model

In the previous sections, we have detailed a solution to the two ancillary steps required for a constructive treatment of the inequivalent irreducible representations of the unimodular unitary groups patterned precisely after Casimir's treatment of SU_2.

The group SU_3 is the logical next step to discuss by the Racah–Wigner techniques of SU_2, and we shall treat SU_3 as the prototype for the general case of SU_n. This is very reasonable since SU_3 does, in fact, exhibit properties typical of the general case, unlike SU_2 which is highly special.

The SU_3 group is of considerable interest for physics. The three-dimensional isotropic harmonic oscillator has SU_3 as its symmetry group, for example.[33] Since harmonic-oscillator wavefunctions are a customary set of basis functions for perturbation treatments of nuclear structure, the explicit connection (in three dimensions) with SU_3 is a valuable tool—this was the starting point of Elliott's work.[34] More recently, the elementary-particle theorists have speculated on the possibility that SU_3 is of fundamental significance in classifying the elementary-particle plethora. Hopefully, therefore, the present discussion of SU_3 may be valuable for more than the calculational aspects which motivated the work originally.

C. The Structure Constants and Symmetric Coupling Coefficients

The defining representation of SU_3 is the 3×3 representation whose generators are

$$h_1 = (1/2\sqrt{3}) \begin{bmatrix} 1 & & \\ & 0 & \\ & & -1 \end{bmatrix}, \tag{1a}$$

$$h_2 = (1/6) \begin{bmatrix} 1 & & \\ & -2 & \\ & & 1 \end{bmatrix}, \tag{1b}$$

and

$$e_\alpha = e_{(13)} \cdot 6^{-\frac{1}{2}}, \tag{1c}$$

$$e_\beta = e_{(12)} \cdot 6^{-\frac{1}{2}}, \tag{1d}$$

$$e_{\bar\beta} = e_{(23)} \cdot 6^{-\frac{1}{2}}, \tag{1e}$$

together with their "negatives",

$$e_{-\alpha} = e_{(31)} \cdot 6^{-\frac{1}{2}}, \cdots .$$

The commutators of these x_A lead to the root vectors

$(i\gamma^\gamma)$

i \ γ	α	$-\alpha$	β	$-\beta$	$\bar\beta$	$-\bar\beta$
1	1	−1	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
2	0	0	$\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{3}$

$\times 3^{-\frac{1}{2}}$, (2)

and to the familiar vector diagram of SU_3. The only remaining structure constant to determine is: $(\beta\bar\beta\alpha) = 6^{-\frac{1}{2}}$; all others may be obtained from the ones give here and above, using $g_{AB} = \delta_A{}^B$ and the fact that (ABC) is totally antisymmetric.

The symmetric coupling coefficients are the anticommutators, and these are determined to be

$[\gamma i^\gamma]$

i \ γ	$\pm\alpha$	$\pm\beta$	$\pm\bar\beta$
1	0	$\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{3}$
2	1	$-\frac{1}{2}$	$-\frac{1}{2}$

$\times \frac{1}{3}$ (3)

(This might be used to define vectors also; note that then these new vectors would be perpendicular to the old ones.)

$[ijk]$:

$$[112] = \tfrac{1}{3}; \qquad [222] = -\tfrac{1}{3}; \tag{4}$$

all other $[ijk]$ not equivalent to these are zero. (Note that $[ABC]$ is completely symmetric.)

The remaining independent $[ABC]$ to be defined is $[\beta\bar\beta\alpha] = 6^{-\frac{1}{2}}$.

D. The Invariant Operators for SU_3 and the Labeling Operator

It follows from the general discussion of the invariants, that the two invariants for SU_3 are ex-

[33] J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940); G. A. Baker, *ibid.* **30**, 1119 (1956); H. V. McIntosh, Am. J. Phys. **27**, 620 (1959); H. A. Buchdahl, *ibid.* **30**, 829 (1962); P. H. E. Meijer and T. Tanaka, *ibid.* **31**, 161 (1963); E. Merzbacher, *ibid.* **31**, 549 (1963).

[34] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128, 562 (1958).

plicitly:

$$I_2 = \sum_A X_A X_{-A} = H_1^2 + H_2^2 + E_\alpha E_{-\alpha} + E_{-\alpha}E_\alpha + E_\beta E_{-\beta} + E_{-\beta}E_\beta + E_{\bar\beta}E_{-\bar\beta} + E_{-\bar\beta}E_{\bar\beta}, \quad (5)$$

and

$$I_3 \equiv \sum_{ABC} [ABC]X_A X_B X_C = \sum_{ijk}[ijk]H_iH_jH_k + 3\sum_{i\alpha}[i\alpha - \alpha]H_iE_\alpha E_{-\alpha} + \sum_{\alpha\beta\gamma\neq 0}[\alpha\beta\gamma]E_\alpha E_\beta E_\gamma. \quad (6)$$

(Note that this latter form uses the fact that $\sum_i (i\alpha^\alpha)[i\alpha^\alpha] = 0$, i.e., the "perpendicularity" mentioned above.)

It is useful to write I_3 in terms of subsidiary operators (to be defined below), and the complete form for I_3 which results upon substituting for $[ABC]$ is not readily interpretable otherwise. The specific form for the H's alone, however, is of considerable interest. This is

$$\sum_{ijk}[ijk]H_iH_jH_k = \tfrac{1}{3}H_2(3H_1^2 - H_2^2). \quad (7)$$

Let us next introduce the labeling operator defined by the Casimir invariant of the SU_2 subgroup generated by E_α, $E_{-\alpha}$, and H_1. This is the operator: $\mathbf{\Lambda}^2 = 3(E_\alpha E_{-\alpha} + E_{-\alpha}E_\alpha + H_1^2)$, where the normalization is chosen to make $\mathbf{\Lambda}^2 \to \lambda(\lambda + 1)$ with λ integral or half-integral.

The remaining operators H_2, $E_{\pm\beta}$, $E_{\pm\bar\beta}$ may be classified by their transformation properties under this SU_2 subgroup. As indicated previously—in the discussion of the symmetry groups of the SU_3 vector diagram—the operator H_2 is a scalar under this SU_2, while $E_{\pm\beta}$ and $E_{\pm\bar\beta}$ constitute a *bispinor*. Explicitly, one finds that:

(a) E_β behaves as the $\tau = \frac{1}{2}$ component of the tensor operator $T_{\frac{1}{2}}^{\ \tau}$,

(b) $E_{-\bar\beta}$ behaves as the $\tau = -\frac{1}{2}$ component of $T_{\frac{1}{2}}^{\ \tau}$,

(c) $E_{\bar\beta}$ behaves as the $\tau = \frac{1}{2}$ component of the tensor operator $\bar T_{\frac{1}{2}}^{\ \tau}$,

(d) $E_{-\beta}$ behaves as the $\tau = -\frac{1}{2}$ component of $\bar T_{\frac{1}{2}}^{\ \tau}$.

Since, however, E_β and $E_{-\beta}$ are Hermitian conjugates (and similarly for $E_{\bar\beta}$, $E_{-\bar\beta}$), we see that $T_{\frac{1}{2}}^{\ \tau}$ and $\bar T_{\frac{1}{2}}^{\ \tau}$ are Hermitian adjoint operators. We choose the appropriate phase convention for time-reversal, and define

$$(T_{\frac{1}{2}}^{\ \tau})^\dagger \equiv (-)^{\frac{1}{2}-\tau}\bar T_{\frac{1}{2}}^{\ -\tau}, \quad (8)$$

which agrees with the E operators, up to a phase.

Using the bispinor operator $T_{\frac{1}{2}}$, one can now construct two bilinear Hermitian operators, S (scalar) and $\mathbf{V}$ (vector), employing the Wigner coefficients. Thus for the scalar operator one finds

$$S \equiv \sum_\tau C^{\frac{1}{2}\ \frac{1}{2}\ 0}_{\tau\ -\tau}\, T_{\frac{1}{2}}^{\ \tau}\bar T_{\frac{1}{2}}^{\ -\tau} = 2^{-\frac{1}{2}}\sum_\tau (-)^{\frac{1}{2}-\tau}T_{\frac{1}{2}}^{\ \tau}\bar T_{\frac{1}{2}}^{\ -\tau} = 2^{-\frac{1}{2}}\sum_\tau T_{\frac{1}{2}}^{\ \tau}(T_{\frac{1}{2}}^{\ \tau})^\dagger = 2^{-\frac{1}{2}}(E_\beta E_{-\beta} + E_{-\bar\beta}E_{\bar\beta}). \quad (9)$$

(Note the convenience of the phase choice, which removes the usual minus sign of the antisymmetric spinor combination. Note also that an *order* has been specified in defining these operators.)

For the vector operator $\mathbf{V}$, one has the result

$$V^q \equiv \sum_{\tau,\bar\tau} C^{\frac{1}{2}\ \frac{1}{2}\ 1}_{\tau\ \bar\tau\ q}\, T_{\frac{1}{2}}^{\ \tau}\,\bar T_{\frac{1}{2}}^{\ \bar\tau} = \sum_{\tau,\bar\tau} C^{\frac{1}{2}\ \frac{1}{2}\ 1}_{\tau\ \bar\tau\ q}\cdot(-)^{\frac{1}{2}+\bar\tau}\cdot T_{\frac{1}{2}}^{\ \tau}(T_{\frac{1}{2}}^{\ -\bar\tau})^\dagger = \begin{cases} -E_\beta E_{\bar\beta}, & q = 1, \\ 2^{-\frac{1}{2}}(E_\beta E_{-\beta} - E_{-\bar\beta}E_{\bar\beta}), & q = 0, \\ E_{-\bar\beta}E_{-\beta}, & q = -1. \end{cases} \quad (10)$$

From this definition one sees that the vector operator $\mathbf{V}$ obeys the rule $(V^q)^\dagger = (-)^q V^{-q}$, similar to the spherical harmonics. S, of course, is a Hermitian scalar.

The invariant operator I_3 can now be given in a more useful form:

$$I_3 = \tfrac{1}{6}H_2(H_2^2 - 1) - \tfrac{1}{2}H_2I_2 + \tfrac{1}{6}\mathbf{\Lambda}^2(3H_2 - 1) + \sqrt{2}\mathbf{\Lambda}\cdot\mathbf{V}. \quad (11)$$

Similarly, the Casimir invariant for SU_3 takes the form

$$I_2 = \tfrac{1}{3}\mathbf{\Lambda}^2 + 2\sqrt{2}S + H_2(H_2 - 1). \quad (12)$$

[It is worth remarking that these results for I_2 and I_3 are not explicitly dependent on H_1 (all H_1 dependence is absorbed in the H_1 invariant operators $\mathbf{\Lambda}^2$ and $\mathbf{\Lambda}\cdot\mathbf{V}$.) This will be seen to be an expression of the fact that the SU_3 Wigner coefficients *factor* into an SU_2 Wigner coefficient multiplied by an M_1 independent part[15] (Wigner–Eckart theorem). *This factorization, however, will prove to be a general decomposition property of the complete chain of SU_n Wigner coefficients.*]

E. The Matrices of the Generators

The matrices of the generators E_α, $E_{-\alpha}$, H_1, and H_2 can be accomplished readily. H_1 and H_2 are diagonal so that one has [the (diagonal) I_2, I_3 designation is omitted for brevity]

$$\langle\lambda'\mu'm'\,|\sqrt{6}E_\alpha|\,\lambda\mu m\rangle = \delta_\lambda^{\lambda'}\delta_m^{m'}\cdot\delta_{\mu'}^{\mu+1}\cdot\sqrt{(\lambda-\mu)(\lambda+\mu+1)}, \quad (13a)$$

$$\langle \lambda'\mu'm' \,|\sqrt{6}E_{-\alpha}|\, \lambda\mu m\rangle = \delta_\lambda^{\lambda'}\delta_m^{m'}\delta_{\mu'}^{\mu-1}\sqrt{(\lambda+\mu)(\lambda-\mu+1)}. \quad (13b)$$

The evaluation of the $E_{\dot\alpha\beta}$, $E_{\dot\alpha\dot\beta}$ is more difficult, and for this purpose consider the bispinor operator $T_{\frac12}^{\tau}$. The matrix elements of this operator are

$$\langle \lambda'\mu'm' \,|T_{\frac12}^{\tau}|\, \lambda\mu m\rangle = \langle \lambda'm' \,||T||\, \lambda m\rangle C_{\mu\ \tau\ \mu'}^{\lambda\ \frac12\ \lambda'}, \quad (14)$$

where the reduced matrix element $\langle \lambda'm' \,||T||\, \lambda m\rangle$ vanishes unless $m' = m + \frac12$, as required by the commutation rules.

In order to determine the required two reduced matrix elements, one uses the two invariants I_2 and I_3, expressed in terms of diagonal operators and the bilinear operators S and $\mathbf{V}$. That is:

$$2\sqrt{2}S = I_1 - \tfrac13\Lambda^2 - H_2(H_2-1) = 2\sum_\tau T_{\frac12}^{\tau}(T_{\frac12}^{\tau})^\dagger, \quad (15)$$

and

$$\sqrt{2}\Lambda\cdot\mathbf{V} = I_3 - \tfrac16 H_2(H_2^2-1) - \tfrac16\Lambda^2(3H_2-1) - \tfrac12 H_2 I_2 = \sqrt{2}\sum_{\tau,q} C_{\tau\ q-\tau\ q}^{\frac12\ \frac12\ 1}(-1)^{\frac12-\tau}\cdot\Lambda_{-q}T_{\frac12}^{\tau}(T_{\frac12}^{\tau-q})^\dagger. \quad (16)$$

Taking matrix elements of these two equations, one finds that only the diagonal matrix element is nonzero. For the scalar operator one finds

$$2\sum_{\tau,\lambda',\mu',m'} |\langle \lambda\mu m \,|T_{\frac12}^{\tau}|\, \lambda'\mu'm'\rangle|^2 = 2\sum_{\lambda'=\lambda\pm\frac12} |\langle \lambda m \,||T||\, \lambda' m-\tfrac12\rangle|^2 = I_2 - \tfrac13\lambda(\lambda+1) - m(m-1). \quad (15')$$

for the vector operator, the result is

$$\sqrt{2}\lambda(\lambda+1)\sum_{\tau,q,\lambda',\mu'} (-)^{\frac12-\tau} C_{\tau\ \frac12\ q-\tau}^{\frac12\ \frac12\ 1} C_{\mu+q-q\ q}^{\lambda\ 1\ \lambda} C_{\mu'\ \tau\ \mu+q}^{\lambda'\ \frac12\ \lambda} C_{\mu'\ \tau-q\ \mu}^{\lambda'\ \frac12\ \lambda} |\langle \lambda m \,||T||\, \lambda' m-\tfrac12\rangle|^2 = I_3 - \tfrac16 m(m^2-1) - \tfrac16\lambda(\lambda+1)(3m-1) - \tfrac12 m I_2. \quad (16')$$

The sum in Eq. (16′) is readily evaluated by the Racah–Wigner techniques and leads to the result

$$[2\lambda(\lambda+1)]^{\frac12}\sum(\cdots) = \sum_{\lambda'}[6\lambda(\lambda+1)(2\lambda+1)]^{\frac12} \times W(\lambda 1\lambda'\tfrac12;\lambda\tfrac12)\cdot|\langle \lambda\ m \,||T||\, \lambda'\ m-\tfrac12\rangle|^2 = \{(\lambda+1)\,|\langle \lambda\ m \,||T||\, \lambda-\tfrac12\ m-\tfrac12\rangle|^2 - \lambda\,|\langle \lambda\ m \,||T||\, \lambda+\tfrac12\ m-\tfrac12\rangle|^2\}. \quad (16'')$$

The two reduced matrix elements are now easily evaluated from Eqs. (15′) and (16′):

$$\langle \lambda\ m \,||T||\, \lambda-\tfrac12\ m-\tfrac12\rangle = [6(2\lambda+1)]^{-\frac12}\cdot[6I_3 + 3I_2(\lambda+m) - (\lambda+1+m)(\lambda+m-1)(\lambda+m)]^{\frac12}, \quad (17a)$$

$$\langle \lambda\ m \,||T||\, \lambda+\tfrac12\ m-\tfrac12\rangle = [6(2\lambda+1)]^{-\frac12}\cdot[-6I_3 + 3I_2(\lambda+1-m) - (\lambda+2-m)(\lambda+1-m)(\lambda-m)]^{\frac12}. \quad (17b)$$

In our above expressions for the reduced matrix elements, let us make the substitutions for I_2 and I_3:

$$I_2 = \tfrac13(p^2+q^2-pq+3p), \quad (18)$$

$$I_3 = 2^{-1}\cdot 3^{-4}(p-2q)(2p+3-q)(p+q+3). \quad (19)$$

Thus, with no loss of generality, we have changed our two unknown quantities to the two variables p and q. The advantage of our new form for the unknown quantities I_2 and I_3 is that one may now completely factor the reduced matrix elements into the forms

$$\langle \lambda+\tfrac12\ m+\tfrac12 \,||T||\, \lambda\ m\rangle = \tfrac{1}{18}\{[p-2q+3\lambda+3m+3] \times [p+q+3\lambda+3m+6] \cdot[2p-q-3\lambda-3m]\}^{\frac12}\cdot[\lambda+1]^{-\frac12}, \quad (20a)$$

$$\langle \lambda-\tfrac12 m+\tfrac12 \,||T||\, \lambda m\rangle = \tfrac{1}{18}\{[-p+2q+3\lambda-3m] \times [p+q-3\lambda+3m+3] \cdot[2p-q+3\lambda-3m+3]\}^{\frac12}\cdot[\lambda]^{-\frac12}. \quad (20b)$$

For the Hermitian adjoint operators one has

$$\langle \lambda-\tfrac12 m-\tfrac12 \,||\bar T||\, \lambda m\rangle = \tfrac{1}{18}\{[p-2q+3\lambda+3m] \times [p+q+3\lambda+3m+3] \cdot[2p-q-3\lambda-3m+3]\}^{\frac12}\cdot[\lambda]^{-\frac12}, \quad (21a)$$

$$\langle \lambda+\tfrac12 m-\tfrac12 \,||\bar T||\, \lambda m\rangle = -\tfrac{1}{18}\{[-p+2q+3\lambda-3m+3] \times [p+q-3\lambda+3m] \cdot[2p-q+3\lambda-3m+6]\}^{\frac12}\cdot[\lambda+1]^{-\frac12}. \quad (21b)$$

F. The Range of the Labeling Quantum Numbers

Let us recall now the familiar example of angular momentum, i.e., the group SU_2. At a similar stage in the analysis we would have found that

$$I_2 \equiv j(j+1), \tag{22a}$$

$$\langle j'm' \,|H_1|\, jm\rangle = \delta_i{}^{i'}\,\delta_m{}^{m'}\,m, \tag{22b}$$

and

$$\langle j'm' \,|J_\pm|\, jm\rangle = \delta_i{}^{i'}\,\delta_m{}^{m\pm1}[(j\mp m)(j\pm m+1)]^{\frac{1}{2}}. \tag{22c}$$

At this stage, j and m are simply variables whose properties are yet to be determined. From the commutation relations and the fact that J_-J_+ and J_+J_- are each positive definite, one next proved that $(j-m)(j+m+1) \geq 0$, and $(j+m)(j-m+1) \geq 0$. This in turn implied *four* boundary points $[m = \pm j_1 \pm (j+1)]$ whose interior region then implied that $2j$ = positive integer and $j > m > -j$.

An analogous argument carries over to all SU_n groups. In particular, for SU_3, we may now determine the range at the labeling quantum numbers λ, m, μ from the vanishing of the matrix elements for fixed $[p, q]$. The labeling quantum-number space for SU_3 is, however, now *three-dimensional*, and bounded by planes, in contrast to the SU_2 case which was one-dimensional and bounded by the two points $J_z = \pm J$. This rapid increase in dimensionality is a general property; for SU_n, the space of the labeling quantum numbers (i.e., all but $[p, q, \cdots]$) is $\frac{1}{2}[n(n-1)]$-dimensional.

The quantum number μ is limited (by the SU_2 group) to the range $\lambda \leq \mu \leq \lambda$.

Consider now the quantum numbers λ and m. The operators $E_{\pm\beta}$ and $E_{\pm\bar{\beta}}$ shift the values of λ by $\pm\frac{1}{2}$, and the values of m by $\pm\frac{1}{2}$. (The operators $E_{\pm\beta}$, $E_{\pm\bar{\beta}}$ also shift the values of μ, but this need not be considered explicitly, since the region of μ is bounded by $\pm\lambda$.) The reduced matrix elements for T allow one to introduce four "shift operators": $S(\pm, \pm) \Rightarrow \lambda \to \lambda \pm \frac{1}{2}$; $m \to m \pm \frac{1}{2}$.

Now let us consider operating on an arbitrary state with E_β. If we consider the case where E_β raises both λ and m, then its matrix element will be the first of the above four matrix elements, that is, the matrix element belonging to $S(+\ +)$. Since this matrix element must be positive definite, our "raising" operation must terminate in order that we shall not get an imaginary result for our matrix element. Thus by setting the three factors in the matrix element each separately equal to zero, we find three constraints giving three boundary lines. From each of the three remaining matrix elements, i.e., the three shift operators $S(+\ -)$, $S(-\ +)$, and $S(-\ -)$, we arrive at three further constraints for each operator. The intersections of these twelve boundary lines yields an interior region bounded by the parallelogram whose vertices are the four points whose (λ, m) values are

(a) $(\frac{1}{2}p, \frac{1}{6}(p-2q))$,

(b) $(\frac{1}{2}q, -\frac{1}{6}(2p-q))$,

(c) $(\frac{1}{2}(p-q), \frac{1}{6}(p+q))$,

(d) $(0, \frac{1}{3}(2q-p))$.

In determining the interior region, we necessarily had to take p and q to be positive integers with $p \geq q$.

The point (a) is the only point from which we cannot move except by simultaneously lowering *both* indices (λ and m). We shall call the state represented by this point the maximum state (with respect to λ and m). We may now determine the p and q values from this maximum state (exactly as one determines j from the maximum m value for SU_2). That is,

$$p = 2\lambda_{\max}, \tag{23a}$$

$$q = \lambda_{\max} - 3m_a. \tag{23b}$$

(Note that m_a does not denote the maximum m value, but rather the m value of the maximum state, hence our notation; however, $\lambda_a = \lambda_{\max}$.)

To complete the designation of the maximum state, one reintroduces the quantum number μ and sets $\mu_a = \lambda_{\max}$. Finally—just as in the angular-momentum paradigm—we designate the representation by the p, q values of the maximum state, and henceforth refer to the representation by $[p, q]$ in place of I_2 and I_3. The integers p and q are precisely the Young symmetry pattern designation.

There is a simple construction to determine the allowed (λ, m) values for a representation labeled by $[p, q]$. The idea is to exploit the fact that $\lambda = 0$ is always a boundary point; this suggests that the parallelogram be plotted as a rectangle with $\lambda = 0$ being its lower right-hand vertex. One then gets the array of possible λ values:

$$\begin{array}{cccc}
\frac{1}{2}p & \frac{1}{2}(p-1) & \cdots & \frac{1}{2}q \\
\frac{1}{2}(p-1) & \frac{1}{2}(p-2) & \cdots & \frac{1}{2}(q-1) \\
\cdots & \cdots & & \cdots \\
\frac{1}{2}(p-q+1) & \frac{1}{2}(p-q) & \cdots & \frac{1}{2} \\
\frac{1}{2}(p-q) & \frac{1}{2}(p-q-1) & \cdots & 0.
\end{array}$$

The lines of equal m value are now 45°, linking λ states that differ by *integers*.

In this form, the results we have obtained (from

the matrices of the generators) appear as a special case of Weyl's general "branching theorem"[9] for SU_n. One advantage of the present detailed treatment is that it assigns the proper (λ, m) values to this array, and thus leads to a unique description of the state vectors.

G. The Number of States Belonging to $[pq]$

For SU_2, the calculation of the number of states belonging to the label $[p]$ was quite direct—being simply the number of points (separated by unit distance) lying within the boundaries $-\frac{1}{2}p \leq J_z \leq \frac{1}{2}p$. The calculation for SU_3 is similar in principle; we need only find the number of points within, and on, the boundary of a three-dimensional region in $(\lambda\mu m)$ space.

It is easier to proceed from the $(p - q + 1) \times (q + 1)$ array given in Sec. I.F, above. For every value of λ in this array there are $2\lambda + 1$ states. If we sum the number of states belonging to the rth row of this array one finds that the rth row contains $(p - q + 1)[r + \frac{1}{2}(p - q)]$ states.

Summing next over the number of rows $(r = 1, 2, \cdots q + 1)$, one then finds that the total number of states belonging to $[pq]$ is $\frac{1}{2}[(p-q+1)(p+2)(q+1)]$. This is the dimension of the representation $[pq]$, and, once again, is but a special case of the general dimension formula for SU_n given by Weyl.[35]

H. Some Properties of the Invariants

Although not directly relevant to the present discussion, it is useful nonetheless to discuss some properties of the invariant operators here.

The invariants I_2 and I_3 have been evaluated in terms of usual labels $[pq]$ for the irreducible representations. Let us consider, however, the invariants in terms of the magnetic quantum numbers of the maximum state, i.e., the highest weight (M_1, M_2). For this one has

$$I_2 \to M_1^2 + M_2^2 + (2/\sqrt{3})M_1, \tag{24a}$$

$$I_3 \to \tfrac{1}{3}M_2[(\sqrt{3}M_1 + 1)^2 - M_2^2]. \tag{24b}$$

Consider next only those terms in the operators I_2 and I_3 which involve the H_i alone. That is,

$$K_2 \equiv \sum_i H_i^2, \tag{25a}$$

$$K_3 \equiv \sum_{ijk} [ijk]H_iH_jH_k. \tag{25b}$$

These terms are easily evaluated:

$$K_2 = H_1^2 + H_2^2 \to M_1^2 + M_2^2, \tag{26a}$$

$$K_3 = \tfrac{1}{3}H_2(3H_1^2 - H_2^2) \to \tfrac{1}{3}M_2(3M_1^2 - M_2^2). \tag{26b}$$

[35] Reference 9, p. 381.

Comparing these expressions with the eigenvalues of I_2 and I_3 (given above), one sees that the two expressions differ only by the substitution $M_1 \to M_1 + 1/\sqrt{3}$, aside from an additive constant. Expressing this more precisely, one may assert: the eigenvalues of I_n may be obtained from the eigenvalues of K_n (operating on the state vector of highest weight) by the replacement of $M_1 \to M_1 + 1/\sqrt{3}$, $M_2 \to M_2$ in the expression $K_n(M_1, M_2) - K_n(1/\sqrt{3}, 0)$.

In other words, the evaluation of the invariants I_n may be inferred from the form of the terms in I_n involving the H_i alone. Using Weyl's determination of the characters of the SU_n group, Racah has asserted that the result we have just demonstrated for SU_3 is *true in general.* If we denote the highest weight by the vector $\mathbf{M}$, and denote by $\mathbf{R}$ the vector whose components are

$$R_i \equiv \tfrac{1}{2} \sum_{\gamma+\nu\epsilon} (\gamma i^\gamma), \tag{27}$$

then the general result is

$$I_n(\mathbf{M}) = K_n(\mathbf{M} + \mathbf{R}) - K_n(\mathbf{R}), \tag{28}$$

where $I_n(\mathbf{M})$ is the nth invariant expressed as a function of the highest weight $\mathbf{M}$, and $K_n(\mathbf{\Lambda})$ is that part of the nth invariant involving the H_i only, with $H_i \to \Lambda_i$. Using the explicit form for the invariant operators I_n, a purely infinitesimal proof of this theorem is now possible.

This remarkable result is very helpful in checking the evaluation of the invariant operators.

I. Generalization of the Method to any SU_n

Let us summarize what has been accomplished in the preceding discussion of SU_3. By means of purely infinitesimal techniques we have determined that there exist finite-dimensional unitary representations of the SU_3 group labeled by two integers p, q. These representations are defined on basis functions $|pq; \lambda\mu m\rangle$, which form an orthonormal basis uniquely labeled by the five quantum numbers: p, q for the representations; λ = labeling group; and μ_1, m_2 magnetic quantum numbers. The representations themselves are to be defined explicitly in the form

$$D^{[pq]}(\alpha_1 \cdots \alpha_8) = \exp \sum_A i\alpha_A X_{-A}^{[pq]}, \tag{29}$$

where the generators X_A have the explicit matrix form given in Eqs. (13), (14), (20), and (21). (Although this form is indeed explicit, it nonetheless requires much further discussion, which will be given subsequently.)

The fact that the labeling furnished by the canonical subgroup decomposition uniquely speci-

fies every state of a given irreducible representation is thus obtained as a corollary to the determination of the matrices of the generators, as mentioned earlier in Sec. I.A.

Let us now discuss the generalization of these results to all SU_n. We shall presuppose that SU_{n-1} has already been treated, and moreover that the Wigner coefficient for combining the fundamental representation and its conjugate are known for SU_{n-1}. [There is no difficulty in defining this (vector) Wigner coefficient in the general case, since it is in fact no more than an appropriate piece of the symmetric coupling coefficient of SU_n; the scalar Wigner coefficient is obvious.]

The generators of SU_n break up under the subgroup decomposition $SU_n \supset U_1 \times SU_{n-1}$ into four operator types:

(a) the generators of SU_{n-1} [$n(n-2)$ in number],
(b) the generator H_n,
(c) the generators that transform like the fundamental representation of SU_{n-1} ($n-1$ in number),
(d) the generators that transform like the conjugate to the fundamental representation of SU_{n-1} ($n-1$ in number).

[In the SU_3 example, $(E_\beta, E_{-\beta})$ belonged to (c); $E_{-\beta}, E_\beta$ belonged to (d); together these formed a bispinor under SU_2.]

We combine the generators in (c) with those in (d), via the Wigner coefficient of SU_{n-1}, to form two bilinear operators: a scalar S, and a "vector" $\mathbf{V}$ which has components V_A in SU_{n-1}. The invariants $I_2 \cdots I_n$ may now be expressed in terms of the invariants of SU_{n-1} and the two bilinear operators S and $\{V_A\}$. It is essential to note at this point that there are exactly $n-1$ linearly independent equations expressing the matrix elements of the bilinear operators S and $\{V_A\}$; in particular—as explicitly seen for the SU_3 example—this matrix element has $n-1$ independent components that are obtained by taking a single square root of an nth-order polynomial.

Just as for the SU_2 and SU_3 examples, the bounding hyperplanes result from setting the reduced matrix elements corresponding to the $2(n-1)$ shift operators to zero. This results in $2(n-1)$ hyperplanes in all, each occurring n times (corresponding to a nth-order polynomial). The interior region bounded by $2(n-1)$ hyperplanes then determines the range of the new labeling quantum numbers of the reduced matrix elements.

In this way one establishes in a direct (but laborious!) manner that the invariants I_n and the canonical labeling scheme above serve to uniquely label both irreducible representations, and every state belonging to a given irreducible representation.

The generalization which we have sketched for all SU_n is but an explicit carrying out of a program initiated by Casimir and Van der Waerden (for SU_2) and extended by Racah to all semisimple Lie groups.

II. EXPLICIT DETERMINATION OF THE MATRICES OF THE GENERATORS FOR ALL U_n USING YOUNG TABLEAUX AND BOSON OPERATOR TECHNIQUES

A. Introductory Survey

We have in part I of this paper shown in detail how a purely algebraic technique using only the infinitesimal generators and their commutators can be utilized to determine explicitly the matrices of the generators of the unitary groups. The essential ideas in this procedure were two: (a) the invariant operators I_n; (b) the Weyl subgroup decomposition $SU_n \supset U_1 \times SU_{n-1}$ for the labeling operators. To be sure, only the SU_3 case was really carried out in detail, and the generalization to all SU_n was merely sketched. It is clear that such a general procedure could be carried through, and thus provide a constructive proof of the sufficiency of the purely algebraic infinitesimal approach.

In point of fact, as the preceding work illustrates, such a procedure would be almost intolerably tedious and cumbersome. Although there is considerable interest, and merit, in many of the intermediate results of the algebraic approach for the actual determination of the matrices of the generators, and of the irreducible unitary bases connected with the representations, there are enormous advantages in the classical approach based upon integral procedures (even though this basis may be only implicit). The advantages of the algebraic approach, which are by no means lacking, are most apparent only in the construction of general invariants on the group ("Racah coefficients", etc.), and in similar problems where explicit reference to the basis states is not required. Similarly, algebraic methods are essential in extending the definition of the functions defined by the group to admit noninteger parameters.

Accordingly, we shall briefly sketch the current status of these alternative methods, and apply such methods to an explicit determination of the matrices of the generators of the general unitary group.

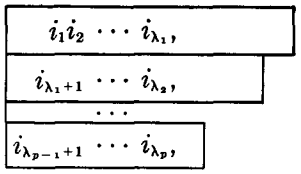
B. The Gelfand and Weyl Basis Vectors

Let us begin by noting that there is no sharp separation in the literature between the infinitesimal

and integral approaches; the classic researches of Schur, Cartan, and Weyl[36] employed both techniques where expedient. Perhaps the sharpest distinction between our purely algebraic-constructive approach and the classic methods lies in the application of the theory of the symmetric group to determining the structure of the unitary groups. Let us sketch the basis for this connection.

The fundamental (defining) representation of the unitary group, U_n, has as its carrier space the n-dimensional vector space A. The direct-product space of p such spaces, $A^{(1)} \times A \cdots \times A^{(p)} = B$, is the carrier space of the pth-rank tensors. Weyl noted that transformation induced by the operations of U_n commuted with transformations permuting the p vector spaces among themselves (the transformations induced by the operations of S_p). The transformations of this latter group are, however, completely described by the *Young symmetry patterns* defined by the partitions $[\lambda]$ of p, each pattern uniquely denoting an irreducible subspace of B. Each *Young tableau* (the Young pattern filled lexically with the integers denoting the p vector spaces) defines an operator, the Young symmetrizer, which projects the direct product space into the invariant subspace defined by the Young tableau.

This subject is completely standard,[37] and has received a classic and beautifully lucid treatment in the Princeton lectures by Weyl.[18] To fix the notation, consider the irreducible representation of the group U_n defined by the Young pattern $[\lambda] = (\lambda_1 \geq \lambda_2 \geq \cdots \lambda_0 \geq 0)$, $\sum_1^p \lambda_i = p$. The bases for this representation are the pth-rank tensors whose components are $T_{i_1 \cdots i_p}$ where i_j runs over 1 to n. The Young symmetrizer $Y_{[\lambda]}$ is the operator associated with the Young tableau,



and is the sum of products of the operators denoting permutations of the rows (denoted P_i) multiplied by the operators of the permutations of the columns, with the sign + for positive and − for negative permutations (denoted by N_i).

Thus $Y_{[\lambda]} = \sum_{i,j} N_i P_i$. The order NP is essential (PN would be equally valid but is distinct).

The basis vectors of the irreducible representation $[\lambda]$ of U_n are determined in a one-to-one correspondence with the lexical Young tableau (with repetitions) in which the indices i_j of the Young symmetrizer tableau have been assigned numerical values (1 to n). The highest weight associated with the irreducible representation $[\lambda]$ is simply $[\lambda]$ itself. The weight associated with each basis vector is $[w_i]$ where w_i is the number of indices with component i in the lexical Young tableau denoting the state. Weyl's methods provide a beautifully compact and elegant determination of the representations.

A major extension of Weyl's methods was initiated by Schwinger,[23] and considerably developed by Friedrichs,[25] Bargmann,[24] Moshinsky,[26] and Helmers,[27] among others, who noted that the p vectors of the carrier space B may be mapped onto the boson creation operators, $A_i^j \rightleftarrows a_i^{j*}$, where the superscript j runs over 1 to p and the subscript i runs over 1 to n. The boson operators obey the commutation rules: $[a_i^j, a_{i'}^{j'*}] = \delta_i^{j'} \delta_{i'}^{j'}$, with all other combinations commuting. For typographic convenience let us denote the destruction operator with a bar, $\bar{a}$, and omit the star from the creation operator.

Because of the boson nature of the a_i^j, symmetrizing is now unnecessary in the Young tableau defining the states. *The basis vectors defined by Weyl now become explicit boson operators operating on the vacuum ket* $|0\rangle$. It is convenient now to drop the superscripts and introduce the antisymmetrized product operators

$$a_{i_1 i_2 \cdots i_k} \equiv \sum \epsilon(i_1 i_2 \cdots i_k) a_{i_1}^1 a_{i_2}^2 \cdots a_{i_k}^{(k)}. \quad (30)$$

Thus, for example, the state of the representation [21] of the group U_3 defined by the tableau $\left(\begin{smallmatrix} 1 & 3 \\ 2 & \end{smallmatrix}\right)$ having weight (111) is given by

$$|[21]; 111\rangle = M^{-\frac{1}{2}} a_{12} a_3 |0\rangle. \quad (31)$$

Proper normalization is provided by the numerical factor (to be discussed below) denoted as the "measure" associated with the tableau defining the state.

We shall designate these states as "states in the Weyl basis" and the associated tableau as "Weyl basis tableaux."

It is not surprising that in dealing with the Weyl basis it is most expedient to drop the unimodular restriction and use the Weyl infinitesimal generators E_{ij} of the unitary group U_n. These generators obey

[36] Weyl's *The Classical Groups*, reference 18, is the classic exposition of the subject.

[37] D. E. Littlewood, *The Theory of Group Characters* (Oxford University Press, Oxford, England, 1950); J. S. Lomont, *Application of Finite Groups* (Academic Press Inc., New York, 1959); G. de B. Robinson, *Representation Theory of the Symmetric Group* (The University of Toronto Press, Toronto, 1961); M. Hammermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1962).

the rule

$$[E_{ij}, E_{kl}] = \delta_{jk} E_{il} - \delta_{il} E_{kj}. \tag{32}$$

The Schwinger method introduces the explicit operator realization of the E_{ij} by

$$E_{ij} = \sum_{l=1}^{p} a_i^{(l)} \bar{a}_j^{(l)}, \tag{33}$$

which are easily seen to satisfy the desired commutation rules.

The Weyl basis offers many advantages, an essential simplicity being the most attractive, yet it is not wholly satisfactory. The reason is simply that the Weyl basis is not orthogonal except for the simplest case, U_2. (This may at first appear surprising since the Weyl basis is but an expression of the Young symmetrizer, and these define primitive idempotents, i.e., $Y_{[\lambda]} Y_{[\lambda']} = \delta_{\lambda\lambda'} Y_{[\lambda]}$. The difficulty is that the orthogonality of the basis requires $Y_{[\lambda]}^{-1} Y_{[\lambda']} = \delta_{\lambda\lambda'} Y_{[\lambda]}$.) The lack of orthogonality is unimportant for determining characters, which was the purpose for which Weyl invented this basis.

One may surmise this lack of orthogonality on different grounds by noting that the weights do not provide a unique state labeling, and the Weyl basis supplements the weight of the state by the lexical requirement—which has no invariant operator significance.

As discussed earlier, the state-labeling problem is solved (for the unitary groups) by the canonical subgroup decomposition $U_n \supset U_1 \times U_{n-1}$, which is equivalent to using the branching law theorem of Weyl.[9] In terms of this, a unique labeling scheme of a unitary (orthonormal) basis may be easily devised from the tableau defining the Weyl basis. Consider the general lexical Weyl basis tableau. This has the form:

First row: $m_{11}1$'s, followed by
$(m_{12} - m_{11})2$'s, $\cdots$, $(m_{1,n} - m_{1,n-1})n$'s.
Second row: $m_{22}2$'s, followed by
$(m_{23} - m_{22})3$'s, $\cdots$, $(m_{2,n} - m_{2,n-1})n$'s.
$\cdots$
kth row: $m_{kk}k$'s, followed by
$(m_{k,k+1} - m_{k,k})(k+1)$'s, $\cdots$, $(m_{k,n} - m_{k,n-1})n$'s.

This tableau has the Young pattern defined by the partition $[\lambda] = (m_{1n}, m_{2n}, \cdots m_{nn})$.

The orthonormal state to be associated with this Weyl basis tableau will be denoted by the symbol (m), where

$$(m) \equiv \begin{bmatrix} m_{1n} & m_{2n} & \cdots & m_{nn} \\ & m_{1,n-1} & \cdots & m_{n-1,n-1} \\ & & \cdots & \\ & & m_{11} & \end{bmatrix}. \tag{34}$$

As we shall prove in a moment, the positive (including 0) integers m_{ij} obey the condition $m_{i+1,j+1} \leq m_{ij} \leq m_{i,j+1}$. *In terms of the triangular pattern the m's lie between the integers directly above,* as implied by the scheme.

We shall denote the state vectors in this unitary representation as states "in the Gelfand basis" since Gelfand[29] originated the (m) scheme.

The meaning to be assigned to the labels in the Gelfand basis can be understood most directly in terms of the Weyl basis tableau by which we defined the m_{ij}. Consider such a tableau. The weight E_{11} is m_{11}. Next by operating with E_{12}—which raises 2's into 1's (hence the name transfer operators used by Bargmann and Moshinsky)—we arrive at the highest nonvanishing lexical tableau. (Note that the $2 \to 1$ at the boundary.) The weights E_{11} and E_{22} of the new tableau are m_{12} and m_{22}, respectively. Repeating the process we arrive at the desired interpretation: the $(m_{1i}, m_{2i} \cdots, m_{ii})$ are the highest weights of the subgroup U_i contained in the canonical decomposition $U_{i+1} \supset U_1 \times U_i$.

Alternatively, we could have *defined* the symbols (m_{ij}) in terms of the eigenvalues of the invariant operators $(I_1^{(i)} I_2^{(i)} \cdots I_i^{(i)})$ belonging to the subgroup decomposition. (Note that the unit operator $I_1 = \sum_1^i E_{jj}$ is now included since the unimodular condition is not imposed.)

It is essential to point out that, although the Gelfand basis is in a one-to-one correspondence with a lexical Weyl basis tableau, the associated Gelfand basis vector and Weyl basis vector are *not* necessarily equivalent.

C. The Determination of the U_2 and U_3 Generator Matrices

In order to carry out the general determination of the matrices of the generators of the general unitary group, it is very valuable to treat first the examples of the special cases $n = 2$ and $n = 3$.

The case $n = 2$ (under the unimodular restriction) was first treated with the boson calculus by Schwinger and is, of course, closely related to Weyl's own "$\xi - \eta$ calculus." The U_2 case is particularly convenient to discuss since *the Weyl and Gelfand basis vectors are identical.*

The general basis vector (m) is thus

$$|(m)\rangle = \left| \begin{bmatrix} m_{12} & m_{22} \\ & m_{11} \end{bmatrix} \right\rangle$$
$$\equiv M^{-\frac{1}{2}} (a_{12})^{m_{22}} (a_1)^{m_{11} - m_{22}} (a_2)^{m_{12} - m_{11}} |0\rangle. \tag{35}$$

(This expression is immediately evident from the Weyl basis tableau.)

The normalization is given by

$$M = [(m_{12}+1)!m_{22}!][(m_{12}+1-m_{22})!]^{-1} \times (m_{11}-m_{22})! \times (m_{12}-m_{11})! \quad (36)$$

The normalization associated with the state vectors in the Weyl basis can be calculated directly from the definition of the state as a boson operator. What makes this situation much more interesting is the fact that this calculation need never be carried out! *The required normalization may be written down in general directly from inspection of the associated Weyl basis tableau.* It is, of course, no surprise that the normalization—or measure as we shall denote it—is a function of the tableau; what is interesting is the simplicity of the connection in terms of the recently invented concept of the "hook" of a Young-pattern node.[38] (We shall generalize this concept to the Weyl basis tableau in a subsequent paper, where it is more appropriate.)

It is immediate now to determine the matrices of the generators using the general-state Equation (34) and the operator-realization Equation (35). One finds first

$$\begin{aligned} E_{11}\,|(m)\rangle &= m_{11}\,|(m)\rangle, \\ E_{22}\,|(m)\rangle &= (m_{22}+m_{12}-m_{11})\,|(m)\rangle. \end{aligned} \quad (37)$$

Operating on the general state with E_{21} one finds that

$$\begin{aligned} &E_{21}\,|(m)\rangle \\ &= [E_{21}, (a_{12})^{m_{22}}a_1^{m_{11}-m_{22}}a_2^{m_{12}-m_{11}}][M(m)]^{-\frac{1}{2}}\,|0\rangle \\ &= [M(m)]^{-\frac{1}{2}}(m_{11}-m_{22})a_{12}^{m_{22}}a_1^{m_{11}-m_{22}-1}a_2^{m_{12}-m_{11}+1}\,|0\rangle \\ &= (m_{11}-m_{22})[M(m)/M(m')]^{-\frac{1}{2}}\,|(m')\rangle, \end{aligned} \quad (38)$$

where

$$(m') = \begin{bmatrix} m_{12} & & m_{22} \\ & m_{11}-1 & \end{bmatrix}.$$

It follows upon using Eq. (36) for $M(m)$ that

$$\langle (m')|\,E_{21}\,|(m)\rangle = \delta_{m'_{11}}^{m_{11}-1}\cdot[(m_{11}-m_{22})(m_{12}-m_{11}+1)]^{\frac{1}{2}}. \quad (39)$$

To put this result in more familiar form, let us note that the $(a_{12})^{m_{22}}$ represent m_{22} "inert pairs" under the operators (SU_2 generators) E_{12}, E_{21}, and $8^{-\frac{1}{2}}(E_{11}-E_{22}) \equiv H_1$. Thus if one effectively omits these states by shifting the (m) eigenvalues down by m_{22} and then identifies the new m

$$\begin{aligned} m_{11} - \tfrac{1}{2}m_{12} &\equiv M, \\ \tfrac{1}{2}m_{12} &\equiv J, \end{aligned} \quad (40)$$

then the usual form results.

The SU_3 system is handled in a similarly economical manner. (It is convenient to put $m_{33}=0$ and use a unimodular system.) Those SU_3 states which belong to maximum weight in the U_2 subgroup are equivalent in both the Gelfand and Weyl bases. Such states have then general definition

$$\left|\begin{array}{ccc|c} & m_{12} & & |0 \\ m_{22} & & m_{23}-m_{22} & \end{array}\right|\,m_{13}-m_{12}\Big|\Big\rangle = \left|\begin{bmatrix} m_{13} & m_{23} & 0 \\ & m_{12} & m_{22} \\ & & m_{12} \end{bmatrix}\right\rangle \equiv M^{-\frac{1}{2}} \cdot (a_{12})^{m_{22}}\cdot(a_{13})^{m_{23}-m_{22}}\cdot(a_1)^{m_{12}-m_{23}}(a_3)^{m_{13}-m_{12}}\,|0\rangle, \quad (41)$$

with

$$M = \frac{(m_{12}+1)!m_{22}!}{(m_{12}-m_{22}+1)!} \cdot\frac{(m_{13}-m_{22}+1)!(m_{23}-m_{22})!}{(m_{13}-m_{23}+1)!} \cdot(m_{12}-m_{23})!\cdot(m_{13}-m_{12})!. \quad (42)$$

This state can be written down directly from the Weyl basis tableau, *including the normalization.*

The general state of the Gelfand basis is then obtained by using the operator $(E_{21})^{m_{11}-m_{12}}$ to carry m_{11} from m_{12} in Eq. (42) to the general value m_{11}. The result of this operation is

$$\left|\begin{bmatrix} m_{13} & m_{23} & 0 \\ & m_{12} & m_{22} \\ & & m_{11} \end{bmatrix}\right\rangle = N\cdot(a_{12})^{m_{22}}(a_{13})^{m_{23}-m_{22}} \cdot(a_1)^{m_{11}-m_{23}}\cdot(a_2)^{m_{12}-m_{11}}\cdot(a_3)^{m_{13}-m_{12}} \cdot{}_2F_1(m_{22}-m_{23}, m_{11}-m_{12}, m_{11}-m_{23}+1; (a_1a_{23}/a_2a_{13}))\,|0\rangle, \quad (43)$$

where:

$$N = \left[\frac{(m_{11}-m_{22})!}{(m_{11}-m_{23})!}\cdot\frac{(m_{12}-m_{23})!}{(m_{12}-m_{22})!}\cdot\frac{(m_{12}-m_{22}+1)!}{(m_{12}+1)!m_{22}!} \cdot\frac{(m_{13}-m_{23}+1)!}{(m_{12}-m_{11})!(m_{11}-m_{23})!(m_{23}-m_{22})!(m_{13}-m_{22}+1)!(m_{13}-m_{12})!}\right]^{\frac{1}{2}}, \quad (44)$$

[38] G. de B. Robinson, reference 37, p. 44.

and ${}_2F_1(a, b, c; x)$ denotes the hypergeometric function. (Note that this function is a polynomial for the parameters here, and that, moreover, the negative powers of the boson operators do *not* actually appear in the explicit polynomial form. It is also useful to note that the explicit polynomial form corresponds to a sum which includes *nonlexical* Weyl basis tableaux.)

The use of the hypergeometric function is not just a convenient symbolism. One frequently needs the contiguous relations and Kummer transformations which the ${}_2F_1$ notation makes evident.

The matrices of the generators now follow from these state vectors. For the diagonal operators one has

$$\begin{aligned} E_{11} &\to m_{11}, \\ E_{22} &\to m_{12} + m_{22} - m_{11}, \\ E_{33} &\to m_{13} + m_{23} - m_{12} - m_{22}. \end{aligned} \tag{45}$$

[The Gelfand basis is chosen such that the operator $E_{ii} \to \sum_i (m_{ii} - m_{i,i-1})$.]

For the operator E_{32}, one finds, by commutation with the boson operator of the general state, that

$$\begin{aligned} E_{32} \left| \begin{pmatrix} m_{13} & m_{23} & 0 \\ & m_{12} & m_{22} \\ & & m_{11} \end{pmatrix} \right\rangle \\ = \frac{N}{N'} m_{22} \left(\frac{m_{11} - m_{22} + 1}{m_{12} - m_{22} + 1} \right) \cdot \left| \begin{pmatrix} m_{13} & m_{23} & 0 \\ & m_{12} & m_{22} - 1 \\ & & m_{11} \end{pmatrix} \right\rangle \\ + (m_{12} - m_{11}) \left(\frac{m_{12} + 1}{m_{12} - m_{22} + 1} \right) \frac{N}{N''} \\ \times \left| \begin{pmatrix} m_{13} & m_{23} & 0 \\ & m_{12} - 1 & m_{22} \\ & & m_{11} \end{pmatrix} \right\rangle, \end{aligned} \tag{46}$$

where N'', refer to the normalization integral evaluated for the new parameters.

In order to obtain this relation one needs the contiguous relation identity

$$\begin{aligned} & m_{22}(a_2 a_{13})\ {}_2F_1(m_{22} - m_{23}, m_{11} - m_{12}, m_{11} - m_{23} + 1; x) \\ & + (m_{12} - m_{11})(a_3 a_{12}) \\ & \cdot {}_2F_1(m_{22} - m_{23}, m_{11} - m_{12} + 1, m_{11} - m_{23} + 1; x) \\ & = m_{22} \cdot \left(\frac{m_{11} - m_{22} + 1}{m_{12} - m_{22} + 1} \right) (a_2 a_{13}) \\ & \cdot {}_2F_1(m_{22} - 1 - m_{23}, m_{11} - m_{12}, m_{11} - m_{23} + 1; x) \\ & + (m_{12} - m_{11}) \left(\frac{m_{12} + 1}{m_{12} - m_{22} + 1} \right) (a_3 a_{12}) \\ & \cdot {}_2F_1(m_{22} - m_{23}, m_{11} - m_{12} + 1, m_{11} - m_{23} + 1; x) \\ & + (a_1 a_{23} - a_2 a_{13} + a_3 a_{12}) \cdot [\cdots]. \end{aligned} \tag{47}$$

The left side of this identity arises upon commutation with E_{32}. (Note that the last term on the right-hand side *vanishes identically*. The occurrence of such expressions is the real difficulty in working with these boson-operator states, for it will be quickly seen that commutation with the identity $\sum a_i a_{jk}\, \epsilon_{ijk} = 0$ can produce invalid relations involving nonlexical Weyl basis operators.)

Introducing the explicit form of the normalization, one finds, for the matrices of the generator E_{32}, the result

$$\begin{aligned} E_{32} \left| \begin{pmatrix} m_{13} & m_{23} & m_{33} \\ & m_{12} & m_{22} \\ & & m_{11} \end{pmatrix} \right\rangle \\ = \left[\frac{(m_{12} - m_{11})(m_{13} - m_{12} + 1)(m_{23} - m_{12})(m_{33} - m_{12} - 1)}{(m_{12} - m_{22} + 1)(m_{12} - m_{22})} \right]^{\frac{1}{2}} \left| \begin{pmatrix} m_{13} & m_{23} & m_{33} \\ & m_{12} - 1 & m_{22} \\ & & m_{11} \end{pmatrix} \right\rangle \\ + \left[\frac{(m_{22} - m_{11} - 1)(m_{13} - m_{22} + 2)(m_{23} - m_{22} + 1)(m_{33} - m_{22})}{(m_{12} - m_{22} + 2)(m_{12} - m_{22} + 1)} \right]^{\frac{1}{2}} \left| \begin{pmatrix} m_{13} & m_{23} & m_{33} \\ & m_{12} & m_{22} - 1 \\ & & m_{11} \end{pmatrix} \right\rangle. \end{aligned} \tag{48}$$

This result is seen to correspond exactly to the results obtained in the earlier[39] [Eqs. (21a) and (21b)] upon using the correspondence

$$\begin{aligned} p &= m_{13} - m_{33}, \quad q = m_{23} - m_{33}, \quad \lambda = \tfrac{1}{2}(m_{12} - m_{22}), \\ \mu &= m_{11} - \tfrac{1}{2}(m_{12} + m_{22}), \quad m = \tfrac{1}{2}(m_{12} + m_{22}) - \tfrac{1}{3}(m_{13} + m_{23} + m_{33}). \end{aligned} \tag{49}$$

[39] By Eq. (14) we must multiply the expressions (21a) and (21b) by the respective Wigner coefficients $C^{\lambda}_{\mu-\frac{1}{2}\,\mu-\frac{1}{2}}$ and $C^{\lambda}_{\mu-\frac{1}{2}\,\mu-\frac{1}{2}}$ in order to obtain the entire matrix element corresponding to Eq. (48).

(It should be noted that this result corrects the misprinted result of reference 29.)

Repetition of the formulas for the remaining matrices in the Gelfand notation seems unnecessary.

D. Generalization to Obtain the Matrices of All U_n

In principle, the method illustrated for U_3 can be generalized immediately to the case for arbitrary n. There is no inherent limitation to the method whereby normalized states for U_n of highest weight in U_{n-1} are written out immediately in terms of the equivalent Weyl and Gelfand basis functions and then lowered by the operators of the U_{n-1} subgroup to obtain the general basis vector explicitly. (This procedure is not as complicated as it might appear, since the functions analogous to the ${}_2F_1$ of the U_3 case show considerable symmetry in their structure.) The matrices of the generators are then obtained directly, exactly as for U_3.

This direct, or "brute force" method is, however, quite unnecessary for the determination of the generator matrices, and the complete determination results from considering only the very few Weyl basis tableaux relevant to the calculation.

The simplicity to which we refer is the simplicity "built into" the generator matrices by explicitly maintaining the canonical subgroup factorization in evidence at all stages. It follows from this that the matrices of the generators *factorize* in the familiar pattern of the Racah–Wigner calculus into two parts: (a) the reduced matrix element of the operator E_{ii} of U_n; and (b) a Wigner coefficient of the subgroup U_{n-1}. The factor in (a) is independent of the parameters of the U_{n-2} subgroup (the analogue of magnetic quantum numbers in the angular-momentum paradigm). The factor in (b) is a well-defined Wigner coefficient since it refers to the reduction of $[\lambda] \times [10 \cdots 0]$ in the U_{n-1} group, and this direct product is always multiplicity-free.

These considerations show that in a recursive calculation from U_{n-1} to U_n only the matrices E_{ni} with $1 \leq i \leq n$ need be considered, all other generator matrices being known by hypothesis from the U_{n-1} calculation. Of the n "new" matrices, E_{nn} is, however, diagonal with the eigenvalue $\sum_{i=1}^{n} m_{i,n} - \sum_{i=1}^{n-1} m_{i,n-1}$. Moreover, all the $E_{n,i}$ with $i < n-1$ may be calculated from the single matrix element $E_{n,n-1}$ and the matrix of $E_{n-1,i}$, using the commutator relation $E_{n,i} = [E_{n,n-1}, E_{n-1,i}]$.

Thus the entire constructive determination of the generator matrices depends upon knowing *the single new matrix* $E_{n,n-1}$ at each step in a chain determination through the subgroup U_{n-1}. The determination of $E_{n,n-1}$ is, however, as we shall show, determined directly from the measures (M) associated with two Weyl basis tableaux.

We shall accordingly give the determination of $E_{n,n-1}$ in detail and simply state the remaining generator matrices. The matrix $E_{n,n-1}$ itself illustrates clearly the general structure of the desired answers.

Let us now determine the reduced matrix element of $E_{n,n-1}$ for the group U_n. Since reduced matrix elements in U_n are independent of the parameters of the U_{n-2} subgroup, we may choose these parameters to correspond to the state of highest U_{n-1} weight. This state belongs to equivalent Weyl and Gelfand bases and has the explicit form

$$\left|\begin{array}{ccccccc} m_{1n} & & m_{2n} & \cdots & m_{n-1,n} & & m_{nn} = 0 \\ & m_{1,n-1} & & m_{2,n-1} & \cdots & m_{n-1,n-1} & \\ & = & & = & & & \\ & & m_{1,n-2} & & m_{2,n-2} & \cdots & \\ & & = & & = & & \\ & & & m_{1,1} & = & m_{1,n-1} & \end{array}\right\rangle$$

$$= \left|\begin{array}{ccccc} m_{1,n-1} & & & 0 \cdots 0 & m_{1,n} - m_{1,n-1} \\ m_{2,n-1} & & \cdots 0 \cdots & & m_{2,n} - m_{2,n-1} \\ \cdots & & & & \\ m_{n-1,n-1} & m_{n-1,n} - m_{n-1,n-1} & & & \end{array}\right\rangle$$

$$= [M(m)]^{-\frac{1}{2}} \cdot (a_{1,2\cdots,n-1})^{m_{n-1,n-1}} \cdot (a_{1,\cdots n-2,n})^{m_{n-1,n}-m_{n-1,n-1}}$$
$$\cdot (a_{1,2,\cdots,n-2})^{m_{n-2,n-1}-m_{n-1,n}} \cdots (a_{1,n})^{m_{2,n}-m_{2,n-1}} \cdot (a_1)^{m_{1,n-1}-m_{2,n}} \cdot (a_n)^{m_{1,n}-m_{1,n-1}} |0\rangle. \quad (50)$$

The boson operator given in the above equation appears rather complicated in its dependence on the (m) parameters. In terms of the Weyl basis tableau, however, the correspondence is much more easily seen. [Note also that we have put $m_{n,n} = 0$ (unimodular restriction) for convenience. The corresponding boson operator is $(a_{1,2,\cdots,n})^{m_{n,n}}$ which commutes with all $E_{i<j}$ and thereby drops out of all results. One may reinsert $m_{n,n} \neq 0$ by letting $m_{ij} \to m_{ij} - m_{n,n}$ $(i, j \neq n, n)$ in all final results.]

The calculation of the reduced matrix element of $E_{n,n-1}$ is easily seen now to be simply

$$\langle (m') | \, |E_{n,n-1}| \, | (m) \rangle = [M(m')/M(m)]^{\frac{1}{2}} \cdot m_{n-1,n-1}, \tag{51}$$

where (m') has $m'_{ij} = m_{ij}$ except for $m'_{n-1,n-1} = m_{n-1,n-1} - 1$.

We have yet to specify the measure associated with the Weyl tableau of (m) [Eq. (50)]. This measure can be calculated directly from the definition [that $\langle (m) \rangle$ be normalized] and has the form

$$M(m) = \frac{\prod_{i=1}^{n-1} (m_{i,n-1} + n - i - 1)!}{\prod_{i=1}^{n-2} (m_{i,n-1} - m_{n-1,n-1} + n - i - 1)!} \times \frac{\prod_{i=1}^{n-1} (m_{in} - m_{n-1,n-1} + n - i - 1)!}{\prod_{i=1}^{n-2} (m_{in} - m_{n-1,n} + n - i - 1)!} \times \cdots \times (m_{1,n-1} - m_{2,n})! \times (m_{1,n} - m_{1,n-1})!. \tag{52}$$

This result, as mentioned, can be obtained by direct calculation, although the method is quite tedious to be sure. What is far more interesting is that the answer may be written out immediately by defining the *hook structure* of the Weyl basis tableau. The hook length of an (ij) node in a Young pattern[40] is defined by

$$h_{ij} \equiv \lambda_i - j + \lambda_j' - i + 1,$$

where $[\lambda]$ is the partition and $[\lambda']$ the conjugate partition of the Young pattern. To define the hook structure of a Weyl basis tableau requires further extension of the concept than apparently is in the current literature, and we hope to be permitted to discuss this subject in detail in a subsequent paper.

The ratio of the two measures $M(m)/M(m')$ is much simpler than the actual measures themselves. The result follows from the explicit form given above and is

$$\frac{M(m)}{M(m')} = m_{n-1,n-1} \cdot \frac{\prod_{i=1}^{n-2} (m_{i,n-1} - m_{n-1,n-1} + n - i)}{\prod_{i=1}^{n-1} (m_{in} - m_{n-1,n-1} + n - i)}. \tag{53}$$

The simplicity of this result is a consequence of the fact that only the *change* in the hook structure caused by a shift in the boundary between the $(n-1)$'s and the n's in the bottom row of the Weyl basis tableau is involved. In other words, we may write out the hook changes at the relevant boundary *directly from the tableau*, and thereby obviate the evaluation of $M(m)$ itself.

It follows that the matrix element of $E_{n,n-1}$ corresponding to $(m) \to (m')$ is given by

$$\langle (m') | \, |E_{n,n-1}| \, | (m) \rangle = \left[(m_{n-1,n-1}) \cdot \frac{\prod_{i=1}^{n-1} (m_{i,n} - m_{n-1,n-1} + n - i)}{\prod_{i=1}^{n-2} (m_{i,n-1} - m_{n-1,n-1} + n - i)} \right]^{\frac{1}{2}}, \tag{54}$$

where $(m') = (m)$ except $m'_{n-1,n-1} = m_{n-1,n-1} - 1$.

It remains now to calculate the Wigner coefficient in U_{n-1} corresponding to the matrix element of $E_{n,n-1}$ evaluated in U_{n-1}. For this calculation we may choose the most convenient values of the parameters m_{in}, since the reduced matrix element has already determined the dependence on these variables. Moreover, the desired Wigner coefficient is independent of the parameters of the subgroup U_{n-3}, since $[E_{n,n-1}, E_{ij}] = 0$ for $i, j \leq n-2$. Thus we may choose these parameters in the most convenient way also.

The optimal choice of the free parameters would be one yielding a Weyl basis vector equivalent to a Gelfand basis vector. This is in fact possible.

Let us choose the parameters for the initial state $(\bar{m})$ in the following way:

$$(\bar{m}) \equiv \begin{bmatrix} m_{1,n-1} & m_{1,n-1} & m_{2,n-1} & \cdots & m_{n-2,n-1} & 0 \\ & m_{1,n-1} & m_{2,n-1} & \cdots & m_{n-2,n-1} & m_{n-1,n-1} \\ & & m_{1,n-2} & m_{2,n-2} & \cdots & m_{n-2,n-2} \\ & & = & = & \cdots & \\ & & & \cdots & & \\ & & & m_{1,n-2} & & \end{bmatrix}. \tag{55}$$

[40] G. de B. Robinson, reference 37, pp. 44 and 166. On page 44 Robinson notes explicitly the "striking result involving hook lengths, which is of fundamental importance."

This set of parameters corresponds to the Weyl basis tableau:

$m_{1,n-2}$	$0 \cdots$	$\cdots 0$		$m_{1,n-1} - m_{1,n-2}$
$m_{2,n-2}$	$0 \cdots$	$\cdots 0$	$m_{2,n-1} - m_{2,n-2}$	$m_{1,n-1} - m_{2,n-1}$
	$\cdots$		$m_{2,n-1} - m_{3,n-1}$	
	$\cdots$			
$m_{n-2,n-2}$	$m_{n-2,n-1} - m_{n-2,n-2}$	$m_{n-3,n-1} - m_{n-2,n-1}$		
$m_{n-1,n-1}$	$m_{n-2,n-1} - m_{n-1,n-1}$			

(56)

The characteristic feature of this tableau—which determined its choice—is the fact that in all except the last row $(n - 1)$'s cannot be changed into n's without the tableau vanishing (nonlexical).

First we must prove that this Weyl basis vector corresponds to a Gelfand basis vector. To prove this requires that one demonstrate that the Weyl basis vector corresponds to an eigenstate of the U_{n-1} system in the canonical subgroup decomposition. Now in U_{n-2}, the tableau corresponds to a maximal state and hence is "sharp" in U_{n-2}. If we delete all n's from the tableau, the resulting state (in U_{n-1}) is clearly sharp (since it corresponds to maximal U_{n-2}, which has been demonstrated sufficient). But adding the n's back to the tableau changes every $a_{ij\cdots;}$ to an $a_{ij\cdots;n}$ with the single exception of the operator $(a_{1,2,\cdots n-1})^{m_{n-1,n-1}}$ which remains unchanged. This operator, however, commutes with all operators E_{ij}, $i \neq j$, $i, j \leq n - 1$, and may be disregarded for the U_{n-1} subgroup. Noting that all properties in U_{n-1} of the operators $a_{ijk\cdots;}$ are preserved under $a_{ijk\cdots;} \rightarrow a_{ijk\cdots;n}$, one sees that indeed the state above is sharp in U_{n-1} and hence corresponds to a Gelfand state.

The desired matrix element is now easily determined, since the $E_{n,n-1}$ operator affects only the boson operator $(a_{1,2,\cdots n-1})^{m_{n-1,n-1}}$ and shifts $m_{n-1,n-1} \rightarrow m_{n-1,n-1} - 1$, bringing down the factor $m_{n-1,n-1}$. (This may be read off the Weyl basis tableau.) Thus the desired matrix element is

$$\langle(\bar{m}')| E_{n,n-1} |(\bar{m})\rangle = [M(\bar{m}')/M(\bar{m})]^{\frac{1}{2}} \cdot (m_{n-1,n-1}). \quad (57)$$

The shift in the measure of the initial state $(\bar{m})$ under the change in $m_{n-1,n-1} \rightarrow m_{n-1,n-1} - 1$ may be calculated from changes in the hook structure of the $(\bar{m})$ tableau. Thus one finds:

$$M(\bar{m})/M(\bar{m}') = (m_{n-1,n-1}) \cdot \prod_{i=1}^{n-2} (m_{i,n-2} - m_{n-1,n-1} + n - i - 1). \quad (58)$$

To calculate the Wigner coefficient, we must normalize this result by its value when $(\bar{m}) = (m)$ confined to U_{n-1} (the initial state for the reduced matrix-element calculation).

The desired Wigner coefficient $[\lambda] \times [1]$ in U_{n-1} corresponding to the operator $E_{n,n-1}$ with quantum numbers $(m) \rightarrow (m')$ (restricted to U_{n-1}) is then found to be

$$\left[\frac{\prod_{i=1}^{n-2} (m_{i,n-2} - m_{n-1,n-1} + n - i - 1)}{\prod_{i=1}^{n-2} (m_{i,n-1} - m_{n-1,n-1} + n - i - 1)}\right]^{\frac{1}{2}}. \quad (59)$$

Let us note that the explicit evaluation of the fundamental Wigner coefficient $[\lambda] \times [1]$ for *general* U_n has not been previously reported in the literature. The result given in Eq. (59) is, to be sure, only one component of the general result, which is in fact a supermatrix array corresponding to the subgroup decomposition of the Wigner coefficients. The fact that a supermatrix array occurs is intimately connected with the fact that the group is *non*-simply reducible. We prefer, however, to discuss this matter systematically in a subsequent paper.

To complete the argument one multiplies the reduced matrix element Eq. (54) and the Wigner coefficient Eq. (59) to obtain the matrix of $E_{n,n-1}$ corresponding to $(m) \rightarrow (m')$. The structure of the final answer is quite clear, and one sees that *the matrix element may in fact be read off the hook structure changes in the Weyl basis tableaux.*

We shall utilize this simple structure to obviate the calculation of the remaining $(n - 2)$ components of the matrix $E_{n,n-1}$. For completeness of thesis, however, let us note that the reduced matrix elements are readily calculated from the state of maximal U_{n-1} weight after shifting by a lowering operator in U_{n-2}. This shift eliminates the vanishing U_{n-1} Wigner coefficient (which occurs for the maximal U_{n-1} weight) and enables the reduced matrix element to be calculated. Similar considerations

hold for the optimal state used for the Wigner-coefficient calculation.

The final results for the matrix elements of $E_{n,n-1}$ are

$$E_{n,n-1}\left|\begin{bmatrix} m_{1,n} & m_{2,n} & \cdots & m_{nn} \\ & m_{1,n-1} & \cdots & m_{i,n-1} \cdots m_{n-1,n-1} \\ & & \cdots & \\ & & m_{11} & \end{bmatrix}\right\rangle$$

$$= \sum_{i=1}^{n-1}\left\{\left[\frac{\prod_{j=1}^{n-2}(m_{j,n-2}-m_{i,n-1}-j+i)}{\prod_{\substack{j=1\\ j\neq i}}^{n-1}(m_{j,n-1}-m_{i,n-1}-j+i)}\right]^{\frac{1}{2}}\right.$$

$$\times\left.\left[(-)\frac{\prod_{j=1}^{n}(m_{j,n}-m_{i,n-1}-j\pm i+1)}{\prod_{\substack{j=1\\ j\neq i}}^{n-1}(m_{j,n-1}-m_{i,n-1}-j+i+1)}\right]^{\frac{1}{2}}\right\}$$

$$\times\left|\begin{bmatrix} m_{1,n} & m_{2,n} & \cdots & m \\ m_{1,n-1} & \cdots & m_{i,n-1}-1 & \cdots m_{n-1,n-1} \\ & & \cdots & \\ & & m_{11} & \end{bmatrix}\right\rangle. \quad (60)$$

These results agree with those given in reference 29, with the important exception that the factor $(-)^{k+1}$ appearing in reference 29 is in error. Let us note that this result for the matrix element of $E_{n,n-1}$ explicitly shows the factorization into (reduced matrix) $\times$ (Wigner coefficient), where the $(n-1)$ reduced matrix elements are given by the second factor in Eq. (60), and the $(n-1)$ Wigner coefficients are the first factor.

It is quite striking that these results—when interpreted in terms of hook structure changes induced by the operations of $E_{n,n-1}$ *on the tableau*—can be obtained without requiring the intermediary determination of an explicit basis. This property is of great potential significance since it indicates that the matrix elements may be defined on tableaux directly in terms of hook structures, with the transformations appearing as operations on tableaux. If this property holds true for more general operators than the generators, it affords the interesting possibility of complete mechanization of such matrix-element calculations by digital computers.

The complete result for the matrices of the generators $E_{n,k}(n > k)$ then follows from the use of the commutator relations and the matrix of the generator $E_{n,n-1}$. The general matrix element factors into two parts: the reduced matrix element [given in Eq. (60)]; and the *general* Wigner coefficient, which in turn factors in accord with the subgroup decomposition. The explicit result is given by

$$\langle(m')|\,E_{n,k}\,|(m)\rangle = \begin{bmatrix} n \\ i_1 : n-1 \end{bmatrix}$$

$$\times\left\{\prod_{l=1}^{n-k-1}\begin{bmatrix} i_l & : n-l \\ i_{l+1} : n-l-1 \end{bmatrix}\times\begin{bmatrix} i_{n-k} : k \\ k-1 \end{bmatrix}\right\}, \quad (61)$$

where: (1) the new Gelfand basis (m') is specified by the $(n-k)$ indices $i_1, \cdots, i_{n-k}$ and $(m') = (m)$ except for $m'_{i_j,n-j} = m_{i_j,n-j} - 1$; (2) the reduced matrix element given by the second factor in Eq. (60) is denoted by the symbol

$$\begin{bmatrix} n \\ i_1 : n-1 \end{bmatrix};$$

(3) the general Wigner coefficient is the factor in curly brackets, consisting of the U_k Wigner coefficient

$$\begin{bmatrix} i_{n-k} : k \\ k-1 \end{bmatrix},$$

given for U_n by the first factor in Eq. (60), and the "reduced Wigner coefficient"

$$\prod_{l=1}^{n-k-1}\begin{bmatrix} i_l & : n-l \\ i_{l+1} : n-l-1 \end{bmatrix}$$

defined by

$$\begin{bmatrix} i_l & : n-l \\ i_{l+1} : n-l-1 \end{bmatrix} = S(i_{l+1}-i_l)$$

$$\times\left[\frac{\prod_{i=1}^{n-l-1}(m_{i,n-l-1}-m_{i_l,n-l}-i+i_l)}{\prod_{\substack{i=1\\ i\neq i_l}}^{n-l}(m_{i,n-l}-m_{i_l,n-l}-i+i_l)}\right.$$

$$\times\left.\frac{(-)\prod_{i=1}^{n-l}(m_{i,n-l}-m_{i_{l+1},n-l-1}-i+i_{l+1}+1)}{\prod_{\substack{i=1\\ i\neq i_{l+1}}}^{n-l-1}(m_{i,n-l-1}-m_{i_{l+1},n-l-1}-i+i_{l+1}+1)}\right]^{\frac{1}{2}}$$

$$\times[(m_{i_l,n-l}-m_{i_{l+1},n-l-1}-i_l+i_{l+1}+1)$$

$$\times(m_{i_l,n-l}-m_{i_{l+1},n-l-1}-i_l+i_{l+1})]^{-\frac{1}{2}}, \quad (62)$$

where $S(x)$ is the *sign* of x and $S(0)$ is defined to be $+1$.

This result for $\langle(m')|\,E_{n,k}\,|(m)\rangle$ was given by Gelfand and Zetlin, but our result above differs in two essential respects: (1) an undetermined $(\pm)$ sign in reference (29) is determined explicitly here; and (2) *the significance of the structure of the result* (factorization into reduced matrix elements, ...) *is made explicit.*

A Class of Null Flat-Space Coordinate Systems*

EZRA T. NEWMAN AND T. W. J. UNTI
Physics Department, University of Pittsburgh, Pittsburgh, Pennsylvania
(Received 21 June 1963)

A new coordinate system, intrinsically attached to an arbitrary timelike world line, is investigated in flat-space time. The Maxwell field tensor associated with the field of an arbitrarily moving charged particle assumes a particularly simple form in this, its intrinsic coordinate system. This reference frame is expected to be useful in General Relativity, in asymptotic studies of radiation, and equations of motion.

I. INTRODUCTION

RECENTLY there has been considerable interest in the asymptotic behavior of the metric and Riemann tensors at spatial infinity, in spaces which are asymptotically flat. The investigations can be roughly divided into two approaches. The first approach[1-3] is based upon a special choice of coordinate system, the "Bondi-type" coordinates. In this system, the limiting value of the metric tensor at spatial infinity is the Minkowski metric, or its polar coordinate version. Using Bondi-type coordinates, the behavior of the Riemann tensor can be derived and the remaining coordinate freedom can be investigated.[1,3] The second approach[4,5] assumes a specific asymptotic behavior of two physical components of the empty-space Riemann tensor, from which the behavior of the remaining components of the Riemann tensor can be deduced. It can be shown[5] (assuming certain topological restrictions) that it is then possible to introduce the Bondi-type coordinates.

Although the Bondi-type coordinates have proved to be extremely useful, it is plausible of course that other coordinate systems, "tailor made" for special cases, might be more useful in those special cases. As an example, it is seen that the metric tensor in most of the Robinson–Trautman[6] metrics, is singular at $r = \infty$ although the Riemann tensor vanishes there. Yet if the Bondi-type coordinates were introduced, the desirable closed form of these solutions would be lost. In other words, occasionally there are coordinate systems intrinsic to a problem other than asymptotically flat coordinates.

In the present paper we investigate a new coordinate system in flat space. This reference system is intrinsically attached to an arbitrary timelike world line. The metric associated with these coordinates contains a term linear in r, and, hence, increases linearly as r approaches infinity.

We exhibit the Maxwell field tensor of the field of an arbitrarily moving charged particle, to demonstrate its simple form in this, its intrinsic coordinate system.

In a future paper we plan to show how this coordinate system and associated metric can be used in General Relativity to study asymptotically certain radiation problems and equations of motion.

II. DERIVATION OF A SPECIAL FLAT-SPACE LINE ELEMENT

In this section we transform from a Minkowski coordinate system (y^μ) to a reference frame (x^μ) attached to an arbitrary timelike world line. Figure 1 shows this world line, whose proper time is denoted by u and whose coordinates in Minkowski space are $y^\mu = \xi^\mu(u)$.

Any point in Minkowski space lies on one of the

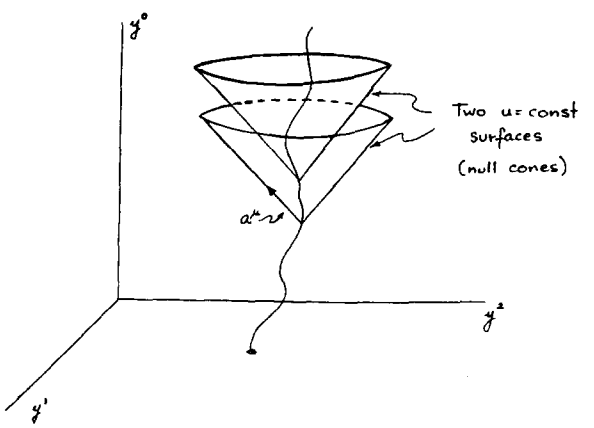


FIG. 1. Null surfaces emanating from an arbitrary timelike world line. (One spatial dimension is suppressed.)

* This work has been supported by the Office of Aerospace Research, United States Air Force.

[1] H. Bondi, Proc. Roy. Soc. (London) **269**, 21 (1962).
[2] R. K. Sachs, Proc. Roy. Soc. (London) **270**, 103 (1962).
[3] R. K. Sachs, Phys. Rev. **128**, 2851 (1962).
[4] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[5] E. Newman and T. Unti, J. Math. Phys. **3**, 891 (1962).
[6] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **265**, 463 (1962).

future cones emanating from the world line. The Minkowski coordinates can be expressed in terms of four parameters associated with the null surfaces in the following manner.[7]

Consider the geodesic line $y^\mu = a^\mu r + \xi^\mu(u)$ (satisfying the geodesic equation $d^2y^\mu/dr^2 = 0$), with a^μ an arbitrary null vector and r the affine parameter along the geodesic; the origin of r being chosen at $\xi^\mu(u)$. As a^μ sweeps out all null directions, the geodesic lines form the null cone; i.e., if we let a^μ be a function of two coordinates on a sphere, x^i $(i = 2, 3)$, u remaining constant,

$$y^\mu = a^\mu(x^i)r + \xi^\mu(u)$$

is the parametric form of the null cone with parameters x^i labeling the geodesics and r indicating where y^μ lies on the geodesic. If we now allow u to vary (with a gain in generality, we can let a^μ depend on u as well as x^i), the equation

$$y^\mu = a^\mu(u, x^i)r + \xi^\mu(u) \tag{1}$$

can be looked upon as the coordinate transformation from the Minkowski coordinates y^μ to $x^0 = u$, $x^1 = r$ and x^i.

Since a^μ is a null vector, we have (denoting $\partial/\partial u$ by a dot and $\partial/\partial x^i$ by comma i),

$$a^\mu a_\mu = 0, \tag{2}$$

$$a^\mu \dot{a}_\mu = 0, \tag{3}$$

$$a^\mu a_{\mu,i} = 0, \tag{4}$$

where summation is performed with the Minkowski metric $\eta_{\mu\nu} = \{1, -1, -1, -1\}$. In addition, the velocity vector of the world line satisfies

$$\dot{\xi}^\mu \xi_\mu = 1, \tag{5}$$

since u is proper time. r, being an affine parameter, is defined only up to a linear transformation. This arbitrariness may be used to normalize the projection of a^μ on the velocity vector ξ^μ,

$$a^\mu \xi_\mu = 1. \tag{6}$$

The metric tensor in the new coordinate system obtained by transformation (1) on the $\eta_{\mu\nu}$ is

$$g_{00} = 1 + 2\dot{a}^\mu \xi_\mu r + \dot{a}^\mu \dot{a}_\mu r^2, \quad g_{01} = 1, \quad g_{0k} = \dot{a}^\mu a_{\mu,k} r^2,$$
$$g_{11} = 0, \quad g_{1k} = 0, \quad g_{mn} = a^\mu{}_{,m} a_{\mu,n} r^2. \tag{7}$$

At this point, the labeling x^i on the hypersurface is arbitrary, in other words we have the coordinate freedom

$$u = u',$$
$$r = r', \tag{8}$$
$$x^i = x^i(u', x^{i'}).$$

Part of this freedom can be used up by making the two-dimensional metric g_{ij} conformal to the Euclidean plane,[8] i.e.,

$$a^\mu{}_{,i} a_{\mu,j} = -\delta_{ij}/2P^2. \tag{9}$$

P is defined by this equation. The remaining freedom is the analytic transformation

$$\zeta = \zeta(u', \zeta'), \tag{10}$$

where $\zeta = x^2 + ix^3$, $\zeta' = x^{2'} + ix^{3'}$.

It will be possible (proved later) to choose a transformation, Eq. (10), such that

$$a^\mu{}_{,i} \dot{a}_\mu = 0. \tag{11}$$

Equation (11) with Eqs. (2), (3), and (4) lead to $\dot{a}_\mu \propto a_\mu$. With these results, the metric Eq. (7) becomes

$$g_{00} = 1 + 2\dot{a}_\mu \xi^\mu r = 1 - 2(\dot{P}/P)r,$$
$$g_{01} = 1, \quad g_{0k} = 0,$$
$$g_{11} = 0, \quad g_{1k} = 0, \tag{12}$$
$$g_{mn} = a^\mu{}_{,m} a_{\mu,n} r^2 = -(r^2/2P^2)\delta_{mn},$$

where now P is defined as

$$P = g(u)\zeta\bar{\zeta} + M(u)\zeta + \bar{M}(u)\bar{\zeta} + h(u), \quad \zeta = x^2 + ix^3. \tag{13}$$

The $g(u)$, $M(u)$, and $h(u)$ are functions of u, related (as will be shown shortly) to the acceleration of the world line.

A method of showing that Eq. (11) can always be satisfied is as follows. First choose a transformation (10) such that the P in Eq. (9) takes the form[6]

$$P = (1 + \tfrac{1}{4}\zeta\bar{\zeta})/\sqrt{2}.$$

In general, $a^\mu{}_{,i}\dot{a}_\mu$ will be different from zero but its dependence on the acceleration can be worked out explicitly. Now do another transformation (10) demanding that $g_{0i} = 0$. This leads to the following differential equation:

$$\partial\zeta/\partial u = A(u) - \alpha(u)\zeta - \tfrac{1}{4}\bar{A}(u)\zeta^2. \tag{14}$$

The functions $A(u)$ and $\alpha(u)$ are defined by considering a spacelike triad λ^μ_A $(A = 1, 2, 3)$ defined at one point of the world line, orthogonal to ξ^μ,

[7] The range and summation conventions used here are: lower-case greek indices 0, 1, 2, 3; lower-case latin indices 2, 3.

[8] L. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Ginn and Company, Boston, Massachusetts, 1949), Chap. 2, Sec. 40.

and Fermi-propagated along the world line. Then

$$A(u) = (\lambda_{1\mu} + i\lambda_{2\mu})\ddot{\xi}^{\mu}, \qquad \alpha(u) = \lambda_{3\mu}\ddot{\xi}^{\mu}, \qquad (15)$$

where $\ddot{\xi}^{\mu}$ is the acceleration of the world line.

Equation (14), being the generalized Ricatti equation, always possesses solutions. The solution $\zeta = \zeta(u, \zeta')$ (where ζ' is a complex constant of integration), used as a coordinate transformation leads to $g_{0i} = 0$, or that Eq. (11) is satisfied.

By a theorem of Euler's,[9] ζ can be stated as an explicit function of ζ' (though not of u) and the transformation can be carried out in detail on the metric, yielding Eqs. (12) and (13). The functions g, M, and h of u can now be stated as functions of the acceleration;

$$g(u) = \kappa(\zeta_0 - \zeta_2)(\bar{\zeta}_0 - \bar{\zeta}_2)(4 + \zeta_1\bar{\zeta}_1),$$
$$M(u) = \kappa(\zeta_0 - \zeta_2)(\bar{\zeta}_2 - \bar{\zeta}_1)(4 + \bar{\zeta}_0\zeta_1),$$
$$h(u) = \kappa(\zeta_2 - \zeta_1)(\bar{\zeta}_2 - \bar{\zeta}_1)(4 + \zeta_0\bar{\zeta}_0),$$

where

$$\kappa \equiv [32(\zeta_1 - \zeta_0)(\bar{\zeta}_1 - \bar{\zeta}_0)(\zeta_0 - \zeta_2) \times (\bar{\zeta}_0 - \bar{\zeta}_2)(\zeta_2 - \zeta_1)(\bar{\zeta}_2 - \bar{\zeta}_1)]^{-\frac{1}{2}}.$$

ζ_0, ζ_1, and ζ_2 are three particular solutions (having no dependence on ζ') of the Ricatti equation (14) which contains the accelerations.

If the functions $\zeta_k(u)$ $(k = 0, 1, 2)$ are regarded as the fundamental quantities and are given, we can obtain the $g(u)$, $M(u)$, and $h(u)$ as above, as well as the triad components of the acceleration. With the definitions

$$\dot{\zeta}_k \equiv d\zeta_k/du, \qquad k = 0, 1, 2,$$
$$p_k \equiv (1 - \tfrac{1}{4}\zeta_k^2),$$
$$t_k \equiv i(1 + \tfrac{1}{4}\zeta_k^2),$$
$$D \equiv \begin{vmatrix} p_0 & \zeta_0 & t_0 \\ p_1 & \zeta_1 & t_1 \\ p_2 & \zeta_2 & t_2 \end{vmatrix},$$

we have

$$\ddot{\xi}^{\mu}\lambda_{1\mu} = D^{-1}\begin{vmatrix} \dot{\zeta}_0 & \zeta_0 & t_0 \\ \dot{\zeta}_1 & \zeta_1 & t_1 \\ \dot{\zeta}_2 & \zeta_2 & t_2 \end{vmatrix},$$

$$\ddot{\xi}^{\mu}\lambda_{2\mu} = D^{-1}\begin{vmatrix} p_0 & \zeta_0 & \dot{\zeta}_0 \\ p_1 & \zeta_1 & \dot{\zeta}_1 \\ p_2 & \zeta_2 & \dot{\zeta}_2 \end{vmatrix},$$

$$\ddot{\xi}^{\mu}\lambda_{3\mu} = -D^{-1}\begin{vmatrix} p_0 & \dot{\zeta}_0 & t_0 \\ p_1 & \dot{\zeta}_1 & t_1 \\ p_2 & \dot{\zeta}_2 & t_2 \end{vmatrix}.$$

It should be pointed out at this point, that the transformation (1) can be explicitly written out in the following form[10]:

$$y^{\mu} = \xi^{\mu}(u) + a^{\mu}(\zeta, \bar{\zeta}, u)r,$$
$$a^{\mu} = \frac{b^{\mu}(\zeta, \bar{\zeta})}{\sqrt{2}\,P(\zeta, \bar{\zeta}, u)},$$
$$b^{\mu} = (1 + \tfrac{1}{4}\zeta\bar{\zeta}, 1 - \tfrac{1}{4}\zeta\bar{\zeta}, \tfrac{1}{2}(\zeta + \bar{\zeta}), (\zeta - \bar{\zeta})/2i),$$
$$\sqrt{2}\,P = [\xi^0 - \xi^1 + \tfrac{1}{4}(\xi^0 + \xi^1)\zeta\bar{\zeta} - \tfrac{1}{2}(\xi^2 - i\xi^3)\zeta - \tfrac{1}{2}(\xi^2 + i\xi^3)\bar{\zeta}],$$
$$\xi^{\mu} = (d/du)\xi^{\mu}(u).$$

The only allowed coordinate freedom on the resulting metric (12) [keeping the timelike line $y^{\mu} = \xi^{\mu}(u)$ unchanged] is the six-parameter transformation

$$\zeta = (a\zeta' + b)/(c\zeta' + d), \qquad ad - bc = 1,$$

which is isomorphic to the homogeneous Lorentz group.

The metric (12) may also be derived by other methods (e.g., by specializing the Robinson–Trautman[6] metrics or solving the equations obtained by Newman and Penrose,[4] specialized to flat-space time), although the geometric significance is not as clear as in the present derivation.

As a simple application of these coordinates, it is seen that the Lienard–Weichart potentials in electrodynamics assume a simple form in the new coordinate system. The vector potentials become

$$A_0 = (e/4\pi)(1/r - \dot{P}/P),$$
$$A_1 = (e/4\pi)(1/r),$$
$$A_i = 0,$$

where e is the charge. The electromagnetic field tensor is then

$$F_{01} = -e/4\pi r,$$
$$F_{0i} = -(e/4\pi)(\dot{P}/P)_{,i},$$

the remaining components being zero. Thus, the field tensor is split into a pure Coulomb field, and terms independent of r.

[9] L. Euler, Novye Comm. Acad. Petrop. VIII (1760–1761) [1763]. For a later reference, see Watson, *Theory of Bessel Functions* (Cambridge University Press, New York, 1952), Sec. 4.21.

[10] We are indebted to I. Robinson for having pointed out this simplification.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

Interaction *contra* Classical Relativistic Hamiltonian Particle Mechanics*

DOUGLAS G. CURRIE†‡
University of Rochester, Rochester, New York
(Received 5 April 1963)

The relativistic Hamiltonian formalism is outlined and discussed for classical particles. The implications of the requirement that the coordinates of an event transform according to the Lorentz transformation law are discussed and expressed in a form, called the world-line conditions, which may be considered in the relativistic Hamiltonian formalism. It is then shown that the world-line conditions imply that there is no interaction in the relativistic Hamiltonian formalism; that is, the motion of any pair of particles described by the relativistic Hamiltonian formalism consists of straight-line motion. In other words, if the events which compose the world lines of the particles transform according to the Lorentz transformation law, and the path of the particle is not a straight line, then this phenomena cannot be described in terms of a relativistic two-particle Hamiltonian formalism. The experimental basis for a determination of the transformation properties of an event is considered, and the relationship of the experiment to the applicability of such a formalism is discussed.

I. INTRODUCTION

In order to describe the dynamics of relativistic systems, Dirac[1] proposed an extension of the Hamiltonian formalism to make it compatible with special relativity. The postulates involved in the extension were made plausible by the consideration of commonly accepted physical ideas concerning relativity and relativistic invariance. Since the essential point is the consideration of the set of "fundamental quantities", which have a structure similar to that of the set of the generators of infinitesimal Lorentz transformations and are identified with these generators, this extension of the Hamiltonian formalism will be called the Relativistic Generator Formalism to distinguish it from the Hamiltonian formalism, which properly involves only time translation. This formalism, by its very structure, guarantees that the physical laws of the system will be relativistically invariant. The operator analog to the classical Relativistic Generator Formalism has been developed and is widely employed in quantum mechanics and quantum field theory. In order to facilitate a detailed investigation of the formalism, only the classical particle system discussed by Dirac in his original paper will be considered. The system will be further simplified by considering the mutual interaction of two particles (not the interaction of two particles and a field).

An intuitive outline of the paper will be presented in this introduction. To allow a lucid development of the general structure, some assumptions which are to be discussed in later sections will remain implicit. Briefly, the dynamics of the particles are described by ten "fundamental quantities" which are functions of the particle positions and the conjugate variables with respect to the Poisson bracket, which compose "phase space". These conjugate variables will be given the name "particle momenta", but no identification between these variables and the so-called "physical momentum" is required. The fundamental quantities have Poisson bracket relations which are isomorphic with the commutator bracket relations of the generators of infinitesimal Lorentz transformations. Thus a particular theory in the Relativistic Generator Formalism is defined by ten functions over phase space which are the fundamental quantities, or generators of infinitesimal transformations. The functions have the Poisson bracket relations

$$\begin{aligned} &[\mathfrak{P}_i, \mathfrak{P}_j] = 0, \quad [\mathfrak{J}_i, \mathfrak{P}_j] = \epsilon_{ijk}\mathfrak{P}_k, \quad [\mathfrak{J}_i, \mathfrak{J}_j] = \epsilon_{ijk}\mathfrak{J}_k, \\ &[\mathfrak{H}, \mathfrak{P}_i] = 0, \qquad [\mathfrak{H}, \mathfrak{J}_i] = 0, \\ &[\mathfrak{K}_i, \mathfrak{P}_j] = \delta_{ij}\mathfrak{H}, \qquad [\mathfrak{K}_i, \mathfrak{J}_j] = \epsilon_{ijk}\mathfrak{K}_k, \\ &[\mathfrak{K}_i, \mathfrak{H}] = \mathfrak{P}_i, \qquad [\mathfrak{K}_i, \mathfrak{K}_j] = -\epsilon_{ijk}\mathfrak{J}_k. \end{aligned} \tag{1}$$

In general, the change in a physical quantity ξ (represented by a function over phase space), due to an infinitesimal transformation of magnitude b is given by

$$\xi' = \xi + b[\xi, \mathfrak{B}] = \xi - b[\mathfrak{B}, \xi], \tag{2}$$

where the fundamental quantity which generates the transformation is denoted generically by $\mathfrak{B}$. In particular, the change in the position of the nth particle q_i^n is given by

* This paper is based, in part, on a thesis submitted to the University of Rochester in partial fulfillment of the requirements for the Ph.D. degree.
† National Science Foundation Cooperative Fellow.
‡ Present Address: Palmer Physical Laboratory, Princeton University, Princeton, New Jersey.
[1] P. A. M. Dirac, Rev. Mod. Phys. 21, 392 (1949).

$$q_i^{n\prime} = q_i^n + b[q_i^n, \mathfrak{B}]. \tag{3}$$

Let the fundamental quantity $\mathfrak{P}_j$, for example, generate an infinitesimal space translation of magnitude a_i parallel to the q_i axis. Then the transformed particle position is

$$q_i^{n\prime} = q_i^n + a_j[q_i^n, \mathfrak{P}_j], \tag{4}$$

with no summation over j. But a space translation of magnitude a_j in the j direction simply increases the component of the position by a_j and leaves the other components unchanged, as indicated by the equation

$$q_i^{n\prime} = q_i^n + a_j\delta_{ij}. \tag{5}$$

Comparing (4) to (5) yields the translation world-line condition

$$[q_i^n, \mathfrak{P}_j] = \delta_{ij}. \tag{6}$$

In the same manner, consider the rotation generated by $\mathfrak{J}_i$, which is a rotation about the q_i axis, whose magnitude is measured by θ_i. The change in the position of a particle due to the infinitesimal transformation generated by the fundamental quantity $\mathfrak{J}_i$ is

$$q_i^{n\prime} = q_i^n + \theta_j[q_i^n, \mathfrak{J}_j]. \tag{7}$$

On the other hand, the transformation which corresponds to $\mathfrak{J}_1$ is a rotation about the q_1 axis, and the change in the position due to a rotation through a small angle θ_1, about the q_1 axis, is

$$\begin{aligned} q_1^{n\prime} &= q_1^n, \\ q_2^{n\prime} &= q_2^n \cos\theta_1 - q_3^n \sin\theta_1 = q_2^n - \theta_1 q_3^n + \cdots, \\ q_3^{n\prime} &= q_3^n \cos\theta_1 + q_2^n \sin\theta_1 = q_3^n + \theta_1 q_2^n + \cdots. \end{aligned} \tag{8}$$

These equations, and the equations for the rotations about the q_2 and q_3 axis may be summarized by

$$q_i^{n\prime} = q_i^n + \theta_j\epsilon_{ijk}q_k^n. \tag{9}$$

Comparing (7) to (9) yields the rotation world-line condition

$$[q_i^n, \mathfrak{J}_j] = \epsilon_{ijk}q_k^n. \tag{10}$$

For the time translation generated by the fundamental quantity $\mathfrak{H}$, the transformed position is

$$q_i^{n\prime} = q_i^n + \tau[q_i^n, \mathfrak{H}]. \tag{11}$$

Since $\mathfrak{H}$ corresponds to a time translation, the transformed position is the position of the particle at the new time t' given by $t' = t - \tau$. Considering the position as a function of time, the event represented

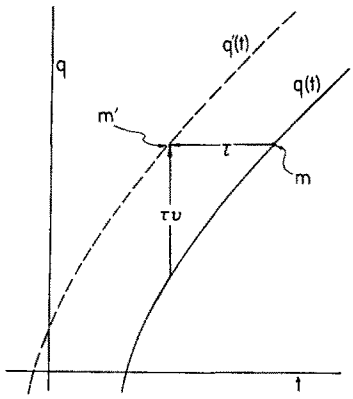


FIG. 1. Time translation.

by the new position function at a time t' is the same as the old position function at a time t, but expressing t' in terms of t and expanding the position function in powers of τ yields

$$q_i^n(t) = q_i^{n\prime}(t') = q_i^{n\prime}(t-\tau) = q_i^{n\prime}(t) - \tau\dot{q}_i^{n\prime}(t) + \cdots,$$

or, since $\tau\dot{q}_i^{n\prime} = \tau\dot{q}_i^n$ if terms of higher than second order in τ are dropped,

$$q_i^{n\prime}(t) = q_i^n(t) + \tau\dot{q}_i^n(t) = q_i^n(t) + \tau v_i^n(t), \tag{12}$$

where the symbols $v_i^n(t)$ and $\dot{q}_i^n(t)$ are defined to be $dq_i^n(t)/dt$. This derivation may be expressed in a more intuitive manner with the aid of Fig. 1. The transformation in question carries an event at time t (m for example) to an event (m') at time $t' = t - \tau$ with the same spatial coordinates (q). However, the change in position function on a time slice is desired, so $q_i^{n\prime}(t)$ is equal to $q_i^n(t) + \tau v_i^n(t)$. Comparing (11) to (12) yields the expression

$$[q_i^n, \mathfrak{H}] = \dot{q}_i^n = dq_i^n/dt = v_i^n. \tag{13}$$

The analysis for the Lorentz transformation is slightly more involved. The position in a frame which is moving with a velocity $\tanh(\alpha_j)$ is determined by the fundamental quantity $\mathfrak{K}_j$ to be

$$q_i^{n\prime} = q_i^n + \alpha_j[q_i^n, \mathfrak{K}_j]. \tag{14}$$

Due to the lack of invariance of simultaneity, the position of the particle in the new frame at time zero will not be the same event on the world line as the event at which the time is zero in the original frame. Denoting the time of the event m' by t', t' may be determined by the Lorentz transformation of the coordinates of an event to be $t' = t\cosh\alpha_j - q_j\sinh\alpha_j$. To simplify the expressions, we shall consider the change due to the Lorentz transformation of the position for time zero, rather than an arbitrary time. For a very small Lorentz trans-

formation, that is, $\alpha_i \ll 2\pi$, t' is given by $-q_i\alpha_i$. This is the value of the transformed time for any value of q_i, but since the value of the transformed time in the vicinity of the particle which is being considered is desired, the significant t' is $-q_i^n\alpha_i$. Thus the transformed position at time equal to zero may be written as the old position function of the transformed time t', or, given the transformed time from above, this implies

$$q_i^n(0) = q_i^{n'}(t') = q_i^{n'}(-q_i^n\alpha_i). \tag{15}$$

Expanding the function q_i in powers of α_i yields the expression

$$q_i^n(0) = q_i^{n'}(0) - q_i^n\alpha_i\dot{q}_i^{n'}(0) = q_i^{n'}(0) - \alpha_i q_i^n(0)v_i^n(0),$$

since v_i^n equals $v_i^{n'}$ due to dropping second-order terms in α_i; so

$$q_i^n(0) = q_i^{n'}(0) + \alpha_i q_i^n(0)v_i^n(0). \tag{16}$$

Comparing (14) to (16) yields the acceleration world-line condition, and using (13) to eliminate v_i^n yields

$$[q_i^n, \mathfrak{K}_i] = q_i^n v_i^n = q_i^n[q_i^n, \mathfrak{H}]. \tag{17}$$

The ten functions, denoted by $\mathfrak{H}$, $\mathfrak{P}$, $\mathfrak{K}$, and $\mathfrak{J}$, define a theory. These functions are not independent, as they must satisfy the Poisson bracket relations of (1) or the Lorentz condition. They must further satisfy the world-line conditions (6), (10), and (17). We might now ask whether there exist any set of functions which satisfy all these conditions. There is one obvious example, in which the functions are defined by

$$\begin{aligned} \mathfrak{P}_i &\equiv p_i^1 + p_i^2, & \mathfrak{J}_i &\equiv \epsilon_{ijk}(q_j^1 p_k^1 + q_j^2 p_k^2), \\ \mathfrak{H} &\equiv \omega^1 + \omega^2, & \mathfrak{K}_i &\equiv q_i^1\omega^1 + q_i^2\omega^2, \end{aligned} \tag{18}$$

where $\omega^n \equiv (p_i^n p_i^n + (m_n)^2)^{\frac{1}{2}}$. A simple calculation demonstrates that these fundamental quantities satisfy the Lorentz condition and the world-line conditions. The velocity is p_i^n/ω^n and the acceleration, defined as $[[q_i^n, \mathfrak{H}], \mathfrak{H}]$ vanishes. Therefore, the above-defined set of fundamental quantities is said to define a theory which describes free particles.

It will be shown in a later section that the most general form of the "kinematic" generators $\mathfrak{P}_i$ and $\mathfrak{J}_i$, consistent with the world-line conditions, is the standard form as given for the free theory.

Now let us consider whether there exists any set of acceptable functions for which the acceleration is not zero. As indicated by its definition, the acceleration is a well-defined function of the derivatives of the function $\mathfrak{H}$. But the most general form which the function $\mathfrak{H}$ may take is restricted by the world-line conditions and the Lorentz conditions, which are strong enough (as will be demonstrated in a later section) to imply that there is no set of functions which give nonzero accelerations. In the language used to describe the kinematic generators, the world-line conditions are strong enough to reduce the fundamental quantities $\mathfrak{H}$ and $\mathfrak{K}$ to the "standard form" given in the free-particle theory.

Before considering the usefulness of any formalism to describe nature, some discussion must be made concerning the phenomena which it is to describe. In other words, an idealized example of the experimental situation which the formalism is to describe will now be discussed.

Consider a classical scattering experiment conducted in a gravity-free laboratory. Small bodies (charged pith balls) are projected (by a spring-gun mechanism) toward a common point. They approach, scatter without contact, and recede from one another. Their positions at various instants are recorded by a camera taking many photographs. Only their positions are measured (not their momentum and energy). It is also assumed that measurements made during repetitions of the same experiment indicate that, for reasonable levels of illumination, the intensity and color of the light which illuminates the pith balls does not affect their motion to any observable extent.

An attempt might be made to describe the results of such an experiment by the consideration of a model of point particles interacting by a potential, or in the frame of the observer, by a Hamiltonian which would describe the interaction. Our present interest is whether such an experiment could be explained (the positions of the particles predicted) in terms of a model employing the Relativistic Generator Formalism for two particles. Consistent with the usual form of classical mechanical models, it shall be assumed that the effect of the measuring apparatus (the light illuminating the particles in the above model) may be reduced continuously to zero.

Before proceeding further, there is another experimental question to be considered—that of the transformation of the coordinates of an event determined by one observer to the coordinates determined by an observer moving with respect to the first observer. In order to be able to speak with more precision, a specific type of event will be considered, for example, a flash of light from a particle, or the light reflected from a particle. In this way, an accelerated "event" is an event which comes from an accelerated

particle. While this type of an event is not all inclusive, it will serve to establish some of the ideas, and is the type of event which is historically significant in the study of classical particle mechanics. At the basis of the consideration of this type of event is the assumption that the method of measurement (the emission of light or the reflection of light) does not affect the motion of the particles. The present experimental evidence indicates that the coordinates of a free event (the determination of the position of a free particle at a given time) transform according to the Lorentz transformation law—in particular, the motion of "free" photons and free mesons. On the other hand, there seems to be no evidence as to the transformation properties of an accelerated event. It is not known whether the coordinates of an accelerated "event" measured by two observers in relative motion can be correlated by the Lorentz transformation law using the relative velocity and position of the observers. For the main part of this paper, it will be assumed that when this experiment is performed, the result will be that the coordinates of the accelerated event transform according to the Lorentz transformation law. Other possible transformation laws as well as their consequences are discussed later.

Thomas[2] reaches a conclusion which appears to be the same as the conclusion proved in this paper, so a short discussion will be given to his proof. Consider two particles, interacting with each other, as indicated in Fig. 2. At the point δ, consider the possibility of the particle undergoing an external interaction which would change its trajectory from 2 to $\bar{2}$. The acceleration of particle 1 at the point α, as determined by the Relativistic Generator Formalism will be a function of the position and momentum of particle 1 at α, and the position and momentum of particle 2 at m. The acceleration of the first particle at the event α is now determined in a second, moving frame whose coordinate axes are indicated by the dashed lines. In this frame, if the external interaction does not take place, the acceleration is determined by the position and momentum of particle 2 at m', while if the external interaction does take place, the acceleration of particle 1 at α is determined by the position and momentum of particle 2 at $\bar{m}'$. Thus the acceleration of particle 1 at α depends upon whether or not the external interaction takes place. On the other hand, in the original frame, the acceleration of particle 1 at α did not depend upon whether the external

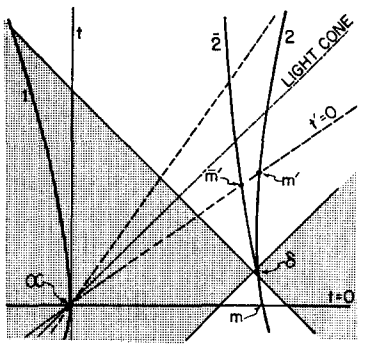


FIG. 2. Effect of casual interference.

interaction took place. From this apparent contradiction, Thomas concludes that noninvariant world lines must be employed. On the other hand, the very fact that an external interaction may take place at δ implies that at this point one cannot describe the motion with only a two-particle formalism, but some account of the external agency must be made. For this reason, at the point δ it is the two-particle formalism which is insufficient, rather than the assumed transformation properties. In the same manner, in order to describe the interaction of two particles in terms of their positions and momentum at one instant, neither particle must be acted upon by an external agency. Since the instant is not invariant, every point on the second world line which may be made simultaneous with the point at which the external agency acts on the first particle, lies in a region in which the two-particle formalism will be insufficient. This region is the set of all points which are spacelike with respect to the region in which the external agency acts. The region in which the external agency acts will be called the interference region (the point δ in our example), and the points which are spacelike with respect to the interference region will be called the shadow of the interference. Some of these points have been discussed by Havas[3] in connection with the study of equation of motion of particles. The result of Thomas may now be stated more precisely: the two-particle Relativistic Generator Formalism, and the usual transformation properties of events are not compatible in the shadow of the interference. Since the two-particle Relativistic Generator Formalism no longer is applicable due to the external interference, no statement can be made as to the transformation properties in this region based upon such an argument. In the present paper, however, we deal with the possible use of the two-particle relativistic

[2] L. H. Thomas, Phys. Rev. 85, 868 (1952).

[3] P. Havas and J. Plebanski, Bull. Am. Phys. Soc. 5, 433 (1961).

Hamiltonian formalism outside the shadow of the interference (i.e., where the particles are affected only by each other, and not by some external agency). In other words, what pairs of world lines may be described by the Relativistic Generator Formalism? This question is being asked before considerations of causality, etc. are made. If the formalism can describe some motion, then other questions may be entertained.[4]

At this point, two issues might well be re-emphasized. Implicit in the use of the Hamiltonian formalism is the assumption that the initial data at one instant (position and momenta in the present case) determine, through the Hamiltonian, all higher derivatives at that one instant. From this set of derivatives, the entire history and future of the system may be constructed. In the conventional terminology, this might be described as an "instantaneous propagation of the interaction," but such an appellation is misleading, since it implies that external interactions will be considered to determine the speed of propagation of a disturbance. The following discussion and proofs do not consider or make use of the more powerful restrictions which might be implied by the consideration of such external interactions and the requirement of a finite velocity of propagation for such a disturbance. Thus the question at issue might be formulated: Can the motion of two particles be described by a two-particle Hamiltonian scheme, i.e., *as if* the interaction were propagated in an instantaneous manner?

In this paper, no arguments have been presented to extend these results from statements concerning the interaction of a pair of particles to statements concerning the interaction of two particles and a field (i.e., particles with the interaction "transmitted" by the field). In particular, the use of a dynamic field to transmit the interaction would imply the specification of the initial conditions of the field, as well as the particle variables. Such a situation requires a modified formulation, although there may well be analogous results.

II. RELATIVISTIC GENERATOR FORMALISM

The extension of the Hamiltonian formalism to apply to relativistic systems proposed by Dirac,[1] will be defined in a formal, semiaxiomatic manner. Intuitively, this formalism guarantees the relativistic invariance of physical laws by expressing the laws in terms of a set of ten "fundamental quantities" whose Poisson bracket relations have the same structure as the Lie algebra of the Lorentz group.[5] These fundamental quantities are assumed to exist, and are identified with the generators of infinitesimal Lorentz transformations. Thus the change in a dynamical quantity (represented by a function over phase space) in going from one frame to another frame is assumed to be given by the Poisson bracket of the dynamical quantity with the corresponding fundamental quantity. It is also assumed that the change of a dynamical quantity with time in a given frame is determined by the Poisson bracket of the quantity with the fundamental quantity H. The concept of guaranteeing relativistic invariance by the association of the laws with the representations of the Lorentz group was proposed by Wigner,[7] although he developed this concept only for free-particle systems, or equivalently, to the motion of the center of mass of a system (an irreducible system). Thomas[8] also discussed but did not develop some of these ideas for general systems from more of a geometric point of view. An explicit extension of this formalism to the internal coordinates of a complex system (a reducible system), which shall be called the Relativistic Generator Formalism, was introduced by Dirac, who presented the postulates upon which the extension rests and the physical reasoning which argues that such a set of postulates provides an extension which is the correct method of guaranteeing the relativistic invariance of physical laws. Thomas[2] discusses the question of the compatibility of the extension proposed by Dirac and the traditional Lorentz transformation law for the coordinates of an event in special relativity. His position is that if these two sets of ideas are not consistent for the description of interacting systems, the Relativistic Generator Formalism, rather than the geometric aspects of special relativity, is to be retained. Thomas[2] and later Bakamjian and Thomas[9] developed the formalism for classical systems, and obtained sets of nontrivial fundamental quantities which satisfy the Lorentz conditions. Recently, Foldy[10] has developed further the Relativistic Generator Formalism with a discussion of its application

[4] See Appendix.

[5] A more detailed discussion and justification of the Relativistic Generator Formalism may be found in reference 6. This work considers the geometric properties within the context of the relativistic generator formalism, and so, while leading to considerations of transformations in the formalisms, it does not leave one in the position of being able to judge the value of the formalism itself.

[6] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).

[7] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[8] L. H. Thomas, Rev. Mod. Phys. **17**, 182 (1945).

[9] B. Bakamjian and L. H. Thomas, Phys. Rev. **92**, 1300 (1953).

[10] L. L. Foldy, Phys. Rev. **122**, 275 (1961).

to quantum mechanics. The general philosophy of the Relativistic Generator Formalism in classical mechanics, and its relationship to relativistic invariance are well illustrated and discussed by Sudarshan.[11,12] Specific examples of the use of this formalism in its conventional sense to obtain an interaction theory may be found in the work of Sudarshan,[12] Foldy,[10] and Bakamjian and Thomas.[9]

Dirac, in his presentation of the extension of the Hamiltonian Formalism, discusses two logically distinct structures. The first consists of an axiom system which comprises the Relativistic Generator Formalism. The second part of Dirac's discussion consists in establishing a correspondence between the elements of the axiom system (like the fundamental quantities) and physical or empirical concepts (like the transformation of measurements performed in different frames). It is by such a correspondence that deductions made within the axiom system may be related to experimental results. The two parts shall be considered separately to facilitate the separation of mathematical and physical aspects.

Formally, a classical, relativistic, two-particle theory in the Relativistic Generator Formalism is defined by a set of ten real functions (called fundamental quantities) of the twelve variables q_i^1, q_i^2 and p_i^1, p_i^2, ($i = 1$, 2, or 3), which are called the individual variables. These functions satisfy the Poisson bracket relations of (1), where the Poisson bracket of two arbitrary functions A and B of the individual variables is defined by

$$[A, B] \equiv \sum_{i,n} \frac{\partial A}{\partial q_i^n}\frac{\partial B}{\partial p_i^n} - \frac{\partial A}{\partial p_i^n}\frac{\partial B}{\partial q_i^n}. \tag{19}$$

The fundamental quantities are assumed to be differentiable functions of the individual variables.

The product of two fundamental quantities, or a fundamental quantity and one of the variables, is the ordinary product of functions. Thus the definition of the Poisson bracket implies that $[A, BC] = B[A, C] + [A, B]C$, so the Poisson bracket defines a derivation.

This defines the formal structure of the Relativistic Generator Formalism. Any set of functions satisfying the Lorentz conditions defines a theory in the formalism. Some relationships will now be developed within the formal structure. Defining the exponential function by

$$e^{a[\ ,A]}B \equiv B + a[B, A] + \tfrac{1}{2}a^2[[B, A]A] + \cdots \equiv \exp(a[\ , A])B \equiv e^{-a[A,\ ]}B, \tag{20}$$

the function f_a is defined by $\exp(a[\ , A])f$, where A represents one of the fundamental quantities. Thus f_a consists of a one-parameter family of functions, labeled by the parameter a. The parameters τ, a_i, θ_i, and α_i will be associated with the fundamental quantities $\mathfrak{H}$, $\mathfrak{P}_i$, $\mathfrak{J}_i$, $\mathfrak{K}_i$, respectively. The derivative of f with respect to a may be defined as follows and, from the properties of the Poisson bracket, equals

$$\partial f/\partial a \equiv \lim_{a\to 0} 1/a(f_a - f) = [f, A]. \tag{21}$$

More explicitly, let us write this out fully for the case of t and $\mathfrak{H}$:

$$\begin{aligned}\frac{\partial f}{\partial t} = [f, \mathfrak{H}] &= \frac{\partial f}{\partial q_i^1}[q_i^1, \mathfrak{H}] + \frac{\partial f}{\partial q_i^2}[q_i^2, \mathfrak{H}] \\ &\quad + \frac{\partial f}{\partial p_i^1}[p_i^1, \mathfrak{H}] + \frac{\partial f}{\partial p_i^2}[p_i^2, \mathfrak{H}] \\ &= \dot{q}_i^1\frac{\partial f}{\partial q_i^1} + \dot{q}_i^2\frac{\partial f}{\partial q_i^2} + \dot{p}_i^1\frac{\partial f}{\partial p_i^1} + \dot{p}_i^2\frac{\partial f}{\partial p_i^1}.\end{aligned} \tag{22}$$

In order to simplify later equations, a new set of variables, called the collective variables, will be defined by

$$\begin{aligned} Q_i &\equiv \tfrac{1}{2}(q_i^1 + q_i^2), \\ q_i &\equiv \tfrac{1}{2}(q_i^1 - q_i^2) = \tfrac{1}{2}(-1)^{\bar{n}}(q_i^n - q_i^{\bar{n}}), \\ P_1 &\equiv p_i^1 + p_i^2, \\ p_i &\equiv p_i^1 - p_i^2 = (-1)^{\bar{n}}(p_i^n - p_i^{\bar{n}}). \end{aligned} \tag{23}$$

These equations may be solved to obtain the inverse equations

$$q_i^n = Q_i + (-1)^{\bar{n}}q_i, \qquad p_i^n = \tfrac{1}{2}P_i + \tfrac{1}{2}(-1)^{\bar{n}}p_i. \tag{24}$$

The individual variables, in terms of which the collective variables are defined, are a canonical set with respect to the Poisson bracket of (19). From this it may easily be shown that the collective variables form a canonical set. Since the collective variables are canonical, the Poisson bracket defined in terms of the collective variables is identical to the Poisson bracket defined in terms of the individual variables.

For reasons which will appear later, we shall be interested in restricted transformations, which are defined as transformations of the variables which do not affect the positions. Working with the collective variables, the most general restricted transformation has the form

$$\begin{aligned} \bar{q}_i &= q_i, \qquad \bar{p}_i = p_i + f_i(q_i, p_i, Q_i, P_i), \\ \bar{Q}_i &= Q_i, \qquad \bar{P}_i = P_i + F_i(q_i, p_i, Q_i, P_i). \end{aligned} \tag{25}$$

[11] E. C. G. Sudarshan, *1961 Brandeis University Summer Institute, Lectures in Theoretical Physics* (W. A. Benjamin Inc., New York, 1962), Vol. 1.

[12] E. C. G. Sudarshan, "Hamiltonian Dynamics of Relativistic Particles" NYO 9680 (1961) (unpublished); and "Principles of Classical Mechanics" NYO 10250 (1963) (unpublished).

If the restricted transformation is required to be canonical, that is, $[\bar{q}_i, \bar{p}_j] = \delta_{ij}$ and $[\bar{Q}_i, \bar{P}_j] = \delta_{ij}$ with all other Poisson brackets of the variables vanishing, certain restrictions upon the functions f_i and F_i are implied. Thus the nonvanishing Poisson brackets imply

$$\partial f_i/\partial p_j = 0, \qquad \partial F_i/\partial P_j = 0, \tag{26}$$

respectively, for all values of i and j. This in turn implies that f_i is independent of p_j and F_i is independent of P_j. In the same manner, the vanishing of $[\bar{q}_i, \bar{P}_j]$ and $[\bar{Q}_i, \bar{p}_j]$ implies that F_i is independent of p_i and f_i independent of P_i, so both f_i and F_i are functions of the positions only. The vanishing of $[\bar{p}_i, \bar{p}_j]$ and $[\bar{P}_i, \bar{P}_j]$ imply that the curl of f_i with respect to q_i, and the curl of F_i with respect to Q_j vanish, which implies, by the use of Stokes theorem, that there exist functions $\bar{f}$ and $\bar{F}$, the gradients of which are f_i and F_i, that is,

$$f_i = \partial\bar{f}/\partial q_i = [\bar{f}, p_i] = -[p_i, \bar{f}],$$
$$F_i = \partial\bar{F}/\partial Q_i = [\bar{F}, P_i] = -[P_i, \bar{F}]. \tag{27}$$

Finally, the vanishing of $[\bar{p}_i, \bar{P}_j]$ implies, by the Jacobi identity,

$$\partial/\partial q_i\, \partial/\partial Q_i(\bar{f} - \bar{F}) = 0, \tag{28}$$

so that $\bar{f}(q_i, Q_i) = \bar{F}(q_i, Q_i) + \sigma(q_i) + \tau(Q_i)$. Since only the derivative of $\bar{f}$ with respect to q_i ever appears in the transformation equations for the variables, the function $\tau(Q_i)$ does not affect the transformation of the variables, and may be dropped. The most general restricted canonical transformation is thus defined by the two functions $\bar{F}(q_i, Q_i)$ and $\sigma(q_i)$, where $\bar{f} = \bar{F} + \sigma$. On the other hand, every such set of functions defines a restricted canonical transformation of the variables, such that

$$\bar{P}_i = P_i + F_i = P_i + [\bar{F}, P_i] = P_i + \partial\bar{F}/\partial Q_i,$$
$$\bar{p}_i = p_i + f_i = p_i + [\bar{f}, p_i] = p_i + [\bar{F}, p_i] \tag{29}$$
$$+ [\sigma, p_i] = p_i + \partial\bar{F}/\partial q_i + \partial\sigma/\partial q_i.$$

Several theories in the Relativistic Generator Formalism will now be considered. The first in the "free theory" which was defined in the previous section, for which both the acceleration and the change in momentum vanished. The vanishing of either could be used as a criteria to specify that a set of fundamental quantities describes a free theory. Although the latter criteria is perhaps more conventional, in this paper the former shall be used since it is the position and its changes with respect to time which are being measured. Thus a particle will be considered free as long as the world line of the particle is straight in four-space. Although the distinction between the two criteria is inconsequential for this theory, the next theory to be considered has vanishing acceleration, without a vanishing change in momentum.

This theory is defined by a set of generators which are the same as for the free theory, except that

$$\omega^n \to \omega^n_\lambda \equiv (\{p^n_i + (-1)^{\bar{n}} q_i\lambda(q_k q_k)\}\{p^n_i + (-1)^{\bar{n}} q_i\lambda\} + (m_n)^2)^{\frac{1}{2}} \equiv (p^n_{\lambda i} p^n_{\lambda i} + m_n^2)^{\frac{1}{2}}. \tag{30}$$

For this theory, a calculation with the Poisson brackets gives, the velocity to be $p^n_{\lambda i}/\omega^n_\lambda$ and the acceleration vanishes. The time rate of change of the momentum is $(\delta_{ij} + q_i q_j\, \partial\lambda/\partial q^2)v^n_i$. Thus in terms of the change of momentum the motion is not free, since the momentum changes with time. On the other hand, the velocity is constant, so the world lines of the particles are straight lines. Since we are considering the position as the quantity which is measured, and are not attempting to measure the momentum, our attention is directed to the straight world lines, and the motion will be called free. This theory also obeys the world-line conditions mentioned in the introduction.

Up to now in this section, the formal part of the Relativistic Generator Formalism has been defined, discussed, and developed. However, in order that the formal structure be useful in physics, a relationship between the structure and empirical concepts and operations must be established. One part of the correspondence has been implicitly introduced already, the quantity q^n is identified as the position of the nth particle measured in the manner which will be discussed in the next section. As previously mentioned, the extra restriction of identifying the p^n with some other observable, called the momentum of the individual particles, will not be attempted. The primary identification is the identification of the fundamental quantities with the generators of infinitesimal Lorentz transformations. As thus stated, the identification may be intuitively reasonable, but it lacks mathematical precision, so the idea of an infinitesimal Lorentz transformation will be developed more precisely and then a more precisely defined identification will be made.

Dirac in his article is rather free with identifications. In addition to the mentioned identifications, the fundamental quantities are identified with such quantities as the total momentum and the total energy, and the p^n with the individual particle momentum. If the latter identifications are made (with the assumption that the energy–momentum trans-

forms as a four-vector) in addition to the previously mentioned ones, it can quickly be shown that the only theory that is admissible is the free theory.[13] If we postulate that some quantity is both the generator of translation, and the value of some measurable quantity (momentum), then the postulate of our logical system implies some experimental knowledge, i.e., the postulate may not be made "in all possible worlds". It is logically possible that the axiom system is not only wrong when compared to nature, but lies beyond the domain of verifiability and meaning. Thus to ensure a meaningful development, the identifications, if they be multiple, should first be made singly to see if this is valid, and then an attempt be made to combine them. As a side note, however, it should be noted that the interest in a momentum identification arises in quantum theory, where it is common to prepare and measure states which are essentially eigenstates of momentum (in and out scattering states). The eigenstates of position (upon which translation and rotation would be defined) are not of much practical use so it is the second identification which assumes greater importance, the first identification serving mainly to justify the introduction of the Poisson bracket relations of the fundamental quantities.

The correspondence of Dirac may now be precisely stated as the identification of the parameter a_i with the spatial separation of space-translated frames, τ with the relative time translation, θ_i with the relative rotation about the q_i axis, and α_i with the tangent of the relative velocity of the frames. Dirac presents physical reasons why such an identification is reasonable.

III. GEOMETRIC ASPECTS OF DYNAMICS

In the next section, a formal structure for the geometric aspects of measurement will be developed. In order to clarify what this represents, and to distinguish between definition, postulate, and empirical concepts, some of the ideas of position and time measurement and descriptions in other frames will be discussed. The object of this discussion is to demonstrate the relationship of the formal structure, which will be presented in the next section, to dynamics and the actual procedure used to determine the position, as a function of time, of an artificial satellite, or one of the moons of Jupiter.

The method of measurement to be used will be called the optical method, and the position and time so determined, the optical position and optical time, or the optical coordinates. The primary instrument used to determine the optical position is the theodolite, an instrument which measures the angle of incidence of an incoming ray, or pulse of light with respect to a frame of reference defined by the base of the physical instrument. The theodolite is also equipped with a phototube which denotes the instant at which a pulse of light arrives. Some type of clock is used to determine from the phototube signal the time of arrival of the pulse. An observation station consists of two theodolites mounted on the ends of a metal rod which is taken as the standard of length. Thus if a flash of light is observed, the angle of incidence at both theodolites is measured as well as the time of arrival at each theodolite. From this set of six numbers (two angles and a time of arrival determined at each theodolite), four numbers are selected: both angles from the first theodolite, and one from the second, and the time of arrival at the first theodolite. These four numbers are called the initial optical data, and characterize an event. Questions as to whether they are the "true" time of arrival, or angles are meaningless.

From the set of four numbers, a function of the initial optical data is defined which is called the optical position. The form of the function is that which would be obtained if it were assumed that light travels in straight lines, and space is Euclidian. This formula, using the initial optical data, may be employed to give a set of three numbers called the optical position of the event with no assumptions necessary concerning the nature of space and light. The optical position will have its familiar properties, however, only if light propagates in straight lines in a Euclidian space. In the same manner the time, or the fourth coordinate, of the event is defined as the time of arrival of the pulse at the first theodolite, corrected by an interval calculated as if light traveled at a uniform velocity in a straight line from the optical position of the event to the first theodolite. Thus from the four numbers of the initial optical data, four numbers, called the optical coordinates of the event, are calculated by the given formula.

A second observation station, composed in the same manner as the previous one, is located at a different point in space. The question of the equality of the rods and the equivalence of the clocks is resolved by measuring the position of the ends of the rod of the new station from the old station. If the separation, so determined, is constant and equal to the length assigned to the rod of the original station, the rods will be said to be equal. Clock synchronization is achieved with light signals and using the convention that the propagation of light occurs

[13] See Appendix of reference 6.

with the same velocity in each direction. A second, later synchronization assures that the rates are equal. The detailed considerations of measurement theory have been rather crude, since a proper consideration would be much more lengthy; and this is not intended as an investigation of geometry. It is hoped that the points skipped here will not confuse the later developments.

In discussing the formulas which were used to determine the optical coordinates of the event from the initial optical data, it was illuminating to assume the convention that the space was flat and light followed a null geodesic in order to "derive" the formula which defines the coordinates in terms of the initial optical data. With two stations, this assumption becomes more than conventional, for the assumption implies that, given the relative separation and orientation of the observation stations, and the optical coordinates determined in one frame, the optical coordinates in the second frame may be calculated. The agreement of the measured vs calculated values for the optical coordinates determined by the second observer is required to maintain the assumption of the Euclidian nature of space and geodesic propagation of light. Thus these properties become testable premises for two stations.

Let us now return to the question of the transformation properties of free vs accelerated events. The part of this question which interests us takes the experimental form of being able to find points (pair of events) which, to one observer, appear coincident and simultaneous, while to a second observation station, they are no longer coincident. If the observed events have this complicated property, the analysis is much complicated, so we shall make the previously mentioned assumption that this is not the experimental situation. In later terminology this is equivalent to assuming that the strong world-line condition holds. It is unfortunate that, in the interest of brevity, this assumption must be made, since as will soon be seen, if interaction is described by the Relativistic Generator Formalism, then the experimental situation must be of this strange type. The remainder of the paper will show that this is the case, and a general discussion of the situation will be made at the end.

Now proceeding with the assumption that events which are simultaneous and coincident in one frame are so in all frames, where this property is called the strong world-line condition, the geometry of space and the propagation of light may be considered. The convention that light follows a null geodesic will be made, so the empirical question is the geometry of space.[14] The determination of the optical coordinates of events by the two observation stations may now be compared to decide whether or not these positions may be related by the Euclidian transformation, using the relative position and orientation of the stations determined by optical observation. If the positions are related by the Euclidian transformation, then the structure of space as determined by the optical measurements is compatible with the assumption of Euclidian geometry. By the observation of many events by many observers, we obtain the experimental result that the space containing the events which are measured by optical measurements, can, to a certain level of accuracy, be assigned a Euclidian structure.

Since all the measurements by different stations have, up to now, been by stations which are at rest with respect to one another, the geometry of space has been considered, but not the geometry of space–time. To complete the investigation of the geometry of space–time, the data obtained by observation stations which are moving with respect to one another must be used. These stations observe one another, and are said to be equivalent if the rest lengths of the standard bars are the same. The rest length of a bar is defined by the usual formula result which gives the rest length in terms of initial optical data in Lorentzian space–time. Thus it has the usual meaning in Lorentz space–time, and is well defined regardless of the geometry of space–time. The relative velocity, position, and orientation of the frames may also be determined. They may then observe other events and determine the geometry of four-space, that is, the geometry which is compatible with the coordinates of events as determined by the optical measurements. It might for example be Lorentzian, Galiliean, or near a heavy mass, Schwarzschildian. We shall assume that the experiments will indicate that the geometry so determined in our space–time region of interest is Lorentzian. If the geometry as determined by free events is Lorentzian, then the assumption of the strong world-line condition may be rephrased in the following form; let the events which originate from free particles determine the Lorentzian geometry, and let the events from accelerated particles determine some other geometry. The strong world-line condition now says that the geometry determined by the accelerated particles is also Lorentzian.

Having explained what will be meant by the "position" of an event, and by the assumption that

[14] H. Reichenbach, *The Philosophy of Space and Time* (Dover Publications, Inc., New York, 1957).

space is Lorentzian, we shall now explain what will be meant by a particle, or more properly, a particle explanation of observations of events, in a nonquantum theory.

In general, the concept of a particle is a construct by which certain kinds of observations may be conveniently explained. The type of observations which suggest a particle for their explanation consists of a set of events (which might be flashes, or "reflections") which are all timelike with respect to one another. By timelike, is meant that the time order given to the flashes by all observation stations is the same. If, rather than assuming the strong world-line condition, space had been assumed Lorentzian only with respect to events which originate on unaccelerated bodies, then for accelerated events there might be no time-ordering properties. We now consider directly the transformation properties of events originated by accelerated particles (that is, members of sets of events which are timelike with respect to each other in the present frame and not on a straight line). If the coordinates of the events which compose a curved world line transform according to the Lorentz transformation laws, that is, in the same manner as the coordinates of the events upon a straight world lines, then the strong world-line condition is said to be satisfied for these events. This implies that the transformation properties of an event are independent of the adjacent events on the world line (whether the segment of the world line is straight or curved), or of the forces acting upon the particle which originates the event. If this is true for all world lines, that is, for all events which are to be considered in the theory, then the extended strong world-line condition is satisfied, which means that the coordinates of every event, regardless of the world line it is on, transform in the same manner as the free events. If the transformation properties of the coordinates of an event are the same for identical world lines, that is, the transformation properties of the event are implied by the world line to which it belongs, then the set of events which compose the world line is said to obey the weak world-line condition. Intuitively, this means that if two different kinds of forces produce the same world line, then the coordinate transformation of the world-line events does not depend upon the type of force. If this is true for all trajectories, then the extended weak world-line condition is satisfied, which intuitively means that, if two force fields produce the same trajectories, they cannot be distinguished by the transformation properties of the coordinates of events. It should be noted that the idea of dynamics has now been extended. Originally, dynamics was to determine or describe the motion of particles in one frame. If we limit ourselves to this point of view, then the world line conditions have nothing to do with dynamics. On the other hand, if we wish to have a theory which also explains the transformation properties as one goes from frame to frame, then this extension may prove to be very involved, especially if the experimental situation should be such that the weak world-line condition is not satisfied. Then two different extended theories may have the same dynamics (predict the same trajectories), but predict different transformation properties. The extended weak world-line condition implies that if two types of dynamics produce the same motion, they are indistinguishable. The satisfaction of the strong world-line condition implies that the transformation properties of coordinates of an event do not depend upon the motion of the source of the event. For the bulk of the paper, the strong world-line condition will be assumed. The possibility of it not being maintained is discussed in the final sections.

In a given frame, we may have a position function, which is the position expressed as a function of time. A given infinite set of events will determine a position function. In a different frame, these events will determine a different position function. The next section will consider relationships among the generators of infinitesimal Lorentz transformations, and between the generators and the position functions. In this section, the intention has been to exhibit the motivation behind the mathematical assumptions which are used in the next section. The use of the phrase "world-line condition" will henceforth imply the extended strong world-line condition.

IV. GEOMETRIC FORMALISM

This section consists of the derivation of commutation relations among the generators of infinitesimal Lorentz transformations in a realization of a degree four, and relations between these generators and certain subsets of the realization space. The reasons for an interest in such a mathematical structure, as well as its relation to physical theories, are discussed in the previous and following sections. In this section, the intention is to state all hypotheses and to make a semblance of a proper mathematical proof of the results.

Let M represent a four-dimensional vector space over the real numbers. In a given basis, denoted by X_μ ($\mu = 0, 1, 2, 3$) or t and X_i ($i = 1, 2, 3$), t will

be called the time, and X_i the position of the point, and the time and position collectively will be called the coordinates of the point. An interval is defined by the "metric" tensor: $g^{00} = -g^{11} = -g^{22} = -g^{33} = 1$.

Consider a set of transformations upon M defined by $X'_\mu = \Lambda'_\mu X_\nu + a_\mu$ where the a_μ are real numbers, and Λ'_μ is a real matrix which obeys the relations (i) $\Lambda^{\mu\nu}\Lambda'_\mu = g^{\nu\sigma}$, (ii) $\Lambda^0_0 > 0$, and (iii) Det $\Lambda'_\mu = 1$. The set of all such transformations will be denoted by P, and called the Poincaré set of transformations (or the inhomogeneous Lorentz set of transformations). By conventional methods, it can be shown that the Poincaré set is a ten-parameter, connected Lie group, with a realization[15] of degree four in the space M. For the present purposes, the Poincaré group may most easily be studied by considering separately several sets of one-parameter subgroups.

The first of these is the time-translation subgroup, defined by $X'_i = X_i$, $t' = t - \tau$. An obvious calculation shows that time translation, for any value of τ, is an element of the Poincaré set, and that τ is a canonical parameter for the Poincaré group. The same is true for the transformations and parameters to be introduced in the next paragraph. Now let us express this transformation in the form[16]

$$X'_i = e^{-\tau\mathcal{H}}X_i, \qquad t' = e^{-\tau\mathcal{H}}t. \tag{31}$$

For small values of τ, these equations take the form

$$X'_i = X_i - \tau\mathcal{H}X_i = X_i, \quad t' = t - \tau\mathcal{H}t = t - \tau, \tag{32}$$

where $\mathcal{H}$ is the generator of an infinitesimal translation.[17] By comparing the equations with the ones obtained from the definition of the transformation, the result of the infinitesimal generator acting upon the various coordinates may be determined to be $\mathcal{H}X_i = 0$ and $\mathcal{H}t = 1$.

In the same manner, a space translation (st) in the j direction of magnitude a_j, a space rotation (sr) about the j axis of magnitude θ_j, and a Lorentz rotation (Lr) in the $X_j - t$ plane of magnitude α_j, for small values of the parameters, are defined by

$$\begin{aligned} X_i^{(\mathrm{st})} &= X_i + \delta_{ij}a_j, & t^{(\mathrm{st})} &= t, \\ X_i^{(\mathrm{sr})} &= X_i + \theta_j\epsilon_{ijk}X_k, & t^{(\mathrm{sr})} &= t, \\ X_i^{(\mathrm{Lr})} &= X_i - \delta_{ij}\alpha_j t, & t^{(\mathrm{Lr})} &= t - \alpha_j X_j. \end{aligned} \tag{33}$$

Denoting the generators of infinitesimal space translations, space rotations, and Lorentz rotation by $\mathcal{P}_i$, $\mathcal{J}_i$, $\mathcal{K}_i$, respectively, and comparing with (33), the result of the generators of infinitesimal transformations acting upon the coordinates is $\mathcal{P}_iX_j = -\delta_{ij}$, $\mathcal{P}_it = 0$, $\mathcal{J}_iX_j = \epsilon_{ijk}X_k$, $\mathcal{J}_it = 0$, $\mathcal{K}_iX_j = \delta_{ij}t$, and $\mathcal{K}_it = X_i$. The generators may also be expressed as derivatives with respect to the coordinates in M space, having the form

$$\mathcal{P}_i = -\partial/\partial X_i, \qquad \mathcal{H} = \partial/\partial t, \tag{34}$$

$$\begin{aligned} \mathcal{J}_i &= \epsilon_{ijk}X_k\,\partial/\partial X_j = \epsilon_{ilm}X_l\mathcal{P}_m, \\ \mathcal{K}_i &= t\,\partial/\partial X_i + X_i\,\partial/\partial t = X_i\mathcal{H} - t\mathcal{P}_i. \end{aligned} \tag{35}$$

Commutation relations between these generators may be calculated to be

$$\begin{aligned} [\mathcal{P}_i, \mathcal{P}_j]_c &\equiv \mathcal{P}_i\mathcal{P}_j - \mathcal{P}_j\mathcal{P}_i = 0, \\ [\mathcal{H}, \mathcal{P}_i]_c &= 0, \qquad [\mathcal{H}, \mathcal{J}_i]_c = 0, \end{aligned} \tag{36}$$

$$[\mathcal{P}_i, \mathcal{J}_j]_c = \epsilon_{ijk}\mathcal{P}_k, \tag{37}$$

$$\begin{aligned} [\mathcal{J}_i, \mathcal{J}_j]_c &= \epsilon_{ijk}\mathcal{J}_k, \qquad [\mathcal{J}_i, \mathcal{K}_j]_c = \epsilon_{ijk}\mathcal{K}_k, \\ [\mathcal{P}_i, \mathcal{K}_j]_c &= -\delta_{ij}\mathcal{H}, \end{aligned} \tag{38}$$

$$[\mathcal{K}_i, \mathcal{H}]_c = \mathcal{P}_i, \qquad [\mathcal{K}_i, \mathcal{K}_j]_c = -\epsilon_{ijk}\mathcal{J}_k.$$

For example,

$$[\mathcal{P}_i, \mathcal{P}_j]_c \equiv \mathcal{P}_i\mathcal{P}_j - \mathcal{P}_j\mathcal{P}_i = \frac{\partial}{\partial X_i}\frac{\partial}{\partial X_j} - \frac{\partial}{\partial X_j}\frac{\partial}{\partial X_i} = 0, \tag{39}$$

and

$$\begin{aligned} [\mathcal{P}_i, \mathcal{K}_j]_c &= \mathcal{P}_i(X_j\mathcal{H} - t\mathcal{P}_j) - (X_j\mathcal{H} - t\mathcal{P}_j)\mathcal{P}_i \\ &= -\delta_{ij}\mathcal{H} + X_j\mathcal{P}_i\mathcal{H} - t\mathcal{P}_i\mathcal{P}_j - X_j\mathcal{H}\mathcal{P}_i + t\mathcal{P}_j\mathcal{P}_i \\ &= -\delta_{ij}\mathcal{H}. \end{aligned} \tag{40}$$

Let us now consider a set of points of M. An element of the Poincaré set transforms each point to a new point, and thus transforms such a set into a new set. A set of points of M will be called a world line, and denoted w if it has the following properties:

(i) the square of the interval defined by the "metric" tensor between any two elements of w is positive;

(ii) the set w is closed and there is a point between any pair of points (i.e., w is homeomorphic to real line);

(iii) the set w is unbounded (i.e., it extends infinitely forward and back in time).

The set of all w is denoted W.

Due to the above properties, the set of points of M which comprise a world line may be specified by three continuous functions whose slopes are less than unity. Each of these functions specify one of

[15] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1931), p. 14.

[16] The sign convention used here differs from the more common usage (see reference 17), but in order to maintain the usual sign convention for (i) Dirac's canonical transformation, (ii) Lorentz conditions, and (iii) the same sign for $B\varphi$ as $[\mathfrak{B}, \varphi]$, this convention is necessary.

[17] P. M. Cohn, *Lie Groups* (Cambridge University Press, Cambridge, England, 1957), p. 66.

the three "spacelike" coordinates of the world line as a function of the "timelike" coordinate. The set of three functions will be called the position function (χ), and each individual function will be called a component of the position function [χ_i or $\chi_i(t)$]. To every position function (set of three functions with slope less than unity) there is an element of W. The set of all analytic position functions will be denoted $\mathfrak{X}$.

An element of the Poincaré group maps the coordinates of a point of M into a new set of coordinates. This may be considered to be either the transformation to a new frame of reference while we consider the "same" point, or the mapping of each point of M into a new point. The latter interpretation shall be used, since it facilitates the discussion of the ideas involved. Thus an element of the Poincaré group maps a point of M into a new point of M. In the same manner, an element of the Poincare group will map a set of points in M to a new set of points. The new set of points into which a world line w is mapped will be denoted w'. Since the transformations of the Poincaré group are continuous and do not change the value of the interval between two points, the set w' satisfies the conditions required of a world line if these conditions were satisfied by w. Thus the set of world lines W is closed under the Poincaré set of transformations.

Let the position function associated with the world line w' be denoted χ'. Since an element of the Poincaré group transforms w to w', by the above association, the same element of the Poincaré set may be said to transform χ to χ'. Thus the result of an element of the Poincaré group acting upon the set of position functions may be defined. By the above procedures, the following equation acquires a well-defined meaning:

$$\chi' = e^{-b\mathfrak{G}}\chi = \chi - b\mathfrak{G}\chi + \tfrac{1}{2}b^2\mathfrak{G}\mathfrak{G}\chi - \cdots, \tag{41}$$

where $\mathfrak{G}$ represents one of the generators of infinitesimal Lorentz transformations, and b its parameter. For a small value of the parameter, the new function will not differ greatly from the old position function, so the new position function may be written as the sum of the old position function, and another function, in the form $\chi_i'(\ ) = \chi_i(\ ) + b\eta_i(\ )$. By comparison, the definition of η_i is $-\mathfrak{G}\chi_i$.

The next object is to calculate $\mathfrak{G}\chi_i(\ )$ for the various particular generators. The operator $e^{-b\mathfrak{G}}$, acting upon a point m of M, yields the new point m' of M. If the point m lies upon a world line with position function $\chi_i(\ )$, then the coordinates of the point are given by $t = t_0$, $X_i = \chi_i(t_0)$. The coordinates of m' are given by $t' = e^{-b\mathfrak{G}}$, $t_0 = t_0 - b\mathfrak{G}t_0 \equiv t_0 - b\bar{\mathfrak{G}}$, and $X_i' = X_i - b\mathfrak{G}X_i \equiv X_i - b\hat{\mathfrak{G}}_i$, where the functions $\hat{\mathfrak{G}}$ and $\bar{\mathfrak{G}}_i$ are defined by these relations. On the other hand, w' is a world line, and m' is a point of this world line. If $\chi_i'(\ )$ is the position function of w', then the position of m' is $X_i' = \chi_i'(t')$. In order to determine the function $\chi_i'(\ )$ expressed in terms of the function $\chi_i(\ )$, the expressions for t' and X_i', in terms of t_0, $\chi_i(t_0)$, are substituted into the equation for X_i', yielding

$$\begin{aligned} X_i' &= X_i - b\bar{\mathfrak{G}}_i(\chi_i(t_0), t_0) = \chi_i(t_0) - b\bar{\mathfrak{G}}_i = \chi_i'(t') \\ &= \chi_i'(t_0 - b\hat{\mathfrak{G}}(\chi_i(t_0), t_0)) = \chi_i'(t_0) - b\dot{\chi}_i(t_0)\hat{\mathfrak{G}}. \end{aligned} \tag{42}$$

Expressing $\chi_i'(\ )$ as $\chi_i + b\eta_i$ as discussed above, and dropping terms of second order or higher in b (that is, considering a very small transformation),

$$\chi_i(t_0) - b\bar{\mathfrak{G}}_i = \chi_i(t_0) + b\eta_i(t_0) - b\dot{\chi}_i(t_0)\hat{\mathfrak{G}} + \cdots, \tag{43}$$

$$\eta_i(t_0) = -\bar{\mathfrak{G}}(\chi_i(t_0), t_0) + \dot{\chi}_i(t_0)\hat{\mathfrak{G}}(\chi_i(t_0), t_0). \tag{44}$$

Since the value of t_0 was arbitrary (no special properties of its numerical value were used), this equation is true for all values of t_0; thus

$$\mathfrak{G}\chi_i(t) = -\eta_i(t) = \bar{\mathfrak{G}}_i(\chi_i(t), t) - \dot{\chi}_i(t)\hat{\mathfrak{G}}(\chi_i(t), t), \tag{45}$$

$$\mathfrak{G}\chi_i(t) = \mathfrak{G}X_i - \dot{\chi}_i(t)\mathfrak{G}t. \tag{46}$$

The change in the position function due to the infinitesimal Lorentz transformation may now be determined,[18]

$$\begin{aligned} \mathcal{P}_i\chi_i &= \mathcal{P}_iX_i - \dot{\chi}\mathcal{P}_it = -\delta_{ij} - \dot{\chi}0 = -\delta_{ij}, \\ \mathcal{H}\chi_i &= \mathcal{H}X_i - \dot{\chi}_i\mathcal{H}t = 0 - \dot{\chi}_i1 = -\dot{\chi}_i, \\ \mathcal{J}_i\chi_i &= \epsilon_{ijk}X_k = \epsilon_{ijk}\chi_k, \\ \mathcal{K}_i\chi_i &= \delta_{ij}t - \dot{\chi}_iX_j = \delta_{ij}t - \dot{\chi}_i\chi_j \\ &= \chi_i\mathcal{H}\chi_i - t\mathcal{P}_i\chi_i = (\chi_i\mathcal{H} - t\mathcal{P}_i)\chi_i. \end{aligned} \tag{47}$$

Thus the position function in a Lorentz-transformed frame can be obtained by the expression of the Lorentz generator $\mathcal{K}_i$ in terms of $\mathcal{H}$ and $\mathcal{P}_i$, which appeared for the transformation of the coordinates of a point in the new frame. In other words, the decomposition of the Lorentz transformation on a point into a space translation and time translation can also be achieved in a comparatively simple manner for the position functions.

The usual geometric results of special relativity can be quickly derived in the above formalism, as, for instance, the Lorentz contraction of a rod (composed of uniformly moving particles) in a moving

[18] M. H. Pryce [Proc. Roy. Soc. (London) **A195**, 621 (1948)] has obtained results which are essentially these equations, but he uses them only for the motion of the center of mass.

frame, and the relativistic composition of velocities for a uniformly moving particle. Since it is not limited to uniformly moving particles however, these results are easily generalized to determine the position function of a particle in a transformed frame.

If a description at one instant is desired (as will occur in the Hamiltonian formalism), it is convenient to choose this instant to be $t = 0$, and then the above equations become

$$\mathcal{P}_i\chi_i = -\delta_{ii}, \quad \mathcal{J}_i\chi_i = \epsilon_{iik}\chi_k, \quad \mathcal{K}_i\chi_i = \chi_i\mathcal{H}\chi_i, \quad (48)$$

which are called the geometric world-line conditions.

V. DYNAMIC POSITION

The Relativistic Generator Formalism has been defined, and some theorems within this formal structure have been proved in Sec. II. Another formal structure was defined and developed in the previous section, and it was intuitively identified with empirical concepts and operations in Sec. III. In order that the results derived in the Relativistic Generator Formalism may be related to the results of experiments, some explicit connection between the formalism and empirical concepts must be made. This may be done through a correspondence established between the Relativistic Generator Formalism and the geometric formalism, and an identification between the geometric formalism and empirical concepts. Part of the correspondence is explicitly mentioned in Dirac's paper, that is, functions (fundamental quantities) are assumed to exist which correspond to the generators of infinitesimal Lorentz transformations, in that they generate the change which functions defined on phase space undergo in going to a transformed frame. To be consistent with this picture, these fundamental quantities are required to have Poisson bracket relations isomorphic to the commutation relations of the geometric generators.

In a classical particle theory, however, the change of certain functions (the particle positions) must have a well-defined form if the particle position is to be identified with the series of events which are observed as the trajectory of the particle. In the geometric formalism, the position function (or the geometric position function) was a function of time which gave the position of the particle at each instant. For the correspondence, it is assumed that there is a function defined over phase space (the dynamic position function) which has the value of the position of the nth particle for every state of the system represented by a point in phase space (pure state). If no such function can be defined over phase space which gives the position of each particle for any pure state of the classical particle system, then the Relativistic Generator Formalism would seem to be a rather inadequate method of description for classical particle dynamics.

If the dynamic position function ($\mathfrak{x}$) is to be identified with the position of the particle, there are certain relationships it must satisfy in order to justify such an identification. In the previous section these relationships for the geometric formalism were developed. Thus from the requirement that in a space-translated frame, the change in the position function equals the change in position of the origin of the frame of reference, the geometric translation world-line condition $\mathcal{P}_i\chi_i = -\delta_{ii}$ was obtained. Since $\mathfrak{P}_i$ is to generate a space translation, and $\mathfrak{x}$ is to represent the particle position in the Relativistic Generator Formalism, the dynamic world-line condition is $[\mathfrak{P}_i, \mathfrak{x}_j] = -\delta_{ij}$. If the function $\mathfrak{P}_i$ is to transform the position function for each particle to the form which it would have in the translated frame, then $[\mathfrak{x}_i^n, \mathfrak{P}_j] = \delta_{ij}$ must be satisfied. This is the minimum requirement that $\mathfrak{P}_i$ transform all functions of phase space to the form which they would have in the transformed frame.

In the same manner, if $\mathfrak{J}_i$ is to transform all functions to the form in which they would have a rotated frame, then since the transformation of a point under rotation is assumed known, the geometric rotation world-line condition is expressed by $\mathcal{J}_i\chi_j^n = \epsilon_{ijk}\chi_k^n$ and the dynamic rotation world-line condition is $[\mathfrak{J}_i, \mathfrak{x}_j^n] = \epsilon_{ijk}\mathfrak{x}_k$. For the Lorentz transformation if the transformation properties of an event are known to be according to the Lorentz transformation law, then by the derivation of Sec. IV, the geometric acceleration world-line condition is $\mathcal{K}_i\chi_j^n = \chi_i^n\mathcal{H}\chi_j^n - t\mathcal{P}_i\chi_j^n$, and the dynamic acceleration world-line condition is thus $[\mathfrak{K}_i, \mathfrak{x}_j^n] = \mathfrak{x}_j^n[\mathfrak{H}, \mathfrak{x}_j^n] - t[\mathfrak{P}_i, \mathfrak{x}_j^n]$.

Let us consider the role of the t in this equation. As mentioned in the previous section, it is normally set equal to zero. The value of the parameter actually indicates the time separation between the time about which the Lorentz rotation takes place, and the time at the point on the geometric world line, just as X_i^n is the spatial distance between the position about which the Lorentz rotation takes place and the position of the point on the world line. Thus, in setting t to zero, invariance is not being required for all possible Lorentz transformations, but only those which "rotate" about the instant for which the function of phase space gives the values of the dynamical quantities. A seemingly more restrictive form of relativistic invariance could be obtained by

requiring that the acceleration world-line condition be true for all values of t, but this appears to give essentially the same results, since the position at the later time in the Relativistic Generator Formalism is determined by $e^{-\tau[\mathfrak{P},\,]}\,\mathfrak{x}$.

Since the functions are combined by ordinary multiplication, the geometric position function of the various particles obviously commute, thus implying the usual vanishing of $[\mathfrak{x}_i^n,\ \mathfrak{x}_j^m]$. It is then plausible to assume that a canonical transformation may be made on q_i^n (in terms of which the Poisson bracket and the fundamental quantities are defined) to $\mathfrak{x}_i^n$ so that the Poisson brackets and generators are defined in terms of $\mathfrak{x}_i^n$. Changing the notation (denoting this $\mathfrak{x}_i^n$ by q_i^n), the dynamic world-line conditions now have the form

$$[q_i^n,\ \mathfrak{P}_j] = \delta_{ij}, \qquad [q_i^n,\ \mathfrak{I}_j] = \epsilon_{ijk}q_k^n, \qquad [q_i^n,\ \mathfrak{K}_j] = q_i^n[q_i^n,\ \mathfrak{H}], \tag{49}$$

where the fundamental quantities are functions of q_i^n and p_i^n, which are a canonical set.

VI. KINEMATIC FUNDAMENTAL QUANTITIES

The fundamental quantities $\mathfrak{P}_i$ and $\mathfrak{I}_i$ are often considered on a slightly different footing than $\mathfrak{H}$ and $\mathfrak{K}_i$. It is conventionally assumed that these fundamental quantities have a "standard form" given by $\mathfrak{P}_i = p_i^1 + p_i^2 = P_i$ and $\mathfrak{I}_i = \epsilon_{ijk}(q_j^1p_k^1 + q_j^2p_k^2)$. The reasoning which justifies such an assumption is based upon the usual identification of these fundamental quantities with the linear and angular momentum and, due to the assumed conservation of these quantities for all acceptable interactions, their form should not be dependent upon the form of the interaction. Thus they should have the above "free" form, and are referred to as the kinematic fundamental quantities. The assumption that they have the standard form enters the formal structure as a postulate, motivated by the above considerations, and the interpretation of these quantities as the linear and angular momentum.

Since the world-line conditions are intended to express, at least in part, the role of interpretation in the dynamics, one might hope to prove from the world-line conditions that the kinematic fundamental quantities have the standard form, rather than requiring a separate postulate. This will now be done.

As previously discussed, we shall be concerned only with the position, velocity, and acceleration, so that the change brought about in the individual particle momentum by a restricted canonical transformation is not considered to be physically significant since it does not affect any predictions of the theory. Thus it will be shown that the restrictions upon the kinematic fundamental quantities expressed by the Lorentz relations and the world-line conditions imply that there exists a restricted canonical transformation which brings the fundamental quantities to the standard form.

The most general form for $\mathfrak{P}_i$ may be expressed as $\mathfrak{P}_i = p_i^1 + p_i^2 + h_i(q_i^1, q_i^2, p_i^1, p_i^2)$ where h_i are three arbitrary differentiable functions over phase space. The world-line conditions $[q_i^1,\ \mathfrak{P}_j] = \delta_{ij}$, since p_i^n is canonical related to q_i^n, imply that $\partial\mathfrak{P}_i/\partial p_j^n = \delta_{ij}$ for n equal to 1 or 2. This in turn implies that h_i may be written as a function of q_i^1 and q_i^2 or q_i and Q_i, independent of p_i^1 and p_i^2.

The Lorentz condition $[\mathfrak{P}_i,\ \mathfrak{P}_j] = 0$ implies that the curl of $\mathfrak{P}_i$ with respect to Q_i vanishes, which in turn implies that there exists a function $\tilde{h}$ such that

$$h_i = \partial\tilde{h}/\partial Q_i = [\tilde{h}, P_i]. \tag{50}$$

A restricted canonical transformation may now be defined (see Sec. II) by $F \equiv h$, $\sigma \equiv 0$, and as shown in Sec. II, the new variables defined by this transformation are

$$\bar{P}_i = P_i + [\tilde{h}, P_i] = P_i + h_i, \qquad \bar{p}_i = p_i + [\tilde{h}, p_i]. \tag{51}$$

But referring to the most general form for the fundamental quantity $\mathfrak{P}_i$, we see that $\mathfrak{P}_i = p_i^1 + p_i^2 + h_i = P_i + h_i = \bar{P}_i = \bar{p}_i^1 + \bar{p}_i^2$. Thus in terms of the new variables, $\mathfrak{P}_i$ is in standard form.

In the same manner, the most general form for $\mathfrak{I}_i$ is $\mathfrak{I}_i = \epsilon_{ijk}(q_j^1p_k^1 + q_j^2p_k^2) + \rho_i(q_i^1, q_i^1, p_i^2, p_i^2)$. The world-line conditions for $\mathfrak{I}_i$ imply that ρ_i are independent of the individual particle momenta. The Lorentz condition $[\mathfrak{P}_i,\ \mathfrak{I}_j] = \epsilon_{ijk}\mathfrak{P}_k$ implies that ρ_i are independent of the mean position, that is, they are functions of the relative position only. The Lorentz condition $[\mathfrak{I}_i,\ \mathfrak{I}_j] = \epsilon_{ijk}\mathfrak{I}_k$ implies, in vector notation, $(\mathbf{q}\cdot\nabla)\boldsymbol{\rho} = \nabla(\mathbf{q}\cdot\boldsymbol{\rho})$. By the use of Lemma 2 of Keller,[19] there exists a function $\mathfrak{R}(q)$ such that

$$\rho_i = \epsilon_{ijk}q_j(\partial\mathfrak{R}/\partial q_k) = \epsilon_{ijk}q_j[\mathfrak{R}, p_k]. \tag{52}$$

A restricted canonical transformation defined by $f = \mathfrak{R}$, $\sigma = 0$ now brings $\mathfrak{I}_i$ to standard form, and leaves $\mathfrak{P}_i$ in standard form.[20,21]

---

[19] J. B. Keller, Commun. Pure Appl. Math. 14, 77 (1961).

[20] D. G. Currie, Thesis, University of Rochester (unpublished) contains a more detailed form of this calculation, as well as the calculation of the following section.

[21] These vector operations of dot and scalar products consider the quantities as three-dimensional vectors, not as twelve-dimensional vectors.

VII. THE NO-INTERACTION THEOREM

This section consists of the proof of a mathematical theorem, and a short discussion of the extension of the theorem. The following statement will be proved: Given a set of differentiable fundamental quantities which obey the Lorentz conditions and the world-line conditions, the acceleration vanishes.

Let us define the subset $\mathcal{O}$ of the phase space as the set of points for which $\epsilon_{ijk}q_ip_j^1p_k^2$ vanishes. If a function, differentiable in all variables, has a certain form for all points in phase space except in $\mathcal{O}$, then it seems reasonable to assume that its extension into $\mathcal{O}$ is unique. However, at a later point a more general class of functions will be considered so notice will be taken when a step in the proof is true only outside of $\mathcal{O}$. As mentioned, the fundamental quantities are, for this proof, the differentiable functions from twelve real variables to the real line.

In the previous section, it was shown that the world line conditions involving $\mathfrak{P}_i$ and $\mathfrak{J}_i$ imply that $\mathfrak{P}_i$ and $\mathfrak{J}_i$ may be transformed to the standard form. It will now be shown that the third world-line condition implies that $\mathfrak{H}$ and $\mathfrak{K}_i$ may be transformed to standard (i.e., free) form. The Lorentz condition

$$[\mathfrak{H}, \mathfrak{P}_i] = [\mathfrak{H}, P_i] = \partial\mathfrak{H}/\partial Q_i = 0 \tag{53}$$

implies that the function $\mathfrak{H}$ is independent of the mean coordinate Q_i. By taking the Poisson bracket of the third world-line condition with the positions, and using the Jacobi identity, the relations

$$q_i[q_j^1, [q_k^2, \mathfrak{H}]] = 0, \qquad [q_i^1[q_j^2, \mathfrak{K}_k]] = 0 \tag{54}$$

may be derived, which imply that $\mathfrak{H}$ and $\mathfrak{K}_i$ may be written as

$$\begin{aligned} \mathfrak{H} &= h^1(p_i^1, q_i) + h^2(p_i^2, q_i), \\ \mathfrak{K}_i &= k_i^1(p_i^1, q_i, Q_i) + k_i^2(p_i^1, q_i, Q_i). \end{aligned} \tag{55}$$

The deduction for $\mathfrak{H}$ is true only if some component of q_i is nonzero, that is, $\mathfrak{H}$ may be written in this form only outside of $\mathcal{O}$. Using the third world-line condition directly, k_i^n equals $q_i^nh^n$. The following symbols are defined by

$$\rho^n \equiv p_i^np_i^n, \quad \xi^n \equiv p_i^nq, \quad \kappa \equiv q_iq_i, \quad r \equiv p_i^{\bar{n}}p_i^{\bar{n}}. \tag{56}$$

Repeated use of the Lorentz conditions involving $\mathfrak{J}_i$, and the reduction of $\mathfrak{J}_i$ to standard form, shows the functions $\mathfrak{H}$ and $\mathfrak{K}_i$ may be written as a function of the above symbols in the form

$$\begin{aligned} \mathfrak{H} &= \beta^1(\rho^1, \xi^1, \kappa) + \beta^2(\rho^2, \xi^2, \kappa), \\ \mathfrak{K}_i &= q_i^1\beta^1 + q_i^2\beta^2. \end{aligned} \tag{57}$$

By using the above form for the generators, $[\mathfrak{K}_i, \mathfrak{H}] = \mathfrak{P}_i$ has the form

$$\begin{aligned} &p_i^1\partial/\partial\rho^1((\beta^1)^2 - \rho^1) + p_i^2\partial/\partial\rho^2((\beta^2)^2 - \rho^2) \\ &+ \tfrac{1}{2}q_i\{\partial(\beta^1)^2/\partial\xi^1 + \partial(\beta^2)^2/\partial\xi^2 + 4[\beta^1, \beta^2]\} = 0. \end{aligned} \tag{58}$$

Defining the vectors κ_i^1 and κ_i^2 as the "vector cross product" of p_i^2 and p_i^1 with q_i, respectively, and taking the "dot product" of κ_i^n with the above equation, we have

$$\begin{aligned} \partial/\partial\rho^n((\beta^n)^2 - \rho^n) &= 0, \\ \partial(\beta^1)^2/\partial\xi^1 + \partial(\beta^2)^2/\partial\xi^2 + 4[\beta^1, \beta^2] &= 0, \end{aligned} \tag{59}$$

for the part of phase space outside of $\mathcal{O}$. The first two equations imply that β^n may be written in the form $\beta^n = (\rho^n + w^n)^{\frac{1}{2}}$ where w^n is an arbitrary function of ξ^n and κ. The third equation may be written in the form

$$r\left(\frac{\partial\beta^1}{\partial\xi^1}\frac{\partial\beta^2}{\partial\rho^2} + \frac{\partial\beta^1}{\partial\rho^1}\frac{\partial\beta^2}{\partial\xi^2}\right) = \tfrac{1}{4}f(\xi^1, \xi^2, \rho^1, \rho^2, \kappa). \tag{60}$$

The derivative with respect to r implies[22] that the coefficient of r must vanish, and the coefficient in terms of ρ^n and w^n may be expressed as

$$\partial w^1/\partial\xi^1 = -\partial w^2/\partial\xi^2. \tag{61}$$

Defining λ as the derivative of w^1 with respect to ξ^1, it may easily be demonstrated from (61) that λ is independent of p^1 and p^2, and w^n has the form $w^n = (-1)^{\bar{n}}\lambda(\kappa) + \tau^n(\kappa)$. Using this form for w^n and substituting into (58), τ^n has the form $\tau^n = \frac{1}{4}\kappa(\lambda)^2 + (m_n)^2$ so that β^n may be written as

$$\beta^n = ((p_i^n + \tfrac{1}{2}(-1)^{\bar{n}}q_i\lambda)(p_i^n + \tfrac{1}{2}(-1)^{\bar{n}}q_i\lambda) + (m_n)^2)^{\frac{1}{2}}. \tag{62}$$

A direct calculation shows that the Poisson bracket of β^1 with β^2 vanishes. The acceleration may be calculated directly and it too vanishes. Thus the theorem is proved.

A restricted canonical transformation, which does not affect the vanishing of the acceleration, brings the generators to the free form.

In order to extend the theorem to functions more general than the class of differentiable functions, two difficulties are encountered. The first is a method of defining the Poisson bracket for two distributions, since this will, formally at least, involve the product of distributions, which is not well defined in the conventional treatments. The second difficulty is that the theorem cannot be "continued" into the region $\mathcal{O}$ for the more general class of functions, so a more comprehensive form of the proof must be used.

[22] The variable r is independent of ρ^1, ρ^2, ξ^1, ξ^2, as long as the point in question lies outside of $\mathcal{O}$.

Some preliminary calculations indicate that such an extension will be possible. For example, consider an $\mathfrak{H}$ of the form $\mathfrak{H} = \mathfrak{H}_0 + \mathfrak{V}(q)$ where $\mathfrak{V}(q)$ is any function of q, including delta functions and their derivatives (which gives rise to contact interactions); then the proof can be extended to show that such a $\mathfrak{V}$ cannot produce acceleration or scattering. That is, in any set of fundamental quantities of which $\mathfrak{H}$ of this form is a member, the Lorentz conditions, and the world-line conditions imply that $\mathfrak{V}(q)$ vanishes. This particular form of $\mathfrak{H}$ is sufficiently restricted so the difficulties of products of delta functions are escaped.

VIII. DISCUSSION AND SUMMARY

It has been shown that if the coordinates of the points which compose the world line of a particle transform according to the Lorentz transformation law, then the Relativistic Generator Formalism formulated in terms of this position can only describe the motion of free particles.

Historically, in order to "explain" the astronomical (optical) observation of the planets and satellites, a certain set of concepts has proved useful and gradually been developed. The model which was successfully used to explain the data employed the abstraction of point particles moving in a Euclidian space in which light (by which the position measurements were made) traveled in straight lines. In this model, the position of the particle was the optical position previously discussed and if a pair of events appeared coincident and simultaneous, i.e. to both occur at the same space–time point, they would appear coincident and simultaneous to all uniformly moving observers. This is the strong world-line condition. Thus a series of events composing a world line could be transformed from one frame to another by transforming the coordinates of each event without a knowledge of the "structure" of the events, or the forces on the particle which was the source of the events.

A law of motion for a pair of particles consisted of an equation expressing the acceleration of one of the particles in terms of the position and velocity of the two particles at that instant. The optical observations of planets and satellites could be used to confirm or deny a hypothesized equation of motion. With regard to the transformation properties of the coordinates of an event, it was assumed that if the coordinates of the event were measured in one frame, then given the relative position and velocity of a second frame, one could calculate the coordinates which an observer in the second frame would measure. That is, the transformation properties of the event were assumed to depend only upon the coordinates of the events and the relationship between the frames. A law of motion was said to be invariant if, in one frame, a pair of particles obeyed the law, then this motion (this pair of world lines) viewed from another frame would also satisfy the law.

This view developed with the belief that sufficiently accurate experiments would reveal that the transformation of the coordinates of an event were governed by the Galileian transformation. Within this framework, the law of motion was expressed either in terms of a force field, or the Hamiltonian formalism. Since it is now believed that when these experiments are finally performed, the transformation of the coordinates will be governed by the Lorentz transformation law, the problem is to express the laws of motion in a form which is invariant under the Lorentz transformation rather than the Galiliean transformation. Since the use of a force field did not seem practicable, Dirac attempted to extend the Hamiltonian formalism to a form which would be compatible with the Lorentz transformation. The discussion by which Dirac justified his method of extension is quite persuasive, and has led to the view that this extension must be the proper foundation upon which to construct dynamics. This view is expressed by Thomas who, when he feels that the transformation properties of the events composing world lines are in conflict with the extension of the Hamiltonian formalism, retains the formalism and introduces "noninvariant" world lines. In this line of thought and in opposition to the view taken in this paper, in a joint paper with Jordan and Sudarshan[5] the view is taken that the Dirac extension is the correct method of describing classical interaction, with a discussion of relativistic invariance. Having set the groundwork in this manner, the proof of Sec. VII is there discussed, and from that viewpoint one reaches the conclusion that there is something strange about the world-line conditions and that some modification of the traditional ideas of measurement must be made.

In order to maintain the Newtonian viewpoint,[23] the extended strong world-line condition must be valid, and it cannot hold if we have interaction de-

[23] See J. L. Synge, *Relativity: the special theory* (North-Holland Publishing Company, Amsterdam, 1958) for a discussion of a similar form of the Newtonian viewpoint, especially p. 6: "the history of a particle is a continuous sequence of events" which transforms according to the Lorentz transformation law.

scribed by the Relativistic Generator Formalism. The difference is not a formal one, but an experimental one. Present data does not seem to bear on this question, since the data appears to concern only free particles, and "free" photons! Thus the possibility of "strange" transformations cannot be ruled out. The possibility of "strange" transformation properties for classical particles was mentioned by Thomas,[2] and discussed by Sudarshan.[12]

In order to clarify the distinction between the two views, an experiment is proposed, the results of which would distinguish which viewpoint correctly describes the empirical world. Two observation stations are moving with an appreciable velocity with respect to each other. The secondary observer triggers, simultaneously in his frame, four flashbulbs, one placed at his origin, and one at a unit a distance in the x, y, and z directions. After one second, by his clock, he triggers another flashbulb at his origin. From the optical coordinates of these flashes, the primary observer can determine the relative position, orientation, velocity, and units of distance and time that the secondary observer is using. The secondary observer also has the following device. A flashbulb is mounted upon the rim of a rapidly rotating wheel, so that the flashbulb is moving in a circle, and undergoes an acceleration of constant magnitude and rotating direction. The second observer then triggers the flashbulb and determines the position and time (the coordinates) of the flash. The primary observer, using his measurements and the previously determined relationship between the frames, may calculate, by the Lorentz transformation, the Newtonian prediction of the coordinates determined by the secondary observer. If the predictions are indeed what the secondary observer measured, then the Lorentz transformation law for an accelerated event is verified for the event whose source is the accelerated particle (flashbulb). Therefore a two-particle Relativistic Generator Formalism of particle interaction seems untenable unless other experiments indicate that the weak world-line condition is not satisfied. On the other hand, if the prediction does not agree with the measurements of the secondary observer, then the strong world-line conditions, and the Newtonian view are not tenable.

Since the use of the Relativistic Generator Formalism for particles requires a rather drastic change in the intuitive ideas of space and time (the relaxation of the strong world-line condition allows the constructions of intuitively discomforting paradoxes, mainly due to the nonpreservation of coincidence and simultaneity of events), it would seem that the most reasonable approach would be to first attempt an explanation of the interaction of particles in some manner which does less violence to the intuitive basis of measurement. This in turn implies that the most productive approach is the reexamination of the extension made by Dirac to determine at what point the seemingly innocent assumptions introduce a powerful, but hidden additional postulate which implies that the conventional ideas of measurement are no longer acceptable. Upon proper examination, this postulate may be found to be without sufficient justification, and a more reasonable postulate at this point might allow a formalism similar to the Relativistic Generator Formalism which is compatible with the current ideas of measurement and position in classical physics. Thus while the use of the Relativistic Generator Formalism for particles is not, at present, in conflict with experimental findings, it does not have an obvious direct connection with the empirical world (since the strong world-line condition and Lorentz transformation law for all events may not be used), and it seems to require the alteration of some of the groundwork of classical physics. For this reason, some formalism which preserves the usual connection with the empirical world (allows the use of the world-line condition) would seem worth developing.

Both the Relativistic Generator Formalism and the Newtonian particle and measurement concepts might be compatible if extra, unobserved variables, in the form of a field are introduced.[1] The proof would then not apply, which seems to be related to the fact that due to the radiation in the system, the particle coordinates are not sufficient for a complete specification of the final state of the system. On the other hand, the only example of an interaction by classical fields is that of Dirac's classical electron,[24] interacting by a massless vector field, and the extension to other fields by Havas.[25] These models violate casuality, and a predeterminism much stronger than the determinism of LaPlace is implied.

In summary, it has been shown that in the description of the interaction of two classical particles, the Relativistic Generator Formalism is not compatible with conventional ideas of world lines in space–time, and it is further suggested on an intuitive basis that the formalism, rather than the conventional ideas of world lines and particles, may be in need of reexamination.

---

[24] P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1939).

[25] P. Havas, Phys. Rev. **87**, 305 (1952).

ACKNOWLEDGMENT

The guidance and encouragement of E. C. G. Sudarshan, many helpful discussions with T. F. Jordan, and a discussion of the manuscript with A. S. Wightman are all gratefully acknowledged.

APPENDIX

The logical structure of the general considerations of this paper may be illustrated more clearly by the consideration of a one-dimensional model. As in three dimensions, arguments are first presented which suggest, from general physical considerations, that the Relativistic Generator Formalism should be a useful method of describing the interaction of relativistic classical particles. Thus the formal structure defined by three functions which have the Poisson bracket relations

$$[\mathfrak{H}, \mathfrak{P}] = 0, \qquad [\mathfrak{K}, \mathfrak{P}] = \mathfrak{H}, \qquad [\mathfrak{K}, \mathfrak{H}] = \mathfrak{P}$$

shall be considered.

To relate the formalism to the concept of the position for the classical particles (and coordinate of an event on its world line), geometric considerations imply that the above fundamental quantities should obey the world line conditions

$$[q^n, \mathfrak{P}] = 1, \qquad [q^n, \mathfrak{K}] = q^n[q^n, \mathfrak{H}].$$

At this point, general physical arguments might be raised, such as: The Hamiltonian formulation implies instantaneous propagation of signals, and is therefore unsuitable for particle interaction, or; Since fields are fundamental, they must be included, and any formulation which does not include them is doomed. Such objections are out of place at this point since, given the above, the present task is to find the mathematical implications of the formalism. These arguments might imply that work on the above formulation would be wasted, but we shall be satisfied with attempting to develop this formalism. If what appear to be proper descriptions can finally be found, then these objections may be considered as to the usefulness of the representations.

The mathematical question to be considered is whether there exists any set of three functions which obey these conditions (Lorentz and world-line conditions) and have nonzero acceleration. In contrast to the three-dimensional case, there are such functions in one dimension, so some of the logical relationships may be exhibited more clearly. The generators defined by

$$\mathfrak{P} \equiv p^1 + p^2, \qquad \mathfrak{H} = p^1\eta - p^2\eta = p\eta,$$
$$\mathfrak{K} = (q^1p^1 - q^2p^2)\eta = (qP - Qp)\eta,$$

where $\eta \equiv (1 + e/(q)^2)^{\frac{1}{2}}$, obey the Lorentz conditions and the world-line conditions. The acceleration is given by $(-1)^n e/(q)^3$, which is obviously not zero. The equation of the world line for one of the particles is given by $q^n = (-1)^{\bar{n}}((t - t_0)^2 - e)^{\frac{1}{2}} + Q_0^n$, again, obviously not straight.[26]

Thus generators have been found which obey the conditions and imply nonzero accelerations. Now, after this, we can ask questions regarding the applicability of this set of generators to particle motion.

Let us for a moment consider the previously mentioned objections. The question of instantaneous propagation, in the face of relativity, has an answer that was discussed in connection with the Thomas proof. Questions of causality, and the speed of interaction arise only in the case of some action which may or may not occur, i.e., an external agency. The extension of the present formalism to include other interactions has not been presented, so cannot be a topic of discussion at this level. As to the other objection, if this formalism can describe interaction and has a particle interpretation, then it would seem to be an admissible description as an alternative to the use of a field description.

Returning to the consideration of the particle interpretation of these generators (which satisfy the conditions and imply nonzero acceleration), further development demonstrates that the particle interpretation is not very satisfying. The velocities of the particles are functions of the positions of the particles, and thus are not independent initial data. Thus a modification of these generators (the modified one-dimensional generators) which have the proper number of initial conditions must be considered.[27] These modified generators (if they exist) with a suitable particle interpretation, could then be used to describe interaction.

Let us now note, however, that in three dimensions, it is not a question of a suitable particle interpretation, nor a question of causality which rendered the formalism unsuitable; instead, it was the logically prior difficulty that the mathematical problem had no solution, i.e., there were no generators which satisfied the Lorentz and world-line conditions, and which had nonzero acceleration.

In the three-dimensional case, a mathematical

[26] The reason for the difference between the one- and three-dimension case lies in the possibility of an H which is a function of p and not P and, at the same time, may be split into a pair of functions which depend upon p^1 and p^2 separately. The rotational symmetry in three dimensions prevents such a situation. It would be prevented in one dimension if we retained the remanent of rotational symmetry (reflection symmetry) in the one-dimensional model.

[27] The question of the existence of the modified set of generators is at present unknown.

definition of a formalism which has been suggested as useful in the description of relativistic dynamics was presented. The first question to be considered is whether the formalism admits any solutions which are consistent with the definition of the formalism (yes, free theory), and whether any of the admissible solutions have other than straight world lines (no!). If such solutions did exist, then they might be investigated further to see if they were interesting in other related aspects, for instance, the possible introduction of other interactions to test causality and speed of propagation of the interaction.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

The Remainder in Mayer's Fugacity Series*

OLIVER PENROSE†
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 11 July 1963)

Upper and lower bounds are obtained for the remainder after a finite number of terms of the expansions in powers of fugacity z for the pressure p, the s-particle distribution functions, the density ρ, and the fugacity coefficient z/ρ, for a system of particles with two-body interactions. The interaction potential must be either nonnegative or else have a hard core and decrease faster than r^{-3} at large distances. The results hold for all positive z, and apply to lattice gases as well as to fluids.

For nonnegative potentials, the results imply that successive partial sums of each of these fugacity expansions provide alternate upper and lower bounds, valid even if the series diverges, on the physical quantity the expansion represents.

1. INTRODUCTION

A METHOD has been proposed by Elliott Lieb[1] for obtaining upper and lower bounds on the distribution and thermodynamic functions for a classical system of particles with two-body interactions. Lieb concentrated on the case where the interaction potential $\varphi(\mathbf{r})$ is nonnegative and satisfies the condition

$$\int |e^{-\varphi(\mathbf{r})/kT} - 1|\, d^\nu\mathbf{r} < \infty, \tag{1.1}$$

where ν is the number of space dimensions and the integral is over all space.

In the present paper we shall show that for such potentials Lieb's method leads to a very simple result: the expansions in powers of the fugacity z for the pressure p, for all distribution functions, including the number density ρ, and also for the fugacity coefficient z/ρ, have what may be called the *alternating bound property*. This means that, for any positive value of z, successive partial sums of the series give alternate upper and lower bounds.

We shall also obtain the corresponding result for *hard-core* potentials, which may take negative values, but satisfy the conditions[2]

$$\varphi(\mathbf{r}) = +\infty \quad \text{for} \quad r < a, \tag{1.2}$$

$$|\varphi(\mathbf{r})| \leq Ar^{-\nu-\epsilon} \quad \text{for} \quad r \geq a, \tag{1.3}$$

where r denotes the length of the vector $\mathbf{r}$, and a, A, and ϵ are positive constants.

The calculations will be carried out in the notation appropriate to a fluid, but they do not presume either spherical symmetry of the function $\varphi(\mathbf{r})$ or continuity of the set of points available to the particles. Thus the results may be applied also to lattice gases, simply by reinterpreting integrals like (1.1) as sums over lattice sites.[2]

2. BASIC INEQUALITIES

Lieb's method[1] is based on an estimate of the remainder term R_L in the expansion

$$\prod_{i=1}^{N}(1 + f_i) = 1 + \sum_{i=1}^{N} f_i + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} f_i f_j + \cdots \equiv \sum_{l=0}^{L-1} A_l + R_L, \tag{2.1}$$

* Work supported by the U. S. Air Force Office of Scientific Research under Grant No. 62-64.
† On leave of absence from Imperial College, London, England.
[1] E. Lieb, J. Math. Phys. 4, 671 (1963)
[2] O. Penrose, J. Math. Phys. 4, 1312 (1963).

definition of a formalism which has been suggested as useful in the description of relativistic dynamics was presented. The first question to be considered is whether the formalism admits any solutions which are consistent with the definition of the formalism (yes, free theory), and whether any of the admissible solutions have other than straight world lines (no!). If such solutions did exist, then they might be investigated further to see if they were interesting in other related aspects, for instance, the possible introduction of other interactions to test causality and speed of propagation of the interaction.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

The Remainder in Mayer's Fugacity Series*

OLIVER PENROSE†
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 11 July 1963)

Upper and lower bounds are obtained for the remainder after a finite number of terms of the expansions in powers of fugacity z for the pressure p, the s-particle distribution functions, the density ρ, and the fugacity coefficient z/ρ, for a system of particles with two-body interactions. The interaction potential must be either nonnegative or else have a hard core and decrease faster than r^{-3} at large distances. The results hold for all positive z, and apply to lattice gases as well as to fluids.

For nonnegative potentials, the results imply that successive partial sums of each of these fugacity expansions provide alternate upper and lower bounds, valid even if the series diverges, on the physical quantity the expansion represents.

1. INTRODUCTION

A METHOD has been proposed by Elliott Lieb[1] for obtaining upper and lower bounds on the distribution and thermodynamic functions for a classical system of particles with two-body interactions. Lieb concentrated on the case where the interaction potential $\varphi(\mathbf{r})$ is nonnegative and satisfies the condition

$$\int |e^{-\varphi(\mathbf{r})/kT} - 1|\, d^\nu\mathbf{r} < \infty, \tag{1.1}$$

where ν is the number of space dimensions and the integral is over all space.

In the present paper we shall show that for such potentials Lieb's method leads to a very simple result: the expansions in powers of the fugacity z for the pressure p, for all distribution functions, including the number density ρ, and also for the fugacity coefficient z/ρ, have what may be called the *alternating bound property*. This means that, for any positive value of z, successive partial sums of the series give alternate upper and lower bounds.

We shall also obtain the corresponding result for *hard-core* potentials, which may take negative values, but satisfy the conditions[2]

$$\varphi(\mathbf{r}) = +\infty \quad \text{for} \quad r < a, \tag{1.2}$$

$$|\varphi(\mathbf{r})| \leq Ar^{-\nu-\epsilon} \quad \text{for} \quad r \geq a, \tag{1.3}$$

where r denotes the length of the vector $\mathbf{r}$, and a, A, and ϵ are positive constants.

The calculations will be carried out in the notation appropriate to a fluid, but they do not presume either spherical symmetry of the function $\varphi(\mathbf{r})$ or continuity of the set of points available to the particles. Thus the results may be applied also to lattice gases, simply by reinterpreting integrals like (1.1) as sums over lattice sites.[2]

2. BASIC INEQUALITIES

Lieb's method[1] is based on an estimate of the remainder term R_L in the expansion

$$\prod_{i=1}^{N}(1 + f_i) = 1 + \sum_{i=1}^{N} f_i + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} f_i f_j + \cdots \equiv \sum_{l=0}^{L-1} A_l + R_L, \tag{2.1}$$

* Work supported by the U. S. Air Force Office of Scientific Research under Grant No. 62-64.
† On leave of absence from Imperial College, London, England.
[1] E. Lieb, J. Math. Phys. 4, 671 (1963)
[2] O. Penrose, J. Math. Phys. 4, 1312 (1963).

where L is a nonnegative integer, and

$$A_l \equiv \sum_{i_1=1}^{N-l+1} \sum_{i_2=i_1+1}^{N-l+2} \cdots \sum_{i_l=i_{l-1}+1}^{N} f_{i_1} f_{i_2} \cdots f_{i_l} \qquad (2.2)$$

is the sum of the $N!/l!(N-l)!$ different possible products of l different factors f_i. We interpret A_0 as 1, and any summation whose lower limit exceeds its upper limit as zero, so that the sum in (2.1) vanishes if $L = 0$, and A_l vanishes if $l > N$. The numbers $f_1, \cdots f_N$ are values taken by the Ursell–Mayer function[3]

$$f(\mathbf{r}) \equiv e^{-\varphi(\mathbf{r})/kT} - 1. \qquad (2.3)$$

For a nonnegative potential, they lie in the range $-1 \leq f_i \leq 0$, and Lieb[1,4] showed that R_L then has the sign $(-)^L$. We now derive the corresponding result for hard-core potentials.

Using Lagrange's form[5] for the remainder in the Taylor expansion of $\prod (1 + sf_i)$ in powers of s, and then setting $s = 1$, one may demonstrate the existence of a number t in the range $0 \leq t \leq 1$ such that

$$R_L = \frac{1}{L!}\left(\frac{d}{dt}\right)^L \prod_{i=1}^{N}(1 + tf_i) = \sum_{\{i\}} f_{i_1} \cdots f_{i_L} \prod_j{}' (1 + tf_j), \qquad (2.4)$$

where $\sum_{\{i\}}$ is the summation used in (2.2) and $\prod_j'$ is a product over all values of j satisfying the conditions $1 \leq j \leq N$, $j \neq i_1$, $j \neq i_2$, $\cdots$ $j \neq i_L$.

To put (2.4) into a convenient form we define, for any real quantity Q, the two nonnegative quantities

$$[Q]_\pm \equiv \tfrac{1}{2}(|Q| \pm Q) = \max(\pm Q, 0). \qquad (2.5)$$

If P, P_+, and P_- are any quantities satisfying both $P \leq +P_+$ and $P \geq -P_-$, or in a more compact notation,

$$P \lesseqgtr \pm P_\pm, \qquad (2.6)$$

then (2.5) implies

$$PQ \lesseqgtr \pm P_+[Q]_\pm \pm P_-[Q]_\mp. \qquad (2.7)$$

Note that P_+ and $[P]_+$ need not be equal.

The result (2.7) can be used to simplify (2.4). Suppose the f_i's satisfy the conditions

$$1 + f_i \geq 0 \qquad \text{for all } i, \qquad (2.8)$$

$$\prod_{j=1}^{N}(1 + [f_i]_+) \leq u, \qquad (2.9)$$

where u is a constant. Then the product over j in (2.4) lies between 0 and u so that (2.7) gives

$$R_L \lesseqgtr \pm u \sum_{\{i\}} [f_{i_1} \cdots f_{i_L}]_\pm \quad (L = 0, 1, \cdots). \qquad (2.10)$$

If none of the f_i's is positive, u can be taken as 1 and (2.10) implies that R_L has the sign $(-)^L$; this fact is the starting point of Lieb's paper.

3. THE TRUNCATED KIRKWOOD–SALSBURG EQUATION

The s-particle density distribution function for a grand canonical ensemble is defined, for all positive integers s, by

$$n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) = \sum_{N=0}^{\infty} z^{s+N} \int \cdots \int e_{s+N} \times (\mathbf{x}_1 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_N)\, d^\nu\mathbf{y}_1 \cdots d^\nu\mathbf{y}_N/N!\, \Xi(z), \qquad (3.1)$$

where

$$e_r(\mathbf{x}_1 \cdots \mathbf{x}_r) \equiv \exp\left[-\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} \varphi(\mathbf{x}_i - \mathbf{x}_j)/kT\right] \qquad (3.2)$$

for all $r \geq 1$, $\Xi(z)$ is the grand partition function defined by

$$\Xi(z) \equiv 1 + \sum_{N=1}^{\infty} \frac{z^N}{N!} \int \cdots \int e_N(\mathbf{y}_1 \cdots \mathbf{y}_N)\, d^\nu\mathbf{y}_1 \cdots d^\nu\mathbf{y}_N, \qquad (3.3)$$

and all integrations are over the spatial region V occupied by the system. The temperature T will be treated as a constant, but functional dependence on z will be shown explicitly.

To make use of the inequalities derived in Sec. 2, we use (3.2) to write the integrand of (3.1) in the form

$$e_{s+N}(\mathbf{x}_1 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_N) = e_{s+N-1}(\mathbf{x}_2 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_N) \times \prod_{i=2}^{s} e_2(\mathbf{x}_1, \mathbf{x}_i) \prod_{j=1}^{N}(1 + f_j), \qquad (3.4)$$

where

$$f_j \equiv f(\mathbf{x}_1 - \mathbf{y}_j) \qquad (j = 1 \cdots N), \qquad (3.5)$$

and $f(\mathbf{r})$ is the Mayer–Ursell function defined in (2.3). With the f_j's defined in this way, the condition (2.8) for the validity of (2.10) is satisfied for any $\varphi(\mathbf{r})$. The left side of the other condition, (2.9), has the form

$$\exp\left[-(kT)^{-1} \sum_i \varphi(\mathbf{x}_1 - \mathbf{y}_i)\right], \qquad (3.6)$$

where the sum ranges over a selection from the

---

[3] T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), Chap. 5.

[4] See also G. Pólya and G. Szegö, *Aufgaben und Lehrsätze der Analysis* (Julius Springer Verlag, Berlin, 1925), Vol. II, Part V, Chap. 3, No. 163.

[5] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, New York, 1927), 4th ed., p. 96.

numbers 1, 2, $\cdots N$. Both for nonnegative and for hard-core potentials, such sums have[2] a lower bound $-2\Phi'$ independent of $\mathbf{x}_1, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \cdots$, so that (2.9) holds with

$$u = \exp 2\Phi'/kT \geq 1, \tag{3.7}$$

provided, in the case of hard-core potentials, that no two of the points $\mathbf{y}_i$ in (3.6) are separated by a distance less than a.

Substitution of (2.1), (2.2), and (2.10) into (3.4) yields an inequality which holds without any proviso, since both sides vanish if any of the points y_i are too close together. This inequality may then be substituted into (3.1) to give, for real positive z,

$$\begin{aligned} n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) &\leqslant z\left\{\prod_{i=2}^{s} e_2(\mathbf{x}_1, \mathbf{x}_i)\right\} \\ &\times \left\{\sum_{l=0}^{L-1} \frac{1}{l!} \int \cdots \int n_{s-1+l}(\mathbf{x}_2 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_l \mid z)\right. \\ &\times f_1 \cdots f_l \, d\mathbf{y}_1 \cdots d\mathbf{y}_l \pm \frac{u}{L!} \int \cdots \int n_{s-1+L} \\ &\left. \times (\mathbf{x}_2 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_L \mid z)[f_1 \cdots f_L]_+ \, d\mathbf{y}_1 \cdots d\mathbf{y}_L\right\}. \end{aligned} \tag{3.8}$$

This inequality[6] holds for all $L \geq 0$, and for all $s \geq 1$ if the convention $n_0 = 1$ is adopted to deal with the $s = 1$, $l = 0$ term. If $L \to \infty$ and the remainder term tends to zero, (3.8) reduces to the Kirkwood–Salsburg equation[7]; for finite L, (3.8) provides upper and lower bounds for the remainder after L terms of the infinite series in the Kirkwood–Salsburg equation.

By precisely analogous methods, one may derive truncated forms of the Mayer–Montroll equation[8] and the equation[9] of which both the Mayer–Montroll equation and the Kirkwood–Salsburg equation are special cases. Alternatively, all these inequalities can be derived from the theory of nonuniform systems.[10]

4. THE REMAINDER IN THE FUGACITY SERIES FOR THE DISTRIBUTION FUNCTIONS

A formal solution of the Kirkwood–Salsburg equation is provided by the fugacity series

$$n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) = \sum_{k=0}^{\infty} n_{s,k}(\mathbf{x}_1 \cdots \mathbf{x}_s) z^{s+k}, \qquad (s = 0, 1, \cdots), \tag{4.1}$$

whose radius of convergence has[2] a positive lower bound, independent of s and V. Since $n_0 = 1$, the coefficients $n_{s,k}$ for $s = 0$ are

$$n_{0,k} = \delta_{0,k} \begin{cases} \equiv 1 & \text{if } k = 0 \\ \equiv 0 & \text{if } k \neq 0 \end{cases}. \tag{4.2}$$

For $s \geq 1$ they satisfy[2] a recurrence relation obtained by making $L \to \infty$ in (3.8) and then substituting from (4.1) and equating coefficients of powers of z:

$$\begin{aligned} &n_{s,k}(\mathbf{x}_1 \cdots \mathbf{x}_s) \\ &= \left\{\prod_{i=2}^{s} e_2(\mathbf{x}_1, \mathbf{x}_i)\right\} \sum_{l=0}^{k} \frac{1}{l!} \int \cdots \int d\mathbf{y}_1 \cdots d\mathbf{y}_l \\ &\quad \times n_{s-1+l,k-l}(\mathbf{x}_2 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_l) f_1 \cdots f_l. \end{aligned} \tag{4.3}$$

The central result of this paper is obtained by using the inequality (3.8) to study the remainder term in the series (4.1). The remainder after L ($= 0, 1, \cdots$) terms of that series is $O(z^{s+L})$ and is therefore conveniently written $m_{s,L} z^{s+L} \equiv m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) z^{s+L}$:

$$n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) = \sum_{l=0}^{L-1} n_{s,l}(\mathbf{x}_1 \cdots \mathbf{x}_s) z^{s+l} + m_{s,L} z^{s+L}, \tag{4.4}$$

where

$$m_{s,L} = O(1) \quad \text{as} \quad z \to 0. \tag{4.5}$$

Since $n_0 \equiv 1$ the quantities $m_{s,L}$ for $s = 0$ are given by an equation analogous to (4.2),

$$m_{0,L} = \delta_{0,L}. \tag{4.6}$$

For $s \geq 1$, they can be estimated using a recurrence relation analogous to (4.3), obtained by substituting (4.4) into (3.8). This substitution yields

$$\begin{aligned} &\sum_{l=0}^{L-1} n_{s,l}(\mathbf{x}_1 \cdots \mathbf{x}_s) z^{s+l} + m_{s,L} z^{s+L} \leqslant \left\{\prod_{i=2}^{s} e_2(\mathbf{x}_1, \mathbf{x}_i)\right\} \\ &\times \left\{\sum_{l=0}^{L-1} \frac{1}{l!} \int \cdots \int \left[\sum_{k=0}^{L-l-1} n_{s-1+l,k}\right.\right. \\ &\left. \times (\mathbf{x}_2 \cdots \mathbf{x}_s, \mathbf{y}_1 \cdots \mathbf{y}_l) z^{s+l+k} + m_{s-1+l,L-l} z^{s+L}\right] \\ &\times f_1 \cdots f_l \, d\mathbf{y}_1 \cdots d\mathbf{y}_l \pm \frac{u}{L!} \int \cdots \int m_{s-1+L,0} z^{s+L} \\ &\left. \times [f_1 \cdots f_L]_+ \, d\mathbf{y}_1 \cdots d\mathbf{y}_L\right\}. \end{aligned} \tag{4.7}$$

By (4.3), the terms in z^s, z^{s+1}, $\cdots z^{s+L-1}$ cancel, and on dividing by z^{s+L}, we obtain the desired recurrence relation

[6] A particular case of (3.8) with the lower sign, $s = 2$, $L = 1$, and nonnegative potentials, is discussed by E. Lieb [reference 1, Eq. (2.15)].

[7] J. G. Kirkwood and Z. W. Salsburg, Discussions Faraday Soc. **15**, 28 (1953); T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), p. 251.

[8] J. E. Mayer and E. Montroll, J. Chem. Phys. **9**, 2 (1941).

[9] J. E. Mayer, J. Chem. Phys. **15**, 187 (1947), Eq. (54′).

[10] J. L. Lebowitz and J. K. Percus, J. Math. Phys. **4**, 1495 (1963) (following paper).

$$m_{s,L} \lesseqgtr \left\{ \prod_{i=2}^{s} e_2(\mathbf{x}_1, \mathbf{x}_i) \right\} \times \left\{ \sum_{l=0}^{L-1} \frac{1}{l!} \int \cdots \int m_{s-1+l,L-l} f_1 \cdots f_l \, d\mathbf{y}_1 \cdots d\mathbf{y}_l \pm \frac{u}{L!} \int \cdots \int m_{s-1+L,0} [f_1 \cdots f_L]_{\mp} \, d\mathbf{y}_1 \cdots d\mathbf{y}_L \right\}, \tag{4.8}$$

which holds for all $s \geq 1$, $L \geq 0$.

Since neither the starting condition (4.6) nor the recurrence relation (4.8) contains z explicitly, the upper and lower bounds calculated from them will be independent of z. Rather than attempt a "best possible" solution of the recurrence relation, we shall simply[11] look for upper and lower bounds on the $m_{s,L}$'s that are independent of $\mathbf{x}_1 \cdots \mathbf{x}_s$ as well as z. Such bounds may be written

$$m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) \lesseqgtr \pm m_{s,L,\mp}, \tag{4.9}$$

where the quantities $m_{s,L,+}$ and $m_{s,L,-}$ depend, unlike $m_{s,L}$, on s and L only. Substituting from (4.9) into (4.8), and using (2.7), we deduce

$$m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) \lesseqgtr \pm \left\{ \prod_{i=2}^{s} e_2(\mathbf{x}_1, \mathbf{x}_i) \right\} \left\{ \sum_{l=0}^{L-1} \frac{1}{l!} \times \left(m_{s-1+l,L-l,+} \int \cdots \int [f_1 \cdots f_l]_{\mp} \, d\mathbf{y}_1 \cdots d\mathbf{y}_l + m_{s-1+l,L-l,-} \int \cdots \int [f_1 \cdots f_l]_{\mp} \, d\mathbf{y}_1 \cdots d\mathbf{y}_l \right) + \frac{u}{L!} m_{s-1+L,0,+} \int \cdots \int [f_1 \cdots f_L]_{\mp} \, d\mathbf{y}_1 \cdots d\mathbf{y}_L \right\}. \tag{4.10}$$

Since $m_{0,K,+} \geq 0$ for all K by (4.6) and (4.9), it can be shown inductively that all the upper bounds on the quantities $m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z)$ derived from (4.10) are ≥ 0, and all the lower bounds are ≤ 0; therefore we may without loss of generality take all the quantities $m_{s,L,\mp}$ in (4.9) to be nonnegative. Since the integrands in (4.10) are also nonnegative, the inequalities are preserved if we extend the range of the integrations beyond the region V to the whole of ν-dimensional space. These integrals over all space simplify, by (2.5) and (3.5), to

$$\int \cdots \int [f_1 \cdots f_l]_{\mp} \, d\mathbf{y}_1 \cdots d\mathbf{y}_l = \frac{1}{2} \int \cdots \int (|f_1 \cdots f_l| \pm f_1 \cdots f_l) \, d\mathbf{y}_1 \cdots d\mathbf{y}_l = \frac{1}{2} (B^l \pm \bar{B}^l) \equiv B_{\mp}^{(l)}, \tag{4.11}$$

where

$$B \equiv \int_{\text{all space}} |e^{-\varphi(\mathbf{r})/kT} - 1| \, d^\nu \mathbf{r}, \tag{4.12}$$

$$\bar{B} \equiv \int_{\text{all space}} [e^{-\varphi(\mathbf{r})/kT} - 1] \, d^\nu \mathbf{r}. \tag{4.13}$$

A further simplification of (4.10) is made possible by the fact that the second factor in braces is nonnegative. Therefore, the first factor in braces, $\prod e_2(\mathbf{x}_1, \mathbf{x}_i)$, may be replaced by its upper bound u, so that (4.10) implies a pair of inequalities of the form (4.9), with

$$m_{s,L,\mp} \equiv u \sum_{l=0}^{L-1} \frac{1}{l!} \{ m_{s-1+l,L-l,+} B_{\mp}^{(l)} + m_{s-1+l,L-l,-} B_{\pm}^{(l)} \} + \frac{u^2}{L!} m_{s-1+L,0,+} B_{\mp}^{(L)} \qquad (s = 1, 2, \cdots). \tag{4.14}$$

For hard-core potentials, the upper bound u on $\prod e_2(\mathbf{x}_1, \mathbf{x}_i)$ does not apply if two or more of the points $\mathbf{x}_2 \cdots \mathbf{x}_s$ have a separation less than a; but in this case both $n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z)$ and $m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z)$ vanish identically for all z, and therefore (4.9) is again satisfied, simply because the $m_{s,L,\mp}$'s are nonnegative.

The recurrence relation (4.14) defines the quantities $m_{s,L,\mp}$ for all $s \geq 1$, $L \geq 0$, provided definitions of $m_{0,L,\mp}$ are supplied. The best nonnegative values for $m_{0,L,\mp}$ are, by (4.6) and (4.9),

$$m_{0,L,+} = \delta_{0,L} \quad \text{and} \quad m_{0,L,-} = 0 \quad (L = 0, 1, \cdots). \tag{4.15}$$

5. SOLUTION OF THE RECURRENCE RELATION

The recurrence relation (4.14) can be simplified by transforming to new sets of unknowns $q_{s,L}$ and $\bar{q}_{s,L}$ (proportional to the sum and difference of $m_{s,L,+}$ and $m_{s,L,-}$) defined implicitly for $s = 0, 1, \cdots$ and $L = 0, 1, \cdots$ by

$$m_{s,L,\mp} \equiv \tfrac{1}{2} u^{s+L} \{ B^L q_{s,L} \pm \bar{B}^L \bar{q}_{s,L} \}. \tag{5.1}$$

Substituting into (4.14) and (4.15), and using (4.11), we find that the $q_{s,L}$'s and $\bar{q}_{s,L}$'s satisfy the same recurrence relations and starting conditions, and are therefore identical. Hence (5.1) simplifies to

$$m_{s,L,\mp} = u^{s+L} B_{\mp}^{(L)} q_{s,L}, \tag{5.2}$$

where the set of quantities $q_{s,L}$ $(= \bar{q}_{s,L})$ is defined by

$$q_{0,L} = \delta_{0,L}, \tag{5.3}$$

[11] This simplification is not obligatory, at least for small L. For example, J. Groeneveld [Phys. Letters 3, 50 (1962)] obtains the configuration-dependent upper bound $m_{s,0}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) \leq u^s e_s(\mathbf{x}_1 \cdots \mathbf{x}_s)$.

$$q_{s,L} = \sum_{l=0}^{L-1} \frac{1}{l!}\, q_{s-1+l,L-l} + \frac{u}{L!}\, q_{s-1+L,0} \qquad (s = 1, 2, \cdots), \tag{5.4}$$

for all $L \geq 0$.

When $u = 1$, the solution of (5.3) and (5.4) is

$$q_{s,L} = s(s + L)^{L-1}/L!. \tag{5.5}$$

For $u > 1$, the general solution is complicated, but it can be verified that (5.3) and (5.4) imply

$$q_{s,L} \leq s(s + L)^{L-1}u^{s+L}/L!. \tag{5.6}$$

The first few $q_{s,L}$'s are

$$\begin{aligned} q_{0,0} &= 1, \\ q_{1,0} &= u, \\ q_{2,0} &= u^2, \\ q_{3,0} &= u^3, \\ q_{0,1} &= 0, \\ q_{1,1} &= u^2, \\ q_{2,1} &= u^3 + u^2, \\ q_{3,1} &= u^4 + u^3 + u^2, \\ q_{0,2} &= 0, \\ q_{1,2} &= \tfrac{1}{2}(u^3 + 2u^2), \\ q_{2,2} &= \tfrac{1}{2}(u^4 + 3u^3 + 4u^2), \text{ etc.} \end{aligned} \tag{5.7}$$

Combining (4.9), (5.2), and (5.5) or (5.6), gives our basic set of upper and lower bounds on the remainder terms in the fugacity expansions (4.1):

$$\begin{aligned} m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) &\lesseqgtr \pm u^{s+L} B_{\pm}^{(L)} q_{s,L} \\ &\lesseqgtr \pm u^{2(s+L)} B_{\pm}^{(L)} s(s + L)^{L-1}/L!, \end{aligned} \tag{5.8}$$

for all $s \geq 1$ and $L \geq 0$.

For $s = 1$, this result can be improved on by setting $s = 1$ in (4.8) and substituting from (4.9), (4.11), (5.2), and (5.4). Since the product over j in (4.8) degenerates to 1, we obtain

$$\begin{aligned} m_{1,L}(\mathbf{x}_1 \mid z) &\lesseqgtr \pm u^L B_{\pm}^{(L)} q_{1,L} \\ &\lesseqgtr \pm u^{1+2L} B_{\pm}^{(L)} (1 + L)^{L-1}/L!, \end{aligned} \tag{5.9}$$

for all $L \geq 0$.

For nonnegative potentials, the quantities $\bar{B}$ and B are equal and opposite, by (4.12) and (4.13), so that

$$B_{+}^{(L)} = 0 \quad \text{if } L \text{ is odd,}$$
$$\text{and } B_{-}^{(L)} = 0 \quad \text{if } L \text{ is even.} \tag{5.10}$$

This implies, by (5.8), that $m_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z)$ has the sign $(-)^L$. It follows at once from (4.4) that for nonnegative potentials the series (4.1) has the *alternating bound property* defined in Sec. 1.

6. MODIFIED FUGACITY SERIES

Stronger inequalities can be obtained if the system is spatially uniform. One way to make the system uniform is to make the region V infinitely large. Another is to make V a cube, and make the interaction between particles at $\mathbf{x}_i$ and $\mathbf{x}_j$ periodic in $\mathbf{x}_i - \mathbf{x}_j$ with unit cell V; this is best done by defining this interaction as $\varphi(\mathbf{r}_{ij})$ where $\mathbf{r}_{ij}$ is the vector from $\mathbf{x}_i$ to the nearest point congruent to $\mathbf{x}_j$ under translations with unit cell V.

For a uniform system the number density is a function of z only:

$$n_1(\mathbf{x}_1 \mid z) = \rho(z) \equiv \mathcal{V}^{-1} \int d^\nu \mathbf{x}_1 n_1(\mathbf{x}_1 \mid z), \tag{6.1}$$

where $\mathcal{V} \equiv \int d^\nu \mathbf{x}$ is the volume of the region V; and we may seek a solution of the Kirkwood–Salsburg equation by using, instead of (4.1), the expansion

$$n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) = \rho \sum_{k=0}^{\infty} n'_{s,k}(\mathbf{x}_1 \cdots \mathbf{x}_s) z^{s+k-1} \qquad (s = 1, 2, \cdots). \tag{6.2}$$

Since $n_1 = \rho$, the coefficients for $s = 1$ are

$$n'_{1,k}(\mathbf{x}_1) = \delta_{0,k} \qquad (k = 0, 1, \cdots). \tag{6.3}$$

Recurrence relations for the $n'_{s,k}$'s when $s \geq 2$ can be obtained, just as in the derivation of (4.3), by substituting in the Kirkwood–Salsburg equation. These recurrence relations turn out to be identical with those [Eq. (4.3)] satisfied by the $n_{s,k}$'s, the only difference being that the initial condition is now (6.3) instead of (4.2). By analogy with (4.4), the remainder after L terms of Eq. (6.2) will be written $m'_{s,L}\, \rho z^{s+L-1} \equiv m'_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z)\, \rho z^{s+L-1}$, so that

$$n_s(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) = \rho \times \left[\sum_{l=0}^{L-1} n'_{s,l}(\mathbf{x}_1 \cdots \mathbf{x}_s) z^{s+l-1} + m'_{s,L} z^{s+L-1} \right], \tag{6.4}$$

where

$$m'_{s,L} = O(1) \quad \text{as} \quad z \to 0. \tag{6.5}$$

Setting $s = 1$ in (6.4), and using (6.1) and (6.3), we find

$$m'_{1,L} = \delta_{0,L} \qquad (L = 0, 1, \cdots). \tag{6.6}$$

The $m'_{s,L}$'s for $s \geq 2$ can be studied by precisely

the same methods as were used in Sec. 4 to study the $m_{s,L}$'s for $s \geq 1$. They have upper and lower bounds analogous to the bounds in (4.9):

$$m'_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) \lesseqgtr \pm m'_{s,L,\pm}, \tag{6.7}$$

where the quantities $m'_{s,L,\pm}$ for $s \geq 2$ satisfy the same recurrence relation (4.14) as do the $m_{s,L,\pm}$'s. However, instead of (4.15), the starting conditions are now

$$m'_{1,L,+} = \delta_{0,L} \text{ and } m'_{1,L,-} = 0 \quad (L = 0, 1, \cdots). \tag{6.8}$$

Comparing (6.8) with (4.15) and using the recurrence relation (4.14) both for $m_{s,L,\pm}$ and $m'_{s,L,\pm}$, we find that these two sets of quantities are related by

$$m'_{s,L,\pm} = m_{s-1,L,\pm} \quad (s = 1, 2, \cdots; \quad L = 0, 1, \cdots). \tag{6.9}$$

Combining (6.7), (6.9), (5.2), and (5.6) gives our basic set of upper and lower bounds on the remainder terms in the expansion (6.2):

$$m'_{s,L}(\mathbf{x}_1 \cdots \mathbf{x}_s \mid z) \lesseqgtr \pm u^{s-1+L} B_{\pm}^{(L)} q_{s-1,L}$$
$$\lesseqgtr \pm u^{2(s-1+L)} B_{\pm}^{(L)} (s-1)(s-1+L)^{L-1}/L!. \tag{6.10}$$

For nonnegative potentials, it follows from (5.10) that $m'_{s,L}$ has the sign $(-)^L$, and therefore that the series (6.2) has the alternating bound property.

7. THERMODYNAMIC INEQUALITIES

Upper and lower bounds on thermodynamic quantities such as the density ρ and pressure p follow in a simple way from the foregoing results. For a finite system, which need not be uniform, the mean density $\rho(z, V)$ has, by (4.4) and (5.9), the sets of upper and lower bounds

$$\rho(z, V) \equiv \mathcal{V}^{-1} \int d^\nu \mathbf{x}_1 n_1(\mathbf{x}_1 \mid z)$$
$$\lesseqgtr \sum_{l=1}^{L-1} l b_l(V) z^l + L b_L^{(\pm)} z^L, \tag{7.1}$$

for $L = 1, 2, \cdots$, where

$$l b_l(V) \equiv \mathcal{V}^{-1} \int d^\nu \mathbf{x}_1 n_{1,l-1}(\mathbf{x}_1 \mid z), \tag{7.2}$$

$$L b_L^{(\pm)} \equiv \pm u^{2L-1} B_{\pm}^{(L-1)} L^{L-1}/L!. \tag{7.3}$$

The grand partition function therefore has the bounds

$$\mathcal{V}^{-1} \ln \Xi = \int_0^z \rho \, dz/z \lesseqgtr \sum_{l=1}^{L-1} b_l(V) z^l + b_L^{(\pm)} z^L. \tag{7.4}$$

In the simplest case $L = 1$, the pair of inequalities (7.1) reduces to $\rho \geq 0$ and Groeneveld's result[12] $\rho \leq uz$, and the pair (7.4) reduces to $0 \leq \ln \Xi \leq \mathcal{V} z u$, which can also be derived by a simpler method.[2]

Taking the limit $V \to \infty$ in (7.1) and (7.4), we obtain upper and lower bounds for the remainder after L terms in Mayer's fugacity series[3] for the thermodynamic density $\rho(z)$ and the thermodynamic pressure $p(z)$:

$$\rho(z) \lesseqgtr \sum_{l=1}^{L-1} l b_l z^l + L b_L^{(\pm)} z^L, \tag{7.5}$$

$$p(z)/kT \lesseqgtr \sum_{l=1}^{L-1} b_l z^l + b_L^{(\pm)} z^L, \tag{7.6}$$

where

$$b_l \equiv \lim_{V \to \infty} b_l(V). \tag{7.7}$$

For nonnegative potentials, $b_L^{(+)}$ vanishes for even L, and $b_L^{(-)}$ for odd L, by (5.10) and (7.3). Therefore, the fugacity expansions (7.1), (7.4), (7.5), and (7.6) have the alternating bound property for nonnegative potentials.

For spatially uniform systems, we can obtain another set of upper and lower bounds on $\rho(z)$ by using the truncated series from Sec. 6 instead of those from Sec. 4. Setting $s = 1$ in (3.8), substituting from (6.1) and (6.4) with k written for l and $L - l$ for L, and arranging by powers of z, we obtain

$$\rho \lesseqgtr z + \rho \left[\sum_{l=1}^{L-1} a_l z^l + r_L^{(\pm)} z^L \right], \tag{7.8}$$

for $L = 1, 2, \cdots$, where[13]

$$a_l \equiv \sum_{l=1}^{l} \frac{1}{l!} \int \cdots \int n'_{l,l-l}$$
$$\times (\mathbf{y}_1 \cdots \mathbf{y}_l) f_1 \cdots f_l \, d\mathbf{y}_1 \cdots d\mathbf{y}_l, \tag{7.9}$$

$$r_L^{(\pm)} \equiv \sum_{l=1}^{L-1} \frac{1}{l!} \int \cdots \int m'_{l,L-l}$$
$$\times (\mathbf{y}_1 \cdots \mathbf{y}_l) f_1 \cdots f_l \, d\mathbf{y}_1 \cdots d\mathbf{y}_l$$
$$\pm \frac{u}{L!} \int \cdots \int m'_{L,0}(\mathbf{y}_1 \cdots \mathbf{y}_L)$$
$$\times [f_1 \cdots f_L]_{\pm} \, d\mathbf{y}_1 \cdots d\mathbf{y}_L. \tag{7.10}$$

The remainder estimates provided by (7.10) can be simplified [using first (2.7), (6.10), and (4.11), and afterwards (5.6)] to

$$r_L^{(\pm)} \lesseqgtr a_L^{(\pm)} \equiv \sum_{l=1}^{L-1} \frac{1}{l!} \{ \pm u^{L-1} B_+^{(L-l)} q_{l-1,L-l} B_{\pm}^{(l)}$$

[12] J. Groeneveld (reference 11).

[13] The coefficients a_l can be expressed in terms of b_l's by algebraic formulas which are discussed by Lieb (reference 1), and by Lebowitz and Percus (reference 10).

$$\pm u^{L-1}B_-^{(L-l)}q_{l-1,L-l}B_+^{(l)}\} \pm \frac{u^L}{L!}\,q_{L-1,0}B_+^{(L)}$$

$$= \pm u^{L-1}B_+^{(L)}\left\{\sum_{l=1}^{L-1}\frac{q_{l-1,L-1}}{l!} + \frac{uq_{L-1,0}}{L!}\right\} \leqslant \pm u^{2L-1}B_+^{(L)}$$

$$\times \sum_{l=1}^{L}\frac{(l-1)(L-1)^{L-l-1}}{l!\,(L-l)!}$$

$$= \pm u^{2L-1}B_+^{(L)}\,\frac{(L-1)^{L-1}}{L!}. \tag{7.11}$$

The results (7.8) and (7.11) may be combined to give a remainder theorem for the fugacity expansion of the fugacity coefficient z/ρ:

$$z/\rho \gtrless 1 - \sum_{l=1}^{L-1} a_l z^l - a_L^{(*)}z^L \quad (L = 1, 2, \cdots). \tag{7.12}$$

For nonnegative potentials, (5.10) implies that $a_L^{(+)}$ vanishes for odd L, and $a_L^{(-)}$ for even L, so that the infinite series corresponding to (7.12) has the alternating bound property. Lieb[1] demonstrated the existence of a set of polynomials giving alternating upper and lower bounds on z/ρ. He conjectured that the coefficient of z^l in the Lth polynomial would equal a_l for sufficiently large L. The alternating bound property shows that Lieb's conjecture is true in its simplest possible form, since the coefficient of z^l is always a_l, unless it vanishes trivially.

By combining (7.12) with the identity $p/kT = \int_0^z \rho\, dz/z$ we obtain the further inequalities for $p(z)$:

$$p/kT \lesseqgtr \int_0^z dz\left[1 - \sum_{l=1}^{L-1} a_l z^l - a_L^{(*)}z^L\right]^{-1}, \tag{7.13}$$

which are valid provided that the quantity in square brackets is positive throughout the range of integration. By (7.12), this condition is always satisfied for the lower bounds on p/kT, but for the upper bounds it is violated[1] for large z.

Other thermodynamic quantities can be studied by similar methods. For example, the energy can be expressed in terms of the two-particle distribution function and so can be bounded above and below with the help of (5.8) or (6.10) for $s = 2$. The free energy per unit volume equals $\rho kT \log [z(h^2/2\pi mkT)^{\frac{3}{2}}] - p$ and can therefore be bounded above and below using the bounds on ρ and p derived in this section.

8. DISCUSSION

The upper and lower bounds obtained here are valid for all z, but their usefulness depends on the value of z. If z is small (less than $1/u^2eB$), the bounds on the remainders given in (7.5) and (7.6) tend to zero with increasing L; therefore, the most accurate estimates of $\rho(z)$ and $p(z)$ are obtained by using as many terms of the fugacity series as possible and using the results of this paper to estimate the small truncation error. On the other hand, for large z, the upper and lower bounds on the remainder move apart as L increases, so that small values of L are best.

For quick reference, some bounds on p and ρ for small L, obtained from (7.5), (7.6), (7.12), and (7.13), are collected here:

$$L = 1 : 0 \leq \rho \leq uz, \tag{8.1}$$

$$0 \leq p/kT \leq uz, \tag{8.2}$$

$$z/(1 + uB_+z) \leq \rho \leq z/(1 - uB_-z), \tag{8.3}$$

$$[\log (1 + uB_+z)]/uB_+ \leq p/kT$$

$$\leq -[\log (1 - uB_-z)]/uB_-, \tag{8.4}$$

$$L = 2 : z - u^3B_+z^2 \leq \rho \leq z + u^3B_-z^2, \tag{8.5}$$

$$z - \tfrac{1}{2}u^3B_+z^2 \leq p/kT \leq z + \tfrac{1}{2}u^3B_-z^2, \quad \text{etc.}, \tag{8.6}$$

where u is defined in (3.7), and

$$B_\pm \equiv B_\pm^{(1)} = \int [e^{-\varphi(\mathbf{r})/kT} - 1]_\pm\, d^\nu\mathbf{r} \tag{8.7}$$

is the contribution of the $\pm$ve part of the potential $\varphi(\mathbf{r})$ to the integral (4.12). The right sides of (8.3) and (8.4) hold only if $z < 1/uB_-$. For nonnegative potentials, the inequalities of this section reduce, since $B_- = 0$ and $u = 1$, to some of those found already by Lieb.[1]

For an illustration of the use of this type of inequality, the reader is referred to the following paper by Lebowitz and Percus,[10] where the approximate equation of state found by Reiss, Frisch, and Lebowitz for the hard-sphere gas is tested against some of Lieb's upper and lower bounds.

ACKNOWLEDGMENTS

I am indebted to Dr. E. Lieb for informing me of his results before publication, and to Dr. J. Lebowitz for stimulating discussions.

Integral Equations and Inequalities in the Theory of Fluids*

J. L. LEBOWITZ
Belfer Graduate School of Science, Yeshiva University, New York, New York
AND
J. K. PERCUS
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 11 June 1963)

Using the method of functional Taylor expansion developed previously, an extensive set of equations is obtained for the distribution functions and Ursell functions in a classical fluid. These include in a systematic way many previously derived relations, e.g., Mayer–Montroll and Kirkwood–Salsburg equations. By terminating the Taylor expansion after a finite number of terms and retaining the remainder, we also obtain inequalities for the distribution functions and thermodynamic parameters of the fluid. For the case of positive interparticle potentials, we recover the inequalities first found by Lieb. For nonpositive potentials, new inequalities (some also obtained by Penrose) are derived. These inequalities are applied to the case of a hard-sphere fluid in three dimensions where they are compared with the results of machine computations and approximate theories. Different inequalities, not obtainable from the above considerations, and some properties of the fugacity expansions, are also derived.

I. INTRODUCTION

RECENTLY, several authors have derived a number of exact results for classical equilibrium systems. Lieb[1] found a series of inequalities for the thermodynamic parameters and distribution functions when the pair potential $\phi(r) \geq 0$, and indicated how these may be extended, partially at least, to more general potentials. Independently of this, Groeneveld,[2] Penrose,[3] and Ruelle[4] obtained rigorous bounds for the radius of convergence of the Mayer series, in powers of the fugacity, for the pressure and distribution functions. The latter two authors used as their starting points the Mayer–Montroll and the Kirkwood–Salsburg[5] sets of equations for the distribution functions. In the present paper, we develop a wide class of new integral equations for the various correlation functions of interest in fluids. One set of these equations is equivalent to those obtained by Mayer[6] of which the Mayer–Montroll and Kirkwood–Salsburg are special cases. Another set of equations includes as a special case two equations recently derived by Green.[7] The same equations also yield in a very natural way the inequalities found by Lieb[1] and Penrose as well as some new inequalities.

The technique used in this paper is an extension of our previous work on the correlation functions and thermodynamic parameters of nonuniform classical fluids.[8] We consider a system represented by a grand canonical ensemble with a chemical potential μ and reciprocal temperature β. Each particle is subject to an external potential $U(\mathbf{r})$, so that the system is characterized by a point function

$$\begin{aligned}\gamma(\mathbf{r}) &= \tfrac{3}{2} \ln (2\pi m/\beta h^2) + \beta\mu - \beta U(\mathbf{r}) \\ &= \ln z - \beta U(\mathbf{r}),\end{aligned} \tag{1.1}$$

where z is the fugacity. Any function of interest in this ensemble ψ may be expanded (formally at least) in a functional Taylor series (with or without remainder) in the deviation of $\gamma(\mathbf{r})$ from some reference value $\gamma_0(\mathbf{r})$. Any other function $\omega(\mathbf{r})$ which is uniquely related to $\gamma(\mathbf{r})$ can equally well serve to characterize the system, and expansions can be made in the deviation of $\omega(\mathbf{r})$ from its corresponding reference value $\omega_0(\mathbf{r})$. In particular, in I and II,[8] we were concerned principally with expansions in $\gamma(\mathbf{r})$ or in the density $n(\mathbf{r})$.

Once we have decided on an appropriate "independent variable" $\omega(\mathbf{r})$ to expand in, it is convenient to introduce a parameter α such that

$$\begin{aligned}\omega(\mathbf{r}, \alpha) &= \omega_0(\mathbf{r}) + \alpha[\omega(\mathbf{r}) - \omega_0(\mathbf{r})], \\ &= \omega_0(\mathbf{r}) + \alpha\Delta\omega(\mathbf{r}).\end{aligned} \tag{1.2}$$

The function to be expanded, ψ, which is a functional of $\omega(\mathbf{r})$, $\psi[\omega(\mathbf{r})]$, may now be considered to be simply a function of α, $\psi(\alpha)$. The expansion is then

* Supported by the Air Force Office of Scientific Research under Grant No. 62-64, the National Aeronautics and Space Administration, Grant NSG 227-62, and the U. S. Atomic Energy Commission under Contract AT(30-1)-1480.

[1] E. Lieb, J. Math. Phys. 4, 671 (1963).
[2] J. Groeneveld, Phys. Letters 3, 50 (1962).
[3] O. Penrose, J. Math. Phys. 4, 1312 (1963).
[4] D. Ruelle, Ann. Phys. (N. Y.) (to be published).
[5] J. E. Mayer and E. Montroll, J. Chem. Phys. 9, 2 (1941); J. G. Kirkwood and Z. Salsburg, Discussions Faraday Soc. 15, 28 (1953).
[6] J. E. Mayer, J. Chem. Phys. 15, 187 (1947).
[7] H. S. Green, Nucl. Fusion 1, 69 (1961).
[8] J. L. Lebowitz and J. K. Percus, J. Math. Phys. 4, 116, 248 (1963) (hereafter referred to as I and II, respectively).

a simple Taylor series in α,

$$\psi(\alpha = 1) = \sum_{j=0}^{\ell} \frac{1}{\ell!} \frac{d^j\psi(\alpha)}{d\alpha^j}\bigg|_{\alpha=0} + \frac{1}{\ell!}\int_0^1 (1-\alpha)^\ell \frac{d^{\ell+1}\psi(\alpha)}{d\alpha^\ell}\, d\alpha,$$

$$= \psi[\omega_0] + \int \frac{\delta\psi[\omega]}{\delta\omega(\mathbf{r}_1)}\bigg|_0 \Delta\omega(\mathbf{r}_1)\, d\mathbf{r}_1 + \frac{1}{2!}\iint \frac{\delta^2\psi[\omega]}{\delta\omega(\mathbf{r}_1)\delta\omega(\mathbf{r}_2)}\bigg|_0 \Delta\omega(\mathbf{r}_1)\Delta\omega(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2 + \cdots + \frac{1}{\ell!}\int_0^1\int\cdots\int (1-\alpha)^\ell \frac{\delta^{\ell+1}\psi[\omega]}{\delta\omega(\mathbf{r}_1,\alpha)\cdots\delta\omega(\mathbf{r}_\ell,\alpha)} \times \Delta\omega(\mathbf{r}_1)\cdots\Delta\omega(\mathbf{r}_\ell)\, d\mathbf{r}_1\cdots d\mathbf{r}_\ell\, d\alpha, \quad (1.3)$$

and the burden of the physics lies in suitable choice of ψ, ω, and ω_0.

II. DISTRIBUTION FUNCTIONS

Let us denote the N-body coordinate-space Boltzmann factor by $e_N(\mathbf{r}_1, \cdots, \mathbf{r}_N) \equiv e_N(\mathbf{r}^N)$. Thus,

$$e_N(\mathbf{r}^N) = \exp\Big[-\beta \sum_{i<j\le N} \phi(\mathbf{r}_{ij})\Big], \quad (2.1)$$

for particles interacting via the pair potential $\phi(\mathbf{r}_{ij})$. The grand partition function has the form[9]

$$\Xi[\gamma] = \sum_{N=0}^{\infty} \frac{1}{N!}\int e_N(\mathbf{r}^N) \prod_{i=1}^{N} e^{\gamma(\mathbf{r}_i)}\, d\mathbf{r}^N, \quad (2.2)$$

and the ensemble weight factor of (2.2) determines all expectation values. Hence, if

$$F_s^N(\mathbf{y}^N) = \sum_{i_1<i_2\cdots i_s\le N} F_s(\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \cdots, \mathbf{y}_{i_s}), \quad (2.3)$$

then

$$\langle F_s^N(\mathbf{y}^N)\rangle = \Xi[\gamma]^{-1} \sum_{s}^{\infty} \frac{1}{N!} \times \int F_s^N(\mathbf{r}^N) e_N(\mathbf{r}^N) \prod_{1}^{N} e^{\gamma(\mathbf{r}_i)}\, d\mathbf{r}^N. \quad (2.4)$$

Since the distribution functions $n_s(\mathbf{y}^s, [\gamma])$ are defined by

$$\langle F_s^N(\mathbf{y}^N)\rangle = \frac{1}{s!}\int F_s(\mathbf{y}^s) n_s(\mathbf{y}^s, [\gamma])\, d\mathbf{y}^s, \quad (2.5)$$

it readily follows from (2.4) [on counting the terms in (2.3)] that

$$n_s(\mathbf{y}^s, [\gamma]) = \Xi[\gamma]^{-1} \prod_{i=1}^{s} e^{\gamma(\mathbf{y}_i)} \sum_{0}^{\infty} \frac{1}{N!} \times \int e_{N+s}(\mathbf{y}^s, \mathbf{r}^N) \prod_{1}^{N} e^{\gamma(\mathbf{r}_i)}\, d\mathbf{r}^N. \quad (2.6)$$

We also have from (2.3) and (2.5) the alternative definition

$$n_s(\mathbf{y}^s, [\gamma]) = \Big\langle \sum_{i_1\ne i_2\cdots\ne i_s} \delta(\mathbf{y}_1 - \mathbf{r}_{i_1})\cdots\delta(\mathbf{y}_s - \mathbf{r}_{i_s})\Big\rangle, \quad (2.7)$$

precisely the distribution of s distinct particles. It is often of value to introduce as well the distributions $\hat{n}_s$ in which[8] the arguments are allowed to refer to identical particles:

$$\hat{n}_s(\mathbf{y}^s, [\gamma]) = \Big\langle\Big(\sum_i \delta(\mathbf{y}_1 - \mathbf{r}_i)\Big)\cdots\Big(\sum_i \delta(\mathbf{y}_s - \mathbf{r}_i)\Big)\Big\rangle. \quad (2.8)$$

It is a simple matter to derive the relations

$$\hat{n}_1(\mathbf{y}) = n_1(\mathbf{y}),$$

$$\hat{n}_2(\mathbf{y}_1, \mathbf{y}_2) = n_2(\mathbf{y}_1, \mathbf{y}_2) + n_1(\mathbf{y}_1)\delta(\mathbf{y}_1 - \mathbf{y}_2), \cdots. \quad (2.9)$$

Both sequences of distributions possess generating functions. If $\lambda(\mathbf{y})$ is a suitably well-behaved test function, then according to (2.8) and (2.2),

$$\int \hat{n}_s(\mathbf{y}^s, [\gamma]) \prod_{1}^{s} \lambda(\mathbf{y}_i)\, d\mathbf{y}^s = \Xi[\gamma]^{-1} \sum_{0}^{\infty} \frac{1}{N!} \times \int e_N(\mathbf{r}^N) \prod_{i=1}^{N} e^{\gamma(\mathbf{r}_i)} \Big(\sum_{1}^{N} \lambda(\mathbf{r}_i)\Big)^s d\mathbf{r}^N, \quad (2.10)$$

leading at once to the relation

$$\sum_{0}^{\infty} \frac{1}{s!}\int \hat{n}_s(\mathbf{y}^s, [\gamma]) \prod_{1}^{s} \lambda(\mathbf{y}_i)\, d\mathbf{y}^s = \Xi[\gamma + \lambda]/\Xi[\gamma]. \quad (2.11)$$

Since (2.11) is itself a functional power series in λ, we have by direct comparison with (1.3),

$$\hat{n}_s(\mathbf{y}^s, [\gamma]) = \Xi[\gamma]^{-1} \frac{\delta^s\Xi[\gamma]}{\delta\gamma(\mathbf{y}_1)\cdots\delta\gamma(\mathbf{y}_s)}. \quad (2.12)$$

On the other hand, from (2.1) and (2.2), using a test function denoted by $\Delta e^{\gamma(\mathbf{y})}$,

$$\int n_s(\mathbf{y}^s, [\gamma]) \prod_{1}^{s} \Delta e^{\gamma(\mathbf{y}_i)}\, d\mathbf{y}^s = \Xi[\gamma]^{-1} \sum_{0}^{\infty} \frac{1}{N!}\int e_N(\mathbf{r}^N) \prod_{1}^{N} e^{\gamma(\mathbf{r}_i)} \times \sum_{i_1\ne i_2\cdots\ne i_s\le N} \Delta e^{\gamma(\mathbf{r}_{i_1})}\cdots\Delta e^{\gamma(\mathbf{r}_{i_s})}\, d\mathbf{r}^N. \quad (2.13)$$

Hence, if $\Delta e^{\gamma} \equiv e^{\gamma+\Delta\gamma} - e^{\gamma}$,

$$\sum_{0}^{\infty} \frac{1}{s!}\int n_s(\mathbf{y}^s, [\gamma]) \prod_{1}^{s} \Delta e^{\gamma(\mathbf{y}_i)}\, d\mathbf{y}^s = \Xi[\gamma + \Delta\gamma]/\Xi[\gamma], \quad (2.14)$$

from which, noting that $\Delta e^{\gamma} = e^{\gamma}(e^{\Delta\gamma} - 1)$,

$$n_s(\mathbf{y}^s, [\gamma]) = \Xi[\gamma]^{-1} \prod_{1}^{s} e^{\gamma(\mathbf{y}_i)} \frac{\delta^s\Xi[\gamma]}{\delta e^{\gamma(\mathbf{y}_1)}\cdots\delta e^{\gamma(\mathbf{y}_s)}}. \quad (2.15)$$

[9] See, for example, T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), p. 233.

Finally, we may similarly consider the Ursell[10] distributions $\mathfrak{F}_s$ and $\hat{\mathfrak{F}}_s$ associated with n_s and $\hat{n}_s$. Indeed, they are most simply defined by relating their generating functions.[11] With the above notation, one defines

$$\sum_1^\infty \frac{1}{s!}\int \hat{\mathfrak{F}}_s(\mathbf{y}^s, [\gamma]) \prod_1^s \lambda(\mathbf{y}_i)\, d\mathbf{y}^s$$
$$= \ln \sum_0^\infty \frac{1}{s!}\int \hat{n}_s(\mathbf{y}^s, [\gamma]) \prod_1^s \lambda(\mathbf{y}_i)\, d\mathbf{y}^s$$
$$= \ln \Xi[\gamma + \lambda] - \ln \Xi[\gamma], \tag{2.16}$$

$$\sum_1^\infty \frac{1}{s!}\int \mathfrak{F}_s(\mathbf{y}^s, [\gamma]) \prod_1^s \Delta e^{\gamma(\mathbf{y}_i)}\, d\mathbf{y}^s$$
$$= \ln \sum_0^\infty \frac{1}{s!}\int n_s(\mathbf{y}^s, [\gamma]) \prod_1^s \Delta e^{\gamma(\mathbf{y}_i)}\, d\mathbf{y}^s$$
$$= \ln \Xi[\gamma + \Delta\gamma] - \ln \Xi[\gamma]. \tag{2.17}$$

By expansion, the relations

$$\mathfrak{F}_1(\mathbf{y}) = n_1(\mathbf{y}),$$
$$\mathfrak{F}_2(\mathbf{y}_1, \mathbf{y}_2) = n_2(\mathbf{y}_1, \mathbf{y}_2) - n_1(\mathbf{y}_1)n_1(\mathbf{y}_2), \tag{2.18}$$
$$\mathfrak{F}_3(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) = n_3(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) - n_2(\mathbf{y}_1, \mathbf{y}_2)n_1(\mathbf{y}_3)$$
$$- n_2(\mathbf{y}_2, \mathbf{y}_3)n_1(\mathbf{y}_1) - n_2(\mathbf{y}_3, \mathbf{y}_1)n_1(\mathbf{y}_2)$$
$$+ 2n_1(\mathbf{y}_1)n_1(\mathbf{y}_2)n_1(\mathbf{y}_3), \cdots$$

are obtained, and precisely the same relations hold between the $\hat{\mathfrak{F}}_s$ and $\hat{n}_s$. As a consequence of (2.16) and (2.17), we now have

$$\hat{\mathfrak{F}}_s(\mathbf{y}^s, [\gamma]) = \frac{\delta^s \ln \Xi[\gamma]}{\delta\gamma(\mathbf{y}_1)\cdots\delta\gamma(\mathbf{y}_s)},$$
$$\mathfrak{F}_s(\mathbf{y}^s, [\gamma]) = \prod_1^s e^{\gamma(\mathbf{y}_i)} \frac{\delta^s \ln \Xi[\gamma]}{\delta e^{\gamma(\mathbf{y}_1)}\cdots\delta e^{\gamma(\mathbf{y}_s)}}. \tag{2.19}$$

The characteristic property of the Ursell distributions is that they vanish whenever their arguments decompose into two independent sets. To see this, suppose that the total volume Ω is divided into two volumes $\Omega = \Omega_1 + \Omega_2$ such that

$$\hat{n}_{a+b}(\mathbf{y}_1, \cdots \mathbf{y}_a, \mathbf{y}_{a+1}, \cdots \mathbf{y}_{a+b}) = \hat{n}_a(\mathbf{y}_1, \cdots \mathbf{y}_a)\hat{n}_b(\mathbf{y}_{a+1}, \cdots \mathbf{y}_b),$$

when $\mathbf{y}_1, \cdots \mathbf{y}_a$ are in Ω_1, $\mathbf{y}_{a+1}, \cdots \mathbf{y}_b$ in Ω_2 (and similarly under permutations of $\hat{n}_{a+b}$). Then since

$$\int \hat{n}_s(\mathbf{y}^s)\lambda(\mathbf{y}_1)\cdots\lambda(\mathbf{y}_s)\, d\tau(\mathbf{y}_1)\cdots d\tau(\mathbf{y}_s) = \sum_a \binom{s}{a}$$
$$\times \int \hat{n}_a(\mathbf{y}^a)\lambda(\mathbf{y}_1)\cdots\lambda(\mathbf{y}_a)\, d\tau_1(\mathbf{y}_1)\cdots d\tau_1(\mathbf{y}_a)$$
$$\times \int \hat{n}_{s-a}(\mathbf{y}_{a+1}, \cdots, \mathbf{y}_s)$$
$$\times \lambda(\mathbf{y}_{a+1})\cdots\lambda(\mathbf{y}_s)\, d\tau_2(\mathbf{y}_{a+1})\cdots d\tau_2(\mathbf{y}_s), \tag{2.20}$$

it follows that

$$\ln \Xi[\gamma + \lambda] - \ln \Xi[\gamma] = \ln \sum \frac{1}{a!}$$
$$\times \int \hat{n}_a(\mathbf{y}^a)\lambda(\mathbf{y}_1)\cdots\lambda(\mathbf{y}_a)\, d\tau_1(\mathbf{y}_1)\cdots d\tau_1(\mathbf{y}_a)$$
$$+ \ln \sum \frac{1}{b!}\int \hat{n}_b(\mathbf{y}^b)$$
$$\times \lambda(\mathbf{y}_1)\cdots\lambda(\mathbf{y}_b)\, d\tau_2(\mathbf{y}_1)\cdots d\tau_2(\mathbf{y}_b) \tag{2.21}$$

has no component $\hat{\mathfrak{F}}_s(\mathbf{y}^s)$ in which some particles are in Ω_1 and the others in Ω_2. The same of course holds for $\mathfrak{F}_s(\mathbf{y}^s)$.

III. BASIC INTEGRAL EQUATIONS

If one holds s particles fixed, the k-particle distribution becomes a conditional $(k + s)$-particle distribution. In a classical grand ensemble, particles may be fixed by placing their force fields at fixed points. Thus, the higher-order distributions are related to lower-order distributions with external potentials, a relation upon which many developments in classical statistical mechanics are predicated.

To be explicit, and somewhat pedantic, we have from (2.6),

$$n_s(\mathbf{y}^s, [\gamma]) = \Xi[\gamma]^{-1} \prod e^{\gamma(\mathbf{y}_i)}$$
$$\times \sum \frac{1}{N!}\int e_N(\mathbf{r}^N) \prod e^{\gamma(\mathbf{r}_i)} \frac{e_{N+s}(\mathbf{r}^N\mathbf{y}^s)}{e_N(\mathbf{r}^N)}\, d\mathbf{r}^N, \tag{3.1}$$

but for two-body forces,

$$e_{N+s}(\mathbf{r}^N\mathbf{y}^s)/e_N(\mathbf{r}^N) = e_s(\mathbf{y}^s) \exp\left[-\beta \sum \phi(\mathbf{r}_i, \mathbf{y}_j)\right].$$

Hence,

$$n_s(\mathbf{y}^s, [\gamma]) = e_s(\mathbf{y}^s) \prod_1^s e^{\gamma(\mathbf{y}_i)} \Xi\left[\gamma - \beta \sum_1^s \phi_{\mathbf{y}_i}\right] \Big/ \Xi[\gamma], \tag{3.2}$$

where

$$\phi_{\mathbf{y}}(\mathbf{r}) \equiv \phi(\mathbf{r}, \mathbf{y}).$$

By virtue of (2.15), Eq. (3.2) yields the functional differential equation

$$\delta^s \Xi[\gamma]/\delta e^{\gamma(\mathbf{y}_1)}\cdots\delta e^{\gamma(\mathbf{y}_s)} = e_s(\mathbf{y}^s)\Xi\left[\gamma - \beta \sum_1^s \phi_{\mathbf{y}_i}\right], \tag{3.3}$$

which is our literal starting point. If (3.3) is differentiated k times, it generalizes to

[10] See, for example, J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), p. 137.

[11] See, for example, J. L. Lebowitz and J. K. Percus, Phys. Rev. **122**, 1675 (1961).

$$\frac{\delta^{k+s}\Xi[\gamma]}{\delta e^{\gamma(\mathbf{x}_1)}\cdots\delta e^{\gamma(\mathbf{x}_k)}\delta e^{\gamma(\mathbf{y}_1)}\cdots\delta e^{\gamma(\mathbf{y}_s)}} = \frac{e_{k+s}(\mathbf{y}^s\mathbf{x}^k)}{e_k(\mathbf{x}^k)}\frac{\delta^k\Xi\left[\gamma-\beta\sum_1^s\phi_{\mathbf{y}_j}\right]}{\delta e^{\gamma(\mathbf{x}_1)-\beta\sum\phi(\mathbf{x}_1\mathbf{y}_j)}\cdots\delta e^{\gamma(\mathbf{x}_k)-\beta\sum\phi(\mathbf{x}_k\mathbf{y}_j)}}, \tag{3.4}$$

where we have made use of the relation

$$e_s(\mathbf{y}^s)\exp\left[-\beta\sum\phi(\mathbf{y}_j,\mathbf{x}_i)\right] = e_{s+k}(\mathbf{y}^s\mathbf{x}^k)/e_k(\mathbf{x}^k).$$

Equation (3.4), in fact, when written as[12]

$$\frac{n_{k+s}(\mathbf{x}^k\mathbf{y}^s,[\gamma])}{n_s(\mathbf{y}^s,[\gamma])} = n_k\left(\mathbf{x}^k,\left[\gamma-\beta\sum_1^s\phi_{\mathbf{y}_j}\right]\right), \tag{3.5}$$

is precisely the conditional distribution relation alluded to above.

We now wish to relate distributions of different orders for the same system. This requires elimination of the external potential which enters into such as (3.5) and may be accomplished by turning on the external potential in a series expansion of the form (1.3). To this end, we set

$$\begin{aligned} e^{\gamma(\mathbf{r}|0)} &= z \\ e^{\gamma(\mathbf{r}|\alpha)} &\equiv z\left\{1+\alpha\left(\exp\left[-\beta\sum_1^s\phi(\mathbf{y}_i,\mathbf{r})\right]-1\right)\right\} \\ &\equiv z\{1+\alpha f(\mathbf{y}^s;\mathbf{r})\} \\ e^{\gamma(\mathbf{r}|1)} &\equiv z\exp\left[-\beta\sum_1^s\phi(\mathbf{y}_i,\mathbf{r})\right], \end{aligned} \tag{3.6}$$

$f(\mathbf{y}^s;\mathbf{r})$ being a generalized Mayer f function, and follow the corresponding transition of a quantity we shall refer to as $G_{k,s}(\alpha)$:

$$\begin{aligned} G_{k,s}(\mathbf{r}^k\mathbf{y}^s|0) &= \Xi(0)z^{-k}n_k(\mathbf{r}^k), \\ G_{k,s}(\mathbf{r}^k\mathbf{y}^s|\alpha) &\equiv \delta^k\Xi(\alpha)/\delta e^{\gamma(\mathbf{r}_1|\alpha)}\cdots\delta e^{\gamma(\mathbf{r}_k|\alpha)}, \\ &= \Xi(\alpha)\prod_1^k e^{-\gamma(\mathbf{r}_i|\alpha)}n_k(\mathbf{r}^k|\alpha), \\ G_{k,s}(\mathbf{r}^k\mathbf{y}^s|1) &= \Xi(0)z^{-(k+s)}n_{k+s}(\mathbf{r}^k\mathbf{y}^s)e_k(\mathbf{r}^k)/e_{k+s}(\mathbf{r}^k\mathbf{y}^s). \end{aligned} \tag{3.7}$$

Here $\Xi(\alpha)$ denotes $\Xi[\gamma(\ |\alpha)]$, etc.

Direct application of (1.3) now yields

$$\begin{aligned} G_{k,s}(1) &= \sum_{i=0}^{\ell}\frac{1}{j!}\int\frac{\delta^jG_{k,s}(\alpha)}{\delta e^{\gamma(\mathbf{r}_1)}\cdots\delta e^{\gamma(\mathbf{r}_j)}}\Big|_{\alpha=0} \\ &\times\prod_1^j[zf(\mathbf{y}^s;\mathbf{r}_i)]\,d\mathbf{r}^j \\ &+\int_0^1\int\frac{(1-\alpha)^\ell}{\ell!}\frac{\delta^{\ell+1}G_{k,s}(\alpha)}{\delta\exp[\gamma(\mathbf{r}_1|\alpha)]\cdots\delta\exp[\gamma(\mathbf{r}_{\ell+1}|\alpha)]} \\ &\times\prod_1^{\ell+1}[zf(\mathbf{y}^s;\mathbf{r}_i)]\,d\mathbf{r}^{\ell+1}\,d\alpha, \end{aligned} \tag{3.8}$$

[12] See, e.g., J. K. Percus, Phys. Rev. Letters 8, 462 (1962).

or employing

$$\delta^jG_{k,s}(\mathbf{x}^k\mathbf{y}^s|\alpha)/\delta e^{\gamma(\mathbf{r}_1|\alpha)}\cdots\delta e^{\gamma(\mathbf{r}_j|\alpha)} = G_{k+j,s}(\mathbf{x}^k\mathbf{r}^j,\mathbf{y}^s|\alpha) \tag{3.9}$$

and the definition of $G_{k,s}$,

$$\begin{aligned} \frac{n_{k+s}(\mathbf{x}^k\mathbf{y}^s)e_k(\mathbf{x}^k)}{e_{k+s}(\mathbf{x}^k\mathbf{y}^s)z^s} &= \sum_{i=0}^{\ell}\frac{1}{j!}\int n_{k+j}(\mathbf{x}^k\mathbf{r}^j)\prod_{i=1}^j f(\mathbf{y}^s;\mathbf{r}_i)\,d\mathbf{r}^j \\ &+ z^{\ell+1}\int_0^1\frac{(1-\alpha)^\ell}{\ell!}\frac{\Xi(\alpha)}{\Xi(0)}\prod_1^k e^{-\gamma(\mathbf{x}_i|\alpha)} \\ &\times\left\{\int\prod_1^{\ell+1}[f(\mathbf{y}^s;\mathbf{r}_i)e^{-\gamma(\mathbf{r}_i|\alpha)}]n_{k+\ell+1}(\mathbf{x}^k\mathbf{r}^{\ell+1}|\alpha)\,d\mathbf{r}^{\ell+1}\right\}d\alpha \\ &= N_{k,s}^{(\ell)}+R_{k,s}^{(\ell)}. \end{aligned} \tag{3.10}$$

If this series converges, so that the remainder term $R_{k,s}^{(\ell)}$ vanishes when ℓ goes to infinity, then for any choice of $k\geq 0$, we get a set of recursive equations for the distributions *in the original system without external potential*, as s is varied. Similarly, for any choice of $s\geq 1$ (the choice $s=0$ leads to an identity), we obtain a set of equations for different k. In particular, the choice $k=0$ recovers the Mayer–Montroll equations, while $s=1$ gives the Kirkwood–Salsburg equations. Equation (3.10), with $\ell=\infty$, is equivalent to Eq. (54′) of Mayer.[6]

A different but equivalent set of integral equations may be obtained by instead expanding $G_{k,s}(0)$ about its value at $\alpha=1$ [corresponding to setting $\alpha=1-\delta$ in (3.6) and expanding about $\delta=0$]. This procedure leads to the equations

$$\begin{aligned} z^sn_k(\mathbf{x}^k) &= \sum_{i=0}^{\ell}\frac{(-1)^i}{j!}\int n_{k+s+j}(\mathbf{x}^k\mathbf{y}^s\mathbf{r}^j)\prod_{i=1}^j f(\mathbf{y}^s;\mathbf{r}_i) \\ &\times e_{k+j}(\mathbf{x}^k\mathbf{r}^j)/e_{k+s+j}(\mathbf{x}^k\mathbf{y}^s\mathbf{r}^j)\,d\mathbf{r}^j \\ &+\frac{z^{k+\ell+s+1}(-1)^{\ell+1}}{\Xi(0)}\int\prod_1^{\ell+1}f(\mathbf{y}^s;\mathbf{r}_i) \\ &\times\int_0^1 G_{k+\ell+1,s}(\mathbf{x}^k\mathbf{r}^{\ell+1}\mathbf{y}^s|\alpha)\frac{\alpha^\ell}{\ell!}\,d\alpha\,d\mathbf{r}^{\ell+1}. \end{aligned} \tag{3.11}$$

It should be noted that the left-hand side of (3.11) contains, in contrast to (3.10), z raised to a positive power. Eq. (3.11) is formally the inverse or "solution" of (3.10).

The fugacity (z) expansion[3] implicit in (3.10) may be avoided if we consider instead the expansion of the function

$$\tilde{G}_{k,s}(\alpha) = \prod_{i=1}^{k} e^{-\gamma(\mathbf{r}_i|\alpha)} \mathfrak{F}_k(\mathbf{r}^k|\alpha) = \frac{\delta^k \ln \Xi(\alpha)}{\delta e^{\gamma(\mathbf{r}_1|\alpha)} \cdots \delta e^{\gamma(\mathbf{r}_k|\alpha)}}. \tag{3.12}$$

For $k = 0$, $\tilde{G}_{0,s}(\alpha) = \ln \Xi(\alpha)$, we find

$$\ln\left[\frac{n_s(\mathbf{y}^s)}{z^s e_s(\mathbf{y}^s)}\right] = \sum_{j=0}^{\ell} \frac{1}{j!} \int \mathfrak{F}_j(\mathbf{r}^j) \prod_{i=1}^{j} f(\mathbf{y}^s; \mathbf{r}_i)\, d\mathbf{r}^j + z^{\ell+1} \int_0^1 \frac{(1-\alpha)^\ell}{\ell!} \times \int \tilde{G}_{\ell+1,s}(\mathbf{r}^{\ell+1}\mathbf{y}^s|\alpha) \prod_{i=1}^{\ell+1} f(\mathbf{y}^s; \mathbf{r}_i)\, d\mathbf{r}^j\, d\alpha, \tag{3.13}$$

while for $k \geq 1$,

$$\tilde{F}_{k/s}(\mathbf{x}^k; \mathbf{y}^s) \frac{e_s(\mathbf{y}^s) e_k(\mathbf{x}^k)}{e_{s+k}(\mathbf{y}^s\mathbf{x}^k)} = \sum_{0}^{\ell} \frac{1}{j!} \int \mathfrak{F}_{k+j}(\mathbf{x}^k\mathbf{r}^j) \prod_{1}^{j} f(\mathbf{y}^s; \mathbf{r}_i)\, d\mathbf{r}^j + z^{k+\ell+1} \int_0^1 \frac{(1-\alpha)^\ell}{\ell!} \int \tilde{G}_{k+\ell+1,s}(\mathbf{x}^k\mathbf{r}^{\ell+1}\mathbf{y}^s|\alpha) \times \prod_{1}^{\ell+1} f(\mathbf{y}^s; \mathbf{r}_i)\, d\mathbf{r}^j\, d\alpha. \tag{3.14}$$

Here $\tilde{F}_{k/s}(\mathbf{x}^k; \mathbf{y}^s)$ is the Ursell function for k particles when s particles are fixed, and corresponds to replacing the densities $n_i(\mathbf{x}^i)$ appearing in $\mathfrak{F}_k$ by the conditional densities $n_{i+s}(\mathbf{x}^i\mathbf{y}^s)/n_s(\mathbf{y}^s)$. Thus,

$$\tilde{F}_{1/1}(\mathbf{x}; \mathbf{y}) = n_2(\mathbf{x}\mathbf{y})/n_1(\mathbf{y}), \tag{3.15}$$

$$\tilde{F}_{2/1}(\mathbf{x}_1\mathbf{x}_2; \mathbf{y}) = n_3(\mathbf{x}_1\mathbf{x}_2\mathbf{y})/n_1(\mathbf{y}) - n_2(\mathbf{x}_1\mathbf{y})n_2(\mathbf{x}_2\mathbf{y})/n_1(\mathbf{y})^2.$$

Equation (3.14) may now be used to develop a virial expansion in the density for the Ursell functions in a manner similar to the fugacity expansion obtainable from (3.10).[3] Unfortunately, however, the left-hand side of (3.14) is not linear in the F's—see, e.g., (3.15)—and this will lead to quite complicated recurrence relations for the coefficients in the density expansion. For completeness, we indicate how the relation between $\mathfrak{F}_k$ and $\tilde{F}_{k/1}$ may be obtained. If, for a test function $g(\mathbf{x})$, we set

$$\mathfrak{F}_k = \int \mathfrak{F}_k(\mathbf{x}^k) \prod_{1}^{k} g(\mathbf{x}_i)\, d\mathbf{x}^k/k!,$$

$$n_k = \int n_k(\mathbf{x}^k) \prod_{1}^{k} g(\mathbf{x}_i)\, d\mathbf{x}^k/k!,$$

$$\tilde{F}_{k/1}(\mathbf{y}) = \int \tilde{F}_{k/1}(\mathbf{x}^k\mathbf{y}) \prod_{1}^{k} g(\mathbf{x}_i)\, d\mathbf{x}^k/k!, \tag{3.16}$$

$$n_{k/1}(\mathbf{y}) = \int n_{k+1}(\mathbf{x}^k\mathbf{y}) \prod_{1}^{k} g(\mathbf{x}_i)\, d\mathbf{x}^k/(k!n_1(\mathbf{y})),$$

then from the generating function relation (2.17), we have

$$\sum_{1}^{\infty} \mathfrak{F}_k = \ln\left(1 + \sum_{1}^{\infty} n_k\right) = \ln\left[1 + \left(\int g(x)\, \delta/\delta g(x)\, dx\right)^{-1} \sum_{1}^{\infty} k n_k\right] = \ln\left[1 + \left(\int g\, \delta/\delta g\right)^{-1} \cdot \int g(\mathbf{y}) n_1(\mathbf{y}) \sum_{0}^{\infty} n_{k/1}(\mathbf{y})\, d\mathbf{y}\right],$$

so that

$$\sum_{1}^{\infty} \mathfrak{F}_k = \ln\left[1 + \left(\int g\, \delta/\delta g\right)^{-1} \int g(\mathbf{y}) n_1(\mathbf{y}) \times \exp\left(\sum_{1}^{\infty} \tilde{F}_{k/1}(\mathbf{y})\right) d\mathbf{y}\right], \tag{3.17}$$

which can be expanded out. Equation (3.13) and the first of Eqs. (3.14), without a remainder term, were also obtained, using a different method, by Green.[7]

IV. INEQUALITIES FOR POSITIVE POTENTIAL

We have already indicated that (3.10) may be used, when $\ell \to \infty$, to obtain series expansions of the distribution functions in powers of z. The purpose of this section is to obtain rigorous inequalities on the distribution functions, valid for all values of z, by securing upper and lower bounds on the remainder term $R_{k,s}^{(\ell)}$ in (3.10). We repeat for convenience

$$\frac{n_{k+s}(\mathbf{x}^k\mathbf{y}^s)e_k(\mathbf{x}^k)}{e_{k+s}(\mathbf{x}^k\mathbf{y}^s)z^s} = N_{k,s}^{(\ell)} + R_{k,s}^{(\ell)}. \tag{4.1}$$

It is simplest to consider first the case of positive potential, $\phi(\mathbf{y}, \mathbf{r}) > 0$, so that $-1 \leq f(\mathbf{y}^s; \mathbf{r}) \leq 0$. This was first studied by a very different method by Lieb,[1] who obtained the bounds given in Eq. (4.10). In this case, since $n_{k+\ell+1}$ is always nonnegative, the remainder $R_{k,s}^{(\ell)}$ has the sign $(-1)^{\ell+1}$. Thus, $R_{k,s}^{(\ell)} \geq 0$ for ℓ odd, $R_{k,s}^{(\ell)} \leq 0$ for ℓ even, and we have at once the result

$$\frac{n_{k+s}(\mathbf{x}^k\mathbf{y}^s)e_k(\mathbf{x}^k)}{e_{k+s}(\mathbf{x}^k\mathbf{y}^s)z^s} \begin{cases} \geq \\ \leq \end{cases} \begin{cases} N_{k,s}^{(\ell)}(\mathbf{x}^k\mathbf{y}^s) & \ell \text{ odd} \\ & \ell \text{ even} \end{cases}, \tag{4.2}$$

with eventual steady decrease in the interval hemmed in by successive bounds, if the series $N_{k,s}^{(\ell)}$ converges as $\ell \to \infty$, but rigorous bounds under all circumstances.

A stronger inequality for the same value of ℓ is obtained by noting that $G_{m,n}(\alpha)$ is a monotonically decreasing function of α, for $\phi \geq 0$, since

$$\partial G_{m,n}(\mathbf{x}^m\mathbf{y}^n|\alpha)/\partial\alpha = z \int f(\mathbf{y}^n; \mathbf{r}) G_{m+1,n}(\mathbf{x}^m\mathbf{r}\mathbf{y}^n|\alpha)\, d\mathbf{r} \leq 0 \quad \text{for} \quad f \leq 0. \tag{4.3}$$

Thus, $0 \leq G_{m,n}(1) \leq G_{m,n}(\alpha) \leq G_{m,n}(0)$. Using $G_{m,n}(1)$ as a lower bound in the remainder, (4.2) is then strengthened to

$$R^{(\ell)}_{k,s}(x^k y^s) \left.\begin{matrix}\geq \\ \leq\end\right\} \frac{1}{\ell+1!}\frac{1}{z^s}\int n_{k+\ell+s+1}(\mathbf{x}^k\mathbf{r}^{\ell+1}\mathbf{y}^s) \times \frac{e_{k+\ell+1}(\mathbf{x}^k\mathbf{r}^{\ell+1})}{e_{k+\ell+s+1}(\mathbf{x}^k\mathbf{r}^{\ell+1}\mathbf{y}^s)} \prod_1^{\ell+1} f(\mathbf{y}^s;\mathbf{r}_i)\,d\mathbf{r}^{\ell+1} \quad \begin{matrix}\ell \text{ odd} \\ \ell \text{ even}\end{matrix}\Big\}. \tag{4.4}$$

On the other hand, using $G_{m,n}(0)$ as an upper bound succeeds only in converting the inequality for ℓ into the original one for $\ell+1$. Still stronger inequalities can be obtained by noting that $G(\alpha)$ is convex, since successive derivatives with respect to α alternate in sign.

By setting $e^{\gamma(\mathbf{r}|0)}$ in (3.6) equal to z, we have implicitly removed any external potential from the interior of our system. Thus, if $\phi(\mathbf{y},\mathbf{r}) = \phi(\mathbf{y}-\mathbf{r})$ and we adopt periodic boundary conditions, the system will be uniform and the constant density $\rho = n_1(\mathbf{r})$ one of its thermodynamic parameters. We, henceforth, restrict our attention to uniform systems (although this restriction is for most of our purposes inessential). We now proceed to obtain actual bounds upon the distributions and not merely relations between them. As a prototype, (4.2) may be truncated at $\ell = 0$ to yield

$$\frac{n_{k+s}(\mathbf{x}^k\mathbf{y}^s)}{e_{k+s}(\mathbf{x}^k\mathbf{y}^s)} \leq z^s \frac{n_k(\mathbf{y}^k)}{e_k(\mathbf{y}^k)}, \tag{4.5}$$

and, in particular,

$$n_s/e_s \leq \rho z^{s-1} \leq z^s. \tag{4.5'}$$

The last form of the inequality was obtained originally by Groeneveld.[2]

Successively finer inequalities now involve truncating at higher ℓ and eliminating all distributions but one. The process may, however, be carried out in numerous ways. Consider first the Kirkwood–Salsburg sequence $s = 1$, $k = 0, 1, 2, \cdots$. We have from (4.4),

$$\frac{\rho}{z} \leq 1 + z\int f(\mathbf{r})\frac{n_2(\mathbf{r})}{z^2}e^{\beta\phi(\mathbf{r})}\,d\mathbf{r} \leq 1, \tag{4.6a}$$

$$\frac{\rho}{z} \geq 1 + z\int \frac{\rho}{z} f(\mathbf{r})\,d\mathbf{r} + \frac{z^2}{2!}\int f(\mathbf{r}_1-\mathbf{y})f(\mathbf{r}_2-\mathbf{y}) \times \frac{n_3(\mathbf{r}_1\mathbf{r}_2\mathbf{y})}{z^3}\frac{e_2(\mathbf{r}_1\mathbf{r}_2)}{e_3(\mathbf{r}_1\mathbf{r}_2\mathbf{y})}\,d\mathbf{r}_1\,d\mathbf{r}_2 \geq 1 + \rho\int f(\mathbf{r})\,d\mathbf{r}, \tag{4.6b}$$

$$\frac{\rho}{z} \leq 1 + z\int \frac{\rho}{z} f(\mathbf{r})\,d\mathbf{r} + \frac{z^2}{2!}\int f(\mathbf{r}_1-\mathbf{y})f(\mathbf{r}_2-\mathbf{y})\frac{n_2(\mathbf{r}_1\mathbf{r}_2)}{z^2}\,d\mathbf{r}_1\,d\mathbf{r}_2 + \frac{z^3}{3!}\int fff \frac{n_4}{z^4}\frac{e_3}{e_4}\,d\mathbf{r}_1\,d\mathbf{r}_2\,d\mathbf{r}_3 \leq 1 + \rho\int f(\mathbf{r})\,d\mathbf{r} + \frac{z^2}{2!}\int f(\mathbf{r}_1-\mathbf{y})f(\mathbf{r}_2-\mathbf{y})\frac{n_2(\mathbf{r}_1,\mathbf{r}_2)}{z^2}\,d\mathbf{r}_1\,d\mathbf{r}_2, \tag{4.6c}$$

$$\frac{n_2(\mathbf{xy})}{z^2}e^{\beta\phi(\mathbf{x}-\mathbf{y})} \leq \frac{\rho}{z} + z\int f(\mathbf{r}-\mathbf{y})\frac{n_3(\mathbf{xry})}{z^3}\frac{e_2(\mathbf{xr})}{e_3(\mathbf{xry})}\,d\mathbf{r} \leq \rho/z, \tag{4.7a}$$

$$\frac{n_2(\mathbf{xy})}{z^2}e^{\beta\varphi} \geq \frac{\rho}{z} + z\int f(\mathbf{r}-\mathbf{y})\frac{n_2(\mathbf{xr})}{z^2}\,d\mathbf{r} + \frac{z^2}{2!}\int f(\mathbf{r}_1-\mathbf{y})f(\mathbf{r}_2-\mathbf{y})\frac{n_4(\mathbf{xr}_1\mathbf{r}_2\mathbf{y})}{z^4}\frac{e_3(\mathbf{xr}_1\mathbf{r}_2)}{e_4(\mathbf{xr}_1\mathbf{r}_2\mathbf{y})}\,d\mathbf{r}_1\,d\mathbf{r}_2 \geq \frac{\rho}{z} + z\int f(\mathbf{r}-\mathbf{y})\frac{n_2(\mathbf{x},\mathbf{r})}{z^2}\,d\mathbf{r}, \tag{4.7b}$$

$$\frac{n_3(\mathbf{x}_1\mathbf{x}_2\mathbf{y})}{z^3}e^{\beta[\phi(\mathbf{x}_1-\mathbf{y})+\phi(\mathbf{x}_2-\mathbf{y})]} \leq \frac{n_2(\mathbf{x}_1\mathbf{x}_2)}{z^2} + z\int f(\mathbf{r}-\mathbf{y})\frac{n_4(\mathbf{x}_1\mathbf{x}_2\mathbf{ry})}{z^4}\frac{e_3(\mathbf{x}_1\mathbf{x}_2\mathbf{r})}{e_4(\mathbf{x}_1\mathbf{x}_2\mathbf{ry})}\,d\mathbf{r} \leq n_2(\mathbf{x}_1,x_2)/z^2. \tag{4.8}$$

We can now also obtain bounds on ρ/z, n_2/z^2, etc., which do not contain any distribution functions but only ρ and z and explicitly known quantities. Perhaps the simplest consistent (but not necessarily "best") manner of doing this is to *consider only the second (weaker) inequalities* in Eqs. (4.6)–(4.8). Then we utilize (4.7*a*) to eliminate n_2 from (4.6c), (4.7b), and (4.8c). This yields (4.9c), (4.15b), and (4.16), ... etc., —the key to the maintenance of the inequality in each case being that $f < 0$. We thus find

$$\rho \leq z, \qquad \rho \geq z + \rho z\int f(\mathbf{r})\,d\mathbf{r}, \tag{4.9a, b}$$

$$\rho \leq z + \rho\Big\{z\int f(\mathbf{r})\,d\mathbf{r} + \frac{z^2}{2!}\int e^{-\beta\varphi(\mathbf{r})}f(\mathbf{y})f(\mathbf{r}-\mathbf{y})\,d\mathbf{r}\,d\mathbf{y}\Big\}. \tag{4.9c}$$

This process was carried out in detail by Penrose,[13] and leads to a general set of inequalities, Lieb's inequalities are of the form

$$\rho \left.\begin{matrix}\leq \\ \geq\end\right\} z + \rho\sum_{i=1}^{\ell} a_i z^i \quad \begin{matrix}\ell \text{ even} \\ \ell \text{ odd}\end{matrix}\Big\}, \tag{4.10}$$

which, as long as $\sum_1^\ell a_i z^i < 1$, can be written as

[13] O. Penrose, J. Math. Phys. 4, 1488 (1963) (previous paper).

$$\rho \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} z \Big/ \Big(1 - \sum_1^{\ell} a_i z^i\Big). \tag{4.11}$$

The coefficients a_i in (4.10) and (4.12) may be found in terms of the more customary quantities b_k occurring in the fugacity expansions of pressure and density,

$$\begin{aligned} \beta p &= \sum_1^{\infty} b_k z^k, \\ \rho &= \sum_1^{\infty} k b_k z^k. \end{aligned} \tag{4.12}$$

Since the a_i do not depend[13] upon ℓ, one need only take the $\ell = \infty$ limit of (4.11) and equate coefficients:

$$z \Big/ \Big(1 - \sum_1^{\infty} a_i z^i\Big) = \sum_1^{\infty} k b_k z^k. \tag{4.13}$$

Making use of a well-known formula[14] for the power series representing the quotient of two power series, one finds

$$a_i = (-1)^{i+1} \times \begin{vmatrix} 2b_2 & 1 & 0 & \cdots & 0 \\ 3b_3 & 2b_2 & 1 & \cdots & 0 \\ \cdot & 3b_3 & 2b_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & 1 \\ (j+1)b_{j+1} & jb_j & (j-1)b_{j-1} & & 2b_2 \end{vmatrix}. \tag{4.14}$$

The same process which led to (4.9), leads directly to the analogous sequence

$$n_2(\mathbf{xy})/z^2 \leq (\rho/z) e^{-\beta\phi(\mathbf{x}-\mathbf{y})}, \tag{4.15a}$$

$$\frac{n_2(\mathbf{x}, \mathbf{y})}{z^2} \geq \frac{\rho}{z}\Big[e^{-\beta\phi(\mathbf{x}-\mathbf{y})} + z\int e^{-\beta\phi(\mathbf{x}-\mathbf{r})} f(\mathbf{r}-\mathbf{y})\, d\mathbf{r}\Big], \cdots, \tag{4.15b}$$

$$\frac{n_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})}{z^3} \leq \rho e_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}), \cdots. \tag{4.16}$$

In general, one derives, in this fashion, inequalities of the form[13]

$$n_k(\mathbf{x}^k) \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \rho z^{k-1} \sum_{j=0}^{\ell} a_{k,j}(\mathbf{x}^k) z^j \quad \begin{matrix} \ell \text{ even} \\ \ell \text{ odd} \end{matrix}, \tag{4.17}$$

and the $a_{k,j}$ can again be obtained by letting $\ell \to \infty$, then equating (4.17) to the usual fugacity expansion of n_k:

$$n_k(\mathbf{x}^k) = \rho z^{k-1} \sum_0^{\infty} a_{k,j}(\mathbf{x}^k) z^j = z^k \sum_0^{\infty} n_{k,j}(\mathbf{x}^k) z^j,$$

or

$$\sum_0^{\infty} a_{k,j}(\mathbf{x}^k) z^j = \sum_0^{\infty} n_{k,j}(\mathbf{x}^k) z^j \Big/ \sum_0^{\infty} (j+1) b_{j+1} z^j. \tag{4.18}$$

A somewhat simpler set of inequalities, which however are not always as strong as (4.11) and (4.17), is available from the Mayer–Montroll equations, i.e., $k = 1$ in (4.1). For this purpose, we iterate as before, but also eliminate $\rho = n_1(\mathbf{r})$ in terms of z and f. This yields directly[13]

$$n_k(\mathbf{x}^k) \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} z^k \sum_0^{\ell} n_{k,j}(\mathbf{x}^k) z^j \quad \begin{matrix} \ell \text{ even} \\ \ell \text{ odd} \end{matrix}, \quad k \geq 1. \tag{4.19}$$

In particular, for $k = 1$, since $p = \int (\rho/z)\, dz$ we have both

$$\rho \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \sum_0^{\ell-1} j b_j z^j \quad \begin{matrix} \ell \text{ odd} \\ \ell \text{ even} \end{matrix},$$

and

$$\beta p \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \sum_0^{\ell-1} b_j z^j. \quad \begin{matrix} \ell \text{ odd} \\ \ell \text{ even} \end{matrix} \tag{4.20}$$

V. INEQUALITIES FOR GENERAL POTENTIAL

When the pair potential $\phi(\mathbf{r})$ is not everywhere positive (or zero) we must obtain bounds for both the product of the f's and $n(\alpha)$ in the remainder in (4.1). Separating the potential into a positive and negative part:

$$\begin{aligned} \phi(\mathbf{r}) &= \phi_+(\mathbf{r}) + \phi_-(\mathbf{r}), \\ \phi_+ \geq 0, \quad \phi_- \leq 0, \quad \phi_-\phi_+ &= 0, \end{aligned} \tag{5.1}$$

we can thereby obtain also for f

$$f(\mathbf{y}^s; \mathbf{r}) \equiv f_+(\mathbf{y}^s; \mathbf{r}) - f_-(\mathbf{y}^s; \mathbf{r}), \tag{5.2}$$

where $f_+ \geq 0$, $f_- \geq 0$, $f_+f_- = 0$. Correspondingly, in the factor $\prod f$ in (3.10), we may collect separately the positive and negative terms:

$$\prod_1^{\ell+1} [f_+(\mathbf{y}^s; \mathbf{r}) - f_-(\mathbf{y}^s; \mathbf{r})] = F_+(\mathbf{y}^s; \mathbf{r}^{\ell+1}) - F_-(\mathbf{y}^s; \mathbf{r}^{\ell+1}), \tag{5.3}$$

with $F_+ \geq 0$, $F_- \geq 0$.

To estimate the quantity $n_{k+\ell+1}(\mathbf{x}^k \mathbf{r}^{\ell+1} \mid \alpha)\ \Xi(\alpha)$ in the remainder of (4.1), we recall [Eq. (3.2)] that

$$\Xi[\gamma] \prod_1^{t} e^{-\gamma(\mathbf{x}_i)} n_i(\mathbf{x}^t, [\gamma]) = e_t(\mathbf{x}^t) \Xi\Big[\gamma - \beta \sum_1^{t} \phi_{\mathbf{x}_i}\Big]. \tag{5.4}$$

[14] See, for example, I. S. Gradstein and I. M. Riezhic, *Collection of Formulas* (Four Continent Publishing Company, Inc., New York, 1962).

Now in the present instance, $t = k + \ell + 1$, and in the integrand of $\Xi[\gamma - \beta \sum_1^s \phi_{x_i}]$ we have (with $0 \leq \alpha \leq 1$)

$$\prod_1^N e^{\gamma(\mathbf{r}_i|\alpha)} = z^N \prod_1^N (1 + \alpha f(\mathbf{y}^s; \mathbf{r}_i)) \leq z^N \prod_1^N (1 + f_+(\mathbf{y}^s; \mathbf{r}_i)) = z^N \exp\left[-\beta \sum_{i=1}^N \sum_{i=1}^s \phi_-(\mathbf{y}_i - \mathbf{r}_i)\right]. \tag{5.5}$$

It has been shown by Penrose[3] that when the particles have hard cores and $\phi(\mathbf{r})$ falls off sufficiently rapidly as $r \to \infty$, then in the domain of nonvanishing Boltzmann factor,

$$\sum_{j=1}^N \phi(\mathbf{y} - \mathbf{r}_i) \geq -2\Phi', \qquad N = 1, 2, \cdots, \tag{5.6}$$

where Φ' is some constant. The same argument shows that

$$\sum_{i=1}^N \phi_-(\mathbf{y} - \mathbf{r}_i) \geq -2\Phi'. \tag{5.7}$$

It follows then from (5.4) that

$$\Xi[\gamma] \prod_1^t e^{-\gamma(\mathbf{x}_i)} n_t(\mathbf{x}^t, [\gamma]) \leq \Xi[\ln z] z^{-t} n_t(\mathbf{x}^t, [\ln z]) e^{2\beta s \Phi'}, \tag{5.8}$$

or

$$\left(\frac{\Xi(\alpha)}{\Xi(0)}\right) \prod_1^k e^{-\gamma(\mathbf{x}_i|\alpha)} \prod_1^{\ell+1} e^{-\gamma(\mathbf{r}_i|\alpha)} n_{k+\ell+1}(\mathbf{x}^k \mathbf{r}^{\ell+1}|\alpha) \leq z^{-k-\ell-1} e^{2\beta s \Phi'} n_{k+\ell+1}(\mathbf{x}^k \mathbf{r}^{\ell+1}), \tag{5.8'}$$

and, hence (performing the α integration), that the remainder term of (4.1) is bounded from both sides:

$$R_{k,s}^{(\ell)} \begin{matrix} \leq \\ \geq \end{matrix} \Bigg\} \pm e^{2\beta s \Phi'} \int \frac{n_{k+\ell+1}(\mathbf{x}^k \mathbf{r}^{\ell+1})}{(\ell+1)!} \begin{matrix} F_+(\mathbf{y}^s \mathbf{r}^{\ell+1}) \\ F_-(\mathbf{y}^s \mathbf{r}^{\ell+1}) \end{matrix} \Bigg\} d\mathbf{r}^{\ell+1}, \tag{5.9}$$

F_+ being used for the upper bound, F_- for the lower.

Equations (4.1) and (5.9) may be treated by successive elimination in the fashion of (4.6)–(4.9) to obtain somewhat more complicated bounds on the distribution functions.[13] In the special case $k = 0$, $s = 1$, one has directly

$$\rho/z \leq 1 + \rho e^{2\beta\Phi'} \int f_+(\mathbf{r})\, d\mathbf{r}, \qquad \rho/z \geq 1 - \rho e^{2\beta\Phi'} \int f_-(\mathbf{r})\, d\mathbf{r}. \tag{5.10}$$

For positive potentials, where $\Phi' = 0$, $f_+ = 0$, $f_- = -f$, (5.9) implies the previous (4.2).

It is to be noted that, by the same reasoning used to derive (5.8), one may obtain

$$n_{k+s}(\mathbf{x}^k \mathbf{y}^s) \leq z^s \frac{e_{k+s}(\mathbf{x}^k \mathbf{y}^s)}{e_k(\mathbf{x}^k)} n_s(\mathbf{y}^s) e^{2s\beta\Phi'} \leq z^{k+s} e_{k+s}(\mathbf{x}^k \mathbf{y}^s) e^{2(k+s)\beta\Phi'}. \tag{5.11}$$

The last inequality was first obtained by Groeneveld.[2] As a special case, one has the simple

$$\rho \leq z e^{2\beta\Phi'}. \tag{5.12}$$

VI. APPLICATION TO HARD SPHERES

In this section, we shall evaluate explicitly our rigorous inequalities on the thermodynamic parameters and radial distribution function for a gas of hard spheres of diameter a, and compare them with machine computations and approximate theories.

Consider first the bounds (4.11) and (4.20) for the function $\rho(z)$. For hard spheres of diameter a, the first five virial coefficients are known:

$$B = \tfrac{2}{3}\pi a^3, \quad C = \tfrac{5}{8}B^2, \quad D = 0.2869\, B^3, \quad E = 0.115\, B^4 \tag{6.1}$$

(E is only known to within 15%), thus yielding the corresponding irreducible cluster integrals

$$\beta_1 = -2B, \quad \beta_2 = -\tfrac{15}{16}B^2, \quad \beta_3 = -0.3825\, B^3, \quad \beta_4 = -0.144\, B^4, \tag{6.2}$$

and connected cluster integrals, in units of B,

$$b_2 = -1, \quad b_3 = 1.689, \quad b_4 = -3.555, \quad b_5 = 8.467. \tag{6.3}$$

We note parenthetically the relation of (6.3) to rigorous upper bounds which have been derived[2]: $|b_3| \leq 2$, $|b_4| \leq \frac{16}{3}$, $|b_5| \leq \frac{50}{3}$.

To determine the coefficients a_i, we employ (4.14) in the form

$$1 - \sum_1^\infty a_j z^j = 1 \Big/ \left(1 + \sum_2^\infty b_k z^{k-1}\right), \tag{6.4}$$

and find that,

$$a_1 = -2, \quad a_2 = 1.063, \quad a_3 = -1.950, \quad a_4 = 4.617. \tag{6.5}$$

We, therefore, have the following upper bounds $\rho^{(i)}(z)$ and lower bounds $\rho_{(i)}(z)$ on ρ [from (4.12) and (4.20)]:

$$\rho^{(1)}(z) = z, \quad \rho^{(2)}(z) = z, \quad \rho^{(3)}(z) = z/(1 + 2z - 1.063z^2), \quad \rho^{(4)}(z) = z - 2z^2 + 5.063z^3, \tag{6.6}$$

$$\rho^{(5)}(z) = z/(1 + 2z - 1.063z^2 + 1.950z^3 - 4.617z^4),$$
$$\rho^{(6)}(z) = z - 2z^2 + 5.063z^3 - 14.222z^4 + 42.335z^5,$$
$$\rho_{(1)}(z) = z/(1 + 2z), \qquad \rho_{(2)}(z) = z - 2z^2,$$
$$\rho_{(3)}(z) = z/(1 + 2z - 1.063z^2 + 1.950z^3), \qquad (6.7)$$
$$\rho_{(4)}(z) = z - 2z^2 + 5.063z^3 - 14.222z^4.$$

The fractional expressions are valid until the denominators change sign.

The lowest upper bound of (6.6) and greatest lower bound of (6.7) are plotted in Fig. 1, for the range $0 \leq z \leq 3$. They bracket ρ to within 1% for $z < 0.25$ ($\rho < 0.17$), to within 25% for $z < 1.0$ ($\rho < 0.5$), and thereafter diverge rapidly (close packing occurs at $\rho \sim 3$ in these units). We may use the bounds of (6.6) and (6.7) to test approximate equations of state. The approximation for hard spheres obtained on the basis of thermodynamic arguments by Reiss, Frisch, and Lebowitz,[15] and by Wertheim[16] and Thiele[16] from the exact solution of the Percus–Yevick[16] approximate integral equation for the radial distribution function,

$$\beta_p = \rho[1 + \tfrac{1}{4}\rho + (\tfrac{1}{4}\rho)^2]/(1 - \tfrac{1}{4}\rho)^3, \qquad (6.8)$$

is in very good agreement with machine computations[17] of the pressure over the whole range of "fluid" densities, $\rho < 1.6$. From (6.8), the fugacity z is obtained by means of

$$\ln z = \ln \rho + \int_0^\rho \frac{\partial(\beta p - \rho)}{\partial \rho} \frac{d\rho}{\rho}, \qquad (6.9)$$

or

$$z = \frac{\rho}{1 - \frac{1}{4}\rho} \exp\left\{\frac{\frac{1}{4}\rho}{(1 - \frac{1}{4}\rho)^3}[7 - \tfrac{13}{2}(\tfrac{1}{4}\rho) + \tfrac{5}{2}(\tfrac{1}{4}\rho)^2]\right\}, \qquad (6.10)$$

which, when plotted on the same graph as the bounds, fits snugly in the center for $z < 1$.

For the radial distribution function $g(r, z) = n_2(r, z)/\rho^2$, we find from Eqs. (4.17)–(4.18)

$$\begin{aligned} \rho g(r) &\leq z a_{2,0}(r) = z n_{2,0} = z e^{-\beta\varphi(r)}, \\ \rho g(r) &\geq z a_{2,0}(r) + z^2 a_{2,1}(r) \\ &= (z + 2z^2) n_{2,0}(r) + z^2 n_{2,1}(r) \\ &= e^{-\beta\varphi}[z + (g_1 - 2)z^2], \end{aligned} \qquad (6.11)$$

$$\rho g(r) \leq z a_{2,0}(r) + z^2 a_{2,1}(r) + z^3 a_{2,2}(r)$$
$$= e^{-\beta\varphi(r)}[z + (g_1 - 2)z^2 + (3b_3 - 4g_1 + g_2)z^3],$$

[15] H. Reiss, H. Frisch, and J. Lebowitz, J. Chem. Phys. **31**, 369 (1959).

[16] M. Wertheim, Phys. Rev. Letters **10**, 321 (1963); E. Thiele, J. Chem. Phys. **38**, 1959 (1963); J. K. Percus and G. J. Yevick, Phys. Rev. **100**, 1 (1958).

[17] B. J. Alder and T. A. Wainwright, J. Chem. Phys. **33**, 1439 (1960).

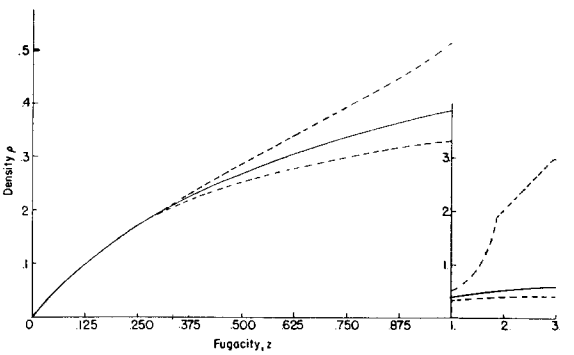


FIG. 1. Bounds on the density. Dotted lines are least upper bound and greatest lower bound for $\rho(z)$ obtained from (6.6) and (6.7); solid line is $\rho(z)$ given by R–F–L and P–Y, Eq. (6.10).

where we have used (4.18) and the relation

$$g(r) = e^{-\beta\varphi(r)}[1 + g_1\rho + g_2\rho^2 + \cdots]. \qquad (6.12)$$

On the other hand, we have from (4.19)

$$\begin{aligned} \rho g(r) &\leq (z/\rho) z n_{2,0}(r), \\ \rho g(r) &\geq (z/\rho)(z n_{2,0}(r) + z^2 n_{2,1}(r)), \end{aligned} \qquad (6.13)$$

which may be completed to ρ-independent inequalities for $\rho g(r)$ by application of (4.20). For a hard-sphere gas, the values of g_k and $n_{2,k}$ for $k = 0, 1, 2$ are known from the work of Nijboer and Van Hove.[18] In Fig. 2, we have plotted the lowest upper bound and greatest lower bound of (6.11) [(6.12) turns out to be inferior in each case] for $z = \frac{1}{4}$ and $z = \frac{1}{2}$ and the value of $\rho g(r)$ obtained by Wertheim,[16] for $z = \frac{1}{4}$, $r \leq 2$.

One can also look, e.g., at (4.6) from the viewpoint of integral inequalities on $\rho g(r)$. Thus, from the first and third equations of (4.6) (dropping the last term in the latter), one has

$$\begin{aligned} 1 &\leq \frac{z}{\rho} + \int f(r) e^{\beta\phi(r)} \rho g(r)\, d\mathbf{r}, \\ 1 &\leq \frac{z}{\rho} + z \int f(r)\, d\mathbf{r} \\ &\quad + \frac{1}{2} z \int f(r_1) f(r_2) \rho g(\mathbf{r}_1 - \mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2, \end{aligned} \qquad (6.14)$$

or in the case of hard spheres,

$$\begin{aligned} &\int_{|r| \leq a} \rho g(r) e^{\beta\phi(r)}\, d\mathbf{r} \leq \frac{z}{\rho} - 1, \\ &\frac{1}{2}\left\{\int_{|r_1| \leq a, |r_2| \leq a, |r_{12}| \geq a} \rho g(\mathbf{r}_1 - \mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2\right\} \\ &\qquad \geq \tfrac{4}{3}\pi a^3 + \frac{1}{z} - \frac{1}{\rho}. \end{aligned} \qquad (6.15)$$

[18] B. R. A. Nijboer and L. Van Hove, Phys. Rev. **85**, 777 (1952).

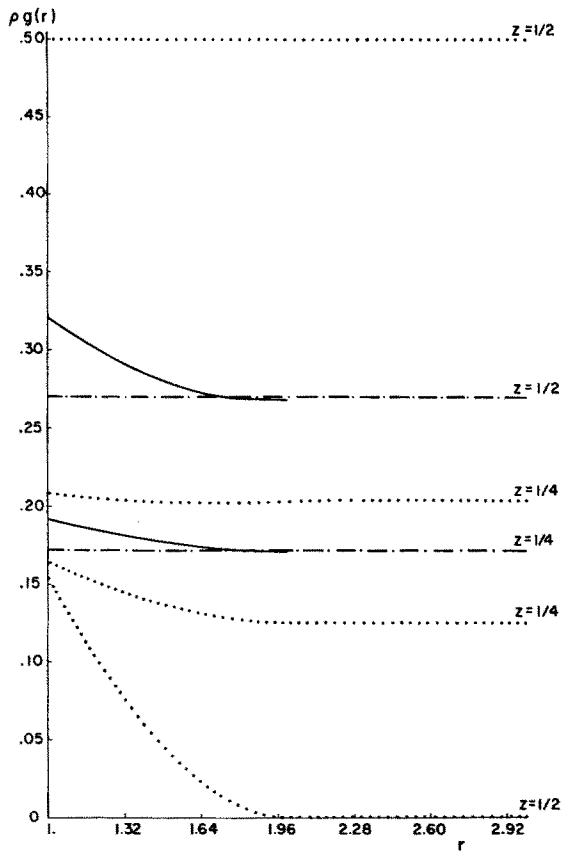


FIG. 2. Bounds on $\rho g(r)$. Dotted lines represent upper and lower bounds of $\rho g(r)$ for $z = \frac{1}{2}$ and $z = \frac{1}{4}$. Solid lines represent solution of the P–Y equation for the same values of z for $r \leq 2$. Straight dot-dash lines represent values of ρ for $z = \frac{1}{4}, \frac{1}{2}$ from (6.10), and are thus the asymptotic values of $\rho g(r)$ for $z = \frac{1}{4}, \frac{1}{2}$.

VII. OTHER INEQUALITIES

In this section we describe briefly some other bounds on the thermodynamic functions which are not obtainable directly from the general development given above. First, it is known that, for all physically reasonable potentials, $p(z)$ and $\rho(z)$ are nondecreasing functions of z.[19] When $\phi(r) \geq 0$, it was also shown by Groeneveld that $p(z_1 + z_2) \leq p(z_1) + p(z_2)$. We shall now show further that when the forces between the particles are purely repulsive, $\phi'(r) \leq 0$, then $p(z)/z$ is a monotonically decreasing function of z [since $\phi(r)$ goes to zero as $r \to \infty$, $\phi'(r) \leq 0$ implies $\phi \geq 0$]. Recalling that $\rho = \partial\beta p/\partial \ln z$, we have, on taking the derivative of $\beta p/z$,

$$(d/dz)(\beta p/z) = (1/z^2)[\rho - \beta p] = \frac{1}{z^2}\frac{1}{6}\int r\beta\phi'(r)n_2(r)\,d\mathbf{r}. \tag{7.1}$$

The last equality follows from the virial theorem and applies to all potentials. For purely repulsive forces, however, $d(\beta p/z)/dz$ is clearly negative.

For more general forces, but with potentials such that (5.6) and the sequel are valid, we may instead obtain bounds on the pressure. Since f can always be decomposed into a difference of functions decreasing monotonically to zero, we write

$$f(r) = f^+(r) - f^-(r), \tag{7.2}$$

$$\frac{d}{dr}f^+ \geq 0, \qquad \frac{d}{dr}f^- \geq 0,$$

$$\left(\frac{d}{dr}f^+\right)\left(\frac{d}{dr}f^-\right) = 0. \tag{7.3}$$

We then find from the virial theorem and (5.11) that

$$\beta p = \rho + \frac{1}{6}\int e^{\beta\phi(r)} r f'(r) n_2(r)\,d\mathbf{r}$$

$$\leq \rho + \frac{1}{6}\int e^{\beta\phi(r)} r \frac{df^+}{dr} n_2(r)\,d\mathbf{r}$$

$$\leq \rho\left[1 + \frac{1}{6} z e^{2\beta\Phi'} \int r \frac{df^+}{dr}\,d\mathbf{r}\right].$$

Hence

$$\beta p \leq \rho\left[1 - \frac{1}{2} z e^{2\beta\Phi'} \int f^+(r)\,d\mathbf{r}\right], \tag{7.4}$$

where $\int f^+(r)\,d\mathbf{r} \leq 0$. Eliminating z or ρ from (7.4) by virtue of (5.10) thus yields

$$\beta p \leq \rho\left[1 - \frac{1}{2}\frac{\rho e^{2\beta\Phi'} \int f^+(r)\,d\mathbf{r}}{1 - \rho e^{2\beta\Phi'} \int f_-(r)\,d\mathbf{r}}\right] \tag{7.5}$$

$$\beta p \leq z \frac{1 - \frac{1}{2} z e^{2\beta\Phi'} \int f^+(r)\,d\mathbf{r}}{1 - z e^{2\beta\Phi'} \int f_+(r)\,d\mathbf{r}}. \tag{7.6}$$

For repulsive forces, (7.5) reduces to

$$\beta p \leq \rho(1 - |b_2|\,\rho)/(1 - 2\,|b_2|\,\rho). \tag{7.7}$$

Lower bounds on p may be obtained in a similar fashion.

Bounds of the form (7.5)–(7.7), unfortunately, become useless at high density. Nonetheless, the monotonicity of $\rho(z)$ alone helps establish some bounds on $p(z)$. We have

$$\beta p(z) = \beta p(z_0) + \int_{z_0}^{z} \frac{\rho}{z}\,dz \geq \beta p(z_0) + \rho(z_0) \ln (z/z_0), \tag{7.8}$$

$$\beta p(z) \leq \beta p(z_0) + \rho(z) \ln (z/z_0), \qquad z \geq z_0.$$

Now for potentials with a hard core, $\rho(z)$ has an upper bound ρ_c, the close-packing density. An upper bound to $p(z)$ is then given by

$$\beta p(z) \leq \beta p(z_0) - \rho_c \ln z_0 + \rho_c \ln z, \qquad z \geq z_0. \tag{7.9}$$

[19] L. Van Hove, Physica 15, 951 (1949); D. Ruelle, Helv. Phys. Acta 36, 183 (1963).

We see from (7.9) that $\beta p(z) - \rho_c \ln z$ is a decreasing function of z and that $\lim_{z\to\infty} (\beta p(z)/\ln z) = \rho_c$. This suggests that we write the relation between p and z in the form

$$z = B(\beta p)e^{\beta p/\rho_c}. \tag{7.10}$$

Then $B(\beta p)$ starts out as βp, increases monotonically with p, and satisfies the equation

$$(d/d\beta p) \ln B(\beta p) = 1/\rho(p) - 1/\rho_c \geq 0. \tag{7.11}$$

One may obtain a power series expansion for $B(\beta p)$ from inversion of the series $\sum b_l z^l$. In the special case of a one-dimensional gas of hard rods, one has $B(\beta p) = \beta p$.

Finally, we shall mention one other type of inequality, which compares two systems with different potentials. Let the potential $\phi(r; \alpha)$ have the form

$$\phi(r; \alpha) = \begin{cases} \infty & r < a \\ \alpha u(r) & r \geq a \end{cases}. \tag{7.12}$$

Then the canonical partition function for N particles in a volume V becomes

$$\begin{aligned} Z_N(\alpha) &= \frac{\lambda^N}{N!} \int_{\Omega_N(V)} e^{-\beta\alpha \sum_{i>j} u(r_{ij})}\, d\mathbf{r}^N \\ &= \frac{\lambda^N}{N!} \int_{\Omega_N(V)} (e^{-\frac{1}{2}\beta\alpha \sum_{i>j} u(r_{ij})})^2\, d\mathbf{r}^N, \end{aligned} \tag{7.13}$$

where λ is a constant and $\Omega_N(V)$ the complement of the excluded volume. It follows now from Schwartz's inequality that

$$Z_N(\alpha)Z_N(-\alpha) \geq \left(\frac{\lambda^N}{N!} \int_{\Omega_N} d\mathbf{r}^N\right)^2 = Z_N(0)^2, \tag{7.14}$$

where $Z_N(0)$ is simply the partition function for hard spheres. Since the free energy per particle is given by

$$F_1(\rho; \alpha) = -\lim_{\substack{N\to\infty \\ V\to\infty \\ N/V=\rho}} \frac{\ln Z_N(V, \alpha)}{\beta N}, \tag{7.15}$$

it follows from (7.14) that

$$F_1(\rho; \alpha) + F_1(\rho; -\alpha) \leq 2F_1(\rho; 0). \tag{7.16}$$

The relation (7.16) may be written in terms of the pressure

$$p(\rho; \alpha) = \rho^2 \frac{\partial F_1(\rho; \alpha)}{\partial \rho} = -\frac{\partial F_1(v; \alpha)}{\partial v}, \tag{7.17}$$

where $v = \rho^{-1}$ is the volume per particle, and becomes

$$\int_v^\infty \left[\beta p(v'; \alpha) + \beta p(v'; -\alpha) - \frac{2}{v'}\right] dv'$$

$$\leq 2 \int_v^\infty \left[\beta p(v'; 0) - \frac{1}{v'}\right] dv'. \tag{7.18}$$

It is interesting to note that when $u(r)$ represents a very long-range weak potential, such as the type considered by Kac, Uhlenbeck, and Hemmer,[20]

$$u(r) = -\gamma^\nu u(\gamma r), \qquad \int u(r)\, d\mathbf{r} = 1, \tag{7.19}$$

where $\nu = 1, 2, 3$ is the dimensionality of the space, then in the limit $\gamma \to 0$

$$\beta p(\rho; \alpha) = \beta p(\rho; 0) + \alpha\rho^2, \tag{7.20}$$

for $\alpha < 0$, while for $\alpha > 0$, (7.20) has to be supplemented by the Maxwell rule for the two-phase region. Thus (7.18) becomes an equality for volume v outside the two-phase region.

ACKNOWLEDGMENTS

We would like to thank Dr. E. Lieb, Dr. G. Stell, and especially Dr. O. Penrose for many very valuable discussions and comments. We should also like to thank J. Steadman for computing the graphs.

APPENDIX A. FUGACITY EXPANSION OF THE URSELL FUNCTION

We may use the methods developed in this paper to obtain directly the fugacity expansion, with or without a remainder, of the Ursell function $\mathfrak{F}_s(\mathbf{r}^s; z)$. We consider a "turning-on process" which takes $\exp \gamma(\mathbf{y} \mid \alpha)$ from an initial value zero to a final value z, to wit,

$$e^{\gamma(\mathbf{y}|\alpha)} = \alpha z. \tag{A1}$$

Thus expanding $\mathfrak{F}_s(\mathbf{r}^s, [\gamma]) \prod_1^s e^{-\gamma(\mathbf{r}_i|\alpha)}$ at $\alpha = 1$ about its reference value at $\alpha = 0$, we obtain with the help of (2.19)

$$\begin{aligned} \frac{\mathfrak{F}_s(\mathbf{r}^s; z)}{z^s} &= \sum_{j=0}^{t} \frac{z^j}{j!} \int U_{s+j}(\mathbf{r}^s\mathbf{x}^j)\, d\mathbf{x}^j \\ &+ \int_0^z dz' \frac{(z - z')^t}{t!(z')^{s+t+1}} \int \mathfrak{F}_{s+t+1}(\mathbf{r}^s\mathbf{x}^{t+1}; z')\, d\mathbf{x}^{t+1}, \end{aligned} \tag{A2}$$

where

$$U_k(\mathbf{y}^k) = \lim_{z\to 0} \frac{\mathfrak{F}_k(\mathbf{y}^k; z)}{z^k} \tag{A3}$$

is the kth Mayer cluster function.[2]

When $t \to \infty$, (A2) yields the fugacity series for F_s. When $s = 1$, (A2) reduces to (4.13b). It was shown by Groeneveld[2] that for a positive potential, $\phi(r) \geq 0$, the U_k alternate in sign, $(-1)^k U_k \geq 0$. Thus, the coefficients in the fugacity expansion of the Ursell functions alternate in sign when $\phi \geq 0$.

[20] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963).

However, unlike the case of the expansion of $\mathfrak{F}_1 = n$, or the n_k in general, stopping the series after a finite number of terms does not yield an upper or lower bound on $\mathfrak{F}_s$ for $s > 1$, i.e., the remainder term in (A2) alternates in sign only for $s = 1$.

We may however still obtain bounds on the remainder term in (A2) by decomposing $\mathfrak{F}_{s+\ell+1}$ into its positive and negative parts, which are products of n's, and using the bounds on the n's obtained in this paper. Unfortunately, the bounds thus derived do not have the property of approaching zero when the s particles in $\mathfrak{F}_s$ are separated into groups which are far removed from each other. This is not surprising since these bounds hold also when the system is split into two phases, in which case the Ursell functions do not approach zero asymptotically.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

Solution of Linear Integral Equations Using Padé Approximants

J. S. R. CHISHOLM*
CERN, Geneva, Switzerland
(Received 21 June 1963)

It is shown that the exact solution of a nonhomogeneous linear integral equation with a kernel K of rank n is given by forming the Padé approximant $P(n, n)$ from the first $2n$ terms of the perturbation series solution. It follows that for a compact kernel K, the solution is $\lim_{n\to\infty} P(n, n)$; this gives meaning to a large class of perturbation series when the perturbation is large. The possible extension of this result to wider classes of equations is discussed.

I. INTRODUCTION

THE method of Padé approximants has been applied with considerable success to the solution of a variety of problems[1-3] in which a divergent or an apparently divergent series requires interpretation. Gammel[4] has suggested the use of this technique to solve a strong coupling problem in quantum field theory—the calculation of the magnetic moments of the nucleons from the perturbation series. This suggestion has led the author to study the realm of validity of the Padé method in perturbation theory, in particular in potential theory, and then to broaden the scope of the theory to a general study of linear integral equations.

Given a series

$$s = \sum_{r=0}^{\infty} \lambda^r m_r, \tag{1.1}$$

where m_r are complex in general, the Padé approximant

$$P(\alpha, \beta) \equiv \left(\sum_{p=0}^{\alpha} \lambda^p a_p\right)\left(\sum_{q=0}^{\beta} \lambda^q b_q\right)^{-1} \tag{1.2}$$

is defined by the identity in λ,

$$\left(\sum_{q=0}^{\beta} \lambda^q b_q\right)\left(\sum_{r=0}^{\infty} \lambda^r m_r\right) = \sum_{p=0}^{\alpha} \lambda^p a_p + O(\lambda^{\alpha+\beta+1}). \tag{1.3}$$

Thus a_p and b_q are given by

$$\sum_{q=0}^{\min(r,\beta)} b_q m_{r-q} = a_r \qquad (r = 0, 1, \cdots, \alpha), \tag{1.4}$$

and

$$\sum_{q=0}^{\min(r,\beta)} b_q m_{r-q} = 0 \qquad (r = \alpha + 1, \cdots, \alpha + \beta). \tag{1.5}$$

In general we can choose, say, $b_0 = 1$, and (1.4) then defines $b_1, \cdots, b_\beta$ uniquely; from (1.3), a_r and hence $P(\alpha, \beta)$ are defined uniquely. Further, it is clear that the formal expansion of (1.2) as a power series in λ will agree with (1.1) to $O(\lambda^{\alpha+\beta})$.

II. THE SOLUTION OF INTEGRAL EQUATIONS WITH KERNELS OF FINITE RANK

We shall first study the nonhomogeneous linear integral equation

* On leave of absence from the School of Mathematics, Trinity College, Dublin, Ireland.

[1] G. A. Baker and J. L. Gammel, J. Math. Anal. and Appls. **2**, 21 (1961).
[2] G. A. Baker, Phys. Rev. **124**, 768 (1961).
[3] C. Lovelace and D. Masson, Nuovo Cimento **26**, 472 (1962).
[4] J. L. Gammel (private communication).

However, unlike the case of the expansion of $\mathfrak{F}_1 = n$, or the n_k in general, stopping the series after a finite number of terms does not yield an upper or lower bound on $\mathfrak{F}_s$ for $s > 1$, i.e., the remainder term in (A2) alternates in sign only for $s = 1$.

We may however still obtain bounds on the remainder term in (A2) by decomposing $\mathfrak{F}_{s+\ell+1}$ into its positive and negative parts, which are products of n's, and using the bounds on the n's obtained in this paper. Unfortunately, the bounds thus derived do not have the property of approaching zero when the s particles in $\mathfrak{F}_s$ are separated into groups which are far removed from each other. This is not surprising since these bounds hold also when the system is split into two phases, in which case the Ursell functions do not approach zero asymptotically.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

Solution of Linear Integral Equations Using Padé Approximants

J. S. R. CHISHOLM*
CERN, Geneva, Switzerland
(Received 21 June 1963)

It is shown that the exact solution of a nonhomogeneous linear integral equation with a kernel K of rank n is given by forming the Padé approximant $P(n, n)$ from the first $2n$ terms of the perturbation series solution. It follows that for a compact kernel K, the solution is $\lim_{n\to\infty} P(n, n)$; this gives meaning to a large class of perturbation series when the perturbation is large. The possible extension of this result to wider classes of equations is discussed.

I. INTRODUCTION

THE method of Padé approximants has been applied with considerable success to the solution of a variety of problems[1-3] in which a divergent or an apparently divergent series requires interpretation. Gammel[4] has suggested the use of this technique to solve a strong coupling problem in quantum field theory—the calculation of the magnetic moments of the nucleons from the perturbation series. This suggestion has led the author to study the realm of validity of the Padé method in perturbation theory, in particular in potential theory, and then to broaden the scope of the theory to a general study of linear integral equations.

Given a series

$$s = \sum_{r=0}^{\infty} \lambda^r m_r, \tag{1.1}$$

where m_r are complex in general, the Padé approximant

$$P(\alpha, \beta) \equiv \left(\sum_{p=0}^{\alpha} \lambda^p a_p\right)\left(\sum_{q=0}^{\beta} \lambda^q b_q\right)^{-1} \tag{1.2}$$

is defined by the identity in λ,

$$\left(\sum_{q=0}^{\beta} \lambda^q b_q\right)\left(\sum_{r=0}^{\infty} \lambda^r m_r\right) = \sum_{p=0}^{\alpha} \lambda^p a_p + O(\lambda^{\alpha+\beta+1}). \tag{1.3}$$

Thus a_p and b_q are given by

$$\sum_{q=0}^{\min(r,\beta)} b_q m_{r-q} = a_r \qquad (r = 0, 1, \cdots, \alpha), \tag{1.4}$$

and

$$\sum_{q=0}^{\min(r,\beta)} b_q m_{r-q} = 0 \qquad (r = \alpha + 1, \cdots, \alpha + \beta). \tag{1.5}$$

In general we can choose, say, $b_0 = 1$, and (1.4) then defines $b_1, \cdots, b_\beta$ uniquely; from (1.3), a_r and hence $P(\alpha, \beta)$ are defined uniquely. Further, it is clear that the formal expansion of (1.2) as a power series in λ will agree with (1.1) to $O(\lambda^{\alpha+\beta})$.

II. THE SOLUTION OF INTEGRAL EQUATIONS WITH KERNELS OF FINITE RANK

We shall first study the nonhomogeneous linear integral equation

* On leave of absence from the School of Mathematics, Trinity College, Dublin, Ireland.

[1] G. A. Baker and J. L. Gammel, J. Math. Anal. and Appls. **2**, 21 (1961).
[2] G. A. Baker, Phys. Rev. **124**, 768 (1961).
[3] C. Lovelace and D. Masson, Nuovo Cimento **26**, 472 (1962).
[4] J. L. Gammel (private communication).

$$f(x) = g(x) + \lambda \int K(x, y)f(y)\, dy, \qquad (2.1)$$

when the kernel $K(x, y)$ is of finite rank n, so that it is expressible in the form

$$K(x, y) = \sum_{i=1}^{n} \phi_i(x)\bar{\psi}_i(y), \qquad (2.2)$$

the basis functions $\phi_1(x), \cdots, \phi_n(x)$ being linearly independent. Then the solution of the equation is of the form[5]

$$f(x) = g(x) + \sum_{i=1}^{n} h_i\phi_i(x), \qquad (2.3)$$

where the n coefficients h_i satisfy the set of n linear equations

$$\sum_{i=1} (\delta_{ij} - \lambda k_{ij})h_j = g_i, \qquad (2.4)$$

with

$$k_{ij} = (\phi_i, \psi_i) \equiv \int \phi_i(y)\bar{\psi}_i(y)\, dy \qquad (2.5)$$

and

$$g_i = (g, \psi_i) \equiv \int g(y)\bar{\psi}_i(y)\, dy. \qquad (2.6)$$

Thus, using the representation with finite basis $\{\phi_1, \cdots \phi_n\}$, the integral equation reduces to a finite algebraic system. Equation (2.4) can be written in matrix form as

$$(\mathsf{I} - \lambda\mathsf{k})\mathbf{h} = \mathbf{g}. \qquad (2.7)$$

If $\mathsf{T} = (T_{ki})$ is any nonsingular $(n \times n)$ matrix, we can define a new basis $\{\Phi_k\}$ $(k = 1, \cdots, n)$ by

$$\Phi_k = \sum_{i=1}^{n} \phi_i T_{ik},$$

or

$$\mathbf{\Phi} = \boldsymbol{\phi}\mathsf{T}. \qquad (2.8)$$

This transformation simply defines a new representation, and we can express the solution in this new representation by defining

$$\bar{\mathbf{\Psi}} = \mathsf{T}^{-1}\bar{\boldsymbol{\psi}}, \qquad (2.9)$$

$$\mathbf{H} = \mathsf{T}^{-1}\mathbf{h}, \qquad (2.10)$$

and

$$\mathbf{G} = (g, \mathbf{\Psi}) = \mathsf{T}^{-1}\mathbf{g}. \qquad (2.11)$$

[5] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, (Blackie & Sons, London) pp. 161–2.

Then Eqs. (2.3) and (2.7) become

$$f = g + \sum_{i=1}^{n} H_i\Phi_i, \qquad (2.12)$$

and

$$\mathsf{T}^{-1}(\mathsf{I} - \lambda\mathsf{k})\mathsf{T}\mathbf{H} = \mathbf{G}. \qquad (2.13)$$

For simplicity, we now assume that the eigenvalues of k are nondegenerate, so that k can be diagonalized. The theorem which we shall now prove is true even if we do not make this assumption, and its proof is a generalization of that given here, based on a reduction of k to canonical form. Since T is an arbitrary nonsingular $(n \times n)$ matrix, we can choose it to make $\mathsf{T}^{-1}(\mathsf{I}-\lambda\mathsf{k})T$ diagonal. If κ_j $(j=1, 2, \cdots, n)$ are the eigenvalues of k, (2.13) can then be solved to give

$$H_j = G_j/(1 - \lambda\kappa_j) \qquad (j = 1, 2, \cdots, n). \qquad (2.14)$$

So far we have used representations in which the kernel $K(x, y)$ is represented by an $(n \times n)$ matrix. The representations which are convenient to use in practice may not be of this form; generally, we must consider a basis $\{\phi_l'\}$ of linearly independent functions (often orthogonal), in which l ranges through the positive integers 1, 2, 3, $\cdots$. In this representation,

$$K(x, y) = \sum_{l=1}^{\infty} \phi_l'(x)\bar{\psi}_l'(y), \qquad (2.15)$$

and is represented by the infinite matrix

$$k_{lm}' = \int \phi_l'(y)\bar{\psi}_m'(y)\, dy. \qquad (2.16)$$

If we regard $K = \int dy\, K(x, y)$ as an operator in Hilbert space, then since it is of finite rank, it is a particular type of "completely continuous" or "compact" operator.[6] Loosely speaking, compact operators are those that can in some representation be approximated by a sequence of finite matrices; for kernels of finite rank, the operator is *exactly* represented by a finite matrix, as in (2.4) and (2.5). We may therefore use known properties of compact operators; in particular, we know that the operator K is uniformly approximated as $p \to \infty$ by the sequence of $(p \times p)$ matrices (k_{lm}') $(l, m = 1, \cdots, p)$, operating on the space spanned by $\{\phi_1', \cdots, \phi_p'\}$.

We can now extend the set $\{\Phi_1, \cdots, \Phi_n\}$ to form a complete basis $\{\Phi_i\}$ $(i = 1, 2, 3, \cdots)$; we choose $\Phi_{n+1}, \Phi_{n+2}, \cdots$, to be an orthonormal set, each normal to the set $\{\Phi_1, \cdots, \Phi_n\}$. Then the solution (2.14) can be expressed as

[6] Reference 5, p. 204.

$$\begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_n \\ \hline H_{n+1} \\ H_{n+2} \\ \vdots \end{pmatrix} = \left(\begin{array}{cccc:ccc} \frac{1}{1-\lambda\kappa_1} & 0 & \cdots & 0 & & & \\ 0 & \frac{1}{1-\lambda\kappa_2} & \cdots & 0 & & & \\ \cdot & \cdot & & \cdot & & & \\ 0 & 0 & \cdots & \frac{1}{1-\lambda\kappa_n} & & & \\ \hdashline & & & & 0 & 0 & \cdots \\ & & & & 0 & 0 & \cdots \\ & & & & & \cdots & \end{array}\right) \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \\ \hline G_{n+1} \\ G_{n+2} \\ \vdots \end{pmatrix}, \tag{2.17}$$

Except when $\lambda = \kappa_i^{-1}$ $(i = 1, \cdots, n)$, the matrix in (2.17), denoted by R, defines a completely continuous operator; so in any representation the matrix is uniformly approximated by the $(p \times p)$ matrices formed from the first p rows and columns. We may therefore treat this matrix as though it were finite.

Now suppose that the bases $\{\Phi_i\}$ and $\{\phi_i'\}$ are linked by the transformation U, so that $g(x)$ is represented in practice by

$$\mathbf{g}' = \mathsf{U}\mathbf{G}. \tag{2.18}$$

Then the matrix R in (2.17) transforms to

$$\mathsf{r}' = \mathsf{URU}^{-1}. \tag{2.19}$$

The elements of this matrix are linear combinations of the elements of R, that is, of $(1 - \lambda\kappa_i)^{-1}$ $(i = 1, \cdots, n)$. Thus as functions of λ they are of the form of Padé approximants $P(n - 1, n)$, with denominator

$$\prod_{i=1}^{n} (1 - \lambda\kappa_i).$$

Thus the solution

$$\mathbf{f}' = (\mathsf{I} + \mathsf{r}')\mathbf{g}' \tag{2.20}$$

has components of the form $P(n, n)$ as functions of λ.

Now Eq. (2.1) can be solved, formally at least, by perturbation methods. Symbolically the solution is

$$f = [1 + \lambda K + \lambda^2 K^2 + \cdots]g,$$

in any representation; the components of f are given formally as power series in λ, of form (2.1). Suppose that we calculate the power series for a component up to $O(\lambda^{2n})$ and then form the diagonal approximant $P(n, n)$ by (2.3); we know that this approximant is *uniquely* determined by the first $2n$ terms of the power series. However, the exact solution (2.20) is of the form of $P(n, n)$, and must satisfy (1.3); hence $P(n, n)$ is the exact solution. Thus we have

Theorem I. When the kernel $K(x, y)$ in the integral equation

$$f(x) = g(x) + \lambda \int K(x, y)f(y)\, dy$$

is of finite rank n, then the exact solution

$$f(x) = \sum_{k=0,1,\cdots} f_k'\phi_k'(x)$$

in terms of any basis $\{\phi_k'\}$ is given by forming the Padé approximants $P(n, n)$ from the first $2n$ terms of the perturbation series expansion for f_k'.

For a kernel of finite rank n, the homogeneous equation formed by putting $g(x) \equiv 0$ in (2.1) has solutions for n values of λ only. From (2.7), these n eigenvalues are the roots of the determinental equation

$$||\mathsf{I} - \lambda\mathsf{k}|| = 0,$$

and thus are the inverses κ_i^{-1} of the eigenvalues of k. Hence the singularities of the H_i in (2.14), and thus of the components (3.18) of f in any representation, are given by the eigenvalues $\lambda = \kappa_i^{-1}$ $(j = 1, 2, \cdots, n)$ of the homogeneous equation. So as a corollary to the theorem, we know that the singularities of the approximants $P(n, n)$ calculated from the perturbation series give us the eigenvalues of the homogeneous integral equation.

III. EXTENSIONS OF THEOREM I

Since any compact kernel is the limit of a sequence of kernels of finite rank, we might expect that in general the solution of (2.1) when $K(x, y)$ is compact would be given by

$$\lim_{n\to\infty} P(n, n)$$

in any representation. To prove this, we consider a kernel $K^{(u)}(x, y)$ of rank u which approximates

the compact $K(x, y)$. If $\kappa_i^{(u)}$ is an eigenvalue of $K^{(u)}$ with (normalized) eigenfunction $f_i^{(u)}$, then

$$\int dy K^{(u)}(x, y) f_i^{(u)}(y) - \kappa_i^{(u)} f_i^{(u)}(y) = 0.$$

Therefore

$$\int dy K(x, y) f_i^{(u)}(y) - \kappa_i^{(u)} f_i^{(u)}(x) = \int dy [K(x, y) - K^{(u)}(x, y)] f_i^{(u)}(y);$$

as $u \to \infty$ this tends to zero, so that the eigenvalues of $K^{(u)}$ all approach eigenvalues of K as $u \to \infty$. Therefore if we consider the region R of the λ plane formed by cutting out small circles around the singular points of the operator $(I - \lambda K)^{-1}$, we know that for large enough n, there will be no singular points of $(I - \lambda K^{(u)})^{-1}$ in R.

Now let us consider the Padé approximant $P(n, n)$ formed from the first $(2n + 1)$ terms of the perturbation solution of (2.1), and the approximant $P^{(u)}(n, n)$ formed from the first $(2n + 1)$ terms of the perturbation solution of

$$f(x) = g(x) + \lambda \int K^{(u)}(x, y) f(y)\, dy. \tag{3.1}$$

The series solutions, in form (1.1), are formally

$$s = \sum_{r=0}^{\infty} \lambda^r K^r g,$$

and

$$s^{(u)} = \sum_{r=0}^{\infty} \lambda^r [K^{(u)}]^r g,$$

and we know that $[K^{(u)}]^r g \to K^r g$ uniformly as $u \to \infty$. We take the Padé approximants for kernels K and $K^{(u)}$, respectively, as

$$P(n, n) = \left(\sum_{p=0}^{n} \lambda^p a_p \right) \left(\sum_{q=0}^{n} \lambda^q b_q \right)^{-1}, \tag{3.2}$$

and

$$P^{(u)}(n, n) = \left(\sum_{p=0}^{n} \lambda^p a_p^{(u)} \right) \left(\sum_{q=0}^{n} \lambda^q b_q^{(u)} \right)^{-1}. \tag{3.3}$$

Except when the coefficients $m_r = K^r g$ in (1.5) "accidentally" have zero determinant, Eqs. (1.4) and (1.5) can be solved for (a_p, b_q) and $(a_p^{(u)}, b_q^{(u)})$; and since $[K^{(u)}]^r g \to K^r g$, we know that $a_p^{(u)} \to a_p$ and $b_q^{(u)} \to b_q$ as $u \to \infty$. Hence in the region R, where the denominators of (3.2) and (3.3) cannot be zero, $P^{(u)}(n, n) \to P(n, n)$ as $u \to \infty$. Near to the poles, the difference $P(n, n) - P^{(u)}(n, n)$ is proportional to a power of λ. If this difference is to tend uniformly to zero in R, the radii of the circles around the poles in the λ plane must be made proportional to λ. Then by choosing n large enough, we can make the difference as small as we please arbitrarily near to all poles lying within $|\lambda| \leq M$, where M is *any* fixed bound; in this sense, $P^{(n)}(n, n)$ approximates $P(n, n)$ in the whole λ plane as $n \to \infty$.

Since K is uniformly approximated by K_n as $n \to \infty$, the operator $(I - \lambda K)^{-1}$ is uniformly approximated by $(I - \lambda K^{(n)})^{-1}$ in the region R, where it is nonsingular. Thus the solution $f = (I - \lambda K)^{-1} g$ of (2.1) is approximated by the solution $f^{(n)} = (I - \lambda K^{(n)})^{-1} g$ of (3.1) with $u = n$; once again, the proportionality of the radii of the circles to λ ensures uniform approximation within R; this proportionality reflects the fact that we cannot hope to represent more than n poles of a function by using an approximant with n poles. But by Theorem I, the solution of (3.1) is exactly

$$f^{(n)} = P^{(n)}(n, n),$$

and we have just shown that $P^{(n)}(n, n)$ approximates $P(n, n)$ as $n \to \infty$. Hence,

$$f = (I - \lambda K)^{-1} g = \lim_{n \to \infty} P(n, n).$$

As $n \to \infty$, the region R can be taken to exclude as little of the region $|\lambda| \leq M$ of the λ plane as we wish where M is as large as we wish. Thus we have

Theorem II. The solution of (2.1) when the kernel $K(x, y)$ is compact is

$$\lim_{n \to \infty} P(n, n),$$

where $P(n, n)$ are the diagonal Padé approximants formed from the first $(2n + 1)$ terms of the perturbation solution of (2.1).

Theorem II enables us to apply the Padé method to the solution of a wide class of problems. For example, it has been shown[7] that the Schrödinger equation with a potential which is a superposition of Yukawa potentials can be reduced to an integral equation with a compact kernel. Whether the Padé method is preferable to other methods, such as the Fredholm method or the "K matrix" method, will depend upon the nature of the problem.

It seems that the Padé method should be applicable to a wider class of kernels than that of compact kernels, for the following reasons:

[7] D. I. Fivel, R. Brown, B. Lee, and R. F. Sawyer (preprint 1963).

(a) the justification of the Padé method is analytically similar to that of the Fredholm method, and it is known[8] that this method extends beyond the class of completely continuous K.

(b) the Padé method gives us a continued-fraction representation of the operator $(I - \lambda K)^{-1}$. It is well known[9] that functions with branch cuts and even isolated essential singularities can be represented by continued fractions, and applications of this technique are common in electrical engineering.[10] It would therefore be surprising if this technique of solving integral equations was limited to compact kernels, which give no limit points of singularities (i.e., only poles) in the finite part of the λ plane.

(c) it is clear from the work of Baker, Gammel, and Wills[1,11] and of Lovelace and Masson[3] that the realm of validity of continued-fraction expansions is wider than that compassed by the present analytic theory.

The hypothesis put forward by Baker, Gammel, and Wills[11] concerning the realm of validity of $P(n, n)$ approximants is a strong challenge to broaden the basis of the analytic theory of continued fractions. It may be that this broadening could take place within the framework of functional analysis, based on a theory of convergence of spectra; it would then immediately result in an extension of the range of applicability of the Padé technique to the solution of wider classes of integral and other types of equation.

While this paper was being prepared, the author learned[12] that a theorem concerning the application of Padé methods to potential theory has been proved by G. Baker.

ACKNOWLEDGMENTS

The author has benefitted greatly from many discussions and communications with Dr. John Gammel on the subject of Padé approximants; this work was motivated by Dr. Gammel's suggestion of using this technique to interpret the perturbation series. The relevance of the concept of complete continuity was pointed out by Dr. Lovelace, with whom the author has had several very illuminating discussions. The paper has been improved in several ways as a result of comments by Professor J. L. B. Cooper. The program of work was initiated while the author was in the Mathematics Department, University College, Cardiff, Wales, and during visits to Atomic Energy Research Establishment, Harwell, England, as consultant.

[8] Reference 5, p. 205.

[9] H. S. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand, Inc., Princeton, New Jersey), Chaps. 18 and 19.

[10] B. Gross and E. P. Braga, *Singularities of Linear System Functions* (Elsevier Publishing Company, Inc., Houston, Texas), Chap. 4.

[11] G. A. Baker, J. L. Gammel, and J. G. Wills, J. Math. Anal. and Appls. **2**, 21 (1961).

[12] J. L. Gammel (private communication).

Remarks on the Type of Von Neumann Algebras of Local Observables in Quantum Field Theory*

RICHARD V. KADISON†
Mathematics Department, Columbia University, New York, New York
(Received 30 July 1963)

The von Neumann algebras of local observables associated with certain regions of space–time are believed to be factors. We show that these algebras are not of finite type. The commutant of the tensor product of two semifinite von Neumann algebras is analyzed with the aid of this result. The factors in question have the vacuum state as separating and cyclic vector. It is shown that a factor of type I_∞ with I_∞ commutant, and a subfactor of type I_∞ with I_∞ relative commutant have a common separating and cyclic vector. This settles negatively some conjectures aimed at proving that these factors are not of type I. An argument of Araki's showing that the factors associated with certain regions are not of type I is presented in simplified form.

I. INTRODUCTION

SOME attention has been given recently to the algebras of local observables associated with regions of space–time by a quantum field theory.[1–6] For certain regions, these von Neumann algebras are believed to be factors in the sense of Murray and von Neumann.[7] The question of the types[8] of the factors occurring is of some importance in this connection. Making use of the cyclicity and separating properties of the vacuum state for these factors, we show (Theorem 1) that they are of infinite type. This same result makes possible a direct proof (avoiding Hilbert algebras) of the known result[9] $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)' = \mathfrak{R}_1' \otimes \mathfrak{R}_2'$ when $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are semifinite von Neumann algebras (i.e., have no portion of type *III*). Section IV is devoted to this proof.

The strong separating and cyclicity properties of the vacuum state relative to the various factors seem to rule out their being of type I. The basic question is:

If $\mathfrak{R}$, $\mathfrak{R}'$, $\mathfrak{R}_1$, and $\mathfrak{R}_1' \cap \mathfrak{R}$ are factors of type I_∞ and $\mathfrak{R}_1 \subseteq \mathfrak{R}$, can $\mathfrak{R}$ and $\mathfrak{R}_1$ have a joint generating and separating vector? (1.1)

In Sec. V we analyze cyclic and separating vectors for factors of type I_∞ with I_∞ commutants, and reduce some variants of (1.1)[10] to (1.1). In Sec. VI, we construct such a joint generating and cyclic vector (settling the associated conjectures[10] negatively).

The final section contains a simplified form of an argument of Araki's.[11] The uniqueness of the vacuum state as a translation invariant, together with the fact that it is separating, is used to show that the factor associated with a certain region of space–time is not of type I.

Question 1.1 arose in a conversation (October 1962) with A. S. Wightman (Theorem 1 was proved during this conversation).

II. NOTATION

As we have done in the introduction, we denote by $\mathfrak{R}'$ the set of (bounded) operators commuting with all the operators of $\mathfrak{R}$ ($\mathfrak{R}'$ is called *the commutant* of $\mathfrak{R}$). We use the symbol and terminology for an orthogonal projection operator interchangeably with the symbol and terminology for its range (the closed subspace on which it projects). If $\mathfrak{R}$ is a family of operators and N a set of vectors, $[\mathfrak{R}N]$ will denote the closed subspace spanned by vectors of the form Ax with A in $\mathfrak{R}$ and x in N (so that, by the convention just adopted, $[\mathfrak{R}N]$ will also denote the orthogonal projection operator on this subspace).

* This research was supported in part by the National Science Foundation under Grant No. NSF-G 19022.
† Alfred P. Sloan Fellow.
[1] H. Araki, J. Math. Phys. 4, 1343 (1963).
[2] H. Araki and E. J. Woods, J. Math. Phys. 4, 637 (1963).
[3] R. Haag, Proceedings of the Midwest Conference on Theoretical Physics, Minneapolis, Minnesota, 1961.
[4] R. Haag and B. Schroer, J. Math. Phys. 3, 248 (1962).
[5] I. E. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, Rhode Island, 1963).
[6] M. Guénin and B. Misra, "On the von Neumann algebra generated by the field operators" (mimeographed note, Institute of Theoretical Physics, Geneva).
[7] F. J. Murray and J. von Neumann, Ann. Math. 37, 116 (1936).
[8] See reference 7, especially pp. 171–172.
[9] J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1957), p. 102, Proposition 14.
[10] See reference 6 (listed there as Conjectures B1 and B2).
[11] See reference 1, especially Lemmas 10.1—10.3; and reference 2, especially Lemmas 4.1 and 4.2.

III. INFINITE TYPE

The von Neumann algebra of local observables associated with a bounded open region $\mathcal{O}$ of space–time by a quantum field is a factor which has the vacuum state ψ_0 as a separating and cyclic vector. If $\mathcal{O}_0$ is an open subregion of $\mathcal{O}$ with boundary at positive distance from the boundary of $\mathcal{O}$, its factor $\mathfrak{R}_0$ is a *proper* subfactor of $\mathfrak{R}$ (again with ψ_0 as separating and cylic vector). We prove

Theorem 1. The factor $\mathfrak{R}$ is of infinite type.

This will be accomplished by establishing:

Lemma 2. If $\mathfrak{R}_0$ is a proper sub von Neumann algebra of the von Neumann algebra $\mathfrak{R}$, and x is a separating and cyclic vector for both $\mathfrak{R}$ and $\mathfrak{R}_0$, then $\mathfrak{R}$ (and $\mathfrak{R}_0$) are of infinite type.

Remark. Although the proof is somewhat simpler in the factor case, it seems worthwhile to establish this lemma for arbitrary von Neumann algebras. We shall do this.

Proof: We assume that $\mathfrak{R}$ is finite and show that $\mathfrak{R}_0 = \mathfrak{R}$. Assuming $\mathfrak{R}$ finite, $[\mathfrak{R}x]$ $(=\mathfrak{K}$, the underlying Hilbert space) is finite in $\mathfrak{R}'$[12]; so that $\mathfrak{R}'$ (and, similarly, $\mathfrak{R}_0'$) are finite. Let D, D', D_0, and D_0' denote the center-valued dimension functions on $\mathfrak{R}$, $\mathfrak{R}'$, $\mathfrak{R}_0$, and $\mathfrak{R}_0'$, respectively,[13] each normalized so that the identity operator I has dimension I. Since $\mathfrak{K} = [\mathfrak{R}x] = [\mathfrak{R}'x] = [\mathfrak{R}_0x] = [\mathfrak{R}_0'x]$, $D'([\mathfrak{R}y]) = D([\mathfrak{R}'y])$ and $D_0'([\mathfrak{R}_0y]) = D_0([\mathfrak{R}_0'y])$, for each y in $\mathfrak{K}$, by virtue of the Coupling Theorem.[14] In particular, with P a central projection in $\mathfrak{R}$, $[\mathfrak{R}_0Px] = P = [\mathfrak{R}'Px] \subseteq [\mathfrak{R}_0'Px]$; so that $[\mathfrak{R}_0'\mathfrak{R}_0Px] \subseteq [\mathfrak{R}_0'Px] \subseteq [\mathfrak{R}_0'\mathfrak{R}_0Px]$ and $D_0'([\mathfrak{R}_0Px]) = D_0'(P) = D_0([\mathfrak{R}_0'Px]) = [\mathfrak{R}_0'Px]$ (since $[\mathfrak{R}_0'Px]$ is $[\mathfrak{R}_0'\mathfrak{R}_0Px]$, a central projection in $\mathfrak{R}_0$).[13] Now $[\mathfrak{R}_0'Px]D_0'(P) = D_0'([\mathfrak{R}_0'Px]P) = D_0'([\mathfrak{R}_0'Px])$, so that $P \geq [\mathfrak{R}_0'Px]$. Thus $P = [\mathfrak{R}_0'Px] \in \mathfrak{R}_0$; and the center of $\mathfrak{R}$ is contained in that of $\mathfrak{R}_0$. By the same token, the center of $\mathfrak{R}_0'$ is contained in that of $\mathfrak{R}$. Uniqueness of the (normalized) dimension function now implies that D_0 is the restriction of D to $\mathfrak{R}_0$; and D' is the restriction of D_0' to $\mathfrak{R}'$.

Let E be a projection in $\mathfrak{R}$ and y be Ex. Then $E = [\mathfrak{R}'y]$, so that $D(E) = D'([\mathfrak{R}y])$. Since $[\mathfrak{R}y] \in \mathfrak{R}' \subseteq \mathfrak{R}_0'$, $D(E) = D'([\mathfrak{R}y]) = D_0'([\mathfrak{R}y]) \geq D_0'([\mathfrak{R}_0y]) = D_0([\mathfrak{R}_0'y]) = D([\mathfrak{R}_0'y]) \geq D([\mathfrak{R}'y]) = D(E)$. Thus $D(E) = D([\mathfrak{R}_0'y])$; and, since $E \leq [\mathfrak{R}_0'y]$, $E = [\mathfrak{R}_0'y] \in \mathfrak{R}_0$. Hence each projection in $\mathfrak{R}$ lies in $\mathfrak{R}_0$; and $\mathfrak{R} = \mathfrak{R}_0$, contradicting the hypotheses.

[12] This is a consequence of Lemma 9.3.3 of reference 7 (as in reference 9, p. 242, Proposition 3, or Lemma 3.3.4 of reference 13).

[13] R. Kadison, Ann. Math. **66**, 304 (1957), see Chap. III.

[14] See reference 13, Theorem 3.3.8.

IV. TENSOR PRODUCTS

Lemma 2 is the key to[9]

Theorem 3. If $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are semifinite von Neumann algebras, then $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)' = \mathfrak{R}_1' \otimes \mathfrak{R}_2'$.

For the proof of this, we shall want:

Lemma 4. If $\mathfrak{R}_0$ and $\mathfrak{R}$ are von Neumann algebras such that $\mathfrak{R}_0 \subseteq \mathfrak{R}$, the center of $\mathfrak{R}$ is contained in that of $\mathfrak{R}_0$, and $\{E_\alpha'\}$ is a family of projections in $\mathfrak{R}'$ with union I such that $\mathfrak{R}E_\alpha' = \mathfrak{R}_0E_\alpha'$ (or, dually, $E_\alpha'\mathfrak{R}'E_\alpha' = E_\alpha'\mathfrak{R}_0'E_\alpha'$) for each α; then $\mathfrak{R}_0 = \mathfrak{R}$.

Proof: Since von Neumann algebras are generated by their projections, it suffices to show that each projection E in $\mathfrak{R}$ lies in $\mathfrak{R}_0$. By assumption, for each α there is an A_0 in $\mathfrak{R}_0$ such that $EE_\alpha' = A_0E_\alpha'$. Let F_0 be the range projection of A_0. Then F_0 lies in $\mathfrak{R}_0$.[15] Now F_0E_α' $(= E_\alpha'F_0)$ and A_0E_α' $(= E_\alpha'A_0)$ are both projections with $\{E_\alpha'A_0x\}$ dense in their ranges; so that $A_0E_\alpha' = F_0E_\alpha' = EE_\alpha'$. With T' in $\mathfrak{R}'$, $T'EE_\alpha' = ET'E_\alpha' = T'F_0E_\alpha' = F_0T'E_\alpha'$; so that $EP_\alpha = F_0P_\alpha$, where P_α is the central carrier of E_α' (relative to $\mathfrak{R}'$).[16] Since the center of $\mathfrak{R}$ is contained in the center of $\mathfrak{R}_0$, F_0P_α lies in $\mathfrak{R}_0$. Moreover,

$$E(\bigvee_\alpha P_\alpha) \geq E(\bigvee_\alpha E_\alpha') = E \cdot I = E,$$

so that

$$E = E(\bigvee_\alpha P_\alpha) = \bigvee_\alpha EP_\alpha = \bigvee_\alpha F_0P_\alpha$$

lies in $\mathfrak{R}_0$.

If $E_\alpha'\mathfrak{R}'E_\alpha' = E_\alpha'\mathfrak{R}_0'E_\alpha'$, then $\mathfrak{R}E_\alpha' = \mathfrak{R}_0E_\alpha'$ for each α[17]; and from the preceding, $\mathfrak{R} = \mathfrak{R}_0$.

Proof of Theorem 3: With A_1' in $\mathfrak{R}_1'$ and A_2' in $\mathfrak{R}_2'$, $A_1' \otimes A_2'$ commutes with each $A_1 \otimes A_2$ in $\mathfrak{R}_1 \otimes \mathfrak{R}_2$ so that $A_1' \otimes A_2'$ lies in $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$. Thus $\mathfrak{R}_1' \otimes \mathfrak{R}_2' \subseteq (\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$. The problem resides in establishing the reverse inclusion.

Suppose E_1' and E_2' are projections in $\mathfrak{R}_1'$, $\mathfrak{R}_2'$, respectively, such that

$$[(\mathfrak{R}_1E_1') \otimes (\mathfrak{R}_2E_2')]' = (\mathfrak{R}_1E_1')' \otimes (\mathfrak{R}_2E_2')'. \tag{4.1}$$

Then

$$\begin{aligned}(E_1' \otimes E_2')(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'(E_1' \otimes E_2') &= (E_1'\mathfrak{R}_1'E_1') \otimes (E_2'\mathfrak{R}_2'E_2') \\ &= (E_1' \otimes E_2')(\mathfrak{R}_1' \otimes \mathfrak{R}_2')(E_1' \otimes E_2'). \end{aligned} \tag{4.2}$$

With $\mathfrak{R}_1$ and $\mathfrak{R}_2$ Abelian, E_1', E_2' as above and cyclic; $\mathfrak{R}_1E_1'$, $\mathfrak{R}_2E_2'$ and $(\mathfrak{R}_1E_1') \otimes (\mathfrak{R}_2E_2')$ are maximal Abelian since each is Abelian and has a cyclic

[15] The range projection F_0 commutes with R'. Cf. J. von Neumann, Math. Ann. **102**, 370 (1929).

[16] See reference 13, especially Sec. 3.1.

[17] See reference 7, Lemma 11.3.2, and reference 9, p. 18, Proposition 1.

vector.[18] Thus (4.1), and, hence, (4.2) hold, in this case. Since the union of projections $E_1' \otimes E_2'$ in $\mathfrak{R}_1' \otimes \mathfrak{R}_2'$, with E_1', E_2' cyclic, is I; $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)' = \mathfrak{R}_1' \otimes \mathfrak{R}_2'$, from Lemma 4, when $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are Abelian—once we note that $\mathfrak{R}_1 \otimes \mathfrak{R}_2$, being Abelian, is its own center as well as that of $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$ and is contained in $\mathfrak{R}_1' \otimes \mathfrak{R}_2'[\subseteq (\mathfrak{R}_1 \otimes \mathfrak{R}_2)']$ and hence in the center of $\mathfrak{R}_1' \otimes \mathfrak{R}_2'$.

For arbitrary von Neumann algebras $\mathfrak{R}_1$, $\mathfrak{R}_2$ with centers $\mathfrak{C}_1$ and $\mathfrak{C}_2$, respectively, the center $\mathfrak{C}$ of $\mathfrak{R}_1 \otimes \mathfrak{R}_2$ is $\mathfrak{C}_1 \otimes \mathfrak{C}_2$. In fact, $\mathfrak{C}_1 \otimes \mathfrak{C}_2 \subseteq \mathfrak{C}$; while $\mathfrak{R}_1 \otimes I \subseteq \mathfrak{R}_1 \otimes \mathfrak{R}_2 \subseteq \mathfrak{C}'$ and $\mathfrak{R}_1' \otimes I \subseteq \mathfrak{R}_1' \otimes \mathfrak{R}_2' \subseteq (\mathfrak{R}_1 \otimes \mathfrak{R}_2)' \subseteq \mathfrak{C}'$. Now, $\mathfrak{R}_1$ and $\mathfrak{R}_1'$ generate $\mathfrak{C}_1'$; so that $\mathfrak{C}_1' \otimes I \subseteq \mathfrak{C}'$. Similarly $I \otimes \mathfrak{C}_2' \subseteq \mathfrak{C}'$. Thus $\mathfrak{C}_1' \otimes \mathfrak{C}_2' = (\mathfrak{C}_1 \otimes \mathfrak{C}_2)' \subseteq \mathfrak{C}'$; and $\mathfrak{C}_1 \otimes \mathfrak{C}_2 \supseteq \mathfrak{C}$. It follows that $\mathfrak{C}_1 \otimes \mathfrak{C}_2 = \mathfrak{C}$. We conclude that $\mathfrak{R}_1' \otimes \mathfrak{R}_2'$ and $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$ have the same center (viz. $\mathfrak{C}_1 \otimes \mathfrak{C}_2$, the center of $\mathfrak{R}_1 \otimes \mathfrak{R}_2$).

Combining this last conclusion with (4.1), (4.2), the comment that $P(\vee_\gamma G_\gamma) = \vee_\gamma PG_\gamma$ when $PG_\gamma = G_\gamma P$ for each γ, and Lemma 4, we see that it suffices to prove

$$[(\mathfrak{R}_1 E_\alpha') \otimes (\mathfrak{R}_2 F_\beta')]' = (\mathfrak{R}_1 E_\alpha')' \otimes (\mathfrak{R}_2 F_\beta)', \qquad (4.3)$$

for all α and β, where $\{E_\alpha'\}$ and $\{F_\beta'\}$ are families of projections in $\mathfrak{R}_1'$ and $\mathfrak{R}_2'$, respectively, with union I. With $\mathfrak{R}_1$ and $\mathfrak{R}_2$ semifinite, $\mathfrak{R}_1'$ and $\mathfrak{R}_2'$ are[19]; and each is generated by its finite cyclic projections. If E' and F' are finite cyclic projections in $\mathfrak{R}_1'$ and $\mathfrak{R}_2'$, respectively, $(\mathfrak{R}_1 E')'(= E' R_1' E')$ and $(\mathfrak{R}_2 F')'$ are finite; and their commutants have cyclic vectors. We may assume, therefore, that $\mathfrak{R}_1'$ and $\mathfrak{R}_2'$ are finite; and that $\mathfrak{R}_1$ and $\mathfrak{R}_2$ have cyclic vectors.

Since $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)' = \mathfrak{R}_1' \otimes \mathfrak{R}_2'$ is equivalent to $\mathfrak{R}_1 \otimes \mathfrak{R}_2 = (\mathfrak{R}_1' \otimes \mathfrak{R}_2')'$, and the finite cyclic projections in $\mathfrak{R}_1$, $\mathfrak{R}_2$ have union I, it suffices to prove $(\mathfrak{R}_1' E)' \otimes (\mathfrak{R}_2' F)' = [(\mathfrak{R}_1' E) \otimes (\mathfrak{R}_2' F)]'$, for all such projections E and F. But now $(\mathfrak{R}_1' E)'$, $(\mathfrak{R}_2' F)'$, $\mathfrak{R}_1' E$, and $\mathfrak{R}_2' F$ are all finite and $\mathfrak{R}_1' E$, $\mathfrak{R}_2' F$ have cyclic vectors. We may assume that $\mathfrak{R}_1$, $\mathfrak{R}_2$, $\mathfrak{R}_1'$, and $\mathfrak{R}_2'$ are finite and $\mathfrak{R}_1$, $\mathfrak{R}_2$ have cyclic vectors x and y, respectively. For each vector z, $D_1([\mathfrak{R}_1' z]) \leq D_1([\mathfrak{R}_1' x])$.[14] But $D_1(I - [\mathfrak{R}_1' z]) = I - D_1([\mathfrak{R}_1' z]) \geq D_1([\mathfrak{R}_1' x]) - D_1([\mathfrak{R}_1' z])$, so that there is a partial isometry V in $\mathfrak{R}_1$ with initial space $[\mathfrak{R}_1' x]$ and final space $V([\mathfrak{R}_1' x]) = [\mathfrak{R}_1' Vx]$ containing $[\mathfrak{R}_1' z]$. Now $[\mathfrak{R}_1 Vx] \supseteq [\mathfrak{R}_1 V^* Vx] = [\mathfrak{R}_1 x]$; so that each cyclic projection in $\mathfrak{R}_1$ is contained in a projection $[\mathfrak{R}_1' w]$, with w cyclic for $\mathfrak{R}_1$. Hence the union of such projections in $\mathfrak{R}_1$ is I. Since the same is true for $\mathfrak{R}_2$, it suffices to prove $(\mathfrak{R}_1'[\mathfrak{R}_1' x] \otimes \mathfrak{R}_2'[\mathfrak{R}_2' y])' = ([\mathfrak{R}_1' x]\mathfrak{R}_1[\mathfrak{R}_1' x]) \otimes ([\mathfrak{R}_2' y]\mathfrak{R}_2[\mathfrak{R}_2' y])$, for all cyclic vectors x for $\mathfrak{R}_1$ and y for $\mathfrak{R}_2$. But $\mathfrak{R}_1'[\mathfrak{R}_1' x]$ and $[\mathfrak{R}_1' x]\mathfrak{R}_1[\mathfrak{R}_1' x]$ are finite with x as cyclic vector for each, while $\mathfrak{R}_2'[\mathfrak{R}_2' y]$ and $[\mathfrak{R}_2' y]\mathfrak{R}_2[\mathfrak{R}_2' y]$ are finite with y as cyclic vector for each.

We may assume $\mathfrak{R}_1$, $\mathfrak{R}_2$, $\mathfrak{R}_1'$, $\mathfrak{R}_2'$ are finite with x a cyclic vector for $\mathfrak{R}_1$, $\mathfrak{R}_1'$, and y a cyclic vector for $\mathfrak{R}_2$, $\mathfrak{R}_2'$. In this case, $x \otimes y$ is cyclic for $\mathfrak{R}_1 \otimes \mathfrak{R}_2$ and $\mathfrak{R}_1' \otimes \mathfrak{R}_2'$ $[\subseteq (\mathfrak{R}_1 \otimes \mathfrak{R}_2)']$; hence for $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$. The product of the center-valued traces[20] on $\mathfrak{R}_1$ and $\mathfrak{R}_2$ extends to a (finite) center-valued trace on $\mathfrak{R}_1 \otimes \mathfrak{R}_2$,[21] so that $\mathfrak{R}_1 \otimes \mathfrak{R}_2$ is finite. Since $\mathfrak{R}_1 \otimes \mathfrak{R}_2$ has a cyclic vector, $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$ is finite. From Lemma 2, $\mathfrak{R}_1' \otimes \mathfrak{R}_2' = (\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$.

Remark. The formula for $(\mathfrak{R}_1 \otimes \mathfrak{R}_2)'$ has not been proved for $\mathfrak{R}_1$ and $\mathfrak{R}_2$ factors of type III.

V. JOINT CYCLIC AND SEPARATING VECTOR—REDUCTION OF THE PROBLEM

The presumption that the cyclic and separating vector of (1.1) does not exist can be cast as a conjecture in many forms. Two variants of this due to Guénin and Misra[6] are listed as:

B_1: If $\mathfrak{R}_1$ is a proper subfactor of $\mathfrak{R}$, both are factors of type I_∞, and ψ is a separating and cyclic vector for both $\mathfrak{R}_1$ and $\mathfrak{R}$, then each minimal projection in $\mathfrak{R}_1$ is finite relative to $\mathfrak{R}$.

B_2: If $\mathfrak{R}_1$ is a proper subfactor of $\mathfrak{R}$ unitarily equivalent to $\mathfrak{R}$, ψ is a separating and cyclic vector for both $\mathfrak{R}_1$ and $\mathfrak{R}$, and $\mathfrak{R}$ is the von Neumann algebra generated by $\mathfrak{R}_1$ and $\mathfrak{R}_1' \cap \mathfrak{R}$; then each finite projection in $\mathfrak{R}_1$ is finite relative to $\mathfrak{R}$.

Under the hypothesis of B_1, $\mathfrak{R}_1' \cap \mathfrak{R}$ is a factor of type I_n (n possibly ∞). The dimension of a minimal projection in $\mathfrak{R}_1$ relative to $\mathfrak{R}$ is n. To see this, note that the situation does not change if we replace $\mathfrak{R}$ by a von Neumann algebra isomorphic to it. Assume, for the moment, that $\mathfrak{R}$ is all bounded operators on some (separable) Hilbert space—so that $\mathfrak{R}_1$ is then a I_∞ factor on this space with I_n commutant $\mathfrak{R}_1'$ $(= \mathfrak{R}_1' \cap \mathfrak{R})$. If E is a minimal projection in $\mathfrak{R}_1$, the mapping $A_1' \to A_1' E$ is an isomorphism (since $\mathfrak{R}_1'$ is a factor) of $\mathfrak{R}_1'$ onto the algebra of all bounded operators acting on E (by minimality of E)— which algebra is, accordingly, of type I_n. Thus E is n-dimensional (with $\mathfrak{R}$ all bounded operators), i.e., E has dimension n relative to $\mathfrak{R}$.

Conjecture B_1 becomes then: $\mathfrak{R}_1' \cap \mathfrak{R}$ cannot be of type I_∞ with ψ a cyclic and separating vector

[18] See reference 15. This can be made to follow from reference 7, Lemma 9.3.3, or reference 9, p. 242, Proposition 3.

[19] This follows from the references of 12, or explicitly in reference 9, p. 101, Corollaire 1.

[20] See reference 9, p. 267, Théorème 3, or R. Kadison, Proc. Am. Math. Soc. 12, 973 (1961).

[21] See reference 9, p. 56, Théorème 2.

for the factor $\mathfrak{R}$ of type I_∞ with type I_∞ commutant, and for the subfactor $\mathfrak{R}_1$ of type I_∞ of $\mathfrak{R}$—i.e., B_1 asserts that (1.1) has a negative answer. Now if $\mathfrak{R}$ and $\mathfrak{S}$ are factors of type I_∞ (on separable Hilbert spaces $\mathfrak{H}$ and $\mathfrak{K}$) each with commutant of type I_∞, each has a separating and cyclic vector[22] and they are unitarily equivalent[23]; viz. there is a unitary transformation U of $\mathfrak{H}$ onto $\mathfrak{K}$ such that the mapping $A \to UAU^{-1}$ of bounded operators on $\mathfrak{H}$ into bounded operators on $\mathfrak{K}$ maps $\mathfrak{R}$ *-isomorphically onto $\mathfrak{S}$. If $\mathfrak{R}_1$ and $\mathfrak{S}_1$ are subfactors of $\mathfrak{R}$ and $\mathfrak{S}$, respectively, of type I_∞, each with commutant relative to $\mathfrak{R}$ and $\mathfrak{S}$ ($\mathfrak{R}_1' \cap \mathfrak{R}$ and $\mathfrak{S}_1' \cap \mathfrak{S}$) of type I_∞, then each has absolute commutant of type I_∞ and so has its own cyclic and separating vector, from the preceding remarks. Moreover, $U\mathfrak{R}_1U^{-1}$ is a type I_∞ subfactor of $\mathfrak{S}$ with relative commutant $(U\mathfrak{R}_1U^{-1})' \cap \mathfrak{S}$ of type I_∞. Again, from the preceding remarks (representing $\mathfrak{S}$ as all bounded operators on some separable space), there is a unitary operator V in $\mathfrak{S}$ such that $VU\mathfrak{R}_1U^{-1}V^{-1} = \mathfrak{S}_1$. Thus VU is a unitary transformation of $\mathfrak{H}$ onto $\mathfrak{K}$ carrying $\mathfrak{R}$ onto $\mathfrak{S}$, $\mathfrak{R}_1$ onto $\mathfrak{S}_1$, and, hence a separating and cyclic vector for $\mathfrak{R}$ and $\mathfrak{R}_1$, if one exists, onto such a vector for $\mathfrak{S}$ and $\mathfrak{S}_1$. Thus, if one such pair $\mathfrak{R}$, $\mathfrak{R}_1$ has a joint separating and cyclic vector, all such pairs do (all being unitarily equivalent to $\mathfrak{R}$ and $\mathfrak{R}_1$).

We have noted that each of $\mathfrak{R}$ and $\mathfrak{R}_1$ has its own cyclic and separating vector. The problem is whether one vector will serve as such for both of them. Suppose x is such a vector. In any event, $\mathfrak{R}$ and $\mathfrak{R}_1$, being of type I_∞ with (absolute) commutant of type I_∞, are unitarily equivalent, as noted above. Further, $\mathfrak{R}$ and $\mathfrak{R}_1$ being factors of type I_∞ implies[24] that $\mathfrak{R}$ is unitarily equivalent to the tensor product of $\mathfrak{R}_1$ and $\mathfrak{R}_1' \cap \mathfrak{R}$—in particular, $\mathfrak{R}$ is generated by $\mathfrak{R}_1$ and $\mathfrak{R}_1' \cap \mathfrak{R}$ [and $(\mathfrak{R}_1' \cap \mathfrak{R})' \cap \mathfrak{R} = \mathfrak{R}_1$]. As noted, the minimal projections of $\mathfrak{R}_1$, which are certainly finite in $\mathfrak{R}_1$, have dimension ∞ relative to $\mathfrak{R}$, with $\mathfrak{R}_1' \cap \mathfrak{R}$ of type I_∞. Thus the example constructed in this and the next section, to show that (1.1) has an affirmative answer, settles both conjectures B_1 and B_2 negatively.

We begin by constructing a factor $\mathfrak{R}$ of type I_∞ and a subfactor $\mathfrak{R}_1$ of type I_∞ with $\mathfrak{R}_1' \cap \mathfrak{R}$ of type I_∞ (which pair will be a "canonical form" for all pairs, by virtue of the preceding remarks). Let $\mathfrak{H}$ be a (fixed) separable Hilbert space, $\mathfrak{B}(\mathfrak{H})$ the algebra of all bounded operators on $\mathfrak{H}$, $\mathfrak{H}'$ the direct sum $\mathfrak{H} \oplus \mathfrak{H} \oplus \cdots$ of $\mathfrak{H}$ with itself a countable number of times, and $\mathfrak{H}''$ the same, with $\mathfrak{H}'$ in place of $\mathfrak{H}$. With T an operator on $\mathfrak{H}$, let $T^\sim$ be the operator on $\mathfrak{H}'$ defined by $T^\sim(x') = (Tx_1, Tx_2, \cdots)$, where x' [$= (x_1, x_2, \cdots)$] is a vector in $\mathfrak{H}'$. Similarly, if $\bar{T}$ is an operator on $\mathfrak{H}'$, we can associate with it an operator $\bar{T}^\sim$ on $\mathfrak{H}''$. In terms of (infinite) matrices with operator entries, $T^\sim$ is the matrix with all off-diagonal entries 0 and each diagonal entry equal to T. Viewed as infinite (operator entry) matrices, the operators on $\mathfrak{H}''$ are infinite matrices each of whose entries is an infinite matrix with entries operators on $\mathfrak{H}$. Thus $\mathfrak{B}(\mathfrak{H})^{\sim\sim}$ is an "infinite copy" of $\mathfrak{B}(\mathfrak{H})^\sim$; and $\mathfrak{B}(\mathfrak{H}')^\sim$, an infinite copy of $\mathfrak{B}(\mathfrak{H}')$, contains $\mathfrak{B}(\mathfrak{H})^{\sim\sim}$. Both are factors of type I_∞ with commutants of type I_∞. Denote $\mathfrak{B}(\mathfrak{H})^{\sim\sim}$ by $\mathfrak{R}_1$ and $\mathfrak{B}(\mathfrak{H}')^\sim$ by $\mathfrak{R}$. The matrices representing operators in $(\mathfrak{B}(\mathfrak{H}')^\sim)'$ have scalar multiples of the identity operator on $\mathfrak{H}'$ as entries. Moreover, $(\mathfrak{B}(\mathfrak{H})^{\sim\sim})' \cap \mathfrak{B}(\mathfrak{H}')^\sim$ $(= \mathfrak{R}_1' \cap \mathfrak{R})$ consists of operators whose matrix representation has each principal (diagonal) infinite matrix block [i.e., operator in $\mathfrak{B}(\mathfrak{H}')$] equal to one infinite matrix, all of whose entries are scalar multiples of the identity operator on $\mathfrak{H}$, and all nonprincipal blocks equal to 0. Thus $\mathfrak{R}_1' \cap \mathfrak{R}$ is and infinite copy of a factor of type I_∞ [viz. $(\mathfrak{B}(\mathfrak{H})^\sim)'$], and is itself a factor of type I_∞.

In the notation of tensor products of Hilbert spaces, $\mathfrak{H}'$ can be identified with $\mathfrak{H} \otimes \mathfrak{H}$, and $\mathfrak{H}''$ with $\mathfrak{H} \otimes \mathfrak{H} \otimes \mathfrak{H}$, $\mathfrak{B}(\mathfrak{H})^\sim$ with $\mathfrak{B}(\mathfrak{H}) \otimes I$, $\mathfrak{B}(\mathfrak{H}')$with $\mathfrak{B}(\mathfrak{H}) \otimes \mathfrak{B}(\mathfrak{H})$, $\mathfrak{B}(\mathfrak{H})^{\sim\sim}$ $(= \mathfrak{R}_1)$ with $\mathfrak{B}(\mathfrak{H}) \otimes I \otimes I$, $\mathfrak{B}(\mathfrak{H}')^\sim (= \mathfrak{R})$ with $\mathfrak{B}(\mathfrak{H}) \otimes \mathfrak{B}(\mathfrak{H}) \otimes I (= \mathfrak{B}(\mathfrak{H}') \otimes I)$, and $\mathfrak{R}_1' \cap \mathfrak{R}$ with $I \otimes \mathfrak{B}(\mathfrak{H}) \otimes I$.

In the development which follows, we shall derive conditions on the set of coordinates of a vector x' in $\mathfrak{H}'$ under which it is a separating vector for $\mathfrak{B}(\mathfrak{H})^\sim$ and conditions under which it is a cyclic vector for $\mathfrak{B}(\mathfrak{H})^\sim$.

Definition 5. A set of vectors $\{x_i\}$ in $\mathfrak{H}$ is said to be an L_2 set when $\sum_{i=1}^\infty \|x_i\|^2 < \infty$. An L_2 set of vectors $\{x_i\}$ in $\mathfrak{H}$ will be said to be L_2-independent when $\sum_{i=1}^\infty \alpha_i x_i = 0$, for α_i with $\sum_{i=1}^\infty |\alpha_i|^2 < \infty$, implies $\alpha_i = 0$ for all j.

Remark 6. The L_2 sets are precisely the possible sets of coordinates of vectors in $\mathfrak{H}'$.

Remark 7. Note that with $\{x_i\}$ an L_2 set, and $\sum_{i=1}^\infty |\alpha_i|^2 < \infty$, $\sum_{i=1}^\infty \alpha_i x_i$ converges absolutely, for $\sum_{i=1}^\infty |\alpha_i| \, \|x_i\| \leq (\sum_{i=1}^\infty |\alpha|^2)^{\frac{1}{2}} (\sum_{i=1}^\infty \|x_i\|^2)^{\frac{1}{2}}$ (by Cauchy–Schwarz).

Lemma 8. The set of vectors $\{x_i\}$ in $\mathfrak{H}$ is L_2-independent if and only if there exists a Hilbert–Schmidt operator T on $\mathfrak{H}$ which is one–one [i.e., *null*

[22] See reference 7, p. 182, Theorem X; or E. L. Griffin, Jr., Trans. Am. Math. Soc. 75, 471 (1953), especially Lemma 1.2.8; or reference 13, Lemma 3.3.6.

[23] See reference 9, p. 233, Théorème 3.

[24] See reference 7, Lemma 3.2.4.

space (0)] and an orthonormal basis $\{y_i\}$ for $\mathfrak{H}$ such that $Ty_i = x_i$, for all j.

Proof: By a unitary equivalence, we may assume that $\mathfrak{H}$ is l_2 (sequence Hilbert space) and $x_i = (\alpha_{1i}, \alpha_{2i}, \cdots)$. Let T be the operator on $\mathfrak{H}$ (so represented) which corresponds to the matrix (α_{ki}) relative to the orthonormal basis $\{y_i\}$, where y_i has jth coordinate 1 and all other coordinates 0. Then T is a Hilbert–Schmidt operator if and only if $\{x_i\}$ is an L_2 set, for $\sum_{i=1}^{\infty} ||x_i||^2 = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} |\alpha_{ki}|^2 =$ Trace (T^*T).

With $\{x_i\}$ an L_2 set, and $z = (\beta_1, \beta_2, \cdots)$, $0 = Tz = (\sum_{i=1}^{\infty} \beta_i\alpha_{1i}, \sum_{i=1}^{\infty} \beta_i\alpha_{2i}, \cdots)$, if and only if $\sum_{i=1}^{\infty} \beta_i\alpha_{ki} = 0$, for all k. This last is the case, if and only if $\sum_{i=1}^{\infty} \beta_i x_i = 0$. Thus T is one–one if and only if $\{x_i\}$ is L_2-independent.

Lemma 9. The vector $x' = (x_1, x_2, \cdots)$ in $\mathfrak{H}'$ is cyclic for $\mathfrak{B}(\mathfrak{H})^{\sim}$ if and only if $\{x_i\}$ is L_2-independent.

For the proof of this, we shall need the following two remarks:

Remark 10. If $\bar{T}$ in $\mathfrak{B}(\mathfrak{H}')$ has $\alpha_{ki}I$ as k, jth entry, and z' in $\mathfrak{H}'$ has all coordinates 0 except the jth, which is some unit vector z in $\mathfrak{H}$, then $\bar{T}z' = (\alpha_{1i}z, \alpha_{2i}z, \cdots)$; so that $||\bar{T}z'||^2 = \sum_{k=1}^{\infty} |\alpha_{ki}|^2 \leq ||\bar{T}||^2\ ||z'||^2 = ||\bar{T}||^2$. Thus all columns of this special $\bar{T}$ are "square summable". Applying this to $\bar{T}^*$, we conclude that all rows of $\bar{T}$ are "square summable".

Remark 11. If $\bar{T}$ has $\beta_i I$ as entry in the first row and jth column, and 0 at all other entries, where $\beta^2 = \sum_{i=1}^{\infty} |\beta_i|^2 < \infty$, then, with $z' = (z_1, z_2, \cdots)$, $\bar{T}z' = (\sum_{i=1}^{\infty} \beta_i z_i, 0, 0, \cdots)$. Thus

$$||\bar{T}z'|| = \left|\left|\sum_{i=1}^{\infty} \beta_i z_i\right|\right| \leq \sum_{i=1}^{\infty} |\beta_i|\ ||z_i|| \leq \left(\sum_{i=1}^{\infty} |\beta_i|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{\infty} ||z_i||^2\right)^{\frac{1}{2}} = \beta\ ||z'||,$$

which establishes both the convergence of $\sum_{i=1}^{\infty} \beta_i z_i$, so that $\bar{T}$ is a well-defined linear operator on $\mathfrak{H}'$, and the boundedness of $\bar{T}$.

Proof of Lemma 9: The vector x' is cylic for $\mathfrak{B}(\mathfrak{H})^{\sim}$ if and only if it is separating for $(\mathfrak{B}(\mathfrak{H})^{\sim})'$.[25] From our earlier comments about the matrix form of an operator $\bar{T}$ in $(\mathfrak{B}(\mathfrak{H})^{\sim})'$, we know that the k, jth entry is $\alpha_{ki}I$, with α_{ki} some scalar. Thus $\bar{T}x' = 0$, if and only if $\sum_{i=1}^{\infty} \alpha_{ki}x_i = 0$, for all k. From Remark 10, $\sum_{i=1}^{\infty} |\alpha_{ki}|^2 < \infty$, for all k. Thus if $\{x_i\}$ is L_2-independent, $\alpha_{ki} = 0$, for all k, j, $\bar{T} = 0$; and x' is separating for $(\mathfrak{B}(\mathfrak{H})^{\sim})'$ and cyclic for $\mathfrak{B}(\mathfrak{H})^{\sim}$.

[25] See reference 9, p. 6, Proposition 5 (note: "totalisateur" replaces "cyclic").

On the other hand, if x' is cyclic for $\mathfrak{B}(\mathfrak{H})^{\sim}$, so, separating for $(\mathfrak{B}(\mathfrak{H})^{\sim})'$, and $\sum_{i=1}^{\infty} \beta_i x_i = 0$, with $\sum_{i=1}^{\infty} |\beta_i|^2 < \infty$, then $\bar{T}$, with $\beta_i I$ as 1, jth entry and all entries 0 in rows other than the first, is bounded, by Remark 11. Hence $\bar{T}$ lies in $(\mathfrak{B}(\mathfrak{H})^{\sim})'$ (by virtue of its matrix form). But $\bar{T}x' = (\sum_{i=1}^{\infty} \beta_i x_i, 0, 0, \cdots) = 0$; so that $\bar{T} = 0$ and $\beta_i = 0$, for all j. Thus $\{x_i\}$ is L_2-independent.

Lemma 12. The vector $x' = (x_1, x_2, \cdots)$ is separating for $\mathfrak{B}(\mathfrak{H})^{\sim}$ if and only if its set of coordinates $\{x_k\}$ spans $\mathfrak{H}$.

Proof: We have $T^{\sim}x' = 0$ if and only if $Tx_k = 0$, for all k, which is the case if and only if T annihilates the subspace M of $\mathfrak{H}$ spanned by $\{x_k\}$. Now T annihilating M is equivalent to T (and hence $T^{\sim}$) being 0, if and only if $M = \mathfrak{H}$.

From Lemmas 9 and 12, we see that $\mathfrak{R}_1$ and $\mathfrak{R}$ have a joint cyclic and separating vector if and only if there is an L_2 set $\{x_k'\}$ in $\mathfrak{H}'$ which spans $\mathfrak{H}'$—so that $x'' = (x_1', x_2', \cdots)$ in $\mathfrak{H}''$ is separating for $\mathfrak{R}$ (and *a fortiori* for $\mathfrak{R}_1$)—such that $\{x_{ki}\}$ is L_2-independent, where $x_k' = (x_{k1}, x_{k2}, \cdots)$—so that x'' is cyclic for $\mathfrak{R}_1$ (and *a fortiori* for $\mathfrak{R}$).

It is useful to view the desired construction in intrinsic form (say in our fixed Hilbert space $\mathfrak{H}$). We ask for a sequence $E_1, E_2, \cdots$ of mutually orthogonal, infinite-dimensional subspaces of $\mathfrak{H}$, an isometry V_i of E_i onto E_1, and an L_2 set $\{x_k\}$ which spans $\mathfrak{H}$ such that $\{V_iE_ix_k\}$ is L_2-independent (in E_1). In this formulation, E_1 replaces $\mathfrak{H}$, $\mathfrak{H}$ replaces $\mathfrak{H}'$ (as a direct sum of the E_i or E_1 with itself a countable number of times by virtue of the isometric identification V_i of E_i with E_1), x_k replaces x_k', and $V_iE_ix_k$ replaces x_{ki}. It is in this form that we establish the existence of a joint separating and cyclic vector, in the next section.

VI. THE CONSTRUCTION

We state the result being proved explicitly as:

Theorem 13. If $\mathfrak{R}$ is a factor of type I_∞ acting on the separable Hilbert space $\mathfrak{H}$, $\mathfrak{R}'$ is of type I_∞ and $\mathfrak{R}_1$ is a subfactor of $\mathfrak{R}$ of type I_∞ with relative commutant $\mathfrak{R}_1' \cap \mathfrak{R}$ a factor of type I_∞, then there is a vector x in $\mathfrak{H}$ which is cyclic and separating for both $\mathfrak{R}_1$ and $\mathfrak{R}$.

Proof: For the purposes of this construction (and from the discussion of the preceding section), we may take $\mathfrak{H}$ in the specific representation $L_2([0, 1])$ (relative to Lebesgue measure). Following the required construction as outlined at the end of the last section, we take x_k to be the function $\gamma \to \gamma^k/k$ (actually, the equivalence class of all square-summable functions which differ from this function at most on a set of measure 0). As E_k,

we choose the subspace of $\mathcal{H}$ consisting of those functions which vanish almost everywhere (a.e.) outside of $[2^{-k}, 2^{-(k-1)}]$. Let $\bar{x}_k = kx_k$.

Note that the transformation U_k' defined by $(U_k'f)(\gamma) = 2^{-(k-1)/2}f(\gamma/2^{k-1})$, for continuous f in E_k, maps this set of functions isometrically onto the set of continuous functions in E_1. Denote by U_k the (unique) extension of U_k' to E_k mapping E_k isometrically onto E_1. Note also that $U_iE_ix_k = k^{-1}2^{-(i-1)(k+\frac{1}{2})}E_1\bar{x}_k$. Let W_k be the operator on $\mathcal{H}$ defined by $W_kf = f_k \cdot f$, for f in $\mathcal{H}$, where f_k is 0 on $[0, \frac{1}{2}]$, 1 on $[\frac{1}{2}, 1-2^{-(k+1)})$, and -1 on $[1-2^{-(k+1)}, 1]$. We note that each W_k maps E_1 isometrically onto itself. Finally, we take V_k to be W_kU_k.

To see that the choices satisfy the desired conditions, observe that $\{x_k\}$ spans $\mathcal{H}$ by virtue of the Weierstrass Polynomial Approximation Theorem. Suppose $\sum_{i,k=1}^{\infty}\alpha_{ik}V_iE_ix_k = 0$, with $\sum_{i,k=1}^{\infty}|\alpha_{ik}|^2 < \infty$. Then $0 = \sum_{i,k=1}^{\infty}\alpha_{ik}k^{-1}2^{-(i-1)(k+\frac{1}{2})}W_iE_1\bar{x}_k = \sum_{i,k=1}^{\infty}\beta_{ik}W_iE_1\bar{x}_k$, with $\sum_{i,k=1}^{\infty}|\beta_{ik}| < \infty$. Thus $0 = \sum_{i=1}^{\infty}W_iy_i$, where $y_i = \sum_{k=1}^{\infty}\beta_{ik}E_1\bar{x}_k$. Now y_i is the (equivalence class of the) restriction to $[\frac{1}{2}, 1]$ of g_i, where

$$g_i(\gamma) = \sum_{k=1}^{\infty}\beta_{ik}\gamma^k, \tag{6.1}$$

so that g_i is analytic on the open unit disk $\mathfrak{D}$ in the plane of complex numbers (since $\sum_{k=1}^{\infty}|\beta_{ik}| < \infty$). Since f_ig_i is in the equivalence class W_iy_i,

$$\text{l.i.m.}\sum_{i=1}^{n}f_ig_i = 0 \tag{6.2}$$

(i.e., the sum $\sum_{i=1}^{\infty}f_ig_i$ converges in L_2 to 0). But f_ig_i is g_i on $[\frac{1}{2}, \frac{3}{4})$; so that $g = \sum_{i=1}^{\infty}g_i$ is 0 (a.e) on $[\frac{1}{2}, \frac{3}{4})$. Since g is analytic on $\mathfrak{D}$, g is 0 on $\mathfrak{D}$. Define g_0 to be 0; and note that f_k is -1 on the interval $[1 - 2^{-(k+1)}, 1 - 2^{-(k+2)})$ $(= a)$, while f_j is 1 on a, for $j = k+1, k+2, \cdots$. Suppose we have established that $g_0, \cdots, g_{k-1}$ are 0; so that $\sum_{i=k}^{\infty}g_i$ is 0. Then, from (6.2), $\lim[\sum_{i=k+1}^{n}g_i - g_k] = 0$ on a; so that $\lim\sum_{i=k}^{n}g_i = 2g_k$ on a. Since g_k is analytic on $\mathfrak{D}$, $g_k = 0$. By induction, each g_k is 0. From (6.1), $\beta_{ik} = 0$, for all j and k. It follows that $\{V_iE_ix_k\}$ is L_2-independent; and the proof is complete.

VII. REGIONS WITH FACTORS NOT OF TYPE *I* (ARAKI)

Araki[11] shows that the von Neumann algebra of local observables associated with a certain region is a factor not of type *I*. He considers the region $\mathcal{O}$ of space–time, the coordinates of whose points satisfy $|x_0| < |x_1|$, $x_1 > 0$, x_2, and x_3, arbitrary (and also the interior of the set of points spacelike with respect to these—for the purpose of the commutant). He notes that $\mathcal{O}$ is invariant under translations in x_2 and x_3, and that the unitary operators associated with such translations have the vacuum ψ_0 as unique invariant state. The von Neumann algebra $\mathfrak{R}$ associated with $\mathcal{O}$ is a factor which has ψ_0 as separating and cyclic vector. From this data, we conclude that $\mathfrak{R}$ is not of type *I*. In fact:

Proposition 14. If $\mathfrak{R}$ is a factor acting on the Hilbert space $\mathcal{H}$, U is a unitary operator which induces a nontrivial automorphism of $\mathfrak{R}$, ψ_0 is separating for $\mathfrak{R}$, and ψ_0 spans the eigenspace for U corresponding to the eigenvalue 1, then $\mathfrak{R}$ is not of type I.

Proof: If $\mathfrak{R}$ is of type *I*, then $U = VW'$ with V in $\mathfrak{R}$ and W' in $\mathfrak{R}'$ (both unitary). (This is well-known: each automorphism of a type *I* factor is inner, as noted in Sec. V; and if V in $\mathfrak{R}$ induces the same automorphism as U, then $W' = V^{-1}U$ commutes with $\mathfrak{R}$.) Since $VW' = W'V$; U, V and W' commute. Thus $V\psi_0 = VU\psi_0 = UV\psi_0$, and by uniqueness, $V\psi_0 = a\psi_0$ with $|a| = 1$. Since ψ_0 is separating for $\mathfrak{R}$, $V = aI$. Thus U $(= aW')$ is in $\mathfrak{R}'$ and induces the trivial (identity) automorphism of $\mathfrak{R}$.

ACKNOWLEDGMENT

It is a pleasure to record our gratitude to A. S. Wightman for many informative conversations on the relation of von Neumann algebras to quantum field theory. Our thanks are also due to H. Araki, E. J. Woods, M. Guénin, and B. Misra for the privilege of seeing prepublication copies of results.

Perturbation Approach to the Electrostatic Problem for Irregularly Shaped Conductors

VICTOR A. ERMA
Giannini Scientific Corporation, Santa Ana, California
(Received 12 June 1963)

A simple and straightforward perturbation method for treating the electrostatic problem of a charged, irregularly shaped conductor is presented. The perturbation solution is generated starting from the zero-order solution for a charged sphere. The method consists of expanding the boundary condition in a Taylor series, which in effect transforms the boundary condition at the irregular boundary into a succession of boundary conditions to be satisfied at the surface of a sphere. The simplicity of the formalism consists further in applying, in a consistent manner, sufficient rather than necessary conditions on the successive correction potentials. First- and second-order expressions for the potential, surface charge density, and capacitance of irregularly shaped conductors, are derived explicitly, and an elementary theorem for the first-order capacitance is obtained. A perturbation expansion for the capacitance valid to all orders is presented. The application of the method is illustrated by calculating the capacitance of several irregularly shaped conductors. Possible generalizations to more complicated boundary-value problems are indicated.

I. INTRODUCTION

MANY of the problems encountered in mathematical physics can be stated in terms of a boundary-value problem, specified by a partial differential equation and the boundary conditions to be satisfied by a certain function on a given closed surface. Unfortunately, in practice, few such boundary-value problems can be solved exactly. This may be due either to the partial differential equation being too complicated to be solved exactly, or to the shape of the boundary surface being such that the boundary conditions cannot be expressed simply in a coordinate system in which the partial differential equation is separable. However, given the exact solution of a particular problem it is possible to generate new solutions for more complex problems by various approximate techniques. Thus, for example, if for a fixed boundary surface the more complicated partial differential equation differs only slightly from one which can be solved exactly, an approximate solution may be found by so-called *perturbation techniques*.[1] The thought lies near, therefore, that analogous perturbation techniques may be developed to treat the converse case where the partial differential equation remains unchanged, but the boundary surface is altered somewhat from one for which an exact solution is available. Some work in this direction has been previously reported.[2] However, the methods developed to date have been exclusively concerned with eigenvalue problems. They are subject to a number of severe restrictions,[2] and encounter significant difficulties of convergence and consistency, particularly in higher orders.[3]

It is our aim here to develop a particularly simple and straightforward perturbation method based on a Taylor expansion of the boundary condition at the perturbed boundary. In the way of illustration, we shall treat the electrostatic problem for irregularly shaped conductors; i.e., we address ourselves to the problem of finding the potential (and other related quantities of interest such as the surface charge density and capacitance) due to a charged perfect conductor in vacuum.[4]

[1] R. Courant, and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I, p. 343.

[2] P. M. Morse, and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Part II, p. 1052.

[3] G. D. Wassermann, *Proc. Cambridge Phil. Soc.*, 44, 251 (1947).

[4] A number of approximate methods for calculating the capacitance of some irregularly shaped conductors have been previously developed and are described by G. Polya and G. Szegö in *Isoperimetric Inequalities in Mathematical Physics*, Annals of Mathematics Studies No. 27 (Princeton University Press, Princeton, New Jersey, 1951). The underlying idea of the methods of Polya and Szegö is the observation that a physical quantity, such as the capacitance, can frequently be approximated by calculating a much more easily computed geometrical quantity associated with the body in question. The approximations themselves are obtained by means of variational principles, i.e., the original boundary-value problem to be solved is replaced by a variational problem which when solved yields upper and/or lower bounds for the capacitance. On the other hand, a perturbation technique such as developed here attacks the complete boundary-value problem directly, and thus yields solutions not only for the capacitance but also for the charge density distribution over the conductor and the potential outside the conductor, for which no variational principles exist. Moreover, in contrast to a variational calculation whose success frequently depends on the physical intuition and guessing ability of the calculator, it is a relatively straightforward matter to extend the perturbation technique to higher-order approximations as well as to more complicated potential problems, such as problems involving prescribed external fields or the presence of several conductors. Finally, while any particular variational principle

The perturbation solution will be generated from the exactly solvable zero-order problem of a charged conducting sphere of radius a. Accordingly, we assume that the equation of the perturbed boundary surface in the spherical coordinates (r, θ, ϕ) can be written in the general form

$$r_s = a[1 + \epsilon F_1(\theta, \phi) + \epsilon^2 F_2(\theta, \phi) + \cdots], \quad (1)$$

where a is an arbitrary constant, ϵ is a smallness parameter, and the $F_n(\theta, \phi)$ are arbitrary, single-valued, continuous functions satisfying the conditions

$$F_n(\theta, 0) = F_n(\theta, 2\pi); \quad |\epsilon F_n(\theta, \phi)| < 1, \quad \begin{matrix} 0 \leq \theta \leq \pi, \\ 0 \leq \phi \leq 2\pi. \end{matrix} \quad (2)$$

Two remarks should be made concerning the representation of arbitrary boundary surfaces by means of Eqs. (1) and (2). First, it would appear at first sight that, in view of the restriction (2), Eq. (1) is capable only of describing irregular surfaces which do not deviate excessively from a spherical shape. However, we must recall that the value a of the radius of the unperturbed sphere and the location of the center of the spherical coordinate system may be defined arbitrarily. Hence, it is clear that all irregular shapes, for which it is possible to locate the center of the coordinate system in such a way that the radius vector to all points on the surface is single-valued,[5] can be described by Eqs. (1) and (2). For example, all simply connected, convex shapes fall into this class. This of course does not alter the fact that the most accurate results with the least amount of labor will be obtained for shapes which are almost spherical.

Secondly, inasmuch as any arbitrary boundary surface may in principle be described with $F_1(\theta, \phi)$ alone, it may seem redundant to include the higher-order terms in Eq. (1). However, as we shall illustrate below, it is frequently more convenient and natural in practice to define the smallness parameter such that the equation of the boundary surface takes the form (1) which explicitly separates the various orders.

In Sec. II we shall treat the first-order theory and give a simple theorem for the capacitance of irregularly shaped conductors to first order. Based on the results of first-order theory, the second-order theory will be developed in Sec. III. Section V is devoted to deriving a perturbation expansion for the capacitance valid to all orders for the special case of cylindrically symmetric surfaces. A number of examples are worked out in the way of illustration in Sec. IV, and possibilities for generalizing the approach to more complicated problems are indicated in the Conclusion.

II. FIRST-ORDER THEORY

We begin with the zero-order problem of a conducting sphere of radius a which is at a potential V_0 and carries a total charge Q. Its solution is well-known to be

$$V = Q/(4\pi\epsilon_0 r), \quad r > a; \qquad \sigma = Q/(4\pi a^2), \quad C \equiv Q/V_0 = 4\pi\epsilon_0 a, \quad (3)$$

where ϵ_0 is the permittivity of the vacuum, and σ and C represent the surface charge density and capacitance, respectively.

For the surface described by (1), we shall write the potential outside the boundary surface to first order in the form

$$V(r, \theta, \phi) = V_s(r) + \epsilon V_1(r, \theta, \phi), \quad (4)$$

where V_s represents the zero-order solution, and V_1 is the first-order correction to the potential. However, instead of assuming V_s to be given by the potential of (3), we shall write V_s in the form

$$V_s = A/r, \quad (5)$$

where the constant A shall purposely remain undetermined for the time being. This amounts to retaining the form but not the exact value of the unperturbed potential.

If we let V_0 represent the constant potential at the surface of the conductor, the boundary condition to be satisfied by V is

$$V(r_s) = V_0 = V_s(r_s) + \epsilon V_1(r_s, \theta, \phi), \quad (6)$$

where r_s is given by Eq. (1). We now expand the terms on the right-hand side (henceforth abbreviated by RHS) of (6) to order ϵ in a Taylor series about the unperturbed boundary $r = a$, obtaining

$$V_0 = V_s(a) + \epsilon[aV_s'(a)F_1(\theta, \phi) + V_1(a, \theta, \phi)], \quad (7)$$

where the prime denotes differentiation with respect to the argument. The above equation requires that the RHS of (7) be independent of both θ and ϕ, i.e., that the partial derivatives of the bracket with respect to θ and ϕ vanish. However, instead

for the capacitance is unique, a perturbation solution may be generated from a number of different exactly solvable cases. Thus, although the approximate solutions for the capacitance obtained here are in some particulars similar to those given by Polya and Szegö, the perturbation technique developed here is more versatile and generally applicable than the corresponding variational methods.

[5] Known in mathematical parlance as surfaces which admit of a *radial single-valued explicit representation.*

of applying this *necessary* condition, we shall apply the much simpler *sufficient* condition that the bracket itself vanish. It now becomes clear why the constant A entering into V_s was left undetermined, since any constant terms inside the bracket are to be considered as incorporated in A. That this procedure of applying a sufficient condition in place of a necessary one in the manner described leads to a consistent and unique result in all orders, will be demonstrated further below.

Applying the stated condition to Eq. (7), we obtain, in view of (5),

$$V_0 = A/a, \tag{8}$$

$$V_1(a, \theta, \phi) = (A/a)F_1(\theta, \phi). \tag{9}$$

The first of these equations merely serves to define A, while the second represents a boundary condition to be satisfied by the first-order correction potential on the surface of a sphere of radius a.

Since outside the conductor V must satisfy Laplace's equation, and since V_s itself satisfies Laplace's equation, it is clear from (4) that V_1 must also separately satisfy Laplace's equation. Consequently, we may write for V_1 the general solution of Laplace's equation appropriate to our problem, which is[6]

$$V_1(r, \theta, \phi) = A \sum_{n,m=0}^{\infty} r^{-(n+1)}(A_{nm} \cos m\phi + B_{nm} \sin m\phi)P_n^m(x), \tag{10}$$

where A_{nm} and B_{nm} are undetermined coefficients, and $P_n^m(x)$ is the associated Legendre function of the argument $x = \cos\theta$. The coefficients A_{nm} and B_{nm} may be determined in the standard manner by substituting (10) into (9) and making use of the appropriate orthogonality properties.[6] Thus, we find

$$A_{nm} = \frac{a^n(2n+1)(n-m)!}{2\pi(1+\delta_{m0})(n+m)!} \times \int_0^{2\pi}\int_{-1}^{1} F_1(\theta, \phi) \cos m\phi P_n^m(x)\, dx\, d\phi, \tag{11}$$

$$B_{nm} = \frac{a^n(2n+1)(n-m)!}{2\pi(n+m)!} \times \int_0^{2\pi}\int_{-1}^{1} F_1(\theta, \phi) \sin m\phi P_n^m(x)\, dx\, d\phi, \tag{12}$$

where δ_{m0} is the Kronecker delta which has the value unity when $m = 0$, and vanishes otherwise. With (11) and (12), the coefficients of (10) may thus be determined for any arbitrary boundary perturbation function $F_1(\theta, \phi)$. It follows from the orthogonality properties of the trigonometric and associated Legendre functions that, if F_1 can be expressed as a finite series of spherical harmonics, the series (10) for $V_1(r, \theta, \phi)$ will likewise be finite.

To complete the solution, it remains to relate the undetermined constant A to the specified value of Q. This is most easily done by observing that regardless of the shape of the conductor, the leading term of the potential must approach $Q/(4\pi\epsilon_0 r)$ as $r \to \infty$. Consequently, Q can be simply defined to be $4\pi\epsilon_0$ times the coefficient of the r^{-1} term of expression (4). In view of (5) and (10) we can thus write

$$Q = 4\pi\epsilon_0 A(1 + \epsilon A_{00}), \tag{13}$$

and in view of (8), the capacitance to first order is given by

$$C \equiv Q/V_0 = 4\pi\epsilon_0 a(1 + \epsilon A_{00}), \tag{14}$$

where A_{00} is given explicitly by Eq. (11) as

$$A_{00} = \frac{1}{4\pi}\int_0^{2\pi}\int_{-1}^{1} F_1(\theta, \phi)\, dx\, d\phi. \tag{15}$$

The result (14) may be given a simple physical interpretation which is contained in the following theorem:

Theorem. The capacitance of an irregularly shaped conductor whose surface is described by Eq. (1) can be written to first order as

$$C = 4\pi\epsilon_0 \bar{r}_s, \tag{16}$$

where $\bar{r}_s$ is the mean radius of the conductor averaged over its surface.

The theorem may be proved by computing $\bar{r}_s$ and comparing the result with Eqs. (14) and (15). It is valid only to first order; no analogous theorem can be derived to second order.

We next derive a first-order expression for the surface charge density on the conductor, which in general can be written as

$$\sigma = \epsilon_0 \mathbf{E}\cdot\mathbf{n}|_{r_s}, \tag{17}$$

where $\mathbf{E}$ is the electric field, and $\mathbf{n}$ is the unit outward normal at the surface of the conductor. The latter can be written as

$$\mathbf{n} = \nabla h/|\nabla h||_{r_s}; \qquad h \equiv r - r_s(\theta, \phi), \tag{18}$$

where $r = r_s(\theta, \phi)$ is the equation of the surface in spherical coordinates. Substituting Eq. (1) to first order into (18) and carrying out the indicated

[6] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950).

operations, we obtain

$$\mathbf{n} = \left[1 + \frac{\epsilon^2 a^2}{r_s^2}\left(\frac{\partial F_1}{\partial \theta}\right)^2 + \frac{\epsilon^2 a^2}{r_s^2 \sin^2\theta}\left(\frac{\partial F_1}{\partial \phi}\right)^2\right]^{-\frac{1}{2}} \times \left(\mathbf{e}_r - \frac{\epsilon a}{r_s}\frac{\partial F_1}{\partial \theta}\mathbf{e}_\theta - \frac{\epsilon a}{r_s \sin\theta}\frac{\partial F_1}{\partial \phi}\mathbf{e}_\phi\right), \quad (19)$$

where $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ are unit vectors in the direction of increasing values of the coordinates (r, θ, ϕ), respectively. Expanded to first order in ϵ, expression (19) becomes

$$\mathbf{n} = \mathbf{e}_r - \epsilon\frac{\partial F_1}{\partial \theta}\mathbf{e}_\theta - \frac{\epsilon}{\sin\theta}\frac{\partial F_1}{\partial \phi}\mathbf{e}_\phi. \quad (20)$$

The electric field is obtained from $\mathbf{E} = -\nabla V$, which from Eq. (4) is found to be

$$\mathbf{E} = -\left(\frac{\partial V_s}{\partial r} + \epsilon\frac{\partial V_1}{\partial r}\right)\mathbf{e}_r - \frac{\epsilon}{r}\frac{\partial V_1}{\partial \theta}\mathbf{e}_\theta - \frac{\epsilon}{r\sin\theta}\frac{\partial V_1}{\partial \phi}\mathbf{e}_\phi. \quad (21)$$

The first-order surface charge density is then obtained by substituting (20) and (21) into (17), and keeping only terms of order ϵ. In this way we find

$$\sigma = -\epsilon_0[(\partial V_s/\partial r) + \epsilon(\partial V_1/\partial r)]_{r_s}. \quad (22)$$

Substituting expressions (5) and (10) into (22), the latter becomes

$$\sigma = \epsilon_0 A\left[\frac{1}{r_s^2} + \epsilon\sum_{n,m=0}^{\infty}\frac{(n+1)}{r_s^{n+2}} \times (A_{nm}\cos m\phi + B_{nm}\sin m\phi)P_n^m(x)\right]. \quad (23)$$

If we then substitute expression (13) for A, and Eq. (1) for r_s, and expand to order ϵ, we finally obtain the following explicit expression for the surface charge density to first order:

$$\sigma = \frac{Q}{4\pi a^2}\left[1 - \epsilon A_{00} - 2\epsilon F_1(\theta,\phi) + \epsilon\sum_{n,m=0}^{\infty}\frac{(n+1)}{a^n} \times (A_{nm}\cos m\phi + B_{nm}\sin m\phi)P_n^m(x)\right]. \quad (24)$$

We note parenthetically that the first-order expansion (20) for $\mathbf{n}$, and hence also expression (24) for σ, are valid only if in addition to condition (2), the function $F_1(\theta, \phi)$ also satisfies the condition

$$|\epsilon\nabla F_1| < 1; \quad 0 \le \theta \le \pi, \quad 0 \le \phi \le 2\pi. \quad (25)$$

The special case of cylindrically symmetric boundary perturbations, for which $F_1 = F_1(\theta)$ is a function of θ alone, corresponds to setting $m = 0$ in the above formulas. The expressions for the first-order correction potential and the capacitance then take the simpler form

$$V_1 = A\sum_{n=0}^{\infty}\frac{A_n}{r^{n+1}}P_n(x); \quad A_n = \frac{(2n+1)a^n}{2}\int_{-1}^{1}F_1(\theta)P_n(x)\,dx, \quad (26)$$

$$C = 4\pi\epsilon_0 a(1 + \epsilon A_0), \quad (27)$$

where $P_n(x)$ is the nth Legendre polynomial of the argument $x = \cos\theta$.

III. SECOND-ORDER THEORY

On the basis of the first-order results obtained in the preceding section, we can now extend the method successively to second and higher orders in an analogous manner. Thus, we write the second-order potential in the form

$$V(r,\theta,\phi) = V_s(r) + \epsilon V_1(r,\theta,\phi) + \epsilon^2 V_2(r,\theta,\phi), \quad (28)$$

where now

$$V_s(r) = B/r, \quad (29)$$

with B again an undefined constant which in general differs from A by order ϵ. The boundary condition which must be satisfied by V on the irregular boundary is now

$$V(r_s,\theta,\phi) = V_0 = V_s(r_s) + \epsilon V_1(r_s,\theta,\phi) + \epsilon^2 V_2(r_s,\theta,\phi), \quad (30)$$

where r_s is given by Eq. (1). As before, we now expand the RHS of (30) to order ϵ^2 in a Taylor series about $r = a$, obtaining

$$V_0 = V_s(a) + \epsilon[V_s'(a)aF_1(\theta,\phi) + V_1(a,\theta,\phi)] + \epsilon^2[V_s'(a)aF_2(\theta,\phi) + \tfrac{1}{2}V_s''(a)a^2F_1^2(\theta,\phi) + (\partial V_1/\partial r)|_a\, aF_1(\theta,\phi) + V_2(a,\theta,\phi)]. \quad (31)$$

The potential V_1 is given as before by expression (10), but with A now replaced by B, i.e.,

$$V_1 = B\sum_{n,m=0}^{\infty}\frac{1}{r^{n+1}} \times (A_{nm}\cos m\phi + B_{nm}\sin m\phi)P_n^m(x), \quad (32)$$

where the coefficients A_{nm}, B_{nm} are given by Eqs. (11) and (12). Consequently, the first bracket on the RHS of (31) vanishes identically as before. In order to assure the overall independence of the RHS of (31) on the variables θ and ϕ, we shall now again impose the sufficient condition that the second bracket also vanish identically. Any constant terms eliminated thereby are assumed to be in-

corporated in the first term. Consequently, Eq. (31) becomes

$$V_0 = B/a, \tag{33}$$

$$V_2(a, \theta, \phi) = -[V_s'(a)aF_2(\theta, \phi) + \tfrac{1}{2}V_s''(a)a^2F_1^2(\theta, \phi) + (\partial V_1/\partial r)|_a\, aF_1(\theta, \phi)]. \tag{34}$$

Again, Eq. (33) thus serves to define B, while (34) represents the boundary condition to be satisfied by $V_2(r, \theta, \phi)$ on a sphere of radius a, since all terms on the RHS of (39) are now explicitly known. Substituting expressions (29) and (32) into (34), we obtain explicitly

$$V_2(a, \theta, \phi) = \frac{B}{a}\left[F_2(\theta, \phi) - F_1^2(\theta, \phi) + \sum_{k,l=0}^{\infty} \frac{(k+1)}{a^k}(A_{kl}\cos l\phi + B_{kl}\sin l\phi)P_k^l(x)F_1(\theta, \phi)\right]. \tag{35}$$

Inasmuch as V, V_s, and V_1 all satisfy Laplace's equation, the same must be true of V_2, so that V_2 may be written in the form

$$V_2(r, \theta, \phi) = B\sum_{n,m=0}^{\infty} r^{-(n+1)} \times (C_{nm}\cos m\phi + D_{nm}\sin m\phi)P_n^m(x), \tag{36}$$

where the coefficients C_{nm} and D_{nm} are determined in the usual manner by

$$C_{nm} = \frac{a^{n+1}(2n+1)(n-m)!}{2\pi B(1+\delta_{m0})(n+m)!} \times \int_0^{2\pi}\int_{-1}^{1} V_2(a, \theta, \phi)\cos m\phi P_n^m(x)\, dx\, d\phi, \tag{37}$$

$$D_{nm} = \frac{a^{n+1}(2n+1)(n-m)!}{2\pi B(n+m)!} \times \int_0^{2\pi}\int_{-1}^{1} V_2(a, \theta, \phi)\sin m\phi P_n^m(x)\, dx\, d\phi, \tag{38}$$

with $V_2(a, \theta, \phi)$ given explicitly by (35). We note that since $V_2(a, \theta, \phi)$ contains B as a multiplying factor, the coefficients C_{nm}, D_{nm} are independent of B. Also, if both F_1 and F_2 can be expanded in a finite series of spherical harmonics, the potential V_2 has likewise only a finite number of terms.

We now again determine B by the requirement that as $r\to\infty$, the leading term of the total potential must approach $Q/(4\pi\epsilon_0 r)$. Collecting the coefficients of the spherically symmetric r^{-1} terms, we obtain in view of Eqs. (28), (29), (32), and (36):

$$Q = 4\pi\epsilon_0 B(1 + \epsilon A_{00} + \epsilon^2 C_{00}), \tag{39}$$

which serves to relate B to Q. Substituting (33) into (39), the capacitance to second order is obtained as

$$C \equiv (Q/V_0) = 4\pi\epsilon_0 a(1 + \epsilon A_{00} + \epsilon^2 C_{00}). \tag{40}$$

The coefficient A_{00} is given by (15); C_{00} can be found explicitly by substituting (35) into (37) for the special case $m = n = 0$, which yields

$$C_{00} = \frac{1}{4\pi}\int_0^{2\pi}\int_{-1}^{1}\left[F_2 - F_1^2 + \sum_{k,l=0}^{\infty}\frac{(k+1)}{a^{2k}} \times (A_{kl}\cos l\phi + B_{kl}\sin l\phi)P_k^l(x)F_1(\theta, \phi)\right] dx\, d\phi. \tag{41}$$

The integrals of the terms occurring in the sum are the same as those defined by Eqs. (11) and (12). Substituting these into (41), C_{00} becomes explicitly

$$C_{00} = \frac{1}{4\pi}\left\{\pi\sum_{k,l=0}^{\infty}\frac{2(k+1)(k+l)!}{(2k+1)a^{2k}(k-l)!} \times [(1+\delta_{l0})A_{kl}^2 + B_{kl}^2] + \int_0^{2\pi}\int_{-1}^{1}(F_2 - F_1^2)\, dx\, d\phi\right\}. \tag{42}$$

Thus, only the coefficients A_{kl}, B_{kl} of the first-order potential are required to find the capacitance to second order.

We shall next derive an expression for the surface charge density σ valid to second order and demonstrate the consistency to second order by showing that the surface integral of σ yields the same value of Q as that given by Eq. (39). The surface charge density is given by expression (17), as before. The electric field at the surface now becomes

$$\mathbf{E} = -\left(\frac{\partial V_s}{\partial r} + \epsilon\frac{\partial V_1}{\partial r} + \epsilon^2\frac{\partial V_2}{\partial r}\right)_{r_s}\mathbf{e}_r - \frac{\epsilon}{r_s}\left(\frac{\partial V_1}{\partial\theta} + \epsilon\frac{\partial V_2}{\partial\theta}\right)_{r_s}\mathbf{e}_\theta - \frac{\epsilon}{r_s\sin\theta}\left(\frac{\partial V_1}{\partial\phi} + \epsilon\frac{\partial V_2}{\partial\phi}\right)_{r_s}\mathbf{e}_\phi. \tag{43}$$

Similarly, expression (19) takes the form

$$\mathbf{n} = \frac{\mathbf{e}_r - \frac{\epsilon a}{r_s}\left(\frac{\partial F_1}{\partial\theta} + \epsilon\frac{\partial F_2}{\partial\theta}\right)\mathbf{e}_\theta - \frac{\epsilon a}{r_s\sin\theta}\left(\frac{\partial F_1}{\partial\phi} + \epsilon\frac{\partial F_2}{\partial\phi}\right)\mathbf{e}_\phi}{\left[1 + \frac{\epsilon^2a^2}{r_s^2}\left(\frac{\partial F_1}{\partial\theta} + \epsilon\frac{\partial F_2}{\partial\theta}\right)^2 + \frac{\epsilon^2a^2}{r_s^2\sin^2\theta}\left(\frac{\partial F_1}{\partial\phi} + \epsilon\frac{\partial F_2}{\partial\phi}\right)^2\right]^{\frac{1}{2}}}. \tag{44}$$

If we now take the dot product of expressions (43) and (44), substitute expressions (29), (32), and (36) for V_s, V_1, and V_2, carry out the indicated differentiations and keep only terms to order ϵ^2, we obtain

$$\sigma = \frac{\epsilon_0 B}{r_s^2}\Bigg[1 + \epsilon\sum_{n,m=0}^{\infty}\frac{(n+1)}{r_s^n} \times (A_{nm}\cos m\phi + B_{nm}\sin m\phi)P_n^m(x) - \frac{\epsilon^2}{2}\left(\frac{\partial F_1}{\partial\theta}\right)^2 - \frac{\epsilon^2}{2\sin^2\theta}\left(\frac{\partial F_1}{\partial\phi}\right)^2 + \epsilon^2\sum_{n,m=0}^{\infty}\frac{(n+1)}{a^n}(C_{nm}\cos m\phi + D_{nm}\sin m\phi)P_n^m(x) + \epsilon^2\sum_{n,m=0}^{\infty}\frac{1}{a^n}(A_{nm}\cos m\phi + B_{nm}\sin m\phi)\frac{dP_n^m(x)}{d\theta}\frac{\partial F_1}{\partial\theta} + \epsilon^2\sum_{n,m=0}^{\infty}\frac{m}{a^n}(-A_{nm}\sin m\phi + B_{nm}\cos m\phi)\frac{P_n^m(x)}{(1-x^2)}\frac{\partial F_1}{\partial\phi}\Bigg]. \tag{45}$$

An explicit expression for σ valid to order ϵ^2 may be trivially obtained from (45) by expanding the two terms containing r_s consistently to order ϵ^2.

The total charge is obtained by integrating expression (45) over the surface of the conductor, i.e.,

$$Q = \int_S \sigma\, ds = \int_{-1}^{1}\int_0^{2\pi}\sigma r_s^2\left[1 + \frac{\epsilon^2}{\sin^2\theta}\left(\frac{\partial F_1}{\partial\phi}\right)^2\right]^{\frac{1}{2}} \times\left[1 + \epsilon^2\left(\frac{\partial F_1}{\partial\theta}\right)^2\right]^{\frac{1}{2}} d\phi\, dx, \tag{46}$$

where the surface area element on the RHS of (46) has been written only to second order in ϵ. Substituting (45) into (42), and expanding to order ϵ^2, we thus obtain

$$Q = B\epsilon_0\int_{-1}^{1}\int_0^{2\pi} d\phi\, dx\Bigg\{1 + \epsilon\sum_{n,m=0}^{\infty}\frac{(n+1)}{a^n} \times (A_{nm}\cos m\phi + B_{nm}\sin m\phi)P_n^m(x) - \epsilon^2\sum_{n,m=0}^{\infty}\frac{n(n+1)}{a^n}(A_{nm}\cos m\phi + B_{nm}\sin m\phi)P_n^m(x)F_1(\theta,\phi) + \epsilon^2\sum_{n,m=0}^{\infty}\frac{(n+1)}{a^n}(C_{nm}\cos m\phi + D_{nm}\sin m\phi)P_n^m(x) + \epsilon^2\sum_{n,m=0}^{\infty}\frac{1}{a^n}(A_{nm}\cos m\phi + B_{nm}\sin m\phi)\frac{dP_n^m(x)}{d\theta}\frac{\partial F_1}{\partial\theta} + \epsilon^2\sum_{n,m=0}^{\infty}\frac{m}{a^n}(-A_{nm}\sin m\phi + B_{nm}\cos m\phi)\frac{P_n^m(x)}{(1-x^2)}\frac{\partial F_1}{\partial\phi}\Bigg\}. \tag{47}$$

We first note that of the first and third infinite series on the RHS of (47), only the $m = n = 0$ terms contribute, due to the orthogonality of the spherical harmonics. We next integrate the terms of the fourth infinite series by parts with respect to x; recalling that $x = \cos\theta$, we then have

$$\int_{-1}^{1}\frac{dP_n^m(x)}{d\theta}\frac{\partial F_1}{\partial\theta}\,dx = \int_{-1}^{1}(1-x^2)\frac{dP_n^m(x)}{dx}\frac{\partial F_1}{\partial x}\,dx = -\int_{-1}^{1}F_1\frac{d}{dx}\left[(1-x^2)\frac{dP_n^m(x)}{dx}\right]dx. \tag{48}$$

Similarly, we integrate the terms of the fifth series by parts with respect to ϕ, obtaining

$$\int_0^{2\pi}(-A_{nm}\sin m\phi + B_{nm}\cos m\phi)\frac{\partial F_1}{\partial\phi}\,d\phi = m\int_0^{2\pi}(A_{nm}\cos m\phi + B_{nm}\sin m\phi)F_1\,d\phi. \tag{49}$$

If we now substitute (48) and (49) into expression (47), the latter reduces to

$$Q = 4\pi\epsilon_0 B(1 + \epsilon A_{00} + \epsilon^2 C_{00}) - \epsilon^2 B\epsilon_0\sum_{n,m=0}^{\infty}\int_{-1}^{1}\int_0^{2\pi}\frac{1}{a^n}(A_{nm}\cos m\phi + B_{nm}\sin m\phi)F_1\Bigg\{\frac{d}{dx}\left[(1-x^2)\frac{dP_n^m(x)}{dx}\right] + n(n+1)P_n^m(x) - \frac{m^2}{(1-x^2)}P_n^m(x)\Bigg\}d\phi\,dx. \tag{50}$$

However, the integrands of the last term on the RHS of (50) vanish identically by virtue of the associated Legendre equation[6]; consequently (50) gives a result for Q which is consistent with our previous expression (39).

It is also possible to give a more general argument for the consistency of our method to arbitrary order on the basis of the uniqueness theorem. It is clear that by applying the *sufficient* condition of setting the successive terms on the RHS of the Taylor-expanded boundary condition [Eqs. (7), (31), and so on] equal to zero, while leaving the constant of the zero-order solution undetermined, we guarantee that the potential satisfies the boundary condition at the surface of the perturbed boundary. Similarly, by indentifying $Q/(4\pi\epsilon_0)$ with the sum of the coefficients of the r^{-1} terms, we automatically guarantee the correct behavior at infinity. Accordingly, the method described, consistently yields a solution of Laplace's equation which satisfies the prescribed

boundary condition on a closed regular surface and has the proper behavior at infinity, and must therefore be unique by virtue of the uniqueness theorem for harmonic functions.[7] While we have dealt explicitly only with the first- and second-order perturbations, it is clear that higher-order expressions for the potential, charge density, and capacitance can be successively obtained in an exactly analogous manner.

For ready reference, we here give without further comment the corresponding second-order formulas for the case where F_1 and F_2 are functions of θ alone. We then have

$$\begin{aligned} V_2(r,\theta) &= B\sum_{n=0}^{\infty} C_n r^{-(n+1)} P_n(x); \\ C_n &= \frac{(2n+1)a^{n+1}}{2B}\int_{-1}^{1} V_2(a,\theta)P_n(x)\,dx, \\ V_2(a,\theta) &= \frac{B}{a}\left[F_2(\theta) - F_1^2(\theta) + \sum_{k=0}^{\infty}\frac{(k+1)}{a^k}A_k P_k(x)F_1(\theta)\right], \end{aligned} \tag{51}$$

where the first-order coefficients are given by (26). The capacitance is

$$C = 4\pi\epsilon_0 a(1 + \epsilon A_0 + \epsilon^2 C_0), \tag{52}$$

with

$$C_0 = \sum_{k=0}^{\infty}\frac{(k+1)A_k^2}{(2k+1)a^{2k}} + \frac{1}{2}\int_{-1}^{1}(F_2 - F_1^2)\,dx. \tag{53}$$

IV. EXAMPLES

For purposes of illustrating the application of the above method, we first choose to calculate the capacitance of an arbitrary ellipsoid to second order, inasmuch as the result may be compared to an exact expression obtained by other means. We denote the semiaxes of the ellipsoid by a, $a(1+\epsilon)$, and $a(1+\beta\epsilon)$, where ϵ is a smallness parameter and β is arbitrary, as long as $|\beta\epsilon| < 1$. By expanding to order ϵ^2, the equation of the ellipsoid may then be put into the form (1), with

$$F_1(\theta,\phi) = \sin^2\theta\sin^2\phi + \beta\cos^2\theta, \tag{54}$$

$$F_2(\theta,\phi) = \tfrac{3}{2}(\sin^4\theta\sin^4\phi + 2\beta\sin^2\theta\cos^2\theta\sin^2\phi + \beta^2\cos^4\theta - \sin^2\theta\sin^2\phi - \beta^2\cos^2\theta). \tag{55}$$

The coefficient A_{00} of expression (40) for the capacitance may be found directly from Eq. (11), which yields

$$A_{00} = \tfrac{1}{3}(1+\beta). \tag{56}$$

[7] O. D. Kellogg, *Foundations of Potential Theory* (Dover Publications, Inc., New York, 1953) p. 216.

In order to determine C_{00}, we must find the remaining coefficients A_{nm} and B_{nm} by means of Eqs. (11) and (12). The only nonvanishing coefficients are found to be

$$A_{20} = (\tfrac{1}{3}a^2)(2\beta - 1); \qquad A_{22} = -\tfrac{1}{6}a^2. \tag{57}$$

If F_1, F_2, and the above coefficients are substituted into (42), we obtain

$$C_{00} = -\tfrac{1}{45}(1 - \beta + \beta^2). \tag{58}$$

Together with (56) and (58), Eq. (40) then yields the following second-order result for the capacitance of the ellipsoid:

$$C = 4\pi\epsilon_0 a[1 + \tfrac{1}{3}\epsilon(1+\beta) - \tfrac{1}{45}\epsilon^2(1 - \beta + \beta^2)]. \tag{59}$$

This may be compared with the exact result given by Smythe[6] who finds the capacitance of an ellipsoid to be given by the elliptic integral

$$C = 8\pi\epsilon_0\left\{\int_0^{\infty}[(a^2+\theta)(b^2+\theta)(c^2+\theta)]^{-\frac{1}{2}}\,d\theta\right\}^{-1}. \tag{60}$$

After substituting $b = a(1+\epsilon)$, $c = a(1+\beta\epsilon)$, expanding to order ϵ^2, and integrating, we are led to a second-order result in exact agreement with (59).

For purposes of further illustration, we have calculated the capacitance of a number of other irregularly shaped conductors. The boundaries considered, together with the corresponding results, are presented in Fig. 1. For reasons of simplicity, the examples chosen correspond to boundary perturbations which could be expressed by $F_1(\theta)$ alone. As can be seen from the results, the successive correction terms to the capacitance decrease quite rapidly, for the most part.

V. THE CAPACITANCE TO ARBITRARY ORDER

Inasmuch as our perturbation method yields a particularly simple expression for the capacitance of an irregularly shaped conductor, since it involves only the $n = m = 0$ coefficients to each order, we shall here derive an expression for the capacitance valid to all orders in ϵ. The derivation shall be carried out for the case where the boundary perturbation is a function of θ alone; the generalization to perturbations which are functions of both θ and ϕ is obvious. We shall not here repeat in detail the justification of the steps of the method.

We begin by writing the equation of the perturbed boundary in the form

$$\begin{aligned} r_s &= a\left(1 + \sum_{n=1}^{\infty}\epsilon^n F_n(\theta)\right); \\ &\left|\sum_{n=1}^{\infty}\epsilon^n F_n(\theta)\right| < 1, \qquad 0 \le \theta \le \pi. \end{aligned} \tag{61}$$

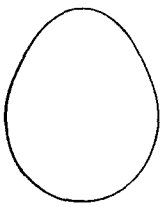
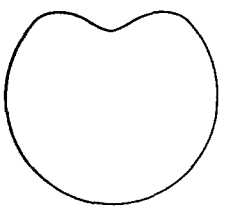
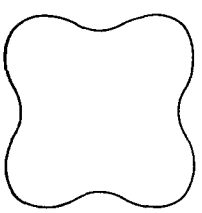
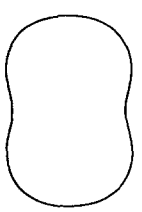
BOUNDARY SHAPE	PERTURBING FUNCTION	CAPACITANCE
 $\epsilon = 0.4$	$F_1(\theta) = \cos^2\theta - \cos\theta$ $= x^2 - x$	$C = 4\pi\epsilon_0 a\left(1+\frac{1}{3}\epsilon+\frac{23}{45}\epsilon^2\right)$
 $\epsilon = 0.2$	$F_1(\theta) = \begin{cases} -(1+\cos 4\theta), & 0\le\theta\le\pi/4 \\ 0, & \pi/4\le\theta\le\pi \end{cases}$	$C = 4\pi\epsilon_0 a\left[1-\frac{(7-4\sqrt{2})}{15}\epsilon\right]$ $\approx 4\pi\epsilon_0 a(1-0.0357\epsilon)$
 $\epsilon = 0.2$	$F_1(\theta) = -\cos 4\theta$ $= -(8x^4 - 8x^2 + 1)$	$C = 4\pi\epsilon_0 a\left(1+\frac{1}{15}\epsilon+\frac{18,944}{11,025}\epsilon^2\right)$ $\approx 4\pi\epsilon_0 a\left(1+\frac{1}{15}\epsilon+1.72\epsilon^2\right)$
 $\epsilon = 0.4$	$F_1(\theta) = -\sin^4\theta$ $= -(1 - 2x^2 + x^4)$	$C = 4\pi\epsilon_0 a\left(1-\frac{8}{15}\epsilon+\frac{2,816}{11,025}\epsilon^2\right)$ $\approx 4\pi\epsilon_0 a\left(1-\frac{8}{15}\epsilon+0.255\epsilon^2\right)$

FIG. 1. Perturbation calculations of the capacitance of several cylindrically symmetric, irregularly shaped conductors. The polar equation $r_s = a(1 + \epsilon F_1(\theta))$ of the surface of the conductor is plotted in the first column; the corresponding function $F_1(\theta)$ is given in the second column ($x = \cos\theta$), and the capacitances calculated are presented in the third column.

Similarly, the potential is written as

$$V = B\left(\frac{1}{r} + \sum_{n=1}^{\infty} \epsilon^n V_n(r, \theta)\right), \tag{62}$$

where the nth-order correction potential can be expanded in the form

$$V_n(r, \theta) = \sum_{k=0}^{\infty} \frac{A_{nk}}{r^{k+1}} P_k(x); \qquad x = \cos\theta. \tag{63}$$

The charge Q on the conductor is then given by

$$Q = 4\pi\epsilon_0 B\left(1 + \sum_{n=1}^{\infty} \epsilon^n A_{n0}\right). \tag{64}$$

If the constant potential on the surface is denoted by V_0, the boundary condition to be satisfied by V is $V(r_s) = V_0$. Accordingly, we have

$$V_0 = B\left[\frac{1}{r_s} + \sum_{n=1}^{\infty} \epsilon^n \sum_{k=0}^{\infty} \frac{A_{nk}}{r_s^{k+1}} P_k(x)\right]. \tag{65}$$

It is clear that if (61) is substituted into (65), and the resulting expression expanded as an infinite power series in ϵ, the result can be written formally in the form

$$V_0 = B/a + B \sum_{q=1}^{\infty} \epsilon^q B_q(x). \tag{66}$$

As before, we then set

$$V_0 = B/a, \tag{67}$$

and require

$$B_q(x) \equiv 0; \qquad q = 1, 2, 3, \cdots. \tag{68}$$

In order to obtain $B_q(x)$ explicitly, we substitute Eq. (61) into (65) and expand the result in a power series in ϵ. On the basis of the binomial theorem, we may write

$$\begin{aligned}\frac{1}{r_s} &= \frac{1}{a} + \sum_{l=1}^{\infty} (-1)^l \left(\sum_{n=1}^{\infty} \epsilon^n F_n\right)^l \\ &= \frac{1}{a} + \sum_{l=1}^{\infty} (-1)^l \sum_{n_1, n_2, \cdots n_l = 1}^{\infty} \epsilon^{S(n_l)} F_{n_1} F_{n_2} \cdots F_{n_l},\end{aligned} \tag{69}$$

where the second sum on the RHS of (69) represents and l-fold infinite sum, and where

$$S(n_l) = n_1 + n_2 + \cdots + n_l. \tag{70}$$

Similarly,

$$\begin{aligned}\frac{1}{r_s^{k+1}} = \frac{1}{a} \sum_{l=0}^{\infty} \frac{(-1)^l}{a^k} \binom{k+l}{l} \\ \times \sum_{n_1, n_2, \cdots n_l = 1}^{\infty} \epsilon^{S(n_l)} F_{n_1} F_{n_2} \cdots F_{n_l},\end{aligned} \tag{71}$$

where

$$\binom{k+l}{l} = \frac{(k+l)!}{k!\, l!}. \tag{72}$$

Substituting (71) and (69) into (65), we obtain

$$\begin{aligned}V_0 = B/a + B\Bigg\{ &\sum_{l=1}^{\infty} (-1)^l \\ &\times \sum_{n_1, n_2, \cdots n_l = 1}^{\infty} \epsilon^{S(n_l)} F_{n_1} F_{n_2} \cdots F_{n_l} \\ &+ \sum_{n=1}^{\infty} \epsilon^n \sum_{k=0}^{\infty} \frac{A_{nk}}{a^k} P_k(x) \sum_{l=0}^{\infty} (-1)^l \binom{k+l}{l} \\ &\times \sum_{n_1, n_2, \cdots n_l = 1}^{\infty} \epsilon^{S(n_l)} F_{n_1} F_{n_2} \cdots F_{n_l} \Bigg\}.\end{aligned} \tag{73}$$

Comparing with (66), we see that in order to find $B_q(x)$ we must isolate the coefficient of ϵ^q on the RHS of (75). If this is done, and the condition (68) applied, we find

$$\begin{aligned}B_q(x) = 0 = &\sum_{l=1}^{q} (-1)^l \sum_{S(n_l) = q} F_{n_1} F_{n_2} \cdots F_{n_l} \\ &+ \sum_{n=1}^{q} \sum_{k=0}^{\infty} \frac{A_{nk}}{a^k} P_k(x) \sum_{l=0}^{(q-n)} (-1)^l \binom{k+l}{l} \\ &\times \sum_{S(n_l) = (q-n)} F_{n_1} F_{n_2} \cdots F_{n_l},\end{aligned} \tag{74}$$

where $\sum_{S(n_l)=q}$ represents an l-fold sum over all values $n_1, n_2, \cdots, n_l = 1, 2, 3 \cdots$ such that $S(n_l) = q$. The coefficients A_{nk} may now be successively determined by means of Eq. (74). Thus, if we isolate the term corresponding to $n = q$ in Eq. (74), the latter may be written in the form

$$\sum_{k=0}^{\infty} \frac{A_{qk}}{a^k} P_k(x) = C_q(x), \tag{75}$$

where

$$\begin{aligned}C_q(x) = -\Bigg\{ &\sum_{l=1}^{q} (-1)^l \sum_{S(n_l) = q} F_{n_1} F_{n_2} \cdots F_{n_l} \\ &+ \sum_{n=1}^{(q-1)} \sum_{k=0}^{\infty} \frac{A_{nk} P_k(x)}{a^k} \sum_{l=0}^{(q-n)} (-1)^l \binom{k+l}{l} \\ &\times \sum_{S(n_l) = q-n} F_{n_1} F_{n_2} \cdots F_{n_l} \Bigg\},\end{aligned} \tag{76}$$

which contains coefficients A_{nk} up to order $n = (q-1)$ only. In view of (75), the coefficients A_{qk} are then obtained in the usual manner by

$$A_{qk} = \tfrac{1}{2}(2k+1) a^k \int_{-1}^{1} C_q(x) P_k(x)\, dx. \tag{77}$$

The only coefficients entering into the capacitance are those corresponding to $k = 0$, which are given explicitly by

$$A_{q0} = \tfrac{1}{2} \int_{-1}^{1} C_q(x). \tag{78}$$

Finally, in cases where the boundary-surface perturbation is expressed in terms of the single function $F_1(\theta)$, expression (76) takes the much simpler form

$$\begin{aligned}C_q(x) = -\Bigg[&(-1)^q F_1^q + \sum_{n=1}^{(q-1)} \sum_{k=0}^{\infty} \frac{A_{nk}}{a^k} P_k(x) (-1)^{(q-n)} \\ &\times \binom{k+q-n}{q-n} F_1^{(q-n)} \Bigg].\end{aligned} \tag{79}$$

VI. CONCLUSION

In the preceding we have developed a very simple and direct perturbation approach for solving a specific boundary-value problem with a perturbed

boundary shape. The principle of the method is twofold. First, by expanding the boundary condition in a Taylor series about the unperturbed boundary, the boundary condition to be satisfied at the perturbed surface is transformed into a succession of boundary conditions at the unperturbed surface. Secondly, by applying sufficient rather than necessary conditions in a consistent manner, the perturbation formalism is appreciably simplified.

The specific problem which we treated was the electrostatic problem for irregularly shaped charged conductors in vacuum, starting with the zero-order solution for a charged sphere. It is clear that the same method may be applied with only minor modifications to more complicated problems of electrostatics, such as a conductor in an external field, several conductors, etc. This merely requires starting with a different expression for the unperturbed potential V_s. Similarly, it is by no means necessary to generate the perturbation expansion from a sphere as the zero-order boundary; in practical cases it may be more convenient to express the irregular boundary as a perturbation of some other boundary, such as a cylinder for example, for which an exact solution V_s is available.

Finally, we remark that, while we have addressed our attention to the electrostatic problem which consists essentially of solving Laplace's equation with Dirichlet boundary conditions, analogous perturbation methods should be applicable to more complicated partial differential equations with more general boundary conditions. An analogous treatment of the scattering of electromagnetic radiation from soft objects is under investigation.

Theory of Transport Coefficients in Low-Density Gases*

JAMES A. MCLENNAN, JR.
Department of Physics, Lehigh University, Bethlehem, Pennsylvania
AND
ROBERT J. SWENSON
Faculté des Sciences de l'Université Libre de Bruxelles, Bruxelles, Belgium
(Received 22 April 1963)

The evaluation of the correlation-function expressions for transport coefficients is discussed. It is shown that for the case of a low-density nondegenerate monatomic gas with short-range interactions, results are obtained which are in agreement with those usually found from the Boltzmann equation. The analysis makes use of a generalized master equation, and of the factorization theorem of Kac. While the main concern is with low-density gases, the methods developed have a wider range of application.

I. INTRODUCTION

THE theory of transport coefficients has traditionally been based on the Boltzmann equation.[1] However, in recent years, considerable attention has been focused on a radically different and more general approach. As was first shown by Green,[2] it is possible to obtain expressions for the thermal conductivity, viscosity, and coefficients of diffusion which are generally valid, and which thus do not suffer from the limitations inherent in the Boltzmann equation. These expressions take the form of integrals of time-dependent correlation functions. Green's results have subsequently been derived by other authors,[3] and have also been extended to provide a general statistical description of hydrodynamic processes.[4] In addition, Kubo,[5] Nakano,[6] and others have obtained a correlation-function expression for the electrical conductivity.

The generality of the correlation-function expressions is compensated for by complexity; although formally simple in appearance, they involve the dynamics of a system with many degrees of freedom. Consequently it is necessary to develop suitable approximation methods in order to obtain results of practical value. Here we wish to consider the evaluation of the transport coefficients for a low-density nondegenerate gas of point molecules, and to show that the Green formulas give, for this case, the same results as the Boltzmann equation. We have two motives for such a calculation: First, it is clearly a necessary preliminary to any extension into regions where the Boltzmann equation is not valid. Second, we have occasionally encountered a suspicion that the Green formulas are not completely determined, but that somewhere in the calculation it might become necessary, in order to obtain a well-defined result, to introduce an assumption akin to coarse graining or molecular chaos, that is, an assumption of the sort familiar in derivations of the Boltzmann equation. We wish to show that this is not the case, but that the results usually obtained from the Boltzmann equation follow in a straightforward manner on taking the low-density limit, no other assumptions or approximations being necessary.

The evaluation of the correlation-function expressions has been considered previously by a number of authors. Dicussions of the electrical conductivity have been given by Kohn and Luttinger,[7] Lax,[8] Chester and Thellung,[9] Montroll and Ward,[10] Verboven,[11] Langer,[12] Matsudaira,[13] and Fujita and Abe.[14] The lattice contribution to the thermal con-

* Research supported by the National Science Foundation.

[1] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, New York, 1952).

[2] M. S. Green, J. Chem. Phys. **20**, 1281 (1952); **22**, 398 (1954).

[3] H. Mori, Phys. Rev. **112**, 1829 (1958); R. Kubo, M. Yokota, and S. Nakajima, J. Phys. Soc. Japan **12**, 1203 (1957); J. G. Kirkwood in *Rendiconti della Scuola Internazionale di Fisica "Enrico Fermi," Corso X* (Nicola Zanichelli, Bologna, 1959); E. W. Montroll, *ibid.*

[4] J. A. McLennan, Phys. Fluids **3**, 493 (1960); **4**, 1319 (1961); *Advances in Chemical Physics*, edited by I. Prigogine (Interscience Publishers, Inc., New York, 1963), Vol. V.

[5] R. Kubo, J. Phys. Soc. Japan, **12**, 570 (1957).

[6] H. Nakano, Progr. Theoret. Phys. (Kyoto) **15**, 77 (1956); **17**, 145 (1957).

[7] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957); **109**, 1892 (1958).

[8] M. Lax, Phys. Rev. **109**, 1921 (1958).

[9] G. V. Chester and A. Thellung, Proc. Phys. Soc. (London) **73**, 745 (1959).

[10] E. W. Montroll and J. Ward, Physica **25**, 423 (1959).

[11] E. Verboven, Physica **26**, 1091 (1960).

[12] J. Langer, Phys. Rev. **120**, 714 (1960); **124**, 1003 (1961); **127**, 5 (1962); **128**, 110 (1962).

[13] N. Matsudaira, Progr. Theoret. Phys. (Kyoto) **25**, 153 (1961).

[14] S. Fujita and R. Abe, J. Math. Phys. **3**, 350 (1962).

ductivity of solids has been treated by Schieve and Peterson,[15] Hardy,[16] and Schieve and Hardy.[17] With a few exceptions, these discussions deal with a one-particle problem, with dissipation being introduced through randomly distributed scattering centers, and as a consequence they do not touch on the second question mentioned in the previous paragraph.

Transport in low-density gases has been discussed by M. S. Green[2] and by H. S. Green,[18] who assume the validity of the Boltzmann equation. Mori[19] has given a discussion based on the master equation; his treatment is approximate except for Maxwellian molecules. Recently Fujita[20] has treated the same problem without the introduction of those assumptions (other than low density) on which the Boltzmann and master equations are based. Our discussion is very similar in its aim to that of Fujita. It differs in that we use a generalized master equation,[21] and consequently do not require a diagrammatic expansion or asymptotic time integration; in addition, we give an explicit treatment of the factorization problem.

II. REDUCTION TO TWO-BODY OPERATORS

The thermal conductivity, viscosity, and coefficients of diffusion can all be treated in a similar manner; we will work with the thermal conductivity as a typical example. Green's formula for the thermal conductivity is

$$\lambda = \frac{1}{3VkT^2}\int_{-\infty}^{0} e^{\epsilon t}\, dt\langle \mathbf{S}\cdot\mathbf{S}(t)\rangle_0, \qquad (1)$$

where V is the volume, k Boltzmann's constant, T the Kelvin temperature, $\mathbf{S}$ and $\mathbf{S}(t)$ are fluxes whose detailed form will be given below, and the brackets $\langle\ \rangle_0$ denote an equilibrium average.

The factor $e^{\epsilon t}$ has been introduced to represent the turning on of an interaction with external reservoirs which bring about the temperature gradient. In order to obtain a thermal conductivity which is independent of ϵ, that is, independent of the past history of the system, it is necessary to take the limit $\epsilon \to 0$. However, this limit must be preceded by the limit of infinite volume,[22] and consequently Eq. (1) is to be interpreted as

$$\lambda = \lim_{\epsilon\to 0}\lim_{V\to\infty}\frac{1}{3VkT^2}\int_{-\infty}^{0} e^{\epsilon t}\, dt\langle \mathbf{S}\cdot\mathbf{S}(t)\rangle_0. \qquad (2)$$

The limit $V \to \infty$ is to be taken with the intensive parameters such as density and temperature held constant.

It is to be emphasized that the limits of Eq. (2) are in no sense essential to the definition of thermal conductivity. For finite V and nonvanishing ϵ, there is a dependence of λ on ϵ and V. The dependence on V can be attributed to boundary effects; for small samples such boundary effects are well known.[23] For large samples the dependence on V is certainly extremely small. The dependence on ϵ is a reflection of the fact, which should not be unexpected, that the thermal conductivity is not strictly a material property, but depends upon the manner in which the nonequilibrium state has been prepared. It is expected that the dependence on ϵ will show a plateau behavior (in the same sense used by Kirkwood[22]): over a wide range of values for ϵ, λ will be sensibly independent of ϵ. However, we will not attempt to give a detailed estimate of the error involved in applying results obtained from the limiting form (2) to a finite system.

We remark that the low-density limit can be taken only after the limit $V \to \infty$. If the density is made vanishingly small with finite volume, a Knudsen gas is obtained and the concept of thermal conductivity is no longer applicable.

The average denoted by $\langle\ \rangle_0$ is to be taken over a canonical (or grand canonical) ensemble,

$$\rho = (1/Z)e^{-H/kT}. \qquad (3)$$

For a low-density gas the potential energy contributions to ρ can be neglected,

$$\rho = (1/Z_0)e^{-H_0/kT}. \qquad (4)$$

Here H_0 is the free-particle Hamiltonian, and Z_0

[15] W. C. Schieve and R. L. Peterson, Phys. Rev. **126**, 1458 (1962).
[16] R. J. Hardy, thesis, Lehigh University, 1962 (unpublished).
[17] W. C. Schieve and R. J. Hardy, Bull. Am. Phys. Soc. **8**, 15 (1963).
[18] H. S. Green, J. Math. Phys. **2**, 344 (1961).
[19] H. Mori, Phys. Rev. **111**, 694 (1958); also H. Nishimura and H. Mori, Progr. Theoret. Phys. (Kyoto) **26**, 967 (1961).
[20] S. Fujita, J. Math. Phys. **3**, 359 (1962).
[21] L. Van Hove, Physica **21**, 517 (1955); **23**, 441 (1957); I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962). A simplified derivation, not based on perturbation expansions, has been given by R. J. Swenson, J. Math. Phys. **3**, 1017 (1962). It is important for our purposes to recognize that the latter derivation yields a generalized master equation which is exact for finite systems.
[22] It can be shown (see the last item of reference 4) that one obtains a vanishing result if the order of the limits is reversed. This was first recognized by Kirkwood, reference 3; J. Chem. Phys. **14**, 180 (1946); **15**, 72 (1947), the point being that ". . . there will exist Poincaré cycle periods within which each orbit will be traversed in the reverse sense to any desired degree of accuracy, canceling its initial contribution . . .".
[23] As an example, the effect of boundary scattering of phonons on the thermal conductivity of a solid has been discussed by H. B. G. Casimir, Physica **5**, 495 (1938); see also P. Carruthers, Rev. Mod. Phys. **33**, 92 (1961).

the sum-over-states

$$Z_0 = [(2\pi\hbar)^3/(2\pi mkT)^{\frac{3}{2}}V]^N, \tag{5}$$

m being the mass of a particle and N the total number of particles. The flux $\mathbf{S}$ is

$$\mathbf{S} = \sum_i \mathbf{s}(\mathbf{p}_i), \qquad \mathbf{s}(\mathbf{p}) = [(p^2/2m) - \tfrac{5}{2}kT]\mathbf{p}, \tag{6}$$

where $\mathbf{p}_i$ is the momentum of particle i. (In general, $\mathbf{S}$ contains additional contributions from the potential energy, but these are negligible at low density.) Although we will ignore the effects of degeneracy, it is nevertheless convenient to use the formalism of quantum theory. $\mathbf{S}$ then becomes an operator, but since it depends only on the momenta, it is diagonal in momentum representation. The quantity $\mathbf{S}(t)$ is given by

$$\mathbf{S}(t) = e^{iHt/\hbar}\mathbf{S}e^{-iHt/\hbar}. \tag{7}$$

Denote the free-particle momentum eigenstates by

$$|\mathbf{P}\rangle = |\mathbf{p}_1, \cdots, \mathbf{p}_N\rangle, \tag{8}$$

with $\mathbf{P}$ used as an abbreviation for the N momenta $\mathbf{p}_1, \cdots, \mathbf{p}_N$. Since ρ and $\mathbf{S}$ are diagonal in this representation, Eq. (1) becomes

$$\lambda = (1/3VkT^2) \sum_{\mathbf{P}} \sum_{\mathbf{P}'} \rho(\mathbf{P})\mathbf{S}(\mathbf{P})\cdot\mathbf{S}(\mathbf{P}')I(\mathbf{P}, \mathbf{P}'), \tag{9}$$

where

$$I(\mathbf{P}, \mathbf{P}') = \int_{-\infty}^{0} e^{\epsilon t}\, dt\langle\mathbf{P}\,|e^{(i/\hbar)Ht}|\,\mathbf{P}'\rangle \times \langle\mathbf{P}'\,|e^{-(i/\hbar)Ht}|\,\mathbf{P}\rangle, \tag{10}$$

and

$$\rho(\mathbf{P}) = \langle\mathbf{P}\,|\rho|\,\mathbf{P}\rangle, \qquad \mathbf{S}(\mathbf{P}) = \langle\mathbf{P}\,|\mathbf{S}|\,\mathbf{P}\rangle. \tag{11}$$

One's first reaction at this point is to apply a perturbation expansion. However, the terms in such an expansion are inversely proportional to powers of ϵ, and it is necessary to perform a resummation before the limit $\epsilon \to 0$ can be taken. Such a procedure is possible, and indeed is the method adopted by Fujita[20] and others. However we will proceed differently, using an analysis based on the generalized master equation.[21]

Let one of the factors in Eq. (10), say the second, be represented by a contour integral,

$$\langle\mathbf{P}'\,|e^{-(i/\hbar)Ht}\,|\mathbf{P}\rangle = -\frac{1}{2\pi i}\int_C dz e^{-(i/\hbar)zt}\langle\mathbf{P}'|\,R_z\,|\mathbf{P}\rangle, \tag{12}$$

where R_z is the resolvent operator,

$$R_z = (H - z)^{-1}, \tag{13}$$

and the contour C encloses the poles of $\langle\mathbf{P}'|\,R_z\,|\mathbf{P}\rangle$. Since these all lie on the real axis, we may take C to consist of the two lines $(-\infty - i\eta, \infty - i\eta)$ and $(\infty + i\eta, -\infty + i\eta)$. If $\mathrm{Im}\, z > -\hbar\epsilon$, or $\eta < \hbar\epsilon$, the time integral in Eq. (10) can be interchanged with the contour integral to yield

$$I(\mathbf{P}, \mathbf{P}') = \frac{\hbar}{2\pi}\int_C dz\langle\mathbf{P}\,|R_{z+i\hbar\epsilon}|\,\mathbf{P}'\rangle\langle\mathbf{P}'\,|R_z|\,\mathbf{P}\rangle. \tag{14}$$

The contour consists of one line lying between a row of poles along the real axis and a row displaced $\hbar\epsilon$ below the real axis, plus a line above the real axis. Since the integrand decreases as $1/|z|^2$ for large $|z|$, the upper line can be removed to infinity and ignored; we are then left with

$$I(\mathbf{P}, \mathbf{P}') = \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE X(\mathbf{P}, \mathbf{P}'), \tag{15}$$

where

$$X(\mathbf{P}, \mathbf{P}') = \langle\mathbf{P}\,|R_{E-i\eta+i\hbar\epsilon}|\,\mathbf{P}'\rangle\langle\mathbf{P}'\,|R_{E-i\eta}|\,\mathbf{P}\rangle. \tag{16}$$

It is convenient to put $\eta = \frac{1}{2}\hbar\epsilon$, which is consistent with the restriction $\eta < \hbar\epsilon$, and we then have

$$X(\mathbf{P}, \mathbf{P}') = \langle\mathbf{P}\,|R_{E+i\eta}|\,\mathbf{P}'\rangle\langle\mathbf{P}'\,|R_{E-i\eta}|\,\mathbf{P}\rangle. \tag{17}$$

The quantity X is conveniently determined from an identity related to the generalized master equation.[21] Introducing

$$X_{ll'}(\mathbf{P}, \mathbf{P}') = \langle\mathbf{P}'\,|R_l|\,\mathbf{P}\rangle\langle\mathbf{P}\,|R_{l'}|\,\mathbf{P}'\rangle, \tag{18}$$

we have for the identity in question

$$\begin{aligned}(l - l')X_{ll'}(\mathbf{P}, \mathbf{P}') &= F_{ll'}(\mathbf{P})\delta_{\mathbf{PP}'} \\ &+ \sum_{\mathbf{P}''}[F_{ll'}(\mathbf{P})W_{ll'}(\mathbf{P}, \mathbf{P}'')X_{ll'}(\mathbf{P}'', \mathbf{P}') \\ &- F_{ll'}(\mathbf{P}'')W_{ll'}(\mathbf{P}'', \mathbf{P})X_{ll'}(\mathbf{P}, \mathbf{P}')].\end{aligned} \tag{19}$$

Here $\delta_{\mathbf{PP}'}$ is a Kronecker delta, and the quantities F and W will be defined below. For the case at hand, Eq. (19) becomes

$$\begin{aligned}-2i\eta X(\mathbf{P}, \mathbf{P}') &= F(\mathbf{P})\delta_{\mathbf{PP}'} \\ &+ \sum_{\mathbf{P}''}[F(\mathbf{P})W(\mathbf{P}, \mathbf{P}'')X(\mathbf{P}'', \mathbf{P}') \\ &- F(\mathbf{P}'')W(\mathbf{P}'', \mathbf{P})X(\mathbf{P}, \mathbf{P}')].\end{aligned} \tag{20}$$

Here for simplicity we have dropped the subscripts l and l' on F and W, remembering that they take on the values $l = E - i\eta$, $l' = E + i\eta$.

The utility of the identity (20) will be found in the fact that the low-density limit (or, more generally, the dependence on density) of the quantities F and W can be readily put into a convenient form;

X is then left as the solution to Eq. (20) rather than being determined directly.

We now proceed to determine the low-density limit of F and W. These quantities are defined as follows. Divide the resolvent into a diagonal part D plus a nondiagonal part (diagonal and nondiagonal referring to the free-particle representation),

$$R = (1 + DU)D. \tag{21}$$

For simplicity, the argument z has been dropped. Define a diagonal operator G by

$$D = (H_0 + G - z)^{-1}. \tag{22}$$

It follows[21] that U and G satisfy the equations

$$\begin{aligned} G &= [H_1 + H_1DU]_{\mathrm{d}}, \\ U &= -[H_1 + H_1DU - GDU]_{\mathrm{nd}}, \end{aligned} \tag{23}$$

where the subscripts d and nd denote the diagonal and nondiagonal parts, respectively, and

$$H_1 = H - H_0 \tag{24}$$

is the potential energy of interaction. F is defined by

$$F_{ll'}(\mathbf{P}) = D_l(\mathbf{P}) - D_{l'}(\mathbf{P}), \tag{25}$$

and W satisfies the equation

$$\begin{aligned} W_{ll'}(\mathbf{P}', \mathbf{P}) &= U_l(\mathbf{P}, \mathbf{P}')U_{l'}(\mathbf{P}', \mathbf{P}) \\ &\quad - \sum_{\mathbf{P}''} W_{ll'}(\mathbf{P}', \mathbf{P}'')D_l(\mathbf{P}'')D_{l'}(\mathbf{P}'') \\ &\quad \times U_l(\mathbf{P}, \mathbf{P}'')U_{l'}(\mathbf{P}'', \mathbf{P}). \end{aligned} \tag{26}$$

Here

$$D_l(\mathbf{P}) = \langle \mathbf{P} \,|D_l|\, \mathbf{P}\rangle, \tag{27}$$

and

$$U_l(\mathbf{P}, \mathbf{P}') = \langle \mathbf{P} \,|U_l|\, \mathbf{P}'\rangle. \tag{28}$$

At this point we need an expansion in two-body scattering operators. Such expansions have been discussed by a number of authors,[24] but for the sake of completeness we include a brief derivation. Combine the two Eqs. (23) into[25]

$$U = (G - H_1)(1 + DU). \tag{29}$$

On introducing an operator K defined by

$$K = H_1 - KDH_1, \tag{30}$$

we find

$$(1 - KD)U = [(1 - KD)G - K](1 + DU), \tag{31}$$

or

$$U = -K + (1 - KD)G(1 + DU). \tag{32}$$

Dividing Eq. (31) by $(1 - KD)$ and taking the diagonal part, we find an equation for G,

$$G = [(1 - KD)^{-1}K(1 + DU)]_{\mathrm{d}}. \tag{33}$$

Equations (32) and (33) determine U and G in terms of K. Assume H_1 to be a sum of two-body potentials[26]

$$H_1 = \sum_{\mu} v_{\mu}, \tag{34}$$

where μ denotes a pair of particles, the sum extends over all pairs, and v_{μ} will be assumed to depend only on the distance between particles. Define scattering operators t_{μ} by

$$t_{\mu} = v_{\mu} - t_{\mu}Dv_{\mu}. \tag{35}$$

(Note that the t_{μ} are actually not two-body operators, since the whole system is involved in D. However this complication will disappear in the low-density limit.) Introduce, in addition, the quantities K_{μ} defined by

$$K_{\mu} = v_{\mu} - v_{\mu}DK. \tag{36}$$

Then

$$K = \sum_{\mu} K_{\mu}, \tag{37}$$

and

$$\begin{aligned} K_{\mu} - t_{\mu}DK_{\mu} &= (1 - t_{\mu}D)v_{\mu}(1 - DK) \\ &= t_{\mu}(1 - DK), \end{aligned}$$

or

$$\begin{aligned} K_{\mu} &= t_{\mu} - t_{\mu}D(K - K_{\mu}) \\ &= t_{\mu} - t_{\mu}D \sum_{\lambda \neq \mu} K_{\lambda}. \end{aligned} \tag{38}$$

Iteration of Eq. (38) and substitution into Eq. (37) gives a t-matrix expansion for K; this combined with Eqs. (32) and (33) yields the desired expansions for U and G.

We now suppose t_{μ} to be proportional to a force range, or scattering length.[27] Since we ignore degeneracy, the thermal de Broglie wavelength is not available as a parameter, and so the only quantities

[24] K. Watson, Phys. Rev. **103**, 489 (1956); A. J. F. Siegert and E. Teramoto, Phys. Rev. **110**, 1232 (1958); J. Weinstock, Phys. Rev. **126**, 341 (1962); R. J. Swenson, J. Math. Phys. **4**, 544 (1963).

[25] Our procedure here is suggested by that of W. Heitler, *The Quantum Theory of Radiation* (Oxford University Press, London, 1954), 3rd ed., Sec. 16 and Appendix 4.

[26] We will impose periodic boundary conditions in order not to have to include wall forces in the Hamiltonian.

[27] Cf., e.g., J. M. J. Van Leeuwen and A. S. Reiner, Physica **27**, 99 (1961).

with the dimension of length are the mean distance between particles and the scattering length. For dimensional reasons, low density is then equivalent to small scattering length, or small t_μ. (Note that our analysis is not applicable to the Coulomb interaction, which is not characterized by a range. In addition the existence of bound states invalidates our analysis since then the t-matrix no longer describes a simple scattering process, and the effects of internal excitation and dissociation must be taken into account.) For small t_μ, the first iterate of Eq. (38) suffices,

$$K_\mu = t_\mu, \qquad K = \sum_\mu t_\mu. \tag{39}$$

Equations (32) and (33) give, to lowest order in K,

$$G = K_d, \qquad U = -K + G = -K_{nd}. \tag{40}$$

On retaining the lowest-order nonvanishing terms of F and W, we now find

$$F_{ll'}(\mathbf{P}) = d_l(\mathbf{P}) - d_{l'}(\mathbf{P}), \tag{41}$$

and

$$W_{ll'}(\mathbf{P}', \mathbf{P}) = U_l(\mathbf{P}, \mathbf{P}')U_{l'}(\mathbf{P}', \mathbf{P}) = \langle \mathbf{P} | K_{l\,nd} | \mathbf{P}' \rangle \langle \mathbf{P}' | K_{l'\,nd} | \mathbf{P} \rangle. \tag{42}$$

Here d_l is the noninteracting resolvent,

$$d_l = (H_0 - l)^{-1}, \tag{43}$$

and

$$d_l(\mathbf{P}) = \langle \mathbf{P} | d_l | \mathbf{P} \rangle = (E_\mathbf{P} - l)^{-1}, \tag{44}$$

where $E_\mathbf{P}$ is the noninteracting energy associated with state $|\mathbf{P}\rangle$,

$$E_\mathbf{P} = \sum_i p_i^2/2m. \tag{45}$$

For the particular values of l and l' needed, F reduces to

$$F(\mathbf{P}) = -2i\eta/[(E_\mathbf{P} - E)^2 + \eta^2]. \tag{46}$$

In the limit $\eta \to 0$, this becomes

$$\lim_{\eta \to 0} F(\mathbf{P}) = -2\pi i \delta(E_\mathbf{P} - E). \tag{47}$$

It is premature at this point to take the limit, but in anticipation we write

$$F(\mathbf{P}) = -2\pi i \delta_\eta(E_\mathbf{P} - E). \tag{48}$$

At this point we summarize briefly by combining Eqs. (9), (15), (20), and (48) to obtain for λ

$$\lambda = \frac{1}{3VkT^2} \sum_\mathbf{P} \rho(\mathbf{P})\mathbf{S}(\mathbf{P}) \times \int_{-\infty}^{\infty} dE\ \delta_\eta(E_\mathbf{P} - E)\mathfrak{S}(\mathbf{P}), \tag{49}$$

where $\mathfrak{S}(\mathbf{P})$ is defined by

$$\frac{\hbar}{2\pi} \sum_{\mathbf{P}'} X(\mathbf{P}, \mathbf{P}')\mathbf{S}(\mathbf{P}') = \delta_\eta(E_\mathbf{P} - E)\mathfrak{S}(\mathbf{P}), \tag{50}$$

and obeys the equation

$$\mathbf{S}(\mathbf{P}) - \epsilon\mathfrak{S}(\mathbf{P}) = \frac{2\pi}{\hbar} \sum_{\mathbf{P}'} \delta_\eta(E_{\mathbf{P}'} - E) \times [W(\mathbf{P}', \mathbf{P})\mathfrak{S}(\mathbf{P}) - W(\mathbf{P}, \mathbf{P}')\mathfrak{S}(\mathbf{P}')]. \tag{51}$$

Here we have reintroduced $\epsilon = 2\eta/\hbar$. It is convenient to shift the origin of the E integration by an amount $E_\mathbf{p}$; we then find for Eqs. (49) and (51) [with a redefined $\mathfrak{S}(\mathbf{P})$]

$$\lambda = \frac{1}{3VkT^2} \sum_\mathbf{P} \rho(\mathbf{P})\mathbf{S}(\mathbf{P}) \int_{-\infty}^{\infty} dE\ \delta_\eta(E)\mathfrak{S}(\mathbf{P}), \tag{52}$$

$$\mathbf{S}(\mathbf{P}) - \epsilon\mathfrak{S}(\mathbf{P}) = \frac{2\pi}{\hbar} \sum_{\mathbf{P}'} \delta_\eta(E_{\mathbf{P}'} - E_\mathbf{P} - E) \times [W(\mathbf{P}', \mathbf{P})\mathfrak{S}(\mathbf{P}) - W(\mathbf{P}, \mathbf{P}')\mathfrak{S}(\mathbf{P}')]. \tag{53}$$

We now return to the quantity W. In the sum of Eq. (53), the term $\mathbf{P}' = \mathbf{P}$ actually gives no contribution, and consequently the restriction nd in Eq. (42) can be dropped:

$$W_{ll'}(\mathbf{P}', \mathbf{P}) = \langle \mathbf{P} | K_l | \mathbf{P}' \rangle \langle \mathbf{P}' | K_{l'} | \mathbf{P} \rangle. \tag{54}$$

Furthermore, at this point we have

$$K = \sum_\mu t_\mu, \tag{55}$$

with

$$t_\mu = v_\mu - t_\mu d_l v_\mu. \tag{56}$$

Denote t_μ for $l = E + E_\mathbf{P} - i\eta$, $l' = E + E_\mathbf{P} + i\eta$ by t_μ^+ (recall the change of variables $E - E_\mathbf{P} \to E$),

$$t_\mu^+ = v_\mu - t_\mu^+ d_{E+E_\mathbf{P}+i\eta} v_\mu = v_\mu - t_\mu^+ (H_0 - E_\mathbf{P} - E \pm i\eta)^{-1} v_\mu. \tag{57}$$

It is convenient to change notation, denoting the pair of molecules by i, j rather than μ,

$$t_{ij}^+ = v_{ij} - t_{ij}^+ (H_0 - E_\mathbf{P} - E \pm i\eta)^{-1} v_{ij}. \tag{58}$$

A sum over μ such as (55) becomes

$$K = \sum_{i<j} t_{ij}\ . \tag{59}$$

The matrix elements of t_{ij} ,taken between free-particle states, will be diagonal in all momenta except those for particles i and j. Furthermore, the terms in H_0 referring to particles other than i and j can be replaced by their eigenvalues. Thus

$$\langle \mathbf{P} | t_{ij}^+ | \mathbf{P}' \rangle = \langle \mathbf{p}_i, \mathbf{p}_j | t_{ij}^+ | \mathbf{p}_i', \mathbf{p}_j' \rangle \delta_{ij}(\mathbf{P}, \mathbf{P}'), \tag{60}$$

where $|\mathbf{p}_i, \mathbf{p}_j\rangle$ denotes the momentum eigenvectors for the two particles i and j, $\delta_{ij}(\mathbf{P}, \mathbf{P}')$ is defined by

$$\delta_{ij}(\mathbf{P}, \mathbf{P}') = \prod_{k\neq i,j} \delta_{\mathbf{p}_k,\mathbf{p}_k'}, \tag{61}$$

and $l_{ij}^{\pm}$ satisfy the equations

$$l_{ij}^{\pm} = v_{ij} - l_{ij}^{\pm}(h_{ij} - \epsilon_{ij} - E \pm i\eta)^{-1}v_{ij}. \tag{62}$$

Here h_{ij} is the noninteracting Hamiltonian for the two particles i and j, and ϵ_{ij} is its eigenvalue for the state $|\mathbf{P}\rangle$,

$$\epsilon_{ij} = (p_i^2 + p_j^2)/2m. \tag{63}$$

On substituting the above results into Eq. (54), we obtain W as a sum of two-particle quantities,

$$W_{ll'}(\mathbf{P}', \mathbf{P}) = \sum_{i<j} \delta_{ij}(\mathbf{P}, \mathbf{P}')\langle \mathbf{p}_i, \mathbf{p}_j| \, l_{ij}^{+} \, |\mathbf{p}_i', \mathbf{p}_j'\rangle \times \langle \mathbf{p}_i', \mathbf{p}_j'| \, l_{ij}^{-} \, |\mathbf{p}_i, \mathbf{p}_j\rangle. \tag{64}$$

Since l_{ij}^{+} and l_{ij}^{-} are Hermitian conjugates, one can also write the above result as

$$W_{ll'}(\mathbf{P}', \mathbf{P}) = \sum_{i<j} \delta_{ij}(\mathbf{P}, \mathbf{P}') \times |\langle \mathbf{p}_i, \mathbf{p}_j| \, l_{ij}^{+} \, |\mathbf{p}_i', \mathbf{p}_j'\rangle|^2. \tag{65}$$

The above matrix element contains a momentum-conserving Kronecker delta. Put

$$\langle \mathbf{p}_i, \mathbf{p}_j| \, l_{ij}^{+} \, |\mathbf{p}_i', \mathbf{p}_j'\rangle = \frac{1}{V} T_{ij} \, \delta^{\mathrm{Kr}}(\mathbf{p}_i + \mathbf{p}_j - \mathbf{p}_i' - \mathbf{p}_j'). \tag{66}$$

The differential scattering cross section (in the relative system of coordinates) is given by[28]

$$\sigma_{ij} = (m^2/16\pi^2\hbar^4) \lim_{\eta\to 0} \lim_{V\to\infty} |T_{ij}|^2. \tag{67}$$

The cross section and T_{ij} are, of course, functions of the initial and final momenta, but for simplicity we have not explicitly indicated this dependence. [Actually it is necessary to put $E = 0$, $\epsilon_{ij} = \epsilon_{ij}'$ in Eq. (67). However in the equations to follow these restrictions will be automatically imposed when the limits $V \to \infty$, $\eta \to 0$, are taken.] As in Eq. (2), the two limits must be taken in the order indicated; otherwise the existence of boundaries will have an effect on the scattering process.[22,29] At this point, before the limits are taken, we write

$$|T_{ij}|^2 = (16\pi^2\hbar^4/m^2)\sigma_{ij}(E, \eta). \tag{68}$$

[28] Cf., e.g., E. Merzbacher, *Quantum Mechanics* (John Wiley & Sons, Inc., New York, 1961), Chap. 21.

[29] In order for Eq. (68) to give a close approximation to the cross section, it is necessary that $1/\epsilon$ be large compared to the time taken by a collision, but less than the Poincaré recurrence time for a two-body collision. Within this range for ϵ, the thermal conductivity will take on its plateau value.

We then have for W

$$W_{ll'}(\mathbf{P}', \mathbf{P}) = (16\pi^2\hbar^4/m^2V^2) \sum_{i<j} \delta_{ij}(\mathbf{P}, \mathbf{P}') \times \delta^{\mathrm{Kr}}(\mathbf{p}_i + \mathbf{p}_j - \mathbf{p}_i' - \mathbf{p}_j')\sigma_{ij}(E, \eta). \tag{69}$$

Since we have considered interactions v_{ij} which depend only on the distance between particles, $\sigma_{ij}(E, \eta)$ satisfies the condition of detailed balancing: it is unaffected by an interchange of the primed and unprimed momenta. Consequently W satisfies

$$W_{ll'}(\mathbf{P}', \mathbf{P}) = W_{ll'}(\mathbf{P}, \mathbf{P}'). \tag{70}$$

We can now rewrite Eq. (53) in a form involving only two-body operators,

$$\mathbf{S}(\mathbf{P}) - \epsilon\mathfrak{S}(\mathbf{P}) = \Omega\mathfrak{S}(\mathbf{P}), \tag{71}$$

where

$$\Omega = \frac{1}{V} \sum_{i<j} \Omega_{ij}, \tag{72}$$

with

$$\Omega_{ij} = (32\pi^3\hbar^3/m^2V) \sum_{\mathbf{p}_i',\mathbf{p}_j'} \delta_\eta(\epsilon_{ij}' - \epsilon_{ij} - E) \times \delta^{\mathrm{Kr}}(\mathbf{p}_i + \mathbf{p}_j - \mathbf{p}_i' - \mathbf{p}_j')\sigma_{ij}(E, \eta)[1 - A_{ij}]. \tag{73}$$

Here we have introduced the symbol A_{ij} which converts $\mathbf{p}_i$, $\mathbf{p}_j$, into $\mathbf{p}_i'$, $\mathbf{p}_j'$, leaving the remaining momenta unchanged:

$$A_{ij}\mathfrak{S}(\mathbf{P}) = \mathfrak{S}(\mathbf{p}_1, \cdots, \mathbf{p}_{i-1}, \mathbf{p}_i', \mathbf{p}_{i+1}, \cdots, \mathbf{p}_{j-1}, \mathbf{p}_j', \mathbf{p}_{j+1}, \cdots, \mathbf{p}_N). \tag{74}$$

Certain properties of Ω_{ij} follow immediately: it is symmetric in i and j,

$$\Omega_{ij} = \Omega_{ji}. \tag{75}$$

When operating on a function which is independent of $\mathbf{p}_i$ and $\mathbf{p}_j$, Ω_{ij} gives zero,

$$\Omega_{ij}f(\mathbf{p}_k) = 0, \qquad k \neq i, j. \tag{76}$$

For large V (or in the classical limit), sums over momenta can be replaced by integrals according to

$$[(2\pi\hbar)^3/V] \sum_{\mathbf{p}} \to \int d^3\mathbf{p}, \tag{77}$$

and in addition

$$[V/(2\pi\hbar)^3]\delta^{\mathrm{Kr}}(\mathbf{p}) \to \delta(\mathbf{p}). \tag{78}$$

Here, and in the following, we will make the above replacements, but with the understanding that $\int d^3\mathbf{p}$ stands for an integral, and $\delta(\mathbf{p})$ for a Dirac delta function, only after the limit $V \to \infty$ (or $\hbar \to 0$)

has been performed. We therefore write for Ω_{ij}

$$\Omega_{ij}\mathfrak{f} = \frac{4}{m^2}\int d^3\mathbf{p}_i' \int d^3\mathbf{p}_j' \, \delta_\eta(\epsilon_i' + \epsilon_j' - \epsilon_i - E)$$
$$\delta(\mathbf{p}_i + \mathbf{p}_j - \mathbf{p}_i' - \mathbf{p}_j')\sigma_{ij}(E, \eta)[1 - A_{ij}]. \quad (79)$$

We observe that Ω_{ij} is independent of volume for large volume; it was for this reason that the factor $1/V$ was introduced in Eq. (72).

It is readily verified that for $E = 0$ and $\eta \to 0$, Ω_{ij} is self-adjoint in the sense that

$$\int d^3\mathbf{p}_i \int d^3\mathbf{p}_i f(\mathbf{p}_i) f(\mathbf{p}_j) F \Omega_{ij} G$$
$$= \int d^3\mathbf{p}_i \int d^3\mathbf{p}_j f(\mathbf{p}_i) f(\mathbf{p}_j) G \Omega_{ij} F. \quad (80)$$

Here F and G are arbitrary functions of the momenta, and $f(\mathbf{p})$ is the Maxwell–Boltzmann distribution function,

$$f(\mathbf{p}) = (1/2\pi mkT)^{\frac{3}{2}} \exp(-p^2/2mkT). \quad (81)$$

We emphasize that the property (80) does not hold for nonvanishing E or η.

III. FACTORIZATION

If Eq. (71) is combined directly with (52), the result for λ is identical with what would be obtained from an unfactorized Boltzmann equation. More specifically, the combination $\mathbf{A}(\mathbf{p}_1) + \mathbf{A}(\mathbf{p}_2)$ appearing in the Boltzmann operator of Eq. (104) below would be replaced by a two-particle function $\mathbf{A}(\mathbf{p}_1, \mathbf{p}_2)$. A factorized result, containing only one-particle functions, can be obtained with the aid of the factorization theorem of Kac.[30] In order to be able to take advantage of Kac's analysis, we proceed as follows. The solution to Eq. (71) can be written as

$$\mathfrak{S}(\mathbf{P}) = \int_{-\infty}^{0} e^{\epsilon t}\mathfrak{S}(t)\, dt, \quad (82)$$

with $\mathfrak{S}(t)$ defined by the differential equation

$$\partial\mathfrak{S}(t)/\partial t = \Omega\mathfrak{S}(t), \quad (83)$$

and the initial condition

$$\mathfrak{S}(0) = \mathbf{S}(\mathbf{P}). \quad (84)$$

The solution to Eqs. (83) and (84) can be written

$$\mathfrak{S}(t) = e^{\Omega t}\mathbf{S}(\mathbf{P}), \quad (85)$$

and Eq. (52) becomes

$$\lambda = \frac{1}{3VkT^2}\int_{-\infty}^{0} e^{\epsilon t}\, dt \sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{S}(\mathbf{P})$$
$$\times \int_{-\infty}^{\infty} dE\, \delta_\eta(E) e^{\Omega t}\mathbf{S}(\mathbf{P}). \quad (86)$$

In view of Eq. (6), this can be written

$$\lambda = \frac{n}{3kT^2}\int_{-\infty}^{0} e^{\epsilon t}\, dt \sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)$$
$$\times \int_{-\infty}^{\infty} dE\, \delta_\eta(E) e^{\Omega t}\mathbf{S}(\mathbf{P}), \quad (87)$$

where $n = N/V$ is the density.

On expanding the exponential in Eq. (87) one will encounter expressions of the form

$$\sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty} dE\, \delta_\eta(E)\Omega^r\mathbf{S}(\mathbf{P}),$$
$$r = 0, 1, 2, \cdots. \quad (88)$$

Consider the case $r = 1$. We have

$$\Omega\mathbf{S}(\mathbf{P}) = \frac{1}{V}\sum_{i<j}\Omega_{ij}\sum_{k=1}^{N}\mathbf{s}(\mathbf{p}_k)$$
$$= \frac{1}{V}\sum_{i<j}\Omega_{ij}[\mathbf{s}(\mathbf{p}_i) + \mathbf{s}(\mathbf{p}_j)]. \quad (89)$$

When the average of (88) is performed, only the terms for $i = 1$ will give a contribution since the average of $\mathbf{s}(\mathbf{p}_1)$ vanishes. Hence

$$\sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty} dE\, \delta_\eta(E)\Omega\mathbf{S}(\mathbf{P})$$
$$= \frac{1}{V}\sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty} dE\, \delta_\eta(E)$$
$$\times \sum_{i=2}^{N}\Omega_{1i}[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_i)]$$
$$= \frac{N-1}{V}\sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty} dE\, \delta_\eta(E)$$
$$\times \Omega_{12}[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_2)]$$
$$= \frac{N-1}{V}\left[\frac{(2\pi\hbar)^3}{V}\right]^2 \sum_{\mathbf{p}_1,\mathbf{p}_2} f(\mathbf{p}_1)f(\mathbf{p}_2)\mathbf{s}(\mathbf{p}_1)$$
$$\times \int_{-\infty}^{\infty} dE\, \delta_\eta(E)\Omega_{12}[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_2)]. \quad (90)$$

On substituting for Ω_{12} from Eq. (79) we find

$$\lim_{V\to\infty}\sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty} dE\, \delta_\eta(E)\Omega\mathbf{S}(\mathbf{P})$$
$$= \frac{4n}{m^2}\int d^3\mathbf{p}_1 \int d^3\mathbf{p}_2 f(\mathbf{p}_1)f(\mathbf{p}_2)\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty} dE\, \delta_\eta(E)$$

[30] M. Kac in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (University of California Press, Berkeley, 1956). See also R. Brout, Physica **22**, 509 (1956).

$$\times \int d^3\mathbf{p}_1' \int d^3\mathbf{p}_2' \, \delta_\eta(\epsilon_{12}' - \epsilon_{12} - E)$$
$$\times \delta(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_1' - \mathbf{p}_2')\sigma_{12}(E, \eta)$$
$$\times [\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_2) - \mathbf{s}(\mathbf{p}_1') - \mathbf{s}(\mathbf{p}_2')]. \quad (91)$$

We can now take the limit $\epsilon \to 0$, or $\eta = \frac{1}{2}\hbar\epsilon \to 0$, and integrate over E to obtain

$$\lim_{\eta\to 0}\lim_{V\to\infty} \sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1) \int_{-\infty}^{\infty} dE \, \delta_\eta(E)\Omega\mathbf{S}(\mathbf{P})$$
$$= n \int d^3\mathbf{p}_1 f(\mathbf{p}_1)\mathbf{s}(\mathbf{p}_1) I[\mathbf{s}(\mathbf{p}_1)], \quad (92)$$

where I is a linearized Boltzmann collision operator defined by

$$I[\mathbf{s}(\mathbf{p}_1)] = \frac{4}{m^2} \int d^3\mathbf{p}_2 \int d^3\mathbf{p}_1' \, f(\mathbf{p}_2)$$
$$\times \int d^3\mathbf{p}_2' \, \delta(\epsilon_{12}' - \epsilon_{12}) \, \delta(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_1' - \mathbf{p}_2')$$
$$\times \sigma_{12}[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_2) - \mathbf{s}(\mathbf{p}_1') - \mathbf{s}(\mathbf{p}_2')]. \quad (93)$$

Now consider higher values of r in Eq. (88). (Actually it is only for $r \geq 2$ that factorization is a problem.) For $r = 2$ we have

$$\Omega^2\mathbf{S}(\mathbf{P}) = \frac{1}{V^2} \sum_{i<j} \sum_{k<l} \Omega_{ij}\Omega_{kl}[\mathbf{s}(\mathbf{p}_k) + \mathbf{s}(\mathbf{p}_l)]. \quad (94)$$

There is no contribution unless at least one of k, l is equal to i or j, and so

$$\Omega^2\mathbf{S}(\mathbf{P}) = \frac{1}{V^2} \sum_{i<j} \Omega_{ij}\Big\{ \sum_{l=i+1}^{N} \Omega_{il}[\mathbf{s}(\mathbf{p}_i) + \mathbf{s}(\mathbf{p}_l)]$$
$$+ \sum_{l=j+1}^{N} \Omega_{jl}[\mathbf{s}(\mathbf{p}_j) + \mathbf{s}(\mathbf{p}_l)]$$
$$+ \sum_{k=1}^{i-1} \Omega_{ik}[\mathbf{s}(\mathbf{p}_i) + \mathbf{s}(\mathbf{p}_k)]$$
$$+ \sum_{k=1}^{j-1} \Omega_{jk}[\mathbf{s}(\mathbf{p}_j) + \mathbf{s}(\mathbf{p}_k)]\Big\}$$
$$= \frac{1}{V^2} \sum_{i<j} \Omega_{ij}\{ \sum_{k\neq i} \Omega_{ik}[\mathbf{s}(\mathbf{p}_i) + \mathbf{s}(\mathbf{p}_k)]$$
$$+ \sum_{k\neq j} \Omega_{jk}[\mathbf{s}(\mathbf{p}_j) + \mathbf{s}(\mathbf{p}_k)]\}. \quad (95)$$

We separate the above into two parts,

$$\Omega^2\mathbf{S}(\mathbf{P}) = \frac{1}{V^2} \sum_{i<j} \sum_{k\neq i,j} \Omega_{ij}\{\Omega_{ik}[\mathbf{s}(\mathbf{p}_i) + \mathbf{s}(\mathbf{p}_k)]$$
$$+ \Omega_{jk}[\mathbf{s}(\mathbf{p}_j) + \mathbf{s}(\mathbf{p}_k)]\}$$
$$+ \frac{2}{V^2} \sum_{i<j} \Omega_{ij}^2[\mathbf{s}(\mathbf{p}_i) + \mathbf{s}(\mathbf{p}_j)]. \quad (96)$$

Remarking again that the average of $\mathbf{s}(\mathbf{p})$ vanishes, it is seen that nonvanishing contributions to (88) will be obtained only when at least one of the above indices i, j, k is equal to 1. Since $j \geq 2$, the only possibilities are $i = 1$ or $k = 1$. It is a consequence of Eqs. (80) and (76) that terms with $k = 1$ but $i \neq 1$ will give no contribution when the limit $V \to \infty$, $\eta \to 0$ is taken. Thus, as before, the only nonvanishing contributions are obtained from terms with $i = 1$, for which we have

$$\Omega^2\mathbf{S}(\mathbf{P})|_{i=1} = \frac{1}{V^2} \sum_{\substack{j,k=2\\ j\neq k}}^{N} \Omega_{1j}\{\Omega_{1k}[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_k)]$$
$$+ \Omega_{jk}[\mathbf{s}(\mathbf{p}_j) + \mathbf{s}(\mathbf{p}_k)]\}$$
$$+ \frac{2}{V^2} \sum_{j=2}^{N} \Omega_{1j}^2[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_j)]. \quad (97)$$

On performing the average of (88) we find

$$\sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1) \int_{-\infty}^{\infty} dE \, \delta_\eta(E)\Omega^2\mathbf{S}(\mathbf{P})$$
$$= \frac{(N-1)(N-2)}{V^2} \int d^3\mathbf{p}_1 \int d^3\mathbf{p}_2 \int d^3\mathbf{p}_3$$
$$\times f(\mathbf{p}_1)f(\mathbf{p}_2)f(\mathbf{p}_3)\mathbf{s}(\mathbf{p}_1) \int_{-\infty}^{\infty} dE \, \delta_\eta(E)$$
$$\times \Omega_{13}\{\Omega_{12}[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_2)] + \Omega_{23}[\mathbf{s}(\mathbf{p}_2) + \mathbf{s}(\mathbf{p}_3)]\}$$
$$+ \frac{2(N-1)}{V^2} \int d^3\mathbf{p}_1 \int d^3\mathbf{p}_2 f(\mathbf{p}_1)f(\mathbf{p}_2)\mathbf{s}(\mathbf{p}_1)$$
$$\times \int_{-\infty}^{\infty} dE \, \delta_\eta(E)\Omega_{12}^2[\mathbf{s}(\mathbf{p}_1) + \mathbf{s}(\mathbf{p}_2)]. \quad (98)$$

We see immediately that the last term above is of order N/V^2 and therefore vanishes in the limit N, $V \to \infty$. In the first term, the integral over $\mathbf{p}_2$ converts Ω_{12} and Ω_{23} into Boltzmann operators; the integral over $\mathbf{p}_3$ then has the same effect on Ω_{13}. The result is

$$\lim_{\eta\to 0}\lim_{V\to\infty} \sum_{\mathbf{P}} \rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1) \int_{-\infty}^{\infty} dE \, \delta_\eta(E)\Omega^2\mathbf{S}(\mathbf{P})$$
$$= n^2 \int d^3\mathbf{p}_1 f(\mathbf{p}_1)\mathbf{s}(\mathbf{p}_1) I^2[\mathbf{s}(\mathbf{p}_1)]. \quad (99)$$

The procedure is readily generalized to arbitrary r. In the expression

$$\Omega^r\mathbf{S}(\mathbf{P}) = \frac{1}{V^r} \sum_{i_r<j_r} \cdots \sum_{i_1<j_1} \Omega_{i_rj_r} \cdots$$
$$\times \Omega_{i_1j_1}[\mathbf{s}(\mathbf{p}_{i_1}) + \mathbf{s}(\mathbf{p}_{j_1})], \quad (100)$$

nonvanishing contributions are obtained only when $\Omega_{i_1j_1}$ is coupled to $\Omega_{i_2j_2}$ through a pairing of indices,

$\Omega_{i_2 j_2}$ is similarly coupled to $\Omega_{i_3 j_3}$, and so on through to $\Omega_{i_r j_r}$. This pairing of $r-1$ sets of indices reduces the number of summations in Eq. (100) from $2r$ to $r+1$. The requirement that $i_r = 1$ gives one more reduction, leaving r independent summations. Divide the result into two parts. In the first, or dominant part, impose the restriction that none of the r summation indices be equal [corresponding to the restriction $k \neq j$ in Eq. (97)]. The remaining part [corresponding to the second term of Eq. (97)] will contain no more than $r-1$ summations. Its contribution to Eq. (88) will therefore be of order N^{r-1}/V^r, and thus vanishes in the limit $N, V \to \infty$. In the contribution of the dominant part, one may, proceeding from right to left, integrate over momenta in the manner leading from Eq. (98) to Eq. (99). Each such integration converts an Ω_{ij} operator into a Boltzmann operator, with the result

$$\lim_{\eta\to 0}\lim_{V\to\infty}\sum_{\mathbf{P}}\rho(\mathbf{P})\mathbf{s}(\mathbf{p}_1)\int_{-\infty}^{\infty}dE\;\delta_\eta(E)\Omega^r\mathbf{S}(\mathbf{P}) = n^r\int d^3\mathbf{p}_1 f(\mathbf{p}_1)\mathbf{s}(\mathbf{p}_1)I^r[\mathbf{s}(\mathbf{p}_1)]. \quad (101)$$

The above argument can be restated as follows. The dominant contribution is obtained by allowing a given pair of particles to scatter no more than once. Rescatterings, when a given pair scatter twice or more, are placed in the remainder, which is of order $1/V$.

On combining Eqs. (101) and (87) we find

$$\lambda = \frac{n}{3kT^2}\int_{-\infty}^{0}dt\int d^3\mathbf{p}_1 f(\mathbf{p}_1)\mathbf{s}(\mathbf{p}_1)e^{tI}\mathbf{s}(\mathbf{p}_1), \quad (102)$$

which can be written

$$\lambda = \frac{1}{3kT^2}\int d^3\mathbf{p}_1 f(\mathbf{p}_1)\mathbf{s}(\mathbf{p}_1)\cdot\mathbf{A}(\mathbf{p}_1), \quad (103)$$

where $\mathbf{A}(\mathbf{p}_1)$ is a solution to the integral equation

$$I[\mathbf{A}(\mathbf{p}_1)] = \mathbf{s}(\mathbf{p}_1). \quad (104)$$

Equations (103) and (104) are identical to the results usually obtained from the Boltzmann equation.[1]

We note that the time integral of Eq. (102) actually exists since $\mathbf{s}(\mathbf{p})$ is orthogonal to the summational invariants of I. Remark further that the solution to Eq. (104) is arbitrary to the extent that one can add to $\mathbf{A}$ any linear combination of the summational invariants, but this arbitrariness has no effect on the thermal conductivity as given by Eq. (103).

The above procedure is clearly applicable to other transport coefficients. Indeed, nothing we have done depends on the particular form of $\mathbf{s}(\mathbf{p})$, so long as it is orthogonal to the summational invariants.

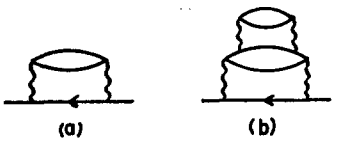


FIG. 1. Typical diagrams occurring in Fujita's analysis.

IV. DISCUSSION

In Fujita's analysis, the result (103), (104) is obtained directly, without the use of a factorization theorem. The reason for this is not immediately evident, and we devote some comment to this point here. Using a diagrammatic expansion, Fujita associates with each diagram a factor $c^m t^n$, $m, n = 0, 1, 2, \cdots$, where c is the density (we revert to his notation) and t the time. In the "ct limit", appropriate to low density, only diagrams of order $(ct)^m$ are retained, those of order $c^n(ct)^m$ being neglected. A typical diagram retained by Fujita is given in Fig. 1(a). However, there are, in addition, diagrams typified by Fig. 1(b) which should, according to the ct criterion, be retained, but were neglected by Fujita. These diagrams correspond to rescatterings of the sort mentioned above, and thus their neglect is indeed justified in the limit $N, V \to \infty$.

While the preceding discussion has been primarily concerned with low-density gases, we wish to emphasize that our method, based on the generalized master equation, is directly applicable to other systems and seems to offer the most economical and rigorous approach. In this connection we have already cited the work of Chester and Thellung[9] and of Verboven[11] on electrical conductivity; similar methods have also been used by Hardy[16] in a discussion of the lattice contribution to the thermal conductivity of solids. In addition, our discussion is readily extended to the problem of transport in gases of moderate density, as we hope to show in a future publication. In this case it is necessary to take account of the potential energy contributions to the equilibrium density matrix and to the flux operators; this is readily done with the aid of the generalizations of the Van Hove equation obtained by Janner[31] and by Peterson and Quay.[32]

The treatment we have given has a number of points of similarity with the derivation of the Boltzmann equation via a master equation, as outlined by Kac[30] and discussed further by Brout.[30] Indeed, Eq. (83) above is very nearly a master

[31] A. Janner, Helv. Phys. Acta 35, 47 (1962).
[32] R. L. Peterson and P. M. Quay, J. Math. Phys. (to be published).

equation, suitably modified for the problem at hand. However, the following point should be noted. A master equation can be written only for finite V, since for an infinite system (of nonvanishing density), Ω becomes a meaningless infinite sum of vanishing operators: each of the terms Ω_{ij}/V in the sum (72) goes to zero, while the number of terms in the sum becomes infinite. If V is held finite then, as already noted, the limit $\epsilon \to 0$ cannot be taken, and with nonvanishing ϵ, one has an "incomplete" master equation involving scattering cross sections "off the energy shell." Nevertheless, as we have seen, the proof of the factorization theorem still goes through with such an incomplete master equation. (It is of course clear that the application of the usual master equation is legitimate for processes which do not vary significantly over times of the order of the time taken by a collision and for which Poincaré recurrences can be ignored.)

ACKNOWLEDGMENTS

We are indebted to Dr. S. Fujita and to Dr. R. J. Hardy for a number of valuable discussions. One of us (R.J.S.) also wishes to thank Professor I. Prigogine for his hospitality at the University of Brussels, and the National Academy of Sciences–National Research Council for the grant of a Postdoctoral Fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

Determinantal Solution of the Fredholm Equation with Green's Function Kernel*

HENRY BRYSK
Physics Department, Adelphi University, Garden City, New York
(Received 16 April 1963)

The Fredholm integral equation with a Green's function type of kernel has been transformed by Drukarev into an equivalent Volterra equation. It is now proven that the Neumann series solution of the Volterra equation yields the determinantal solution of the Fredholm equation. Thus, a Born approximation technique suffices to obtain the full Fredholm solution.

I. INTRODUCTION

IN the Green's function formulation, scattering problems lead to a Fredholm integral equation for the partial wave. The Born approximation consists in solving the Fredholm equation by means of a Neumann series (power series in the strength parameter λ). This perturbation expansion has a limited radius of convergence, determined by the lowest eigenvalue λ_1 of the homogeneous equation. On the other hand, Fredholm theory gives a solution convergent for all regular λ in terms of the ratio of two power series in λ. The terms in the Fredholm expansion are integrals of determinants; the kernel appears as the element of the determinant. The zeros of the denominator determine the eigenvalues. This last fact has been used to ascertain the radius of convergence of the Born approximation,[1,2] but direct evaluation of the determinantal solution has usually been shunned because of presumed computational difficulty.

Drukarev[3] has transformed the Fredholm equation for the wavefunction into a Volterra equation for an auxilliary function. The wavefunction is expressed as the auxilliary function divided by an integral over the auxilliary function. The Neumann series solution of the Volterra equation converges for all λ. An expression for the wavefunction is then obtained as the ratio of two power series in λ.[4]

The present paper proves that, after a rearrangement to split off the free term from the Drukarev solution, the ratio of power series obtained from the Neumann series solution of the Volterra equation coincides with the determinantal solution of the Fredholm equation. Thus, the computational technique normally applied to the evaluation of the Born approximation (with limited radius of convergence) suffices, with the Drukarev transformation, to calculate the Fredholm determinantal solution (valid for all regular λ).

* Work supported by the National Science Foundation.
[1] R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).
[2] Walter Kohn, Rev. Mod. Phys. **26**, 292 (1954).
[3] G. F. Drukarev, Zhur. Eksp. i Teoret. Fiz. **25**, 139 (1953).
[4] G. F. Drukarev, Vestn. Leningr. Univ. Ser. Fiz. i Khim. **22**, 65 (1958).

equation, suitably modified for the problem at hand. However, the following point should be noted. A master equation can be written only for finite V, since for an infinite system (of nonvanishing density), Ω becomes a meaningless infinite sum of vanishing operators: each of the terms Ω_{ij}/V in the sum (72) goes to zero, while the number of terms in the sum becomes infinite. If V is held finite then, as already noted, the limit $\epsilon \to 0$ cannot be taken, and with nonvanishing ϵ, one has an "incomplete" master equation involving scattering cross sections "off the energy shell." Nevertheless, as we have seen, the proof of the factorization theorem still goes through with such an incomplete master equation. (It is of course clear that the application of the usual master equation is legitimate for processes which do not vary significantly over times of the order of the time taken by a collision and for which Poincaré recurrences can be ignored.)

ACKNOWLEDGMENTS

We are indebted to Dr. S. Fujita and to Dr. R. J. Hardy for a number of valuable discussions. One of us (R.J.S.) also wishes to thank Professor I. Prigogine for his hospitality at the University of Brussels, and the National Academy of Sciences–National Research Council for the grant of a Postdoctoral Fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 12 DECEMBER 1963

Determinantal Solution of the Fredholm Equation with Green's Function Kernel*

HENRY BRYSK
Physics Department, Adelphi University, Garden City, New York
(Received 16 April 1963)

The Fredholm integral equation with a Green's function type of kernel has been transformed by Drukarev into an equivalent Volterra equation. It is now proven that the Neumann series solution of the Volterra equation yields the determinantal solution of the Fredholm equation. Thus, a Born approximation technique suffices to obtain the full Fredholm solution.

I. INTRODUCTION

IN the Green's function formulation, scattering problems lead to a Fredholm integral equation for the partial wave. The Born approximation consists in solving the Fredholm equation by means of a Neumann series (power series in the strength parameter λ). This perturbation expansion has a limited radius of convergence, determined by the lowest eigenvalue λ_1 of the homogeneous equation. On the other hand, Fredholm theory gives a solution convergent for all regular λ in terms of the ratio of two power series in λ. The terms in the Fredholm expansion are integrals of determinants; the kernel appears as the element of the determinant. The zeros of the denominator determine the eigenvalues. This last fact has been used to ascertain the radius of convergence of the Born approximation,[1,2] but direct evaluation of the determinantal solution has usually been shunned because of presumed computational difficulty.

Drukarev[3] has transformed the Fredholm equation for the wavefunction into a Volterra equation for an auxilliary function. The wavefunction is expressed as the auxilliary function divided by an integral over the auxilliary function. The Neumann series solution of the Volterra equation converges for all λ. An expression for the wavefunction is then obtained as the ratio of two power series in λ.[4]

The present paper proves that, after a rearrangement to split off the free term from the Drukarev solution, the ratio of power series obtained from the Neumann series solution of the Volterra equation coincides with the determinantal solution of the Fredholm equation. Thus, the computational technique normally applied to the evaluation of the Born approximation (with limited radius of convergence) suffices, with the Drukarev transformation, to calculate the Fredholm determinantal solution (valid for all regular λ).

* Work supported by the National Science Foundation.
[1] R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).
[2] Walter Kohn, Rev. Mod. Phys. **26**, 292 (1954).
[3] G. F. Drukarev, Zhur. Eksp. i Teoret. Fiz. **25**, 139 (1953).
[4] G. F. Drukarev, Vestn. Leningr. Univ. Ser. Fiz. i Khim. **22**, 65 (1958).

II. DETERMINANTAL SOLUTION

The Fredholm equation

$$\varphi(x) - \lambda \int_a^b K(x, s)\varphi(s)\, ds = f(x) \tag{1}$$

has a solution in the form

$$\varphi(x) = f(x) + \lambda \int_a^b \Gamma(x, s; \lambda) f(s)\, ds, \tag{2}$$

where the resolvent

$$\Gamma(x, s; \lambda) = D(x, s; \lambda)/D(\lambda) \tag{3}$$

can be expressed for all regular λ (i.e., for any λ which is not an eigenvalue of the homogeneous equation) as a ratio of power series in λ:

$$D(x, s; \lambda) = \sum_{n=0}^{\infty} (-1)^n (n!)^{-1} B_n(x, s)\lambda^n, \tag{4}$$

$$D(\lambda) = \sum_{n=0}^{\infty} (-1)^n (n!)^{-1} c_n \lambda^n, \tag{5}$$

where

$$B_0(x, s) = K(x, s), \qquad c_0 = 1, \tag{6}$$

and the higher-order coefficients are integrals of determinants connected by the recursion relations

$$c_{n+1} = \int_a^b B_n(s, s)\, ds, \tag{7}$$

$$B_n(x, s) = c_n K(x, s) - n \int_a^b K(x, t) B_{n-1}(t, s)\, dt. \tag{8}$$

The resolvent can be written as a ratio of power series in λ provided that the kernel $K(x, s)$ and the free term $f(x)$ are both square integrable. The recursion relations are proven under the somewhat more restrictive condition that $\int_a^b |K(x, s)|^2\, ds$ and the corresponding x integral are bounded.[5]

III. DRUKAREV TRANSFORMATION

If the kernel is of the Green's function type,

$$K(x, s) = V(s) f(r_<) g(r_>), \tag{9}$$

where $r_<$ and $r_>$ denotes the smaller and larger, respectively, of x and s, and f is the same function as appears in the free term, the Fredholm equation can be rewritten as

$$\varphi(x) = f(x)\left[1 + \lambda \int_a^b g(s) V(s)\varphi(s)\, ds\right] - \lambda \int_a^x [f(x)g(s) - g(x)f(s)] V(s)\varphi(s)\, ds. \tag{10}$$

[5] S. G. Mikhlin, *Integral Equations* (Pergamon Press, Inc., New York, 1957).

The expression

$$N = 1 + \lambda \int_a^b g(s) V(s)\varphi(s)\, ds \tag{11}$$

is a constant. If it is factored out of the unknown function,

$$\varphi(x) = Nu(x), \tag{12}$$

Eq. (10) reduces to the Volterra equation[3]

$$u(x) = f(x) - \lambda \int_a^x [f(x)g(s) - g(x)f(s)] V(s)u(s)\, ds. \tag{13}$$

The value of N is obtained upon substitution of Eq. (12) into Eq. (11):

$$N = \left[1 - \lambda \int_a^b g(s) V(s)u(s)\, ds\right]^{-1}. \tag{14}$$

The Volterra equation can be solved in terms of a Neumann series

$$u(x) = \sum_{n=0}^{\infty} \lambda^n u^{(n)}(x) \tag{15}$$

that converges for all λ. Explicitly,

$$u^{(0)}(x) = f(x), \tag{16}$$

$$u^{(n+1)}(x) = -\int_a^x [f(x)g(s) - g(x)f(s)] V(s)u^{(n)}(s)\, ds. \tag{17}$$

When Eq. (15) is combined with Eqs. (12) and (14), $\varphi(x)$ is expressed as the ratio of two power series in λ[4].

IV. PROOF OF EQUIVALENCE

The first step in the comparison is to recast Eq. (12), with the Neumann series substituted for $u(x)$, into the form of Eq. (2):

$$\varphi(x) = N \sum_{n=0}^{\infty} \lambda^n u^{(n)}(x) = f(x) + N\left[N^{-1}(N - 1)f(x) + \sum_{n=1}^{\infty} \lambda^n u^{(n)}(x)\right]. \tag{18}$$

From Eq. (14),

$$N^{-1}(N - 1) = \lambda \int_a^b g(s) V(s)u(s)\, ds = \lambda \sum_{n=0}^{\infty} \lambda^n \int_a^b g(s) V(s)u^{(n)}(s)\, ds. \tag{19}$$

Combining like powers of λ,

$$\varphi(x) = f(x) + \lambda N \sum_{n=0}^{\infty} \lambda^n \left[f(x) \int_a^b g(s) V(s) \times u^{(n)}(s)\, ds + u^{(n+1)}(x)\right]. \tag{20}$$

However, Eq. (17) can be rewritten as

$$u^{(n+1)}(x) = \int_a^b K(x, s)u^{(n)}(s)\, ds - f(x) \int_a^b g(s)V(s)u^{(n)}(s)\, ds, \quad (21)$$

so that finally,

$$\varphi(x) = f(x) + \lambda \frac{\sum_{n=0}^{\infty} \lambda^n \int_a^b K(x, s)u^{(n)}(s)\, ds}{1 - \sum_{n=0}^{\infty} \lambda^{n+1} \int_a^b g(s)V(s)u^{(n)}(s)\, ds}. \quad (22)$$

The equivalence of the two expansions will be demonstrated if it is proven that the corresponding terms in Eqs. (2) and (22) are identical, i.e.,

$$(-1)^n(n!)^{-1} \int_a^b B_n(x, s)f(s)\, ds = \int_a^b K(x, s)u^{(n)}(s)\, ds, \quad (23)$$

and

$$(-1)^{n+1}[(n+1)!]^{-1}c_{n+1} = -\int_a^b g(s)V(s)u^{(n)}(s)\, ds. \quad (24)$$

For $n = 0$, Eq. (23) follows immediately on substitution of Eqs. (6) and (16), while Eqs. (6), (7), and (16) yield

$$c_1 = \int_a^b K(s, s)\, ds = \int_a^b V(s)f(s)g(s)\, ds = \int_a^b g(s)V(s)u^{(0)}(s)\, ds. \quad (25)$$

The proof, then, proceeds by induction: Assume that Eqs. (23) and (24) hold for some n. According to Eq. (8),

$$(-1)^{n+1}[(n+1)!]^{-1} \int_a^b ds\, f(s)B_{n+1}(x, s) = (-1)^{n+1}[(n+1)!]^{-1}c_{n+1} \int_a^b ds\, f(s)K(x, s) + (-1)^n(n!)^{-1} \int_a^b ds\, f(s) \int_a^b dt\, K(x, t)B_n(t, s). \quad (26)$$

According to Eq. (21),

$$\int_a^b ds\, K(x, s)u^{(n+1)}(s) = -\int_a^b ds\, f(s)K(x, s) \int_a^b dt\, g(t)V(t)u^{(n)}(t) + \int_a^b ds\, K(x, s) \int_a^b dt\, K(s, t)u^{(n)}(t). \quad (27)$$

The first terms on the right of Eqs. (26) and (27) are equal because of Eq. (24). The second terms on the right are integrals over the kernel of the two sides of Eq. (23) (apart from an interchange of order of integration). Thus, the left sides are equal, proving Eq. (23) for $n + 1$ and hence for all n. Consider now Eqs. (26) and (27) for an arbitrary n. Since Eq. (23) holds for all n, the left sides of the equations are equal, as are the second terms on the right. Then the first terms on the right must be equal, proving Eq. (24) for all n.

ACKNOWLEDGMENT

It is a pleasure to thank Herbert Kurss for several helpful discussions.

External-Mass Singularity

NOBORU NAKANISHI*†
Department of Physics, University of Illinois, Urbana, Illinois
(Received 2 July 1963)

It is shown by a concrete example that the scattering amplitude may have an external-mass singularity, which does not correspond to a threshold for any energy variable. If the external mass exceeds the external-mass singularity, the scattering amplitude (even for the forward scattering) is no longer a boundary value of an analytic function of energy variables in the usual sense. This would mean a deadlock of the S-matrix theory based on the analyticity. In the Appendix an example of a threshold in the non-Euclidean case is given.

RECENTLY, the S-matrix theory based on the analyticity[1] has been recognized as a powerful approach to the theory of strong interactions. In such an approach it is naturally assumed[2] that the transition amplitude can be analytically continued as a function of energy variables,[3] and that its physical sheet is unambiguously defined at least for the two-particle scattering process. It is known[4] that the Green's function is always a boundary value of an analytic function of all independent squares of external momenta and their sums, but the same property does not necessarily hold for the transition amplitude because the external masses are now *fixed* on the mass shells. The purpose of the present note is to give a perturbation-theoretical example in which even the elastic scattering amplitude indeed has no physical sheet in the usual sense.

Consider a Feynman graph G having N internal lines and n independent circuits. For simplicity we assume that all particles are scalar and all couplings are direct. Then the transition amplitude corresponding to G is proportional to

$$\int_0^1 dx_1 \cdots \int_0^1 dx_N\, \delta\Big(1 - \sum_{i=1}^{N} x_i\Big) \times [U^2(V - i\epsilon)^{N-2n}]^{-1}, \qquad (1)$$

where U is a nonnegative function of x_i, V being a linear function of energy variables and external masses. Let us consider the two-particle scattering. Then V is given by[5]

$$V = \sum_{i=1}^{N} x_i m_i^2 - \zeta_A M_a^2 - \zeta_B M_b^2 - \zeta_C M_c^2 - \zeta_D M_d^2 - \zeta_{AB} s - \zeta_{AC} t - \zeta_{AD} u, \qquad (2)$$

with

$$M_a^2 + M_b^2 + M_c^2 + M_d^2 = s + t + u. \qquad (3)$$

Here m_i is an internal mass, M_a, M_b, M_c, and M_d the four external masses, s, t, and u the invariant energy variables. The parametric functions ζ are nonnegative.

The prescription of finding the singularities of (1) is well known.[6] The leading singularity may be given by the solution of the simultaneous equations

$$\partial V/\partial x_i = 0, \qquad (i = 1, 2, \cdots, N). \qquad (4)$$

If this point is a minimum of V,[7] i.e.,

$$\sum_{i,j} \frac{\partial^2 V}{\partial x_i\, \partial x_j} \Delta x_i\, \Delta x_j \geq 0, \qquad (5)$$

then it is a threshold. As for the other singularities, some of Feynman parameters are put equal to zero, and the problem is reduced to finding the leading singularity of the corresponding reduced graph.

The above prescription is quite adequate if not only the energy variables but also the external masses are considered on the equal footing. But in the S-matrix theory, the latter are fixed, and the scattering amplitude is considered as a boundary

* On leave of absence from the Institute for Advanced Study, Princeton, New Jersey. Present address: Brookhaven National Laboratory, Upton, Long Island, New York.

† This work was supported by the National Science Foundation.

[1] See, for example, G. F. Chew, *S-Matrix Theory of Strong Interactions* (W. A. Benjamin Company, Inc., New York, 1961). This work contains further references.

[2] H. P. Stapp, Phys. Rev. **125**, 2139 (1962); Rev. Mod. Phys. **34**, 390 (1962).

[3] "Momentum transfer" can be considered as an energy variable according to the substitution law.

[4] D. Hall and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat-Fys. Medd. **31**, No. 5 (1957); H. Araki, Progr. Theoret. Phys. (Kyoto) Suppl. **18**, 83 (1961).

[5] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 337 (1961). See also Y. Nambu, Nuovo Cimento **6**, 1064 (1957); K. Symanzik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958).

[6] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **22**, 128 (1959); L. D. Landau, Nucl. Phys. **13**, 181 (1959); J. D. Bjorken, (unpublished, Stanford University, 1959).

[7] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. **18**, 1 (1961), Part II; J. Tarski, J. Math. Phys. **1**, 149 (1960), Appendix A.

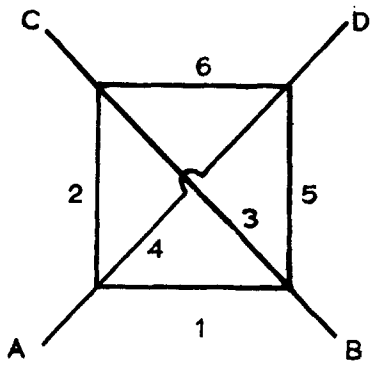


FIG. 1. Simplest symmetric graph.

value of an analytic function of the former only. This means that the infinitesimal imaginary part $(-i\epsilon)$ in (1) must be included in s or t as its imaginary part [for definiteness, we eliminate u by using (3)]. Then the sign of the coefficient $\xi \equiv \zeta_{AB} - \zeta_{AD}$ (or $\zeta_{AC} - \zeta_{AD}$) of s (or t) has an important meaning, because it determines from which side the boundary value should be taken. Thus the Feynman integral (1) should be divided at $\xi = 0$. In this sense, a surface

$$\zeta_{AB} = \zeta_{AC} = \zeta_{AD} \tag{6}$$

in the Feynman parametric space plays a special role, that is to say, (6) is a sort of end point of the Feynman parametric integral. If the denominator function V has a minimal zero point on the surface (6) for certain values of external masses, it will correspond to a new type of singularity. We shall call it "external-mass singularity".

If $\zeta_{AD} \equiv 0$, (6) gives $\zeta_{AB} = \zeta_{AC} = 0$. According to the explicit expression of ζ_{AB},[8] $\zeta_{AB} = 0$ means that all x_i on a path connecting one of external lines A and B with one of C and D must vanish. Likewise, $\zeta_{AC} = 0$ yields a result that either A or C is connected with either B or D by a path of $x_i = 0$. Therefore, the reduced graph becomes a self-energy graph of an external particle. Thus the external-mass singularity in this case is nothing but the mass value, above which the external particle decays spontaneously. This is a normal threshold for an external mass, and it is avoided in the S-matrix theory by requiring the stability condition.

[8] According to the expression given in Sec. 2 of N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 337 (1961), we easily see that $\zeta_{AB}(\neq 0)$ vanishes if and only if $H \cap S$ is not empty for any intermediate state S of the s channel, where H denotes the totality of the lines of $x_i = 0$. If H contains no path which connects A or B with C or D, when every line of H is shrunk to a point, the reduced graph must contain at least one intermediate state of the s channel. It is of course an intermediate state S in the original graph, and thus $H \cap S$ is empty.

When $\zeta_{AD} \not\equiv 0$, a different situation will happen, namely, an external-mass singularity may appear even under the stability conditions. To see this, we consider the simplest example shown in Fig. 1. The parametric functions are given by

$$\begin{aligned}
\zeta_A &= x_1x_2x_4(x_3 + x_5 + x_6)/U, \\
\zeta_B &= x_1x_3x_5(x_2 + x_4 + x_6)/U, \\
\zeta_C &= x_2x_3x_6(x_1 + x_4 + x_5)/U, \\
\zeta_D &= x_4x_5x_6(x_1 + x_2 + x_3)/U, \\
\zeta_{AB} &= x_2x_3x_4x_5/U, \qquad \zeta_{AC} = x_1x_3x_4x_6/U, \\
\zeta_{AD} &= x_1x_2x_5x_6/U,
\end{aligned} \tag{7}$$

with

$$\begin{aligned}
U = {} & x_1x_2x_3 + x_1x_2x_5 + x_1x_2x_6 + x_1x_3x_4 \\
& + x_1x_3x_6 + x_1x_4x_5 + x_1x_4x_6 + x_1x_5x_6 \\
& + x_2x_3x_4 + x_2x_3x_5 + x_2x_4x_5 + x_2x_4x_6 \\
& + x_2x_5x_6 + x_3x_4x_5 + x_3x_4x_6 + x_3x_5x_6.
\end{aligned} \tag{8}$$

Hence (6) is satisfied when

$$x_1x_6 = x_2x_5 = x_3x_4 \equiv y. \tag{9}$$

Then by putting

$$x_0 \equiv x_1x_2x_3/y, \tag{10}$$

the V function is reduced to

$$\begin{aligned}
x_0V = {} & x_0x_1m_1^2 + x_0x_2m_2^2 + x_0x_3m_3^2 + x_1x_2m_4^2 \\
& + x_1x_3m_5^2 + x_2x_3m_6^2 - (x_0x_1x_2M_a^2 + x_0x_1x_3M_b^2 \\
& + x_0x_2x_3M_c^2 + x_1x_2x_3M_d^2)(x_0 + x_1 + x_2 + x_3)^{-1}.
\end{aligned} \tag{11}$$

For simplicity, we consider the case

$$\begin{aligned}
M_a = M_b = M_c = M_d \equiv M, \\
m_1 = m_6, \quad m_2 = m_5, \quad m_3 = m_4.
\end{aligned} \tag{12}$$

Then it is easy to see that (11) has the minimal zero point at

$$M^2 = 2(m_1^2 + m_2^2 + m_3^2), \tag{13}$$

with

$$x_0 = x_1 = x_2 = x_3. \tag{14}$$

The external mass given by (13) satisfies the stability condition if m_1^2, m_2^2, m_3^2 satisfy the trianglar inequalities. Thus (13) may be called an anomalous threshold for external masses.

Now, if M^2 becomes larger than (13), the scattering amplitude (1) always has the imaginary part regardless of the values of s and t and the signs of their coefficients. More explicitly, the scattering amplitude

as a function of an energy variable s is then represented as

$$\int_{-\infty}^{\infty} ds' \frac{\rho_1(s', t)}{s' - s - i\epsilon} - \int_{-\infty}^{\infty} ds' \frac{\rho_2(s', t)}{s' - s + i\epsilon} \qquad (15)$$

if we operate $\sum_i \partial/\partial m_i^2$ in order to avoid the ultraviolet divergence. The spectral functions $\rho_1(s, t)$ and $\rho_2(s, t)$ are easily shown to be positive-definite[9] (therefore nonvanishing) for *any* real values of s and t. Thus the double cut covers all physical regions. In other words, the scattering amplitude no longer has a physical sheet as an analytic function of energy variables as far as M^2 is fixed to a value larger than (13). Hence, in this case, no dispersion relation in the ordinary sense holds even for the forward scattering, as was previously noticed by the present author.[10]

With the present experimental knowledge it seems that any particle which is stable against strong interactions does not lie on the branch cut of the external-mass singularity. Particle-number conservation laws usually prevent its appearing. But if a pseudoscalar strangeness-zero meson which is heavier than 340 MeV is found, then the S-matrix theory based on the analyticity[1] will meet a fatal difficulty. Conversely, requirement of the absence of external-mass singularities gives some restriction for masses of stable particles. It might be an interesting problem to check, without using perturbation theory, whether or not the dispersion relation for the above-mentioned particle contradicts the unitarity requirement.

ACKNOWLEDGMENT

The author would like to thank Professor K. Nishijima and Professor G. M. Almy for the warm hospitality extended to him at the University of Illinois.

APPENDIX

It may be interesting to see the relation between our external-mass singularity and the usual leading singularity of Fig. 1. Eqs. (9) and (14) read

$$x_1 = x_2 = x_3 = x_4 = x_5 = x_6. \qquad (A1)$$

We can easily solve Eq. (4) by inserting (A1) and (13) to get

$$s = 8(m_2^2 + m_3^2 - m_1^2), \quad t = 8(m_1^2 + m_3^2 - m_2^2), \quad u = 8(m_1^2 + m_2^2 - m_3^2). \qquad (A2)$$

Eqs. (13) and (A2) give the intersection of the above two singularity surfaces. The matrix $(\partial^2 V/\partial x_i \partial x_j)$ is easily computed and we find that its eigenvalues are

$$0,\ m_1^2,\ m_2^2,\ m_3^2,$$
$$\tfrac{1}{2}(m_1^2 + m_2^2 + m_3^2) \pm [m_1^4 + m_2^4 + m_3^4 - m_1^2 m_2^2 - m_2^2 m_3^2 - m_3^2 m_1^2]^{\frac{1}{2}}, \qquad (A3)$$

apart from a positive numerical factor. The appearance of zero is a consequence of general theory.[11] On account of (5) and (A3), the critical point given by (13) and (A2) becomes a threshold if and only if m_1, m_2, m_3 satisfy the triangular inequalities.

It may be noteworthy that external momenta can be Euclidean[12] at this point if and only if m_1^2, m_2^2, m_3^2 satisfy the triangular inequalities. Thus we get an example of a threshold in the non-Euclidean case if $(m_1 + m_2)^2 > m_3^2 > m_1^2 + m_2^2$.

[9] This is because the integrand of the Feynman integral for $\rho_i(s, t)$ is a positive-definite distribution with a nontrivial support in the integration region for any values of s and t. If the integrand were not so, we could not immediately conclude the existence of the double cut.

[10] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **26**, 337 (1961), Appendix B.

[11] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. **18**, 1 (1961), Part II.

[12] Here "Euclidean" means that the square of any linear combination of effective external momenta (in the reduced graph) is not negative. (In this sense a normal threshold belongs to the Euclidean case because the effective external momenta is timelike.) It was questioned in reference 11 whether or not thresholds can appear in the non-Euclidean case. Within the author's knowledge, no explicit example of a threshold in the non-Euclidean case has been given so far.