JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 9, NUMBER 4 APRIL 1968

# Dyadic Analysis of Spatially Homogeneous World Models*

F. B. ESTABROOK, H. D. WAHLQUIST, AND C. G. BEHR†
*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California*

(Received 11 May 1967)

The dyadic formalism is applied to cosmological models, and leads to a convenient set of first-order ordinary differential equations. The Bianchi–Behr type of any model is shown to be constant in time, regardless of the state of the matter content. The case of perfect fluid matter content is formulated. Type V models and Type VIII and IX models with incoherent matter are discussed, and some consistent subtypes delineated. The Gödel Hamiltonian for symmetric Type IX models is derived and generalized.

## I. INTRODUCTION

In 1950 Taub[1] considered spatially homogeneous empty world models. These space-times admit locally a three-parameter group of motions which is simply transitive on each member of a one-parameter family of geodesically parallel, spacelike, three-dimensional hypersurfaces covering the manifold. He introduced a mathematical formalism to construct such world models as solutions of Einstein's field equations. The formalism is based on geometrical results which go back to Lie, Ricci, Bianchi, Fubini, and Cartan; it uses Gaussian coordinates based on the family of homogeneous hypersurfaces, and introduces a basis set of three linearly independent vector fields $\mathbf{e}^a$ in each hypersurface. These vector fields have the properties that (a) their inner products, say $\gamma^{ab}$ $(a, b = 1, 2, 3)$, are constants within any given hypersurface, (b) they have vanishing Lie derivatives with respect to the geodesic gradient vector field normal to the hypersurfaces, and (c) their curls in the hypersurfaces are given by a set of structure constants $C^a_{bc}$ which algebraically characterize the group of motions. The fundamental form of the 3-spaces then becomes $d\sigma^2 = \gamma_{ab}\mathbf{e}^a \cdot d\mathbf{x}\mathbf{e}^b \cdot d\mathbf{x}$, and for the space–time $ds^2 = d\sigma^2 - dt^2$. The matrix $\gamma_{ab}$ is inverse to $\gamma^{ab}$; i.e., it is formed from the inner products of the dual basis $\mathbf{e}_a$.

According to Bianchi's analysis of three-parameter Lie groups,[2] there are nine algebraically inequivalent types of three-dimensional spaces admitting a simply transitive group of motions. Within each of these types the $\gamma^{ab}$ are arbitrary, while the $C^a_{bc}$ are determined up to a nonsingular constant affine transformation. Assuming the "Bianchi type" of the hypersurfaces to remain the same throughout the entire four-dimensional manifold, the $C^a_{bc}$ can be taken as a canonical set of constants, ones and zeros. Einstein's field equations give in the most general case a very complicated system of six coupled second-order ordinary differential equations for the $\gamma^{ab}$ as functions only of the time coordinate $t$ that labels the hypersurfaces.

Heckmann and Schücking[3] have reformulated Taub's method for spatially homogeneous universes with incoherent matter. Previously known special world models of this type are Einstein's static universe, Friedmann's isotropic cosmologies, and Gödel's

* Sponsored by the National Aeronautics and Space Administration under Contract No. NAS7-100.

† National Research Council Postdoctoral Resident Research Associate, supported by NASA.

[1] A. Taub, Ann. Math. **53**, 472 (1951).

[2] I. Bianchi, Mem. Soc. Ital. Sci., Ser. **IIIa**, 11 (1897).

[3] O. Heckmann and E. Schücking, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), Chap. 11.

rotating universe. Heckmann and Schücking refer to this and other previous work, and find a few additional solutions containing expansion and shear, but no rotation. Behr[4] has studied certain more general cases.

Behr[5] has introduced a classification of spatially homogeneous cosmologies differing somewhat from that based on Bianchi type. Let $A^{il} = \frac{1}{2}C^{i}_{jk}\epsilon^{jkl}$. The vanishing or nonvanishing of $A^{[il]}$, and the rank and absolute value of the signature of $A^{(il)}$ are invariant under constant affine transformation. By these criteria 10 inequivalent types are distinguished. Specializing the basis $\mathbf{e}^a$, and taking account of the Jacobi identity, $A^{il}$ can be put into the form

$$A^{il} = \begin{pmatrix} a & n & 0 \\ -n & b & 0 \\ 0 & 0 & c \end{pmatrix}, \quad \text{where} \quad nc = 0.$$

TABLE I. Bianchi–Behr classification of homogeneous cosmological models.

| Model type | $n^2$ | Sign $a$ | Sign $b$ | Sign $c$ | Bianchi type |
|---|---|---|---|---|---|
| I | 0 | 0 | 0 | 0 | I |
| II | 0 | + | 0 | 0 | II |
| $VII_0$ | 0 | + | + | 0 | VII |
| $VI_0$ | 0 | + | − | 0 | VI |
| IX | 0 | + | + | + | IX |
| VIII | 0 | + | + | − | VIII |
| V | + | 0 | 0 | 0 | V |
| IV | + | + | 0 | 0 | IV |
| $VII_h$ | + | + | + | 0 | VII |
| $VI_h$ | + | + | − | 0 | VI, III |

In Table I we summarize the Bianchi–Behr types in terms of these parameters. In two cases, denoted by Behr as Types $VI_h$ and $VII_h$, the type is subclassified by a continuous invariant parameter $h = n^2/ab$. Behr has further obtained coordinate components of the specialized $\mathbf{e}^a$'s expressed in terms of $a$, $b$, $c$, and $n$, so that the invariant differential form $\mathbf{e}^a \cdot d\mathbf{x}$ is explicitly known for each of the cosmological models.

In the present paper we recast the cosmological problem in terms of the dyadic formalism. A superficially quite different mathematical formulation results: In particular, we now deal with sets of first-order differential equations. This, plus the fact that all quantities in this formalism have immediate physical interpretation, gives rise to the hope that further exact solutions or at least further easily analyzable subcases can be found.

[4] C. G. Behr, Z. Astrophys. **54**, 268 (1962); **60**, 286 (1965); Astron. Abhandl. Hbg. Sternwarte **7**, 249 (1965).
[5] C. G. Behr (to be published).

We recapitulate the general dyadic approach to general relativity in Sec. II, and then specialize the equations to the cosmological problem. In Sec. III the Bianchi–Behr classification is found to emerge naturally in the dyadic formalism. We prove that—unless a singular state occurs—the cosmological principle alone implies constancy of the Bianchi–Behr type (even to the parameter $h$) of the homogeneous spacelike hypersurfaces throughout space–time, regardless of the state and evolution of the matter content. We consider the matter content in Sec. IV, and derive dyadic expressions for the acceleration, rotation, shear, and expansion of the matter congruence in the case of perfect fluid. As an example of the dyadic approach, the equations for Type V models with incoherent matter are explicitly discussed in Sec. V. In Sec. VI we consider cosmologies of Types VIII and IX with incoherent matter, and are able to recover and generalize some elegant results previously announced without proof by Gödel.

## II. DYADIC EQUATIONS

As a physical theory, 3 + 1 Riemannian geometry takes a convenient form as four asymmetric first-order dyadic differential equations[6,7]:

$$\nabla\mathbf{a} - (\dot{\mathbf{S}} + \boldsymbol{\omega} \times \mathbf{S} - \mathbf{S} \times \boldsymbol{\omega}) + (\dot{\boldsymbol{\Omega}} + \boldsymbol{\omega} \times \boldsymbol{\Omega}) \times \mathbf{I}$$
$$= \mathbf{S} \cdot \mathbf{S} - \boldsymbol{\Omega} \times \mathbf{S} - \mathbf{S} \times \boldsymbol{\Omega} + \boldsymbol{\Omega}\boldsymbol{\Omega} - (\Omega^2)\mathbf{I} - \mathbf{a}\mathbf{a}$$
$$+ \mathbf{A} + \mathbf{T} + \tfrac{1}{3}(\rho - 2\,\mathrm{Tr}\,\mathbf{T} - \Lambda)\mathbf{I}, \quad (1)$$

$$\nabla\boldsymbol{\Omega} + \mathbf{S} \times \nabla = -2\mathbf{a}\boldsymbol{\Omega} + (\mathbf{a} \cdot \boldsymbol{\Omega})\mathbf{I} + \mathbf{B} + \mathbf{t} \times \mathbf{I}, \quad (2)$$

$$\dot{\mathbf{N}} + \mathbf{S}^* \cdot \mathbf{N} = \mathbf{S}^{*T} \times (\nabla + \mathbf{a})$$
$$+ [(\nabla + \mathbf{a}) \cdot (\boldsymbol{\Omega} - \boldsymbol{\omega})]\mathbf{I}, \quad (3)$$

$$\nabla \times \mathbf{N} = -\tfrac{1}{2}\mathbf{N}^T \overset{\times}{\times} \mathbf{N} + \mathbf{E} + (\boldsymbol{\Omega} \cdot \mathbf{S}^*) \times \mathbf{I}, \quad (4)$$

where

$$\mathbf{S}^* = \mathbf{S} - (\boldsymbol{\Omega} - \boldsymbol{\omega}) \times \mathbf{I}, \quad \mathbf{S}^{*T} = \mathbf{S} + (\boldsymbol{\Omega} - \boldsymbol{\omega}) \times \mathbf{I}, \quad (5)$$

and

$$-\mathbf{E} = \mathbf{A} - \mathbf{T} + \tfrac{1}{3}(\mathrm{Tr}\,\mathbf{T} - 2\rho - \Lambda)\mathbf{I}$$
$$+ \tfrac{1}{2}\mathbf{S} \overset{\times}{\times} \mathbf{S} + \boldsymbol{\Omega}\boldsymbol{\Omega} + \boldsymbol{\Omega}\boldsymbol{\omega} + \boldsymbol{\omega}\boldsymbol{\Omega}. \quad (6)$$

This completely general formulation is based on an arbitrary timelike reference congruence kinematically described by three-vectors of acceleration and vorticity $\mathbf{a}$ and $\boldsymbol{\Omega}$, and by a symmetric rate-of-strain dyadic $\mathbf{S}$. The trace of $\mathbf{S}$ is the expansion, the trace-free part the shear, of the reference congruence. The 3-space

[6] F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **5**, 1629 (1964).
[7] H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **7**, 894 (1966).

orthogonal to the timelike reference congruence is spanned by an orthonormal spacelike triad reference vector frame having angular velocity $\boldsymbol{\omega}$ and nine triad rotation coefficients $\mathbf{N}$. It is a quotient space ($V_4^3$) unless $\boldsymbol{\Omega} = 0$; in this case the antisymmetric part of Eq. (4) is a familar differential identity implying a 3-metric, Eq. (6) is a Gauss equation for this induced metric, and $\mathbf{E}$, its Einstein curvature dyadic, satisfies the 3-dimensional Bianchi identity $\boldsymbol{\nabla} \cdot \mathbf{E} = 0$. The operator $(\dot{})$ is differentiation along the timelike congruence; the operator $\boldsymbol{\nabla}$ is covariant differentiation in the quotient space. The usual vector operations $\cdot$ and $\times$ serve conveniently to make explicit the consequences of the signature of space–time, which here leads to a positive-definite quotient space.

Of the 36 first-order equations in (1), (2), (3), and (4), 20 serve to define the Riemann tensor components[8]: $\mathbf{T}$, $\mathbf{t}$, and $\rho$ are, in Einstein's theory, identified as the symmetric stress dyadic, momentum density vector, and energy density (in relativistic units); $\mathbf{A}$ and $\mathbf{B}$ are symmetric and traceless, respectively, the so-called "electric" and "magnetic" local components of the Weyl tensor. $\Lambda$ is the cosmological constant.

The timelike reference congruence we adopt in the following, to reduce the space–time of a world model, is the normal congruence of the homogeneous subspaces. The cosmological principle requires any scalar formed from the induced metrical structure on any of these, or from its second fundamental form, or from the timelike normal congruence *to be constant in the 3-space*, and so *not* to allow any geometric or kinematic means of intrinsic identification of the points in the 3-space. Thus first the angular velocity of the reference congruence $\boldsymbol{\Omega} = 0$; moreover, since the acceleration $\mathbf{a}$ of this congruence is then expressible as the gradient of a scalar potential $\phi$,[7] and since all scalars are allowed to be functions only of time, we immediately have also $\mathbf{a} = 0$. This is the dyadic proof that the reference congruence is geodesic; i.e., that the homogeneous subspaces are geodesically parallel. (In general, this reference congruence will *not* coincide with the world lines of the matter.) Under these circumstances the dyadic formulation becomes quite simple: The remaining variables of the problem are just the symmetric rate-of-strain dyadic $\mathbf{S}$ of the reference congruence (which is here also the second fundamental form of the 3-spaces) and the asymmetric Ricci rotation dyadic $\mathbf{N}$ that describes the intrinsic metric geometry of the 3-spaces.

[8] There is a slight notation change here from Refs. 6 and 7. There the cosmological term in the field equations, if any, was incorporated in $T_{\mu\nu}$ and so in $\mathbf{T}$ and $\rho$. Here we write it explicitly, and $\mathbf{T}$, $\mathbf{t}$, and $\rho$ describe just the local physics.

In the world models considered in the present paper, there is always the possibility of three intrinsically defined orthonormal reference vector congruences in the homogeneous 3-spaces. Adopting these, the components of any dyadic quantity appear as intrinsic scalar fields. In Ref. 6 we defined a differentiation operation $\mathbf{D}$ which operates only on the components of dyadic quantities. As all intrinsic scalars are to be constant in the 3-spaces, the differentiation operator $\mathbf{D}$ applied to any quantity gives zero. So to obtain the covariant 3-space derivatives of any vector $\mathbf{V}$ or dyadic $\mathbf{M}$ we use the identities[6]

$$\boldsymbol{\nabla}\mathbf{V} = \mathbf{D}\mathbf{V} - \mathbf{N} \times \mathbf{V}, \tag{7}$$

$$\boldsymbol{\nabla} \cdot \mathbf{M} = \mathbf{D} \cdot \mathbf{M} - \mathbf{N}^T \dot{\times} \mathbf{M} - 2\mathbf{n} \cdot \mathbf{M}, \tag{8}$$

$$\boldsymbol{\nabla} \times \mathbf{M} = \mathbf{D} \times \mathbf{M} + (\operatorname{Tr} \mathbf{N})\mathbf{M} - \mathbf{N}^T \cdot \mathbf{M} - \mathbf{N}^{T\times}_{\times}\mathbf{M}, \tag{9}$$

putting, henceforth, $\mathbf{D}\mathbf{V} = 0$ and $\mathbf{D}\mathbf{M} = 0$. We use the notation that $\mathbf{N} = \mathbf{N}^S - \mathbf{n} \times \mathbf{I}$ where $\mathbf{N}^S$ is symmetric. $\mathbf{N}^T = \mathbf{N}^S + \mathbf{n} \times \mathbf{I}$.

From (2), we can solve for $\mathbf{B}$ and $\mathbf{t}$ in terms of $\mathbf{N}$ and $\mathbf{S}$:

$$\mathbf{B} - \mathbf{t} \times \mathbf{I} = -(\operatorname{Tr} \mathbf{N})\mathbf{S} + \mathbf{N}^T \cdot \mathbf{S} + \mathbf{N}^{T\times}_{\times}\mathbf{S}. \tag{10}$$

This equation is traceless. Equation (4) gives $\mathbf{E}$ in terms of $\mathbf{N}$:

$$\mathbf{E} = (\operatorname{Tr} \mathbf{N})\mathbf{N} - \mathbf{N}^T \cdot \mathbf{N} - \tfrac{1}{2}\mathbf{N}^{T\times}_{\times}\mathbf{N}, \tag{11}$$

the antisymmetric part of which is an identity which must be satisfied by $\mathbf{N}$,

$$\mathbf{N} \dot{\times} \mathbf{N} = 0, \tag{12}$$

or

$$\mathbf{F} \cdot \mathbf{n} = 0, \tag{13}$$

where we define

$$\mathbf{F} = \mathbf{N}^S - (\operatorname{Tr} \mathbf{N})\mathbf{I}. \tag{14}$$

$\mathbf{n}$ is thus an eigenvector of the symmetric dyadic $\mathbf{F}$, with zero eigenvalue. (This is the dyadic guise taken by the Jacobi identity of the underlying 3-parameter isometry group!) We henceforth use $\mathbf{F}$ and $\mathbf{n}$ in preference to $\mathbf{N}$. The trace of (6) gives $\rho$, and from (11) the 3-space curvature is described by

$$\mathbf{E} = -2\mathbf{F} \cdot \mathbf{F} + (\operatorname{Tr} \mathbf{F})\mathbf{F} - \mathbf{n} \times \mathbf{F} + \mathbf{F} \times \mathbf{n} - \tfrac{1}{2}[\tfrac{1}{2}(\operatorname{Tr} \mathbf{F})^2 - \mathbf{F}:\mathbf{F} + 2n^2]\mathbf{I}. \tag{15}$$

The dyadic $\mathbf{A}$ is now given, in terms of $\mathbf{F}$, $\mathbf{n}$, $\mathbf{S}$, and $\mathbf{T}$ by the trace-free part of (6).

There remain six equations in (1) and nine equations in (3), which give the time behavior of the variables $\mathbf{S}$, $\mathbf{F}$, and $\mathbf{n}$. These are the first-order differential equations which state the mathematical problem at

hand, solutions of which are relativistic homogeneous cosmologies:

$$\dot{\mathbf{S}} + \boldsymbol{\omega} \times \mathbf{S} - \mathbf{S} \times \boldsymbol{\omega} = -2\mathbf{T} - (\mathrm{Tr}\,\mathbf{S})\mathbf{S} - 2\mathbf{F}\cdot\mathbf{F} + (\mathrm{Tr}\,\mathbf{F})\mathbf{F} - \mathbf{n} \times \mathbf{F} + \mathbf{F} \times \mathbf{n} + [\mathrm{Tr}\,\mathbf{T} + \tfrac{1}{2}\Lambda + \tfrac{1}{4}(\mathrm{Tr}\,\mathbf{S})^2 - \tfrac{1}{4}\mathbf{S}:\mathbf{S} + \tfrac{3}{4}\mathbf{F}:\mathbf{F} - \tfrac{3}{8}(\mathrm{Tr}\,\mathbf{F})^2 + \tfrac{1}{2}n^2]\mathbf{I}, \quad (16)$$

$$\dot{\mathbf{F}} + \boldsymbol{\omega} \times \mathbf{F} - \mathbf{F} \times \boldsymbol{\omega} = \mathbf{S}\cdot\mathbf{F} + \mathbf{F}\cdot\mathbf{S} - (\mathrm{Tr}\,\mathbf{S})\mathbf{F}, \quad (17)$$

$$\dot{\mathbf{n}} + \boldsymbol{\omega} \times \mathbf{n} = -\mathbf{n}\cdot\mathbf{S}. \quad (18)$$

In these equations $\boldsymbol{\omega}$ may be chosen arbitrarily (Ref. 6) and, as we will see, $\mathbf{T}$ is to be given by local physical considerations to complete the differential set.

We do not labor the essential equivalence of the dyadic approach and the approach described in the Introduction (Ref. 5), but this equivalence is perhaps obvious after one recognizes the algebraic identity of the components of the dyadic $\mathbf{N} - (\mathrm{Tr}\,\mathbf{N})\mathbf{I} = \mathbf{F} - \mathbf{n} \times \mathbf{I}$ and the tensor density $A^{il}$ belonging to the basis here being used. In the dyadic approach this vector basis $\mathbf{u}^a$ in the 3-spaces is once and for all time *taken orthonormal*, so that for it the inner products $\gamma^{ab} =$ diag (1, 1, 1) everywhere, rather than being functions of time. Moreover, the field equations are cast as *first*-order equations for the variables in $\mathbf{N}$, which now *are* functions of time, together with the variables in $\mathbf{S}$, the components of the second fundamental form of the immersed homogeneous 3-spaces. In the present formulation, the final integrations to obtain the metric explicitly as a function of holonomic coordinates are deferred until all the triad Ricci rotation coefficients $\mathbf{N}$ are known—the *essential* mathematical problem being the determination of these latter. Local physical quantities of interest are always algebraic functions of $\mathbf{S}$ and $\mathbf{N}$. We will only briefly discuss such a final integration to determine a metric form in the next section.

## III. CONSERVATION OF BIANCHI–BEHR TYPE

We may now make a choice of intrinsic triad reference vectors in each successive homogeneous 3-space. Such a choice determines $\boldsymbol{\omega}$. The choice which usually appears to be most convenient for the present problem is to everywhere diagonalize $\mathbf{F}$; furthermore, using Eq. (13), we choose the direction of $\mathbf{n}$:

$$F^{ab} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}, \quad n^a = (0, 0, n), \quad nc = 0. \quad (19)$$

Inserting this into (17) and (18), we write in components first the differential equations for the eigenvalues of $\mathbf{F}$ and the magnitude of $\mathbf{n}$:

$$\dot{a} = (S_{11} - S_{22} - S_{33})a, \quad (20)$$

$$\dot{b} = (-S_{11} + S_{22} - S_{33})b, \quad (21)$$

$$\dot{c} = (-S_{11} - S_{22} + S_{33})c, \quad (22)$$

$$\dot{n} = -S_{33}n, \quad (23)$$

and next the equations which give $\boldsymbol{\omega}$:

$$\omega_1 n = S_{23}n, \quad (24)$$

$$\omega_2 n = -S_{13}n, \quad (25)$$

$$\omega_1(b - c) = S_{23}(b + c), \quad (26)$$

$$\omega_2(c - a) = S_{31}(c + a), \quad (27)$$

$$\omega_3(a - b) = S_{12}(a + b). \quad (28)$$

We may from the form of Eqs. (20)–(23) immediately draw the conclusions that in any given cosmological solution, irrespective of the physical content (described by the stress $\mathbf{T}$, which we have not yet specified), (a) either $\mathbf{n}$ vanishes or does not vanish for all time, and (b) that $a$, $b$, and $c$ are each either zero, positive, or negative *and remain so* for all time, unless a singularity occurs and our differential analysis fails. Thus we arrive precisely at the type of classification scheme already given by Behr (Table I). Type $\mathrm{VI}_h$ and $\mathrm{VII}_h$ homogeneous 3-spaces are further classified by the value of $h$, where

$$h = n^2/ab. \quad (29)$$

Bianchi Type III is the subcase of Type $\mathrm{VI}_h$ when $h = -1$. From Eqs. (20)–(23) we quickly find

$$\dot{h} = 0 \quad (30)$$

so that, again, this parameter classifies globally such cosmological models.

We wish to call attention to the structure we have found for the differential equations for $a$, $b$, $c$, and $n$—that the setting of any of these equal to zero implies that its first, and higher, derivatives are zero, and hence that it *remains* zero so long as all quantities are bounded. By setting such a variable equal to zero, we consistently reduce by one the number of differential equations in the set, and so have a classifiable, distinct subtype of cosmological model at hand for solution. Again, it may happen that if several variables of a set (for example, for perfect fluids, the three variables in $\mathbf{t}$) are simultaneously set equal to zero, that an equal number of the first-order differential

equations are identically satisfied, and that the set of relations continues to remain zero as time progresses. Such relations have been discussed by Levi-Civita,[9] who denotes them "invariant relations" or "invariant sets" with respect to the set of $n$ ordinary first-order differential equations being considered. Such relations serve to separate and classify the congruence of trajectories describing the general solution, in the $n$-dimensional space of dependent variables, by means of subspaces which are themselves not crossed by trajectories. We will denote such variables and sets as IR variables and IR sets. We regard the classifying of solutions of a complete set of first-order differential equations as equivalent to the recognition of IR variables and IR sets.

We further wish to remark that the present result, the conservation of Bianchi–Behr type, is a consequence of the cosmological postulate *without reference to the field equations* (i.e., irrespective of the state of the matter content). It is purely a consequence of the 3 + 1 Riemannian geometry and is contained precisely in our dyadic Eq. (3). The set of relations (3) is not at all easy or natural to express in more conventional tensorial terminology, and yet is vital for the completeness of the total set of relations in Eqs. (1), (2), (3), and (4), that they together describe the geometry of 3 + 1 Riemannian space–time. It is an advantage of the dyadic formalism that the set in Eq. (3) is explicit.

So, then, adopting the Bianchi–Behr classification, we are left with the problem of integration of complete first-order sets consisting of six equations (16) and of four equations (20)–(23), where from one to four of the latter vanish identically for any given type. [In some cases, however, this situation may be improved upon further by a *different* choice of $\boldsymbol{\omega}$ in Eqs. (16)–(18)—we will illustrate this in Sec. V.] After integration of the first-order set giving $\mathbf{F}$, $\mathbf{n}$, $\boldsymbol{\omega}$, and $\mathbf{S}$, the local physics is everywhere known for the model. However, it sometimes may be desired to go further: to obtain an explicit space–time metric as in Sec. I. Indeed we already have at hand a set of coordinate components of invariant triads $\mathbf{e}^a$ for each Bianchi type (Ref. 5) (these are functions of particular canonical coordinates $x^\alpha$, where $\alpha = 1, 2, 3$, and *constant* structure parameters, say $F_0^{ab}$ and $n_{0a}$, or, if specialized, just $a_0$, $b_0$, $c_0$, and $n_0$). But now how are their inner products to be found? We introduce a matrix of affine transformation $C^a_{\cdot b}$ connecting the invariant basis $\mathbf{e}^a$ and the orthonormal basis $\mathbf{u}^a$,

$$\mathbf{e}^a = C^a_{\cdot b}\mathbf{u}^b, \quad \text{so} \quad \mathbf{e}^a \cdot \mathbf{e}^b = \gamma^{ab} = C^a_{\cdot c}C^b_{\cdot d}\delta^{cd}. \tag{31}$$

$\dot{\mathbf{u}}^a = 0$ by definition, and $\dot{\mathbf{e}}^a = -\boldsymbol{\omega} \times \mathbf{e}^a - \mathbf{S} \cdot \mathbf{e}^a$ because the $\mathbf{e}^a$ have the property of being comoving with the reference congruence (Ref. 7). Thus, for the unfolding affine matrix $C^a_{\cdot b}$ and its transposed inverse $C^{\cdot b}_a$ we find the equivalent differential equations

$$\begin{aligned} &\dot{C}^a_{\cdot b} - \epsilon_{bcd}\omega_d C^a_{\cdot c} + S_{bc}C^a_{\cdot c} = 0, \\ &\dot{C}^{\cdot b}_a - \epsilon_{bcd}\omega_d C^{\cdot c}_a - S_{bc}C^{\cdot c}_a = 0, \\ &|\det C^a_{\cdot b}|^{\cdot} = -(\operatorname{Tr}\mathbf{S})\,|\det C^a_{\cdot b}|. \end{aligned} \tag{31'}$$

It is these coupled linear equations which finally must be integrated, so that the $\gamma_{ab}(t)$ can be calculated. Nine integration constants are at first sight required, but the solutions $C^a_{\cdot b}(t)$, as is easily shown from Eqs. (17) and (18), will satisfy tensor-density-type transformation equations under the affine change of basis:

$$n_b(t) = C^a_{\cdot b}n_{0a}, \quad F^{ab}(t) = |\det C^e_{\cdot f}|\, C^a_{\cdot c}C^b_{\cdot d}F_0^{cd}. \tag{32}$$

Thus, *knowing* the constants $n_{0a}$ and $F_0^{cd}$ for a given Bianchi–Behr type, Eq. (32) fixes a number of the integration constants. The extent to which the integration constants remain arbitrary expresses the so-called automorphism group of the invariant basis.

In practice, since we usually take $\mathbf{F}$ diagonal, and $\mathbf{n}$ aligned, and since the canonical 3-space metrics given by Behr in Ref. 5 are similarly specialized in the forms $F_0^{ab} = \operatorname{diag}(a_0, b_0, c_0)$ and $n_{0a} = (0, 0, n_0)$, the arbitrariness allows the $C^a_{\cdot b}$ to be taken quite special, *ab initio*. For example, in Type IX, we may write the matrix of $C^a_{\cdot b}$ as just

$$C^a_{\cdot b} = \begin{pmatrix} a_0^{+\frac{1}{2}} & 0 & 0 \\ 0 & b_0^{+\frac{1}{2}} & 0 \\ 0 & 0 & c_0^{+\frac{1}{2}} \end{pmatrix} \times \begin{pmatrix} 3 \times 3 \text{ orthog-} \\ \text{onal matrix:} \\ 3 \text{ parameters} \end{pmatrix} \begin{pmatrix} a^{-\frac{1}{2}} & 0 & 0 \\ 0 & b^{-\frac{1}{2}} & 0 \\ 0 & 0 & c^{-\frac{1}{2}} \end{pmatrix} f, \tag{33}$$

where $f = (a\,b\,c/a_0\,b_0\,c_0)^{\frac{1}{2}}\,|\det \text{orthogonal matrix}|^{-1}$, and there are but three functions to be found from the last integration.

## IV. THE MATTER CONTENT: PERFECT FLUID, INCOHERENT MATTER

We consider now the choice of matter content for these cosmological models. We already have, from Eqs. (6) and (10), the local energy and momentum

[9] T. Levi-Civita and U. Amaldi, *Lezioni di Meccanica Razionale* (Nicola Zanichelli, Bologna, 1926), Vol. II, pp. 339–353; cf. also J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), p. 215.

densities $\rho$ and $\mathbf{t}$ (as seen by an observer whose congruence is the reference congruence with $\mathbf{a} = \boldsymbol{\Omega} = 0$):

$$\rho = \tfrac{1}{4}\{(\mathrm{Tr}\,\mathbf{S})^2 - \mathbf{S}:\mathbf{S} - \mathbf{F}:\mathbf{F} + \tfrac{1}{2}(\mathrm{Tr}\,\mathbf{F})^2 - 6n^2 - 2\Lambda\}, \tag{34}$$

$$\mathbf{t} = \tfrac{1}{2}\{\mathbf{F} \dot{\times} \mathbf{S} - (\mathrm{Tr}\,\mathbf{S})\mathbf{n} + 3\mathbf{n}\cdot\mathbf{S}\}. \tag{35}$$

Furthermore, we may directly calculate from these (or use the contracted Bianchi identities in dyadic notation) the conservation laws,

$$\dot{\rho} = -(\mathrm{Tr}\,\mathbf{S})\rho + \mathbf{T}:\mathbf{S} + 2\mathbf{n}\cdot\mathbf{t}, \tag{36}$$

$$\dot{\mathbf{t}} + \boldsymbol{\omega}\times\mathbf{t} = -(\mathrm{Tr}\,\mathbf{S})\mathbf{t} - \mathbf{t}\cdot\mathbf{S} - 3\mathbf{n}\cdot\mathbf{T} + (\mathrm{Tr}\,\mathbf{T})\mathbf{n} - \mathbf{F}\dot{\times}\mathbf{T}. \tag{37}$$

It remains to specify the expression of the stress dyadic $\mathbf{T}$ in terms of $\mathbf{t}$ and $\rho$, for insertion in Eq. (16). For a perfect fluid, for example, we have

$$\mathbf{T} = -p\mathbf{I} - (p+\rho)^{-1}\mathbf{tt}, \tag{38}$$

where $p$ is the pressure. A noteworthy result of this relation is that Eq. (37) then becomes such that $\mathbf{t} = 0$ implies $\dot{\mathbf{t}} = 0$. We recognize the same structure as we met in Eqs. (17) and (18): $\mathbf{t}$ is an IR set, and subtypes with $\mathbf{t} = 0$ may be formulated for cosmologies with perfect fluid matter content.

The velocity 4-vector of a fluid is the timelike eigenvector of its energy–momentum tensor. If the reference tetrad components of this unit 4-vector are written as $\lambda^r = \gamma(1, \mathbf{V})$, where $\gamma = (1 - \mathbf{V}\cdot\mathbf{V})^{-\frac{1}{2}}$, we define the fluid 3-velocity $\mathbf{V}$. In a proper frame moving with this velocity, a proper observer will measure pressure $p$ and proper density, say ${}_P\rho$. From (38), ${}_P\rho$ and $\mathbf{V}$ are then found to be related to $\rho$ and $\mathbf{t}$ by

$$\rho = \gamma^2[{}_P\rho + pV^2], \quad \mathbf{V} = \mathbf{t}/(p+\rho). \tag{39}$$

It is usual in cosmology to describe the kinematics of the matter content in the comoving or proper frame; i.e., by the acceleration ${}_Pa_r$, rotation ${}_P\Omega_{rs}$, shear ${}_P\sigma_{rs}$, and expansion ${}_P\theta$ of the $\lambda^r$ congruence:

$$\begin{aligned} &{}_Pa_r = \lambda_{r;s}\lambda^s, \quad {}_P\Omega_{rs} = \lambda_{[r;s]} + {}_Pa_{[r}\lambda_{s]}, \\ &{}_P\sigma_{rs} = \lambda_{(r;s)} + {}_Pa_{(r}\lambda_{s)} - \tfrac{1}{3}{}_P\theta(g_{rs} + \lambda_r\lambda_s), \quad {}_P\theta = \lambda^r_{\ ;r}. \end{aligned} \tag{40}$$

We can describe this kinematics of the matter congruence in dyadic language, and also use clocks moving on the reference congruence, by introducing vectors $\bar{\mathbf{a}}$ and $\bar{\boldsymbol{\Omega}}$, a symmetric dyadic $\bar{\boldsymbol{\sigma}}$, and a scalar $\bar{\theta}$ according to

$$\begin{aligned} &{}_Pa_r = \gamma(-\bar{\mathbf{a}}\cdot\mathbf{V}, \bar{\mathbf{a}}), \quad {}_P\Omega_{rs} = \gamma\begin{pmatrix} 0 & -\bar{\boldsymbol{\Omega}}\times\mathbf{V} \\ \bar{\boldsymbol{\Omega}}\times\mathbf{V} & -\bar{\boldsymbol{\Omega}}\times\mathbf{I} \end{pmatrix}, \\ &{}_P\sigma_{rs} = \gamma\begin{pmatrix} \mathbf{V}\cdot\bar{\boldsymbol{\sigma}}\cdot\mathbf{V} & -\bar{\boldsymbol{\sigma}}\cdot\mathbf{V} \\ -\mathbf{V}\cdot\bar{\boldsymbol{\sigma}} & \bar{\boldsymbol{\sigma}} \end{pmatrix}, \quad {}_P\theta = \gamma\bar{\theta}. \end{aligned} \tag{41}$$

To calculate these, we expand Eq. (40) in the orthonormal reference frame, thus introducing *its* kinematic quantities $\mathbf{S}$, $\mathbf{a}$, and $\boldsymbol{\Omega}$—the last two, however, are zero in the present case. We use Eq. (7) to calculate $\nabla\mathbf{V}$, and find $\dot{\mathbf{V}}$ from (36) and (37):

$$\begin{aligned} \dot{\mathbf{V}} + \boldsymbol{\omega}\times\mathbf{V} = {}&\mathbf{V}\cdot\mathbf{F}\times\mathbf{V} + \underline{\mathbf{n}}\cdot\mathbf{VV} - V^2\mathbf{n} \\ &- \mathbf{V}\cdot\mathbf{S} + \mathbf{V}\cdot\mathbf{S}\cdot\mathbf{VV} - [\dot{p}/(p+\rho)]\mathbf{V}. \end{aligned} \tag{42}$$

The final result is

$$\begin{aligned} \bar{\mathbf{a}} &= -[\gamma^3\dot{p}/(p+\rho)]\mathbf{V}, \\ \bar{\boldsymbol{\Omega}} &= -\tfrac{1}{2}\mathbf{V}\cdot\mathbf{F} - \tfrac{1}{2}\mathbf{n}\times\mathbf{V}, \\ \bar{\boldsymbol{\sigma}} &= \tfrac{1}{2}(\mathbf{Vn} + \mathbf{nV}) - \mathbf{n}\cdot\mathbf{VI} + \tfrac{1}{2}(\mathbf{V}\times\mathbf{F} - \mathbf{F}\times\mathbf{V}) \\ &\quad + \mathbf{S} - [\gamma^4\dot{p}/(p+\rho)]\mathbf{VV} - \tfrac{1}{3}\bar{\theta}(\mathbf{I} + \gamma^2\mathbf{VV}), \\ \bar{\theta} &= -2\mathbf{n}\cdot\mathbf{V} + \mathrm{Tr}\,\mathbf{S} - \mathbf{V}\cdot\mathbf{S}\cdot\mathbf{V} - [\gamma^2V^2\dot{p}/(p+\rho)]. \end{aligned} \tag{43}$$

To repeat, these are the quantities which describe the proper kinematics of the matter congruence in 3-dimensional language based on the reference congruence. Using (39), they are calculable from $\mathbf{S}$, $\mathbf{F}$, $\mathbf{n}$, and $p$. $\dot{p}$ must be given by local physical considerations to complete the first-order set of equations for these variables.

The following may each be shown to be an IR set or variable: $\mathbf{V}$; $\gamma$; $\mathbf{V}\cdot\mathbf{F}$; $\mathbf{t}\cdot\mathbf{F}$; $\mathbf{n}\times\mathbf{V}$; $\mathbf{t}\cdot\mathbf{F}\cdot\mathbf{t}$; $\bar{\boldsymbol{\Omega}}$. For example,

$$\begin{aligned} \dot{\bar{\boldsymbol{\Omega}}} + \boldsymbol{\omega}\times\bar{\boldsymbol{\Omega}} &= [\mathbf{S} + \{\mathbf{n}\cdot\mathbf{V} + \mathbf{V}\cdot\mathbf{S}\cdot\mathbf{V} - \mathrm{Tr}\,\mathbf{S} - \dot{p}/(p+\rho)\}\mathbf{I}]\cdot\bar{\boldsymbol{\Omega}}, \\ \dot{\gamma} &= -\mathbf{V}\cdot\mathbf{S}\cdot\mathbf{V}\gamma - [V^2\gamma^3\dot{p}/(p+\rho)], \\ (\mathbf{t}\cdot\mathbf{F}\cdot\mathbf{t})^{\cdot} &= 3(2\mathbf{n}\cdot\mathbf{V} - \mathrm{Tr}\,\mathbf{S})\mathbf{t}\cdot\mathbf{F}\cdot\mathbf{t}. \end{aligned} \tag{44}$$

Thus the presence or absence of any of these may be used to classify cosmologies with perfect fluid content. We have not, however, found any IR sets involving the matter shear and expansion $\bar{\boldsymbol{\sigma}}$ and $\bar{\theta}$. When $\mathbf{n} = 0$, the last of Eq. (44) leads immediately, with Eqs. (20)–(22), to a first integral of angular momentum

$$\mathbf{t}\cdot\mathbf{F}\cdot\mathbf{t}/(\det F)^3 = \text{const.} \tag{45}$$

As an example of completion, we consider now, and henceforth in this paper, the case of incoherent (or dustlike) matter, $p = 0$. The right sides of the first-order differential set of equations [(17)–(18)] can be shown to be analytical functions of the dependent variables when $\rho$ [given by Eq. (14)] is nonzero. The fundamental theorem for such a set guarantees the existence, uniqueness, and analyticity of solutions so long as all quantities remain finite. Thus the mathematical requirements which we previously invoked in discussing invariant relations are clearly satisfied. In this incoherent case, we have, naturally, the density

$\rho$ as an IR variable:

$$\dot{\rho} = \{-\mathrm{Tr}\,\mathbf{S} - \mathbf{V}\cdot\mathbf{S}\cdot\mathbf{V} + 2\mathbf{n}\cdot\mathbf{V}\}\rho. \quad (46)$$

From (44) and (46) we find an integral

$$\gamma^3\mathbf{V}\cdot\mathbf{F}\cdot\mathbf{V}/\rho = \text{const.} \quad (47)$$

In vector-free cosmologies ($\boldsymbol{n} = 0$), as we will see in Sec. VI, this may be combined with the integral of angular momentum to yield a local conservation law for energy. We will in the next two sections consider, respectively, Type V, and Types VIII and IX, with $p = 0$, to illustrate these various relations.

## V. TYPE V COSMOLOGIES

To illustrate the above considerations, let us put $\mathbf{F} = 0$ and $p = 0$. This we know that we may consistently do, since $\mathbf{F}$ is an IR set from Eq. (17). Let us now choose $\boldsymbol{\omega}$ so that $\mathbf{S}$ is everywhere diagonal. The three off-diagonal equations in (16) give $\boldsymbol{\omega}$. We are left with six first-order differential equations for solution: the diagonal equations in Eqs. (16) and (18).

For incoherent matter, with $\Lambda = 0$, the right-hand sides of these equations are homogeneous of the second order in the variables $S_{11}, S_{22}, S_{33}, n_1, n_2, n_3$. The only nonquadratic terms are those from the matter tensor: $\mathbf{T} = -\rho^{-1}\mathbf{tt}$, $\mathbf{t}$ and $\rho$ being quadratic functions of the variables, according to Eqs. (31) and (32). The equations nevertheless appear impossible for explicit solution. Rotation is in general present. We note that the integrals (45) and (47) become empty for such a special Bianchi–Behr type.

The rotation-free case occurs when the IR vector $\mathbf{V}\times\mathbf{n}$ vanishes. Now we find $\boldsymbol{\omega} = 0$ and $\mathbf{V}$ and $\mathbf{n}$ to be eigenvectors of $\mathbf{S}$—we take them to be in the 3-direction. With the new variables $\theta = S_{11} + S_{22} + S_{33}$ to describe the expansion and $\sigma = S_{11} - S_{22}$ and $\tau = -S_{11} - S_{22} + 2S_{33}$ to describe the shear (of the reference congruence!), we have a set of four equations for solution:

$$\begin{aligned}
\dot{\sigma} &= -\theta\sigma,\\
\dot{\tau} &= -\theta\tau + \rho^{-1}\tau^2 n^2,\\
\dot{\theta} &= -\tfrac{1}{2}\theta^2 - \tfrac{1}{4}\rho^{-1}\tau^2 n^2 + \tfrac{3}{2}\Lambda - \tfrac{3}{8}\sigma^2 - \tfrac{1}{8}\tau^2 + \tfrac{3}{2}n^2,\\
\dot{n} &= -\tfrac{1}{3}\theta n - \tfrac{1}{3}\tau n,
\end{aligned} \quad (48)$$

where

$$4\rho = \tfrac{2}{3}\theta^2 - \tfrac{1}{2}\sigma^2 - \tfrac{1}{6}\tau^2 - 6n^2 - 2\Lambda. \quad (49)$$

One first integral has been found: $n = CV(1 - V^2)^{-\frac{1}{2}}$, where $2\rho V = \tau n$ and $C = \text{const.}$

We note that $\sigma$ and $\tau$ are separately IR variables. The subcase $\sigma = 0$ is, interestingly, of Petrov type D. The subcase $\tau = 0$, which is the case when all of $\mathbf{t} = 0$, is especially simple. Now two integrations may be performed, and in the remaining equation the substitution $n = R^{-1}$ leads immediately to the "generalized Friedmann equation" given by Heckmann and Schücking (Ref. 3, p. 445). Indeed, from Eq. (11) we have the curvature of the 3-spaces to be $\mathbf{E} = -n^2\mathbf{I}$, isotropic for all Type V models, so $R$ is of course seen to be the Gaussian curvature.

## VI. TYPE VIII AND IX COSMOLOGIES

In Eqs. (24)–(28) we insert $\mathbf{n} = 0$ and $p = 0$ and solve for $\boldsymbol{\omega}$. The equations for $a, b, c$ are (20)–(22); for the components of $\mathbf{S}$ we refer to (16). Let us denote these last by

$$S_{ab} = \begin{pmatrix} \phi & v & \tau \\ v & x & \sigma \\ \tau & \sigma & \psi \end{pmatrix}. \quad (50)$$

The components of $\mathbf{t}$ are given by (35): $2t_1 = (b - c)\sigma$, and 2 cyclic permutations. The complete set of 9 first-order equations for a Type VIII or IX cosmology with incoherent matter is

$$\dot{a} = (\phi - x - \psi)a \quad (51)$$

together with 2 cyclic permutations;

$$\begin{aligned}
\dot{\phi} &+ 2\tau^2(c + a)/(c - a) - 2v^2(a + b)/(a - b)\\
&= -(\phi + x + \psi)\phi + (-a + b + c)a\\
&\quad + \tfrac{1}{4}\rho^{-1}\sigma^2(b - c)^2 - \tfrac{1}{4}\rho^{-1}\tau^2(c - a)^2\\
&\quad - \tfrac{1}{4}\rho^{-1}v^2(a - b)^2 + \tfrac{1}{2}\Lambda\\
&\quad + \tfrac{1}{2}(\phi x + x\psi + \psi\phi - \sigma^2 - \tau^2 - v^2)\\
&\quad + \tfrac{3}{8}(a^2 + b^2 + c^2 - 2ab - 2bc - 2ca)
\end{aligned} \quad (52)$$

together with 2 cyclic permutations;

$$\begin{aligned}
\dot{\sigma} &+ 2v\tau(bc - a^2)/(a - b)(c - a)\\
&\quad + \sigma(x - \psi)(b + c)/(b - c)\\
&= \tfrac{1}{2}\rho^{-1}v\tau(c - a)(a - b) - (\phi + x + \psi)\sigma
\end{aligned} \quad (53)$$

together with 2 cyclic permutations.

By cyclic permutations we mean simultaneous permutation of the 3 triplets $(abc)$, $(\phi x\psi)$, and $(\sigma\tau v)$. We recall from (34) that

$$\begin{aligned}
4\rho &= 2(\phi x + x\psi + \psi\phi - \sigma^2 - \tau^2 - v^2)\\
&\quad - \tfrac{1}{2}(a^2 + b^2 + c^2) + ab + bc + ca - 2\Lambda.
\end{aligned} \quad (54)$$

As we have seen, there are precisely two first integrals of these equations, which can be found because of the special form of the matter terms. The first of these is a *proper energy integral*,

$$8\rho/(\gamma abc) = 2H, \quad (55)$$

where $\dot{H} = 0$. $\gamma = (1 - V^2)^{-\frac{1}{2}} = (1 - \rho^{-2}t^2)^{-\frac{1}{2}}$, as

before, is the Lorentz factor relating interval along the matter congruence to interval along the reference congruence.

The second integral expresses conservation of angular-momentum density in the proper frame:

$$a(b-c)^2\sigma^2 + b(c-a)^2\tau^2 + c(a-b)^2v^2 = \tfrac{1}{4}K^2H^2a^3b^3c^3, \quad (56)$$

where $\dot{K} = 0$.

We now specialize to the *symmetric case* treated by Gödel[10] and Behr[4]: It is clear from the above that any two off-diagonal terms in **S**, say $\tau$ and $v$, are an IR set, and by setting them simultaneously equal to zero we consistently reduce the set of simultaneous differential equations by two, i.e., $\tau$ and $v$ then *stay* zero. Now **t** and **V** are only in the 1-direction, and we are left with just

$$\dot{\sigma} + (x-\psi)\sigma(b+c)/(b-c) = -(\phi + x + \psi)\sigma, \quad (57)$$

$$\left.\begin{aligned}
\dot{\phi} &= -(\phi + x + \psi)\phi + (-a+b+c)a \\
&\quad + \tfrac{1}{4}\rho^{-1}\sigma^2(b-c)^2 + \tfrac{1}{2}\Lambda \\
&\quad + \tfrac{1}{2}(\phi x + x\psi + \psi\phi - \sigma^2) \\
&\quad + \tfrac{3}{8}(a^2+b^2+c^2) - \tfrac{3}{4}(ab+bc+ca), \\
\dot{x} &= 2\sigma^2(b+c)/(b-c) - (\phi + x + \psi)x \\
&\quad + (a-b+c)b - \tfrac{1}{4}\rho^{-1}\sigma^2(b-c)^2 + \tfrac{1}{2}\Lambda \\
&\quad + \tfrac{1}{2}(\phi x + x\psi + \psi\phi - \sigma^2) \\
&\quad + \tfrac{3}{8}(a^2+b^2+c^2) - \tfrac{3}{4}(ab+bc+ca), \\
\dot{\psi} &= -2\sigma^2(b+c)/(b-c) - (\phi + x + \psi)\psi \\
&\quad + (a+b-c)c - \tfrac{1}{4}\rho^{-1}\sigma^2(b-c)^2 + \tfrac{1}{2}\Lambda \\
&\quad + \tfrac{1}{2}(\phi x + x\psi + \psi\phi - \sigma^2) \\
&\quad + \tfrac{3}{8}(a^2+b^2+c^2) - \tfrac{3}{4}(ab+bc+ca),
\end{aligned}\right\} \quad (58)$$

$$\left.\begin{aligned}
\dot{a} &= (\phi - x - \psi)a, \\
\dot{b} &= (-\phi + x - \psi)b, \\
\dot{c} &= (-\phi - x + \psi)c.
\end{aligned}\right\} \quad (59)$$

Eq. (57) may be omitted, and $\sigma$ eliminated from (58) by virtue of the integral (56), which now reads

$$a(b-c)^2\sigma^2 = \tfrac{1}{4}K^2H^2a^3b^3c^3. \quad (60)$$

We will write (55) again as

$$8\rho/abc - 2\gamma H = 0, \quad (61)$$

or explicitly,

$$4(abc)^{-1}[\phi x + x\psi + \psi\phi - \tfrac{1}{4}K^2H^2a^2b^3c^3(b-c)^{-2} - \tfrac{1}{4}(a^2+b^2+c^2) + \tfrac{1}{2}(ab+bc+ca) - \Lambda] - 2H[1 + K^2bc]^{\frac{1}{2}} = 0. \quad (62)$$

[10] K. Gödel, Proc. Intern. Congr. Math., 1950, Vol. I, pp. 175–181.

Scrutiny of this last expression has enabled us to discover a connection between the present formulation of the symmetric Type VIII and IX cosmologies, and the Lagrangian announced by Gödel.[10] If we introduce variables $p_\alpha$ and $q_\alpha$ ($\alpha = 1, 2, 3$) by setting

$$q_1 = (bc)^{-1}, \quad \text{and cyclic permutation,} \quad (63)$$

$$p_1 = 2q_1^{\frac{1}{2}}(x+\psi)q_1^{-1} \quad \text{and cyclic permutation,} \quad (64)$$

where

$$q = q_1q_2q_3, \quad (65)$$

the expression (62) takes the functional form

$$\begin{aligned}
&\mathcal{H}(q_\alpha, p_\alpha, T, -H) \\
&\quad \equiv 2q^{-\frac{1}{2}}[\tfrac{1}{4}(q_1q_2p_1p_2 + q_2q_3p_2p_3 + q_3q_1p_3p_1) \\
&\qquad - \tfrac{1}{8}(q_1^2p_1^2 + q_2^2p_2^2 + q_3^2p_3^2) \\
&\qquad - \tfrac{1}{2}K^2H^2q_2q_3(q_2-q_3)^{-2} - 2\Lambda q \\
&\qquad - \tfrac{1}{2}(q_1^2 + q_2^2 + q_3^2) + (q_1q_2 + q_2q_3 + q_3q_1)] \\
&\qquad - 2H[1 + K^2q_1^{-1}]^{\frac{1}{2}} = 0,
\end{aligned} \quad (66)$$

and *this is an "energy equation,"* in the terminology of Synge,[11] for a symmetric canonical formulation of our set of first-order equations (58) and (59); that is, we take as four canonically conjugate pairs $\{q_i, p_i\} \equiv \{q_1, p_1; q_2, p_2; q_3, p_3; T, -H\}$; the equations $\partial\mathcal{H}/\partial p_i = \dot{q}_i$ and $\partial\mathcal{H}/\partial q_i = -\dot{p}_i$ ($\cdot$ being a total derivative with regard to an *unspecified* independent variable, say $w$, again in Synge's notation) give first ($i = 1, 2, 3$) exactly (58) and (59) and *also* give, ($i = 4$),

$$2[1 + K^2q_1^{-1}]^{\frac{1}{2}} = \dot{T} = \gamma \quad \text{and} \quad \dot{H} = 0. \quad (67)$$

Thus we see that $w$ is to be interpreted as proper time *along the reference* congruence (cosmological time $t$); $T$ is interpretable as the ignorable coordinate conjugate to the energy constant $H$.

Gödel's Lagrangian is apparently the same as the above symmetric Hamiltonian formulation, if the sign of the "kinetic energy" terms (those quadratic terms involving the $p_i$) is reversed, and the $\dot{q}_i$ used instead of the $p_i$. Such a formulation, however, obscures the interpretation of the $\cdot$ differentiation, the fact that numerically $\mathcal{H} = 0$, and the constancy of $H$ (which indeed is apparently normalized to the value $+1$ by Gödel). We also have included the cosmological constant $\Lambda$.

**ACKNOWLEDGMENT**

The authors wish to acknowledge fruitful conversations with Professor E. Schücking, and to thank him for his encouragement.

[11] J. L. Synge in *Encylopedia of Physics* (Springer-Verlag, Berlin, 1960), Vol. III/1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Alpha-Function Technique for Two-Center Integrals*

R. R. SHARMA†
*Department of Physics, University of California, Riverside, California*

(Received 30 December 1966)

A closed form has been derived for the function $\alpha_l(NML \mid a, r)$ introduced by Löwdin for two-center integrals in molecules and solid states. This expression is general and applied to all values of $l$, $L$, and $M$.

## I. INTRODUCTION

Usually, two- or multiple-center integrals are involved when we deal with the structure and properties of systems of atoms, molecules, and complexes. Expressions for quantities such as electronic energy, molecular-dipole and -quadrupole moments, fine- and hyperfine-coupling constants, transition probabilities, scattering coefficients, nuclear-magnetic shielding constants, spin-Hamiltonian parameters, and various other quantities, require the evaluation of electronic integrals of one- and two-electron operators. These integrals are computed with the help of molecular orbitals formed out of atomic orbitals with the origin at different centers. The methods of evaluation of such integrals are based on either (a) the classical expansion in spherical harmonics, first used by Coolidge[1] and subsequently developed by Landshoff,[2] Löwdin,[3] Barnett and Coulson[4] or (b) the transformation of the integrand into a prolate-spheroidal coordinate system, as developed by Roothan, Ruedenberg, Jaunzemis, Wahl, Cade, and others.[5] The latter method is convenient only when the atomic orbitals involved in the integrals to be in the form of Slater-type orbitals (STO). Using Slater-type orbitals, a number of closed forms for two-center integrals involving one- and two-electron operators have been described by Roothan *et al.*[5] and Geller and Griffith.[6] These methods are not suitable when numerically tabulated Hartree–Fock functions or analytical wavefunctions, which cannot be expressed as sums of exponentials, are used for the computation of two-center integrals. In such cases one has to use Löwdin's α-function method.[7] This method is of type (a) and uses an expansion of the atomic orbitals (say $\Phi_{NLM}$) about one center in terms of functions measured from the other center. Consequently a two-center integral is reduced to a sum of one-center integrals which can be easily evaluated. The expansion of the wavefunction around one center in terms of functions about the other center leads to a sum of products of a radial part [designated as an α function, that is, $\alpha_l(NLM \mid a, r)$, "$a$" being the distance between the two centers] and an angular part $Y_l^M(\theta, \varphi)$, where $l$ runs from zero to infinity and $r$, $\theta$, and $\varphi$ are the coordinates with respect to the new center about which the wavefunction $\Phi_{NLM}$ is expanded. This means that a wavefunction involving a definite angular momentum $L$ about first center is equivalent to a combination of various angular momenta on the other center. The α-function technique of Löwdin has recently been used by Ikenberry and Das[8] for the evaluation of nuclear magnetic shielding constants in alkali halides, by Knox[9] for the calculation of excited-state wavefunctions, excitation energies, and oscillator strengths for argon, and by Smith[10] for the investigation of the $g$ factor of hydrogen and alkali atoms trapped in rare gas solids. The same technique has also been adopted recently by us[11] for the investigation of the overlap contributions to the zero-field splitting parameters $D$ and $E$ occurring in the spin Hamiltonian of paramagnetic ions. It has been found in these cases that the $\alpha_l(NLM \mid a, r)$ functions are required not only for a few of the smaller values of $l$, $L$, and $M$ considered earlier by Löwdin, but also for higher values of $l$, $L$, and $M$. The lack of availability of the algebraic form of $\alpha_l(NLM \mid a, r)$ for higher values of $l$, $L$, and $M$ is

* Supported in part by National Science Foundation and Advanced Research Project Agency.

† Present address: Department of Physics, Purdue University, Lafayette, Indiana.

[1] A. S. Coolidge, Phys. Rev. **42**, 189 (1932).

[2] R. Landshoff, Z. Physik **102**, 201 (1936); Phys. Rev. **52**, 246 (1937).

[3] P.-O. Löwdin, Arkiv Mat. Fys. Astr. **35A**, No. 30 (1948).

[4] M. P. Barnett and C. A. Coulson, Phil. Trans. Roy. Soc. (London) **A243**, 221 (1951).

[5] A. C. Wahl, P. E. Cade, and C. C. J. Roothan, J. Chem. Phys. **41**, 2578 (1964); C. C. J. Roothan, J. Chem. Phys. **19**, 1445 (1951); K. Ruedenberg, J. Chem. Phys. **19**, 1459 (1951); K. Ruedenberg, C. C. J. Roothan, and W. Jaunzemis, J. Chem. Phys. **24**, 201 (1956); C. C. J. Roothan, J. Chem. Phys. **24**, 947 (1956).

[6] Murray Geller and Richard W. Griffith, J. Chem. Phys. **40**, 2309 (1964).

[7] P.-O. Löwdin, Advan. Phys. **5**, 1 (1956).

[8] D. Ikenberry and T. P. Das, J. Chem. Phys. **43**, 2199 (1965); Phys. Rev. **138**, A822 (1965).

[9] R. S. Knox, Phys. Rev. **110**, 375 (1958).

[10] D. Y. Smith, Phys. Rev. **131**, 2056 (1963).

[11] R. R. Sharma, T. P. Das, and R. Orbach, Phys. Rev. **149**, 257 (1966); R. R. Sharma, T. P. Das, and R. Orbach, Phys. Rev. **155**, 338 (1967).

particularly felt in the calculation of properties of molecules and solids involving $d$ and $f$ electrons. The purpose of this article is to present a closed and general expression for the $\alpha$ function which applies to any values of $l$, $L$, and $M$. Some of the standard mathematical relations used in this work have also been listed. A general asymptotic expression for the $\alpha$ function, for small values of $r$ valid for all values of $l$, $L$, and $M$, has also been presented. Finally, we have substituted different values of $l$, $L$, and $M$ in our general expression of the $\alpha$ function, and the results are compared with Löwdin's earlier expressions as a check of the correctness of our formula. It should be remarked that Coulson and Barnett's $\zeta$-function technique is really a special case of $\alpha$ functions, appropriate to the case of hydrogenic wavefunctions—that is, those expressible as a single exponential.

## II. DERIVATION OF A CLOSED FORM FOR THE $\alpha$ FUNCTION

We are interested in deriving explicit forms for the $\alpha_l$ which hold for general values of $l$, $L$, and $M$. Löwdin[7] had derived such expressions for a few of the smaller values of $l$, $L$, and $M$. According to our definition of $\alpha_l(NLM \mid a, r)$, we write

$$\Phi(NLM \mid R, \Theta, \Phi) = \sum_l (1/r)\alpha_l(NLM \mid a, r) Y_l^M(\theta, \varphi), \tag{1}$$

where $\Phi(NLM \mid R, \Theta, \Phi)$ is the wavefunction of an electron on the atom $B$ to be expanded at the center $A$ which is at a distance "$a$" away from the center $B$ (Fig. 1). The polar coordinates $r$, $\theta$, and $\varphi$ are the coordinates of the electron with respect to the center $A$ as origin, while $R$, $\Theta$, and $\Phi$ are the coordinates with respect to $B$ as origin, the axes being disposed in the manner shown in Fig. 1.

According to Löwdin's definition[7] of the $\alpha$ function, which we denote by $\alpha_l^0(NLM \mid a, r)$, we have

$$\Phi(NLM \mid R, \Theta, \Phi) = k_{LM} \sum_{l=0}^{\infty} \alpha_l^0(NLM \mid a, r) P_L^M(\cos\theta) \begin{cases} \cos M\varphi \\ \sin M\varphi \end{cases}, \tag{2a}$$

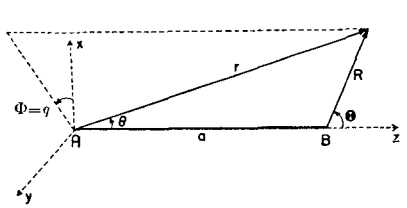



FIG. 1. The systems of axes used for the expansion of a wavefunction centered at $B$ in terms of the $\alpha_l(NLM \mid a, r)$ functions centered at $A$.

where

$$k_{LM} = \left(\epsilon_M \frac{2L+1}{4\pi} \frac{(L-M)!}{(L+M)!}\right)^{\frac{1}{2}}, \tag{2b}$$

$$\epsilon_0 = 1, \tag{2c}$$

$$\epsilon_\nu = 2 \quad (\nu \geq 1), \tag{2d}$$

and Löwdin employs real spherical harmonics instead of the usual complex ones that we have used.

We can relate $\alpha_l$ defined by (1) to Löwdin's definition $\alpha_l^0$ [Eq. (2)] by the relation

$$\alpha_l(NLM \mid a, r) = K_{lLM} r \alpha_l^0(NLM \mid a, r), \tag{3a}$$

where

$$K_{lLM} = \left(\frac{(2L+1)(L-M)!(l+M)!}{(2l+1)(l-M)!(L+M)!}\right)^{\frac{1}{2}}. \tag{3b}$$

The expansion of $\Phi(NLM \mid R, \Theta, \Phi)$ about $A$ yields[7]

$$\alpha_l^0(NLM \mid a, r) = (2\pi/\epsilon_M) k_{lM}^2 \int_0^\pi (f_{NL}(R)/R) P_L^M(\cos\Theta) \times P_l^M(\cos\theta) \sin\theta \, d\theta, \tag{4a}$$

where $f_{NL}(R)$ is $R$ times the radial part of

$$\Phi(NLM \mid R, \Theta, \Phi),$$

that is,

$$\Phi(NLM \mid R, \Theta, \Phi) = \begin{cases} \dfrac{f_{NL}(R)}{R} Y_L^M(\Theta, \Phi) & (4b) \\ k_{LM} \dfrac{f_{NL}(R)}{R} P_L^M(\cos\Theta) \begin{cases} \cos M\Phi \\ \sin M\Phi \end{cases} & (4c) \end{cases}$$

and the coordinates of the electron in one system $(r, \theta, \varphi)$ are related to the other system $(R, \Theta, \Phi)$ by the relations

$$R^2 = a^2 + r^2 - 2ar\cos\theta, \tag{5a}$$

$$\Phi = \varphi, \tag{5b}$$

$$-R\cos\Theta + r\cos\theta = a, \tag{5c}$$

$$R\sin\Theta = r\sin\theta. \tag{5d}$$

Changing the variable of integration in (4a) from $\theta$ to $R$ with the help of the relations (5), we have

$$\alpha_l^0(NLM \mid a, r) = (2\pi/\epsilon_M)(k_{lM}^2/ar) \int_{|a-r|}^{(a+r)} f_{NL}(R) P_L^M(Z_1) P_l^M(Z_2) \, dR, \tag{6a}$$

where

$$Z_1 = -(a^2 + R^2 - r^2)/2aR = \cos\Theta \tag{6b}$$

and

$$Z_2 = (a^2 + r^2 - R^2)/2ar = \cos\theta. \tag{6c}$$

In order to simplify expression (6a) we need certain mathematical relations appropriate for Legendre

polynomials. These are

$$P_l^M(Z_2) = \frac{2^M\pi^{\frac{1}{2}}(Z_2^2-1)^{-M/2}(-1)^{-M/2}}{\Gamma(\frac{1}{2}-l/2-M/2)\Gamma(1+l/2-M/2)} \times F(-l/2-M/2, \tfrac{1}{2}+l/2-M/2; \tfrac{1}{2}, Z_2^2) - \frac{\pi^{\frac{1}{2}}2^{M+1}Z_2(Z_2^2-1)^{-M/2}(-1)^{-M/2}}{\Gamma(\frac{1}{2}+l/2-M/2)\Gamma(-l/2-M/2)} \times F(\tfrac{1}{2}-l/2-M/2, 1+l/2-M/2; \tfrac{3}{2}, Z_2^2), \quad (7a)$$

where $F(a, b; c, z)$ is the hypergeometric function defined as

$$F(a, b; c, z) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n n!} z^n, \quad c \neq 0, -1, -2, \cdots, \quad (7b)$$

$$(a)_0 = 1, \quad (7c)$$

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} = a(a+1)\cdots(a+n-1), \quad n = 1, 2, 3, \cdots. \quad (7d)$$

In relations (7a) and (7d), $\Gamma(x)$ is the standard Gamma function.

$$P_L^M(Z_1) = (1/2^L)(1/L!)(1-Z_1^2)^{M/2}\frac{d^{L+M}}{dZ_1^{L+M}}(Z_1^2-1)^L, \quad (8a)$$

$$\frac{d^{L+M}}{dZ_1^{L+M}}(Z_1^2-1)^L = \frac{(2L)!}{L!}\sum_{r'=0}^{\text{qu.}(L-M)/2}(-1)^{r'}p_{r'}Z_1^{L-M-2r'}, \quad (8b)$$

$$p_{r'} = \frac{(L!)^2(2L-2r')!}{(2L)!(L-r')!r'!(L-2r'-M)!}, \quad (8c)$$

where qu. $(L-M)/2$ stands for "quotient of $(L-M)/2$" (that is, the integral part of $(L-M)/2$; for example, $\frac{3}{2}$ has quotient 1).

Making use of (6b) and (6c), we have

$$(1-Z_1^2)/(1-Z_2^2) = r^2/R^2. \quad (9)$$

Further,

$$Z_1^m = \left(\frac{a^2-r^2}{2aR}\right)^2\sum_{q=0}^{m}\binom{m}{q}\left(\frac{R^2}{a^2-r^2}\right)^q, \quad (10a)$$

$$Z_2^m = \left(\frac{a^2+r^2}{2ar}\right)^m\sum_{q=0}^{m}(-1)^q\binom{m}{q}\left(\frac{R^2}{a^2+r^2}\right)^q, \quad (10b)$$

where

$$\binom{m}{q} = \frac{m!}{q!(m-q)!} \quad (10c)$$

and

$$P_L^M(-x) = (-1)^{L+M}P_L^M(x). \quad (10d)$$

Now if we define

$$A_{Ll}^M(R) = P_L^M(Z_1)P_l^M(Z_2), \quad (11a)$$

then Eq. (6a) becomes

$$\alpha_l^0(NLM \mid a, r) = (2\pi/\epsilon_M)(k_{LM}^2/ar)\int_{|a-r|}^{(a+r)} f_{NLM}(R)A_{Ll}^M(R)\,dR. \quad (11b)$$

Making use of the expansion (7a) for $P_l^M(Z_2)$, the expression (8) for $P_L^M(Z_1)$ and the relations (9) and (10) one can put the expression (11a) for $A_{Ll}^M(R)$ into the form

$$A_{Ll}^M = \frac{(-1)^L}{2^{L-M}}(r/a)^M \times \sum_{\substack{r'=\text{qu.}(L-M)/2 \\ q=L-2r'-M \\ q'=2n+1}}^{n=\infty}\sum_{n,r',q,q'=0} Cg(r)\left[B_1 - B_2\frac{(1+r^2/a^2)}{(2r/a)}\right] \times (R/a)^{-L+2r'+2q+2q'}, \quad (12a)$$

where

$$C = \frac{(-1)^{r'+q'}(2n)!(2L-2r')!}{n!q'!(2n+1-q')!q!(L-2r'-M-q)!(L-r')!r'!}, \quad (12b)$$

$$g(r) = \frac{(1-r^2/a^2)^{L-2r'-M-q}}{2^{L-2r'-M}}\frac{(1+r^2/a^2)^{2n-q'}}{(2r/a)^{2n}}, \quad (12c)$$

$$B_1 = \frac{\pi^{\frac{1}{2}}(-l/2-M/2)_n(\frac{1}{2}+l/2-M/2)_n(2n-q'+1)}{\Gamma(\frac{1}{2}-l/2-M/2)\Gamma(1+l/2-M/2)(1/2)_n}, \quad (12d)$$

$$B_2 = \frac{2\pi^{\frac{1}{2}}(\frac{1}{2}-l/2-M/2)_n(1+l/2-M/2)_n(2n+1)}{\Gamma(\frac{1}{2}+l/2-M/2)\Gamma(-l/2-M/2)}. \quad (12e)$$

The lower limits of $n, r', q$, and $q'$ in the summations in (12a) are zero, but the higher limits for $n, r', q$, and $q'$ are $\infty$, quotient of $(L-M)/2$, $L-2r'-M$, and $2n+1$, respectively. Equation (12a) gives only a finite number of terms because the factorials, Gamma functions, and $(a)_n$ make the terms vanish after a certain number of terms.

We can now simplify $B_1$ and $B_2$ as defined in (12d) and (12e) with the help of standard relations:

$$\Gamma(z)\Gamma(1-z) = \pi/\sin\pi z, \quad (13a)$$

$$2^{2z-1}\Gamma(z)\Gamma(z+\tfrac{1}{2}) = \Gamma(\tfrac{1}{2})\Gamma(2z). \quad (13b)$$

Thus

$$B_1 = \frac{\sin\{\frac{1}{2}\pi(1-l-M)\}(l+M-1)!}{\frac{1}{2}(l-M)!\left[\frac{1}{2}(l+M)-1\right]!2^{l+M-1}} \frac{(-l/2-M/2)_n(\frac{1}{2}+l/2-M/2)_n(2n-q'+1)}{(\frac{1}{2})_n}, \tag{14a}$$

$$B_2 = \frac{\sin\{\frac{1}{2}\pi(-l+M)\}}{\frac{1}{2}(l-M-1)!2^{l+M-1}} \frac{(l+M)!}{\frac{1}{2}(l+M-1)!} \frac{(\frac{1}{2}-l/2-M/2)_n(1+l/2-M/2)_n}{(\frac{3}{2})_n} \cdot (2n+1). \tag{14b}$$

It is evident from (14a) and (14b) that

$$B_1 = 0 \quad \text{if} \quad l+M = \text{odd integer}, \tag{15a}$$

$$B_2 = 0 \quad \text{if} \quad l+M = \text{even integer}. \tag{15b}$$

Next, we substitute (12a) for $A_{Ll}^{M}$ in (11b) with $C$, $g(r)$, $B_1$, and $B_2$ as defined by (12b), (12c), (14a), and (14b), and we define the new running index $s$ as

$$s = r' + q + q'. \tag{16}$$

Finally, we obtain

$$\alpha_l^0(NLM \mid a, r) = E \sum_{s=0}^{l+L} D_{Ll}^{M,s} \int_{|a-r|}^{(a+r)} f_{NLM}(R)(R)^{-L+2s}\, dr, \tag{17a}$$

where

$$E = \begin{cases} \dfrac{(2l+1)}{a^{2-L}} \dfrac{(-1)^L}{2^{(2L+l-M)}} \dfrac{\cos\{[(l+M)/2]\pi\}(l+M-1)!}{(l-M/2)![(l+M)/2-1]!} \left(\dfrac{r}{a}\right)^{M-1} \dfrac{(l-M)!}{(l+M)!} & \text{if } l+M = \text{even integer}, \quad (17b) \\ \dfrac{2l+1}{a^{2-L}} \dfrac{(-1)^L}{2^{(2L+l+1-M)}} \dfrac{\sin\{[(l+M)/2]\pi\}(l-M)!}{[(l-M-1)/2]!\,[(l+M-1)/2]!} \left(\dfrac{r}{a}\right)^{M-2} \left(1+\dfrac{r^2}{a^2}\right) & \text{if } l+M = \text{odd integer} \quad (17c) \end{cases}$$

and

$$D_{L,l}^{M,s} = (1/a^{2s}) \sum_{n,r',q'=0}^{n_0,r_0',q_0'} CC_1 (1-r^2/a^2)^{L-r'-M-s+q'} \frac{(1+r^2/a^2)^{2n-q'}}{(r/a)^{2n}} \tag{17d}$$

with

$n_0$ = quotient of $(l+M)/2$,

$r_0'$ = quotient of $(L-M)/2$,

$q_0'$ = $(2n)$ or $(2n+1)$ according as $(l+M)$ is even or odd,

$$C = \frac{(-1)^{r'+q'}(2n)!(2L-2r')!}{n!q'!(2n+1-q')!(s-q'-r')!(L-r'-M-s+q')!(L-r')!r'!2^{2(n-r')}}, \tag{17e}$$

and

$$C_1 = \begin{cases} (-l/2-M/2)_n(\frac{1}{2}+l/2-M/2)_n(2n-q'+1)/(\frac{1}{2})_n & \text{if } l+M = \text{even integer}, \quad (17f) \\ (\frac{1}{2}-l/2-M/2)_n(1+l/2-M/2)_n(2n+1)/(\frac{3}{2})_n & \text{if } l+M = \text{odd integer}. \quad (17g) \end{cases}$$

The expression (17a) is a general expression and can, therefore, be used for any values of $l$, $L$, and $M$. Knowing $\alpha_l^0$ from (17a) we can determine $\alpha_l$ from (3a), that is,

$$\alpha_l(NLM \mid a, r) = K_{lLM} r \alpha_l^0(NLM \mid a, r), \tag{18}$$

where $K_{lLM}$ is already defined in (3b).

In order to check our expressions (17) we compare the expressions for a few $\alpha$ functions from the expressions (17) with those obtained by Löwdin.

### A. Derivation of $\alpha_1^0(N10 \mid a, r)$

In this case,

$$l = 1, \quad L = 1, \quad M = 0. \tag{19}$$

Since $l+M$ is odd, we use (17c) for $E$ and (17g) for $C_1$. Substituting (19) in (17c), (17d), (17e), and (17g) for $E$, $D_{L,l}^{M,s}$, $C$, and $C_1$, we obtain

$$E = \frac{3}{16a^2} \frac{a^2}{r^2}\left(1+\frac{r^2}{a^2}\right), \tag{20a}$$

$$D_{1,2}^{1,s} = \sum_{n,r'q'=0}^{n_0=0,r_0'=0,q_0'=1} CC_1 \left(1-\frac{r^2}{a^2}\right)^{1-s+q'} \left(1+\frac{r^2}{a^2}\right)^{-q'}, \tag{20b}$$

$$C = \frac{2(-1)^{q'}}{(1-q')!q'!(s-q')!(1-s+q')!}, \tag{20c}$$

$$C_1 = 1. \tag{20d}$$

Hence, from (17a) and (20), we have

$$\alpha_1^0(N10 \mid a, r) = -\frac{3}{8r^2}\left[\left(1-\frac{r^4}{a^4}\right)\int f_{N10}(R)\left(\frac{R}{a}\right)^{-1} dR + \frac{2r^2}{a^2}\int f_{N10}(R)\left(\frac{R}{a}\right) dR - \int f_{N10}(R)\left(\frac{R}{a}\right)^3 dR\right]. \quad (21)$$

The expression (21) agrees with the corresponding result obtained by Löwdin.

Similarly, by substituting $l = 1, L = 0$, and $M = 0$, in (17), we have

$$\alpha_1^0(N00 \mid a, r) = \frac{3}{4r^2}\left[\left(1+\frac{r^2}{a^2}\right)\int f_{N00}(R)\, dR - \int f_{N00}(R)\left(\frac{R}{a}\right)^2 dR\right]. \quad (22)$$

Also, for $l = 0$, $L = 0$, and $M = 0$, we have

$$\alpha_0^0(N00 \mid a, r) = \frac{1}{2ar}\int_{|a-r|}^{(a+r)} f_{N00}(R)\, dR. \quad (23)$$

For $L = 1$, $l = 1$, and $M = 1$ we get

$$\alpha_1^0(N11 \mid a, r) = -\frac{3}{8r^2}\left[\left(\frac{1}{2}-\frac{r^2}{a^2}+\frac{1}{2}\frac{r^4}{a^4}\right)\int f(R)\frac{a}{R}\, dR - \left(1+\frac{r^2}{a^2}\right)\int f(R)\frac{R}{a}\, dR + \frac{1}{2}\int f(R)\left(\frac{R}{a}\right)^3 dR\right]. \quad (24)$$

Expressions (22), (23), and (24) also agree with Löwdin's expressions.

### B. Asymptotic Form of α Functions for Small r

For numerical evaluation of the α function [Eq. (17a)] we have to compute the integral occurring in Eq. (17a), which we denote by I:

$$\mathrm{I} = \int_{|a-r|}^{(a+r)} f_{NLM}(R) R^{-L+2s}\, dR. \quad (25)$$

For small $r$ we can write

$$f_{NLM}(a \pm r) = f_{NLM}(a) \pm r f'_{NLM}(a) \cdots. \quad (26)$$

Retaining only the first term in (26) and substituting in (25), after integration we have

$$\mathrm{I} = \begin{cases} f_{NLM}(a)\left\{\dfrac{(a+r)^{-L+2s+1} - (|a-r|)^{-L+2s+1}}{-L+2s+1}\right\} & \text{if } L-2s \neq 1, \quad (27a) \\ f_{NLM}(a)\ln\left\{\dfrac{a+r}{|a-r|}\right\} & \text{if } L-2s = 1. \quad (27b) \end{cases}$$

For small values of $r$, since the integration (25) occurring in (17a) cannot be performed very accurately, we therefore use the relations (27a) or (27b), accordingly as $L - 2s \neq 1$ or $L - 2s = 1$.

### C. Value of the α Function for r = 0

For $r \to 0$ Eq. (5a) yields

$$R \xrightarrow[r=0]{} a, \quad (28)$$

and (6b) gives, for $r = 0$,

$$z_1 = -1 = \cos\Theta. \quad (29)$$

Making use of

$$P_L^M(-1) = (-1)^{L+M}$$

and

$$\int P_l^M(\cos\theta)\sin\theta\, d\theta = 2\delta_{l,0}\delta_{M,0},$$

from (4a), (2b), and (2c) for $r = 0$, we finally obtain

$$\alpha_l^0(NLM \mid a, 0) = \frac{f_{NLM}(a)}{a}(-1)^L\delta_{l,0}\delta_{M,0}. \quad (30)$$

The expression (30) has also been quoted by Löwdin, and acts as a good numerical check in actual computations.

In summary, we have obtained a general expression for $\alpha_l(NLM \mid a, r)$ in Eq. (17) which permits the evaluation of two-center integrals involving wavefunctions of any $L$. This expression includes as special cases those derived by Löwdin for smaller special values of $l$, $L$, and $M$. It is hoped that the availability of this general expression for α will make it convenient to use numerical Hartree–Fock atomic wavefunctions in molecular and solid-state problems, as has been the case more recently with analytic functions.

### ACKNOWLEDGMENT

The author is highly grateful to Professor T. P. Das for useful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Renormalization and Composite Effects in the Lee Model

MICHAEL M. BROIDO*
*Rutherford College, University of Kent, Canterbury, Kent, England*

(Received 22 May 1967)

We consider the version of the Lee model with relativistic kinematics. The mass renormalization of the $V$ particle, described in purely field-theoretic terms, is a nonlocal effect. We discuss the composite limit of the model. The natural choice of composite field is nonlocal in the elementary constituents. In the composite limit, $Z_1$ does not vanish. The Hilbert-space formalism of the composite theory is not equivalent to that of an $N$–$\theta$ theory with a four-particle interaction. All these results are cutoff independent.

## 1. INTRODUCTION

In a recent analysis,[1] Ynduráin pointed out that when the Lee model is analyzed with relativistic kinematics, the point-coupling limit of the model is better behaved than when nonrelativistic kinematics are used.[2,3] The charge renormalization remains finite, and no ghost states appear. These features make the model particularly suitable for the investigation of various effects of the renormalization program.

In the present paper, we wish to discuss the following questions. Is the renormalization a local process? What happens to the theory in the limit $Z_V \to 0$ (composite $V$ particle)?

To make these questions more precise, we introduce appropriate linguistic conventions. By the "Lee model," we will understand the Lee model with relativistic kinematics, as discussed in Ref. 1. By the "point limit" of the Lee model, we understand that calculations in the Lee model are to be performed with a cutoff, which will be removed in the end.

We discuss the $N$–$\theta$ sector of the model only. We show that an operator may be introduced whose Fourier components create physical one $V$-particle states (eigenstates of the local Hamiltonian) and that this operator is formed in a nonlocal fashion from the bare $V$, $N$, $\theta$ "fields" (Fourier transforms of the creation operators), even in the point limit of the model. We christen this object the totally renormalized $V$ field.

We study the composite limit of the model by letting $Z_V \to 0$, as has been done in the conventional form of the Lee model by various authors.[4,5] We show that the nonrelativistic nature of the model causes difficulties in the definition of a vertex-function renormalization constant. However we get round these, $Z_1$ does not tend to zero with $Z_3$, contrary to what is observed in the conventional model. We discuss the Green's function equations for the composite $V$, and note that the totally renormalized $V$ field gives a particularly simple form. We consider the possibility that the composite $V$ model is equivalent to a model without a $V$, with the $N$ and $\theta$ having a four-field interaction. Although the $S$ matrices are equal, as follows from the general Green's function theory of composite particles,[6] the two theories cannot be completely equivalent in the details of their Hilbert space formalism. This arises from a combination of renormalization considerations with the relativistic kinematics, and contrasts with the situation observed in simpler models.[7]

We discuss the effective coupling constant of the composite model, illustrating ideas of Ref. 6. A suggestion about the composite limit of a model with two $V$ particles[5] is shown to be specious. We point out that in the present model, $Z_3 = 0$ does not arise from "kinematic" considerations, as has recently been suggested.

## 2. THE LEE MODEL

The free Hamiltonian is

$$H_0 = \int E_V(\mathbf{p}) V^*(\mathbf{p}) V(\mathbf{p})\, d\mathbf{p} + \int E_N(\mathbf{p}) N^*(\mathbf{p}) N(\mathbf{p})\, d\mathbf{p} + \int \omega(\mathbf{p}) a^*(\mathbf{p}) a(\mathbf{p})\, d\mathbf{p}. \quad (1)$$

Here, $V$, $N$, $a$ are annihilation operators for the fermions $V$ and $N$ and the boson $\theta$. $E_V$, $E_N$, $\omega$ are the corresponding energy functions, which in the relativistic model have the form

$$E_V(\mathbf{p}) = (M^2 + \mathbf{p}^2)^{\frac{1}{2}},$$
$$E_N(\mathbf{p}) = (m^2 + \mathbf{p}^2)^{\frac{1}{2}}, \quad (2)$$
$$\omega(\mathbf{p}) = (\mu^2 + \mathbf{p}^2)^{\frac{1}{2}},$$

* Present address: Magdalen College, Oxford, England.

[1] F. J. Ynduráin, J. Math. Phys. 7, 1133 (1966).

[2] T. D. Lee, Phys. Rev. 95, 1329 (1954).

[3] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row Publ., Inc., New York, 1961).

[4] J. C. Houard, Ann. Inst. Henri Poincaré 2, 105 (1965), and references quoted there.

[5] P. K. Srivastava and S. R. Choudhoury, Nuovo Cimento 39, 650 (1965).

[6] M. M. Broido and J. G. Taylor, Phys. Rev. 147, 993 (1966).

[7] K. Sekine, Nucl. Phys. 76, 513 (1965).

where $M$, $m$, $\mu$ are the masses of $V$, $N$, $\theta$ particles, respectively. The interaction Hamiltonian is

$$H_I = \int d\mathbf{p}\,\delta E_V(\mathbf{p}) V^*(\mathbf{p}) V(\mathbf{p}) + g\int d\mathbf{p}\,d\mathbf{k}[h(\mathbf{p},\mathbf{k})V^*(\mathbf{p})N(\mathbf{p}-\mathbf{k})a(\mathbf{k}) + \text{H.c.}], \quad (3)$$

where $\delta E_V(p)$ is the energy renormalization of the $V$ particle:

$$\delta E_V(\mathbf{p}) = g^2\int d\mathbf{k}\,\frac{|h(\mathbf{p},\mathbf{k})|^2}{E_N(\mathbf{p}-\mathbf{k}) + \omega(\mathbf{k}) - E_V(\mathbf{p})}, \quad (4)$$

$$h(\mathbf{p},\mathbf{k}) = f(\mathbf{p},\mathbf{k})[E_V(\mathbf{p})E_N(\mathbf{p}-\mathbf{k})\omega(\mathbf{k})]^{-\frac{1}{2}}; \quad (5)$$

$f(p, k)$ is a cutoff function taking the value unity in the point-coupling limit. We retain it, following the conventions set up in the Introduction.

The tractability of the model arises largely because the $N$ and $\theta$ particles do not need renormalization. Then if we demote the *physical* one-particle $V$ state by $|V(\mathbf{p})\rangle_d$, it is given by

$$|V(\mathbf{p})\rangle_d = Z_V^{\frac{1}{2}}(p)\Big\{V^*(\mathbf{p})\rangle_0 + g\int d\mathbf{k}\Phi(\mathbf{p},\mathbf{k})N^*(\mathbf{p}-\mathbf{k})a^*(\mathbf{k})\rangle_0\Big\}, \quad (6)$$

where the state function $\Phi(\mathbf{p}, \mathbf{k})$ is, in the relativistic case,

$$\Phi(\mathbf{p},\mathbf{k}) = \frac{1}{(2\pi)^{\frac{3}{2}}}\frac{h(\mathbf{p},\mathbf{k})}{E_V(\mathbf{p}) - E_N(\mathbf{p}-\mathbf{k}) - \omega(\mathbf{k})}. \quad (7)$$

Equation (6) is the solution of the one-particle problem and tells us just what we want to know, namely that if we regard the one-particle physical $V$ state as created by a creation operator $W^*(\mathbf{p})$, the Fourier transforms

$$W(\mathbf{x}) = (2\pi)^{-\frac{3}{2}}\int d\mathbf{p}e^{i\mathbf{p}\mathbf{x}}W(\mathbf{p})E_V^{\frac{1}{2}}(\mathbf{p}),$$

$$V(\mathbf{x}) = (2\pi)^{-\frac{3}{2}}\int d\mathbf{p}e^{i\mathbf{p}\mathbf{x}}V(\mathbf{p})E_V^{\frac{1}{2}}(\mathbf{p})Z_V^{\frac{1}{2}}(\mathbf{p}),$$

$$N(\mathbf{x}) = (2\pi)^{-\frac{3}{2}}\int d\mathbf{p}e^{i\mathbf{p}\mathbf{x}}N(\mathbf{p})E_N^{\frac{1}{2}}(\mathbf{p}),$$

and so forth, are connected by the relation

$$W(\mathbf{x}) = V(\mathbf{x}) + \int N(\mathbf{x}-\mathbf{y})\theta(\mathbf{x}-\mathbf{z})F(\mathbf{y},\mathbf{z})\,d\mathbf{y}\,d\mathbf{z}, \quad (8)$$

where

$$F(\mathbf{y},\mathbf{z}) = \int F(\mathbf{p},\mathbf{k})\exp\{i(\mathbf{p}\mathbf{y}+\mathbf{k}\mathbf{z})\}\,d\mathbf{p}\,d\mathbf{k} \quad (9)$$

and

$$F(\mathbf{p},\mathbf{k}) = \frac{gf(\mathbf{p},\mathbf{k})Z_V^{\frac{1}{2}}(\mathbf{p})}{E_V(\mathbf{p}) - E_N(\mathbf{p}-\mathbf{k}) - \omega(\mathbf{k})}. \quad (10)$$

Then $W(\mathbf{x})$, as given by Eq. (8), is our totally renormalized $V$ field. $N(\mathbf{x})$ and $\theta(\mathbf{x})$ are of course already totally renormalized. Note that $F(\mathbf{p}, \mathbf{k})$ is in fact *independent* of the coupling constant $g$.

Let us consider the properties of $W(\mathbf{x})$. Most strikingly, it is *not point coupled*, even in the point limit of the theory ($f = 1$). Point coupling of $W$ would require $F(\mathbf{p}, \mathbf{k})$ to be a function of $(\mathbf{p} - \mathbf{k})$. (There is also a nonlocality in the $V$ term due to the $p$ dependence of $Z_V$, a dependence whose effects we will discuss in detail below, but this particular nonlocality is less important.) We see from (10) that this condition on $F(\mathbf{p}, \mathbf{k})$ cannot be achieved by any sensible choice of $f(\mathbf{p}, \mathbf{k})$. For instance, if we choose $f(\mathbf{p}, \mathbf{k})$ so that $F(\mathbf{p}, \mathbf{k}) = 1$, the integrals involved in the $(N$–$\theta)$-sector $S$ matrix, which are convergent even for $f = 1$, go badly divergent. Thus *the totally renormalized field cannot be point coupled if the original interaction Hamiltonian is.*

As we can see from the spectral representation for $Z_V$, it is of the form

$$Z_V(\mathbf{p}) = \int |Y(\mathbf{p},\mathbf{k})|^2\,d\mathbf{k},$$

where, if the masses are neglected, $Y$ is a function of $(\mathbf{p} - \mathbf{k})$ only. Hence in this approximation $Z_V(\mathbf{p})$ is constant. Provided that $M < \mu$, the neglect of the masses then causes $F(\mathbf{p}, \mathbf{k})$ to become a continuous function of $(\mathbf{p} - \mathbf{k})$ in the point limit, and a direct manipulation yields the approximate, local formula

$$W(\mathbf{x}) = V(\mathbf{x}) + g\nabla N(\mathbf{x})\cdot\nabla\theta(\mathbf{x}).$$

We do not know what to make of this odd relation.

## 3. RENORMALIZATION CONSTANTS

Only the $V$ particle requires renormalization. The mass renormalization is logarithmically divergent in the point limit, as we see from Eq. (4) (compared with the linear divergence in the conventional version). We note that $\delta M$ is proportional to $g^2$.

The "wavefunction renormalization constant" $Z_V$ appearing in Eq. (6) is now no longer a constant as in the static case, but is a function $Z_V(\mathbf{p})$ of the $V$ three-momentum. This need not cause any trouble provided we carefully consider the relation between this $Z_V(\mathbf{p})$ and the wavefunction renormalization constant, say $Z_3$, of a fully relativistically covariant quantum field theory. (After all, the main interest of the Lee model is as a "model" of just this class of theories.) Such a $Z_3$ is certainly a Lorentz scalar (and may be computed[6] in terms of Green's functions). If our model were covariant, the $Z_V(\mathbf{p})$ of Eq. (6) would be a function of the appropriate four-momentum squared,

and $Z_3$ would be related to $Z_V(M^2)$ and would renormalize the complete propagator of the theory. But the model is quite fundamentally noncovariant, independently of the relativistic choice of energies (or of any other relevant form factors). With this proviso, the function $Z_V(\mathbf{p})$ will play the role of the wavefunction renormalization constant, particularly in discussing the possibility of a composite $V$ below.

Certainly $Z_V(\mathbf{p})$ is well defined and finite in the point limit and is proportional to $g^{-2}$ (whereas in the static model it is logarithmically divergent).

There is not general agreement in the literature about a vertex-function renormalization constant. We feel that it is most natural, following Ref. 1, to define the renormalized coupling constant $g_r$ as the residue in the pole at $M^2$ in the $N$–$\theta$ scattering amplitude $A(s)$. In the point limit this gives[1]

$$g_r^2 = \frac{g^2}{8M} \times \left\{1 + \frac{g^2}{2M}\int_0^\infty \frac{\mathbf{p}^2\,d\mathbf{p}}{E_N(\mathbf{p})\omega(\mathbf{p})[E_N(\mathbf{p}) - \omega(\mathbf{p}) - M]^2}\right\}^{-1}. \tag{11}$$

If we could define a vertex-function renormalization constant $Z_1$ by

$$g_r^2 = Z_1^{-1} Z_V g^2 \tag{12}$$

(in the conventional fashion), clearly $Z_1$ would be a finite function of $g$, tending as $g \to \infty$ to a finite, nonzero limit. Actually this analogy is not very useful because of the $\mathbf{p}$ dependence of $Z_V$, but it does show us that in no sense at all can $Z_1 = 0$ arise through the limit $g \to \infty$, and suggests strongly that this will also be the case in a full relativistic theory. We will refer to this in our remarks on the composite limit of the model below.

It is interesting to observe that for $M = 0$, the renormalized coupling constant given by (11) is finite and is independent of the bare one. This latter property has been conjectured for quantum electrodynamics (the photon corresponding to the massless $V$), on the basis of perturbation theory.[8]

## 4. THE COMPOSITE $V$ LIMIT

In accordance with the general theory[6] we let $V$ become composite by letting $Z_V(\mathbf{p}) \to 0$. In this process, we must regard $Z_V(\mathbf{p})$ as dependent on appropriate further variables, and discuss how to vary these. Such variables are the renormalized quantities $g_r$ and $M$. (We could also include the $N$ and $\theta$ masses $m$ and $\mu$, but we will not be interested in the possible variation of the results with these masses.) Equation (11) is a locally 1–1 relation between $g_r$ and $g$, and thus we can meaningfully talk about letting $Z_V(\mathbf{p}) \to 0$ by letting $g \to \infty$ (as it will, by the previous section, uniformly on compact sets). Then $g_r$ is given by

$$(g_r^{\max})^2 = \frac{1}{4\pi}\left\{\int_0^\infty \frac{\mathbf{p}^2\,d\mathbf{p}}{E_N(\mathbf{p})\omega(\mathbf{p})[E_N(\mathbf{p}) - \omega(\mathbf{p}) - M]^2}\right\}^{-1} \tag{13}$$

so that we can vary the renormalized mass of the composite provided we also vary the renormalized coupling constant by Eq. (13). We notice that the connection expressed by Eq. (13) is locally 1–1, as is required for a correct particle interpretation. This condition fails for the photon in quantum electrodynamics[9] and is responsible for the failure of attempts to describe quantum electrodynamics as a composite theory.[9]

We see that the dressed 1-particle $V$ state, given by Eq. (6), remains well defined. The first term in curly brackets tends to zero, but the second does not because $Z_V^{\frac{1}{2}} g$ remains finite. Similarly, the totally renormalized $V$ field, Eq. (8), remains well defined and just as non-point-coupled as ever.

Thus we see that although we could perfectly have defined the composite $V$ field by the local choice

$$W^{\text{local}}(x) = N(x)\theta(x), \tag{14}$$

we immediately obtain a nonlocality when the mass renormalization is performed. This is characteristically field-theoretic effect; the $V$ is off its mass shell in $N$–$\theta$ scattering by an amount given by Eq. (7).

It is generally considered that the deuteron cannot be a composite in the sense of field theory because it has internal structure. On the other hand, it is known[10] that the success of the effective-range approximation can be explained in terms of $Z = 0$ theory. We see that there is nothing contradictory about all this; the internal structure is reintroduced by the renormalization of the local composite field.

We can perform a Green's function analysis of the local composite field $W^{\text{local}}$ according to the methods given in detail in Ref. 6. The vertex-function equation will be a Bethe–Salpeter equation involving *mass-renormalized* Green's functions of $W^{\text{local}}$, and we see at once that these will be precisely the Green's functions of $W$ (which will of course need no mass renormalization). In this sense, the totally renormalized field is forced on us by the renormalization program.

Unfortunately we cannot make any interesting

[8] M. Gell-Mann and F. E. Low, Phys. Rev. **95**, 1300 (1954).

[9] M. M. Broido, Phys. Rev. **157**, 1444 (1967).

[10] S. Weinberg, Phys. Rev. **137**, B672 (1965).

assertion of this sort about charge renormalization, because the renormalization of the $VN\theta$ vertex is entirely due to bubbles, i.e., to what are generally called mass-renormalization effects. Thus a detailed discussion of charge-renormalization effects and locality will have to await the introduction of models in which the charge renormalization is less trivial (from a graphical point of view).

## 5. IS THE COMPOSITE MODEL EQUIVALENT TO A MODEL WITHOUT A $V$?

We wish to investigate the possibility of the complete equivalence of the Lee model with composite $V$ and a model with Hamiltonian

$$K = K_0 + K_I,$$

$$K_0 = \int E_N(\mathbf{p})N^*(\mathbf{p})N(\mathbf{p})\,d\mathbf{p} + \int \omega(\mathbf{p})a^*(\mathbf{p})a(\mathbf{p})\,d\mathbf{p}, \tag{15}$$

$$K_I = \lambda' \int a^*(\mathbf{k})N^*(\mathbf{p}-\mathbf{k})h(\mathbf{p},\mathbf{k})h(\mathbf{p},\mathbf{k}') \times N(\mathbf{p}-\mathbf{k}')a(\mathbf{k}')\,d\mathbf{k}\,d\mathbf{k}'\,d\mathbf{p}. \tag{16}$$

By complete equivalence we mean, of course, equivalence of the entire Hilbert-space formalism and not merely of the $S$ matrices. The latter is already guaranteed by the work of Ref. 6 (at least for the point limit).

We will see that the appearance of the function $h(p, k)$ in (16) is quite a general necessary condition for the equivalence. Namely, whatever the exact form of the interaction Hamiltonian $K_I$, the requirement that the dressed one-particle $V$ state, given in terms of $N$ and $\theta$ by the composite limit of Eq. (6), i.e., by

$$|V(p)\rangle_d = \left\{\lim_{g\to\infty} Z_V^{\frac{1}{2}} g\right\} \int d\mathbf{k}\Phi(\mathbf{p},\mathbf{k})N^*(\mathbf{p}-\mathbf{k})a^*(\mathbf{k})\rangle_0, \tag{17}$$

should be an eigenstate of $K$ with eigenvalue $E_V(p)$, reads

$$\begin{aligned} K_I|V(\mathbf{p})\rangle_d &= \left\{\lim_{g\to\infty} Z_V^{\frac{1}{2}} g\right\} \int [E_V^{\frac{1}{2}}(\mathbf{p}) - E_N(\mathbf{p}-\mathbf{k}) - \omega(\mathbf{k})] \\ &\quad \times \Phi(\mathbf{p},\mathbf{k})N^*(\mathbf{p},\mathbf{k})a^*(\mathbf{k})\rangle_0\,d\mathbf{k} \\ &= \left\{\lim_{g\to\infty} Z_V^{\frac{1}{2}} g\right\} \int h(\mathbf{p},\mathbf{k})N^*(\mathbf{p}-\mathbf{k})a^*(\mathbf{k})\rangle_0\,d\mathbf{k}. \end{aligned} \tag{18}$$

This equation shows already why we must have $h(p, k)$ in $K_I$ as given by Eq. (16). Then use of (16) yields

$$K_I|V(\mathbf{p})\rangle_d = -\lambda' \int Q(\mathbf{p},\mathbf{k})N^*(\mathbf{p}-\mathbf{k})a^*(\mathbf{k})\rangle_0\,d\mathbf{k}, \tag{19}$$

where

$$Q(\mathbf{p},\mathbf{k}) = h(\mathbf{p},\mathbf{k}) \int h(\mathbf{p},\mathbf{k}')\Phi(\mathbf{p},\mathbf{k}')\,d\mathbf{k}'. \tag{20}$$

Clearly the expressions (18) and (19) cannot be equal unless the integral in (20) is a constant. Since it is not, the equivalence is impossible. This was what we set out to prove.

A less restrictive form of equivalence would be to require only that the rest state of the composite should have energy eigenvalue $M$. Then the equality of (18) and (19) has to hold only for $p = 0$. The integral in (20) is precisely the mass-shift integral, so that we obtain the condition

$$\lambda' = \lim_{g\to\infty} -\frac{gZ_V^{-\frac{1}{2}}(0)}{\delta M}. \tag{21}$$

This condition was derived in Ref. 6 as a general equivalence condition (for the two systems of Green's functions) in any Green's function system. It also arises in the nonlocal Hamiltonian model of Ref. 7 where the *entire* formal apparatus *is* equivalent to the $g \to \infty$ limit of a simple elementary model. Thus we see that although the equivalence of the Hilbert-space formalisms of elementary and composite models holds only in quite special cases, the condition (21) on the coupling constants is an extremely general condition.

## 6. MISCELLANEOUS REMARKS

Dispersion relations for the Lee model were written down in Ref. 1. They are very similar to those of the Zacharaisen model. We have pointed out that $Z_1 \to 0$ does not occur in the composite limit of the Lee model. This suggests that where it occurs in the Zacharaisen model,[11] it is due to the approximations used (elastic unitarity and perturbation theory).

It has been asserted[5] that in the Lee model with two composite $V$ particles, only one $V$ becomes composite when both $Z_V$ tend to zero. This happened because the authors of Ref. 5 ignored the solution of their equations corresponding to our condition (21). (At the bottom of page 653 of their paper, let $\alpha_1 \to 0$, $g_2 \to 0$, with $g_2^2/\alpha_1$ finite in the limit.) This point will be discussed in detail in a forthcoming publication.[12]

Finally, it has been suggested[13] that the condition $Z_3 = 0$ is a "necessary condition for composite particles" arising in a purely kinematic fashion. This is clearly not the case in the model we consider here. Indeed, when we consider the criterion, Eq. (3) of

[11] See, for instance, N. G. Deshpande and S. A. Bludman, Phys. Rev. **143**, 1239 (1966), and references quoted there.

[12] M. M. Broido and J. G. Taylor, Phys. Rev. **161**, 1301 (1967).

[13] H. M. Fried and Y. S. Jin, Phys. Rev. Letters **17**, 1152 (1966).

Ref. 13, in the composite Lee model, we find $Z^{-1} = 0$, contradicting what we know to be the case (namely $Z = 0$). We believe that this unfortunate conclusion is due to these authors' use of perturbation theory. The failure of perturbation theory in composite models is hardly surprising when one considers that in all known soluble models, the composite limit corresponds to $g \to \infty$.

ACKNOWLEDGMENT

My thanks are due to Dr. J. G. Taylor for pointing out to me the momentum-dependence of the wave-function renormalization "constant" $Z_V(\mathbf{p})$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Some Properties of the Contact between Theory and Experiment

PAUL A. BENIOFF*
*Argonne National Laboratory, Argonne, Illinois*

(Received 21 June 1967)

This paper uses the sample probability-space description to review, under general conditions, the measurement process whereby an empirical expectation value is obtained for comparison either with other values or with values calculated from a theory. The emphasis here is on some of the conditions which must be satisfied by general sequences of single measurements, about which an observer may have relatively little knowledge, in order that such sequences yield suitable expectation values. In particular, sequences are considered for which the requirements that the single measurements of the sequence be independent and made of the same physical quantity on ensembles of identically prepared systems are not necessarily satisfied. The differences between sequences which satisfy these requirements and those which do not are discussed in terms of the implications or meaning of the resultant expectation value as a point of contact between theory and experiment.

## I. INTRODUCTION

In science in general and particularly in quantum mechanics, the basic method of making contact between theory and experiment is by means of expectation values obtained from sequences of single measurements. From a sequence of $N$ single measurements in quantum mechanics, the mean $M_N$ of the $N$ empirical results is determined. As $N$ goes to infinity, the sequences of empirical means is supposed to converge to a limit value which is equated to or compared with an expectation value $\mathrm{Tr}\,\rho O$ obtained from theory.[1]

There are also some requirements or conditions of acceptability which are usually imposed on a sequence in order that it yield a limit empirical mean suitable as a point of contact between theory and experiment. The requirements are that each single measurement in the sequence is made of the same physical quantity on a system obtained from an ensemble prepared under identical relevant conditions.[2] Also the single measurements are usually required to be statistically independent of one another.

Although these requirements are usually considered to be satisfied for sequences of measurements in quantum mechanics, it is possible to construct examples which do not satisfy one or more of these requirements. For example, consider two states, $\rho$ and $\rho'$, prepared by two apparatus, $A$ and $A'$, each of which separately satisfy the "same preparation condition" requirement. Then one can consider a sequence of single measurements made on an ensemble prepared by using both apparatus, $A$ and $A'$, in some definite proportion. Then the sequence does not satisfy the "same preparation condition" requirement.

From this example it might appear that one does not need to impose any requirements as conditions of acceptability on sequences of single measurements. However, it is quite clear that any arbitrary sequence of single measurements is not acceptable. The question then arises regarding what some minimum conditions are which a sequence must satisfy in order to yield a limit empirical mean suitable for comparing with a theoretical calculation.

This paper is mainly concerned with this question. In particular, we consider here the description of sequences of single measurements which do not satisfy the "independence," "same physical quantity," and "same preparation condition" requirements.

* This work was performed under the auspices of the U.S. Atomic Energy Commission.

[1] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton University Press, Princeton, N.J., 1955), Chap. IV.

[2] J. M. Jauch, Helv. Phys. Acta 37, 293 (1964).

Ref. 13, in the composite Lee model, we find $Z^{-1} = 0$, contradicting what we know to be the case (namely $Z = 0$). We believe that this unfortunate conclusion is due to these authors' use of perturbation theory. The failure of perturbation theory in composite models is hardly surprising when one considers that in all known soluble models, the composite limit corresponds to $g \to \infty$.

ACKNOWLEDGMENT

My thanks are due to Dr. J. G. Taylor for pointing out to me the momentum-dependence of the wave-function renormalization "constant" $Z_V(\mathbf{p})$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Some Properties of the Contact between Theory and Experiment

PAUL A. BENIOFF*
*Argonne National Laboratory, Argonne, Illinois*

(Received 21 June 1967)

This paper uses the sample probability-space description to review, under general conditions, the measurement process whereby an empirical expectation value is obtained for comparison either with other values or with values calculated from a theory. The emphasis here is on some of the conditions which must be satisfied by general sequences of single measurements, about which an observer may have relatively little knowledge, in order that such sequences yield suitable expectation values. In particular, sequences are considered for which the requirements that the single measurements of the sequence be independent and made of the same physical quantity on ensembles of identically prepared systems are not necessarily satisfied. The differences between sequences which satisfy these requirements and those which do not are discussed in terms of the implications or meaning of the resultant expectation value as a point of contact between theory and experiment.

## I. INTRODUCTION

In science in general and particularly in quantum mechanics, the basic method of making contact between theory and experiment is by means of expectation values obtained from sequences of single measurements. From a sequence of $N$ single measurements in quantum mechanics, the mean $M_N$ of the $N$ empirical results is determined. As $N$ goes to infinity, the sequences of empirical means is supposed to converge to a limit value which is equated to or compared with an expectation value $\mathrm{Tr}\,\rho O$ obtained from theory.[1]

There are also some requirements or conditions of acceptability which are usually imposed on a sequence in order that it yield a limit empirical mean suitable as a point of contact between theory and experiment. The requirements are that each single measurement in the sequence is made of the same physical quantity on a system obtained from an ensemble prepared under identical relevant conditions.[2] Also the single measurements are usually required to be statistically independent of one another.

Although these requirements are usually considered to be satisfied for sequences of measurements in quantum mechanics, it is possible to construct examples which do not satisfy one or more of these requirements. For example, consider two states, $\rho$ and $\rho'$, prepared by two apparatus, $A$ and $A'$, each of which separately satisfy the "same preparation condition" requirement. Then one can consider a sequence of single measurements made on an ensemble prepared by using both apparatus, $A$ and $A'$, in some definite proportion. Then the sequence does not satisfy the "same preparation condition" requirement.

From this example it might appear that one does not need to impose any requirements as conditions of acceptability on sequences of single measurements. However, it is quite clear that any arbitrary sequence of single measurements is not acceptable. The question then arises regarding what some minimum conditions are which a sequence must satisfy in order to yield a limit empirical mean suitable for comparing with a theoretical calculation.

This paper is mainly concerned with this question. In particular, we consider here the description of sequences of single measurements which do not satisfy the "independence," "same physical quantity," and "same preparation condition" requirements.

* This work was performed under the auspices of the U.S. Atomic Energy Commission.

[1] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton University Press, Princeton, N.J., 1955), Chap. IV.

[2] J. M. Jauch, Helv. Phys. Acta 37, 293 (1964).

Such a description applies to any sequences for which an observer either does not or can not know that these requirements hold.

As a framework for discussion of general sequences of single measurements, probability theory is used. The fact that there are noncommuting observables in quantum mechanics is not immediately relevant to this work because the concern is with the properties of a single sequence of single measurements, not in comparing limit means obtained from different sequences. The fact that many textbook discussions of probability theory are limited mainly to independent, identically distributed[3] sequences of single measurements is not, as is well known,[4] an intrinsic limitation of probability theory.

Some general aspects of a description of a sequence of single measurements within a probability theory framework are discussed. In particular, the minimum ergodic conditions imposed by probability theory on a sequence are considered. We are also interested in the relationship between what an observer knows about a sequence and the consequences or implications of the result obtained from the sequence. It will be seen that for sequences for which the same physical quantity and same relevant preparation condition requirements are satisfied, the basic meaning or implication of a comparison between theory and experiment is stronger than it is for sequences for which an observer cannot know that these requirements are satisfied. This can be seen from the fact that, for the former type of sequence, the limit expectation value found empirically also applies to each single measurement of the sequence, whereas in the latter type, the limit value can not be applied to each single measurement.

A basic aspect of the use of probability theory as a framework for describing sequences of single measurements is that probability theory begins with or requires the assignment of a probability measure to each event generated by the sequence of single events. However, this does not imply that one must have prior knowledge of the values of the probabilities of the events of the sequence. If this were so, then probability theory would be essentially useless. As has been pointed out,[5] one can profitably discuss within this framework sequences of measurements which are not actually comtemplated or for which the experimental conditions are essentially impossible to realize on earth.

At this point, it seems worthwhile to review very briefly some of the reasons discussed elsewhere[6] why one must consider sequences of single measurements for which an observer cannot know that the same observable and same relevant preparation condition requirements are satisfied. The reason for presenting a review is that if it is possible for an observer to arrange all sequences of physical measurements such that he knows that these requirements are satisfied, then the extensions reviewed here would be of academic interest only so far as physics is concerned.

One consequence of a description of the measurement process, which includes these requirements, is that an observer must always know how to construct a sequence of single measurements which satisfies these requirements. That is, he must already know how to decide empirically whether or not each single measurement of a sequence is made of the same physical quantity on systems prepared under the same relevant conditions.

Now according to quantum mechanics, such empirical information can only be obtained as expectation values from other sequences of single measurements made of the same observables on ensembles prepared under the same relevant conditions. But then, in order to know that these requirements are satisfied means that still other sequences which satisfy these requirements have to have been done, etc. Thus one sees that the description of the measurements of an expectation value, as given in quantum mechanics, leads to an infinite regression in that any sequence of single measurements, which is to yield an expectation value, always implies previous sequences of single measurements by which one knows that the requirements are satisfied. Such a description does not allow one to start the process of acquiring empirical knowledge. The reason is that for sequences made at the start of this process, an observer cannot know what the relevant conditions are for identical preparation or how to ensure that a measurement apparatus measures the same observable for each single measurement of a sequence.

Furthermore, these arguments indicate that this inadequacy of the quantum-mechanical description

[3] An independent, identically distributed sequence is one in which there is no statistical correlation between the results of the single measurements of the sequence and for which the probability of occurrence of any single measurement result is independent of the place in the sequence at which the result occurs. Such sequences are slightly more general than those for which the single measurements are independent and made of the same physical quantity on an ensemble of systems which is prepared under the same relevant conditions. This can be seen from the following sequence which satisfies the former but not the latter requirements: (1) a flip of a coin $P$ (heads) $= \frac{1}{2}$, (2) a proton passed through a Stern–Gerlach apparatus $P$ (spin up) $= \frac{1}{2}$, etc.

[4] M. Loeve, *Probability Theory* (D. Van Nostrand Co., Inc., Princeton, N.J., 1965), 3rd ed.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1965), 2nd ed., Vol. 1, Introduction and Chap. I.

[6] P. A. Benioff (unpublished).

of the measurement process is especially pronounced for the very fundamental measurements made at the start of the process of acquiring empirical knowledge. Consider, for example, the basic property of all sequences of single measurements in that each single measurement must occupy a different space–time region. This means that before one can know that an ensemble of systems is prepared under identical conditions, he must know that the necessarily different space–time labels associated with each preparation are irrelevant variables. How does one describe the sequence of single measurements by which this fundamental knowledge is acquired? Similar questions can be asked with respect to how one describes the sequences of single measurements by which the validity of an invariance principle for space–time transformations is empirically decided.[6,7]

From this brief review, it appears that one must consider more general types of sequences of single measurements than are considered in quantum mechanics. Since probability theory can handle quite general types of sequences, it seems worthwhile to give a review of the probability-theory description of general sequences, as such material does not seem to be easily available. Such a review can also yield clues regarding the basic properties of knowledge acquisition which will be discussed in future work.

Finally, this review is restricted to those single measurements which have an arbitrarily large but finite number of outcomes. This is not an essential simplification and is used because all single measurements actually made have a finite number of observable outcomes. For the same reason it is assumed that the numbers $x_1, x_2, \cdots, x_n$ which have already been assigned to the outcomes of each single measurement are all finite. Again this is not an essential simplification. We shall in this review essentially follow Loeve.[4]

## II. REVIEW

### A. Sample-Space Description

In probability theory, any measurement can be associated with an induced probability space $(\Omega, A, P)$ where $\Omega$ is the space of all possible outcomes of the measurement, $A$ is the Boolean $\sigma$ algebra of all measurable subsets of $\Omega$, and $P$ is a probability measure defined on $A$.[8,9] In our case, the certain event $\Omega$ in the space for each single measurement is $[x_1, \cdots, x_n]$, $A$ consists of all subsets of $\Omega$, and, for any event $E$ in $A$, $PE$ is the probability of occurrence of that event.

Similarly, a sequence of $N$ single measurements is described by a Cartesian product of the $N$ measurable spaces $(\Omega_j, A_j)$ associated with the $j$th single measurement with $j = 1, 2, \cdots, N$, and a probability measure defined on the product space. An infinite sequence of single measurements is then described by the probability space $(\Omega, A, P)$, where $\Omega = (\prod_{j=1}^{\infty} \Omega_j)$, $A = (\prod_{j=1}^{\infty} A_j)$, and $P$ is a probability measure on $A$. This infinite-product space is often referred to as the sample probability space[8,9] or the phase space of an experiment.[10] In this space the points of $\Omega$, $(x_\alpha, x_\beta, \cdots)$, describe the possible trajectories of a whole experiment or equivalently give the possible infinite sequences of outcomes. The events $E_{\alpha j}$, $\alpha = 1, 2, \cdots, n$; $j = 1, 2, \cdots$, as measurable subsets of $A$, denote the events "outcome $x_\alpha$ occurred on the $j$th single measurement." These events are explicitly given in $(\Omega, A)$ as

$$E_{\alpha j} = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_{j-1} \times [x_\alpha]_j \times \Omega_{j+1} \times \cdots . \quad (1)$$

An equivalent way to generate the sample space is to consider the set of events $E_{\alpha j}$, $\alpha = 1, \cdots, n$; $j = 1, 2, \cdots$ as the generators of a Boolean $\sigma$ algebra $A$ of events on a space $\Omega$.[11] This is done by taking all countable intersections, unions, and complements among the $E_{\alpha j}$. The $E_{\alpha j}$ are defined relative to the empty set $\Phi$ and the certain event $\Omega$ by $\Omega = \bigcup_{\alpha=1}^{n} E_{\alpha j}$ for each $j$ (this says that each single measurement will yield some outcome) and $\Phi = E_{\alpha j} \cap E_{\beta j}$ for any $j$ and $\alpha \neq \beta$ (this says that two outcomes cannot occur simultaneously in any single measurement). The points of $\Omega$ are obtained from the infinite intersections of the $E_{\alpha j}$ for different $j$ and correspond to the possible infinite sequences of outcomes. Again $(\Omega, A, P)$ is the sample probability space with $P$ defined on $A$.

The random variable $X_j$ (as an element of a space of measurable functions from $\Omega$ to Borel sets of the real line) which represents the $j$th single measurement is given by $(j = 1, 2, \cdots)$

$$X_j = \sum_{\alpha=1}^{n} x_\alpha I E_{\alpha j}, \quad (2)$$

where $IE_{\alpha j}$, the indicator random variable for the event $E_{\alpha j}$, has the value 1 for all points $\omega \in E_{\alpha j}$ and is equal to 0 otherwise. The random variable $\bar{X}_N$ representing the mean of the first $N$ single measurements is defined as

$$\bar{X}_N = \frac{1}{N} \sum_{j=1}^{N} X_j . \quad (3)$$

One can also define on the event algebra $A$ a

[7] E. P. Wigner, Nuovo Cimento **3**, 517 (1963); R. F. Houtappel, H. Van Dam, and E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965); H. Ekstein, Phys. Rev. **153**, 1397 (1967).

[8] Reference 4, Chap. III, pp. 362, 363.

[9] Reference 5, Chaps. I, V, and VIII.

[10] U. Uhlhorn, *Rendiconto Scuola Internazionale Fisica Enrico Fermi, Varenna, Italy, 1960; Course 14: Ergodic Theory* (Academic Press Inc., New York, 1962), pp. 195–206.

[11] Reference 4, Sec. 26.

measurable set transformation $T$ by the relation

$$E_{\alpha,j+1} = TE_{\alpha j} \tag{4}$$

for all $\alpha = 1, \cdots, n$ and $j = 1, 2, \cdots$. Also for any event $E = E_{\alpha j} \cap E_{\beta k} \cap \cdots \cap E_{\epsilon n}$ one has

$$TE = E_{\alpha,j+1} \cap E_{\beta,k+1} \cap \cdots \cap E_{\epsilon,n+1}. \tag{5}$$

If desired, $T$ can be considered as a time translation by one unit and as such depends on the interval $j + 1 - j$ and not on $j + 1$ and $j$ separately. One also has

$$E_{\alpha,j+p} = T^p E_{\alpha j}, \tag{6}$$

with $p = 0, 1, \cdots$ and $T^0 = 1$. Also for any event $E$, consisting of arbitrary intersections and unions of different $E_{\alpha j}$, $T^pE$ shifts each index $j$ to $j + p$.

This transformation also induces a transformation $T'$ on the space of random variables given by

$$T'IE_{\alpha j} = ITE_{\alpha j} = IE_{\alpha,j+1} \tag{7}$$

for all $j$ and $\alpha = 1, \cdots, n$. Thus one has for $p = 0, 1, \cdots$

$$X_{j+p} = (T')^p X_j, \tag{8}$$

and Eq. (2) becomes

$$\bar{X}_N = \frac{1}{N} \sum_{j=1}^{N} (T')^{j-1} X_1 = \bar{T}'_N X_1, \tag{9}$$

where the transformation $\bar{T}'_N$ is defined by

$$\bar{T}'_N = \frac{1}{N} \sum_{j=1}^{N} (T')^{j-1}. \tag{10}$$

It is worthwhile at this point to discuss some aspects of what has been presented so far. First of all, the events $E_{\alpha j}$ and the random variables $X_j$ were first introduced and then the transformations $T$ and $T'$ were defined on the given events and random variables. However, as Loeve has noted,[12] one can use the transformations to define the sequence of single measurements, given the first one. Thus, if $E_{\alpha 1}$ for $1 \leq \alpha \leq n$ is the event "outcome $\alpha$ occurred on the first measurement," the event which corresponds to the *same* outcome of the $j$th single measurement as occurred in the first single measurement is *defined* as $T^{j-1}E_{\alpha 1}$. Similarly, the random variable $X_j$ representing the $j$th single measurement, which is a "repetition" of the first single measurement, is defined by $(T')^{j-1}X_1$.

Also, it is quite important to note that the set transformation $T$ is neither invertible nor necessarily measure preserving. The noninvertibility of $T$ can be seen from the fact that the inverse of $T$, if defined,[13] when applied to any event $E_{\alpha j}$ would give $E_{\alpha,j-1}$. However, this inverse when applied to the events $E_{\alpha 1}$ for any $\alpha$ is not defined as the $E_{\alpha 1}$, for $1 \leq \alpha \leq n$ describe the first events which can occur. As a result $T$ has no inverse. This lack of an inverse corresponds to the description of a sequence of single measurements as having a beginning but no end. If one prefers to think of the sequence of single measurements as a sampling of a hypothetical doubly infinite sequence extending both into the infinite past and future, then the probability-theory description given above can be easily changed to describe this situation. In this case $T$ has an inverse, since the index $j$ labeling any event $E_{\alpha j}$ now can have negative and zero values instead of positive values only. In this case the $T^p$ for $p = 0, 1, \cdots$ form a group of transformations rather than a semigroup.

In our opinion the fact that a sequence of single measurements has a beginning (but no end) is important. In particular, the fact that an observer can always add on more single measurements after an arbitrary finite sequence has been completed but that he cannot add on any before the first one, is an important aspect of the knowledge-acquisition process. That is, this process has a beginning. For this reason the description given above in which $T$ does not have an inverse is to be preferred.

The not necessarily measure-preserving property of $T$ means that for any event $E$ in $A$ the probability of the event $TE$ need not equal that of the event $E$. That is

$$PTE \neq PE \tag{11}$$

must be allowed for. This possibility arises because we specifically want to include sequences of single measurements which are not identically distributed. As was noted earlier, sequences of single measurements made on the same system or made on an ensemble of identically prepared systems which are described by measure-preserving transformations are special cases of the more general description reviewed here. It should also be noted that we do *not* require that the single measurements be statistically independent.

The expectation value of the mean of the first $N$ single measurements is obtained from Eqs. (2) and (3) as

$$\langle \bar{X}_N \rangle = \sum_{\alpha=1}^{n} x_\alpha \frac{1}{N} \sum_{j=1}^{N} PE_{\alpha j}, \tag{12}$$

where $\langle IE_{\alpha j} \rangle = PE_{\alpha j}$. Using Eq. (4) this result can also be written as

$$\langle \bar{X}_N \rangle = \sum_{\alpha=1}^{n} x_\alpha P_N E_{\alpha 1}, \tag{13}$$

[12] Reference 4, Secs. 30, 31.

[13] By the inverse of $T$ is meant the actual inverse of $T$ also defined as a set transformation, not a point transformation $T^{-1}$ corresponding to $T$. This use (Ref. 12) of $T$ rather than $T^{-1}$ to denote a set transformation is different from popular usage.

where the probability measure $P_N$ is defined in terms of $P$ by[13]

$$P_N E = \frac{1}{N} \sum_{j=1}^{N} P T^{j-1} E \tag{14}$$

for every $E$ in $A$.

Also the limit probability measure $\bar{P}$, if it exists, is defined by

$$\bar{P}E = \lim_{N \to \infty} P_N E \tag{15}$$

for every event $E$ in $A$. If this limit probability exists, then the limiting expectation of the random variable $\bar{X}_N$ is given from Eqs. (2) and (9) as

$$\lim_{N \to \infty} \langle \bar{T}'_N X_1 \rangle = \sum_{\alpha=1}^{n} x_\alpha \bar{P} E_{\alpha 1}. \tag{16}$$

The basic conditions which must be satisfied if a sequence of single measurements is to yield a number which can serve as a point of contact between theory and experiment is that the associated sequence of empirical means under consideration converge to a limit which can be uniquely represented as an expectation value. These conditions are equivalent to the requirements that the sequence of single measurements be ergodic, or that the sequence of random variables $\bar{X}_N$ $N = 1, \cdots$ satisfy an ergodic theorem and the ergodic hypothesis. In our case an appropriate ergodic theorem states that if a limit probability measure $\bar{P}$ exists on $A$ then, as $N \to \infty$, the random-variable sequence converges almost surely. That is, if Eq. (15) holds for every $E$ in $A$, then from Eqs. (2), (9)[12]

$$\text{a.s.} \lim_{N \to \infty} \bar{X}_N = \bar{T}' X_1 = \bar{E}^{\Gamma} X_1, \tag{17}$$

where $\bar{E}^{\Gamma} X_1$ is defined as the conditional expectation of $X_1$, given the sub $\sigma$ algebra $\Gamma$ of invariant events. (An event $C \in \Gamma$ is invariant if $TC = C$.) By almost sure convergence of $\bar{X}_N$ to $\bar{E}^{\Gamma} X_1$ is meant that $\bar{X}_N(\omega) \to \bar{E}^{\Gamma}_{\omega} X_1$ in the Cauchy sense for every point $\omega$ of $\Omega$ except possibly on a subset $M$ of $\Omega$ for which $\bar{P}M = 0$. (One also has $\bar{P}M = 0$ because $T$ must be a null-preserving transformation.[12])

Ergodic theorems, which, in our case, are entirely equivalent to Eq. (17), are the mean ergodic theorems[14] which say that for $r \geq 1$, if a limit probability $\bar{P}$ exists on $A$, then[13]

$$\lim_{N \to \infty} \left[ \int_{\Omega} |\bar{T}'_N X_1(\omega) - \bar{T}' X_1(\omega)|^r \, dP(\omega) \right]^{1/r} = 0. \tag{18}$$

In particular for $r = 2$, Eq. (18) is equivalent to mean-square convergence of the sequence of empirical means. For the special case of $T$ being invertible and measure-preserving, Eqs. (17) and (18) are Birkhoff's individual ergodic theorem[15] and Von Neumann's mean ergodic theorem ($r = 2$),[16] respectively.

The desired result has not yet been achieved because $\bar{T}' X_1 = \bar{E}^{\Gamma} X_1$ still need not represent, almost surely, a single numerical result. In fact, if $\Gamma$ is generated by a countable partition $[C]$ of the sure event $\Omega$, then from Eq. (2), $\bar{E}^{\Gamma} X_1$ is defined up to a $\bar{P}$ equivalence by[17,18]

$$\bar{E}^{\Gamma} X_1 = \sum_{\alpha=1}^{n} x_\alpha \sum_{C} \bar{P}_C E_{\alpha 1} IC, \tag{19}$$

where $IC$ is the indicator random variable for the invariant event $C$ and $\bar{P}_C E_{\alpha 1}$ is the limit conditional probability for the event $E_{\alpha 1}$ given that the event $C$ has occurred. The statement "up to a $\bar{P}$ equivalence" refers to the fact that $\bar{P}_C E_{\alpha 1}$ is undefined for any event $C$ for which $\bar{P}C = 0$.

The ergodic limit of Eq. (19) shows the problem more clearly in that the limit of the empirical mean sequences may be equal to any one of the values $\sum_\alpha x_\alpha \bar{P}_C E_{\alpha 1}$ for different $C$ and in general one may not know which $C$ of the partition to choose. This would be the case if the structure of the partition $[C]$ or which points $\omega$ of $\Omega$ belong to which $C$ were not known. An equivalent statement is that the space $\Omega$ is almost surely decomposable into invariant subspaces $C$ and that the infinite sequence of single measurements almost surely occupies only one of the subspaces $C$. In particular, this means that if the infinite sequence were repeated, then there is a nonzero probability $P$ that a different limit value would be obtained. This can be easily seen from Eq. (17) by recalling that a given infinite sequence represents a particular infinite sequence of outcomes and hence a specific point of $\Omega$. If $\omega$ labels the first infinite sequence and $\omega'$ the repetition of the sequence, then by Eq. (17) the limit values obtained for the first and second sequences would be given by $\sum_\alpha x_\alpha \bar{P}_{C_1} E_{\alpha 1}$ and $\sum_\alpha x_\alpha \bar{P}_{C_2} E_{\alpha 2}$, where $\omega \in C_1$ and $\omega' \in C_2$.

To avoid these difficulties, the requirement that the sequence of single measurements is metrically transitive is imposed. This is equivalent to the requirements that the transformation $T$ be $P$ indecomposable or that the only invariant events in $\Gamma$ are almost surely $\Phi$ and $\Omega$.[12] In this case $\bar{E}^{\Gamma} X_1 = \bar{T}' X_1$ reduces to an invariant degenerate random variable whose value is

[14] Convergence in $r$th mean and almost sure convergence are equivalent in our case because the random variables discussed here are all bounded; see Ref. 12.

[15] G. D. Birkhoff, Proc. Natl. Acad. Sci. **17**, 656 (1931).

[16] J. Von Neumann, Proc. Natl. Acad. Sci. **18**, 70, 263 (1932).

[17] Reference 4, Sec. 24.

[18] Eq. (19) can also be extended to include noncountable paritions of $\Omega$; see Ref. 12.

given by

$$\bar{T}'X_1 = \sum_{\alpha=1}^{n} x_\alpha \bar{P}E_{\alpha 1}. \tag{20}$$

A degenerate random variable $\bar{X}$ is one whose value on $\Omega$ is almost surely a constant equal to $\langle \bar{X} \rangle$.

### B. Discussion

With this we come to the desired result that a sequence of single measurements yields, in the limit, a number which corresponds to a unique expectation value. However, there are many other interesting aspects of this process which are worth considering at this point.

#### 1. *Invariance Properties of $\bar{P}$ and $\bar{T}'$*

First of all, the invariance properties of $\bar{P}$ and $\bar{T}'$ under $T$ and $T'$, respectively, should be noted. In fact it is easy to see from the definitions of $\bar{P}$ and $\bar{T}'$ that for any event $E$, $\bar{P}TE = \bar{P}E$ and $\bar{P}C = PC$ for any invariant event $C$.[12] Also the ergodic limit operator $\bar{T}'$ is invariant under $T'$ or commutes with $T'$ so that $T'\bar{T}'X = \bar{T}'T'X = \bar{T}'X$. This can be seen from the facts that $\bar{T}'X$ is a sum over invariant events [Eq. (19)] and that $\bar{P}$ is an invariant measure. The invariance of $\bar{P}$ and $\bar{T}'$ occurs because $T$ satisfies the semigroup or group multiplication property. That is, the translation $T_m$ from events of the $j$th single measurement to events of the $(j+n)$th single measurement is equal to $T^m$ for $m = 0, 1, 2, \cdots$.

These properties of $\bar{P}$ and $\bar{T}'$ mean that the limit expectation values obtained from an infinite sequence of single measurements are independent of where the sequence is started. That is, the fact that $\bar{P}E_{\alpha j} = \bar{P}E_{\alpha 1}$ and $\bar{T}'X_j = \bar{T}'X_1$, for any $j$, means that one could discard an arbitrary initial segment of the sequence without affecting the limit results or, equivalently, the sequence could be started with the $j$th single measurement as well as the first without affecting the limit results. If $T$ refers to a time translation by one unit, then this invariance means that the expectation values are independent of what time $t_j$ the sequence begins.

#### 2. *Comparison between Theory and Experiment*

The previous discussion has shown that the description of the limit experiment, which consists of the determination of the mean of an infinite ergodic sequence of single measurements, is described by an invariant degenerate random variable $\bar{T}'X_1$ where

$$\bar{T}'X_1 = \langle \bar{T}'X_1 \rangle, \tag{21}$$

almost surely. The result obtained from this experiment is an expectation value given uniquely by Eq. (21) or (16) and as such can be compared directly with a theoretical number. This suggests an alternate definition of the conditions under which the result of an experiment is suitable for comparison with theory. Namely, that any experiment, whether it consists of a finite or infinite number of single measurements, which can be described by an invariant degenerate random variable, gives a result which can be directly compared with theory.[19] For any such experiment, even if it consists of one single measurement, gives almost surely a single unique result which is invariant under repetition and any repeated sequence of experiments described by invariant degenerate random variables is trivially ergodic. However, for essentially all measurements made in physics, the random variables $X_j$, $j = 1, 2, \cdots$ describing each single measurement of a sequence are neither invariant ($X_{j+1} \neq X_j$) nor degenerate [Eq. (2)]. In this case, any experiment which is described by invariant degenerate random variables is generated only by an infinite ergodic sequence of single measurements.

As is the case for the $X_j$, the random variables describing the means of the first $N$ single measurements $\bar{X}_N$ [Eqs. (3) or (9)] for any finite $N$ are also neither invariant nor degenerate. Thus the empirical mean $M_N$ obtained from $N$ single measurements can not strictly be compared with theory. The lack of invariance of $\bar{X}_N$ under $T'$ means that $M_N$ is not invariant under a change in the starting point of the sequence of $N$ single measurements. Also, the lack of degeneracy of $\bar{X}_N$ means that $M_N$ can be any one of many possible values taken on by $\bar{X}_N$ and consequently $M_N$ is neither the almost surely unique expectation value, $\langle \bar{T}'X_1 \rangle$ nor $\langle \bar{X}_N \rangle$ [Eq. (13)].

On the other hand, it is well known that one acquires knowledge by making comparisons between means of a finite number of single measurements and theoretical expectation values. Of course, such a comparison involves an approximation which states how "close" the empirical mean, which is one of the values of $\bar{X}_N$, is to the limit expectation value. The well-known point[20] we wish to stress again[6] is that all such approximation statements relate an empirical result to a calculated expectation value and as such are also strictly valid only in the limit of an infinite

[19] This definition includes some trivial types of single-measurement sequences such as a sequence of flips of a two-headed coin, etc., for which each single measurement is described by an invariant degenerate random variable. However, for all these cases the result of each single measurement is known in advance and thus need not be considered at all. As a result, this type of measurement will not be considered further.

[20] A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill Book Co., New York, 1965), Chap. 8.

ergodic sequence. In fact, such statements can be easily constructed using the tools given in the last section by considering an infinite sequence of repetitions of a sequence of $N$ single measurements.

### 3. *The $L_\infty(\Omega, A, P)$ Space of Random Variables*

In the discussion given so far, we have concentrated on the properties of a sequence of random variables $[X_j]$ generated from one random variable $X_1$ [Eqs. (2), (3), and (8)]. However, the discussion can easily be extended to cover properties of all $L_r(\Omega, A, P)$ spaces containing all bounded random variables defined on the sample probability space $(\Omega, A, P)$. Since in actual measurements the numbers associated with the outcomes of each single measurement are always finite, we restrict them to the space $L_\infty(\Omega, A, P)$ of almost surely bounded random variables.[21]

If one now considers sequences of random variables generated by repeated applications of $T$ to any random variable in $L_\infty$ [Eqs. (1)–(15)], then one has the same ergodic theorems as before but applying to every random variable in $L_\infty$.[12] That is, if the limit measure $\bar{P}E$, defined by Eq. (15), exists for every event $E$ in $A$, then the ergodic theorems Eqs. (17) and (18) hold for every random variable in $L_\infty$. Equation (17) becomes

$$\text{a.s.} \lim_{N\to\infty} \bar{T}'_N X = \bar{T}'X = \bar{E}^{\Gamma} X \qquad (22)$$

for every $X$ in $L_\infty$. Furthermore, if the sequence of single measurements is metrically transitive, or the ergodic hypothesis holds, then every random variable $\bar{T}'X$ is both invariant under $T'$ and degenerate with its value given almost surely by $\langle \bar{X} \rangle$, where $\langle \bar{X} \rangle$ is taken with respect to the limit measure $\bar{P}$.

The main reason for this change of emphasis from one random variable to every random variable in $L_\infty(\Omega, A, P)$ is as follows: If the ergodic conditions are met, i.e., if the limit measure $\bar{P}$ exists and $P$ is $T$ indecomposable,[12] then the one sequence of single measurements, which generates the sample space $(\Omega, A, P)$, is sufficient to yield empirical determinations of the expectation values of every random variable in $L_\infty(\Omega, A, P)$. For example, from the *one* sequence, an observer can empirically determine the value of $\bar{P}E$ for every $E$ in $A$ merely by setting $X = IE$. This holds also no matter how complex $E$ is [Eq. (5)] or how much dependence there is among the different single measurements of the sequence.

This fact is, of course, well known in quantum mechanics for sequences of independent single measurements made of the same observable on an ensemble of identically prepared systems. In this case, if the observable $O = \sum O_\alpha P_\alpha$ is being measured, it is clear that, besides empirically determining $\langle O \rangle$ from the one sequence, the expectation values of each observable in the algebra generated by the set of projection operators $[P_\alpha]$ are determined. Of course, this requires that the observer know the actual result sequence (which corresponds to a point $\omega$ in $\Omega$) associated with the sequence of single measurements.

### 4. *The Ergodic Properties*

The previous discussion has shown us that the requirement, that an infinite sequence be ergodic, must be satisfied in order that the result obtained can be compared with any theoretical expectation value. It is worthwhile to stress at this point how really basic this requirement is to the process of knowledge acquisition. In particular, if the sequence does not satisfy an ergodic theorem, then none of the associated sequences of empirical means will converge. Such a sequence of single measurements, if carried out, will not give new knowledge as it does not yield any result which can be compared with other results or with theory. These points can be seen by consideration of an example of an attempted probability measurement by a sequence of single measurements which does not satisfy an ergodic theorem.

In this example there are only two outcomes 0 and 1 for each single measurement and the random variable for the outcome 1 for the $j$th single measurement is given by $X_j = IE_{1j}$ in Eq. (2). Now suppose the sequence of single measurements is such that the probability of events $E_{1j}$ is given by $PE_{1j} = K$ for $2^n \leq j < 2^{n+1}$, $n = 0, 1, 2, \cdots$ and $K = \frac{2}{3}$ for $n$ even and $K = \frac{1}{3}$ for $n$ odd. For such a sequence, the average probability for outcome 1 for the first $N$ measurements, $P_N E_{11}$ [Eqs. (4), (13), and (14)] oscillates between values of $\frac{5}{9}$ and $\frac{4}{9}$ as $N$ is increased. As a result the limit measure $\bar{P}$ of Eq. 15 does not exist and ergodic theorems of Eq. (17) or (18) are not satisfied. It is also clear that the empirical mean $M_N$, obtained from the first $N$ single measurements, is meaningless even as an approximate measure of $\bar{P}E_{1j}$ as the latter does not exist.

However, it must also be kept in mind that a segment of $N$ single measurements can be part of many different infinite sequences of single measurements, some of which may be ergodic. Thus while $M_N$ is useless as a measurement of $\bar{P}$ in the sequence described, it may be useful as a measurement of $\bar{P}_N E_{11}$ in another sequence which consists of an infinite repetition of the first $N$ single measurements. In particular, if, in the new sample probability space

[21] Reference 4, Sec. 9.4: The results obtained here are essentially the same for any $L_r$ space with $r \geq 1$; see Ref. 12.

of this new sequence, the measurements are independent of one another and the transformation $T$ on the new space is measure preserving, then this new sequence is ergodic[12] and does give knowledge of the values of $P_N E_{11}$.

Another way to stress the really basic aspect of the ergodic theorem requirement is to note that if a sequence of single measurements did not satisfy an ergodic theorem, it would not be a measurement or even an observation. Consider, for example, the formation of an image of an object on a grid of photosensitive detectors behind a lens (the eye of an observer, for instance). In order to see or register anything, a large number of photons must be scattered off the object through the lens into the grid. Now an image is formed only if the relative frequency of firing of each detector in the grid converges to a limit which is proportional to the intensity of the scattered light. (For the eye, the observer would see the object.) On the other hand, if the relative frequency of firing of each detector did not approach a limit or satisfy an ergodic theorem, then one would not have an image but instead would have a meaningless jumble of light flashes or detector firings.

It should be noted that this requirement does not mean that the image cannot vary with time. Rather, it means that the variation with time be sufficiently slow with respect to the incident photon current. In this case the sequence of grid firings, obtained during a time in which the photon current has not changed appreciably, can be considered as an initial segment of another independent, identically distributed sequence of infinite repetitions of the initial segment. In this case the initial segment can be a good approximation to the limit image.

The above discussion has shown the importance of the ergodic theorem requirement on the sequence of random variables describing a sequence of single measurements. However, it is also important that the ergodic hypothesis be satisfied. The discussion preceding Eq. (20) showed that a $P$-decomposable sequence, rather than occurring in the entire space $\Omega$, occurs, almost surely, in one of the invariant subspaces $C$.

Now if one can tell from the outcome of the single measurements in which invariant subspace the sequence lies, then either the probability $P$ can be replaced by the conditional probability $P_C$ on the original sample space, or a new sample space, in which $C$ is the certain event and $P$ is the probability measure, is used to describe the sequence. In either case the ergodic hypothesis is satisfied[12] and the sequence of measurements yields an expectation value which can be compared with theory.[22] This procedure is equivalent to discarding all but one of the terms in the $C$ sum of Eq. (19) and replacing $IC$ by 1.

Unfortunately, the general situation may not be as pleasant as this. Besides the possibility that an observer may not know the invariant subspaces of $\Omega$, there is a worse difficulty. This is that even if one knows some of the invariant subspaces and that the sequence lies in one of them, it may turn out that the invariant subspace containing the sequence is even further $P$ decomposable into still smaller invariant subspaces. This would be the situation if the probability measure in the sample probability space is mostly unknown to an observer. The reason for this is that the decomposability of a space is always considered relative to the points of the certain event $\Omega$ outside the $P$-null events. Thus, if one does not know which part of the space $\Omega$ is $P$ null, one can not know whether or not it is almost surely decomposable under $T$.

Furthermore, any experiments by which one would hope to determine whether or not the sequence was metrically transitive, such as determining the $P$-null structure of $(\Omega, A, P)$, are infinite sequences of single measurements which give expectation values. Thus the problem of knowing whether or not the original sequence is metrically transitive is transferred to these new sequences. Further attempts at answering the problem merely transfer it to still other sequences, giving an infinite regression but no solution.

A similar basic problem concerns how an observer knows that a sequence of single measurements satisfies an ergodic theorem. It is clear that he cannot directly verify the convergence of the empirical mean sequence, because he can at most obtain a finite number of terms. As is well known, knowledge of an arbitrary initial finite segment of an infinite sequence gives no information about the convergence properties of the sequence.

From this discussion, it would appear that the problem of whether or not sequences of single measurements satisfy an ergodic theorem and are metrically transitive would be of central importance to the usual application of probability theory to the measurement process. Yet these problems are usually not discussed in textbooks on probability and statistics. The basic

[22] As a simple example, consider a sequence in which the first single measurement consists of randomly selecting a coin out of a box of nickels and dimes and flipping it. All succeeding single measurements consist of repeated flips of the selected coin. For such a sequence, the points $\omega$ of the certain event $\Omega$ in the sample probability space are the infinite sequences of the four outcomes, $nh$, $dh$, $nt$, and $dt$ where $d$ = dime, $h$ = heads, etc. For this case $\Omega$ is almost surely $P$ and $T$ decomposable into the invariant subspaces $C_d$ and $C_n$ where $C_d$ and $C_n$ correspond to the respective selection of a dime and a nickel in the first single measurement and measured limit probability is $\bar{P}_{C_d}$ or $P_{C_n}$, respectively.

reason for this is that essentially all measurements in science are assumed to fall into a special class in which the single measurements in a sequence are statistically independent and made of the same physical quantity on either the same system or on an ensemble of systems prepared under the same relevant conditions.[2] A sequence of random variables describing a sequence of single measurements, which satisfy these requirements, is an independent, identically distributed sequence and as such automatically satisfies both an ergodic theorem and the ergodic hypothesis.[12] For this case the convergence property is usually described in terms of the weak or strong law of large numbers or mean-square convergence and the ergodic hypothesis in terms of the Borel zero–one law.[23]

In conclusion, it should be noted that there are several papers which discuss ergodicity, especially in connection with the measurement process in quantum mechanics.[10,24] The basic difference between our discussion and that of these other papers is that these other papers first assume a basic quantum dynamical description of each system or of the measuring apparatus. Then the ergodic aspects arise in the description of a sequence of measurements made on one system or of macroscopic systems whose microscopic properties are at best only partly known. Our discussion is different in that no physical dynamical description of the behavior of single systems is postulated. The probability theory description of a sequence of single measurements requires the existence of a probability measure on the events in a sample space. But it is irrelevant to the theory whether or not these events are connected by any physical dynamics.

## III. FURTHER ASPECTS OF THE CONTACT BETWEEN THEORY AND EXPERIMENT

### A. Empirical Determination of $\bar{P}$ and $P$

It is of interest to compare the consequences, with respect to the theory–experiment contact point, of the minimum ergodic requirements with those of the stronger independence and identical distribution requirements. It is to be expected that as the requirements on the sequences become more stringent, or that the amount of prior knowledge an observer has about the sequence increases, then the implication or meaning of "comparison" or "agreement" between theory and experiment becomes stronger. Such is indeed the case.

It is necessary to digress slightly at this point to recall that with respect to events $E$, described in the sample space $(\Omega, A, P)$, $PE$ and $\bar{P}E$ are probabilities associated respectively with one measurement described by $E$ and with a limit average over an infinite ensemble of such unit measurements [Eqs. (14) and (15)]. Now if each single measurement of a sequence is performed on a different system, then, for the events $E_{\alpha j}$ (for all $\alpha$ and $j$), $P$ and $\bar{P}$ can be thought of as the measures associated with single systems and with infinitely large ensembles of systems, respectively. Similarly, for events of the type $E_{\alpha 1} \cap E_{\beta 2} \cap \cdots \cap E_{\epsilon m}$, $P$ is the measure associated with a single $m$-body correlation measurement on a single set of $m$ systems, whereas $\bar{P}$ is associated with an average of successive $m$-body correlation measurements on an infinite ensemble of systems. Thus $P$ is associated with a single measurement on a single finite set of one or more single systems, whereas $\bar{P}$ is always associated with an infinite ensemble of systems.

Now let us first consider the consequences of the ergodic requirements. From the discussion given earlier, it was seen that the limit means obtained from an infinite sequence of single measurements are associated with the measure $\bar{P}$. This means that for any sequence of single measurements which are known to satisfy the ergodic requirements, the limit ensemble probability measure $\bar{P}$ is directly observable or empirically measurable. However, the measure $P$ associated with single measurements is not directly observable or empirically measurable. That is, given any event $E$, the sequence of single measurements does not give one the value of $PE$. This can be seen directly from the relation between $P$ and $\bar{P}$ [Eqs. (14) and (15)] where for general non-measure-preserving transformations there is no way to relate $PE$ to $\bar{P}E$, since each term in the $j$ sum of Eq. (14) can be different.[25]

On the other hand, if one also knows for any sequence that $T$ is measure preserving or that the sequence is stationary,[12] then the terms in the $j$ sum of Eq. (14) are all equal. In this case, one has $P = \bar{P}$, which has the consequence that both the limit-ensemble probability and the single-measurement probabilities are directly empirically measurable. That is, for any event $E$, the sequence gives one empirical values of $\bar{P}E$ which can be set equal to $PE$.

From these considerations, one arrives at the important point that the contact between theory and experiment, or the meaning of agreement between theory and experiment given by a stationary sequence

[23] Reference 4, pp. 228–230; Ref. 5, Chaps. VIII, X.

[24] A. Daneri, A. Loinger, and G. M. Prosperi, Nuovo Cimento **44B**, 119 (1966); Nucl. Phys. **33**, 297 (1962); G. Ludwig, *Rendiconto Scuola Internazionale Fisica Enrico Fermi, Verenna, Italy, 1960; Course 14: Ergodic Theory* (Academic Press Inc., New York, 1962), pp. 57–132.

[25] At present, it is not clear how much empirical knowledge of the measure $P$ can be obtained from general ergodic sequences. Although $P$ cannot be directly determined, it is clear that at least some information about $P$ can be obtained, other than that implied by the ergodic requirements. It is hoped to study this point in future work.

is stronger or implies more than does the contact given by a nonstationary ergodic sequence. This occurs because whatever one learns from a comparison of a limit empirical mean obtained from a stationary sequence to a theoretical number applies both to the limit ensemble and to each single measurement. However, whatever one learns from a similar comparison made for a more general ergodic sequence applies directly to the limit ensemble only.

Now, if each single measurement consists of a preparation and measurement procedure and it is known that the sequence of single measurements is ergodic and stationary, then an observer can obtain from the sequence empirical knowledge of the probabilities associated with each single system. However, this is usually not sufficient for physical measurements because these single-system probabilities may not represent an ensemble of independent or noninteracting systems. This is taken care of by the additional requirement that the ensemble of single measurements be statistically independent of one another.

In this case, then, for any sequence which is known to be independent and stationary or identically distributed,[3,12] knowledge of the values of the limit measure $\bar{P}$ also implies knowledge of the independent or noninteracting single-system probabilities. In this case, the implications or consequences of agreement or comparison between theory and experiment are quite strong compared to that for sequences which are known only to be ergodic.

The cases discussed above represent rather extreme cases of a large amount of knowledge, the independence and identical-distribution (or stationarity) requirements and a small amount of knowledge (the ergodic requirements only) which one may have about a sequence. There are also many intermediate cases. Consider, for example, a sequence which is known to satisfy the ergodic requirements and which is also known to be Markovian. In this case, the limit probabilities $\bar{P}E_{\alpha 1}$ for each $\alpha$ are obtained from the transition matrix only[26] and as such are independent of the initial probabilities $PE_{\alpha 1}$. Thus one sees that a measurement of the values of $\bar{P}E_{\alpha 1}$ gives no information about the values of $PE_{\alpha 1}$. On the other hand, it is known that the values of $PE_{\alpha j}$ converge exponentially to $\bar{P}E_{\alpha 1}$ as $j \to \infty$.[26] Therefore, for Markov sequences of single measurements, one sees that knowledge of the values of $\bar{P}E_{\alpha 1}$ also implies knowledge, to an accuracy $\epsilon$, of all but $M(\epsilon)$ values of $PE_{\alpha j}$ for $j = 1, 2, \cdots$, where $M(\epsilon)$ is an exponential function depending on the convergence rate of the sequence.

These considerations of the effect that different amounts of prior knowledge about a sequence of single measurements have on the consequences of the empirical results, stress the importance of understanding the dynamics of the knowledge acquisition process. They show an interesting feedback aspect in that the more prior knowledge one has about a sequence, the stronger the consequences are of the empirical results. If one knows only that a sequence is ergodic, then the empirical results imply less in that they give complete knowledge of the limit measure $\bar{P}$ and, at most, partial knowledge about the measure $P$.[25] On the other hand, if one knows relatively more about a sequence, i.e., that it is independent and identically distributed, then the same empirical results imply more in that they give complete knowledge both of the limit ensemble and single-system measures.

This feedback aspect means that one must be careful about making arbitrary assumptions of prior knowledge at the start of any measurement sequence. It may well be that such assumptions have nontrivial consequences for the measurement process, and possibly even for physics.[6] Thus it may be that if one arbitrarily assumes that the preparation and measurement procedures used to construct a single measurement sequence give an independent identically distributed sequence, then the empirical results obtained from this and other sequences may not contradict and might even support the assumption. On the other hand, if one does not make this assumption, then exactly the same empirical results, which now are weaker in that they give at best only partial knowledge of $P$, may say nothing about whether the procedures used yield independent, identically distributed sequences.

### B. Connection with Quantum Mechanics

Here we shall only indicate very briefly how contact might be made with quantum mechanics and leave for future work a more detailed discussion. In the previous discussion, it was seen that direct contact between a physical theory and experiment is made by means of limit empirical means obtained from infinite sequences of single measurements. Furthermore, this contact gives one direct knowledge of the measure $\bar{P}$. Since quantum mechanics is a physical theory which provides expectation values for comparison with experiment, it is natural to equate $\bar{P}E_\alpha$ to $\mathrm{Tr}\rho P_\alpha$. That is, from Eq. (20) one writes for ergodic sequences

$$\bar{T}'X = \sum_\alpha x_\alpha \bar{P}E_\alpha = \sum_\alpha x_\alpha \mathrm{Tr}\rho P_\alpha = \mathrm{Tr}\,\rho O, \quad (23)$$

where the eigenvalue expansion of the observable $O$ is given by $\sum_\alpha x_\alpha P_\alpha$. The subscript $j$ has been left off the event to indicate the independence of $\bar{P}E_\alpha$ from $j$.

[26] Reference 4, Sec. 27.

It is clear from this equation that $\mathrm{Tr}\rho O$ represents the infinite ensemble of single measurements or that $\rho$ represents an infinite ensemble of single systems.[27] This occurs because of the direct relation between $\mathrm{Tr}\rho P_\alpha$ and the limit ensemble measure $\bar{P}$. Furthermore, since the previous discussion has shown that one cannot directly relate $\bar{P}$ to $P$ for general ergodic sequences,[25] one cannot use $\mathrm{Tr}\rho O$ as a description of any single measurement of the sequences. Equivalently, one cannot directly use $\rho$ in this case to describe a single system.

On the other hand, any sequence for which it is known that the single measurements are independent and made of the same physical quantity on an ensemble of identically prepared systems is a sequence for which $P = \bar{P}$. In this case, $\mathrm{Tr}\rho O$ refers both to the infinite ensemble of single measurements as well as to each single measurement. In this case, it would seem that $\rho$ can indeed represent a single system.

These aspects may throw some light on the controversy regarding whether a state in quantum mechanics represents an infinite ensemble of systems or a single system.[27] The preceding discussion suggests that, if one can show that, starting from general ergodic sequences, it is possible to acquire sufficient knowledge of the physical world to enable him to construct independent identically distributed sequences, then $\rho$ can represent either a limit ensemble or a single system. However, if such knowledge is not rigorously attainable, then $\rho$ can exactly represent an infinite ensemble only. The problem of whether such knowledge is attainable or not is not trivial in quantum mechanics as other work has shown.[6]

## IV. CONCLUSION

In this work, we have seen, within a probability-theory framework, that there are minimum ergodic conditions which a sequence of single measurements must satisfy to serve as a point of contact between a physical theory and experiment. That is, the sequence must satisfy an ergodic theorem and be metrically transitive. Although these are much weaker conditions than the usually assumed independence and identical distribution, they indicate that, as is the case with quantum mechanics,[6,28] probability theory does not give a sufficiently complete framework for the description of the process of measurement or knowledge acquisition. For one thing, the question is left open of how an observer is to know that a sequence is ergodic. This question is particularly relevant to the description of the basic sequences of single measurements by which one tests for the homogeneity of space–time.

In spite of these difficulties, such a discussion, as has been given here, can help to clarify the basic problems of measurement. Also, it does suggest other approaches to the understanding of the basic dynamics of the knowledge acquisition process and how it can be relevant to physics.

## ACKNOWLEDGMENT

The author wishes to thank Dr. Jose Moyal for reading this manuscript and offering helpful suggestions.

[27] G. Ludwig, "Solved and Unsolved Problems of the Measurement Process", translated by E. Wilip from *Werner Heisenberg and Contemporary Physics*, F. Bopp, Ed. (F. Vieweg and Sohn, Braunschweig, 1961), pp. 150–181; H. Ekstein, Ergeb. Exakt. Naturwiss. **37**, 150 (1965); D. Bohm and J. Bubb, Rev. Mod. Phys. **38**, 453 (1966).

[28] M. M. Yanase, Am. J. Phys. **32**, 208 (1964).

# Structure of the Phonon Propagator for an Anharmonic Crystal

A. A. MARADUDIN
*Department of Physics, University of California, Irvine, California*
AND
I. P. IPATOVA
*A. F. Ioffe Physico-Technical Institute, Leningrad K-21, U.S.S.R.*

(Received 27 June 1966)

The phonon propagator for an arbitrary crystal is the analytic continuation to the complex $z$ plane of the Fourier coefficient of the imaginary time correlation function $D_{jj'}(\mathbf{k};u) = \langle TA_{\mathbf{k}j}(u)A^{+}_{\mathbf{k}j'}(0)\rangle$, where $A_{\mathbf{k}j}$ is the field operator for phonons with wavevector $\mathbf{k}$ and polarization or branch index $j$. Considering $\mathbf{D}(\mathbf{k};u)$ as a $3r \times 3r$ matrix whose elements are labeled by $j$ and $j'$ ($j,j' = 1, 2, \cdots, 3r$), where $r$ is the number of atoms in a primitive unit cell of the crystal, the restrictions imposed on the form of this matrix by the symmetry and structure of the crystal are determined here. In particular, it is proved that the element $D_{jj'}(\mathbf{k};u)$ vanishes unless $j$ and $j'$ label normal modes of vibration which transform according to the same row of the same irreducible multiplier representation of the point group of the wavevector $\mathbf{k}$, $G_0(\mathbf{k})$. As a corollary to this result it follows that if no two modes labeled by the wavevector $\mathbf{k}$ exist whose frequencies are different, but whose associated eigenvectors transform according to the same irreducible multiplier representation of $G_0(\mathbf{k})$, the matrix $D_{jj'}(\mathbf{k};u)$ is diagonal in $j$ and $j'$.

In a recent paper[1] it was shown that the phonon propagator $D_{jj'}(\mathbf{k};z)$ for an anharmonic crystal of the rocksalt structure is diagonal in the phonon branch indices $j$ and $j'$ when $\mathbf{k} = 0$. This result was obtained by showing that the proper self-energy matrix $P_{jj'}(\mathbf{k};z)$ has the same property and invoking the Dyson equation[2]

$$D^{-1}_{jj'}(\mathbf{k};z) = \delta_{jj'}D^{(0)-1}_{j}(\mathbf{k};z) - P_{jj'}(\mathbf{k};z) \qquad (1)$$

which relates the matrices $\mathbf{P}(\mathbf{k};z)$ and $\mathbf{D}(\mathbf{k};z)$. In this equation, $D^{(0)}_{j}(\mathbf{k};z)$ is the free-phonon propagator,

$$D^{(0)}_{j}(\mathbf{k};z) = \frac{2\omega_j(\mathbf{k})}{\beta\hbar}\frac{1}{\omega^2_j(\mathbf{k}) - z^2}, \qquad (2)$$

where $\omega_j(\mathbf{k})$ is the frequency of the normal mode of a harmonic crystal described by the wavevector $\mathbf{k}$ and the branch index $j$, and $\beta = (k_BT)^{-1}$. In Eqs. (1) and (2) $z$ is a complex variable.

That $P_{jj'}(\mathbf{0};z)$ is diagonal in $j$ and $j'$ was proven by a term-by-term examination of the perturbation series for the proper self-energy. This, however, is a rather unsatisfactory way of demonstrating what must be a consequence of the symmetry and structure of the crystal. In the present note we present a group theoretic analysis of the structure of the phonon propagator $D_{jj'}(\mathbf{k};z)$ as a matrix in the indices $j$ and $j'$ for an arbitrary crystal, when the wavevector $\mathbf{k}$ refers to one of the points of symmetry in the first Brillouin zone of the crystal, i.e., when the point group of the wavevector $\mathbf{k}$, $G_0(\mathbf{k})$ contains more than the identity. Because the type of calculation we describe here is somewhat similar to the determination of space group selection rules for two-phonon infrared and Raman processes, the methods employed here may be of use in the context of the latter problem as well.

The phonon propagator $D_{jj'}(\mathbf{k};z)$ is the analytic continuation to the complex $z$ plane of the Fourier coefficient[2]

$$D_{jj'}(\mathbf{k};i\omega_l) = \frac{1}{\beta}\int_0^{\beta} du\, e^{-i\hbar\omega_l u}D_{jj'}(\mathbf{k};u), \qquad \omega_l = 2\pi l/\beta\hbar, \quad (3)$$

of the imaginary time correlation function

$$D_{jj'}(\mathbf{k};u) = \langle TA_{\mathbf{k}j}(u)A^{+}_{\mathbf{k}j'}(0)\rangle, \quad -\beta < u < \beta. \quad (4)$$

In Eq. (4) $A_{\mathbf{k}j}$ is the phonon-field operator for the normal mode of the harmonic crystal described by the wavevector $\mathbf{k}$ and the branch index $j$. It is given in terms of the phonon creation and destruction operators $b^{+}_{\mathbf{k}j}$ and $b_{\mathbf{k}j}$ by

$$A_{\mathbf{k}j} = b_{\mathbf{k}j} + b^{+}_{-\mathbf{k}j} = A^{+}_{-\mathbf{k}j}. \qquad (5)$$

The operator $A_{\mathbf{k}j}(u)$ is defined by

$$A_{\mathbf{k}j}(u) = e^{uH}A_{\mathbf{k}j}(0)e^{-uH}, \qquad (6)$$

where $H$ is the crystal Hamiltonian. The operator $T$ orders a product of $u$ dependent operators from right to left in the order of increasing arguments. Finally, the angular brackets $\langle\cdots\rangle$ denote an average carried out with respect to the canonical ensemble described by the Hamiltonian $H$.

For our purposes it is convenient to rewrite the correlation function $D_{jj'}(\mathbf{k};u)$ in a different form. The relation between the displacement $\mathbf{u}(l\kappa)$ of the

[1] I. P. Ipatova, A. A. Maradudin, and R. F. Wallis, Fiz. Tverd. Tela **8**, 1064 (1966); [Sov. Phys.—Solid State **8**, 850 (1966)].
[2] A. A. Maradudin and A. E. Fein, Phys. Rev. **128**, 2589 (1962).

$\kappa$th atom in the $l$th unit cell of the crystal and the operator $A_{\mathbf{k}j}$ is given by

$$u_\alpha(l\kappa) = \left(\frac{\hbar}{2NM_\kappa}\right)^{\frac{1}{2}} \sum_{\mathbf{k}j} \frac{e_\alpha(\kappa \mid \mathbf{k}j)}{(\omega_j(\mathbf{k}))^{\frac{1}{2}}} e^{i\mathbf{k}\cdot\mathbf{x}(l)} A_{\mathbf{k}j}. \quad (7a)$$

Here, $\alpha$ labels the Cartesian axes, $N$ is the number of unit cells in the crystal, and $M_\kappa$ is the mass of the $\kappa$th kind of atom in a primitive unit cell. $\omega_j(\mathbf{k})$ is the frequency of the normal mode of the harmonic crystal which is described by the wavevector $\mathbf{k}$, and the phonon branch index $j$, which assumes the values $1, 2, \cdots, 3r$, where $r$ is the number of atoms in a primitive unit cell of the crystal. $\mathbf{e}(\kappa \mid \mathbf{k}j)$ is the associated unit polarization vector. The vector $\mathbf{x}(l)$ is the position vector of the origin of the $l$th unit cell in the crystal, and we will denote the position of the $\kappa$th ion in a primitive unit cell relative to the origin in the unit cell by $\mathbf{x}(\kappa)$. This origin must be chosen in such a way that the position vector of the atom $(l\kappa)$ is given by $\mathbf{x}(l) + \mathbf{x}(\kappa) \equiv \mathbf{x}(l\kappa)$. The allowed values of $\mathbf{k}$ are determined by the cyclic boundary condition, and they are uniformly distributed throughout the first Brillouin zone of the crystal with a density $V/(2\pi)^3$, where $V$ is the volume of the crystal.

The eigenvectors $\{\mathbf{e}(\kappa \mid kj)\}$ and the frequencies $\{\omega_j(\mathbf{k})\}$ are related through the eigenvalue equation[3]

$$\sum_{\kappa'\beta} C_{\alpha\beta}(\kappa\kappa' \mid \mathbf{k}) e_\beta(\kappa' \mid \mathbf{k}j) = \omega_j^2(\mathbf{k}) e_\alpha(\kappa \mid \mathbf{k}j), \quad (8)$$

where $\mathbf{C}(\mathbf{k})$ is a $3r \times 3r$ Hermitian matrix called the Fourier-transformed dynamical matrix. The explicit expression for the elements of $\mathbf{C}(\mathbf{k})$ will not be required in what follows.

The eigenvectors $\{\mathbf{e}(\kappa \mid kj)\}$ satisfy the orthonormality and closure conditions[4]

$$\sum_{\kappa\alpha} e_\alpha(\kappa \mid \mathbf{k}j) e_\alpha^*(\kappa \mid \mathbf{k}j') = \delta_{jj'}, \quad (9a)$$

$$\sum_{j} e_\alpha(\kappa \mid \mathbf{k}j) e_\beta^*(\kappa' \mid \mathbf{k}j) = \delta_{\kappa\kappa'}\delta_{\alpha\beta}. \quad (9b)$$

The relation inverse to Eq. (7) is readily found to be

$$A_{\mathbf{k}j} = \left(\frac{2\omega_j(\mathbf{k})}{\hbar N}\right)^{\frac{1}{2}} \sum_{l\kappa\alpha} (M_\kappa)^{\frac{1}{2}} e_\alpha^*(\kappa \mid \mathbf{k}j) e^{-i\mathbf{k}\cdot\mathbf{x}(l)} u_\alpha(l\kappa). \quad (7b)$$

With this result the correlation function $D_{jj'}(\mathbf{k}; u)$ can be written equivalently as

$$D_{jj'}(\mathbf{k}; u) = \frac{2}{\hbar N} [\omega_j(\mathbf{k})\omega_{j'}(\mathbf{k})]^{\frac{1}{2}} \sum_{ll'} \sum_{\kappa\kappa'} \sum_{\alpha\beta} (M_\kappa M_{\kappa'})^{\frac{1}{2}} \times e_\alpha^*(\kappa \mid \mathbf{k}j) e_\beta(\kappa' \mid \mathbf{k}j') e^{-i\mathbf{k}\cdot(\mathbf{x}(l)-\mathbf{x}(l'))} \times \langle T u_\alpha(l\kappa; u) u_\beta(l'\kappa'; 0)\rangle. \quad (10)$$

[3] See, for example, M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford University Press, Oxford, 1954), p. 297. Note that the matrix which we have denoted by $\mathbf{C}(\mathbf{k})$ is called $\mathbf{D}(\mathbf{k})$ by these authors.

[4] Reference 3, p. 298.

To determine the structure of $D_{jj'}(\mathbf{k}; z)$, regarded as a $3r \times 3r$ matrix in the branch indices $j$ and $j'$, it suffices to study the correlation function $D_{jj'}(\mathbf{k}; u)$. The operations of evaluating its Fourier coefficient and continuing the result to the complex $z$ plane cannot alter this structure.

An operation of the space group $G$ of the crystal can be written in the Seitz[5] notation as $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ and is defined by its effect on the position vector $\mathbf{x}(l\kappa)$,

$$\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}\mathbf{x}(l\kappa) = \mathbf{S}\mathbf{x}(l\kappa) + \mathbf{v}(S) + \mathbf{x}(m) = \mathbf{x}(LK). \quad (11)$$

The matrix $\mathbf{S}$ is a $3 \times 3$ real, orthogonal matrix which describes a proper or improper rotation. $\mathbf{x}(m)$ is a lattice translation vector, and $\mathbf{v}(S)$ is a translation through less than any primitive translation vector of the crystal. Space groups for which $\mathbf{v}(S)$ is zero for every rotation $\mathbf{S}$ are called symmorphic. The second equality in Eq. (11) expresses the fact that because the operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ is one which restores the crystal to itself the lattice site $(l\kappa)$ must be taken into an equivalent site which we label by $(LK)$. Where a more explicit notation is not required, we use the convention of labeling by capital letters the site into which a given site is transformed by the operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$.

With each operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ we associate a linear operator $O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})$ which is defined through its effect when applied to a scalar function of $\mathbf{x}(l\kappa)$:

$$O(\{\mathbf{S} \mid \mathbf{v}(S + \mathbf{x}(m)\})f(\mathbf{x}(l\kappa)) = f(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}^{-1}\mathbf{x}(l\kappa)), \quad (12a)$$

where we have explicitly that

$$\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}^{-1} = \{\mathbf{S}^{-1} \mid -\mathbf{S}^{-1}\mathbf{v}(S) - \mathbf{S}^{-1}\mathbf{x}(m)\}. \quad (12b)$$

When the crystal is subjected to a symmetry operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ which sends the lattice site $(l\kappa)$ into the site $(LK)$, the displacement vector $\mathbf{u}(l\kappa)$ associated with this site is both rotated in the same sense as the crystal and transferred to the site $(LK)$. Its law of transformation can therefore be expressed in the form

$$O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})\mathbf{u}(\mathbf{x}(l\kappa))O^{-1}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) = \mathbf{S}\mathbf{u}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}^{-1}\mathbf{x}(l\kappa)),$$

or, in view of Eq. (11), as

$$O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})\mathbf{u}(LK)O^{-1}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) = \mathbf{S}\mathbf{u}(l\kappa). \quad (13)$$

[5] F. Seitz, Ann. Math. 37, 17 (1936).

Because the symmetry operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ sends the crystal into itself, the crystal Hamiltonian must be invariant under the application of the operator $O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})$, i.e.,

$$O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})HO^{-1}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) = H. \tag{14}$$

In fact, this result together with Eq. (13) enables one to obtain the independent nonzero elements of the force constant tensors which appear when the crystal potential energy is expanded in powers of the atomic displacements.

Combining Eqs. (13) and (14) we obtain a useful result starting from the identity

$$\begin{aligned}\langle u_\alpha(l\kappa; u)u_\beta(l'\kappa'; 0)\rangle &= \langle u_\alpha(l\kappa; u)u_\beta(l'\kappa'; 0)O^{-1}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) \\ &\quad \times O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})\rangle \\ &= \langle O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})u_\alpha(l\kappa; u) \\ &\quad \times O^{-1}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) \\ &\quad \times O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})u_\beta(l'\kappa'; 0) \\ &\quad \times O^{-1}(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})\rangle \\ &= \sum_{\mu\nu} S_{\alpha\mu}S_{\beta\nu}\langle u_\mu(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}^{-1}\mathbf{x}(l\kappa); u) \\ &\quad \times u_\nu(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}^{-1}\mathbf{x}(l'\kappa'); 0)\rangle. \end{aligned} \tag{15}$$

The second equation is a consequence of the cyclic invariance of the trace, while the third is a consequence of the fact that according to Eq. (14), the operator $O(\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})$ commutes with the crystal Hamiltonian. Using Eq. (11) we can write finally that

$$\begin{aligned}\langle u_\alpha(LK; u)u_\beta(L'K'; 0)\rangle &= \sum_{\mu\nu} S_{\alpha\mu}S_{\beta\nu}\langle u_\mu(l\kappa; u)u_\nu(l'\kappa'; 0)\rangle. \end{aligned} \tag{16}$$

Before discussing the restrictions on the form of the correlation function $D_{jj'}(\mathbf{k}; u)$ imposed by the symmetry and structure of a given crystal, we establish several general properties of this function which hold for any crystal.

We note first, that because $\omega_j(\mathbf{k})$ and $e_\alpha(\kappa \mid \mathbf{k}j)$ are periodic functions of $\mathbf{k}$, with the periodicity of the reciprocal lattice,[6]

$$\omega_j(\mathbf{k} + 2\pi\mathbf{b}) = \omega_j(\mathbf{k}), \tag{17a}$$

$$\mathbf{e}(\kappa \mid \mathbf{k} + 2\pi\mathbf{b}j) = \mathbf{e}(\kappa \mid \mathbf{k}j), \tag{17b}$$

where $\mathbf{b}$ is an arbitrary translation vector of the reciprocal lattice, the correlation function $D_{jj'}(\mathbf{k}; u)$ also has the same periodicity in $\mathbf{k}$,

$$D_{jj'}(\mathbf{k} + 2\pi\mathbf{b}; u) = D_{jj'}(\mathbf{k}; u). \tag{18}$$

If we write $D_{jj'}(\mathbf{k}; u)$ as

$$\begin{aligned} D_{jj'}(\mathbf{k}; u) &= \theta(u)\frac{1}{Z}\sum_{mn} e^{-\beta E_m}e^{u(E_m-E_n)}\langle m| A_{\mathbf{k}j} |n\rangle \\ &\quad \times \langle n| A_{\mathbf{k}j'} |m\rangle \\ &\quad + \theta(-u)\frac{1}{Z}\sum_{mn} e^{-\beta E_m}e^{-u(E_m-E_n)} \\ &\quad \times \langle m| A_{-\mathbf{k}j'} |n\rangle\langle n| A_{\mathbf{k}j} |m\rangle, \end{aligned} \tag{19}$$

where $\theta(u)$ is the Heaviside unit step function, and where $E_m$ is the energy of the eigenstate $|m\rangle$ and $Z$ is the crystal partition function, the following two results follow immediately if we recall Eq. (5):

$$D_{jj'}(-\mathbf{k}; u) = D_{j'j}(\mathbf{k}; -u), \tag{20}$$

$$D^*_{jj'}(\mathbf{k}; u) = D_{j'j}(\mathbf{k}; u). \tag{21}$$

If we now invoke time-reversal symmetry and time-translation invariance, we obtain in addition that

$$\begin{aligned} D_{jj'}(\mathbf{k}; u) &= \langle TA^+_{\mathbf{k}j'}(0)A_{\mathbf{k}j}(-u)\rangle \\ &= \langle TA_{-\mathbf{k}j'}(u)A_{\mathbf{k}j}(0)\rangle \\ &= D_{j'j}(-\mathbf{k}; u). \end{aligned} \tag{22}$$

Combining Eqs. (20)–(22) we see that in general $D_{jj'}(\mathbf{k}; u)$ is an element of a Hermitian matrix and is an even function of $u$, with the property that

$$D_{jj'}(-\mathbf{k}; u) = D^*_{jj'}(\mathbf{k}; u). \tag{23}$$

In the special case that the point group of the crystal contains the inversion, i.e., when the crystal possesses a center of inversion, the matrix $D_{jj'}(\mathbf{k}; u)$ is not just Hermitian, it is real and symmetric. To show this, let us denote the inversion by $\mathbf{I}$, and label the site into which $(l\kappa)$ is taken by the space-group operation $\{\mathbf{I} \mid \mathbf{v}(I) + \mathbf{x}(m)\}$ by $(\bar{l}\bar{\kappa})$. Then with the conventional choice of phases[7,8]

$$e^*_\alpha(\bar{\kappa} \mid \mathbf{k}j) = e^{-i\mathbf{k}\cdot[\mathbf{x}(\bar{\kappa})+\mathbf{x}(\kappa)]}e_\alpha(\kappa \mid \mathbf{k}j), \tag{24}$$

together with the relation

$$\mathbf{I}\mathbf{x}(l) = \mathbf{x}(\bar{l}) + \mathbf{x}(\bar{\kappa}) - \{\mathbf{I} \mid \mathbf{v}(I) + \mathbf{x}(m)\}\mathbf{x}(\kappa), \tag{25}$$

which follows directly from Eq. (11), we obtain from Eqs. (10) and (16)

$$\begin{aligned} D_{jj'}(\mathbf{k}; u) &= \frac{2}{\hbar N}(\omega_j(\mathbf{k})\omega_{j'}(\mathbf{k}))^{\frac{1}{2}}\sum_{ll'}\sum_{\kappa\kappa'}\sum_{\alpha\beta}(M_{\bar{\kappa}}M_{\bar{\kappa}'})^{\frac{1}{2}} \\ &\quad \times e_\alpha(\bar{\kappa} \mid \mathbf{k}j)e^{-i\mathbf{k}\cdot[\mathbf{x}(\bar{\kappa})+\mathbf{x}(\kappa)]} \\ &\quad \times e^*_\beta(\bar{\kappa}' \mid \mathbf{k}j')e^{i\mathbf{k}\cdot[\mathbf{x}(\bar{\kappa}')+\mathbf{x}(\kappa')]} \\ &\quad \times e^{i\mathbf{k}\cdot[\mathbf{x}(\bar{l})+\mathbf{x}(\bar{\kappa})+\mathbf{x}(\kappa)-\mathbf{x}(\bar{l}')-\mathbf{x}(\bar{\kappa}')-\mathbf{x}(\kappa')]} \\ &\quad \times \langle Tu_\alpha(\bar{l}\bar{\kappa}; u)u_\beta(\bar{l}'\bar{\kappa}'; 0)\rangle. \end{aligned} \tag{26}$$

[6] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press Inc., New York, 1963), p. 77. See also Ref. 8 below.

[7] M. Lax, Phys. Rev. **138**, A793 (1965).

[8] A. A. Maradudin and S. H. Vosko, Rev. Mod. Phys. **40**, 1 (1968).

Replacing $(l\kappa)$ and $(l'\kappa')$ by $(\tilde{l}\tilde{\kappa})$ and $(\tilde{l}'\tilde{\kappa}')$ as summation variables we have finally that

$$D_{jj'}(\mathbf{k};u) = D_{jj'}(-\mathbf{k};u). \tag{27}$$

Combining Eqs. (23) and (27) we see that $D_{jj'}(\mathbf{k};u)$ is real for crystals possessing a center of inversion. From Eq. (21) it follows that $D_{jj'}(\mathbf{k};u)$ is symmetric in $j$ and $j'$ in this case. Finally, from Eq. (23) we find that $D_{jj'}(\mathbf{k};u)$ is an even function of $\mathbf{k}$ for such crystals (as well as being even in $u$).

We now turn to the determination of the restrictions imposed on $D_{jj'}(\mathbf{k};u)$ by the symmetry and structure of a given crystal.

When the result given by Eq. (16) is substituted into Eq. (10) the latter becomes

$$\begin{aligned} D_{jj'}(\mathbf{k};u) = {} & (2/\hbar N)(\omega_j(\mathbf{Sk})\omega_{j'}(\mathbf{Sk}))^{\frac{1}{2}} \\ & \times \sum_{ll'}\sum_{\kappa\kappa'}\sum_{\alpha\beta}\sum_{\mu\nu}(M_K M_{K'})^{\frac{1}{2}} e^*_\alpha(\kappa \mid \mathbf{k}j) e_\beta(\kappa' \mid \mathbf{k}j') \\ & \times e^{-i\mathbf{Sk}\cdot[\mathbf{Sx}(l)-\mathbf{Sx}(l')]}\langle Tu_\mu(LK;u)u_\nu(L'K';0)\rangle \\ & \times S_{\mu\alpha}S_{\nu\beta}. \end{aligned} \tag{28}$$

In writing this result we have used the fact that $\omega_j(\mathbf{k})$ considered as a function of $\mathbf{k}$ has the full point symmetry of the crystal,[8]

$$\omega_j(\mathbf{Sk}) = \omega_j(\mathbf{k}). \tag{29}$$

In addition, we have noted that because the symmetry operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ takes an atom of type $\kappa$ into an atom of type $K$, which must be the same kind of atom as $\kappa$, we must have that

$$M_\kappa = M_K. \tag{30}$$

If now we use the result, which follows from Eq. (11), and of which Eq. (25) is a special case, that

$$\mathbf{Sx}(l) = \mathbf{x}(L) + \mathbf{x}(K) - \{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}\mathbf{x}(\kappa), \tag{31}$$

we obtain

$$\begin{aligned} D_{jj'}(\mathbf{k};u) = {} & \frac{2}{\hbar N}[\omega_j(\mathbf{Sk})\omega_{j'}(\mathbf{Sk})]^{\frac{1}{2}} \sum_{ll'}\sum_{\kappa\kappa'}\sum_{\mu\nu}(M_\kappa M_{\kappa'})^{\frac{1}{2}} \\ & \times \left\{\sum_{\kappa_1\alpha} \mathcal{S}^*_{\mu\alpha}(\kappa\kappa_1 \mid \mathbf{k}; \{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) e^*_\alpha(\kappa_1 \mid \mathbf{k}j)\right\} \\ & \times \left\{\sum_{\kappa_2\beta} \mathcal{S}_{\nu\beta}(\kappa'\kappa_2 \mid \mathbf{k}; \{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) e_\beta(\kappa_2 \mid \mathbf{k}j')\right\} \\ & \times e^{-i\mathbf{Sk}\cdot[\mathbf{x}(l)-\mathbf{x}(l')]}\langle Tu_\mu(l\kappa;u)u_\nu(l'\kappa';0)\rangle. \end{aligned} \tag{32}$$

In writing this result we have introduced the $3r \times 3r$ matrix $\mathcal{S}(\mathbf{k};\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\})$ whose elements are given explicitly by

$$\begin{aligned} \mathcal{S}_{\alpha\beta}(\kappa\kappa' \mid \mathbf{k}; \{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) = {} & S_{\alpha\beta}\delta[\kappa, K_0(\kappa';S)] \\ & \times e^{i\mathbf{Sk}\cdot[\mathbf{x}(\kappa)-\{\mathbf{S} \mid \mathbf{v}(S)+\mathbf{x}(m)\}\mathbf{x}(\kappa')]}. \end{aligned} \tag{33}$$

Moreover, we have made more explicit the fact that $K$ is the label of the atom into which the atom $\kappa$ is carried by the space-group operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ by writing

$$K = K_0(\kappa;S). \tag{34}$$

That only the rotation $\mathbf{S}$ is required to relate $K$ to $\kappa$ uniquely is a consequence of the fact that a pure translation of the crystal leaves the basis label $\kappa$ of the site $(l\kappa)$ unchanged, and of the fact that the translation vector $\mathbf{v}(S)$ is uniquely specified once $\mathbf{S}$ is given.

It follows from Eq. (4.8) of Ref. 8 that as long as the operation $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ is not one of the operations of the space group of the wavevector $\mathbf{k}$, $G_\mathbf{k}$, i.e., as long as

$$\mathbf{Sk} \neq \mathbf{k} - 2\pi\mathbf{b}(\mathbf{k};\mathbf{S}), \tag{35}$$

where $\mathbf{b}(\mathbf{k};\mathbf{S})$ is a translation vector of the reciprocal lattice, then

$$\begin{aligned} & \left\{\sum_{\kappa_1\alpha} \mathcal{S}^*_{\mu\alpha}(\kappa\kappa_1 \mid \mathbf{k}; \{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) e^*_\alpha(\kappa_1 \mid \mathbf{k}j)\right\} \\ & \quad \times \left\{\sum_{\kappa_2\beta} \mathcal{S}_{\nu\beta}(\kappa'\kappa_2 \mid \mathbf{k}; \{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}) e_\beta(\kappa_2 \mid \mathbf{k}j')\right\} \\ & \quad = e^*_\mu(\kappa \mid \mathbf{Sk}j) e_\nu(\kappa' \mid \mathbf{Sk}j'). \end{aligned} \tag{36}$$

Combining Eqs. (32) and (36) we find that $D_{jj'}(\mathbf{k};u)$ has the point symmetry of the crystal,

$$D_{jj'}(\mathbf{k};u) = D_{jj'}(\mathbf{Sk};u), \quad \mathbf{Sk} \neq \mathbf{k} - 2\pi\mathbf{b}(\mathbf{k};\mathbf{S}). \tag{37}$$

Equation (37) tells us that if we know the form of $D_{jj'}(\mathbf{k};u)$ for a given value of $\mathbf{k}$, we know it for all values of the wavevector obtained by applying the operations of the point group of the crystal (i.e., of the crystal class) to $\mathbf{k}$. That is, we know $D_{jj'}(\mathbf{k};u)$ for all other wavevectors in the star of $\mathbf{k}$. It only remains to determine the structure of $D_{jj'}(\mathbf{k};u)$ at a given value of $\mathbf{k}$.

To achieve this we must do two things. The first is to restrict the operations $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$ of the space group $G$ of the crystal which appear in Eqs. (32) and (33) to those which comprise the subgroup of $G$ called the space group of the wavevector $\mathbf{k}$, $G_\mathbf{k}$. This latter group consists of all operations $\{\mathbf{R} \mid \mathbf{v}(R) + \mathbf{x}(m)\}$ of the space group $G$ whose rotational elements $\{\mathbf{R}\}$ have the property that they leave the vector $\mathbf{k}$ invariant, modulo $2\pi$ times a translation vector of the reciprocal lattice[9]:

$$\mathbf{Rk} = \mathbf{k} - 2\pi\mathbf{b}(\mathbf{k};\mathbf{R}). \tag{38}$$

This equation defines the reciprocal lattice vector

[9] To distinguish the rotational operations in the group $G_\mathbf{k}$ from those of the full group $G$, we denote the former by $\{\mathbf{R}\}$ and continue to denote the latter by $\{\mathbf{S}\}$.

$\mathbf{b}(\mathbf{k};\mathbf{R})$, which clearly is a function of both the given wavevector $\mathbf{k}$ and of the particular rotation $\mathbf{R}$ which we are considering. It should also be clear from the property of the first Brillouin zone for a crystal that no two points in it can differ by more than $2\pi$ times a translation vector of the reciprocal lattice, that $\mathbf{b}(\mathbf{k};\mathbf{R})$ can be nonzero only when $\mathbf{k}$ lies on the boundary of the zone.[10] Finally, we note that the collection of rotational elements $\{\mathbf{R}\}$ itself forms a group, which we call the point group of the wavevector $\mathbf{k}$ and denote by $G_0(\mathbf{k})$.[10] The order of this point group, which is one of the 32 crystallographic point groups, will be denoted by $h$.

When the symmetry operations appearing in the right-hand side of Eq. (32) are those of the group $G_{\mathbf{k}}$, this equation takes the form

$$D_{jj'}(\mathbf{k};u)=\frac{2}{\hbar N}[\omega_j(\mathbf{k})\omega_{j'}(\mathbf{k})]^{\frac{1}{2}}\sum_{ll'}\sum_{\kappa\kappa'}\sum_{\mu\nu}(M_\kappa M_{\kappa'})^{\frac{1}{2}}\times\left\{\sum_{\kappa_1\alpha}\mathcal{S}^*_{\mu\alpha}(\kappa\kappa_1\,|\,\mathbf{k};\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\})e^*_\alpha(\kappa_1\,|\,\mathbf{k}j)\right\}\times\left\{\sum_{\kappa_2\beta}\mathcal{S}_{\nu\beta}(\kappa'\kappa_2\,|\,\mathbf{k};\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\})e_\beta(\kappa_2\,|\,\mathbf{k}j')\right\}\times e^{-i\mathbf{k}\cdot[\mathbf{x}(l)-\mathbf{x}(l')]}\langle Tu_\mu(l\kappa;u)u_\nu(l'\kappa';0)\rangle, \tag{39}$$

where now

$$\mathcal{S}_{\alpha\beta}(\kappa\kappa'\,|\,\mathbf{k};\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\})=R_{\alpha\beta}\delta(\kappa,K_0(\kappa';R))\times e^{i\mathbf{k}\cdot[\mathbf{x}(\kappa)-\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\}\mathbf{x}(\kappa')]}. \tag{40}$$

The reason that the exponent on the right-hand side of Eq. (40) is as simple as it is, is that

$$\mathbf{x}(K')-\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\}\mathbf{x}(\kappa')$$

is a translation vector of the crystal, as can be seen directly from Eq. (31).

The set of $3r$-dimensional matrices $\mathcal{S}(\mathbf{k};\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\})$ can be shown to provide a representation of the group $G_{\mathbf{k}}$.[8,11] However, we will not exploit this property of these matrices. Instead, we prefer to work with a new set of $3r$-dimensional matrices $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$, which is obtained from the $\mathcal{S}(\mathbf{k};\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\})$ by the relation

$$T_{\alpha\beta}(\kappa\kappa'\,|\,\mathbf{k};\mathbf{R})=e^{i\mathbf{k}\cdot(\mathbf{v}(R)+\mathbf{x}(m))}\times\mathcal{S}_{\alpha\beta}(\kappa\kappa'\,|\,\mathbf{k};\{\mathbf{R}\,|\,\mathbf{v}(R)+\mathbf{x}(m)\})=R_{\alpha\beta}\delta(\kappa,K_0(\kappa';R))e^{i\mathbf{k}\cdot[\mathbf{x}(\kappa)-\mathbf{R}\mathbf{x}(\kappa')]}. \tag{41}$$

The matrices $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$ stand in a one-to-one correspondence with the elements of the point group $G_0(\mathbf{k})$. However, they do not provide an ordinary representation of this group, but rather provide a multiplier representation of it,[12,13] in the sense that the multiplication rule for any two matrices $\mathbf{T}(\mathbf{k};\mathbf{R}_i)$ and $\mathbf{T}(\mathbf{k};\mathbf{R}_j)$ is[8]

$$\mathbf{T}(\mathbf{k};\mathbf{R}_i)\mathbf{T}(\mathbf{k};\mathbf{R}_j)=\phi(\mathbf{k};\mathbf{R}_i,\mathbf{R}_j)\mathbf{T}(\mathbf{k};\mathbf{R}_i\mathbf{R}_j), \tag{42a}$$

$$\phi(\mathbf{k};\mathbf{R}_i,\mathbf{R}_j)=\exp 2\pi i\mathbf{b}(\mathbf{k};\mathbf{R}_i^{-1})\cdot\mathbf{v}(\mathbf{R}_j). \tag{42b}$$

We see from Eqs. (42) that the matrices $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$ provide an ordinary representation of $G_0(\mathbf{k})$ for an arbitrary crystal if $\mathbf{k}$ lies inside the first Brillouin zone, because then $\mathbf{b}(\mathbf{k};\mathbf{R}_i^{-1})$ vanishes. The matrices $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$ also provide an ordinary representation of $G_0(\mathbf{k})$ for all wavevectors $\mathbf{k}$ inside or on the boundary of the first Brillouin zone for a crystal with a symmorphic space group, because for such crystals $\mathbf{v}(R_j)$ vanishes for each operation $\mathbf{R}_j$.

Substituting Eq. (41) into Eq. (39) we obtain

$$D_{jj'}(\mathbf{k};u)=\frac{2}{\hbar N}(\omega_j(\mathbf{k})\omega_{j'}(\mathbf{k}))^{\frac{1}{2}}\sum_{ll'}\sum_{\kappa\kappa'}\sum_{\mu\nu}(M_\kappa M_{\kappa'})^{\frac{1}{2}}\times\left\{\sum_{\kappa_1\alpha}T^*_{\mu\alpha}(\kappa\kappa_1\,|\,\mathbf{k};\mathbf{R})e^*_\alpha(\kappa_1\,|\,\mathbf{k}j)\right\}\times\left\{\sum_{\kappa_2\beta}T_{\nu\beta}(\kappa'\kappa_2\,|\,\mathbf{k};\mathbf{R})e_\beta(\kappa_2\,|\,\mathbf{k}j')\right\}\times e^{-i\mathbf{k}\cdot[\mathbf{x}(l)-\mathbf{x}(l')]}\langle Tu_\mu(l\kappa;u)u_\nu(l'\kappa';0)\rangle. \tag{43}$$

The second thing we must do to determine the structure of $D_{jj'}(\mathbf{k};u)$ is to generalize our notation for labeling the branches of the phonon spectrum. Up to now it has been sufficient to label them by a single index $j$ $(=1,2,\cdots,3r)$ and to assign the $3r$ frequencies for a given $\mathbf{k}$ to these branches by a convention such as $\omega_j^2(\mathbf{k})\leq\omega_{j+1}^2(\mathbf{k})$. However, in order to proceed further in our discussion, we must label the branches of the phonon spectrum in a way that displays explicitly the possible degeneracies of the modes and their symmetry properties. Among the $3r$ branches of the phonon spectrum corresponding to a given value of $\mathbf{k}$ there may be several which are degenerate. To take account of this possibility we could replace $j$ by a double index $(\sigma\lambda)$, where $\sigma$ labels the *distinct* frequencies, while $\lambda$ differentiates among the $f_\sigma$ linearly independent eigenvectors $\mathbf{e}(\kappa\,|\,\mathbf{k}\sigma\lambda)$ $(\lambda=1,2,\cdots,f_\sigma)$ associated with the frequency $\omega_\sigma(\mathbf{k})$. Thus the normal mode labeled by $\mathbf{k}$ and $\sigma$ is assumed to be $f_\sigma$-fold degenerate.

The matrix $\mathbf{T}(\mathbf{k};\mathbf{R})$ can be shown[8] to commute with the Fourier transformed dynamical matrix $\mathbf{C}(\mathbf{k})$ for each operation $\mathbf{R}$ in the point group $G_0(\mathbf{k})$. From Eq. (8) we see that this fact has the consequence that

[10] G. F. Koster, in *Solid State Physics*, F. Seitz and D. Turnbull, Eds., (Academic Press Inc., New York, 1957), Vol. 5, p. 173; in particular, see p. 223.

[11] H. W. Streitwolf, Phys. Stat. Solidi **5**, 383 (1964).

[12] G. Ya. Liubarskii, *The Application of Group Theory in Physics* (Pergamon Press, Inc., New York, 1960), p. 95.

[13] P. Rudra, J. Math. Phys. **6**, 1273 (1965).

if the $3r$-component vector $\mathbf{e}(\mathbf{k}\sigma\lambda)$ is an eigenvector of $\mathbf{C}(\mathbf{k})$ corresponding to the eigenvalue $\omega_\sigma^2(\mathbf{k})$, then the vector $\mathbf{T}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma\lambda)$ is also an eigenvector of $\mathbf{C}(\mathbf{k})$ with the same eigenvalue $\omega_\sigma^2(\mathbf{k})$. However, because there are only $f_\sigma$ linearly independent eigenvectors $\{\mathbf{e}(\mathbf{k}\sigma\lambda)\}$ $(\lambda = 1, 2, \cdots, f_\sigma)$ corresponding to the eigenvalue $\omega_\sigma^2(\mathbf{k})$, the vector $\mathbf{T}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma\lambda)$ must be some linear combination of these $f_\sigma$ eigenvectors:

$$\mathbf{T}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma\lambda) = \sum_{\lambda'=1}^{f_\sigma} \tau_{\lambda'\lambda}^{(\sigma)}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma\lambda'). \tag{44}$$

The $hf_\sigma \times f_\sigma$ matrices $\{\boldsymbol{\tau}^{(\sigma)}(\mathbf{k};\mathbf{R})\}$ stand in a one-to-one correspondence with the operations $\{\mathbf{R}\}$ of the point group $G_0(\mathbf{k})$, and can be shown[8] to provide a unitary multiplier representation of this group. In fact, in the absence of accidental degeneracy the representation of $G_0(\mathbf{k})$ provided by the matrices $\{\boldsymbol{\tau}^{(\sigma)}(\mathbf{k};\mathbf{R})\}$ is irreducible.[14]

However, in the most general case the index $\sigma$ cannot be used to label the irreducible representations of $G_0(\mathbf{k})$ contained in the representation of this group provided by the matrices $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$. The index $\sigma$ has been introduced to label the distinct eigenvalues of $\mathbf{C}(\mathbf{k})$. It can happen that the eigenvectors corresponding to two distinct eigenvalues $\omega_\sigma^2(\mathbf{k})$ and $\omega_{\sigma'}^2(\mathbf{k})$ $(\sigma \neq \sigma')$ which have the same degeneracy, i.e., $f_\sigma = f_{\sigma'}$, are transformed into the same linear combinations of each other under multiplication by the matrix $\mathbf{T}(\mathbf{k};\mathbf{R})$. That is, we have the relations

$$\mathbf{T}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma\lambda_1) = \sum_{\lambda_1'=1}^{f_\sigma} \tau_{\lambda_1'\lambda_1}^{(\sigma)}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma\lambda_1'), \tag{45a}$$

$$\mathbf{T}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma'\lambda_2) = \sum_{\lambda_2'=1}^{f_{\sigma'}} \tau_{\lambda_2'\lambda_2}^{(\sigma')}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}\sigma'\lambda_2') \tag{45b}$$

for each $\mathbf{R}$ in $G_0(\mathbf{k})$, where $\boldsymbol{\tau}^{(\sigma)}(\mathbf{k};\mathbf{R}) = \boldsymbol{\tau}^{(\sigma')}(\mathbf{k};\mathbf{R})$. In other words, the two sets of eigenvectors $\{\mathbf{e}(\mathbf{k}\sigma\lambda)\}$ and $\{\mathbf{e}(\mathbf{k}\sigma'\lambda)\}$ associated with unequal frequencies $\omega_\sigma(\mathbf{k})$ and $\omega_{\sigma'}(\mathbf{k})$, respectively, transform according to the same irreducible multiplier representation of $G_0(\mathbf{k})$. To allow for such possibilities, we replace the single index $\sigma$ by a double index $(sa)$, where $s$ labels the irreducible representations of $G_0(\mathbf{k})$ contained in the representation $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$, while $a$ $(= 1, 2, \cdots, c_s)$ is a repetition index which distinguishes the different $f_s$-fold degenerate eigenvalues $\omega_{s\alpha}^2(\mathbf{k})$ whose associated eigenvectors $\{\mathbf{e}(\mathbf{k}sa\lambda)\}$ transform according to the $s$th irreducible representation. The numbers $f_s$ and $c_s$ must therefore satisfy the relation

$$\sum_s f_s c_s = 3r. \tag{46}$$

From the preceding discussion we are led to the conclusion that in its most general form Eq. (44) can be written as

$$\mathbf{T}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}sa\lambda) = \sum_{\lambda'=1}^{f_s} \tau_{\lambda'\lambda}^{(s)}(\mathbf{k};\mathbf{R})\mathbf{e}(\mathbf{k}sa\lambda'), \quad \lambda = 1, 2, \cdots, f_s, \quad a = 1, 2, \cdots, c_s. \tag{47}$$

The matrices $\boldsymbol{\tau}^{(s)}(\mathbf{k};\mathbf{R})$ have been tabulated by Kovalev[15] for $\mathbf{k}$ vectors corresponding to symmetry points in the first Brillouin zone for crystals belonging to all 230 space groups.

The determination of which irreducible multiplier representations of $G_0(\mathbf{k})$ are contained in the representation $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$ is carried out by the use of the character orthogonality theorem which yields the result that[13]

$$c_s = \frac{1}{h}\sum_{\mathbf{R}} \chi^{(s)}(\mathbf{k};\mathbf{R})^*\chi(\mathbf{k};\mathbf{R}) \tag{48}$$

where

$$\chi^{(s)}(\mathbf{k};\mathbf{R}) = \operatorname{Tr}\boldsymbol{\tau}^{(s)}(\mathbf{k};\mathbf{R}), \tag{48'}$$

$$\chi(\mathbf{k};\mathbf{R}) = \operatorname{Tr}\mathbf{T}(\mathbf{k};\mathbf{R}). \tag{48''}$$

Substituting Eq. (47) into Eq. (43), we obtain finally the conditions imposed on $D_{sa\lambda;s'a'\lambda'}(\mathbf{k};u)$ by the symmetry and structure of a crystal:

$$D_{sa\lambda;s'a'\lambda'}(\mathbf{k};u) = \sum_{\lambda_1\lambda_2} \tau_{\lambda_1\lambda}^{(s)}(\mathbf{k};\mathbf{R})^*\tau_{\lambda_2\lambda'}^{(s')}(\mathbf{k};\mathbf{R}) \times D_{sa\lambda_1;s'a'\lambda_2}(\mathbf{k};u). \tag{49}$$

When the representation matrices $\{\boldsymbol{\tau}^{(s)}(\mathbf{k};\mathbf{R})\}$ of the irreducible multiplier representations $s$ and $s'$ of $G_0(\mathbf{k})$ contained in the representation $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$ are substituted into the right-hand side of Eq. (49) for each of the operations $\mathbf{R}$ of $G_0(\mathbf{k})$, the resulting equations determine the independent nonzero elements of the matrix $D_{sa\lambda;s'a'\lambda'}(\mathbf{k};u)$ and any relations among them.

Somewhat less detailed information about the structure of this matrix is obtained if we divide both sides of Eq. (49) by $h$, the order of $G_0(\mathbf{k})$, and then sum both sides over the elements $\mathbf{R}$ of $G_0(\mathbf{k})$. The orthogonality of the representation matrices expressed

[14] V. Heine, *Group Theory in Quantum Mechanics* (Pergamon Press, Inc., New York, 1960), p. 44.

[15] O. V. Kovalev, *Irreducible Representations of the Space Groups* (Academy of Sciences of the Ukrainian S.S.R., Kiev, 1961) [English transl.: Gordon and Breach Science Publishers, Inc., New York, 1964]. It has been pointed out to one of the authors (A.A.M.) by Dr. J. Zak that the irreducible representations associated with severak $\mathbf{k}$ vectors corresponding to symmetry points on the boundary of the first Brillouin zone have been omitted from Kovalev's tables. For example, the symmetry points for the cubic system missing from Kovalev's book are (1) simple cubic: $(k_x, \pi/a, k_z)$, $(k_x, \pi/a, 0)$, $(\pi/a, k_y, k_z)$; (2) face-centered cubic: $(k_x, 2\pi/a, k_z)$, $(k_x, 2\pi/a, k_z)$, $(\pi/a, k_y, (2\pi/a) - k_y)$; (3) body centered cubic: $(\pi/a, \pi/a, 0)$, $(k_x, (2\pi/a) - k_x, 0)$, where $a$ is the lattice parameter. These omissions will be rectified in a forthcoming book by Casher, Gluck, Gur, and Zak.

by[13]

$$\sum_{\mathbf{R}} \tau^{(s)}_{\lambda_1\lambda}(\mathbf{k};\mathbf{R})^* \tau^{(s')}_{\lambda_2\lambda'}(\mathbf{k};\mathbf{R}) = (h/f_s)\delta_{ss'}\delta_{\lambda_1\lambda_2}\delta_{\lambda\lambda'} \quad (50)$$

yields the result that

$$D_{sa\lambda;s'a'\lambda'}(\mathbf{k};u) = \delta_{ss'}\delta_{\lambda\lambda'}(1/f_s)\sum_{\lambda_1} D_{sa\lambda_1;sa'\lambda_1}(\mathbf{k};u). \quad (51)$$

We can express this result in the conventional notation by the statement that unless $j$ and $j'$ label modes which transform according to the same row of the same irreducible representation of $G_0(\mathbf{k})$, the matrix element $D_{jj'}(\mathbf{k};u)$ vanishes. As a corollary to this result we also see from Eq. (51) that if no irreducible multiplier representation of $G_0(\mathbf{k})$ appears more than once in the reduction of the representation $\{\mathbf{T}(\mathbf{k};\mathbf{R})\}$ (in which case we can suppress the repetition indices $a$ and $a'$), then $D_{jj'}(\mathbf{k};u)$ is diagonal in the indices $j$ and $j'$.

It should be pointed out that the results given by Eqs. (49) and (51) hold for any value of $\mathbf{k}$ inside or on the boundary of the first Brillouin zone of the crystal. However, at a general point of the zone the point group of the wavevector $\mathbf{k}$, $G_0(\mathbf{k})$ consists of only the identity $\epsilon$. There is only one irreducible representation of this group, and it is one-dimensional with $\chi^{(1)}(\mathbf{k};\epsilon) = 1$. Meanwhile $\boldsymbol{\chi}(\mathbf{k};\mathbf{R})$ is $3r$, so that according to Eq. (48) $c_1 = 3r$. From Eqs. (49) and (51) we see that there is therefore no simplification in the structure of the matrix $\mathbf{D}(\mathbf{k})$ which is required by symmetry for a general value of $\mathbf{k}$. The results given by Eqs. (49) and (51) can predict simplifications in the structure of $\mathbf{D}(\mathbf{k})$ only when $\mathbf{k}$ is a point of symmetry inside or on the boundary of the first Brillouin zone, i.e., when $\mathbf{k}$ is a point for which $G_0(\mathbf{k})$ consists of more than the identity.

There is a final degeneracy of $D_{jj'}(\mathbf{k};z)$ which is not predicted by the preceding treatment. This is that for $z \neq 0$, $D_{jj'}(\mathbf{0};z)$ vanishes if either $j$ or $j'$, or both, refer to an acoustic branch of the phonon spectrum. This conclusion follows from the fact that as a consequence of infinitesimal translational symmetry $\omega_j(\mathbf{0})$ vanishes if $j$ refers to one of the three acoustic branches.[16] As a result of this, $D_j^{(0)}(\mathbf{0};z)$ vanishes for $z \neq 0$, for $j$ an acoustic branch. From the Dyson equation (1) it is seen that if $P_{jj'}(\mathbf{0};z)$ is finite, the vanishing of $D_j^{(0)}(\mathbf{0};z)$ implies the vanishing of $D_{jj'}(\mathbf{0};z)$. In fact it is shown in Appendix A of Ref. 1 that infinitesimal translational invariance forces $P_{jj'}(\mathbf{0},z)$ to vanish if either $j$ or $j'$, or both, refer to an acoustic branch, from which our original statement follows.

To conclude this note we return to the problem which prompted the investigation described in it. We consider the structure of the phonon propagator corresponding to the wavevector $\mathbf{k} = 0$ for a crystal of the rocksalt structure. If we neglect for the moment the macroscopic electric field associated with the longitudinal optical modes of long wavelengths, the point group of this wavevector is $O_h$. Because no operation of the space group of the rocksalt structure $O_h^5$ can interchange the two sublattices, the matrix element $T_{\alpha\beta}(\kappa\kappa' \mid \mathbf{0};\mathbf{R})$ takes the simple form in the present case

$$T_{\alpha\beta}(\kappa\kappa' \mid \mathbf{0};\mathbf{R}) = R_{\alpha\beta}\delta(\kappa,\kappa'). \quad (52)$$

The matrices $\{\mathbf{T}(\mathbf{0};\mathbf{R})\}$ clearly provide an ordinary representation of the point group $O_h$. From Eq. (52) it follows that

$$\chi(\mathbf{0};\mathbf{R}) = 2\sum_{\alpha} R_{\alpha\alpha} = 2(2\cos\phi \pm 1), \quad (53)$$

where $\phi$ is the angle through which the rotation described by the matrix $\mathbf{R}$ is carried out, while the $+$ sign applies if $\mathbf{R}$ describes a proper rotation and the $-$ sign applies if the rotation is improper. By the use of Eq. (53) we obtain the following character table:

| $\mathbf{R}$ | $E$ | $8C_3$ | $3C_2$ | $6C_4$ | $6C_2'$ | $I$ | $8S_6$ | $3\sigma_h$ | $6S_4$ | $6\sigma_d$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\chi(\mathbf{0};\mathbf{R})$ | 6 | 0 | −2 | 2 | −2 | −6 | 0 | 2 | −2 | 2 |

The reduction of the representation of $G_0(\mathbf{0})$ provided by the matrices $\{\mathbf{T}(\mathbf{0};\mathbf{R})\}$ yields $2\Gamma_{15}$, where $\Gamma_{15}$ is a conventional name for the (three-dimensional) polar-vector irreducible representation of the point group $O_h$.[10] One irreducible representation $\Gamma_{15}$ clearly corresponds to the three acoustic modes, while the second corresponds to the three optical modes. The fact that $\Gamma_{15}$ appears twice in this reduction ordinarily would imply that $D_{jj'}(\mathbf{0};z)$ is not diagonal in $j$ and $j'$, according to the discussion following Eq. (51). However, in the present case the diagonal terms of $D_{jj'}(\mathbf{0};z)$ for which $j$ refers to an acoustic branch, and the off-diagonal terms for which either $j$ or $j'$ refers to an acoustic branch vanish. We therefore find that the only nonzero elements of $D_{jj'}(\mathbf{0};z)$ are the three diagonal elements $D_{jj}(\mathbf{0};z)$ where $j$ refers to one of the three optical branches. Moreover, because $\Gamma_{15}$ is a three-dimensional irreducible representation,

[16] Reference 6, p. 13.

these three nonzero diagonal elements are all equal. Thus we have reached the conclusion that for an ionic crystal of the rocksalt structure, the phonon propagator $D_{jj'}(\mathbf{0}; z)$ is diagonal in $j$ and $j'$, that the only nonzero diagonal elements are the three associated with the optical branches, and that they are all equal.

However, this conclusion is not correct because we have ignored the macroscopic electric field which accompanies the long-wavelength longitudinal optical vibrations. It is now well known[17] that this field splits the triple degeneracy of the optical modes at $\mathbf{k} = 0$ by raising the frequency of the longitudinal mode above the frequency of the (now) doubly degenerate transverse optical modes. The degeneracy of the acoustic modes at $\mathbf{k} = 0$ is not lifted by the macroscopic electric field, and $D_{jj'}(\mathbf{0}; z)$ still vanishes if $j$ and $j'$, or both, refers to an acoustic branch. Therefore, $D_{jj'}(\mathbf{0}; z)$ remains diagonal in $j$ and $j'$, and the only nonzero diagonal elements are still those associated with the optical branches. However, now the two diagonal elements associated with the doubly degenerate transverse optical modes are equal, and are different from the diagonal element which is associated with the longitudinal optical mode. This is the result obtained in Ref. 1.

It is the particularly simple form of the phonon-propagator matrix at $\mathbf{k} = 0$ in ionic crystals of the rocksalt structure which enables the theory of the fundamental lattice vibration absorption in such crystals to be developed free of the algebraic complications which would otherwise arise from the necessity of having to solve a matrix Dyson equation for the propagator.[1]

[17] Reference 3, Sec. 7.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Zero-Mass Representations of the Poincaré Group in an $O(3, 1)$ Basis*

N. MUKUNDA†
*Physics Department, Syracuse University, Syracuse, New York*

(Received 5 May 1967)

Unitary irreducible representations of the Poincaré group, corresponding to zero mass and finite helicity, are reduced with respect to the subgroup of homogeneous Lorentz transformations. The action of the energy–momentum operators in the basis suited to this reduction is examined.

## INTRODUCTION

Several recent papers in elementary particle physics have considered in one way or another the possibility of embedding the Poincaré group in a larger noncompact group.[1] By working within one representation of the larger group, one may be able to correlate several different representations of the Poincaré group in a useful way. Though the immediate physical motivation in each case may be different, it would be useful to know in general how to decompose unitary representations of such a large group into parts irreducible under the Poincaré group. This is part of the general problem of reducing the unitary representations of a given noncompact group into unitary irreducible representations (UIR's) of a noncompact subgroup, and of trying to understand the representation of the whole group expressed in the basis suited to this reduction.

In previous papers, we have considered the reduction of unitary representations of $O(2, 1)$ and $O(3, 1)$ with respect to the subgroups $O(1, 1)$ and $O(2, 1)$, respectively.[2] We have also examined some of the properties of the generators of the whole group when they act on irreducible representations of the relevant subgroup. In the present note, we examine the UIR's of the Poincaré group corresponding to vanishing mass and finite helicity, and carry out the reduction of these UIR's into UIR's of the homogeneous Lorentz group $O(3, 1)$. In particular, we examine the nature of the generators in the basis made up of UIR's of $O(3, 1)$.

* Work supported in part by the U.S. Atomic Energy Commission.

† Present address: Tata Institute of Fundamental Research, Bombay, India.

[1] Representative of the varied nature of the possibilities are: C. Fronsdal, in *Proceedings of the Third Coral Gables Conference on Symmetry Principles at High Energy* (W. H. Freeman & Company, San Francisco, 1966); H. Ruegg, W. Ruhl, and T. S. Santhanam, Helv. Phys. Acta **40**, 9 (1967); E. H. Hoffmann, Phys. Rev. Letters **16**, 210 (1966), and Commun. Math. Phys. **4**, 237 (1967); Y. Nambu, Progr. Theoret. Phys. **37**, 368 (1966), and Phys. Rev. **160**, 1171 (1967); N. Mukunda, E. C. G. Sudarshan, and A. Bohm, Phys. Letters **24B**, 301 (1967).

[2] N. Mukunda, J. Math. Phys. **8**, 2210 (1967); **9**, 50 (1968). We refer to these as (A) and (B), respectively.

these three nonzero diagonal elements are all equal. Thus we have reached the conclusion that for an ionic crystal of the rocksalt structure, the phonon propagator $D_{jj'}(\mathbf{0}; z)$ is diagonal in $j$ and $j'$, that the only nonzero diagonal elements are the three associated with the optical branches, and that they are all equal.

However, this conclusion is not correct because we have ignored the macroscopic electric field which accompanies the long-wavelength longitudinal optical vibrations. It is now well known[17] that this field splits the triple degeneracy of the optical modes at $\mathbf{k} = 0$ by raising the frequency of the longitudinal mode above the frequency of the (now) doubly degenerate transverse optical modes. The degeneracy of the acoustic modes at $\mathbf{k} = 0$ is not lifted by the macroscopic electric field, and $D_{jj'}(\mathbf{0}; z)$ still vanishes if $j$ and $j'$, or both, refers to an acoustic branch. Therefore, $D_{jj'}(\mathbf{0}; z)$ remains diagonal in $j$ and $j'$, and the only nonzero diagonal elements are still those associated with the optical branches. However, now the two diagonal elements associated with the doubly degenerate transverse optical modes are equal, and are different from the diagonal element which is associated with the longitudinal optical mode. This is the result obtained in Ref. 1.

It is the particularly simple form of the phonon-propagator matrix at $\mathbf{k} = 0$ in ionic crystals of the rocksalt structure which enables the theory of the fundamental lattice vibration absorption in such crystals to be developed free of the algebraic complications which would otherwise arise from the necessity of having to solve a matrix Dyson equation for the propagator.[1]

[17] Reference 3, Sec. 7.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Zero-Mass Representations of the Poincaré Group in an $O(3, 1)$ Basis*

N. MUKUNDA†
*Physics Department, Syracuse University, Syracuse, New York*

(Received 5 May 1967)

Unitary irreducible representations of the Poincaré group, corresponding to zero mass and finite helicity, are reduced with respect to the subgroup of homogeneous Lorentz transformations. The action of the energy–momentum operators in the basis suited to this reduction is examined.

## INTRODUCTION

Several recent papers in elementary particle physics have considered in one way or another the possibility of embedding the Poincaré group in a larger noncompact group.[1] By working within one representation of the larger group, one may be able to correlate several different representations of the Poincaré group in a useful way. Though the immediate physical motivation in each case may be different, it would be useful to know in general how to decompose unitary representations of such a large group into parts irreducible under the Poincaré group. This is part of the general problem of reducing the unitary representations of a given noncompact group into unitary irreducible representations (UIR's) of a noncompact subgroup, and of trying to understand the representation of the whole group expressed in the basis suited to this reduction.

In previous papers, we have considered the reduction of unitary representations of $O(2, 1)$ and $O(3, 1)$ with respect to the subgroups $O(1, 1)$ and $O(2, 1)$, respectively.[2] We have also examined some of the properties of the generators of the whole group when they act on irreducible representations of the relevant subgroup. In the present note, we examine the UIR's of the Poincaré group corresponding to vanishing mass and finite helicity, and carry out the reduction of these UIR's into UIR's of the homogeneous Lorentz group $O(3, 1)$. In particular, we examine the nature of the generators in the basis made up of UIR's of $O(3, 1)$.

* Work supported in part by the U.S. Atomic Energy Commission.

† Present address: Tata Institute of Fundamental Research, Bombay, India.

[1] Representative of the varied nature of the possibilities are: C. Fronsdal, in *Proceedings of the Third Coral Gables Conference on Symmetry Principles at High Energy* (W. H. Freeman & Company, San Francisco, 1966); H. Ruegg, W. Ruhl, and T. S. Santhanam, Helv. Phys. Acta **40**, 9 (1967); E. H. Hoffmann, Phys. Rev. Letters **16**, 210 (1966), and Commun. Math. Phys. **4**, 237 (1967); Y. Nambu, Progr. Theoret. Phys. **37**, 368 (1966), and Phys. Rev. **160**, 1171 (1967); N. Mukunda, E. C. G. Sudarshan, and A. Bohm, Phys. Letters **24B**, 301 (1967).

[2] N. Mukunda, J. Math. Phys. **8**, 2210 (1967); **9**, 50 (1968). We refer to these as (A) and (B), respectively.

It should be pointed out that reductions of UIR's of the Poincaré group with respect to $O(3, 1)$ have been carried out in the past.[3] However, it appears that the nature of the generators of the Poincaré group in this basis has not been considered previously.

In Sec. 1, we recall the form of the principal series of UIR's of the homogeneous Lorentz group $O(3, 1)$. These are the only UIR's with which we are concerned here. Section 2 contains the relevant material on the UIR's of the Poincaré group corresponding to zero mass and finite helicity (lightlike representations). In Sec. 3, we carry out the reduction of these UIR's of the Poincaré group into UIR's of $O(3, 1)$. Lastly, in Sec. 4, we consider the action of the generators of the Poincaré group in the $O(3, 1)$ basis.

## 1. PRINCIPAL SERIES OF UIR'S OF $O(3, 1)$

As is well known, the homogeneous Lorentz group $O(3, 1)$ has for it's covering group the group $SL(2, C)$ of all complex unimodular matrices in two dimensions.[4] The Lie algebra of $O(3, 1)$ [equivalently, of $SL(2, C)$] is spanned by the six elements $J_j$, $K_j$ $(j = 1, 2, 3)$ obeying the commutation relations

$$\begin{aligned} [J_j, J_k] &= i\epsilon_{jkl}J_l, \\ [J_j, K_k] &= i\epsilon_{jkl}K_l, \\ [K_j, K_k] &= -i\epsilon_{jkl}J_l, \end{aligned} \tag{1.1}$$

$J_j$ are the generators of spatial rotations while the $K_j$ generate pure Lorentz transformations. There are two Casimir invariants for $O(3, 1)$,

$$\mathcal{C}_1 = K_jK_j - J_jJ_j, \quad \mathcal{C}_2 = K_jJ_j. \tag{1.2}$$

In a UIR of $O(3, 1)$, $J_m$ and $K_j$ are self-adjoint operators. There are two kinds of UIR's of $O(3, 1)$, namely the principal series, and the supplementary series.[5] Here, we only need the former; UIR's of the principal series may be labeled in the form $\{j_0, \rho\}$; $j_0$ is a nonnegative integer or half-odd integer, while $\rho$ is any real number. Different pairs $\{j_0, \rho\}$ denote inequivalent UIR's. $\mathcal{C}_1$ and $\mathcal{C}_2$ may be expressed as

$$\mathcal{C}_1 = 1 + \rho^2 - j_0^2, \quad \mathcal{C}_2 = \rho j_0. \tag{1.3}$$

The explicit structure of the UIR $\{j_0, \rho\}$ may be exhibited by introducing a basis in Hilbert space made up of orthonormal eigenvectors of $J^2$ and $J_3$. In any UIR of $O(3, 1)$, each finite-dimensional UIR of the compact subgroup $O(3)$ generated by $J_j$ can appear no more than once. The members of this orthonormal basis are $|j, m\rangle$ and obey

$$\begin{aligned} \langle j'm' \,|\, jm\rangle &= \delta_{j'j}\delta_{m'm}, \\ J^2|jm\rangle = j(j+1)|jm\rangle, &\quad J_3|jm\rangle = m|jm\rangle. \end{aligned} \tag{1.4}$$

In the UIR $\{j_0, \rho\}$, $j$ takes on the values

$$j = j_0, j_0 + 1, j_0 + 2, \cdots, \infty \tag{1.5}$$

while, of course, for each $j$, $m = -j, -j + 1, \cdots, j$. Introducing the spherical components $J_m$, $K_m$, $m = +1, 0, -1$, of the generators, their matrix elements are given by

$$\begin{aligned} \langle j'm' |\, J_M \,| jm\rangle &= \delta_{j'j}(j(j+1))^{\frac{1}{2}} C^{j\ 1\ j'}_{m\ M\ m'}, \\ \langle j'm' |\, K_M \,| jm\rangle &= \langle j' \,\|\, K \,\|\, j\rangle\, C^{j\ 1\ j'}_{m\ M\ m'}, \end{aligned}$$

$$\begin{aligned} &\langle j+1 \,\|\, K \,\|\, j\rangle \\ &= -i[((j+1)^2 - j_0^2)((j+1)^2 + \rho^2)/(j+1)(2j+3)]^{\frac{1}{2}}, \\ &\langle j \,\|\, K \,\|\, j\rangle = \rho j_0[j(j+1)]^{-\frac{1}{2}}, \\ &\langle j-1 \,\|\, K \,\|\, j\rangle = -i[(j^2 - j_0^2)(j_0^2 + \rho^2)/j(2j-1)]^{\frac{1}{2}}. \end{aligned} \tag{1.6}$$

## 2. ZERO-MASS REPRESENTATIONS OF THE POINCARÉ GROUP

The ten generators of the Poincaré group $J_j$, $K_j$, $P_j$, $H$ obey the commutation rules (1.1) and the following:

$$\begin{aligned} [J_j, P_k] &= i\epsilon_{jkl}P_l, \quad [J_j, H] = 0, \\ [K_j, P_k] &= i\delta_{jk}H, \quad [K_j, H] = iP_j, \\ [P_j, P_k] &= [P_j, H] = 0. \end{aligned} \tag{2.1}$$

Unitary irreducible representations of the Poincaré group

$$H^2 = P_jP_j, \tag{2.2}$$

corresponding to zero mass, finite helicity, and positive energy may be labeled by the helicity $s$ of the representation. Here, $s$ is an integer or half-odd integer, either positive or negative or zero. Such UIR's are most easily constructed in a Hilbert space $\mathcal{H}$ of momentum wavefunctions in which the generators $H$ and $P_j$ are diagonal.[6] Vectors $f$ in $\mathcal{H}$ correspond to complex-valued functions $f(\mathbf{p})$ of a three-dimensional vector $\mathbf{p}$, with the inner product $(f, g)$ and norm

[3] I. S. Shapiro, Sov. Phys —Doklady **1**, 91 (1956); Chou Kuang-Chao and L. G. Zastavenko, Zh. Eksperim. i Teor. Fiz. **35**, 1417 (1958) [Soviet Phys.—JETP **8**, 990 (1959)].

[4] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (The Macmillan Company, New York, 1963).

[5] See, for instance, Ref. 4, p. 200.

[6] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962).

$\|f\|$ being given by

$$(f, g) = \int \frac{d^3\mathbf{p}}{\omega} f^*(\mathbf{p}) g(\mathbf{p}), \quad \omega = |\mathbf{p}|,$$
$$\|f\| = (f,f)^{\frac{1}{2}} < \infty. \tag{2.3}$$

Let us denote the Cartesian components of $\mathbf{p}$ by $p_j$, $j = 1, 2, 3$. Then the generators are represented as follows:

$$\begin{aligned}
J_1 &= -i\left(p_2 \frac{\partial}{\partial p_3} - p_3 \frac{\partial}{\partial p_2}\right) + \frac{s p_1}{(\omega + p_3)}, \\
J_2 &= -i\left(p_3 \frac{\partial}{\partial p_1} - p_1 \frac{\partial}{\partial p_3}\right) + \frac{s p_2}{(\omega + p_3)}, \\
J_3 &= -i\left(p_1 \frac{\partial}{\partial p_2} - p_2 \frac{\partial}{\partial p_1}\right) + s, \\
K_1 &= -i\omega \frac{\partial}{\partial p_1} - \frac{s p_2}{(\omega + p_3)}, \\
K_2 &= -i\omega \frac{\partial}{\partial p_2} + \frac{s p_1}{(\omega + p_3)}, \\
K_3 &= -i\omega \partial / \partial p_3, \\
P_j &= p_j, \quad H = \omega.
\end{aligned} \tag{2.4}$$

The effect of a finite Poincaré transformation on a wavefunction $f(\mathbf{p})$ has been given by Moses.[7]

## 3. REDUCTION WITH RESPECT TO $O(3,1)$

We wish to reduce the representations of the Poincaré group given in Sec. 2 into irreducible representations of $O(3, 1)$.[8] [For definiteness, we will consider the case with nonnegative integral $s$.] To this end, we change from the Cartesian variable $p_j$ to spherical-polar coordinates $\omega$, $\theta$, $\varphi$:

$$\begin{aligned}
p_1 = \omega \sin\theta \cos\varphi, \quad p_2 = \omega \sin\theta \sin\varphi, \\
p_3 = \omega \cos\theta,
\end{aligned} \tag{3.1}$$

and simultaneously define a new wavefunction $\tilde{f}(\omega, \theta, \varphi)$ to represent the vector $f$ which was previously represented by the wavefunction $f(\mathbf{p})$:

$$f \to \tilde{f}(\omega, \theta, \varphi) = e^{is\varphi} f(\mathbf{p}). \tag{3.2}$$

Then the scalar product becomes

$$(f, g) = \int_0^\infty \omega \, d\omega \int_0^{2\pi} d\varphi \int_0^\pi \sin\theta \, d\theta \, \tilde{f}^*(\omega, \theta, \varphi) \tilde{g}(\omega, \theta, \varphi) \tag{3.3}$$

and the effect of the generators on the functions $f(\omega, \theta, \varphi)$ is given by

$$\begin{aligned}
J_1 &= i \sin\varphi \frac{\partial}{\partial\theta} + i \cos\varphi \cot\theta \frac{\partial}{\partial\varphi} + s \frac{\cos\varphi}{\sin\theta}, \\
J_2 &= -i \cos\varphi \frac{\partial}{\partial\theta} + i \sin\varphi \cot\theta \frac{\partial}{\partial\varphi} + s \frac{\sin\varphi}{\sin\theta}, \\
J_3 &= -i \frac{\partial}{\partial\varphi}, \\
K_1 &= -i \sin\theta \cos\varphi \omega \frac{\partial}{\partial\omega} - i \cos\theta \cos\varphi \frac{\partial}{\partial\theta} \\
&\quad + i \frac{\sin\varphi}{\sin\theta} \frac{\partial}{\partial\varphi} + s \sin\varphi \cot\theta, \\
K_2 &= -i \sin\theta \sin\varphi \omega \frac{\partial}{\partial\omega} - i \cos\theta \sin\varphi \frac{\partial}{\partial\theta} \\
&\quad - i \frac{\cos\varphi}{\sin\theta} \frac{\partial}{\partial\varphi} - s \cos\varphi \cot\theta, \\
K_3 &= -i \cos\theta \omega \frac{\partial}{\partial\omega} + i \sin\theta \frac{\partial}{\partial\theta}, \\
P_1 &= \omega \sin\theta \cos\varphi, \quad P_2 = \omega \sin\theta \sin\varphi, \\
P_3 &= \omega \cos\theta, \quad H = \omega.
\end{aligned} \tag{3.4}$$

The structure of the generators $J_j$, $K_j$ above is almost exactly that which was obtained for them in (B); there, we were dealing with UIR's of $O(3, 1)$ of the principal series realized in a Hilbert space of functions on the unit sphere. To cast $J_j$ and $K_j$ into exactly the same form as in (B), we essentially have to perform a Fourier transformation with respect to the variable $\ln \omega$. We carry this out as follows. Let $x$ be related to $\omega$ by

$$\omega = e^x \tag{3.5}$$

so that $0 \leq \omega < \infty$ corresponds to $-\infty < x < \infty$. We first replace the wavefunction $\tilde{f}(\omega, \theta, \varphi)$ by a new one, $f_1(x, \theta, \varphi)$:

$$f_1(x, \theta, \varphi) = \omega \tilde{f}(\omega, \theta, \varphi), \tag{3.6}$$

and then express $f_1(x, \theta, \varphi)$ as a Fourier transform of a function $f_2(\rho, \theta, \varphi)$:

$$f_1(x, \theta, \varphi) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{i\rho x} f_2(\rho, \theta, \varphi) \, d\rho. \tag{3.7}$$

Thus all the functions $f(\mathbf{p})$, $\tilde{f}(\omega, \theta, \varphi)$, $f_1(x, \theta, \varphi)$, $f_2(\rho, \theta, \varphi)$ related to one another by Eqs. (3.2), (3.6), and (3.7) represent the same vector $f$ in $\mathcal{H}$. The scalar product has the form

$$\begin{aligned}
(f, g) &= \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\varphi \int_0^\pi \sin\theta \, d\theta \, f_1^*(x, \theta, \varphi) g_1(x, \theta, \varphi) \\
&= \int_{-\infty}^{\infty} d\rho \int_0^{2\pi} d\varphi \int_0^\pi \sin\theta \, d\theta \, f_2^*(\rho, \theta, \varphi) g_2(\rho, \theta, \varphi).
\end{aligned} \tag{3.8}$$

[7] H. E. Moses, Ann. Phys. (New York) **41**, 158 (1967).

[8] It should be pointed out that we are considering UIR's of the proper inhomogeneous Lorentz group only, without including the operations of space and/or time reflection.

Now we can express the effect of the generators $J_j$, $K_j$ on wavefunctions $f_2(\rho, \theta, \varphi)$ by the differential operators

$$\begin{aligned}
J_1 &= i \sin\varphi \frac{\partial}{\partial\theta} + i\cos\varphi\cot\theta\frac{\partial}{\partial\varphi} + \frac{s\cos\varphi}{\sin\theta},\\
J_2 &= -i\cos\varphi\frac{\partial}{\partial\theta} + i\sin\varphi\cot\theta\frac{\partial}{\partial\varphi} + \frac{s\sin\varphi}{\sin\theta},\\
J_3 &= -i\frac{\partial}{\partial\varphi},\\
K_1 &= (\rho + i)\sin\theta\cos\varphi - i\cos\theta\cos\varphi\frac{\partial}{\partial\theta}\\
&\quad + i\frac{\sin\varphi}{\sin\theta}\frac{\partial}{\partial\varphi} + s\sin\varphi\cot\theta,\\
K_2 &= (\rho + i)\sin\theta\sin\varphi - i\cos\theta\sin\varphi\frac{\partial}{\partial\theta}\\
&\quad - i\frac{\cos\varphi}{\sin\theta}\frac{\partial}{\partial\varphi} - s\cos\varphi\cot\theta,\\
K_3 &= (\rho + i)\cos\theta + i\sin\theta\frac{\partial}{\partial\theta}.
\end{aligned} \tag{3.9}$$

These coincide exactly with the generators of $O(3, 1)$ as given in (B) (except for a trivial change of sign of the $K_j$). Evaluating $\mathcal{C}_1$, and $\mathcal{C}_2$, from (3.9) we get

$$\mathcal{C}_1 = 1 + \rho^2 - s^2, \quad \mathcal{C}_2 = \rho s. \tag{3.10}$$

Thus by representing the vectors $f$ of $\mathcal{H}$ by the wavefunctions $f_2(\rho, \theta, \varphi)$, we have achieved the reduction of the UIR of the Poincaré group into UIR's of the subgroup $O(3, 1)$. The variable $\rho$ in the function $f_2(\rho, \theta, \varphi)$ is the same parameter that appeared in Sec. I in labeling the principal series of UIR's of $O(3, 1)$. The result of these considerations is the following.

*Theorem:* The unitary irreducible representation of the Poincaré group corresponding to zero mass, positive energy, and finite nonnegative integral helicity $s$ reduces into a direct integral of UIR's of the subgroup $O(3, 1)$, belonging to the principal series of UIR's of $O(3, 1)$; each UIR of the type $\{s, \rho\}$, for every $\rho$ in the range $-\infty < \rho < \infty$, appears exactly once.

It is obvious that if $s$ is not necessarily integral and nonnegative, we merely replace $s$ by $|s|$ in the statement of the above theorem.

The Casimir invariant of the $O(3)$ group $J^2$ is

$$J^2 = -\left[\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\left\{\frac{\partial^2}{\partial\varphi^2} - 2is\cos\theta\frac{\partial}{\partial\varphi} - s^2\right\}\right]. \tag{3.11}$$

Orthonormal eigenfunctions of $J^2$ and $J_3$ are well known.[9] They are the $D$ functions of angular-momentum theory

$$\begin{aligned}
J^2 D_j^{m,s}(\varphi, \theta, 0) &= j(j + 1)D_j^{m,s}(\varphi, \theta, 0),\\
J_3 D_j^{m,s}(\varphi, \theta, 0) &= m D_j^{m,s}(\varphi, \theta, 0).
\end{aligned} \tag{3.12}$$

With the help of these functions, we can construct a basis in the Hilbert space $\mathcal{H}$ of the UIR of the Poincaré group, corresponding to the states $|j, m\rangle$ appearing in Eq. (1.4). Namely, we define a set of "ideal" vectors in $\mathcal{H}$,

$$\Psi_{\rho jm} \tag{3.13}$$

whose wavefunctions written in terms of the variables $x$, $\theta$, $\varphi$ are

$$\Psi_{\rho jm} \to (2\pi)^{-\frac{1}{2}} e^{i\rho x}[(2j + 1)/4\pi]^{\frac{1}{2}} D_j^{m,s}(\varphi, \theta, 0). \tag{3.14}$$

Using the expression (3.8) for the scalar product, we establish

$$(\Psi_{\rho' j' m'}, \Psi_{\rho jm}) = \delta(\rho' - \rho)\delta_{j'j}\delta_{m'm}. \tag{3.15}$$

From the completeness properties of the $D_{j'}$ functions over the unit sphere, and the Fourier expansion theorem for square-integrable functions, we see that every normalizable vector $f$ in $\mathcal{H}$ may be "expanded" in the $\Psi_{\rho jm}$:

$$f = \sum_{j=s}^{\infty}\sum_{m=-j}^{j}\int_{-\infty}^{\infty} d\rho f_{jm}(\rho)\Psi_{\rho jm}. \tag{3.16}$$

The scalar product has the form

$$(f, g) = \sum_{j=s}^{\infty}\sum_{m=-j}^{j}\int_{-\infty}^{\infty} d\rho f_{jm}^*(\rho) g_{jm}(\rho). \tag{3.17}$$

The vectors $\Psi_{\rho jm}$ are "ideal vectors" and are not normalizable. Apart from phase factors, these states $\Psi_{\rho jm}$ for fixed $\rho$ are precisely the states $|jm\rangle$ introduced in (1.4) as a basis within the UIR $\{s, \rho\}$ of $O(3, 1)$. Acting on a wavefunction $f_{jm}(\rho)$ corresponding to a vector $\mathbf{f}$ in their domain, the ten operators $J_j$, $K_j$ leave $\rho$ unaltered and (except for some trivial phase factors) connect different values of $j, m$ with one another according to the matrix elements (1.6). We have

$$\begin{aligned}
(J_M f)_{jm}(\rho) &= [j(j + 1)]^{\frac{1}{2}} C_{m-M\ M\ m}^{\ j\ 1\ j} f_{j,m-M}(\rho),\\
&(K_M f)_{jm}(\rho)\\
&= -C_{m-M\ M\ m}^{\ j-1\ 1\ j}[(j^2 - s^2)/j(2j + 1)]^{\frac{1}{2}}\\
&\quad \times (\rho - ij) f_{j-1,m-M}(\rho)\\
&\quad + C_{m-M\ M\ m}^{\ j\ 1\ j}[\rho s/(j(j + 1))^{\frac{1}{2}}] f_{j,m-M}(\rho)\\
&\quad + C_{m-M\ M\ m}^{\ j+1\ 1\ j}[((j + 1)^2 - s^2)/(j + 1)(2j + 1)]^{\frac{1}{2}}\\
&\quad \times (\rho - i(j + 1)) f_{j+1,m-M}(\rho).
\end{aligned} \tag{3.18}$$

[9] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1957), Chap. 4.

Finally, we also have

$$(\mathcal{C}_1 f)_{jm}(\rho) = (1 + \rho^2 - s^2) f_{jm}(\rho),$$
$$(\mathcal{C}_2 f)_{jm}(\rho) = \rho s f_{jm}(\rho). \tag{3.19}$$

## 4. THE ENERGY-MOMENTUM OPERATORS

At the end of the last section, we have seen how the generators $J_j$, $K_j$ of the homogeneous Lorentz group act on a vector $\mathbf{f}$ given by the wavefunction $f_{jm}(\rho)$. Now we see how the energy–momentum operators $H$ and $P_j$ act on these wavefunctions.

Let us begin by writing the generators $P_j$ and $H$ in terms of $x, \theta, \varphi$; their effect on a wavefunction $f_1(x, \theta, \varphi)$ is given by

$$P_1 = e^x \sin\theta \cos\varphi, \quad P_2 = e^x \sin\theta \sin\varphi,$$
$$P_3 = e^x \cos\theta, \quad H = e^x. \tag{4.1}$$

Since we already know the effect of $K_j$, and since one can obtain $P_j$ in terms of the commutator of $K_j$ and $H$, it suffices to consider here the operator $H$ alone.

Let $\mathbf{f}$ be a vector in the domain of $H$. Its wavefunction $f_1(x, \theta, \varphi)$ then obeys

$$\|f\|^2 = \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta \, d\theta \, |f_1(x, \theta, \varphi)|^2 < \infty,$$
$$\|Hf\|^2 = \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta \, d\theta \, e^{2x} |f_1(x, \theta, \varphi)|^2 < \infty. \tag{4.2}$$

Clearly, (4.2) imposes severe restrictions on the behavior of $f_1(x, \theta, \varphi)$ as $x \to \infty$, and suggests that the Fourier transform of $f_1(x, \theta, \varphi)$, namely $f_2(\rho, \theta, \varphi)$, can be analytically continued into the upper-half complex $\rho$ plane. This is exactly the situation discussed in detail in (A), where we were concerned with UIR's of $O(2, 1)$. Using exactly the arguments given there, we arrive at the following conclusions[10]: If a vector $\mathbf{f}$ is in the domain of $H$, then for each $j, m$, the function

$$f_{jm}(\rho)$$

is the boundary value (in the sense of the limit in the mean) of an analytic function $\psi_{jm}(\zeta)$ of $\zeta = \rho + i\eta$, as $\eta \to 0_+$. $\psi_{jm}(\zeta)$ is analytic in the strip $0 < \eta < 1$ (at least); for each fixed $\eta$ in this range, we have

$$\sum_{j=s}^{\infty} \sum_{m=-j}^{j} \int_{-\infty}^{\infty} d\rho \, |\psi_{jm}(\rho + i\eta)|^2 < \infty. \tag{4.3}$$

As $\eta \to 1_-$, $\psi_{jm}(\zeta)$ approaches (in the limit in the mean sense) a square-integrable function of $\rho$, which we can write as

$$f_{jm}(\rho + i). \tag{4.4}$$

We have

$$(Hf)_{jm}(\rho) = f_{jm}(\rho + i) \tag{4.5}$$

and

$$\sum_{j=s}^{\infty} \sum_{m=-j}^{j} \int_{-\infty}^{\infty} d\rho \, |f_{jm}(\rho + i)|^2 < \infty. \tag{4.6}$$

The correspondence between vectors $\mathbf{f}$ in $\mathcal{H}$ and the wavefunctions $f_{jm}(\rho)$ is, in any case, only up to sets of measure zero; that is to say, given a vector $\mathbf{f}$, each member of the sequence of functions of $\rho$, $f_{jm}(\rho)$ is determined only up to sets of measure zero. Nonetheless, if $\mathbf{f}$ is in the domain of $H$, the functions $f_{jm}(\rho)$ (known only up to sets of measure zero), suffice to determine unique analytic continuations into the upper-half complex $\rho$ plane; calling these analytic functions $\psi_{jm}(\zeta)$ as above, the boundary values $f_{jm}(\rho)$ and $f_{jm}(\rho + i)$ are *limits in the mean* of $\psi_{jm}(\zeta)$ as $\mathrm{Im}\,\zeta \to 0_+$ and $1_-$, respectively. Thus from the functions $f_{jm}(\rho)$, we can determine (up to sets of vanishing measure) the functions $f_{jm}(\rho + i)$, and $f_{jm}(\rho + i)$ is the wavefunction of the normalizable vector $Hf$.

As is to be expected, it is meaningless to talk of the effect of the operators $P_j$, $H$ on the ideal, nonnormalizable states $\Psi_{\rho jm}$. They may only act on linear combinations $\mathbf{f}$ of the $\Psi_{\rho jm}$, if the "coefficients of the linear combination" $f_{jm}(\rho)$ possess all the properties mentioned above and permit the requisite analytic continuation in $\rho$. In this way, the energy–momentum operators "connect" different UIR's of $O(3, 1)$ to one another.[11]

We finally compute the effect of the operators $P_j$ on a vector $\mathbf{f}$ in their domain. Using spherical components $P_m$, the relation

$$[K_M, H] = iP_M, \tag{4.7}$$

and Eqs. (3.18) and (4.5), we find

$$\begin{aligned}(P_M f)_{jm}(\rho) &= C^{j-1}_{m-M\,1\,M\,m}[(j^2 - s^2)/j(2j+1)]^{\frac{1}{2}} f_{j-1,m-M}(\rho + i) \\ &\quad - C^{j}_{m-M\,1\,M\,m}(s/(j(j+1))^{\frac{1}{2}}) f_{j,m-M}(\rho + i) \\ &\quad - C^{j+1}_{m-M\,1\,M\,m}[((j+1)^2 - s^2)/(j+1)(2j+1)]^{\frac{1}{2}} \\ &\quad \times f_{j+1,m-M}(\rho + i).\end{aligned} \tag{4.8}$$

### ACKNOWLEDGMENTS

The author would like to thank Professor E. C. G. Sudarshan for useful discussions.

[10] The relevant theorems on Fourier and Laplace transforms may be found in D. V. Widder, *The Laplace Transform* (Princeton University Press, Princeton, N.J., 1941), Chap. VI, Sec. 8.

[11] This situation is quite similar to that encountered in the papers of Ref. 2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Current Formalism. I. Ordering Theorems for Currents*†

J. G. WRAY‡
*Physics Department, Syracuse University, Syracuse, New York*

(Received 11 July 1967)

The basic assumptions of the theory are strong unitarity, Bogoliubov causality, completeness of the in (out) fields on a unique vacuum, and Poincaré invariance. The algebra of Π functions, which are integral operators in the form of tempered distributions, is the main technical tool developed here. These Π functions can be regarded as generalized step functions where the combination law is such that products with $\delta$ functions and their derivatives are well defined. An application of this algebra of Π functions in the context of the above assumptions leads to the current-formalism representations of $n$th-order functional derivatives of the current and $S_{\text{op}}$. These are alternatively called the $R_P$ and $P$ product representations, respectively, the counterparts to the $R$ and $\Phi$ product representations of the $n$th-order derivative of the field and $S_{\text{op}}$ in the field formalism. In a subsequent paper these relations are used to derive the integro-differential equations of Pugh [R. E. Pugh, Ann. Phys. (N.Y.) **23**, 335 (1961)] in a form amenable to a diagrammatic analysis. The perturbation series is then shown to be unique and finite with no cutoffs and a number of parameters that is independent of (increases with) the order of expansion for renormalizable (nonrenormalizable) interactions.

## 1. INTRODUCTION

Asymptotic quantum field theory (AQFT) is characterized by the requirement that the interacting fields become free fields for large times. In combination with other reasonable assumptions, this has led to the reduction formulas of LSZ[1] and a representation of the $S_{\text{op}}$ on the mass shell (m.s.) in terms of $\Phi$ products of interacting fields. Its primary success was to provide a basis for the study of analyticity. As the reduction formulas provided a representation of only the m.s. $S_{\text{op}}$, it was not yet suited for dynamical calculations. However, by assuming that the $\Phi$ product be the off m.s. extension of the $S_{\text{op}}$ and that microcausality holds, an infinite set of integro-differential equations for the point functions was derived.[1] The perturbation-theoretic solution of these equations was demonstrated to exist at least formally (up to convergence of the expansion), but the question of uniqueness and of attendant boundary conditions was not answered.

Subsequent to LSZ, several proposals have been made for a finite formulation of field theory along similar lines.[2] The work of Pugh[3] was based on a strong unitarity condition and a "dynamical assumption." This "dynamical assumption" was later shown to be equivalent to strong Bogoliubov causality.[4] Strong unitarity and Bogoliubov causality are two physically motivated constraints that can be applied directly to the $S_{\text{op}}$ without the auxiliary construct of a field. Thus the interpolating field need never enter the formalism. If the field is introduced, the assumptions serve to specify the off m.s. extension of the $S_{\text{op}}$ as the $\Phi$ product of fields and the off m.s. extension of the field itself as an $R$ product of fields. This result has been reproduced in the context of a weak free-field equation[5] as introduced by Chen.[6] It will be demonstrated in the present work that without the field one is led to representations of the off m.s. $S_{\text{op}}$ and current by $P$ and $R_P$ products, respectively, of currents. The $P$ and $R_P$ products are formed analogously to $T$ and $R$ products, only with the $\theta$ function replaced by the Π function. The latter can be regarded as a generalization of the $\theta$ function.

The Π function and its further generalization, the $\Pi^{[N]}$ function, are the primary technical tools of this development. They form an algebra over a subspace of the space of tempered distributions and have the advantage over $\theta$ functions in that they can form products with a wider class of distributions, in particular, Dirac $\delta$ functions and their derivatives. They are identified in the theory as multipliers on *current* products and, in fact, Π multiplication on *field* products in general does not exist. The complete utilization of these functions leads to the elimination of the interacting field and enables the generalization of operator expressions for current and $S_{\text{op}}$ derivatives as functionals of currents. These are the current analogs of the $\theta$ and $R$ product representations of the $S_{\text{op}}$ and interpolating fields, respectively.

The validity for physics of the operator expressions

---

* Based in part on the Ph.D. thesis Syracuse University, 1966, during which time the author was a NASA Trainee.

† Work supported in part by a NSF research grant.

‡ Present address: Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

[1] H. Lehmann, R. Symanzik, and W. Zimmermann, Nuovo Cimento **6**, 1122 (1957).

[2] M. Muraskin and K. Nishijima, Phys. Rev. **122**, 331 (1961); B. V. Medvedev, Zh. Eksp. Teor. Fiz. **48**, 1479 (1965)[Sov. Phys. JETP **21**, 989 (1965)]; V. Ya. Fainberg, Zh. Eksp. Teor. Fiz. **47**, 2285 (1964)[Sov.-Phys. JETP **20**, 1529 (1965)]; Ref. 3; and others.

[3] R. E. Pugh, Ann. Phys. (N.Y.) **23**, 335 (1963).

[4] T. W. Chen, F. Rohrlich, and M. Wilner, J. Math. Phys. **7**, 1365 (1966).

[5] J. G. Wray, Ph.D. thesis, Syracuse University, 1966.

[6] T. W. Chen, Ann. Phys. (N.Y.) **42**, 476 (1967).

derived here is demonstrated *a posteriori* in a subsequent paper[7] where a perturbation expansion reproduces the results of renormalized Feynman–Dyson (FD) theory to all orders computed with a $\varphi^3(L_{\text{int}})$ interaction. Further, the formalism leads naturally to physically motivated boundary conditions for the integro-differential equations, resulting in unique matrix elements. It also leads to a program of diagrammatics. In the present work we will be content to demonstrate that, at least formally, unrenormalized FD is a solution of the equations.

The main assumptions and notation are covered in Sec. 2. In Sec. 3, ordering theorems with respect to the $\Pi(\Pi^{[N]})$ functions are derived. In Sec. 4 the analogs of the representation of the field by retarded products and of the $S_{\text{op}}$ by $\Phi$ products are derived for currents. Then, in the same context it is demonstrated that the Feynman–Dyson algorithm is a formal solution of these equations. The extension of these arguments with possible application to nonrenormalizable theories is facilitated in the manner of Chen's operator theory. A summary of the results is given in Sec. 5.

## 2. ASSUMPTIONS

The assumptions of AQFT have evolved somewhat over the years, but by and large contain statements of Lorentz invariance, unitarity, causality, and the asymptotic condition. The latter condition is not necessary here, since the interpolating field is absent from the formalism. In addition there is the necessary apparatus for constructing the space of physical states, in this case a Hilbert space. A recent review article[8] includes a statement of the assumptions and discusses their motivation. Here we will be content to state them briefly as a vehicle to introduce some of the notation of the theory.

*Assumption* (*i*): The theory is required to be Lorentz invariant. In this regard the metric of Minkowski space will have signature + 2.

*Assumption* (*ii*): There exists a representation of the free fields that is complete in the sense that all operators in the theory can be expressed in terms of them. Since only Hermitian scalar fields are to be considered, this means that all admissible operators $F$ will have the representation

$$F = \sum_{n=0}^{\infty} \frac{1}{n!} \int dx_1 \cdots dx_n f_n(x_1 \cdots x_n) :a_1 \cdots a_n: . \tag{2.1}$$

In this expression

$$a^{\text{in}}(x_i) \equiv a(x_i) \equiv a_{x_i} \equiv a_i \tag{2.2}$$

is a free Hermitian scalar field satisfying the free-field commutation relations; the coefficient functions, $f_n(x_1 \cdots x_n)$, are $c$-number distributions, symmetric under permutation of the 4 vectors $(x_1 \cdots x_n)$, and the integration is over all space–time. In other words, $F$ is an operator-valued distribution on the Hilbert space, $\mathcal{H}$, spanned by all polynomials in the creation operator $a_\alpha^*$ acting on the unique vacuum $|0\rangle$. Here

$$a_\alpha^* \equiv i \int d\sigma^\mu a(x) \overleftrightarrow{\partial_\mu} f_\alpha(x),$$

and $f_\alpha(x)$ is a normalizable, positive-frequency solution of the Klein–Gordon equation and will frequently be referred to as a mass shell (m.s.) test function. For this work it will be sufficient to restrict the coefficient function $f_n(x_1 \cdots x_n)$ to $\mathcal{S}'$, the space of tempered distributions.

The functional derivative with respect to a free field will be employed extensively throughout this work. The $m$th derivative of $F$, for example, is defined by

$$\frac{\delta^m F}{\delta a_{y_1} \cdots \delta a_{y_m}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} \int dx_1 \cdots dx_n \times f_{n+m}(y_1 \cdots y_m, x_1 \cdots x_n) :a_{x_1} \cdots a_{x_n}: \tag{2.3}$$

and is seen to free the $f_{m+n}$ from the m.s. constraint of the free fields in $m$ variables. Realizations of this derivative are thoroughly discussed in the literature.[9]

As pointed out by Pugh,[10] it is necessary to discriminate between strong ($\overset{\text{s}}{=}$) and weak ($\overset{\text{w}}{=}$) operator equations with respect to the functional derivative. A strong equation is valid even after an arbitrary number of derivatives have been taken, whereas a weak equation is not. Chen[11] helped alleviate a confusing situation by observing that the free-field equation is most naturally interpreted as a weak equation.

$$K_x a_x \equiv (\Box_x - m^2) a_x \overset{\text{w}}{=} 0, \tag{2.4}$$

with

$$\frac{\delta}{\delta a_y} K_x a_x \equiv K_x \frac{\delta a_x}{\delta a_y} = K_x \delta(x - y). \tag{2.5}$$

This convention will also be adopted here.

*Assumption* (*iii*): The $S_{\text{op}}$ can be defined as a unitary transformation between $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$, spanned

[7] J. G. Wray, J. Math. Phys. **9**, 552 (1968), following article.
[8] F. Rohrlich, *Perspective in Modern Physics*, R. E. Marshak, Ed. (Interscience Publishers, Inc., New York, 1966).
[9] F. Rohrlich, J. Math. Phys. **5**, 324 (1964); F. Rohrlich and M. Wilner, J. Math. Phys. **7**, 482 (1966).
[10] R. E. Pugh, J. Math. Phys. **6**, 740 (1965).
[11] Reference 6, p. 479.

by $\{a^{\text{in (out)}}_{(x)}\}$,

$$a^{\text{out}}_{(x)} \overset{S}{=} S^* a^{\text{in}}_{(x)} S, \quad S^* \overset{S}{=} S^{-1}. \tag{2.6}$$

Unitarity has been assumed as a strong equality, though for renormalizable interactions it follows as a consequence of weak unitarity and the remaining axioms.[12]

The stability of the vacuum and single-particle state under $S$ is also required.

*Assumption (iv)*: It will be required that the currents satisfy strong Bogoliubov causality,

$$\frac{\delta J_x}{\delta a_y} \overset{S}{=} 0, \quad \text{for} \quad x \sim y \quad \text{and} \quad x^0 < y^0, \tag{2.7}$$

and

$$J_x \equiv S^* \frac{i\delta S}{\delta a_x}, \tag{2.8}$$

where $x \sim y$ means that $(x - y)$ is spacelike. Bogoliubov causality, a restriction on the singularities appearing at the origin of the light cone, and the other assumptions imply the "dynamical" equation[13]

$$(1 - B_{xy})(i\delta J_x/\delta a_y) \overset{S}{=} \Pi_{xy}[J_x, J_y]. \tag{2.9}$$

The coefficients appearing in (2.9) are the tempered distributions,

$$\Pi_{xy} \equiv \Pi_{xy}(xy; \xi\eta) \equiv K_x K_y \theta_{xy} \Delta_A(x - \xi)\Delta_R(y - \eta), \tag{2.10}$$

for

$$\theta_{xy} \equiv \theta(x - y) = \begin{cases} 1, & x^0 > y^0 \\ 0, & x^0 < y^0, \end{cases} \tag{2.11}$$

and

$$B_{xy} = 1 - \Pi_{xy} - \Pi_{yx}. \tag{2.12}$$

The product with $\Pi_{xy}(xy; \xi\eta)$ is defined as a convolution in the right set of variables, $(\xi\eta)$, such that, for example

$$\Pi_{xy}[J_x, J_y] \equiv \int d\xi\, d\eta \Pi_{xy}(xy; \xi\eta)[J_\xi, J_\eta]. \tag{2.13}$$

These are the same operator coefficients first discovered by Pugh[14]; they satisfy

$$\left.\begin{aligned} \Pi^2_{xy} &= \Pi_{xy}, \\ \Pi_{xy}\Pi_{yx} &= 0, \\ B^2_{xy} &= B_{xy}, \\ \Pi_{xy} B_{xy} &= 0. \end{aligned}\right\} \tag{2.14}$$

The connection with the notation of Pugh is simply

$$P_R(x, y) = \Pi_{xy}, \quad P_A(x, y) = \Pi_{yx}.$$

Equation (2.9) can be seen to arise from the axioms of the theory with the additional restriction that $(i\delta J_1/\delta a_2)$ have only $\delta$- and $K\delta$-function singularities at the origin of the light cone. If this is true, then one can form the product $(\Pi_{xy} + \Pi_{yx})(i\delta J_x/\delta a_y)$, since $\Pi_{xy}\delta(x - y)$ and $\Pi_{xy}K_x\delta(x - y)$ are well defined and in fact vanish. From strong unitarity we obtain

$$(i\delta J_x/\delta a_y) - (i\delta J_y/\delta a_x) \overset{S}{=} [J_x, J_y], \tag{2.15}$$

which upon forming the product with $\Pi_{xy}$ becomes

$$\Pi_{xy}(i\delta J_x/\delta a_y) \overset{S}{=} \Pi_{xy}[J_x, J_y], \tag{2.16}$$

where

$$\Pi_{xy}(i\delta J_y/\delta a_x) \overset{S}{=} 0. \tag{2.17}$$

Equation (2.17) follows since with the restriction to $\delta$- and $K\delta$-function singularities at the origin of the light cone and with strong Bogoliubov causality the common support of its two factors vanishes. Further, since $\Pi_{yx}(i\delta J_x/\delta a_y) \overset{S}{=} 0$ in the same manner, Eq. (2.16) can be written

$$(\Pi_{xy} + \Pi_{yx})(i\delta J_x/\delta a_y) \overset{S}{=} \Pi_{xy}[J_x, J_y],$$

or

$$(1 - B_{xy})(i\delta J_x/\delta a_y) \overset{S}{=} \Pi_{xy}[J_x, J_y].$$

To conclude this section, we mention the following technical points. Only Heisenberg operators will be used and bound states have been excluded from the scattering states by construction. The compatibility of the assumptions has not been proved. However, the term-by-term existence of perturbation expansions (convergence of the series is not known) for nontrivial systems is suggestive that they are compatible.

## 3. OPERATOR ORDERING

In this section the $P$ and $P^{[N]}$ ordering of currents will be introduced. In addition, ordering theorems and domains of validity will be discussed. The presentation is parallel to that for $T$ ordering.[15]

### General Remarks

The ordered product of $n$ operators has the following general properties. For the operator $O$ defined by

$$O(A_1 \cdots A_n) \equiv (A_1 \cdots A_n)_{\text{ordered}}, \tag{3.1}$$

we have:

(1) $O$ is idempotent, $O^2 = 0$.

(2) $O$ is symmetric (for Bose fields) under permutation of the operators to be ordered.

[12] T. W. Chen, Nuovo Cimento **45**, A533 (1966).

[13] This equation was first proposed by Pugh[10] and later shown to follow from the assumptions.[5,6] This later work is patterned after similar work[4] which was for a representation of the dynamical equation possessing only formal significance.

[14] R. E. Pugh, J. Math. Phys. **7**, 376 (1966).

[15] F. Rohrlich and J. G. Wray, J. Math. Phys. **7**, 1697 (1966).

(3) If the $n$ operators are independent of each other, the ordering process will be linear. But usually this is not the case and in general ordering is a nonlinear operation such that $O(A + B) \neq O(A) + O(B)$ unless $A$ and $B$ are independent.

In general, the $n$ objects to be ordered will be operator-valued distributions. For the work reported here, the class of tempered distributions $\mathcal{S}'$ will be large enough so that the ordered products and their respective ordering theorems are in general rigorous and differentiable only as functionals to be folded with test functions from $\mathcal{S}$, the test function space of tempered distributions.

The $T$ product of fields, labeled by their space–time points, is defined by

$$T_+(A_1 \cdots A_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \theta_{1 \cdots n} A_1 \cdots A_n, \quad (3.2)$$

and

$$T_-(A_1 \cdots A_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \theta_{1 \cdots n} A_n \cdots A_1, \quad (3.3)$$

where

$$\theta_{1 \cdots n} \equiv \theta_{12}\theta_{23} \cdots \theta_{n-2,n-1}\theta_{n-1,n}, \quad (3.4)$$

$$\theta_{ij} \equiv \theta(x_i - x_j), \quad (3.5)$$

and

$$A_i \equiv A(x_i) \equiv A_{x_i}. \quad (3.6)$$

The $T_+(T_-)$ orderings thus define ordering with decreasing (increasing) times from left to right.

The idempotency of the $T_\pm$ ordering is guaranteed by the multiplicative relations satisfied by $\theta_{ij}$ as a distribution,

$$\theta_{ij}\theta_{1 \cdots i \cdots j \cdots n} = \theta_{1 \cdots i \cdots j \cdots n}, \quad (3.7)$$

and

$$\theta_{ji}\theta_{1 \cdots i \cdots j \cdots n} = 0, \quad \text{for} \quad 1 \leq i < j \leq n. \quad (3.8)$$

In addition, there is the important completeness relation,

$$\theta_{ij} + \theta_{ji} = 1. \quad (3.9)$$

These expressions are reviewed because it will be necessary to prove analogous relations for the distribution $\Pi_{ij}$ before it can properly qualify as the elemental coefficient around which the $P$-ordering concept is to be built. In this regard it is also important to note that the product of $\theta_{ij}$'s as defined by (3.4) is manifestly Abelian and associative. This, for example, will not be a trivial observation for the product of the $\Pi_{ij}$'s.

The $P$ product is constructed analogously to the $T$ product with $\theta_{ij} \to \Pi_{ij}$. It arises in the formalism as the ordering operation to be associated with currents and will therefore be defined on products of currents.

We have for $P_\pm$ products, respectively,

$$P_+(J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{1 \cdots n} J_1 \cdots J_n, \quad (3.10)$$

and

$$P_-(J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{1 \cdots n} J_n \cdots J_1, \quad (3.11)$$

where

$$\Pi_{1 \cdots n} \equiv \Pi_{12}\Pi_{23} \cdots \Pi_{n-2,n-1}\Pi_{n-1,n}, \quad (3.12)$$

$$\Pi_{ij} \equiv \Pi_{ij}(x_i x_j; y_i y_j)$$
$$\equiv K_{x_i} K_{x_j} \theta_{x_i x_j} \Delta_A(x_i - y_i)\Delta_R(x_j - y_j), \quad (3.13)$$

and

$$J_i \equiv J_{x_i} \equiv J(x_i). \quad (3.14)$$

Let us now define the space $\mathcal{F}'$, or indicating the four-vector variables explicitly, $\mathcal{F}'_{1 \cdots m} \equiv \mathcal{F}'_{x_1 \cdots x_m}$,

$$\mathcal{F}'_{1 \cdots m} \equiv \{\varphi_{1 \cdots m}\pi_{1 \cdots m}; \quad \text{all} \quad \varphi_{1 \cdots m} \in \mathcal{S}_{1 \cdots m}, \quad \text{all} \quad \pi_{1 \cdots m} \in \boldsymbol{\pi}_{1 \cdots m}\}. \quad (3.15)$$

In (3.15) $\mathcal{S}_{1 \cdots m}$ is the space of test functions of tempered distributions in the $m$-four vectors $(x_1 \cdots x_m)$ and $\boldsymbol{\pi}'_{1 \cdots m}$ is the linear space of functionals formed from all possible products of the $\Pi_{ij}$ $(i, j = 1, \cdots, m)$ and unity. That $\mathcal{F}' \subset \mathcal{S}'$ follows from $\boldsymbol{\pi}' \subset \mathcal{S}'$, the statement that all elements of $\boldsymbol{\pi}'$ are tempered distributions. $\mathcal{F}'$ is important since it exhausts the space of all left multipliers of the $\Pi_{ij}$ products of this theory with the exception of mass shell (m.s.) test functions. But since the $\Pi_{ij}$ vanish on the m.s., multiplicative relations to be established on $\mathcal{F}'$ become trivial on m.s. test functions.

The $\Pi_{ij}$ are tempered distributions that form an Abelian, associative algebra on $\mathcal{F}'$. The product law for a chain of $\Pi_{ij}$ (3.12) is a convolution for each repeated index. From (3.13) we see that each $\Pi_{ij}$ has a left and right set of variables running over the indices $i, j$.

$$(x_i x_j) = \text{left set},$$
$$(y_i y_j) = \text{right set}.$$

The convolution is formed from the right set of one $\Pi_{ij}$, proceeding to the right to the left set of the $\Pi_{kl}$ containing the first repetition of that index. Thus, for example, the product for the repeated index 1 in a particular chain is

$$\overbrace{\Pi_{23}\Pi_{21}(x_2 x_1; y_2 u_1)}^{\text{convolution}}\overbrace{\Pi_{45}\Pi_{61}(x_6 u_1; y_6 v_1)}^{\text{convolution}}\Pi_{78}\Pi_{19}(v_1 x_9; y_1 y_9), \quad (3.16)$$

where only those variables necessary to demonstrate the convolution product have been explicitly stated. This product is always defined, since the variables of

the left set have bounded support.[16] Associativity of the product is simply the statement that the result is independent of the order of performing the convolutions; i.e., for the example (3.16), the result is independent of the order of integration in $u_1, v_1$. This is demonstrated explicitly in Appendix A. The commutativity of the product

$$[\Pi_{ij}, \Pi_{kl}] = 0 \quad \text{on} \quad \mathcal{F}'_{(i\neq j, k\neq l)}, \quad \text{for } i,j,k,l = 1, \cdots m < \infty \tag{3.17}$$

is proved in Appendix B.

The idempotency of the $P$ ordering is guaranteed by the following multiplicative relations satisfied by the $\Pi_{ij}$ on $\mathcal{F}'$ (Appendix A)

$$\Pi_{ij}\Pi_{1\cdots i\cdots j\cdots n} = \Pi_{1\cdots i\cdots j\cdots n}, \tag{3.18}$$

and

$$\Pi_{ji}\Pi_{1\cdots i\cdots j\cdots n} = 0, \quad \text{for} \quad 1 \leq i < j \leq n. \tag{3.19}$$

These are the analogs of the $\theta$ products, (3.7) and (3.8). The analog of completeness is significantly different,

$$\Pi_{ij} + \Pi_{ji} = 1 - B_{ij}, \tag{3.20}$$

and can be taken as the defining relation for $B_{ij}$. The physical significance of these operators may become clear as the theory develops.

### $P$-Ordering Theorems

In this section the basic ordering theorem for $P$ products will be derived. As previously discussed, all $P$-ordered products and their associated theorems are rigorous relations only as linear functionals on the left space of multipliers $\mathcal{F}'$ (3.15). All relations proved in this section will use (3.18)–(3.20) and the following properties of the $\Pi_{ij}$, all of which are proved in Appendixes A and B,

$$\begin{aligned} &[\Pi_{ij}, \Pi_{kl}] = 0, \quad (i \neq j, k \neq l), \\ &B_{ij}\Pi_{1\cdots i\cdots l}\Pi_{k\cdots j\cdots n} = 0. \end{aligned} \tag{3.21}$$

The basic $P$-ordering theorem relates the $P$-ordered product of $n-1$ operators to that of $n$. From the definition of the $P$ ordered product, we have

$$P_+(J_1\cdots J_n) = \sum_{\substack{\text{perm}\\(1\cdots n)}} \Pi_{1:\cdots:n} J_1\cdots J_n, \tag{3.22}$$

which can be written

$$\begin{aligned} P_+(J_1\cdots J_n) = \sum_{\substack{\text{perm}\\(2\cdots n)}} \Pi_{2\ldots n}\{&\Pi_{12}(J_1J_2\cdots J_n) \\ &+ \Pi_{213}(J_2J_1J_3\cdots J_n) + \cdots + \Pi_{l,1,l+1} \\ &\times (J_2\cdots J_lJ_1J_{l+1}\cdots J_n) + \cdots + \Pi_{n1}(J_2\cdots J_nJ_1)\}. \end{aligned} \tag{3.23}$$

Also, we have

$$\begin{aligned} \Pi_{213} &= \Pi_{21}\Pi_{13} = \Pi_{23}\Pi_{21}\Pi_{13} \\ &= \Pi_{23}\Pi_{21}(1 - B_{13} - \Pi_{31}) \\ &= \Pi_{23}\Pi_{21} - \Pi_{23}\Pi_{21}B_{13} - \Pi_{23}\Pi_{21}\Pi_{31} \\ &= \Pi_{23}(\Pi_{21} - \Pi_{31}), \end{aligned} \tag{3.24}$$

since

$$\Pi_{23}\Pi_{21}B_{13} = 0$$

and

$$\Pi_{23}\Pi_{21}\Pi_{31} = \Pi_{231}\Pi_{21} = \Pi_{231} = \Pi_{23}\Pi_{31}.$$

In general, this analysis leads to

$$\Pi_{lmn} = \Pi_{ln}(\Pi_{lm} - \Pi_{nm}) \tag{3.25}$$

such that (3.23) becomes

$$\begin{aligned} &P_+(J_1\cdots J_n) \\ &\quad = \sum_{\substack{\text{perm}\\(2\cdots n)}} \Pi_{2\ldots n}\{(1 - B_{12} - \Pi_{21})(J_1J_2\cdots J_n) \\ &\qquad + \Pi_{23}(\Pi_{21} - \Pi_{31})(J_2J_1J_3\cdots J_n) + \cdots \\ &\qquad + \Pi_{l,l+1}(\Pi_{l1} - \Pi_{l+1,1})(J_2\cdots J_lJ_1J_{l+1}\cdots J_n) \\ &\qquad + \cdots + \Pi_{n1}(J_2\cdots J_nJ_1)\} \\ &\quad = \sum_{\substack{\text{perm}\\(2\cdots n)}} \Pi_{2\ldots n}\{(1 - B_{12} - \Pi_{21})(J_1J_2\cdots J_n) \\ &\qquad + (\Pi_{21} - \Pi_{31})(J_2J_1J_3\cdots J_n) + \cdots \\ &\qquad + (\Pi_{l1} - \Pi_{l+1,1})(J_2\cdots J_lJ_1J_{l+1}\cdots J_n) + \cdots \\ &\qquad + \Pi_{n1}(J_2\cdots J_nJ_1)\}. \end{aligned} \tag{3.26}$$

By rearranging terms, (3.26) can be written

$$\begin{aligned} P_+(J_1\cdots J_n) = \sum_{\substack{\text{perm}\\(2\cdots n)}} \Pi_{2\ldots n}\{&(1 - B_{12})(J_1J_2\cdots J_n) \\ &+ \Pi_{21}([J_2, J_1]J_3\cdots J_n) \\ &+ \cdots + \Pi_{l1}(J_2\cdots [J_l, J_1]J_{l+1}\cdots J_n) \\ &+ \Pi_{l+1,1}(J_2\cdots J_l[J_{l+1}, J_1]\cdots J_n) \\ &+ \cdots + \Pi_{n1}(J_2\cdots J_{n-1}[J_n, J_1])\} \end{aligned}$$

or

$$\begin{aligned} P_+(J_1\cdots J_n) = (1 - B_{12})J_1P_+(J_2\cdots J_n) \\ + [P_+(J_2\cdots J_n), J_1]_{R_P}, \end{aligned} \tag{3.27}$$

for[17]

$$[P_+(J_2\cdots J_n), J_1]_{R_P} \equiv \sum_{l=2}^{n} P'_+(J_2\cdots [J_l, J_1]_{R_P}\cdots J_n), \tag{3.28}$$

with

$$[J_l, J_1]_{R_P} \equiv \Pi_{l1}[J_l, J_1]. \tag{3.29}$$

Equation (3.27) is the basic ordering theorem which,

[16] V. Gorgé, Syracuse University Research Report SU-66-03 (1966).

[17] The (′) appearing on the right-hand side of (3.28) means the $P_+$ ordering with respect to all currents but $J_1$.

with its variations, can be summarized

$$P_{\pm}(J_1 \cdots J_n) = (1 - B_{12})J_1 P_{\pm}(J_2 \cdots J_n) \pm [P_{\pm}(J_2 \cdots J_n), J_1]_{\substack{R_P \\ A_P}}, \quad (3.30)$$

and

$$P_{\pm}(J_1 \cdots J_n) = (1 - B_{12})P_{\pm}(J_2 \cdots J_n)J_1 \pm [J_1, P_{\pm}(J_2 \cdots J_n)]_{\substack{R_P \\ A_P}}, \quad (3.31)$$

where

$$[P_{\pm}(J_2 \cdots J_n), J_1]_{\substack{R_P \\ A_P}} = [J_1, P_{\pm}(J_2 \cdots J_n)]_{\substack{A_P \\ R_P}}. \quad (3.32)$$

In correspondence with the $T$-product ordering theorems, both the Jacobi-type identities and the factorization formulas for $P$ products can be proved. However, only the equivalence of the $R_P$ product and $R_P$ commutator are of importance to this work. In Appendix C it is proved that

$$R_P = R'_P, \quad (3.33)$$

where $R_P$ is the $R_P$ product

$$i^n R_P(J_0; J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{01 \cdots n}[\cdots [J_0, J_1], J_2], \cdots J_n], \quad (3.34)$$

and $R'_P$, the $R_P$ commutator,

$$i^n R'_P(J_0; J_1 \cdots J_n) \equiv [\cdots [J_0, J_1]_{R_P}, J_2]_{R_P}, \cdots J_n]_{R_P}. \quad (3.35)$$

### $P^{[N]}$-Ordering Theorems

Chen[6] has extended this formulation of field theory to a class of interactions larger than the renormalizable interactions through the introduction, via the current Eq. (2.9), of a generalization of $\Pi_{ij}$, $\Pi^{[N]}_{ij}$ ($\Pi_{ij} = \Pi^{[1]}_{ij}$). The new current equation is[18]

$$(1 - B^{[N]}_{xy}) \frac{i\delta J_x}{\delta a_y} \overset{s}{=} \Pi^{[N]}_{xy}[J_x, J_y], \quad (3.36)$$

for

$$\Pi^{[N]}_{xy} \equiv (K_x K_y)^N \theta_{xy} \Delta_A^{(N-1)}(x - \xi)\Delta_R^{(N-1)}(y - \eta), \quad (3.37)$$

and[19]

$$\Delta^{(N)}_{A(R)}(x - \xi) \equiv \int du_1 \cdots du_N \Delta_{A(R)}(x - u_1)\Delta_{A(R)} \times (u_1 - u_2) \cdots \Delta_{A(R)}(u_{N-1} - u_N)\Delta_{A(R)}(u_N - \xi) = \frac{1}{(2\pi)^4} \int_{C_{A(R)}} \frac{d^4 p e^{ip(x-\xi)}}{(p^2 + m^2)^{N+1}}. \quad (3.38)$$

[18] The notation here is related to that of Chen's by the correspondence,

$$P^{(n)}_R(x, y) = \Pi^{[N]}_{xy}, \quad P^{(n)}_A(x, y) = \Pi^{[N]}_{yx}.$$

[19] V. Gorgé and F. Rohrlich, Syracuse University Research Report SU-66-10 (1966).

$B^{[N]}_{xy}$ is defined by

$$1 - B^{[N]}_{xy} \equiv \Pi^{[N]}_{xy} + \Pi^{[N]}_{yx}. \quad (3.39)$$

The motivation for this move is based on the fact that $\Pi^{[N]}_{xy}$ exists in product with $K^r_x \delta(x - y)$, $r < 2N$. Thus the current commutator and derivative of (3.36) can have singularities at the origin of the light cone of the order $K^r_x \delta(x - y)$, $r < 2N$.

The same multiplicative relations hold as for $\Pi_{xy}$ and $B_{xy}$

$$(\Pi^{[N]}_{xy})^2 = \Pi^{[N]}_{xy}, \qquad \Pi^{[N]}_{xy}\Pi^{[N]}_{yx} = 0. \quad (3.40)$$

In fact the $\Pi^{[N]}_{xy}$ form an Abelian, associative algebra on $\mathcal{F}^{[N]\prime}$ in the same manner that the $\Pi_{xy}$ form an Abelian, associative algebra on $\mathcal{F}'$ defined by (3.15). $\mathcal{F}^{[N]\prime}$, or, with the dependent variables $(x_1 \cdots x_m)$ explicitly indicated, $\mathcal{F}^{[N]\prime}_{1 \cdots m}$, is defined by

$$\mathcal{F}^{[N]\prime}_{1 \cdots m} \equiv \{\varphi_{1 \cdots m}\pi^{[N]}_{1 \cdots m}; \quad \text{all} \quad \pi^{[N]}_{1 \cdots m} \in \boldsymbol{\pi}^{[N]\prime}_{1 \cdots m}, \quad \text{all} \quad \varphi_{1 \cdots m} \in \mathfrak{S}_{1 \cdots m}\}, \quad (3.41)$$

for $N$ finite. Here $\mathfrak{S}_{1 \cdots m}$ is again the test-function space for tempered distributions in the $m$ four-vectors $(x_1 \cdots x_m)$, and $\boldsymbol{\pi}^{[N]\prime}_{1 \cdots m}$ is the linear space spanned by all polynomials of the $\Pi^{[N]}_{ij}$ $(i, j = 1, \cdots, m)$. That $\mathcal{F}^{[N]\prime} \subset \mathfrak{S}'$ for $N$ finite follows from the fact that all $\pi^{[N]}_{1 \cdots m} \in \boldsymbol{\pi}^{[N]\prime}_{1 \cdots m}$ are tempered distributions.

The product law, and proofs of existence, associativity, and commutativity proceed in the same manner as for $\Pi_{xy}$ only with the superscript $[N]$ added to all spaces and objects entering the proofs. Thus, chains of $\Pi^{[N]}_{xy}$ can be used to define $P^{[N]}$ ordering with corresponding ordering theorems valid on $\mathcal{F}^{[N]\prime}$. We have, therefore,

$$P^{[N]}_{+}(J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi^{[N]}_{1 \cdots n} J_1 \cdots J_n, \quad (3.42)$$

$$P^{[N]}_{-}(J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi^{[N]}_{1 \cdots n} J_n \cdots J_1, \quad (3.43)$$

for

$$\Pi^{[N]}_{1 \cdots n} \equiv \Pi^{[N]}_{12}\Pi^{[N]}_{23} \cdots \Pi^{[N]}_{n-1,n}. \quad (3.44)$$

The ordering theorems of importance here are

$$P^{[N]}_{\pm}(J_1 \cdots J_n) = (1 - B^{[N]}_{12})J_1 P^{[N]}_{\pm}(J_2 \cdots J_n) \pm [P^{[N]}_{\pm}(J_2 \cdots J_n), J_1]_{\substack{R_P{}^N \\ A_P{}^N}}, \quad (3.45)$$

and

$$P^{[N]}_{\pm}(J_1 \cdots J_n) = (1 - B^{[N]}_{12})P^{[N]}_{\pm}(J_2 \cdots J_n)J_1 \pm [J_1, P^{[N]}_{\pm}(J_2 \cdots J_n)]_{\substack{R_P{}^N \\ A_P{}^N}}, \quad (3.46)$$

for

$$[P_{\pm}^{[N]}(J_2 \cdots J_n), J_1]_{\substack{R_P{}^N \\ A_P{}^N}} \equiv \sum_{l=2}^{n} P_{\pm}^{[N]'}(J_2 \cdots [J_l, J_1]_{\substack{R_P{}^N \\ A_P{}^N}} \cdots J_n), \quad (3.47)$$

$$[J_l, J_1]_{R_P{}^N} \equiv \Pi_{l1}^{[N]}[J_l, J_1], \quad (3.48)$$

$$[J_l, J_1]_{A_P{}^N} \equiv -\Pi_{l1}^{[N]}[J_l, J_1], \quad (3.49)$$

and

$$[J_1, P_{\pm}^{[N]}(J_2 \cdots J_n)]_{\substack{R_P{}^N \\ A_P{}^N}} = [P_{\pm}^{[N]}(J_2 \cdots J_n), J_1]_{\substack{A_P{}^N \\ R_P{}^N}}. \quad (3.50)$$

The equality of the $R_P^{[N]}$ commutator and product can also be proved on $\mathcal{F}^{[N]'}$ by substituting $\Pi \to \Pi^{[N]}$ in the notation for the proof of $R'_P = R_P$ in Appendix C. Thus we have on $\mathcal{F}^{[N]'}$,

$$R_P^{[N]'} = R_P^{[N]}, \quad (3.51)$$

for

$$i^n R_P^{[N]'}(J_0; J_1 \cdots J_n) \equiv [\cdots [J_0, J_1]_{R_P{}^N}, J_2]_{R_P{}^N}, \cdots J_n]_{R_P{}^N}, \quad (3.52)$$

and

$$i^n R_P^{[N]}(J_0; J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{01 \cdots n}^{[N]}[\cdots [J_0, J_1], J_2], \cdots J_n]. \quad (3.53)$$

The rewards and shortcomings of this innovation by Chen will be explored in the next sections where generalized current equations are derived and in a subsequent paper[7] where an $S$-matrix formalism is developed. The corresponding field formalism is also being developed.[20] The current formalism, however, is enough to establish an integral equation for $S$-matrix elements, providing a framework for discussing the important question of boundary conditions in perturbation expansion.

## 4. THE CURRENT FORMALISM

In this section the representations for the off m.s. current and $S_{\text{op}}$ will be given in terms of $R_P$ and $P$ products of currents, respectively. These are the analogs of the $R$- and $\Phi$-product representations, respectively, in the field formalism. The physical content of these equations gains heuristic support in that they are formally satisfied by the Feynman–Dyson $S$ operator as will be proved. The question of the finiteness of the theory and a rigorous perturbation expansion are developed in a subsequent paper.[7] A generalization in the manner of Chen[6] of the off m.s. current and of the $S_{\text{op}}$ equations is also carried out.

[20] F. Rohrlich (to be published).

### Generalized Current Equation

The axioms of the theory imply as a strong equality that the first derivative of the current is given by Eq. (2.9),

$$(1 - B_{12})(i\delta J_1/\delta a_2) \overset{s}{=} \Pi_{12}[J_1, J_2]. \quad (4.1)$$

Now since the weak free-field formalism enables the commutation of operator and coordinate differentiation, (4.1) can be extended to the generalized current equation, valid on $\mathcal{F}'$,

$$\left[\prod_{i>i=0}^{n}(1 - B_{ij})\right]\frac{\delta^n J_0}{\delta a_1 \cdots \delta a_n} \overset{s}{=} R'_P(J_0; J_1 \cdots J_n). \quad (4.2)$$

This restriction is demanded by the use of an Abelian product for the $\Pi_{ij}$ which has been proved only on $\mathcal{F}'$.

The proof of (4.2) proceeds by induction. For $n = 1$ we regain Eq. (4.1). Now assume it for $n$ and it will be proved for $n + 1$. We have, by assumption,

$$\left[\prod_{k=0}^{n}(1 - B_{k,n+1})\right]\frac{\delta}{\delta a_{n+1}}\left[\prod_{j>i=0}^{n}(1 - B_{ij})\right]\frac{\delta^n J_0}{\delta a_1 \cdots \delta a_n} \overset{s}{=} \left[\prod_{k=0}^{n}(1 - B_{k,n+1})\right]\frac{\delta}{\delta a_{n+1}} R'_P(J_0; J_1 \cdots J_n). \quad (4.3)$$

But because the $(1 - B_{k,n+1})$, $(k = 0, 1, \cdots, n)$ commute on $\mathcal{F}'$ and are idempotent, Eq. (4.3) becomes

$$\left[\prod_{i>j=0}^{n+1}(1 - B_{ij})\right]\frac{\delta^{n+1} J_0}{\delta a_1 \cdots \delta a_{n+1}} \overset{s}{=} \left[\prod_{k=0}^{n}(1 - B_{k,n+1})\right] \times \sum_{l=0}^{n} R'_P\left[J_0; J_1 \cdots J_{l-1}(1 - B_{ln+1})\frac{\delta J_l}{\delta a_{n+1}} \cdots J_n\right]. \quad (4.4)$$

Now with (4.1), Eq. (4.4) can be written

$$\begin{aligned}\left[\prod_{i>j=0}^{n+1}(1 - B_{ij})\right]\frac{\delta^{n+1} J_0}{\delta a_1 \cdots \delta a_{n+1}} &\overset{s}{=} \left[\prod_{k=0}^{n}(1 - B_{k,n+1})\right](-i) \\ &\quad \times \sum_{l=0}^{n} R'_P(J_0; J_1 \cdots [J_l, J_{n+1}]_{R_P} \cdots J_n) \\ &\overset{s}{=} \left[\prod_{k=0}^{n}(1 - \dot{B}_{k,n+1})\right](-i)[R'_P(J_0; J_1 \cdots J_n), J_{n+1}]_{R_P} \\ &\overset{s}{=} \left[\prod_{k=0}^{n}(1 - B_{k,n+1})\right]R'_P(J_0; J_1 \cdots J_{n+1}), \end{aligned} \quad (4.5)$$

where the last step follows by definition of the $R_P$ commutator. That the coefficient function

$$\left[\prod_{k=0}^{n}(1 - B_{k,n+1})\right] \to 1$$

in (4.5) follows by (3.21), since all products of $B_{k,n+1}$ $(k = 0, 1, \cdots, n)$ contained in this coefficient act on

$R'_P(J_0; J_1 \cdots J_{n+1})$ which has a $\Pi_{ij}$ in every variable for every term. This concludes the proof of (4.2), since Eq. (4.5) becomes

$$\left[\prod_{i>j=0}^{n+1}(1 - B_{ij})\right]\frac{\delta^{n+1}J_0}{\delta a_1 \cdots \delta a_{n+1}} \overset{s}{=} R'_P(J_0; J_1 \cdots J_{n+1}). \tag{4.6}$$

Further, since the $R_P$ commutator and product are equal on $\mathcal{F}'$ (Appendix C), this also proves

$$\left[\prod_{i>j=0}^{n}(1 - B_{ij})\right]\frac{\delta^{n}J_0}{\delta a_1 \cdots \delta a_{n}} \overset{s}{=} R_P(J_0; J_1 \cdots J_{n}). \tag{4.7}$$

### The Generalized $S_{\text{op}}$ Equation

The operator equation for the second derivative of the $S_{\text{op}}$ follows from the current equation (4.1) and unitarity,

$$\begin{aligned}(1 - B_{12})S^*\frac{i^2\delta^2 S}{\delta a_1 \delta a_2} &\overset{s}{=} (1 - B_{12})\left(\frac{i\delta J_1}{\delta a_2} + J_2 J_1\right)\\ &\overset{s}{=} \Pi_{12}[J_1, J_2] + (1 - B_{12})J_2 J_1\\ &\overset{s}{=} P_+(J_1 J_2).\end{aligned} \tag{4.8}$$

That this result generalizes on $\mathcal{F}'$ to

$$\left[\prod_{i>j=1}^{n}(1 - B_{ij})\right]J_{1\cdots n} \overset{s}{=} P_+(J_1 \cdots J_n), \tag{4.9}$$

for

$$J_{1\cdots n} \equiv S^*\frac{i^n\delta^n S}{\delta a_1 \cdots \delta a_n}, \tag{4.10}$$

follows by induction.

For $n = 2$ we regain Eq. (4.8). Now assume it for $n$ and prove it for $n + 1$. Thus we have by assumption

$$\begin{aligned}&\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\frac{i\delta}{\delta a_{n+1}}\left[\prod_{i>j=1}^{n}(1 - B_{ij})\right]J_{1\cdots n}\\ &\quad\overset{s}{=}\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\frac{i\delta}{\delta a_{n+1}}P_+(J_1 \cdots J_n).\end{aligned} \tag{4.11}$$

Now

$$\begin{aligned}&\frac{i\delta}{\delta a_{n+1}}\left[\prod_{i>j=1}^{n}(1 - B_{ij})\right]J_{1\cdots n}\\ &\overset{s}{=}\left[\prod_{i>j=1}^{n}(1 - B_{ij})\right]\frac{i\delta}{\delta a_{n+1}}J_{1\cdots n}\\ &\overset{s}{=}\left[\prod_{i>j=1}^{n}(1 - B_{ij})\right](-J_{n+1}J_{1\cdots n} + J_{1\cdots n+1})\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}&\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\frac{i\delta}{\delta a_{n+1}}P_+(J_1 \cdots J_n)\\ &\quad\overset{s}{=}\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\\ &\quad\times\sum_{l=1}^{n}P_+\left[J_1 \cdots (1 - B_{l,n+1})\frac{i\delta J_l}{\delta a_{n+1}} \cdots J_n\right],\end{aligned} \tag{4.13}$$

but only on $\mathcal{F}'$ since we have commuted $(1 - B_{l,n+1})$ with the $P$ ordering. By the current equation (4.1), we obtain for (4.13)

$$\begin{aligned}&\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\frac{i\delta}{\delta a_{n+1}}P_+(J_1 \cdots J_n)\\ &\overset{s}{=}\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\sum_{l=1}^{n}P_+(J_1 \cdots [J_l, J_{n+1}]_{R_P} \cdots J_n)\\ &\overset{s}{=}\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right][P_+(J_1 \cdots J_n), J_{n+1}]_{R_P}.\end{aligned} \tag{4.14}$$

As in the proof of the generalized current equation, the coefficient in (4.14)

$$\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right] \to 1$$

by (3.21) since all products of the $B_{k,n+1}$ $(k = 1, \cdots, n)$ act on $[P_+(J_1 \cdots J_n), J_{n+1}]_{R_P}$ which has $\Pi_{ij}$ in every variable for every term. With this fact plus Eqs. (4.12) and (4.14) we have

$$\begin{aligned}&\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]\left[\prod_{i>j=1}^{n}(1 - B_{ij})\right]\\ &\qquad\times(-J_{n+1}J_{1\cdots n} + J_{1\cdots n+1})\\ &\qquad\overset{s}{=}[P_+(J_1 \cdots J_n),\ J_{n+1}]_{R_P},\end{aligned} \tag{4.15}$$

or

$$\begin{aligned}&\left[\prod_{i>j=1}^{n+1}(1 - B_{ij})\right]J_{1\cdots n+1}\\ &\qquad\overset{s}{=}\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]J_{n+1}P_+(J_1 \cdots J_n)\\ &\qquad\quad+ [P_+(J_1 \cdots J_n), J_{n+1}]_{R_P},\end{aligned} \tag{4.16}$$

where the assumption for $n$ of the equation to be proved has been used in the last step. But by (3.21) and (2.12),

$$\begin{aligned}&\left[\prod_{k=1}^{n}(1 - B_{k,n+1})\right]J_{n+1}P_+(J_1 \cdots J_n)\\ &\qquad\overset{s}{=}(1 - B_{n+1,1})J_{n+1}P_+(J_1 \cdots J_n),\end{aligned} \tag{4.17}$$

such that by the ordering theorem (3.30), Eq. (4.16) becomes

$$\left[\prod_{i>j=1}^{n+1}(1 - B_{ij})\right]J_{1\cdots n+1} \overset{s}{=} P_+(J_1 \cdots J_{n+1}),$$

concluding the proof.

The formal solution represented by the Feynman–Dyson (FD) $S$ operator

$$S_{\text{FD}} = (e^{-iH})_+ \tag{4.18}$$

is a solution of (4.2) and (4.9). This follows since the proofs of these expressions were inductive and were

therefore consequences of the $m = 1$ case. But for $m = 1$ the proof has already been given by Pugh.[21]

In connection with $S_{\rm FD}$, however, it is interesting to see that a mapping relation previously derived for fields[22] has its analog for currents based on $S_{\rm FD}$. This relation is (Appendix D),

$$T_+(J_1 \cdots J_m) = S^*_{\rm FD}(j_1 \cdots j_m S_{\rm FD})_+, \quad (4.19)$$

for

$$J_1 = S^*_{\rm FD}(i\delta S_{\rm FD}/\delta a_1) = S^*_{\rm FD}(j_1 S_{\rm FD})_+, \quad (4.20)$$

and

$$j_1 \equiv \delta H/\delta a_1. \quad (4.21)$$

In addition the current has an $R$-product representation (Appendix D)

$$J_x = \sum_{l=0}^{\infty} \frac{1}{l!} \int d\xi_1 \cdots d\xi_l R(j_x; h_1 \cdots h_l), \quad (4.22)$$

for

$$H = \int d\xi h_\xi. \quad (4.23)$$

These formal expressions may have value as heuristic aids to determine relations based on the $P$-ordering concept which have more than just formal significance when applied to currents.

### Current Formalism According to Chen

The innovation of Chen discussed in Sec. 3,

$$(1 - B^{[N]}_{xy})(i\delta J_x/\delta a_y) \stackrel{s}{=} \Pi^{[N]}_{xy}[J_x, J_y] \quad (4.24)$$

and, equivalently,

$$(1 - B^{[N]}_{xy})J_{xy} \stackrel{s}{=} P^{[N]}_+(J_x J_y), \quad (4.25)$$

can be extended on $\mathcal{F}^{[N]'}$ in the same way as in the cases for $N = 1$, Eqs. (4.2) and (4.9). The proofs are identical with those for $\Pi_{xy}$ and $B_{xy}$ except for their replacement by $\Pi^{[N]}_{xy}$ and $B^{[N]}_{xy}$, so that only the results will be presented here. The current Eq. (4.24), generalizes to

$$\left[\prod_{i>j=0}^{n}(1 - B^{[N]}_{ij})\right] \frac{\delta^n J_0}{\delta a_1 \cdots \delta a_n} \stackrel{s}{=} R^{[N]'}_P(J_0; J_1 \cdots J_n) \quad (4.26)$$

$$\stackrel{s}{=} R^{[N]}_P(J_0; J_1 \cdots J_n), \quad (4.27)$$

the $R^{[N]}_P$ commutator and product, respectively. For $N = 1$ these results reduce to (4.2) and (4.7). The $S_{\rm op}$ equation, (4.25), generalizes to

$$\left[\prod_{i>j=1}^{n}(1 - B^{[N]}_{ij})\right] J_{1\ldots n} \stackrel{s}{=} P^{[N]}_+(J_1 \cdots J_n), \quad (4.28)$$

which reduces for $N = 1$ to Eq. (4.9). The notation and ordering theorems for these proofs was developed in Sec. 3.

## 5. CONCLUSION

The identification of the $\Pi$-($\Pi^{[N]}$-) function algebra is the most important technical contribution of this work. This algebra is associative and Abelian on the space $\mathcal{F}'(\mathcal{F}^{[N]'}) \subset \mathcal{S}'$ and its elements arise in convolution products with currents. The $\Pi(\Pi^{[N]})$ functions permit multiplication with certain distributions such as Dirac $\delta$ functions and their derivatives in contrast to $\theta$ functions which do not have this property. Coupled with the fact that its product with fields does not in general exist, we find the $\Pi(\Pi^{[N]})$ functions appearing as the natural elements for constructing ordering theorems of currents. Thus its first major application leads to the current representations of the derivatives of the current and the $S_{\rm op}$, respectively. Specifically, these are the results (4.2) and (4.9) with respect to the $\Pi$ function and (4.27) and (4.28) with respect to the $\Pi^{[N]}$ function.

Additional applications of the $\Pi$-functional algebra are developed in a subsequent paper.[7] Here the integral equation of Pugh[1] is derivable without ever introducing the construct of an interpolating field. The equations take on a particularly simple form, however, enabling the identification of diagrams and a successful solution to the boundary-condition problems in perturbation expansion, as will be shown. Thus in the current formalism an ordered product of currents analogous to time ordering has been generated. It has the advantage, however, of rigorous rather than just formal existence for the case of renormalizable interactions. For nonrenormalizable interactions it is demonstrated[7] that the perturbation expansion exists term by term and is unique, but at the expense of introducing additional parameters with increasing order of expansion.

## ACKNOWLEDGMENTS

The author wishes to express his appreciation to Professor F. Rohrlich for his encouragement and suggestions during the course of this research. Also, helpful discussions with Drs. T. W. Chen, M. Wilner, M. Wellner, and V. Gorgé are gratefully acknowledged.

## APPENDIX A: CHAINS OF $\Pi_{ij}$

The $\Pi_{ij}$ and products formed from them are tempered distributions which constitute an Abelian, associative algebra on $\mathcal{F}'$ as defined in Eq. (3.15). In this appendix the product will be demonstrated to be associative and various multiplicative relations will

[21] Reference 10, p. 745.
[22] Reference 15, p. 1704.

be proved. Commutativity of the product is proved in Appendix B.

For the product

$$\Pi_{23}\Pi_{45}\Pi_{13}\Pi_{78}\Pi_{39}$$

$$= \Pi_{23}(x_2x_3; y_2u_3)\Pi_{45}\Pi_{13}(x_1u_3; y_1v_3)\Pi_{78}\Pi_{39}(v_3x_9; y_3y_9), \quad \text{(A1)}$$

(convolution, convolution)

associativity is independence of the order of integration over $u_1$ and $v_1$. This can be seen by explicitly carrying out the convolution of any one product. Take, for example,

$$\Pi_{23}\Pi_{13} = \int \Pi_{23}(x_2x_3; y_2u_3)\, du_3 \Pi_{13}(x_1u_3; y_1y_3), \quad \text{(A2)}$$

or

$$\Pi_{23}\Pi_{13} = K_{x_2}K_{x_3}\theta_{x_2x_3}\int \Delta_A(x_2 - y_2)\, du_3 \Delta_R(x_3 - u_3) \times K_{x_1}K_{u_3}\theta_{x_1u_3}\Delta_A(x_1 - y_1)\Delta_R(u_3 - y_3). \quad \text{(A3)}$$

Integrating by parts twice leaves only the volume term, which is simply a convolution over a Dirac $\delta$ function,

$$\begin{aligned}\Pi_{23}\Pi_{13} &= K_{x_2}K_{x_3}\theta_{x_2x_3}\Delta_A(x_2 - y_2) \\ &\quad \times \int du_3[K_{u_3}\Delta_R(x_3 - u_3)] \\ &\quad \times K_{x_1}\theta_{x_1u_3}\Delta_A(x_1 - y_1)\Delta_R(u_3 - y_3) \\ &= -K_{x_2}K_{x_3}\theta_{x_2x_3}\Delta_A(x_2 - y_2)\int du_3\delta(x_3 - u_3) \\ &\quad \times K_{x_1}\theta_{x_1u_3}\Delta_A(x_1 - y_1)\Delta_R(u_3 - y_3) \\ &= -K_{x_2}K_{x_3}K_{x_1}\theta_{x_2x_3}\theta_{x_1x_3}\Delta_A(x_1 - y_1) \\ &\quad \times \Delta_A(x_2 - y_2)\Delta_R(x_3 - y_3). \quad \text{(A4)}\end{aligned}$$

Now since all convolution products appearing in the chain are of this form, only the volume terms contribute after integration by parts. But the volume terms are all convolutions over Dirac $\delta$ functions and lead to unambiguous results independent of the order of integration.

That the surface terms in (A4) vanish can be seen in the following way. Explicitly, we have

$$\begin{aligned}\Pi_{23}\Pi_{13}\Big|_{\substack{\text{surface}\\ \text{terms}}} &= -\left(\lim_{u_3^0 \to \infty} - \lim_{u_3^0 \to -\infty}\right) \\ &\quad \times K_{x_2}K_{x_3}\theta_{x_2x_3}\Delta_A(x_2 - x_3) \\ &\quad \times \int d^3u_3 \Delta_R(x_3 - u_3) \\ &\quad \times \overleftrightarrow{\partial}_{u_3^0} K_{x_1}\theta_{x_1u_3}\Delta_A(x_1 - y_1)\Delta_R(u_3 - y_3).\end{aligned}$$

This term vanishes because the factors

$$\begin{aligned}&\text{(i)} \quad K_{x_1}\theta_{x_1u_3}\Delta_A(x_1 - y_1)\Delta_R(u_3 - y_3) \\ &\qquad = -K_{x_1}\theta_{y_1x_1u_3y_3}\Delta(x_1 - y_1)\Delta(u_3 - y_3) \\ &\text{(ii)} \quad K_{x_1}\dot{\theta}_{x_1u_3}\Delta_A(x_1 - y_1)\Delta_R(u_3 - y_3) \\ &\qquad = -K_{x_1}\delta(x_1^0 - u_3^0)\theta_{y_1x_1}\theta_{u_3y_3}\Delta(x_1 - y_1)\Delta(u_3 - y_3), \\ &\text{(iii)} \quad K_{x_1}\theta_{x_1u_3}\Delta_A(x_1 - y_1)\dot{\Delta}_R(u_3 - y_3) \\ &\qquad = -K_{x_1}\theta_{y_1x_1u_3y_3}\Delta(x_1 - y_1)\dot{\Delta}(u_3 - y_3),\end{aligned}$$

have no contribution in the $\lim_{u_3^0 \to \pm\infty}$ because of their respective $\theta$ functions. The differential operator $K_{x_1}$ will not affect this limit since in each of the differentiated terms the argument $u_3^0$ will still be bounded above and below by $y_1^0$, $y_3^0$, respectively, by surviving $\theta$ functions.

The multiplicative and linear relations between $\Pi_{12}$, $\Pi_{21}$, and $B_{12}$ have been summarized in Eqs. (2.12) and (2.14). Products or chains of two or more $\Pi_{ij}$ $(i, j = 1, \cdots, m; m \geq 2)$ have properties which are generalizations of those for $m = 2$. The particular chain

$$\Pi_{\alpha_1 \cdots \alpha_m} \equiv \Pi_{\alpha_1\alpha_2}\Pi_{\alpha_2\alpha_3}\cdots\Pi_{\alpha_{m-1}\alpha_m} \quad \text{(A5)}$$

for $\{\alpha_i\}$ some set of integers, has already been introduced in the text. Since, as it is proved in Appendix B, the $\Pi_{ij}$ form an Abelian algebra on $\mathcal{F}'$,

$$[\Pi_{ij}, \Pi_{kl}] = 0, \quad \text{(A6)}$$

the ordering of the $\Pi_{ij}$ in (A5) is immaterial. In addition, (A6) enables the reduction of any chain to products of factors of the form (A6). For example,

$$\Pi_{32}\Pi_{12}\Pi_{13} = \Pi_{12}\Pi_{13}\Pi_{32} = \Pi_{12}\Pi_{132}, \quad \text{(A7)}$$

where (A6) and (A5) have been used, respectively. Thus, to characterize properties of an arbitrary chain, it is sufficient to discuss only those of the form (A5).

Useful properties of chains to be proved on $\mathcal{F}'$ are the following. A particular ordering of the integers has been chosen for convenience and does not restrict the generality of the results.

$$\begin{aligned}&\text{(i)} \quad \Pi_{1\cdots m}^* \equiv (\Pi_{1\cdots m})_{\substack{\text{complex}\\ \text{conjugate}}} = \Pi_{1\cdots m}, \\ &\text{(ii)} \quad \Pi_{1\cdots m}^2 = \Pi_{1\cdots m}, \\ &\text{(iii)} \quad \Pi_{1\cdots m} = (-)^m K_{x_1}\cdots K_{x_m}\theta_{x_1\cdots x_m}\Delta_A(x_1 - y_1) \\ &\qquad \cdots \Delta_A(x_{m-1} - y_{m-1})\Delta_R(x_m - y_m), \\ &\text{(iv)} \quad \Pi_{ij}\Pi_{1\cdots i\cdots j\cdots m} = \Pi_{1\cdots i\cdots j\cdots m}\Pi_{ij} \\ &\qquad = \Pi_{1\cdots i\cdots j\cdots m}, \\ &\text{(v)} \quad \Pi_{1\cdots i\cdots j\cdots m}\Pi_{1\cdots j\cdots i\cdots m} = 0, \\ &\text{(vi)} \quad B_{ij}\Pi_{1\cdots i\cdots j\cdots m} = \Pi_{1\cdots i\cdots j\cdots m}B_{ij} = 0, \\ &\text{(vii)} \quad B_{ij}\Pi_{1\cdots i\cdots l}\Pi_{k\cdots j\cdots m} = \Pi_{1\cdots i\cdots l} \\ &\qquad \times \Pi_{k\cdots j\cdots m}B_{ij} = 0,\end{aligned}$$

and

$$\Pi_{1\dots i\dots l}B_{ij}\Pi_{k\dots j\dots m}=0.$$

The proofs are, in order:

(i) $$\Pi^*_{1\dots m}=\Pi^*_{12}\Pi^*_{23}\cdots\Pi^*_{m-1,m}$$
$$=\Pi_{12}\Pi_{23}\cdots\Pi_{m-1,m}=\Pi_{1\dots m} \quad \text{(A8)}$$

by virtue of $\Pi^*_{ij}=\Pi_{ij}$.

(ii) $$\Pi^2_{1\dots m}=(\Pi_{12}\Pi_{23}\cdots\Pi_{m-1,m})^2$$
$$=\Pi^2_{12}\Pi^2_{23}\cdots\Pi^2_{m-1,m}$$
$$=\Pi_{12}\Pi_{23}\cdots\Pi_{m-1,m}=\Pi_{1\dots m}, \quad \text{(A9)}$$

where use has been made of (A6) and $\Pi^2_{ij}=\Pi_{ij}$.

(iii) To prove

$$\Pi_{1\dots m}=(-)^mK_{x_1}\cdots K_{x_m}\theta_{x_1\cdots x_m}$$
$$\times\Delta_A(x_1-y_1)\cdots\Delta_A(x_{m-1}-y_{m-1})\Delta_R(x_m-y_m) \quad \text{(A10)}$$

will involve some calculation. First we have, explicitly indicating the convolutions to be performed,

$$\Pi_{1\dots m}=\Pi_{12}\Pi_{23}\Pi_{34}\cdots\Pi_{m-1,m}$$
$$=\int\Pi_{12}(x_1x_2;\,y_1u_2)\,du_2\Pi_{23}(u_2x_3;\,y_2u_3)$$
$$\times\,du_3\Pi_{34}(u_3x_4;\,y_3u_4)\,du_4\cdots\Pi_{m-1,m}$$
$$\times(u_{m-1}x_m;\,y_{m-1}y_m). \quad \text{(A11)}$$

But as we have seen [Eq. (A4)] after twice integrating by parts, each of these integrals has only a volume-term contribution, which is a convolution over a Dirac $\delta$ function. Effectively, the $\Delta_I(x-u)$ of the left member of any convolution product combines with $K_\mu$ of the right member to yield $-\delta(x-u)$. Thus the result can be written down at once

$$\Pi_{1\dots m}=\int K_{x_1}K_{x_2}\theta_{x_1x_2}\Delta_A(x_1-y_1)\Delta_R(x_2-u_2)$$
$$\times\,du_2K_{u_2}K_{x_3}\theta_{u_2x_3}\Delta_A(u_2-y_2)\cdot\Delta_R(x_3-u_3)$$
$$\times\,du_3\cdots du_{m-1}K_{u_{m-1}}K_{x_m}\theta_{u_{m-1}x_m}$$
$$\times\Delta_A(u_{m-1}-y_{m-1})\Delta_R(x_m-y_m)$$
$$=(-)^{m-2}\int K_{x_1}K_{x_2}\theta_{x_1x_2}\Delta_A(x_1-y_1)$$
$$\times\,\delta(x_2-u_2)\cdot du_2K_{x_3}\theta_{u_2x_3}\Delta_A(u_2-y_2)$$
$$\times\,\delta(x_3-u_2)\cdot du_3\cdots du_{m-1}$$
$$\times K_{x_m}\theta_{u_{m-1}x_m}\Delta_A(u_{m-1}-y_{m-1})\Delta_R(x_m-y_m)$$
$$=(-)^mK_{x_1}\cdots K_{x_m}\theta_{x_1\cdots x_m}$$
$$\times\Delta_A(x_1-y_1)\cdots\Delta_A(x_{m-1}-y_{m-1})$$
$$\Delta_R(x_m-y_m). \quad \text{Q.E.D.}$$

(iv) To prove

$$\Pi_{ij}\Pi_{1\dots i\dots j\dots m}=\Pi_{1\dots i\dots j\dots m}\Pi_{ij}$$
$$=\Pi_{1\dots i\dots j\dots m}, \quad \text{(A12)}$$

we first note that

$$\Pi_{ij}\Pi_{1\dots i\dots j\dots m}=\Pi_{ij}(\Pi_{12}\Pi_{23}\cdots\Pi_{m-1,m})$$
$$=[\Pi_{ij},\,\Pi_{12}](\Pi_{23}\cdots\Pi_{m-1,m})$$
$$+\,\Pi_{12}[\Pi_{ij},\,\Pi_{23}](\Pi_{34}\cdots\Pi_{m-1,m})$$
$$+\cdots+(\Pi_{12}\cdots\Pi_{m-2,m-1})[\Pi_{ij},\,\Pi_{m-1,m}]$$
$$+\,\Pi_{1\dots m}\Pi_{ij},$$

which by (A4) becomes

$$\Pi_{ij}\Pi_{1\dots m}=\Pi_{1\dots m}\Pi_{ij}, \quad \text{(A13)}$$

proving the first part of (A12). The remaining expression can be proved explicitly by calculation. From (A10) we have

$$\Pi_{ij}\Pi_{1\dots i\dots j\dots m}$$
$$=K_{x_i}K_{x_j}\theta_{x_ix_j}\int du_i\,du_j\Delta_A(x_i-u_i)$$
$$\times\Delta_R(x_j-u_j)\cdot(-)^mK_{x_1}\cdots K_{u_i}\cdots K_{u_j}\cdots K_{x_m}$$
$$\times\theta_{x_1\cdots u_i\cdots u_j\cdots x_m}\Delta_A(x_1-y_1)\cdots\Delta_A$$
$$\times(u_i-y_i)\cdots\Delta_A(u_j-y_j)\cdots\Delta_R(x_m-y_m).$$

Once again, integration by parts twice in both 4-variables $u_1$, $u_2$ has only volume contributions over the Dirac $\delta$ functions $\delta(u_i-x_i)$, $\delta(u_j-x_j)$, respectively, such that

$$\Pi_{ij}\Pi_{1\dots i\dots j\dots m}=(-)^mK_{x_1}\cdots K_{x_m}\theta_{x_1\cdots x_m}$$
$$\times\Delta_A(x_1-y_1)\cdots\Delta_A(x_{m-1}-y_{m-1})\Delta_R(x_m-y_m)$$
$$=\Pi_{1\dots m},$$

concluding the proof.

(v) We have

$$\Pi_{1\dots i\dots j\dots m}\Pi_{1\dots j\dots i\dots m}$$
$$=\Pi_{1\dots i\dots j\dots m}\Pi_{ij}\Pi_{ji}\Pi_{1\dots j\dots i\dots m}=0, \quad \text{(A14)}$$

where the first step follows from (A12) and second from (2.14).

(vi) This proof follows again from (A12), since

$$B_{ij}\Pi_{1\dots i\dots j\dots m}=B_{ij}\Pi_{ij}\Pi_{1\dots i\dots j\dots m}=0 \quad \text{(A15)}$$

by (2.14). The remaining term can be proved in the same way or just by the fact that

$$[B_{ij},\,\Pi_{1\dots m}]=[(1-\Pi_{ij}-\Pi_{ji}),\,\Pi_{1\dots m}]=0 \quad \text{(A16)}$$

by (A13), such that (A15) implies

$$\Pi_{1\dots i\dots j\dots m}B_{ij}=0. \quad \text{(A17)}$$

(vii) This result does not follow from the previous relations so that a calculation must be made. Again one of the three statements implies the others by

(A16). To prove the first relation, we note that

$$B_{ij}\Pi_1\cdots{}_i\cdots{}_l\Pi_k\cdots{}_j\cdots{}_m = B_{ij}\Pi_{1i}\Pi_{jm}\Pi_1\cdots{}_i\cdots{}_l\Pi_k\cdots{}_j\cdots{}_m \quad \text{(A18)}$$

by (A12). But

$$B_{ij}\Pi_{1i}\Pi_{jm} = K_{x_i}K_{x_j}\theta_{x_ix_j}\int du_i\,du_j[\Delta(x_i-u_i)\times\Delta_R(x_j-u_j)-\Delta_R(x_i-u_i)\Delta(x_j-u_j)]\times K_{x_1}K_{u_i}K_{u_j}K_{x_m}\theta_{x_1u_i}\theta_{u_jx_m}\Delta_A(x_1-y_1)\times\Delta_R(u_i-y_i)\Delta_A(u_j-y_j)\Delta_R(x_m-y_m)=0, \quad \text{(A19)}$$

since integrating twice by parts leaves only the volume terms which vanish by

$$K_{u_i}K_{u_j}[\Delta(x_i-u_i)\Delta_R(x_j-u_j)-\Delta_R(x_i-u_i)\Delta(x_j-u_j)]=0. \quad \text{(A20)}$$

Thus (A18) vanishes by (A19) and the relation is proved. The proof here and that for (A15) are equivalent to saying that the Fourier transform of $[\Pi_1\cdots{}_m]$ vanishes whenever one of its variables is put on the mass shell. That is, since the right set of variables in the Fourier transform of $B(x_1x_2;\,y_1y_2)$ contains one m.s. $\delta$ function in each term

$$\text{F.T. }[B(x_1x_2;\,y_1y_2)] = \tilde{B}(p_1p_2;\,q_1q_2) \sim A(p_1p_2q_1q_2)\delta(q_1^2+m^2)+C(p_1p_2q_1q_2)\delta(q_2^2+m^2), \quad \text{(A21)}$$

the product (A18) can be viewed as taking the factor F.T. $[\Pi_1\cdots{}_m]$ to the m.s. in one variable.

## APPENDIX B: COMMUTATIVITY OF THE $\Pi_{ij}$

In this appendix it will be proved that

$$[\Pi_{ij},\Pi_{kl}]=0 \quad \text{on} \quad \mathcal{F}'(i\neq j,\,k\neq l), \quad \text{(B1)}$$

or more precisely that

$$f_1\cdots{}_m[\Pi_{ij},\Pi_{kl}]=0, \quad i,j,k,l=1,\cdots,m \quad (i\neq j,\,l\neq k) \quad \text{for all } f_1\cdots{}_m\in\mathcal{F}'_1\cdots{}_m \quad \text{(B2)}$$

and $\mathcal{F}'_1\cdots{}_m$ defined by Eq. (3.15).

If we let $(i,j,k,l)$ run over $(1,2,3,4)$ for simplicity, we must contend with the three cases

(i) $(ij)=(kl)$: $[\Pi_{12},\Pi_{12}]$, $[\Pi_{12},\Pi_{21}]$,

(ii) $i=k, j\neq l$: $[\Pi_{12},\Pi_{23}]$, $[\Pi_{12},\Pi_{32}]$,

(iii) $(ij)\neq(Kl)$: $[\Pi_{12},\Pi_{34}]$.

Four of these vanish directly as a result of carrying out their products:

$$[\Pi_{12},\Pi_{12}]=\Pi_{12}^2-\Pi_{12}^2=0, \quad \text{(B3)}$$

$$[\Pi_{12},\Pi_{21}]=\Pi_{12}\Pi_{21}-\Pi_{21}\Pi_{12}=0-0=0, \quad \text{(B4)}$$

$$[\Pi_{12},\Pi_{34}]=0, \quad \text{(B5)}$$

since there is no convolution between dissimilar indices and

$$[\Pi_{12},\Pi_{32}]=\Pi_{12}\Pi_{32}-\Pi_{32}\Pi_{12} = K_{x_1}K_{x_2}\theta_{x_1x_2}\Delta_A(x_1-\xi_1)\times\int d\eta_2\Delta_R(x_2-\eta_2)K_{x_3}K_{\eta_2}\theta_{x_3\eta_2}\times\Delta_A(x_3-y_3)\Delta_R(\eta_2-\xi_2) - K_{x_3}K_{x_2}\theta_{x_3x_2}\Delta_A(x_3-\xi_3)\times\int d\eta_2\Delta_R(x_2-\eta_2)K_{x_1}K_{\eta_2}\theta_{x_1\eta_2}\times\Delta_A(x_1-\xi_1)\Delta_R(\eta_2-\xi_2),$$

which upon twice integrating by parts becomes[23]

$$[\Pi_{12},\Pi_{32}] = -K_{x_1}K_{x_2}K_{x_3}\theta_{x_1x_2}\theta_{x_3x_2}\times[\Delta_A(x_1-\xi_1)\Delta_A(x_3-\xi_1)\Delta_R(x_2-\xi_2) - \Delta_A(x_1-\xi_1)\Delta_A(x_3-\xi_3)\Delta_R(x_2-\xi_2)] = 0. \quad \text{(B6)}$$

The fifth combination $[\Pi_{12}\,\Pi_{23}]$ does not vanish as a consequence of carrying out the indicated products, but it is still the zero distribution on $\mathcal{F}'_1\cdots{}_m$ as defined by (B2). Explicitly we have

$$[\Pi_{12},\Pi_{23}] = \Pi_{12}\Pi_{23}-\Pi_{23}\Pi_{12} = -K_{x_1}K_{x_2}K_{x_3}\theta_{x_1x_2x_3}\Delta_A(x_1-\xi_1)\times[\Delta_A(x_2-\xi_2)-\Delta_R(x_2-\xi_2)]\Delta_R(x_3-\xi_3) = K_{x_1}K_{x_2}K_{x_3}\theta_{x_1x_2x_3}\Delta_A(x_1-\xi_1)\times\Delta(x_2-\xi_2)\Delta_R(x_3-\xi_3)\neq 0. \quad \text{(B7)}$$

However, $[\Pi_{12},\Pi_{23}]$ is nil potent. This follows by

$$[\Pi_{12},\Pi_{23}]^2=[\Pi_{12},\Pi_{23}](\Pi_{12}\Pi_{23}-\Pi_{23}\Pi_{12}),$$

where

$$[\Pi_{12},\Pi_{23}]\Pi_{12} = K_{x_1}K_{x_2}K_{x_3}\theta_{x_1x_2x_3}\int d\eta_1\,d\eta_2\times\Delta_A(x_1-\eta_1)\Delta(x_2-\eta_2)\Delta_R(x_3-\xi_3)\times K_{\eta_1}K_{\eta_2}\theta_{\eta_1\eta_2}\Delta_A(\eta_1-\xi_1)\Delta_R(\eta_2-\xi_2)=0, \quad \text{(B8)}$$

since twice integrating by parts in $\eta_2$ leaves a vanishing volume term by virtue of

$$K_{\eta_2}\Delta(x_2-\eta_2)=0.$$

In the same manner,

$$[\Pi_{12},\Pi_{23}]\Pi_{23}=0, \quad \text{(B9)}$$

[23] The product formed between any two indexes attached to different $\Pi_{ij}$ is carried out explicitly in Appendix A. There it is demonstrated that only volume terms contribute after twice integrating by parts.

proving

$$[\Pi_{12}, \Pi_{23}]^2 = 0. \tag{B10}$$

To prove that $[\Pi_{12}, \Pi_{23}]$ vanishes on $\mathcal{F}'_{1\cdots m}$, we first must introduce two subspaces of $\mathcal{F}'_{1\cdots m}$. The distributions $\Pi_{12}\Pi_{23}$, $\Pi_{23}\Pi_{12}$ are idempotent. For $\Pi_{12}\Pi_{23}$, we have

$$\begin{aligned}(\Pi_{12}\Pi_{23})^2 &= \Pi_{12}\Pi_{23}\Pi_{12}\Pi_{23}\\ &= [\Pi_{12}, \Pi_{23}]\Pi_{12}\Pi_{23} + \Pi_{23}\Pi_{12}\Pi_{12}\Pi_{23}\\ &= \Pi_{23}\Pi_{12}\Pi_{23} = [\Pi_{23}, \Pi_{12}]\Pi_{23} + \Pi_{12}\Pi_{23}\Pi_{23}\\ &= \Pi_{12}\Pi_{23},\end{aligned}$$

where use has been made of (B8), (B9), and the idempotency of $\Pi_{ij}$. The proof for $(\Pi_{32}\Pi_{12})^2 = \Pi_{32}\Pi_{12}$ goes through in the same manner.

The eigenvalue-zero eigenspaces of $\Pi_{12}\Pi_{23}$ and $\Pi_{23}\Pi_{12}$ can be constructed as follows

$$\mathcal{G}'_{1\cdots m} \equiv \{f_{1\cdots m}(1 - \Pi_{12}\Pi_{23});\ \text{all } f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}\} \tag{B11}$$

and

$$\mathcal{H}'_{1\cdots m} \equiv \{f_{1\cdots m}(1 - \Pi_{32}\Pi_{12});\ \text{all } f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}\}, \tag{B12}$$

where

$$g_{1\cdots m}\Pi_{12}\Pi_{32} = 0 \quad \text{for all} \quad g_{1\cdots m} \in \mathcal{G}'_{1\cdots m} \tag{B13}$$

and

$$h_{1\cdots m}\Pi_{23}\Pi_{12} = 0 \quad \text{for all} \quad h_{1\cdots m} \in \mathcal{H}'_{1\cdots m}. \tag{B14}$$

These are necessarily subspaces of $\mathcal{F}'_{1\cdots m}$ since, for example, all vectors

$$f_{1\cdots m}(1 - \Pi_{12}\Pi_{23}) \in \mathcal{F}'_{1\cdots m} \quad (\text{all } f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}) \tag{B15}$$

by construction. Further, we have

$$\begin{aligned}&g_{1\cdots m}\Pi_{23}\Pi_{12}\\ &\quad = g_{1\cdots m}(1 - \Pi_{12}\Pi_{23})\Pi_{23}\Pi_{12} \qquad \text{[by (B13)]}\\ &\quad = g_{1\cdots m}(\Pi_{23}\Pi_{12} - \Pi_{12}\Pi_{23}\Pi_{12})\\ &\quad = g_{1\cdots m}(\Pi_{23}\Pi_{12} - [\Pi_{12}, \Pi_{23}]\Pi_{12} - \Pi_{23}\Pi_{12}\Pi_{12})\\ &\quad = 0 \quad \text{for all} \quad g_{1\cdots m} \in \mathcal{G}'_{1\cdots m},\end{aligned} \tag{B16}$$

where the last step follows from (B8) and $\Pi_{12}$ idempotent. But this implies

$$\mathcal{G}'_{1\cdots m} \subset \mathcal{H}'_{1\cdots m}. \tag{B17}$$

In the same manner it can be shown that

$$\mathcal{G}'_{1\cdots m} \supset \mathcal{H}'_{1\cdots m}, \tag{B18}$$

which, together with (B17) implies

$$\mathcal{G}'_{1\cdots m} = \mathcal{H}'_{1\cdots m}. \tag{B19}$$

Now, with the aid of (B19), Eq. (B2) can be proved for the last combination $[\Pi_{12}, \Pi_{23}]$. The proof is by contradiction and starts with the assumption

$$f_{1\cdots m}[\Pi_{12}, \Pi_{23}] \neq 0, \quad \text{for some} \quad f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}. \tag{B20}$$

Equation (B20) can also be expressed as

$$\begin{aligned}&f_{1\cdots m}[\Pi_{12}, \Pi_{23}]\\ &\quad = f_{1\cdots m}(\Pi_{12}\Pi_{23} - \Pi_{23}\Pi_{12})\\ &\quad = f_{1\cdots m}(\Pi_{12}\Pi_{23} + [\Pi_{12}, \Pi_{23}]\Pi_{12} - \Pi_{12}\Pi_{23}\Pi_{12})\\ &\quad = f_{1\cdots m}(\Pi_{12}\Pi_{23} - \Pi_{12}\Pi_{23}\Pi_{23}\Pi_{12})\\ &\quad = f_{1\cdots m}\Pi_{12}\Pi_{23}(1 - \Pi_{23}\Pi_{12}) \neq 0,\end{aligned} \tag{B21}$$

using (B8) and the idempotency of $\Pi_{ij}$. This in turn implies that $f_{1\cdots m}\Pi_{12}\Pi_{23}$ has a nonvanishing component in $\mathcal{H}'_{1\cdots m}$ for some $f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}$, since

(i) $f_{1\cdots m}\Pi_{12}\Pi_{23} = f'_{1\cdots m} \in \mathcal{F}'_{1\cdots m}$, by construction,

(ii) $f'_{1\cdots m}(1 - \Pi_{23}\Pi_{12}) \in \mathcal{H}'_{1\cdots m}$, by construction for $f'_{1\cdots m} \in \mathcal{F}'_{1\cdots m}$.

But with (B19) this implies that $f_{1\cdots m}\Pi_{12}\Pi_{23}$ has a nonvanishing component in $\mathcal{G}'_{1\cdots m}$ for some $f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}$. This is a contradiction, since

$$f_{1\cdots m}\Pi_{12}\Pi_{23}(1 - \Pi_{12}\Pi_{23}) = 0, \quad \text{for all} \quad f_{1\cdots m} \in \mathcal{F}'_{1\cdots m} \tag{B22}$$

and (B23) implies

$$f_{1\cdots m}\Pi_{12}\Pi_{23}(1 - \Pi_{12}\Pi_{23}) \neq 0, \quad \text{for some} \quad f_{1\cdots m} \in \mathcal{F}'_{1\cdots m}. \tag{B23}$$

Thus the assumption (B20) is wrong and in fact

$$f_{1\cdots m}[\Pi_{12}, \Pi_{23}] = 0 \quad \text{for all} \quad f_{1\cdots m} \in \mathcal{F}'_{1\cdots m} \tag{B24}$$

concluding the proof of (B2).

## APPENDIX C: EQUALITY OF THE $R_P$ PRODUCT AND THE $R_P$ COMMUTATOR

The $R_P$ product is defined by

$$i^n R_P(J_0; J_1 \cdots J_n) \equiv \sum_{\substack{\text{perm}\\(1\cdots n)}} \Pi_{01\cdots n}[\cdots [J_0, J_1], J_2], \cdots, J_n] \tag{C1}$$

and the $R_P$ commutator by

$$i^n R'_P(J_0; J_1 \cdots J_n) \equiv [\cdots [J_0, J_1]_{R_P}, J_2]_{R_P}, \cdots, J_n]_{R_P}. \tag{C2}$$

The proof of their equivalence follows from the

generating form for $R_P$, valid on $\mathcal{F}'$,

$$iR_P(J_0; J_1 \cdots J_{n+1}) = [R_P(J_0; J_1 \cdots J_n), J_{n+1}]_{R_P}, \tag{C3}$$

which implies

$$R_P = R'_P \quad \text{on} \quad \mathcal{F}'. \tag{C4}$$

Equation (C3) is proved analogously to the proof of $R = R'$,[15]

$$\begin{aligned} &i^n[R_P(J_0; J_1 \cdots J_n), J_{n+1}]_{R_P} \\ &\quad = \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{01\ldots n}[\cdots [J_0, J_1], J_2], \cdots, J_n], J_{n+1}]_{R_P} \\ &\quad = \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{01\ldots n} \sum_{j=0}^{n} [\cdots [J_0, J_1], \cdots, J_{j-1}], \\ &\qquad\qquad \times [J_j, J_{n+1}]_{R_P}], J_{j+1}], \cdots, J_n], \end{aligned}$$

$$\begin{aligned} &i^n[R_P(J_0; J_1 \cdots J_n), J_{n+1}]_{R_P} \\ &\quad = \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{01\ldots n}\{(\Pi_{0,n+1} - \Pi_{1,n+1}) \\ &\qquad \times [\cdots [J_0, J_{n+1}], J_1], \cdots, J_n] \\ &\qquad + \cdots + (\Pi_{l,n+1} - \Pi_{l+1,n+1}) \\ &\qquad \times [\cdots [J_0, J_1], \cdots, J_l], J_{n+1}], J_{l+1}], \cdots, J_n] \\ &\qquad + \cdots + \Pi_{n,n+1}[\cdots [J_0, J_1], J_2], \cdots, J_n], J_{n+1}]\}. \end{aligned} \tag{C5}$$

But, from the properties of the $\Pi_{1\ldots n}$ (Appendix A), we have

$$\begin{aligned} &\Pi_{01\ldots n}(\Pi_{l,n+1} - \Pi_{l+1,n+1}) \\ &\qquad = \Pi_{01\ldots n}\Pi_{l,l+1}(\Pi_{l,n+1} - \Pi_{l+1,n+1}) \\ &\qquad = \Pi_{01\ldots n}\Pi_{l,l+1}\Pi_{l,n+1}(1 - \Pi_{l+1,n+1}) \\ &\qquad = \Pi_{01\ldots n}\Pi_{l,l+1}\Pi_{l,n+1}(\Pi_{n+1,l+1} + B_{l+1,n+1}) \\ &\qquad = \Pi_{01\ldots n}\Pi_{l,n+1,l+1}, \end{aligned} \tag{C6}$$

since

$$\Pi_{l,l+1}\Pi_{l,n+1}B_{n+1,l+1} = 0. \tag{C7}$$

Thus (C5) becomes, using (C6),

$$\begin{aligned} &[i^n R_P(J_0; J_1 \cdots J_n), J_{n+1}]_{R_P} \\ &\quad = \sum_{\substack{\text{perm} \\ (1 \cdots n)}} \Pi_{01\ldots n}\{\Pi_{0,n+1,1}[\cdots [J_0, J_{n+1}], J_1], \cdots, J_n] \\ &\qquad + \cdots + \Pi_{l,n+1,l+1} \\ &\qquad \times [\cdots [J_0, J_1], \cdots, J_l], J_{n+1}], J_{l+1}], \cdots, J_n] \\ &\qquad + \cdots + \Pi_{n,n+1}[\cdots [J_0, J_1], J_2], \cdots, J_n], J_{n+1}]\} \\ &\quad = \sum_{\substack{\text{perm} \\ (1 \cdots n+1)}} \Pi_{01\ldots n+1}[\cdots [J_0, J_1], J_2], \cdots, J_{n+1}] \\ &\quad = i^{n+1}R_P(J_0; J_1 \cdots J_{n+1}). \qquad \text{Q.E.D.} \end{aligned} \tag{C8}$$

## APPENDIX D: CURRENT-MAPPING THEOREM IN FEYNMAN–DYSON THEORY

In this appendix it will be proved that the Feynman–Dyson (FD) current satisfies the relations

$$J_x = \sum_{l=0}^{\infty} \frac{1}{l!} \int (d\xi) R(j_x h_1 \cdots h_l) \tag{D1}$$

$$\equiv \sum_{l=0}^{\infty} \frac{1}{l!} \int d\xi_1 \cdots d\xi_l R(j_x h_{\xi_1} \cdots h_{\xi_l}), \tag{D2}$$

and

$$T_+(J_1 \cdots J_m) = S^*(j_1 \cdots j_m S)_+, \tag{D3}$$

with $S = S_{\mathrm{FD}}$, Eq. (4.18). The relevant ordering theorems for $T$-ordered products to be used are[24]

$$\begin{aligned} &T_\pm(j_{x_1} \cdots j_{x_m} h_{y_1} \cdots h_{y_n}) \\ &\qquad = j_{x_1} T_\pm(j_{x_2} \cdots j_{x_m} h_{y_1} \cdots h_{y_n}) \\ &\qquad\quad \pm [T_\pm(j_{x_2} \cdots j_{x_m} h_{y_1} \cdots h_{y_n}), j_{x_1}]_{\substack{R \\ A}}, \end{aligned} \tag{D4}$$

$$\begin{aligned} &T_\pm(j_{x_1} \cdots j_{x_m} h_{y_1} \cdots h_{y_n}) \\ &\qquad = T_\pm(j_{x_2} \cdots j_{x_m} h_{y_1} \cdots h_{y_n}) j_{x_1} \\ &\qquad\quad \pm [j_{x_1}, T_\pm(j_{x_2} \cdots j_{x_m} h_{y_1} \cdots h_{y_n})]_{\substack{R \\ A}}, \end{aligned} \tag{D5}$$

$$(j_1 \cdots j_m S)_\pm = j_1(j_2 \cdots j_m S)_\pm \pm [(j_2 \cdots j_m S)_\pm, j_1]_{\substack{r \\ a}} \tag{D6}$$

$$= (j_2 \cdots j_m S)_\pm j_1 \pm [j_1, (j_2 \cdots j_m S)_\pm]_{\substack{r \\ a}}, \tag{D7}$$

for

$$\begin{aligned} &[(j_2 \cdots j_m S)_\pm, j_1]_{\substack{r \\ a}} \\ &\equiv \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int (d\xi)[T_\pm(j_{x_2} \cdots j_{x_m} h_{\xi_1} \cdots h_{\xi_n}), j_{x_1}]_{\substack{R \\ A}}, \end{aligned} \tag{D8}$$

and

$$[(j_2 \cdots j_m S)_\pm, j_1]_{\substack{r \\ a}} = [j_1, (j_2 \cdots j_m S)_\pm]_{\substack{a \\ r}}. \tag{D9}$$

In addition we will use

$$\begin{aligned} &[j_x, T_+(h_1 \cdots h_n)]_R \\ &\quad = \sum_{i=1}^{n} T_+(h_1 \cdots h_{i-1}, h_{i+1} \cdots h_n)[j_x, h_i]_R \\ &\qquad + \sum_{i>j=1}^{n} T_+(h_1 \cdots h_{i-1}, h_{i+1} \cdots h_{j-1}, h_{j+1} \cdots h_n) \\ &\qquad \times [[j_x, h_j]_R, h_i]_R \\ &\qquad + \cdots + [\cdots [j_x, h_1]_R, h_2]_R, \cdots, h_n]_R. \end{aligned} \tag{D10}$$

The $R$-product representation for $J_x$, (D1), will be proved first. We have for $S = S_{\mathrm{FD}}$,

$$\begin{aligned} SJ_x &\equiv \frac{i\delta S}{\delta a_x} = (j_x S)_+ \\ &\equiv \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int (d\xi) T_+(j_x h_1 \cdots h_n). \end{aligned} \tag{D11}$$

[24] Reference 15, p. 1704.

But by (D5) and (D10),

$$T_+(j_x h_1 \cdots h_n) = T_+(h_1 \cdots h_n) j_x + [j_x, T_+(h_1 \cdots h_n)]_R,$$

$$\begin{aligned} &T_+(j_x h_1 \cdots h_n) \\ &\quad = T_+(h_1 \cdots h_n) j_x + \sum_{i=1}^{n} T_+(h_1 \cdots h_{i-1}, h_{i+1} \cdots h_n) \\ &\qquad \times [j_x, h_i]_R + \sum_{i>j=1}^{n} T_+(h_1 \cdots h_{i-1}, h_{i+1} \cdots h_{j-1}, \\ &\qquad h_{j+1} \cdots h_n)[[j_x, h_i]_R, h_j]_R \\ &\qquad + \cdots + [\cdots [j_x, h_1]_R, h_2]_R, \cdots, h_n]_R. \end{aligned} \tag{D12}$$

The symmetry under the integral of (D11) in the variables $(\xi_1 \cdots \xi_n)$ allows (D12) to become

$$\begin{aligned} &T_+(j_x h_1 \cdots h_n) \\ &\xrightarrow[i n(1 \cdots n)]{\text{Symmetry}} T_+(h_1 \cdots h_n) j_x + \sum_{l=1}^{n} \frac{n!}{l!\,(n-l)!} T_+(h_{l+1} \cdots h_n) \\ &\qquad \times [\cdots [j_x, h_1]_R, h_2]_R, \cdots, h_l]_R \\ &\longrightarrow \sum_{l=0}^{n} \frac{n!\, i^l}{l!\,(n-l)!} T_+(h_{l+1} \cdots h_n) R(j_x; h_1 \cdots h_l), \end{aligned} \tag{D13}$$

where the last step follows by definition of the $R$ product. Thus for Eq. (D11) we have

$$\begin{aligned} SJ_x = &\sum_{n=0}^{\infty} \sum_{l=0}^{n} \frac{(-i)^n i^l n!}{n!\; l!\,(n-l)!} \\ &\times \int (d\xi) T_+(h_{l+1} \cdots h_n) R(j_x; h_1 \cdots h_l). \end{aligned} \tag{D14}$$

Assuming uniform convergence so that the summations in (D14) can be interchanged, we obtain

$$\begin{aligned} SJ_x = &\sum_{l=0}^{\infty} \frac{1}{l!} \sum_{n=l}^{\infty} \frac{(-i)^{n-l}}{(n-l)!} \\ &\times \int (d\xi) T_+(h_{l+1} \cdots h_n) R(j_x; h_1 \cdots h_l), \end{aligned}$$

$$SJ_x = S \sum_{l=0}^{\infty} \frac{1}{l!} \int (d\xi) R(j_x; h_1 \cdots h_l),$$

or

$$J_x = \sum_{l=0}^{\infty} \frac{1}{l!} \int (d\xi) R(j_x; h_1 \cdots h_l). \qquad \text{Q.E.D.}$$

With this result we can readily prove by induction the mapping relation, (D3). It is true for $n = 1$ by construction. If it is assumed for $n$ it can be proved for $n + 1$. We have, by (D4) with the arguments replaced by $(J_1 \cdots J_{n+1})$,

$$\begin{aligned} &T_+(J_1 \cdots J_{n+1}) \\ &\qquad = J_1 T_+(J_2 \cdots J_{n+1}) + [T_+(J_2 \cdots J_{n+1}), J_1]_R. \end{aligned} \tag{D15}$$

Now, using the implicit $T$-product ordering theorems, (D6) and (D7), we have for $j$ Hermitian

$$\begin{aligned} &J_1 T_+(J_2 \cdots J_{n+1}) \\ &\quad = S^*(j_1 S)_+ S^*(j_2 \cdots j_{n+1} S)_+ \\ &\quad = (j_1 S^*)_- S S^*(j_2 \cdots j_{n+1} S)_+ \\ &\quad = \{S^* j_1 - [S^*, j_1]_r\}(j_2 \cdots j_{n+1} S)_+ \\ &\quad = S^*(j_1 \cdots j_{n+1} S)_+ - S^*[(j_2 \cdots j_{n+1} S)_+, j_1]_r \\ &\qquad - [S^*, j_1]_r (j_2 \cdots j_{n+1} S)_+ \\ &\quad = S^*(j_1 \cdots j_{n+1} S)_+ - [S^*(j_2 \cdots j_{n+1} S)_+, j_1]_r \\ &\quad = S^*(j_1 \cdots j_{n+1} S)_+ - [T_+(J_2 \cdots J_{n+1}), j_1]_r, \end{aligned} \tag{D16}$$

where the last step follows by assumption. Thus (D15) becomes

$$\begin{aligned} &T_+(J_1 \cdots J_{n+1}) \\ &\qquad = S^*(j_1 \cdots j_{n+1} S)_+ - [T_+(J_2 \cdots J_{n+1}), j_1]_r \\ &\qquad\quad + [T_+(J_2 \cdots J_{n+1}), J_1]_R. \end{aligned} \tag{D17}$$

Now if we take the $\theta_{1 \ldots n+1}$ projection of (D17), we obtain

$$\begin{aligned} &\theta_{1 \ldots n+1} T_+(J_1 \cdots J_{n+1}) \\ &\quad = \theta_{1 \ldots n+1} \{S^*(j_1 \cdots j_{n+1} S)_+ - [T_+(J_2 \cdots J_{n+1}), j_1]_r \\ &\qquad\qquad + [T_+(J_2 \cdots J_{n+1}), J_1]_R\}. \end{aligned} \tag{D18}$$

But

$$\theta_{1 \ldots n+1} T_+(J_1 \cdots J_{n+1}) = \theta_{1 \ldots n+1} S^*(j_1 \cdots j_{n+1} S)_+, \tag{D19}$$

since

$$\begin{aligned} &\theta_{1 \ldots n+1}[T_+(J_2 \cdots J_{n+1}), J_1]_R \\ &\qquad = \theta_{1 \ldots n+1} \sum_{i=2}^{n+1} \theta_{i1} T_+(J_2 \cdots [J_i, J_1] \cdots J_{n+1}) = 0, \end{aligned} \tag{D20}$$

by (3.8), and

$$\begin{aligned} &\theta_{1 \ldots n+1}[T_+(J_2 \cdots J_{n+1}), j_1]_r \\ &\quad = \theta_{1 \ldots n+1} \sum_{i=2}^{n+1} T_+(J_2 \cdots [J_i, j_1]_r \cdots J_n) = 0, \end{aligned} \tag{D21}$$

by

$$\begin{aligned} &\theta_{1 \ldots n+1}[J_{x_i}, j_{x_1}] \\ &\qquad = \theta_{1 \ldots n+1} \sum_{l=0}^{\infty} \frac{1}{l!} \int (d\xi)[R(j_{x_i}; h_1 \cdots h_l), j_{x_1}]_R \\ &\qquad = \theta_{1 \ldots n+1} \theta_{i1} \sum_{l=0}^{\infty} \frac{1}{l!} \int (d\xi)[R(j_{x_i}; h_1 \cdots h_l), j_{x_1}]_R \\ &\qquad = 0. \end{aligned} \tag{D22}$$

In proving (D22) we have used the fact that $R(j_x; h_1 \cdots h_l) = 0$ for $t_x < t_1, \cdots, t_l$ which implies

$$[R(j_x; h_1 \cdots h_l), j_y]_R = 0, \quad \text{for} \quad t_x < t_y. \quad \text{(D23)}$$

Proving (D19) also proves the contention, since both $T_+(J_1 \cdots J_{n+1})$ and $S^*(j_1 \cdots j_{n+1}S)_+$ are symmetric in $(1 \cdots n+1)$. That is, for an arbitrary projection $\theta_{\alpha_1 \cdots \alpha_{n+1}}$, we have

$$\theta_{\alpha_1 \cdots \alpha_{n+1}} T_+(J_1 \cdots J_{n+1}) = \theta_{\alpha_1 \cdots \alpha_{n+1}} S^*(j_1 \cdots j_{n+1}S)_+$$

and therefore that

$$T_+(J_1 \cdots J_{n+1}) = S^*(j_1 \cdots j_{n+1}S)_+. \quad \text{Q.E.D.} \quad \text{(D24)}$$

From (D17) we also see that

$$[T_+(J_2 \cdots J_{n+1}), J_1]_R = [T_+(J_2 \cdots J_{n+1}), j_1]_r. \quad \text{(D25)}$$

In addition, taking the Hermitian adjoint of (D24) yields

$$T_-(J_1 \cdots J_n) = (j_1 \cdots j_n S^*)_- S. \quad \text{(D26)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Current Formalism. II. The $S$ Matrix in Perturbation Theory*†

J. G. WRAY‡
*Department of Physics, Syracuse University, Syracuse, New York*

(Received 11 July 1967)

This work has accomplished in the context of asymptotic quantum field theory the following objectives. (1) The $S$-matrix equations of Pugh [R. E. Pugh, Ann. Phys. (N.Y ) **23,** 335 (1961)] and their generalization in the manner of Chen [T. W. Chen, Ann. Phys. (N.Y.) **42,** 476 (1967)] are derived without the aid of an interacting field. (2) A diagrammatic representation of these integrodifferential equations is demonstrated. (3) The problem of boundary conditions for a self-interacting system is solved in perturbation theory. This leads to a finite, divergence free, no cutoff expansion in the physical coupling constant. For renormalizable interactions, the only additional parameter is the physical mass, whereas for nonrenormalizable interactions, uniqueness of the expansion requires additional parameters with increasing order of expansion. (4) The success of the perturbation expansion serves as *a posteriori* justification of the formulation in CF.I., [J. G. Wray, J. Math. Phys. **9,** 537 (1968)] upon which the present work is built. The Π or "generalized step" function and its associated algebra plays the principal technical role in facilitating this work. The crucial Π-ordering relations and theorems developed in CF.I. are reviewed here.

## 1. INTRODUCTION

In the present work the general formalism of CF.I. is used to generate the integro-differential equations of Pugh,[1] facilitate a choice of boundary conditions for finite, unique solutions in perturbation expansion, and discuss the extension due to Chen[2] for nonrenormalizable interactions. A model based on a single self-interacting Hermitian scalar field is worked out in detail for the $\varphi^3$ interaction with reference to a Lagrangian. The generalization to charge, spin, and many-field problems is to be carried out in a later work. The work is motivated by the belief that a finite formulation of quantum field theory exists which contains only physical parameters and no divergences. The approach of Pugh[1] succeeded in yielding finite results to any order of perturbation expansion for renormalizable interactions. However, the work was not completely self-consistent, a problem rectified in the present work. Chen[2] has extended the ideas of Pugh to an operator formalism which can support a perturbation expansion up to any given order for nonrenormalizable interactions. However, by extrapolating to an $S$-matrix formalism, it is demonstrated here that the number of parameters increases with the order of expansion.

The assumptions and notation of the theory were introduced in CF.I.[3] We review the essential points of this work in the present section. The assumptions are the familiar ones, with the notable absence of an asymptotic condition:

(i) Lorentz invariance.
(ii) Strong unitarity.
(iii) Strong Bogoliubov causality.

This is because the interpolating field never appears explicitly, since all expressions are written as functionals of currents and higher derivatives of the $S_{\text{op}}$.

* Based in part on the Ph.D. thesis "The Current Formalism in Asymptotic Quantum Field Theory," Syracuse University, December, 1966, during which time the author was a NASA Trainee.

† This work was supported in part by the National Science Foundation.

‡ Present address: Department of Physics, University of Pittsburgh, Pennsylvania 15213.

[1] R. E. Pugh, Ann Phys. (N.Y.) **23,** 335 (1963).

[2] T. W. Chen, Ann. Phys. (N.Y.) **42,** 476 (1967).

[3] J. G. Wray, J. Math. Phys. **9.** 537 (1968), preceding article.

In proving (D22) we have used the fact that $R(j_x; h_1 \cdots h_l) = 0$ for $t_x < t_1, \cdots, t_l$ which implies

$$[R(j_x; h_1 \cdots h_l), j_y]_R = 0, \quad \text{for} \quad t_x < t_y. \quad \text{(D23)}$$

Proving (D19) also proves the contention, since both $T_+(J_1 \cdots J_{n+1})$ and $S^*(j_1 \cdots j_{n+1}S)_+$ are symmetric in $(1 \cdots n+1)$. That is, for an arbitrary projection $\theta_{\alpha_1 \cdots \alpha_{n+1}}$, we have

$$\theta_{\alpha_1 \cdots \alpha_{n+1}} T_+(J_1 \cdots J_{n+1}) = \theta_{\alpha_1 \cdots \alpha_{n+1}} S^*(j_1 \cdots j_{n+1}S)_+$$

and therefore that

$$T_+(J_1 \cdots J_{n+1}) = S^*(j_1 \cdots j_{n+1}S)_+. \quad \text{Q.E.D.} \quad \text{(D24)}$$

From (D17) we also see that

$$[T_+(J_2 \cdots J_{n+1}), J_1]_R = [T_+(J_2 \cdots J_{n+1}), j_1]_r. \quad \text{(D25)}$$

In addition, taking the Hermitian adjoint of (D24) yields

$$T_-(J_1 \cdots J_n) = (j_1 \cdots j_n S^*)_- S. \quad \text{(D26)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Current Formalism. II. The $S$ Matrix in Perturbation Theory*†

J. G. WRAY‡
*Department of Physics, Syracuse University, Syracuse, New York*

(Received 11 July 1967)

This work has accomplished in the context of asymptotic quantum field theory the following objectives. (1) The $S$-matrix equations of Pugh [R. E. Pugh, Ann. Phys. (N.Y ) **23,** 335 (1961)] and their generalization in the manner of Chen [T. W. Chen, Ann. Phys. (N.Y.) **42,** 476 (1967)] are derived without the aid of an interacting field. (2) A diagrammatic representation of these integrodifferential equations is demonstrated. (3) The problem of boundary conditions for a self-interacting system is solved in perturbation theory. This leads to a finite, divergence free, no cutoff expansion in the physical coupling constant. For renormalizable interactions, the only additional parameter is the physical mass, whereas for nonrenormalizable interactions, uniqueness of the expansion requires additional parameters with increasing order of expansion. (4) The success of the perturbation expansion serves as *a posteriori* justification of the formulation in CF.I., [J. G. Wray, J. Math. Phys. **9,** 537 (1968)] upon which the present work is built. The Π or "generalized step" function and its associated algebra plays the principal technical role in facilitating this work. The crucial Π-ordering relations and theorems developed in CF.I. are reviewed here.

## 1. INTRODUCTION

In the present work the general formalism of CF.I. is used to generate the integro-differential equations of Pugh,[1] facilitate a choice of boundary conditions for finite, unique solutions in perturbation expansion, and discuss the extension due to Chen[2] for nonrenormalizable interactions. A model based on a single self-interacting Hermitian scalar field is worked out in detail for the $\varphi^3$ interaction with reference to a Lagrangian. The generalization to charge, spin, and many-field problems is to be carried out in a later work. The work is motivated by the belief that a finite formulation of quantum field theory exists which contains only physical parameters and no divergences. The approach of Pugh[1] succeeded in yielding finite results to any order of perturbation expansion for renormalizable interactions. However, the work was not completely self-consistent, a problem rectified in the present work. Chen[2] has extended the ideas of Pugh to an operator formalism which can support a perturbation expansion up to any given order for nonrenormalizable interactions. However, by extrapolating to an $S$-matrix formalism, it is demonstrated here that the number of parameters increases with the order of expansion.

The assumptions and notation of the theory were introduced in CF.I.[3] We review the essential points of this work in the present section. The assumptions are the familiar ones, with the notable absence of an asymptotic condition:

(i) Lorentz invariance.
(ii) Strong unitarity.
(iii) Strong Bogoliubov causality.

This is because the interpolating field never appears explicitly, since all expressions are written as functionals of currents and higher derivatives of the $S_{\text{op}}$.

* Based in part on the Ph.D. thesis "The Current Formalism in Asymptotic Quantum Field Theory," Syracuse University, December, 1966, during which time the author was a NASA Trainee.

† This work was supported in part by the National Science Foundation.

‡ Present address: Department of Physics, University of Pittsburgh, Pennsylvania 15213.

[1] R. E. Pugh, Ann Phys. (N.Y.) **23,** 335 (1963).

[2] T. W. Chen, Ann. Phys. (N.Y.) **42,** 476 (1967).

[3] J. G. Wray, J. Math. Phys. **9.** 537 (1968), preceding article.

Some technical points involving the mathematical apparatus of the theory are the following.

All operators of the theory will be defined on the Hilbert space $H$ spanned by all polynomials of the smeared free (in) field,

$$a^{\text{in}}(x) \equiv a(x) \equiv a_x, \tag{1.1}$$

acting on a unique vacuum. The unitarily equivalent space spanned by the out-fields define the $S_{\text{op}}$,

$$a^{\text{out}}(x) = S^* a^{\text{in}}(x) S. \tag{1.2}$$

The functional derivative is carried out with respect to the in-fields and is a means of expressing an operator with some variables extrapolated off the mass shell (m.s.). That is, for $F$ an operator on $H$ with the representation

$$F = \sum_{n=0}^{\infty} \frac{1}{n!} \int d\xi_1 \cdots d\xi_n f_n(\xi_1 \cdots \xi_n) :a_{\xi_1} \cdots a_{\xi_n}:, \tag{1.3}$$

the functional derivative is

$$\frac{\delta F}{\delta a_x} = \sum_{n=0}^{\infty} \frac{1}{n!} \int d\xi_1 \cdots d\xi_n f_{n+1}(x\xi_1 \cdots \xi_n) :a_{\xi_1} \cdots a_{\xi_n}:. \tag{1.4}$$

In fact, carrying on in this manner we see that

$$f_n(x_1 \cdots x_n) = \left\langle \frac{\delta^n F}{\delta a_{x_1} \cdots \delta a_{x_n}} \right\rangle_0 \tag{1.5}$$

for $\langle \cdots \rangle_0$, the vacuum expectation value. However, it is necessary to discriminate between equations that remain valid after functional differentiation and those that do not. These are called strong and weak equations, respectively, with the notation

$$\overset{\text{s}}{=} \rightarrow \text{strong equality},$$

$$\overset{\text{w}}{=} \rightarrow \text{weak equality}.$$

For example, the free fields will be taken to satisfy a weak free-field equation

$$Ka_x \overset{\text{w}}{=} 0, \tag{1.6}$$

for

$$K \equiv \Box - m^2. \tag{1.7}$$

This will enable the commutation of functional and coordinate differentiation. For a more detailed analysis of these points, the reader is referred to CF.I. and the references noted there.

An additional technical aspect concerns the distribution

$$\Pi_{xy} \equiv \Pi_{xy}(xy; \xi\eta) \equiv K_x K_y \theta_{xy} \Delta_A(x - \xi) \Delta_R(y - \eta), \tag{1.8}$$

for

$$\theta_{xy} \equiv \theta(x - y) = \begin{cases} 1, & x^0 > y^0 \\ 0, & x^0 < y^0. \end{cases} \tag{1.9}$$

The product formed between $\Pi_{xy}$ at the same and different space–time points is associative and Abelian on the space $\mathcal{F}' \subset \mathcal{S}'$. This space $\mathcal{F}'$ is defined in CF.I. and, with the exception of m.s. test functions, exhausts the space of left multipliers of the $\Pi_{xy}$ products necessary for this work. But since the $\Pi_{xy}$ vanish on the m.s., multiplicative relations of the $\Pi_{xy}$ established on $\mathcal{F}'$ are trivially satisfied on m.s. test functions. The product law is a convolution such that, for example,

$$\Pi_{xyz} \equiv \Pi_{xy}\Pi_{yz} = \int du \Pi_{xy}(xy; \xi u)\Pi_{yz}(uz; \eta\zeta). \tag{1.10}$$

Chains of

$$\begin{aligned} &\Pi_{ij}(\Pi_{ij} \equiv \Pi_{x_i x_j}), \\ &\Pi_{1 \cdots m} \equiv \Pi_{12}\Pi_{23} \cdots \Pi_{m-1,m}, \end{aligned} \tag{1.11}$$

are formed in the manner of Eq. (1.10) and we are led to the $P$-ordering operation constructed in analogy to the $T$-ordering operation with $\theta_{ij} \rightarrow \Pi_{ij}$,

$$P_+(J_1 \cdots J_m) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots m)}} \Pi_{1 \cdots m} J_1 \cdots J_m. \tag{1.12}$$

It is further proved in CF.I. that the current and $S_{\text{op}}$ derivatives have the current representations on $\mathcal{F}'$,

$$\prod_{i>j=1}^{m} (1 - B_{ij}) \frac{\delta^m J_0}{\delta a_1 \cdots \delta a_m} \overset{\text{s}}{=} R_P(J_0; J_1 \cdots J_m), \tag{1.13}$$

and

$$\prod_{i>j=1}^{m} (1 - B_{ij}) S^* \frac{i^m \delta^m S}{\delta a_1 \cdots \delta a_m} \overset{\text{s}}{=} P_+(J_1 \cdots J_m), \quad [\text{all } m = 1, 2, \cdots], \tag{1.14}$$

for the $R_P$ product given in analogy to the $R$ product,

$$i^m R_P(J_0; J_1 \cdots J_m) \equiv \sum_{\substack{\text{perm} \\ (1 \cdots m)}} \Pi_{01 \cdots m}[\cdots [J_0, J_1], J_2], \cdots J_m]. \tag{1.15}$$

That these equations have at least formal significance follows from the demonstration that the Feynman–Dyson algorithm is a formal solution. In the present work the equations are solved in perturbation expansion for renormalizable interactions with only physical parameters and no divergent expressions. Only self-interacting Hermitian scalar fields will be considered.

It is noted that the advantage of $P$ ordering over $T$ ordering lies in the ability of the $\Pi_{ij}$ function to see $\delta$ functions with up to three time derivatives. This is

adequate for the solution of renormalizable interactions in perturbation expansion, but not for nonrenormalizable interactions. In an effort to present a consistent formulation of nonrenormalizable interactions, a generalization in the manner of Chen has also been presented in CF.I.

$$\prod_{i>j=1}^{m}(1-B_{ij}^{[N]})\frac{\delta^m J_0}{\delta a_1\cdots\delta a_m}\overset{s}{=}R_P^{[N]}(J_0;J_1\cdots J_m), \tag{1.16}$$

and

$$\prod_{i>j=1}^{m}(1-B_{ij}^{[N]})S^*\frac{i^m\delta^m S}{\delta a_1\cdots\delta a_m}\overset{s}{=}P_+^{[N]}(J_1\cdots J_m). \tag{1.17}$$

The $R_P^{[N]}$ and $P_+^{[N]}$ products are identical to the $R_P$ and $P_+$ products with

$$\Pi_{ij}\to\Pi_{ij}^{[N]},$$

while

$$\Pi_{xy}^{[N]}\equiv\Pi_{xy}^{[N]}(xy;\xi\eta)$$
$$\equiv(K_xK_y)^N\theta_{xy}\Delta_A^{(N-1)}(x-\xi)\Delta_R^{(N-1)}(y-\eta), \tag{1.18}$$

and

$$\Delta_{A(R)}^{(N-1)}(x)=\frac{1}{(2\pi)^4}\int_{C_{A(R)}}\frac{d^4p\,e^{ipx}}{(p^2+m)^N}. \tag{1.19}$$

The present work serves to demonstrate that at least in perturbation expansion this framework is not adequate. A finite unique result requires a number of parameters that increases with the order of expansion for nonrenormalizable interactions.

The next section deals with the formulation for renormalizable interactions with a sample calculation in $\varphi^3$ theory. Nonrenormalizable interactions are discussed in Sec. 3 with a summary in Sec. 4.

## 2. THE *S*-MATRIX THEORY OF RENORMALIZABLE INTERACTIONS

The *S*-matrix formalism in asymptotic quantum field theory for renormalizable interactions[1] and its extension to nonrenormalizable interactions[2] are both incomplete. The first because boundary conditions on the integral equation were stated incorrectly, and the latter because these boundary conditions were never stated. In this section, boundary conditions will be proposed for $\varphi^3$ theories, with reference to the interaction Lagrangian, for a self-interacting, scalar, Hermitian field. A sample calculation is carried out in perturbation expansion.

### General Remarks

The *S*-matrix equation for the $n$th order $m$-point function is given by[4]

$$(1-B_{1\ldots m})\omega_{1\cdots m}^{(n)}=\lambda_{1\cdots m}^{(n)}, \tag{2.1}$$

where

$$B_{1\ldots m}\equiv B_{1\ldots m}(x_1\cdots x_m;y_1\cdots y_m)$$
$$\equiv -K_{x_1}\cdots K_{x_m}\sum_{k=1}^{m}\theta_{x_kx_1}\cdots\theta_{x_kx_{k-1}}\theta_{x_kx_{k+1}}\cdots\theta_{x_kx_m}$$
$$\times\Delta_R(x_k-y_k)\prod_{\substack{i=1\\ i\neq k}}^{m}\Delta(x_i-y_i), \tag{2.2}$$

and

$$\omega_{1\cdots m}^{(n)}=\left\langle\frac{i^m\delta^m S}{\delta a_1\cdots\delta a_m}\right\rangle_0^{(n)}. \tag{2.3}$$

The product $B_{1\ldots m}\omega_{1\cdots m}^{(n)}$ is formed by

$$B_{1\ldots m}\omega_{1\cdots m}^{(n)}\equiv B_{1\ldots m}\omega^{(n)}(x_1\cdots x_m)$$
$$=\int B_{1\ldots m}(x_1\cdots x_m;y_1\cdots y_m)$$
$$\times dy_1\cdots dy_m\omega^{(n)}(y_1\cdots y_m). \tag{2.4}$$

The inhomogeneity of (2.1), $\lambda_{1\cdots m}^{(n)}$, is completely determined by the solutions of order $l<n$. The general solution is given by

$$\omega_{1\cdots m}^{(n)}=\lambda_{1\cdots m}^{(n)}+\chi_{1\cdots m}^{(n)}, \tag{2.5}$$

where $\chi_{1\cdots m}^{(n)}$ is a solution of the homogeneous equation

$$(1-B_{1\ldots m})\chi_{1\cdots m}^{(n)}=0. \tag{2.6}$$

The physical solution is determined by fixing $\chi_{1\cdots m}^{(n)}$ with appropriate boundary conditions.

It is interesting to note that the scattering amplitude on the mass shell (or with one leg off the mass shell) is completely contained in the homogeneous term, $\chi_{1\cdots m}^{(n)}$. This is a consequence of the fact that

$$B_{1\ldots m}|_{\text{m.s.(m.s.}-1)}\to 1, \tag{2.7}$$

which implies

$$(1-B_{1\ldots m})\omega_{1\cdots m}^{(n)}|_{\text{m.s.(m.s.}-1)}$$
$$=\lambda_{1\cdots m}^{(n)}|_{\text{m.s.(m.s.}-1)}=0,$$

or that

$$\omega_{1\cdots m}^{(n)}|_{\text{m.s.(m.s.}-1)}=(\lambda_{1\cdots m}^{(n)}+\chi_{1\cdots m}^{(n)})|_{\text{m.s.(m.s.}-1)}$$
$$=\chi_{1\cdots m}^{(n)}|_{\text{m.s.(m.s.}-1)}, \tag{2.8}$$

where m.s. (m.s. − 1) indicates that all momenta (all momenta but one) of the Fourier transform have been restricted to the mass shell. This further implies that for more than one leg off the m.s., the partition of $\omega_{1\cdots m}^{(n)}$ into $\lambda_{1\cdots m}^{(n)}+\chi_{1\cdots m}^{(n)}$ is not in general a Lorentz-invariant separation. To make this statement clear, let us look at the *invariant* homogeneous

[4] Reference 1, p. 342.

solutions in momentum space,

$$\tilde{\chi}(p_1 \cdots p_m) = \left\{\delta\left(\sum_{i=1}^{m} p_i\right) F(p_1 \cdots p_m)\Big|_{\text{all } p_i^2=-m^2};\right. \\ \left. \delta\left(\sum_{i=1}^{m} p_i\right) P(p_1 \cdots p_m)\Big|_{\text{all } p_i^2\neq -m^2}\right., \tag{2.9}$$

where $F(p_1 \cdots p_m)|_{\text{all } p_i^2=-m^2}$ is any invariant tempered distribution constructed from the m.s. momenta, and $P(p_1 \cdots p_m)$ is a polynomial in the invariants formed from the off-m.s. momenta. This polynomial is further restricted by the constraint $l_j + l_k < 4$, where $l_j$, $l_k$ are the powers of any two moments in any one term. Now for the particular case when all momenta are off the m.s., the requirement that both $\lambda_{1\cdots m}^{(n)}$ and $\chi_{1\cdots m}^{(n)}$ be invariants would restrict $\chi_{1\cdots m}^{(n)}$ to its polynomial solutions. But by (2.8) this would imply that the scattering amplitude on the m.s. (m.s. $-$ 1) would become a polynomial in momentum space (point support in coordinate space), or vanish. But only the first-order vertex and two-point function can have these characteristics, respectively. Thus, other than for these two exceptions, $\chi_{1\cdots m}^{(n)}$, and therefore $\lambda_{1\cdots m}^{(n)}$ cannot be invariants with all legs off the m.s. The noninvariant solutions of the homogeneous equation, however, afford a large enough class of functionals to support interactions. For some legs on the m.s. the same discussion will hold with respect to the invariant solutions of the homogeneous equations for the $B$ operator with some legs on the m.s. When all or all but one of the legs are on the m.s., the separation (2.5) becomes trivial by Eq. (2.8). Thus the homogeneous solutions completely determine the m.s. scattering amplitudes and are determined up to symmetric, invariant polynomials of restricted degree ($l_j + l_k < 4$; for $l_j$, $l_k$ the power of any two momenta in any one factor) by the requirement that

$$\omega_{1\cdots m}^{(n)} = \lambda_{1\cdots m}^{(n)} + \chi_{1\cdots m}^{(n)}$$

be Lorentz invariant. It will be shown just how these polynomials can be determined by physically motivated boundary conditions in the next section.

In view of the above discussion, the requirement placed on the $n$th-order vertex function[5]

$$B_{1\cdots m}\underset{\text{vertex}}{\omega_{1\cdots m}^{(n)}} = 0, \quad n > 1,$$

leads to contradictions. This follows, since with (2.7) we have

$$B_{1\cdots m}\underset{\text{vertex}}{\omega_{1\cdots m}^{(n)}}\Big|_{\text{m.s.(m.s.}-1)} = \underset{\text{vertex}}{\omega_{1\cdots m}^{(n)}}\Big|_{\text{m.s.(m.s.}-1)} = 0, \quad n > 1,$$

or

$$\underset{\text{vertex}}{\omega_{1\cdots m}}\Big|_{\text{m.s.(m.s.}-1)} = \sum_{n=0}^{\infty} g^n \underset{\text{vertex}}{\omega_{1\cdots m}^{(n)}}\Big|_{\text{m.s.(m.s.}-1)} \\ = g\underset{\text{vertex}}{\omega_{1\cdots m}^{(1)}}\Big|_{\text{m.s.(m.s.}-1)},$$

a situation similar to that of the Lee model, where the first-order approximation is taken to be the exact vertex.

### Diagrams in $S$-Matrix Theory

It is convenient at this point to generate $\lambda_{1\cdots m}^{(n)}$, the inhomogeneity of the integral equation (2.1), directly from the current formalism, CF.I. Let us define $\Lambda_{1\cdots m}$ as the $(1 - B_{1\cdots m})$ projection of $J_{1\cdots m}$,

$$\Lambda_{1\cdots m} \equiv (1 - B_{1\cdots m})J_{1\cdots m}, \tag{2.10}$$

for

$$J_{1\cdots m} \equiv S^* \frac{i^m \delta^m S}{\delta a_1 \cdots \delta a_m}. \tag{2.11}$$

The useful characteristic of the $(1 - B_{1\cdots m})$ operator as far as perturbation theory is concerned is that it projects out the terms linear in $J_x$ and its derivatives. Thus $\Lambda_{1\cdots m}$ is entirely nonlinear in the currents and current derivatives. This can be seen explicitly using the identity[6]

$$J_{1\cdots m} \overset{s}{=} [J_l + i(\delta/\delta a_l)]J_{1\cdots l-1,l+1\cdots m}, \tag{2.12}$$

and the relation proved in Appendix A,

$$B_{1\cdots m} = \prod_{i>j=1}^{m} B_{ij}. \tag{2.13}$$

The product used in (2.13) is, for example,

$$B_{123} = B_{12}B_{13}B_{23} \\ = \int du_1\, du_2\, du_3 B_{12}(x_1x_2; u_1u_2)B_{13}(u_1x_3; y_1u_3) \\ \times B_{23}(u_2u_3; y_2y_3). \tag{2.14}$$

The ordering of the pairs is immaterial, since we have $[B_{ij}, B_{kl}] =$ zero distribution on

$$\mathcal{F}'(i,j,k,l = 1, \cdots, m; i \neq j, k \neq l), \tag{2.15}$$

since

$$B_{ij} = 1 - (\Pi_{ij} + \Pi_{ji}) \tag{2.16}$$

and the $\Pi_{ij}$ commute on $\mathcal{F}'$.

With (2.13), $(1 - B_{1\cdots m})$ can be expressed

$$(1 - B_{1\cdots m}) = 1 - \prod_{i>j=1}^{m} B_{ij} \\ = 1 - B_{12} + \left(1 - \prod_{\substack{i>j=1 \\ (ij)\neq 1,2}}^{m} B_{ij}\right) B_{12},$$

[5] Reference 1, p. 345.

[6] F. Rohrlich, *Perspective in Modern Physics*, R. E. Marshak, Ed. (Interscience Publishers Inc., New York, 1966).

which upon further expansion yields

$$
\begin{aligned}
&(1 - B_{1\cdots m}) \\
&= (1 - B_{12}) + (1 - B_{13})B_{12} \\
&\quad + (1 - B_{14})B_{12}B_{13} + \cdots + (1 - B_{1m})\prod_{i=2}^{m-1} B_{1i} \\
&\quad + (1 - B_{23})\prod_{i=2}^{m} B_{1i} + \cdots + (1 - B_{m-1,m})\prod_{\substack{i>j=1 \\ j\neq m-1}}^{m} B_{ij}.
\end{aligned}
\tag{2.17}
$$

Now,

$$
\begin{aligned}
&\Lambda_{1\cdots m} \\
&\quad \overset{s}{=} (1 - B_{1\cdots m})J_{1\cdots m} \\
&\quad \overset{s}{=} \Big\{(1 - B_{12}) + (1 - B_{13})B_{12} \\
&\qquad + (1 - B_{14})B_{12}B_{13} + \cdots + (1 - B_{1m})\prod_{i=2}^{m-1} B_{1i} \\
&\qquad + (1 - B_{23})\prod_{i=2}^{m} B_{1i} + \cdots \\
&\qquad + (1 - B_{m-1,m})\prod_{\substack{i>j=1 \\ j\neq m-1}}^{m} B_{ij}\Big\}J_{1\cdots m},
\end{aligned}
\tag{2.18}
$$

which can be seen to be nonlinear in terms $J_{1\cdots l}$, $1 \leq l < m$, and their derivatives. This follows, since for any $r, s = 1, \cdots, m$ $(r \neq s)$,

$$
\begin{aligned}
(1 - B_{rs})J_{1\cdots m} &\overset{s}{=} (1 - B_{rs})\prod_{\substack{j=1 \\ j\neq r,s}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)J_{rs} \\
&\overset{s}{=} \prod_{\substack{j=1 \\ j\neq r,s}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)(1 - B_{rs})J_{rs} \\
&\overset{s}{=} \prod_{\substack{j=1 \\ j\neq r,s}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P(J_rJ_s)
\end{aligned}
\tag{2.19}
$$

is nonlinear in $J_{1\cdots l}$ $(l \geq 1)$ and their derivatives.

Explicitly, with (2.19), (2.18) can be written

$$
\begin{aligned}
\Lambda_{1\cdots m} \overset{s}{=} &\prod_{j=3}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P(J_1J_2) \\
&+ B_{12}\prod_{\substack{j=2 \\ j\neq 1,3}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P(J_1J_3) \\
&+ B_{12}B_{13}\prod_{\substack{j=2 \\ j\neq 4}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P(J_1J_4) + \cdots \\
&+ \prod_{i=2}^{m-1} B_{1i}\prod_{j=2}^{m-1}\left(J_j + \frac{i\delta}{\delta a_j}\right)P(J_1J_m) \\
&+ \prod_{i=2}^{m} B_{1i}\prod_{\substack{j=1 \\ j\neq 2,3}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P(J_2J_3) + \cdots \\
&+ \prod_{\substack{i>j=1 \\ j\neq m-1}}^{m} B_{ij}\prod_{k=2}^{m-2}\left(J_k + \frac{i\delta}{\delta a_k}\right)P(J_{m-1}J_m),
\end{aligned}
\tag{2.20}
$$

proving the nonlinearity of $\Lambda_{1\cdots m}$ for all $m$. Thus in perturbation theory, $\Lambda_{1\cdots m}^{(n)}$ is determined by $J_{1\cdots m}^{(k)}$ all $k < n$, and the inhomogeneity of the integral equation (2.1) is obtained by taking the vacuum expectation value of $\Lambda_{1\cdots m}$,

$$\lambda_{1\cdots m} = \langle\Lambda_{1\cdots m}\rangle_0,$$

which in $n$th order becomes

$$\lambda_{1\cdots m}^{(n)} = \langle\Lambda_{1\cdots m}\rangle_0^{(n)}. \tag{2.21}$$

Next it will be proved that $\lambda_{1\cdots m}^{(n)}$ can be partitioned into terms labeled by the particular Feynman diagram to which they contribute. We have

$$\lambda_{1\cdots m}^{(n)} = \sum_{D\,\epsilon\,\{\text{all diagrams}\}} \lambda_{1\cdots m}^{(n)D}, \tag{2.22}$$

for

$$\lambda_{1\cdots m}^{(n)D} \equiv \text{all terms} \in \lambda_{1\cdots m}^{(n)} \text{ that contribute only to the diagram } D. \tag{2.23}$$

Equation (2.22) can in turn be used to define the integral equation for $\omega_{1\cdots m}^{(n)D}$,

$$(1 - B_{1\cdots m})\omega_{1\cdots m}^{(n)D} = \lambda_{1\cdots m}^{(n)D}, \tag{2.24}$$

where $\omega_{1\cdots m}^{(n)D} \equiv$ solution corresponding to the diagram $D$.

This is consistent with (2.1), since

$$\omega_{1\cdots m}^{(n)} = \sum_{D\,\epsilon\,\{\text{all diagrams}\}} \omega_{1\cdots m}^{(n)D} \tag{2.25}$$

satisfies

$$(1 - B_{1\cdots m})\omega_{1\cdots m}^{(n)} = \lambda_{1\cdots m}^{(n)}$$

by virtue of (2.22).

The piece of $\lambda_{1\cdots m}^{(n)}$ associated with one particular Feynman graph, $\lambda_{1\cdots m}^{(n)D}$, is identified in the formal limit of unrenormalized Feynman–Dyson (FD) theory. That is, $\lambda_{1\cdots m}^{(n)}$ partitions into terms that can be identified by the diagram to which they contribute in unrenormalized FD, a formal solution of Eq. (2.1). Thus, $\lambda_{1\cdots m}^{(n)D}$ is uniquely defined for each diagram. The uniqueness follows, since formally the only difference between the renormalized and unrenormalized solutions is a difference in boundary conditions. But at least if the theory is restricted to a one-vertex model, different boundary conditions, as it will be shown, cannot change the topology or graph characterization of $\lambda_{1\cdots m}^{(n)D}$ and therefore of $\omega_{1\cdots m}^{(n)D}$. It remains to be shown that

(i) The term $\lambda_{1\cdots m}^{(n)D}$ characterizes the unrenormalized Feynman graph $D$.

(ii) Boundary conditions cannot change the topology already presented by $\lambda_{1\cdots m}^{(n)D}$.

A typical term in the expression for $\lambda_{1\cdots m}^{(n)}$ given

by the vacuum of Eq. (2.20) is of the form

$$\mathsf{B}\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\left(J_j+\frac{i\delta}{\delta a_j}\right)P(J_lJ_k)\right\rangle_0^{(n)}$$
$$=\mathsf{B}\Pi_{lk}\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\left(J_j+\frac{i\delta}{\delta a_j}\right)J_lJ_k\right\rangle_0^{(n)}$$
$$+\mathsf{B}\Pi_{kl}\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\left(J_j+\frac{i\delta}{\delta a_j}\right)J_kJ_l\right\rangle_0^{(n)}, \quad (2.26)$$

with $\mathsf{B}$ some $c$-number coefficient composed of a suitable product of $B_{ij}$'s. Further, the vacuum products of (2.26),

$$\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\left(J_j+\frac{i\delta}{\delta a_j}\right)J_lJ_k\right\rangle_0^{(n)},$$

for example, can be expressed as a linear combination of factors, at least bilinear in $\delta^{(\alpha_i)}\mathcal{S}$, defined as

$$\delta^{(\alpha_i)}\mathcal{S}\equiv\left[\prod_{j\in\{\alpha_i\}}\frac{i\delta}{\delta a_j}\right]\mathcal{S}, \quad (2.27)$$

$$\mathcal{S}=S \quad \text{or} \quad S^*. \quad (2.28)$$

That is,

$$\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\left(J_j+\frac{i\delta}{\delta a_j}\right)J_lJ_k\right\rangle_0^{(n)}$$
$$=\sum_{r=2}^{\max}\sum_{\text{sets }\{\alpha_i\}}C\{\alpha_i\}\langle\delta^{(\alpha_1)}\mathcal{S}\delta^{(\alpha_2)}\mathcal{S}\cdots\delta^{(\alpha_r)}\mathcal{S}\rangle_0^{(n)}, \quad (2.29)$$

for the sets, $\{\alpha_i\}$, made up from the integers $1,\cdots,m$. For example, in the simplest case, $m=2$, we have

$$\langle J_1J_2\rangle_0^{(n)}=\left\langle S^*\frac{i\delta S}{\delta a_1}S^*\frac{i\delta S}{\delta a_2}\right\rangle_0^{(n)}$$
$$=\left\langle\frac{-i\delta S^*}{\delta a_1}\frac{i\delta S}{\delta a_2}\right\rangle_0^{(n)}, \quad (2.30)$$

which is of the form of Eq. (2.29). Expanding $\mathcal{S}$ in terms of point functions, the typical term of (2.29) becomes

$$\langle\delta^{(\alpha_1)}\mathcal{S}\cdots\delta^{(\alpha_k)}\mathcal{S}\rangle_0$$
$$=\sum_{r_1,r_2\cdots r_k=1}^{\infty}\frac{(-i)^{\sum_1^k r_j}}{r_1!\,r_2!\cdots r_k!}(dx_1\cdots dx_{r_1})$$
$$\times(dy_1\cdots dy_{r_2})\cdots(dz_1\cdots dr_{r_k})$$
$$\times[s(x_1\cdots x_{r_1}x_{r_1+1}\cdots x_{r_1+\alpha_1})$$
$$\times s(y_1\cdots y_{r_2}y_{r_2+1}\cdots y_{r_2+\alpha_2})\cdots$$
$$\times s(z_1\cdots z_{r_k}z_{r_k+1}\cdots z_{r_k+\alpha_k})]^{(n)}$$
$$\times\langle:a_{x_1}\cdots a_{x_{r_1}}::a_{y_1}\cdots a_{y_{r_1}}:\cdots:a_{z_1}\cdots a_{z_{r_k}}:\rangle_0, \quad (2.31)$$

where

$$s(x_1\cdots x_\beta)\equiv\begin{cases}\omega(x_1\cdots x_\beta), & \text{for } \mathcal{S}=S\\ (-)^\beta\omega^*(x_1\cdots x_\beta), & \text{for } \mathcal{S}=S^*.\end{cases} \quad (2.32)$$

The point functions $s(x_1\cdots x_\beta)$ appearing in (2.31) are all of order $l<n$, since $k\geq 2$ ($k=$ number of point functions appearing in any one product). Thus, if we assume that a diagram representation exists for all point functions of order $l<n$, then (2.31) gives us a natural program for labeling each of its terms by the topology represented by one and only one graph. The vacuum products of (2.31),

$$\langle:a_{x_1}\cdots a_{x_{r_1}}::a_{y_1}\cdots a_{y_{r_2}}:\cdots:a_{z_1}\cdots a_{z_{r_k}}:\rangle_0,$$

reduce to all possible products of $\Delta_+(x_i-y_j)$, which saturate the internal (integration) variables $(x_1\cdots x_{r_1})(y_1\cdots y_{r_2})\cdots(z_1\cdots z_{r_k})$ and connect only variables of different sets. The topology of $n$th order is to be identified by associating a line with each $\Delta_+(x_i-y_j)$. This serves to connect two inserted parts represented by a diagram from each of two-point functions $s(x_1\cdots x_\beta)$ of (2.31). Thus, since each product of $\Delta_+(x_i-y_j)$ will represent a unique pairing of all internal variables between different sets, and all diagrams for $s^{(l)}(x_1\cdots x_\beta)$, $(l<n)$ are known by assumption, this prescription does assign a unique diagram label to each term of (2.31). Now by repeating this analysis for all of $\lambda_{1\cdots m}^{(n)}$, one can group those terms labeled by the same diagram. This grouping then defines $\lambda_{1\cdots m}^{(n)D}$ as the collection of all terms of $\lambda_{1\cdots m}^{(n)}$ labeled by the same diagram $D$.

To prove that this topology has physical content, it must be demonstrated first that apart from introducing new vertexes, the homogeneous solutions cannot alter the topology as defined. This will be adequate, since we are interested here only in single-vertex models. As explained earlier, $\lambda_{1\cdots m}^{(n)D}$ can be used to define $\omega_{1\cdots m}^{(n)D}$ through the integral equation

$$(1-B_{1\ldots m})\omega_{1\cdots m}^{(n)D}=\lambda_{1\cdots m}^{(n)D},$$

with the general solution

$$\omega_{1\cdots m}^{(n)D}=\lambda_{1\cdots m}^{(n)D}+\chi_{1\cdots m}^{(n)D}. \quad (2.33)$$

As previously discussed, $\chi_{1\cdots m}^{(n)D}$ is first of all the Lorentz-invariant completion of $\lambda_{1\cdots m}^{(n)D}$. But from (2.26) the structure of $\lambda_{1\cdots m}^{(n)D}$ is seen to be that of manifestly invariant vacuum products,

$$\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\left(J_j+\frac{i\delta}{\delta a_j}\right)J_lJ_k\right\rangle_0^{(n)},$$

with manifestly noninvariant $c$-number operator

coefficients represented by $B\Pi_{lk}$. The role of $\chi_{1\cdots m}^{(n)D}$ is therefore seen to be that of adding appropriate corrections to these coefficients. Thus, up to invariant polynomial solutions of the homogeneous equation, $\chi_{1\cdots m}^{(n)D}$ will necessarily have the same topology as $\lambda_{1\cdots m}^{(n)D}$, since the topology as it has been defined is all contained in the invariant factors

$$\left\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m} \left(J_j + \frac{i\delta}{\delta a_j}\right) J_l J_k \right\rangle_0^{(n)}.$$

The polynomial terms of $\chi_{1\cdots m}^{(n)D}$ are then to be fixed by boundary conditions in such a way that they cannot introduce new topological structures in the form of additional vertexes. Thus the homogeneous term of (2.33) is so restricted that its diagram topology is the same as $\lambda_{1\cdots m}^{(n)D}$, uniquely determining the topology of $\omega_{1\cdots m}^{(n)D}$ to be that of $\lambda_{1\cdots m}^{(n)D}$.

That the diagram topology defined for $\lambda_{1\cdots m}^{(n)D}$ is in fact a physical partition of $\lambda_{1\cdots m}^{(n)}$ follows from its formal identification with FD in the limit of the unrenormalized solutions. The integral corresponding to any $n$th order $m$-point graph $D$ in unrenormalized FD can be found to correspond to $\lambda_{1\cdots m}^{(n)D}$ in the following way. Each term of $\lambda_{1\cdots m}^{(n)D}$ is determined up to operator coefficients $B\Pi_{kl}$ (symbolically) by a particular graph with inserted diagrams of order $t < n$, connected by lines associated with $\Delta_+$ functions. But up to coefficients of appropriate $\theta$-function products, this same graph can be generated in the unrenormalized FD integral by judicious substitutions,

$$\Delta_c(x_1 - x_2) \to \theta_{x_1x_2}\Delta_+(x_1 - x_2) + \theta_{x_2x_1}\Delta_+(x_2 - x_1)$$

for those lines not contained in the inserted parts of the graph of the $\lambda_{1\cdots m}^{(n)D}$ term. Now since $\chi_{1\cdots m}^{(n)D}$ is equipped to manipulate these coefficients and since it has already been proved that the unrenormalized FD is a formal solution of the equations (CF.I.), it is concluded that

$$\lambda_{1\cdots m}^{(n)D} \xrightarrow{\text{(formally)}} \underset{\text{(FD)}}{\omega_{1\cdots m}^{(n)D}},$$

when the boundary conditions appropriate to unrenormalized FD are applied.

This section has presented a program for generating exact $S$-matrix equations directly from the current formalism. If one resorts to a perturbation expansion, we have demonstrated a diagrammatic approach to the effect that in the limit of unrenormalized FD it reproduces Feynman diagrams. It has also been shown that boundary conditions play a crucial role in determining the topology by defining the vertex function (or functions) of the theory. In the next section physically motivated boundary conditions will be established that lead to the anticipated result of a two-parameter solution to renormalizable theories in perturbation expansion.

### Boundary Conditions in $\varphi^3$ Theory

The problem of establishing physically motivated boundary conditions (b.c.) in a perturbation expansion that leads to a unique solution for the $S$ matrix in terms of the two parameters,

$m$ = renormalized mass,
$g$ = renormalized coupling constant,

has been discussed to some extent in the last two sections. It has been shown that the inhomogeneity of the integral equation $\lambda_{1\cdots m}^{(n)}$ is not in general Lorentz invariant such that the requirements:

(i) Lorentz invariance,
(ii) symmetry of $\omega_{1\cdots m}$ under permutation of its arguments,

determine $\omega_{1\cdots m}^{(n)}$ up to invariant symmetric $\chi_{1\cdots m}^{(n)}$ terms. These are symmetric, invariant polynomials of the transform momenta $p_1 \cdots p_m$ with the restriction that

$$l_j + l_k < 4, \tag{2.34}$$

for $l_j, l_k$ the exponents of any two momenta in any one factor. Thus the role of b.c.'s is that of fixing these polynomials.

A first attempt might be to choose the b.c.

$$\lim_{-p_i^2\to\infty} \tilde{\omega}^{(n)}(p_1 \cdots p_m) \to 0 \quad \text{for all } m \text{ with } \quad n > 1. \tag{2.35}$$

This uniquely determines all solutions, since when a $\tilde{\chi}(p_1 \cdots p_m)$ is determined such that (2.35) holds, only polynomials also satisfying (2.35) can be added. But there are no polynomials satisfying (2.35), so that the solution is unique. However, there are rather serious problems with this solution. Supposedly, there are at least two phenomenologically determined parameters $m$ and $g$ imbedded in the solution. But the b.c. (2.35) has completely determined the theory. There is no arbitrariness left with which to match the parameters with measurements. That is, it is impossible at this point, for example, to require that $\tilde{\omega}_{(p_1p_2)}^{(n)}$ have a zero on the mass shell (m.s.), to fix the mass at the physical value, or to demand that the form factor at some value of its momenta equal the coupling constant, fixing that parameter. Beyond these contradictions it becomes apparent after computation to just 2nd order that the b.c. (2.35) can be satisfied only in a formal sense, since divergent terms necessarily appear. Thus it is clear that not just any b.c. can be applied. In fact, the requirements of the theory, once the first Born term is specified, appear to determine

uniquely the boundary conditions of a consistent theory.

The set of boundary conditions leading to finite results for $\varphi^3$ theory will now be developed. For the 2-point function, the b.c. ($m$ = observed mass),

$$\lim_{-p_1^2\to m^2}\frac{\tilde{\omega}^{(n)}(p_1p_2)}{(p_1^2+m^2)}=0, \tag{2.36}$$

fixes all polynomials and introduces the physical mass shell by requiring that the 2-point function have a 2nd-order zero at $p_1^2=-m^2$. It is a unique result, since there are no polynomials satisfying (2.34) that can have 2nd-order zeros in $p^2$. It is noted that a first-order zero would not have led to unique results, since symmetric invariant polynomials can satisfy (2.34) and have a first-order zero in $p^2$. However, Eq. (2.36) is equivalent to the statement

$$B_{12}\omega_{12}=0, \quad \text{for all } n, \tag{2.37}$$

since $B_{12}$ on any functional that approaches the mass shell as $(p_1^2+m^2)^2$ is always zero. But it is shown in Appendix B that (2.37) is a consequence of the stability of the vacuum and single-particle states. Thus the 2-point function is completely determined by the theory and no additional constraint is necessary.

For the 3-point function, the situation is slightly more complicated. This is because the coupling constant is defined in terms of the proper vertex only, necessitating an identification of proper diagrams before b.c.'s can be applied. The identification of diagrams was discussed in the last section. Improper diagrams are simply connected, while proper diagrams are those remaining.[7] Thus, defining the proper part of $\tilde{\omega}_{123}^{(n)}$ and $\tilde{\lambda}_{123}^{(n)}$, the Fourier transforms of $\omega_{123}^{(n)}$ and $\lambda_{132}^{(n)}$, we have

$$\tilde{\omega}_P^{(n)}(p_1p_2p_3)\equiv\text{proper part of }\tilde{\omega}^{(n)}(p_1p_2p_3), \tag{2.38}$$

$$\tilde{\lambda}_P^{(n)}(p_1p_2p_3)\equiv\text{proper part of }\tilde{\lambda}^{(n)}(p_1p_2p_3), \tag{2.39}$$

satisfying the equation

$$(1-B_{123})\tilde{\omega}_P^{(n)}(p_1p_2p_3)=\tilde{\lambda}_P^{(n)}(p_1p_2p_3). \tag{2.40}$$

The coupling constant is defined by the exact proper vertex function at a particular value of its variables. The point (in phase space) at which the variables are fixed is arbitrary, and we choose the definition

$$\tilde{\omega}_P(p_1p_2p_3)\Big|_{\substack{p_1^2=p_2^2=-m^2\\ p_3^2=0}}\equiv\delta\Big(\sum_{i=1}^3 p_i\Big)\Big|_{\substack{p_1^2=p_2^2=-m^2\\ p_3^2=0}}g. \tag{2.41}$$

Of course, since $\tilde{\omega}_P(p_1p_2p_3)$ is symmetric, the asymmetry between 1, 2, and 3 in (2.41) is only apparent.

[7] J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1959), p. 206.

Now in perturbation theory with

$$\tilde{\omega}_P(p_1p_2p_3)=\sum_{n=0}^{\infty}g^n\tilde{\omega}_P^{(n)}(p_1p_2p_3), \tag{2.42}$$

and

$$g\tilde{\omega}^{(1)}(p_1p_2p_3)\equiv g\delta\Big(\sum_{i=1}^3 p_i\Big), \tag{2.43}$$

Eq. (2.41) implies the boundary condition

$$\text{(iii)}\quad \lim_{p_3^2\to 0}\tilde{\omega}_P^{(n)}(p_1p_2p_3)\Big|_{p_1^2=p_2^2=-m^2}\to 0, \quad n>1. \tag{2.44}$$

Thus with symmetry in $(p_1p_2p_3)$, we have the requirement that $\tilde{\omega}_P^{(n)}(p_1p_2p_3)$ vanishes whenever two legs go on the mass shell and the square of the third goes to zero.

However, (2.44) does not uniquely specify $\tilde{\omega}_P^{(n)}(p_1p_2p_3)$, since there exists invariant symmetric polynomials which satisfy both constraints (2.34) and (2.44). So the question is raised: Is there any additional condition that will fix these polynomials and still introduce no new parameters? The answer is in the affirmative and is given by the high-energy bound[8]

$$\text{(iv)}\quad \lim_{-p_i^2\to\infty}\frac{\tilde{\omega}_P^{(n)}(p_1p_2p_3)}{-p_i^2}\to 0, \quad \text{all } i. \tag{2.45}$$

Additional polynomials are forced to be constants by the high-energy condition (2.45) and the constants are forced to vanish by the zero condition of (2.44). Terms linear in a particular momentum are converted to bilinear ones, e.g., by

$$p_1\cdot p_2\xrightarrow{\delta(\sum_{i=1}^3 p_i)}-p_1\cdot(p_1+p_3)=-p_1^2-p_1\cdot p_3,$$

and therefore do not satisfy the high-energy limit (2.44), i.e.,

$$\lim_{-p_1^2\to\infty}\frac{p_1\cdot p_2}{-p_1^2}\not\to 0.$$

Improper vertex functions are completely determined by requiring that the simply connected parts of

[8] This point was further illuminated by Wilner, who suggested that Eq. (2.45) might actually follow from the LSZ theorem on the vertex. That this is indeed so can be seen by the following. The LSZ theorem says that

$$\lim_{-p_3^2\to\infty}\frac{\omega(p_1p_2p_3)}{(-p_3^2)^{\frac{1}{2}}}\Big|_{p_1^2=p_2^2=-m^2}=0$$

for a $\varphi^3$ interaction. This determines $\chi$ up to terms that are nonvanishing on the $p_1$ and $p_2$ mass shells and independent of $p_3$, i.e.,

$$\chi\sim A+(p_1^2+m^2)B+(p_2^2+m^2)C.$$

But since $\chi$ must be symmetric, we must have

$$\chi\sim A+B\sum_{i=1}^3(p_i^2+m^2),$$

which satisfies the LSZ theorem for $B=0$. Thus $\chi\sim A$, which is determined to be $A=0$ by (2.44), which satisfies the requirement of self-consistency for the definition of the coupling constant.

an improper vertex $\tilde{\omega}^{(n)}_{\text{improper}}(p_1p_2p_3)$ must satisfy the boundary conditions of the proper diagrams to which they correspond. Uniqueness is guaranteed, since additional polynomials must reflect the boundary conditions satisfied by the simply connected proper parts. But we have already seen that these conditions completely determine all polynomials.

For the $m$-point function, $m \geq 4$, the only boundary condition which introduces no new parameters into the theory, yet fixes the polynomial terms in $\chi_{1\cdots m}$, is the one chosen by Pugh,[5]

$$\text{(v)} \quad \lim_{-p_i^2\to\infty} \tilde{\omega}^{(n)}(p_1 \cdots p_m) \to 0, \quad \text{for all } n \text{ and } m > 4. \quad (2.46)$$

That this condition and indeed all high-energy boundary conditions are compatible with existence requirements is best expressed in the language of subtractions. Equation (2.46) is simply the statement that no subtractions are needed in primitive $m$-point diagrams for $m \geq 4$ to all orders in $\varphi^3$ theory. Likewise, (2.45) is the statement that one subtraction is needed to all orders for the primitive diagram of the vertex in $\varphi^3$ theory. This subtraction is necessitated not by existence requirements, however, but rather by the introduction of a coupling constant and the associated self-consistency requirements or boundary conditions. The two-point function satisfies Eq. (2.36), which is equivalent to the requirement of two subtractions, one more than is necessary for existence.

Thus, in summary, the constraints sufficient for a two-parameter solution of the integral equation in $\varphi^3$ theory are:

(i) Lorentz invariance.

(ii) Symmetry of $\omega^{(n)}_{1\cdots m}$ under permutation of its arguments.

$$\text{(iii)} \quad \lim_{p_3^2\to 0} \tilde{\omega}^{(n)}_P(p_1p_2p_3)\big|_{p_1^2=p_2^2=-m^2} = 0, \quad (n > 1)$$

and

$$\tilde{\omega}^{(1)}(p_1p_2p_3) = g\delta\left(\sum_{i=1}^{3} p_i\right).$$

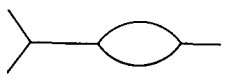
$$\text{(iv)} \quad \lim_{-p_i^2\to\infty} \frac{\tilde{\omega}^{(n)}_I(p_1p_2p_3)}{-p_i^2} \to 0, \quad n > 1.$$

$$\text{(v)} \quad \lim_{-p_i^2\to\infty} \tilde{\omega}^{(n)}(p_1 \cdots p_m) \to 0 \quad \text{for} \quad m \geq 4, \quad \text{all } n.$$

An example is worked out in the next section in order to demonstrate the efficacy of this program.

### Calculation in $\varphi^3$ Theory

To demonstrate the techniques, the contribution to the 3rd-order vertex theory represented by the diagram $D$,

$D \equiv$  (2.47)

will be derived. Starting from current-operator expressions, the integro-differential equation for the 3-point function will be obtained. In perturbation expansion the contribution to the 3rd-order graph $D$ will be identified. Then a unique solution is found by applying the boundary conditions. The homogeneous solution will be determined first as the Lorentz invariant completion of the inhomogeneous term and second by the boundary conditions applicable to an improper vertex function as discussed in the last section.

The equation containing the graph $D$ is the 3-point equation in 3rd order,

$$(1 - B_{123})\omega^{(3)}_{123} = \lambda^{(3)}_{123}. \quad (2.48)$$

Before perturbation expansion, the inhomogeneity can be represented by the vacuum of the operator reduction

$$\begin{aligned}\Lambda_{123} &\equiv (1 - B_{123})J_{123} \\ &= \left(1 - \prod_{i>j=1}^{3} B_{ij}\right)J_{123} \\ &= \left\{\prod_{i>j=1}^{3}(1 - B_{ij}) + \sum_{\substack{\text{cycle}\\(123)}} [B_{12}(1 - B_{23})]\right\}J_{123}.\end{aligned} \quad (2.49)$$

But from (1.14) we have

$$\prod_{i>j=1}^{3}(1 - B_{ij})J_{123} = P(J_1J_2J_3), \quad (2.50)$$

and

$$\begin{aligned}B_{12}(1 - B_{23})J_{123} &= B_{12}(1 - B_{23})(J_1 + i\delta/\delta a_1)J_{23} \\ &= B_{12}(J_1 + i\delta/\delta a_1)(1 - B_{23})J_{23} \\ &= B_{12}(J_1 + i\delta/\delta a_1)P(J_2J_3),\end{aligned} \quad (2.51)$$

such that (2.49) becomes

$$\Lambda_{123} = P(J_1J_2J_3) + \sum_{\substack{\text{cycle}\\(123)}} [B_{12}(J_1 + i\delta/\delta a_1)P(J_2J_3)]. \quad (2.52)$$

Now the inhomogeneity of (2.48) is simply

$$\lambda^{(3)}_{123} = \langle\Lambda_{123}\rangle^{(3)}_0 = \left\langle\left\{P(J_1J_2J_3) + \sum_{\substack{\text{cycle}\\(123)}} B_{12}[(J_1 + i\delta/\delta a_1)P(J_2J_3)]\right\}\right\rangle^{(3)}_0. \quad (2.53)$$

To evaluate (2.53) the solutions of order less than three must first be computed. The first-order assumption determines the vertex and must be a solution of the homogeneous equation, since the inhomogeneity necessarily vanishes. To establish the $\varphi^3$ theory, we assume in coordinate space

$$\omega^{(1)}(x_1x_2x_3) = -\delta(x_1 - x_2)\delta(x_2 - x_3), \quad (2.54)$$

and

$$\omega^{(1)}(x_1 \cdots x_l)\big|_{l\neq 3} = 0, \quad (2.55)$$

such that

$$J_x^{(1)} = \left(S^* \frac{i\delta S}{\delta a_x}\right)^{(1)} = \tfrac{1}{2}{:}a_x^2{:}. \tag{2.56}$$

The second-order solutions are uniquely determined by the theory:

(i) 2-point function:

$$\omega^{(2)}(x_1x_2) = \langle P(J_{x_1}J_{x_2})\rangle_0^{(2)} \tag{2.57}$$

$$= K_{x_1}K_{x_2}\int d\mu^2\rho(\mu^2)\Delta_c(x_1 - x_2\,|\,\mu^2), \tag{2.58}$$

for

$$\rho(\mu^2) = (i/16\pi^2)\theta(\mu^2 - 4m^2)(\mu^2 - m^2)^{-2} \times [(\mu^2 - 4m^2)/\mu^2]^{\frac{1}{2}}; \tag{2.59}$$

(ii) 4-point function:

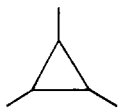
$$\omega^{(2)}(x_1 \cdots x_4) = \tfrac{1}{2}i \sum_{\text{pairs}} [\delta(x_1 - x_2)\delta(x_3 - x_4)\Delta_c(x_1 - x_3)]; \tag{2.60}$$

(iii) $l$-point function:

$$\omega^{(2)}(x_1 \cdots x_l)\big|_{l\neq 2,4} = 0. \tag{2.61}$$

Using (2.56) the factors containing three currents in (2.53) can be seen to contribute only to the triangle graph. Using the rules for defining the diagram topology of a particular factor, we obtain

$$\langle J_1J_2J_3\rangle_0^{(3)} = \tfrac{1}{8}\langle {:}a_1^2{::}a_2^2{::}a_3^2{:}\rangle_0 = (-i)^3\Delta_+(x_1 - x_2)\Delta_+(x_2 - x_3)\Delta_+(x_1 - x_3)$$

$\in$  (2.62)

Thus the contributions to the graph $D$ are all contained in the factors of (2.53) with only two currents. Representing the contributions to this graph by $\lambda^D$, we have

$$\lambda^D \in \left\langle \sum_{\substack{\text{cycle}\\(123)}} \left[ B_{12}\frac{i\delta}{\delta a_3} P(J_2J_3)\right]\right\rangle_0^{(3)}. \tag{2.63}$$

The typical vacuum product we must look at is, for example,

$$X_{123} \equiv \left\langle \frac{i\delta}{\delta a_1}(J_2J_3)\right\rangle_0^{(3)} = \left\langle \frac{i\delta}{\delta a_1}\left[S^*\frac{i\delta S}{\delta a_2}S^*\frac{i\delta S}{\delta a_3}\right]\right\rangle_0^{(3)} = \left\langle \frac{i\delta}{\delta a_1}\left[(-i)\frac{\delta S^*}{\delta a_2}\,i\frac{\delta S}{\delta a_3}\right]\right\rangle_0^{(3)},$$

$$X_{123} = i\left\langle \frac{\delta}{\delta a_1}\left[\frac{\delta S^{*(2)}}{\delta a_2}\frac{\delta S^{(1)}}{\delta a_3} + \frac{\delta S^{*(1)}}{\delta a_2}\frac{\delta S^{(2)}}{\delta a_3}\right]\right\rangle_0. \tag{2.64}$$

Now, from (2.54)–(2.61) and the definition of $S$, we have

$$S^{(1)} = \frac{-i}{3!}\int d\xi\,{:}a_\xi^3{:}, \tag{2.65}$$

$$S^{(2)} = \frac{(-i)^4}{4!}\frac{i}{2}\int d\xi_1 \ldots d\xi_4 \times \sum_{\substack{\text{cycle}\\(1\cdots 4)}} [\delta(\xi_1 - \xi_2)\delta(\xi_3 - \xi_4) \times \Delta_c(\xi_1 - \xi_3)]{:}a_{\xi_1}\cdots a_{\xi_4}{:} + \tfrac{1}{2}(-i)^2\int d\xi_1\,d\xi_2\omega^{(2)}(\xi_1\xi_2){:}a_{\xi_1}a_{\xi_2}{:}$$

$$= \int d\xi_1\,d\xi_2[\tfrac{1}{8}i\Delta_c(\xi_1 - \xi_2){:}a_{\xi_1}^2a_{\xi_2}^2{:} - \tfrac{1}{2}\omega^{(2)}(\xi_1\xi_2){:}a_{\xi_1}a_{\xi_2}{:}], \tag{2.66}$$

$$\delta S^{(1)}/\delta a_{x_1} = -\tfrac{1}{2}i{:}a_{x_1}^2{:}, \tag{2.67}$$

$$\delta S^{(2)}/\delta a_{x_1} \overset{\text{w}}{=} 0 \nearrow = \int d\xi[\tfrac{1}{2}i\Delta_c(x_1 - \xi){:}a_{x_1}a_\xi^2{:} - \omega^{(2)}(x_1\xi)a_\xi], \tag{2.68}$$

$$\delta^2S^{(1)}/\delta a_{x_1}\delta a_{x_2} = -i\delta(x_1 - x_2)a_{x_1}, \tag{2.69}$$

$$\delta^2S^{(2)}/\delta a_{x_1}\delta a_{x_2} = i\Delta_c(x_1 - x_2){:}a_{x_1}a_{x_2}{:} + \tfrac{1}{2}i\delta(x_1 - x_2) \times \int d\xi\Delta_c(x_1 - \xi){:}a_\xi^2{:} - \omega^{(2)}(x_1x_2). \tag{2.70}$$

Substituting these results into (2.64) yields

$$X_{x_1x_2x_3} = i\left\langle\left\{\left[(-i)\Delta_c(x_1 - x_2){:}a_{x_1}a_{x_2}{:} - \tfrac{1}{2}i\delta(x_1 - x_2) \times \int d\xi\Delta_c(x_1 - \xi){:}a_\xi^2{:}\right](-\tfrac{1}{2}i){:}a_{x_3}^2{:} + \tfrac{1}{2}i{:}a_{x_2}^2{:}\left[i\Delta_c(x_1 - x_3){:}a_{x_1}a_{x_3}{:} + \tfrac{1}{2}i\delta(x_1 - x_3)\int d\xi\Delta_c(x_1 - \xi){:}a_\xi^2{:}\right]\right\}\right\rangle_0. \tag{2.71}$$

The vacuum products of (2.71) then reduce to

$$X_{x_1x_2x_3} = i\left[\Delta_c(x_1 - x_2)\Delta_+(x_1 - x_3)\Delta_+(x_2 - x_3) + \delta(x_1 - x_2)\int d\xi\Delta_c(x_1 - \xi)\Delta_+^2(\xi - x_3)\right] + i\left[\Delta_c(x_1 - x_3)\Delta_+(x_2 - x_1)\Delta_+(x_2 - x_3) + \delta(x_1 - x_3)\int d\xi\Delta_c(x_1 - \xi)\Delta_+^2(x_2 - \xi)\right]. \tag{2.72}$$

But only the 2nd and 4th terms have the topology of the graph $D$, (2.47), the remainder having that of the triangle graph, (2.62). If these two terms are now

labeled $X^D_{123}$, the equation for the diagram $D$ becomes, using (2.53), (2.64), and (2.72),

$$(1 - B_{123})\omega^{(3)D}_{(x_1x_2x_3)} = \lambda^{(3)D}_{(x_1x_2x_3)}, \tag{2.73}$$

for

$$\lambda^{(3)D}_{(x_1x_2x_3)} = \sum_{\substack{\text{cycle}\\(123)}} \{B_{12}[\Pi_{23}X^D_{(x_1x_2x_3)} + \Pi_{32}X^D_{(x_1x_2x_3)}]\} \tag{2.74}$$

$$= i \sum_{\substack{\text{cycle}\\(123)}} B_{12}\Big\{\Pi_{23}\int d\xi\Delta_c(x_1 - \xi)[\delta(x_1 - x_2) \times \Delta^2_+(\xi - x_3) + \delta(x_1 - x_3)\Delta^2_+(x_2 - \xi)] + \Pi_{32}\int d\xi\Delta_c(x_1 - \xi)[\delta(x_1 - x_3) \times \Delta^2_+(\xi - x_2) + \delta(x_1 - x_2)\Delta^2_+(x_3 - \xi)]\Big\}. \tag{2.75}$$

Let us define the quantity

$$Y_{(x_1x_2x_3)} \equiv B_{12}\Pi_{23}\delta(x_1 - x_2)\int d\xi\Delta_c(x_1 - \xi)\Delta^2_+(\xi - x_3)$$

$$= B_{12}\Pi_{23}\delta(x_1 - x_2) \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)}\Delta_+(x_2 - x_3 \,|\, \mu^2), \tag{2.76}$$

where $\rho(\mu^2)$ is given by (2.59). But (2.76) can be decomposed further by the separation,

$$Y_{(x_1x_2x_3)} = B_{12}\{\delta(x_1 - x_2)\Pi_{23} + [\Pi_{23}, \delta(x_1 - x_2)]\} \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)}\Delta^2_+(x_2 - x_3 \,|\, \mu^2) \tag{2.77}$$

$$= \delta(x_1 - x_2)K_{x_2}K_{x_3}\theta_{x_2x_3} \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)^3}\Delta_+(x_2 - x_3 \,|\, \mu^2) + \chi^{\mathrm{I}}_{(x_1x_2x_3)}, \tag{2.78}$$

for

$$\chi^{\mathrm{I}}_{(x_1x_2x_3)} \equiv B_{12}[\Pi_{23}, \delta(x_1 - x_2)] \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)}\Delta_+(x_2 - x_3 \,|\, \mu^2). \tag{2.79}$$

The term separated off in (2.78), $\chi^{\mathrm{I}}_{(x_1x_2x_3)}$, is shown to be a solution of the homogeneous equation at the end of this section. The first term of (2.78) has absorbed the multiplier $B_{12}$ since it is an eigenvalue-one eigensolution of $B_{12}$. The factor $\Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)$ of $\Pi_{23}$ has been absorbed under the integral over $\mu^2$ as an increase in the power of the denominator $(m^2 - \mu^2)$.

Proceeding in the same manner, it can be shown that

$$\chi^{\mathrm{II}}_{(x_1x_2x_3)} \equiv B_{12}\Pi_{23}\delta(x_1 - x_3)\int d\xi\Delta_c(x_1 - \xi)\Delta^2_+(x_2 - \xi) \tag{2.80}$$

is also a solution of the $(1 - B_{123})$ homogeneous equation such that with (2.80), (2.78), and (2.76) Eq. (2.75) becomes

$$\lambda^D_{(x_1x_2x_3)} = i \sum_{\substack{\text{cycle}\\(123)}} \Big\{\delta(x_1 - x_2)K_{x_2}K_{x_3} \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)^3}[\theta_{x_2x_3}\Delta_+(x_2 - x_3 \,|\, \mu^2) + \theta_{x_3x_2}\Delta_+(x_3 - x_2 \,|\, \mu^2)] + \chi^{\mathrm{III}}_{(x_1x_2x_3)}\Big\}, \tag{2.81}$$

or

$$\lambda^D_{(x_1x_2x_3)} = i \sum_{\substack{\text{cycle}\\(123)}} \Big\{\delta(x_1 - x_2)K_{x_2}K_{x_3} \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)^3}\Delta_c(x_2 - x_3 \,|\, \mu^2) + \chi^{\mathrm{III}}_{(x_1x_2x_3)}\Big\}, \tag{2.82}$$

for

$$\chi^{\mathrm{III}}_{(x_1x_2x_3)} = \chi^{\mathrm{I}}_{(x_1x_2x_3)} + \chi^{\mathrm{I}}_{(x_1x_3x_2)} + \chi^{\mathrm{II}}_{(x_1x_2x_3)} + \chi^{\mathrm{II}}_{(x_1x_3x_2)}. \tag{2.83}$$

Thus $\lambda^D_{(x_1x_2x_3)}$ is partitioned into an invariant term plus a solution of the homogeneous equation $\chi^{\mathrm{III}}_{(x_1x_2x_3)}$. The particular solution of (2.73),

$$\omega^{(3)D}_{(x_1x_2x_3)} = \lambda^{(3)D}_{(x_1x_2x_3)} + \chi_{(x_1x_2x_3)}, \tag{2.84}$$

is obtained from the general solution with b.c.'s to determine $\chi_{(x_1x_2x_3)}$ up to invariant symmetric solutions of the homogeneous equation, $\chi^{\mathrm{IV}}_{(x_1x_2x_3)}$,

$$\chi_{(x_1x_2x_3)} = -\sum_{\substack{\text{cycle}\\(123)}} \chi^{\mathrm{III}}_{(x_1x_2x_3)} + \chi^{\mathrm{IV}}_{(x_1x_2x_3)}, \tag{2.85}$$

such that with (2.82) and (2.84),

$$\omega^{(3)D}_{(x_1x_2x_3)} = i \sum_{\substack{\text{cycle}\\(123)}} \Big\{\delta(x_1 - x_2)K_{x_2}K_{x_3} \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2 - \mu^2)^3}\Delta_c(x_2 - x_3 \,|\, \mu^2)\Big\} + \chi^{\mathrm{IV}}_{(x_1x_2x_3)}. \tag{2.86}$$

Since the graph $D$ is an improper vertex part, the invariant homogeneous solution $\chi^{\mathrm{IV}}_{(x_1x_2x_3)}$ is to be fixed by requiring that $\omega^{(3)D}_{(x_1x_2x_2)}$ satisfy the b.c.'s implied by its simply connected factors. Thus, since this graph factors into a 1st-order vertex and a 2nd-order self-energy part, we have the b.c. in momentum space

$$\lim_{-p_3^2 \to m^2} \tilde{\omega}^{(1)}(p_1p_2p_3)\tilde{\Delta}_c(p_3 \,|\, m^2)\tilde{\omega}^{(2)}(p_3p_4)$$

$$= C \lim_{-p_3^2 \to m^2} \delta\Big(\sum_{i=1}^{3} p_i\Big)\frac{1}{p_3^2 + m^2 - i\epsilon}(p_3^2 + m^2)^2\delta(p_3 + p_4)$$

$$= C \lim_{-p_3^2 \to m^2} \delta\Big(\sum_{i=1}^{3} p_i\Big)\delta(p_3 + p_4)(p_3^2 + m^2), \tag{2.87}$$

for $C$ a factor independent of $p_3^0$. But the first term of (2.86) does not satisfy this b.c., since it $\to 0$ as $(p_3^2 + m^2)^2$ when $p_3^2 \to -m^2$, thus demanding the following polynomial for $\tilde{\chi}^{\mathrm{IV}}_{p_1p_2p_3}$, where

$$FT[\chi^{\mathrm{IV}}_{(x_1x_2x_3)}] = \tilde{\chi}^{\mathrm{IV}}_{(p_1p_2p_3)}. \tag{2.88}$$

The Fourier transform of (2.86) yields

$$\tilde{\lambda}^{(3)D}_{(p_1p_2p_3)} = \sum_{l=1}^{3}\left\{\delta\left(\sum_{i=1}^{3}p_i\right)(p_l^2+m^2)^2 \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2-\mu^2)^3}\Delta_c(p_l\,|\,\mu^2)\right\} + \tilde{\chi}^{\mathrm{IV}}_{(p_1p_2p_3)}, \tag{2.89}$$

which can be written

$$\begin{aligned}
&\tilde{\lambda}^{(3)D}_{(p_1p_2p_3)} \\
&= \sum_{l=1}^{3}\left\{\delta\left(\sum_{i=1}^{3}p_i\right)(p_l^2+m^2) \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2-\mu^2)^3}(p_l^2+\mu^2+m^2-\mu^2)\Delta_c(p_l\,|\,\mu^2)\right\} \\
&\quad + \tilde{\chi}^{\mathrm{IV}}_{(p_1p_2p_3)} \\
&= \sum_{l=1}^{3}\left\{\delta\left(\sum_{i=1}^{3}p_i\right)(p_l^2+m^2) \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2-\mu^2)^2}\Delta_c(p_l\,|\,\mu^2)\right\} + \tilde{\chi}^{\mathrm{V}}_{(p_1p_2p_3)},
\end{aligned} \tag{2.90}$$

for

$$\tilde{\chi}^{\mathrm{V}}_{(p_1p_2p_3)} \equiv \tilde{\chi}^{\mathrm{IV}}_{(p_1p_2p_3)} + \sum_{l=1}^{3}\left\{\delta\left(\sum_{i=1}^{3}p_i\right)(p_l^2+m^2)\int d\mu^2 \frac{\rho(\mu^2)}{(m^2-\mu^2)^3}\right\}. \tag{2.91}$$

The additional contribution to the homogeneous term (2.91) is just an invariant polynomial solution. The remaining invariant part of (2.90) now satisfies the b.c. (2.87) and we have

$$\tilde{\chi}^{\mathrm{V}}_{(p_1p_2p_3)} = 0. \tag{2.92}$$

The solution, uniquely determined by the b.c. is, therefore,

$$\tilde{\omega}^{(3)D}_{(p_1p_2p_3)} = \sum_{l=1}^{3}\left\{\delta\left(\sum_{i=1}^{3}p_i\right)(p_l^2+m^2) \times \int d\mu^2 \frac{\rho(\mu^2)}{(m^2-\mu^2)^2}\Delta_c(p_l\,|\,\mu^2)\right\}. \tag{2.93}$$

It remains only to prove the contention made earlier that the term separated off in (2.78), $\chi^{\mathrm{I}}_{(x_1x_2x_3)}$, is a solution of the homogeneous equation,

$$(1 - B_{123})\chi^{\mathrm{I}}_{(x_1x_2x_3)} = 0.$$

This follows since the factor $B_{12}[\Pi_{23}, \delta(x_1 - x_2)]$ has only point support in time, with the number of time derivatives satisfying the familiar constraint, $\eta_i + \eta_j < 4$ $(i, j = 1, 2, 3)$, for $\eta_i$, $\eta_j$, the exponents of any two time derivatives in any one term. This can be seen as follows. The $B_{12}$ can be dropped, since it commutes with the $\Pi_{23}$ and reduces to the unit multiplier on $\delta(x_1 - x_2)$. Thus we have for $f_{x_1x_2x_3}$ the factors to the right of the commutator in (2.79),

$$\begin{aligned}
&[\Pi_{23}, \delta(x_1 - x_2)]f_{x_1x_2x_3} \\
&= \Pi_{23}\delta(x_1 - x_2)f_{x_1x_2x_3} - \delta(x_1 - x_2)\Pi_{23}f_{x_1x_2x_3} \\
&= K_{x_2}K_{x_3}\theta_{x_2x_3} \\
&\quad \times \int d\xi_2\, d\xi_3 \Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)\delta(x_1 - \xi_2)f_{x_1\xi_2\xi_3} \\
&\quad - \delta(x_1 - x_2)K_{x_2}K_{x_3}\theta_{x_2x_3} \\
&\quad \times \int d\xi_2\, d\xi_3 \Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)f_{x_1\xi_2\xi_3}.
\end{aligned}$$

Now commute the $K_{x_2}K_{x_3}$ with the $\theta_{x_2x_3}$ to form the factors

$$\begin{aligned}
&[\Pi_{23}, \delta(x_1 - x_2)]f_{x_1x_2x_3} \\
&= \theta_{x_2x_3}\delta(x_1 - x_2)f_{x_1x_2x_3} - \delta(x_1 - x_2)\theta_{x_2x_3}f_{x_1x_2x_3} \\
&\quad + [K_{x_2}K_{x_3}, \theta_{x_2x_3}] \\
&\quad \times \int d\xi_1\, d\xi_2 \Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)\delta(x_1 - \xi_2)f_{x_1\xi_2\xi_3} \\
&\quad - \delta(x_1 - x_2)[K_{x_2}K_{x_3}, \theta_{x_2x_3}] \\
&\quad \times \int d\xi_1\, d\xi_2 \Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)f_{x_1\xi_2\xi_3}.
\end{aligned}$$

But these last two terms are homogeneous solutions of both $(1 - B_{12})$ and $(1 - B_{23})$ on $\mathcal{F}'$, e.g.,

$$\begin{aligned}
&B_{12}[K_{x_2}K_{x_3}, \theta_{x_2x_3}] \\
&\quad \times \int d\xi_2\, d\xi_3 \Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)\delta(x_1 - \xi_2)f_{x_1\xi_2\xi_3} \\
&= B_{12}\{\Pi_{23}\delta(x_1 - x_2) - \theta_{x_2x_3}\delta(x_1 - x_2)\}f_{x_1x_2x_3} \\
&= \{\Pi_{23}B_{12}\delta(x_1 - x_2) - B_{12}\delta(x_1 - x_2)\theta_{x_2x_3}\}f_{x_1x_2x_3} \\
&= [K_{x_2}K_{x_3}, \theta_{x_2x_3}] \\
&\quad \times \int d\xi_2\, d\xi_3 \Delta_A(x_2 - \xi_2)\Delta_R(x_3 - \xi_3)\delta(x_1 - \xi_2)f_{x_1\xi_2\xi_3}.
\end{aligned}$$

Further, we have

$$(1 - B_{23})[K_{x_2}K_{x_3}, \theta_{x_2x_3}] = 0,$$

since $[K_{x_2}K_{x_3}, \theta_{x_2x_3}]$ only has point support in time and satisfies the derivative constraint, $\eta_2 + \eta_3 < 4$. These terms are, however, also homogeneous solutions of $(1 - B_{13})$, since point support in $(x_1^0 - x_3^0)$ is guaranteed by their being homogeneous solutions of $(1 - B_{12})$ and $(1 - B_{23})$. The derivative requirement is satisfied

since there are no derivatives in $x_1^0$, while the derivatives in $x_3^0$ already satisfy the constraint by virtue of its being a homogeneous solution of $(1 - B_{23})$. Thus by (2.13), $[\Pi_{23}, \delta(x_1 - x_2)]f_{x_1x_2x_3}$ is proved a solution of the homogeneous equation in $(1 - B_{123})$.

### Discussion

The boundary conditions for the $\varphi^3$ and $\varphi^4$ models[9] follow naturally from an understanding of the roles played by $\chi_{1\cdots m}$, the homogeneous solution. First, it is the Lorentz-invariant completion of the inhomogeneous term of the integral equation, $\lambda^{(n)}_{1\cdots m}$. Then the invariant polynomial solutions of the homogeneous equation served as subtractions. The number of subtractions is determined essentially by the order of the zeros at finite values of the invariants formed from the momenta or the degree of polynomials appearing for large momenta. Where the subtractions are to be made is determined by the position of the zeros. The number of subtractions in this formalism is necessarily limited to two, since the polynomial solutions of the homogeneous equation are limited in degree by Eq. (2.34). The program of Chen lifts this restriction, but encounters other difficulties as will be discussed in the next section.

The major advantage of deriving the $S$-matrix equations from the current formalism has been the ability to make clear distinction between the noninvariant operator coefficients $B_{ij}\cdots\Pi_{kl}$ and the invariant vacuum products of currents and their derivatives. Since the $B_{ij}\cdots\Pi_{kl}$ coefficients play a role somewhat analogous to $\theta$-function coefficients in FD, it was possible to identify the Feynman graph topology of the integral equation itself. One other advantage is essentially that of any operator formalism. Before applying perturbation theory, one is at liberty to use operator identities to simplify a term as much as possible, reducing it to a minimum number of factors. For example, doing perturbation theory in the form

$$S^*\frac{i\delta S}{\delta a_1}S^*\frac{i\delta S}{\delta a_2}$$

is much more strenuous than the equivalent form (under unitarity),

$$-\frac{i\delta S^*}{\delta a_1}\frac{i\delta S}{\delta a_2}.$$

There is no inherent problem in extending this work to physical models such as quantum electrodynamics and this will probably be done in the near future.

[9] The $\varphi^4$ interaction is discussed in Ref. 3.

## 3. FINITE $S$-MATRIX THEORY

The work of Chen[2] has extended the $S_{\text{op}}$ theory to a wider class of interactions. However, the theory is by no means complete. Though Chen has proved the existence of solutions, there is no mention of what or how many supplementary conditions and parameters must ultimately enter the theory to specify a particular interaction to some order in perturbation expansion. These questions are best discussed in the context of $S$-matrix theory, where the supplementary constraints take the form of boundary conditions on the solutions of an integral equation. The next section, therefore, develops the $S$-matrix theory that is the natural extension of Chen's $S_{\text{op}}$ theory. Then the boundary conditions for self-interacting Hermitian scalar fields represented by the Lagrangian $\varphi^r$ $(r > 4)$ will be formulated. It is found that, though the program is finite to any order of perturbation theory, the traditional difficulty of nonrenormalizable perturbation theories is still present. The number of parameters necessary to determine the solution is an increasing function of the order in perturbation expansion. These parameters enter the theory through boundary conditions.

### Finite $S$-Matrix Theory

By finite $S$-matrix theory, we mean the extension of the Pugh integral equation[4] for the $m$-point function and appropriate boundary conditions in the manner that Chen[2] has extended the $S_{\text{op}}$ formalism. The program is developed in detail, since there are nontrivial steps involving the generalization of $B_{12}^{[N]} \equiv 1 - \Pi_{12}^{[N]} - \Pi_{21}^{[N]}$ to an $m$-point object $B_{1\cdots m}^{[N]}$ and its associated integral equation,

$$(1 - B^{[N]}_{1\cdots m})\omega_{1\cdots m} = \lambda^{[N]}_{1\cdots m}. \tag{3.1}$$

Equation (3.1) must reduce to the nonperturbative form of (2.1) for $N = 1$,

$$(1 - B_{1\cdots m})\omega_{1\cdots m} = \lambda_{1\cdots m}. \tag{2.13}$$

The logical canditate for $B^{[N]}_{1\cdots m}$ is the direct generalization of (2.13),

$$B_{1\cdots m} = \prod_{i>j=1}^{m} B_{ij} \quad \text{on} \quad \mathcal{F}'_{1\cdots m}, \tag{3.2}$$

which for $N \neq 1$ is taken to be

$$B^{[N]}_{1\cdots m} \equiv \prod_{i>j=1}^{m} B^{[N]}_{ij} \quad \text{on} \quad \mathcal{F}'^{[N]}_{1\cdots m}. \tag{3.3}$$

By this definition $B^{[N]}_{1\cdots m}$ is idempotent with an eigenspace to the right which is the natural generalization of the $N = 1$ case. The idempotency is proved by using the commutativity and idempotency of the

$B_{ij}^{[N]}$, on $\mathcal{F}'^{[N]}$,

$$(B_{1\cdots m}^{[N]})^2 = \left(\prod_{i>j=1}^{m} B_{ij}^{[N]}\right)^2 = \prod_{i>j=1}^{m} (B_{ij}^{[N]})^2 = \prod_{i>j=1}^{m} B_{ij}^{[N]} = B_{1\cdots m}^{[N]}. \quad (3.4)$$

The right eigenspace of $B_{1\cdots m}^{[N]}$ is defined by the set of $\chi_{1\cdots m}^{[N]}$ satisfying the homogeneous equation

$$(1 - B_{1\cdots m}^{[N]})\chi_{1\cdots m}^{[N]} = 0. \quad (3.5)$$

A necessary and sufficient condition for $\chi_{1\cdots m}^{[N]}$ to be a solution of (3.5) is that it satisfy

$$(1 - B_{ij}^{[N]})\chi_{1\cdots m}^{[N]} = 0, \quad \text{for all} \quad i > j = 1, \cdots, m. \quad (3.6)$$

That it is sufficient follows from the definition of $B_{1\cdots m}^{[N]}$, Eq. (3.3). That it is necessary follows from the commutativity and idempotency of the $B_{ij}^{[N]}$ on $\mathcal{F}'^{[N]}$. The proof of necessity is by contradiction. Assume (3.6) is not satisfied by one of the $B_{ij}^{[N]}$ $(i < j = 1 \cdots m)$, say $B_{lk}^{[N]}$. Then we have

$$\chi_{1\cdots m}^{[N]} \neq B_{lk}^{[N]}\chi_{1\cdots m}^{[N]} = B_{lk}^{[N]}\left(\prod_{i>j=1}^{m} B_{ij}^{[N]}\right)\chi_{1\cdots m}^{[N]}, \quad (3.7)$$

by assumption and Eqs. (3.5) and (3.3). But

$$B_{lk}^{[N]}\left(\prod_{i>j=1}^{m} B_{ij}^{[N]}\right) = (B_{lk}^{[N]})^2\left(\prod_{\substack{i>j=1 \\ (i,j\neq l,k)}}^{m} B_{ij}^{[N]}\right) = B_{lk}^{[N]}\left(\prod_{\substack{i>j=1 \\ (i,j\neq l,k)}}^{m} B_{ij}^{[N]}\right) = \prod_{i>j=1}^{m} B_{ij}^{[N]} \quad (3.8)$$

such that (3.7) becomes

$$\chi_{1\cdots m}^{[N]} \neq B_{lk}^{[N]}\chi_{1\cdots m}^{[N]} = \left(\prod_{i>j=1}^{m} B_{ij}^{[N]}\right)\chi_{1\cdots m}^{[N]} = \chi_{1\cdots m}^{[N]}, \quad (3.9)$$

a contradiction. Thus (3.6) is also necessary.

Now the solution of

$$(1 - B_{ij}^{[N]})\chi_{ij}^{[N]} = 0 \quad (3.10)$$

is given in momentum space by the set of functionals,[10]

$$\tilde{\chi}_{ij}^{[N]} = \{P(p_i^0, p_j^0)Q(p_i^0 + p_j^0, \vec{p}_i, \vec{p}_j)\}, \quad (3.11)$$

for $P(p_i^0, p_j^0)$ a polynomial restricted by

$$l_i + l_j < 4N, \quad (3.12)$$

where $l_i$, $l_j$ are the exponents of the two momenta in any one term, while $Q$ is an arbitrary distribution restricted to $\mathcal{S}'$. Thus the solution of (3.6) and therefore (3.5) is necessarily of the form

$$\tilde{\chi}_{1\cdots m}^{[N]} = \left\{P(p_1^0 \cdots p_m^0)Q\left[\left(\sum_{i=1}^{m} p_i^0\right), \vec{p}_1 \cdots \vec{p}_m\right]\right\}, \quad (3.13)$$

for $P(p_1^0 \cdots p_m^0)$ a polynomial restricted by (3.12) only with $i, j$ running over all momenta and $Q$ again an arbitrary tempered distribution.

Now in the framework of operator theory, the projection operator $(1 - B_{1\cdots m}^{[N]})$ can facilitate the derivation of the $m$-point equation (3.1) for $N \neq 1$ in the same manner as demonstrated for $N = 1$ in Sec. 2. Thus following Eqs. (2.10)–(2.21) we have the operator form

$$\begin{aligned}\Lambda_{1\cdots m}^{[N]} &\equiv (1 - B_{1\cdots m}^{[N]})J_{1\cdots m} \\ &\overset{s}{=} \left(1 - \prod_{i>j=1}^{m} B_{ij}^{[N]}\right)J_{1\cdots m} \\ &\overset{s}{=} \Big\{(1 - B_{12}^{[N]}) + (1 - B_{13}^{[N]})B_{12}^{[N]} \\ &\quad + (1 - B_{14}^{[N]})B_{12}^{[N]}B_{13}^{[N]} + \cdots \\ &\quad + (1 - B_{m-1,m}^{[N]})\left(\prod_{i>j=1}^{m} B_{ij}^{[N]}\right)\Big\}J_{1\cdots m},\end{aligned} \quad (3.14)$$

with

$$\begin{aligned}(1 - B_{lk}^{[N]})J_{1\cdots m} &\overset{s}{=} (1 - B_{lk}^{[N]})\left[\prod_{\substack{j=1 \\ j\neq l,k}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)\right]J_{lk} \\ &\overset{s}{=} \prod_{\substack{j=1 \\ j\neq l,k}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)(1 - B_{lk}^{[N]})J_{lk} \\ &\overset{s}{=} \prod_{\substack{j=1 \\ j\neq l,k}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P^{[N]}(J_iJ_k),\end{aligned} \quad (3.15)$$

where the last step follows by (1.17), such that (3.14) becomes

$$\begin{aligned}\Lambda_{1\cdots m}^{[N]} \overset{s}{=} &\prod_{j=3}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P^{[N]}(J_1J_2) \\ &+ B_{12}^{[N]}\prod_{\substack{j=1 \\ j\neq 1,3}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P^{[N]}(J_1J_3) \\ &+ B_{12}^{[N]}B_{13}^{[N]}\prod_{\substack{j=1 \\ j\neq 1,4}}^{m}\left(J_j + \frac{i\delta}{\delta a_j}\right)P^{[N]}(J_1J_4) + \cdots \\ &+ \prod_{\substack{i>j=1 \\ j\neq m-1}}^{m} B_{ij}^{[N]}\prod_{k=1}^{m-2}\left(J_k + \frac{i\delta}{\delta a_k}\right)P^{[N]}(J_{m-1}J_m).\end{aligned} \quad (3.16)$$

As a consistency check, we see that this result reduces to (2.20) for $N = 1$. In addition, it is nonlinear in $J$ and its derivatives and therefore nonlinear in $\mathcal{S} = S$ or $S^*$ and its derivative so that in perturbation expansion $\Lambda_{1\cdots m}^{[N](n)}$ ($n$th order) is determined by all solutions of order $l < n$. The integral equation for the $m$-point function, Eq. (3.1), is to be identified with the vacuum of (3.14) yielding

$$(1 - B_{1\cdots m}^{[N]})\omega_{1\cdots m} = \lambda_{1\cdots m}^{[N]} = \langle\Lambda_{1\cdots m}^{[N]}\rangle_0, \quad (3.17)$$

[10] Equation (3.11) is the Fourier transform of the results published in Ref. 2, p. 9.

with $\Lambda^{[N]}_{1\cdots m}$ given by (3.16) and

$$\omega_{1\cdots m} = \langle J_{1\cdots m}\rangle_0. \tag{3.18}$$

The general solution is accordingly

$$\omega_{1\cdots m} = \lambda^{[N]}_{1\cdots m} + \chi^{[N]}_{1\cdots m},$$

with $\chi^{[N]}_{1\cdots m}$ a solution of the homogeneous equation.

The structure of this theory corresponds with the $N = 1$ theory of Sec. 2. Again we have the relation

$$B^{[N]}_{1\cdots m}\Big|_{\text{m.s.(m.s.}-1)} = 1 \tag{3.19}$$

and the mass shell (one leg off the mass shell) amplitudes are again contained completely in the homogeneous terms,

$$\omega_{1\cdots m}\Big|_{\text{m.s.(m.s.}-1)} = \chi^{[N]}_{1\cdots m}\Big|_{\text{m.s.(m.s.}-1)}. \tag{3.20}$$

Thus by the same argument as for $N = 1$, the separation of $\omega_{1\cdots m}$ into two parts is necessarily a noninvariant separation. That is, the *invariant* $\chi^{[N]}_{1\cdots m}$ are restricted to the set

$$\chi^{[N]}_{1\cdots m}\Big|_{\text{invariant}} = \Big\{P(p_1\cdots p_m)\delta\Big(\sum_{i=1}^{m} p_i\Big);\ \delta\Big(\sum_{i=1}^{m} p_i\Big)F(p_1\cdots p_m)\Big|_{\text{m.s.(m.s.}-1)}\Big\}, \tag{3.21}$$

for $P(p_1\cdots p_m)$ a polynomial in the invariants formed from the momenta $(p_1\cdots p_m)$ restricted by (3.12) and $F(p_1\cdots p_m)_{\text{m.s.(m.s.}-1)} \in \mathfrak{S}'$. Thus an invariant off m.s. separation would allow only the polynomial solutions of $\chi^{[N]}_{1\cdots m}$. But this would restrict $\omega_{1\cdots m/ms}$ to polynomials by Eq. (3.20). Therefore $\chi^{[N]}_{1\cdots m}$ is again determined up to invariant polynomials as the Lorentz invariant completion of $\lambda^{[N]}_{1\cdots m}$. These polynomials must then be fixed by appropriate boundary conditions, a question to be discussed in the next section.

The discussion on diagrammatics for $N = 1$ will also be valid here with $N \neq 1$. This follows since $\lambda^{[N](n)}_{1\cdots m}$ ($n$th order) is composed of the same invariant vacuum products of currents and their derivatives, but with different noninvariant coefficients characterized this time by the explicit appearence of the superscript $[N]$. A typical term, e.g., is

$$\begin{aligned}\mathsf{B}^{[N]}&\Big\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\Big(J_j + \frac{i\delta}{\delta a_j}\Big)P^{[N]}(J_lJ_k)\Big\rangle_0^{(n)}\\ &= \mathsf{B}^{[N]}\Pi^{[N]}_{kl}\Big\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\Big(J_j + \frac{i\delta}{\delta a_j}\Big)J_kJ_l\Big\rangle_0^{(n)}\\ &+ \mathsf{B}^{[N]}\Pi^{[N]}_{lk}\Big\langle \prod_{\substack{j=1\\ j\neq l,k}}^{m}\Big(J_j + \frac{i\delta}{\delta a_j}\Big)J_lJ_k\Big\rangle_0^{(n)},\end{aligned} \tag{3.22}$$

with $\mathsf{B}^{[N]}$ some $c$-number coefficient composed of a suitable product of $B^{[N]}_{ij}$'s. But the identification of diagrams was all carried out within the vacuum product such that the discussion is not altered by the presence of different coefficient functions, $\mathsf{B}^{[N]}$. The uniqueness argument also remains, once boundary conditions are chosen to prevent the polynomial solutions of $\chi^{[N]}_{1\cdots m}$ from introducing new diagram topology in the form of vertices. Whether or not the boundary-condition program can be carried out is another question, and will be discussed next.

### Boundary Conditions

In perturbation expansion the equation to be solved is

$$(1 - B^{[N]}_{1\cdots m})\omega^{(n)}_{1\cdots m} = \lambda^{[N](n)}_{1\cdots m}, \tag{3.23}$$

with the general solution

$$\omega^{(n)}_{1\cdots m} = \lambda^{[N](n)}_{1\cdots m} + \chi^{[N](n)}_{1\cdots m}, \tag{3.24}$$

for $\chi^{[N](n)}_{1\cdots m}$ a solution of the homogeneous Eq. (3.5). As already discussed, the requirement that $\omega^{(n)}_{1\cdots m}$ be Lorentz invariant (and symmetric) determines $\chi^{[N](n)}_{1\cdots m}$ up to invariant (symmetric) polynomials in momentum space satisfying the restriction (3.12). It is therefore necessary to introduce boundary conditions (b.c.) in order to determine these polynomials and obtain a unique solution. These b.c.'s enable finite unique results for theories of the form $\varphi^r$ $(r > 4)$, but at the expense of having a number of parameters that is an increasing function of the order in perturbation expansion.

We have established the Feynman diagrams as the topology of a one-vertex problem to any order of the perturbation expansion. Thus the degree of divergence of a particular graph in FD can be used as a guide indicating the number of "subtractions" that must enter in order that the formalism remain finite. For $\varphi^r$ theories the primitively divergent $m$-point graphs in $n$th order are given by[11]

$$\text{for}\quad \begin{aligned} &0 < K \leq -4,\\ &K = -m + (r-4)n. \end{aligned} \tag{3.25}$$

$K$ is the degree of divergence such that, for $K = 0, 1, 2$, etc., the integral is called logarithmically, linearly, quadratically, etc., divergent. Thus for each $m$-point function there exists an order $n$ for which a new primitive divergence appears in FD theory for $\varphi^r$, $r > 4$. Therefore, the b.c. on $\tilde{\omega}^{(n)}(p_1\cdots p_m)$ must be such as to allow subtracted results to appear for $n$ large enough such that $\tilde{\omega}^{(n)}(p_1\cdots p_m)$ becomes primitively divergent in FD. That is, for every $m$ there

[11] Reference 7, p. 205.

exists an $N$ such that for $n > N$, the b.c. leading formally to FD,

$$\lim_{-p_i^2 \to \infty} \tilde{\omega}^{(n)}(p_1 \cdots p_m) \to 0, \tag{3.26}$$

no longer yields finite results. But the only alternative is

$$\lim_{-p_i^2 \to \infty} \frac{\tilde{\omega}^{(n)}(p_1 \cdots p_m)}{(-p_i^2)^l} \to 0, \quad \text{for some } l \geq 1. \tag{3.27}$$

This follows since $l < 1$ determines the same polynomials as $l = 0$; i.e., the requirement that $\chi_{1\cdots m}^{[N](n)}$ be the Lorentz-invariant completion of $\lambda_{1\cdots m}^{[N](n)}$ determines it up to invariant polynomials that must all vanish for the high-energy bounds given by $l \leq 0$. The b.c. (3.27) for $l \geq 1$, on the other hand, determines the solution up to invariant polynomials which are of degree $k$, in the $p_i^0$ component of any one 4-momenta $(p_i; i = 1, \cdots, m)$, where $0 \leq k < 2l$, such that they still satisfy the bound. Now to fix these polynomials it will be necessary to introduce an additional b.c. It must further be a bound for some finite value of the invariant arguments since the $\lim_{-p_i^2 \to \infty}$ has already been exhausted. Thus for some point in the phase space spanned by the $m$ four-vectors $(p_1, \cdots, p_m)$, e.g.,

$$\{p_1 \cdots p_m\} \to \{\alpha_1 \cdots \alpha_m\},$$

where

$$p_i^\mu \to \alpha_i^\mu \quad (i = 1, \cdots, m; \mu = 0, 1, 2, 3), \tag{3.28}$$

$\tilde{\omega}^{(n)}(p_1 \cdots p_m)$ must be fixed in such a way as to determine these polynomials. A sufficient condition would be the requirement that these polynomials have an $l$th order zero at this phase space point. This is so, since invariant polynomials of degree $k < 2l$ in any of the $p_i^0$ $(i = 1, \cdots, m)$ cannot have an $l$th order zero. For example, a boundary condition of this form that uniquely determines the solution (i.e., fixes all polynomials of degree $k < 2l$ in $p_i^0$; $i = 1, \cdots, m$) is

$$\lim_{p_1 \to \alpha_1} \frac{\tilde{\omega}^{(n)}(p_1 \alpha_2 \cdots \alpha_m) - \gamma_m^{(n)}}{(p_1^2 - \alpha_1^2)^{l-1}} = 0, \tag{3.29}$$

for

$$\lim_{p_1 \to \alpha_1} \equiv \lim_{p_1^0 \to \alpha_1^0} \lim_{p_1^1 \to \alpha_1^1} \lim_{p_1^2 \to p_1^2} \lim_{p_1^3 \to p_1^3}.$$

This is because any invariant polynomial satisfying (3.29) must contain the factor $(p_1^2 - \alpha_1^2)^l$, which is of degree $k = 2l$ in $p_1^0$ violating the high-energy constraint (3.27).

The $\gamma_m^{(n)}$ of (3.29) can be reduced to one parameter for all orders through some solution of

$$\gamma_m = \sum_{n=0}^{\infty} g^n \gamma_m^{(n)}, \tag{3.30}$$

such as

$$\gamma_m = \gamma_m^{(l)}, \quad 0 = \gamma_m^{(n)} \quad (n \neq l), \tag{3.31}$$

where $\gamma_m$ is the measured value of

$$\tilde{\omega}(p_1 \cdots p_m)\big|_{\{p_i\} \to \{\alpha_i\}}.$$

Thus a finite unique solution for $\tilde{\omega}_{1\cdots m}^{(n)}$, in the case of a nonrenormalizable interaction of the form

$$L_{\text{int}} = \varphi^r \quad (r > 4),$$

is possible in this framework, but at the expense of introducing at least one parameter for each point function. Thus, since the number of different point functions that necessarily enter the theory is an increasing function of the order of perturbation, so also is the number of parameters required.

Nothing has really been said about the numerical value of the superscript $N$. By (3.12) it determines the degree of polynomial and therefore the number of subtractions entering the theory. Since these are an increasing function of the order of perturbation, calculations to all order would require that $N \to \infty$. But since the number of parameters must increase indefinitely with $N$, the question of existence of

$$\lim_{N \to \infty} B_{1\cdots m}^{[N]}$$

as a distribution in some space becomes academic, For calculations to some finite order, $N$ must be taken large enough to accommodate all the subtractions necessary to that order.

### Discussion

In the preceding sections a program has been demonstrated that can yield finite results to any given order of perturbation expansion for quantum field theories, consistent with the claims of Chen. However, the price for uniqueness is a number of parameters increasing with the order of perturbation. This breakdown of the perturbation expansion does not exclude the possible existence of solutions in closed form or even some other approximation scheme involving, for example, a different parameter of expansion than the one chosen here.

In this regard the work of Güttinger and Pfaffelhuber[12] is interesting. They have produced a one-parameter subtraction convention, applicable to nonrenormalizable interactions, and yielding finite results unique up to infinite series of polynomials (in momentum space). In addition, the claim is made that these polynomials need not be determined, since

[12] W. Güttinger and E. Pfaffelhuber, University of Munich Preprint, 66/562-TH. 660, 1966 (unpublished).

they do not contribute to the scattering amplitudes on the mass shell. There is an inconsistency in this argument, however, when applied to an iterative solution based on Born approximation in the context of the work presented here. That is, if one assumes that to some order of iteration (perturbation), all amplitudes including these polynomials are known, but only on the m.s., then one does not have enough information to generate the next iteration. This follows since without knowledge of the off-m.s. amplitudes of lower order, one cannot carry out the necessary integrals. Thus we have a contradiction and the Born term does not determine the higher-order parts.

## 4. SUMMARY

The Π-functional algebra introduced in CF.I. has been applied to the derivation of the integral equation of Pugh[4] without the construction of an interacting field. In perturbation expansion a diagram representation of the integral equation can be constructed so that it coincides with that of Feynman–Dyson in the formal limit of the unrenormalized solution. This enables a physical construction of boundary conditions and a completion of the theory at least for the $\varphi^3$ and $\varphi^4$ models. The arguments, however, are general enough and can be extended to more complicated renormalizable models. This program is also directly generalizable to the distributions of Chen[2] for which no field representation has yet been derived. For this extension it is further demonstrated that finite unique results require a number of parameters that is an increasing function of the order of perturbation expansion.

There are many problems that remain to be solved. The situation as it stands with respect to nonrenormalizable interactions is not very satisfactory. A solution that in principle requires an arbitrarily large number of parameters, even though finite and unique, cannot be considered meaningful as a theory. However, it can be argued that this is simply due to an expansion in terms of parameters for which no expansion exists. There may still be exact solutions, but how are they to be found? Now that a perturbation expansion for renormalizable interactions exists on a term-by-term basis, the question of its convergence may be intelligently approached.

The algebra of the Π-functions should be generalized to higher-spin systems and many-field systems, and the finite theory of quantum electrodynamics should be completed with respect to the boundary conditions. The implicit $P$-ordering concept must be explored for the possible existence of an algorithm for the renormalized theory, the analog of the formal expression $(e^{-iH})_+$.

## ACKNOWLEDGMENTS

The author wishes to express his appreciation to Professor F. Rohrlich for his encouragement and suggestions during the course of this research. Also, helpful discussions with Drs. T. W. Chen, M. Wilner, M. Wellner, and V. Gorgé are gratefully acknowledged.

## APPENDIX A: THE $B_{ij}$ FACTORIZATION OF $B_{1\dots m}$

Here it will be proved that $B_{1\dots m}$ has the factorization

$$B_{1\dots m} = \prod_{i>j=1}^{m} B_{ij} \quad \text{on} \quad \mathcal{F}'_{1\dots m}. \tag{A1}$$

The restriction to $\mathcal{F}'_{1\dots m}$ comes about through the use of relations such as

$$[B_{iz}, B_{kl}] = 0, \tag{A2}$$

which have been proved only on $\mathcal{F}'$. The defining representation for $B_{1\dots m}$ is given in the text, Eq. (2.2).

To prove (A1) we note that

$$(1 - B_{1\dots m}) \prod_{i>j=1}^{m} B_{ij} = \prod_{i>j=1}^{m} B_{ij} - B_{1\dots m}. \tag{A3}$$

This is so since

$$\begin{aligned} B_{1\dots m}B_{lk} &= B_{1\dots m}(B_{lk}|_{\text{m.s.(m.s.}-1)}), \\ &\qquad (1 \le l \ne k \le m), \\ &= B_{1\dots m}, \end{aligned} \tag{A4}$$

by

$$B_{lk}|_{\text{m.s.(m.s.}-1)} = 1, \tag{A5}$$

where m.s. (m.s. − 1) means all legs on the mass shell (all legs but one on the mass shell). Now this implies

$$\begin{aligned} B_{1\dots m}\prod_{i>j=1}^{m} B_{ij} &= B_{1\dots m}B_{12}B_{13}\cdots B_{lk}\cdots B_{m-1,m} \\ &= B_{1\dots m}B_{13}\cdots B_{lk}\cdots B_{m-1,m}, \\ B_{1\dots m}\prod_{i>j=1}^{m} B_{ij} &= B_{1\dots m}B_{lk}\cdots B_{m-1,m} \\ &= B_{1\dots m}B_{m-1,m} \\ &= B_{1\dots m}. \end{aligned} \tag{A6}$$

A useful representation for $(1 - B_{1\dots m})$ can be obtained from a decomposition of the unit distribution. One starts with

$$\begin{aligned} 1 \equiv \delta(1)\cdots\delta(m) &= (-)^m(K)_m\Delta_R(1)\cdots\Delta_R(m) \\ &= (-)^m(K)_m\left(\sum_{k=1}^{m}\theta_{1k}\theta_{2k}\cdots\theta_{kk-1}\theta_{kk+1}\cdots\theta_{mk}\right) \\ &\quad \times \Delta_R(1)\cdots\Delta_R(m), \end{aligned} \tag{A7}$$

for

$$\begin{aligned}\delta(i) &\equiv \delta(x_i - y_i)\\ (K)_m &\equiv K_{x_1}\cdots K_{x_m}\\ \Delta_R(i) &\equiv \Delta_R(x_i - y_i),\end{aligned} \tag{A8}$$

and

$$\theta_{ik} = \theta(x_i - x_k).$$

Now with the substitution

$$\Delta_R(i) = \Delta(i) + \Delta_A(i), \quad i \neq k, \tag{A9}$$

followed by appropriate grouping of terms, we obtain

$$\begin{aligned}1 = (-)^m(K)_m \sum_{k=1}^{m} \{&(\theta_{1k}\theta_{2k}\cdots\theta_{kk-1}\theta_{kk+1}\cdots\theta_{mk})\\ &\times(\Delta(1)+\Delta_A(1))\cdots(\Delta(k-1)+\Delta_A(k-1))\\ &\times(\Delta(k+1)+\Delta_A(k+1))\cdots(\Delta(m)+\Delta_A(m))\Delta_R(k)\}.\end{aligned}$$

A further regrouping of terms yields

$$\begin{aligned}1 = B_{1\cdots m} + (-)^m(K)_m \sum_{k=1}^{m} \Big\{&(\theta_{1k}\theta_{2k}\cdots\theta_{kk-1}\\ &\times\theta_{kk+1}\cdots\theta_{mk})\sum_{l=1}^{m-1}\sum_{(\text{sets})}[(\Delta_A)^l(\Delta)^{m-l-1}]\Delta_R(k)\Big\}.\end{aligned} \tag{A10}$$

Here $B_{1\cdots m}$ has been identified as a particular separation of terms with one inhomogeneous $\Delta_R$ and $m-1$ homogeneous $\Delta$ functions, while

$$\begin{aligned}\sum_{\text{sets}}[(\Delta_A)^l(\Delta)^{m-l-1}] \equiv &\sum_{\substack{\text{perm}\\(\alpha_1\cdots\alpha_{m-1})}} \frac{1}{l!\,(m-l-1)!}[\Delta_A(\alpha_1)\cdots\Delta_A(\alpha_l)\\ &\times\Delta(\alpha_l+1)\cdots\Delta(\alpha_{m-1})],\end{aligned} \tag{A11}$$

for

$$(\alpha_1\cdots\alpha_{m-1}) = (1,\cdots,k-1,k+1,\cdots,m).$$

From (A10) we have the desired representation,

$$(1 - B_{1\cdots m}) = \sum_{k=1}^{m}\sum_{l=1}^{m-1}\sum_{\text{sets}} X_{lk}, \tag{A12}$$

for

$$\begin{aligned}X_{lk} = (-)^m(K)_m&\theta_{1k}\\ &\times\theta_{2k}\cdots\theta_{kk-1}\theta_{kk+1}\cdots\theta_{mk}(\Delta_A)^l(\Delta)^{m-l-1}\Delta_R(k).\end{aligned}$$

Now look at one term in the triple sum of (A12)

$$\begin{aligned}X_{lm} = (-)^m(K)_m&\theta_{1m}\theta_{2m}\cdots\theta_{m-1,m}\Delta_A(1)\cdots\Delta_A(l)\\ &\times\Delta(l+1)\cdots\Delta(m-1)\Delta_R(m).\end{aligned} \tag{A13}$$

This is an eigenvector to the left of $\Pi_{1m} + \Pi_{m1} = 1 - B_{1m}$, since

$$\begin{aligned}X_{lm}&(\Pi_{1m}+\Pi_{m1})\\ &= (-)^m(K)_m\theta_{1m}\theta_{2m}\cdots\theta_{m-1,m}\\ &\quad\times\int d\xi_1\,d\xi_2\Delta_A(x_1-\xi_1)\cdots\Delta_A(l)\\ &\quad\times\Delta(l+1)\cdots\Delta(m-1)\Delta_R(x_m-\xi_m)\\ &\quad\times K_{\xi_1}K_{\xi_m}\{\theta_{\xi_1\xi_m}\Delta_A(\xi_1-y_1)\Delta_R(\xi_m-y_m)\\ &\quad+\theta_{\xi_m\xi_1}\Delta_A(\xi_m-y_m)\Delta_R(\xi_1-y_1)\},\end{aligned}$$

which becomes, after twice integrating by parts,

$$\begin{aligned}X_{lm}&(\Pi_{1m}+\Pi_{m1})\\ &= (-)^m(K)_m\theta^2_{x_1x_m}\theta_{2m}\cdots\theta_{m-1,m}\\ &\quad\times[\Delta_A(x_1-y_1)\Delta_A(2)\cdots\Delta_A(l)\\ &\quad\times\Delta(l+1)\cdots\Delta(m-1)\Delta_R(x_m-y_m)]\\ &\quad+(-)^m(K)_m\theta_{x_1}\theta_{x_m}\theta_{x_mx_1}\theta_{2m}\cdots\theta_{m-1,m}\\ &\quad\times[\Delta_R(x_1-y_1)\Delta_A(2)\cdots\Delta_A(l)\\ &\quad\times\Delta(l+1)\cdots\Delta(m-1)\Delta_A(x_m-y_m)]\\ &= (-)^m(K)_m\theta_{1m}\theta_{2m}\cdots\theta_{m-1,m}\\ &\quad\times[\Delta_A(1)\cdots\Delta_A(l)\Delta(l+1)\cdots\Delta(m-1)\Delta_R(m)]\\ &= X_{lm},\end{aligned} \tag{A14}$$

for $\theta_{1m}\theta_{m1} = 0$ and $\theta_{1m}\theta_{1m} = \theta_{1m}$.

In the same manner it follows that

$$\begin{aligned}X_{lm} &= X_{lm}(\Pi_{1m}+\Pi_{m1})(\Pi_{2m}+\Pi_{m2})\cdots(\Pi_{lm}+\Pi_{ml})\\ &= X_{lm}(1-B_{1m})(1-B_{2m})\cdots(1-B_{lm}).\end{aligned} \tag{A15}$$

But this implies

$$\begin{aligned}X_{lm}&\prod_{i>j=1}^{m} B_{ij}\\ &= X_{lm}(1-B_{1m})(1-B_{2m})\cdots(1-B_{lm})\prod_{i>j=1}^{m}B_{ij}\\ &= 0, \quad l \geq 1,\end{aligned} \tag{A16}$$

by

$$\prod_{i>j=1} B_{ij} = B_{lm}\prod_{i>j=1}^{m} B_{ij} \quad \text{all} \quad l = 1, 2, \cdots, m-1,$$

and

$$(1 - B_{lm})B_{lm} = 0.$$

Now, since this analysis can be made for an arbitrary term of (A12) and since $l \geq 1$ for all terms, we have

$$(1 - B_{1\cdots m})\prod_{i>j=1}^{m} B_{ij} = \sum_{k=1}^{m}\sum_{l=1}^{m-1}\sum_{\text{sets}} X_{lk}\prod_{i>j=1}^{m} B_{ij} = 0. \tag{A17}$$

Equations (A3) and (A17) imply

$$\prod_{i>j=1}^{m} B_{ij} = B_{1\cdots m},$$

concluding the proof.

## APPENDIX B: REMARKS ON THE TWO-POINT FUNCTION

In this appendix it will be demonstrated that

$$B_{12}\omega_{12} = 0, \tag{B1}$$

for the case of renormalizable theories. Equation (B1) was first proposed by Pugh,[13] but the proof was based on $\omega(p_1p_2)|_{\text{m.s.}} = 0$. This is not enough since, for example, $(p_1^2 + m^2)|_{\text{m.s.}} = 0$, but is a solution of the homogeneous equation and therefore does not

[13] Reference 1, p. 347.

satisfy (B1). The proof does follow, however, from the stability of the vacuum and single-particle state.

We have for $B_{12}\omega_{12}$

$$B_{12}\omega_{12} = K_{x_1}K_{x_2}\theta_{x_1x_2}\int d^4y_1\, d^4y_2[\Delta(x_1-y_1)\Delta_R(x_2-y_2) - \Delta_R(x_1-y_1)\Delta(x_2-y_2)]\cdot K_{y_1}K_{y_2}\langle\Phi(A_{y_1}A_{y_2})\rangle_0. \tag{B2}$$

Integration by parts twice in $y_1$ yields

$$\begin{aligned} B_{12}\omega_{12} = {} & K_{x_1}K_{x_2}\theta_{x_1x_2} \\ & \times\int d^4y_1\, d^4y_2\{K_{y_1}[\Delta(x_1-y_1)\Delta_R(x_2-y_2) \\ & - \Delta_R(x_1-y_1)\Delta(x_2-y_2)]\}\cdot K_{y_2}\langle\Phi(A_{y_1}A_{y_2})\rangle_0 \\ & - \left(\lim_{y_1^0\to\infty} - \lim_{y_1^0\to-\infty}\right)K_{x_1}K_{x_2}\theta_{x_1x_2} \\ & \times\int d^3y_1\, d^4y_2[\Delta(x_1-y_1)\Delta_R(x_2-y_2) \\ & - \Delta_R(x_1-y_1)\Delta(x_2-y_2)] \\ & \times\overleftrightarrow{\partial}_{y_1^0}K_{y_2}\langle\Phi(A_{y_1}A_{y_2})\rangle_0, \end{aligned}$$

which equals

$$\begin{aligned} B_{12}\omega_{12} = {} & K_{x_1}K_{x_2}\theta_{x_1x_2}\int d^4y_2\Delta(x_2-y_2)K_{y_2}\langle\Phi(A_{x_1}A_{y_2})\rangle_0 \\ & - K_{x_1}K_{x_2}\theta_{x_1x_2} \\ & \times\int d^3y_1\, d^4y_2\Delta(x_1-y_1)\Delta_R(x_2-y_2)\overleftrightarrow{\partial}_{y_1^0}K_{y_2} \\ & \times\langle\{a^{\text{out}}_{y_1}A_{y_2}|_{y_1^0=0} - A_{y_2}a^{\text{in}}_{y_1}|_{y_1^0=-\infty}\}\rangle_0 \\ & - K_{x_1}K_{x_2}\theta_{x_1x_2}\int d^3y_1\, d^4y_2\Delta(x_1-y_1) \\ & \times\Delta(x_2-y_2)\overleftrightarrow{\partial}_{y_1^0}K_{y_2}\langle A_{y_2}a^{\text{in}}_{y_1}\rangle_0|_{y_1^0=-\infty} \\ = {} & K_{x_1}K_{x_2}\theta_{x_1x_2}\int d^4y_2[\Delta(x_2-y_2)K_{y_2}\langle\Phi(A_{x_1}A_{y_2})\rangle_0 \\ & - \Delta_R(x_2-y_2)\cdot\langle(a^{\text{out}}_{x_1}J_{y_2} - J_{y_2}a^{\text{in}}_{x_1})\rangle_0 \\ & - \Delta(x_2-y_2)\langle J_{y_2}a^{\text{in}}_{x_1}\rangle_0]. \end{aligned} \tag{B3}$$

The last term of (B3) vanishes by the stability of the single-particle state, since

$$\int d^4y_2\Delta(x_2-y_2)\langle 0|\, J_{y_2} = \langle 0|\,(a^{\text{in}}_{x_2} - a^{\text{out}}_{x_2}) = 0. \tag{B4}$$

This leaves the first two terms which, after integration by parts in the first term and substitution of fields for the integral over the current in the second term, becomes

$$\begin{aligned} & B_{12}\omega_{12} \\ & = K_{x_1}K_{x_2}\theta_{x_1x_2}\{-\langle(a^{\text{out}}_{x_2}A_{x_1} - A_{x_1}a^{\text{in}}_{x_2})\rangle_0 - i\Delta(x_2-x_1) \\ & \quad + \langle[a^{\text{out}}_{x_1}(A_{x_2} - a^{\text{in}}_{x_2}) - (A_{x_2} - a^{\text{in}}_{x_2})a^{\text{in}}_{x_1}]\rangle_0\}. \end{aligned} \tag{B5}$$

But $\langle(a^{\text{out}}_{x_2}A_{x_1} - A_{x_1}a^{\text{in}}_{x_2})\rangle_0 = \langle[a_{x_2}, A_{x_1}]\rangle_0$ by (B4) for $a^{\text{in}}_x = a_x$. Further, only the single-particle state contributes to a sum over intermediate states

$$\begin{aligned} \langle 0|\, a_{x_1}A_{x_2}\,|0\rangle &= \langle 0|\, a_{x_1}\,|1\rangle\langle 1|\, A_{x_2}\,|0\rangle \\ &= \langle 0|\, a_{x_1}\,|1\rangle\langle 1|\, a_{x_2}\,|0\rangle, \end{aligned}$$

such that

$$\langle(a^{\text{out}}_{x_2}A_{x_1} - A_{x_1}a^{\text{in}}_{x_2})\rangle_0 = [a_{x_2}, a_{x_1}], \tag{B6}$$

and (B5) becomes

$$\begin{aligned} B_{12}\omega_{12} = {} & K_{x_1}K_{x_2}\theta_{x_1x_2}\{-[a_{x_2}, a_{x_1}] - i\Delta(x_2-x_1) \\ & + \langle[a_{x_1}(a_{x_2} - a_{x_2}) - (a_{x_2} - a_{x_2})a_{x_1}]\rangle_0 = 0. \end{aligned} \tag{B7}$$

The result was proved independently by Wilner[14] in momentum space using the assumption of a spectrum gap between the single particle state and continuum for the spectral representation of the 2-point function. There it becomes clear that (B1) holds, because of the 2nd-order zero on the mass shell as expressed in the 2-point boundary condition, Eq. (2.36).

[14] M. Wilner (private communication).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Point Transformation of Classical Hard-Core Potential*

MARTIN J. COOPER†
*Brandeis University, Waltham, Massachusetts*
AND
*National Bureau of Standards, Washington, D.C.*

(Received 1 August 1967)

The method of an extended canonical point transformation is used to reformulate the singular repulsions in a classical hard-sphere gas as equivalent velocity-dependent interactions. The approach provides a Hamiltonian in which the repulsions appear as nonlocal potential interactions between the particles and may therefore be treated within any of the conventional perturbation methods of many-body analysis. Application of the technique to obtain a kinetic equation for a hard-sphere gas is outlined.

## 1. INTRODUCTION AND MOTIVATION

Considerable difficulties arise from attempts to apply any of the usual techniques of analysis in describing the effects of the strong short-ranged repulsions between the particles of a classical many-body system. If these repulsive forces are idealized (and mathematically simplified) by assuming that the individual particles are structureless elastic spheres of diameter $\sigma$, the interactions are too singular to permit any meaningful perturbation approximation. Moreover, for a van der Waals-type system the combination of weak long-ranged attractions and the hard-core repulsions appears to be inherently untreatable within any formal expansion type of development. Consequently, one usually finds that a comparatively naive approach based upon some self-consistent insight into a particular formalism results in the best possible analysis.

The present article summarizes the efforts of the author to provide a comprehensive method which may be employed to describe the repulsive interactions within *any* of the conventional techniques of many-body physics. The central idea is to reduce the effects of the hard cores into equivalent but more regular potential interactions while retaining the essential Hamiltonian description of the particle dynamics. Following a suggestion of Gross, this is accomplished by performing an extended canonical point transformation upon the dynamical coordinates so as to express the repulsions as equivalent velocity-dependent forces between the individual particles. Such interactions are well known in classical mechanics and can be studied within any statistical-mechanical formalism without difficulty. The analysis in fact becomes a novel exercise in determining the properties of a many-body system whose Hamiltonian contains a simple momentum-dependent potential.

The approach is developed by recognizing that the hard-core repulsions imply the existence of certain excluded-volume regions in the coordinate phase space by prohibiting any two particles from approaching closer than a distance $\sigma$ apart. Such restrictions are essentially kinematic (holomorphic) constraints upon the relative physical motions of the individual particles,[1] and can be eliminated by making the coordinate representation dependent upon the relative positions of the particles. In order that this modification of the coordinates leaves the description of the particle dynamics unaltered, the conjugate-momentum variables must be correspondingly modified. This is accomplished by means of the extended point transformation of the coordinates and momenta. It is the spatial dependence of the canonical momenta which reformulates the kinematics constraints into the simple nonlocal velocity-dependent interactions between the particles.[2]

This technique of reducing singular coordinate repulsions between the particles of a many-body system into more regular momentum-dependent potential interactions was first proposed by Bohm and Gross[3] and developed by Eger and Gross[4–6] to study the equilibrium properties of a system composed of hard-core bosons. They were able to effect a reformulation of the hard-sphere scattering for binary

* The research reported in this paper is based upon a Ph.D. thesis presented at Brandeis University, Waltham, Massachusetts (1966) and supported in part by the Office of Naval Research under Contract NONR 1677-04.

† The author is currently NBS-NRC Postdoctoral Associate at the National Bureau of Standards, Washington, D.C., 20234.

[1] This is analogous to the restriction which F. Dyson [Phys. Rev. **102**, 1217 (1956)] noted in attempting to distinguish between kinematical and dynamical interactions of a magnetic-spin system where a form of kinematic constraint "arises from the fact that more than $2S$ units of reversed spin cannot be attached to the same atom simultaneously . . . there is therefore a certain statistical hindrance to any dense packing of spin waves within a region."

[2] Bohm and Pines have pointed out that it is such a canonical reformulation of the $2N$ individual dynamical coordinates into the collective-coordinate representation which in effect replaces position-dependent potentials by equivalent momentum-dependent interactions. See D. Pines and D. Bohm, Phys. Rev. **85**, 338 (1952).

[3] D. Bohm and E. P. Gross (unpublished). See Refs. 4–6.

[4] M. Eger and E. P. Gross, Ann. Phys. (N.Y.) **24**, 63 (1963).

[5] M. Eger and E. P. Gross, Nuovo Cimento **34**, 1225 (1964).

[6] M. Eger and E. P. Gross, J. Math. Phys. **7**, 578 (1966).

collisions that in fact reproduces the usual phase-shift analysis, by choosing the transformation with regard to a metric-potential term which appears as the result of noncommutativity of the quantum-mechanical momentum and position operators.[7] A similar approach has been put forth by Luban,[8] who was able to explicitly construct an operator defined over all relative distances between two particles which plays the role of a Hamiltonian such that the subsequent eigenvalue problem has as its only solutions precisely the eigenfunctions for two hard spheres. Both of these developments clearly exhibit the nonlocal aspects of hard-core repulsions in the wavelike characterization of the quantum particles and unlike the pseudopotential method[9] provide for a completely equivalent but more regular (Fourier analyzable) Hermitian Hamiltonian over the entire phase space.

In this paper a similar analysis is completed for a classical hard-particle system. A general form of the exact transformation is first obtained for the two-body problem where the method is easily shown to reproduce the details of the classical hard-sphere scattering. Generalizing this procedure to the full many-body system, it is noted that the appropriate transformation involves determining the Jacobian of the coordinate transformation relating all $N-1$ relative positions of the particles. Since the reduction of any such a form is not practical (let alone possible), the transformation is taken to be linear in the conjugate momentum. The result is a Hamiltonian which retains many of the features of a point-particle form. Exhibited in this manner, it is immediately seen that the real advantage of this canonical reformulation permits the strong repulsions to be studied within any of the standard perturbation formalisms. By assuming that the collisions may be decomposed into a simple superposition of binary encounters, the virial expansion is easily reproduced. To illustrate application of the approach, the properties of a van der Waals system are discussed by developing a mean-field approximation valid for both the hard-core repulsions as well as the weaker long-ranged attractions. A description of the approach to equilibrium for such a hard-sphere gas is outlined using Zwanzig's projection-operator formalism.

[7] An excellent exposition on the canonical elimination of the hard core for quantum particles is contained in an article by J. S. Bell, "Many-Body Problem" in *Bergen Lectures* (W. A. Benjamin, Co., Inc., New York, 1961). In the quantum-mechanical problem, the nonlocal aspect of the hard-core repulsions is immediately obvious due to the wavelike characterization of the particles. For the classical scattering of two hard spheres the nonlocal nature of the interaction is by no means as apparent.

[8] M. Luban, Phys. Rev. **138**, A1028 (1965).

[9] See K. Huang, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1963).

## 2. TWO-PARTICLE TRANSFORMATION

The phase-plane motion of two intersecting hard-core particles serves to outline the qualitative behavior of the scattering process and to motivate the basic features of the transformation. For identical but distinguishable rigid spheres of mass $m$ and diameter $\sigma$, the scattering process is described in the relative coordinate system ($\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, $\mathbf{p} = \mathbf{p}_1 - \mathbf{p}_2$, $\mu$ reduced mass) by the Hamiltonian

$$H(\mathbf{r}, \mathbf{p}) = (1/2\mu)\mathbf{p}^2 + V(r). \tag{2.1}$$

The interaction potential $V(r)$ is assumed to depend only upon the magnitude of the relative coordinate $|\mathbf{r}|$. The geometric constraint that no two particles occupy the same region in space is expressed by restricting the domain (of definition) of $r$ such that $r \geq \sigma$ and excluding the region $0 \leq r \leq \sigma$. For unbounded physical motions, the incident particle is scattered symmetrically about the apse of the orbit passing a distance $r_0$ from the scattering center. When the distance of closest approach is greater than the size of the hard-core $r_0 \geq \sigma$ the particle motion is reversed continuously at the turning point of the orbit as the radial momentum vanishes identically $p_r(r_0) = 0$. If, however, the particle strikes the scattering center $r_0 < \sigma$ in an elastic collision, the motion undergoes an abrupt reversal of direction, experiencing a discontinuous change in the radial momentum by an amount $\Delta p_r = -2p_r$ $(r = \sigma_+)$.

The restriction of the physical motion to the region $r \geq \sigma$ may be eliminated by simply choosing a new relative coordinate variable $\mathbf{R}$ defined over all values $|\mathbf{R}| \geq 0$ and expressing $\mathbf{r}$ as a continuous function of $\mathbf{R}$. The simple form $r = S(R) = R + \sigma$ serves to illustrate the technique. If, however, the transformation is to leave the description of the motion unaltered, it is necessary to effect a compensating change in the conjugate-momentum variable $\mathbf{p}$ so that the new momentum $\mathbf{P}$ possesses a zero at the value of $R$ corresponding to the turning point of the motion, $P_R(r_0) = 0$.

This change of variables from $(\mathbf{r}, \mathbf{p})$ to $(\mathbf{R}, \mathbf{P})$ can be made rigorous by means of a simple canonical transformation, the invariance of the Hamiltonian formalism preserving the dynamical description of the motion.[10] From the generator of such transformations $\mathcal{G}(\mathbf{p}, \mathbf{R})$, specification of the relationship $\mathbf{r} = \mathbf{S}(\mathbf{R})$ dictates the functional form of

$$\partial\mathcal{G}(\mathbf{p}, \mathbf{R})/\partial\mathbf{p} = \mathbf{r} = \mathbf{S}(\mathbf{R}). \tag{2.2}$$

[10] See any standard text on classical mechanics such as H. Goldstein, *Classical Mechanics* (Addison-Wesley Publ. Co., Reading, Mass., 1950).

It follows by integrating with respect to $\mathbf{p}$ that

$$\mathcal{G}(\mathbf{p}, \mathbf{R}) = \mathbf{p} \cdot \mathbf{S}(\mathbf{R}) + \mathcal{G}_0(\mathbf{R}). \tag{2.3}$$

The conjugate momentum $\mathbf{P}$ is then determined as a function of the variables $(\mathbf{p}, \mathbf{R})$ as

$$\mathbf{P} = \frac{\partial \mathcal{G}(\mathbf{p}, \mathbf{R})}{\partial \mathbf{R}} = \mathbf{p} \cdot \frac{\partial \mathbf{S}(\mathbf{R})}{\partial \mathbf{R}}. \tag{2.4}$$

[A simple gauge transformation has been used to eliminate the additional term $\partial \mathcal{G}_0(\mathbf{R})/\partial \mathbf{R}$ which arises as a constant of integration.]

Since the scattering depends only upon the magnitude of the relative coordinate $r$, one may choose $\mathbf{S}(\mathbf{R})$ to be a spherically symmetric function. The angular coordinates $(\theta, \phi)$ then undergo the simple identity transformation to $(\Theta, \Phi)$ provided $\mathbf{S}(\mathbf{R})$ has the components $\{S(R), \Theta, \Phi\}$. From the generator of the transformation

$$\mathcal{G}(\mathbf{p}, \mathbf{R}) = p_r S(R) + p_\theta \Theta + p_\phi \Phi, \tag{2.5}$$

it follows that the conjugate momenta are given by

$$\mathbf{P} = \frac{\partial \mathcal{G}(\mathbf{p}, \mathbf{R})}{\partial \mathbf{R}}, \tag{2.6a}$$

$$\begin{pmatrix} P_R \\ P_\Theta \\ P_\phi \end{pmatrix} = \begin{pmatrix} p_r \dfrac{\partial S(R)}{\partial R} \\ p_\theta \\ p_\phi \end{pmatrix}, \tag{2.6b}$$

corresponding to the inverse coordinate transformation

$$\mathbf{r} = \partial \mathcal{G}(\mathbf{p}, \mathbf{R})/\partial \mathbf{p}, \tag{2.7a}$$

$$\begin{pmatrix} r \\ \theta \\ \phi \end{pmatrix} = \begin{pmatrix} S(R) \\ \Theta \\ \Phi \end{pmatrix}. \tag{2.7b}$$

If the coordinate $\mathbf{R}$ is to be defined over all $R \geq 0$, the transformation must be chosen such that the origin $R = 0$ corresponds to the point $r = \sigma$. This is achieved by setting $S(R = 0) = \sigma$. The conjugate radial momentum $P_R$ will then exhibit a true turning point for some value $R_0 > 0$ corresponding to a simple scattering about $r_0 > \sigma$ or must vanish identically at $R = 0$. From the identity $P_R = p_r[\partial S(R)/\partial R]$ it follows that $S'(0) = 0$.

For asymptotically large values of the radial coordinate $r \to \infty$, the scattering remains unaffected by the presence of the hard core, suggesting that $S(R)$ reduces to the simple identity transform for large $R$

$$S(R \gg 0) = R, \tag{2.8}$$

$$\partial S(R \gg 0)/\partial R = 1. \tag{2.9}$$

Combining the limiting values for large and small values of $R$, the function $S(R)$ is seen to be a simple monotonically increasing function of $R$ satisfying the conditions

$$\text{(c-1)} \quad S(R = 0) = \sigma, \qquad \text{(c-3)} \quad S(R \gg 0) = R,$$
$$\text{(c-2)} \quad S'(R = 0) = 0, \qquad \text{(c-4)} \quad S'(R \gg 0) = 1.$$

Within the requirement that the particular form of the transformation be chosen to preserve the description of the individual scattering events, any further specification of the form of $S(R)$ is unnecessary (see Appendix).

These properties of the function $S(R)$ can be shown to be consistent with those obtained directly from the equations of motion. Since the scattering Hamiltonian in radial coordinates

$$H(\mathbf{r}, \mathbf{p}) = \frac{1}{2\mu}\left[p_r^2 + \frac{1}{r^2} p_\theta^2 + \frac{1}{r^2 \sin^2 \theta} p_\phi^2\right] + V(r > \sigma) \tag{2.10}$$

remains invariant under a canonical change of variables, a direct substitution for $(\mathbf{r}, \mathbf{p})$ in terms of $(\mathbf{R}, \mathbf{P})$ transforms $H(\mathbf{r}, \mathbf{p})$ into

$$H(\mathbf{R}, \mathbf{P}) = \frac{1}{2\mu}\left[\left(\frac{P_R}{S'(R)}\right)^2 + \left(\frac{P_\Theta}{S(R)}\right)^2 + \left(\frac{P_\Phi}{S \sin \Theta}\right)^2\right] + V(S(R)). \tag{2.11}$$

Instead of attempting a complete description of the scattering in terms of quadratures, it is sufficient to consider the dynamics of the equivalent one-dimensional reduction by expressing the radial momentum $P_R$ as a function of the total energy ($H$ = constant $E$)

$$P_R = \frac{\partial S(R)}{\partial R}\left(2\mu\{E - V(S(R))\} - \left[\left(\frac{P_\Theta}{S}\right)^2 + \left(\frac{P_\Phi}{S \sin \Theta}\right)^2\right]\right)^{\frac{1}{2}}. \tag{2.12}$$

The term $[(P_\Theta/S(R))^2 + (P_\Phi/S \sin \Theta)^2]$ is the modified centrifugal potential barrier which scatters the incident particle about a (continouus) turning point at the zero of the radical. If the radial kinetic energy is sufficient to overcome this finite centrifugal barrier, the incident particle is turned around at the origin $R = 0$ by requiring that $\partial S(R = 0)/\partial R$ vanish identically.[11]

From the properties of $S(R)$ it is seen that the transformed static potential $V(S(R))$ remains unaltered at the end points $r = \sigma$ and $r \to \infty$, while for the

[11] Unlike the usual centrifugal barrier which diverges as $r^{-2}$ at the origin ($r = 0$), the modified form $[S(R)]^{-2}$ is finite at $R = 0$. The kinetic energy, however, remains bounded by requiring that $[P_R(R)/S(R)]$ be finite everywhere.

intermediate values of the radial coordinate $r = S(R) < R$, the potential is altered by the smearing out of the hard-core repulsion over the entire space. In the limit of a vanishingly small core size $\sigma \to 0$, these results conveniently reproduce the usual expressions for radial scattering as $S(R)$ reduces to the identity transformation $S(R) = R$.

## 3. MANY-PARTICLE TRANSFORMED HAMILTONIAN

The analysis of the previous section clearly demonstrates that it is possible to eliminate the geometric restriction on the relative coordinate $r \geq \sigma$ for the two-particle scattering Hamiltonian through a canonical reformulation of the dynamical variables. Through a generalization of this technique one can effect a similar transformation for a system composed of $N$ such particles which will eliminate the $(N-1)$ constraints from the many-body Hamiltonian

$$H = \frac{1}{2m}\sum_{i=1}^{N}\mathbf{p}_i^2 + \frac{\lambda}{2}\sum_{i\neq j}^{N} V(r_{ij} > \sigma). \tag{3.1}$$

However, as this involves the simultaneous knowledge of the relative positions between all $N$ particles, it is not possible to explicitly construct the generator $\mathcal{G}_N$ of any such transformation. Instead (having exhibited that such a reformulation exists for any two particles), it is asserted that for the set of dynamical variables

$$\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N\}, \quad \mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_N\},$$

all $r_{ij} \geq \sigma$, there exists a transformation $\mathcal{G}_N(\mathbf{p}, \mathbf{R})$ to the set

$$\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_N\}, \quad \mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_N\}$$
$$\exists\, \mathcal{G}_N(\mathbf{p}, \mathbf{R}) \ni \mathcal{G}_N{:}[\mathbf{r}, \mathbf{p}] \Rightarrow [\mathbf{R}(\mathbf{r}), \mathbf{P}(\mathbf{r}, \mathbf{p})],$$

where $\mathcal{G}_N$ is to be chosen from the full canonical group so as to preserve the description of the individual scattering events. The most general such transformation possible will then relate the variables as

$$\mathbf{R}_i = \mathbf{R}_i(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N),$$
$$\mathbf{P}_i = \mathbf{P}_i(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N, \mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_N),$$

for which it is immediately apparent that no manageable algebraic manipulation can be performed. It is therefore necessary to impose certain realistic restrictions upon $\mathcal{G}_N$ in order to obtain a tractable form. For this purpose, it is assumed that the transformation:

(i) reduces to the simple identity form in the limit of point-particles, the hard core vanishing identically ($\sigma = 0$),

(ii) depends only upon the relative coordinates between the particles (translational invariance), and

(iii) relates the momenta $\mathbf{P}$ linearly with $\mathbf{p}$.

This can be achieved by a generator of the form

$$\mathcal{G}_N(\mathbf{p}, \mathbf{R}) = \sum_{i=1}^{N} \mathbf{p}_i \cdot \mathbf{S}_i(\mathbf{R}, \sigma), \tag{3.2}$$

with

$$\mathbf{S}_i(\mathbf{R}, \sigma = 0) = \mathbf{R}_i, \tag{3.3}$$

since

$$\mathbf{P}_i = \frac{\partial \mathcal{G}_N}{\partial \mathbf{R}_i} = \sum_{j=1}^{N} \mathbf{p}_j \frac{\partial \mathbf{S}_j}{\partial \mathbf{R}_i}, \tag{3.4}$$

$$\mathbf{r}_i = \frac{\partial \mathcal{G}_N}{\partial \mathbf{p}_i} = \mathbf{S}_i(\mathbf{R}, \sigma). \tag{3.5}$$

Under such a canonical transformation, the original Hamiltonian $H(\mathbf{r}, \mathbf{p})$ becomes

$$\mathcal{G}_N{:}H(\mathbf{r}, \mathbf{p}) \Rightarrow H(\mathbf{R}, \mathbf{P}) \equiv \mathcal{H}(\mathbf{p}, \mathbf{R}), \tag{3.6}$$

where

$$\mathcal{H}(\mathbf{P}, \mathbf{R}) = \frac{1}{2m}\sum_{i,j}^{N} \mathbf{P}_i \mathbf{A}_{ij}(R) \mathbf{P}_j + \frac{\lambda}{2}\sum_{i\neq j} \tilde{V}(ij).$$

The modified potential interactions have been denoted by

$$\tilde{V}(ij) = V(|\mathbf{S}_i(\mathbf{R}) - \mathbf{S}_j(\mathbf{R})|). \tag{3.7}$$

The dyadic

$$\mathbf{A}_{ij}(\mathbf{R}) = \sum_{\alpha=1}^{3N} \left(\frac{\partial S_i}{\partial R_\alpha}\right)^{-1} \left(\frac{\partial S_j}{\partial R_\alpha}\right)^{-1} \tag{3.8}$$

is a spatially-dependent quantity which by coupling the various components of the different particle momenta gives rise to the velocity-dependent interactions which prohibit particle motions into regions of $r_{ij} < \sigma$.

If the system is translationally invariant, the repulsive interactions can be separated from the usual kinetic-energy terms $[(1/2m)\sum \mathbf{P}_i^2)]$ by setting

$$\mathbf{A}_{ij}(R) = \delta_{ij} + \sigma \mathbf{G}_{ij}(\mathbf{R}), \tag{3.9}$$

where $\delta_{ij}$ defines the Kronecker delta (unit dyadic). The metric $\mathbf{G}_{ij}(\mathbf{R})$ forms a $3N \times 3N$ matrix whose elements depend upon the relative positions of all $N$ particles in the system. If only pairwise interactions occur (binary collisions), $\mathbf{G}_{ij}(\mathbf{R})$ will depend upon relative coordinate $R_{ij}$ only.

Expressed in this manner, the transformed Hamiltonian

$$\mathcal{H}(\mathbf{P}, \mathbf{R}) = \frac{1}{2m}\sum_{i=1}^{N} \mathbf{P}_i^2 + \frac{\sigma}{2m}\sum_{i\neq j}^{N} \mathbf{P}_i \mathbf{G}_{ij}(\mathbf{R}) \mathbf{P}_j + \frac{\lambda}{2}\sum_{i\neq j} \tilde{V}(ij) \tag{3.10}$$

closely resembles the usual point-particle form—the quadratic dependence upon the momenta being a characteristic of the linear point transformation.

Although it contains momentum-dependent interactions ($\sum \mathbf{P}_i \mathbf{G}_{ij} \mathbf{P}_j$) in addition to the modified potential term, $H$ is no longer restricted by the set of $N-1$ constraints (all $r_{ij} > \sigma$).

Associated with the general transformation which carries the $2N$ dynamical variables $(\mathbf{r}, \mathbf{p})$ into the set $(\mathbf{R}, \mathbf{P})$ is the Jacobian

$$J \equiv \frac{\partial(\mathbf{R}, \mathbf{P})}{\partial(\mathbf{r}, \mathbf{p})} = \det \begin{vmatrix} \partial\mathbf{R}/\partial\mathbf{r} & \partial\mathbf{R}/\partial\mathbf{p} \\ \partial\mathbf{P}/\partial\mathbf{r} & \partial\mathbf{P}/\partial\mathbf{p} \end{vmatrix}. \quad (3.11)$$

The nonvanishing of this functional determinant is both the necessary and sufficient condition that the mapping be unique and reversible. By taking the transformation from the full canonical group so as to preserve the properties of the Hamiltonian formalism insures that this Jacobian is identically unity. Furthermore, since any canonical transformation leaves the extent in the phase space an (integral) invariant quantity

$$\int d\mathbf{r} \int d\mathbf{p} \Rightarrow \int d\mathbf{R} \int d\mathbf{P}[\partial(\mathbf{r}, \mathbf{p})/\partial(\mathbf{R}, \mathbf{P})] = \int d\mathbf{R} \int d\mathbf{P},$$

the average of a physical observable determined over an appropriate distribution function $\rho(H)$ remains unaltered

$$\int d\mathbf{r} \int d\mathbf{p} \rho(H(\mathbf{r}, \mathbf{p})) \Rightarrow \int d\mathbf{R} \int d\mathbf{P} \rho(H(\mathbf{R}, \mathbf{P})). \quad (3.12)$$

By the particular choice

$$\mathfrak{S}_N : (\mathbf{r}, \mathbf{p}) \Rightarrow \{\mathbf{R}(\mathbf{r}), \mathbf{P}(\mathbf{r}, \mathbf{p})\},$$

it follows from the group property of the transformation

$$(\mathbf{r}, \mathbf{p}) \Rightarrow (\mathbf{R}, \mathbf{p}) \Rightarrow \{\mathbf{R}(\mathbf{r}), \mathbf{P}(\mathbf{r}, \mathbf{p})\}$$

that the Jacobian then obeys a product law

$$J = \det \partial\mathbf{R}/\partial\mathbf{r} \cdot \det \partial\mathbf{P}/\partial\mathbf{p}.$$

Because of this, the integral of the canonical distribution function $\rho = e^{-\beta H}$ over the phase space can be written as

$$\int d\mathbf{R} \int d\mathbf{p} J_R \cdot \rho(\mathcal{H}(\mathbf{R}, \mathbf{p})).$$

This allows the Jacobian of the coordinate transformation

$$J_R = \partial(\mathbf{r}, \mathbf{p})/\partial(\mathbf{R}, \mathbf{p})$$

to be interpreted (in the manner of Kirkwood) as an equivalent "temperature-dependent configurational potential energy"

$$\vartheta(R) = -\beta^{-1} \log J_R(R). \quad (3.13)$$

(A further discussion of this property is provided in Sec. 4.)

## 4. STATISTICAL MECHANICS OF VELOCITY-DEPENDENT INTERACTIONS

To provide further motivation for this replacement of the hard-core repulsions by equivalent momentum-dependent interactions, consider the properties of an $N$-particle system with the Hamiltonian

$$\mathcal{H} = \sum_{\alpha,\nu}^{3N} \mathbf{p}_\alpha A_{\alpha\nu}(\mathbf{R}) \mathbf{p}_\nu. \quad (4.1)$$

[May,[12] noting that velocity-dependent interactions are often used in nuclear physics to simulate the effects of nuclear forces, has pointed out that "it is quite reasonable to expect a certain similarity between the velocity-dependent potential and the hard-core repulsions" between the particles of a many-body system. Postulating a Hamiltonian of the form Eq. (4.1), he considered the partition function and developed the cluster expansion for such a system.]

Because this Hamiltonian is quadratic in the momentum, it follows immediately from the classical equipartition theorem

$$\langle \mathcal{H} \rangle_\beta = \left\langle \sum_\gamma^{3N} p_\gamma \frac{\partial}{\partial p_\gamma} \mathcal{H} \right\rangle_\beta = \frac{3N}{2\beta}, \quad (4.2)$$

that the energy is simply

$$E = \tfrac{3}{2}kT \quad (\beta = 1/kT).$$

The specific heat $C_v = \frac{3}{2}k$ is therefore independent of the temperature and thus identical to that of a collection of $N$ simple noninteracting hard particles.

The detailed statistical properties of the system may be obtained from the $N$-particle distribution function $\rho(\mathcal{H})$. Preforming the $3N$-fold integration over the momentum phase space[13] determines the spatial-distribution function

$$\rho(\mathbf{R}) = \int d\mathbf{p} e^{-\beta \mathcal{H}(\mathbf{p}, \mathbf{R})} \quad (4.3)$$

$$= (\pi/\beta)^{\frac{3}{2}N} [\det A(\mathbf{R})]^{-\frac{1}{2}}. \quad (4.4)$$

Exponentiating the $\det^{\frac{1}{2}}$ form, it is seen that the hard-core repulsions give rise to a potential of average force for which the interaction energy is

$$\vartheta(\mathbf{R}) = -(1/2\beta) \log \det A(\mathbf{R}). \quad (4.5)$$

[12] R. May, Nucl. Phys. **62**, 177 (1965).

[13] $H$ can be expressed as the sum of squares by a simple rotation of the momentum space, thereby reducing the integral to the product-independent integrals

$$I = \int_{-\infty}^{+\infty} dp_1 \cdots \int dp_N \exp\left(-\sum_{ij}^{N} p_i A_{ij} p_j\right)$$

$$= \prod_{n=1}^{N} \int_{-\infty}^{+\infty} dx_n \exp(-\tfrac{1}{2} a_n^2 x_n^2) = \prod_{n=1}^{N} (2\pi/a_n)^{\frac{1}{2}},$$

with $a_1 a_2 \cdots a_N = \det A$, and hence the identity.

[Compare this result with Eq. (3.14) of Sec. 3.] By developing a formal expansion for the log det $A$ it is possible to represent this potential energy as the sum of individual interactions between clusters of particles.[14]

The partition function $Z$ is found by carrying out the remaining integrations over the coordinate space

$$Z = \int d\mathbf{R} \int d\mathbf{p} e^{-\beta \mathcal{H}(\mathbf{R},\mathbf{p})} = (\pi/\beta)^{\frac{3}{2}N} \int \frac{d\mathbf{R}}{[\det A(\mathbf{R})]^{\frac{1}{2}}}. \tag{4.6}$$

The resulting configurational integral is seen to be independent of the temperature—a property unique to the excluded volume regions of a collection of hard particles. Since the actual evaluation of this configurational integral is not possible, it is necessary to resort to the usual procedure of decomposing the integrand into irreducible cluster diagrams. In fact, noting that the integrand can be expressed as the Jacobian of the coordinate transformation

$$\int d\mathbf{R} J_R = \int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N \prod_{i \neq j}^{N} \frac{\partial(r_{ij})}{\partial(R_{ij})}, \tag{4.7}$$

the standard results for the virial coefficients of a hard-sphere gas are easily reproduced.[12,14]

## 5. APPLICATION TO A VAN DER WAALS GAS

The previous reduction of the hard-core repulsions into simple momentum-dependent interactions served only to postpone the actual analysis of the many-body problem. However, in doing so, this approach provides an equivalent Hamiltonian in which the singular repulsions may be treated as simple interactions between the particles rather than kinematic constraints upon their relative motions. This is especially useful in studying the properties of a classical van der Waals system since it permits the strong repulsion to be described in the same manner as the weaker long-ranged attractions.

For example, a standard procedure in many-body analysis is to replace the direct interactions between the particles by an effective potential obtained in some mean-field approximation. To lowest order, this procedure requires only finding the average value of the direct interparticle potential $V(r_{ij})$ normalized over the entire volume $\Omega$

$$V_0 = \Omega^{-1} \int d\mathbf{r} V(\mathbf{r}).$$

(In the collective coordinate representation this is equivalent to determining the small wavenumber $\mathbf{k}$ behavior of the Fourier transform of the potential.) But the average value of such a pathological interaction as the hard core $V(r < \sigma) \to \infty$, $V(r > \sigma) = 0$ does not exist and therefore certain obvious difficulties arise.[15] However, as the individual elements of the matrix $\mathbf{G}_{ij}$ in effect represent the repulsive interactions between the particle pair $i$–$j$, a mean-field type of approximation can be easily carried out for the hard-core interactions. Replacing each $\mathbf{G}_{ij}$ by its average value,

$$G_{ij} \approx G_0 = \Omega^{-1} \int d\mathbf{R} \mathbf{G}_{ij}(\mathbf{R}) \approx \Delta/\Omega, \quad \Delta = \tfrac{4}{3}\pi\sigma^3, \tag{5.1}$$

simplifies the corresponding $\mathbf{A}$ matrix to the form

$$\mathbf{A} = \begin{vmatrix} 1 & G_0 & G_0 & \cdots & G_0 \\ G_0 & 1 & G_0 & \cdots & G_0 \\ G_0 & G_0 & 1 & \cdots & \\ \cdot & & & \cdot & \\ \cdot & & & \cdot & \\ \cdot & & & & 1 \end{vmatrix} = |\delta_{ij} + G_0(1 - \delta_{ij})|, \tag{5.2}$$

with eigenvalues $g_1 \cdots g_N$,

$$g_1 = 1 + (N-1)G_0,$$
$$g_2 = g_3 = \cdots = g_N = 1 - G_0.$$

Relating the det of a matrix to its eigenvalues (det $G = \prod_i g_i$) it follows that the configurational potential energy Eq. (4.5) in such an approximation is

$$\vartheta = (2\beta)^{-1} \log \det (1 - \sigma G)$$
$$\approx (2\beta)^{-1} \left[ \log \left( 1 - \frac{N\Delta}{\Omega} \right) + (N-1) \log (1 - \Delta/\Omega) \right]. \tag{5.3}$$

This is the usual description for the excluded-volume effect in a van der Waals-type gas in which each particle decreases the total accessible volume by exactly its own size.

Using this equivalent representation for the hard-core repulsions it is also possible to attempt a detailed calculation of the dynamical behavior for a hard-core gas. With the effects of the collisions reduced to the momentum-dependent interactions between the particles, one may simply proceed within the usual perturbation formalisms instead of constructing an *ad hoc* collision term based upon arguments of

[14] M. J. Cooper, Brandeis University thesis, 1966 (unpublished).

[15] In the collective coordinate representation, J. Percus and G. Yevick [Phys. Rev. **138**, A1028 (1965)] have shown that the hard-core repulsions can be replaced by an "equivalent potential whose precise value depends upon the approximation used for the long-ranged forces."

continuing molecular chaos of forgetful collisions. The problem of doing so, however, is somewhat more complicated because the interactions are both spatially and velocity dependent.

The temporal evolution of the phase-space distribution function $\rho(\mathbf{r}, \mathbf{p}, t)$ is conveniently summarized in the single Liouville equation

$$i(\partial\rho/\partial t) = \mathfrak{L}\rho, \tag{5.4}$$

where the Liouville operator $\mathfrak{L}$ is obtained from the Poisson bracket with the Hamiltonian

$$\mathfrak{L} = i\{\mathcal{H}, \} = i\left(\frac{\partial\mathcal{H}}{\partial\mathbf{p}}\cdot\frac{\partial}{\partial\mathbf{R}} - \frac{\partial\mathcal{H}}{\partial\mathbf{R}}\cdot\frac{\partial}{\partial\mathbf{p}}\right). \tag{5.5}$$

If only binary collisions occur between the particles, the total Hamiltonian is of the form of Eq. (4.1), where the elements $\mathbf{G}_{ij}$ are functions of the individual $R_{ij}$ only. The corresponding Liouville operator thus consists of the sum of an unperturbed term

$$\mathfrak{L}_0 = -i\sum_{i=1}^{N}\mathbf{p}_i\cdot\frac{\partial}{\partial\mathbf{R}_i} \tag{5.6}$$

plus terms arising from the direct potential interactions

$$\lambda\mathfrak{L}_v = -i\frac{\lambda}{2}\sum_{i\neq j}^{N}\frac{\partial V(R_{ij})}{\partial\mathbf{R}_i}\left(\frac{\partial}{\partial\mathbf{p}_i} - \frac{\partial}{\partial\mathbf{p}_j}\right), \tag{5.7}$$

and the effects of the momentum-dependent interactions

$$\sigma\mathfrak{L}_\sigma = i\sigma\sum_{i\neq j}^{N}\mathbf{p}_i\cdot\frac{\partial\mathbf{G}(R_{ij})}{\partial\mathbf{R}_i}\cdot\mathbf{p}_j\left(\frac{\partial}{\partial\mathbf{p}_i} - \frac{\partial}{\partial\mathbf{p}_j}\right) - 2\mathbf{p}_i\mathbf{G}(R_{ij})\frac{\partial}{\partial\mathbf{R}_j}. \tag{5.8}$$

The coupling of the strong repulsions and the potential interactions are observed in the noncommutativity of the various parts of the perturbed Liouville operator, $\delta\mathfrak{L} = \lambda\mathfrak{L}_v + \sigma\mathfrak{L}_\sigma$.

For a spatially homogeneous system, the momentum-space distribution function $\rho(\mathbf{p}, t)$ has been shown by Zwanzig[16] to obey a master equation of the form

$$\frac{\partial}{\partial t}\rho(\mathbf{p}, t) = -\int_0^t dT\mathcal{K}(T, \lambda, \sigma)\rho(\mathbf{p}, t - T), \tag{5.9}$$

where the kernel $\mathcal{K}$ is the projection of the perturbed Liouville operator over the entire coordinate space

$$\mathcal{K} = \frac{1}{\Omega^N}\int d\mathbf{R}_1\cdots\int d\mathbf{R}_N\delta\mathfrak{L}(\mathbf{R}, \mathbf{p}) \equiv \mathfrak{I}\cdot\delta\mathfrak{L}. \tag{5.10}$$

In the absence of any weak interactions $\lambda = 0$, when the only interactions between the particles occur through direct collisions, this equation becomes

$$\frac{\partial\rho(\mathbf{p}, t)}{\partial t} = -\sigma^2\int_0^t d\tau\mathfrak{I}\,\mathfrak{L}_\sigma e^{-iT(1-\mathfrak{I})\mathfrak{L}}\mathfrak{L}_\sigma\rho(\mathbf{p}, t - T). \tag{5.11}$$

For very short collision intervals, in the long time $t \to \infty$ this may be reduced to a master equation of Markoffiian form ($\rho \sim e^{-\Gamma t}$),

$$\partial\rho(\mathbf{p}, T)/\partial T = -\mathfrak{C}\rho(\mathbf{p}, T), \tag{5.12}$$

$$\mathfrak{C} = \int_0^\infty d\tau\mathfrak{I}\,\mathfrak{L}_\sigma e^{-i\tau\mathfrak{L}_0}\mathfrak{L}_\sigma, \tag{5.13}$$

where the collision operator consists of two terms.[17]

For particles interacting through both weak attractions and hard-core repulsions, the noncommutativity of the various parts of the perturbed Liouville operator give rise to an integro-differential equation preventing any easy solution for $\rho(\mathbf{p}, t)$. However, a kinetic equation for such a van der Waals system can be obtained in which the effects of the collisions are described by the velocity-dependent interactions.[14,18]

## APPENDIX A: PHYSICAL FORM OF TWO-BODY TRANSFORMATION

It has been shown that the conditions on $\mathbf{S}(\mathbf{R})$ are sufficient to insure that the form of the coordinate transformation preserve the mechanical description of the two-particle scattering. Since the canonical nature of the transformation leaves the statistical properties of the system unaltered, it is not possible to attempt any variational procedure to further determine $S(R)$. Instead, physical considerations must be used to motivate a particular choice within the full canonical group. For example, Eger and Gross,[4–6] in using this method to eliminate the difficulties associated with the repulsive interactions in a system of bosons, chose the transformation so as to enable one to play off the various types of interactions in the system, thereby simplifying its description.

Lacking any such criteria for a dilute system of hard-sphere particles, for the purpose of illustration, we choose the two-particle transformation

$$r = S(R) = R + \sigma e^{-R/\sigma}.$$

[16] R. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).

[17] Because the interactions depend upon both the action and angle variables, the collision operator thus obtained is in general a complex quantity. Consequently, the approach to equilibrium will not necessarily be strictly monotonic but will tend to oscillate. A similar result was found by I. Prigogine *et al.* [*Non-Equilibrium Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1962)] for a lattice system of very strongly interacting particles.

[18] L. de Sobrino, Can. J. Phys. **45**, 363 (1967) has recently proposed a kinetic equation for a van der Waals gas in which "just before a collision the correlation between two particles is due exclusively to the decrease of the volume of the phase space due to the hard cores."

This form not only satisfies the boundary conditions on $S(R)$ but in the limit of a vanishing hard core $\sigma \to 0$ reduces to the simple identity $r = S(R, \sigma = 0) = R$. It represents a physically satisfactory transformation in that it falls off rapidly, reducing to $r \approx R$ after several hard-core radii, without the introduction of any arbitrary cutoff. The Jacobian associated with the coordinate transformation forms an equivalent potential

$$\vartheta_R = -\beta^{-1} \log J_R = -\beta^{-1} \log (1 - e^{-R/\sigma}),$$

which is bounded below and vanishes for large values of $R$ such that its integral $\Omega^{-1} \int d\mathbf{R} \vartheta_R$ exists and is finite. It therefore satisfies the necessary and sufficient conditions on the pair potential that a stable configuration exists for the many-body system.[19]

ACKNOWLEDGMENTS

The author gratefully acknowledges the numerous suggestions of Professor D. Falkoff and Professor E. Gross and wishes to express his gratitude to them for their continuing encouragement and guidance provided during the course of this work,

[19] M. Fisher, Arch. Rat. Mech. Anal. 17, 410 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Modified Born Approximation and Elastic Scattering by Weak Central Potentials*

S. K. Dutta, W. A. Wilson, and A. E. S. Green
*Department of Physics and Astronomy, University of Florida, Gainesville, Florida*

(Received 28 July 1967)

When only a few partial waves are substantially phase shifted and yet many partial waves are slightly phase shifted, it is possible to use the direct Born approximation, provided that one projects out the inaccurate lower partial waves and replaces them by accurate theoretical or phenomenological phase shifts. We test this technique for central potentials with two different well strengths, i.e., one which will fail to bind the $1S$ wave, and one which can bind the $1S$ state. We compare numerically generated angular distributions and total cross sections with those obtained from a modified form of the direct Born approximation. The technique would be useful for weak forces, e.g., the nucleon–nucleon and electron–atom interactions, but inapplicable for strong forces, e.g., atom–atom interaction.

## 1. INTRODUCTION

In the analysis of elastic scattering of a particle by a central potential, one sometimes finds that the first few partial waves experience substantial phase shifts, whereas the higher partial waves are only slightly affected. In the case of a weak central potential which gives rise to no bound states, or at most one bound state, the $S$ wave alone undergoes a major phase shift. In such circumstances the scattering amplitude as calculated by the Born approximation is quite reliable for all the partial waves, except the first few—say the $S$ and $P$ waves. When this occurs, it seems possible to adapt the direct Born approximation which is so convenient. The modification consists in projecting out the lower partial-wave components (say the $l = 0$ and 1) from the closed-form expression for the Born elastic-scattering amplitude and replacing them by accurate components. To determine the accurate scattering amplitude we can use the values of phase shifts of the corresponding partial waves, determined either experimentally or by some theoretical calculation. Thus we obtain the modified Born elastic-scattering amplitude in terms of usual Born amplitude, a few subtractive terms representing the lower partial-wave components of the Born amplitude, and a few additive terms representing the exact amplitude for the lower partial waves. The correction terms are important only in the low-energy region; in the high-energy region the total direct Born amplitude should become dominant.[1]

If for a scattering problem there is not a sufficient amount of differential cross-section data available, or if the experimental errors involved in obtaining the differential cross sections are moderately large, a phase-shift analysis may greatly reduce the reliability of the data. In such cases the comparison of theoretical and experimental results may be more meaningful in terms of differential cross sections where the modified Born approach is very convenient. Powell and

* Supported in part by U.S.A.F. Office of Scientific Research.

[1] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London, 1965), 3rd ed., Chap. V.

This form not only satisfies the boundary conditions on $S(R)$ but in the limit of a vanishing hard core $\sigma \to 0$ reduces to the simple identity $r = S(R, \sigma = 0) = R$. It represents a physically satisfactory transformation in that it falls off rapidly, reducing to $r \approx R$ after several hard-core radii, without the introduction of any arbitrary cutoff. The Jacobian associated with the coordinate transformation forms an equivalent potential

$$\vartheta_R = -\beta^{-1} \log J_R = -\beta^{-1} \log (1 - e^{-R/\sigma}),$$

which is bounded below and vanishes for large values of $R$ such that its integral $\Omega^{-1} \int d\mathbf{R} \vartheta_R$ exists and is finite. It therefore satisfies the necessary and sufficient conditions on the pair potential that a stable configuration exists for the many-body system.[19]

ACKNOWLEDGMENTS

The author gratefully acknowledges the numerous suggestions of Professor D. Falkoff and Professor E. Gross and wishes to express his gratitude to them for their continuing encouragement and guidance provided during the course of this work,

[19] M. Fisher, Arch. Rat. Mech. Anal. 17, 410 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Modified Born Approximation and Elastic Scattering by Weak Central Potentials*

S. K. Dutta, W. A. Wilson, and A. E. S. Green
*Department of Physics and Astronomy, University of Florida, Gainesville, Florida*

(Received 28 July 1967)

When only a few partial waves are substantially phase shifted and yet many partial waves are slightly phase shifted, it is possible to use the direct Born approximation, provided that one projects out the inaccurate lower partial waves and replaces them by accurate theoretical or phenomenological phase shifts. We test this technique for central potentials with two different well strengths, i.e., one which will fail to bind the $1S$ wave, and one which can bind the $1S$ state. We compare numerically generated angular distributions and total cross sections with those obtained from a modified form of the direct Born approximation. The technique would be useful for weak forces, e.g., the nucleon–nucleon and electron–atom interactions, but inapplicable for strong forces, e.g., atom–atom interaction.

## 1. INTRODUCTION

In the analysis of elastic scattering of a particle by a central potential, one sometimes finds that the first few partial waves experience substantial phase shifts, whereas the higher partial waves are only slightly affected. In the case of a weak central potential which gives rise to no bound states, or at most one bound state, the $S$ wave alone undergoes a major phase shift. In such circumstances the scattering amplitude as calculated by the Born approximation is quite reliable for all the partial waves, except the first few—say the $S$ and $P$ waves. When this occurs, it seems possible to adapt the direct Born approximation which is so convenient. The modification consists in projecting out the lower partial-wave components (say the $l = 0$ and 1) from the closed-form expression for the Born elastic-scattering amplitude and replacing them by accurate components. To determine the accurate scattering amplitude we can use the values of phase shifts of the corresponding partial waves, determined either experimentally or by some theoretical calculation. Thus we obtain the modified Born elastic-scattering amplitude in terms of usual Born amplitude, a few subtractive terms representing the lower partial-wave components of the Born amplitude, and a few additive terms representing the exact amplitude for the lower partial waves. The correction terms are important only in the low-energy region; in the high-energy region the total direct Born amplitude should become dominant.[1]

If for a scattering problem there is not a sufficient amount of differential cross-section data available, or if the experimental errors involved in obtaining the differential cross sections are moderately large, a phase-shift analysis may greatly reduce the reliability of the data. In such cases the comparison of theoretical and experimental results may be more meaningful in terms of differential cross sections where the modified Born approach is very convenient. Powell and

* Supported in part by U.S.A.F. Office of Scientific Research.

[1] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London, 1965), 3rd ed., Chap. V.

Crasemann[2] have discussed a similar approach, one which treats the first one or two partial waves by other methods, when the Born approximation is inapplicable to them because of their large phase shifts.

## 2. MATHEMATICAL FORMULATION

The exact scattering amplitude, in terms of partial-wave phase shifts, is given by

$$f_e(E, \theta) = \sum_{l=0}^{\infty} (2l + 1) f_{e,l}(E) P_l(\cos \theta), \quad (2.1)$$

where

$$f_{e,l}(E) = (1/k) e^{i\delta_l(E)} \sin \delta_l(E).$$

The Born scattering amplitude may also be placed in the form

$$f_B(E, \theta) = \sum_{l=0}^{\infty} (2l + 1) f_{B,l}(E) P_l(\cos \theta),$$

where

$$f_{B,l}(E) = \frac{1}{2} \int_{-1}^{1} f_B(E, \theta) P_l(\cos \theta)\, d(\cos \theta). \quad (2.2)$$

We suggest a modification of the Born amplitude of the form

$$f_{MB}(E, \theta) = f_B(E, \theta) + f_c(E, \theta), \quad (2.3)$$

where $f_c(E, \theta)$ is the correction term given by

$$f_c(E, \theta) = \sum_{l=0}^{N} (2l + 1) f_{c,l}(E) P_l(\cos \theta),$$

and where

$$f_{c,l}(E) = f_{e,l}(E) - f_{B,l}(E). \quad (2.4)$$

Here $N$ denotes the maximum number of lower partial waves to be corrected.

## 3. THE YUKAWA POTENTIALS

We consider first a spherically symmetric central potential represented by a Yukawa-potential function. We scale our distance so that the potential is in the form

$$U(r) = (2\mu/\hbar^2) V(r) = -\alpha(e^{-r}/r), \quad (3.1)$$

where $\alpha$ is a dimensionless constant whose magnitude characterizes the parameters of the attractive potential. [Usually $U(r) = -\alpha \exp(-\mu r)/r$, but in this paper we have chosen $\mu = 1$ for convenience.] For a central field the scattering amplitude in the Born approximation is given by

$$f_B(E, \theta) = -\frac{1}{K} \int_0^{\infty} r \sin KrU(r)\, dr, \quad (3.2)$$

where

$$K = 2k \sin(\tfrac{1}{2}\theta).$$

[2] J. L. Powell and B. Crasemann, *Quantum Mechanics* (Addison-Wesley Publ. Company, Inc., Reading, Mass., 1961), p. 277.

Substituting (3.5) into (3.6), we get

$$f_B(E, \theta) = \alpha/(1 + K^2). \quad (3.3)$$

Its partial-wave projection is

$$f_{B,l}(E) = (\alpha/2k^2) Q_l[1 + (1/2k^2)], \quad (3.4)$$

where $Q_l(Z)$ is the Legendre function of the second kind. In the case under consideration here, the correction to the first two lower partial waves is sufficient. As a matter of fact, the major contribution to the correction term comes from the $S$ wave. For such as the $S$ and $S$–$P$ cases, the modified Born amplitude is written as

$$f_{MB}(E, \theta) = f_B(E, \theta) + \sum_{l=0}^{N} (2l + 1) \times \{f_{e,l}(E) - f_{B,l}(E)\} P_l(\cos \theta), \quad (3.5)$$

where, for $S$-wave correction, we should have $N = 0$, and for $S$ and $P$ correction, $N = 1$.

### Angular Distributions and Total Cross Sections for Weak Yukawa Potentials

In the numerical calculation reported in this section, we have tested the Yukawa potential given by (3.1) for two different values of $\alpha$, e.g., 1 and 3, corresponding to two different strengths of potential. For the case of $\alpha = 1$ there exists no bound state, whereas $\alpha = 3$ is strong enough to give an $S$-wave bound state. We have calculated to total elastic-scattering cross section and angular distribution for values of $k^2$ ranging from 0.2 to 4.0 by numerically solving the Schrödinger equation for the above potentials (which we designate as the exact value). These exact values have been compared with the corresponding quantities calculated by the Born approximation (3.3), $S$-wave modified Born approximation (3.5), and the $S$- and $P$-wave modified Born approximation (3.5) in order to examine the efficacy of our approach—the partial-wave modification of the Born amplitude. In calculating the exact partial-wave scattering amplitude, we have used the corresponding phase shifts obtained by solving the Schrödinger equation using the potential (3.1).

We first discuss the results for the case $\alpha = 1$. The total elastic-scattering cross section predicted by the Born approximation is lower than the exact value for low values of $k^2$ (those up to $k^2 = 2.0$), beyond which its predictions are almost identical with the exact ones (Fig. 1): If the $S$-wave part is modified in the way discussed above, the modified Born approximation gives a total cross section agreeing well with the exact value. Thus it appears that only one lower partial-wave correction to the

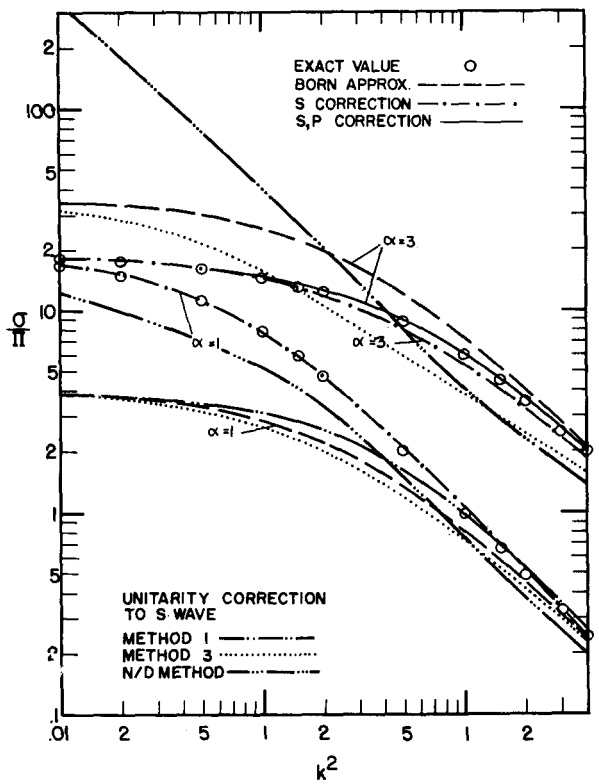



FIG. 1. The plot of total cross section against $k^2$ on log–log scale, showing the exact value, the Born approximation result, and its various modifications for both the cases $\alpha = 1$ and $\alpha = 3$. $k^2$ is dimensionless. Symbols are explained in the figure.

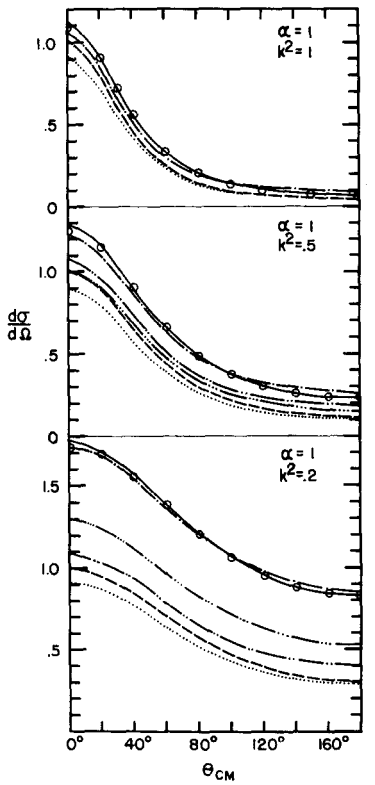



FIG. 2. Angular distribution for the case $\alpha = 1$; the exact value, the Born prediction, and its various modifications. Symbols are the same as in Fig. 1.

Born amplitude is sufficient to predict the exact total cross section in the low-energy region for the case of a very weak central potential. As for the angular distribution up to the values of $k^2 = 2.0$, the Born prediction is always lower than the exact one (Fig. 2). With the $S$-wave modification it gives values very close to the exact. With $S$- and $P$-wave modifications it gives values almost identical to the exact ones, except in the very low-angular region ($\theta < 20°$), where there is a little disagreement which minimizes with the increase of $k^2$. For values of $k^2$ greater than 2.0, the predictions of the Born approximation are fairly close to the exact values and hardly need any modification. Next we discuss the case of $\alpha = 3$ when the central potential is strong enough to give a loosely bound $S$-wave state. In this case the total cross section predicted by the Born approximation is higher than the exact value in the low-energy region (Fig. 1). This difference gets minimized with the increase of $k^2$ and, beyond $k^2 = 2.0$, it is not appreciable. At this point we would like to draw attention to the behavior of the Born total cross section with respect to the behavior of the exact values with the variation of potential strength. In the case of $\alpha = 1$ the Born prediction is lower, while for $\alpha = 3$ it is higher than the exact value in the low-energy region. This difference diminishes with increasing energy, and in the higher-energy region the Born approximation holds good.

In the case of $\alpha = 3$ we see that if we modify the $S$- and $P$-wave part of the Born amplitude in the way discussed above, the total cross section given by the modified Born amplitude agrees closely with the exact one. In the case of angular distribution (Fig. 3), the difference between the Born values and the exact ones is significant both in the shape of the curve and in its magnitude for $k^2 < 1$, beyond which the Born values have the same nature but different magnitude. But in all the cases we see that the modification of $S$ and $P$ waves gives a close fit to the exact values. Thus, in the case of a moderately weak central potential, the first two lower partial-wave modifications of the Born amplitude are sufficient to give a close fit to the exact values of total cross section and angular distribution.

## 4. UNITARIZATION SCHEMES

In this section we consider the following question: Given the Born amplitude, how may an amplitude be

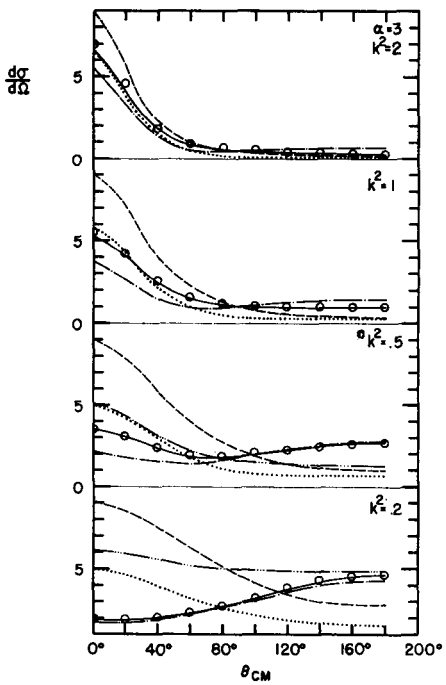



FIG. 3. Angular distribution for the case $\alpha = 3$. Symbols are the same as in Fig. 1.

constructed which satisfies unitarity and some dynamical theory? If the Born amplitude was originally constructed from a potential, then the solution is easy: the potential is reconstructed and a Schrödinger equation is solved. However, in some problems the Born amplitude does not come from a potential, but is derived from the lowest-order Feynman diagrams or simply by postulating it as the "input potential."

One technique which has been quite successful in the past few years is to require that the amplitude satisfy the Mandelstam representation (which serves as the dynamics). The left-hand cut is taken to be the same as the Born amplitude, while the right-hand cut is determined by unitarity. The use of the $N/D$ equations makes this program rather straightforward.

The unitarity condition is

$$\operatorname{Im} f_l(k^2 + i\epsilon) = k\,|f_l(k^2 + i\epsilon)|^2, \tag{4.1}$$

where $k$ is the relative momentum. The amplitude is then written as a ratio

$$f_l(k^2) = N_l(k^2)/D_l(k^2), \tag{4.2}$$

where $N_l(k^2)$ has only left-hand singularities and $D_l(k^2)$ has only right-hand singularities. The unitarity conditions then become

$$\operatorname{Im} D_l(k^2 + i\epsilon) = -kN_l(k^2), \quad k^2 \geq 0. \tag{4.3}$$

Using Eqs. (4.3) and the Cauchy integral theorem, a dispersion relation may be written for $D$:

$$D_l(k^2) = -\frac{1}{\pi}\int_0^\infty dk^2\,\frac{k' N_l(k'^2)}{k'^2 - k^2} + A(k^2), \tag{4.4}$$

where $A(k^2)$ is any real analytic function. It is usually chosen to be equal to one so that $D(\infty) = 1$.

A dispersion relation is written for $N$, using the fact that $N$ must have the same left-hand singularities as $D_l(k^2)b_l(k^2)$ [where $b_l(k^2)$ has the same left-hand cut as $f_l(k^2)$ and no right-hand cut], and we approximate it by the Born amplitude

$$N_l(k^2) = b_l(k^2)D_l(k^2) - \frac{1}{\pi}\int_0^\infty dk^2\,\frac{b_l(k^2)\operatorname{Im} D_l(k^2)}{k'^2 - k^2}. \tag{4.5}$$

The normalization imposed on $N_l(k^2)$ is that for large $k^2$ it must approach $b_l(k^2)\,D_l(k^2)$; that is, for high-energy regions the amplitude and the Born amplitude become equal.

The amplitude which is thus constructed does not satisfy the threshold behavior $f_l(k^2) \alpha k^{2l}$. Equations (4.4) and (4.5) can be modified so that the threshold behavior is satisfied.

Now the question stated at the beginning of the section is modified: Given the Born amplitude, how may an amplitude be constructed which satisfies unitarity? This question may be answered using the concept of geometrical unitarization.[3] Unitarity implies that the partial-wave amplitude may be written as

$$f_l(k^2) = e^{i\delta_l(k^2)}\sin\delta_l(k^2)/k. \tag{4.6}$$

We may describe the content of Eq. (4.6) by the use of Fig. 4. In the complex $f_l(k^2)$ plane the locus of points, which satisfy Eq. (4.6) with $\delta_l(k^2)$ real, is the circle

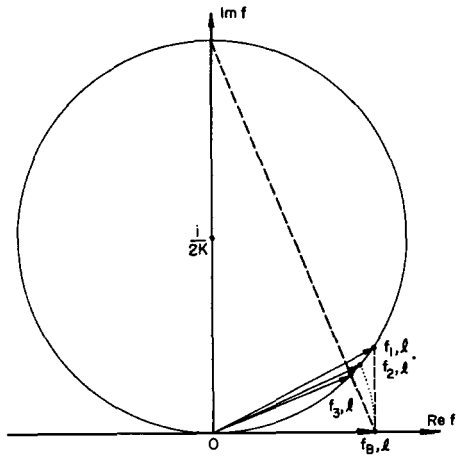



FIG. 4. The geometrical unitarization scheme.

[3] M. J. Moravcsik, Ann. Phys. (N.Y.) 30, 10 (1964).

whose center is at $i/2k$ and whose radius is $1/2k$. Therefore we may say that any amplitude lying on this circle satisfies unitarity.

The Born amplitude is real. Therefore let a partial-wave Born amplitude be represented by the vector $O \rightarrow f_{B,l}$. Any rule which transforms the $O \rightarrow f_{B,l}$ vector into a vector which lies on the circle is an adequate rule for transforming a partial-wave Born amplitude into one which satisfies unitarity. Since there evidently is an infinity of such rules (because we are not imposing the restrictions of dynamics), any given rule must be tested against a dynamical theory in order to give us confidence in the use of that rule for any given Born amplitude. We outline three rules which have been used, and we examine them more carefully for the Born amplitude which comes from a Yukawa potential.

The first rule is to set the real part of the amplitude equal to the Born amplitude. This yields the amplitude given by the vector $O \rightarrow f_{1,l}$ in Fig. 4. In terms of phase shifts this yields

$$\delta_l(k^2) = \tfrac{1}{2} \sin^{-1}(2kf_{B,l}). \tag{4.7}$$

When $f_{B,l} > (2k)^{-1}$, this method obviously does not work.

The second method is to set the magnitude of the amplitude equal to the Born amplitude. This corresponds to rotating the vector $O \rightarrow f_{B,l}$ about the origin until it intersects the circle at $f_{2,l}$. In terms of phase shifts this yields

$$\delta_l(k^2) = \sin^{-1}(kf_{B,l}). \tag{4.8}$$

When $f_{B,l} > k^{-1}$, this rule also does not work.

The third method is based on the $K$ matrix. The $S$ matrix is written as

$$S = \frac{1 + iK}{1 - iK}. \tag{4.9}$$

When $S$ is expanded in powers of the $K$ matrix, this becomes

$$S = 1 + 2iK - 2K^2 + \cdots. \tag{4.10}$$

The leading term in $K$, $2iK$, is set equal to $2ikf_B$, which corresponds to equating the Born amplitude with the leading term in the $S$ matrix. In terms of phase shifts this becomes

$$\delta_l(k^2) = \tan^{-1}(kf_{B,l}). \tag{4.11}$$

This rule may be stated differently: The real part of the inverse of the amplitude is equal to the inverse of the Born amplitude. The geometrical interpretation of this rule is that the amplitude is given by the intersection of the circle and the line which joins the points $f_{B,l}$ and $i/k$.

The three geometrical unitarization methods above are equal to order $kf_{B,l}$ so that they give approximately the same phase shifts in a very weak coupling theory. In a stronger coupling theory ($kf_{B,l} \sim 1$) the three methods differ. (In fact, methods one and two may fail.) This strongly suggests that for a given Born amplitude the method must be checked against dynamics.

## 5. TOTAL CROSS SECTION AND ANGULAR DISTRIBUTION

### A. Modification of Born Amplitude due to Geometrical Unitarity Correction

Next we discuss the effects of application of the unitarity condition to the Born amplitude for the case of a weak Yukawa potential given by Eq. (3.1). In this case also the modified Born amplitude is constructed in the manner discussed in Sec. 2. To calculate the exact partial-wave scattering amplitude which replaces the projected-out partial-wave Born amplitude, we use the phase shifts obtained from the application of the unitarity condition to the Born amplitude. Thus, we are actually replacing the partial-wave Born amplitude by the unitarity-corrected partial-wave Born amplitude which now has an imaginary part.

We have discussed three different methods of applying the unitarity condition and have noted the conditions under which those methods work. We have found that, for the case of very weak potential strength $\alpha = 1$, all the three methods work. In this case the Born approximation gives total cross sections much lower than its exact value in the low-energy region (Fig. 1). Applying the unitarity condition by use of the first method, we find that the modified Born amplitude gives total cross sections slightly higher than the Born prediction and gets closer to the exact value with increase in energy. As for the second method, we see that it does not change the Born prediction at all, and the total cross section given by the modified Born is equal to the Born prediction itself. Applying the third method, it is found that the total cross section given by the modified amplitude is slightly lower than that given by the Born amplitude; thus it gives the worst agreement with the exact value. In all the three cases above the total cross sections given by the $S$-wave-modified Born amplitude and $S$- and $P$-wave-modified Born amplitude have almost equal values and fall on the same line in the graph. This is because the $P$-wave suffers very small phase shifts for such a weak potential in the low-energy region.

Now, for the slightly stronger potential $\alpha = 3$, the first and the second methods of applying the unitarity condition do not work, and so we use only the third method. In this case the Born amplitude predicts total cross sections much higher than the exact value in the low-energy region. By using the third method (applying the unitarity condition), it is seen that the modified amplitude gives total cross sections lower than the Born amplitude as well as the exact value, except in the very low-energy region ($k^2 < 0.1$). In this case also the main correction comes from $S$-wave modification alone; the contributions of $P$-wave correction are very small and occur in the high-energy region.

We have also investigated the effect of the unitarity condition on the angular distribution for the case of low-energy scattering where the Born approximation fails to predict the exact value. We discuss first the case of $\alpha = 3$, when only the third method is applicable. In this case (Fig. 3) the Born approximation gives angular-distribution curves, as we have noted before, which are different from the exact ones, both in shape and magnitude for $k^2 < 1$, beyond which they have the same nature but different magnitude. The $S$-wave-modified Born amplitude lowers the magnitude of the Born prediction for $k^2 < 1$, but still it differs both in shape and magnitude with respect to the exact ones. For $k^2 = 1$ it gives an angular distribution which is lower than the Born value and the exact value in all the angles except the forward direction, where it is closer to the exact value. For higher values of $k^2$ the modified amplitude gives values closer to the exact ones, but then, with increasing energy, the Born prediction itself gets closer to the exact value. When we modify both the $S$ and $P$ wave, we see that the inclusion of the $P$ wave does not change the values significantly, but rather gives almost the same values as $S$-wave modification.

For the case $\alpha = 1$ all the three methods of unitarization work. The Born approximation gives angular distribution lower than the exact one in the low-energy region, but it gets closer to the exact value with an increase in energy (Fig. 2). For the case $k^2 < 1$ the first method gives values slightly higher, and the third method gives values slightly lower, than the Born value; method two gives values almost the same as the Born prediction. The same trend continues with the increase in the value of $k^2$, only the magnitude of the difference is reduced while the Born prediction itself gets closer to the exact value. As in the case of $\alpha = 3$, here also the inclusion of the $P$ wave to the modification contributes practically nothing.

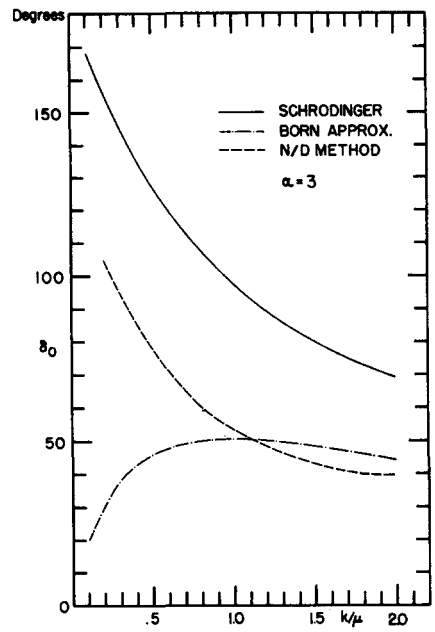



FIG. 5. The $S$-wave phase shifts for the case $\alpha = 3$ due to different methods as explained in the figure.

### B. Modification of Born Amplitude Using $N/D$ Method

While examining the partial-wave modification of the Born amplitude, we have made use of the partial-wave phase shifts as obtained by the application of $N/D$ method to Yukawa-potential scattering. As a matter of fact, Luming[4] has given such partial-wave phase shifts for the Yukawa potential strengths that we have considered in our papers (e.g., $\alpha = 3$ and 1), and we have used his values. In the case of total cross sections, the effect of using those phase shifts to modify the Born amplitude is to increase the Born (total cross section) value in the low-energy region; whereas in the higher-energy region ($k^2 \sim 1$), this modified amplitude gives values agreeing closely with those given by unitarity method three, which is lower than the Born value (Fig. 1). For the potential strength $\alpha = 1$, this increase in the total cross section in the low-energy region ($k^2 < 0.5$) brings it closer to the exact value. For the case $\alpha = 3$, an anomaly occurs in the low-energy region. An examination of Fig. 1 shows that the total cross section using the $N/D$ phase shifts gives a rather high result compared to the Born and exact total cross sections. But yet Fig. 5 shows that the $S$-wave phase shift (which dominates the low-energy cross section) for the $N/D$ is much closer to the exact phase shift than the Born value obtained by the $K$-matrix method. This anomaly

[4] M. Luming, Phys. Rev. 136, B1120 (1964).

occurs because the total cross section in this region ($k = 0.1$–$0.6$) is approximately proportional to $\sin^2 \delta_0$. Furthermore, in this region $\delta_0$ (Born) $\cong \pi - \delta_0$ (exact), so that the total cross sections given by the exact and Born phase shifts will be close compared to the results of the $N/D$ method.

In the higher-energy region ($k^2 \geq 1$) for both the values of $\alpha$, this modified amplitude gives values lower than the exact ones. Thus we find that the modification using $N/D$ phase shifts increases total cross sections in the very low-energy region and lowers its value in the higher-energy region.

Next we discuss the angular distribution given by the modified amplitude using $N/D$ phase shifts. In the case of potential strength $\alpha = 1$ for $k^2 = 0.2$ (Fig. 2), this modification increases the Born value, but still it is lower than the exact one. (Actually it lies midway between the two.) When $k^2 = 0.5$, this modification very slightly increases the Born value; whereas for $k^2 = 1.0$, it gives values lower than the Born prediction and the same as the unitarity method three. In all three cases above, the exact value is always higher than the Born value.

In the case of $\alpha = 3$ for $k^2 = 0.2$ (Fig. 3), the modified amplitude gives values which agree better with the exact one than do the Born values, even though for $\theta < 120°$ the difference in magnitude is quite great. For $\alpha = 3$ and $k^2 = 0.5$, the modification lowers the Born value and makes it closer to the exact one for $\theta < 90°$, beyond which it does not effect the Born value much. With increasing energy values $k^2 = 1.0$ and 2.0, we find that $N/D$ modification and that due to unitarization method three give the same values which are quite close to the exact ones.

## 6. DISCUSSION AND CONCLUSIONS

In this paper we have discussed the method of partial-wave modification of the Born amplitude for elastic scattering by a weak central potential in the low-energy region. In this context we have examined various geometrical unitarization methods and the results of their application have been discussed. It is found that the first two lower partial-wave modifications of the Born amplitude, using the exact phase shifts corresponding to those partial waves, works best in predicting the exact values of total cross sections and angular distributions in the low-energy region. But none of the three different unitarization methods works that well, as we have noted in Sec. 5. Although throughout this paper we have dealt with Yukawa-potential scattering, this method is applicable to any form of potential representing a weak interaction that barely gives a bound state or gives none at all, such as a nucleon–nucleon or electron–light atom in the low-energy region (in which case only the first few lower partial waves undergo quite substantial phase shifts, while others suffer very little phase shift). For nucleon–nucleon scattering due to the extreme long range of the $\pi$ meson, it is necessary to include as many as eighteen partial waves to reproduce the differential cross section. For $d$ waves and higher, the Born approximation applied to the $\pi$-meson potential gives the phase shifts quite well. The heavier mesons greatly complicate and enhance the potential in the short-range region ($r < 1\,fm$) of the $N$–$N$ interaction; consequently, in the low-energy region, the Born approximation is not applicable for $S$ and $P$ waves. Also, in the case of electron scattering by light atoms, it is found that only a few lower partial waves suffer large phase shifts, although, in all, around ten partial waves are affected.

If we want to use the partial-wave method in such cases, we need to consider quite a large number of partial waves, even if most of them (except the lower few) are slightly phase shifted. This is not quite practicable in view of the very large number of terms involved; in addition, all the required phase shifts may not be available, derived either experimentally or phenomenologically, for such a calculation. In the method described in this paper we need know only the first two or three lower partial-wave phase shifts to make the Born approximation applicable in the low-energy region. In many cases involving weak potentials as cited before (e.g., $N$–$N$ and e–He), the experimental values of such phase shifts are available, and the actual calculations for such cases are underway and will be reported soon. Incidentally, it may be noted that the interaction between an electron and a light atom in the case of elastic scattering may be represented by a potential of Yukawa form,[5] which is also used in representing the interaction between two nucleons due to the exchange of mesons.

In the calculation of thermodynamic properties of many-body systems of charged particles, i.e., plasmas, it is shown that the shielded Coulomb (Yukawa or Debye–Hückel) potential[6,7,8] is quite good in describing the effective two-particle quantum interactions in a sea of charged particles (such as a hydrogen plasma).

Recently Swan[9] has discussed various methods of counting the number of bound states in a central

[5] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955). 2nd ed., p. 170.
[6] G. M. Harris, Phys. Rev. **125**, 1131 (1962).
[7] D. Kelley and H. Margenau (unpublished), as reviewed by H. Margenau and M. Lewis, Rev. Mod. Phys. **31**, 569 (1959).
[8] G. Ecker and W. Weizel, Ann. Physik **17**, 126 (1956).
[9] P. Swan, Phys. Rev. **153**, 1379 (1967).

potential and has also given an approximate method of evaluating all the eigenenergies via scattering phase shifts for such a potential. This is helpful in choosing the appropriate strength of potential that barely gives a bound state, which is of interest to this paper. For the potential strengths considered here (e.g., $\alpha = 1$ and $\alpha = 3$), it has been found that the first bound state of zero energy appears at $\alpha = 1.7$.

Gerjuoy and Saxon[10] have studied the application of Schwinger's variational principle to central-field scattering by a Yukawa potential in the intermediate-energy region and have found that it gives better results than the Born approximation. They have also noted that this holds good for most potentials.

In cases where there is a Coulomb potential in addition to a Yukawa potential, the modified Born method does not work. This is simply due to the fact that a partial-wave expansion of the Coulomb amplitude diverges. However, in the case of proton–proton scattering, the method is still applicable if one ignores the Coulomb potential in taking the Born approximation. The lower partial waves may be projected out of the Born amplitude and replaced by exact amplitudes which contain the Coulomb effects. This procedure then ignores the Coulomb effects for the higher partial waves. This seems to be reasonable for $d$ waves and higher for energies above 20 MeV.[11]

In the nucleon–nucleon problem one of the current trends is to take the Yukawa potential (or its corresponding Born term) to represent the force due to the exchange of a meson. A field-theoretic derivation of this potential shows that the potential contains noncentral parts such as spin-orbit forces, velocity dependence, and possibly a repulsive core; however, the general nature of the potential is still of the Yukawa type. Several authors have used a superposition of Yukawa potentials to fit the nucleon–nucleon phase-shift data by varying the masses and coupling constants of the exchanged mesons. The values of the parameters vary according to the particular unitarity scheme which is used.

In order to see easily the dependence on the unitarization scheme, we idealize the situation by varying only the parameters for one meson. Furthermore, we shall ignore the spin dependence of the meson potentials so that we may directly compare the unitarization effects for the simple Yukawa potential.

By examining Fig. 5 we can tell something about the relationship between coupling constants and masses of the exchanged mesons. The figure gives the $S$-wave phase shifts as a function of $k/\mu$ [see parenthetical expression after Eq. (3.1)]. If the mass $\mu$ is increased (decreased) for a fixed value of $k$, the net effect is that the phase-shift curve is expanded (contracted) to the right (left). In the case where many mesons are exchanged (i.e., the potential is a superposition of Yukawas), the phase shift is not a function of $k/\mu$ but is a function of $k$ and $\mu$, independently. However, the shift of the curve due to a variation of $\mu$ is still monotonic. An increase (decrease) in the coupling constant monotonically increases (decreases) the phase shift for a fixed value of $K$.

Therefore, if only $S$ waves are included in an analysis (which usually is not the case), one would expect that the Schrödinger and $N/D$ methods would use about the same masses, but that the $N/D$ method would use a much larger coupling constant. The Born phase obtained by the $K$-matrix method possibly would use a slightly smaller mass, but a larger coupling constant than the Schrödinger phase. When the higher partial waves are considered, approximately the same effects occur. The Born and $N/D$ phases are always lower than the Schrödinger phase, thereby requiring larger coupling constants. Also the Born and $N/D$ phases cross at $k/\mu = 1$, so that the comparison of the coupling constants for the Born and $N/D$ depends on where the phases are fitted to the experimental points. Unfortunately, for the nucleon–nucleon problem this simple analysis is not adequate, since it appears that, in order to fit the experimental data, it is necessary to use the exchange of several mesons. Also it appears that some type of "cutoff" must be used in the calculations. The presence of these two added complications make a direct analysis of the comparison of phenomenologically determined coupling constants and masses quite difficult. Not only the method of imposing a cutoff but also the values of the cutoff parameters strongly influence the values of the phenomenological parameters.

## ACKNOWLEDGMENT

We would like to thank Dr. T. Sawada for making available his code for numerical solution of the Schrödinger equation.

[10] E. Gerjuoy and D. S. Saxon, Phys. Rev. **94**, 478 (1954).
[11] R. Arndt and M. MacGregor, Phys. Rev. **141**, 873 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Green's Function of a Particle in a Uniform External Field*

C. EFTIMIU AND K. W. THOMPSON
*Department of Physics, Indiana University, Bloomington, Indiana*

(Received 15 June 1967)

The exact propagator of a spinless charged particle in a uniform electromagnetic field with $|\mathbf{E}| = |\mathbf{H}|$ and $\mathbf{E} \cdot \mathbf{H} = 0$ is shown to lack the pole corresponding to the mass of the particle.

The purpose of this note is to point out the peculiar analytic properties of the Green's function for a charged particle moving in a uniform electromagnetic field. The electromagnetic field we have considered satisfies the conditions

$$\mathbf{E} \cdot \mathbf{H} = 0, \quad |\mathbf{E}| = |\mathbf{H}| = \text{const}; \tag{1}$$

i.e., it can be derived from an electromagnetic potential of the form

$$A_\mu = a_\mu \mathbf{k} \cdot \mathbf{x} \tag{2}$$

with $\mathbf{a} \cdot \mathbf{k} = 0$, $k^2 = 0$. Since a uniform field with the properties (1) can be viewed as the limiting case of a plane-wave electromagnetic field of very large wavelength, the Green's function for a charged particle could, in principle, be obtained from that corresponding to a radiation field. The latter has recently[1] been worked out in detail, but, rather than involve ourselves in a somewhat delicate limiting process, we prefer to rederive the Green's function directly for this particular case. This is a relatively simple procedure, especially for the case of a spinless charged particle when the Green's function is a solution of the equation

$$[(\partial_\mu + ieA_\mu)^2 + m^2]G(\mathbf{x}, \mathbf{x}') = -\delta(\mathbf{x} - \mathbf{x}'). \tag{3}$$

If then $G(\mathbf{x}, \mathbf{x}')$ is looked for in the form

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{(2\pi)^4} \int d\mathbf{p} \quad e^{-i\mathbf{p} \cdot (\mathbf{x} - \mathbf{x}')} f(\mathbf{k} \cdot \mathbf{x}, \mathbf{k} \cdot \mathbf{x}'), \tag{4}$$

it can be readily seen that $f(y, y')$ satisfies the equation

$$(p^2 - m^2 + e^2A^2 - 2e\mathbf{p} \cdot \mathbf{A})f(y, y') + 2i\mathbf{p} \cdot \mathbf{k}(\partial f/\partial y) = 1, \tag{5}$$

whose solution is

$$f(y, y') = (1/2i\mathbf{p} \cdot \mathbf{k}) \exp[-i(\alpha y + \beta y^2 - \gamma y^3)] \times \int_{y'}^{y} d\eta \exp[i(\alpha\eta + \beta\eta^2 - \gamma\eta^3)], \tag{6}$$

where

$$\alpha = (m^2 - p^2)/2\mathbf{p} \cdot \mathbf{k}, \quad \beta = e\mathbf{p} \cdot \mathbf{a}/2\mathbf{p} \cdot \mathbf{k}, \quad \gamma = e^2a^2/6\mathbf{p} \cdot \mathbf{k}. \tag{7}$$

To make the definition of the Green's function unambiguous we specify that a small negative imaginary part should be added to the mass $m$ in the above equations. This specification corresponds, of course, to the definition of the propagation function in the sense of Feynman.

We use now the representation[2]

$$\exp[-i(\beta y^2 - \gamma y^3)] = \int_{-\infty}^{\infty} ds\, A_{\mathbf{p}}(s) e^{isy} \tag{8}$$

in order to rewrite the solution corresponding to the case $y' \to -\infty$ as[3]

$$f(y, -\infty) = \frac{1}{2i\mathbf{p} \cdot \mathbf{k}} \int_0^\infty d\lambda \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} ds' A_{\mathbf{p}}^*(s') A_{\mathbf{p}}(s) \times \exp[-i\lambda(\alpha - s') + i(s - s')y]. \tag{9}$$

It follows immediately that the Fourier transform

$$F(\mathbf{q}, \mathbf{q}') = \frac{1}{(2\pi)^8} \int d\mathbf{x}\, d\mathbf{x}' e^{i\mathbf{q} \cdot \mathbf{x} - i\mathbf{q}' \cdot \mathbf{x}'} G(\mathbf{x}, \mathbf{x}') \tag{10}$$

of the Green's function constructed by means of the solution (9) is

$$F(q, q') = \frac{1}{(2\pi)^4} \frac{1}{2i\mathbf{q} \cdot \mathbf{k}} \int_0^\infty d\lambda \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} ds' A_{\mathbf{p}}^*(s') \times A_{\mathbf{q}}(s) e^{-i\lambda(\alpha - s)} \delta[\mathbf{q}' - \mathbf{q} - (s' - s)\mathbf{k}], \tag{11}$$

where it is understood that $\mathbf{q}$ replaces $\mathbf{p}$ in the definitions (7). Note also that $A_{\mathbf{q}'}(s) = A_{\mathbf{q}}(s)$ when

$$(\mathbf{q} - \mathbf{q}') \cdot \mathbf{k} = 0.$$

Consequently, in general $\mathbf{q} \neq \mathbf{q}'$, and this fact, which also occurs in the case of a radiation field, has been interpreted in Ref. 1 as indicating the description by means of the Green's function of real absorption or emission processes. Let us concern ourselves, however, with only that part of (11) which describes

* Work supported in part by the U.S. Army Research Office, Durham, North Carolina.

[1] H. R. Reiss and J. H. Eberly, Phys. Rev. **151**, 1058 (1966).

[2] A. I. Nikishov and V. I. Ritus, Zh. Eksp. Teor. Fiz. **46**, 776 (1963) [Sov. Phys.—JETP **19**, 529 (1964)].

[3] When the field (1) is considered as a special kind of radiation field, one should perhaps keep in mind that the limit of large wavelengths is not necessarily interchangeable with the limit $y' \to -\infty$. In the present work we explicitly assume that the source point is relegated to infinity while the particle is kept permanently in the uniform field.

pure propagation; it is given by the expression

$$F_0(\mathbf{q}, \mathbf{q}') = \delta(\mathbf{q} - \mathbf{q}') \times \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} ds \frac{A_q^*(s) A_q(s)}{(\mathbf{q} + s\mathbf{k})^2 - m^2}. \quad (12)$$

The effect of the presence of the electromagnetic field thus becomes evident. If $a = 0$ (i.e., $\beta = 0$, $\gamma = 0$), obviously $A_p(s) \to \delta(s)$, and therefore

$$F(\mathbf{q}, \mathbf{q}') \to \delta(\mathbf{q} - \mathbf{q}')[1/(2\pi)^4][1/(q^2 - m^2)], \quad (13)$$

i.e., the Green's function describes only the propagation of a particle of momentum $\mathbf{q}$, and the presence of the pole at $q^2 = m^2$ shows that the mass of the particle is $m$. Now, for $\mathbf{a} \neq 0$, $A_p(s)$ is a regular function of $s$. Specifically, for $\mathbf{a} \neq 0$,

$$A_p(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dy \exp[-i(sy + \beta y^2 - \gamma y^3)] \quad (14)$$

$$= \pi^{-\frac{1}{2}}(6\gamma)^{-\frac{1}{3}} \exp[is(4\beta/3\gamma) - 4i\gamma(4\beta/3\gamma)^3] \cdot \Phi(\xi), \quad (15)$$

where

$$\Phi(\xi) = (\xi/3\pi)K_{\frac{1}{3}}(\tfrac{2}{3}\xi^{\frac{3}{2}}), \quad \xi = (6\gamma)^{-\frac{1}{3}}[s - (4\beta)^2/6\gamma].$$

It is now apparent that, in contradistinction to the free-particle propagator (13), $F(\mathbf{q}, \mathbf{q}')$ does not exhibit a pole for any value of $\mathbf{q}$. More exactly, the imaginary part of the propagator $1/(m^2 - q^2)$ is $\pi\delta(m^2 - q^2)$ while, when the particle propagates in the presence of the uniform field, the imaginary part coming from (12) is proportional to

$$\left| A_q\left(\frac{m^2 - q^2}{2\mathbf{q} \cdot \mathbf{k}}\right) \right|^2, \quad (16)$$

and this is a bounded function for any value of the argument, as is shown in Fig. 1 where (16) is plotted against

$$\xi_0 = \frac{m^2 - q^2 + 2\mathbf{q} \cdot \mathbf{k}\beta/3\gamma}{2\mathbf{q} \cdot \mathbf{k}\gamma^{\frac{1}{3}}}. \quad (17)$$

This behavior is to be compared not only to the free-particle case but to the case when the particle moves in a radiation field as well. As shown in Ref. 1, the propagator has in this case a pole at $q^2 = m^2 + \Delta m^2$, where $\Delta m^2$ is a "mass shift" due to the presence of the external field. In our case, as can be seen from (12), as soon as the particle finds itself in a uniform field, the pole at $m$ "stretches" to become a "cut" extended along the entire real axis. If one is then willing to look at (16) as giving the "mass distribution" of the particle, one might be tempted to relate the position of the first peak to the mass shift observed in the radiation field case. However, strictly speaking, the absence of a pole in the propagator actually precludes any such interpretation, since it shows that the mass has no definite value, even if some seem more "probable" than others.

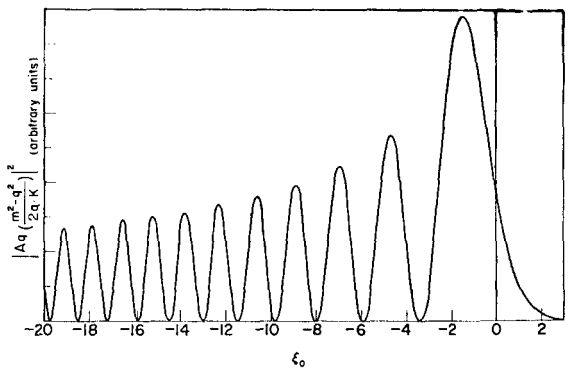



FIG. 1. Structure of $|Aq[(m^2 - q^2)/2q \cdot k]|^2$.

The possible disappearance of the one-particle contribution to the propagator of a particle coupled to an electromagnetic field has also been indicated by Schroer,[4] though in a rather different context. If the phenomenon pointed out by Schroer has the same origin as the above peculiar analytic behavior of the propagator, then it might be that even the appearance of a mass shift in the radiation-field case reflects the same anomalous behavior of the Green's function of a particle in an electromagnetic field—although severe doubts as to its occurrence and interpretation have been raised on apparently different grounds.[5]

## ACKNOWLEDGMENT

We are deeply indebted to Professor R. G. Newton for helpful discussions and interest.

[4] B. Schroer, Fortschr. Physik **11**, 1 (1963). We are indebted to Professor R. G. Newton for calling our attention to this reference.

[5] Z. Fried, A. Baker, and D. Korff, Phys. Rev. **151**, 1040 (1966); P. Stehle and P. G. DeBaryshe, Phys. Rev. **152**, 1135 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Analytic Properties of a Class of Nonlocal Interactions. I

D. GUTKOWSKI and A. SCALIA
*Istituto di Fisica dell'Università, Catania Sezione Siciliana dell'INFN, Gruppo di Catania, Italy*

(Received 8 August 1967)

The analytic properties of the functions $S_l(k)$ for a class of nonlocal interactions are studied in the complex $k$ (wavenumber) plane, for physical angular momenta. The results are compared with those of the local interactions. Two specific examples are discussed.

## 1. INTRODUCTION

Recent studies[1–4] have emphasized the importance of nonlocal, factorable interactions, because of the possibility they offer of finding exact solutions for both bound-state and scattering problems.

Another series of work[5–8] has shown that from the study of finite-rank interactions one can deduce consequences relevant for a more general class of interactions.

A detailed investigation of the analytic properties of nonlocal but finite-rank interactions therefore appears interesting. In particular, one has to expect that the methods based on the study of the integral kernel of the Lippmann–Schwinger equation might lead to explicit results for these interactions.

In this paper we are dealing with the analytic properties of the functions $S_l(k)$ in the complex $k$ plane, for physical angular momenta, for a definite class of interactions, which are separable in the sense defined in Ref. 4 and are also linear combinations of factorable ones.

In Sec. 2 we define the class of interactions we are considering. These will always be invariant against rotation and time reversal and in addition will satisfy some further conditions of a mathematical character, in order to allow computations in specific cases.

In Sec. 3 a study is made of the analytic properties of $S_l(k)$, starting directly from the hypotheses defining the interactions under consideration. Such a study is connected with the study of the integral kernel of the Lippmann–Schwinger (or Schrödinger) equation, considering this kernel as a function of the complex parameter $k$ (wavenumber).

In particular, we shall see that $S_l(k)$ is single valued in the complex $k$ plane, or in other words that there is only one analytic continuation from the physical sheet to the unphysical sheet of the energies. The study will be made starting from functions analogous to Jost ones, and will show the possibility of finding bound and virtual states. The possible existence of $S_l(k)$ poles, which do not correspond to bound states, and are analogous to *Ma* poles[9] for local interactions will be shown for a purely imaginary $k$; the possibility of having bound states for positive values of the energy will also be shown.

In Sec. 4 two explicit examples are given of the methods we have introduced. The former is relative to the Yamaguchi potential,[1] the latter is relative to that potential which gives the "orthogonality scattering"[10] defined by starting from the bound state of the Yamaguchi potential. The analytic properties in the angular-momenta complex plane for the same class of interactions that we have considered in this paper will be studied in a subsequent paper.

## 2. DEFINITION OF THE CONSIDERED CLASS OF INTERACTIONS

We shall consider Hamiltonians $H = H_0 + V$, where $H_0 = p^2/M$ and the matrix elements of $V$ in the momentum representation are given by[11]

$$\langle \mathbf{p}| V |\mathbf{p}'\rangle = -4\pi \frac{\lambda}{M} \sum_{l=0}^{L} \sum_{i=1}^{I_l} \sum_{m=-l}^{l} g_{il}(p) g_{il}(p') Y_l^m(\hat{\mathbf{p}}) Y_l^{m*}(\hat{\mathbf{p}}'). \quad (2.1)$$

$\lambda$ is a real number, $L$ and $I_l$ natural numbers. The form (2.1) ensures the rotational invariance of the interaction.

We shall also make the hypotheses that for any $l$ $(0 \leq l \leq L)$ and for any $i, j$ $(1 \leq i, j \leq I_l)$:

(a) $g_{il}(p)$ is nonsingular and real for real $p$;

(b) there exists a unique analytic continuation of $g_{il}(p)$ into the complex $p$ plane, apart from isolated

[1] Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).
[2] J. T. Cushing, Nuovo Cimento **28**, 819 (1963).
[3] A. N. Mitra and J. D. Anand, Phys. Rev. **130**, 2117 (1963).
[4] G. C. Ghirardi and A. Rimini, J. Math. Phys. **5**, 722 (1964).
[5] F. Coester, Phys. Rev. **133B**, 1516 (1964).
[6] S. Chisholm, J. Math. Phys. **4**, 1506 (1963).
[7] S. Tani, Ann. Phys. (N.Y.) **37**, 411 (1966).
[8] S. Tani, Ann. Phys. (N.Y.) **37**, 451 (1966).

[9] T. Y. Wu and T. Ohmura, *Quantum Theory of Scattering* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962), p. 96.
[10] R. G. Sachs, Phys. Rev. **95**, 1065 (1954).
[11] Equation (2.1) describes positive or negative definite interactions. It would seem at first sight that analysis of Sec. 3 does not need this restriction. However, it turns out that the proof exhibited in (3.17) does not apply to interactions having both attractive and repulsive parts.

singularities and such a continuation is an even, meromorphic function of $p$;

(c) $$\lim_{P\to\infty}\int_{CP} dp\, g_{il}(p)g_{il}(p) = 0,$$

where $CP$ is the half-circle having center in the origin and radius $P$, lying in the upper half-plane.

Hypothesis (a) ensures the time reversal invariance.[1] Hypotheses (b) and (c) allow us to calculate some frequently occurring integrals.

The $g_{il}(p)$ $(1 \leq i \leq I_l)$ have the following properties, which will be employed later:

2P1: If $p$ is purely imaginary, then $g_{il}(p)$ is real. This follows on from the hypotheses (a) and (b), and from the reflection principle of Riemann-Schwarz.[12,13]

2P2: $g_{il}(p)$ has a finite number of poles in the complex $p$ plane.

In fact, if there were an infinite number of poles, either they would have a limit point in the finite plane which contradicts hypothesis (b), or they would accumulate to infinity, contrary to hypothesis (c).

Let $F$ stand for the class of interactions defined by (2.1) and hypotheses (a), (b), and (c). It forms a generalization of the potential considered by Yamaguchi in Ref. 1, which is obtained by putting $L = 0$; $I_0 = 1$; $g_{10}(p) = (p^2 + \beta^2)^{-1}$ in Eq. (2.1).

### 3. ANALYTIC PROPERTIES OF $S_l(k)$

Let us consider the $S$ matrix (on the energy shell) for an interaction of class $F$.

From the usual definition one gets:

$$\begin{aligned}\langle \mathbf{k}|\, S\,|\mathbf{k}'\rangle &= \langle \psi_{\mathbf{k}}^{(-)} \mid \psi_{\mathbf{k}'}^{(+)}\rangle \\ &= \langle \mathbf{k} \mid \mathbf{k}'\rangle - 2\pi i\delta[(k^2/M) - (k'^2/M)]\langle \mathbf{k}|\, T\,|\mathbf{k}'\rangle, \\ &\qquad (|\mathbf{k}| = |\mathbf{k}'|), \quad (3.1)\end{aligned}$$

$$\langle \mathbf{k}|\, T\,|\mathbf{k}'\rangle = \langle \mathbf{k}|\, V\,|\psi_{\mathbf{k}'}^{(+)}\rangle, \quad (3.2)$$

$$|\psi_{\mathbf{k}'}^{(+)}\rangle = |\mathbf{k}'\rangle + (E - H_0 + i\epsilon)^{-1}V\,|\psi_{\mathbf{k}'}^{(+)}\rangle. \quad (3.3)$$

Let us now consider the resolution of the identity

$$\mathbf{1} = \sum_{l,m}\int_0^\infty dk\,|klm\rangle\langle klm|, \quad (3.4)$$

in which the states $|klm\rangle$ are normalized by

$$\langle k'l'm' \mid klm\rangle = \delta(k - k')\delta_{ll'}\delta_{mm'}. \quad (3.5)$$

Thus we have

$$\langle \mathbf{k}' \mid klm\rangle = k^{-1}\delta(k - |\mathbf{k}'|)Y_l^m(\hat{\mathbf{k}}').$$

For an interaction of class $F$ we have [as a consequence of (2.1)]

$$\langle kl_1m_1|\, V\,|k'l_2m_2\rangle = \delta_{l_1l_2}\delta_{m_1m_2}(-)4\pi kk'\frac{\lambda}{M}\sum_{i=1}^{I_{l_1}} g_{il_1}(k)g_{il_2}(k'). \quad (3.6)$$

Let us define $S_l(k)$ by

$$\langle k'l'm'|\, S\,|klm\rangle = \delta_{mm'}\delta_{ll'}\delta(k - k')S_l(k) \quad (3.7)$$

and $\psi_{k'l}^{(+)}(k)$ by

$$\psi_{k'l}^{(+)}(k) = \langle klm \mid \psi_{\mathbf{k}'}^{(+)}\rangle[Y_l^{m*}(\hat{\mathbf{k}}')]^{-1}. \quad (3.8)$$

If we insert the resolution of the identity (3.4), in Eq. (3.1), taking into account (3.2), (3.3), (3.6), (3.7), and the orthogonality of spherical harmonics $Y_l^m(\hat{\mathbf{k}})$, we get

$$\begin{aligned}S_l(k) = 1 - \pi ik\Big\{&-4\pi\lambda\sum_{i=1}^{I_l} g_{il}^2(k) \\ &+ \int_0^\infty dk_1k_1^2(-)4\pi\lambda\sum_{i=1}^{I_l} g_{il}(k)g_{il}(k_1)\frac{1}{k^2 - k_1^2 + i\epsilon} \\ &\times\int_0^\infty dk_4(-)4\pi\lambda k_4\sum_{j=1}^{I_l} g_{jl}(k_1)g_{jl}(k_4)\psi_{kl}^{(+)}(k_4)\Big\}.\end{aligned} \quad (3.9)$$

It is to be noticed that $\psi_{kl}^{(+)}(k_4)$ appearing in (3.9) must satisfy the integral equation, deduced from (3.3)

$$\psi_{kl}^{(+)}(k_4) = \frac{\delta(k_4 - k)}{k} - 4\pi\lambda\int_0^\infty dk_2k_2k_4\sum_{i=1}^{I_l}\frac{g_{il}(k_4)g_{il}(k_2)\psi_{kl}^{(+)}(k_2)}{k^2 - k_4^2 + i\epsilon}, \quad (3.10)$$

whose kernel is of rank $I_l$.

If we put

$$f_{i,j;l}(-k) = 4\pi\int_0^\infty dk_2\frac{k_2^2g_{il}(k_2)g_{jl}(k_2)}{k^2 - k_2^2 + i\epsilon}, \quad (3.11)$$

the solution of (3.10) is given by[14]

$$\psi_{kl}^{(+)}(k_4) = [\delta(k - k_4)/k] + [\det|\delta_{ij} + \lambda f_{i,j;l}(-k)|]^{-1} \times(-)4\pi\lambda\sum_{i,j=1}^{I_l}\frac{g_{jl}(k)g_{il}(k_4)k_4}{k^2 - k_4^2 + i\epsilon}\,d\begin{pmatrix} j \\ i\end{pmatrix}_l, \quad (3.12)$$

[12] Let us remember the principle of reflection of Riemann-Schwarz: "Suppose that a region $D$ of the $Z$ plane has as part of its boundary a segment $l$ of a straight line; and that $w = f(z)$ is an analytic function, regular in $D$ and continuous on $l$, and such that, as $z$ describes $l$, $w$ describes a straight line $\lambda$ in the $w$ plane. Let $z$ be a point of $D$, $z_1$ its reflection in $l$, and let $w_1$ be the reflection of $w$ in $\lambda$. Then $w_1 = w_1(z_1)$ is an analytic continuation of $w$."

[13] E. C. Titchmarsh, *The Theory of Functions* (Clarendon Press, Oxford, 1960), 2nd ed., p. 155.

[14] F. Riesz and B. Sz. Nagy, *Lecon d'analyse fonctionelle* (Gauthier-Villars, Paris, 1965), p. 161.

where $d\begin{pmatrix} j \\ i \end{pmatrix}_l$ is the cofactor in $\det|\delta_{ij} + \lambda f_{i,j;l}(-k)|$ of the element $\delta_{ji} + \lambda f_{j,i;l}(-k)$.

By substituting (3.12) into (3.9), after some lengthy algebra, one gets

$$S_l(k) = \frac{\det|\delta_{ij} + \lambda f_{i,j;l}(-k)| + 4\pi^2 i\lambda k \sum_{i,\nu=1}^{I_l} g_{il}(k) g_{\nu l}(k) d\begin{pmatrix} \nu \\ i \end{pmatrix}_l}{\det|\delta_{ij} + \lambda f_{i,j;l}(-k)|}. \tag{3.13}$$

Let us put

$$D_l(-k) = \det|\delta_{ij} + \lambda f_{i,j;l}(-k)|. \tag{3.14}$$

We want now to show that hypotheses (b) and (c) of Sec. 2 allow us to establish the following propositions:

3P1: (3.14) and (3.11) define uniquely a function $D_l(-k)$ which is meromorphic in the upper half-plane. This function can be continued analytically (apart from a finite number of poles) into the whole complex $k$ plane in a unique way.

We shall again call the function thus continued $D_l(-k)$. Therefore $D_l(k)$ will also be a function which is meromorphic in the whole complex $k$ plane.

3P2: The numerator in (3.13) is $D_l(k)$.

*Proof:* Let us remember 2P2. If $g_{il}(k_2)$, $g_{jl}(k_2)$ have, in the upper half-plane, respectively, $r$ and $s$ poles at the points $q_{i1}, \cdots, q_{ir}$; $q_{j1}, \cdots, q_{js}$, hypotheses (b) and (c) of Sec. 2 allow us to obtain for $f_{i,j;l}(-k)$

$$f_{i,j;l}(-k) = 4\pi^2 i\left\{\sum_{\alpha=1}^{r} \operatorname*{Res}_{k_2=q_{i\alpha}} + \sum_{\beta=1}^{s} \operatorname*{Res}_{k_2=q_{j\beta}} - \tfrac{1}{2}[g_{il}(k) g_{jl}(k) k]\right\}, \tag{3.15}$$

where $\operatorname*{Res}_{k_2=q_{i\alpha}}$ (Res) is the residue, calculated in $q_{i\alpha}(q_{j\beta})$, of the integrand of (3.11) as a function of $k_2$.

From the hypothesis (b) the last term on the right-hand side of (3.15) has the property expressed by 3P1 that we want to prove $D_l(-k)$ to have. It will be enough to prove that the other terms on the right-hand side of (3.15) have the same property. Let $q_i$ be a pole of $g_{il}(k_2)$. If $k_2^2 g_{il}(k_2) g_{jl}(k_2)$ is singular in $q_i$, then according to the hypothesis (b) there exists a natural number $M$ such that in a certain neighborhood of $q_i$ (excluding perhaps $q_i$) the Laurent expression gives

$$k_2^2 g_{il}(k_2) g_{jl}(k_2) = \frac{d_{-M}}{(k_2 - q_i)^M} + \cdots + \frac{d_{-1}}{k_2 - q_i} + d_0 + d_1(k_2 - q_i) + \cdots.$$

Let $k \neq q_i$ in (3.11). The Taylor expansion for $(k^2 - k_2^2)^{-1}$ gives

$$(k^2 - k_2^2)^{-1} = C_0(k) + C_1(k)(k_2 - q_i) + \cdots,$$

where, for any natural $m$,

$$C_m(k) = \frac{1}{m!} \frac{d^m}{dk_2^m} \frac{1}{k^2 - k_2^2}\bigg|_{k_2=q_i}.$$

Therefore

$$\operatorname*{Res}_{k_2=q_i} = C_0(k) d_{-1} + C_1(k) d_{-2} + \cdots + C_{M-1}(k) d_{-M}.$$

One can see that $C_m(k)$ (as functions of $k$) are single-valued, even meromorphic functions, having poles only in $k = q_i$, $k = -q_i$. 3P1 is thus proved.

From (3.15) the element $\delta_{ij} + \lambda f_{i,j;l}(-k)$ can be expressed as

$$\delta_{ij} + \lambda f_{i,j;l}(-k) = a_{i,j;l}(-k) - b_{i,j;l}(-k)$$

with

$$a_{i\,j;l}(-k) = \delta_{ij} + 4\pi^2 i\lambda\left(\sum_{\alpha=1}^{r} \operatorname*{Res}_{k_2=q_{i\alpha}} + \sum_{\beta=1}^{s} \operatorname*{Res}_{k_2=q_{j\beta}}\right) \quad \text{even function in } k,$$

$$b_{i,j;l}(-k) = 2\pi^2 i\lambda g_{il}(k) g_{jl}(k) k \quad \text{odd function in } k.$$

If one expands

$$D_l(-k) = \det|\delta_{ij} + \lambda f_{i,j;l}(-k)| = \det|a_{i,j;l}(-k) - b_{i,j;l}(-k)|$$

as a sum of determinants whose columns are made only with either the functions $a_{i,j;l}(-k)$ or with the functions $b_{i,j;l}(-k)$, from the particular form of $b_{i,j;l}(-k)$, determinants having two or more columns in $b_{i,j;l}(-k)$ vanish. Therefore we shall get

$$\det|\delta_{ij} + \lambda f_{i,j;l}(-k)| = \det|a_{i,j;l}(-k)| - \sum_{i,j=1}^{I_l} b_{i,j;l}(-k)\, d_{(a)}\begin{pmatrix} i \\ j \end{pmatrix}_l,$$

where $d_{(a)}\begin{pmatrix} i \\ j \end{pmatrix}_l$ is the cofactor of the element of the $i$th row and the $j$th column belonging to the matrix having the $j$th column in $b_{i,j;l}(-k)$, and the other in $a_{m,n;l}(-k)$. It is easy to see that the numerator on the right-hand side of (3.13) can be written as

$$\det|a_{i,j;l}(-k) + b_{i,j;l}(-k)|.$$

From the fact that $a_{i,j;l}(-k)$ and $b_{i,j;l}(-k)$ are single-valued meromorphic functions in the complex $k$ plane, and from their parity, 3P2 is proved.

Then the equality

$$S_l(k) = D_l(k)/D_l(-k) \tag{3.16}$$

holds true.

There are remarkable analogies between the case we have dealt with and the case of local potentials:

(a) (3.16) is formally analogous to the expression of $S_l(k)$ in terms of Jost functions for local interactions.[15]

(b) The Jost function $F_l(-k)$ for local interactions is equal to the Fredholm determinant relative to the integral kernel of the Lippmann–Schwinger equation for the $l$th partial wave. It is easy to see that $D_l(-k)$ in $\text{Im}\, k > 0$ is the Fredholm determinant of (3.10).

Let us observe that the proof of the equality between Jost function $F_l(-k)$ and Fredholm determinant of the integral Lippmann–Schwinger equation for the $l$th partial wave is deduced in Ref. 16 for local interactions, starting from assumptions which are different from those we have in our case.

(c) Zeros of $D_l(-k)$ in $\text{Im}\, k > 0$, $\text{Re}\, k = 0$ correspond to the bound-state energies of the system. In fact, one knows from the theory of integral equations[14] that zeros of $\det|\delta_{ij} + \lambda f_{i,j;l}(-k)|$ give all the characteristic values (and only these) of the homogeneous integral equation associated with (3.10), whose solution belongs to $L^2$.

(d) Between $D_l(k)$ and $D_l(-k)$ there is the so-called Hermiticity relationship

$$D_l^*(-k) = D_l(k^*). \tag{3.17}$$

*Proof:* Let us recall

$$D_l(-k) = \det|\delta_{ij} + \lambda f_{i,j;l}(-k)|, \tag{3.14}$$

(3.11), and successive assumptions. $f_{i,j;l}(-k) =$ analytic continuation of the function

$$4\pi \int_0^\infty dk_2 \frac{k_2^2 g_{il}(k_2) g_{jl}(k_2)}{k^2 - k_2^2} \quad (\text{Im}\, k > 0)$$

and the reality of $\lambda$. One sees that $D_l(-k)$ make any nonsingular point of the upper imaginary semiaxis ($\text{Im}\, k > 0$; $\text{Re}\, k = 0$) correspond to a point of the real axis in the $W = D_l(-k)$ plane. So it is also for the points of the lower imaginary semiaxis ($\text{Im}\, k < 0$, $\text{Re}\, k = 0$) according to the expansion of $D_l(-k)$ in $\det|a_{i,j;l}(-k) - b_{i,j;l}(-k)|$ due to the parity of the functions $a_{i,j;l}(-k)$, $b_{i,j;l}(-k)$ we have seen in the proof of 3P1 and 2P1.

Excluding the singular points of $D_l(-k)$, which according to 2P2 and 3P1 are in a finite number, let us apply the reflection principle of Riemann–Schwarz.[12] If $k_1$ is a point of the complex $k$ plane, its reflection in the Im $k$ axis is $-k_1^*$. If $W_1 = D_l(-k_1)$, its reflection in the Re $W$ axis is $W_1^* = D_l^*(-k_1)$, and from the reflection principle

$$W_1^* = D_l^*(-k_1) = D_l(k_1^*).$$

From (3.16) and the proof of 3P1 and 3P2 one sees how "false poles" can arise in $S_l(-k)$. Let us suppose that there exists a pole $q_i$ of $g_{il}(p)$ such that $\text{Re}\, q_i = 0$, $\text{Im}\, q_i > 0$. Then $-q_i = -k$ will be, in general, a pole of $a_{i,j;l}(-k)$ and therefore of $S_l(k)$ in $\text{Im}\, k > 0$, but not a zero of $D_l(-k)$.

Therefore, in $k = q_i$ ($\text{Im}\, q_i > 0$), $S_l(k)$ may have a pole which does not correspond to a bound state.

In both the examples of the following section we shall see that this possibility is verified.

Let us conclude this section by observing that for interactions of class $F$ there are possible bound states at positive energy. The following example will show this fact. Let us consider the potential of class $F$ and rank 1, defined by

$$\langle \mathbf{p}|\, V\, |\mathbf{p}'\rangle = -(\lambda/M) g(p) g(p')$$

with

$$g(p) = (p^2 - k_1^2)/[(p^2 - \beta^2)(p^2 - \beta^{*2})]$$

and $k_1^2$ real and positive.

Thus,

$$\lambda = \left[\int_0^\infty dp \frac{p^2(p^2 - k_1^2)}{(p^2 - \beta^2)^2 (p^2 - \beta^{*2})^2}\right]^{-1}$$

is a real characteristic value of the integral Schrödinger equation

$$\psi(p) = \lambda \int_0^\infty \frac{g(p) g(p') \psi(p') p'^2}{p^2 - k_1^2} dp',$$

which admits the normalizable solution

$$\psi(p) = N/[(p^2 - \beta^2)(p^2 - \beta^{*2})] \epsilon L^2(0, \infty)$$

relative to the value $k_1^2/M$ of the energy.

In the $r$ representation we have

$$\tilde{\psi}(r) = \pi^2 N \exp(-r\, \text{Im}\, \beta) \sin(r\, \text{Re}\, \beta)/r\, \text{Re}\, \beta\, \text{Im}\, \beta,$$

where $\beta$ is the square root of $\beta^2$ with $\text{Im}\, \beta > 0$.

## 4. EXAMPLES

This last section is devoted to the exposition of two examples, in which are applied the results we have deduced in the preceding section.

The first example refers to the potential of class $F$ and rank 1 studied by Yamaguchi in Ref. 1 and defined by

$$\langle \mathbf{p}|\, V\, |\mathbf{p}'\rangle = -(\lambda/M) g(p) g(p'),$$

$$g(p) = (p^2 + \beta^2)^{-1}, \tag{4.1}$$

$\lambda$ being chosen in such a way as to have a bound state

[15] V. De Alfaro and T. Regge, *Potential Scattering* (North-Holland Publ. Co., Amsterdam, 1965) p. 39.

[16] R. G. Newton, J. Math Phys. **1**, 319 (1960).

of energy $-\alpha^2/M$. We have obviously only $s$-wave scattering. Let us take into account that the condition on the bound state gives

$$\frac{1}{\lambda} = 4\pi \int_0^\infty dp \, \frac{p^2 g^2(p)}{\alpha^2 + p^2} = \frac{\pi^2}{\beta(\alpha + \beta)^2}.$$

The calculations of $D_0(-k)$ performed according to (3.14), (3.11), and (3.15) lead to

$$D_0(-k) = (k - i\alpha)[k + i(\alpha + 2\beta)]/(k + i\beta)^2, \quad (4.2)$$

where $\alpha$ and $\beta$ indicate the positive determinations of the square roots of $\alpha^2$ and $\beta^2$.

(3.16) and (4.2) give

$$S_0(k) = \frac{D_0(k)}{D_0(-k)} = \frac{\{(-k - i\alpha)[-k + i(\alpha + 2\beta)]\}/(-k + i\beta)^2}{\{(k - i\alpha)[k + i(\alpha + 2\beta)]\}/(k + i\beta)^2}. \quad (4.3)$$

So we find again that there is only one bound state for $k = i\alpha$ [zero of $D_0(-k)$]. On the other hand $S_0(k)$ has another pole in the upper imaginary semi-axis. This one, not being a zero of the denominator of (4.3), but a pole in its numerator, does not correspond to a bound state. It is analogous to the false poles or *Ma* poles.[9]

Let us notice that the zero of the denominator of (4.3) in $k = -i(\alpha + 2\beta)$ does not correspond to a solution of the homogeneous integral Schrödinger equation, because by definition, $D_0(-k)$ is the Fredholm determinant of this equation only if $\operatorname{Im} k > 0$. This zero therefore corresponds to a virtual state.

The other example is relative to the Hamiltonian $H^M$, defined in the following way.

Let $P$ be the projection operator onto the space of the bound states of $H = H_0 + V$, where $V$ is given by (4.1) (in this case $P$ is defined by a unique bound state), $Q = 1 - P$ the projection operator onto the space orthogonal to the space onto which $P$ projects, the Hamiltonian $H^M$ is given by[17]

$$H^M = H - QVQ = H_0 + V^M. \quad (4.4)$$

The explicit calculation of the projection operators $P$ and $Q$ shows that $V^M$ is an integral kernel of rank two, given by

$$\langle \mathbf{p}| V^M |\mathbf{p}'\rangle = -\frac{\lambda}{M} \sum_{\mu=1}^{2} f_\mu(p) g_\mu(p') \quad (4.5)$$

with

$$f_1(p) = (N^2/\lambda) g(p), \qquad g_1(p') = g(p')/(\alpha^2 + p'^2),$$

$$f_2(p) = \frac{N^2}{\lambda} \frac{g(p)}{\alpha^2 + p^2}, \qquad g_2(p') = g(p')\left[1 - \frac{N^2}{\lambda(\alpha^2 + p'^2)}\right],$$

$$\frac{1}{N^2} = \int d\mathbf{p} \, \frac{g^2(p)}{(\alpha^2 + p^2)^2},$$

and the other symbols defined as in the preceding example.

Let us mark with an upper index $M$ the quantities that are relative to $H^M$.

The calculation of $D_0^M(-k)$, made according to (3.14), (3.11), and (3.15)[18] leads to

$$D_0^M(-k) = (k - i\alpha)[k + i(\alpha + \beta) - (\alpha\beta)^{\frac{1}{2}}] \times [k + i(\alpha + \beta) + (\alpha\beta)^{\frac{1}{2}}] \times [(k + i\beta)^2(k + i\alpha)]^{-1}. \quad (4.6)$$

(3.16) and (4.6) give

$$S_0^M(k) = \frac{\{(-k - i\alpha)[-k + i(\alpha + \beta) - (\alpha\beta)^{\frac{1}{2}}][-k + i(\alpha + \beta) + (\alpha\beta)^{\frac{1}{2}}]\}}{[(-k + i\beta)^2(-k + i\alpha)]} \Big/ \frac{\{(k - i\alpha)[k + i(\alpha + \beta) - (\alpha\beta)^{\frac{1}{2}}][k + i(\alpha + \beta) + (\alpha\beta)^{\frac{1}{2}}]\}}{[(k + i\beta)^2(k + i\alpha)]}.$$

We must notice that also in this case the unique bound state is in $k = i\alpha$, which is a simple zero for $D_0^M(-k)$, while it is a double pole for $S_0^M(k)$. There exists also a false pole in $k = i\beta$ and two poles of $S_0^M(k)$ in $\operatorname{Im} k < 0$ in $k = -i(\alpha + \beta) + (\alpha\beta)^{\frac{1}{2}}$ and in $k = -i(\alpha + \beta) - (\alpha\beta)^{\frac{1}{2}}$.

## ACKNOWLEDGMENTS

Thanks are due to Professor A. Agodi for having suggested this study and for interesting discussions and advice.

Thanks are due to the referee for useful remarks.

[17] As far as the purposes of the present discussion are concerned, the fact that a Hamiltonian like (4.4) allows us, from the knowledge of its bound states, to solve the scattering problem, is of no importance.

[18] It can be easily realized that the proof given of (3.14), (3.11), and (3.15) also holds true for an interaction like (4.5), for which $f_\mu(p) \neq g_\mu(p)$ $(\mu = 1, 2)$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Asymptotic Gravitational Field of the "Electron"*

A. W. MARTIN† AND P. L. PRITCHETT‡
*Institute of Theoretical Physics, Department of Physics, Stanford University, Stanford, California*

(Received 6 September 1967)

The Einstein–Maxwell equations appropriate to the exterior metric and fields of a source characterized by mass, electric charge, and magnetic dipole moment are formulated. Because of the presence of a nonvanishing Poynting vector, the metric tensor must include off-diagonal elements. A rigorous reduction of the metric tensor to three unknown functions is accomplished and the field equations are solved to second order in the gravitational constant. Using the parameters of the electron and proton, we find that the magnetic dipole terms dominate the metric at small distances and that general relativistic effects become important at distances of the order of $10^{-22}$ cm. Possible applications of the asymptotic metric are discussed.

## I. INTRODUCTION

Occasionally in the early literature of general relativity one finds the Reissner–Nordström metric[1] for a point charge referred to as the "gravitational field of the electron." This was, of course, before the discovery of the electron's spin and magnetic dipole moment. While the solution of the Einstein–Maxwell equations is relatively straightforward in the case of a point charge with arbitrary mass, the exact metric for an electric or magnetic dipole with mass is unknown[2] and the determination of the exterior metric for the "electron," that is, a point charge and magnetic dipole moment with mass, is an even more complicated problem.

The purpose of this work is to develop the solution of the Einstein–Maxwell equations for this case in terms of an expansion in the gravitational constant $k$. The results, presented to order $k^2$, provide the asymptotic metric and electromagnetic fields of the point charge–magnetic dipole, being accurate at suitably large distances but not at very small distances. Clearly, no light can be shed on the question of metric singularities through such a perturbative approach, but the magnitude of the gravitational constant is such that for the parameters of the electron, proton, or neutron (or any known particle with electromagnetic properties), the asymptotic metric is accurate to quite small distances.

One application of the approximate metric to be developed, and a motivation for this undertaking, is the question of general relativistic effects in the energy levels of the hydrogen atom or in positronium. The generalization of the Dirac equation to curved space–time is well known[3] and all that is required in a first approximation to the hydrogen-atom problem is the metric of the proton. It is an interesting fact that the magnetic dipole moment distorts the metric as much as does the charge. For this reason the Reissner–Nordström metric, though much simpler, is inappropriate in discussions of the electron or proton. Another possible application of the metric concerns the extension of quantum electrodynamics to Riemannian manifolds.

The first and most obvious complication in the Einstein–Maxwell equations is the fact that the electromagnetic fields of a point charge plus magnetic dipole moment give rise to a nonvanishing Poynting vector. That is, the energy–momentum tensor for the electromagnetic fields contains off-diagonal elements which lead to the necessity of off-diagonal elements in the metric tensor. The simplest metric tensor possible seems at first glance to require four unknown functions in contrast to the two unknowns of many static and axially symmetric problems. When the scalar and vector potentials for the fields are included, the Einstein–Maxwell equations yield coupled, nonlinear partial differential equations in six unknowns. It then turns out that one of the four metric functions can be expressed in terms of the others. This is a nontrivial generalization of the reduction of the diagonal axially symmetric metric tensor in special cases.[4]

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR Contract AF 49(638)-1389.

† Address after September 1, 1967: Department of Physics, Rutgers, The State University, New Brunswick, New Jersey.

‡ N.S.F. Predoctoral Fellow.

[1] H. Reissner, Ann. Physik **50**, 106 (1916); G. Nordström, Verhandl. Koninkl. Ned. Akad. Wetenschap. Afdel. Natuurk. **26**, 1201 (1918).

[2] Solutions of the field equations corresponding to an electric or magnetic dipole moment are known, but these solutions do not contain a mass term. Instead, they contain terms that can be interpreted as mass dipoles. The reason is that the solutions are obtained by means of Weyl's procedure, which cannot lead to the metric of a massive electric or magnetic dipole. See, for example, H. Weyl, Ann. Physik **54**, 117 (1917); H. E. J. Curzon, Proc. London Math. Soc. **23**, 477 (1925); G. E. Tauber, Can. J. Phys. **35**, 477 (1957).

[3] See, for example, W. Pauli, Ann. Physik **18**, 337 (1933); E. Schrödinger, Berlin Ber. **11**, 105 (1932); H. S. Ruse, Proc. Roy. Soc. Edinburgh **57**, 97 (1937).

[4] J. L. Synge, *Relativity: The General Theory* (North-Holland Publ. Co., Amsterdam, 1960), pp. 309–312.

While the possibility of finding exact solutions of these equations seems remote, the development of power-series solutions in powers of the gravitational constant is well defined and straightforward. It is necessary first of all to specify the boundary conditions. We will insist that the electromagnetic fields vanish at infinity and that the metric becomes asymptotically flat. (It would be a simple matter to impose a curved metric at infinity but we will not pursue this possibility here.) We also require that in the limit in which the charge and dipole moment vanish the metric reduces to the exterior Schwarzschild solution (in cylindrical coordinates). Finally, we insist that in the limit in which the gravitational constant vanishes the metric is everywhere flat and the electromagnetic fields become those of a point charge and magnetic dipole alone. This eliminates higher electromagnetic multipole contributions that could in principle be included.

In Sec. II a sufficiently general metric tensor is formulated and Maxwell's equations and the energy–momentum tensor are computed. In Sec. III we construct the Ricci tensor for our metric and present the resulting field equations. We then show that the field equations permit one of the metric functions to be expressed in terms of others, thus simplifying the equations to a small extent. The Einstein–Maxwell equations are solved to second order in the gravitational constant in Sec. IV and the three basic lengths, associated with the mass, charge, and magnetic dipole moment of the source, that characterize the metric are evaluated for the parameters of the electron and proton. It is shown that in both cases the dipole moment terms dominate the metric at small distances and that the asymptotic solutions hold down to distances of the order of $10^{-22}$ cm. Section V is devoted to conclusions.

## II. ELECTROMAGNETIC CONSIDERATIONS

Our purpose in this section is to present the metric tensor and the electromagnetic field tensor for the point charge–magnetic dipole problem and to derive the energy–momentum tensor and the content of Maxwell's equations. Because of the axial symmetry of the problem we work in cylindrical coordinates

$$x^\mu = (x^0, x^1, x^2, x^3) = (ct, r, z, \theta).$$

The metric tensor for this problem is taken to be

$$g_{\mu\nu} = \begin{bmatrix} e^\rho & 0 & 0 & \omega \\ 0 & -e^\lambda & 0 & 0 \\ 0 & 0 & -e^\lambda & 0 \\ \omega & 0 & 0 & -r^2e^{-\nu} \end{bmatrix}, \tag{1}$$

where $\rho$, $\lambda$, $\nu$, and $\omega$ are functions of $r$ and $z$ alone. The justification of this form is of course that the resulting Einstein–Maxwell equations be well defined. The contravariant form of the metric tensor is then

$$g^{\mu\nu} = s^{-1}\begin{bmatrix} r^2e^{-\nu} & 0 & 0 & \omega \\ 0 & -se^{-\lambda} & 0 & 0 \\ 0 & 0 & -se^{-\lambda} & 0 \\ \omega & 0 & 0 & -e^\rho \end{bmatrix},$$

where the auxiliary function $s$ is defined by

$$s = r^2e^{\rho-\nu} + \omega^2.$$

Finally we note that

$$g \equiv \det g_{\mu\nu} = -se^{2\lambda}.$$

The electromagnetic four-potential for the point charge–magnetic dipole problem can be written in this coordinate system as

$$A_\mu = (\Phi, 0, 0, -\psi),$$

where the scalar potential $\Phi$ and the vector potential $\psi$ are functions only of $r$ and $z$. In the flat-space limit these potentials are given by

$$\Phi = qx^{-\frac{1}{2}}, \quad \psi = \mu r^2 x^{-\frac{3}{2}},$$

where $q$ is the electric charge, $\mu$ is the magnetic dipole moment, and we have introduced the convenient measure of spherical distance

$$x = r^2 + z^2. \tag{2}$$

The covariant form of the electromagnetic field tensor (with our sign conventions) is readily seen to be

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu} = \begin{bmatrix} 0 & -\Phi_r & -\Phi_z & 0 \\ \Phi_r & 0 & 0 & -\psi_r \\ \Phi_z & 0 & 0 & -\psi_z \\ 0 & \psi_r & \psi_z & 0 \end{bmatrix},$$

where the comma, as usual, denotes ordinary differentiation and the subscripts $r$ and $z$ on the potential functions denote partial differentiation with respect to those variables.

By construction, the field tensor satisfies Maxwell's "field" equations

$$F_{\mu\nu,\lambda} + F_{\nu\lambda,\mu} + F_{\lambda\mu,\nu} = 0.$$

Maxwell's "source" equations

$$F^{\mu\nu}_{\ ;\nu} = -(4\pi/c)J^\mu,$$

where the semicolon denotes covariant differentiation,

can be written in the absence of sources (i.e., everywhere except at the origin) as

$$[(-g)^{\frac{1}{2}}F^{\mu\nu}]_{,\nu} = [s^{\frac{1}{2}}e^{\lambda}F^{\mu\nu}]_{,\nu} = 0.$$

It is now convenient to introduce some simplifying notation. Let $a$ and $b$ be arbitrary functions of $r$ and $z$. We define the operators

$$\begin{aligned}\square^2 a &= a_{rr} + a_{zz} = \partial^2 a/\partial r^2 + \partial^2 a/\partial z^2,\\ (a, b) &= a_r b_r + a_z b_z,\\ [a, b] &= a_r b_r - a_z b_z,\\ \{a, b\} &= a_r b_z + a_z b_r.\end{aligned}$$

In this notation the two nontrivial Maxwell's equations become

$$\square^2\Phi + 2r^{-1}\Phi_r - (\nu, \Phi) - (s, \Phi)/2s = r^{-2}e^{\nu}[\omega\square^2\psi + (\omega, \psi) - \omega(s, \psi)/2s], \quad (3)$$

$$\square^2\psi + (\rho, \psi) - (s, \psi)/2s = -e^{-\rho}[\omega\square^2\Phi + (\omega, \Phi) - \omega(s, \Phi)/2s]. \quad (4)$$

To obtain the Einstein equations it is necessary to construct the energy–momentum tensor which in our conventions is given by

$$T_{\mu\nu} = (4\pi)^{-1}(F_{\mu}^{\lambda}F_{\lambda\nu} + g_{\mu\nu}F^{\lambda\sigma}F_{\lambda\sigma}/4). \quad (5)$$

The nonvanishing elements of this tensor are computed to be

$$\begin{aligned}T_{00} &= (8\pi s)^{-1}e^{\rho-\lambda}[2s(\Phi, \Phi)e^{-\rho} - r^2(\Phi, \Phi)e^{-\nu} + 2\omega(\Phi, \psi) + (\psi, \psi)e^{\rho}],\\ T_{11} &= -T_{22} = -(8\pi s)^{-1}[r^2[\Phi, \Phi]e^{-\nu} - 2\omega[\Phi, \psi] - [\psi, \psi]e^{\rho}],\\ T_{33} &= (8\pi s)^{-1}r^2e^{-\lambda-\nu}[r^2(\Phi, \Phi)e^{-\nu} - 2\omega(\Phi, \psi) + 2sr^{-2}(\psi, \psi)e^{\nu} - (\psi, \psi)e^{\rho}],\\ T_{12} &= T_{21} = -(4\pi s)^{-1}[r^2\Phi_r\Phi_z e^{-\nu} - \omega\{\Phi, \psi\} - \psi_r\psi_z e^{\rho}],\\ T_{03} &= T_{30} = -(8\pi s)^{-1}e^{-\lambda}[r^2\omega(\Phi, \Phi)e^{-\nu} + 2(s - \omega^2)(\Phi, \psi) - \omega(\psi, \psi)e^{\rho}].\end{aligned}$$

It is readily verified that the trace $T = T_{\mu}^{\mu}$ of the tensor vanishes as required by the structure of the basic definition (5).

## III. GEOMETRICAL CONSIDERATIONS

With the metric tensor (1) the construction of the Ricci tensor $R_{\mu\nu}$ is a straightforward calculation. The nonvanishing elements of this tensor are found to be

$$\begin{aligned}R_{00} &= (2s)^{-1}e^{\rho-\lambda}[s\square^2\rho + \tfrac{1}{2}(\rho, s) + \omega^2(\rho, \rho) - 2\omega(\rho, \omega) + (\omega, \omega)],\\ R_{11} &= -(2s)^{-1}[s\square^2\lambda - \tfrac{1}{2}[\lambda, s] + (s - \omega^2)\rho_r(\nu_r - 2r^{-1}) + s_{rr} - \omega_r^2 - s_r^2/2s],\\ R_{22} &= -(2s)^{-1}[s\square^2\lambda + \tfrac{1}{2}[\lambda, s] + (s - \omega^2)\rho_z\nu_z + s_{zz} - \omega_z^2 - s_z^2/2s],\\ R_{33} &= (2s)^{-1}r^2e^{-\lambda-\nu}[s\square^2\nu + 2sr^{-2} + \tfrac{1}{2}(\nu, s) - \omega^2(\nu, \nu) - 2\omega(\nu, \omega) - (\omega, \omega) - r^{-1}s_r + 4\omega^2r^{-1}(\nu_r - r^{-1}) + 4\omega r^{-1}\omega_r],\\ R_{12} &= R_{21} = -(2s)^{-1}[-\tfrac{1}{2}\{\lambda, s\} + \tfrac{1}{2}(s - \omega^2)\{\rho_r\nu_z + \rho_z(\nu_r - 2r^{-1})\} + s_{rz} - \omega_r\omega_z - s_rs_z/2s],\\ R_{03} &= R_{30} = (2s)^{-1}e^{-\lambda}[s\square^2\omega + \omega(\omega, \omega) - \tfrac{1}{2}(\omega, s) - \omega(s - \omega^2)(\rho, \nu) + 2\omega r^{-1}(s - \omega^2)\rho_r].\end{aligned}$$

A partial check of these equations against known results is afforded by setting $\omega = 0$ in which case the Ricci tensor reduces to that given by Synge.[4]

The Einstein equations with our sign convention read

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = (8\pi k/c^4)T_{\mu\nu},$$

where $k$ is the gravitational constant and $R = R_{\mu}^{\mu}$ is the curvature scalar. Because the trace of the energy–momentum tensor vanishes for our electromagnetic problem, the curvature scalar must also vanish and the Einstein equations reduce to

$$R_{\mu\nu} = (8\pi k/c^4)T_{\mu\nu}.$$

With the elements of the energy–momentum tensor given in Sec. II and the Ricci tensor presented above, the field equations are

$$s\square^2\rho + \tfrac{1}{2}(\rho, s) + \omega^2(\rho, \rho) - 2\omega(\rho, \omega) + (\omega, \omega) = (2k/c^4)[(s + \omega^2)(\Phi, \Phi)e^{-\rho} + 2\omega(\Phi, \psi) + (\psi, \psi)e^{\rho}], \quad (6)$$

$$s\square^2\nu + 2r^{-2}s + \tfrac{1}{2}(\nu, s) - \omega^2(\nu, \nu) - 2\omega(\nu, \omega) - (\omega, \omega) - r^{-1}[s_r - 4\omega^2(\nu_r - r^{-1}) - 4\omega\omega_r] = (2k/c^4)[r^2(\Phi, \Phi)e^{-\nu} - 2\omega(\Phi, \psi) + r^{-2}(s + \omega^2)(\psi, \psi)e^{\nu}], \quad (7)$$

$$s\square^2\omega + \omega(\omega, \omega) - \tfrac{1}{2}(\omega, s) - \omega(s - \omega^2)(\rho, \nu) + 2r^{-1}\omega(s - \omega^2)\rho_r = -(2k/c^4)[r^2\omega(\Phi, \Phi)e^{-\nu} + 2(s - \omega^2)(\Phi, \psi) - \omega(\psi, \psi)e^{\rho}], \quad (8)$$

$$s\square^2\lambda - \tfrac{1}{2}[\lambda, s] + (s - \omega^2)\rho_r(\nu_r - 2r^{-1}) + s_{rr} - \omega_r^2 - s_r^2/2s = (2k/c^4)[r^2[\Phi, \Phi]e^{-\nu} - 2\omega[\Phi, \psi] - [\psi, \psi]e^{\rho}],$$

$$s\square^2\lambda + \tfrac{1}{2}[\lambda, s] + (s - \omega^2)\rho_z\nu_z + s_{zz} - \omega_z^2 - s_z^2/2s = -(2k/c^4)[r^2[\Phi, \Phi]e^{-\nu} - 2\omega[\Phi, \psi] - [\psi, \psi]e^{\rho}],$$

$$\tfrac{1}{2}(s - \omega^2)\{\rho, \nu\} - \tfrac{1}{2}\{\lambda, s\} - r^{-1}(s - \omega^2)\rho_z + s_{rz} - \omega_r\omega_z - s_rs_z/2s = (4k/c^4)[r^2\Phi_r\Phi_z e^{-\nu} - \omega\{\Phi, \psi\} - \psi_r\psi_z e^{\rho}].$$

The first three equations above, together with the Maxwell equations [Eqs. (3) and (4)], are independent of the function $\lambda$ and give five equations in the five unknowns $\rho$, $\nu$, $\omega$, $\Phi$, and $\psi$. The last three equations above, which are not independent, determine $\lambda$ in terms of the other functions.

It is well known[4] that if either $\Phi$ or $\psi$ vanishes so that $\omega = 0$, then the field equations are satisfied by taking $\nu = \rho$. This is the reduction of the axially symmetric metric to two unknowns referred to in Sec. I. We now ask if a similar reduction is possible in the case in which $\omega$ does not vanish. A clue is provided by the observation that the covariant and contravariant forms of the metric tensor will be more similar if $s = r^2$. This requires

$$\nu = \rho - \ln(1 - \omega^2/r^2). \tag{9}$$

It is easily shown that with this choice for $\nu$ the field equation for $\nu$ [Eq. (7)] becomes a linear combination of (6) and (8) and so (9) provides a generalization of the familiar case. Since the auxiliary function $s = r^2$ exactly, the field equations can be simplified somewhat.

Before proceeding to the perturbation-series solution, it is useful to note the exact exterior Schwarzschild solution in this metric. With $\Phi = \psi = 0$ the solution of the field equations that corresponds to a point mass $m$ at the origin and an asymptotically flat metric is simply

$$\rho(r, z) = -2km/c^2x^{\frac{1}{2}},$$
$$\lambda(r, z) = 2km/c^2x^{\frac{1}{2}} - k^2m^2r^2/c^4x^2,$$

where $x$ is defined by (2). As noted in the Introduction, we will insist that the metric for the point charge–magnetic dipole problem reduces to this metric in the limit in which the charge and dipole moment vanish.

## IV. POWER-SERIES SOLUTION

We write the electromagnetic potentials and the metric functions as power-series expansions in the gravitational constant and insert these expansions into the Einstein–Maxwell equations. The resulting equations, together with the boundary conditions, yield the following solutions through order $k^2$:

$$\Phi(r, z) = \frac{q}{x^{\frac{1}{2}}}\left[1 - \frac{km}{c^2x^{\frac{1}{2}}} + \frac{kq^2}{3c^4x} - \frac{k\mu^2(4r^2 - z^2)}{35c^4x^3} + \frac{2k^2m^2}{3c^4x} - \frac{2k^2mq^2}{3c^6x^{\frac{3}{2}}} + \frac{2k^2q^4}{15c^8x^2} + \frac{k^2m\mu^2(5r^2 - 3z^2)}{35c^6x^{\frac{7}{2}}} - \frac{4k^2q^2\mu^2(7r^2 + z^2)}{315c^8x^4} + \frac{2k^2\mu^4(40r^4 - 502r^2z^2 - 487z^4)}{15015c^8x^6}\right],$$

$$\psi(r, z) = \frac{\mu r^2}{x^{\frac{3}{2}}}\left[1 + \frac{km}{2c^2x^{\frac{1}{2}}} + \frac{kq^2}{5c^4x} + \frac{k^2m^2}{5c^4x} - \frac{4k^2mq^2}{15c^6x^{\frac{3}{2}}} + \frac{11k^2q^4}{105c^8x^2} - \frac{k^2m\mu^2(r^2 + 11z^2)}{210c^6x^{\frac{7}{2}}} - \frac{k^2q^2\mu^2(72r^2 - 13z^2)}{1155c^8x^4}\right],$$

$$\rho(r, z) = -\frac{2km}{c^2x^{\frac{1}{2}}} + \frac{kq^2}{c^4x} + \frac{k\mu^2z^2}{c^4x^3} - \frac{2k^2mq^2}{3c^6x^{\frac{3}{2}}} + \frac{k^2q^4}{6c^8x^2} + \frac{2k^2m\mu^2(r^2 - 9z^2)}{35c^6x^{\frac{7}{2}}} + \frac{k^2\mu^4z^4}{6c^8x^6} - \frac{k^2q^2\mu^2(8r^2 - 9z^2)}{35c^8x^4},$$

$$\omega(r, z) = -\frac{kq\mu r^2}{c^4x^2}\left[1 - \frac{4km}{5c^2x^{\frac{1}{2}}} + \frac{7kq^2}{15c^4x} - \frac{k\mu^2(4r^2 - 61z^2)}{105c^4x^3}\right],$$

$$\lambda(r, z) = \frac{2km}{c^2x^{\frac{1}{2}}} - \frac{kq^2z^2}{c^4x^2} - \frac{k\mu^2(r^4 - 6r^2z^2 + 2z^4)}{2c^4x^4} - \frac{k^2m^2r^2}{c^4x^2} + \frac{2k^2mq^2}{3c^6x^{\frac{3}{2}}} - \frac{k^2q^4}{6c^8x^2} - \frac{2k^2m\mu^2(r^2 - 9z^2)}{35c^6x^{\frac{7}{2}}} - \frac{k^2\mu^4z^4}{6c^8x^6} + \frac{k^2q^2\mu^2(4r^2 - 9z^2)}{35c^8x^4},$$

and to this order from the exact relation [Eq. (9)]

$$\nu(r, z) = \rho(r, z) + \frac{k^2q^2\mu^2r^2}{c^8x^4}.$$

In order to see the properties of this asymptotic metric more readily for the special cases of the electron and proton, it is convenient to define three basic lengths, associated with the mass, electric charge, and magnetic dipole moment, respectively. We define

$$R_m = km/c^2, \quad R_q^2 = kq^2/c^4, \quad R_\mu^4 = k\mu^2/c^4,$$

and find for the electron the values

$$R_m = 6.75 \times 10^{-56}\ \text{cm},$$
$$R_q = 1.38 \times 10^{-34}\ \text{cm},$$
$$R_\mu = 5.16 \times 10^{-23}\ \text{cm}.$$

The corresponding lengths for the proton are

$$R_m = 1.24 \times 10^{-52}\ \text{cm},$$
$$R_q = 1.38 \times 10^{-34}\ \text{cm},$$
$$R_\mu = 2.01 \times 10^{-24}\ \text{cm}.$$

In terms of these lengths, the elements of the metric tensor $g_{\mu\nu}$ and the potentials to first order in the gravitational constant are

$$e^{\rho} = 1 - 2R_m/x^{\frac{1}{2}} + R_q^2/x + R_\mu^4 z^2/x^3,$$

$$e^{\lambda} = 1 + 2R_m/x^{\frac{1}{2}} - R_q^2 z^2/x^2 - R_\mu^4(r^4 - 6r^2z^2 + 2z^4)/2x^4,$$

$$r^2 e^{-\nu} = r^2[1 + 2R_m/x^{\frac{1}{2}} - R_q^2/x - R_\mu^4 z^2/x^3],$$

$$\omega = -R_q R_\mu^2 r^2/x^2,$$

$$\Phi = (q/x^{\frac{1}{2}})[1 - R_m/x^{\frac{1}{2}} + R_q^2/3x - R_\mu^4(4r^2 - z^2)/35x^3],$$

$$\psi = (\mu r^2/x^{\frac{3}{2}})[1 + R_m/2x^{\frac{1}{2}} + R_q^2/5x],$$

and the second-order terms are readily computed from the second-order solution given above.

While the "long-range" behavior of the solution is dominated by the mass terms, followed by the contributions of the charge and magnetic dipole moment in that order, it is clear that for the parameters of the electron and proton the metric at small distances is dominated by the dipole–moment terms. At distances of the order of $10^{-23}$ cm, the dipole terms are of order unity and the power-series expansion converges slowly, if at all. For distances larger than this, the asymptotic solution provides an excellent approximation to the metric. Of course, in the case of the proton where the internal structure sets in at distances of the order of $10^{-13}$ cm, the "interior" metric will be substantially different in character.[5] The question of structure in the electron is less clear and the asymptotic metric obtained here presumably holds to distances much smaller than hadron sizes.

Another feature of the metric concerns the distances at which the three contributing factors are comparable in magnitude. For the parameters of the electron, the mass, charge, and dipole terms are roughly equal at distances of $10^{-11}$–$10^{-12}$ cm. For the proton this occurs at distances of the order of $10^{-14}$ cm. If such distances, or smaller ones, are important in an application to a physical problem, it is obviously necessary to include the dipole terms even though the metric is much more complicated than is the Reissner–Nordström solution.

## V. CONCLUSIONS

We have presented the Einstein–Maxwell equations appropriate to the problem of determining the electromagnetic fields and metric tensor for a point source characterized by mass, charge, and magnetic dipole moment. The immediate complication in such a problem is the existence of a nonvanishing Poynting vector which necessitates off-diagonal elements in the metric tensor. A sufficiently general parametrization of the metric tensor is found to involve four unknown functions of the cylindrical-coordinate system variables $r$ and $z$. Together with the two necessary electromagnetic potentials, the Einstein–Maxwell equations then lead to coupled, nonlinear partial differential equations in six unknowns.

As a result of a general nature, we have indicated that one of the metric functions can be solved for in terms of the others. This is a generalization of well-known results for simpler axially symmetric problems and simplifies the remaining field equations to some extent. The possibility of obtaining exact solutions of the equations seems extremely remote, however, and we have utilized expansions in powers of the gravitational constant to construct the asymptotic solutions. These solutions, presented here through second order in the gravitational constant, are characterized by three basic lengths associated with the mass, charge, and magnetic dipole moment of the source.

For the physical parameters of the electron and the proton, it turns out that the length associated with the dipole moment is many orders of magnitude larger than the other two lengths and that general relativistic effects become important at distances of the order of $10^{-22}$–$10^{-23}$ cm. For the proton, of course, this is well within the internal structure and it seems improbable that the curvature of the metric is at all important for the proton. The situation for the electron is much less clear. It is possible that distances of $10^{-22}$ cm are important in certain aspects of electron physics and that general relativistic effects might lead to observable consequences.

Possible applications of the asymptotic metric developed here are considerations of curved space–time effects in the energy levels of the hydrogen atom or those of positronium. It would be amusing, for example, to calculate the general relativistic contribution to the Lamb shift. While it is undoubtedly very small, these considerations suggest that it will be larger than the usual order-of-magnitude estimates indicate. Intimately connected with such applications is the extension of quantum electrodynamics to Riemannian manifolds. We are hopeful that the asymptotic metric of the electron presented here will be of use in the further development of the theory.

[5] For an exact solution of the field equations inside a distribution of mass and charge see, for example, C. F. Kyle and A. W. Martin, Nuovo Cimento **50**, 583 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Gravitational Radiation in an Expanding Universe

S. W. HAWKING
*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England*

(Received 5 July 1967)

Asymptotic expansions are used to study outgoing gravitational radiation in an expanding, dust-filled Friedmann universe of negative curvature. It is found that the interaction with the matter modifies the "peeling-off" behavior. A quantity is defined which is interpreted as the total mass of the source and the disturbance, and which monotonically decreases as gravitational radiation is emitted. The group of coordinate transformations that preserve the asymptotic form of the metric and the boundary conditions is the same as the isometry group of the undisturbed Friedmann model. This may indicate that no physical significance attaches to the extra transformations of asymptotically flat space which are not contained in the inhomogeneous Lorentz group.

## 1. INTRODUCTION

Gravitational radiation in empty, asymptotically flat space has been studied by asymptotic expansions by a number of authors.[1-4] It is found that the Weyl tensor, which represents the gravitational radiation field, exhibits a behavior known as "peeling-off": various different components go as different powers of the affine distance. It is also found that the asymptotic group, the group of coordinate transformations that preserve the asymptotic form of the metric, contains not only the 10-parameter inhomogeneous Lorentz group, but also certain infinite-dimensional "supertranslations."[5] It has been suggested that these might be of significance in connection with elementary-particle symmetries.[6]

Emptiness and asymptotic flatness are reasonable approximations in considering regions of the universe which are small compared to the Hubble radius. However, for cosmological applications one would like to study the propagation of gravitational radiation over very large distances. One would have to include the effects of the presence of matter and of the expansion and curvature of the universe and these would modify the above results. This is particularly important if the asymptotic group has any physical significance, since this significance would presumably attach to the asymptotic group corresponding to the actual universe and not to the one for asymptotically flat, empty space.

It seems reasonable to assume that, on a large scale, our universe is described by one of the Friedmann models.[7] These are all conformally flat (i.e., the Weyl tensor vanishes). We shall consider outgoing gravitational radiation from a bounded source in a model that goes over asymptotically to one of the Friedmann models. We shall interpret the Weyl tensor as the gravitational radiation field and the Ricci tensor as the contribution of the matter to the curvature. It is determined by the Einstein equations:

$$R_{ab} - \tfrac{1}{2}g_{ab}R = -T_{ab},$$

where $T_{ab}$ is the energy-momentum tensor of the matter. At the present epoch of the universe the pressure is much less than the density. We shall therefore adopt the simplifying assumption of zero pressure. Inclusion of a pressure would merely complicate matters by introducing essentially non-gravitational phenomena such as sound waves.

It can be shown[8] that the metric of a Friedmann model with zero pressure can be written in one of the forms:

$$\text{(a)}\quad ds^2 = \Omega^2(t)[dt^2 - d\chi^2 - \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)],$$
$$\Omega = -A(1 - \cos t), \quad A < 0;$$
$$\text{(b)}\quad ds^2 = \Omega^2(t)[dt^2 - d\chi^2 - \chi^2(d\theta^2 + \sin^2\theta\, d\phi^2)],$$
$$\Omega = \tfrac{1}{2}t^2;$$
$$\text{(c)}\quad ds^2 = \Omega^2(t)[dt^2 - d\chi^2 - \sinh^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)],$$
$$\Omega = A(\cosh t - 1), \quad A > 0.$$

The parameter $A^{-1}$ represents the kinetic plus the potential energy of the matter. In case (a) it is insufficient to prevent the universe contracting again to another singularity. This case is unsuitable for a discussion of gravitational radiation by means of asymptotic expansions as one cannot get an infinite distance from the source. Case (b) is the Einstein-de Sitter model in which the energy is just sufficient

[1] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[2] R. K. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962).
[3] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[4] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).
[5] R. K. Sachs, Phys. Rev. **128**, 2851 (1962).
[6] E. T. Newman, Nature **206**, 811 (1965).
[7] H. Bondi, *Cosmology* (Cambridge University Press, New York, 1962).
[8] S. W. Hawking, Astrophys. J. **145**, 544 (1966). This paper contains a different approach to gravitational radiation in an expanding universe.

to prevent the universe contracting again. Norman[9] has used Penrose's conformal technique[10] to study gravitational radiation in this model. However, he encountered a number of difficulties and concluded that the problem could be tackled really only by integration of asymptotic expansions. We shall therefore employ this method. As (b) is a special case and is unstable to small perturbations (such as gravitational radiation), we shall consider radiation in a metric that asymptotically approaches that of case (c), where the energy is more than enough to prevent the universe contracting again. It will be shown that the interaction with the matter modifies the peeling-off behavior and that the asymptotic group does not contain any supertranslations but is simply the isometry group of the undisturbed Friedmann model. The quantity analogous to that interpreted as mass in asymptotically flat space is found to monotonically decrease when there is outgoing radiation.

## 2. NEWMAN-PENROSE FORMALISM

We employ the notation of Newman and Penrose,[3] and Newman and Unti[4] (referred to as NP and NU, respectively). A tetrad $\{l^\mu, n^\mu, m^\mu, \bar{m}^\mu\}$ of null vectors is introduced, where

$$\begin{aligned} l^\mu m_\mu &= n^\mu m_\mu = 0, \\ l^\mu n_\mu &= 1, \quad m^\mu \bar{m}_\mu = -1. \end{aligned} \tag{2.1}$$

Then,

$$g^{\mu\nu} = l^\mu n^\nu + n^\mu l^\nu - m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu. \tag{2.2}$$

Twelve complex combinations of rotation coefficients are defined as follows:

$$\begin{aligned} \rho &= l_{\alpha;\beta} m^\alpha \bar{m}^\beta, & \mu &= -n_{\alpha;\beta} \bar{m}^\alpha m^\beta, \\ \sigma &= l_{\alpha;\beta} m^\alpha m^\beta, & \lambda &= -n_{\alpha;\beta} \bar{m}^\alpha \bar{m}^\beta, \\ \kappa &= l_{\alpha;\beta} m^\alpha l^\beta, & \nu &= -n_{\alpha;\beta} \bar{m}^\alpha n^\beta, \\ \tau &= l_{\alpha;\beta} m^\alpha n^\beta, & \pi &= -n_{\alpha;\beta} \bar{m}^\alpha l^\beta, \\ \epsilon &= \tfrac{1}{2}(l_{\alpha;\beta} n^\alpha l^\beta - m_{\alpha;\beta} \bar{m}^\alpha l^\beta), \\ \gamma &= -\tfrac{1}{2}(n_{\alpha;\beta} l^\alpha n^\beta - \bar{m}_{\alpha;\beta} m^\alpha n^\beta), \\ \alpha &= \tfrac{1}{2}(l_{\alpha;\beta} n^\alpha \bar{m}^\beta - m_{\alpha;\beta} \bar{m}^\alpha \bar{m}^\beta), \\ \beta &= -\tfrac{1}{2}(n_{\alpha;\beta} l^\alpha m^\beta - \bar{m}_{\alpha;\beta} m^\alpha m^\beta). \end{aligned} \tag{2.3}$$

## 3. FIELD EQUATIONS

We introduce a family of null hypersurfaces labeled by a coordinate $u\ (= x^1)$ and take $l_\mu = u_{;\mu}$. This implies that

$$\kappa = 0, \quad \rho = \bar{\rho}, \quad \epsilon = -\bar{\epsilon}, \quad \tau = \bar{\alpha} + \beta. \tag{3.1}$$

As the coordinate $x^2$, we take the affine parameter $r: r_{;\mu} l^\mu = 1$. The coordinates $x^3$ and $x^4$ label the null geodesics in the surfaces of constant $u$. We take the vector $n^\mu$ to be the other null vector orthogonal to the two-surfaces of constant $u$ and $r$. Then,

$$\bar{\pi} = \bar{\alpha} + \beta, \quad \mu = \bar{\mu}. \tag{3.2}$$

The vectors $m^\mu$ and $\bar{m}^\mu$ lie in these two-surfaces. We may choose them so that $\epsilon = 0$. In these coordinates we may express the tetrad vectors as:

$$\begin{aligned} &l_\mu = \delta^1_\mu, \quad l^\mu = \delta^\mu_2, \\ &n^\mu = \delta^\mu_1 + U\delta^\mu_2 + X^i \delta^\mu_i, \quad i = 3, 4, \\ &m^\mu = \xi^i \delta^\mu_i. \end{aligned} \tag{3.3}$$

These components are related to the rotation coefficients by

$$D\xi^i = \rho\xi^i + \sigma\bar{\xi}^i, \tag{3.4}$$

where

$$DX^i = 2\tau\bar{\xi}^i + 2\bar{\tau}\xi^i, \tag{3.5}$$

$$DU = -\gamma - \bar{\gamma}, \tag{3.6}$$

$$\delta X^i - \Delta\xi^i = (\mu + \bar{\gamma} - \gamma)\xi^i + \bar{\lambda}\bar{\xi}^i, \tag{3.7}$$

$$\delta\bar{\xi}^i - \bar{\delta}\xi^i = (\bar{\beta} - \alpha)\xi^i + (\bar{\alpha} - \beta)\bar{\xi}^i, \tag{3.8}$$

$$\delta U = -\bar{\nu}, \tag{3.9}$$

where

$$\begin{aligned} D &= l^\mu \nabla_\mu = \partial/\partial r, \quad \delta = m^\mu \nabla_\mu = \xi^i \partial/\partial x^i, \\ \Delta &= n^\mu \nabla_\mu = U\partial/\partial r + \partial/\partial u + X^i \partial/\partial x^i. \end{aligned} \tag{3.10}$$

The rotation coefficients are related to the components of the Ricci and Weyl tensors by

$$D\rho = \rho^2 + \sigma\bar{\sigma} + \Phi_{00}, \tag{3.11}$$

$$D\sigma = 2\rho\sigma + \Psi_0, \tag{3.12}$$

$$D\tau = 2\tau\rho + 2\bar{\tau}\sigma + \Psi_1 + \Phi_{01}, \tag{3.13}$$

$$D\alpha = (\alpha + \bar{\tau})\rho + \beta\bar{\sigma} + \Phi_{10}, \tag{3.14}$$

$$D\beta = \beta\rho + (\alpha + \bar{\tau})\sigma + \Psi_1, \tag{3.15}$$

$$D\gamma = 2\tau\alpha + 2\bar{\tau}\beta + \tau\bar{\tau} + \Psi_2 - \Lambda + \Phi_{11}, \tag{3.16}$$

$$D\lambda - \bar{\delta}\bar{\tau} = \lambda\rho + \bar{\sigma}\mu + \bar{\tau}^2 + (\alpha - \bar{\beta})\bar{\tau} + \Phi_{20}, \tag{3.17}$$

$$D\mu - \delta\bar{\tau} = \mu\rho + \sigma\lambda + \tau\bar{\tau} - (\bar{\alpha} - \beta)\bar{\tau} + \Psi_2 + 2\Lambda, \tag{3.18}$$

$$D\nu - \Delta\bar{\tau} = 2\bar{\tau}\mu + 2\tau\lambda + (\gamma - \bar{\gamma})\bar{\tau} + \Psi_3 + \Phi_{21}, \tag{3.19}$$

$$\Delta\lambda - \bar{\delta}\nu = (\bar{\gamma} - 3\gamma - 2\mu)\lambda + (3\alpha + \bar{\beta})\nu - \Psi_4, \tag{3.20}$$

[9] D. Norman, Ph.D. thesis, London University, 1964.
[10] R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

$$\delta\rho - \bar{\delta}\sigma = \rho\tau - (3\alpha - \bar{\beta})\sigma - \Psi_1 + \Phi_{01}, \quad (3.21)$$

$$\delta\alpha - \bar{\delta}\beta = \mu\rho - \sigma\lambda + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta - \Psi_2 + \Lambda + \Phi_{11}, \quad (3.22)$$

$$\delta\lambda - \bar{\delta}\mu = \mu\bar{\tau} + (\bar{\alpha} - 3\beta)\lambda - \Psi_3 + \Phi_{21}, \quad (3.23)$$

$$\delta\nu - \Delta\mu = \mu^2 + \lambda\bar{\lambda} + (\gamma + \bar{\gamma})\mu - 2\beta\nu - \bar{\nu}\bar{\tau} + \Phi_{22}, \quad (3.24)$$

$$\delta\gamma - \Delta\beta = \mu\tau - \sigma\nu - (\gamma - \bar{\gamma} - \mu)\beta + \alpha\bar{\lambda} + \Phi_{12}, \quad (3.25)$$

$$\delta\tau - \Delta\sigma = \sigma\mu + \bar{\lambda}\rho + 2\beta\tau - (3\gamma - \bar{\gamma})\sigma + \Phi_{02}, \quad (3.26)$$

$$\Delta\rho - \bar{\delta}\tau = -\rho\mu - \sigma\lambda - 2\alpha\tau + (\gamma + \bar{\gamma})\rho - \Psi_2 - 2\Lambda, \quad (3.27)$$

$$\Delta\alpha - \bar{\delta}\gamma = \rho\nu - (\tau + \beta)\lambda + (\bar{\gamma} - \gamma - \mu)\alpha - \Psi_3, \quad (3.28)$$

where

$$\Phi_{00} = -\tfrac{1}{2}R_{\alpha\beta}l^\alpha l^\beta, \quad \Phi_{22} = -\tfrac{1}{2}R_{\alpha\beta}n^\alpha n^\beta,$$
$$\Phi_{11} = -\tfrac{1}{4}R_{\alpha\beta}(l^\alpha n^\beta + m^\alpha \bar{m}^\beta), \quad \Lambda = R/24,$$
$$\Phi_{01} = \overline{\Phi}_{10} = -\tfrac{1}{2}R_{\alpha\beta}l^\alpha m^\beta, \quad \Phi_{02} = \overline{\Phi}_{20} = -\tfrac{1}{2}R_{\alpha\beta}m^\alpha m^\beta,$$
$$\Phi_{12} = \overline{\Phi}_{21} = -\tfrac{1}{2}R_{\alpha\beta}n^\alpha m^\beta, \quad (3.29)$$

and

$$\Psi_0 = -C_{\alpha\beta\gamma\delta}l^\alpha m^\beta l^\gamma m^\delta,$$
$$\Psi_1 = -C_{\alpha\beta\gamma\delta}l^\alpha n^\beta l^\gamma m^\delta,$$
$$\Psi_2 = -\tfrac{1}{2}C_{\alpha\beta\gamma\delta}(l^\alpha n^\beta l^\gamma n^\delta - l^\alpha n^\beta m^\gamma \bar{m}^\delta), \quad (3.30)$$
$$\Psi_3 = -C_{\alpha\beta\gamma\delta}l^\alpha n^\beta \bar{m}^\gamma n^\delta,$$
$$\Psi_4 = -C_{\alpha\beta\gamma\delta}n^\alpha \bar{m}^\beta n^\gamma \bar{m}^\delta.$$

We also use the Bianchi identities. In the present notation they are[11]:

$$\bar{\delta}\Psi_0 - D\Psi_1 + D\Phi_{01} - \delta\Phi_{00} = (4\alpha - \bar{\tau})\Psi_0 - 4\rho\Psi_1 - \tau\Phi_{00} + 2\rho\Phi_{01} + 2\sigma\Phi_{10}, \quad (3.31)$$

$$\Delta\Psi_0 - \delta\Psi_1 + D\Psi_{02} - \delta\Phi_{01} = (4\gamma - \mu)\Psi_0 - 2(2\tau + \beta)\Psi_1 + 3\sigma\Psi_2 - \bar{\lambda}\Phi_{00} + 2\bar{\alpha}\Phi_{01} + 2\sigma\Phi_{11} + \rho\Phi_{02}, \quad (3.32)$$

$$3(\bar{\delta}\Psi_1 - D\Psi_2) + 2(D\Phi_{11} - \delta\Phi_{10}) + \bar{\delta}\Phi_{01} - \Delta\Phi_{00} = 3\lambda\Psi_0 - 9\rho\Psi_2 - 6\bar{\beta}\Psi_1 - (\mu + 2\gamma + 2\bar{\gamma})\Phi_{00} + (4\bar{\tau} + 2\alpha)\Phi_{01} + 4\beta\Phi_{10} + 2\rho\Phi_{11} + 2\sigma\Phi_{20} - \bar{\sigma}\Phi_{02}, \quad (3.33)$$

$$3(\Delta\Psi_1 - \delta\Psi_2) + 2(D\Phi_{12} - \delta\Phi_{11}) + \bar{\delta}\Phi_{02} - \Delta\Phi_{01} = 3\nu\Psi_0 + 6(\gamma - \mu)\Psi_1 - 9\tau\Psi_2 + 6\sigma\Psi_3 - \bar{\nu}\Phi_{00} - 2\gamma\Phi_{01} - 2\bar{\lambda}\Phi_{10} + 6\tau\Phi_{11} + (4\alpha + \bar{\tau})\Phi_{02} + 2\sigma\Phi_{21}, \quad (3.34)$$

$$3(\bar{\delta}\Psi_2 - D\Psi_3) + D\Phi_{21} - \delta\Phi_{20} + 2(\bar{\delta}\Phi_{11} - \Delta\Phi_{10}) = 6\lambda\Psi_1 - 9\bar{\tau}\Psi_2 - 6\rho\Psi_3 - 2\nu\Phi_{00} + 2\lambda\Phi_{01} - 4\bar{\gamma}\Phi_{10} + 6\bar{\tau}\Phi_{11} + (4\beta + \tau)\Phi_{20} - 2\bar{\sigma}\Phi_{12}, \quad (3.35)$$

$$3(\Delta\Psi_2 - \delta\Psi_3) + D\Phi_{22} - \delta\Phi_{21} + 2(\bar{\delta}\Phi_{12} - \Delta\Phi_{11}) = 6\nu\Psi_1 - 9\mu\Psi_2 + 6(\beta - \tau)\Psi_3 + 3\sigma\Psi_4 - 2\nu\Phi_{01} - 2\bar{\nu}\Phi_{10} + 2\mu\Phi_{11} + 2\lambda\Phi_{02} - \bar{\lambda}\Phi_{20} + 4\alpha\Phi_{12} + 2(\beta + 2\tau)\Phi_{21} - \rho\Phi_{22}, \quad (3.36)$$

$$\bar{\delta}\Psi_3 - D\Psi_4 + \bar{\delta}\Phi_{21} - \Delta\Phi_{20} = 3\lambda\Psi_2 - 2(\alpha + 2\bar{\tau})\Psi_3 - \rho\Psi_4 - 2\nu\Phi_{10} + 2\lambda\Phi_{11} + (2\gamma - 2\bar{\gamma} + \mu)\Phi_{20} + 2(\bar{\tau} - \alpha)\Phi_{21} - \bar{\sigma}\Phi_{22}, \quad (3.37)$$

$$\Delta\Psi_3 - \delta\Psi_4 + \bar{\delta}\Phi_{22} - \Delta\Phi_{21} = 3\nu\Psi_2 - 2(\gamma + 2\mu)\Psi_3 + (4\beta - \tau)\Psi_4 - 2\nu\Phi_{11} - \bar{\nu}\Phi_{20} + 2\lambda\Phi_{12} + 2(\gamma + \mu)\Phi_{21} - \bar{\tau}\Phi_{22}, \quad (3.38)$$

$$D\Phi_{11} - \delta\Phi_{10} - \bar{\delta}\Phi_{01} + \Delta\Phi_{00} + 3D\Lambda = 2(\gamma + \bar{\gamma} - \mu)\Phi_{00} - (2\alpha + \bar{\tau})\Phi_{01} - (2\bar{\alpha} + \tau)\Phi_{10} + 4\rho\Phi_{11} + \bar{\sigma}\Phi_{02} + \sigma\Phi_{20}, \quad (3.39)$$

$$D\Phi_{12} - \delta\Phi_{11} - \bar{\delta}\Phi_{02} + \Delta\Phi_{01} + 3\delta\Lambda = (2\gamma - 3\mu)\Phi_{01} + \bar{\nu}\Phi_{00} - \bar{\lambda}\Phi_{10} + 2(\bar{\beta} - \alpha)\Phi_{02} + 3\rho\Phi_{12} + \sigma\Phi_{21}, \quad (3.40)$$

$$D\Phi_{22} - \delta\Phi_{21} - \bar{\delta}\Phi_{12} + \Delta\Phi_{11} + 3\Delta\Lambda = \nu\Phi_{01} + \bar{\nu}\Phi_{10} - 4\mu\Phi_{11} - \lambda\Phi_{02} - \bar{\lambda}\Phi_{20} + (\bar{\tau} + 2\beta)\Phi_{12} + (\tau + 2\bar{\beta})\Phi_{21} + 2\rho\Phi_{22}. \quad (3.41)$$

## 4. THE UNDISTURBED METRIC

We may express the undisturbed metric (c) as

$$ds^2 = \Omega^2[-du^2 + 2du\,dt - (1 + \tfrac{1}{4}\zeta\bar{\zeta})\sinh^2(t - u)d\zeta d\bar{\zeta}], \quad (4.1)$$

where $u = t - \chi$, $\zeta = x^3 + ix^4$, and $x^3$, $x^4$ are stereographic coordinates. We calculate $r$, the affine parameter, from $l^\alpha = dx^\alpha/dr = g^{\alpha\beta}u_{;\beta}$. This gives

$$r = \int_0^t \Omega^2\,dt' + B(u, x^3, x^4). \quad (4.2)$$

Normally $B$ would be chosen so that $r = 0$ when $t = u$. However, in our case it is more convenient to take it to be zero. This means that surfaces of constant $r$ are surfaces of constant $t$. This may seem rather odd but it should be pointed out that the choice of $B$

[11] The author is indebted to R. G. McLenaghan for these. The complete Bianchi identities appear in *Brandeis Summer Institute in Theoretical Physics, 1964*, S. Deser and K. W. Ford, Eds. (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1964), Vol. I, pp. 350–351.

will not affect the asymptotic order of quantities. It proves easier to perform the calculations with this choice of $B$ but all the results could be transformed back to a more normal coordinate system if desired.

From (4.2),

$$r = A^2[\tfrac{1}{4}\sinh 2t - 2\sinh t + \tfrac{3}{2}t]. \quad (4.3)$$

The matter in the universe is assumed to be pressure-free, so its energy–momentum tensor may be written as $T_{ab} = \eta V_a V_b$, where $\eta$ is the density of the matter and $V_a$ its velocity. For the undisturbed metric,[8]

$$\eta = 6A\Omega^{-3}, \quad V_a = \Omega t_{;a}. \quad (4.4)$$

Now,

$$\Omega = 2^{\frac{1}{2}}s + A - 3A^2[2(2)^{\frac{1}{2}}s]^{-1}\log s + O(s^{-1}), \quad (4.5)$$

where $s^2 = r$. Therefore, if we tried to expand $\eta$ as a series in $s$, the result would be very messy and would involve terms of form $s^{-m}\log^n s$. This would not invalidate it as an asymptotic expansion, but would make it very tedious to handle. For convenience, therefore, we perform the expansions in terms of $\Omega(r)$ which we define to be the same function of $r$ as it is in the undisturbed metric. That is,

$$\Omega = A(\cosh t - 1), \quad (4.6)$$

where $t$ is related to $r$ by (4.3). Then,

$$\begin{aligned} d\Omega/dr &= \Omega^{-1}(1 + 2A\Omega^{-1})^{\frac{1}{2}} \\ &= \Omega^{-1}(1 + A\Omega^{-1} - \tfrac{1}{2}A^2\Omega^{-2} + \tfrac{1}{2}A^3\Omega^{-3} \\ &\qquad - \tfrac{5}{8}A^4\Omega^{-4} + \tfrac{7}{8}A^5\Omega^{-5} - \cdots). \end{aligned} \quad (4.7)$$

As the pressure is zero, the energy–momentum tensor and hence the Ricci tensor have only four independent components. We shall take these as $\Lambda$, $\Phi_{00}$, $\Phi_{01}$ (since $\Phi_{01}$ is complex it represents two components). In terms of these, the other components may be expressed as

$$\begin{aligned} \Phi_{11} &= 3\Lambda + \Phi_{00}^{-1}\Phi_{01}\bar{\Phi}_{01}, \\ \Phi_{22} &= 36\Lambda^2\Phi_{00}^{-1}(1 + \tfrac{1}{6}\Lambda^{-1}\Phi_{00}^{-1}\Phi_{01}\bar{\Phi}_{01})^2, \\ \Phi_{12} &= \bar{\Phi}_{21} = 6\Lambda\Phi_{01}\Phi_{00}^{-1}(1 + \tfrac{1}{6}\Lambda^{-1}\Phi_{00}^{-1}\Phi_{01}\bar{\Phi}_{01})^2, \\ \Phi_{02} &= \Phi_{20} = \Phi_{00}^{-1}\Phi_{01}^2. \end{aligned} \quad (4.8)$$

For the undisturbed metric,

$$\Lambda = \tfrac{1}{4}A\Omega^{-3}, \quad \Phi_{00} = 3A\Omega^{-5}, \quad \Phi_{01} = 0. \quad (4.9)$$

## 5. BOUNDARY CONDITIONS

We wish to consider outgoing gravitational radiation from a bounded source in a metric that asymptotically approaches the undisturbed metric described above. By analogy with the asymptotically flat case,[3,4] we would expect $\Psi_4$ to represent the radiation field and to be $O(r^{-1})$. The Ricci-tensor components, $\Lambda$ and $\Phi_{00}$ should have their undisturbed values plus terms of smaller order. To determine these orders and those of $\Phi_{01}$ and $\Psi_0$, there are two ways in which we may proceed. We may try to find the smallest orders that will permit a general radiation field $\Psi_4$ of order $r^{-1}$. Larger-order terms than these turn out to have their $u$ derivatives dependent only on themselves and not on the $r^{-1}$ coefficient of the radiation field. They thus represent disturbances not created by the radiation field and so are not considered in this paper.

Alternatively, we may proceed by a method of successive approximations. We take the undisturbed values of the rotation coefficients and use them to solve the Bianchi identities as field equations for the Weyl tensor using the asymptotically flat boundary condition[3] $\Psi_0 = O(r^{-5})$. Then, substituting these values of the Weyl tensor in Eqs. (3.4)–(3.28), we calculate the disturbances induced in the rotation coefficients and, using these in Eqs. (3.39)–(3.41), we calculate the disturbances in the Ricci-tensor components. We then substitute them back in the Bianchi identities (3.31)–(3.38) and calculate the corrections to the Weyl-tensor components. Further iteration does not affect the orders of the disturbances. Both these methods indicate that the boundary conditions should be

$$\Lambda = \tfrac{1}{4}A\Omega^{-3} + O(\Omega^{-7}), \quad (5.1)$$

$$\Phi_{00} = 3A\Omega^{-5} + O(\Omega^{-9}), \quad (5.2)$$

$$\Phi_{01} = O(\Omega^{-7}), \quad (5.3)$$

$$\Psi_0 = O(\Omega^{-7}). \quad (5.4)$$

We also assume that these terms are uniformly smooth[3,5]: Differentiation by $\Omega$ lowers the order by one power of $\Omega$ and differentiation by $x^i$ $(i = 3, 4)$ leaves the order unchanged. It is shown that if these boundary conditions hold on one null hypersurface $u = u^0$, they hold on succeeding hypersurfaces and that these conditions are the most severe to permit a general radiation field $\Psi_4$ of order $r^{-1}$.

## 6. INTEGRATION

The procedure is similar to that employed in NP and NU. We first integrate the equations involving radial derivatives and then use the remaining equations to establish relations between the "constants" of integration. The actual calculations are rather tedious though fairly straightforward. Only the main points are given.

We start by integrating Eqs. (3.11) and (3.12) in a

similar manner to NP:

$$\rho = -2\Omega^{-2} - A\Omega^{-3} + \rho^0\Omega^{-4} + (\tfrac{1}{2}A^2 - 2A\rho^0)\Omega^{-5} + (-5A^4/4 + 4A^2\rho^0 - \tfrac{1}{2}(\rho^0)^2 - \tfrac{1}{2}\sigma^0\bar{\sigma}^0)\Omega^{-6} + O(\Omega^{-7}), \quad (6.1)$$

$$\sigma = \sigma^0\Omega^{-4} + O(\Omega^{-5}), \quad (6.2)$$

where $\rho^0(u, x^i)$ and $\sigma^0(u, x^i)$ are constants of integration. Unlike NU, we cannot make $\rho^0$ zero by the transformation $r' = r + \rho^0$ since this would upset the boundary conditions (5.1) and (5.2).

By Eqs. (3.4), (3.13), (3.14), (3.15), (3.31), and the lemma used in NP, it follows that

$$\Psi'_1 = O(\Omega^{-7}), \quad (6.3)$$

$$\tau = \tau^0\Omega^{-4} + O(\Omega^{-5}), \quad (6.4)$$

$$\xi^i = \xi^{i0}\Omega^{-2} - A\xi^{i0}\Omega^{-3} + \tfrac{1}{2}(3A^2\xi^{i0} - \rho^0\xi^{i0} - \bar{\xi}^{i0}\sigma^0)\Omega^{-4} + O(\Omega^{-5}), \quad (6.5)$$

$$\alpha = \bar{\tau} - \bar{\beta} = \alpha^0\Omega^{-2} - A\alpha^0\Omega^{-3} + \tfrac{1}{2}(3A^2\alpha^0 - \rho^0\alpha^0 + \bar{\alpha}^0\bar{\sigma}^0 + \tau^0)\Omega^{-4} + O(\Omega^{-5}). \quad (6.6)$$

We use the lemma again with Eqs. (3.17), (3.18), (3.33), and (3.39) to show that

$$\Psi'_2 = 3A(\tfrac{1}{2}A^2 + \mu^0 - \tfrac{1}{2}\rho^0)\Omega^{-5} + O(\Omega^{-6}\log\Omega), \quad (6.7)$$

$$\lambda = \lambda^0\Omega^{-2} - A(\lambda^0 + \tfrac{1}{2}\bar{\sigma}^0)\Omega^{-3} + O(\Omega^{-4}), \quad (6.8)$$

$$\mu = \tfrac{1}{2}A\Omega^{-1} + \mu_0\Omega^{-2} - A(5A^2/4 + 4\mu^0 - \rho^0)\Omega^{-3} + O(\Omega^{-4}\log\Omega). \quad (6.9)$$

Then by Eqs. (3.5), (3.6), (3.16), (3.35), and (3.40),

$$\Psi'_3 = \Psi'^0_3\Omega^{-4} + O(\Omega^{-5}), \quad (6.10)$$

$$X^i = X^{i0} + O(\Omega^{-4}), \quad (6.11)$$

$$\gamma = \gamma^0 - \tfrac{1}{2}A\Omega^{-1} + \tfrac{1}{4}A^2\Omega^{-2} - A(\tfrac{3}{4}A^2 + \mu^0 - \tfrac{1}{2}\rho^0)\Omega^{-3} + O(\Omega^{-4}), \quad (6.12)$$

$$U = -\tfrac{1}{2}(\gamma^0 + \bar{\gamma}^0)\Omega^2 + O(\Omega). \quad (6.13)$$

By Eq. (3.9), $\nu = O(1)$. Then,

$$\begin{aligned} \Phi_{00,1} &= O(\Omega^{-5}), \quad \Lambda_{,1} = O(\Omega^{-5}), \\ \Phi_{01,1} &= O(\Omega^{-5}), \quad \Psi'_{0,1} = O(\Omega^{-7}), \\ \Psi'_{1,1} &= O(\Omega^{-7}), \quad \rho_{,1} = O(\Omega^{-3}), \\ \sigma_{,1} &= \sigma^0_{,1}\Omega^{-4} + O(\Omega^{-5}), \\ \tau_{,1} &= \tau^0_{,1}\Omega^{-4} + O(\Omega^{-5}), \end{aligned} \quad (6.14)$$

where a comma indicates partial differentiation.

Therefore, by Eqs. (3.19) and (3.37),

$$\nu = \nu^0 + O(\Omega^{-2}), \quad (6.15)$$

$$\Psi'_4 = \Psi'^0_4\Omega^{-2} + O(\Omega^{-3}.) \quad (6.16)$$

Using this in Eqs. (3.18) and (3.36), we have

$$\Psi'_{2,1} = O(\Omega^{-5}), \quad \mu_{,1} = O(\Omega^{-2}). \quad (6.17)$$

Then by the $\Omega^{-1}$ term of Eq. (3.24),

$$\gamma^0 + \bar{\gamma}^0 = -1. \quad (6.18)$$

This gives

$$U = \tfrac{1}{2}\Omega^2 + U^0 - A(A^2 + 2\mu^0 - \rho^0)\Omega^{-1} + O(\Omega^{-2}\log\Omega), \quad \nu^0 = 0. \quad (6.19)$$

From the orthogonality relations (2.1) and (2.2), we have that

$$\begin{aligned} g^{ij} &= -(\xi^i\bar{\xi}^j + \bar{\xi}^i\xi^j), \quad (i, j = 3, 4) \\ &= -(\xi^{i0}\bar{\xi}^{j0} + \bar{\xi}^{i0}\xi^{j0})\Omega^{-4} + O(\Omega^{-5}). \end{aligned} \quad (6.20)$$

As in NU, the coordinates $x^3$ and $x^4$ may be chosen so that the leading term in $g^{ij}$ is a conformally flat metric:

$$g^{ij} = -2P\bar{P}\delta^{ij}\Omega^{-4} + O(\Omega^{-5}), \quad (6.21)$$

where $P(u, x^i) = \xi^{03} = -i\xi^{04}$. This leaves the coordinate freedom $\zeta' = f(u, \zeta)$ where $\zeta = x^3 + ix^4$. We may use this to make $X^{i0}$ zero since by Eq. (3.7), $X^{30} + iX^{40}$ is an analytic function of $\zeta$. We also have from (3.7) that $\gamma^0 = -\frac{1}{2}$ and that $P = (2)^{-\frac{1}{2}}AQ(x^i)e^u$. To obtain agreement with the undisturbed metric (4.1), we take $Q$ to be $(1 + \frac{1}{4}\zeta\bar{\zeta})$. Then,

$$\alpha^0 = (2)^{-\frac{5}{2}}A\bar{\zeta}, \quad \mu^0 = -\tfrac{1}{4}A^2(1 + e^{2u}).$$

We now have from Eq. (3.39) that

$$\Phi_{00,1} = H\Omega^{-7} + O(\Omega^{-8}),$$

where

$$H = \tfrac{3}{2}A(5U^0 - A^2 + 2\rho^0 + A^2e^{2u}). \quad (6.22)$$

If we make the coordinate transformation

$$r' = r + 2(15A)^{-1}\int_{u^0}^{u} H\,du'', \quad (6.23)$$

then $H' = 0$. This transformation does not upset the boundary conditions (5.1)–(5.4) on the initial hypersurface $u = u^0$. It means that on succeeding hypersurfaces we choose the zero of $r$ so that $\Phi_{00} = 3A\Omega^{-5} + O(\Omega^{-8})$. Then by the $\Omega^{-4}$ and $\Omega^{-5}$ terms in Eq. (3.27),

$$\rho^0 = \tfrac{1}{2}A^2(1 - e^{2u}), \quad U^0 = 0. \quad (6.24)$$

Then,

$$\Psi'_2 = \Psi'^0_2\Omega^{-6} + O(\Omega^{-3}). \quad (6.25)$$

We now have

$$\begin{aligned} \Phi_{00,1} &= O(\Omega^{-9}), \quad \Lambda_{,1} = O(\Omega^{-7}), \\ \Phi_{01,1} &= O(\Omega^{-7}), \quad \Psi'_{0,1} = O(\Omega^{-7}). \end{aligned} \quad (6.26)$$

Therefore, if the boundary conditions (5.1)–(5.4) hold

on one null hypersurface, they will hold on succeeding hypersurfaces. The peeling-off behavior is

$$\begin{aligned} &\Psi'_4 = O(r^{-1}), \quad \Psi'_3 = O(r^{-2}), \\ &\Psi'_2 = O(r^{-3}), \quad \Psi'_1 = O(r^{-\frac{5}{2}}), \\ &\Psi'_0 = O(r^{-\frac{3}{2}}). \end{aligned} \tag{6.27}$$

The difference between this and the behavior in empty, asymptotically flat space is in the last two terms. It seems to be caused by the interaction with the matter.

Expressing the remaining quantities in terms of $\sigma^0$ we have

$$\begin{aligned} &\lambda^0 = \tfrac{1}{2}(\bar{\sigma}^0_{,1} - \bar{\sigma}^0), \quad \Psi'^0_4 = 2\lambda^0 - \lambda^0_{,1}, \\ &\tau^0 = \tfrac{1}{2}P^3\overline{\nabla}(\sigma^0 P^{-2}), \quad \Psi'^0_3 = -\tfrac{1}{2}P^3\nabla(\lambda^0 P^{-2}), \\ &\Psi'^0_2 - \bar{\Psi}'^0_2 = P^2\overline{\nabla}(\tau^0 P^{-1}) - P^2\nabla(\bar{\tau}^0 P^{-1}) + \bar{\sigma}^0\bar{\lambda}^0 - \sigma^0\lambda^0, \end{aligned} \tag{6.28}$$

where

$$\nabla = \partial/\partial x^3 + i\partial/\partial x^4;$$

$$\Psi'^0_2 + \bar{\Psi}'^0_2 \text{ is undetermined.}$$

The time derivatives (6.26) all depend on $\sigma^0$. Thus the boundary conditions (5.1)–(5.4) are the most severe to permit a general radiation field $\Psi'^0_4$. The quantity $\lambda^0$ is analogous to the Bondi "news" function.[1,2]

## 7. THE ASYMPTOTIC GROUP

The metric has the form

$$\begin{aligned} &g^{11} = g^{1i} = 0, \quad g^{12} = 1, \quad (i,j = 3,4), \\ &g^{22} = \Omega^2 + O(\Omega^{-2}), \quad g^{2i} = O(\Omega^{-4}), \\ &g^{ij} = -2P^2\delta^{ij}(\Omega^{-4} - A\Omega^{-5}) + O(\Omega^{-6}). \end{aligned} \tag{7.1}$$

The asymptotic group is the group of coordinate transformations that leaves unchanged the above form and that of the boundary conditions (5.1)–(5.4). It can be derived most simply by considering the infinitesimal transformations generated by a vector field $k^\alpha$:

$$x^{\alpha'} = x^\alpha + \epsilon k^\alpha.$$

Then the Lie derivatives of the components of the metric and the Ricci tensor must be

$$\begin{aligned} &\pounds g^{1\alpha} = 0, \quad \pounds g^{22} = O(\Omega^{-2}), \\ &\pounds g^{2i} = O(\Omega^{-4}), \quad \pounds g^{ij} = O(\Omega^{-6}), \\ &\pounds\Lambda = O(\Omega^{-7}), \quad \pounds\Phi_{00} = O(\Omega^{-9}), \\ &\pounds\Phi_{01} = O(\Omega^{-7}). \end{aligned} \tag{7.2}$$

From these it follows that

$$\begin{aligned} &k^1 = k^{01}(x^i), \quad k^2 = O(\Omega^{-2}), \\ &k^j = k^{0j}(x^i) + O(\Omega^{-2}), \end{aligned} \tag{7.3}$$

where

$$\begin{aligned} k^{03}_{,4} &= -k^{04}_{,3}, \\ k^{03}_{,3} &= k^{04}_{,4}, \end{aligned} \tag{7.4}$$

and where $k^1$ and $k^2$ are determined by $k^{0j}$. By (7.4), $k^{03} + ik^{04}$ is an analytic function of $\zeta$ (this is a consequence of the fact that we have reduced the leading term of $g^{ij}$ to a conformally flat form). It follows that $\zeta^{0'}$, the leading term in $\zeta'$, is an analytic function of $\zeta$ in the finite transformations generated by $k^\alpha$. As $\zeta^{0'}$ must have one zero and one pole, it is given by

$$\zeta^{0'} = (a\zeta + b)/(c\zeta + d), \quad ad - bc = 1. \tag{7.5}$$

The transformations of $u$ and $r$ are determined by the six real parameters in (7.5). Thus the asymptotic group is isomorphic to the Moebius group. This is isomorphic to the homogeneous Lorentz group. However, it is also isomorphic to the isometry group of a three-surface of constant negative curvature which is the isometry group of the undisturbed Friedmann model. Thus the asymptotic group is the same as the group of the undisturbed space. There are no supertranslations as in asymptotically flat space.

## 8. THE MASS

Let $l^\alpha$ and $n^\alpha$ be, respectively, the outgoing and the ingoing null vectors orthogonal to a spacelike two-sphere $S$ and let $m^\alpha$ and $\bar{m}^\alpha$ lie in $S$. The quantity

$$M(S) = (4\pi)^{-\frac{3}{2}}\left(\int dS\right)^{\frac{1}{2}}\int(-\Psi'_2 - \sigma\lambda + \Phi_{11} + \Lambda)\,dS \tag{8.1}$$

is uniquely defined. ($\pi$ here has its usual meaning.) By the Gauss–Bonnet theorem [derivable from Eq. (3.22)], $M(S)$ is also equal to $(4\pi)^{-\frac{3}{2}}(\int dS)^{\frac{1}{2}}(2\pi - \int \mu\rho\, dS)$. If we move each point of $S$ an equal distance $dr$ out along the null vectors $l^\alpha = dx^\alpha/dr$, then

$$\begin{aligned} dM/dr = (4\pi)^{-\frac{3}{2}}&\left(\int dS\right)^{\frac{1}{2}}\Big\{-\left(\int dS\right)^{-1}\left(\int \rho\, dS\right) \\ &\times\int(-\Psi'_2 - \sigma\lambda + \Phi_{11} + \Lambda)\,dS \\ &-\int\{\mu(\sigma\bar{\sigma} + \Phi_{00}) - (\alpha + \bar{\beta})\delta\rho \\ &+ \rho[\sigma\lambda + (\bar{\alpha} + \beta)(\alpha + \bar{\beta}) + \Psi'_2 + 2\Lambda]\}\,dS\Big\}. \end{aligned} \tag{8.2}$$

If we choose the scaling of $l^\mu$ so that $r$ is a luminosity parameter (i.e., $\rho = r^{-1}$ and $\int dS = 4\pi r^2$), then

$$\begin{aligned} dM/dr = (4\pi)^{-1}\int[&-\mu r(\sigma\bar{\sigma} + \Phi_{00}) \\ &+ (\alpha + \bar{\beta})(\bar{\alpha} + \beta) + \Phi_{11} + \Lambda]\,dS. \end{aligned} \tag{8.3}$$

We may regard the Ricci-tensor terms $\Phi_{00}$, $\Phi_{11}$, and $\Lambda$ as representing the mass–energy density of matter crossing the outgoing null hypersurface and the term

$\sigma\bar{\sigma}$ as representing the energy density of incoming gravitational radiation.[12] Thus we may interpret $\tilde{M}$, the limit of $M$ as $r \to \infty$, as the total mass of the system. In asymptotically flat, empty space, $\tilde{M}$ coincides with the definition of mass given by Bondi.[1] Moving each point of $S$ an equal distance $u$ along the null vectors $n^\mu$, one gets the conservation law

$$\frac{d\tilde{M}}{du} = -(4\pi)^{-1} \lim_{r\to\infty} \int \lambda\bar{\lambda}\, dS. \tag{8.4}$$

This shows that $\tilde{M}$ monotonically decreases at a rate given by the news function $\lambda$.

In our case $\tilde{M}$ will be infinite since a sphere of infinite radius will enclose an infinite amount of matter. We may divide into $\tilde{M}_1$ which is given by the Ricci-tensor terms $\Phi_{11}$ and $\Lambda$, and $\tilde{M}_2$ which is given by the terms $\Psi_2$ and $\sigma\lambda$. In the undisturbed metric, $\tilde{M}_1$ is infinite and $\tilde{M}_2$ is zero. We may regard $\tilde{M}_2$ as representing the mass of the source plus the mass of the disturbance. It is given by

$$\tilde{M}_2 = -(8\pi A)^{-1} e^{-u} \int P^{-2}(\Psi_2^0 + \sigma^0\lambda^0)\, dx^3\, dx^4. \tag{8.5}$$

It obeys the conservation law (8.4):

$$d\tilde{M}_2/du = -(4\pi A)^{-1} e^{-u} \int P^{-2}\lambda^0\bar{\lambda}^0\, dx^3\, dx^4. \tag{8.6}$$

We may generalize this to obtain a conservation law for energy and momentum.[13] Define $\mathcal{P}^{(\alpha)}$ to be the right-hand side of Eq. (8.5) where the integrand has been multiplied by a weighting function $W^{(\alpha)}(x^i)$ which satisfies

$$\nabla(P^2 \nabla W^{(\alpha)}) = 0. \tag{8.7}$$

Then,

$$\frac{d\mathcal{P}^{(\alpha)}}{du} = -(4\pi A)^{-1} e^{-u} \int P^{-2}\lambda^0\bar{\lambda}^0 W^{(\alpha)}\, dx^3\, dx^4. \tag{8.8}$$

There are four independent solutions of (7.7). They may be chosen as

$$W^{(0)} = 1, \quad W^{(1)} = \frac{\zeta + \bar{\zeta}}{2(1 + \frac{1}{4}\zeta\bar{\zeta})},$$
$$W^{(2)} = \frac{\zeta - \bar{\zeta}}{2i(1 + \frac{1}{4}\zeta\bar{\zeta})}, \quad W^{(3)} = \frac{1 - \frac{1}{4}\zeta\bar{\zeta}}{1 + \frac{1}{4}\zeta\bar{\zeta}}. \tag{8.9}$$

[12] R. Penrose, in *Perspectives in Geometry and Relativity*, B. Hoffman, Ed. (Indiana University Press, Bloomington, Indiana, 1966), pp. 259–274.

[13] R. Penrose, Phys. Rev. Letters **10**, 66 (1963).

We may regard the $\mathcal{P}^{(\alpha)}$ as the components of the total energy–momentum vector of the source and the disturbance.

## 9. OBSERVATION

An observer moving with the matter would find that the frequency was redshifted by a factor $\Omega$. The redshift would also reduce the apparent intensity in his local frame. By measuring the relative accelerations of neighboring particles, he may determine $R_{\alpha\beta\gamma\delta}V^\alpha V^\gamma$, where $V^\alpha$ is his velocity vector. The trace-free part of this represents the "electric" components of the gravitational radiation:

$$E_{\beta\delta} = C_{\alpha\beta\gamma\delta}V^\alpha V^\gamma.$$

In the observer's orthonormal tetrad $e_a^\alpha$ $(a = 1, 2, 3, 4)$ in which $e_1^\alpha = V^\alpha$, $e_2^\alpha = (V^\beta l_\beta)^{-1} l^\alpha - V^\alpha$, $e_3^\alpha = (2)^{-\frac{1}{2}}(m^\alpha + \bar{m}^\alpha)$, and $e_4^\alpha = (2)^{-\frac{1}{2}} i(\bar{m}^\alpha - m^\alpha)$,

$$E_{ab} = \begin{vmatrix} 0 & 0 & 0 & 0 \\ 0 & -2G & E & F \\ 0 & E & G - C & D \\ 0 & F & D & G + C \end{vmatrix}, \tag{9.1}$$

where

$$C + iD = \Psi_4^0\Omega^{-4} + O(\Omega^{-5}),$$
$$E + iF = -(2)^{\frac{1}{2}}\Psi_3^0\Omega^{-5} + O(\Omega^{-6}), \tag{9.2}$$
$$G = 2(\Psi_2^0 + \bar{\Psi}_2^0)\Omega^{-6} + O(\Omega^{-7}).$$

The dominant $\Omega^{-4}$ terms represent the two polarizations of a spin-2 wave. The effect of the redshift is to compress the apparent peeling-off behavior. The $\Psi_1$ and $\Psi_0$ components give, respectively, an $\Omega^{-6}$ term of type similar to that given by $\Psi_3$ and an $\Omega^{-5}$ term similar to $\Psi_4$.

The asymptotic expansions used should be reasonable approximations provided that $\Omega$ is large compared to $A$ and to the value of $\Omega$ at the source. The first condition is satisfied at the present epoch if the average density of the universe is just that of the observed luminous matter (about $10^{-31}$ g/cc). The second is satisfied for sources of large redshift ($z \gg 1$). For nearer sources the behavior approximates that of asymptotically flat space.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Perturbation Expansion for Real-Time Green's Functions*

R. A. CRAIG†
*Department of Physics, University of Oregon, Eugene, Oregon*

(Received 16 November 1966; final manuscript received 14 July 1967)

The development of the time-translation operators in a matrix element of an arbitrary operator is examined. It is noted that we may interpret time as evolving from some remotely early time ($t_0$) to a time in the far future ($t_\infty$) and then back to ($t_0$). Using this interpretation, a perturbation expansion is developed for Green's functions defined along this path and a separation of the two-particle interaction terms into self-energy parts and single-particle Green's function terms is justified for quantities on this path. A connection is established between the real-time Green's functions and the Green's function defined along the path, thereby yielding a perturbation expansion for the real-time functions and a justification of the separation of the interaction terms in the equations of motion for the real-time quantities. The transport equations of Kadanoff and Baym are derived without resorting to an analytic continuation from imaginary times and without the correction terms of Fujita.

## 1. INTRODUCTION

In order to consider transport properties of many-body systems from a microscopic point of view, it is appropriate to work with Green's functions defined for real times. Kadanoff and Baym,[1] using an analytic continuation procedure of Baym and Mermin,[2] have derived equations of motion for the real-time functions. This approach suffers from certain weaknesses. It is not altogether clear that the separation of the two-particle correlation function into real-time single-particle functions and self-energy terms is in fact justified for this situation. In addition, it is difficult to derive expressions for the self-energy functions directly without recourse to the analytic continuation procedure outlined in Kadanoff and Baym. A method for writing down the general expressions for such functions can be derived,[3] but is itself not convenient for investigating the exact functional dependence of such quantities on the interparticle potential and the single-particle Green's functions.

Fujita,[4] using a diagram technique, has investigated systems in the absence of external fields and developed equations similar to those of Kadanoff and Baym but with correction terms. These corrections involve initial-particle correlations which Fujita claims die out in a time of order the collision time of the particles. Clearly this cannot be the entire story, since if, for instance, the system is initially in a single quantum state, the information of initial-particle correlations will be preserved, in principal, indefinitely. Indeed such long-lived correlations are observed in spin-echo experiments in which the system initially is describable not by a single quantum state but rather by an appropriate nonequilibrium ensemble.

The purpose of this work is to remedy the weaknesses of the approach of Kadanoff and Baym, and to include the possibility of initial-particle correlations.[5] A formalism is developed in which it is possible to demonstrate the separation of the potential-energy term in the equation of motion of the real-time Green's functions into self-energy and single-particle functions. In addition, this approach gives a method for writing down these self-energy functions directly.

The formalism derived in this work is similar to that developed by Mills[6] for equilibrium situations in that the final equations are matrix functionals of the time-ordered, anti-time-ordered, advanced and retarded, real-time Green's functions. However, here only real (physical) times are considered; no analytic continuation from imaginary times is needed. We consider the system of fermions interacting instantaneously among themselves—the generalization to include other systems is obvious.

The approach used here is to work directly with the real-time functions and their equations of motion. By considering the behavior of the time-development operator, it is shown that it is appropriate to consider

* Work supported by the National Science Foundation.

† Present address: Department of Physics, University of California, Riverside, Calif.

[1] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (W. A. Benjamin, Inc, New York, 1962).

[2] G. Baym and D. Mermin, J. Math. Phys. **2**, 232 (1961).

[3] R. A. Craig, Ann. Phys. (N.Y.) **40**, 416 (1966).

[4] S. Fujita, J. Math. Phys. **6**, 1877 (1965).

[5] After the completion of this work, the author's attention was brought to the paper of L. V. Keldysh {Zh. Eksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys.—JETP **20**, 1018 (1965)]} in which some of the formalism developed here was derived independently. Keldysch ignores the possibility of initial-particle correlations and prepares the system with an equilibrium-density matrix for infinitely early times so that his work is subject to the same criticism as that of Kadanoff and Baym. In addition Keldysh, in doing his perturbation expansion, makes the rather inconvenient choice of time $t = 0$ for the operators to reduce to those in the absence of interparticle interactions.

[6] R. Mills (to be published). An equilibrium approach similar to that of Mills has been used [V. Korenman, Ann. Phys. (N.Y.) **39**, 72 (1966): Phys. Rev. **154**, 1233 (1967)] to discuss the laser-photon system.

the system as evolving along a path in time which runs from the time $t = -\infty$ to a time $t = \infty$ and then back to $t = -\infty$. This is done in Sec. 2. Such a path lends itself naturally to a matrix description in which the matrix components are related to real-time functions. In this description, a perturbation expansion is developed for the propagator. In Sec. 3, the usual arguments are applied to generate a Dyson equation for the matrix propagator, in this way justifying the separation of the two-particle correlation terms in the equation of motion for the single-particle Green's function into products of Green's functions and self-energy parts and incidently giving a perturbation expansion for the self-energy functions.

## 2. JUSTIFICATION OF PATH APPROACH

Before considering the specific case of the real-time Green's function, we consider the example of a matrix element of an arbitrary operator $\Theta$:

$$\Theta_{mn}(t) = \langle m, t| \Theta |n, t\rangle, \tag{2.1}$$

where the states $|m, t\rangle$, $|n, t\rangle$ are in the Schrödinger picture and obey the equations of motion

$$i\frac{\partial}{\partial t}|m, t\rangle = H_s(t)|m, t\rangle. \tag{2.2}$$

Equation (2.2) may be solved (in a formal sense) in terms of the time-translation operator[7]

$$|m, t\rangle = \mathfrak{U}(t, t_0)|m, t_0\rangle, \tag{2.3a}$$

where

$$i\frac{\partial}{\partial t}\mathfrak{U}(t, t_0) = H_s(t)\mathfrak{U}(t, t_0). \tag{2.3b}$$

If we choose $t_0$ as some remotely early time before which all external fields vanish (eventually we choose $t_0 = -\infty$),[8] then the states are those of a field-free system. We have

$$\Theta_{mn}(t) = \langle m, t_0| \mathfrak{U}(t_0, t)\Theta\mathfrak{U}(t, t_0)|n, t_0\rangle. \tag{2.4}$$

The time translations in Eq. (2.4) can be interpreted in the following way: The states $|m, t_0\rangle$, $|n, t_0\rangle$ are prepared at $(t_0)$ and allowed to develop under the full Hamiltonian. At time $(t)$, the expectation value of the operator is taken between those states which have developed from $|m, t_0\rangle$ and $|n, t_0\rangle$.

An alternative interpretation of the time translations in Eq. (2.4) can be made. We allow the system, initially in the state $|n, t_0\rangle$, to develop to $(t)$; at time $(t)$ we operate on the system with the operator $\Theta$ and then we allow time to develop backwards, returning to $(t_0)$. At $(t_0)$ the projection of the resulting state on $|m, t_0\rangle$ is measured. The former interpretation is more meaningful physically; the latter provides the concept on which this work is based.

Because of the group property of the time-development operators,[7] Eq. (2.4) may be rewritten

$$\Theta_{mn}(t) = \langle m, t_0| \mathfrak{U}(t_0, t_\alpha)\mathfrak{U}(t_\alpha, t) \times \Theta\mathfrak{U}(t, t_0)|n, t_0\rangle, \tag{2.5a}$$

where we can take $(t_\alpha)$ as some latest time after which all external fields vanish (eventually $t_\alpha$ can go to $\infty$). If the matrix element which we are considering is a diagonal element, and if the state is nondegenerate, then by inserting a complete set of states and using the fact that an interaction switched on adiabatically can cause no transitions between levels, we get

$$\begin{aligned}\Theta_{mm}(t) &= \langle m, t_0| \mathfrak{U}(t_0, t_\alpha)|m, t_0\rangle \times \langle m, t_0| \mathfrak{U}(t_\alpha, t)\Theta\mathfrak{U}(t, t_0)|m, t_0\rangle \\ &= e^{i\phi}\langle m, t_0| \mathfrak{U}(t_\alpha, t)\Theta\mathfrak{U}(t, t_0)|m, t_0\rangle \\ &= \frac{\langle m, t_0| \mathfrak{U}(t_\alpha, t)\Theta\mathfrak{U}(t, t_0)|m, t_0\rangle}{\langle m, t_0| \mathfrak{U}(t_\alpha, t_0)|m, t_0\rangle},\end{aligned} \tag{2.6}$$

for adiabatically slowly varying external fields. It is by this reasoning that, in the usual zero-temperature theory, time evolution from $t = -\infty$ to $t = +\infty$ may be considered.[9]

However, in general, we consider states which may be degenerate with others, and external fields which may cause transitions between levels, so that the above arguments are inappropriate. We are therefore constrained to a temporal evolution which runs from $(t_0)$ to $(t_\alpha)$ and then back to $(t_0)$.

The choice of where $(t)$ comes in Eq. (2.5a) is not unique; alternatively we could have written

$$\Theta_{mn}(t) = \langle m, t_0| \mathfrak{U}(t_0, t)\Theta\mathfrak{U}(t, t_\alpha)\mathfrak{U}(t_\alpha, t_0)|n, t_0\rangle. \tag{2.5b}$$

Equation (2.5b) describes a situation in which the state $|n, t_0\rangle$ is allowed to develop from $(t_0)$ to $(t_\alpha)$ and then back to $(t)$; at this time, the operator $\Theta$ acts and the state is allowed to develop to $(t_0)$. At $(t_0)$ the projection of the resulting state on the state $|m, t_0\rangle$ is taken.

[7] See, for instance, S. S. Schweber, *Relativistic Quantum Field Theory* (Row, Peterson and Company, Evanston, Illinois, 1961), pp. 316ff.

[8] This limit does *not* restrict us to the situation in which the external field is turned on in the very remote past, but only that the external field be turned on after $t_0 = -\infty$.

[9] There exist serious questions about the validity of using the usual zero-temperature theory for physical problems. These involve the order of taking the zero temperature and the thermodynamic (volume $\to \infty$) limits. For correct results, the volume should go to infinity first, in which case the near degeneracy of states close to the ground state becomes important. The method used here, it should be noted, can be used when the states involved are degenerate so that all such difficulties are avoided.

The two situations described by (2.5a) and (2.5b) are, of course, identical. However, if we suppose that time development in the forward time direction is governed by different laws than time development in the reverse direction, then the situations are no longer the same.[10] The matrix elements $\Theta_{mn}(t)$ depend on which direction the time $(t)$ is evolving. One method of generating a dynamics, which depends on the direction of temporal development, is to introduce an external potential which possesses this property. In this case, for instance, the equation of motion of a field operator whose Hamiltonian in the absence of external fields is

$$H = \int d^3\mathbf{r}\, \frac{\nabla\psi^\dagger(\mathbf{r})\cdot\nabla\psi(\mathbf{r})}{2m} + \frac{1}{2}\int d^3\mathbf{r}\, d^3\mathbf{r}' V(\mathbf{r},\mathbf{r}')\psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r}) \quad (2.7)$$

would be, upon introducing the external field,

$$i\frac{\partial}{\partial t}\psi(\mathbf{r},t)\Big|_\pm = -\frac{\nabla^2}{2m}\psi(\mathbf{r},t) + \int d^3\mathbf{r}' V(\mathbf{r},\mathbf{r}')\psi^\dagger(\mathbf{r}',t)\psi(\mathbf{r}',t)\psi(\mathbf{r},t) + \int d^3\mathbf{r}' U_\pm(\mathbf{r},\mathbf{r}')\psi(\mathbf{r}'), \quad (2.8)$$

where the $\pm$ refers to the direction of temporal development. It should be emphasized that introducing external fields which differ for time development in the forward and reverse time directions is a bookkeeping artifice. We shall always be concerned only with the physical situation when the fields are the same.

## 3. PERTURBATION EXPANSION OF GREEN'S FUNCTIONS

Before proceeding to develop the ideas of the preceding section further, we wish first to make some observations on actual experimental arrangements. Before applying the external field to produce the nonequilibrium conditions, we start with a system which is not, in general, in a definite eigenstate. Instead the system is in thermal equilibrium with its surroundings at a time which we shall call $t_\beta$. This system is then isolated (in principle) from the outside world and the experiments performed. Alternatively, we prepare the system in some nonequilibrium situation, isolate the system, and perform the experiments. In either case, the statistical expectation value of an operator at some later time is given by

$$\langle\Theta(t)\rangle = \sum_{i,j}\rho_{ij}\langle j, t_\beta|\,\Theta(t)\,|i, t_\beta\rangle. \quad (3.1)$$

[10] J. Schwinger, J. Math. Phys. **2**, 407 (1961).

Here $\rho_{ij}$ are the matrix elements of the initial-density matrix in the representation of eigenstates of the Hamiltonian in the absence of the external field $|i, t_\beta\rangle$. The operator $\theta(t)$ and the *time-independent* states $|i, t_\beta\rangle$ are defined in the Heisenberg representation. Fujita expands the density matrix in states of the noninteracting system. The correction terms he describes have their origin in cross terms between this expansion and the perturbation expansion of the time-development operators. As we see, expressing the initial-density matrix of the system in terms of the exact eigenstates of the undisturbed system preserves initial-particle correlations without Fujita's added terms.

We are interested in the single-particle Green's function

$$ig(1,2) = \langle T\psi(1)\psi^\dagger(2)\rangle = \sum_{i,j}\rho_{ij}\langle j, t_\beta|\,T\,\psi(1)\psi^\dagger(2)\,|i, t_\beta\rangle. \quad (3.2)$$

Here $T$ denotes the conventional time-ordered product in which earlier times are ordered to the right with the appropriate change of sign for permutations of fermions. When we use field operators which are sensitive to the direction of temporal development, as are those defined by Eq. (2.8), we must distinguish between those times which are developing in the forward and those which are developing in the reverse sense. A more fruitful approach in this case is to introduce the quantity $G(1,2)$ defined by

$$iG(1,2) = \langle P\psi(1)\psi^\dagger(2)\rangle, \quad (3.3)$$

where $P$ denotes an ordering of the operators along the temporal path which runs from $(t_0)$ to $(t_\alpha)$ and back to $(t_0)$, and includes a minus sign for each permutation of fermion operators necessary to achieve the proper ordering. The times $t_0(t_\alpha)$ now represent times before (after) which the interparticle interactions vanish. The function $G(1,2)$ contains, as we shall see, all the information on $g(1,2)$, plus additional information on the analytic parts of $g(1,2)$ and $\bar{g}(1,2)$ the anti-time-ordered function. By including the additional information in $G(1,2)$ it is possible to separate the equations of motion for $g(1,2)$ and develop closed forms for the self-energies. Let us consider, for example, the situation when $t_1$ is developing in the forward direction and $t_2$ in the backward. We are then concerned with matrix elements like

$$\langle j, t_\beta|\,\psi^\dagger(2,^-)\psi(1,^+)\,|i, t_\beta\rangle \quad (3.4)$$

which appear in the statistical average. In order to develop a perturbation expansion for quantities of this form, we separate the Hamiltonian into a part

appropriate to a noninteracting system $H_0 + H_{ext}$ and a part containing the interparticle interaction ($H_{int}$). Having made this separation, it is now appropriate to introduce the interaction representation in which operators develop according to the Hamiltonian in which interparticle interactions are neglected

$$i\frac{\partial \Theta_I}{\partial t} = [\Theta_I, (H_0 + H_{ext})_I], \tag{3.5}$$

for an arbitrary operator $\Theta$. In this representation, the matrix elements (3.4) can be written[11]

$$\langle j, t_\beta | \mathcal{U}(t_\beta, t_2)\psi_I^\dagger(2)\mathcal{U}(t_2, t_\alpha)\mathcal{U}(t_\alpha, t_1)\psi_I(1) \times \mathcal{U}(t_1, t_\beta)|i, t_\beta\rangle = \langle j_0 t_0| \mathcal{U}(t_0, t_2)\psi_I^\dagger\mathcal{U}(t_2, t_\alpha)\mathcal{U}(t_\alpha, t_1)\psi_I(1)(t_1, t_0)|i, t_0\rangle, \tag{3.6}$$

where

$$i\frac{\partial}{\partial t_1}\mathcal{U}(t_1, t_2) = [H_{int}(t_1)]_I\mathcal{U}(t_1, t_2) \tag{3.7a}$$

and

$$-i\frac{\partial}{\partial t_2}\mathcal{U}(t_1, t_2) = \mathcal{U}(t_1, t_2)[H_{int}(t_2)]_I. \tag{3.7b}$$

Here $\mathcal{U}(t_1, t_2)$ is a time-evolution operator in the interaction picture. We can rewrite the differential equations for $\mathcal{U}(t_1, t_2)$ in the integral form

$$\mathcal{U}(t_1, t_2) = 1 - i\int_{t_2}^{t_1} dt'[H_{int}(t')]_I\mathcal{U}(t', t_2). \tag{3.8}$$

By iterating this equation, we can get, when $t_1 > t_2$,

$$\begin{aligned}\mathcal{U}(t_1, t_2) = 1 &- i\int_{t_2}^{t_1}[H_{int}(t')]\,dt' \\ &+ (-i)^2\int_{t_2}^{t_1}dt'\int_{t_2}^{t'}dt''[H_{int}(t')H_{int}(t'')] \\ &+ (-i)^3\int_{t_2}^{t_1}dt'\int_{t_2}^{t'}dt''\int_{t_2}^{t''}dt'''[H_{int}(t')H_{int}(t'')H_{int}(t''')] \\ &+ \cdots,\end{aligned} \tag{3.9}$$

which by the usual arguments can be written in the time-ordered form[7]

$$\begin{aligned}\mathcal{U}(t_1, t_2) &= 1 - i\int_{t_2}^{t_1}[H_{int}(t')]\,dt' \\ &+ \frac{(-i)^2}{2!}\int_{t_2}^{t_1}dt'\,dt''T\{H_{int}(t')H_{int}(t'')\} \\ &+ \frac{(-i)^3}{3!}\int_{t_2}^{t_1}dt'\,dt''\,dt'''T\{H_{int}(t')H_{int}(t'')H_{int}(t''')\} \\ &= T\left\{\exp\left[-i\int_{t_2}^{t_1}H_{int}(t')\,dt'\right]\right\}.\end{aligned} \tag{3.10a}$$

[11] Since $t_0$ is a time before which the interparticle interactions vanish, the states $|i, t_0\rangle$ are eigenstates of the noninteracting system.

On the other hand, when $t_2$ is greater than $t_1$, by similar arguments we get

$$\mathcal{U}(t_1, t_2) = T^\dagger\left\{\exp\left[-i\int_{t_2}^{t_1}H_{int}(t')\,dt'\right]\right\}, \tag{3.10b}$$

where $T^\dagger$ is the anti-time-ordering operator which orders operators in the inverse order from $T$. But we note the operation of path ordering ($P$) is just time ordering ($T$) when the times are on the forward leg of the path and anti-time ordering ($T^\dagger$) when on the return leg. Thus it is clear that, since $t_2$ is further along the path than $t_1$ in the expression (3.4), we can write this

$$\langle j, t_0| P\left(\left\{\exp\left[-i\int_{P}^{t_0}H_{int}(s')\,ds'\right]\right\} \times \psi_I(1)\psi_I^\dagger(2)|i, t_0\rangle\right|_{\substack{t_1 \text{on}(+)\\ t_2 \text{on}(-)}}, \tag{3.11}$$

in which the notation $\int_{t_0}^{t_0} ds$ indicates an integration along the path from ($t_0$) to ($t_\alpha$) and back to ($t_0$). The other possible orderings of operators in $G(1, 2)$ give similar results so that we are able to write

$$iG(1, 2) = \left\langle P\left\{\left[\exp\left(-i\int_{P}^{t_0}ds\,H_{int}(s)\right)\right]\psi_I(1)\psi_I^\dagger(2)\right\}\right\rangle_0, \tag{3.12a}$$

where

$$\langle\cdots\rangle_0 = \sum_{i,j}\rho_{ij}\langle j, t_0|\cdots|i, t_0\rangle. \tag{3.12b}$$

The perturbation expansion for $G(1, 2)$ can now be developed in the usual way; the algebra is identical with that of the zero-temperature perturbation theory, only the time-integration path differing. We thus shall not dwell on the details of this development, as all the usual combinatorial arguments apply. The decomposition of time-ordered products equivalent to Wick's theorem follows by the same arguments as applied to the finite-temperature equilibrium theory,[12] so that an expansion of $G(1, 2)$ in powers of the interparticle potential is, except for the unusual temporal treatment, formally identical with the more familiar cases.[13] In the usual way, diagrams with disconnected parts contribute a multiplicative factor $\langle\mathcal{U}(t_0^+, t_0^-)\rangle_0$ to those diagrams without disconnected parts. In the physical limit that temporal development is identical

[12] For a justification of the separation of the noninteracting, many-particle Green's function into single-particle Green's functions by a method applicable here, see V. Ambegaokar, "Astrophysics and the Many Body Problem," in *1962 Brandeis Lectures, Volume 2*, K. W. Ford, Ed. (W. A. Benjamin, Inc., New York, 1963).

[13] A. A. Abrikosov, L. P. Gorkov, I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, translated by Richard A. Silverman (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1963), pp. 68–73.

in the forward and reverse directions, this factor is $\langle \mathcal{U}(t_0^+, t_0^-)\rangle_0 = 1$. It should be observed that the $G_0(1, 2)$, the Green's functions in the absence of the interparticle interaction

$$iG_0(1, 2) = \langle P\psi(1)\psi^\dagger(2)\rangle_0 = \sum_{i,j} \rho_{ij} \langle j, t_0| P\psi(1)\psi^\dagger(2) |i, t_0\rangle \quad (3.13)$$

contain initial-particle correlations through $\rho_{ij}$. Fujita's difficulty of correction terms to the equations of Kadanoff and Baym is replaced, in this case, by the fact that we cannot write down exact expressions for $G_0(1, 2)$. This is only an apparent difficulty however, as we never ask for an analytic form of $G_0(1, 2)$.

We have, therefore, the following rules for constructing an $n$th order contribution:

(a) Form all connected, topologically distinct diagrams with $2n$ vertices and two external points (1) and (2) in which two solid directed lines (one leaving and one entering) and one dotted line meet at each vertex.

(b) With each solid line between points (1′) and (2′) associate a factor $G_0(1', 2')$ where (2′) is the initial point and (1′) the final point of the line.

(c) With each dotted line between the points (1′) and (2′) associate a factor

$$-iV(r_{1'2'})\delta(s_{1'} - s_{2'}).$$

(d) Integrate over all free variables,

$$\int_P^{t_0} ds_i \int d^3\mathbf{x}_i.$$

(e) For fermions only, introduce an over-all factor $(-1)^l$, where $l$ is the number of closed loops in the diagram.

(f) Whenever $G_0(\mathbf{r}, s; \mathbf{r}', s)$ appears, it is to be interpreted as $G_0(\mathbf{r}, s; \mathbf{r}', s^+)$.

We now look at the structure of the diagrams contributing to $G(1, 2)$. Since every term in the expansion, with the exception of the term involving no particle interactions, begins and ends with a bare particle line, it follows that we can write

$$G(1, 2) = G_0(1, 2) + \int d3\, d4\, G_0(1, 3)\Sigma_R(3, 4)G_0(4, 2). \quad (3.14)$$

Here $\Sigma_R(3, 4)$, which plays the role of a reducible self-energy, is the collection of diagrams contributing to $G(1, 2)$ as determined by the above rules but with $G_0(1, 3)$ and $G_0(4, 2)$ deleted. Individual diagrams contributing to $\Sigma_R(3, 4)$ may be classified either as reducible or irreducible. A reducible self-energy diagram is one which, when a single-particle line is removed, reduces to another self-energy diagram. In the standard way,[14] observation of the structure of these diagrams leads to writing Eq. (3.14) in terms of the irreducible self-energy $\Sigma(3, 4)$, getting

$$G(1, 2) = G_0(1, 2) + \int d3\, d4\, G_0(1, 3)\Sigma(3, 4)G(3, 2). \quad (3.15)$$

In Eq. (3.15), $\Sigma(3, 4)$ is the sum of all diagrams which cannot be reduced to two simpler self-energy diagrams by removing one particle line. The particle lines in $\Sigma(1, 2)$ are exact propagators $G(1, 2)$ so that (3.15) is a self-consistent integral equation; as discussed in connection with Eq. (3.13), initial-particle correlations are contained in Eq. (3.15) via the dependence on $G_0(1, 2)$.

The rules for writing down a contribution to $\Sigma(1, 2)$ are similar to those for $iG(1, 2)$. All irreducible diagrams with free vertices at (1) and (2) are drawn; except for an over-all factor of $(i)$ in the case of $\Sigma(1, 2)$ and associating with each solid line a real propagator $iG(1, 2)$, contributions from such a diagram are determined in the same way as for $iG(1, 2)$.

At this point we note that it is possible to characterize the position on the path by two parameters, the time variable and the direction in which time is evolving. We write $\psi(t_1)$ for $t_1$ increasing as $\psi_+(t_1)$ and for $t_1$ decreasing as $\psi_-(t_1)$ with an identical notation for other functions of the path variable. The path-ordered function can therefore be written in the matrix form

$$iG(1, 2)_{12} = \langle P\psi_1(t_1)\psi_2^\dagger(t_2)\rangle, \quad (3.16)$$

where now the path-ordering operator orders all operators with a $(+)$ matrix index to the right of all those with a $(-)$ matrix index. Among the $(+)$ operators, operators with earliest time coordinates are ordered to the right (time-ordered); among the $(-)$ operators, those with the earliest time coordinates are ordered to the left (anti-time ordered). An over-all minus sign is introduced for each permutation of fermion operators necessary to achieve the proper ordering.

Equation (3.15) may be written in the form

$$\int d2G^{-1}(1, 2)G(2, 3) = \delta(\mathbf{r}_1 - \mathbf{r}_3)\delta(s_1 - s_3), \quad (3.17)$$

[14] Reference 13, pp. 85–87. It is to be emphasized that, although the discussions of Ref. 13 are for the special case of a system at zero temperature, the arguments are topological in nature and are equally applicable here.

where

$$G^{-1}(1,2) = G_0^{-1}(1,2) - \Sigma(1,2) \tag{3.18}$$

in which $G_0^{-1}(1,2)$ is the operational inverse of $G_0(1,2)$. When we use the matrix designations of path position, (3.17) becomes

$$\int d^3\mathbf{r}_2 \Big\{ \int_{t_0}^{t_\alpha} dt_2 G^{-1}(1,2)_{1+} G(2,3)_{+3} + \int_{t_\alpha}^{t_0} dt_2 G^{-1}(1,2)_{1-} G(2,3)_{-3} \Big\} = \delta(\mathbf{r}_1 - \mathbf{r}_3)\delta(s_1 - s_3). \tag{3.19}$$

In view of the fact that the transcription of the path-coordinate delta function to the matrix description is given by

$$\delta(s_1 - s_2) = \tau_{12}^3 \delta(t_1 - t_2), \tag{3.20}$$

which may be verified by considering the integral $\int_P ds_1 \delta(s_1 - s_2) = 1$, Eq. (3.19) becomes

$$\int d^3\mathbf{r}_2 \int_{t_0}^{t_\alpha} dt_2 [G^{-1}(1,2)\tau^3 G(2,3)]_{13} = \delta(t_1 - t_3)\delta(\mathbf{r}_1 - \mathbf{r}_3)\tau_{13}^3. \tag{3.21}$$

Here $\tau^3$ is the Pauli-spin matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

If now we define the quantity

$$\tilde{G}(1,2)_{12} = \sum_4 G(1,2)_{14}\tau_{42}^3 \tag{3.22a}$$

and its inverse

$$\tilde{G}^{-1}(1,2)_{12} = \sum_4 G^{-1}(1,2)_{14}\tau_{42}^3, \tag{3.22b}$$

then (3.21) becomes

$$\int d^3\mathbf{r}_2 \int_{t_0}^{t_\alpha} dt_2 [\tilde{G}^{-1}(1,2)\tilde{G}(2,3)]_{13} = \delta(\mathbf{r}_1 - \mathbf{r}_3)\delta(t_1 - t_3)\delta_{13}. \tag{3.23}$$

It would appear at this point that all traces of initial-particle correlations have been eliminated. This is, in fact, not true. The solution to the self-consistent equation (3.15) contains initial-particle correlations through $G_0(1,2)$. When Eq. (3.15) is transformed into a differential equation (3.23) through the application of $G_0^{-1}(1,2)$, then the initial-particle correlations are contained in the boundary conditions. Equation (3.23) is, therefore, a real-time Dyson equation, which contains implicitly through the boundary conditions all the initial-particle correlations.

We now observe that, in the physical limit that temporal development on the (+) and (−) paths is the same, the matrix $\tilde{G}(1,2)$ can be written

$$\tilde{G}(1,2) = \begin{pmatrix} g(1,2) & -g^<(1,2) \\ g^>(1,2) & -\bar{g}(1,2) \end{pmatrix}, \tag{3.24}$$

where $\bar{g}(1,2)$ is the anti-time-ordered Green's function. Therefore, the equation of motion for $\tilde{G}(1,2)$, Eq. (3.23), contains all the information on the equations of motion of $g(1,2)$ and $\bar{g}(1,2)$ [in addition to redundant information on $g^>(1,2)$ and $g^<(1,2)$]. If now we identify $\Sigma^>(1,2)$ and $\Sigma^<(1,2)$ (this may be regarded as a definition of these quantities) via the equations

$$\tilde{\Sigma}(1,2) = \Sigma(1,2)\tau^3 = \begin{pmatrix} [\Sigma^>(1,2)\theta(t_{12}) + \Sigma^<(1,2)\theta(t_{21})] - \Sigma^<(1,2) \\ \Sigma^>(1,2) - [\Sigma^<(1,2)\theta(t_{12}) + \Sigma^>(1,2)\theta(t_{21})] \end{pmatrix}, \tag{3.25}$$

where $\theta(t)$ is the unit step function

$$\theta(t) = \begin{cases} 1 & t > 0, \\ 0 & t < 0, \end{cases}$$

then the equations of motion (3.23) are identical with those of Kadanoff and Baym for $g^>(1,2)$ and $g^<(1,2)$ in terms of $\Sigma^>(1,2)$ and $\Sigma^<(1,2)$.

We have, therefore, justification of the separation of the interaction terms in the equation of motion for the single-particle Green's function into self-energy and single-particle Green's function terms. In addition, we have a method for determining these self-energy functions directly through $\tilde{\Sigma}(1,2)$. To do so, we simply transcribe the rules for the path-ordered self-energy $\Sigma(1,2)$ to the matrix description. The rules for calculating the contribution to $\tilde{\Sigma}(1,2)_{ij}$ are:

(a) Construct all connected, irreducible, primitive diagrams with $2n$ vertices and directed solid lines connecting all vertices except $(1,i)$ for which there is no line leaving, and $(2,j)$ for which there is no line entering. Each vertex is characterized by a space–time coordinate and a matrix variable.

(b) For each solid line from $(3,k)$ to $(4,l)$ a factor

$$i\tilde{G}(3,4)_{kl}.$$

(c) For each dashed interaction, a factor

$$-iU(\mathbf{r}_{34})\delta(t_3 - t_4)\tau_{34}^3.$$

(d) Sum over all free matrix variables, integrate over all free space–time coordinates.

(e) Wherever the quantity $\tilde{G}(\mathbf{r}_3, t_3; \mathbf{r}_4, t_3)$ appears, interpret it as $\tilde{G}(\mathbf{r}_3, t_3; r_4, t_3^+)$.

(f) Introduce an over-all factor $i(-1)^l$ ($i$ for bosons) where $(l)$ is the number of closed loops in the diagram.

Some justification of these rules is in order. Consider a vertex with the coordinates along the path $(\mathbf{r}_3, s_3)$. If we integrate over these coordinates, there is a

factor $(+1)\int_{-\infty}^{\infty} dt_3$ when the time is evolving in the forward direction and a factor $(-1)\int_{-\infty}^{\infty} dt_3$ when time evolves in the backward direction. We can remove these factors by appending a $\tau^3$ to the outgoing line at this vertex, thereby turning $G(3, 4)_{ij}$ into $\tilde{G}(3, 4)_{ij}$. If we replace all $G(3, 4)_{ij}$ by $\tilde{G}(3, 4)_{ij}$ however, one vertex over which no path integration is performed gets a factor of $\tau^3$. But this is precisely the factor in the precise location needed to turn $\Sigma(1, 2)_{ij}$ into $\tilde{\Sigma}(1, 2)_{ij}$. Hence our rules indeed give us $\tilde{\Sigma}(1, 2)_{ij}$.

## 4. DISCUSSION

In the preceding we have shown that, by retaining redundant information on $g^{>}(1, 2)$ and $g^{<}(1, 2)$, the equations of motion for these quantities can be written in a form in which the integrals run over all space and time. Using this approach, and by retaining initial-particle correlations in the unperturbed Green's functions, we have shown that the separation of the interaction terms into self-energy and single-particle Green's function terms is justified without the need for the correction terms of Fujita. In addition, we have devised a perturbation expansion to determine the self-energy functions $\Sigma^{>}(1, 2)$ and $\Sigma^{<}(1, 2)$ appearing in the theory of Kadanoff and Baym. Since both space and time integrations run over all values, the explicit functional dependence of the self-energy functions upon the single-particle Green's function $g^{>}(1', 2')$ and $g^{<}(1', 2')$ may be determined directly via the dependence of $\tilde{\Sigma}_{ij}(1, 2)$ on $\tilde{G}(1', 2')_{i'j'}$.

## ACKNOWLEDGMENTS

The author wishes to thank Professor Joel W. McClure for his kind hospitality during the course of this work. He also wishes to thank Professor Leo P. Kadanoff for the comments which led to this research and Professor Marvin D. Girardeau for a critical reading of the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Extension Theorem and Representations*

A. SANKARANARAYANAN

*Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa*†

(Received 23 February 1967)

An extension theorem prescribing a method for constructing the $(n + 1)$-dimensional Lorentz group $O(n, 1)$ from the $n$-dimensional inhomogeneous rotation group, proved earlier, has been recapitulated and some comments on the theorem are made. The prescription is used to construct the unitary irreducible representations of the Lorentz groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$, and it is found that only a limited number of representations are allowed.

## I. INTRODUCTION

It has been realized in recent years that group theory could profitably be employed in the study of hadrons. This is due to the symmetries possessed by the systems which in turn permit one to classify their spectra and to calculate some experimental numbers. Inquiries along this line have invariably led to noncompact groups as evidenced both in exactly solvable nonrelativistic quantum-mechanical systems and strong-coupling theories. This introduction of noncompact groups has, of course, been necessitated because of the existence of bandlike structures in the spectrum. Meanwhile, it is also noted that such symmetries have a dynamical origin which may throw some light on the role and relevance of noncompact groups and other allied problems. Thus, it is apparent that noncompact groups become a powerful apparatus for exploring symmetries and dynamics of physical systems.

In this paper, an extension theorem pertaining to noncompact groups has been stated and proved. As an application, the theorem is then used to find out the unitary irreducible representations of the noncompact groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$. It is shown that, by using the extension procedure only, a class of unitary irreducible representations can be obtained from the infinitely many otherwise possible for the groups under consideration. Nevertheless, the representations obtained by extension technique are

* Part of this work was done in the Ames Laboratory of the U.S. Atomic Energy Commission.

† Present address: Douglas Advanced Research Laboratories, Huntington Beach, Calif.

factor $(+1)\int_{-\infty}^{\infty} dt_3$ when the time is evolving in the forward direction and a factor $(-1)\int_{-\infty}^{\infty} dt_3$ when time evolves in the backward direction. We can remove these factors by appending a $\tau^3$ to the outgoing line at this vertex, thereby turning $G(3, 4)_{ij}$ into $\tilde{G}(3, 4)_{ij}$. If we replace all $G(3, 4)_{ij}$ by $\tilde{G}(3, 4)_{ij}$ however, one vertex over which no path integration is performed gets a factor of $\tau^3$. But this is precisely the factor in the precise location needed to turn $\Sigma(1, 2)_{ij}$ into $\tilde{\Sigma}(1, 2)_{ij}$. Hence our rules indeed give us $\tilde{\Sigma}(1, 2)_{ij}$.

## 4. DISCUSSION

In the preceding we have shown that, by retaining redundant information on $g^{>}(1, 2)$ and $g^{<}(1, 2)$, the equations of motion for these quantities can be written in a form in which the integrals run over all space and time. Using this approach, and by retaining initial-particle correlations in the unperturbed Green's functions, we have shown that the separation of the interaction terms into self-energy and single-particle Green's function terms is justified without the need for the correction terms of Fujita. In addition, we have devised a perturbation expansion to determine the self-energy functions $\Sigma^{>}(1, 2)$ and $\Sigma^{<}(1, 2)$ appearing in the theory of Kadanoff and Baym. Since both space and time integrations run over all values, the explicit functional dependence of the self-energy functions upon the single-particle Green's function $g^{>}(1', 2')$ and $g^{<}(1', 2')$ may be determined directly via the dependence of $\tilde{\Sigma}_{ij}(1, 2)$ on $\tilde{G}(1', 2')_{i'j'}$.

## ACKNOWLEDGMENTS

The author wishes to thank Professor Joel W. McClure for his kind hospitality during the course of this work. He also wishes to thank Professor Leo P. Kadanoff for the comments which led to this research and Professor Marvin D. Girardeau for a critical reading of the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Extension Theorem and Representations*

A. SANKARANARAYANAN

*Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa*†

(Received 23 February 1967)

An extension theorem prescribing a method for constructing the $(n + 1)$-dimensional Lorentz group $O(n, 1)$ from the $n$-dimensional inhomogeneous rotation group, proved earlier, has been recapitulated and some comments on the theorem are made. The prescription is used to construct the unitary irreducible representations of the Lorentz groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$, and it is found that only a limited number of representations are allowed.

## I. INTRODUCTION

It has been realized in recent years that group theory could profitably be employed in the study of hadrons. This is due to the symmetries possessed by the systems which in turn permit one to classify their spectra and to calculate some experimental numbers. Inquiries along this line have invariably led to noncompact groups as evidenced both in exactly solvable nonrelativistic quantum-mechanical systems and strong-coupling theories. This introduction of noncompact groups has, of course, been necessitated because of the existence of bandlike structures in the spectrum. Meanwhile, it is also noted that such symmetries have a dynamical origin which may throw some light on the role and relevance of noncompact groups and other allied problems. Thus, it is apparent that noncompact groups become a powerful apparatus for exploring symmetries and dynamics of physical systems.

In this paper, an extension theorem pertaining to noncompact groups has been stated and proved. As an application, the theorem is then used to find out the unitary irreducible representations of the noncompact groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$. It is shown that, by using the extension procedure only, a class of unitary irreducible representations can be obtained from the infinitely many otherwise possible for the groups under consideration. Nevertheless, the representations obtained by extension technique are

* Part of this work was done in the Ames Laboratory of the U.S. Atomic Energy Commission.

† Present address: Douglas Advanced Research Laboratories, Huntington Beach, Calif.

physically interesting and relevant to noninvariance groups.

## II. EXTENSION THEOREM

This theorem gives a prescription for constructing the $(n+1)$-dimensional Lorentz group $O(n, 1)$ from the $n$-dimensional inhomogeneous rotation group. This has been achieved by taking certain elements of the universal enveloping algebra of the inhomogeneous rotation group. The universal enveloping algebra is the set of polynomials formed from the generators of the Lie algebra of a Lie group by taking into account the commutation relations of the Lie algebra. It may also be noted that the center of the universal enveloping algebra is generated by the so-called Casimir operators, the values of which essentially characterize the irreducible representation of the group.

*Theorem*[1]*:* If the rotation and translation generators $M_{ab}$ and $p_c$, respectively, of the $n$-dimensional inhomogeneous rotation group $[IO(n)]$ are given, then one can construct the homogeneous Lorentz group of the $(n+1)$-dimensional space; the generators of the latter are given by

$$M_{ab} \quad \text{and} \quad X_c = 1/2(p_a p_a)^{\frac{1}{2}}[M_{cb}, p_b]_+ .$$

*Proof:* The inhomogeneous rotation group in $n$ dimensions is the group of all real linear transformations with determinant $+1$ that leave invariant the quadratic form

$$\sum \delta_{ab}(\Delta x_a)(\Delta x_b), \tag{1}$$

where $(\Delta x_a)$ denotes the displacement between two points in this $n$-dimensional space and the Euclidean metric $\delta_{ab}$ is the Kronecker delta. The transformation is generally written as

$$x_i' = \sum_{j=1}^{n} a_{ij}x_j + b_i, \tag{2}$$

where the coefficients $a_{ij}$ and $b_i$ are real and are characteristic of rotations and translations, respectively. The $a_{ij}$ constitute an orthogonal matrix. Thus the group $IO(n)$ is generated by the $n(n-1)/2$ infinitesimal rotation operators and $n$ infinitesimal translation operators. The number of real parameters of $IO(n)$ is $n(n+1)/2$. As is well known, the infinitesimal operators can be obtained from the Taylor-series expansion of functions in the $n$-dimensional space. The infinitesimal rotation operators $M_{ab}$ are chosen to be Hermitian and are antisymmetric:

$$M_{ab} = -M_{ba}. \tag{3}$$

They obey the commutation relations

$$[M_{ab}, M_{cd}] = i(\delta_{ac}M_{bd} + \delta_{ad}M_{cb} - \delta_{cb}M_{ad} - \delta_{bd}M_{ca}). \tag{4}$$

These commutation relations may easily be obtained from the root diagram of the Lie ring. The infinitesimal translation operators $p_a$ are also chosen to be Hermitian and obey the following commutation relations:

$$[p_a, p_b] = 0, \tag{5}$$

$$[M_{ab}, p_c] = i(\delta_{ac}p_b - \delta_{bc}p_a). \tag{6}$$

It is also to be noted that $p_c^2$ is a Casimir operator of the group as it commutes with the generators $M_{ab}$ and $p_c$. This operator $p_c^2$ must have constant value for an irreducible representation of the group $IO(n)$ which is partially characterized by this value. It is always assumed that $p_c^2$ is greater than zero unless otherwise mentioned.

Now to prove the theorem, one has to show that $M_{ab}$ and $X_c = 1/2(p_a p_a)^{\frac{1}{2}}[M_{cb}, p_b]_+$ give the Lie algebra of the $(n+1)$-dimensional Lorentz group $O(n, 1)$. First of all, the number of parameters for $O(n, 1)$ is $n(n+1)/2$, which is the same as the number of parameters given by $M_{ab}$ [$n(n-1)/2$ parameters] and $X_c$ ($n$ parameters). The Lie algebra of $M_{ab}$ and $X_c$ can be easily calculated by making use of Eqs. (4) to (6) and is given by

$$[M_{ab}, X_c] = i(\delta_{ac}X_b - \delta_{bc}X_a) \tag{7}$$

and

$$[X_a, X_b] = -iM_{ab}. \tag{8}$$

Now Eqs. (4), (7), and (8) can be jointly written as

$$[M_{a'b'}, M_{c'd'}] = i(g_{a'c'}M_{b'd'} + g_{a'd'}M_{c'b'} - g_{b'c'}M_{a'd'} - g_{b'd'}M_{c'a'}), \tag{9}$$

where $a'$, $b'$, $c'$, and $d'$ run from 1 to $(n+1)$; $g_{ab} = \delta_{ab}$ and $g_{n+1,n+1} = -1$, and $M_{n+1,a} = X_a = -M_{a,n+1}$; $M_{n+1,n+1} = 0$. It is also obvious that

$$M_{a'b'} = -M_{b'a'}.$$

Equation (9) gives the Lie algebra of $O(n, 1)$ and the generators are $M_{a'b'}$. Hence the theorem is proved. We now make the following comments.

(1) Although the above "extension" theorem is proved for the inhomogeneous rotation group, *the theorem is true, in general, for a proper rotation group in n dimensions denoted by the generators $M_{ab}$ and a set of quantities, say $q_c$, instead of the translation generators, if $q_c$ transforms like an n-dimensional vector under rotations $M_{ab}$ and forms an Abelian algebra like the translation generators.* This idea has been exploited

[1] A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965).

by the author[2] in constructing the generators of the noninvariance group for the hydrogen atom.

(2) The theorem is true also for pseudo-Euclidean rotations and all one has to do is to use the appropriate reality conditions in accordance with the type of space one is dealing with. That is to say, given the group $IO(m, n)$, one can construct the Lie algebra of the group $O(m + 1, n)$ or $O(m, n + 1)$, depending on the metric element. In particular it is shown elsewhere[3] that one can construct the de Sitter group from the Poincaré group by using the above theorem. It is rather remarkable that by starting with flat space one can go to a space with constant curvature by making use of the theorem. Of course, it is to be noted that the groups obtained for a space of constant curvature and for a flat space have the same number of parameters even though the groups themselves are different both algebraically and topologically. This connection between the Poincaré group and the de Sitter groups [both $O(3, 2)$ and $O(4, 1)$] is perhaps to be traced to the fact that the de Sitter groups have the property of being invariance groups of spaces which are locally isomorphic metrically to Minkowski space-time. It is also suspected that this observation may be true in general.

(3) It has been shown[4] that the subgroups $IO(n, m)$, $O(n + 1, m)$, and $O(n, m + 1)$ of $O(n + 1, m + 1)$ leave invariant a given vector $v$ in the corresponding space if it is lightlike, timelike, or spacelike, respectively. But it is observed earlier that by using the extension theorem one can construct $O(n + 1, m)$ and $O(n, m + 1)$ from $IO(n, m)$. Thus the theorem could be used to find out the little groups. Then these ideas could be exploited in constructing the representations of the various groups as they all have the same complex Lie algebras.

(4) One would have obtained the Lie algebra given by Eq. (9) just by defining $X'_c = (p_a p_a)^{-\frac{1}{2}} M_{cb} p_b$. But this choice is not desirable as it is neither Hermitian nor anti-Hermitian and consequently it will lead to difficulties in interpretation and in defining the Hilbert space and scalar product.

(5) The prescription given by the theorem enables one to construct a basis of the Lie algebra for the Lorentz group $O(m, n + 1)$ or $O(m + 1, n)$ from $IO(m, n)$.[3] Then this could be exploited in studying irreducible representations of the constructed group. We will discuss this elaborately in the section on the representations. But we only mention here that this procedure gives only a limited number of unitary irreducible representations which nevertheless turn out to be physically significant, as we will see later.

(6) It has also been noted[3,5,6] that adding a term $p_a$ to $X_a$ still preserves the commutations and thus the Lie algebra and hence the theorem is still true. In fact this has already been exploited by Böhm[5] to a great extent in defining "dynamical" groups and in deriving mass spectrum. Even though it works well within the framework of Lie-algebra relations, such an additional term (that is, adding $p_a$ to $X_a$) would give undesirable Hermitian or time-reversal properties. To see this clearly, let us take the example $IO(3)$. The generators are chosen to be Hermitian and are given by $\mathbf{J}$ and $\mathbf{p}$, respectively, for the rotations and the translations with the commutation rules

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk} J_k, \\ [J_i, p_j] &= i\epsilon_{ijk} p_k, \\ [p_i, p_j] &= 0. \end{aligned} \tag{10}$$

Under the time-reversal operation $T$,

$$\begin{aligned} \mathbf{p} &\to -\mathbf{p}, \\ \mathbf{J} &\to -\mathbf{J}. \end{aligned} \tag{11}$$

Now one defines the generalized momenta following Böhm and Rosen,

$$\mathbf{X} = \mathbf{p} + \frac{\lambda}{2(p_i p_i)^{\frac{1}{2}}} (\mathbf{p} \times \mathbf{J} - \mathbf{J} \times \mathbf{p}), \tag{12}$$

where $(p_i p_i)$ is assumed to be greater than zero. If $\lambda$ is real, then $\mathbf{X}$ is a well-defined Hermitian operator by Garding's construction.[7] But under time reversal $T$,

$$\mathbf{X} \to -\mathbf{p} + [\lambda/2(p_i p_i)^{\frac{1}{2}}](\mathbf{p} \times \mathbf{J} - \mathbf{J} \times \mathbf{p}). \tag{13}$$

Thus it does not transform like a momentum vector under time reversal in the original space. But one shall make this behave like a momentum vector by making $\lambda$ to be pure imaginary by taking into account the antiunitary nature of time reversal.[8,9] But then the Hermitian character of $\mathbf{X}$ is lost. Thus it leads to the question of whether one can define a Hermitian "generalized momentum" in a consistent way without violation of the invariance under discrete transformations. It is to be remarked that in the above criticism it is tacitly assumed that the physical space and the group space are the same. If this is not so, then the situation is not clear. Anyway, in these days of breakdowns of discrete symmetries, perhaps one does not

[2] A. Sankaranarayanan, Nuovo Cimento **44**, 193 (1966).
[3] A. Sankaranarayanan and R. H. Good, Jr., Phys. Rev. **140**, B509 (1965); J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966).
[4] H. Bacry, Ann. Henri Poincaré **11**, 327 (1965); P. M. Mathews, Nuovo Cimento **45A**, 527 (1966).
[5] A. Böhm, Phys. Rev. **145**, 1212 (1966).
[6] J. Rosen, Boston University (report of work prior to publication).
[7] L. Garding, Proc. Natl. Acad. Sci. U.S. **33**, 331 (1947).
[8] E. P. Wigner, Göttingen Nachr. 546 (1932).
[9] L. L. Foldy, Phys. Rev. **102**, 568, (1956).

have to worry about this. Again, if these dynamical groups signify real physics, it is perhaps an indication of the violation of time-reversal invariance in nature!

### III. EXTENSION THEOREM AND THE REPRESENTATIONS OF $O(2,1)$ FROM $IO(2)$

The group $IO(2)$ may be considered as the group of rigid motions of a plane which is the semidirect product of the one-parameter rotation group and two-parameter translation group. The generators for the rotation and the translation are denoted by $J_3$, $P_1$, and $P_2$, respectively, and their Lie algebra is

$$\begin{aligned}[J_3, P_\pm] &= \pm P_\pm, \\ [P_+, P_-] &= 0,\end{aligned} \tag{14}$$

where $P_\pm = P_1 \pm iP_2$. As is obvious, the generators $P_1$ and $P_2$ form an Abelian invariant subgroup. One can proceed to find the unitary irreducible representation of this group either by making the translation subgroup or by making the rotation subgroup diagonal. Anyway, for our purpose, we will use the latter method.

We will use the Lie algebra given by Eq. (14) to find the unitary irreducible representation of the group $IO(2)$ which is noncompact; then the unitary irreducible representations are infinite dimensional.

One can use the irreducible representations of a Lie algebra if the generators are defined on a dense set of Hilbert space, since a valid connection between the invariant subspaces of the representations of the algebra and the representations of the group exists. This is, of course, the consequence of the fact that there exists a dense set of analytic vectors for every unitary representation.[10]

We will work out the unitary irreducible representations of this group on discrete Hilbert space which is defined as the linear vector space over the orthonormal basis $|m\rangle$ where $m$ is countable and discrete and a vector $\Phi = \sum_m \Phi_m |m\rangle$ belongs to this discrete Hilbert space if and only if the norm $\|\Phi\|$ is finite.

The operator $J_3$ generates space rotation on the plane and thus the subgroup generated by $J_3$ is Abelian and compact. Supposing that a unitary irreducible representation is given, the eigenvalues of the generators are real. The spectrum of $J_3$ is discrete as it is compact and one can diagonalize $J_3$. Then, by virtue of Eq. (14) it is seen that the real eigenvalues of $J_3$ differ by unity and $P_+$ and $P_-$ raises or lowers, respectively, the eigenvalue of $J_3$ by one unit. These representations are in general multivalued. The space corresponding to each eigenvalue of $J_3$ is one-dimensional and the Hilbert space $H$ on which the unitary irreducible representations of $IO(2)$ are defined is the direct sum of these one-dimensional spaces. Thus the eigenvectors of $J_3$ form a basis for $H$. Then

$$J_3 |m\rangle = m |m\rangle. \tag{15}$$

By operating the commutator $[J_3, P_+] = P_+$ on the vector $|m\rangle$ we get

$$P_+ |m\rangle = C'_{m+1} |m+1\rangle, \tag{16}$$

where $C'_{m+1}$ may be a complex number. We exclude the possibility of this being zero. Again from the commutator $[J_3, P_-] = -P_-$ we obtain

$$P_- |m\rangle = C''_m |m-1\rangle, \tag{17}$$

where again we assume $C''_m$ to be a nonzero complex number. By properly normalizing the vectors it can be shown that

$$C'_m = -C''_m, \tag{18}$$

for every $m$. This phase convention is chosen for future convenience. From the commutator $[P_+, P_-] = 0$ we get

$$C'^2_{m+1} - C'^2_m = 0. \tag{19}$$

This reflects the fact that the constant does not depend on $m$. Now the Casimir operator of the group is given by $P_+P_-$ which acts on $|m\rangle$ to give

$$P_+P_- |m\rangle = -C'^2_m |m\rangle = \mu^2 |m\rangle. \tag{20}$$

Thus the meaning of the constant $C$ becomes clear. So, for the complete classification of $IO(2)$ one has only to determine the form of the eigenspectrum of $J_3$ and its relation with the Casimir operator of the group.

Generally, the eigenvectors can be represented by $|\mu, m\rangle$ and then we get

$$\begin{aligned}J_3 |\mu, m\rangle &= m |\mu, m\rangle, \\ P_\pm |\mu, m\rangle &= \pm i\mu |\mu, m \pm 1\rangle,\end{aligned} \tag{21}$$

where $\mu$ is always assumed to be positive. The explicit matrix elements for the generators $J_3$, $P_1$, and $P_2$ are given by

$$\begin{aligned}\langle \mu, m' | J_3 | \mu, m\rangle &= m\delta_{m'm}, \\ \langle \mu, m' | P_1 | \mu, m\rangle &= (i\mu/2)[\delta_{m',m+1} - \delta_{m',m-1}], \\ \langle \mu, m' | P_2 | \mu, m\rangle &= (\mu/2)[\delta_{m',m+1} + \delta_{m',m-1}].\end{aligned} \tag{22}$$

As is obvious from the range of $m$, the unitary irreducible representations of the group are infinite-dimensional and each unitary irreducible representation is characterized by the Casimir invariant $P_1^2 + P_2^2 = \mu^2$ and the value of $J_3$. It is also to be noted in Eq. (22) that $P_1$ and $P_2$ have nonzero matrix

[10] E. Nelson, Ann. Math. 70, 572 (1959).

elements only between the consecutive eigenvalues of $J_3$, and $P_1$ has all its matrix elements pure imaginary and $P_2$ all real; this fact preserves the unitary nature of the representation. We have here avoided the representations in which all the matrix elements of $P_1$ and $P_2$ are zero, since such a case is of no interest as the "extension" theorem cannot be applied.

We now turn to discuss the "extension" of $IO(2)$ and define the operators

$$X_i = [1/2(P_1^2 + P_2^2)^{\frac{1}{2}}][M_{ij}, p_j]_+. \tag{23}$$

As we said earlier, $(P_1^2 + P_2^2)$ is the Casimir operator of $IO(2)$ and it is assumed to be nonzero for our purposes. Then we get

$$\begin{aligned} X_1 &= (1/2\mu)[J_3P_2 + P_2J_3], \\ X_2 &= -(1/2\mu)[J_3P_1 + P_1J_3]. \end{aligned} \tag{24}$$

The commutation of $X_1$ and $X_2$ with $J_3$ is given by

$$\begin{aligned} [X_1, X_2] &= -iJ_3, \\ [X_1, J_3] &= -iX_2, \\ [X_2, J_3] &= iX_1. \end{aligned} \tag{25}$$

Thus Eq. (25) generates the Lie algebra of the $(2 + 1)$ Lorentz group $O(2, 1)$. Now we are interested in finding out which unitary irreducible representations of $O(2, 1)$ are obtained from a given $\mu^2$ unitary irreducible representation of $IO(2)$. Construct

$$\begin{aligned} X_\pm &= X_1 \pm iX_2 \\ &= \mp(i/2\mu)[J_3, P_\pm]_+. \end{aligned} \tag{26}$$

By operating $X_\pm$ on the eigenvectors $|\mu, m\rangle$ we get

$$\begin{aligned} X_\pm |\mu, m\rangle &= \mp(i/2\mu)[J_3, P_\pm]_+ |\mu, m\rangle \\ &= (m \pm \tfrac{1}{2}) |\mu, m \pm 1\rangle. \end{aligned} \tag{27}$$

The Casimir operator of the group $O(2, 1)$ is given by

$$C = -\tfrac{1}{2}(X_+X_- + X_-X_+) + J_3^2. \tag{28}$$

When $C$ operates on the vector $|\mu, m\rangle$ we get

$$C |\mu, m\rangle = -\tfrac{1}{4} |\mu, m\rangle. \tag{29}$$

Thus, we can get ∂ unitary irreducible representation of $O(2, 1)$ characterized by the eigenvalue $(-\frac{1}{4})$ of the Casimir operator. It belongs to the eigenvalues of $J_3$ varying by steps of 1 from $-\infty$ to $+\infty$.

The unitary irreducible representations of $O(2, 1)$ deduced here by using the "extension" theorem are to be identified with the following representations obtained by Bargmann[11]:

(i) For $m$ integer, one obtains the member of Bargmann's continuous class $C_q^0$ with $q = \frac{1}{4}$ and $m = 0, \pm 1, \cdots$.

(ii) For $m = \frac{1}{2}$ odd integer, the unitary irreducible representation belongs to Bargmann's discrete class $D_{\frac{1}{2}}^+$ with $m = \frac{1}{2}, \frac{3}{2}, \cdots$, and $D_{\frac{1}{2}}^-$ with $m = -\frac{1}{2}, -\frac{3}{2}, \cdots$. It is interesting to note here that the operators $X_\pm$ vanish when $m = \mp\frac{1}{2}$ as is obvious from Eq. (27).

(iii) For all other $m$'s the representations of $O(2, 1)$ is neither single-valued nor double-valued and hence corresponds to none of Bargmann's classes as he considers only the single- and double-valued ones. Then writing

$$C = -\tfrac{1}{4} = K(K - 1), \tag{30}$$

we get

$$K = \tfrac{1}{2}.$$

Thus, the unitary irreducible representation we get corresponds to $D_{\frac{1}{2}}^\pm$ in the notation of Bargmann.[11] Hence, it turns out that the operators constructed from $IO(2)$ by means of the "extension" theorem give uniquely the unitary irreducible representation $D_{\frac{1}{2}}^\pm$. It is well known from Bargmann's work that for a fixed value of $K$ there exists an infinite set of representations for the $O(2, 1)$ group and all these representations differ from one another in an arbitrary number $\lambda$ which occurs in choosing the proper basis vectors of the representation. Thus we shall write

$$\begin{aligned} &J_3 |K, m\rangle = (\lambda + m) |K, m\rangle, \\ &X_\pm |K, m\rangle \\ &= [K(1 - K) + (\lambda + m)(\lambda + m \pm 1)]^{\frac{1}{2}} |K, m \pm 1\rangle. \end{aligned} \tag{31}$$

In our case $K(K - 1) = -\frac{1}{4}$. Then when $C$ acts on $|K, m\rangle$ with the choice of eigenvalues given by Eq. (31) we again get

$$\begin{aligned} C |K = \tfrac{1}{2}, m\rangle &= [J_3^2 - \tfrac{1}{2}(X_+X_- + X_-X_+)] |K = \tfrac{1}{2}, m\rangle \\ &= -\tfrac{1}{4} |K = \tfrac{1}{2}, m\rangle. \end{aligned} \tag{32}$$

This agrees with our original result given by Eq. (29). It is interesting to write down the matrix elements of $X_1$ and $X_2$ and to compare them with those of $P_1$ and $P_2$ of $IO(2)$. They are given by

$$\begin{aligned} &\langle K, m'| X_1 |K, m\rangle \\ &\quad = \tfrac{1}{2}\{[\tfrac{1}{4} + (\lambda + m)(\lambda + m + 1)]^{\frac{1}{2}}\delta_{m', m+1} \\ &\qquad + [\tfrac{1}{4} + (\lambda + m)(\lambda + m - 1)]^{\frac{1}{2}}\delta_{m', m-1}\}, \\ &\langle K, m'| X_2 |K, m\rangle \\ &\quad = -i/2\{[\tfrac{1}{4} + (\lambda + m)(\lambda + m + 1)]^{\frac{1}{2}}\delta_{m', m+1} \\ &\qquad - [\tfrac{1}{4} + (\lambda + m)(\lambda + m - 1)]^{\frac{1}{2}}\delta_{m', m-1}\}. \end{aligned} \tag{33}$$

Like the matrix elements of $P_1$ and $P_2$ given by Eq.

[11] V. Bargmann, Ann. Math. **48**, 568 (1947).

(22), the matrix elements of $X_1$ and $X_2$ have nonvanishing matrix elements only just above and just below the main diagonal. It is, of course, a reflection of the fact that both $(P_1, P_2)$ and $(X_1, X_2)$ transform like vectors under the rotation $J_3$. Again, to preserve the unitary nature of the representations, the matrix elements of $X_1$ and $X_2$ are real and purely imaginary, respectively, as in the case of Eq. (22). Both the matrix elements $X_1$ and $X_2$ have different numbers in each element, whereas for $P_1$ and $P_2$ we have the same number $\mu/2$.

The group $O(2, 1)$ is infinitely connected as it is topologically homeomorphic to the direct product of the circle and the Euclidean plane. Then, one might expect to obtain both "single-valued" and "multivalued" irreducible representations of $O(2, 1)$ for $C = -\frac{1}{4}$. This, of course, depends on the fact how one starts in Eq. (21) with "single-valued" ($m =$ integer) or "multivalued" ($m \neq$ integer) irreducible representation of $IO(2)$. The generators of $O(2, 1)$ constructed by using the "extension" theorem are exactly what one needs for the group $O(2, 1)$ when $O(2, 1)$ is interpreted as the noninvariance group of the one-dimensional harmonic oscillator. Thus the generators of $O(2, 1)$ given by Eqs. (24) and (25) completely describe the behavior of a one-dimensional oscillator where energy is associated with the compact subgroup (i.e., rotation in the plane). The energy spectrum for the special value of the Casimir operator $C = -\frac{1}{4}$ (corresponding to the unitary irreducible representation $D^{+}_{\frac{1}{2}}$) is $J_3 = (n + \frac{1}{2})$. Thus the group defined by Eq. (25) contains precisely the complete information about the one-dimensional oscillator, including energy levels, degeneracy, and the Hilbert space. Again this group $O(2, 1)$ is the noninvariance group of the hydrogen atom in unidimension whose bound-state energy spectrum is nondegenerate and discrete.

## IV. REPRESENTATIONS OF $O(3, 1)$ FROM $IO(3)$

The group $IO(3)$ is generated by translation generators $p_i$ and rotation generators $J_i$ and their Lie algebra is given by

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk}J_k, \\ [J_i, p_j] &= i\epsilon_{ijk}p_k, \\ [p_i, p_j] &= 0. \end{aligned} \tag{34}$$

As is obvious from the commutation relations, the $p_i$ form an Abelian invariant subgroup and the $J_i$ form the rotation subgroup. Thus $IO(3)$ is the semidirect product of these two subgroups. As in the case of $IO(2)$, one can find the unitary irreducible representations either by making translation-generator eigenstates or by constructing rotation-group eigenstates. Now all the considerations made earlier about the Hilbert space $H$ are applicable here and we construct unitary irreducible representations of $IO(3)$ in terms of Hermitian matrices on $H$. The Lie algebra given by Eq. (34) is rewritten in the following form for convenience in terms of $J_\pm$, $J_3$, $P_\pm$, and $P_3$:

$$\begin{aligned} [J_3, J_\pm] &= \pm J_\pm, \\ [J_+, J_-] &= 2J_3, \\ [J_3, P_\pm] &= [P_3, J_\pm] = \pm P_\pm, \\ [J_+, P_-] &= [P_+, J_-] = 2P_3, \end{aligned} \tag{35}$$

and the rest of the commutators are zero. The representation space of $IO(3)$ is denoted by $H$ which can be written as a direct sum of $H_j$,

$$H = \sum_j \oplus H_j, \tag{36}$$

where $H_j$ are all irreducible with respect to the rotation group $O(3)$ and each value of $j$ occurs at most once. We choose the unitary irreducible representations of $O(3)$ in such a way that $J_3$ is diagonal and the matrix elements are given by

$$\langle j, m' | J_3 | j, m \rangle = m\delta_{m'm}, \tag{37}$$

where the eigenvalues $m$ run through $-j$ to $+j$ and the $|j, m\rangle$ represent the basis vectors of each subspace and for a fixed $j$ they form a canonical basis for an irreducible representation of $O(3)$. The matrix element $\langle j, m' | J_\pm | j, m \rangle \neq 0$ only for $m' = m \pm 1$ and for all other "$m$" values this matrix element will be zero. This reflects that the "$m$" values in a given representation differ by integers. With an arbitrary choice of phase these matrix elements are

$$\begin{aligned} \langle j, m | J_+ | j, m-1 \rangle &= \langle j, m-1 | J_- | j, m \rangle \\ &= [j(j+1) - m(m-1)]^{\frac{1}{2}}, \end{aligned} \tag{38}$$

$$J^2 | j, m \rangle = j(j+1) | j, m \rangle. \tag{39}$$

Now since $\mathbf{P}$ transforms as a vector under rotations, it follows that the reduced matrix elements $\langle j | \mathbf{P} | j' \rangle = 0$ unless $j' = j$ or $j \pm 1$ and the matrix elements of $\mathbf{P}$ are fixed by its vector nature. The Casimir operators of the group are given by

$$C_1 = \mathbf{P} \cdot \mathbf{P}, \tag{40}$$

$$C_2 = \mathbf{J} \cdot \mathbf{P} = \mathbf{P} \cdot \mathbf{J}, \tag{41}$$

and in each irreducible representation they are given by a numerical constant. The unitary irreducible representations are worked out by literally following the Harish-Chandra–Neumark method[12] and taking into account the commutation relation between $\mathbf{J}$ and $\mathbf{P}$.

[12] Harish-Chandra, Proc. Roy. Soc. (London) **189A**, 372 (1947); M. A. Neumark, Am. Math. Soc. Transl., Ser. 2, **6**, 379 (1957).

The matrix elements of **P** are given by

$$\langle j', m' | P_3 | j, m\rangle = \frac{\mu}{2(j+1)}\left[\frac{(j+h+1)(j-h+1)(j+m+1)(j-m+1)}{(j+\frac{3}{2})(j+\frac{1}{2})}\right]^{\frac{1}{2}} \times \delta_{j',j+1}\delta_{m'm} + \frac{m\mu h}{j(j+1)}\delta_{j',j}\delta_{m',m} + \frac{\mu}{2j}\left[\frac{(j+h)(j-h)(j+m)(j-m)}{(j+\frac{1}{2})(j-\frac{1}{2})}\right]^{\frac{1}{2}}\delta_{j',j-1}\delta_{m',m},$$

$$\langle j', m' | P_\pm | j, m\rangle = \frac{\mu h}{j(j+1)}[(j \pm m + 1)(j \mp m)]^{\frac{1}{2}}\delta_{j',j}\delta_{m',m\pm1} \mp \frac{\mu}{2(j+1)}\left[\frac{(j+h+1)(j-h+1)(j\pm m+1)(j\pm m+2)}{(j+\frac{3}{2})(j+\frac{1}{2})}\right]^{\frac{1}{2}}\delta_{j',j+1}\delta_{m',m\pm1} \pm \frac{\mu}{2j}\left[\frac{(j+h)(j-h)(j\mp m)(j\mp m-1)}{(j+\frac{1}{2})(j-\frac{1}{2})}\right]^{\frac{1}{2}}\delta_{j',j-1}\delta_{m',m\pm1}.$$

The Casimir operators give

$$\mathbf{P}\cdot\mathbf{P}\,|\mu, j, m\rangle = \mu^2\,|\mu, j, m\rangle \tag{43}$$

and

$$\mathbf{J}\cdot\mathbf{P}\,|\mu, j, m\rangle = \mu h\,|\mu, j, m\rangle, \tag{44}$$

where $\mu$ is always assumed to be nonzero and real and the minimum value of $j$ is always equal to the value of $h$. Thus,

$$j_{\min} = h. \tag{45}$$

Then the whole Hilbert space, in which the unitary irreducible representations are defined, is given by

$$H = \sum_{j=j_{\min}=h}^{\infty} \oplus H_j, \tag{46}$$

and it is obvious that the unitary irreducible representations are infinite-dimensional as there is no upper limit on the value of $j$. Here again, we do not consider the case where $\mathbf{P}\cdot\mathbf{P}$ takes the eigenvalue zero since such a situation is not pertinent for using the "extension" theorem in the construction of the Lie algebra of $O(3, 1)$.

A Lie algebra isomorphic to that of $O(3, 1)$ is obtained by applying the "extension" theorem on $IO(3)$. The new generators are

$$K_i = [1/2(\mathbf{P}\cdot\mathbf{P})^{\frac{1}{2}}]\epsilon_{ijk}[J_k, P_j]_+. \tag{47}$$

In vector notation it reads

$$\mathbf{K} = (1/2\mu)[\mathbf{P}\times\mathbf{J} - \mathbf{J}\times\mathbf{P}]. \tag{48}$$

The commutation relations between **J** and **K** are given by

$$\begin{aligned}\mathbf{J}\times\mathbf{J} &= i\mathbf{J},\\ \mathbf{K}\times\mathbf{K} &= -i\mathbf{J},\\ \mathbf{J}\times\mathbf{K} + \mathbf{K}\times\mathbf{J} &= 2i\mathbf{K}.\end{aligned} \tag{49}$$

Thus **J** and **K** generate an algebra isomorphic to the Lie algebra of $O(3, 1)$. It now remains to be seen what unitary irreducible representations are obtained from the $(\mu, m)$ unitary irreducible representation of $IO(3)$ by the "extension" procedure. To achieve this end, let us construct $K_\pm$ and $K_3$, and they are given by

$$\begin{aligned}K_\pm &= \pm(i/2\mu)\{[J_\pm, P_3]_+ - [J_3, P_\pm]_+\},\\ K_3 &= (1/2\mu)[(J_2P_1 + P_1J_2) - (J_1P_2 + P_2J_1)].\end{aligned} \tag{50}$$

By operating $K_\pm$ and $K_3$ on the canonical basis vectors $|j, m\rangle$ we obtain

$$\begin{aligned}K_\pm\,|j, m\rangle &= \pm\frac{i}{2}\left[\frac{(j+h+1)(j-h+1)(j\pm m+1)(j\pm m+2)}{(j+\frac{3}{2})(j+\frac{1}{2})}\right]^{\frac{1}{2}}|j+1, m\pm1\rangle \pm \frac{i}{2}\left[\frac{(j+h)(j-h)(j\mp m)(j\mp m-1)}{(j+\frac{1}{2})(j-\frac{1}{2})}\right]^{\frac{1}{2}}|j-1, m\pm1\rangle,\\ K_3\,|j, m\rangle &= -\frac{i}{2}\left[\frac{(j+h+1)(j-h+1)(j+m+1)(j-m+1)}{(j+\frac{3}{2})(j+\frac{1}{2})}\right]^{\frac{1}{2}}|j+1, m\rangle + \frac{i}{2}\left[\frac{(j+h)(j-h)(j-m)(j+m)}{(j+\frac{1}{2})(j-\frac{1}{2})}\right]|j-1, m\rangle.\end{aligned} \tag{51}$$

Here, since **K** transforms as a vector under rotation, one would expect that $\langle j|\,\mathbf{K}\,|j'\rangle \neq 0$ for $j' = j$ or $j \pm 1$. But from Eq. (51) it is apparent that we do not have nonzero matrix elements for $j' = j$. It will become obvious sooner that this is the reason that one could not obtain all the unitary irreducible representations of $O(3, 1)$ by using the "extension" process. Now to obtain the unitary irreducible representations of the algebra (49), we will explicitly write down the irreducible representation of the Lorentz group[12] in Hilbert space. They are given by

$$
\begin{aligned}
J_3\,|j, m\rangle &= m\,|j, m\rangle,\\
J_\pm\,|j, m\rangle &= [j(j+1) - m(m \pm 1)]^{\frac{1}{2}}\,|j, m \pm 1\rangle,\\
F_3\,|j, m\rangle &= -\frac{imhc}{j(j+1)}\,|j, m\rangle\\
&\quad - \frac{i}{2(j+1)}\left[\frac{(j+h+1)(j-h+1)(j+m+1)(j-m+1)(j+c+1)(j-c+1)}{(j+\frac{3}{2})(j+\frac{1}{2})}\right]^{\frac{1}{2}}\\
&\quad \times |j+1, m\rangle + \frac{i}{2j}\left[\frac{(j+h)(j-h)(j-m)(j+m)(j+c)(j-c)}{(j+\frac{1}{2})(j-\frac{1}{2})}\right]^{\frac{1}{2}}|j-1, m\rangle, \qquad (52)\\
F_\pm\,|j, m\rangle &= \pm\frac{i}{2(j+1)}\left[\frac{(j+h+1)(j-h+1)(j\pm m+1)(j\pm m+2)(j+c+1)(j-c+1)}{(j+\frac{3}{2})(j+\frac{1}{2})}\right]^{\frac{1}{2}}\\
&\quad \times |j+1, m \pm 1\rangle - \frac{ihc}{j(j+1)}\,[(j \mp m)(j \pm m + 1)]^{\frac{1}{2}}\,|j, m \pm 1\rangle\\
&\quad \pm \frac{i}{2j}\left[\frac{(j+h)(j-h)(j\mp m)(j\mp m-1)(j+c)(j-c)}{(j+\frac{1}{2})(j-\frac{1}{2})}\right]^{\frac{1}{2}}|j-1, m \pm 1\rangle,
\end{aligned}
$$

where the generators of the Lorentz group are given by **J** and **F** and $j$, $h$, and $m$ have the same meaning as in Eq. (42). These irreducible representations of the Lorentz group become unitary:

(i) if $c$ is purely imaginary and $h$ is an arbitrary nonnegative integral or half-integral number,

(ii) if $c$ is a real number in the interval $0 \leq c \leq 1$ and $h = 0$.

The unitary irreducible representations belonging to (i) are known as the principal series of representations and those belonging to (ii) are known as the supplementary series of representations. When we let $c = 0$ in Eq. (52) we get precisely Eq. (51) giving the matrix elements of $K_\pm$ and $K_3$ and the matrix elements of $J_\pm$ and $J_3$ are the same in both the cases. *Thus, the Lorentz group $O(3, 1)$ obtained by the "extension" gives only those unitary irreducible representations for which $c$ is zero.* This result could also be easily obtained just by operating the Casimir operators of $O(3, 1)$ on the basis vectors $|j, m\rangle$. The Casimir operators for this group are

$$C_1 = \mathbf{J}^2 - \mathbf{K}^2 \qquad (53)$$

and

$$C_2 = \mathbf{J} \cdot \mathbf{K}. \qquad (54)$$

From the definition of **K** it is easily seen that

$$C_2 = (1/2\mu)[\mathbf{J} \cdot (\mathbf{P} \times \mathbf{J}) - \mathbf{J} \cdot (\mathbf{J} \times \mathbf{P})] = 0 \qquad (55)$$

and

$$C_1 = \mathbf{J}^2 - \{\mathbf{J}^2(\mathbf{P}^2/\mu^2) + (\mathbf{P}^2/\mu^2) - [(\mathbf{J} \cdot \mathbf{P})^2/\mu^2]\}. \qquad (56)$$

Now by making use of Eqs. (39), (43), and (49) one gets

$$\langle j, m|\,C_1\,|j, m\rangle = h^2 - 1. \qquad (57)$$

The structure of the generators defining the Lie algebra of $O(3, 1)$ is such that $C_2$ becomes identically zero and $C_1$ takes the value $(h^2 - 1)$. To understand what unitary irreducible representations are obtained by "extension," we should compare the values of $C_1$ and $C_2$ with that of Neumark. Then we get

$$\begin{aligned} C_1 &= h^2 - 1 = h^2 + c^2 - 1,\\ C_2 &= 0 = ihc. \end{aligned} \qquad (58)$$

Thus, it is seen once again that the unitary irreducible representations obtained here are those for which $c = 0$.

## V. REPRESENTATIONS OF $O(4, 1)$ FROM THE POINCARÉ GROUP $IO(3, 1)$

In this section we shall construct a Lie algebra basis for the de Sitter group $O(4, 1)$ from the Poincaré group[3] and then work out what unitary irreducible representations are obtainable for the de Sitter group from the Poincaré group by the "extension" procedure. The Poincaré group is generated by the rotation and translation generators $M_{\mu\nu}$ and $p_\mu$, respectively, and they obey the Lie algebra

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\delta_{\mu\rho}M_{\nu\sigma} + \delta_{\mu\sigma}M_{\rho\nu} - \delta_{\rho\nu}M_{\mu\sigma} - \delta_{\nu\sigma}M_{\rho\mu}) \qquad (59)$$

and

$$[M_{\mu\nu}, p_\rho] = i(\delta_{\mu\rho}p_\nu - \delta_{\nu\rho}p_\mu),$$
$$[p_\mu, p_\nu] = 0, \tag{60}$$

where $\mu$, $\nu$, $\rho$, and $\sigma$ run from 1 to 4 and we use the Euclidean metric and thus the time components are purely imaginary. As is well known from the works of Wigner[13] and Bargmann,[14] the unitary irreducible representations of this group are characterized by the two Casimir operators

$$C_1 = p_\mu p_\mu; \quad C_1 |m, j\rangle = -m^2 |m, j\rangle \tag{61}$$

and

$$C_2 = w_\mu w_\mu; \quad C_2 |m, j\rangle = j(j+1) |m, j\rangle, \tag{62}$$

where $w_\mu$ is the Pauli–Lubanski operator given by

$$w_\mu = -(i/2m)\epsilon_{\mu\nu\rho\sigma}M_{\nu\rho}p_\sigma. \tag{63}$$

As had already been noted in Ref. 3, the basis for de Sitter algebra is given by $M_{\mu\nu}$ and

$$X_\mu = (1/2m)[M_{\mu\nu}, p_\nu]_+. \tag{64}$$

It is easily seen from Eqs. (59) and (60) that

$$[X_\mu, X_\nu] = -iM_{\mu\nu} \tag{65}$$

and

$$[M_{\mu\nu}, X_\rho] = i(\delta_{\mu\rho}X_\nu - \delta_{\nu\rho}X_\mu). \tag{66}$$

Thus, Eqs. (59), (65), and (66) constitute a basis for the de Sitter algebra $O(4, 1)$.

To find out what unitary irreducible representations of $O(4, 1)$ are produced from the Poincaré group by using this particular basis $M_{\mu\nu}$ and $X_\mu$, we can start, as we did earlier in the case of $O(3, 1)$, either by explicitly writing down the matrix element of each generator or by directly looking at the structure of the Casimir operators and operating them on the basis vectors of the original group. In this case we shall adopt the second method for simplicity. There are two Casimir operators for this group and are given by[15,16]

$$C_1 = \tfrac{1}{2}M_{a'b'}M_{a'b'} \tag{67}$$

and

$$C_2 = W_{a'}W_{a'}, \tag{68}$$

where $a'$ and $b'$ run from 1 to 5;

$$M_{a'b'} = M_{\mu\nu} \quad \text{where } a \text{ and } b \text{ run from 1 to 4},$$
$$M_{\mu 5} = X_\mu = -M_{5\mu} \tag{69}$$

and

$$W_{a'} = -(i/4)\epsilon_{a'b'c'd'e'}M_{b'c'}M_{d'e'}. \tag{70}$$

Equation (70) is the five-dimensional analog of the Pauli–Lubanski operator and $\epsilon_{a'b'c'd'e'}$ is the five-dimensional completely antisymmetric tensor. To evaluate the Casimir operators $C_1$ and $C_2$ on the basis vectors of the Poincaré group, one can take advantage of Wigner's little-group technique. From our definition of $X_\mu$ we assumed the existence of a rest system and all our generators are covariantly defined. Then the easiest way to evaluate $C_1$ and $C_2$ is to compute the same in the rest frame by using the appropriate $M_{\mu\nu}$ and $p_\mu$. Let us denote $M_{\mu\nu}$ by $M^{(0)}_{\mu\nu}$ in the rest system and then write

$$M^{(0)}_{ij} = \epsilon_{ijk}J_k, \tag{71}$$

where $J_k$ is the spin angular momentum and

$$M^{(0)}_{i4} = iK_i. \tag{72}$$

Since $p^{(0)}_\mu = (0, 0, 0, im)$ and

$$X^{(0)}_\mu = (1/2m)[M^{(0)}_{\mu\nu}, p^{(0)}_\nu]_+,$$

we explicitly get

$$X^{(0)}_i = -K_i,$$
$$X^{(0)}_4 = 0. \tag{73}$$

Now writing manifestly $C_1$ and $C_2$ in terms of the rest-system quantities we get

$$C_1 = \tfrac{1}{2}M^{(0)}_{ij}M^{(0)}_{ij} + M^{(0)}_{i4}M^{(0)}_{i4} + X^{(0)}_i X^{(0)}_i$$
$$= \mathbf{J}\cdot\mathbf{J} - \mathbf{K}\cdot\mathbf{K} + \mathbf{K}\cdot\mathbf{K} = \mathbf{J}\cdot\mathbf{J} \tag{74}$$

and

$$\begin{aligned} C_2 &= (M^{(0)}_{12}X^{(0)}_3 + M^{(0)}_{31}X^{(0)}_2 + M^{(0)}_{23}X^{(0)}_1)^2 \\ &\quad + (M^{(0)}_{24}X^{(0)}_3 - M^{(0)}_{34}X^{(0)}_2)^2 \\ &\quad + (M^{(0)}_{14}X^{(0)}_3 - M^{(0)}_{34}X^{(0)}_1)^2 \\ &\quad + (M^{(0)}_{14}X^{(0)}_2 - M^{(0)}_{24}X^{(0)}_1)^2 \\ &\quad + (M^{(0)}_{14}M^{(0)}_{23} + M^{(0)}_{24}M^{(0)}_{31} + M^{(0)}_{34}M^{(0)}_{12})^2 \\ &= (\mathbf{J}\cdot\mathbf{K})^2 + \mathbf{J}\cdot\mathbf{J} - (\mathbf{J}\cdot\mathbf{K})^2 = \mathbf{J}\cdot\mathbf{J}. \end{aligned} \tag{75}$$

Thus, when $C_1$ and $C_2$ act on the states $|m, j\rangle$ it results

$$C_1 |m, j\rangle = C_2 |m, j\rangle = j(j+1) |m, j\rangle, \tag{76}$$

where $j$ is the angular-momentum eigenvalue. Thus, once again, we get only a fraction of the possible unitary irreducible representations of the $O(4, 1)$ group.[15,16]

In general it is not clear how one can construct unitary irreducible representations of a noncompact group from the noncompact subgroup of the original group. This problem is important especially when the quantum numbers of the maximal compact subgroup of the noncompact group do not suffice to completely label the basis vectors of a unitary irreducible representation of the whole noncompact group. Recently, this problem has been attacked by Mukunda for the cases $O(2, 1)$ and $O(3, 1)$ from $O(1, 1)$ and $O(2, 1)$, respectively.[17]

In the case considered here, where the homogeneous part of the Poincaré group is the noncompact group $O(3, 1)$, the above-mentioned problem comes to the fore. From the definition of the generators $X_\mu$ and

[13] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[14] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1948).
[15] L. H. Thomas, Ann. Math. **42**, 113 (1941).
[16] T. D. Newton, Ann. Math. **51**, 730 (1950).
[17] N. Mukunda, Syracuse University report No. NYO-3399-103, 106.

their structures both for negative and positive energies, it is obvious that the Casimir operator values [Eqs. (74) and (75)] do not change. Thus, every proper unitary irreducible representation of the de Sitter group $O(4, 1)$ obtained here is perhaps built from the representations of the Poincaré group corresponding to both positive and negative energies.

## VI. CONCLUSION

We have given a prescription to construct the homogeneous Lorentz group in $(n + 1)$ dimensions from an inhomogeneous rotation group in $n$ dimensions. Then this prescription of constructing a new basis for the homogeneous Lorentz group has been utilized to study the unitary irreducible representations of the groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$. It is found out that in the $O(2, 1)$ case one gets only those unitary irreducible representations corresponding to the Casimir-operator value $C = -\frac{1}{4}$ ($D^{\pm}_{\frac{1}{2}}$ and $C^{0}_{\frac{1}{4}}$ in Bargmann's notation) and in the $O(3, 1)$ case one gets unitary irreducible representations corresponding to the Casimir operator $C_2 = 0$ and $C_1 = h^2 - 1$.[12] The "extension" technique gives only those unitary irreducible representations corresponding to the Casimir-operator values $C_1 = j(j + 1) = C_2$ for the de Sitter group $O(4, 1)$. Even though we are unable to generalize our result to the case $O(n, 1)$, it is expected that one does not get all the unitary irreducible representations by the "extension" procedure. Nevertheless, it is relevant to point not that the representations obtained this way correspond to physically interesting cases and if one is interested in limiting the number of representation for a particular group by defining the generators in a special way (this is the case for the noninvariance groups), the above procedure is useful.[18] Anyway, for groups of the type considered here one can use the "master analytic representation" to obtain the representations by exploiting the fact that the Lie algebras involved have the same complex Lie algebras.[19]

It is to be mentioned that by defining generators of the type given by Eq. (12)

$$X_a = \lambda p_a + [1/2(p_c p_c)^{\frac{1}{2}}][M_{ab}, p_b]_+,$$

with a variable parameter $\lambda$, one can obtain a much wider class of representations than those obtained here for the groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$.[20] But, in the foregoing discussion $\lambda$ is chosen to be zero so as to maintain the time-reversal invariance. Thus, it transpires that this time-reversal invariance is responsible for eliminating the other unobtained representations. If one is primarily interested in obtaining the representations of the group $O(n, 1)$ from $IO(n)$, one can then use the above $X_a$ with variable $\lambda$ without worrying about the time-reversal invariance, but one has to impose this condition if $X_a$ is to be identified with some physical operator.[3]

## ACKNOWLEDGMENT

The author wishes to thank the referee for his valuable suggestions.

[18] N. Mukunda, L. S. O'Raifeartaigh, and E. C. G. Sudarshan Phys. Letters **15**, 1041 (1965).

[19] E. C. G. Sudarshan (private communication).

[20] This point has been particularly drawn to the attention of the author by E. C. G. Sudarshan.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Screening in the Schrödinger Theory of Scattering

JOHN D. DOLLARD*
*Department of Mathematics, University of Rochester, Rochester, New York*

(Received 19 June 1967)

The time-dependent Schrödinger theory of scattering is studied rigorously with the potential $V(\mathbf{x})$ replaced by $e^{-\epsilon|\mathbf{x}|}V(\mathbf{x})$. Sufficient conditions are given that the Møller wave matrices $\Omega\pm$ be obtainable from this theory as $\epsilon \to 0$. The conjecture that this theory can be used to define a reasonable $S$ matrix when the $\Omega\pm$ do not exist is false for the Coulomb potential.

## INTRODUCTION

The time dependent Schrödinger theory of scattering takes on its simplest form if the potentials occurring in the Hamiltonian are "short range" in character. In the case of "short-range" potentials, one has available[1] theorems which guarantee the existence of the Møller wave matrices, defined in the usual way.

* The author wishes to acknowledge the support during this work of NSF Grant GP 7453.

[1] J. M. Cook, J. Math. & Phys. **36**, 82 (1957); J. M. Jauch and I. Zinnes, Nuovo Cimento **II**, 553 (1959); M. N. Hack, Nuovo Cimento **13**, 231 (1959).

their structures both for negative and positive energies, it is obvious that the Casimir operator values [Eqs. (74) and (75)] do not change. Thus, every proper unitary irreducible representation of the de Sitter group $O(4, 1)$ obtained here is perhaps built from the representations of the Poincaré group corresponding to both positive and negative energies.

## VI. CONCLUSION

We have given a prescription to construct the homogeneous Lorentz group in $(n + 1)$ dimensions from an inhomogeneous rotation group in $n$ dimensions. Then this prescription of constructing a new basis for the homogeneous Lorentz group has been utilized to study the unitary irreducible representations of the groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$. It is found out that in the $O(2, 1)$ case one gets only those unitary irreducible representations corresponding to the Casimir-operator value $C = -\frac{1}{4}$ ($D^{\pm}_{\frac{1}{2}}$ and $C^{0}_{\frac{1}{4}}$ in Bargmann's notation) and in the $O(3, 1)$ case one gets unitary irreducible representations corresponding to the Casimir operator $C_2 = 0$ and $C_1 = h^2 - 1$.[12] The "extension" technique gives only those unitary irreducible representations corresponding to the Casimir-operator values $C_1 = j(j + 1) = C_2$ for the de Sitter group $O(4, 1)$. Even though we are unable to generalize our result to the case $O(n, 1)$, it is expected that one does not get all the unitary irreducible representations by the "extension" procedure. Nevertheless, it is relevant to point not that the representations obtained this way correspond to physically interesting cases and if one is interested in limiting the number of representation for a particular group by defining the generators in a special way (this is the case for the noninvariance groups), the above procedure is useful.[18] Anyway, for groups of the type considered here one can use the "master analytic representation" to obtain the representations by exploiting the fact that the Lie algebras involved have the same complex Lie algebras.[19]

It is to be mentioned that by defining generators of the type given by Eq. (12)

$$X_a = \lambda p_a + [1/2(p_c p_c)^{\frac{1}{2}}][M_{ab}, p_b]_+,$$

with a variable parameter $\lambda$, one can obtain a much wider class of representations than those obtained here for the groups $O(2, 1)$, $O(3, 1)$, and $O(4, 1)$.[20] But, in the foregoing discussion $\lambda$ is chosen to be zero so as to maintain the time-reversal invariance. Thus, it transpires that this time-reversal invariance is responsible for eliminating the other unobtained representations. If one is primarily interested in obtaining the representations of the group $O(n, 1)$ from $IO(n)$, one can then use the above $X_a$ with variable $\lambda$ without worrying about the time-reversal invariance, but one has to impose this condition if $X_a$ is to be identified with some physical operator.[3]

## ACKNOWLEDGMENT

The author wishes to thank the referee for his valuable suggestions.

[18] N. Mukunda, L. S. O'Raifeartaigh, and E. C. G. Sudarshan Phys. Letters **15**, 1041 (1965).

[19] E. C. G. Sudarshan (private communication).

[20] This point has been particularly drawn to the attention of the author by E. C. G. Sudarshan.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Screening in the Schrödinger Theory of Scattering

JOHN D. DOLLARD*
*Department of Mathematics, University of Rochester, Rochester, New York*

(Received 19 June 1967)

The time-dependent Schrödinger theory of scattering is studied rigorously with the potential $V(\mathbf{x})$ replaced by $e^{-\epsilon|\mathbf{x}|}V(\mathbf{x})$. Sufficient conditions are given that the Møller wave matrices $\Omega\pm$ be obtainable from this theory as $\epsilon \to 0$. The conjecture that this theory can be used to define a reasonable $S$ matrix when the $\Omega\pm$ do not exist is false for the Coulomb potential.

## INTRODUCTION

The time dependent Schrödinger theory of scattering takes on its simplest form if the potentials occurring in the Hamiltonian are "short range" in character. In the case of "short-range" potentials, one has available[1] theorems which guarantee the existence of the Møller wave matrices, defined in the usual way.

* The author wishes to acknowledge the support during this work of NSF Grant GP 7453.

[1] J. M. Cook, J. Math. & Phys. **36**, 82 (1957); J. M. Jauch and I. Zinnes, Nuovo Cimento **II**, 553 (1959); M. N. Hack, Nuovo Cimento **13**, 231 (1959).

For a "long-range" potential such as the Coulomb potential, one can still[2] give a presentable scattering theory, but the theorems involved are more complicated, because the usual method of defining the Møller wave matrices does not work. In these circumstances, it is attractive to conjecture that one might be able to avoid the complications due to "long-range" potentials by considering the theory in some sense as the limit of theories in which the interactions are better behaved. In a previous paper[3] we studied the method of adiabatic switching in this context, and concluded that it was not possible to view the theory of Coulomb scattering as the limit of adiabatically switched Coulomb-scattering theories, and that therefore the usual method of adiabatic switching was unlikely to produce useful results for other long-range potentials. In the present paper, which may be considered as a sequel to Ref. 3, we will consider the problem of viewing a nonrelativistic scattering theory as the limit of "screened" scattering theories. The "screened" scattering theories are obtained by introducing in the original theory a screening function, or spatial cutoff, in order to obtain a better-behaved interaction. Since similar procedures are of interest in quantum field theory, it is hoped that the results obtained here will shed some light on those procedures at the same time that they give insight into the structure of nonrelativistic scattering theory.

In detail, the procedure to be investigated in this paper is the following:

Given the Hamiltonian

$$H = H_0 + V(\mathbf{x}) = -\Delta/2m + V(\mathbf{x}) \tag{1}$$

($\Delta$ is the Laplacean, $V$ a real multiplicative operator), define a scattering matrix as follows:

(a) Replace $H$ by the screened Hamiltonian

$$H_\epsilon = H_0 + e^{-\epsilon|\mathbf{x}|}V(\mathbf{x}) = H_0 + V_\epsilon(\mathbf{x}). \tag{2}$$

(b) Solve the Schrödinger equation ($\hbar = 1$)

$$i\partial\psi(t)/\partial t = H_\epsilon\psi(t), \tag{3}$$

by writing

$$\psi(t) = e^{-iH_\epsilon t}\psi(0). \tag{4}$$

(c) Form the operators

$$\Omega_\epsilon(t) = e^{iH_\epsilon t}e^{-iH_0 t} \tag{5}$$

and prove the existence of the "screened Møller wave matrices"

$$\Omega_\epsilon^\pm = \lim_{t\to\pm\infty} \Omega_\epsilon(t). \tag{6}$$

(d) Prove the existence of the operators

$$\Omega_0^\pm = \lim_{\epsilon\to 0} \Omega_\epsilon^\pm \tag{7}$$

and use them to define the scattering matrix

$$S_0 = (\Omega_0^+)^*\Omega_0^-, \tag{8}$$

or alternatively, if it is suspected that the limits in (7) do not exist,

(d′) Define the screened $S$ matrix $S_\epsilon$ by

$$S_\epsilon = (\Omega_\epsilon^+)^*\Omega_\epsilon^- \tag{9}$$

and prove that the limit

$$S_0' = \lim_{\epsilon\to 0} S_\epsilon \tag{10}$$

exists. Then define the scattering matrix as $S_0'$. [Naturally, if the strong limits in (6) exist, then the limit in (10) exists and $S_0 = S_0'$].

This procedure is offered as a substitute for the usual method, which is as follows:

(a) Solve the Schrödinger equation without switching

$$i\partial\psi(t)/\partial t = H\psi(t), \tag{11}$$

by writing

$$\psi(t) = e^{-iHt}\psi(0). \tag{12}$$

(b) Form the operator

$$\Omega(t) = e^{iHt}e^{-iH_0 t}. \tag{13}$$

(c) Prove the existence of the operators

$$\Omega^\pm = \lim_{t\to\pm\infty} \Omega(t) \tag{14}$$

and use them to define the scattering matrix

$$S = (\Omega^+)^*\Omega^-. \tag{15}$$

It is hoped that:

(1) When $\Omega^\pm$ of (14) exists, then $\Omega_0^\pm$ of (7) also exists, and

$$\Omega^\pm = \Omega_0^\pm,$$

so that the screened theory yields the usual Møller wave matrices as $\epsilon \to 0$.

(2) When $\Omega^\pm$ of (14) does not exist, then $\Omega_0^\pm$ of (7) or at least $S_0'$ of (14) can still be defined, and used to define a reasonable $S$ matrix for the theory.

In the following, we investigate the method of screening for potential scattering as outlined above, and the truth of statements (1) and (2). In passing, we shall mention some results on $n$-body scattering.

Before proceeding with the analysis, the author would like to note that, although the results proved below concern the screening procedure in which $V(\mathbf{x})$ is replaced by $e^{-\epsilon|\mathbf{x}|}V(\mathbf{x})$, they can easily be extended to other types of screening procedures. For instance, if $g$ is a real function of $|\mathbf{x}|$ which belongs to Schwartz's space[4] $S$ of $C^\infty$ functions which, along

[2] J. Dollard, J. Math. Phys. 5, 729 (1964).
[3] J. Dollard, J. Math. Phys. 7, 802 (1966).
[4] L. Schwartz, *Théorie des distributions* (Hermann and Cie., Paris, 1957), Vol. II, pp. 89ff.

with all their derivatives, die off for large values of $|\mathbf{x}|$ faster than any inverse power of $|\mathbf{x}|$, and if $g(0) = 1$, then all the results proved below hold for the screening procedure in which $V(\mathbf{x})$ is replaced by $g(\epsilon\,|\mathbf{x}|)V(\mathbf{x})$. [In this case, it would be necessary to replace $e^{-\epsilon(s/m)}(-\Delta)^{\frac{1}{2}}$ by $g[(\epsilon s/m)(-\Delta)^{\frac{1}{2}}]$ in Eq. (37).] Some examples of such functions $g$ are: $g(|\mathbf{x}|) = e^{-|\mathbf{x}|^2}$, $g(|\mathbf{x}|) = \cos|\mathbf{x}|^2\, e^{-|\mathbf{x}|^2}$.

## I. EXISTENCE OF THE SCREENED MØLLER WAVE MATRICES

*Orientation:* We work in the Hilbert space $\mathfrak{L}^2(R^3)$ of square-integrable functions of one three-vector variable $\mathbf{x}$. The $\mathfrak{L}^2$ norm of the function $f$ will be denoted by $\|f\|$. We consider a Hamiltonian of the type (1) acting on this space, where $V$ satisfies either or both of *Kato's conditions*[5]:

($K_1$) $V(\mathbf{x})$ is square-integrable, or

($K_2$) $V(\mathbf{x})$ is locally square-integrable and bounded for large $|\mathbf{x}|$; i.e., the integral of $|V(\mathbf{x})|^2$ over any bounded set is finite, and there exist constants $M$ and $R$ such that $|V(\mathbf{x})| \leq M$ when $|\mathbf{x}| \geq R$.

Denote by $\mathfrak{D}(\mathcal{O})$ the domain of the operator $\mathcal{O}$. Then if $V$ satisfies ($K_1$) or ($K_2$), it follows from Kato's work that $\mathfrak{D}(V) \supseteq \mathfrak{D}(H_0)$ and that there exist non-negative constants $\alpha$ and $\beta$ such that if $f \in \mathfrak{D}(H_0)$, then

$$\|Vf\| \leq \alpha\,\|H_0 f\| + \beta\,\|f\|. \tag{16a}$$

Also, $\alpha$ can be chosen as small as desired. It follows that $H$ is self-adjoint with $\mathfrak{D}(H) = \mathfrak{D}(H_0)$. After a little juggling of (16a), one concludes that there exist constants $\alpha'$, $\beta'$, ($\alpha'$ as small as desired) such that if $f \in \mathfrak{D}(H_0)$,

$$\|Vf\| \leq \alpha'\,\|Hf\| + \beta'\,\|f\|. \tag{16b}$$

In the sequel, we shall make use of the fact that $\mathfrak{D}(H_0)$ contains Schwartz's space[4] $\mathcal{S}$, mentioned above.

By inspection, if $V(\mathbf{x})$ satisfies ($K_1$) or ($K_2$), then $e^{-\epsilon|\mathbf{x}|}V(\mathbf{x})$ satisfies ($K_1$). This fact establishes the existence of the switched Møller wave matrices, because it is known[1] that the (usual) Møller wave matrices exist for potentials satisfying ($K_1$). For future reference, we mention that if $V$ satisfies ($K_1$) or ($K_2$), then in order to establish the existence of the (strong) limits in (14), it suffices to prove the convergence of the integrals ($t_0$ and $t_0'$ are real numbers)

$$\int_{t_0}^{\infty} \|Ve^{-iH_0 t}\varphi\|\,dt, \quad \int_{-\infty}^{t_0} \|Ve^{-iH_0 t}\varphi\|\,dt \tag{17}$$

for all $\varphi$ belonging to a set dense in $\mathfrak{L}^2$. This set may be conveniently chosen as Schwartz's space $\mathcal{S}$. [If $\varphi \in \mathcal{S}$, then $e^{-iH_0 t}\varphi \in \mathcal{S} \subseteq \mathfrak{D}(H_0) \subseteq \mathfrak{D}(V)$, so the integrands are well defined.] The convergence of the first integral in (17) establishes the existence of $\Omega^+$, and one has

$$\|(\Omega^+ - \Omega(t))\varphi\| = \left\| \int_t^{\infty} \frac{\partial}{\partial t'}(\Omega(t')\varphi)\,dt' \right\|$$

$$= \left\| i\int_t^{\infty} e^{iHt'} V e^{-iH_0 t'}\varphi\,dt' \right\| \leq \int_t^{\infty} \|Ve^{-iH_0 t'}\varphi\|\,dt' \tag{18}$$

and the similar equation for $\Omega^-$. The right-hand side of (18) approaches 0 as $t \to +\infty$. If $V$ satisfies ($K_1$), the integrals in (17) always converge, as was shown by Cook.[1]

*Results in the n-body case:* In the $n$-body case, we need only redefine $H_0$ and $V$ by

$$H_0 = \sum_{k=1}^{n} -\frac{\Delta_k}{2m_k},$$

$$V = \sum_{j=1}^{n} V_{0j}(\mathbf{x}_j) + \sum_{1 \leq i < j \leq n} V_{ij}(\mathbf{x}_i - \mathbf{x}_j), \tag{19}$$

and assume that each $V_{ij}(0 \leq i < j \leq n)$ satisfies ($K_1$) or ($K_2$). Naturally, we also replace $\mathfrak{L}^2$ in one three-vector variable by $\mathfrak{L}^2$ in the $n$ three-vector variables $\mathbf{x}_1 \cdots \mathbf{x}_n$. Then defining

$$H_\epsilon = H_0 + \sum_{j=1}^{n} e^{-\epsilon|\mathbf{x}_j|} V_{0j}(\mathbf{x}_j) + \sum_{1 \leq i < j \leq n} e^{-\epsilon|\mathbf{x}_i - \mathbf{x}_j|} V_{ij}(\mathbf{x}_i - \mathbf{x}_j), \tag{20}$$

we find that the limits corresponding to (6) (and their analogs for channels other than the channel in which all particles are asymptotically free) exist.

Having established the existence of the switched Møller wave matrices for all potentials satisfying ($K_1$) or ($K_2$), we now turn to the more difficult question of the existence of the limit in Eq. (7).

## II. CONVERGENCE OF $\Omega_\epsilon^\pm$ AS $\epsilon \to 0$

In this section we do two things. First, for a large class of theories in which the limit $\Omega^\pm$ in (14) is known to exist, we show that the limit $\Omega^\pm$ in (7) also exists and equals $\Omega^\pm$. In other words, we show that in a large class of theories in which the usual methods of scattering theory apply, the method of screening is also applicable and gives the same results. We then show that the method of screening is *not* applicable if $V(\mathbf{x})$ is a Coulomb potential.

In all cases with which the author is familiar[1] in which the limit (14) is known to exist, the potentials involved have satisfied either ($K_1$) or ($K_2$) and the existence of the limit in (14) has been or can be

[5] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).

proved by establishing the convergence of the integrals in (17). [It has been remarked that $(K_1)$ by itself implies the convergence of the integrals in (17). However, there are potentials not satisfying $(K_1)$ for which the integrals in (17) converge and $\Omega^{\pm}$ exist, as shown by Jauch and Zinnes.[1] In fact, convergence can be proved for any potential satisfying $(K_2)$ which falls off like $|\mathbf{x}|^{\gamma}$ for large $|\mathbf{x}|$, with $\gamma < -1$. Note that this requirement excludes Coulomb potentials, for which we would have $\gamma = -1$.] We now show that for any potential satisfying $(K_1)$ or $(K_2)$ and for which the integrals in (17) converge, the method of screening is applicable; namely we have:

*Theorem 1*: Suppose that $V(\mathbf{x})$ satisfies $(K_1)$ or $(K_2)$. Suppose that the integrals in (17) converge for all $\varphi \in \mathcal{S}$. Let $V_\epsilon(\mathbf{x}) = e^{-\epsilon|\mathbf{x}|}V(\mathbf{x})$ and define $H$, $H_\epsilon$, $\Omega_\epsilon(t)$, $\Omega(t)$ as in (1), (2), (5), and (13). Then the strong limits

$$\lim_{t\to\pm\infty} \Omega(t) = \Omega^{\pm} \tag{21}$$

and

$$\lim_{t\to\pm\infty} \Omega_\epsilon(t) = \Omega_\epsilon^{\pm} \tag{22}$$

exist, and furthermore we have

$$\lim_{\epsilon\to 0} \Omega_\epsilon^{\pm} = \Omega^{\pm}. \tag{23}$$

*Proof:* We have already noted that under the hypotheses of the theorem the strong limits in (21) and (22) exist. (As the strong limits of unitary operators, $\Omega^{\pm}$ and $\Omega_\epsilon^{\pm}$ are, of course, isometric.) It remains to prove (23). In order to do this we first remark that the convergence in (22) is uniform in $\epsilon$. This can be established as follows: We have, for $\varphi \in \mathcal{S}$,

$$\|(\Omega_\epsilon^{+} - \Omega_\epsilon(t))\varphi\| \leq \int_t^{\infty} \|V_\epsilon e^{-iH_0 t}\varphi\|\, dt \leq \int_t^{\infty} \|Ve^{-iH_0 t}\varphi\|\, dt, \tag{24}$$

since $|V_\epsilon(\mathbf{x})| = e^{-\epsilon|\mathbf{x}|}\,|V(\mathbf{x})| \leq |V(\mathbf{x})|$. Now the right-hand side of (24) is clearly independent of $\epsilon$, and approaches 0 as $t \to +\infty$ since the first integral in (17) converges. Thus by taking $t$ large enough, the left-hand side of (24) can be made as small as desired, independent of $\epsilon$. Since $\mathcal{S}$ is dense in $\mathfrak{L}^2$, $\Omega_\epsilon^{+}$ is isometric and $\Omega(t)$ is unitary, this result extends immediately to any $\varphi \in \mathfrak{L}^2$: For all $\varphi \in \mathfrak{L}^2$, $\Omega_\epsilon(t)\varphi$ converges to $\Omega_\epsilon^{+}\varphi$ as $t \to +\infty$, uniformly in $\epsilon$. The proof for uniform convergence to $\Omega_\epsilon^{+}$ as $t \to -\infty$ is no different.

We next show that for any fixed $t$ we have, in the sense of strong convergence,

$$\lim_{\epsilon\to 0} \Omega_\epsilon(t) = \Omega(t). \tag{25}$$

In order to prove (25), it suffices to show that

$$\lim_{\epsilon\to 0} e^{iH_\epsilon t} = e^{iHt}, \tag{26}$$

and since the operators in (26) are unitary, we need only show that (26) holds on $\mathcal{S}$. If $\varphi \in \mathcal{S}$ we have

$$\begin{aligned}\|(e^{iH_\epsilon t} - e^{iHt})\varphi\| &= \|(1 - e^{-iH_\epsilon t}e^{iHt})\varphi\| \\ &= \left\| -\int_0^t \frac{\partial}{\partial t'}(e^{-iH_\epsilon t'}e^{iHt'}\varphi)\, dt' \right\| \\ &= \left\| i\int_0^t e^{-H_\epsilon t'}(V_\epsilon - V)e^{iHt'}\varphi\, dt' \right\| \\ &\leq \int_0^t \|(V_\epsilon - V)e^{iHt'}\varphi\|\, dt'. \end{aligned} \tag{27}$$

Now the expression $\|(V_\epsilon - V)e^{iHt'}\varphi\|$ is bounded in $t$. This can be seen using (16b) as follows:

$$\begin{aligned}\|(V_\epsilon - V)e^{iHt'}\varphi\| &\leq \|V_\epsilon e^{iHt'}\varphi\| + \|Ve^{iHt'}\varphi\| \\ &\leq 2\,\|Ve^{iHt'}\varphi\| \leq 2(\alpha'\,\|He^{iHt'}\varphi\| + \beta'\,\|e^{iHt'}\varphi\|) \\ &= 2(\alpha'\,\|H\varphi\| + \beta'\,\|\varphi\|). \end{aligned} \tag{28}$$

Here we have also made use of the fact that $|V_\epsilon(\mathbf{x})| \leq |V(\mathbf{x})|$ and the fact that $\varphi \in \mathcal{S} \subseteq \mathfrak{D}(H)$. It should also be clear (e.g., from Lebesgue's dominated convergence theorem) that for each fixed value of $t'$ we have

$$\lim_{\epsilon\to 0} \|(V_\epsilon - V)e^{iHt'}\varphi\| = 0. \tag{29}$$

Because of (28) and (29) we can apply Lebesgue's dominated convergence theorem to the right-hand side of (27) to conclude that

$$\lim_{\epsilon\to 0} \int_0^t \|(V_\epsilon - V)e^{iHt'}\varphi\|\, dt' = 0, \tag{30}$$

completing the proof of (26) and thus of (25).

We now present the proof of (23). Let $\varphi \in \mathfrak{L}^2$ and let $\eta > 0$ be given. We must show that when $\epsilon$ is small enough we have

$$\|(\Omega_\epsilon^{\pm} - \Omega^{\pm})\varphi\| < \eta. \tag{31}$$

Since the argument for $\Omega^{-}$ is no different that that for $\Omega^{+}$, we present only the latter: For any $t$ we have

$$\begin{aligned}\|(\Omega_\epsilon^{+} - \Omega^{+})\varphi\| &\leq \|(\Omega_\epsilon^{+} - \Omega_\epsilon(t))\varphi\| \\ &\quad + \|(\Omega(t) - \Omega^{+})\varphi\| + \|(\Omega_\epsilon(t) - \Omega(t))\|. \end{aligned} \tag{32}$$

Since $\Omega(t)$ converges to $\Omega^{+}$ and $\Omega_\epsilon(t)$ converges uniformly in $\epsilon$ to $\Omega_\epsilon^{+}$ as $t \to +\infty$, we can choose a $t$ so large that each of the first two terms on the right-hand side of (32) is less than $\eta/3$, independent of the value of $\epsilon$. Fixing the value of $t$, we can then, by (25), choose $\epsilon$ so small that the third term on the right-hand side of (32) is less then $\eta/3$, and this establishes

$$\|(\Omega^{+} - \Omega_\epsilon^{+})\varphi\| < \eta \tag{33}$$

for small enough $\epsilon$, as desired. This finishes the proof of Theorem 1.

Thus we see that when the potential satisfies the hypotheses of Theorem 1, the screening procedure can be carried through and gives results identical with those of the usual theory, as was to be hoped. [Compare statement (1) of the Introduction.]

*Result in the n-body case:* Theorem 1 remains true in the $n$-body case if $H$ is replaced by $H_0 + V$ of (19), $H_\epsilon$ is given by (20), and the obvious replacements are made for $\Omega(t)$ and $\Omega_\epsilon(t)$ [e.g., $\Omega(t) = e^{iHt}e^{-iH_0t}$, with $H_0$ given by (19)]. Theorem 1 then states that the Møller wave matrix for the channel in which all particles are asymptotically free in the usual theory agrees with the limit as $\epsilon \to 0$ of the Møller wave matrix for the channel in which all particles are asymptotically free in the screened theory.

We now turn to an investigation of the screening procedure for the Coulomb potential, taking

$$H = -\Delta/2m + e_1e_2/|\mathbf{x}|. \tag{34}$$

The analysis of the screening procedure for this potential is quite analogous to the analysis of the adiabatic switching procedure for this potential given in Ref. 3. It is also rather messy, and consists principally of a rather tedious checking of details analogous to those given in Ref. 3 for the case of adiabatic switching. For this reason, the results of the analysis are given without proof.

Just as in Ref. 3, we introduce the matrices $\Omega_c^\pm$ which correctly give the Coulomb scattering matrix elements. These cannot be obtained as the limits of $\Omega(t)$ of (13) with $H$ given by (34), but they can be obtained by a different method, described in Refs. 2 and 3. Since the Coulomb potential satisfies $(K_2)$, the screened Møller wave matrices $\Omega_{c\epsilon}^\pm$ for the Coulomb potential exist, and we might hope that they converge, as $\epsilon \to 0$, to $\Omega_c^\pm$, or at least that the screened $S$ matrix converges to $(\Omega_c^+)^*\Omega_c^-$. However, this is not the case, as shown by:

*Theorem 2*: Let $\Omega_{c\epsilon}^\pm$ be the screened Møller wave matrices for the Coulomb potential $e_1e_2/|\mathbf{x}|$. Let

$$S_{c\epsilon} = (\Omega_{c\epsilon}^+)^*\Omega_{c\epsilon}^- \tag{35}$$

be the screened $S$ matrix for the Coulomb potential. Then $\Omega_{c\epsilon}^\pm$ and $S_{c\epsilon}$ converge weakly to zero as $\epsilon \to 0$.

Theorem 2 implies that $\Omega_{c\epsilon}^\pm$ and $S_{c\epsilon}$ do not converge strongly as $\epsilon \to 0$. For $\Omega_{c\epsilon}^\pm$ this is obvious, because the operators $\Omega_{c\epsilon}^\pm$ are clearly isometries and their strong limit, if it existed, would have to be an isometry. It would also have to equal their weak limit, which is 0, and not an isometry. Actually, the same argument can be applied to $S_{c\epsilon}$, because $S_{c\epsilon}$ is also an isometry, although this latter statement is not so obvious. We refer the reader who wishes to convince himself that $S_{c\epsilon}$ is isometric to the work of Ikebe[6] who has given a general discussion of potential-scattering theory which covers the case at hand.

Theorem 2 also shows that the weak limits of $\Omega_{c\epsilon}^\pm$ and $S_{c\epsilon}$ are totally useless in defining an $S$ matrix for the Coulomb-scattering problem, so that the approach to this problem via the method of screening outlined in the Introduction is fruitless.

For the interested reader, we include a brief account of the behavior of $\Omega_{c\epsilon}^\pm$ when $\epsilon$ is small: As in the similar situation in Ref. 3, the difficulty with these operators is that they reproduce too faithfully the logarithmic distortion caused by the long range of the Coulomb potential. Namely, as $\epsilon \to 0$, the difference between $\Omega_{c\epsilon}^\pm$ and $\Omega_c^\pm P_\epsilon^\pm$ tends (strongly) to zero, where

$$P_\epsilon^\pm = e^{\pm iL_\epsilon} \tag{36}$$

and

$$L_\epsilon = \frac{me_1e_2}{(-\Delta)^{\frac{1}{2}}}\int_{t_0}^{\infty} \frac{e^{-\epsilon(s/m)}(-\Delta)^{\frac{1}{2}}}{s}\,ds + \frac{me_1e_2}{(-\Delta)^{\frac{1}{2}}}\log\left(\frac{-2t_0\Delta}{m}\right), \tag{37}$$

where $t_0 > 0$ and $\Delta$ is the Laplacean. As $\epsilon \to 0$, the integral in the definition of $L_\epsilon$ diverges logarithmically, and $P_\epsilon^\pm$ "oscillates itself to death," so to speak.

In conclusion, then, we have shown that for a large class of potentials for which the usual methods of scattering theory are applicable, the method of screening is also applicable and gives correct results. However, for the Coulomb potential the method produces no information. As discussed in Ref. 3, the Coulomb potential is one of the "mildest" possible long-range potentials. The fact that the method of screening breaks down for the Coulomb potential makes it seem unlikely that this method can be used to produce reasonable results for other long-range potentials.

[6] T. Ikebe, Arch. Ratl. Mech. Anal. 5, 1 (1960).

# Note on the Geometric Theory of Neutrinos

C. D. COLLINSON
*Department of Applied Mathematics, University of Hull, Hull, England*

(Received 8 October 1967)

The geometric theory of neutrinos proposed by Penney is investigated. It is shown, from the conditions imposed, that the space–time is conformally flat. All solutions of the field equations are found and attention is paid to the limiting process described by Penney.

## 1. INTRODUCTION

Penny[1] has shown that any space–time which satisfies the usual Einstein–Maxwell field equations for a *null* electromagnetic field, together with the condition

$$R_{ab;c} = 0, \tag{1.1}$$

reproduces the physics of a neutrino field. Here $R_{ab}$ is the Ricci tensor of the space-time and the semicolon denotes covariant differentiation. Such fields are characterized by Penney as the limit, as $R^{cd}R_{cd} \to 0$, of the algebraically general electromagnetic fields satisfying

$$R_{ab;c} = (R_{fg}R^{fg})_{;c}R_{ab}. \tag{1.2}$$

The usual Rainich[2] conditions are imposed on the space-time, namely,

$$R^{e}{}_{c} = 0, \tag{1.3}$$

$$R_{f}{}^{a}R^{f}{}_{b} = \tfrac{1}{4}\delta_{b}{}^{a}R_{fg}R^{fg}, \tag{1.4}$$

$$R_{00} \geq 0, \tag{1.5}$$

where the metric tensor takes the diagonal form $(+1, -1, -1, -1)$ in a local Cartesian coordinate system.

It is shown, in Sec. 2, that the conditions (1.2) to (1.5) imply that the space–time is conformally flat. All space–times satisfying the conditions are found in Sec. 3 and the limiting process is discussed in Sec. 4.

## 2. SOME CONSEQUENCES OF THE CONDITIONS IMPOSED ON THE SPACE-TIME

Contracting Eq. (1.2) with $R^{ab}$ gives

$$(\tfrac{1}{2} - R_{ab}R^{ab})(R_{fg}R^{fg})_{;c} = 0.$$

Hence either $(R_{fg}R^{fg})_{;c} = 0$ or $R_{ab}R^{ab} = \frac{1}{2}$. The last condition implies that $(R_{fg}R^{fg})_{;c} = 0$ and so in both cases Eq. (1.2) yields

$$R_{ab;c} = 0. \tag{2.1}$$

This condition is equivalent to, and subsequently replaces, condition (1.2). From (2.1)

$$R_{ab;cd} - R_{ab;dc} = R^{e}{}_{acd}R_{eb} + R^{e}{}_{bcd}R_{ea} = 0,$$

where $R^{e}{}_{acd}$ is the Riemann tensor of the space-time. With condition (1.3), the Weyl tensor $C^{e}{}_{acd}$ is given by

$$C^{e}{}_{acd} = R^{e}{}_{acd} - \tfrac{1}{2}[\delta^{e}_{d}R_{ac} - \delta^{e}_{c}R_{ad} + g_{ac}R^{e}{}_{d} - g_{ad}R^{e}{}_{c}].$$

Hence, using (1.4),

$$C^{e}{}_{acd}R_{eb} = R^{e}{}_{acd}R_{eb} - \tfrac{1}{2}[R_{db}R_{ac} - R_{cb}R_{ad} + \tfrac{1}{4}R^{ef}R_{ef}(g_{ac}g_{bd} - g_{ad}g_{bc})].$$

Since the square bracket is antisymmetric in $a$ and $b$ it follows that

$$C^{e}{}_{acd}R_{eb} + C^{e}{}_{bcd}R_{ea} = 0. \tag{2.2}$$

For an algebraically general electromagnetic field, this equation implies that

$$C_{abcd} = 0. \tag{2.3}$$

This is easily verified by using a null tetrad[3] $l^a$, $n^a$, $m^a$, and $\bar{m}^a$ with $l^a$ and $n^a$ principal null vectors of the electromagnetic field.[4] Then

$$R_{eb} = |A|^2\,(l_e n_b + n_e l_b - \tfrac{1}{2}g_{eb})$$

and Eq. (2.2) becomes

$$C^{e}{}_{acd}l_e n_b + C^{e}{}_{acd}n_e l_b + C^{e}{}_{bcd}l_e n_a + C^{e}{}_{bcd}n_e l_a = 0.$$

Contracting with $l^a m^b$ gives

$$C^{e}{}_{bcd}l_e m^b = 0.$$

It follows from this that

$$C_{abcd}l^a m^b l^c m^d = C_{abcd}l^a m^b l^c n^d = C_{abcd}l^a m^b n^c \bar{m}^d = 0$$

and, because of the symmetry between $l^a$ and $n^a$,

$$C_{abcd}n^a m^b n^c m^d = C_{abcd}n^a m^b n^c l^d = 0.$$

The vanishing of the above five independent complex tetrad components of $C_{abcd}$ is sufficient to prove that

[1] R. Penney, J. Math. Phys. **6**, 1309 (1965).
[2] G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).
[3] E. Newman and R. Penrose, J. Math. Phys. **3**, 565 (1962).
[4] J. L. Synge, *Relativity, the Special Theory* (North-Holland Publ. Co, Amsterdam, 1956), p. 325.

the Weyl tensor is zero. The space-time is therefore conformally flat.[5]

### 3. SOLUTION OF THE FIELD EQUATIONS

In this section all conformally flat space-times satisfying conditions (1.3), (1.4), and (2.1) are found. The metric is assumed to have the form

$$g^{ij} = \phi\eta^{ij} \tag{3.1}$$

and Cartesian coordinates $(t, x, y, z)$ are chosen in the flat space so that $\eta^{ij} = \eta_{ij} =$ diagonal $(+1, -1, -1, -1)$.

The Christofel symbols and Ricci tensor of the space-time can now be written

$$\Gamma^a_{bc} = -\tfrac{1}{2}\phi^{-1}[\delta^a_b\phi_{,c} + \delta^a_c\phi_{,b} - \eta^{ad}\eta_{bc}\phi_{,d}] \tag{3.2}$$

and

$$4\phi^2R_{ab} = -2\phi[2\phi_{,ab} + \eta_{ab}\phi_{,c}^{\ c}] + [2\phi_{,a}\phi_{,b} + 4\eta_{ab}\phi_{,c}\phi^{,c}], \tag{3.3}$$

where indices are raised by the flat space metric $\eta^{ab}$. Condition (1.3) gives

$$2\phi\phi_{,c}^{\ c} = 3\phi_{,c}\phi^{,c}, \tag{3.4}$$

so that

$$4\phi^2R_{ab} = -4\phi\phi_{,ab} + 2\phi_{,a}\phi_{,b} + \eta_{ab}\phi_{,c}\phi^{,c}. \tag{3.5}$$

Condition (2.1) now becomes

$$4\phi R_{ab;c} = -4\phi_{,abc} + \eta_{ab}(\phi^{-1}\phi_{,d}\phi^{,d})_{,c} + \eta_{bc}(\phi^{-1}\phi_{,d}\phi^{,d})_{,a} + \eta_{ca}(\phi^{-1}\phi_{,d}\phi^{,d})_{,b} = 0. \tag{3.6}$$

The integrability condition for this is

$$\eta_{bc}(\phi^{-1}\phi_{,d}\phi^{,d})_{,ae} + \eta_{ca}(\phi^{-1}\phi_{,d}\phi^{,d})_{,be} - \eta_{be}(\phi^{-1}\phi_{,d}\phi^{,d})_{,ac} - \eta_{ea}(\phi^{-1}\phi_{,d}\phi^{,d})_{,bc} = 0. \tag{3.7}$$

Contracting on $ae$ gives

$$4(\phi^{-1}\phi_{,d}\phi^{,d})_{,bc} = \eta_{bc}(\phi^{-1}\phi_{,d}\phi^{,d})_{,e}^{\ ,e}. \tag{3.8}$$

Equation (3.7) is satisfied by virtue of Eq. (3.8). The general solution of Eq. (3.8) can be written as

$$\phi^{-1}\phi_{,d}\phi^{,d} = -\tfrac{1}{2}A\mathbf{x}\cdot\mathbf{x} + \boldsymbol{\alpha}\cdot\mathbf{x} + B, \tag{3.9}$$

where $A$, $B$, and the $\alpha^a$ are constants, and the dot product is defined by

$$\mathbf{v}\cdot\mathbf{w} = \eta_{ab}v^aw^b.$$

Substituting (3.9) into (3.6) gives

$$4\phi_{,abc} = \eta_{ab}(-A\eta_{cd}x^d + \alpha_c) + \eta_{bc}(-A\eta_{ad}x^d + \alpha_a) + \eta_{ca}(-A\eta_{bd}x^d + \alpha_b).$$

The general solution of this equation is

$$4\phi = -\tfrac{1}{8}A(\mathbf{x}\cdot\mathbf{x})^2 + \tfrac{1}{2}(\boldsymbol{\alpha}\cdot\mathbf{x})(\mathbf{x}\cdot\mathbf{x}) + \boldsymbol{\beta}\cdot\mathbf{x} - C + T_{ab}x^ax^b, \tag{3.10}$$

where $\beta^a$, $C$, and $T_{ab}$ are constants. Substituting (3.10) into (3.4) gives

$$3B = T_{ab}\eta^{ab}. \tag{3.11}$$

Further relations among the constants are obtained by substituting (3.10) into (3.9) and comparing powers of the coordinates. These relations are

$$\boldsymbol{\alpha}\cdot\boldsymbol{\alpha} = -2AB, \tag{3.12}$$

$$\boldsymbol{\beta}\cdot\boldsymbol{\beta} = 4BC, \tag{3.13}$$

$$\eta^{ad}T_{db}\alpha^b = -\tfrac{1}{2}A\beta^a + B\alpha^a, \tag{3.14}$$

$$\eta^{ad}T_{db}\beta^b = -C\alpha^a + B\beta^a, \tag{3.15}$$

and

$$4T_{fa}T_{gb}\eta^{fg} - 4BT_{ab} - 2AC\eta_{ab} + (\boldsymbol{\beta}\cdot\boldsymbol{\alpha})\eta_{ab} - \alpha^c\beta^d(\eta_{ca}\eta_{db} + \eta_{da}\eta_{cb}) = 0. \tag{3.16}$$

Finally, the condition (1.4) is identically satisfied and

$$4R_{fg}R^{fg} = B^2 + AC - \boldsymbol{\alpha}\cdot\boldsymbol{\beta}. \tag{3.17}$$

### 4. THE LIMITING PROCESS

The only question to be answered now is whether the limit as $R_{fg}R^{fg} \to 0$ does in fact yield a null electromagnetic field. Consider the case $A \neq 0$. A change of origin $x^a \to x^a + 2\alpha^a/A$ sets $\alpha^a = 0$. Equations (3.12) and (3.14) then yield $B = \beta^a = 0$. Substituting into (3.10) gives

$$4\phi = -\tfrac{1}{8}A(\mathbf{x}\cdot\mathbf{x})^2 - C + T_{ab}x^ax^b, \tag{4.1}$$

where

$$T_{ab}\eta^{ab} = 0 \tag{4.2}$$

and

$$4T_{fa}T_{gb}\eta^{fg} - 2AC\eta_{ab} = 0. \tag{4.3}$$

Of course (4.1) can be further simplified by a Lorentz transformation and also by a change of scale $x^a \to \lambda x^a$. In the limit as $R_{fg}R^{fg} = \frac{1}{2}AC$ tends to zero, the conditions (4.2) and (4.3) imply[6] that there exist constants $k_a$ and $v_a$, satisfying

$$k_av_b\eta^{ab} = k_ak_b\eta^{ab} = 0, \quad v_av_b\eta^{ab} = 1,$$

such that

$$T_{ab} = k_ak_b. \tag{4.4}$$

Substituting (4.4) and (4.1) into (3.5) yields

$$32\phi^2R_{ab} = 2A\eta_{ac}\eta_{bd}x^cx^d[k_fx^fk_gx^g - C] - 4k_ak_b[-\tfrac{1}{8}A(\mathbf{x}\cdot\mathbf{x})^2 - C] - A(\mathbf{x}\cdot\mathbf{x})k_fx^f[k_b\eta_{ac}x^c + k_a\eta_{bc}x^c]. \tag{4.5}$$

In the limit as $C \to 0$,

$$R_{ab} = F_{ac}F_{bd}g^{cd}$$

[5] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, N.J., 1925), p. 92.

[6] J. A. Wheeler, *Geometrodynamics* (Academic Press Inc., New York, 1962), p. 247.

with

$$F_{ab} = (A/4\phi^3)^{\frac{1}{2}}[K_aV_b - K_bV_a]/(\eta_{cd}x^cx^d), \quad (4.6)$$

where

$$K_a = (k_fx^f)\eta_{ac}x^c - \tfrac{1}{2}(x^cx^d\eta_{cd})k_a$$

and

$$V_a = (v_fx^f)\eta_{ac}x^c - \tfrac{1}{2}(x^cx^d\eta_{cd})v_a.$$

A short calculation yields $F_{ab;c} = 0$ and so Maxwell's equations are trivially satisfied. A similar conclusion is arrived at when $A = 0$.

The work presented here was motivated by the need to investigate the nature of the limiting process. For a null electromagnetic field, Eq. (2.2) does not necessarily imply that the space-time is conformally flat. This suggests that not every null electromagnetic field satisfying (1.1) can be found as the limit of a nonnull electromagnetic field satisfying (2.1). In fact, all null electromagnetic fields satisfying $R_{ab;c} = 0$ can be found (this, of course, has nothing to do with the geometrical approach). A short calculation shows that the principal vector of such an electromagnetic field must be geodesic, hypersurface-orthogonal, and expansion-free. Such fields have been discussed by Kundt[7] and it can easily be shown that, for null electromagnetic fields with $R_{ab;c} = 0$, the metric can be put in the form

$$ds^2 = -dz\,d\bar{z} - 2\,du\,dv - H\,du^2, \quad (4.7)$$

where

$$\partial^3H/\partial z\partial\bar{z}\partial u = \partial H/\partial v = 0. \quad (4.8)$$

The condition for the space-time (4.7) to be conformally flat is

$$\partial^2H/\partial z^2 = 0.$$

This is not identically satisfied by those space-times satisfying Eq. (4.8) and so not all neutrino fields can be obtained from the geometric conditions given by Penney.

The space-times obtained in Sec. 3 have also been obtained, in a different form, by Bertotti[8] and Stephani.[9]

[7] W. Kundt, Z. Physik **163**, 77 (1961).
[8] B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[9] H. Stephani, Commun. Math. Phys. **5**, 337 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Comments on Separability Operators, Invariance Ladder Operators, and Quantization of the Kepler Problem in Prolate-Spheroidal Coordinates*

H. M. KIEFER AND D. M. FRADKIN†
*Department of Physics, Wayne State University, Detroit, Michigan*

(Received 10 July 1967)

The Schrödinger equation for the hydrogen atom separates in three coordinate systems: spherical, parabolic, and prolate spheroidal. The separability operators associated with the separation constants for these three systems are exhibited and discussed. Also, for these systems, the invariance ladder operators which transform a simultaneous eigenfunction of the separability operators into a different simultaneous eigenfunction of the same energy are discussed with reference to the elements of the $O_4$ Lie algebra. Quantization of the Kepler problem in terms of prolate spheroidal coordinates is accomplished and discussed.

## I. INTRODUCTION

The application of symmetry groups to elementary particles has revived interest in the structural properties of the Schrödinger equation for the hydrogen atom.[1] The Coulomb problem is one with known solutions for which the invariance group ($O_4$) of the Hamiltonian and its relationship to the energy degeneracy have been exhaustively studied. Moreover, the larger noninvariance, noncompact group [the de Sitter $O(4, 1)$ group] connects bound states belonging to different energy levels.[2] Such a connection offers an algebraic approach that might be applicable to the energy (mass) level structure of the elementary particles.

* Supported in part by the National Science Foundation under Grant No. GP7695.
† Assisted by a Wayne State University Faculty Research Fellowship.

[1] For example, see M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966) and references therein.

[2] E. C. G. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, Phys. Letters **19**, 322 (1963); H. Bacry, Nuovo Cimento **41A**, 222 (1966); M. Y. Han, Nuovo Cimento **42B**, 367 (1966); R. H. Pratt and T. F. Jordan, Phys. Rev. **148**, 1276 (1966); R. Musto, Phys. Rev. **148**, 1274 (1966).

with

$$F_{ab} = (A/4\phi^3)^{\frac{1}{2}}[K_aV_b - K_bV_a]/(\eta_{cd}x^cx^d), \quad (4.6)$$

where

$$K_a = (k_fx^f)\eta_{ac}x^c - \tfrac{1}{2}(x^cx^d\eta_{cd})k_a$$

and

$$V_a = (v_fx^f)\eta_{ac}x^c - \tfrac{1}{2}(x^cx^d\eta_{cd})v_a.$$

A short calculation yields $F_{ab;c} = 0$ and so Maxwell's equations are trivially satisfied. A similar conclusion is arrived at when $A = 0$.

The work presented here was motivated by the need to investigate the nature of the limiting process. For a null electromagnetic field, Eq. (2.2) does not necessarily imply that the space-time is conformally flat. This suggests that not every null electromagnetic field satisfying (1.1) can be found as the limit of a nonnull electromagnetic field satisfying (2.1). In fact, all null electromagnetic fields satisfying $R_{ab;c} = 0$ can be found (this, of course, has nothing to do with the geometrical approach). A short calculation shows that the principal vector of such an electromagnetic field must be geodesic, hypersurface-orthogonal, and expansion-free. Such fields have been discussed by Kundt[7] and it can easily be shown that, for null electromagnetic fields with $R_{ab;c} = 0$, the metric can be put in the form

$$ds^2 = -dz\,d\bar{z} - 2\,du\,dv - H\,du^2, \quad (4.7)$$

where

$$\partial^3H/\partial z\partial\bar{z}\partial u = \partial H/\partial v = 0. \quad (4.8)$$

The condition for the space-time (4.7) to be conformally flat is

$$\partial^2H/\partial z^2 = 0.$$

This is not identically satisfied by those space-times satisfying Eq. (4.8) and so not all neutrino fields can be obtained from the geometric conditions given by Penney.

The space-times obtained in Sec. 3 have also been obtained, in a different form, by Bertotti[8] and Stephani.[9]

[7] W. Kundt, Z. Physik **163**, 77 (1961).
[8] B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[9] H. Stephani, Commun. Math. Phys. **5**, 337 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Comments on Separability Operators, Invariance Ladder Operators, and Quantization of the Kepler Problem in Prolate-Spheroidal Coordinates*

H. M. KIEFER AND D. M. FRADKIN†
*Department of Physics, Wayne State University, Detroit, Michigan*

(Received 10 July 1967)

The Schrödinger equation for the hydrogen atom separates in three coordinate systems: spherical, parabolic, and prolate spheroidal. The separability operators associated with the separation constants for these three systems are exhibited and discussed. Also, for these systems, the invariance ladder operators which transform a simultaneous eigenfunction of the separability operators into a different simultaneous eigenfunction of the same energy are discussed with reference to the elements of the $O_4$ Lie algebra. Quantization of the Kepler problem in terms of prolate spheroidal coordinates is accomplished and discussed.

## I. INTRODUCTION

The application of symmetry groups to elementary particles has revived interest in the structural properties of the Schrödinger equation for the hydrogen atom.[1] The Coulomb problem is one with known solutions for which the invariance group ($O_4$) of the Hamiltonian and its relationship to the energy degeneracy have been exhaustively studied. Moreover, the larger noninvariance, noncompact group [the de Sitter $O(4, 1)$ group] connects bound states belonging to different energy levels.[2] Such a connection offers an algebraic approach that might be applicable to the energy (mass) level structure of the elementary particles.

* Supported in part by the National Science Foundation under Grant No. GP7695.
† Assisted by a Wayne State University Faculty Research Fellowship.

[1] For example, see M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966) and references therein.

[2] E. C. G. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, Phys. Letters **19**, 322 (1963); H. Bacry, Nuovo Cimento **41A**, 222 (1966); M. Y. Han, Nuovo Cimento **42B**, 367 (1966); R. H. Pratt and T. F. Jordan, Phys. Rev. **148**, 1276 (1966); R. Musto, Phys. Rev. **148**, 1274 (1966).

The invariance group for a Hamiltonian is connected with the existence of a coordinate system in which the Schrödinger equation is separable. The separation constants which occur are the eigenvalues of certain operators which have the separated product solutions as their eigenfunctions. These (separability) operators commute with the Hamiltonian and are related to the elements of the Lie algebra characterizing the problem. For a Coulomb potential the Schrödinger equation separates in three coordinate systems: spherical, parabolic, and prolate spheroidal. The relevant separability operators for the first two systems are well known, but to the authors' knowledge, the operators associated with the prolate spheroidal coordinate system have not been exhibited. In this paper we discuss these operators and the related invariance ladder operators. We also obtain the semiclassical quantization conditions on the eigenvalues of the prolate spheroidal description by quantizing the action integrals in the manner of Bohr-Sommerfeld.

## II. REVIEW OF SPHERICAL AND PARABOLIC SEPARATIONS

The space–time separated (i.e., time-independent) Schrödinger equation for the Coulomb problem is

$$H\Psi = E\Psi, \tag{1}$$

where the Hamiltonian

$$H = (p^2/2M) - (\lambda/r) \tag{2}$$

is the operator associated with the separation constant $E$. Here, $\mathbf{p} = -i\hbar\nabla$, $\lambda$ is the strength of the interaction ($\lambda$ is positive for the attractive case), and $r = [x^2 + y^2 + z^2]^{\frac{1}{2}}$. Also, we adopt the caret notation for the vector cross product, as in $\mathbf{L} = \mathbf{r} \wedge \mathbf{p}$. Equation (1) is separable in spherical coordinates:

$$\begin{aligned} \mathbf{r}\cdot\hat{k} &= r\cos\theta, \\ \mathbf{r}\cdot\hat{\imath} &= r\sin\theta\cos\phi, \\ \mathbf{r}\cdot\hat{\jmath} &= r\sin\theta\sin\phi, \end{aligned} \tag{3}$$

where $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ are a right-handed triad of mutually orthogonal unit vectors, with $\hat{k}$ chosen as the direction of the polar axis. The (commuting) operators, of which the product solutions are eigenfunctions, corresponding to the separation constants are

$$H, \quad L^2, \quad \text{and} \quad \hat{k}\cdot\mathbf{L}. \tag{4}$$

The three operators $\mathbf{L}$ are the generators of $SU_2$ (or $O_3$) and $L^2$ is the Casimir operator of this group.

In parabolic coordinates (confocal parabolas of revolution),

$$\begin{aligned} \mu &= r + \mathbf{r}\cdot\hat{k}, \\ \nu &= r - \mathbf{r}\cdot\hat{k}, \\ \tan\phi &= (\mathbf{r}\cdot\hat{\jmath})/(\mathbf{r}\cdot\hat{\imath}). \end{aligned} \tag{5}$$

Equation (1) also separates. As first shown by Bargmann,[3] the (commuting) operators associated with the separation constants, of which product solutions are eigenfunctions, are

$$H, \quad \hat{k}\cdot\mathbf{L}, \quad \text{and} \quad \hat{k}\cdot\mathbf{R}, \tag{6}$$

where $\mathbf{R}$ is the Runge–Lenz vector operator[4]

$$\mathbf{R} = \tfrac{1}{2}(\mathbf{p} \wedge \mathbf{L} - \mathbf{L} \wedge \mathbf{p}) - M\lambda\mathbf{r}/r. \tag{7}$$

The algebra of the related three operators

$$\mathbf{A} = (-2MH)^{-\frac{1}{2}}\mathbf{R}, \tag{8}$$

which is Hermitian when acting on bound states, together with the three operators $\mathbf{L}$, closes on $O_4$:

$$\begin{aligned} [L_i, L_j] &= i\hbar\epsilon_{ijk}L_k, \\ [L_i, A_j] &= i\hbar\epsilon_{ijk}A_k, \\ [A_i, A_j] &= i\hbar\epsilon_{ijk}L_k. \end{aligned} \tag{9}$$

In contrast to the case for separation in spherical coordinates, the separability operators [Eq. (6)] for the parabolic case do not contain a Casimir operator. The Casimir operator for $O_4$ may be obtained by considering the linear combinations

$$\mathbf{G}^{(\rho)} = (1/\sqrt{2})(\mathbf{L} + \rho\mathbf{A}), \quad \rho = \pm 1. \tag{10}$$

One sees that $O_4$ becomes $SU_2 \times SU_2$, since

$$[G_i^{(\rho)}, G_j^{(\rho')}] = \delta_{\rho\rho'}i\hbar\epsilon_{ijk}G_k^{(\rho)}, \quad \text{no sum on } \rho. \tag{11}$$

The Casimir operator is[5]

$$\begin{aligned} G^2 \equiv \mathbf{G}^{(1)}\cdot\mathbf{G}^{(1)} &= \mathbf{G}^{(-1)}\cdot\mathbf{G}^{(-1)} \\ &= \tfrac{1}{2}[-\hbar^2 - (\lambda^2 M/2H)]. \end{aligned} \tag{12}$$

Invariance ladder operators are defined to be those operators (which commute with the Hamiltonian) which transform a state of specified energy into a different state of the same energy. For the spherical-coordinate separability operators $H$, $L^2$, and $L_3$,[6] the basic invariance ladder operators are[7]

$$\begin{aligned} L_\epsilon &= L_1 + i\epsilon L_2, \\ \mathfrak{L}_\epsilon &= A_3\{(L^2 + \hbar^2/4)^{\frac{1}{2}} + \epsilon\hbar/2\} + i\epsilon(\mathbf{A} \wedge \mathbf{L})_3, \end{aligned} \tag{13}$$

where $\epsilon = \pm 1$. These operators satisfy the commutation relations

$$\begin{aligned} [H, L_\epsilon] &= [L^2, L_\epsilon] = 0, \\ [L_3, L_\epsilon] &= \epsilon\hbar L_\epsilon, \end{aligned} \tag{14}$$

---

[3] V. Bargmann, Z. Physik **99**, 576 (1936).

[4] C. Runge, *Vektoranalysis* (B. G. Teubner, Leipzig, 1919), Vol. 1, p. 70; W. Lenz, Z. Physik **24**, 197 (1923); see also W. Pauli, Z. Physik **36**, 336 (1926); V. Fock, Z. Physik **98**, 145 (1935).

[5] Here we use the fact that $\mathbf{R}\cdot\mathbf{R} = 2MH(L^2 + \hbar^2) + \lambda^2M^2$.

[6] Here the unit vector $\hat{k}$ is taken in the 3-direction.

[7] The operator $\mathfrak{L}_\epsilon$ may be obtained most simply through a consideration of a linear combination of the elements of the Lie algebra that commute with $L_3$. It may also be obtained as a product of operators derived by the factorization method [see L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951)], or for $\mathfrak{L}_+$, by a simplification of the expression $(L_-)^{\ell+1-m}A_+(L_+)^{\ell-m}$.

and

$$[H, \mathfrak{L}_\epsilon] = [L_3, \mathfrak{L}_\epsilon] = 0,$$
$$[L^2, \mathfrak{L}_\epsilon] = 2\epsilon\hbar\mathfrak{L}_\epsilon\{(L^2 + \hbar^2/4)^{\frac{1}{2}} + \epsilon\hbar/2\}. \quad (15)$$

Using the fact that the eigenvalues associated with $L^2$ and $L_3$ are, respectively, $\ell(\ell+1)\hbar^2$ and $m\hbar$, and denoting a wavefunction of specified energy by $\Psi_{\ell,m}$, it follows that[8]

$$L_\epsilon\Psi_{\ell,m} = \Psi_{\ell,m+\epsilon},$$
$$\mathfrak{L}_\epsilon\Psi_{\ell,m} = \Psi_{\ell+\epsilon,m}. \quad (16)$$

For the parabolic-coordinate separability operators $H$, $A_3$, and $L_3$,[6] the basic invariance ladder operators are

$$G_\epsilon^{(\rho)} = G_1^{(\rho)} + i\epsilon G_2^{(\rho)}, \quad (17)$$

where $\mathbf{G}^{(\rho)}$ has been defined in Eq. (10), and $\rho$ and $\epsilon$ are independently $\pm 1$. These operators satisfy the commutation relations

$$[H, G_\epsilon^{(\rho)}] = 0,$$
$$[L_3, G_\epsilon^{(\rho)}] = \epsilon\hbar G_\epsilon^{(\rho)}, \quad (18)$$
$$[A_3, G_\epsilon^{(\rho)}] = \epsilon\rho\hbar G_\epsilon^{(\rho)}.$$

In the parabolic system, the eigenvalues associated with $A_3$ and $L_3$ are, respectively, $m'\hbar$, and $m\hbar$, so denoting a wavefunction of specified energy by $\Psi_{m',m}$, it follows that[8]

$$G_\epsilon^{(\rho)}\Psi_{m',m} = \Psi_{m'+\epsilon\rho,m+\epsilon}. \quad (19)$$

For these wavefunctions, the distribution of allowable $m'$, $m$ eigenvalues is such[9] that, for a given energy, it is impossible to change one eigenvalue by unity while keeping the other one fixed.

## III. SEPARATION IN PROLATE-SPHEROIDAL COORDINATES

The Schrödinger equation [Eq. (1)] separates in prolate-spheroidal coordinates:

$$\xi = a^{-1}(r + |\mathbf{r} - \mathbf{a}|),$$
$$\eta = a^{-1}(r - |\mathbf{r} - \mathbf{a}|), \quad (20)$$
$$\tan\phi = (\mathbf{r}\cdot\hat{j})/(\mathbf{r}\cdot\hat{\imath}).$$

This coordinate system involves ellipses and hyperbolas of revolution associated with the foci at $\mathbf{r} = 0$ and $\mathbf{r} = \mathbf{a}$. Here $\hat{\imath}, \hat{j}, \hat{k}$ are the orthonormal triad mentioned previously, and $\mathbf{a} = a\hat{k}$ where $a$ is an arbitrary length. The wavefunction $\Psi$ may be expressed as

$$\Psi(\mathbf{r}) = \Psi_1(\xi)\Psi_2(\eta)\Psi_3(\phi), \quad (21)$$

and the separated equations are

$$\partial^2\Psi_3(\phi)/\partial\phi^2 = -m^2\Psi_3(\phi), \quad (22)$$

$$\left\{\hbar^2\left[\frac{\partial}{\partial\xi}(\xi^2-1)\frac{\partial}{\partial\xi} - \frac{m^2}{(\xi^2-1)}\right] + \frac{a^2ME}{2}\xi^2 + a\lambda M\xi\right\}\Psi_1(\xi) = s\Psi_1(\xi), \quad (23)$$

$$\left\{\hbar^2\left[\frac{\partial}{\partial\eta}(1-\eta^2)\frac{\partial}{\partial\eta} - \frac{m^2}{(1-\eta^2)}\right] - \frac{a^2ME}{2}\eta^2 - a\lambda M\eta\right\}\Psi_2(\eta) = -s\Psi_2(\eta), \quad (24)$$

where $m^2$ and $s$ are separation constants. The operator having $m$ as its eigenvalue is obvious, and the operator having $s$ as its eigenvalue may be found by judiciously substituting for $m^2$ and $E$ (when acting to the extreme right) in terms of the associated operators. In this fashion we find that the (commuting) operators associated with the separation constants, of which the product solutions are eigenfunctions,[10] are

$$H, \quad S, \quad \text{and} \quad \hat{k}\cdot\mathbf{L}, \quad (25)$$

where

$$S = L^2 - \mathbf{a}\cdot\mathbf{R} + \tfrac{1}{2}a^2MH. \quad (26)$$

Here $\mathbf{R}$ is the Runge–Lenz vector defined by Eq. (7). Alternatively, we may write the operator $S$ in terms of the $O_4$ Casimir operator $G^2$ and generators $\mathbf{A}$ [see Eqs. (8) and (12)]. It is found that

$$S = G^2 - [\mathbf{A} + (-\tfrac{1}{2}MH)^{\frac{1}{2}}\mathbf{a}]\cdot[\mathbf{A} + (-\tfrac{1}{2}MH)^{\frac{1}{2}}\mathbf{a}]. \quad (27)$$

The classical equivalent to the operator $S$ is especially simple *if* one chooses $\mathbf{a}$ to be the vector from the force center to the other focus occurring in the problem. Classically,

$$\mathbf{R} = \mathbf{p}\times\mathbf{L} - \lambda M\mathbf{r}/r,$$
$$R^2 = 2MEL^2 + \lambda^2M^2. \quad (28)$$

Setting

$$\mathbf{R}\cdot\mathbf{r} = \cos\theta, \quad (29)$$

the orbit equation is

$$1/r = (\lambda M/L^2)[1 + (R/\lambda M)\cos\theta]. \quad (30)$$

The vectors $\mathbf{r}_{\min}$ and $\mathbf{r}_{\max}$ (associated with $r_{\min}$ and $r_{\max}$) occur when $\theta = 0$ and $\pi$, respectively, so

$$\mathbf{r}_{\min} = [L^2/(\lambda M + R)]\mathbf{R}/R,$$
$$\mathbf{r}_{\max} = -[L^2/(\lambda M - R)]\mathbf{R}/R. \quad (31)$$

Consequently, our choice of $\mathbf{a}$ as the vector from the force center to the other focus means that

$$\mathbf{a} = \mathbf{r}_{\max} + \mathbf{r}_{\min} = \mathbf{R}/(ME), \quad (32)$$

[8] The ladder operators given here have *not* been normalized to generate normalized wavefunctions from normalized wavefunctions.

[9] See for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., p. 89.

[10] In particular, $S(\Psi_1\Psi_2\Psi_3) = s(\Psi_1\Psi_2\Psi_3)$.

or equivalently

$$(-\tfrac{1}{2}MH)^{\frac{1}{2}}\mathbf{a} = -\mathbf{A}. \tag{33}$$

For this assignment of $\mathbf{a}$, Eq. (27) indicates that $S$ is just equal to the Casimir quantity $G^2$.

However, in setting up the prolate-spheroidal coordinate system, $\mathbf{a}$ is not restricted to be any particular vector; so, in general, the operator $S$ has no simple interpretation. One sees from Eq. (26) that the operator $S$ is just a linear combination of the separability operators that occur in the spherical and parabolic cases (the operators $\hat{k}\cdot\mathbf{L}$ and $H$ being the same). We have attempted to obtain the invariance ladder operators for the prolate-spheroidal eigenfunctions—by the techniques sketched in Ref. 7—without success. Such operators exist, since one must surely be able to go from one state to any other state. If the invariance ladder operators could be written in closed form, the eigenfunctions could be obtained by integrating the equation

$$Q\Psi_{s(\max)} = 0,$$

where $Q$ is the raising operator for $s$, to give $\Psi_{s(\max)}$. Then other solutions could be obtained by suitable "lowering" and "raising" operators. Solutions of this problem would not only have general interest, but would also be useful in the two-Coulomb center problem (e.g., the ionized hydrogen molecule in the Born–Oppenheimer approximation), which separates only in prolate-spheroidal coordinates.

## IV. QUANTIZATION IN TERMS OF PROLATE-SPHEROIDAL COORDINATES

In order to obtain the invariance ladder operators for the prolate spheroidal eigenfunctions one must find the eigenvalue spectrum of $S$. We have attempted to solve this problem by invoking the Bohr–Sommerfeld quantization rules in the equivalent classical problem.

The Lagrangian for the classical Coulomb problem in prolate-spheroidal coordinates is

$$L = \frac{Ma^2}{8}\left[\frac{(\xi^2-\eta^2)}{(1-\eta^2)}\dot{\eta}^2 + \frac{(\xi^2-\eta^2)}{(\xi^2-1)}\dot{\xi}^2 + (\xi^2-1)\times(1-\eta^2)\dot{\phi}^2\right] + \frac{2\lambda}{a(\xi+\eta)}. \tag{34}$$

Using the related Hamiltonian and the separated Hamilton–Jacobi equations, we can express the canonical momenta by

$$p_\xi = \pm\left[\frac{-s}{\xi^2-1} - \frac{p_\phi^2}{(\xi^2-1)^2} + \frac{\lambda Ma\xi}{(\xi^2-1)} + \frac{EMa^2\xi^2}{2(\xi^2-1)}\right]^{\frac{1}{2}} \tag{35}$$

and

$$p_\eta = \pm\left[\frac{s}{(1-\eta^2)} - \frac{p_\phi^2}{(1-\eta^2)^2} - \frac{\lambda Ma\eta}{(1-\eta^2)} - \frac{EMa^2\eta^2}{2(1-\eta^2)}\right]^{\frac{1}{2}}, \tag{36}$$

where the separation constant $E$ is the constant energy in this problem, $p_\phi$ is the constant momentum canonical to the ignorable coordinate $\phi$, and the separation constant $s$ is the eigenvalue of $S$. For $x$ being either $\xi$ or $\eta$, the corresponding momentum may be uniformly written

$$p_x = \pm[\tfrac{1}{2}EMa^2x^2 + \lambda Max - s - p_\phi^2(x^2-1)^{-1}]y^{-\frac{1}{2}}, \tag{37}$$

where the quartic $y$ is

$$y = \tfrac{1}{2}EMa^2x^4 + \lambda Max^3 - (s + \tfrac{1}{2}EMa^2)x^2 - \lambda Max + (s - p_\phi^2). \tag{38}$$

The sign that must be chosen for $p_x$ depends on whether the variable $x$ is increasing or decreasing. The analysis which determines the sign is given in the Appendix.

Applying the Bohr–Sommerfeld quantization procedure to the action integrals, we obtain

$$\begin{aligned} mh &= \oint p_\phi\, d\phi = 2\pi p_\phi, \\ nh &= \oint p_\xi\, d\xi, \\ n'h &= \oint p_\eta\, d\eta. \end{aligned} \tag{39}$$

Here $m$, $n$, and $n'$ are integers, $h$ is Planck's constant, and the integration is to be carried out over a complete period of motion. The turning points of the paths involving $p_x$ ($x$ being either $\xi$ or $\eta$) are those points at which $dp_x/dx$ becomes infinite. The only points[11] at which this occurs are the four roots of the quartic $y$, which are also the values at which $p_x$ equals zero. The definitions of $\xi$ and $\eta$ imply that $-1 \le \eta \le 1$ and that $1 \le \xi$; therefore we see that the two smallest roots of $y$ are turning points corresponding to the $\eta$ integration and that the two largest roots of $y$ are turning points for the $\xi$ integration. Thus we may label the roots ($p_\phi \neq 0$) in the following fashion:

$$-1 < \alpha_4 < \alpha_3 < 1 < \alpha_2 < \alpha_1. \tag{40}$$

Using the fact that turning points for the paths occur on the $x$ axis and that $p_x$ is symmetric with respect to the $x$ axis, we find that

$$\begin{aligned} nh &= 2\int_{\alpha_2}^{\alpha_1} p_x\, dx, \\ n'h &= 2\int_{\alpha_4}^{\alpha_3} p_x\, dx. \end{aligned} \tag{41}$$

[11] The variables $\xi$ and $\eta$ obtain the values $\pm 1$ only for the special case $p_\phi = 0$. In the present analysis, we consider only the situations for which $p_\phi \neq 0$.

Choosing the appropriate sign for $p_x$, we may directly do the integration in terms of elliptic integrals. The results are

$$\tfrac{1}{2}[-\tfrac{1}{2}EMa^2(\alpha_1-\alpha_3)(\alpha_2-\alpha_4)]^{\frac{1}{2}}nh$$
$$=\Big[cK(k)+\tfrac{1}{2}EMa^2(\alpha_1-\alpha_3)(\alpha_2-\alpha_4)E(k)$$
$$+\lambda Ma(\alpha_2-\alpha_3)\Pi_1(\rho_1,k)-\frac{p_\phi^2(\alpha_2-\alpha_3)}{(1+\alpha_2)(1+\alpha_3)}$$
$$\times\Pi_1(\rho_1',k)-\frac{p_\phi^2(\alpha_2-\alpha_3)}{(\alpha_2-1)(1-\alpha_3)}\Pi_1(\rho_1'',k)\Big], \quad (42)$$

where

$$c=\alpha_3\lambda Ma+2p_\phi^2(1-\alpha_3^2)^{-1}-2s$$
$$-\tfrac{1}{2}EMa^2(\alpha_1\alpha_2+\alpha_3\alpha_4),$$
$$k^2=[(\alpha_1-\alpha_2)(\alpha_3-\alpha_4)]/[(\alpha_1-\alpha_3)(\alpha_2-\alpha_4)],$$
$$\rho_1=-(\alpha_1-\alpha_2)/(\alpha_1-\alpha_3),$$
$$\rho_1'=[(1-\alpha_3)(\alpha_1-\alpha_2)]/[(\alpha_2-1)(\alpha_1-\alpha_3)],$$
$$\rho_1''=-[(1+\alpha_3)(\alpha_1-\alpha_2)]/[(1+\alpha_2)(\alpha_1-\alpha_3)],$$

and the notation of the Batemen project[12] is used for the complete elliptic integrals K, E, and $\Pi_1$ of the first, second, and third kinds, respectively.

The expression for $n'$ may be obtained from Eq. (42) by means of the following replacements: $n\to -n'$, $\alpha_1\to\alpha_3$, $\alpha_3\to\alpha_1$, $\alpha_2\to\alpha_4$, and $\alpha_4\to\alpha_2$. Under this replacement, $k$ is invariant. One may relate the complete elliptic integrals of the third kind (the $\Pi_1$'s) that appear in the expressions for $n$ and $n'$. Let us designate the $\rho$'s for which the preceding replacement has been made by the subscript 2, e.g.,

$$\rho_2=-(\alpha_3-\alpha_4)/(\alpha_3-\alpha_1).$$

Now it is a fact that

$$(1+\nu_1)\Pi_1(\nu_1,k)+(1+\nu_2)\Pi_1(\nu_2,k)$$
$$=K(k)+\tfrac{1}{2}\pi(-\nu_1\nu_2/k^2)^{\frac{1}{2}}, \quad (43)$$

*if* the pair $(\nu_1,\nu_2)$ satisfies the condition

$$\nu_1\nu_2(1-k^2)=-k^2(1+\nu_1)(1+\nu_2). \quad (44)$$

One may show that $(\rho_1,\rho_2)$, $(\rho_1',\rho_2')$, $(\rho_1'',\rho_2'')$ satisfy the $(\nu_1,\nu_2)$ condition; hence the corresponding $\Pi_1$'s are related.

One may obtain the energy quantization by laboriously adding $n$ and $n'$, using the preceding equations together with expressions relating the coefficients of the quartic to the roots. A somewhat more transparent technique for accomplishing the same result, by use of complex variables, is presented in the Appendix. One finds that the energy $E$ is given by

$$E=-2(\pi\lambda/h)^2M(n+n'+|m|)^{-2}, \quad (45)$$

in terms of the quantum numbers for separation in prolate-spheroidal coordinates.

The roots are functions of the coefficients of the quartic, i.e., functions of the separation constants $E$, $p_\phi$, and $s$. These functions may be explicitly determined by solving the quartic, but the resulting expressions are prohibitively complicated. However, in principle the $\alpha$'s are known in terms of $E$, $p_\phi$, $s$. Since $p_\phi=mh/(2\pi)$ and $E$ is also known as a function of the quantum numbers $n$, $n'$, and $m$ [from Eq. (45)], appropriate substitutions can be made in Eq. (42) to give a functional relation for $s$ involving the three quantum numbers. Thus $s$ is *implicitly* determined. It appears unlikely that the eigenvalue spectrum of $S$ (i.e., the values of $s$) can be written in closed form. Hence it appears that the ladder operators, which transform a single prolate-spheroidal state to another, are formidable nonpolynomial functions that also cannot be written in closed form.[13]

*Note added in proof:* The relation of the two-centered Kepler problem to separation in spheroidal coordinates has been recently discussed by C. A. Coulson and A. Joseph, Intern. J. Quant. Chem. **1**, 337 (1967).

## APPENDIX

The squares of the canonical momenta are

$$p_\xi^2=(\xi^2-1)^{-1}[\tfrac{1}{2}EMa^2\xi^2+\lambda Ma\xi-s$$
$$-p_\phi^2(\xi^2-1)^{-1}] \quad (A1)$$

and

$$p_\eta^2=(1-\eta^2)^{-1}[-\tfrac{1}{2}EMa^2\eta^2-\lambda Ma\eta+s$$
$$-p_\phi^2(1-\eta^2)^{-1}]$$
$$=(\eta^2-1)^{-1}[\tfrac{1}{2}EMa^2\eta^2+\lambda Ma\eta-s$$
$$-p_\phi^2(\eta^2-1)^{-1}]. \quad (A2)$$

Thus one may uniformly write (for $x$ being either $\xi$ or $\eta$)

$$p_x=\pm[\tfrac{1}{2}EMa^2x^2+\lambda Max-s-p_\phi^2(x^2-1)^{-1}]y^{-\frac{1}{2}}, \quad (A3)$$

where, as before, the quartic $y$ is

$$y=(x^2-1)[+\tfrac{1}{2}EMa^2x^2+\lambda Max-s$$
$$-p_\phi^2(x^2-1)^{-1}]. \quad (A4)$$

[12] *Bateman Manuscript Project, Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 2, p. 317.

[13] Similar analytical difficulties are associated with the solutions of the separated equations, Eqs. (23) and (24). Even for the free particle ($\lambda=0$), the solutions involve Lame or spheroidal wave-functions. [See, for example, W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949), p. 158; C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford California, 1957).]

The appropriate sign in Eq. (A3) must be carefully determined to correspond to the variable and the integration path used for that variable in computing the action. For the variable $\xi$, where the path of integration is from[14] $\alpha_2$ to $\alpha_1$, it follows from the Lagrangian that

$$p_\xi = (Ma^2/4)(\xi^2 - \eta^2)(\xi^2 - 1)^{-1}\dot{\xi}. \quad \text{(A5)}$$

But $(\xi^2 - \eta^2)(\xi^2 - 1) > 0$, and as $\xi$ goes from $\alpha_2$ to $\alpha_1$, $\dot{\xi} > 0$. Hence it follows that $p_\xi$ must be taken as positive. Moreover, since $p_\xi^2 > 0$, Eq. (A1) implies that

$$[\tfrac{1}{2}EMa^2\xi^2 + \lambda Ma\xi - s - p_\phi^2(\xi^2 - 1)^{-1}] \geq 0. \quad \text{(A6)}$$

Therefore $y \geq 0$ (where $x = \xi$), the factor in $p_x$ multiplying $y^{-\frac{1}{2}}$ is $\geq 0$, and one must choose the positive sign for $p_x$ in Eq. (A3) for computing the action

$$nh = 2\int_{\alpha_2}^{\alpha_1} p_x\, dx. \quad \text{(A7)}$$

For the variable $\eta$, where the path of integration is from $\alpha_4$ to $\alpha_3$, it follows from the Lagrangian that

$$p_\eta = \tfrac{1}{4}Ma^2(\xi^2 - \eta^2)(1 - \eta^2)^{-1}\dot{\eta}. \quad \text{(A8)}$$

Here $(\xi^2 - \eta^2)(1 - \eta^2)^{-1} > 0$, and as $\eta$ goes from $\alpha_4$ to $\alpha_3$, $\dot{\eta} > 0$. Hence it follows that $p_\eta$ must be taken as positive. However, since $p_\eta^2 > 0$ (and $|\eta| \leq 1$), Eq. (A2) implies that

$$[\tfrac{1}{2}EMa^2\eta^2 + \lambda Ma\eta - s - p_\phi^2(\eta^2 - 1)^{-1}] \leq 0. \quad \text{(A9)}$$

Again, therefore, $y \geq 0$ (where $x = \eta$), but the factor in $p_x$ multiplying $y^{-\frac{1}{2}}$ is $\leq 0$, so one must choose the negative sign for $p_x$ in Eq. (A3) for computing the action

$$n'h = 2\int_{\alpha_4}^{\alpha_3} p_x\, dx. \quad \text{(A10)}$$

Now consider the complex extension of $y(x)$, namely $y(z)$, where

$$\begin{aligned} y(z) &= (-\tfrac{1}{2}EMa^2)(\alpha_1 - z)(z - \alpha_2)(z - \alpha_3)(z - \alpha_4) \\ &= (z^2 - 1)[\tfrac{1}{2}EMa^2z^2 + \lambda Maz - s \\ &\qquad - p_\phi^2(z^2 - 1)^{-1}], \end{aligned} \quad \text{(A11)}$$

which is positive when $z$ is real and $\alpha_2 < z < \alpha_1$. For a discussion of $[y(z)]^{+\frac{1}{2}}$, let us for convenience choose branch cuts for the displayed $z$ factors which run from the branch points to infinity along the real $z(+)$ direction. Furthermore, let us choose phases so that $(\alpha_1 - z)$ has zero phase when $z$ is real and less than $\alpha_1$, and $(z - \alpha_2)$, $(z - \alpha_3)$, $(z - \alpha_4)$ have zero phases on the upper banks of their cuts when $z$ is real and greater than $\alpha_2$, $\alpha_3$, and $\alpha_4$, respectively.[15] Thus, on the upper bank of the cuts, $y(z)$ has zero phase for $z$ real in the region where $\alpha_2 < z < \alpha_1$, and $y(z)$ has a phase of $2\pi$ for $z$ real in the region where $\alpha_4 < z < \alpha_3$. Consequently, for

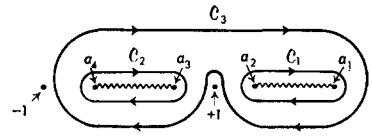



FIG. 1. Contours of integration for action integrals and energy quantization.

$$\begin{aligned} p(z) = [\tfrac{1}{2}EMa^2z^2 + \lambda Maz - s \\ - p_\phi^2(z^2 - 1)^{-1}][y(z)]^{-\frac{1}{2}}, \end{aligned} \quad \text{(A12)}$$

the action integrals may be obtained from the contour integrals (see Fig. 1)

$$\begin{aligned} \int_{C_1} p(z)\, dz &= nh, \\ \int_{C_2} p(z)\, dz &= n'h. \end{aligned} \quad \text{(A13)}$$

The fact that $n$ and $n'$ are real implies that $E$ is negative. (Thus $-\frac{1}{2}EMa^2$ is taken to have zero phase.) Energy quantization can be obtained directly by integrating over contour $C_3$ (see Fig. 1) which encloses the contours $C_1$ and $C_2$, but excludes the poles at $z = 1$, $-1$, and infinity. Integration may be performed by the method of residues by considering the path enclosing the rest of the complex plane, the direction of integration now being in the clockwise direction.[16] The result is

$$(n + n')h = -2\pi\,|p_\phi| - 2\pi\lambda[M/(-2E)]^{\frac{1}{2}}, \quad \text{(A14)}$$

where the first term on the right is the sum of the contributions from the poles at $z = \pm 1$ and $z = -1$, and the second term on the right is the contribution from the pole at infinity. Consequently, using the fact that $2\pi p_\phi = mh$, we obtain the expression for the energy $E$,

$$E = -2(\pi\lambda/h)^2M(n + n' + |m|)^{-2}, \quad \text{(A15)}$$

in terms of the quantum number $n$, $n'$, and $m$.

[14] See Eq. (40) and the discussion preceding it for a description of the roots $\alpha$.

[15] In Fig. 1, only the resultant cuts in the $z$ plane for rational functions of $z$ and $[y(z)]^{\frac{1}{2}}$ are shown.

[16] For an example employing this technique, see H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Co., Inc., Reading Mass., 1950), p. 302.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Existence, Uniqueness, and Convergence of the Solutions of Models in Kinetic Theory

CARLO CERCIGNANI
*Massachusetts Institute of Technology, Cambridge, Massachusetts**

(Received 23 March 1967)

A recently proved theorem of existence and uniqueness for the linearized Boltzmann equation is extended to two- and three-dimensional domains and general boundary conditions. The proof is valid for collision operators having a finite collision frequency, which can arise either from an angular or a radial cutoff or by assuming a model equation. Finally, convergence of the solutions of kinetic models to solutions of the actual Boltzmann equation is shown to hold for the boundary-value problems considered in this paper.

## 1. INTRODUCTION

In a previous paper[1] a simple proof of existence and uniqueness in the large for the linearized Boltzmann equation with finite collision frequency was presented. The proof applied to rigid-sphere molecules as well as molecules interacting with any hard potential, provided that an angular cutoff is introduced in the collision term, as suggested by Grad.[2] It was, however, suggested[1] that the proof could be extended to linearized operators with a radial, rather than an angular, cutoff. The collision operators with hard cutoff potentials have been studied,[3] and the above suggestion was found to be essentially true; i.e., we can show that the collision operator for cutoff potentials enjoys properties which can constitute the basis of a proof of the same kind as the one previously given.[1] The space domain was previously[1] assumed to be a slab, and the boundary conditions had a simplified form, i.e., the distribution function was supposed to be given at each wall for emerging molecules.

In this paper we want to extend the proof to two-dimensional and three-dimensional domains of finite but arbitrary size and general boundary conditions. We also want to discuss the question of convergence of the solutions of kinetic models to solutions of the actual Boltzmann equation. It will be shown that, for the boundary-value problems considered in this paper, such convergence applies under the assumption of reasonable convergence properties of the model collision operator to the actual collision operator.

## 2. PROPERTY OF THE FREE-MOLECULAR OPERATOR FOR GENERAL BOUNDARY CONDITIONS

In this section we want to show that, for a class (to be specified below) of reasonable homogeneous boundary conditions at the boundary $\partial R$ of a bounded domain $R$, the following inequality holds:

$$\int \left|\boldsymbol{\xi}\cdot\frac{\partial f}{\partial \mathbf{x}}\right|^2 \beta(\xi)\,d\boldsymbol{\xi}\,d\mathbf{x} \geq \frac{\gamma}{d^2}\int \xi^2\beta(\xi)\,|f|^2\,d\boldsymbol{\xi}\,d\mathbf{x}. \quad (2.1)$$

Here $\beta(\xi)$ is any given positive function of the molecular speed such that the following integrals exist:

$$\int \omega(\xi)\beta(\xi)\xi\,d\boldsymbol{\xi} \quad \text{and} \quad \int \omega(\xi)[\beta(\xi)]^{-1}\xi\,d\boldsymbol{\xi}, \quad (2.2)$$

where $d$ is the maximum chord which can be drawn in $R$, $\gamma$ is a numerical constant (depending upon $\beta$ and $R$), and $f$ is the linearized distribution function related to the full distribution $F$ by

$$F = F_M(1 + \omega^{\frac{1}{2}}f). \quad (2.3)$$

Here $F_M$ is the relevant Maxwellian and $\omega$ is the same Maxwellian in nondimensional form

$$\omega = (2\pi)^{-\frac{3}{2}}\exp[-\xi^2/2].$$

Molecular velocities are measured in $(RT_0)^{\frac{1}{2}}$ units, $T_0$ being the temperature of $F_M$ and $R$ the gas constant. In Eq. (2.1) $f$ is assumed to satisfy the condition

$$\int_{\xi\cdot\mathbf{n}>0} f\omega^{\frac{1}{2}}\,|\boldsymbol{\xi}\cdot\mathbf{n}|\,d\boldsymbol{\xi}\,d\mathbf{y} = 0, \quad (2.4)$$

where, if $\mathbf{y}\in\partial R$, $\mathbf{n}$ is the inward normal at $\mathbf{y}$ and $d\mathbf{y}$ is a surface element. Equation (2.4) can be always satisfied by introducing a suitable density in $F_M$. $f$ is also such that the integral in the right-hand side of Eq. (2.1) does exist and is finite.

Our boundary conditions will be rather general. We shall assume that

$$f(\boldsymbol{\xi};\mathbf{y}) = \int_{\xi'\cdot\mathbf{n}<0} A(\boldsymbol{\xi}'\to\boldsymbol{\xi};y)f(\boldsymbol{\xi}';\mathbf{y})\,d\boldsymbol{\xi}' \equiv Af, \quad (\boldsymbol{\xi}\cdot\mathbf{n}>0;\ \mathbf{y}\in\partial R). \quad (2.5)$$

The kernel $A(\boldsymbol{\xi}'\to\boldsymbol{\xi};\mathbf{y})$ must satisfy the following

* On leave of absence from Applicazioni e Ricerche Scientifiche, Milano, and Università di Milano, Milano (Italy).

[1] C. Cercignani, J. Math. Phys. **8**, 1653 (1967).

[2] H. Grad, in *Rarefied Gas Dynamics*, J. A. Laurmann, Ed. (Academic Press Inc., New York, 1963).

[3] C. Cercignani, Phys. Fluids **10**, 2097 (1967).

requirement:

$$\int_{\xi'\cdot n>0} A(\xi \to \xi'; y)\omega^{\frac{1}{2}}(\xi')\,|\xi'\cdot n|\,d\xi' = |\xi\cdot n|\,\omega^{\frac{1}{2}}(\xi), \quad (\forall y \in \partial R;\ \xi\cdot n < 0), \quad (2.6)$$

which expresses the mass conservation at the walls. We note that the condition of given $f$ at the boundary, used in the previous paper,[1] does not satisfy this requirement, since the homogeneous form of this condition is $f = 0$, i.e., $A = 0$.

We now place some further restrictions upon the kernel. Primarily, we shall assume that, for any function $\sigma(\xi)$ of the molecular function such that the integrals (2.2) exist with $\beta = \sigma$, we have

$$\int_{\xi\cdot n>0} |\xi\cdot n|\,|Af|^2\,\sigma(\xi)\,d\xi \le C(\sigma) \times \int_{\xi\cdot n<0} |\xi\cdot n|\,|f|^2\,\sigma(\xi)\,d\xi, \quad (2.7)$$

where $C(\sigma)$ is a constant depending upon $\sigma$, which is not larger than 1 for $\sigma(\xi) = 1$. The meaning of Eq. (2.7) for $\sigma = 1$ is that the walls do not destroy entropy in absence of forcing terms. For $\beta \neq 1$, Eq. (2.2) has no evident physical interpretation, but will be shown to hold for the boundary condition of partial diffusion and partial reflection (including total diffusion).

Another property which will be required is connected with the equation

$$f = f_0 + A\tilde{f}, \quad (2.8)$$

where the unknown $f$ varies along the boundary and with $\xi$ $(\xi\cdot n > 0)$, and where

$$\tilde{f}(\xi, y) = f(\xi, \tilde{y}),. \quad (\xi\cdot n > 0;\ y \in \partial R). \quad (2.9)$$

Here $\tilde{y}$ is the conjugate point of $y$ to be defined presently. For a fixed $\xi$ the point conjugate of $x$ (where $x$ can be either inside $R$ or on $\partial R$) is the point $\tilde{y}(x)$ of the boundary such that $x - \tilde{y}$ has the same direction as $\xi$. (If more than one such point is present, the nearest one is to be taken.) We shall assume that the solution of Eq. (2.8) for a given $f_0$ at the boundary is determined up to an additive, constant multiple of $\omega^{\frac{1}{2}}(\xi)$; therefore we have an univocally determined solution $f$ which satisfies both Eq. (2.4) and the inequality

$$\|f\|_B \le k\,\|f_0\|_B, \quad (2.10)$$

provided that

$$(\omega^{\frac{1}{2}}/\beta, f_0)_B = 0, \quad (2.11)$$

where

$$(f, g)_B = \int_{\xi\cdot n>0} \beta(\xi)\,|\xi\cdot n|\,f(\xi, y)g(\xi, y)\,d\xi\,dy, \quad (2.12)$$

$$\|f\|_B^2 = (f,f)_B. \quad (2.13)$$

Later we show that the above requirements are indeed reasonable by showing that the boundary condition of partial diffusion and partial reflection satisfies them.

We start with the equation

$$\xi\cdot(\partial f/\partial x) = g(x, \xi), \quad (x \in R;\ \xi \in \Xi), \quad (2.14)$$

where $\Xi$ is the usual three-dimensional velocity space plus boundary conditions (2.5). Here $g$ is to be regarded as a given square-integrable function of $x$ and $\xi$ with respect to the weight $\beta(\xi)$. $g$ is also required to satisfy

$$\int g(\xi, x)\omega^{\frac{1}{2}}\,d\xi\,dx = 0 \quad (2.15)$$

because of the conservation of mass at the walls, Eq. (2.6). Equation (2.14) gives

$$f(\xi, x) = \frac{1}{\xi}\int_{|x-\tilde{y}|}^{0} g(\xi, x - \Omega t)\,dt + f(\xi, x - \Omega\,|x - \tilde{y}|), \quad (2.16)$$

where $\tilde{y}$ is the point conjugate of $x$ with respect to $\xi$, and $\Omega = \xi/|\xi|$. In particular, if $x = y$ is a point of the boundary (with $\xi\cdot n < 0$),

$$f(\xi, y) = \frac{1}{\xi}\int_{|y-\tilde{y}|}^{0} g(\xi, y - \Omega t)\,dt + \tilde{f}(\xi, y), \quad (\xi\cdot n < 0). \quad (2.17)$$

Now if $\xi\cdot n < 0$ at $y$, then $\xi\cdot n > 0$ at $\tilde{y}$ (the two normals $n$ and $\tilde{n}$ are, of course, differently directed). Therefore Eq. (2.17) gives the arriving distribution at $y$ in terms of the emerging distribution at $\tilde{y}$. If we substitute this into Eq. (2.5) we get Eq. (2.8), where

$$f_0 = Ag_0, \quad g_0 = \frac{1}{\xi}\int_{|y-\tilde{y}|}^{0} g(\xi, y - \Omega t)\,dt. \quad (2.18)$$

Equation (2.11) is satisfied because one can show [by means of Eqs. (2.6), (2.18)] that its left-hand side reduces to the left-hand side of Eq. (2.15), which holds by assumption. As a consequence, we can determine $f$ univocally by requiring that Eq. (2.4) is satisfied, and we have, according to Eq. (2.10),

$$\|f\|_B \le k\,\|f_0\|_B \le kC(\beta)\,\|g_0(-\xi, y)\|_B, \quad (2.19)$$

where Eqs. (2.6), (2.7), and (2.18) have been used. Then, from Eq. (2.16), we have

$$\int \xi^2\beta(\xi)\,|f|^2\,d\xi\,dx \le \int \left|\int_{|x-\tilde{y}|}^{0} g(\xi, x - \Omega t)\,dt\right|^2 \beta(\xi)\,d\xi\,dx + \int |f(\xi, \tilde{y})|^2\,\xi^2\beta(\xi)\,d\xi\,dx = I_1 + I_2. \quad (2.20)$$

Using the Schwartz inequality, we have

$$I_1 \le \int |\mathbf{x} - \tilde{\mathbf{y}}| \left[\int_{|x-\tilde{y}|}^{0} |g(\xi, \mathbf{x} - \boldsymbol{\Omega} t)|^2 \, dt\right] \beta(\xi) \, d\boldsymbol{\xi} \, d\mathbf{x}$$
$$< d \int dt \left[\int \beta(\xi) \, |g(\boldsymbol{\xi}, \mathbf{x} - \boldsymbol{\Omega} t)|^2 \, d\boldsymbol{\xi} \, d\mathbf{x}\right]$$
$$\le d^2 \int |g(\boldsymbol{\xi}, \mathbf{x})|^2 \, \beta(\xi) \, d\boldsymbol{\xi} \, d\mathbf{x}. \tag{2.21}$$

By a change of the integration variables and the use of Eq. (2.9), we have

$$I_2 < \int |\boldsymbol{\Omega} \cdot \mathbf{n}| \, |f(\boldsymbol{\xi}, \tilde{\mathbf{y}})|^2 \, \xi^2 \beta(\xi) \, d\boldsymbol{\xi} \, d\tilde{\mathbf{y}} \, dt$$
$$\le d \int |\boldsymbol{\Omega} \cdot \mathbf{n}| \, |f(\boldsymbol{\xi}, \mathbf{y})|^2 \, \xi^2 \beta(\xi) \, d\boldsymbol{\xi} \, d\mathbf{y}$$
$$\le kC(\beta) d \int |\boldsymbol{\Omega} \cdot \mathbf{n}| \left| \int_{|y-\tilde{y}|}^{0} g(\boldsymbol{\xi}, \mathbf{x} - \boldsymbol{\Omega} t) \, dt \right|^2 \beta(\xi) \, d\boldsymbol{\xi} \, d\mathbf{y}$$
$$\le kC(\beta) d^2 \int |\boldsymbol{\Omega} \cdot \mathbf{n}| \, |g(\boldsymbol{\xi}, \mathbf{x})|^2 \, \beta(\xi) \, d\boldsymbol{\xi} \, d\mathbf{y} \, dt$$
$$= kC(\beta) d^2 \int |g(\boldsymbol{\xi}, \mathbf{x})|^2 \, \beta(\xi) \, d\boldsymbol{\xi} \, d\mathbf{x}. \tag{2.22}$$

By using Eqs. (2.20), (2.21), and (2.22), we conclude that

$$\int \xi^2 |f|^2 \, d\boldsymbol{\xi} \, d\mathbf{x} \le [1 + kC(\beta)] d^2 \int |g|^2 \, d\boldsymbol{\xi} \, d\mathbf{x}, \tag{2.23}$$

which is immediately seen to be Eq. (2.1) with $\gamma = [1 + kC(\beta)]^{-1}$, if Eq. (2.14) is recalled.

Now we want to verify that the assumptions considered above are satisfied by the Maxwell boundary conditions of partly diffuse and partly specularly-reflected molecules. In this case we have

$$A(\boldsymbol{\xi}' \to \boldsymbol{\xi}) = (1 - \alpha) \delta[\boldsymbol{\xi} - \boldsymbol{\xi}' - 2\mathbf{n}(\mathbf{n} \cdot \boldsymbol{\xi}')] + \alpha \, |\boldsymbol{\xi}' \cdot \mathbf{n}| \, [2\pi\omega(\xi)\omega(\xi')]^{\frac{1}{2}},$$
$$(\boldsymbol{\xi}' \cdot \mathbf{n} < 0, \quad \boldsymbol{\xi} \cdot \mathbf{n} > 0, \quad \text{and} \quad 0 < \alpha \le 1). \tag{2.24}$$

First of all, by obvious manipulations and the Schwartz inequality, we have

$$\int_{\xi \cdot n > 0} |\boldsymbol{\xi} \cdot \mathbf{n}| \, |Af|^2 \, \sigma(\xi) \, d\boldsymbol{\xi}$$
$$\le (1 - \alpha)^2 \int |\boldsymbol{\xi} \cdot \mathbf{n}| \, |f|^2 \, \sigma(\xi) \, d\boldsymbol{\xi} + 2\pi\alpha^2 \left[\int_{\xi \cdot n > 0} |\boldsymbol{\xi} \cdot \mathbf{n}| \, \omega(\xi)\sigma(\xi) \, d\boldsymbol{\xi}\right] \times \left[\int_{\xi \cdot n < 0} |\boldsymbol{\xi} \cdot \mathbf{n}| \, \frac{\omega(\xi)}{\sigma(\xi)} \, d\boldsymbol{\xi}\right] \times \left[\int_{\xi \cdot n < 0} |\boldsymbol{\xi} \cdot \mathbf{n}| \, |f(\xi)|^2 \, \sigma(\xi) \, d\boldsymbol{\xi}\right]. \tag{2.25}$$

Equation (2.7) then follows, with

$$C(\sigma) = (1 - \alpha)^2 + 2\pi\alpha^2 \left[\int_{\xi \cdot n > 0} |\boldsymbol{\xi} \cdot \mathbf{n}| \, \omega(\xi)\sigma(\xi) \, d\boldsymbol{\xi}\right] \times \left[\int_{\xi \cdot n < 0} |\boldsymbol{\xi} \cdot \mathbf{n}| \, \frac{\omega(\xi)}{\sigma(\xi)} \, d\boldsymbol{\xi}\right]. \tag{2.26}$$

In particular, if $\sigma(\xi) = 1$, then

$$C(1) = (1 - \alpha)^2 + \alpha^2 \le 1. \tag{2.27}$$

Now we consider Eq. (2.8). We have

$$f(\boldsymbol{\xi}, \mathbf{y}) = f_0(\boldsymbol{\xi}, \mathbf{y}) + (1 - \alpha) f(\boldsymbol{\xi} - 2\mathbf{n}(\mathbf{n} \cdot \boldsymbol{\xi}), \tilde{\mathbf{y}}) + \alpha(2\pi\omega)^{\frac{1}{2}} \int_{\xi' \cdot n < 0} |\boldsymbol{\xi}' \cdot \mathbf{n}| \, [\omega']^{\frac{1}{2}} f(\boldsymbol{\xi}', \mathbf{y}) \, d\boldsymbol{\xi}'. \tag{2.28}$$

Now we split both $f$ and $f_0$ as follows:

$$f = \mu(\mathbf{y})[2\pi\omega(\xi)]^{\frac{1}{2}} + g(\boldsymbol{\xi}, \mathbf{y}),$$
$$f_0 = \mu_0(\mathbf{y})[2\pi\omega(\xi)]^{\frac{1}{2}} + g_0(\boldsymbol{\xi}, \mathbf{y}), \tag{2.29}$$

where $g(\boldsymbol{\xi}, \mathbf{y})$ [respectively, $g_0(\boldsymbol{\xi}, \mathbf{y})$] satisfies

$$\int_{\xi \cdot n > 0} g(\boldsymbol{\xi}, \mathbf{y}) \, |\boldsymbol{\xi} \cdot \mathbf{n}| \, [\omega(\xi)]^{\frac{1}{2}} \, d\boldsymbol{\xi} = 0 \quad (\forall \mathbf{y} \in \partial R). \tag{2.30}$$

The splitting is always possible if we take

$$\mu(\mathbf{y}) = \int f(\boldsymbol{\xi}, \mathbf{y}) \, |\boldsymbol{\xi} \cdot \mathbf{n}| \, [\omega(\xi)]^{\frac{1}{2}} \, d\boldsymbol{\xi} \tag{2.31}$$
$$= \mu_0(\mathbf{y}) + \frac{1}{\pi} \int_{\Omega' \cdot n < 0} |\boldsymbol{\Omega}' \cdot \mathbf{n}| \, \mu(\tilde{\mathbf{y}}) \, d\boldsymbol{\Omega}' + \int_{\xi' \cdot n < 0} |\boldsymbol{\xi}' \cdot \mathbf{n}| \, [\omega(\xi')]^{\frac{1}{2}} g(\boldsymbol{\xi}', \tilde{\mathbf{y}}) \, d\boldsymbol{\xi}', \tag{2.32}$$

$$g(\boldsymbol{\xi}, y) = g_0(\boldsymbol{\xi}, \mathbf{y}) + (1 - \alpha) g(\boldsymbol{\xi} - 2\mathbf{n}(\boldsymbol{\xi} \cdot \mathbf{n}), \tilde{\mathbf{y}}). \tag{2.33}$$

Now Eq. (2.33) can be solved by iteration and easily gives

$$\|g\|_B \le \frac{1}{\alpha} \|g_0\|_B, \quad (0 < \alpha < 1). \tag{2.34}$$

Therefore $g$ can be regarded as known in Eq. (2.25), which can be written

$$\mu(\mathbf{y}) = \nu_0(\mathbf{y}) + \frac{1}{\pi} \int_{\partial R(y)} \frac{|(\mathbf{y} - \mathbf{y}') \cdot \mathbf{n}| \, |(\mathbf{y}' - \mathbf{y}) \cdot \mathbf{n}'|}{|\mathbf{y} - \mathbf{y}'|^4} \mu(\mathbf{y}') \, d\mathbf{y}', \tag{2.35}$$

where

$$\nu_0(\mathbf{y}) = \mu_0(\mathbf{y}) + \int_{\xi' \cdot n < 0} \boldsymbol{\xi}' \cdot \mathbf{n} [\omega(\xi')]^{\frac{1}{2}} g(\boldsymbol{\xi}', \tilde{\mathbf{y}}) \, d\boldsymbol{\xi}'. \tag{2.36}$$

We have written $\mathbf{y}'$ in place of $\tilde{y}$ in Eq. (2.35) and used

the fact that $\mathbf{y} - \mathbf{y}' = \boldsymbol{\Omega}\,|\mathbf{y} - \mathbf{y}'|$,

$$d\boldsymbol{\Omega} = \cos(\mathbf{y}' - \mathbf{y}, \mathbf{n}')\, d\mathbf{y}',$$

and $\mathbf{n}'$ is the normal at $\mathbf{y}'$. $\partial R(\mathbf{y})$ is the part of $\partial R$ which is seen from $\mathbf{y}'$. Therefore it coincides with the whole boundary $\partial R$ for a simply connected region whose boundary has a fixed convexity.

Now we note that the integral equation (2.35) has a symmetric square-integrable kernel for any domain with bounded curvature. This is easily seen, provided that we take into account that both $\mathbf{y}'$ and $\mathbf{y}$ lie on the boundary $\partial R$. Therefore, when $\mathbf{y}' \to \mathbf{y}$, the singularity is much milder than it appears at first glance.

We observe now that $\mu = \text{const}$ is the only solution of Eq. (2.35) for $\nu_0 = 0$. Therefore a necessary and sufficient condition in order to have a solution (determined up to an additive constant) is that $\int \nu_0(\mathbf{y})\, d\mathbf{y} = 0$. Now we have

$$\int_{\partial R} \nu_0(\mathbf{y})\, d\mathbf{y} = \int_{\partial R} \mu_0(\mathbf{y})\, d\mathbf{y} + \int_{\partial R} \int_{\boldsymbol{\xi}' \cdot \mathbf{n} < 0} \boldsymbol{\xi}' \cdot \mathbf{n} [\omega(\xi')]^{\frac{1}{2}} g(\boldsymbol{\xi}', \tilde{\mathbf{y}})\, d\boldsymbol{\xi}'\, d\mathbf{y}. \tag{2.37}$$

The second integral can be transformed as follows:

$$\int_{\boldsymbol{\xi}' \cdot \mathbf{n} < 0} \boldsymbol{\xi}' \cdot \mathbf{n} [\omega(\xi')]^{\frac{1}{2}} g(\boldsymbol{\xi}', \mathbf{y}')\, d\boldsymbol{\xi}'\, d\mathbf{y}' = 0, \tag{2.38}$$

where $\mathbf{y}' = \tilde{\mathbf{y}}$ (accordingly $|\boldsymbol{\xi}' \cdot \mathbf{n}'|\, d\mathbf{y}' = |\boldsymbol{\xi}' \cdot \mathbf{n}|\, d\mathbf{y}$) and Eq. (2.30) have been taken into account. Therefore

$$\int_{\partial R} \nu_0(\mathbf{y})\, d\mathbf{y} = \int_{\partial R} \mu_0(\mathbf{y})\, d\mathbf{y} = \int f_0(\boldsymbol{\xi}, \mathbf{y})\, |\boldsymbol{\xi} \cdot \mathbf{n}|\, [\omega(\xi)]^{\frac{1}{2}}\, d\boldsymbol{\xi}\, d\mathbf{y} = 0 \tag{2.39}$$

if Eq. (2.11) is assumed to be satisfied.

We have, therefore, proved all the properties which we previously required. We note that, although we have used a three-dimensional language throughout, all the considerations apply to the two-dimensional case with little modification. (In the one-dimensional case the treatment is rather trivial.) In particular, $\beta(\xi)$ can now be a function of both the molecular speed and the absolute value $\xi_i$ of the projection of the molecular velocity onto a plane orthogonal to the symmetry axis.

## 3. BASIC EQUATIONS AND OPERATORS

Consider the problem in a bounded region for the separated time equation:

$$sf + \boldsymbol{\xi} \cdot (\partial f / \partial \mathbf{x}) + Lf = 0, \quad (\mathrm{Re}\, s \geq 0, \tag{3.1}$$

where $f$ is the perturbation of a basic Maxwellian, $F_M$ as specified by Eq. (2.3). We assume that the emerging distribution at the boundary is the sum of a given function plus a term related to the incoming distribution by Eq. (2.5), where $A$ satisfies the restrictions which have been pointed out in the previous section.

We shall restrict the problem to suitable collision operators $L$ by requiring that they can be split into two parts as follows:

$$Lf = -Kf + \nu(\xi) f, \tag{3.2}$$

where $\nu(\xi)$ is a multiplication operator and $K$ is a self-adjoint operator such that $\mu^{-\frac{1}{2}} K \mu^{-\frac{1}{2}}$ is a completely continuous operator in the Hilbert space $\mathfrak{H}$ of square-summable functions. Here $\mu = \mu(\xi)$ is a suitable function[4] of $\xi$ such that $\nu(\xi) \leq \mu(\xi)$; it can be taken equal to a multiple of $\nu(\xi)$ for collision operators with angular cutoff or equal to a multiple of some power of $\nu(\xi)$ for collision operators with radial cutoff.[1,3] It can also be shown[3] that, for three-dimensional problems, the first power of $\nu(\xi)$ is sufficient for the validity of the present treatment, even for the case of a radial cutoff—although in this case nothing is known about the complete continuity of $\nu^{-\frac{1}{2}} \mathcal{K} \nu^{-\frac{1}{2}}$. (We shall comment upon this in more detail later.)

One can always choose $\mu(\xi)$ (multiplying it by a constant larger than 1 if necessary) in such a way that one is able to split $Lf$ as follows:

$$Lf = -Hf + \mu f, \tag{3.3}$$

and $H$ is such that

$$0 \leq (f, Hf) \leq (\mu f, f). \tag{3.4}$$

Here parentheses denote, as usual, inner product in the Hilbert space of square-summable functions of $\xi$.

We shall also assume that $\xi/\mu(\xi)$ is bounded. This is necessarily true for radial cutoff and can be made true by suitably changing $\mu(\xi)$ for angular cutoffs.

As a consequence of the splitting in Eq. (3.3), one can rewrite Eq. (2.1) as an "integral" equation:

$$f = f_0 + UHf, \tag{3.5}$$

where

$$f_0 = \exp\left[ -\frac{\mu(\xi) + s}{\xi} (\mathbf{x} - \tilde{\mathbf{y}}) \cdot \boldsymbol{\Omega} \right] \tilde{f}. \tag{3.6}$$

$\tilde{f}$ is the inhomogeneous part of the boundary conditions, and $U$ is the inverse of the operator

$$\boldsymbol{\xi} \cdot (\partial / \partial \mathbf{x}) + [\mu(\xi) + s] \tag{3.7}$$

plus the homogeneous boundary conditions (2.5)

[4] Please note that the $\mu$'s and $\nu$'s used in this section (and in the following) have nothing to do with those used in Sec. 2.

and (2.4) together with all the specifications given in Sec. 2. For a steady problem we have merely to set $s = 0$.

Now we want to show the following lemma.

*Lemma I:* A nonzero numerical constant $\eta$ (depending upon the shape of the boundary) can be found such that, if

$$\rho(\boldsymbol{\xi}) = \{[\mu(\xi) + \operatorname{Re} s]^2 + \eta^2(\xi_1^2/d^2)\}^{\frac{1}{2}} \tag{3.8}$$

($\xi_i$ having been defined at the end, $d$ at the beginning of Sec. 2), then

$$((\rho Ug, Ug)) \leq ((\rho^{-1}g, g)), \tag{3.9}$$

where the double parentheses denote the scalar product in the Hilbert space of square-integrable functions of both $\mathbf{x}$ and $\boldsymbol{\xi}$. In order to show that this equality holds, we note that if we put

$$Ug = h, \tag{3.10}$$

then Eq. (3.9) becomes

$$((\rho h, h)) \leq \left(\left(\rho^{-1}\left\{\xi \cdot \frac{\partial h}{\partial \mathbf{x}} + [\mu(\xi) + s]h\right\},\right.\right.$$
$$\left.\left.\left\{\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}} + [\mu(\xi) + s]h\right\}\right)\right) = \text{rhs}, \tag{3.11}$$

where $h$ satisfies the boundary conditions (2.5). The right-hand side of Eq. (3.11) can now be written as follows:

$$\text{rhs} = \int \rho^{-1}\left|\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\right|^2 d\boldsymbol{\xi}\, d\mathbf{x}$$
$$+ 2\int \rho^{-1}[\mu(\xi) + \operatorname{Re} s]h\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\, d\boldsymbol{\xi}\, d\mathbf{x}$$
$$+ \int |\mu(\xi) + s|^2 \rho^{-1}|h|^2\, d\boldsymbol{\xi}\, d\mathbf{x}. \tag{3.12}$$

Now

$$\int \rho^{-1}[\mu(\xi) + \operatorname{Re} s]h\boldsymbol{\xi} \cdot \frac{\partial h}{\partial x}\, d\boldsymbol{\xi}\, d\mathbf{x}$$
$$= \int h\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\, d\boldsymbol{\xi}\, d\mathbf{x}$$
$$+ \int \{\rho^{-1}[\mu(\xi) + \operatorname{Re} s] - 1\}h\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\, d\boldsymbol{\xi}\, d\mathbf{x}. \tag{3.13}$$

The last term is smaller in absolute value than

$$\frac{\eta}{d}((\xi^2\rho^{-1}h, h))^{\frac{1}{2}}\left(\left(\rho^{-1}\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}, \boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\right)\right), \tag{3.14}$$

where $\eta$ is the constant which appears in $\rho$ and has to be suitably determined.

Therefore, if we apply the Gauss lemma to the first term in the right-hand side of Eq. (3.13) and use Eq. (2.7) with $C(1) = 1$, we obtain

$$\text{rhs} \geq \left[\left(\left(\rho^{-1}\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}, \boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\right)\right)^{\frac{1}{2}} - \eta d^{-1}((\xi^2\rho^{-1}h, h))^{\frac{1}{2}}\right]$$
$$\times \left(\left(\rho^{-1}\boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}, \boldsymbol{\xi} \cdot \frac{\partial h}{\partial \mathbf{x}}\right)\right)^{\frac{1}{2}} + ((\rho^{-1}(\mu + \operatorname{Re} s)^2 h, h)). \tag{3.15}$$

Now consider a closed neighborhood $0 \leq \eta \leq \eta_0$, where $\eta_0$ is some arbitrarily fixed number. Then, for any $\eta$ in this neighborhood, the corresponding $\rho$ satisfies Eq. (2.1) with $\beta = \rho^{-1}$ and some $\gamma = \gamma(\eta) > 0$. Since $\gamma$ depends continuously on $\eta$, it will reach a minimum value $\gamma_0$ which is also positive. We take for $\eta$ the smallest of the two numbers $\gamma_0/2$ and $\eta_0$. Then we can apply Eq. (2.1) twice and obtain

$$\text{rhs} \geq \gamma_0(2d)^{-1}((\xi^2\rho^{-1}h, h)) + ((\rho^{-1}(\mu + \operatorname{Re} s)^2 h, h)),$$
$$= (([(\mu + \operatorname{Re} s)^2 + \gamma_0(2d)^{-1}\xi^2]\rho^{-1}h, h))$$
$$\geq (\rho, h, h), \tag{3.16}$$

as was to be shown.

## 4. EXISTENCE AND UNIQUENESS IN THE LARGE

Now we present a proof of the following theorem.

*Theorem I:* The integral equation form of the Boltzmann equation, Eq. (2.6), has one and only one solution $f$ such that, for any given $\bar{f}$ such that $\rho^{\frac{1}{2}}f_0$ is square-integrable and $\rho^{\frac{1}{2}}f$ is also square-integrable, $\rho$ being defined by Eq. (3.8). This solution $f$ can, in principle, be obtained by a convergent iteration procedure.

The above theorem is an immediately obvious consequence of the contraction-mapping theorem and the following lemma.

*Lemma II:* The operator $UH$ is a contraction operator when acting upon functions belonging to the Hilbert space $\mathcal{K}$ of the functions, which are square-integrable with respect to the weight $\rho$.

To prove this lemma, we use Lemma I to obtain

$$((\rho UHf, UHf)) \leq ((\rho^{-1}Hf, Hf)) = ((Jg, Jg)), \tag{4.1}$$

where

$$J = \rho^{-\frac{1}{2}}H\rho^{-\frac{1}{2}}, \quad g = \rho^{\frac{1}{2}}f. \tag{4.2}$$

Now $J$ is a self-adjoint operator in $\mathcal{K}$ and has a norm smaller than 1. This can be easily shown in the case where $\mu^{-\frac{1}{2}}K\mu^{-\frac{1}{2}}$ is completely continuous (by the same argument as used in Ref. 1). However, in the case of a three-dimensional domain, we can avoid[2] using this

property of complete continuity, provided that $\mu(\xi)$ grows as $\xi$ when $\xi \to \infty$. This allows us to choose $\mu(\xi)$ equal to a multiple of $\nu(\xi)$ for any collision operator with a radial cutoff, as was anticipated in Sec. 3. In any case, under our assumptions we have

$$((\rho UHf, UHf)) \leq \alpha((g, g)) = \alpha((\rho f, f)), \quad (\alpha < 1). \tag{4.3}$$

## 5. CONVERGENCE OF THE SOLUTIONS OF KINETIC MODELS

The above existence and uniqueness theorem shows that a solution does exist for any Knudsen number. The theorem is also constructive in nature, since one can, in principle, write down the solution in the form of a series. However, the complicated character of the operators $U$ and $H$ makes this procedure hopeless from a practical standpoint. From this point of view one has to rely upon the use of model equations which approximate the collision operator $L$ by a simpler operator $L_N$, such that the equations can be handled by analytical or numerical procedures in a satisfactory way.

From the point of view of establishing a sound connection between the present theory and the practical procedure based on kinetic models, it is important to show that, if $L_N \to L$ $(N \to \infty)$ in some sense, then, in some related sense, the solution $f_N$ of the model equations converges to the solution $f$ of the actual Boltzmann equation. This is proved by the following theorem.

*Theorem II:* Let $L$ be a collision operator which can be split as in Eq. (3.3), let $L_N$ be a sequence of operators which can be similarly split (with the same $\mu$ as for $L$), and let

$$\|\rho^{-\frac{1}{2}}(L_N - L)\rho^{-\frac{1}{2}}\| \to 0 \quad \text{as} \quad N \to \infty, \tag{5.1}$$

$\rho$ being given by Eq. (3.8) with appropriate $\mu(\xi)$ and $\eta$.

Then the solution $f_N$ of the integral version of the model equation corresponding to $L_N$ tends (in the $\mathfrak{K}$ norm) to the solution $f$ of the analogous equation corresponding to $L$ with the same boundary conditions. [The norms appearing in Eqs. (5.1) are in the $\mathfrak{K}$ space.]

In order to prove the theorem, we consider the integral version of both the model and the actual equation:

$$f_N = f_0 + UH_N f_N, \tag{5.2}$$

$$f = f_0 + UHf. \tag{5.3}$$

We note that $f_0$ and $U$ are the same in both equations, and both $f$ and $f_N$ exist and are uniquely determined because of the existence and uniqueness theorem proved in Sec. 4.

Now if we put

$$f = f_N + r_N, \tag{5.4}$$

we have

$$r_N = U(H - H_N)f + UH_N r_N. \tag{5.5}$$

Because of Eq. (4.3) applied to $H_N$ we have

$$((\rho UH_N r_N, UH_N r_N)) \leq \alpha_N((\rho r_N, r_N)), \quad (\alpha_N < 1). \tag{5.6}$$

Now because of Eq. (5.1), which implies

$$\|\rho^{-\frac{1}{2}}(H_N - H)\rho^{-\frac{1}{2}}\| \to 0, \quad (N \to \infty), \tag{5.7}$$

we can choose $\alpha_N$ in a way such that $\alpha_N \to \alpha$ as $N \to \infty$, where $\alpha < 1$ is the constant in Eq. (4.3). Since $N = \infty$ is the only accumulation point of the sequence of the integers, it is now obvious that an $\alpha_0 < 1$ exists such that $\alpha_N \leq \alpha_0$ for any $N$. Accordingly,

$$((\rho UH_N r_N, UH_N r_N)) \leq \alpha_0((\rho r_N, \rho_N)), \quad (\alpha_0 < 1), \tag{5.8}$$

and one can apply the contraction-mapping theorem to Eq. (5.5) and deduce the following:

$$\begin{aligned} ((\rho r_N, r_N)) &\leq [1/(1 - \alpha_0)]((\rho U(H - H_N)f, U(H - H_N)f)), \\ &\leq [1/(1 - \alpha_0)]((\rho^{-1}(H - H_N)f, (H - H_N)f)), \\ &\leq [\epsilon_N/(1 - \alpha_0)]((\rho f, f)), \end{aligned} \tag{5.9}$$

where

$$\epsilon_N = \|\rho^{-\frac{1}{2}}(H - H_N)\rho^{-\frac{1}{2}}\| = \|\rho^{-\frac{1}{2}}(L - L_N)\rho^{-\frac{1}{2}}\|.$$

Since $\epsilon_N \to 0$ as $N \to \infty$ because of Eq. (5.1), it follows that $r_N \to 0$ in the $\mathfrak{K}$ norm, as was to be shown.

## 6. CONCLUDING REMARKS

It is felt that the present paper makes a significant contribution to the rigorous foundation of the mathematical theory of the linearized Boltzmann equation. In particular, the theorem concerning the convergence of the solutions of kinetic models should provide a rational basis for the use of such model equations which have proved very useful in the past few years in attacking the transition regime of rarefaction in a systematic way. What is needed now are estimates of the velocity of convergence, which, however, have to be found through methods essentially different from those employed in the present paper.

Another limitation of the present paper is that only

bounded domains have been considered. It is felt that this is not merely a limitation specific to the present paper, but concerns the possibility of treating general problems in unbounded domains with the present kind of approach. This circumstance, in turn, could reflect the difficulty of an uniformly valid linearization in unbounded domains. This matter is, of course, very important for applications of the Boltzmann equation to aerodynamics problems and, therefore, will be investigated in a future paper.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation, Grant No. GP-6084.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Irreducible Representations of the Semidirect-Product Group $K_n = A_n : S_n$ and the Harmonic-Oscillator Shell Model

PETER KRAMER*

*Instituto de Fisica, Universidad de México, México, D.F.*

(Received 24 October 1966)

If $A_n$ denotes the Abelian group of $n \times n$ unitary diagonal matrices and $S_n$ the symmetric group represented by $n \times n$ permutation matrices, the set of elements $ap$ with $a \in A_n$ and $p \in S_n$ form a group $K_n$ which is the semidirect product of $A_n$ and $S_n$. Irreducible representations of $K_n$ and the chains $K_n \supset S_n$ and $U_n \supset K_n$, with $U_n$ being the unitary group in $n$ dimensions, are discussed with applications in the harmonic-oscillator shell model.

## I. INTRODUCTION

In this paper the group $K_n$, generated by all products of unitary diagonal $n \times n$ matrices $a$ with $n \times n$ permutation matrices $p$, is considered. It is shown that all elements of $K_n$ are of the form $a \cdot p$ and, moreover, that $K_n$ is the semidirect product of its invariant Abelian subgroup $A_n$ and its subgroup $S_n$.

No systematic account of the group $K_n$ seems to have been published. There exist finite groups analogous to $K_n$ in which all elements $a$ have finite order $k$. Unimodular groups similar to $K_3$ and its finite counterparts have been studied recently by Fairbairn, Fulton, and Klink.[1] For general $n$ and all elements $a \neq e$ of order $k = 2$, a group with real elements results. This group is known as the hyperoctahedral group.[2] It may be regarded as the symmetry group of the $n$-dimensional Cartesian frame,[2] since its elements transform one system of Cartesian unit vectors

$$e^{(s)} : e_t^{(s)} = \delta_{s,t}(\pm 1), \quad s, t = 1, 2, \cdots n$$

into another. In complete analogy, the group $K_n$ may be regarded as the symmetry group of the unitary frame since its elements transform one system of unitary unit vectors

$$u^{(s)} : u_t^{(s)} = \delta_{s,t} \exp[i\delta^s], \quad s, t = 1, 2, \cdots, n$$

into another.

In Sec. II some properties of $K_n$-like classes and parametrization are discussed first. The construction of irreducible representations (IR) of semidirect-product groups has been discussed by McIntosh[3] and by Altmann[4] based on work of Mackey.[5] The present discussion follows the derivation given by McIntosh. The chain $K_n \supset K_{n-1} \supset \cdots \supset K_1$ is shown to be canonical.

In Sec. III the reduction of IR of $K_n$ in the chain $K_n \supset S_n$ is considered. This requires the reduction of representations of $S_n$ induced by subgroups of the form $S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$. By reciprocity theorems this reduction is related to the simpler problem $S_n \supset S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$, which is therefore considered first. The latter chain has recently been studied by Kaplan[6] and Horie.[7]

---

* Forschungsstipendiat der NATO. Present address: Institut für Theoretische Physik, Tübingen Universität, Deutschland.

[1] W. M. Fairbairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).

[2] H. S. M. Coxeter and W. O. Moser, *Generators and Relations for Discrete Groups* (Springer-Verlag, Berlin, 1965), p. 90.

[3] H. V. McIntosh, J. Mol. Spectry. **5**, 269 (1960); **10**, 51 (1963).

[4] S. L. Altmann, Phil. Trans. Roy. Soc. (London) **255**, 216 (1963).

[5] G. W. Mackey, Proc. Natl. Acad. Sci. US **35**, 537 (1949); Ann. Math. **55**, 101 (1952).

[6] I. G. Kaplan, Zh. Eksp. Teor. Fiz. **41**, 560 and 790 (1961) [Soviet Phys.—JETP **14**, 401 and 568 (1962)].

[7] H. Horie, J. Phys. Soc. Japan **19**, 1783 (1964).

bounded domains have been considered. It is felt that this is not merely a limitation specific to the present paper, but concerns the possibility of treating general problems in unbounded domains with the present kind of approach. This circumstance, in turn, could reflect the difficulty of an uniformly valid linearization in unbounded domains. This matter is, of course, very important for applications of the Boltzmann equation to aerodynamics problems and, therefore, will be investigated in a future paper.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation, Grant No. GP-6084.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Irreducible Representations of the Semidirect-Product Group $K_n = A_n : S_n$ and the Harmonic-Oscillator Shell Model

PETER KRAMER*

*Instituto de Fisica, Universidad de México, México, D.F.*

(Received 24 October 1966)

If $A_n$ denotes the Abelian group of $n \times n$ unitary diagonal matrices and $S_n$ the symmetric group represented by $n \times n$ permutation matrices, the set of elements $ap$ with $a \in A_n$ and $p \in S_n$ form a group $K_n$ which is the semidirect product of $A_n$ and $S_n$. Irreducible representations of $K_n$ and the chains $K_n \supset S_n$ and $U_n \supset K_n$, with $U_n$ being the unitary group in $n$ dimensions, are discussed with applications in the harmonic-oscillator shell model.

## I. INTRODUCTION

In this paper the group $K_n$, generated by all products of unitary diagonal $n \times n$ matrices $a$ with $n \times n$ permutation matrices $p$, is considered. It is shown that all elements of $K_n$ are of the form $a \cdot p$ and, moreover, that $K_n$ is the semidirect product of its invariant Abelian subgroup $A_n$ and its subgroup $S_n$.

No systematic account of the group $K_n$ seems to have been published. There exist finite groups analogous to $K_n$ in which all elements $a$ have finite order $k$. Unimodular groups similar to $K_3$ and its finite counterparts have been studied recently by Fairbairn, Fulton, and Klink.[1] For general $n$ and all elements $a \neq e$ of order $k = 2$, a group with real elements results. This group is known as the hyperoctahedral group.[2] It may be regarded as the symmetry group of the $n$-dimensional Cartesian frame,[2] since its elements transform one system of Cartesian unit vectors

$$e^{(s)} : e_t^{(s)} = \delta_{s,t}(\pm 1), \quad s, t = 1, 2, \cdots n$$

into another. In complete analogy, the group $K_n$ may be regarded as the symmetry group of the unitary frame since its elements transform one system of unitary unit vectors

$$u^{(s)} : u_t^{(s)} = \delta_{s,t} \exp[i\delta^s], \quad s, t = 1, 2, \cdots, n$$

into another.

In Sec. II some properties of $K_n$-like classes and parametrization are discussed first. The construction of irreducible representations (IR) of semidirect-product groups has been discussed by McIntosh[3] and by Altmann[4] based on work of Mackey.[5] The present discussion follows the derivation given by McIntosh. The chain $K_n \supset K_{n-1} \supset \cdots \supset K_1$ is shown to be canonical.

In Sec. III the reduction of IR of $K_n$ in the chain $K_n \supset S_n$ is considered. This requires the reduction of representations of $S_n$ induced by subgroups of the form $S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$. By reciprocity theorems this reduction is related to the simpler problem $S_n \supset S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$, which is therefore considered first. The latter chain has recently been studied by Kaplan[6] and Horie.[7]

---

* Forschungsstipendiat der NATO. Present address: Institut für Theoretische Physik, Tübingen Universität, Deutschland.

[1] W. M. Fairbairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).

[2] H. S. M. Coxeter and W. O. Moser, *Generators and Relations for Discrete Groups* (Springer-Verlag, Berlin, 1965), p. 90.

[3] H. V. McIntosh, J. Mol. Spectry. **5**, 269 (1960); **10**, 51 (1963).

[4] S. L. Altmann, Phil. Trans. Roy. Soc. (London) **255**, 216 (1963).

[5] G. W. Mackey, Proc. Natl. Acad. Sci. US **35**, 537 (1949); Ann. Math. **55**, 101 (1952).

[6] I. G. Kaplan, Zh. Eksp. Teor. Fiz. **41**, 560 and 790 (1961) [Soviet Phys.—JETP **14**, 401 and 568 (1962)].

[7] H. Horie, J. Phys. Soc. Japan **19**, 1783 (1964).

In Sec. IV the group $K_n$ is studied as a subgroup of the unitary group $U_n$ in $n$ dimensions. In particular, it is shown that the bases of irreducible representations (BIR) of $S_n$ constructed by Moshinsky[8] from special bases of $U_n$ result in a natural way as BIR of $K_n$. A physical application of the group $K_n$ results from the group theory of $n$ particles in a common harmonic-oscillator potential as discussed by Bargmann and Moshinsky[9] and by Kretzschmar.[10] The symmetry group of the corresponding Hamiltonian is the unitary group $U_{3n}$ in $3n$ dimensions. A subgroup $\mathfrak{U}_3 \times U_n$ of $U_{3n}$ can be introduced, where $\mathfrak{U}_3$ and $U_n$ are unitary-matrix groups affecting the 3 vector components and the $n$ particle indices, respectively. It is shown in Sec. IV of this paper that, from the use of the chain $U_n \supset K_n$ the concept of a multishell configuration results. Therefore the group $K_n$ appears as the symmetry group of the harmonic-oscillator shell model, and the BIR of the groups in the chain $U_n \supset K_n \supset S_n$ are $n$-particle shell-model states with permutational symmetry. The possible IR of $S_n$ for a fixed IR of $\mathfrak{U}_3$ can be found by using the characters of $K_n$. This offers a new way of extending the results of Elliott[11] and Kretzschmar[10] to arbitrary multishell configurations. Use of the group $K_n$ leads to a factorization of the harmonic-oscillator fractional-parentage coefficients for these configurations.

## II. GROUP $K_n = A_n : S_n$

### A. General Properties—Class Structure

Define the group $A_n$ as the group of unitary diagonal $n \times n$ matrices, $A_n = \{a\}$, where the matrix $a$ has elements

$$a_{st} = \delta_{s,t} \exp [i\alpha^s], \quad s, t = 1, 2, \cdots, n, \quad 0 \leq \alpha^s \leq 2\pi. \tag{1}$$

Similarly, represent the group $S_n$ by $n \times n$ permutation matrices $p$. If the permutation $p$ sends letter $s$ into letter $p(s)$, define the matrix $p$ by specifying its elements

$$p_{st} = \delta_{p(s),t}. \tag{2}$$

Next consider product matrices $ap$ with $a \in A_n$ and $p \in S_n$. Since $pap^{-1}$ has elements

$$(pap^{-1})_{st} = \delta_{s,t} \exp [i\alpha^{p(s)}], \tag{3}$$

it belongs to $A_n$. Then for $a, a' \in A_n$, $p, p' \in S_n$,

$$apa'p' = (apa'p^{-1})(pp') = a''p'',$$

with $a'' = apa'p^{-1} \in A_n$ and $p'' = pp' \in S_n$. The inverse of $ap$ is given by $(p^{-1}a^{-1}p)p^{-1}$, since

$$(p^{-1}a^{-1}p)p^{-1}ap = ap(p^{-1}a^{-1}p)p^{-1} = e.$$

*Theorem II.1:* The set of elements $\{ap\}$ with $a \in A_n$ and $p \in S_n$ form a group $K_n$. $K_n$ is the semidirect product of $A_n$ and $S_n$, i.e., $K_n = A_n : S_n$.

The second statement follows since[3]:

(a) all elements of $K_n$ are of the form $ap$, $a \in A_n$, $p \in S_n$;

(b) $A_n$ is an invariant subgroup of $K_n$, i.e., for $a \in A_n$ and $a'p' \in K_n$

$$(a'p')a(a'p')^{-1} = a'(p'ap'^{-1})a'^{-1} \in A_n;$$

(c) the intersection of $A_n$ and $S_n$ contains only the identity element $e$.

The group $K_n$ has $n$ continuous parameters $\alpha^s$, $s = 1, 2, \cdots, n$, and also has discrete elements; it is a mixed continuous group.[12]

What is the class structure of $K_n$? Transforming the elements $ap$ with an element $b \in A_n$,

$$b_{st} = \delta_{s,t} \exp [i\beta^s],$$

one finds that

$$bapb^{-1} = (bapb^{-1}p^{-1})p,$$

with

$$(bapb^{-1}p^{-1})_{st} = \delta_{s,t} \exp [i(\alpha^s + \beta^s - \beta^{p(s)})].$$

Therefore, new elements of the class of $ap$ are obtained by changing $\alpha^s$ into

$$'\alpha^s = \alpha^s + \gamma^s, \quad \gamma^s = \beta^s - \beta^{p(s)}. \tag{4}$$

These transformations are restricted by

$$\sum_{s=1}^{n} {}'\alpha^s = \sum_{s=1}^{n} (\alpha^s + \gamma^s) = \sum_{s=1}^{n} \alpha^s + \sum_{s=1}^{n} (\beta^s - \beta^{p(s)}) = \sum_{s=1}^{n} \alpha^s. \tag{4'}$$

As a consequence, there are $m - 1$ in-class parameters as defined by Murnaghan[13] for each disjoint $m$ cycle of the permutation $p$. Transforming now $ap$ by an element $q \in S_n$,

$$qapq^{-1} = qaq^{-1}qpq^{-1},$$

one has

$$(qaq^{-1})_{q^{-1}(s),t} = \delta_{q^{-1}(s),t} \exp [i\alpha^s],$$
$$(qpq^{-1})_{q^{-1}(s),t} = \delta_{q^{-1}(p(s)),t}. \tag{5}$$

Therefore, if $p$ sends $s$ into $p(s)$, then $qpq^{-1}$ sends $q^{-1}(s)$ into $q^{-1}(p(s))$ while the multiplying factors

[8] M. Moshinsky, J. Math. Phys. 7, 691 (1966).
[9] V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).
[10] M. Kretzschmar, Z. Phys. **157**, 433 (1960); **158**, 284 (1960).
[11] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128, and 562 (1958).
[12] E. P. Wigner, *Group Theory and its Application to the Quantum Theory of Atomic Spectra* (Academic Press Inc., New York, 1959), p. 89.
[13] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan Books, Washington, 1962), p. 19.

exp $[i\alpha^s]$ are unchanged. The new class elements obtained in this way have the same cycle structure as $p$ but with numbers $s$ replaced by $q^{-1}(s)$ in all cycles, whereas the factors exp $[i\alpha^s]$ "remain fixed" to the same cycle. From the in-class structure of $K_n$, one concludes the following:

*Theorem II.2:* A class of $K_n$ is determined by a definite cycle structure and by the values of $n_c$ class parameters, each belonging to one of the $n_c$ cycles. An element of $K_n$ within this class is determined by a permutation $p$ with the corresponding cycle structure and by the values of $n-n_c$ in-class parameters.

Finally, the eigenvalues of the matrices $ap$ are of interest for later applications. Since the matrix $ap$ is a direct-sum matrix of submatrices belonging to the disjoint cycles of $p$ up to a similarity transformation, it suffices to consider a cyclic $m \times m$ matrix $aq$ where $q$ is of order $m$ so that $q^m = e$. Then

$$(aq)^m = a(qaq^{-1})(q^2aq^{-2}) \cdots (q^{m-1}aq^{1-m})q^m$$

is diagonal with elements

$$\begin{aligned}(aq)^m_{st} &= \delta_{s,t} \exp\left[i(\alpha^s + \alpha^{q(s)} + \alpha^{q^2(s)} + \cdots + \alpha^{q^{m-1}(s)})\right] \\ &= \delta_{s,t} \exp\left[i\sum_{s=1}^{m}\alpha^s\right].\end{aligned}$$

It follows that the eigenvalues $\epsilon$ of $aq$ fulfill

$$\epsilon^m = \exp\left[i\sum_{s=1}^{m}\alpha^s\right].$$

The solutions of this equation can be written as

$$\epsilon_t = \exp\left[\frac{i}{m}\sum_{s=1}^{m}\alpha^s\right] \cdot r_{tm}, \quad t = 1, 2, \cdots, m, \quad (6)$$

where the $r_{tm}$ are the $m$th roots of unity. For each cycle of $p$, one obtains a set of eigenvalues of the type of Eq. (6).

### B. Irreducible Representations of the Group $K_n$

The present construction of irreducible representations of the group $K_n$ follows the discussion of irreducible representations of semidirect product groups given by McIntosh.[3]

First of all, the irreducible representations (IR) of the invariant subgroup $A_n$ or $K_n$ are determined. $A_n$ is the direct sum of $n$ Abelian groups $C_\infty$, and so its IR are the direct-product representations of the $n$ groups $C_\infty$. Since a group $C_\infty$ has the IR

$$\exp[iw\alpha], \quad w \text{ integer},$$

an IR of $A_n$ may be characterized by the weight $w = (w^1, w^2, \cdots, w^n)$ as

$$\langle w|\, a\, |w\rangle = \exp\left[i\sum_{s=1}^{n} w^s\alpha^s\right]. \quad (7)$$

For any weight $w$ and a permutation $p \in S_n$, one can form the new representation

$$p^{-1}[\langle w|\, a\, |w\rangle] = \langle w|\, pap^{-1}\, |w\rangle$$

characterized by the permuted weight

$$(w^{p^{-1}(1)}, w^{p^{-1}(2)}, \cdots w^{p^{-1}(n)}).$$

The set of representations of $A_n$ obtained from $\langle w|\, a\, |w\rangle$ by applying all $p \in S_n$ is called the *star* of the representation. One then defines a subgroup $W$ of $S_n$ which comprises all $h \in S_n$ which transform the IR $\langle w|\, a\, |w\rangle$ of $A_n$ into itself, i.e.,

$$W = \{h\}{:}h^{-1}[\langle w|\, a\, |w\rangle] = \langle w|\, a\, |w\rangle. \quad (8)$$

This group $W$ is often called the little group. It may also be called the group of the weight in analogy to the concept of the group of the wave vector used in the theory of space groups.

Choosing an IR $f_w$ of the group $W$, a product representation

$$\langle w|\, a\, |w\rangle\langle f_w|\, h\, |f_w\rangle$$

can be formed. This is in fact a representation of the semidirect product group $A_n{:}W$ with elements $ah$; since, on defining

$$\langle wf_w|\, ah\, |wf_w\rangle = \langle w|\, a\, |w\rangle\langle f_w|\, h\, |f_w\rangle, \quad (9)$$

one finds for $a, a' \in A_n$, $h, h' \in W$:

$$\begin{aligned}&\langle wf_w|\, aha'h'\, |wf_w\rangle \\ &\quad = \langle wf_w|\, (aha'h^{-1})hh'\, |wf_w\rangle \\ &\quad = \langle w|\, a\, |w\rangle\langle w|\, ha'h^{-1}\, |w\rangle\langle f_w|\, h\, |f_w\rangle\langle f_w|\, h'\, |f_w\rangle \\ &\quad = \langle w|\, a\, |w\rangle\langle f_w|\, h\, |f_w\rangle\langle w|\, a'\, |w\rangle\langle f_w|\, h'\, |f_w\rangle \\ &\quad = \langle wf_w|\, ah\, |wf_w\rangle\langle wf_w|\, a'h'\, |wf_w\rangle. \end{aligned} \quad (10)$$

This representation can be shown to be irreducible.[3]

For the group $K_n$, the group of the weight $W$ can easily be determined. On grouping equal weight components together, one may write

$$\begin{aligned}w = (&\underbrace{w_1, w_1, \cdots, w_1}_{n_1}, \underbrace{w_2, w_2, \cdots, w_2}_{n_2}, \cdots, \\ &\underbrace{w_j, w_j, \cdots, w_j}_{n_j}) \\ \equiv (w_1^{n_1}, w_2^{n_2}, \cdots, w_j^{n_j}) &\equiv (w_l^{n_l}).\end{aligned} \quad (11)$$

This standard weight is clearly left unchanged by all elements of the direct-sum subgroup

$$W = S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$$

of $S_n$, whereas all elements of $S_n$ not contained in this

subgroup change the weight. The IR of $W$ are direct product representations and may be denoted by

$$f_w = f^1 f^2 \cdots f^j, \tag{12}$$

where $f^l = [f_1 f_2 \cdots f_{n_l}]_l$ denotes the partition characterizing the IR of the group $S_{n_l}$. Similarly, the rows of the IR of the group $W$ may be denoted by

$$r_w = r_1 r_2 \cdots r_j, \tag{12'}$$

where $r_l = (r_1 r_2 \cdots r_{n_l})_l$ is the Yamanouchi[14] symbol characterizing the rows of the IR of $S_{n_l}$ in the order used by Horie.[7]

From the particular IR of the subgroup $A_n : W$ of $K_n$ given in Eq. (9), one may induce representations of $K_n$. To do this, choose $k$ right-coset generators $c_m, m = 1, 2, \cdots, k$ of $W$ in $S_n$. For $W = S_{n_1} \oplus S_{n_2}$, a standard choice given by Horie[7] is the set

$$\begin{aligned} c_1 &= e, \\ c_{q+1} &= (s_1, t_1)(s_2, t_2) \cdots (s_q, t_q), \\ 1 &\le q \le \min(n_1, n_2), \\ 1 &\le s_1 < s_2 < \cdots < s_q \le n_1, \\ n_1 + 1 &\le t_1 < t_2 < \cdots < t_q \le n_1 + n_2. \end{aligned} \tag{13}$$

For $W = S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$, first take the coset generators of $S_{n_1} \oplus S_{n_2}$ in $S_{n_1+n_2}$, then the coset generators of $S_{n_1+n_2} \oplus S_{n_3}$ in $S_{n_1+n_2+n_3}$, etc., and multiply all members of the first set by all members of the second set from the right, etc., to obtain in this way a number

$$\begin{aligned} k &= \frac{(n_1+n_2)!}{n_1!\, n_2!} \frac{(n_1+n_2+n_3)!}{(n_1+n_2)!\, n_3!} \cdots \frac{n!}{(n-n_j)!\, n_j!} \\ &= \frac{n!}{n_1!\, n_2! \cdots n_j!} \end{aligned} \tag{14}$$

of distinct right coset generators of $W$ in $S_n$. Then form the induced representation of $K_n$ given by

$$\begin{aligned} &\langle w f_w \bar{r}_w \bar{m} |\, ap \,| w f_w r_w m \rangle \\ &\quad = \langle w f_w \bar{r}_w |\, c_{\bar{m}} a p c_m^{-1} \,| w f_w r_w \rangle \delta(c_{\bar{m}} p c_m^{-1},\ h \in W) \\ &\quad = \langle w |\, c_{\bar{m}} a c_{\bar{m}}^{-1} | w \rangle \langle f_w \bar{r}_w |\, c_{\bar{m}} p c_m^{-1} \,| f_w r_w \rangle \\ &\qquad\qquad \times \delta(c_{\bar{m}} p c_m^{-1},\ h \in W). \end{aligned} \tag{15}$$

The $\delta$ appearing in this expression is one, if $c_{\bar{m}} p c_m^{-1}$ equals some element $h$ of the group of the weight $W$ and zero otherwise. McIntosh[3] gives the proof of the irreducibility of the representations of a semi-direct-product group induced in this way and shows that all irreducible representations are obtained by this method. A proof for the irreducibility of these representations for the case of $K_n$ can also be given by using the characters of the induced representations to be derived in the next section. Therefore one has the following theorem.

*Theorem II.3:* The irreducible representations of the group $K_n$ are given explicitly by the induced representations

$$\begin{aligned} &\langle (w_1^{n_1} w_2^{n_2} \cdots w_j^{n_j}) f^1 f^2 \cdots f^j \bar{r}_1 \bar{r}_2 \cdots \bar{r}_j \bar{m} |\, ap \\ &\qquad \times |(w_1^{n_1} w_2^{n_2} \cdots w_j^{n_j}) f^1 f^2 \cdots f^j r_1 r_2 \cdots r_j m \rangle \\ &\quad = \langle (w_1^{n_1} w_2^{n_2} \cdots w_j^{n_j}) |\, c_{\bar{m}} a c_{\bar{m}}^{-1} \,| (w_1^{n_1} w_2^{n_2} \cdots w_j^{n_j}) \rangle \\ &\qquad \times \langle f^1 \bar{r}_1 |\, h_1 \,| f^1 r_1 \rangle \times \langle f^2 \bar{r}_2 |\, h_2 \,| f^2 r_2 \rangle \times \cdots \\ &\qquad \times \langle f^j \bar{r}_j |\, h_j \,| f^j r_j \rangle \\ &\qquad \times \delta(c_{\bar{m}} p c_m^{-1},\ h_1 \oplus h_2 \oplus \cdots \oplus h_j \\ &\qquad\qquad \in S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}). \end{aligned} \tag{15'}$$

These IR are characterized by the standard weight $w = (w_l^{n_l})$ and by the IR $f^1 f^2 \cdots f^j$ of the group of the weight $W$.

If the bases of the irreducible representation (BIR) $(w, f_w)$ of $A_n : W$ are denoted by

$$|w f_w r_w\rangle,$$

the BIR of the IR $(w, f_w)$ of $K_n$ are given by

$$|w f_w r_w m\rangle = c_m^{-1} |w f_w r_w\rangle.$$

Note that, for $\bar{m} \neq m$,

$$\langle w f_w \bar{r}_w \bar{m} |\, w f_w r_w m \rangle = \langle w f_w \bar{r}_w |\, c_{\bar{m}} c_m^{-1} | w f_w r_w \rangle = 0. \tag{16}$$

For if $c_{\bar{m}} c_m^{-1} \in W$, then $c_{\bar{m}} = h c_m$ for some $h \in W$, so that the cosets $W c_{\bar{m}}$ and $W c_m$ coincide. Therefore $c_{\bar{m}} c_m^{-1} \notin W$, so it changes the weight. In this case the states $|w f_w \bar{r}_w\rangle$ and $c_{\bar{m}} c_m^{-1} |w f_w r_w\rangle$ have different weight so that their scalar product is zero.

### C. Projection Operators and Characters of the Group $K_n$

Using the IR of $K_n$ obtained in the last section, it is possible to write down the Wigner projection operators[15] needed for the determination of BIR of $K_n$. These projection operators involve sums over all group elements; for the mixed continuous group $K_n$ this implies an $n$-fold integral over the continuous parameters $\alpha^1, \alpha^2, \cdots, \alpha^n$ and a sum over all permutations $p \in S_n$.

Denote by $|G|$ the order of a group $G$, and by $|\kappa|$ the dimension of the irreducible representation $\kappa$ of $G$. The order of the mixed continuous groups $K_n$ can be measured by

$$|K_n| = (2\pi)^n n!. \tag{17}$$

[14] T. Yamanouchi, Proc. Phys. Soc. Japan **19**, 436 (1937).

[15] Reference 12, p. 87.

The Wigner projection operator for the IR $(w, f_w)$ of $K_n$ is given by

$$c^{w,f_w}_{r_w m, r_w m} = |w, f_w| \frac{1}{(2\pi)^n} \int da \frac{1}{n!} \times \sum_{p \in S_n} \langle w f_w r_w m | \, ap \, | w f_w r_w m \rangle^* \, ap$$
$$= \frac{1}{(2\pi)^n} \int_0^{2\pi} d\alpha^1 \, d\alpha^2 \cdots d\alpha^n \langle w | \, c_m a c_m^{-1} \, | w \rangle^* \, a \times \frac{|f^1| \cdot |f^2| \cdots |f^j|}{n_1! \, n_2! \cdots n_j!} \times \sum_{h_1 \oplus h_2 \ominus \cdots \ominus h_j \in S_{n_1} \oplus S_{n_2} \ominus \cdots \ominus S_{n_j}} \langle f^1 r_1 | \, h_1 \, | f^1 r_1 \rangle^* \times \langle f^2 r_2 | \, h_2 \, | f^2 r_2 \rangle^* \cdots \langle f^j r_j | \, h_j \, | f^j r_j \rangle^* \times c_m^{-1}(h_1 \oplus h_2 \oplus \cdots \oplus h_j) c_m . \quad (18)$$

The sum over $p \in S_n$ reduces to a sum over $h \in W$ because of the $\delta$'s in the induced representations; to each $p \in S_n$ there corresponds at most one $h \in W$ because of the uniqueness of the coset representative $hc_m$ of $c_{\bar{m}} p$. This leads to the factorization of the Wigner projection operator (18) as discussed in general by McIntosh.[3]

Next consider the characters of the representations (15) of $K_n$ obtained by taking the trace. One finds

$$\chi^{(w,f_w)}(ap) = \sum_{m=1}^{k} \langle w | \, c_m a c_m^{-1} \, | w \rangle \chi^{f_w}(h) \times \delta(c_m p c_m^{-1}, h \in W), \quad (19)$$

or, more explicitly,

$$\chi^{(w,f^1 f^2 \cdots f^j)}(ap) = \sum_{m=1}^{k} \langle w | \, c_m a c_m^{-1} \, | w \rangle \times \chi^{f^1}(h_1) \chi^{f^2}(h_2) \cdots \chi^{f^j}(h_j) \delta(c_m p c_m^{-1}, h_1 \oplus h_2 \oplus \cdots \oplus h_j \in S_{n_1} \oplus S_{n_2} \oplus S_{n_j}). \quad (19')$$

In this and the following section, frequent use will be made of the following theorem.

*Theorem II.4:* The character of a representation of a group $G$ induced by the irreducible representation $\kappa$ of a subgroup $H$ of $G$ is independent of the choice of coset generators of $H$ in $G$. Characters of representations of $G$ induced by the same irreducible representation $\kappa$ of conjugate subgroups $H, H', \cdots$ of $G$ are equal.

*Proof:* The proof of the first part of this theorem is given by Burrow.[16] The second part can be proved by showing that the corresponding induced representations are related by a similarity transformation.

Using this theorem, one can now test the irreducibility of the representations (15) of $K_n$ by calculating

$$\int_{a \in A_n} da \sum_{p \in S_n} \chi^{(\bar{w}, f_{\bar{w}})}(ap) \chi^{(w,f_w)*}(ap). \quad (20a)$$

The integral over the group $A_n$, i.e., over the angles $\alpha^1, \alpha^2, \cdots, \alpha^n$, gives zero unless $\bar{w}$ is obtained from $w$ by a permutation $q \in S_n$. In the latter case, the groups $\overline{W}$ and $W$ of the weights $\bar{w}$ and $w$ are conjugate as also are the groups $A_n : \overline{W}$ and $A_n : W$ from which the representations of $K_n$ were induced. Then it follows from the theorem that

$$\chi^{(\bar{w}, f_{\bar{w}})}(ap) = \chi^{(w,f_w)}(ap).$$

Without loss of generality, one may assume that $\overline{W} = W$ and that the coset generators are identical in both cases. Then

$$\int_{a \in A_n} da \sum_{p \in S_n} \chi^{(\bar{w}, f_{\bar{w}})}(ap) \chi^{(w,f_w)*}(ap)$$
$$= \delta(\bar{w}, w) \sum_{\bar{m}, m} \left[ \int_0^{2\pi} d\alpha^1 \, d\alpha^2 \cdots d\alpha^n \times \exp\left\{ i \sum_{s=1}^{n} (w^{c_{\bar{m}}^{-1}(s)} - w^{c_m^{-1}(s)}) \alpha^s \right\} \times \sum_{p \in S_n} \{ \chi^{f_w}(c_{\bar{m}} p c_{\bar{m}}^{-1}) \delta(c_{\bar{m}} p c_{\bar{m}}^{-1}, h \in W) \times \chi^{f_w}(c_m p c_m^{-1}) \delta(c_m p c_m^{-1}, h \in W) \} \right]$$
$$= \delta(\bar{w}, w)(2\pi)^n \frac{n!}{n_1! \, n_2! \cdots n_j!} \sum_{h \in W} \chi^{f_w}(h) \chi^{f_w *}(h)$$
$$= \delta(\bar{w}, w) \, \delta_{f_w, f_w} (2\pi)^n n!. \quad (20b)$$

Here $\delta(\bar{w}, w)$ equals one if $\bar{w}$ is obtained from $w$ by a permutation; it equals zero otherwise. The integration over $\alpha^1, \alpha^2, \cdots \alpha^n$ gives $\bar{m} = m$, and then the sum over $p \in S_n$ can be changed to a sum over $h \in W$. The result is the one expected for irreducible representations, $(2\pi)^n n!$ being the order of $K_n$.[17]

For later applications, the expressions for the characters of $K_n$ in two special cases are needed.

(a) If all components of the weight $w$ are different from each other, the group of this weight $(w_l)$ is $S_1 \oplus S_1 \oplus \cdots \oplus S_1$. The coset generators are all $q \in S_n$, so that the character of the IR $((w_l), [1]^n)$

[16] M. Burrow, *Representation Theory of Finite Groups* (Academic Press Inc., New York, 1965), pp. 77–80.

[17] Reference 12, p. 87.

of $K_n$ becomes

$$\chi^{(w_1, w_2, \cdots, w_n, [1]^n)}(ap) = \sum_{q \in S_n} \langle w_1, w_2, \ldots, w_n | qaq^{-1} | w_1, w_2, \ldots, w_n \rangle \delta(p, e). \tag{21}$$

(b) If all components of the weight $w$ are equal, i.e., $w = (w^n)$, the group of this weight is $S_n$ and the characters of the IR $(w^n, f)$ of $K_n$ are given by

$$\chi^{(w^n, f)}(ap) = \exp\left[iw \sum_{s=1}^{n} \alpha^s\right] \cdot \chi^f(p), \tag{22}$$

with $f = [f_1 f_2 \cdots f_n]$ denoting an IR of $S_n$.

### D. Chains of Groups $K_n \supset {}_{n-}K_1 \oplus K_1$ and $K_n \supset K_{n_1} \oplus K_{n_2}$

By restricting the elements $ap$ of $K_n$ to direct-sum matrices $a'p' \oplus a''$ containing the $(n-1) \times (n-1)$ matrices $a'p'$ of $K_{n-1}$ and the element $a''$ of $K_1$ along the diagonal, i.e.,

$$a'p' \oplus a'' = \left(\begin{array}{c|c} a'p' & 0 \\ \hline 0 & a'' \end{array}\right),$$

one obtains the elements of the subgroup $K_{n-1} \oplus K_1$ of $K_n$. The IR of this subgroup are the direct-product representations of the IR $(w', f_{w'})$ of $K_{n-1}$ and the IR $(w'', [1])$ of $K_1$. The multiplicity of these IR of $K_{n-1} \oplus K_1$ in an IR $(w, f_w)$ of $K_n$ is given by[18]

$$m[(w, f_w), (w', f_{w'}) \times (w'', [1])] = \frac{1}{(2\pi)^n} \int_{a \in A_{n-1}} da' \int_{a'' \in A_1} da'' \frac{1}{(n-1)!} \times \sum_{p' \in S_{n-1}} \chi^{(w, f_w)}(a'p' \oplus a'') \chi^{(w', f_{w'})^*}(a'p') \chi^{(w'', [1])^*}(a''). \tag{23}$$

Now denoting the weight $w$ by $(w_l^{n_l})$ for the sake of brevity, the integration over the elements of $A_{n-1}$ and $A_1$ gives zero unless $w' = (w_l^{n_l - \delta_{lk}})$ for some $k \leq j$. In this case, one may assume $k = j$. Because of Theorem II.4, this can always be achieved by changing the group of the weight $W$ into an appropriate conjugate subgroup without affecting the characters. For the IR of the group $K_n$, the standard coset generators of the group of the weight

$$W = S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_{j-1}} \oplus S_{n_j}$$

may be chosen. A subset of these coset generators forms a set of coset generators of the group of the weight $W' = S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_{j-1}} \oplus S_{n_j - 1}$ of the IR $(w_l^{n_l - \delta_{lj}})$ of $K_{n-1}$. Using Theorem II.4, one may assume that this set was chosen to determine the character of the IR $(w' f_{w'})$ of $K_{n-1}$. Putting $f_w = f^1 \cdots f^{j-1} f^j$ and $f_{w'} = {}'f^1 \cdots {}'f^{j-1} {}'f^j$, the expression (23) for the multiplicity becomes

$$\begin{aligned} & m[(w_l^{n_l}, f^1 f^2 \cdots f^{j-1} f^j), (w_l^{n_l - \delta_{lj}}, {}'f^1 {}'f^2 \cdots {}'f^{j-1} {}'f^j)(w'', [1])] \\ &= \frac{1}{(2\pi)^n} \sum_{m, m'} \int_0^{2\pi} d\alpha^1 \, d\alpha^2 \cdots d\alpha^{n-1} \\ &\quad \times \exp\left[i \sum_{s=1}^{n-1} (w^{c_m^{-1}(s)} - w^{c_{m'}^{-1}(s)}) \alpha^s\right] \\ &\quad \times \int_0^{2\pi} d\alpha^n \exp[i(w'' - w_j)\alpha^n] \frac{1}{(n-1)!} \\ &\quad \times \sum_{p' \in S_{n-1}} \chi^{f^1 f^2 \cdots f^{j-1} f^j}(c_m p' c_m^{-1}) \delta(c_m p' c_m^{-1}, h \in W') \\ &\quad \times \chi^{{}'f^1, {}'f^2 \cdots {}'f^{j-1}, {}'f^{j*}}(c_{m'} p' c_{m'}^{-1}) \delta(c_{m'} p' c_{m'}^{-1}, h' \in W') \\ &= \delta(w'', w_j) \frac{(n-1)!}{n_1! \cdots n_{j-1}! \, (n_j - 1)!} \frac{1}{(n-1)!} \\ &\quad \times \sum_{h \in W'} \chi^{f^1 f^2 \cdots f^{j-1} f^j}(h) \chi^{{}'f^1 {}'f^2 \cdots {}'f^{j-1} {}'f^{j*}}(h) \\ &= \delta(w'', w_j) \delta(f^j, {}'f^j \times [1]) \delta_{f^1, {}'f^1} \delta_{f^2, {}'f^2} \cdots \delta_{f^{j-1}, {}'f^{j-1}}. \end{aligned} \tag{23'}$$

From the remarks concerning the coset generators, it follows that the sum over $m$, $m'$ reduces to a sum over $m = m'$ after carrying out the integrations. Then the sum over $p' \in S_{n-1}$ can be changed to a sum over $h \in W'$. The result shows that the IR $(w_l^{n_l}, f^1 f^2 \cdots f^j)$ $K_n$ contains the IR $(w_l^{n_l - \delta_{lj}}, f^1 \cdots f^{j-1} {}'f^j)$ of $K_{n-1} \oplus K_1$ once and only once if the IR ${}'f^j$ of $S_{n_j - 1}$ is contained in the IR $f^j$ of $S_{n_j}$. This result is not affected by changing the order of the groups $S_{n_1}, S_{n_2}, \cdots S_{n_{j-1}}$, $S_{n_j}$ and $S_{n_1}, S_{n_2}, \cdots S_{n_{j-1}}, S_{n_j - 1}$ which make up $W$ and $W'$, respectively. Therefore the IR $(w_l^{n_l}, f^1, f^2 \cdots f^j)$ of $K_n$ contains the IR $(w_l^{n_l - \delta_{lh}}, f^1 \cdots f^{h-1} {}'f^h f^{h+1} \cdots f^j)$ of $K_{n-1} \oplus K_1$ once and only once if the IR ${}'f^h$ of $S_{n_h - 1}$ is contained in the IR $f^h$ of $S_{n_h}$. This proves the next theorem.

*Theorem II.5:* The chain of groups $K_n \supset K_{n-1} \oplus K_1$ is multiplicity-free.

Note that the IR $(w'', [1])$ of $K_1$ is completely determined by the IR of $K_n$ and of $K_{n-1}$. Therefore, Theorem II.5 implies that the chain of groups $K_n \supset K_{n-1} \supset \cdots \supset K_1$ is canonical. One could then choose bases which are BIR of all groups in this canonical chain. As will be shown later, the IR of $K_n$, when using this chain, allow us to determine the IR of $K_n$ which transform irreducibly under the groups in the chain

$$K_n \supset K_{n_1} \oplus K_{n_2}, \quad n_2 = n - n_1.$$

[18] Reference 12, p. 86.

This chain is needed in applications to the harmonic-oscillator shell model. The IR of $K_n$ corresponding to the canonical chain $K_n \supset K_{n-1} \supset \cdots \supset K_1$ are different from the IR (15) obtained by induction. While the canonical chain appears mathematically natural, the use of the induced representation leads to the factorization of the projection operators (18) and also simplifies the reduction $K_n \supset S_n$. This is to be discussed in the next section.

## III. CHAIN OF GROUPS $K_n \supset S_n$

The reduction of representations of $K_n$ in the chain $K_n \supset S_n$ will be seen in Sec. IV to be of considerable physical importance and therefore will be discussed in detail there. When restricting $K_n$ to $S_n$, one sees from (15′) that the IR of $K_n$ become reducible representations of $S_n$ induced from the IR $f_w$ of the subgroup $W$ of $S_n$. The reduction of representation of a group $G$ induced from an IR of a subgroup $H$ of $G$ is known to be connected with the reciprocal problem of reducing the IR of $G$ to the IR of $H$. Therefore, the simpler reduction $S_n \supset W$ is studied first and then used in the reduction $K_n \supset S_n$.

### A. Chain of Groups $S_n \supset S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$

This chain of groups has been discussed in particular by Kaplan[6] and by Horie.[7] In short, denote the IR of $W = S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$ by

$$\langle f_w \bar{r}_w | \, h \, | f_w r_w \rangle, \quad h \in W,$$

with $f_w$ and $\Gamma_w$ defined as in Eqs. (12) and (12′). The Young operator[7] for this IR is given by

$$c^{f_w}_{r_w' r_w} = \frac{|f_w|}{|W|} \sum_{h \in W} \langle f_w r_w' | \, h \, | f_w r_w \rangle^* h. \tag{24}$$

There exist transformation brackets which relate the IR

$$\langle f\bar{r} | \, p \, | fr \rangle, \quad p \in S_n$$

of $S_n$ to an equivalent IR which is reduced with respect to the subgroup $W$ of $S_n$, i.e., brackets such that

$$\langle f\bar{r} | \, p \, | fr \rangle = \sum_{\substack{\bar{\varphi} \bar{f}_w \bar{r}_w \\ \varphi f_w r_w}} \langle f\bar{r} \,|\, \bar{\varphi} \bar{f}_w \bar{r}_w \rangle \langle f \bar{\varphi} \bar{f}_w' \bar{r}_w | \, p \, | f \varphi f_w r_w \rangle \times \langle \varphi f_w r_w' \,|\, fr \rangle, \tag{25}$$

where $\bar{\varphi}$, $\varphi$ distinguish between repeated IR of $W$ and where, for $h \in W$,

$$\langle f \bar{\varphi} \bar{f}_w \bar{r}_w | \, h \, | f \varphi f_w \bar{r}_w \rangle = \delta_{\bar{\varphi}\varphi} \delta_{\bar{f}_w f_w} \langle f_w \bar{r}_w | \, h \, | f_w r_w \rangle.$$

These transformation brackets are related to the matrix elements of the Young operator in the representation $f$ of $S_n$ by

$$\langle fr | \, c^{f_w}_{r_w' r_w} \, | fr \rangle = \sum_{\varphi} \langle f\bar{r} \,|\, \varphi f_w r_w' \rangle \langle \varphi f_w r_w \,|\, fr \rangle, \tag{26}$$

as can be verified using Eq. (25). Assuming the transformation brackets to be elements of a unitary matrix, one has

$$\sum_{r} \langle \varphi f_w r_w' \,|\, f\bar{r} \rangle \langle f\bar{r} | \, c^{f_w}_{r_w' r_w} \, | fr \rangle = \langle \varphi f_w r_w \,|\, fr \rangle. \tag{26′}$$

This is a system of linear homogeneous equations for the transformation brackets. The number of possible solutions, which differ only in the label $\varphi$, can easily be determined by summing both sides of Eq. (26) over $\bar{r} = r$ and over $r_w' = r_w$ to obtain

$$\frac{|f_w|}{|W|} \sum_{h \in W} \chi^f(h) \chi^{f_w *}(h) = |f_w| \cdot \sum_{\varphi} 1. \tag{27}$$

On the left-hand side, the multiplicity $m(f, f_w)$ of the IR $f_w$ of $W$ in the IR $f$ of $S_n$ multiplied by $|f_w|$ appears, so that the number of possible values of $\varphi$ is given by

$$\sum_{\varphi} 1 = m(f, f_w). \tag{27′}$$

To solve the system of equations (26) for the transformation brackets the matrix elements of the Young operator (24) of $W$ in the Yamanouchi representation of $S_n$ are needed. These matrix elements can be evaluated by a general method which may be illustrated for the case $W = S_{n_1} \oplus S_{n_2}$. Following Kaplan and Horie, write the Yamanouchi symbol $r = (r_1 r_2 \cdots r_n)$ of the row of the IR of $S_n$ as

$$r = s_1 q_1,$$

where $s_1$ is the Yamanouchi symbol of the row of the IR of $S_{n_1}$ referring to numbers $1, 2, \cdots, n_1$ and where $q_1$ denotes the remainder of $\Gamma$. An element of $W = S_{n_1} \oplus S_{n_2}$ can be written as $h_1 h_2$ with $h_1 \in S_{n_1}$ and $h_1 \in S_{n_2}$ where $h_1$ and $h_2$ are taken to be $n \times n$ matrices. The IR of $W$ can be denoted by $f_w = {}'f^1 {}'f^2$ and its rows by $r_w = r_1 r_2$. The matrix elements of the Young operator of this IR of $W$ in the Yamanouchi representation of $S_n$ is given by

$$\begin{aligned}
&\langle f\bar{r} | \, c^{{}'f^1 {}'f^2}_{r_1' r_2', r_1 r_2} \, | fr \rangle \\
&\quad = \langle f \bar{f}^1 \bar{s}_1 \bar{q}_1 | \, c^{{}'f^1 {}'f^2}_{r_1' r_2', r_1 r_2} \, | f f^1 s_1 q_1 \rangle \\
&\quad = \frac{|{}'f^1| \cdot |{}'f^2|}{n_1! \, n_2!} \sum_{h_1 \in S_{n_1}, h_2 \in S_{n_2}} \langle f \bar{f}^1 \bar{s}_1 \bar{q}_1 | \, h_1 h_2 \, | f f^1 s_1 q_1 \rangle \\
&\qquad \times \langle {}'f^1 r_1 | \, h_1 \, | {}'f^1 r_1' \rangle^* \langle {}'f^2 r_2 | \, h_2 \, | {}'f^2 r_2' \rangle^* \\
&\quad = \delta_{f^1, {}'f^1} \delta_{\bar{s}_1, {}'r_1} \delta_{f^1, {}'f^1} \delta_{s_1, r_1} \\
&\qquad \times \frac{|{}'f^2|}{n_2!} \sum_{h_2 \in S_{n_2}} \langle f' f^1 r_1 \bar{q}_1 | \, h_2 \, | f' f^1 r_1 q_1 \rangle \\
&\qquad \times \langle {}'f^2 r_2 | \, h_2 \, | {}'f^2 r_2' \rangle^*.
\end{aligned} \tag{28}$$

The sum over $h_1 \in S_{n_1}$ can be evaluated since $S_{n_1}$ is a member of the canonical chain. To evaluate the

remaining sum over $h_2 \in S_{n_2}$, introduce a subgroup $S'_{n_2}$ conjugate to $S_{n_2}$ by

$$S_{n_2} = qS'_{n_2}q^{-1}, \quad q^{-1}(n - n_2 + s) = s, \quad s = 1, 2, \cdots, n_2.$$

Then by definition for $h_2 = qh'_2q^{-1} \in S_{n_2}$, $h'_2 \in S'_{n_2}$,

$$\langle {}'f^{2\prime}r_2 | \, h_2 \, | {}'f^2 r_2 \rangle \equiv \langle {}'f^{2\prime}r_2 | \, h'_2 \, | {}'f^2 r_2 \rangle.$$

Introducing $h_2 = qh'_2q^{-1}$ in Eq. (28) and decomposing the IR of $S_n$, one may now perform the sum over $h'_2 \in S'_{n_2}$, since $S'_{n_2}$ is a member of the canonical chain $S_n \supset S_{n-1} \supset \cdots \supset S_1$, to obtain

$$\begin{aligned}\langle f\bar{f}^1\bar{s}_1\bar{q}_1 | \, c'^{f^1,{}'f^1{}'f^2}_{r_1{}'r_2,r_1r_2} \, | f{}'f^1 s_1 q_1 \rangle &= \delta_{f^1,{}'f^1}\delta_{f^1,{}'f^1}\delta_{\bar{s}_1,{}'r_1}\delta_{s_1,r_1}\delta(f, f^1{}'f^2) \\ &\times \sum_{q_2} \langle f'f^1 r_1 \bar{q}_1 | \, q \, | f'f^{2\prime} r_2 q_2 \rangle \\ &\times \langle f'f^2 r_2 q_2 | \, q^{-1} \, | f'f^1 r_1 q_1 \rangle. \quad (28')\end{aligned}$$

Here $r_2q_2$ is defined similarly to $r_1q_1$, but, with respect to $S'_{n_2}$ and $\delta(f, f^1{}'f^2)$, is one if the IR $'f^2$ of $S'_{n_2}$ is contained in the IR $f$ of $S_n$ and zero if otherwise. The matrix elements of the Young operator are now expressed in terms of the matrix elements of the single permutation $q$ in the Yamanouchi representation. It is clear that this nonrecursive method can be generalized in the chain $S_n \supset S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$, so that the coefficients in the system of equations (26′) for the transformation brackets $\langle \varphi f_w r_w \, | \, fr \rangle$ are known. Recursive methods for obtaining the transformation brackets have been discussed by Horie.[7]

It should be noted that the present treatment of the chain $S_n \supset S_{n_1} \oplus S_{n_2}$ is based essentially on the existence of a canonical chain $S_n \supset S_{n-1} \supset \cdots \supset S_1$, and therefore may be applied to other matrix groups which admit a canonical chain $G_n \supset G_{n-1} \supset \cdots \supset G_1$. Since this applies in particular to the group $K_n$, the reduction $K_n \supset K_{n_1} \oplus K_{n_2}$ can be achieved by expressing the matrix elements of the corresponding Young operator in terms of the matrix elements of the permutation $q$ in the representation corresponding to the canonical chain $K_n \supset K_{n-1} \supset \cdots \supset K_1$.

### B. Chain of Groups $K_n \supset S_n$: Reciprocity

The IR of $K_n$ constructed in Sec. IIB, when restricting $K_n$ to $S_n$, become induced representations of $S_n$:

$$\begin{aligned}\langle f_w \bar{r}_w \bar{m} | \, p \, | f_w r_w m \rangle &= \langle f_w \bar{r}_w | \, c_{\bar{m}} p c_m^{-1} \, | f_w r_w \rangle \delta(c_{\bar{m}} p c_m^{-1}, h \in W). \quad (29)\end{aligned}$$

To reduce them to IR of $S_n$, apply the Young operator

$$c^f_{\bar{r}r} = \frac{|f|}{|S_n|} \sum_{p \in S_n} \langle f\bar{r} | \, p \, | fr \rangle^* p \quad (30)$$

for the IR $f$ of $S_n$ and assume that there are transformation brackets $\langle f_w r_w m \, | \, \bar{\varphi} fr \rangle$ which perform the reduction with $\bar{\varphi}$, a multiplicity index. Then

$$\begin{aligned}\langle f_w \bar{r}_w \bar{m} | \, c^f_{\bar{r}r} | f_w r_w m \rangle &= \sum_{\bar{\varphi}} \langle f_w \bar{r}_w \bar{m} \, | \, \bar{\varphi} f \bar{r} \rangle \langle \bar{\varphi} f r \, | f_w r_w m \rangle, \quad (31)\end{aligned}$$

which, analogous to Eq. (26), can be changed into a system of linear homogeneous equations for the transformation brackets. To evaluate the matrix elements of the Young operator, use the $\delta$ in Eq. (29) to change the sum over $p \in S_n$ to a sum over $h \in W$ by putting $h = c_{\bar{m}} p c_m^{-1}$. Then

$$\begin{aligned}\langle f_w \bar{r}_w \bar{m} | \, c^f_{\bar{r}r} \, | f_w r_w m \rangle &= \frac{|f|}{|S_n|} \sum_{h \in W} \langle f_w \bar{r}_w | \, h \, | f_w r_w \rangle \langle f\bar{r} | \, c_{\bar{m}}^{-1} h c_m \, | fr \rangle^* \\ &= \frac{|f|}{|S_n|} \sum_{r'\bar{r}'} \Big[ \langle fr | \, c_m^{-1} \, | fr' \rangle \langle f\bar{r}' | \, c_{\bar{m}} \, | f\bar{r} \rangle \\ &\quad \times \sum_{h \in W} \langle fr' | \, h \, | f\bar{r}' \rangle \langle f_w r_w | \, h \, | f_w \bar{r}_w \rangle^* \Big], \quad (32)\end{aligned}$$

using the unitarity of the representations. The sum appearing in the square brackets can be written in terms of the transformation brackets of the preceding section as

$$\begin{aligned}\sum_{h \in W} \langle fr' | \, h \, | f\bar{r}' \rangle \langle f_w r_w | \, h \, | f_w \bar{r}_w \rangle^* &= \frac{|W|}{|f_w|} \sum_{\varphi} \langle f\bar{r}' \, | \, \varphi f_w r_w \rangle \langle \varphi f_w \bar{r}_w \, | \, f\bar{r} \rangle.\end{aligned}$$

Then the matrix elements of the Young operator become

$$\begin{aligned}\langle f_w \bar{r}_w \bar{m} | \, c^f_{\bar{r}r} \, | f_w r_w m \rangle &= \frac{|f| \cdot |W|}{|S_n| \cdot |f_w|} \sum_{\varphi} \Big[ \Big\{ \sum_{r'} \langle fr | \, c_m^{-1} \, | fr' \rangle \langle fr' \, | \, \varphi f_w r_w \rangle \Big\} \\ &\quad \times \Big\{ \sum_{\bar{r}'} \langle \varphi f_w \bar{r}_w \, | f\bar{r}' \rangle \langle f\bar{r}' | \, c_{\bar{m}} \, | f\bar{r} \rangle \Big\} \Big]. \quad (32')\end{aligned}$$

Before discussing this result any further, we sum over $\bar{r}_w = r_w$ and $\bar{r} = r$, $\bar{m} = m$ and we do the same in Eq. (32). Since both expressions are equal, one finds

$$|f| \sum_{\varphi} 1 = |f| \sum_{\bar{\varphi}} 1, \quad (33)$$

or, in terms of the multiplicities,

$$m(f_w, f) = m(f, f_w). \quad (33')$$

*Theorem III.1 (Frobenius reciprocity theorem)*[19]*:* The multiplicity $m(f_w, f)$ of the IR $f$ of $S_n$ in there presentation induced by the IR $f_w$ of the subgroup $W$ of $S_n$ equals the multiplicity $m(f, f_w)$ of the IR $f_w$ of $W$ in the IR $f$ of $S_n$.

In the present case of $W = S_{n_1} \oplus S_{n_2} \oplus \cdots \oplus S_{n_j}$,

[19] G. Frobenius, Sitzber. Preuss. Akad. **501**, (1898). See also Ref. 16, p. 80.

the multiplicities of IR of $S_n$ are easily found by applying Littlewoods rules[20] to the product of partitions

$$f^1 \times f^2 \times \cdots \times f^j.$$

Now, comparing the two expressions on the right-hand side of Eqs. (31) and (32') it can be seen that the transformation brackets in Eq. (31) can actually be identified as a factor with the curly bracket expressions in Eq. (32'). With an appropriate choice of the factor the result is

$$\langle f_w r_w m \mid \varphi f r\rangle = \left[\frac{|f| \cdot |W|}{|S_n| \cdot |f_w|}\right]^{\frac{1}{2}} \sum_{r'} \langle \varphi f_w r_w \mid f r'\rangle \langle f r' \mid c_m \mid f r\rangle, \quad (34)$$

where $\tilde{\varphi}$ can now be identified with $\varphi$ because of Theorem III.1.

*Theorem III.2 (Extended reciprocity theorem):* The matrices which reduce the representation of $S_n$ induced by the IR $f_w$ of the subgroup $W$ of $S_n$ to IR $f$ of $S_n$ are given in terms of the matrices which reduce the IR $f$ of $S_n$ to IR $f_w$ of $W$ and in terms of the matrices of the coset generators of $W$ in the IR $f$ of $S_n$.

Note that the proof of this theorem employs the orthogonality of the bases $c_m^{-1} \mid f_w r_w\rangle$ of the induced representation.

## IV. GROUP $K_n$ AS A SUBGROUP OF THE UNITARY GROUP $U_n$: APPLICATIONS OF $K_n$

The elements $ap$ of $K_n$ are unitary matrices and therefore form a subgroup of the group $U_n$ of unitary $n \times n$ matrices. The reduction $U_n \supset K_n$ is of general interest since $K_n$ may be used as an intermediate group in the chain $U_n \supset S_n$. Moreover, it will be shown that there is a direct relationship between the chain of groups $U_n \supset K_n \supset S_n$ and the harmonic-oscillator shell model.

### A. Multiplicity of IR of $K_n$ in IR of $U_n$

First recall some properties of the unitary group $U_n$ in $n$ dimensions. An IR of $U_n$ is characterized by a partition $[h_1 h_2 \cdots h_n]$, where the numbers $h_s$ are nonnegative integers fulfilling $h_1 \geq h_2 \geq \cdots h_n \geq 0$. By restricting the unitary $n \times n$ matrices to the direct sums of unitary $(n-1) \times (n-1)$ matrices and an element of value one, a subgroup $U_{n-1} \oplus 1$ of $U_n$ results. The chain of groups $U_n \supset U_{n-1} \supset \cdots \supset U_1$ constructed in this way is canonical.[21,22] The partitions $[h_{1m} h_{2m} \cdots h_{mm}]$ of all subgroups $U_m$, $1 \leq m \leq n-1$ may therefore be used to determine the rows of the IR of $U_n$. These partitions are subject to the restrictions[21,22]

$$h_{1m} \geq h_{1m-1} \geq h_{2m} \geq h_{2m-1} \geq \cdots \geq h_{m-1m-1} \geq h_{mm}. \quad (35)$$

The corresponding BIR are Gel'fand[23] states derived explicitly by Moshinsky.[21] A Gel'fand state is also characterized by a weight

$$w = (w^1, w^2, \cdots, w^n),$$

which determines the IR of the subgroup $A_n$ of $U_n$. The weight components are given in terms of the partition numbers as

$$1 \leq t \leq n-1: \quad w^t = \sum_{s=1}^{t} h_{st} - \sum_{s=1}^{t-1} h_{st-1},$$

$$w^n = \sum_{s=1}^{n} h_s - \sum_{s=1}^{n-1} h_{sn-1}. \quad (36)$$

To find the multiplicity of an IR of $K_n$ in an IR of $U_n$, one may use the characters of both groups. The character of an element $u$ of $U_n$ is a symmetric polynomial in the eigenvalues $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ of $u$. Weyl[24] has given several forms for this polynomial. A particular form of the character of $U_n$ can be given by using the chain of groups $U_n \supset A_n$. The characters of an IR of the group $A_n$ for the diagonal unitary matrix with entries $\epsilon_s$ are given in terms of the weight $w$ as

$$\chi^{(w^1, w^2, \cdots, w^n)}(\epsilon_1 \epsilon_2 \cdots \epsilon_n) = \epsilon_1^{w^1} \epsilon_2^{w^2} \cdots \epsilon_n^{w^n}.$$

Since $A_n$ is a subgroup of $U_n$, the characters of the IR of both groups are related by

$$\chi^{[h_1 h_2 \cdots h_n]}(\epsilon_1 \epsilon_2 \cdots \epsilon_n) = \sum_{w} m([h_1 h_2 \cdots h_n], w) \chi^{w}(\epsilon_1 \epsilon_2 \cdots \epsilon_n)$$

$$= \sum_{(w^1, w^2, \cdots, w^n)} m([h_1 h_2 \cdots h_n], (w^1, w^2, \cdots, w^n)) \times \epsilon_1^{w^1} \epsilon_2^{w^2} \cdots \epsilon_n^{w^n}, \quad (37)$$

where $m([h_1 h_2 \cdots h_n], w)$ denotes the multiplicity of the weight $w = (w^1, w^2, \cdots, w^n)$. This multiplicity is determined by the number of solutions of the inequalities (35) for the given weight with components obtained from (36).

To reduce the IR of $U_n$ to those of $K_n$, one needs the characters of $U_n$ for the elements $ap$ of $K_n$. These characters are then polynomials in the eigenvalues $\epsilon_s$ of the matrices $ap$ which were discussed in Sec. IIA. The multiplicity of the IR $(w, f_w)$ of $K_n$ in

[20] D. E. Littlewood, *The Theory of Group Characters* (Oxford University Press, London, 1940), p. 94.
[21] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[22] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[23] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).
[24] H. Weyl, Math. Z. **23**, 271 (1925).

the IR $[h_1h_2\cdots h_n]$ of $U_n$ is then given by

$$m([h_1h_2\cdots h_n],(w,f_w)) = \frac{1}{(2\pi)^n}\int_{a\in A_n} da\,\frac{1}{n!}\sum_{p\in S_n}\chi^{[h_1h_2\cdots h_n]}(ap)\chi^{(w,f_w)*}(ap). \tag{38}$$

For example, taking a weight $w$, with all components different, from the expressions (21) and (37), for the corresponding characters one finds

$$m([h_1h_2\cdots h_n],(w,[1]^n)) = m([h_1h_2\cdots h_n],w). \tag{39}$$

There is another expression for the characters of $U_n$ given by Weyl[24] which has certain advantages for the reduction $U_n \supset K_n$. It employs the well-known connection between the IR $[h_1h_2\cdots h_n]$ of $U_n$ and the IR $[h_1h_2\cdots h_n]$ of a group $S_N$ where $N = h_1 + h_2 + \cdots + h_n$. Introducing the power sums

$$\sigma_l = \sum_{s=1}^{n}(\epsilon_s)^l, \quad l = 1,2,\cdots,n \tag{40}$$

and denoting the characters of the group $S_N$ for the class with cycle structure $(1^{i_1}2^{i_2}\cdots)$ by

$$\chi^{[h_1h_2\cdots h_n]}(1^{i_1}2^{i_2}\cdots),$$

Weyl's result is

$$\chi^{[h_1h_2\cdots h_n]}(\epsilon_1\epsilon_2\cdots\epsilon_n) = \sum_{(1^{i_1}2^{i_2}\cdots)}\frac{\chi^{[h_1h_2\cdots h_n]}(1^{i_1}2^{i_2}\cdots)}{1^{i_1}2^{i_2}\cdots i_1!\,i_2!\cdots}\sigma_1^{i_1}\sigma_2^{i_2}\cdots. \tag{41}$$

The eigenvalues $\epsilon_{\rho_s}$ of an element $ap$ with $p$ containing an $m$ cycle are given by

$$\epsilon_{\rho_s} = \exp\left[\frac{i}{m}\sum_{s=1}^{m}\alpha^{\rho_s}\right]\cdot r_{sm}, \quad s = 1,2,\cdots,m. \tag{6'}$$

Because of the roots of unity $r_{sm}$, these $\epsilon$'s contribute only to the power sums $\sigma_l$ with $l = \nu\cdot m$, $\nu = 1,2,\cdots$.

As an example of the use of Eq. (41), suppose that the partition $[h_1h_2\cdots h_n]$ of $U_n$ fulfills $h_1 + h_2 + \cdots + h_n = n$ and consider an IR of $K_n$ with $w = (1,1,\cdots,1)$. Then the group of the weight is $S_n$, and the character is given by Eq. (22) with $w = 1$. The $n$-fold integral over $\alpha^1,\alpha^2,\cdots,\alpha^n$ in Eq. (38) singles out those terms in $\chi^{[h_1h_2\cdots h_n]}(ap)$ which are proportional to

$$\exp\left[i\sum_{s=1}^{n}\alpha^s\right].$$

Now suppose that $p$ has the cycle structure $(1^{j_1}2^{j_2}\cdots n^{j_n})$ and consider the $j_m$ $m$-cycles with corresponding angles $\alpha_{\rho_{st}}$, $s = 1,2,\cdots,m$, $t = 1,2,\cdots,j_m$. The contributions of these $j_m$ $m$-cycles to the power sums $\sigma_l$ are of the form

$$\sigma_{\nu m} = m\sum_{t=1}^{j_m}\exp\left[i\nu\sum_{s=1}^{m}\alpha^{\rho_{st}}\right] + \cdots.$$

The only contributions which are not cancelled by the integration must have $\nu = 1$, i.e., must come from the power sum $\sigma_m$. Moreover, $\sigma_m$ has to be taken to the power $j_m$ to give the contribution

$$\sigma_m^{j_m} = m^{j_m}j_m!\exp\left[i\sum_{t=1}^{j_m}\sum_{s=1}^{m}\alpha^{\rho_{st}}\right] + \cdots.$$

This argument applies to all cycles of $p$. Carrying out the sum over $p\in S_n$, one then obtains

$$m([h_1h_2\cdots h_n],(1^n,[f_1f_2\cdots f_n])) = \delta_{[h_1h_2\cdots h_n],[f_1f_2\cdots f_n]}. \tag{42}$$

Using as BIR of $U_n$ the Gel'fand states mentioned earlier in this section, one can express the result as follows.

*Theorem IV.1 (Moshinsky):* The special Gel'fand states[8] with $h_1 + h_2 + \cdots + h_n = n$ and $w = (1,1,\cdots,1)$ are a basis of the irreducible representation $(1^n,[h_1h_2\cdots h_n])$ of $K_n$, and hence a basis of the irreducible representation $[h_1h_2\cdots h_n]$ of $S_n$.

This theorem has been derived by Moshinsky[8] by applying permutations to the special Gel'fand states without using the group $K_n$. He showed, by repeating the argument for all subgroups $U_m$ in the chain $U_n \supset U_{n-1} \supset \cdots \supset U_1$, that the special Gel'fand states correspond to the Yamanouchi basis of $S_n$ derived from the chain $S_n \supset S_{n-1} \supset \cdots \supset S_1$.

### B. Group $K_n$ and the Harmonic-Oscillator Shell Model

The group theory of $n$-particle states in a common harmonic-oscillator potential has been developed in several papers by Bargmann and Moshinsky[9] and by Kretzschmar.[10] The corresponding Hamiltonian had been shown by Baker[25] to be invariant under the unitary group $U_{3n}$ in $3n$ dimensions. By separating transformations affecting the three vector components from those affecting the $n$ particle indices,[9,10] a subgroup $\mathfrak{U}_3 \times U_n$ of $U_{3n}$ is obtained. The IR of $U_{3n}$, $\mathfrak{U}_3$, and $U_n$ are characterized by $[NO^{3n-1}]$, $[h_1h_2h_3]$, and $[h_1h_2h_3O^{n-3}]$, respectively, with $N = h_1 + h_2 + h_3$. The group $\mathfrak{U}_3$ is related to the quadrupole–quadrupole interaction introduced by Elliott[11]; its subgroup $\mathfrak{R}_3$ is related to the total orbital angular momentum.

A complete classification of the part connected with $U_n$ can be given by introducing Gel'fand states corresponding to the chain $U_n \supset U_{n-1} \supset \cdots \supset U_1$.[26] This chain, however, does not lead to states with permutational symmetry as needed for a systematic

[25] G. A. Baker, Jr., Phys. Rev. **103**, 1119 (1956).
[26] P. Kramer and M. Moshinsky, Nucl. Phys. **82**, 241 (1966).

construction of antimetric $n$-particle states by combination of orbital and spin–isospin states. It is clear that the group $K_n$ can be used to construct BIR of the groups in the chain $U_n \supset K_n \supset S_n$.

The IR of $K_n$ are characterized in part by the weight $w = (w^1, w^2, \cdots, w^n)$ up to permutations. For $n$-particle states classified according to the chain $U_{3n} \supset \mathfrak{U}_3 \times U_n$, this weight determines the number of quanta associated with particles $1, 2, \cdots, n$.[25] Therefore, if $w = (w^{n_1} w^{n_2} \cdots w^{n_j})$ denotes the standard weight (11), the numbers $n_1, n_2, \cdots, n_j$ are the occupation numbers of the shells $1, 2, \cdots, j$, characterized by a number $w_1, w_2, \cdots, w_j$ of quanta per particle. Then one has the following theorem.

*Theorem IV.2:* $n$-particle states classified according to the chains of groups

$$U_{3n} \supset \mathfrak{U}_3 \times U_n, \quad U_n \supset K_n \supset S_n$$

correspond to a multishell configuration as defined by Kaplan.[6]

The group $K_n$ may therefore be called the symmetry group of the harmonic-oscillator shell model. To obtain states with permutational symmetry, the reduction $K_n \supset S_n$ discussed in Sec. III can be used. Applications of the method are discussed elsewhere.[27]

The character technique developed in Sec. IVA together with Littlewood's rules admit the determination of the permutational content for an IR

$$[h_1 h_2 h_3 O^{n-3}]$$

of $U_n$ and, correspondingly, $[h_1 h_2 h_3]$ of $\mathfrak{U}_3$ in a multishell configuration. This allows the extension of the results of Elliott[11] and Kretzschmar,[10] who studied the same problem with different methods.

The use of the chain of groups $U_n \supset K_n \supset S_n$ leads to a factorization of the harmonic-oscillator fractional-parentage coefficients as discussed in Ref. 27. To write $n$-particle states with total number of quanta $N$ in terms of products of $n_1$-particle states with total number of quanta $N_1$ and $n_2$ $(= n - n_1)$-particle states with total number of quanta $N_2 = N - N_1$ in a common harmonic-oscillator potential requires finding the transformation brackets which connect the chains of groups

$$U_{3n} \supset \mathfrak{U}_3 \times U_n, \quad U_n \supset K_n \supset S_n, \tag{43}$$

$$U_{3n} \supset U_{3n_1} \oplus U_{3n_2} \supset (\mathfrak{U}_3' \times U_{n_1}) \oplus (\mathfrak{U}_3'' \times U_{n_2}),$$
$$U_{n_1} \supset K_{n_1} \supset S_{n_1}, \quad U_{n_2} \supset K_{n_2} \supset S_{n_2}. \tag{43'}$$

As an intermediate chain one may use

$$U_{3n} \supset \mathfrak{U}_3 \times U_n,$$
$$U_n \supset (U_{n_1} \oplus U_{n_2}) \supset (K_{n_1} \oplus K_{n_2}) \supset (S_{n_1} \oplus S_{n_2}). \tag{43''}$$

By passing first from Eq. (43′) to (43″) and then from (43″) to (43), the over-all transformation bracket factorizes into a general Wigner coefficient of $\mathfrak{U}_3$ and a transformation coefficient connecting different chains of subgroups of $U_n$. This latter coefficient admits two more factorizations by introducing the steps

$$U_n \supset K_n \supset S_n \supset S_{n-1} \supset \cdots \supset S_n, \tag{44}$$

$$U_n \supset K_n \supset S_n \supset S_{n_1} \oplus S_{n_2}, \tag{44'}$$

$$U_n \supset K_n \supset (K_{n_1} \oplus K_{n_2}) \supset (S_{n_1} \oplus S_{n_2}). \tag{44''}$$

The first step from (44) to (44′) requires the use of the results of Sec. III. The second step implies the reduction $K_n \supset K_{n_1} \oplus K_{n_2}$, briefly discussed in Sec. IIIA. Further study of the canonical chain of groups $K_n \supset K_{n-1} \supset \cdots \supset K_1$ is needed to find the corresponding transformation brackets. From the general remarks of this section, one sees that the group $K_n$ in the chain $U_n \supset K_n \supset S_n$ provides an important tool for the construction of $n$-particle states in the harmonic-oscillator shell model, and therefore also for the determination of harmonic-oscillator fractional-parentage coefficients in multishell configurations.

In the chain $U_n \supset S_n$, use of an intermediate group may correspond to different nuclear models.[27] Another such intermediate group is the orthogonal group $O_n$ in $n$ dimensions in the chain $U_n \supset O_n \supset S_n$.[10,26] A simple modification of this chain leads to translational invariant states,[26] much as the present chain leads to shell-model states.

## ACKNOWLEDGMENTS

The present work is part of a program developed in cooperation with Professor M. Moshinsky for the study of group theory as applied to problems of nuclear structure. The author would like to thank the Instituto de Física de la Universidad de México for the hospitality as well as the financial support offered to him during his stay in Mexico. He is indebted to Professor H. V. McIntosh for several clarifying discussions on irreducible representations of semidirect product groups. Finally he would like to thank Professor M. Moshinsky for numerous stimulating discussions on the group theory of harmonic oscillators which initiated the present work.

[27] P. Kramer and M. Moshinsky, in *Group Theory and Its Applications*, E. M. Loebl, Ed. (Academic Press Inc., New York, 1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 4 APRIL 1968

# Dimers on a Rectangular Lattice

R. J. BAXTER
*Research School of Physical Sciences, The Australian National University, Canberra, Australia*

(Received 17 July 1967)

A set of matrix equations is derived which yields the statistical mechanical properties of a system of monomers and dimers on a rectangular lattice in the thermodynamic limit. As the matrices are strictly of infinite dimensionality, the equations cannot be solved directly, but if they are restricted to be of finite and quite small dimensionality, very good approximations to the thermodynamic properties are obtained.

## 1. INTRODUCTION

One problem in statistical mechanics that has received attention in recent years is that of evaluating the number of ways of fitting dimers on to a plane rectangular lattice so that each dimer occupies two adjacent lattice sites and no lattice site is occupied by more than one dimer.

In the case when the lattice is completely filled with dimers, Kasteleyn[1] has shown that the system is equivalent to the Ising model in zero-magnetic field. It follows that the thermodynamic properties of the close-packed dimer system can be obtained exactly, and this has been done by Kasteleyn,[1,2] Temperley and Fisher,[3] and very recently by Lieb.[4] However, no solution of the general problem of an arbitrary density of dimers is as yet known.

In this paper, it is shown that in the thermodynamic limit the properties of the system of dimers and monomers (sites not occupied by a dimer) may be regarded as given by a set of matrix equations, where the matrices are of infinite dimensionality. Although it has not been found possible to solve these equations exactly, their derivation suggests that good approximations to the thermodynamic properties should be obtained by restricting them to be of finite and quite small dimensionality, and this is in fact found to be the case.

## 2. TRANSFER MATRIX

The transfer matrix of a monomer–dimer system has recently been obtained by Lieb.[4] However, for present purposes it is convenient to derive it in a slightly different form, which lends itself more easily to the subsequent variational approach.

Consider a rectangular lattice of $m$ rows and $n$ columns, with cyclic boundary conditions. Then, any arrangement of dimers on this lattice can be defined uniquely by specifying whether each bond, linking two adjacent lattice sites, is or is not occupied by a dimer. This specification can be given a mathematical formulation by associating two sets of parameters $\alpha_{ij}$ and $\beta_{ij}$ with the horizontal and vertical bonds, respectively, such that $\alpha_{ij}$ corresponds to the horizontal bond between sites $(i, j-1)$ and $(i, j)$, and $\beta_{ij}$ corresponds to the vertical bond between sites $(i-1, j)$ and $(i, j)$. If each of these parameters is assigned the value 0 when the corresponding bond is not occupied by a dimer, and the value 1 when it is occupied, then any set of such values of the $\alpha_{ij}$ and $\beta_{ij}$ corresponds to a unique arrangement of dimers on the lattice, and vice versa.

The condition that no lattice site be occupied by more than one dimer can be incorporated into this formalism by noting that the four bonds which have the lattice site $(i, j)$ as an end point correspond to the parameters $\alpha_{ij}$, $\alpha_{i,j+1}$, $\beta_{ij}$, $\beta_{i+1,j}$, so that not more than one parameter of this set can be equal to unity for each site $(i, j)$. Allowing for the fact that each bond occurs in two such sets, it follows that if an "activity" $s^2(t^2)$ is associated with each horizontal (vertical) dimer, then the partition function of the system can be written as

$$Z = \sum_{\{\alpha\},\{\beta\}} \prod_{i,j} K(\alpha_{ij}, \alpha_{i,j+1} \mid \beta_{ij}, \beta_{i+1,j}), \qquad (2.1)$$

where the summation is over all values (0 or 1) of the parameters $\alpha_{ij}$, $\beta_{ij}$, and the function $K(\alpha, \alpha' \mid \beta, \beta')$ is defined by:

$$\begin{aligned} K(\alpha, \alpha' \mid \beta, \beta') &= 1 \quad \text{IF } \alpha, \alpha', \beta, \beta' \text{ ARE ALL ZERO,} \\ &= s \quad \text{IF } \alpha \text{ OR } \alpha' \text{ IS UNITY, THE OTHER 3 PARAMETERS BEING ZERO,} \\ &= t \quad \text{IF } \beta \text{ OR } \beta' \text{ IS UNITY, THE OTHER 3 PARAMETERS BEING ZERO,} \\ &= 0 \quad \text{IF ANY TWO OF THE PARAMETERS ARE UNITY.} \end{aligned} \qquad (2.2)$$

If the single symbol $\tau_i$ is used to denote the $n$ parameters $\beta_{i1}, \cdots, \beta_{in}$ corresponding to the vertical bonds between rows $i-1$ and $i$, Eq. (2.1) can be

[1] P. W. Kasteleyn, J. Math. Phys. **4**, 287 (1963).
[2] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[3] H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961); M. E. Fisher, Phys. Rev. **124**, 1664 (1961).
[4] E. H. Lieb, J. Math. Phys. **8**, 2339 (1967).

written in a more illuminating form as

$$Z = \sum_{r_1, \cdots, r_m} V_{r_1 r_2} V_{r_2 r_3} \cdots V_{r_m r_1}, \tag{2.3}$$

where

$$V_{rr'} = V_{\beta_1 \cdots \beta_n | \beta_1' \cdots \beta_n'} = \sum_{\alpha_1 \cdots \alpha_n} \prod_{j=1}^{n} K(\alpha_j, \alpha_{j+1} \mid \beta_j, \beta_j'), \tag{2.4}$$

the cyclic symmetry implying that $\alpha_{n+1}$ is the same as $\alpha_1$.

It is apparent that the quantities $V_{rr'}$ may be regarded as elements of a $2^n$ by $2^n$ matrix $\mathbf{V}$, so that (2.3) is simply

$$Z = \operatorname{Tr} \mathbf{V}^m, \tag{2.5}$$

and in the limit of $m$ large this implies that

$$Z \sim \Gamma^m, \tag{2.6}$$

where $\Gamma$ is the greatest eigenvalue of $\mathbf{V}$.

The problem of evaluating the partition function of a lattice with infinitely many rows has therefore been reduced to that of determining the greatest eigenvalue of the matrix $\mathbf{V}$, which is known as the transfer matrix. It can be shown that the definition (2.4) of $\mathbf{V}$ is equivalent to that of Lieb,[4] so that it is possible to obtain $\Gamma$ exactly for the close-packed dimer system by following his derivation. In the general case, however, the problem has not as yet been solved by this technique.

In the absence of an exact general solution it is therefore of interest to consider alternative approaches to the problem of evaluating $\Gamma$. One such approach, which relies upon certain properties of the appropriate eigenvector of $\mathbf{V}$, will be presented in Sec. 3.

## 3. VARIATIONAL APPROACH

As the matrix $\mathbf{V}$ is symmetric, it is possible to obtain its greatest eigenvalue from the variational principle

$$\Gamma = \mathbf{x}'\mathbf{V}\mathbf{x}/\mathbf{x}' \cdot \mathbf{x}, \tag{3.1}$$

where the $2^n$-dimensional vector $\mathbf{x}$ is chosen so as to maximize the right-hand side of the equation. If $\mathbf{x}$ is regarded as arbitrary, this procedure is merely a restatement of the eigenvalue problem. However, certain information regarding the form of $\mathbf{x}$ is in fact available, for in the limit of $k$ large,

$$\mathbf{x} \propto \mathbf{V}^k \mathbf{y}, \tag{3.2}$$

where $\mathbf{y}$ is some vector which cannot be orthogonal to $\mathbf{x}$ but is otherwise arbitrary.

If $\mathbf{y}$ is chosen so that each of its elements is unity,[5] then on using the definition (2.4) of $\mathbf{V}$ and writing (3.2) explicitly, it is found that the equation can be written as

$$x(\beta_1, \cdots, \beta_n) \propto \operatorname{Tr} \{G_{\beta_1} G_{\beta_2} \cdots G_{\beta_n}\}, \tag{3.3}$$

where the matrices $\mathbf{G}_\beta$ (i.e., $\mathbf{G}_0$ and $\mathbf{G}_1$) are $2^k$ by $2^k$ matrices with elements

$$G^{\beta}_{\alpha_1 \cdots \alpha_k | \alpha_1' \cdots \alpha_k'} = \sum_{\beta_1 \cdots \beta_k} \prod_{i=1}^{k} K(\alpha_i, \alpha_i' \mid \beta_i, \beta_{i+1}), \tag{3.4}$$

the parameter $\beta_{k+1}$ in the summand of (3.4) being identical to $\beta$.

Allowing $k$ to tend to infinity, the eigenvector $\mathbf{x}$ is therefore of the form (3.3), where $\mathbf{G}_0$ and $\mathbf{G}_1$ are real, symmetric, infinite-dimensional matrices. It follows that if these symmetric matrices are regarded as arbitrary, then (3.3) should provide a good trial function for the variational principle (3.1). In fact, if they are allowed to have infinitely great dimensionality, then the results obtained should be formally exact. Further, and more significantly, if the matrices are restricted to be of finite and quite small dimensionality, then it should still be possible to obtain good approximations to $\Gamma$ and the thermodynamic properties. That this is the case will be shown in Sec. 4.

On substituting the form (3.3) of $\mathbf{x}$ into (3.1), an interesting result is found, namely that the thermodynamic limit of the number $n$ of columns becoming large can be obtained immediately (remember that the limit of the number $m$ of rows becoming large has already been taken). If the $\mathbf{G}_\beta$ are taken to be $r$ by $r$ matrices with elements $G^{\beta}_{\lambda\mu}$, then it is found that

$$\Gamma = \operatorname{Tr} \mathbf{S}^n / \operatorname{Tr} \mathbf{R}^n, \tag{3.5}$$

where $\mathbf{R}$ is the $r^2$ by $r^2$ matrix with elements

$$R_{\lambda\mu|\lambda'\mu'} = \sum_{\beta} G^{\beta}_{\lambda\lambda'} G^{\beta}_{\mu\mu'}, \tag{3.6}$$

and $\mathbf{S}$ is the $2r^2$ by $2r^2$ matrix with elements

$$S_{\alpha\lambda\mu|\alpha'\lambda'\mu'} = \sum_{\beta,\beta'} K(\alpha, \alpha' \mid \beta, \beta') G^{\beta}_{\lambda\lambda'} G^{\beta'}_{\mu\mu'}. \tag{3.7}$$

(Throughout this article, the indices $\alpha$ and $\beta$ are allowed to assume only the values 0 and 1.)

It follows that in the limit of $n$ large,

$$\Gamma \sim \kappa^n, \tag{3.8}$$

where

$$\kappa = \eta/\xi \tag{3.9}$$

and $\xi$ and $\eta$ are the greatest eigenvalues of $\mathbf{R}$ and $\mathbf{S}$, respectively.

If the associated eigenvectors of $\mathbf{R}$ and $\mathbf{S}$ are denoted by $X_{\lambda\mu}$ and $P^{\alpha}_{\lambda\mu}$, then the eigenvalue equations may be written in a convenient form by regarding

[5] If this choice of $\mathbf{y}$ does happen to be orthogonal to the eigenvector $\mathbf{x}$, there are an infinite number of other ways of choosing $\mathbf{y}$ so that $\mathbf{x}$ is of the form (3.3).

these quantities as elements of the matrices $\mathbf{X}$ and $\mathbf{P}_\alpha$. These matrices have the same dimensionality as the $\mathbf{G}_\beta$, so that the equations become

$$\sum_\beta \mathbf{G}_\beta \mathbf{X} \mathbf{G}_\beta = \xi \mathbf{X}, \tag{3.10}$$

$$\sum_{\alpha'\beta\beta'} K(\alpha, \alpha' \mid \beta, \beta') \mathbf{G}_\beta \mathbf{P}_{\alpha'} \mathbf{G}_{\beta'} = \eta \mathbf{P}_\alpha. \tag{3.11}$$

(Use has been made of the fact that the matrices $\mathbf{G}_\beta$ are symmetric.)

To complete the equations, it remains to choose $\mathbf{G}_0$ and $\mathbf{G}_1$ so as to maximize $\Gamma$, and hence $\kappa$. The easiest way to do this is to note that $\mathbf{R}$ and $\mathbf{S}$ are both symmetric matrices, so that the eigenvalue equations (3.10) and (3.11) may be regarded as derived from variational principles for $\xi$ and $\eta$, respectively. Using these principles, it follows from (3.9) that

$$\kappa = \frac{\operatorname{Tr} \mathbf{X}^2 \operatorname{Tr} \sum_{\alpha\alpha'\beta\beta'} K(\alpha, \alpha' \mid \beta, \beta') \mathbf{P}_\alpha \mathbf{G}_\beta \mathbf{P}_{\alpha'} \mathbf{G}_{\beta'}}{\operatorname{Tr} \sum_\beta \mathbf{X} \mathbf{G}_\beta \mathbf{X} \mathbf{G}_\beta \operatorname{Tr} \sum_\alpha \mathbf{P}_\alpha^2} \tag{3.12}$$

When (3.10) and (3.11) are satisfied, the right-hand side of (3.12) is stationary with respect to small variations of $\mathbf{X}$ and of the $\mathbf{P}_\alpha$, so that the condition that the $\mathbf{G}_\beta$ be chosen so as to maximize $\kappa$ implies that (3.12) is stationary with respect to small variations of the $\mathbf{G}_\beta$ when $\mathbf{X}$ and $\mathbf{P}_\alpha$ are held fixed. Defining

$$\xi = \kappa \operatorname{Tr} \sum_\alpha \mathbf{P}_\alpha^2 / \operatorname{Tr} \mathbf{X}^2, \tag{3.13}$$

it follows that

$$\sum_{\alpha\alpha'\beta'} K(\alpha, \alpha' \mid \beta, \beta') \mathbf{P}_{\alpha'} \mathbf{G}_{\beta'} \mathbf{P}_\alpha = \xi \mathbf{X} \mathbf{G}_\beta \mathbf{X}. \tag{3.14}$$

Provided the dimensionality of the five unknown matrices $\mathbf{G}_0$, $\mathbf{G}_1$, $\mathbf{X}$, $\mathbf{P}_0$, $\mathbf{P}_1$ is specified, (3.10), (3.11), and (3.14) form a closed system of equations from which $\kappa$ may be obtained for given values of $s$ and $t$. Further, these equations may all be regarded as derived from a single variational principle, namely that (3.12) be stationary with respect to small variations of the matrices.

One rather unsatisfactory feature of the transfer matrix technique is that it treats the rows of the lattice on a different basis from the columns, and so destroys the symmetry that exists between them. It is therefore very gratifying to find that this symmetry is restored in the above equations. To show this, introduce two further matrices $\mathbf{G}_0^*$ and $\mathbf{G}_1^*$ by defining

$$\mathbf{P}_\alpha = \mathbf{X}^{\frac{1}{2}} \mathbf{G}_\alpha^* \mathbf{X}^{\frac{1}{2}}. \tag{3.15}$$

On substituting this expression for $\mathbf{P}_\alpha$ into (3.12) it becomes apparent that $\kappa$ is unaltered by interchanging $\mathbf{G}_0$ and $\mathbf{G}_1$ with $\mathbf{G}_0^*$ and $\mathbf{G}_1^*$, except that $K(\alpha, \alpha' \mid \beta, \beta')$ is replaced by $K(\beta, \beta' \mid \alpha, \alpha')$. Inspection of the definition (2.2) shows that this is equivalent to interchanging the row and column fugacities $s^2$ and $t^2$, so that $\kappa$ is a symmetric function of these variables.

This symmetry enables a useful simplification to be made when $s$ and $t$ are equal (i.e., when the rows and columns are given equal weight), for then the $\mathbf{G}_\alpha^*$ are the same as the $\mathbf{G}_\alpha$, and Eq. (3.14) is the same as (3.11). As this is the case for which numerical results have been derived, it seems worthwhile to write the simplified equations explicitly. Defining

$$\mathbf{A} = \mathbf{X}^{\frac{1}{2}}, \tag{3.16}$$

$$\mathbf{G} = \mathbf{G}_0, \tag{3.17}$$

$$\mathbf{H} = \mathbf{G}_1, \tag{3.18}$$

(3.10) and (3.11) become

$$\mathbf{G}\mathbf{A}^2\mathbf{G} + \mathbf{H}\mathbf{A}^2\mathbf{H} = \xi \mathbf{A}^2, \tag{3.19}$$

$$\mathbf{G}\mathbf{A}\mathbf{G}\mathbf{A}\mathbf{G} + s\{\mathbf{G}\mathbf{A}\mathbf{G}\mathbf{A}\mathbf{H} + \mathbf{G}\mathbf{A}\mathbf{G}\mathbf{A}\mathbf{G} + \mathbf{H}\mathbf{A}\mathbf{G}\mathbf{A}\mathbf{G}\} = \eta \mathbf{A}\mathbf{G}\mathbf{A}, \tag{3.20}$$

$$s\mathbf{G}\mathbf{A}\mathbf{G}\mathbf{A}\mathbf{G} = \eta \mathbf{A}\mathbf{H}\mathbf{A}. \tag{3.21}$$

It can be seen from the above equations that not only $\mathbf{G}_0$ and $\mathbf{G}_1$, but also $\mathbf{X}$, can be chosen to be real and symmetric. It is further found that $\mathbf{X}$ is positive definite, so that $\mathbf{A}$ is real. Equations (3.19)–(3.21) can therefore be solved for the real symmetric matrices $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{H}$, together with $\xi$, $\eta$, and hence $\kappa$.

One useful feature of the variational principle (3.12) for $\kappa$ is that it enables its derivatives with respect to $s$ and $t$ to be evaluated directly in terms of the matrices. In particular, if $\rho$ is the mean number of dimers per lattice site given by

$$\rho = (2mn)^{-1} \left( s \frac{\partial}{\partial s} + t \frac{\partial}{\partial t} \right) \log Z \tag{3.22}$$

$$= \tfrac{1}{2} \left( s \frac{\partial}{\partial s} + t \frac{\partial}{\partial t} \right) \log \kappa, \tag{3.23}$$

then it follows from (3.12) and the above equations that when $s$ and $t$ are equal,

$$\rho/(2 - \rho) = \operatorname{Tr} (\mathbf{H}\mathbf{A}^2\mathbf{H}\mathbf{A}^2) / \operatorname{Tr} (\mathbf{G}\mathbf{A}^2\mathbf{G}\mathbf{A}^2). \tag{3.24}$$

## 4. NUMERICAL RESULTS

Inspection of Eqs. (3.19)–(3.21) reveals that they are unaffected by applying the same orthogonal transformation to each of the $r$ by $r$ matrices $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{H}$. It follows that it is possible to choose $\mathbf{A}$ to be diagonal, and this representation clearly reduces the amount of numerical work involved in evaluating the various matrix products. Further, if the matrices

$$\mathbf{L} = \mathbf{G}\mathbf{A}\mathbf{G} + s(\mathbf{G}\mathbf{A}\mathbf{H} + \mathbf{H}\mathbf{A}\mathbf{G}), \tag{4.1}$$

$$\mathbf{M} = s\mathbf{G}\mathbf{A}\mathbf{G} \tag{4.2}$$

are defined, then in this representation, Eqs. (3.20) and (3.21) can be written as

$$\mathbf{CQ} = \mathbf{QD}, \tag{4.3}$$

where $\mathbf{C}$ is the $2r$ by $2r$ symmetric matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{L} & \mathbf{M} \\ \mathbf{M} & \mathbf{O} \end{pmatrix}, \tag{4.4}$$

$\mathbf{Q}$ is a $2r$ by $2r$ orthogonal matrix of the form

$$\mathbf{Q} = \xi^{-\frac{1}{2}} \begin{pmatrix} \mathbf{AGA}^{-1} \cdots \\ \mathbf{AHA}^{-1} \cdots \end{pmatrix} \tag{4.5}$$

[using Eq. (3.19) to ensure correct normalization of the columns of $\mathbf{Q}$], and $\mathbf{D}$ is a $2r$ by $2r$ diagonal matrix of the form

$$\mathbf{D} = \eta \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \cdots \end{pmatrix}. \tag{4.6}$$

(The $r$ by $r$ submatrices of $\mathbf{Q}$ and $\mathbf{D}$ indicated by dots do not in principle need to be known, but their existence is relevant to the following discussion.)

In order to complete the equations, it is necessary to specify some normalization of the matrix $\mathbf{A}$ and the matrices $\mathbf{G}$ and $\mathbf{H}$ ($\kappa$ is of course independent of the normalization chosen). If $A_{11}$ and $G_{11}$ are both fixed to be unity, then it was found that the equations could be solved by the following iteration procedure, given some initial guess at $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{H}$:

A. Calculate $\mathbf{L}$, $\mathbf{M}$, and $\mathbf{C}$ from (4.1), (4.2), (4.4).

B. Calculate the eigenvalue matrix $\mathbf{D}$ and the eigenvector matrix $\mathbf{Q}$ of $\mathbf{C}$ from (4.3).

C. Arrange the eigenvalues and eigenvectors in some prescribed order, which must be fixed during the iteration procedure.

D. Using (4.6) and the condition that $A_{11}$ be unity, calculate $\eta$ and $\mathbf{A}$ from the first $r$ diagonal elements of $\mathbf{D}$.

E. Using (4.5) and the condition that $G_{11}$ be unity, calculate $\xi$ and the elements $G_{\lambda\mu}$, $H_{\lambda\mu}$ of $\mathbf{G}$, $\mathbf{H}$ such that $\mu \geq \lambda$ (i.e., the diagonal and top right elements of $\mathbf{G}$, $\mathbf{H}$) from the first $r$ columns of $\mathbf{Q}$.

F. Ensure that $\mathbf{G}$, $\mathbf{H}$ are symmetric by setting the bottom left elements equal to the top right. Then repeat the iteration procedure from stage A until it has effectively converged.

For $s \leq 0.25$ it was found that an adequate initial guess for this iteration procedure was to set all the elements of $\mathbf{A}$, $\mathbf{G}$, $\mathbf{H}$ other than $A_{11}$ and $G_{11}$ equal to zero, and that the $r$ by $r$ solution which maximized $\kappa$ was obtained by arranging the eigenvalues of $\mathbf{C}$ in numerically decreasing order (with the exception of the 3 by 3 case, where the preferred solution was obtained by choosing the first three diagonal elements of $\mathbf{D}$ to be the first, second, and fourth numerically largest eigenvalues). Having obtained solutions for three such values of $s$, it was possible to obtain good initial guesses for successively higher values by numerical extrapolation. For these latter cases it was appropriate to order the eigenvalues of $\mathbf{C}$ so as to correspond to the initial guesses obtained for them. Although the iteration procedure outlined above appeared to always converge, its rate of convergence decreased with increasing $s$ and for $s > 10$ it was preferable to use a Newton–Raphson method to solve the equations.

The calculations were performed on an IBM 360 computer using double-precision floating-point arithmetic accurate to 16 decimal places. A relative error for each of the $3r^2$ equations, represented by (3.19)–(3.21), was defined as the difference between the right- and left-hand sides divided by the absolute sum of all the additive terms in the equation (including the individual additive contributions to the matrix products). The iterations were assumed to have converged only when the relative error for each equation was less than $10^{-10}$. At each iteration, the values of $\kappa$ and $\rho$ obtained from (3.12) and (3.24) were evaluated and no change was observed in their first ten significant figures during the last iteration. It is therefore believed that the numerical values obtained for these quantities are accurate to ten significant figures.

In Table I, the values of $\kappa/s$ obtained from the $r = 1, 2, \cdots, 6$ approximations for $s = 1.0$, 4.0, 10.0, and $\infty$ are given. It is apparent that for each value of $s$, the successive approximations are tending to a limit, and that provided $s$ is not large the convergence is very rapid indeed.

The slowest rate of convergence is obtained when $s$ is infinitely large. This case is of particular interest

TABLE I. The values of $\kappa/s$ obtained from the $r = 1, 2, \cdots, 6$ approximations for $s = 1.0$, 4.0, 10.0, and $\infty$, together with the values obtained by geometric extrapolation from the 4, 5, 6 results.

| $r$ | $\kappa/s$ | | | |
|---|---|---|---|---|
| | $s = 1.0$ | $s = 4.0$ | $s = 10.0$ | $s = \infty$ |
| 1 | 1.937416664 | 1.444670083 | 1.356095932 | 1.299038106 |
| 2 | 1.940215341 | 1.460590906 | 1.381143005 | 1.335033348 |
| 3 | 1.940215344 | 1.460623453 | 1.381458447 | 1.337338271 |
| 4 | 1.940215351 | 1.460629381 | 1.381506501 | 1.337984697 |
| 5 | 1.940215351 | 1.460629397 | 1.381508315 | 1.338250017 |
| 6 | 1.940215351 | 1.460629398 | 1.381508512 | 1.338380390 |
| Extrapolated | 1.940215351 | 1.460629398 | 1.381508536 | 1.338506344 |

as it corresponds to the close-packed dimer system, for which exact results are available.[1-4] In particular, it can be deduced from Eq. (9) of Ref. 1 that

$$\lim_{s\to\infty} \kappa/s = \exp(g/\pi) = 1.338515152, \qquad (4.7)$$

where $g = 1^{-2} - 3^{-2} + 5^{-2} - 7^{-2} + \cdots$ is Catalan's constant.

With this information, inspection of the errors of each approximation indicates that they tend towards a geometric sequence. If it is assumed that

$$(\kappa/s)_{r\text{-approx}} = (\kappa/s)_{\text{exact}} - ay^r, \qquad (4.8)$$

then the $r = 4, 5, 6$ solutions can be used to obtain an extrapolated value of $\kappa/s$. This value is also given in Table I and it is apparent that when $s$ is infinite it agrees with the exact value (4.7) to effectively six significant figures.

In Table II, the estimated true values of $\kappa/s$ and the density $\rho$ for the dimer lattice are given for values of $s^{-1}$ ranging from 0.0 to 5.0. The values of $\kappa/s$ obtained from the $r = 5$ and $r = 6$ approximations agree to 10 significant figures for $s^{-1} \geq 0.3$, while the values of $\rho$ agree to this accuracy for $s^{-1} \geq 0.5$, so that these are believed to be the true values. For larger values of $s$, the values given are obtained by extrapolation from the $r = 4, 5, 6$ approximations and differ from the $r = 6$ approximation only in the last figure given (with the exception of the value of $\kappa/s$ for $s$ infinite).

TABLE II. The values of $\kappa/s$ and $\rho$ for various values of $s^{-1}$.

| $s^{-1}$ | $\kappa/s$ | $\rho$ |
|---|---|---|
| 0.00 | 1.33851 | 0.5000000000 |
| 0.02 | 1.345384 | 0.497088 |
| 0.05 | 1.3580387 | 0.4918108 |
| 0.10 | 1.38150854 | 0.48228188 |
| 0.20 | 1.433197282 | 0.462353025 |
| 0.30 | 1.488879377 | 0.4423290570 |
| 0.40 | 1.547451104 | 0.4226907932 |
| 0.50 | 1.608381870 | 0.4036382364 |
| 0.60 | 1.671357163 | 0.3852640483 |
| 0.80 | 1.802655618 | 0.3506931614 |
| 1.00 | 1.940215351 | 0.3190615546 |
| 1.50 | 2.307013548 | 0.2521316647 |
| 2.00 | 2.700426573 | 0.2003225902 |
| 2.50 | 3.114544308 | 0.1605891249 |
| 3.00 | 3.544765863 | 0.1301534490 |
| 3.50 | 3.987516671 | 0.1067369571 |
| 4.00 | 4.440052963 | 0.08857621602 |
| 4.50 | 4.900283946 | 0.07434949046 |
| 5.00 | 5.366618412 | 0.06308145191 |

It is therefore believed that the values listed are in error by at most $\pm 1$ in the last figure.

It is clearly not possible to deduce rigorously from the above working whether or not the dimer system undergoes a phase transition, but the fact that the successive approximations vary smoothly with $s$ and tend towards the known results at both the high- and low-density limits suggests very strongly that no transition occurs in this system.