On the Two Majorana Representations*

MARKO I. PAVKOVIĆ
Department of Physics & Astronomy, Tel-Aviv University, Ramat-Aviv, Israel

(Received 23 April 1969; Revised Manuscript Received 14 October 1969)

An interesting property of the Majorana representations, symbolized in the Clebsch–Gordan equations $(0, \frac{1}{2}) \otimes (\frac{1}{2}, \frac{3}{2}) = (\frac{1}{2}, 0) \oplus \cdots$ and $(\frac{1}{2}, 0) \otimes (\frac{1}{2}, \frac{3}{2}) = (0, \frac{1}{2}) \oplus \cdots$, is reported. $(\frac{1}{2}, \frac{3}{2})$ refers to the basic spinorial nonunitary representation of the $SL(2, c)$ group in Gel'fand's notation, while $(0, \frac{1}{2})$ and $(\frac{1}{2}, 0)$ are two Majorana representations, again in the same notation. The Majorana representations represent the only solution of the equation $X \otimes (\frac{1}{2}, \frac{3}{2}) = \sum_a \oplus Y_a$, in the sense that, of all the unitary irreducible representations of the $SL(2, c)$ group, only $(0, \frac{1}{2})$, or $(\frac{1}{2}, 0)$ multiplied tensorially with $(\frac{1}{2}, \frac{3}{2})$ will yield a unitary result—at least one of the Y's unitary.

The two Majorana representations are prominent members among the unitary infinite-dimensional irreducible representations of the $SL(2, c)$ group. They possess several exceptional properties of which the most familiar is the feature that, of all the unitary irreducible infinite-dimensional representations of $SL(2, c)$, only the Majorana representations allow the construction of an operator with the relativistic tensor-transformation properties of a 4-vector.[1]

We have found another quite remarkable property of the Majorana representations which does not seem to have been reported in the literature to date. Briefly, if s denotes the basic spinorial representation of $SL(2, c)$, $s = (\frac{1}{2}, \frac{3}{2})$ in Gel'fand's notation, and Y_a, $a = 1, 2, \cdots$, and X are infinite-dimensional irreducible representations of the same group, of which X and at least one of the Y's are unitary, and with the significance of unknowns in the Clebsch–Gordan equation

$$X \otimes s = \sum_a \oplus Y_a, \tag{1}$$

then the only solution of (1) is supplied by the Majorana representations $a = 1, 2$, $X = (0, \frac{1}{2})$, $Y_1 = (\frac{1}{2}, 0)$ or $X = (\frac{1}{2}, 0)$, $Y_1 = (0, \frac{1}{2})$, again in the Gel'fand notation.

We first explain the content of the stated theorem in more detail and then provide the proof. The properties of the $SL(2, c)$ group representations have been intensively investigated in the past, and, in particular, a complete classification of the finite-dimensional nonunitary as well as the infinite-dimensional unitary irreducible representations have been made.[1] In Gel'fand's classification, which is used throughout, the irreducible representations are labeled by the ordered pairs of parameters (k_0, c), whose relationship to the eigenvalues of two invariant operators is

$$\begin{aligned} \mathbf{L}^2 - \mathbf{Q}^2 &= k_0^2 + c^2 - 1, \\ \mathbf{L} \cdot \mathbf{Q} &= -ik_0 c. \end{aligned} \tag{2}$$

Here, k_0 runs through the set of integers or halves of odd integers, and represents the lowest angular-momentum quantum number contained in an irreducible representation, while c is in general a complex number. The six objects L_i and Q_i, $i = 1, 2, 3$, are of course the $SL(2, c)$ group generators which form the Lie algebra commutation relations

$$\begin{aligned} [L_i, L_j] &= i\epsilon_{ij}^k L_k, \\ [L_i, Q_j] &= i\epsilon_{ij}^k Q_k, \\ [Q_i, Q_j] &= -i\epsilon_{ij}^k L_k. \end{aligned}$$

Note that the representation labeled by $(-k_0, -c)$ is indistinguishable from the representation (k_0, c), provided the eigenvalues of the two listed invariant operators are sufficient for specifying uniquely the irreducible representations of $SL(2, c)$. In what follows, the representations (k_0, c) and $(-k_0, -c)$ will be considered equivalent.[1]

The basic spinorial representation is characterized by $s = (\frac{1}{2}, \frac{3}{2})$. It contains only one angular-momentum quantum number $l = \frac{1}{2}$. The conjugate representation is $\dot{s} = (\frac{1}{2}, -\frac{3}{2})$. The group generators in $(\frac{1}{2}, \frac{3}{2})$ are expressible in terms of 2×2 Pauli matrices, as illustrated by

$$\mathbf{L} \to \tfrac{1}{2}\boldsymbol{\sigma},$$

$$\mathbf{Q} \to -\frac{i}{2}\boldsymbol{\sigma}.$$

The relation

$$\left(-\frac{i}{2}\boldsymbol{\sigma}\right)^\dagger = -\left(-\frac{i}{2}\boldsymbol{\sigma}\right)$$

explicitly demonstrates that the representation is not unitary.

Among the unitary infinite-dimensional irreducible representations of $SL(2, c)$, there are two, labeled by $(0, \frac{1}{2})$ and $(\frac{1}{2}, 0)$, which will appear as the solutions of (1). These are two Majorana representations.

The first, $(0, \frac{1}{2})$, belongs to the set of representations of the supplementary series and contains integer angular momenta $l = 0, 1, 2, \cdots$, while the second, $(\frac{1}{2}, 0)$, is a representation of the principal series and contains the fractional l's, $l = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$.

Consider now the problem of the Clebsch–Gordan decomposition of the tensor product of two irreducible representations. Symbolically,

$$(k_0, c) \otimes (k_0', c') = \sum \oplus (k_0'', c''). \tag{3}$$

The case when both irreducible representations on the lhs of (3) are finite dimensional was solved a long time ago.[1] When one or both of the two representations (k_0, c) and (k_0', c') are unitary, the problem is more involved, and has been solved for only some special cases.[2]

It is not our intention to investigate systematically other cases or to calculate the associated Clebsch–Gordan coefficients. Instead, we address ourselves to the simpler problem of determining if and which *unitary* irreducible representations appear in the Clebsch–Gordan expansion of the tensor product of a unitary infinite-dimensional irreducible representation of $SL(2, c)$ with a particular nonunitary finite-dimensional irreducible representation of the same group. The identification of other, nonunitary infinite-dimensional representations of $SL(2, c)$ which may appear on the rhs of (3) is of secondary importance here.

It is not without surprise that we find unitary representations at all in the tensor product involving nonunitary factor. However, the example which is given below clearly illustrates that such situations can occur. Even without reference to the physical applications, it is obvious that the unitary representations in the Clebsch–Gordan expansion deserve special attention.

We now give the solution of (3) for the case $(k_0', c') = (\frac{1}{2}, \frac{3}{2})$. In order to emphasize the role of unknown variables which the unitary irreducible representations of $SL(2, c)$ play in our considerations, we denote the representation (k_0, c) on the lhs of (3) by X. Let $\mathbf{L}$ and $\mathbf{Q}$ be the *Hermitian* generators acting in X. Let $\mathbf{J}$ and $\mathbf{K}$ be the generators acting in the product space $X \otimes (\frac{1}{2}, \frac{3}{2})$:

$$\mathbf{J} = \mathbf{L} + 1/2\boldsymbol{\sigma},$$

$$\mathbf{K} = \mathbf{Q} - i/2\boldsymbol{\sigma}.$$

Then

$$\mathbf{J}^2 - \mathbf{K}^2 = \mathbf{L}^2 - \mathbf{Q}^2 + 3/2 + 2A \tag{4}$$

and

$$\mathbf{J} \cdot \mathbf{K} = \mathbf{L} \cdot \mathbf{Q} - 3i/4 - iA,$$

where

$$A = \tfrac{1}{2}(\boldsymbol{\sigma} \cdot \mathbf{L} + i\boldsymbol{\sigma} \cdot \mathbf{Q}).$$

We will denote the eigenvalues of A by a. A simple relationship exists between the quantities $\mathbf{L}^2 - \mathbf{Q}^2$, $\mathbf{L} \cdot \mathbf{Q}$, and A (see Ref. 3):

$$(A + \tfrac{1}{2})^2 = \tfrac{1}{4}(\mathbf{L}^2 - \mathbf{Q}^2 + 2i\mathbf{L} \cdot \mathbf{Q} + 1). \tag{5}$$

The validity of (5) can easily be verified by direct computation, which includes the relations

$$\boldsymbol{\sigma} \cdot \mathbf{L}\, \boldsymbol{\sigma} \cdot \mathbf{L} = \mathbf{L}^2 - \boldsymbol{\sigma} \cdot \mathbf{L},$$

$$\boldsymbol{\sigma} \cdot \mathbf{Q}\, \boldsymbol{\sigma} \cdot \mathbf{Q} = \mathbf{Q}^2 + \boldsymbol{\sigma} \cdot \mathbf{L},$$

and

$$\tfrac{1}{2}(\boldsymbol{\sigma} \cdot \mathbf{L}\, \boldsymbol{\sigma} \cdot \mathbf{Q} + \boldsymbol{\sigma} \cdot \mathbf{Q}\, \boldsymbol{\sigma} \cdot \mathbf{L}) = \mathbf{L} \cdot \mathbf{Q} - \boldsymbol{\sigma} \cdot \mathbf{Q}.$$

The numerical counterpart of (5), involving the eigenvalues of $\mathbf{L}^2 - \mathbf{Q}^2$, $\mathbf{L} \cdot \mathbf{Q}$, and A, reads

$$(a + \tfrac{1}{2})^2 = \tfrac{1}{4}(c + k_0)^2$$

or

$$a_\pm = -\tfrac{1}{2} \pm \tfrac{1}{2}(c + k_0). \tag{6}$$

The essential step in our considerations is now to determine the conditions which are imposed upon the operator A and consequently upon its two eigenvalues a_+ and a_- by the requirement that the Clebsch–Gordan decomposition of the tensor product $X \otimes (\frac{1}{2}, \frac{3}{2})$ contain unitary pieces. A necessary condition is obviously the hermiticity of the invariant operators $\mathbf{J}^2 - \mathbf{K}^2$ and $\mathbf{J} \cdot \mathbf{K}$ on the corresponding irreducible subspaces. Applied to (4), this hermiticity condition requires that

$$A^\dagger = A$$

and

$$A + \tfrac{3}{4} = 0$$

on at least one irreducible subspace. In the language of eigenvalues, either

$$a_+ = -\tfrac{3}{4}$$

or

$$a_- = -\tfrac{3}{4} \tag{7}$$

or

$$a_+ = a_- = -\tfrac{3}{4}.$$

An inspection of (6) informs us that the case $a_+ = a_- = -\frac{3}{4}$ cannot be realized. Combining (6) and (7), we observe that either

$$c + k_0 = -\tfrac{1}{2}$$

or (8)

$$c + k_0 = +\tfrac{1}{2}.$$

Since k_0 is always real or zero, c in (8) must also be real or zero. But the principal series of the unitary

irreducible representations of $SL(2, c)$ are characterized by purely imaginary c or zero, while the supplementary series are characterized by $k_0 = 0$, c real and $|c| < 1$.[1] Using this information, we conclude that, in the set of unitary infinite-dimensional irreducible representations of $SL(2, c)$, there are only two solutions of (8):

$$k_0 = \pm\tfrac{1}{2}, \quad c = 0 \quad \text{(principal series)}$$

and (9)

$$k_0 = 0, \quad c = \pm\tfrac{1}{2} \quad \text{(supplementary series)}.$$

It remains to be verified that the solutions (9) are also sufficient conditions for the existence of unitary pieces on the rhs of (3). Combining the invariant operators $\mathbf{J}^2 - \mathbf{K}^2$ and $\mathbf{J} \cdot \mathbf{K}$ into non-Hermitian objects

$$\mathbf{J}^2 - \mathbf{K}^2 - 2i\mathbf{JK} + 1 = \mathbf{L}^2 - \mathbf{Q}^2 - 2i\mathbf{LQ} + 1,$$

$$\mathbf{J}^2 - \mathbf{K}^2 + 2i\mathbf{JK} + 1 = (\mathbf{L}^2 - \mathbf{Q}^2 + 2i\mathbf{LQ} + 1) + 4A + 3,$$

one obtains, in the process of diagonalization,

$$(\bar{c} - \bar{k}_0)^2 = (c - k_0)^2$$

and (10)

$$(\bar{c} + \bar{k}_0)^2 = (c + k_0)^2 + 4a + 3.$$

The parameters $\bar{k}_0$ and $\bar{c}$ correspond to the eigenvalues of the invariant operators $\mathbf{J}^2 - \mathbf{K}^2$ and $\mathbf{JK}$. After inserting $a = -\frac{3}{4}$ and the values of k_0 and c from (9) into (10), one obtains

$$\bar{c} - \bar{k}_0 = \epsilon \cdot \tfrac{1}{2}$$

and (11)

$$\bar{c} + \bar{k}_0 = \epsilon' \cdot \tfrac{1}{2},$$

where ϵ and ϵ' can independently assume the values either $+1$ or -1.

The same argument which led from (8) to (9) can now be applied to (11). With the observation that either

$$\bar{k}_0 = 0, \quad \bar{c} = \pm\tfrac{1}{2} \quad \text{(supplementary series)}$$

or

$$\bar{k}_0 = \pm\tfrac{1}{2}, \quad \bar{c} = 0 \quad \text{(principal series)}$$

furnish the solution of (11), we complete the proof of the theorem, symbolically expressed by (1).

It is simple to verify that similar conclusions can be reached if the spinorial representation $(\frac{1}{2}, \frac{3}{2})$ in the equation (3) is replaced by its conjugate representation $(\frac{1}{2}, -\frac{3}{2})$, whose generators are

$$\mathbf{L} \to 1/2\boldsymbol{\sigma},$$

$$\mathbf{Q} \to i/2\boldsymbol{\sigma}.$$

Again, we obtain the same X and Y_1 as in the previous case of the $(\frac{1}{2}, \frac{3}{2})$ representation, namely,

$$(0, \tfrac{1}{2}) \otimes (\tfrac{1}{2}, -\tfrac{3}{2}) = (\tfrac{1}{2}, 0) \oplus \cdots$$

and

$$(\tfrac{1}{2}, 0) \otimes (\tfrac{1}{2}, -\tfrac{3}{2}) = (0, \tfrac{1}{2}) \oplus \cdots.$$

ACKNOWLEDGMENT

It is a pleasure to thank Professor Y. Ne'eman for his interest and encouragement.

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR grant number EOOAR-68-0010, through the European Office of Aerospace Research.

[1] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups* (Pergamon Press, Inc., New York, 1963); M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, Inc., New York, 1964).

[2] (a) A. Z. Dolginov and I. N. Toptygin, Zh. Eksp. Teor. Fiz. 794 (1958) [Sov. Phys.—JETP **8**, 550 (1959)]; (b) A. Z. Dolginov and I. N. Toptygin, Zh. Eksp. Teor. Fiz. 1441 (1959) [Sov. Phys.—JETP **10**, 1022 (1960)]; (c) A. Z. Dolginov and A. N. Moskalev, Zh. Eksp. Teor. Fiz. 1967 (1959) [Sov. Phys.—JETP **10**, 1202 (1960)]; (d) M. A. Naimark, Tr. Moskov. Mat. Obshch. **8**, 121 (1959); **9**, 237 (1960); **10**, 181 (1961); [Am. Math. Soc. Transl.,Ser. 2, **36**, 101–229 (1964)].

[3] Harish-Chandra, Proc. Roy. Soc. (London),Sec. A, **189**, 372 (1947).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Three-Dimensional Linear Transport Theory*

K. M. CASE
Department of Physics, The Rockefeller University, New York, New York 10021
AND
R. D. HAZELTINE†
Department of Physics, The University of Michigan, Ann Arbor, Michigan 48103

(Received 7 August 1969)

A recent technique for extending the singular eigenfunction method in linear transport theory to problems which are not strictly 1-dimensional is compared to a more naive approach based on the Fourier transform. The latter appears to have advantages with regard to simplicity and directness.

INTRODUCTION

The 1-dimensional form of the time-independent one-speed transport equation with isotropic scattering is conventionally solved by means of the "singular-eigenfunction" technique.[1] Recently a method for extending this technique to problems which are not strictly 1-dimensional has been proposed by Kaper.[2] This method, although evidently limited in its applicability to problems in which the boundary *surfaces* are no more complicated than parallel planes, does allow boundary *conditions* which vary over these planes. Thus the "reduced" 3-dimensional form of the transport equation is used throughout, and one may handle, for example, point sources, rather than only plane sources.

The extension of the 1-dimensional theory to this wider class of problems is far from trivial. Indeed, a fairly elaborate mathematical framework, involving the theory of generalized analytic functions,[3] is required. The solutions obtained from this framework are initially in the form of complex 2-dimensional integrals, the reduction of which to more readily useful form is, again, nontrivial.

Our purpose here is to compare the above method to a much more naive approach based on the Fourier transform. We hope to show that the latter has at least as wide a range of applicability as the singular eigenfunction method, while at the same time requiring much less sophisticated mathematics. Moreover, the Fourier-transform method seems in several ways more direct; in particular, it provides solutions directly in a form suitable for evaluation.

Our comparison of the two methods will be effected by examining their respective applications to two simple problems. Thus, in Sec. 1, after some brief remarks on the method of Kaper, we begin with the solutions obtained in Ref. 2 for the infinite-space Green's function and half-space albedo problems. We show how the 2-dimensional integrals which these solutions involve are to be reduced to much simpler contour integrals. The Fourier-transform method, which yields solutions in contour integral form directly, is presented in Sec. 2. In the conclusion, we attempt to summarize the essential differences between the two techniques.

1. THE SINGULAR EIGENFUNCTION SOLUTIONS

The reduced transport equation has the form

$$\left(1 - iB_y\Omega_y - iB_z\Omega_z + \mu\frac{\partial}{\partial x}\right)\hat{N} = \frac{c}{4\pi}\int d\Omega'\hat{N}. \tag{1.1}$$

Here, $\hat{N}(x, B_y, B_z, \mathbf{\Omega})$ is related to the neutron angular density $N(\mathbf{r}, \mathbf{\Omega})$ either by means of the ansatz

$$\hat{N} = e^{-iB_y y - iB_z z}N, \tag{1.2}$$

in which case B_y and B_z are the transverse buckling constants, or by means of a 2-dimensional Fourier transform. In either case, notice that Eq. (1.1) limits us to considering only problems in which the boundaries may be taken to be independent of y and z. The vector $\mathbf{\Omega}$ refers, of course, to the normalized velocity

$$\mathbf{\Omega} = \mathbf{v}/|\mathbf{v}| \tag{1.3}$$

$$= (\mu, (1-\mu^2)^{\frac{1}{2}}\cos\theta, (1-\mu^2)^{\frac{1}{2}}\sin\theta) \tag{1.4}$$

$$= (\mu, \mathbf{\Omega}_{(x)}), \tag{1.5}$$

where, in Eq. (1.5), we have introduced the convention of denoting the y and z components of a vector by a subscript (x).

Kaper[2] observes that, if we define a transformation of variables

$$(\mu, \theta) \to (\xi, \eta) \tag{1.6}$$

by means of the complex variable

$$\zeta = \xi + i\eta \tag{1.7}$$

$$\equiv \mu/(1 - i\mathbf{B}_{(x)}\cdot\mathbf{\Omega}_{(x)}) \tag{1.8}$$

and, furthermore, define a new angular density ψ by

$$\psi(x, \xi, \eta) \equiv (1 - i\mathbf{B}_{(x)} \cdot \mathbf{\Omega}_{(x)})\hat{N}(x, \mathbf{B}_{(x)}, \mathbf{\Omega}), \quad (1.9)$$

then Eq. (1.1) may be written in the form

$$\left(1 + \zeta \frac{\partial}{\partial x}\right)\psi(x, \zeta) = \iint_G g(\zeta')\psi(x, \zeta')\, d\xi'\, d\eta'. \quad (1.10)$$

Here,

$$g(\zeta) = \frac{c}{4\pi}\left|\frac{\partial(\mu, \theta)}{\partial(\xi, \eta)}\right| [1 - i\mathbf{B}_{(x)} \cdot \mathbf{\Omega}_{(x)}(\zeta)] \quad (1.11)$$

and $G = G^+ \cup G^-$ is a certain figure-eight-shaped region of the complex plane. (See Fig. 1, which is reproduced here from Ref. 2 for convenience.) Note that the convenient notation

$$f(\xi, \eta) = f(\zeta) \quad (1.12)$$

is by no means intended to imply that f is an analytic function of ζ.

The evident similarity between Eq. (1.10) and the standard form of the 1-dimensional transport equation[1] is exploited in Ref. 2 by applying to Eq. (1.10) a variation of the singular-eigenfunction technique. Thus, one seeks eigenfunctions of the form

$$\psi_\nu(x, \zeta) = e^{-x/\nu}\phi(\nu, \zeta), \quad (1.13)$$

where $\phi(\nu, \zeta)$ is a generalized function of the complex variable ζ, defined for test functions with support in G. Discrete and continuum "modes" are obtained from Eqs. (1.10) and (1.13), and these are then shown to possess the usual ("half-range" and "full-range") completeness and orthogonality properties. Without attempting to reproduce the work of Kaper here, we wish to remark on a few of the differences between the singular-eigenfunction theory of Eq. (1.10) and the conventional 1-dimensional analysis.

Primary among these differences is the mathematical complexity of the 3-dimensional dispersion function

$$\Lambda(\nu) = 1 + \iint_G \frac{\nu g(\zeta)}{\zeta - \nu}\, d\xi\, d\eta. \quad (1.14)$$

$\Lambda(\nu)$ is analytic "almost nowhere" inside the region G. Only for $\nu \notin G$ can it be written in closed form:

$$\Lambda(\nu) = \hat{\Lambda}(\nu), \quad \nu \notin G, \quad (1.15)$$

$$\hat{\Lambda}(\nu) \equiv 1 - \frac{c}{2(\nu^{-2} - B^2)^{\frac{1}{2}}} \ln \frac{1 + (\nu^{-2} - B^2)^{\frac{1}{2}}}{1 - (\nu^{-2} - B^2)^{\frac{1}{2}}}. \quad (1.16)$$

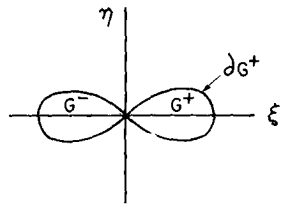


FIG. 1. The region $G = G^+ \cup G^-$ of the complex ζ plane.

$\hat{\Lambda}(\nu)$, like the 1-dimensional dispersion function which it resembles, has the two roots

$$\hat{\Lambda}(\pm\nu_0) = 0, \quad (1.17)$$

where, denoting the known[1] 1-dimensional roots by $\pm L$, we have

$$\nu_0^{-2} = B^2 + L^{-2}. \quad (1.18)$$

However, because of the qualification $\nu \notin G$ on Eq. (1.15), one cannot conclude in general that the $\pm\nu_0$ will be roots of Λ. This technical difficulty (which, as we shall see below, has only a temporary significance) is dealt with in Ref. 2 by defining the function

$$\chi_c(B^2) = 1, \quad 0 < B^2 < 1 - L^{-2},$$
$$= 0, \quad B^2 \geq 1 - L^{-2}. \quad (1.19)$$

Now when $\chi_c(B^2) = 1$, it is clear from Eq. (1.18) that $\pm\nu_0 \notin G$, so that Eq. (1.15) holds and

$$\Lambda(\pm\nu_0) = O, \quad \chi_c(B^2) = 1. \quad (1.20)$$

Equation (1.20) yields, in the usual way, the discrete modes

$$\psi_\pm(x, \zeta) = e^{\mp x/\nu_0}\phi_\pm(\zeta), \quad (1.21)$$

$$\phi_\pm(\zeta) = \pm\nu_0 g(\pm\nu_0)/(\pm\nu_0 - \zeta). \quad (1.22)$$

But it must be observed that these modes do not occur in the case $\chi_c(B^2) = 0$.

Equation (1.10) always possesses a continuum of eigenvalues. Here, the continuum is 2-dimensional: all $\nu \in G$ and the corresponding eigenfunctions

$$\psi_\nu(x, \zeta) = e^{-x/\zeta}\phi(\nu, \zeta), \quad (1.23)$$

$$\phi(\nu, \zeta) = \nu g(\nu)/(\nu - \zeta) + \Lambda(\nu)\delta(\nu - \zeta) \quad (1.24)$$

are generalized analytic functions, with the definitions

$$((\nu - \zeta)^{-1}, \psi(\zeta)) \equiv \iint_G \frac{\psi(\zeta)}{\nu - \zeta}\, d\xi\, d\eta, \quad (1.25)$$

$$(\delta(\nu - \zeta), \psi(\zeta)) \equiv \psi(\nu), \quad \nu \in G,$$
$$\equiv 0, \quad \nu \notin G. \quad (1.26)$$

The formal similarity of Eq. (1.24) to the familiar 1-dimensional continuum modes is perhaps deceptive. Note, for example, that the integral of Eq. (1.25) exists without any principal-value interpretation.

The proof of full-range orthogonality for the eigenfunctions of Eqs. (1.21)–(1.24) is as trivial here as it is in the 1-dimensional theory. However, rather complicated arguments are required to prove their full-range completeness, half-range orthogonality, and half-range completeness. Essential to these arguments is a "theorem" which, ignoring details of rigor, may be stated in the form[4]

$$\frac{\partial}{\partial\xi}\frac{1}{\zeta} = \pi\delta(\zeta). \tag{1.27}$$

In particular, it follows from Eq. (1.27) that

$$\begin{aligned}\frac{\partial\Lambda}{\partial\bar{\zeta}} &= 0, \qquad \zeta \notin G,\\ &= -\pi\zeta g(\zeta), \quad \zeta \in G.\end{aligned} \tag{1.28}$$

This theorem plays a role here somewhat analogous to that of the Plemelj formulas in the conventional theory. It is used in the half-range orthogonality proof, for example, to find a function $X(\zeta)$ which is analytic for $\zeta \notin G^+$, continuous on the boundary ∂G^+ of G^+, and which satisfies

$$\Lambda(\zeta)\frac{\partial X}{\partial\bar{\zeta}} - X(\zeta)\frac{\partial\Lambda}{\partial\bar{\zeta}} = 0, \quad \zeta \in G^+. \tag{1.29}$$

Of course, the degree of complexity of the completeness and orthogonality proofs has little bearing on the task of solving specific problems. Indeed, Kaper shows that this task proceeds in a quite straightforward manner. Given a problem of the general form of Eq. (1.1), with the addition perhaps of inhomogeneous terms, and conditions specified on boundary planes perpendicular to the x axis, we first transform the angular density and angle variables according to Eqs. (1.7)–(1.9). Then, as in the 1-dimensional theory, we expand the unknown transformed density ψ in terms of the eigenfunctions of Eqs. (1.21)–(1.24), and expect the given boundary conditions, together with the appropriate (half-range or full-range) orthogonality relations, to provide the expansion coefficients. For purposes of comparison with a quite different approach to be described below, we wish now to examine the solutions Kaper obtains in this way to two very simple problems.

A. The Infinite-Space Green's Function

The infinite-space Green's function satisfies, for all $\mathbf{r}$,

$$\begin{aligned}(\boldsymbol{\Omega}\cdot\nabla + 1)N_g(\mathbf{r},\boldsymbol{\Omega}) &\\ = \frac{c}{4\pi}\int d\boldsymbol{\Omega}' N_g(\mathbf{r},\boldsymbol{\Omega}') &+ \delta(\mathbf{r})\delta(\boldsymbol{\Omega}-\boldsymbol{\Omega}_0)\end{aligned} \tag{1.30}$$

and is found by Kaper to have the representation

$$N_g(\mathbf{r},\boldsymbol{\Omega}) = \frac{1}{(2\pi)^2}\int d\mathbf{B}_{(x)} e^{-i\mathbf{B}_{(x)}\cdot\mathbf{r}_{(x)}}\frac{\psi_g(x,\zeta)}{1 - i\mathbf{B}_{(x)}\cdot\boldsymbol{\Omega}_{(x)}}, \tag{1.31}$$

$$\begin{aligned}\psi_g(x,\zeta) &= \pm\chi_c(B^2)a_\pm\psi_\pm(x,\zeta)\\ &\pm\iint_{G^+} A(\zeta')\psi_{\zeta'}(x,\zeta)\,d\xi'd\eta', \quad x \gtrless 0,\end{aligned} \tag{1.32}$$

where[5]

$$a_\pm = g(\zeta_0)\left|\frac{\partial(\xi_0,\eta_0)}{\partial(\mu_0,\theta_0)}\right| N_\pm^{-1}\phi_\pm(\zeta_0), \tag{1.33}$$

$$A(\zeta) = g(\zeta_0)\left|\frac{\partial(\xi_0,\eta_0)}{\partial(\mu_0,\theta_0)}\right| [N(\zeta)]^{-1}\phi(\zeta,\zeta_0), \tag{1.34}$$

and the N's are certain quantities defined in Ref. 2.

What is especially interesting for our purposes here is the transformation of the solution afforded by Eqs. (1.32)–(1.34) into a form directly suitable for evaluation. In particular, we wish to express the integral over the region G^+ or G^- [in Eq. (1.32)] in the much more tractable form of a line integral. Assuming for definiteness that $x > 0$, we have

$$\psi_g(x,\zeta) = \chi_c(B^2)\frac{c}{4\pi\mu_0}\frac{\zeta_0}{}\frac{1}{N_+}\phi_+(\zeta_0)\psi_+(x,\zeta) + \frac{c}{4\pi\mu_0}\frac{\zeta_0}{}I_1, \tag{1.35}$$

where

$$I_1 = \iint_{G^+}\frac{\phi(\zeta',\zeta_0)}{N(\zeta')}\psi_{\zeta'}(x,\zeta)\,d\xi'\,d\eta' \tag{1.36}$$

and we have noted that

$$g(\zeta_0)\left|\frac{\partial(\xi_0,\eta_0)}{\partial(\mu_0,\theta_0)}\right| = \frac{c}{4\pi\mu_0}\frac{\zeta_0}{}. \tag{1.37}$$

Making the appropriate substitutions from Eqs. (1.21)–(1.24) and (1.34), and integrating the δ-function terms, we can write Eq. (1.36) as

$$\begin{aligned}I_1 = \chi_+(\zeta)\frac{e^{-x/\zeta}}{(\zeta-\zeta_0)\Lambda(\zeta)} &+ \chi_+(\zeta_0)\frac{e^{-x/\zeta_0}}{(\zeta_0-\zeta)\Lambda(\zeta_0)}\\ &+ \chi_+(\zeta_0)\delta(\zeta-\zeta_0)\frac{e^{-x/\zeta_0}}{\zeta_0 g(\zeta_0)} + I_2,\end{aligned} \tag{1.38}$$

where

$$\begin{aligned}\chi_+(\zeta) &= 1, \quad \zeta \in G^+,\\ &= 0, \quad \zeta \notin G^+,\end{aligned} \tag{1.39}$$

and

$$I_2 = \iint_{G^+} e^{-x/\zeta'}\frac{\zeta' g(\zeta')}{\Lambda^2(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)}\,d\xi'\,d\eta'. \tag{1.40}$$

I_2 can be made to appear much less formidable if we

avail ourselves of the "theorem" (1.27) and its corollary (1.28). These are easily seen to imply that the integrand in Eq. (1.40) may be written in the form

$$\frac{1}{\pi}\frac{\partial}{\partial\zeta'}\left(\frac{e^{-x/\zeta'}}{\Lambda(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)}\right) - \frac{e^{-x/\zeta'}}{(\zeta-\zeta_0)\Lambda(\zeta')}[\delta(\zeta'-\zeta_0)-\delta(\zeta-\zeta_0)]. \quad (1.41)$$

We observe that the δ-function terms in the expression (1.41) are precisely such as to cancel (upon integration over G^+) the first two terms in Eq. (1.38). Furthermore, the integral over G^+ of the first term in (1.41) is readily converted, by means of what is essentially Stokes' theorem, into an integral over the boundary ∂G^+. In this way we find that Eqs. (1.38) and (1.40) reduce to

$$I_1 = \chi_+(\zeta_0)\delta(\zeta-\zeta_0)\frac{e^{-x/\zeta_0}}{\zeta_0 g(\zeta_0)} + \frac{1}{2\pi i}\int_{\partial G^+}\frac{e^{-x/\zeta'}\,d\zeta'}{\Lambda(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)}. \quad (1.42)$$

With regard to the line integral in Eq. (1.42), one further significant manipulation is allowed: Since $\zeta' \in \partial G^+$, we may substitute for $\Lambda(\zeta')$ the much more analytically tractable function $\hat{\Lambda}(\zeta')$ [cf. Eq. (1.16)]. It is clear that the latter function has its only singularities, branch points, at

$$\zeta' = \pm\alpha, \quad (1.43)$$

$$\alpha = (1+B^2)^{-\frac{1}{2}}. \quad (1.44)$$

We choose the cut $l = l^+ + l^-$ as joining these two points along the real axis as in Fig. 2. Now we define a region of the complex plane $\hat{G}_+$ to be the region G^+ with the exclusion of a small neighborhood of the cut l, and observe that the integrand in Eq. (1.42) is analytic throughout the interior of $\hat{G}_+$ except for the following:

(i) There will be a pole at $\zeta' = \zeta(\zeta' = \zeta_0)$ unless $\zeta(\zeta_0) \in G^-$ or $\zeta(\zeta_0) \in l$;

(ii) there will be a pole at $\zeta' = \nu_0$ unless $\chi_c(B^2) = 1$. [It is important to note, from Eq. (1.33) and the known fact[1] that $|L| > 1$, that in any case $\nu_0 \notin l$.]

With these remarks, it is evident that we may write

$$\frac{1}{2\pi i}\int_{\partial G^+}\frac{e^{-x/\zeta'}\,d\zeta'}{\hat{\Lambda}(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)} = \frac{e^{-x/\zeta}}{\hat{\Lambda}(\zeta)(\zeta-\zeta_0)}\hat{\chi}_+(\zeta) + \frac{e^{-x/\zeta_0}}{\hat{\Lambda}(\zeta_0)(\zeta_0-\zeta)}\hat{\chi}_+(\zeta_0) + \frac{e^{-x/\nu_0}[1-\chi_c(B^2)]}{\hat{\Lambda}'(\nu_0)(\nu_0-\zeta)(\nu_0-\zeta_0)} + \frac{1}{2\pi i}\int_{l^{+\prime}}\frac{e^{-x/\zeta'}\,d\zeta'}{\hat{\Lambda}(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)}. \quad (1.45)$$

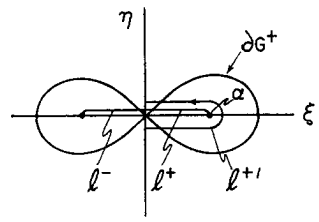


FIG. 2. Cuts and integration contours.

Here, of course,

$$\hat{\chi}_+(\zeta) \equiv 1, \quad \zeta \in \hat{G}_+, \qquad \equiv 0, \quad \zeta \in G^- \cup l^+, \quad (1.46)$$

and $l^{+\prime}$ is a path surrounding the cut l^+ (cf. Fig. 2). Note that we have recovered in Eq. (1.45) both the previously cancelled terms of Eq. (1.38), in somewhat altered form, and the discrete term in the case $\chi_c(B^2) = 0$.

The right-hand side of Eq. (1.45) represents an improvement over the left-hand side for two reasons:

(i) The integration path for the line integral is much simpler. Indeed, when neither ζ nor ζ_0 is on l^+, we may write

$$\frac{1}{2\pi i}\int_{l^+}\frac{e^{-x/\zeta'}\,d\zeta'}{\hat{\Lambda}(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)} = \frac{1}{2\pi i}\int_0^\alpha d\zeta'\,\frac{e^{-x/\zeta'}}{(\zeta'-\zeta)(\zeta'-\zeta_0)}\left(\frac{1}{\hat{\Lambda}^-(\zeta')}-\frac{1}{\hat{\Lambda}^+(\zeta')}\right), \quad (1.47)$$

where the $\hat{\Lambda}^\pm$ are the boundary values of $\hat{\Lambda}$ on its cut. In case either ζ or ζ_0 is on l^+, we avoid the pole by a small semicircle and may write the integral in terms of a principal value plus pole contributions in the conventional way.

(ii) More significantly, we have isolated in Eq. (1.45) the asymptotically dominant contribution to the integral over ∂G^+. Indeed, it is evident from Eqs. (1.18) and (1.44), that the discrete term [which must, regardless of the magnitude of B^2, occur in $\psi(x, \zeta)$] always dominates the integral of Eq. (1.45) for large x.

Substituting Eqs. (1.42) and (1.45) into Eq. (1.35) and reverting to the more physical quantity N_g, we conclude

$$\hat{N}_g(x,\boldsymbol{\Omega}) = \chi_+(\zeta_0)\left|\frac{\partial(\xi_0,\eta_0)}{\partial(\mu_0,\theta_0)}\right|\mu_0^{-1}e^{-x/\zeta_0}\delta(\zeta-\zeta_0) + \frac{c}{4\pi}\frac{\zeta\zeta_0}{\mu\mu_0}\left(\frac{e^{-x/\nu_0}}{\hat{\Lambda}'(\nu_0)(\nu_0-\zeta)(\nu_0-\zeta_0)} + \hat{\chi}_+(\zeta)\frac{e^{-x/\zeta}}{\hat{\Lambda}(\zeta)(\zeta-\zeta_0)} + \hat{\chi}_+(\zeta_0)\frac{e^{-x/\zeta_0}}{\hat{\Lambda}(\zeta_0)(\zeta_0-\zeta)} + \frac{1}{2\pi i}\int_{l^{+\prime}}\frac{e^{-x/\zeta'}\,d\zeta'}{\hat{\Lambda}(\zeta')(\zeta'-\zeta)(\zeta'-\zeta_0)}\right), \quad x>0, \quad (1.48)$$

where the correct functional dependence is ultimately to be obtained by means of Eq. (1.8). We omit this final step and merely draw attention to the fact that Eq. (1.48) involves only $\hat{\Lambda}$,[6] which, unlike Λ, can be written in closed form; there are no 2-dimensional integrals, explicit or implicit, in our final expression for $\hat{N}$.

B. The Albedo Problem

If $\hat{N}_a(x, \mathbf{B}_{(x)}, \mathbf{\Omega})$ satisfies the homogeneous equation (1.1) for $x > 0$ and the boundary condition

$$\hat{N}_a(0, \mathbf{B}_{(x)}, \mathbf{\Omega}) = \delta(\mu - \mu_0)\delta(\theta - \theta_0), \quad \text{for} \quad \mu > 0, \quad \mu_0 < 0, \quad (1.49)$$

then the corresponding ψ_a satisfies Eq. (1.10) for $x > 0$ with[7]

$$\psi_a(0, \zeta) = \frac{\mu_0}{\zeta_0} \frac{\partial(\xi_0, \eta_0)}{\partial(\mu_0, \theta_0)} \delta(\zeta - \zeta_0), \quad \zeta, \zeta_0 \in G^+. \quad (1.50)$$

Using the half-range completeness and orthogonality relations, Kaper obtains from Eqs. (1.10) and (1.50) the solution

$$\psi_a(x, \zeta) = b_+\psi_+(x, \zeta) + \iint_{G^+} B(\zeta')\psi_{\zeta'}(x, \zeta)\, d\xi'\, d\eta', \quad (1.51)$$

where

$$B(\zeta) = \frac{\mu_0}{\zeta_0} \left| \frac{\partial(\xi_0, \eta_0)}{\partial(\mu_0, \theta_0)} \right| \frac{1}{\Lambda(\zeta_0)} \left(\delta(\zeta - \zeta_0) - \frac{\zeta_0 g(\zeta_0)(\zeta_0 - \nu_0)(\zeta - 1)e^{\omega(\zeta_0) - \omega(\zeta)}}{\Lambda(\zeta)(\zeta - \nu_0)(\zeta_0 - 1)(\zeta_0 - \zeta)} \right), \quad (1.52)$$

$$b_+ = \frac{\mu_0}{\Lambda(\zeta_0)} \left| \frac{\partial(\xi_0, \eta_0)}{\partial(\mu_0, \theta_0)} \right| \frac{g(\zeta_0)e^{\omega(\zeta_0)}(\nu_0 - 1)}{g(\nu_0)e^{\omega(\nu_0)}(\zeta_0 - 1)}. \quad (1.53)$$

Here we have used the function

$$\omega(\zeta) = -\frac{1}{\pi} \iint_{G^+} \frac{\partial \Lambda(\zeta')}{\partial \zeta'} \frac{1}{\Lambda(\zeta')} \frac{1}{\zeta' - \zeta}\, d\xi'\, d\eta', \quad (1.54)$$

which arises in the solution of Eq. (1.29). By methods similar to those used in regard to Eqs. (1.38)–(1.45), we may write ω in the form

$$\omega(\zeta) = \hat{\chi}_+(\zeta) \ln \Lambda(\zeta) - \hat{\chi}_+(\zeta) \ln \hat{\Lambda}(\zeta) + \gamma(\zeta) - \ln (\zeta - \alpha)/(\zeta - 1), \quad (1.55)$$

where

$$\gamma(\zeta) \equiv -\frac{1}{2\pi i} \int_{l^+} \frac{\ln \hat{\Lambda}(\zeta')}{\zeta' - \zeta}\, d\zeta'. \quad (1.56)$$

The last term in Eq. (1.56) arises from an integration along the cut of $\ln \hat{\Lambda}$, which cut is taken to be along the real axis inside G and is to be distinguished from the cut of $\hat{\Lambda}$ itself. [Throughout this discussion of the albedo problem we assume for convenience that $\chi_c(B^2) = 1$; the actual value of $\chi_c(B^2)$ is, of course, as irrelevant to the final answer here as it was in the infinite-space case discussed above.]

Our task now is to substitute Eqs. (1.52) and (1.53) into Eq. (1.51), and to reduce the integral over G^+ to a simple branch-cut integral. These manipulations are somewhat lengthier than, but otherwise very similar to, the procedure we performed above for the infinite-space Green's function. Omitting both the detailed calculation and the general result, we state here only the most physically interesting result, namely, that the emergent angular density, which according to Eqs. (1.51)–(1.53) is given by

$$\hat{N}_a(0, \mathbf{B}_{(x)}, \mathbf{\Omega}) = \frac{\psi_a(\zeta)}{1 - i\mathbf{B}_{(x)} \cdot \mathbf{\Omega}_{(x)}}, \quad \mu < 0, \quad (1.57)$$

$$\psi_a(0, \zeta) = b_+\phi_+(0, \zeta) + \frac{\mu_0}{\zeta_0} \left| \frac{\partial(\xi_0, \eta_0)}{\partial(\mu_0, \theta_0)} \right| \frac{1}{\Lambda(\zeta_0)} \times \left(\psi_{\zeta_0}(0, \zeta) - \frac{\zeta_0 g(\zeta_0)e^{\omega(\zeta_0)}(\zeta_0 - \nu_0)}{(\zeta_0 - 1)} \times \iint_{G^+} \frac{(\zeta' - 1)e^{-\omega(\zeta')}\psi_{\zeta'}(0, \zeta)}{(\zeta' - \nu_0)(\zeta_0 - \zeta')\Lambda(\zeta')}\, d\xi'\, d\eta' \right), \quad \zeta \in G^-, \quad (1.58)$$

ultimately reduces to

$$\hat{N}_a(0, \mathbf{B}_{(x)}, \mathbf{\Omega}) = -\frac{c}{4\pi} \frac{\zeta\zeta_0(\zeta_0 - \nu_0)(\zeta - \alpha)e^{\gamma(\zeta_0) - \gamma(\zeta)}}{[\bar{\Lambda}(\zeta_0)](\zeta - \nu_0)(\zeta_0 - \alpha)(\zeta - \zeta_0)}, \quad \mu < 0. \quad (1.59)$$

We have placed the factor $\hat{\Lambda}(\zeta_0)$ in brackets to indicate that it is to be replaced by 1 in the case $\zeta_0 \in l^+$. The fact that Eq. (1.59), like Eq. (1.48), involves $\hat{\Lambda}$ rather than Λ follows essentially from Eq. (1.55).

It is perhaps worth mentioning that the crucial step in obtaining Eq. (1.59) from Eq. (1.58) depends upon the observation that Eq. (1.54) implies

$$\frac{\partial \omega(\zeta)}{\partial \bar{\zeta}} = 0, \quad \zeta \in G^+, \qquad = \frac{-\pi\zeta g(\zeta)}{\Lambda(\zeta)}, \quad \zeta \in G^+. \quad (1.60)$$

Using Eq. (1.60), the surface integral in Eq. (1.58) may be written in terms of some δ-function contributions plus an integral over ∂G^+; since $x = 0$, the contour for the latter may be deformed into a contour at infinity, with some residues. The final result is Eq. (1.59).

Let us consider an alternative formulation.

2. THE FOURIER-TRANSFORM METHOD

A rather general problem in linear transport theory may be stated as follows:

Let V be some (bounded or unbounded) region of 3-dimensional space with boundary S. We are to find that function $\phi(\mathbf{r}, \mathbf{\Omega})$, for $\mathbf{r} \in V$ and $|\mathbf{\Omega}| = 1$, which satisfies

$$(\mathbf{\Omega} \cdot \nabla + 1)\phi(\mathbf{r}, \mathbf{\Omega}) = \frac{c}{4\pi}\rho(\mathbf{r}) + q(\mathbf{r}, \mathbf{\Omega}), \quad \mathbf{r} \in V, \tag{2.1}$$

given the boundary data $\phi_s(\mathbf{r}_s, \mathbf{\Omega})$ and $\mathbf{\Omega}$ inward to V; that is,

$$\phi(\mathbf{r}_s, \mathbf{\Omega}) = \phi_s(\mathbf{r}_s, \mathbf{\Omega}), \quad \mathbf{\Omega} \text{ inward}, \quad \mathbf{r}_s \in S. \tag{2.2}$$

In Eq. (2.1), $\phi(\mathbf{r}, \mathbf{\Omega})$ and $\rho(\mathbf{r})$, where

$$\rho(\mathbf{r}) \equiv \int d\mathbf{\Omega}\phi(\mathbf{r}, \mathbf{\Omega}), \tag{2.3}$$

are the angular density and density, respectively. q represents any external sources which may be present, and, like ϕ_s, is presumed to be given.

In order to express Eq. (2.1) as an integral equation, let $G(\mathbf{r} - \mathbf{r}', \mathbf{\Omega})$, defined for all $\mathbf{r}$ and $\mathbf{r}'$, satisfy

$$(-\mathbf{\Omega} \cdot \nabla + 1)G(\mathbf{r} - \mathbf{r}', \mathbf{\Omega}) = \delta(\mathbf{r} - \mathbf{r}'). \tag{2.4}$$

Then, by a conventional argument, Eqs. (2.1) and (2.4) imply

$$\phi(\mathbf{r}, \mathbf{\Omega}) = \int_V d\mathbf{r}' G(\mathbf{r}' - \mathbf{r}, \mathbf{\Omega})\left(\frac{c}{4\pi}\rho(\mathbf{r}') + q(\mathbf{r}', \mathbf{\Omega})\right) + \mathbf{\Omega} \cdot \int_S \hat{\mathbf{n}}_i \, d\mathbf{r}_s G(\mathbf{r}_s - \mathbf{r}, \mathbf{\Omega})\phi_s(\mathbf{r}_s, \mathbf{\Omega}), \tag{2.5}$$

where $\hat{n}_i$ is the inward normal to V. Note that we have put the known function $\phi_s(\mathbf{r}_s, \mathbf{\Omega})$ in the integrand, instead of $\phi(\mathbf{r}_s, \mathbf{\Omega})$; this is justified by the easily verified [cf. Eq. (2.10) below] fact that

$$G(\mathbf{r}_s - \mathbf{r}, \mathbf{\Omega}) = 0, \quad \text{for} \quad \mathbf{r} \in V, \quad \mathbf{\Omega} \text{ outward}. \tag{2.6}$$

Equation (2.5) holds, of course, only for $\mathbf{r} \in V$, the domain of definition of $\phi(\mathbf{r}, \mathbf{\Omega})$. We now extend this domain by assuming Eq. (2.5) to hold for all $\mathbf{r}$, so that we may take its 3-dimensional Fourier transform. With the convention

$$\tilde{f}(k) = \int f(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \, d\mathbf{r} \tag{2.7}$$

and the definition

$$\tilde{\rho}_V(\mathbf{k}) \equiv \int_V d\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}}\rho(\mathbf{r}), \tag{2.8}$$

the transform of Eq. (2.5) takes the form

$$\tilde{\phi}(\mathbf{k}, \mathbf{\Omega}) = \frac{c}{4\pi}\frac{\tilde{\rho}_V(\mathbf{k})}{1 - i\mathbf{k} \cdot \mathbf{\Omega}} + \frac{\tilde{q}(\mathbf{k}, \mathbf{\Omega})}{1 - i\mathbf{k} \cdot \mathbf{\Omega}} + \frac{\mathbf{\Omega}}{1 - i\mathbf{k} \cdot \mathbf{\Omega}} \cdot \int_S \hat{\mathbf{n}}_i \, d\mathbf{r}_s e^{i\mathbf{k}\cdot\mathbf{r}_s}\phi_s(\mathbf{r}_s, \mathbf{\Omega}). \tag{2.9}$$

In obtaining Eq. (2.9), we used the representation

$$G(\mathbf{r} - \mathbf{r}') = \frac{1}{(2\pi)^3}\int d\mathbf{k} \frac{e^{-i\mathbf{k}\cdot(\mathbf{k}-\mathbf{r}')}}{1 + i\mathbf{k} \cdot \mathbf{\Omega}} \tag{2.10}$$

which follows immediately from Eq. (2.4), and assumed that

$$q(\mathbf{r}, \mathbf{\Omega}) = 0, \quad \mathbf{r} \notin V. \tag{2.11}$$

Upon integrating Eq. (2.9) over all directions $\mathbf{\Omega}$, we obtain the equation

$$\tilde{\rho}(\mathbf{k}) = [1 - \tilde{\Lambda}(\mathbf{k})]\tilde{\rho}_V(\mathbf{k}) + \tilde{B}(\mathbf{k}) + \tilde{\mathbf{Q}}(\mathbf{k}), \tag{2.12}$$

where

$$\tilde{\Lambda}(\mathbf{k}) \equiv 1 - \frac{c}{4\pi}\int \frac{d\mathbf{\Omega}}{1 - i\mathbf{k} \cdot \mathbf{\Omega}} \tag{2.13}$$

is the 3-dimensional dispersion function and

$$\tilde{B}(\mathbf{k}) \equiv \int_S d\mathbf{r}_s e^{i\mathbf{k}\cdot\mathbf{r}_s}\int d\mathbf{r} \frac{\hat{\mathbf{n}}_i \cdot \mathbf{\Omega}}{1 - i\mathbf{k} \cdot \mathbf{\Omega}}\phi_s(\mathbf{r}_s, \mathbf{\Omega}), \tag{2.14}$$

$$\tilde{Q}(\mathbf{k}) \equiv \int d\mathbf{\Omega} \frac{\tilde{q}(\mathbf{k}, \mathbf{\Omega})}{1 - i\mathbf{k} \cdot \mathbf{\Omega}} \tag{2.15}$$

result of course from the *given* boundary and external source contributions, respectively.

A "general" prescription for solving any transport problem of the form (2.1) can now be given:

(i) In some way (perhaps only approximately), we are to solve Eq. (2.12) for $\tilde{\rho}_V(\mathbf{k})$.

(ii) Equation (2.9) then immediately provides $\tilde{\phi}(\mathbf{k}, \mathbf{\Omega})$.

(iii) Finally, we take the inverse Fourier transform of $\tilde{\phi}$.

Of course only step (ii) is trivial. In fact, the practicality of this prescription is in general very questionable. All we hope to show is that for at least that class of problems considered in the previous section, namely, problems in which the boundaries depend upon only one space variable, the method outlined above is indeed workable.

Our hopes for solving Eq. (2.12) rest essentially in the observation that

$$\tilde{\rho}_V(\mathbf{k}) = \int d\mathbf{k}' \tilde{\rho}(\mathbf{k}')\Delta_V(\mathbf{k} - \mathbf{k}'), \tag{2.16}$$

where

$$\Delta_V(\mathbf{k} - \mathbf{k}') \equiv \frac{1}{(2\pi)^3}\int_V d\mathbf{r}' e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}'}, \tag{2.17}$$

so that Eq. (2.12) can be written, in general, as an integral equation for $\tilde{\rho}(\mathbf{k})$.[8] The geometry of a particular problem enters solely through the kernel $\Delta_V(\mathbf{k})$; this will be, for the problems considered here, highly singular (i.e., a generalized function), with the result that our solutions of Eq. (2.12) will depend more on analyticity arguments than on the Fredholm theory. Therefore, before proceeding further, we briefly examine the properties of $\tilde{\Lambda}(\mathbf{k})$ as an analytic function of, say, k_x.

A. The Dispersion Function

For $\mathbf{k}$ real, the integral of Eq. (2.13), like that of Eq. (1.14), may be written in closed form. We thus find the analog of Eq. (1.15):

$$\tilde{\Lambda}(\mathbf{k}) = \Lambda_3(k_x, B), \quad \text{Im}\,(k_x) = 0, \tag{2.18}$$

where

$$B = |\mathbf{k}_{(x)}| \tag{2.19}$$

and

$$\Lambda_3(k, B) \equiv 1 + \frac{ic}{2(k^2 + B^2)^{\frac{1}{2}}} \ln \frac{1 + i(k^2 + B^2)^{\frac{1}{2}}}{1 - i(k^2 + B^2)^{\frac{1}{2}}}. \tag{2.20}$$

The correspondence between these functions and the dispersion function used in Sec. 1 is clear:

$$\tilde{\Lambda}(k_x = -i/\zeta) = \Lambda(\zeta), \tag{2.21}$$

$$\Lambda_3(k = -i/\zeta) = \hat{\Lambda}(\zeta). \tag{2.22}$$

It is again to be remarked that, just as Eq. (1.15) is false for $\zeta \in G$, Eq. (2.18) is invalid outside a certain neighborhood of the real k_x axis. However, since in the present formulation our basic equations are true for real k_x, we may ignore the pathological behavior of $\tilde{\Lambda}(\mathbf{k})$ for general complex k_x and work exclusively with the much more analytically tractable function $\Lambda_3(k_x, B)$, which function is to be taken as defined by Eq. (2.20) for complex k_x also.

We use the notation

$$\Lambda_3(k_x, B) = \Lambda_3(k), \tag{2.23}$$

where, of course, $k = k_x$.

$\Lambda_3(k)$ has its only singularities, branch points, at

$$k = \pm i\beta, \tag{2.24}$$

where

$$\beta \equiv (B^2 + 1)^{\frac{1}{2}} \tag{2.25}$$

and we take the branch cuts $l_\pm$ as extending to $\pm i\infty$ along the imaginary axis (cf. Fig. 3). Because of the resemblance of Eq. (2.20) to the form of the 1-dimensional dispersion function, we may infer that $\Lambda_3(k)$ has only two simple zeros, namely,

$$\Lambda_3(\pm i\kappa_0) = 0, \tag{2.26}$$

where

$$\kappa_0 = (B^2 + L^{-2})^{\frac{1}{2}}. \tag{2.27}$$

Here the $\pm L$ are as in Sec. 1. We may conclude further from the 1-dimensional theory that

(i) $c < 1 \Rightarrow |\kappa_0| > B$ and $\text{Im}\,(\kappa_0) = 0$,
(ii) $c > 1$ and $B > |L^{-1}| \Rightarrow |\kappa_0| < B$ and $\text{Im}\,(\kappa_0) = 0$,
(iii) $c > 1$ and $B < |L^{-1}| \Rightarrow |\kappa_0| < B$ and $\text{Re}\,(\kappa_0) = 0$,

and, perhaps most importantly, that

$$\text{Im}\,(\kappa_0) = 0 \Rightarrow |\kappa_0| < \beta, \tag{2.28}$$

i.e., the roots of $\Lambda_3(k)$ never lie on its cuts. [All these remarks are of course merely the translation, according to Eq. (2.22), of similar facts concerning $\hat{\Lambda}(\zeta)$[2].]

We are now prepared to apply the method outlined to specific transport problems.

B. The Infinite Space Green's Function

It is clear (and in fact well known) that this problem is almost trivially solved by the Fourier transform. Since the region V is all space and we have a point source with direction $\mathbf{\Omega}_0$ at $\mathbf{r}_0$, we find

$$\Delta_V(\mathbf{k}) = \delta(\mathbf{k}), \tag{2.29}$$

$$B(\mathbf{k}) = 0, \tag{2.30}$$

$$\tilde{Q}(\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{r}_0}/(1 - i\mathbf{k}\cdot\mathbf{\Omega}_0). \tag{2.31}$$

Our basic equation (2.12) thus takes the form

$$\tilde{\rho}_G(\mathbf{k}) = [1 - \tilde{\Lambda}(\mathbf{k})]\tilde{\rho}_G(\mathbf{k}) + e^{i\mathbf{k}\cdot\mathbf{r}_0}/(1 - i\mathbf{k}\cdot\mathbf{\Omega}_0) \tag{2.32}$$

or

$$\tilde{\rho}_G(\mathbf{k}) = \frac{1}{\tilde{\Lambda}(\mathbf{k})} \frac{e^{i\mathbf{k}\cdot\mathbf{r}_0}}{1 - i\mathbf{k}\cdot\mathbf{\Omega}_0}, \tag{2.33}$$

i.e., in this translation-invariant case, Eq. (2.12) may be solved algebraically. Equation (2.19) yields

$$\tilde{\phi}_G(\mathbf{k}, \mathbf{\Omega}) = \frac{c}{4\pi} \frac{e^{i\mathbf{k}\cdot\mathbf{r}_0}}{(1 - i\mathbf{k}\cdot\mathbf{\Omega})(1 - i\mathbf{k}\cdot\mathbf{\Omega}_0)\tilde{\Lambda}(\mathbf{k})} + \frac{e^{i\mathbf{k}\cdot\mathbf{r}_0}}{1 - i\mathbf{k}\cdot\mathbf{\Omega}_0} \delta(\mathbf{\Omega} - \mathbf{\Omega}_0). \tag{2.34}$$

The 2-dimensional Fourier transform of the angular density

$$\hat{\phi}_G(x, \mathbf{k}_{(x)}, \mathbf{\Omega}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk_x e^{-ik_x x} \tilde{\phi}_G(\mathbf{k}, \mathbf{\Omega}) \tag{2.35}$$

corresponds to the function $\hat{N}_g$ discussed in Sec. 1. With the convenient abbreviations

$$\omega(\mathbf{\Omega}) \equiv -(i/\mu)(1 - i\mathbf{k}_{(x)}\cdot\mathbf{\Omega}_{(x)}), \tag{2.36}$$

$$\omega_0 = \omega(\mathbf{\Omega}_0), \tag{2.37}$$

Eq. (2.35) takes the form

$$\hat{\phi}_G(x, \mathbf{k}_{(x)}, \boldsymbol{\Omega}) = -\frac{e^{-i\mathbf{k}_{(x)}\cdot\mathbf{r}_{(x)0}}}{2\pi}\int_{-\infty}^{\infty} dk\, \frac{e^{-ik(x-x_0)}}{\mu\mu_0} \times \left(\frac{c}{4\pi}\frac{1}{(k-\omega)(k-\omega_0)\Lambda_3(k)} + \frac{\delta(\boldsymbol{\Omega}-\boldsymbol{\Omega}_0)}{k-\omega_0}\right). \tag{2.38}$$

Here we have used Eq. (2.18). The equivalence of Eqs. (2.38) and (1.48) is easily seen. Indeed, for $x > x_0$, we may close the integration contour of Eq. (2.38) in the lower half-plane by means of the path γ_- which excludes the cut l_- of $\Lambda_3(k)$ (cf. Fig. 3) and find

$$\hat{\phi}_G(x, \mathbf{k}_{(x)}, \boldsymbol{\Omega}) = \hat{\Theta}(\mu_0)\frac{\delta(\boldsymbol{\Omega}-\boldsymbol{\Omega}_0)}{\mu_0} e^{i\mathbf{k}_{(x)}\cdot\mathbf{r}_{(x)0}} e^{-i\omega_0(x-x_0)} + \frac{ice^{i\mathbf{k}_{(x)}\cdot\mathbf{r}_{(x)0}}}{4\pi\mu_0\mu}\left(\frac{e^{-\kappa_0(x-x_0)}}{(i\kappa_0+\omega)(i\kappa_0+\omega_0)\Lambda_3'(-i\kappa_0)} + \hat{\Theta}(\mu_0)\frac{e^{-i\omega_0(x-x_0)}}{(\omega_0-\omega)\Lambda_3(\omega_0)} + \frac{\hat{\Theta}(\mu)e^{-i\omega(x-x_0)}}{(\omega-\omega_0)\Lambda_3(\omega)} - \frac{1}{2\pi i}\int_{\gamma_-} dk\, \frac{e^{-ik(x-x_0)}}{(k-\omega)(k-\omega_0)\Lambda_3(k)}\right), \tag{2.39}$$

where

$$\hat{\Theta}(\mu) \equiv 1, \quad \mu > 0, \quad \omega \notin l_-, \qquad \equiv 0, \quad \mu < 0 \quad \text{or} \quad \omega \in l_-. \tag{2.40}$$

Equation (2.39), with $\mathbf{r}_0 = 0$, differs from Eq. (1.48) only in notation; for example,

$$\omega = -i/\zeta, \qquad \kappa_0 = 1/\nu_0, \tag{2.41}$$

and the integration variables of the branch-cut integrals in Eqs. (2.39) and (1.48) are related by

$$k = -i/\zeta'. \tag{2.42}$$

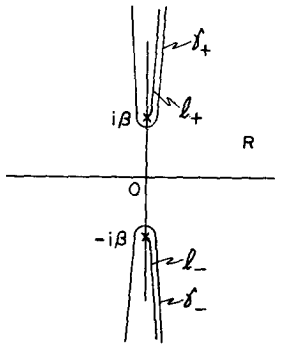


FIG. 3. The region R of the complex k-plane.

C. Half-Space Problems

Somewhat less trivial, but still quite straightforward, is the application of the method based on Eq. (2.12) to problems involving half-spaces.

When V is the region $\{x > 0, \ -\infty < y < \infty, \ -\infty < z < \infty\}$, we have

$$\tilde{\rho}_V(\mathbf{k}) = \frac{1}{(2\pi)^3}\int_0^\infty dx e^{ik_x x}\int_{-\infty}^{\infty} d\mathbf{r}_{(x)} e^{i\mathbf{k}_{(x)}\cdot\mathbf{r}_{(x)}}\rho(x, \mathbf{r}_{(x)}) \tag{2.43}$$

$$= \tilde{\rho}_+(k), \tag{2.44}$$

where $k \equiv k_x$ and we omit reference to the transverse variables $\mathbf{k}_{(x)}$. In Eq. (2.44) we use a common notation: the subscript indicates analyticity in the upper half k-plane, which property is clear from Eq. (2.43). If we similarly use the function

$$\tilde{\rho}_-(k) = \tilde{\rho}(k) - \tilde{\rho}_+(k) \tag{2.45}$$

which is analytic for Im $(k) < 0$, then Eq. (2.12) may be written in the form

$$\Lambda_3(k)\rho_+(k) = -\rho_-(k) + \tilde{B}(k) + \tilde{Q}(k), \quad k \text{ real}. \tag{2.46}$$

Equation (2.46) is of a standard form and may be solved by well-known methods.[9] It is convenient to begin by finding a "Wiener–Hopf factorization" of the function $\Lambda_3(k)$:

$$\Lambda_3(k) = \Lambda_+(k)/\Lambda_-(k), \tag{2.47}$$

where $\Lambda_\pm(k)$ is analytic for Im $(k) \gtrless 0$ and both factors have at most polynomial growth at $k \sim \infty$. We find the $\Lambda_\pm(k)$ by a conventional[10] procedure. First define the region R of the complex plane as being the entire plane with the exclusion of small neighborhoods of the cuts l_+ and l_-. Thus the boundary of R consists of two contours, γ_+ and γ_-, enclosing the lines l_+ and l_-, as in Fig. 3. From our discussion of $\Lambda_3(k)$ above, it is clear that the function

$$L(k) \equiv \ln\left[\Lambda_3(k)(k^2+\beta^2)/(k^2+\kappa_0^2)\right] \tag{2.48}$$

is analytic in R, and that (with the proper branch choice)

$$L(k) \xrightarrow[k\to\infty]{} 0, \quad k \in R. \tag{2.49}$$

Hence we may write

$$L(k) = L_+(k) + L_-(k), \quad k \in R, \tag{2.50}$$

where

$$L_\pm(\ ) = \frac{1}{2\pi i}\int_{\gamma_\mp}\frac{L(k')}{k'-k}\,dk'. \tag{2.51}$$

Equation (2.50) implies that

$$\Lambda_3(k) = [(k^2 + \kappa_0^2)/(k^2 + \beta^2)]e^{L_+(k)+L_-(k)}, \quad k \in R, \tag{2.52}$$

so that the choice

$$\Lambda_+(k) = [(k + i\kappa_0)/(k + i\beta)]e^{L_+(k)}, \tag{2.53}$$

$$\Lambda_-(k) = [(k - i\beta)/(k - i\kappa_0)]e^{-L_-(k)} \tag{2.54}$$

satisfies both the analyticity requirements and Eq. (2.47). Furthermore, it is evident that

$$\Lambda_\pm(k) \xrightarrow[k\to\infty]{} 1. \tag{2.55}$$

We will take the $\Lambda_\pm(k)$ as being defined by Eqs. (2.53) and (2.54) even for $k \notin R$ (i.e., $k \in l_+$ or $k \in l_-$). With this convention, it should be noted that Eq. (2.47) holds only for $k \in R$. It is not hard to show that, for $k \in R$,

$$\frac{1}{2\pi i}\int_{\gamma_-} \frac{\ln[(k'^2 + \beta^2)/(k'^2 + \kappa_0^2)]}{k' - k}\,dk' = 0, \tag{2.56}$$

whence

$$L_+(k) = \Gamma_+(k), \quad \text{for} \quad k \in R, \tag{2.57}$$

where

$$\Gamma_+(k) = \frac{1}{2\pi i}\int_{\gamma_-} \frac{\ln \Lambda_3(k')}{k' - k}\,dk'. \tag{2.58}$$

$\Gamma_+(k)$ is clearly related to the function $\gamma(\zeta)$ of Eq. (1.57), and for this reason will be found useful below.

Further identities and simplifications concerning the $\Lambda_\pm(k)$ are easily deduced[11]; but we wish to return our attention to Eq. (2.48), which may now be written in the form

$$\tilde{\rho}_+(k)\Lambda_+(k) + \tilde{\rho}_-(k)\Lambda_-(k) = [\tilde{B}(k) + \tilde{Q}(k)]\Lambda_-(k), \quad k \text{ real}. \tag{2.59}$$

Thus if we define the function

$$\begin{aligned} F(k) &\equiv \tilde{\rho}_+(k)\Lambda_+(k), \quad \operatorname{Im}(k) > 0, \\ &\equiv -\tilde{\rho}_-(k)\Lambda_-(k), \quad \operatorname{Im}(k) < 0, \end{aligned} \tag{2.60}$$

then $F(k)$ is analytic in the plane cut along the real axis, and has a discontinuity along the cut given by

$$F^+(k) - F^-(k) = [\tilde{B}(k) + \tilde{Q}(k)]\Lambda_-(k). \tag{2.61}$$

It follows in a well-known[9] way that (with properly behaved $\tilde{B}$ and $\tilde{Q}$)

$$F(k) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} dk' \frac{[\tilde{B}(k') + \tilde{Q}(k')]\Lambda_-(k')}{k' - k} + P(k). \tag{2.62}$$

Here $P(k)$ is an entire function which can generally be taken to be zero.[12] Thus, from Eq. (2.60),

$$\tilde{\rho}_+(k) = \frac{1}{2\pi i\Lambda_+(k)}\int_{-\infty}^{\infty} dk' \frac{[\tilde{B}(k') + \tilde{Q}(k')]\Lambda_-(k')}{k' - k - i0} \tag{2.63}$$

and the general half-space problem is solved.

As a particular example we consider again the albedo problem, in which there is no external source

$$\tilde{Q}(k) = 0 \tag{2.64}$$

but a boundary condition of the form

$$\phi_s(0, y, z, \boldsymbol{\Omega}) = \delta(y)\delta(z)\delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_0), \quad \mu, \mu_0 > 0, \tag{2.65}$$

$$\Rightarrow \tilde{B}(k) = i/(k - \omega_0), \quad \operatorname{Im}\omega_0 < 0. \tag{2.66}$$

The integral of Eq. (2.63) is entirely trivial and we find[13]

$$\tilde{\rho}_{a+}(k) = i\Lambda_-(\omega_0)/(k - \omega_0)\Lambda_+(k), \tag{2.67}$$

whence [from Eq. (2.9)]

$$\hat{\phi}_a(\mathbf{k}, \boldsymbol{\Omega}) = \frac{i\delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_0)}{k - \omega_0} - \frac{c}{4\pi\mu\Lambda_+(k)(k - \omega_0)(k - \omega)}. \tag{2.68}$$

To check that Eq. (2.68) agrees with Eq. (1.59), it is only necessary to observe that, for $\mu < 0$, $\hat{\phi}_a$ is analytic in the upper-half $k = k_x$ plane except for a pole at $k = \omega$. Hence, for $x = 0$, the inverse k_x transform involves only a simple residue calculation. Using the explicit forms of the $\Lambda_\pm$ functions, we find the emergent angular density

$$\hat{\phi}_a(0, \mathbf{k}_{(x)}, \boldsymbol{\Omega}) = -\frac{ic}{4\pi\mu}\frac{(\omega_0 + i\kappa_0)(\omega + i\beta)e^{\Gamma_+(\omega_0)-\Gamma_+(\omega)}}{(\omega_0 + i\beta)(\omega + i\kappa_0)(\omega - \omega_0)[\Lambda_3(\omega_0)]}, \quad \mu < 0. \tag{2.69}$$

It is easy to verify that

$$\Gamma_+(-i/\zeta) - \Gamma_+(0) = \gamma(\zeta), \tag{2.70}$$

whence, in view of the relations (2.49) and the fact that $\beta = \alpha^{-1}$, the equivalence of Eqs. (2.69) and (1.60) becomes evident.

We could now go on to consider problems in which the region V is finite in the x direction, i.e., slab problems. It is in fact possible in this case also to derive from Eq. (2.12) a general prescription ("general" in the sense that the particular sources and boundary data need not be specified beforehand)

which provides, at least in the usual wide slab approximation, the desired transformed densities.[11] This prescription involves a combination of function-theoretic techniques with a Fredholm-like iterative procedure. But we will not delve into these, or other still more complicated problems here; it is hoped that the above examples suffice to demonstrate the workability of the technique we have outlined with regard to that class of problems to which the singular eigenfunction technique based on Eq. (1.10) is applicable.

CONCLUSION

Our comparison of the two methods for solving problems in what might be called "quasi-3-dimensional" linear transport theory may be summarized in the following remarks—remarks which can be expected to apply regardless of the particular problem considered.

By expanding in terms of the set of functions $\{e^{ikx}; k \text{ real}\}$ we are led naturally to the dispersion function $\Lambda_3(k)$, which can be expressed in closed form and which has fairly simple analytic properties. On the other hand, use of the set of functions provided by Eqs. (1.21)–(1.24) (the orthogonality and completeness of which is neither well known nor trivial to prove) leads one to the much more mathematically formidable function $\Lambda(\zeta)$; one must then use the theory of generalized analytic functions to eliminate the pathologies of the latter, so that final answers ultimately may be expressed in terms of $\hat{\Lambda}(\zeta) = \Lambda_3(-i/\zeta)$.

A second and, perhaps, more significant difference between the two methods is that, according to the prescription given in Sec. 2, one finds the transformed density $\tilde{\rho}(k)$ [from which the transformed angular density $\tilde{\phi}(\mathbf{k}, \mathbf{\Omega})$ is trivially obtained] first, rather than, as in the singular-eigenfunction technique, determining the angular density first. The result is that the former method is almost entirely free of the mathematical complexities associated with 2-dimensional angular integrals.

Of course both these distinctions are without force in the 1-dimensional $\mathbf{B}_{(x)} = 0$ theory.[14] But when $\mathbf{B}_{(x)} \neq 0$, they have the effect of making the Fourier-transform analysis substantially more elementary and direct.

* This work was supported in part by the National Science foundation.

† Present address: The Institute for Advanced Study, Princeton, N.J. 08540.

[1] See, for example, K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).

[2] H. G. Kaper, J. Math. Phys. **10**, 286 (1969).

[3] I. N. Vekua, *Generalized Analytic Functions* (Addison-Wesley, Reading, Mass., 1962).

[4] Note that Eq. (1.27) is simply the statement that $(1/\pi) \ln \zeta$ is the Green's function for the 2-dimensional Laplacian.

[5] The Jacobians in Eqs. (1.33) and (1.34) were omitted in Ref. 2. These equations also differ from those in the Reference by a trivial factor of 4π, due to different source normalizations.

[6] That final "answers" can *in general* be expressed in terms of Λ only will become evident below; this circumstance allows one to ignore such apparently difficult questions (see Ref. 2) as the possible vanishing of Λ in G.

[7] The factor multiplying the δ function in Eq. (1.50), which is called for by Eqs. (1.9) and (1.49), was omitted in Ref. 2. For this reason the coefficients we give in Eqs. (1.52) and (1.53) differ slightly from those of Kaper.

[8] This fact is exploited,in particular, when the region V is finite in the x direction. (See R. D. Hazeltine, thesis, University of Michigan, 1968.)

[9] See, for example, N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Groningen, The Netherlands, 1953).

[10] See, for example, P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part 1.

[11] R. D. Hazeltine, thesis, University of Michigan, 1968.

[12] Nonzero $P(k)$ clearly implies δ-function singularities in the density $\rho(x)$.

[13] The result, Eq. (2.67), could in this case have been obtained by a simple function-theoretic argument directly from Eq. (2.59) without using the general integral prescription.

[14] That the two methods are equivalent in this case is well known; cf. K. M. Case, Ann. Phys. (N.Y.) 7, 349 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Singular Three-Body Amplitudes in the Theory of the Third Virial Coefficient*

ROGER DASHEN†
Institute for Advanced Study, Princeton, New Jersey 08540
AND
SHANG-KENG MA
Department of Physics and Institute for Pure and Applied Physical Sciences
University of California, San Diego, La Jolla, California 92037

(Received 3 October 1969)

We have shown earlier how the virial coefficients can be expressed in terms of traces taken over products of multiparticle (on-energy-shell) scattering matrices. The traces are to be taken in the angular-momentum representation instead of the momentum representation to avoid the infinite forward N-particle scattering amplitude for $N > 2$ caused by certain singular diagrams. Here, we analyze and sum the contribution of the singular diagrams to the third virial coefficient. It is also shown that one can get the same result if off-shell amplitudes in the momentum representation are utilized.

1. INTRODUCTION

It was shown earlier that, for a gas, the virial coefficients can be expressed in terms of the scattering matrix elements describing the collisions of the constituent particles.[1] For example, the second virial coefficient is expressible in terms of the 2-body S matrix elements and, for calculating the third virial coefficient, it is sufficient to know the 3-body S matrix in addition to the 2-body S matrix.

For the scattering involving three or more particles, there are certain "singular terms" in the S matrix which are characterized by the fact that they give rise to infinite forward scattering amplitudes. The purpose of this paper is to study the contribution of these singular terms to the third virial coefficient.

The virial coefficients can be obtained from the coefficients b_N in the expansion $\Omega \sim \sum_N b_N z^N$ for the grand potential Ω in powers of $z \equiv e^{\beta\mu}$, where β^{-1} and μ are the temperature and the chemical potential, respectively. It was shown in Paper I that, leaving out numerical factors and other details,

$$b_N \sim \mathrm{Tr}\,(e^{-\beta H} - e^{-\beta H_0}) = \int d\epsilon e^{-\beta\epsilon}(4\pi i)^{-1}\,\mathrm{Tr}\,S^{-1}\frac{\overleftrightarrow{\partial}}{\partial\epsilon}S. \qquad (1.1)$$

The trace is taken over the free N-particle states, and S is the S matrix for the N-particle scattering at the center-of-mass energy ϵ. To calculate the third virial coefficient, it is sufficient to know b_2 and b_3.

Since, as is well known, the S matrix contains numerous singularities, one must be careful to specify exactly what is meant by the trace on the right-hand side of (1.1). In Paper I, it was pointed out that the only singularities which really cause any trouble are bound-state poles and another class of poles arising from multiple-scattering diagrams. The bound-state problem was completely solved in I, and is ignored in this paper. We also indicated the solution to the second problem, which we study in more detail here.

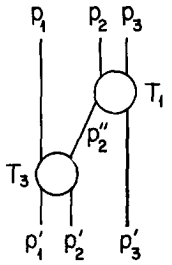


FIG. 1. The double scattering term in the connected 3-body T-matrix.

For $N > 2$ the traces in (1.1) are to be interpreted as sums over states in the angular-momentum representation. That is, one computes the traces in the subspace of each J and then sums over J. Doing things in this order, no serious singularities are encountered and, in fact, the derivation of (1.1) should be understood as first establishing the equality of the two sides for each J and then summing over J. Extreme care must be taken when using any other representation, for example, the momentum representation. This is because the forward N-body scattering amplitude, which is implicitly contained in the rhs of (1.1), is undefined for $N > 2$ owing to certain "singular diagrams" of the type shown in Fig. 1. The forward amplitude blows up due to the vanishing of the intermediate state denominator forced by momentum conservation. The use of the angular momentum gets us around this difficulty. What we do here is study the singular terms for 3-body scattering and sum up their contribution to the third virial coefficient. At the same time we indicate how, with due caution, the momentum representation can be employed.

This paper has the following outline. In Sec. 2, we review the basic aspects of the S matrix formula for b_3 and those of the singular terms. The use of the angular-momentum representation is demonstrated in Sec. 3. The singular terms are then listed and

summed in Sec. 4. In Sec. 5, we show that the results in Sec. 4 can also be obtained by using off-shell amplitudes in the momentum representation. Concluding remarks are given in Sec. 6.

2. REVIEW

A. The Virial Coefficients

Consider a nonrelativistic gas of volume V, temperature β^{-1}, chemical potential μ, and pressure p. The grand potential $\Omega = -pV$ can be written as an expansion in powers of $z \equiv e^{\beta\mu}$,

$$\Omega = \Omega_0 - V\beta^{-1}\lambda^{-3}\sum_{N=2}^{\infty} b_N z^N, \tag{2.1}$$

where

$$\lambda = (m/2\pi\beta)^{-\frac{1}{2}} \tag{2.2}$$

and where $\Omega_0 = -\beta^{-1}\lambda^{-3}V$ is the grand potential for an ideal gas, and m is the mass of a gas molecule. When z is eliminated using (2.1) and the equation for the density n, i.e.,

$$n = -V^{-1}\partial\Omega/\partial\mu, \tag{2.3}$$

one obtains the virial series for the pressure

$$p = n\beta^{-1}\sum_{N=1}^{\infty} a_N(\lambda^3 n)^{N-1}, \tag{2.4}$$

where a_N, the Nth virial coefficient, can be expressed in terms of the b_N in (2.1). One easily verifies that

$$\begin{aligned} a_1 &= 1, \\ a_2 &= -b_2, \\ a_3 &= 4b_2^2 - 2b_3, \end{aligned} \tag{2.5}$$

and so on. To obtain the third virial coefficient a_3, one needs to know b_2 and b_3. The procedure of calculating b_2 is well known.[2] We shall only be interested in b_3 here.

It was shown in Paper I that the coefficients b_N can be obtained from the N-particle scattering-matrix elements. In particular,

$$b_3 = 3^{\frac{3}{2}}\int_0^\infty d\epsilon e^{-\beta\epsilon}(4\pi i)^{-1}\left(\mathrm{Tr}\, AS^{-1}\frac{\overleftrightarrow{\partial}}{\partial\epsilon}S\right)_c, \tag{2.6}$$

which follows from Eq. (5.10) of Paper I. The trace is taken over the free 3-particle states in the center-of-mass (CM) frame and, for reasons mentioned above, in the angular-momentum basis. The quantity ϵ is the total CM energy. The symbol A denotes the antisymmetrization (for fermions) or symmetrization (for bosons) operation to include exchange diagrams due to the identity of particles. The subscript c denotes that only the connected diagrams are kept.

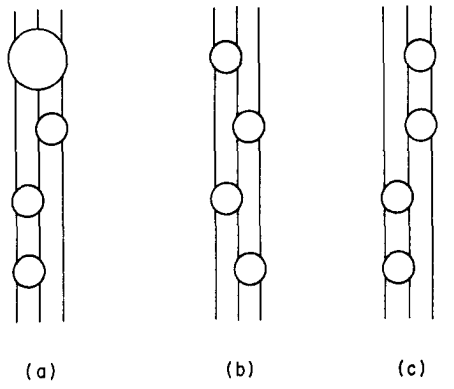


FIG. 2. Some $n = 4$ terms in the expansion (2.9).

In terms of the T matrix, the operator S is given by

$$S(\epsilon) = 1 - 2\pi i\delta(\epsilon - H_0)T(\epsilon). \tag{2.7}$$

It was shown in I that only the on-shell T-matrix elements, i.e., those matrix elements of $T(\epsilon)$ between states with energy ϵ, appear in (2.6).

The trace in (2.6) can also be expressed as a logarithmic derivative

$$(4\pi i)^{-1}\left(\mathrm{Tr}\, AS^{-1}\frac{\overleftrightarrow{\partial}}{\partial\epsilon}S\right)_c = (2\pi)^{-1}\frac{\partial}{\partial\epsilon}\,\mathrm{Im}\,(\mathrm{Tr}\, A\ln S)_c. \tag{2.8}$$

One can expand the logarithm as a series so that (2.8) becomes

$$-(2\pi)^{-1}\frac{\partial}{\partial\epsilon}\sum_{n=1}^{\infty} n^{-1}\,\mathrm{Im}\,[\mathrm{Tr}\, A(2\pi i\delta(\epsilon - H_0)T)^n]_c. \tag{2.9}$$

This expansion will play an important role in our subsequent discussions.

The 3-particle T matrix T contains the connected part T_c and the disconnected terms T_i, $i = 1, 2, 3$. T_i describes the scattering process in which the ith particle moves freely while the other two interact. The terms in the series (2.9) can be represented by diagrams. For example, Fig. 2 shows some terms with $n = 4$.

B. Singular Terms

It is clear that, if the T-matrix elements are nonsingular functions of ϵ, only the on-shell T-matrix elements appear in (2.9) because of the factors $\delta(\epsilon - H_0)$. When the T-matrix elements have singularities, one must be more careful.

According to the analysis in Paper I, there are two kinds of singular terms in T. First, the existence of bound states or composite particles gives rise to poles of the T matrix outside the spectrum of H_0. The bound-state problem was treated in detail in Paper I. The other kind of singular term in T is associated with

the "double scattering" diagram shown in Fig. 1. It is a term in the connected part T_c. It describes the scattering of particles 1 and 2 followed by a scattering of particles 2 and 3. The momentum of 2 in the intermediate state is completely determined by the initial and final momenta. The energy denominator of the intermediate state is given by

$$[\epsilon - \omega(p_1) - \omega(p_2'') - \omega(p_3') + i\eta]^{-1}, \quad (2.10)$$

where $\omega(p) = p^2/2m$. When the T-matrix elements are on-shell, i.e., when $\epsilon = \omega(p_1) + \omega(p_2) + \omega(p_3)$, the term (2.10) blows up for the forward-scattering matrix elements, i.e., for $\mathbf{p}_1 = \mathbf{p}_1'$, $\mathbf{p}_2 = \mathbf{p}_2'$, and $\mathbf{p}_3 = \mathbf{p}_3'$. Therefore, because of the presence of the double scattering term, the forward on-shell T-matrix elements are undefined in the momentum representation.[3] Besides the singular term in T_c, there are terms in the series (2.9) which are not defined in the momentum representation. For example, in Fig. 2(c), the δ function for the middle part of the diagram blows up. From now on all terms not defined in the momentum representation due to the singularities described above are called "singular."

It was shown in Paper I that (2.8) and (2.9) are well defined if one uses the angular-momentum eigenstates instead of the momentum eigenstates. The matrix elements in the angular-momentum representation are continuous superpositions of scattering amplitudes of initial and final momenta in all directions. The singularity in the forward direction is thus smeared out.

The use of the angular-momentum basis is not just a mathematical device to smear out the forward amplitudes. It is in fact a physically reasonable choice from the beginning. Recall that the S-matrix formula for b_N comes from the identity

$$\mathrm{Tr}\,(e^{-\beta H} - e^{-\beta H_0}) = \int d\epsilon e^{-\beta\epsilon}(4\pi i)^{-1}\,\mathrm{Tr}\,S^{-1}\frac{\overleftrightarrow{\partial}}{\partial\epsilon}S. \quad (2.11)$$

Since the total angular momentum J is a conserved quantity, the identity (2.11) can be derived for each J. The S matrix can be calculated for each J independently. One then sums the series in J. This way the forward amplitudes never appear. The formal manipulations in the angular-momentum representation is discussed in the next section.

3. b_3 IN AN ANGULAR-MOMENTUM BASIS

A. General Formulas

We proceed to show how the trace can be evaluated in the angular-momentum representation given the matrix elements in the momentum representation.

In the CM frame, the momenta of the three particles form a triangle. They are fixed by the orientation of the triangle and the lengths of the three sides or, equivalently, the energies of the three particles. Let the orientation of the triangle be specified by a rotation R from some standard orientation. R is given by a set of Eulerian angles. Let the energies of the three particles be $\omega = (\omega_1, \omega_2, \omega_3)$. Thus, the pair (R, ω) is equivalent to the specification of the three momenta $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$.

It is convenient for our discussion to use the angular-momentum eigenstates

$$\psi_{JM\mu\omega}(R, \omega') = [\tfrac{1}{8}\pi^{-2}(2J+1)]^{\frac{1}{2}}\mathfrak{D}^J_{M\mu}(R) \times \delta(\omega_1' - \omega_1)\delta(\omega_2' - \omega_2)\delta(\omega_3' - \omega_3), \quad (3.1)$$

where J is the total angular momentum, M the projection of J along the Z axis, and μ the projection of J along some axis fixed in the momentum triangle. These states have been studied by Omnes.[4]

Let $\mathcal{O}(R\omega, R'\omega')$ denote the matrix element $\langle \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3|\,\mathcal{O}\,|\mathbf{p}_1'\mathbf{p}_2'\mathbf{p}_3'\rangle$ of the operator $\mathcal{O}$. Then, by (3.1), we have the matrix elements of $\mathcal{O}$ in the angular-momentum representation

$$\langle JM\mu\omega|\,\mathcal{O}\,|J'M'\mu'\omega'\rangle = \int dR\,dR'\mathfrak{D}^J_{M\mu}(R)\mathcal{O}(R\omega, R'\omega')\mathfrak{D}^J_{M'\mu'}(R') \times \tfrac{1}{8}\pi^{-2}(2J+1)^{\frac{1}{2}}(2J'+1)^{\frac{1}{2}}, \quad (3.2)$$

where dR stands for $d\alpha\, d\cos\beta\, d\gamma$ in which α, β, and γ are the Eulerian angles specifying R.

If the operator $\mathcal{O}$ is rotationally invariant, we have

$$\mathcal{O}(R\omega, R'\omega') = \mathcal{O}(\omega, R^{-1}R'\omega'). \quad (3.3)$$

In this case, we let $R'' = R^{-1}R'$ and use the orthogonality relations of the $\mathfrak{D}$ functions to reduce (3.2). We obtain

$$\langle JM\mu\omega|\,\mathcal{O}\,|J'M'\mu'\omega'\rangle = \delta_{JJ'}\delta_{MM'}\int dR''\mathfrak{D}^J_{\mu\mu'}(R'')\mathcal{O}(\omega, R''\omega'). \quad (3.4)$$

Since the operators we encounter are all rotationally invariant, (3.4) is very useful in obtaining the angular-momentum representation from the momentum representation. Finally, the sum over states has the form

$$\int d^3p_1\,d^3p_2\,d^3p_3\delta(\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3) = \sum_{J,M,\mu} m^3\int d\omega_1\,d\omega_2\,d\omega_3 \quad (3.5)$$

Substituting (2.8) and (2.9) in (2.6) and using the angular-momentum representation for the trace, we

have

$$b_3 = 3^{\frac{3}{2}}(2\pi)^{-6}\beta\int d\epsilon e^{-\beta\epsilon}(2\pi)^{-1}\,\mathrm{Im}\,(\mathrm{Tr}\,A\ln S)_c$$
$$= 3^{\frac{3}{2}}(2\pi)^{-6}\sum_{J,\mu}(2J+1)Q_{J\mu}, \tag{3.6}$$

where

$$Q_{J\mu} = \int d\epsilon e^{-\beta\epsilon}\tfrac{1}{6}\prod_{i=1}^{3} m\,d\omega_i\beta(2\pi)^{-1}$$
$$\times\,\mathrm{Im}\,\langle JM\mu\omega|\,A\ln S\,|JM\mu\omega\rangle_c$$
$$= -\beta\int\prod_{i=1}^{3} m\,d\omega_i e^{-\beta\omega_i}$$
$$\times\,\tfrac{1}{6}\sum_{n=1}^{\infty} n^{-1}\,\mathrm{Re}\,\langle JM\mu\omega|$$
$$\times\,AT[2\pi i\delta(\omega_1+\omega_2+\omega_3-H_0)T]^{n-1}$$
$$\times\,|JM\mu\omega\rangle_c, \tag{3.7}$$

where the factor $\frac{1}{6}$ is put in to avoid over counting.

B. The Simplest Singular Term

Let us consider the following example involving a singular term to illustrate the use of the angular-momentum basis and the fact that the singular terms are well defined in the angular-momentum representation.

Consider the first term in the sum in (3.7), i.e.,

$$c \equiv \mathrm{Re}\,\langle JM\mu\omega|\,T\,|JM\mu\omega\rangle, \tag{3.8}$$

and suppose that T is given by

$$T = g^2(\omega_1+\omega_2+\omega_3-H_0+i\eta)^{-1}, \tag{3.9}$$

where g is a constant. Equation (3.9) is the singular term given by Fig. 1 with $T_1 = T_3 = g$ and $\epsilon = \omega_1+\omega_2+\omega_3$, i.e., it is on-shell.

For simplicity, we let $m = \frac{1}{2}$ so that $\omega_i = p_i^2$. The relevant momenta are shown in Fig. 3. The matrix elements of T in the momentum representation are

$$T(\omega, R\omega)$$
$$= g^2(\omega_1+\omega_2+\omega_3-\omega_1-\omega_2''-\omega_3'+i\eta)^{-1}$$
$$= g^2[(-\mathbf{p}_1-\mathbf{p}_3)^2-(-\mathbf{p}_3'-\mathbf{p}_1)^2+i\eta]^{-1}$$
$$= g^2[2(\omega_1\omega_3)^{\frac{1}{2}}\hat{\mathbf{p}}_1\cdot(\hat{\mathbf{p}}_3-\hat{\mathbf{p}}_3')+i\eta]^{-1}. \tag{3.10}$$

Thus, by (3.4), we have

$$c = \tfrac{1}{2}g^2(\omega_1\omega_3)^{-\frac{1}{2}}\int dR\mathfrak{D}^J_{\mu\mu}(R)P(\hat{\mathbf{p}}_1\cdot\hat{\mathbf{p}}_3-\hat{\mathbf{p}}_1\cdot\hat{\mathbf{p}}_3')^{-1}, \tag{3.11}$$

where P denotes the principle part. The rotation R brings $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$ to $(\mathbf{p}_1', \mathbf{p}_2', \mathbf{p}_3')$. Notice that the on-shell T-matrix element given by (3.10) has only a pole in the angle variable $\hat{\mathbf{p}}_1\cdot\hat{\mathbf{p}}_3'$, but not in the energy variables.

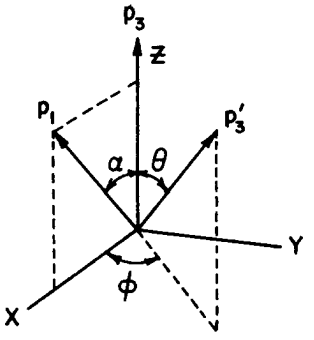


FIG. 3. Geometry of the initial and final momenta. $\mathbf{p}_1$ is in the (x, z) plane.

To carry out the R integral in (3.11), we first integrate over the azimuthal angle around $\mathbf{p}_3'$. This gives 2π for $\mu = 0$ and zero otherwise. One then integrates over the angles θ and φ as shown in Fig. 3. Thus, (3.11) becomes

$$c = \tfrac{1}{2}\delta_{\mu 0}g^2(\omega_1\omega_3)^{-\frac{1}{2}}2\pi\int_{-1}^{1} d\cos\theta$$
$$\times\int_0^{2\pi} d\varphi P_J(\cos\theta)P(\hat{\mathbf{p}}_1\cdot\hat{\mathbf{p}}_3-\hat{\mathbf{p}}_1\cdot\hat{\mathbf{p}}_3')^{-1}, \tag{3.12}$$

where we have made use of the fact that

$$\mathfrak{D}^J_{00}(R) = P_J(\cos\theta). \tag{3.13}$$

The integrals in (3.12) are well defined. The summation over J has been carried out in Paper I. (See the end of Sec. 7 of Paper I.)

4. SINGULAR TERMS

A. Summation

In this section, we list and sum the singular terms in the series

$$-(\mathrm{Tr}\ln S)_c = \sum_{n=1}^{\infty} n^{-1}\{\mathrm{Tr}\,[2\pi i\delta(\epsilon-H_0)T]^n\}_c. \tag{4.1}$$

The trace is always taken in the angular-momentum representation so that all the T-matrix elements are well defined and on-shell. We ignore the symmetrization or antisymmetrization operator A to avoid unnecessary complications. To simplify the notation, let

$$\mathcal{C} \equiv 2\pi i\delta(\epsilon-H_0)T, \tag{4.2}$$

$$\mathcal{C}_{ij} \equiv 2\pi i\delta(\epsilon-H_0)T_iG_0T_j. \tag{4.3}$$

The full T matrix can be written as

$$\mathcal{C} = \sum_i \mathcal{C}_i + \sum_{i\neq j}\mathcal{C}_{ij} + \mathcal{C}', \tag{4.4}$$

where $\mathcal{C}'$ contains no singular term.

The singular terms are shown in Fig. 4. For clarity, we use a wavy line to indicate the factor G_0 in contrast to the $2\pi i\delta(\epsilon - H_0)$ factor indicated by the solid lines. The reader can easily convince himself that there is no other singular term in (4.1).

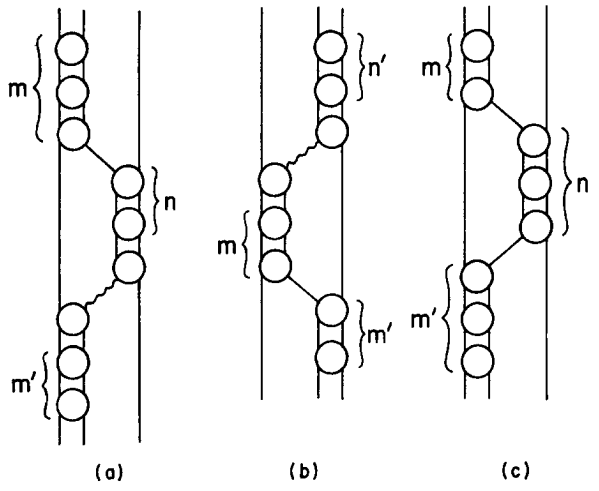


FIG. 4. Singular terms in (2.9).

The contribution of Fig. 4(a) is

$$\sum_{n,m,m'=0}^{\infty} \mathrm{Tr}\, \mathcal{C}_i^m \mathcal{C}_j^n \mathcal{C}_{ji} \mathcal{C}^{m'} (m+n+m'+1)^{-1}$$
$$= \sum_{l,n=0}^{\infty} \mathrm{Tr}\, \mathcal{C}_i^l \mathcal{C}_j^n \mathcal{C}_{ji} (l+1)(l+n+1)^{-1}. \quad (4.5)$$

Similarly, Fig. 4(b) gives

$$\sum_{l,n=0}^{\infty} \mathrm{Tr}\, \mathcal{C}_i^l \mathcal{C}_j^n \mathcal{C}_{ji} n (n+l+1)^{-1}. \quad (4.6)$$

Combining (4.5) and (4.6), one has

$$\sum_{n,l=0}^{\infty} \mathrm{Tr}\, \mathcal{C}_j^n \mathcal{C}_{ji} \mathcal{C}_i^l = \mathrm{Tr}\,(1-\mathcal{C}_j)^{-1} \mathcal{C}_{ji} (1-\mathcal{C}_i)^{-1}. \quad (4.7)$$

Since

$$(1-\mathcal{C}_j)^{-1} = 1 + 2\pi i \delta(\epsilon - H_0) T_j^\dagger, \quad (4.8)$$

Eq. (4.7) reduces to, by (4.3),

$$\mathrm{Tr}\, 2\pi i \delta(\epsilon - H_0) T_j^\dagger G_0 T_i^\dagger, \quad (4.9)$$

which summarizes Fig. 4(a) and 4(b).

The terms shown in Fig. 4(c) give

$$\sum_{n=1,m'=1,m=0}^{\infty} \mathrm{Tr}\, \mathcal{C}_i^m \mathcal{C}_j^n \mathcal{C}_i^{m'} (m+n+m')^{-1}$$
$$= \sum_{l,n=1}^{\infty} \mathrm{Tr}\, \mathcal{C}_i^l \mathcal{C}_j^n l (n+l)^{-1}$$
$$= \tfrac{1}{2} \sum_{l,n=1}^{\infty} \mathrm{Tr}\, \mathcal{C}_i^l \mathcal{C}_j^n$$
$$= \tfrac{1}{2}\, \mathrm{Tr}\, [(1-\mathcal{C}_i)^{-1} - 1][(1-\mathcal{C}_j)^{-1} - 1]$$
$$= \tfrac{1}{2}\, \mathrm{Tr}\, 2\pi i \delta(\epsilon - H_0) T_i^\dagger 2\pi i \delta(\epsilon - H_0) T_j^\dagger. \quad (4.10)$$

Combining (4.9) and (4.10), one obtains

$$\mathrm{Tr}\, 2\pi i \delta(\epsilon - H_0) T_j^\dagger P(\epsilon - H_0)^{-1} T_i^\dagger. \quad (4.11)$$

We have thus collected all singular terms. Only the imaginary part of (4.1) is of interest. Thus, the singular terms can be summarized as

$$\frac{1}{2\pi} \mathrm{Im}\, [\mathrm{Tr} \ln S]_{\mathrm{sing}}$$
$$= -\tfrac{1}{2} \sum_{i \neq j} \mathrm{Tr}\, \delta(\epsilon - H_0) [T_j^\dagger P(\epsilon - H_0)^{-1} T_i^\dagger + \mathrm{c.c.}]$$
$$= -\sum_{i \neq j} (2\pi)^{-1} \mathrm{Tr}\, \mathrm{Re}\, (\mathcal{C}_{ij} - \mathcal{C}_i \mathcal{C}_j). \quad (4.12)$$

B. Evaluation

For each angular momentum J, Eq. (4.12) can be evaluated in the same way as in the example in the previous section. The only complication here is that T_i and T_j also depend on the energies and angles. Instead of (3.11), we have to calculate

$$\langle JM\mu\omega |\, T_1 P(\omega_1 + \omega_2 + \omega_3 - H_0)^{-1} T_3 \,| JM\mu\omega \rangle$$
$$= \tfrac{1}{2}(\omega_1\omega_3)^{-\frac{1}{2}} \int dR \mathfrak{D}^J_{\mu\mu}(R) T_1 T_3 P(\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_3 - \hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_3')^{-1}$$
$$\equiv \sigma_{J\mu}. \quad (4.13)$$

We use the abbreviation $T_1 T_3$ in (4.13) to denote

$$[\langle \mathbf{p}_1 \mathbf{p}_2 \mathbf{p}_3 |\, T_1(\epsilon) \,| \mathbf{p}_1 \mathbf{p}_2'' \mathbf{p}_3' \rangle$$
$$\times \langle \mathbf{p}_1 \mathbf{p}_2'' \mathbf{p}_3' |\, T_3(\epsilon) \,| \mathbf{p}_1' \mathbf{p}_2' \mathbf{p}_3' \rangle]_{\epsilon = \omega_1 + \omega_2 + \omega_3}. \quad (4.14)$$

The contribution to b_3 from the singular terms is thus, by (3.6),

$$b_3^{\mathrm{sing}} = 3^{\frac{3}{2}} (2\pi)^{-6} \sum_{J,\mu} (2J+1) Q^S_{J\mu}, \quad (4.15)$$

where

$$Q^S_{J\mu} = -\tfrac{1}{8}\beta\, \mathrm{Re} \int d\omega_1\, d\omega_3 \int_{-1}^{1} d\lambda$$
$$\times \exp\{-2\beta[\omega_1 + \omega_3 + (\omega_1\omega_3)^{\frac{1}{2}}\lambda]\}$$
$$\times 2(\omega_1\omega_3)^{\frac{1}{2}} \sigma_{J\mu}. \quad (4.16)$$

The mass m is always set equal to $\frac{1}{2}$. We have made use of the fact that

$$\omega_2 = (-\mathbf{p}_1 - \mathbf{p}_3)^2$$
$$= \omega_1 + \omega_3 + 2(\omega_1\omega_3)^{\frac{1}{2}}\lambda,$$

with λ defined by

$$\lambda \equiv \hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_3. \quad (4.17)$$

We proceed to evaluate (4.15). Let R' be the rotation which turns $\hat{\mathbf{p}}_1$ to $\hat{\mathbf{p}}_3$, i.e., the rotation about the y axis of an angle $-\alpha$ (see Fig. 3). The rotation R in (4.13), of course, turns $(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \hat{\mathbf{p}}_3)$ to $(\hat{\mathbf{p}}_1', \hat{\mathbf{p}}_2', \hat{\mathbf{p}}_3')$. In terms of the Eulerian angles, we have

$$R = (\varphi, \theta, \gamma),$$
$$R' = (0, -\alpha, 0). \quad (4.18)$$

Let us define

$$R'' = R'R = (\varphi'', \theta'', \gamma''). \quad (4.19)$$

Then

$$\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_3 = \hat{\mathbf{p}}_3 \cdot R' \hat{\mathbf{p}}_3 = \cos\alpha = \lambda,$$
$$\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_3' = \hat{\mathbf{p}}_3 \cdot R'' \hat{\mathbf{p}}_3 = \cos\theta'' \equiv x. \quad (4.20)$$

Thus, (4.13) becomes

$$\sigma_{J\mu} = \tfrac{1}{2}(\omega_1\omega_3)^{-\frac{1}{2}} \int dR'' \mathfrak{D}^J_{\mu\mu}(R'^{-1}R'') T_1 T_3 P(\lambda - x)^{-1}. \tag{4.21}$$

Substituting (4.21) in (4.16), we have

$$\begin{aligned} Q^S_{J\mu} = & -\beta \tfrac{1}{8} \int d\omega_1 d\omega_3 e^{-2\beta(\omega_1+\omega_3)} \\ & \times \mathrm{Re} \int_{-1}^{1} d\lambda \int_{-1}^{1} dx \int_0^{2\pi} d\varphi'' \, d\gamma'' \mathfrak{D}^J_{\mu\mu}(R'^{-1}R'') T_1 T_3 \\ & \times \exp[-2\beta\lambda(\omega_1\omega_3)^{\frac{1}{2}}] P(\lambda - x)^{-1}. \end{aligned} \tag{4.22}$$

Only the integration along the line $\lambda = x$, $\varphi'' = \gamma'' = 0$ contributes to the sum (4.15). This is because, when one sums over J, μ, the equality

$$\tfrac{1}{8}\pi^{-2} \sum_{J,\mu} (2J+1) \mathfrak{D}^J_{\mu\mu}(R'^{-1}R'') = \delta(\varphi'')\delta(\lambda - x)\delta(\gamma'') \tag{4.23}$$

shows that the formally summed integrand vanishes except on $\lambda = x$, $\varphi'' = \gamma'' = 0$.

Let

$$F(\lambda, x) = \exp[-2\beta\lambda(\omega_1\omega_3)^{\frac{1}{2}}] T_1 T_3|_{\gamma''=\varphi''=0}, \tag{4.24}$$

$$\xi \equiv \tfrac{1}{2}(\lambda + x),$$

$$\zeta \equiv \lambda - x, \tag{4.25}$$

and expand F in powers of ζ:

$$F(\lambda, x) = F(\xi, \xi) + \frac{1}{2}\left(\frac{\partial F}{\partial \lambda} - \frac{\partial F}{\partial x}\right)_{x=\lambda=\xi} \zeta + O(\zeta^2). \tag{4.26}$$

We now substitute (4.26) in (4.22) first and then sum over J and μ using the identity (4.23). The first term in (4.26) does not contribute by symmetry. The second and third terms remain nonsingular after the multiplication by $P(1/\zeta)$. We obtain

$$\begin{aligned} & \sum_{J,\mu} (2J+1) Q^S_{J\mu} \\ & = \mathrm{Re}\, 4\pi^2(-\beta)\tfrac{1}{8} \int d\omega_1 \, d\omega_2 e^{-2\beta(\omega_1+\omega_3)} \\ & \quad \times \int_{-1}^{1} d\xi \left(\frac{\partial F}{\partial \lambda} - \frac{\partial F}{\partial x}\right)_{x=\lambda=\xi} \\ & = 4\pi^2(-\beta)\tfrac{1}{8}\, \mathrm{Re} \int d\omega_1 \, d\omega_3 e^{-2\beta(\omega_1+\omega_3)} \\ & \quad \times \int_{-1}^{1} d\lambda \exp[-2\beta\lambda(\omega_1\omega_3)^{\frac{1}{2}}] \\ & \quad \times \left[-2\beta(\omega_1\omega_3)^{\frac{1}{2}} T_1 T_3 + \left(\frac{\partial}{\partial \lambda} - \frac{\partial}{\partial x}\right) T_1 T_3\right]_{x=\lambda, \varphi''=\gamma''=0}. \end{aligned} \tag{4.27}$$

Substituting (4.27) in (4.15), one obtains the contribution of the singular terms to b_3.

The condition that $x = \lambda$, $\varphi'' = \gamma'' = 0$ is the same as that $\mathbf{p}_i = \mathbf{p}_i'$, $i = 1, 2, 3$. The first term in (4.27) thus involves the product of two forward 2-body scattering amplitudes. The second term in (4.27) involves the derivatives of the 2-body amplitudes with respect to the scattering angles evaluated in the forward direction.

We have thus completed the evaluation of the singular terms. To gain further insight, we discuss in the next section an alternative form of (4.27) and derive the same result by using off-shell amplitudes in the momentum representation.

5. FURTHER DETAILS

A. Alternative Expression for b_3^{sing}

We now proceed to show that these derivatives with respect to angles in (4.27) can be written in terms of the derivatives with respect to the (off-shell) energy variable ϵ.

Let us analyze the dependence of $T_1 T_3$ on x and λ. First, consider the 2-body T matrix T_3. Figure 5(a) shows a general term in the perturbation expansion of T_3 in powers of the 2-body potential. The contribution of this diagram to the matrix element

$$\langle \mathbf{p}_1 \mathbf{p}_2'' \mathbf{p}_3 | \, T_3(\epsilon) \, | \mathbf{p}_1' \mathbf{p}_2' \mathbf{p}_3' \rangle$$

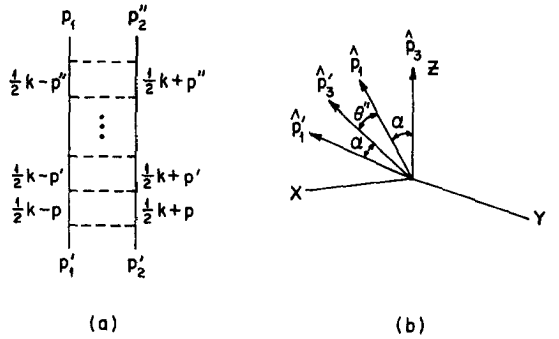
is

$$\begin{aligned} & \int \frac{d^3p}{(2\pi)^3} \frac{d^3p'}{(2\pi)^3} \cdots \frac{d^3p''}{(2\pi)^3} V(\mathbf{p}_1 - \tfrac{1}{2}\mathbf{k} + \mathbf{p}'') \\ & \quad \times \cdots \times V(\mathbf{p}' - \mathbf{p}) V(\mathbf{p}_1' - \tfrac{1}{2}\mathbf{k} + \mathbf{p}) \\ & \quad \times (E' - \tfrac{1}{2}k^2 - 2p''^2)^{-1} \\ & \quad \times \cdots \times (E' - \tfrac{1}{2}k^2 - 2p'^2)^{-1} (E' - \tfrac{1}{2}k^2 - 2p^2)^{-1}, \end{aligned} \tag{5.1}$$

where

$$E' \equiv \epsilon - \omega_3,$$

$$\mathbf{k} \equiv \mathbf{p}_1' + \mathbf{p}_2' = -\mathbf{p}_3' \tag{5.2}$$



(a) (b)

FIG. 5. (a) A general term in the perturbation expansion of T_3. (b) The geometry of the momentum unit vectors for $\varphi'' = \gamma'' = 0$.

are the total energy and the total momentum of the 2-body system. We restrict ourselves to the case where $p_i^2 = p_i'^2 = \omega_i$. Thus $k^2 = \omega_3$. The 2-body potential $V(\mathbf{q})$, as a function of the momentum transfer $\mathbf{q}$, is assumed to be spherically symmetric.

By symmetry, the results after integrating over $\mathbf{p}, \mathbf{p}', \cdots, \mathbf{p}''$ in (5.1) can be a function of only the variables

$$\epsilon, \quad p_1^2 = p_1'^2 = \omega_1, \quad k^2 = \omega_3,$$
$$\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{k}}, \quad \hat{\mathbf{p}}_1' \cdot \hat{\mathbf{k}}, \quad \hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_1'. \tag{5.3}$$

Furthermore, this function is unaltered upon interchanging $\mathbf{p}_1$ and $\mathbf{p}_1'$.[5] This can be verified explicitly by permuting the integration variables in (5.1). We are only interested in the case where $\varphi'' = \gamma'' = 0$. Thus, the relevant momenta are as those shown in Fig. 5(b). Since $\mathbf{k} = -\mathbf{p}_3'$ [see Eq. (5.2)], we have

$$\hat{\mathbf{p}}_1' \cdot \hat{\mathbf{k}} = -\hat{\mathbf{p}}_1' \cdot \hat{\mathbf{p}}_3' = -\cos\alpha = -\lambda,$$
$$\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{k}} = -\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_3' = -\cos\theta'' = -x. \tag{5.4}$$

Notice that $\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_1'$ is a symmetric function of x and λ.

We thus conclude from the above analysis of (5.1) that

$$\langle \mathbf{p}_1\mathbf{p}_2''\mathbf{p}_3' | \, T_3(\epsilon) \, | \mathbf{p}_1'\mathbf{p}_2'\mathbf{p}_3' \rangle = f(\epsilon, \omega_3, \omega_1, \lambda, x) = f(\epsilon, \omega_3, \omega_1, x, \lambda), \tag{5.5}$$

where f is some complicated function. By the same arguments, we arrive at the same conclusion for T_1, i.e.,

$$\langle \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3 | \, T_1(\epsilon) \, | \mathbf{p}_1\mathbf{p}_2''\mathbf{p}_3' \rangle = f(\epsilon, \omega_1, \omega_3, \lambda, x) = f(\epsilon, \omega_1, \omega_3, x, \lambda). \tag{5.6}$$

Since

$$\omega_2 = (-\mathbf{p}_1 - \mathbf{p}_3)^2 = \omega_1 + \omega_3 + 2(\omega_1\omega_3)^{\frac{1}{2}}\lambda, \tag{5.7}$$

we have, for the on-shell value of ϵ,

$$\epsilon = 2[\omega_1 + \omega_3 + (\omega_1\omega_3)^{\frac{1}{2}}\lambda]. \tag{5.8}$$

We now substitute (5.5) and (5.6) for T_1T_3 in the last term of (4.27). Since f is a symmetric function of λ and x for fixed ϵ, ω_1, and ω_3, the only contribution comes from the derivative with respect to ϵ, which depends on λ according to (5.8), i.e.,

$$\left(\frac{\partial}{\partial\lambda} - \frac{\partial}{\partial x}\right) T_1T_3\big|_{x=\lambda,\varphi''=\gamma''=0} = \frac{\partial}{\partial\lambda}\frac{\partial}{\partial\epsilon} f(\epsilon, \omega_1, \omega_3, \lambda, x) f(\epsilon, \omega_3, \omega_1, \lambda, x) = 2(\omega_1\omega_3)^{\frac{1}{2}} \frac{\partial}{\partial\epsilon} T_1^F(\epsilon) T_3^F(\epsilon), \tag{5.9}$$

where ϵ is set equal to its on-shell value (5.8). $T_{1,3}^F(\epsilon)$ are the forward 2-body T-matrix elements given by

$$T_{1,3}^F(\epsilon) = \langle \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3 | \, T_{1,3}(\epsilon) \, | \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3 \rangle = f(\epsilon, \omega_{1,3}, \omega_{3,1}, \lambda, \lambda). \tag{5.10}$$

We have thus expressed the derivatives with respect to the angle variables x and λ in (4.27) in terms of the derivative with respect to the (off-shell) energy variable ϵ.

Substituting (5.9) in (4.27), we have

$$b_3^{\text{sing}} = 3^{\frac{3}{2}}(2\pi)^{-6}\sum_{J,\mu}(2J+1)Q_{J\mu}^S = 3^{\frac{3}{2}}(2\pi)^{-6}(-\beta)^{\frac{1}{2}} \operatorname{Re}\int d\omega_1 \, d\omega_3 e^{-2\beta(\omega_1+\omega_3)} \times 4\pi^2 \int_{-1}^{1} d\lambda \exp\left[-2\beta\lambda(\omega_1\omega_3)^{\frac{1}{2}}\right] 2(\omega_1\omega_3)^{\frac{1}{2}} \times \left(-\beta T_1^F(\epsilon) T_3^F(\epsilon) + \frac{\partial}{\partial\epsilon} T_1^F(\epsilon) T_3^F(\epsilon)\right), \tag{5.11}$$

with $\epsilon = 2[\omega_1 + \omega_3 + (\omega_1\omega_3)^{\frac{1}{2}}\lambda]$.

C. Singular Terms in the Momentum Representation

While the 3-body forward scattering amplitudes do not exist owing to the singular terms, the off-shell forward amplitudes are well defined. We proceed to show that the singular terms can be evaluated in the momentum representation provided one does not set ϵ in the amplitude to its on-shell value.

We can rewrite (4.12) as

$$\frac{1}{2\pi}\operatorname{Im}(\operatorname{Tr}\ln S)_{\text{sing}} = -\sum_{i\neq j}\frac{1}{2\pi}\operatorname{Re}\operatorname{Tr}\tfrac{1}{2}i(G_0 - G_0^\dagger)T_i(G_0 + G_0^\dagger)T_j. \tag{5.12}$$

The imaginary part $i\eta$, going along with the energy variable ϵ, must be kept finite until all the G_0's are combined to give a function which is well defined in the limit $\eta \to 0$.

The matrix elements of interest are

$$\begin{aligned}
&\langle \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3 | \, (G_0 - G_0^\dagger)T_1(G_0 + G_0^\dagger)T_3 \, | \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3 \rangle \\
&= [(\epsilon - \omega_1 - \omega_2 - \omega_3 + i\eta)^{-2} \\
&\quad - (\epsilon - \omega_1 - \omega_2 - \omega_3 - i\eta)^{-2}] T_1^F(\epsilon) T_3^F(\epsilon) \\
&= \left(-2i\frac{\partial}{\partial\epsilon}\operatorname{Im}(\epsilon - \omega_1 - \omega_2 - \omega_3 + i\eta)^{-1}\right) T_1^F(\epsilon) T_3^F(\epsilon) \\
&= 2\pi i\delta'(\epsilon - \omega_1 - \omega_2 - \omega_3) T_1^F(\epsilon) T_3^F(\epsilon).
\end{aligned} \tag{5.13}$$

We have taken the limit $\eta \to 0$ in the last step. This limit is well defined. The forward 2-body amplitudes are well defined and assumed to be smooth

functions of ϵ. We thus have

$$b_3^{\text{sing}} = 3^{\frac{3}{2}}(2\pi)^{-6}\,\text{Re}\,\beta\int d\epsilon e^{-\beta\epsilon}\int d^3p_1\,d^3p_3 \times \tfrac{1}{2}\delta'(\epsilon-\omega_1-\omega_2-\omega_3)T_1^F(\epsilon)T_3^F(\epsilon)$$
$$= 3^{\frac{3}{2}}(2\pi)^{-6}(-\beta)\,\text{Re}\int d^3p_1\,d^3p_3\tfrac{1}{2}e^{-\beta(\omega_1+\omega_2+\omega_3)} \times \Big(-\beta T_1^F(\epsilon)T_3^F(\epsilon) + \frac{\partial}{\partial\epsilon}T_1^F(\epsilon)T_3^F(\epsilon)\Big)_{\epsilon=\omega_1+\omega_2+\omega_3}. \quad (5.14)$$

It takes a trivial change of variables to verify that (5.14) is the same as (5.11).

What we have just demonstrated is the equivalence between using the angular-momentum representation, where only on-shell 3-body T-matrix elements appear, and using the momentum representation, where the off-shell matrix elements are used.

6. CONCLUSIONS

We have discussed the singular terms in b_3 in detail and have shown that their contribution to b_3 is well defined. The nonsingular terms are expected to be well defined in the angular-momentum representation as well as in the momentum representation. Our conclusion is that, once the scattering matrix elements, including both singular and nonsingular terms, are given, b_3 can be calculated by the straightforward application of our formula. We feel that the same conclusion should also hold for b_N, $N > 3$.

ACKNOWLEDGMENT

We are grateful to Professor J. M. Luttinger for his helpful comments. Part of this work was done in the summer of 1969, while one of us (S. M.) was at the Environmental Science Services Administration, Boulder, Colorado. The hospitality of Dr. R. H. Williams is very gratefully acknowledged.

* This research supported in part by the U.S. Atomic Energy Commission, under Contract AT(11-1)-GEN-10, P.A. 11: Report No. UCSD-10P11-98.

† Alfred P. Sloan Foundation Fellow.

[1] R. Dashen, S. Ma, and H. J. Bernstein, Phys. Rev., to be published, hereafter referred to as I.

[2] K. Huang, *Statistical Mechanics* (J. Wiley & Sons, Inc., New York, 1963).

[3] An infinite forward-scattering amplitude is often a result of a long-range interaction. For example, the Coulomb scattering has an infinite forward amplitude. For the double scattering shown in Fig. 1, the particle 2 in the intermediate state may be regarded as a virtual particle exchanged between the two blobs analogous to the exchanged virtual photon in the Coulomb scattering. For forward scattering, the virtual particle approaches its energy shell and becomes a real particle, which can propagate over a long distance and thus provides an effective long-range interaction. This is analogous to the fact that the exchanged virtual photon becomes a real photon for the forward Coulomb scattering. The long-range effect of the double-scattering term has been demonstrated by Iagolnitzer [J. Math. Phys. 6, 1576 (1965)] using wavepackets to describe the incoming particles. It was shown that when particle 3 passes by particles 1 and 2 at a large distance away, the dominant contribution to the collision (involving all three particles) comes from the double scattering term.

[4] R. L. Omnes, Phys. Rev. 134, B1358 (1964).

[5] This is a consequence of the invariance under the product of parity and time reversal.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Generalized Algebraic Mass Formulas. I

E. D. DICKENS, JR.,* AND A. D. LEVINE
Department of Physics, West Virginia University, Morgantown, West Virginia 26506

(Received 10 July 1969)

It is shown that the $SL(2, C)$-Poincaré associative algebraic model of Böhm can be extended without essential difficulty to an $SL(2, C)$-de Sitter model which gives rise to a meson mass formula with both spin and isospin dependence.

INTRODUCTION

To overcome the well-known difficulties[1–6] of purely group-theoretical methods for obtaining mass formulas, Böhm[7–9] considered models based on associative algebras employing the enveloping algebras of the Poincaré group and noncompact internal symmetry groups such as $SL(2, C)$ or $SL(3, C)$. The defining algebras for these models contained non-Lie algebraic operators and employed a coupling first stated by Werle.[10] The meson mass formulas obtained were

$$SL(2, C): \quad m^2 = m_0^2 - bI(I+1), \tag{1}$$

$$SL(3, C): \quad m^2 = z + g[\tfrac{1}{4}Y^2 - I(I+1) + \lambda^2 + 2\lambda], \tag{2}$$

where m_0 and z are constants, b and g are coupling constants (introduced in the Werle coupling), and I, Y, and λ are eigenvalues of the $SU(2)[SL(2, C)]$ or $SU(3)[SL(3, C)]$ Casimir operators.

These associative algebraic structures have been the basis for several very interesting calculations (by Böhm and Sudarshan[11]) of the 2-, 3-, and 4-body leptonic decays of mesons. Of particular interest was the calculation of Cabibbo angles which turned out to be in very good agreement with experiment.

In an earlier work,[12] however, Böhm exhibited a dynamical group model based on the enveloping algebras of the Poincaré and de Sitter groups in which a mass–spin relation of the form[13]

$$m^2 = \lambda^2\alpha^2 - \tfrac{9}{4}\lambda^2 + \lambda^2 S(S+1) \tag{3}$$

was obtained ($\lambda = R^{-1}$, where R = radius of the de Sitter space and $\lambda^2\alpha^2$ is the eigenvalue of one of the de Sitter invariants).

It is the purpose of this note to show that the unification of these two different models[13] (a dynamical group and algebra) into a single algebraic model can be easily obtained, and, therefore, we can obtain a generalized mass formula for mesons which fits the existing data very well.[14] In this paper, however, we will treat the mathematically simple, but physically uninteresting, case of spin and isospin and will retain the more complex $SU(3)$ model (and the data fitting) for a forthcoming paper.[14]

THE COMBINED MODEL

The combined-model algebra D is generated by the enveloping algebras of the de Sitter, Poincaré, and $SL(2, C)$ groups. We have then the Poincaré–de Sitter algebra[15]

$$[P_\mu, P_\nu] = 0, \quad M^2 = P_\mu P^\mu > 0, \quad \mu, \nu = 0, 1, 2, 3, \tag{4}$$

$$[P_\rho, L_{\mu\nu}] = i(g_{\mu\rho}P_\nu - g_{\nu\rho}P_\mu), \tag{5}$$

$$[L_{\mu\nu}, L_{\rho\sigma}] = i(g_{\mu\rho}L_{\nu\sigma} + g_{\nu\sigma}L_{\mu\rho} - g_{\mu\sigma}L_{\nu\rho} - g_{\nu\rho}L_{\mu\sigma}), \tag{6}$$

$$[B_\mu, B_\nu] = i\lambda^2 L_{\mu\nu}, \tag{7}$$

$$[L_{\mu\nu}, B_\rho] = i(g_{\nu\rho}B_\mu - g_{\mu\rho}B_\nu), \tag{8}$$

where

$$B_\mu = P_\mu + \tfrac{1}{2}\lambda(P_\mu P^\mu)^{-\frac{1}{2}}\{P^\rho, L_{\rho\mu}\} \tag{9}$$

such that

$$[B_\mu, P_\nu] = i\lambda(P_\rho P^\rho)^{-\frac{1}{2}}(P_\rho P^\rho g_{\mu\nu} - P_\mu P_\nu) \tag{10}$$

and

$$B_\mu B^\mu = P_\mu P^\mu + \tfrac{9}{4}\lambda^2 + \lambda^2(P_\rho P^\rho)^{-1}P^\rho P^\sigma L_{\rho\mu}L_\sigma^\mu. \tag{11}$$

The $SL(2, C)$ algebra is then

$$[I_i, I_j] = i\epsilon_{ij}{}^k I_k, \qquad i, j, k = 1, 2, 3, \tag{12}$$

$$[I_i, F_j] = i\epsilon_{ij}{}^k F_k, \quad [F_i, F_j] = -i\epsilon_{ij}{}^k I_k. \tag{13}$$

Before stating how these algebras are coupled, it is worthwhile to backtrack a little and rewrite the Poincaré–de Sitter algebras. Let us explicitly introduce the operator

$$X_\mu \equiv M^{-1}P_\mu = P_\mu M^{-1}. \tag{14}$$

Then, since

$$[M^2, L_{\mu\nu}] = \{M, [M, L_{\mu\nu}]\} = 0 \Rightarrow [M, L_{\mu\nu}] = 0, \tag{15}$$

we have

$$[X_\mu, X_\nu] = 0, \tag{16}$$

$$[X_\rho, L_{\mu\nu}] = i(g_{\mu\rho}X_\nu - g_{\nu\rho}X_\mu), \tag{17}$$

$$X_\mu X^\mu = 1, \tag{18}$$

$$B_\mu = P_\mu + \tfrac{1}{2}\lambda\{X^\rho, L_{\rho\mu}\}, \tag{19}$$

$$[B_\mu, X_\nu] = i\lambda(X_\rho X^\rho g_{\mu\nu} - X_\mu X_\nu), \tag{20}$$

$$B_\mu B^\mu = P_\mu P^\mu + \tfrac{9}{4}\lambda^2 + \lambda^2 X^\rho X^\sigma L_{\rho\mu}L^\mu_\sigma, \tag{21}$$

and the others as listed. The reason for this step is that the coupling between the algebras now takes the simple form

$$[L_{\mu\nu}, I_i] = [L_{\mu\nu}, F_i] = 0, \tag{22}$$

$$[X_\mu, I_i] = [X_\mu, F_i] = 0, \tag{23}$$

$$[B_\mu B^\mu, I_i] = [M^2, I_i] = 0, \tag{24}$$

i.e.,

$$[B_\mu, I_i] = [P_\mu, I_i] = 0 \tag{25}$$

and

$$[B_\mu B^\mu, F_i] = [M^2, F_i] = ib\epsilon_i^{jk}\{I_j, F_k\}. \tag{26}$$

As we will show, these commutators will be sufficient to characterize the algebra D.

INVARIANT OPERATORS OF THE ALGEBRA

It is obvious from Eqs. (22)–(26) and Ref. 7 that the invariant operators of $SL(2, C)$, i.e.,

$$R_1 = \mathbf{F}^2 - \mathbf{I}^2, \quad R_2 = \mathbf{I}\cdot\mathbf{F}, \tag{27}$$

are invariant operators of D.

The invariants operators of the de Sitter group, however, are not invariant operators of the algebra because of Eq. (26). We have two invariant operators[12]

$$\lambda^2 Q = B_\mu B^\mu - \tfrac{1}{2}\lambda^2 L_{\mu\nu}L^{\mu\nu} = \lambda^2\alpha^2 \tag{28}$$

and

$$W = \tfrac{1}{4}\lambda^2(\mathbf{M}\cdot\mathbf{N})^2 - V_\mu V^\mu = \beta, \tag{29}$$

$$V_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}B^\nu L^{\rho\sigma}, \tag{30}$$

$$\mathbf{N} = (L_{01}, L_{02}, L_{03}), \quad \mathbf{M} = (L_{23}, L_{31}, L_{12}). \tag{31}$$

From Eqs. (21)–(23), we see immediately that

$$[\lambda^2 Q, I_i] = 0. \tag{32}$$

But

$$[\lambda^2 Q, F_i] = [B_\mu B^\mu, F_i] = [M^2, F_i] = ib\epsilon_i^{jk}\{I_j, F_k\}. \tag{33}$$

Now, we know that

$$[I^2, F_i] = \{I^k, [I_k, F_i]\} = -i\epsilon_{ikl}\{I^k, F^l\}, \tag{34}$$

so we see that the operator

$$A = \lambda^2 Q + bI^2 \tag{35}$$

will commute with all of the operators of D.

We note that W is, in fact, equivalent to $-\omega_\mu\omega^\mu$, where $\omega_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}P^\nu L^{\rho\sigma}$ and is λ independent,[16] but is not, however, an invariant of the algebra D. Instead, we have as the invariant operator the usual Poincaré invariant

$$\Gamma = \Gamma_\mu\Gamma^\mu = M^{-2}\omega_\mu\omega^\mu, \tag{36}$$

where

$$\Gamma_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}X^\nu L^{\rho\sigma}. \tag{37}$$

This can be seen immediately from Eqs. (22) and (23) and from the fact that it is an invariant operator of the Poincaré group. Then,

$$[L_{\mu\nu}, \Gamma] = [P_\mu, \Gamma] = 0, \tag{38}$$

such that

$$[B_\mu, \Gamma] = 0 \tag{39}$$

and

$$[\lambda^2 Q, \Gamma] = 0. \tag{40}$$

Now, by a straightforward calculation, we obtain

$$\Gamma = -X_\beta X_\rho L^{\rho\sigma}L^\beta_\sigma + \tfrac{1}{2}L^{\beta\sigma}L_{\beta\sigma}, \tag{41}$$

but

$$\begin{aligned}\lambda^2 Q &= B_\mu B^\mu - \tfrac{1}{2}\lambda^2 L^{\mu\nu}L_{\mu\nu} \\ &= P_\mu P^\mu + \tfrac{9}{4}\lambda^2 + \lambda^2 X_\rho X_\sigma L^{\sigma\gamma}L^\rho_\gamma - \tfrac{1}{2}\lambda^2 L^{\mu\nu}L_{\mu\nu}\end{aligned} \tag{42}$$

or

$$\lambda^2 Q = P_\mu P^\mu + \tfrac{9}{4}\lambda^2 - \lambda^2\Gamma = \lambda^2\alpha^2. \tag{43}$$

We note that if a dynamical group approach is employed completely, as was done in Ref. 12 (where now $\lambda^2 Q$ is a de Sitter invariant operator), we have Γ invariant and, therefore, we require that $P_\mu P^\mu$ also be invariant, which, consequently, gives no mass splitting (as was incorrectly implied in Ref. 12). In this case, however, $\lambda^2 Q$ is not an invariant operator of D and the mass splitting can be nontrivial.

The maximal commuting system is then

$$A, \Gamma, R_1 = \mathbf{F}^2 - \mathbf{I}^2, \quad R_2 = \mathbf{I}\cdot\mathbf{F}, \mathbf{I}^2, I_3, S_3, P_i, \tag{44}$$

where the first four are invariant operators which label the states, S_3 is the third component of the spin, and we add P_i by convention.

To find the appropriate representation space, we first consider the now classical contraction relation between the Poincaré and de Sitter groups.[17–19]

The contraction is performed in the limit $\lambda \to 0$ and, to obtain the correct representation of the Poincaré group, we require that, in the limit $\lambda \to 0$,

$$\lambda^2\alpha^2 \to m_0^2 \tag{45}$$

since, in the limit $\lambda \to 0$,

$$\lambda^2 Q \to B_\mu B^\mu = P_\mu P^\mu. \tag{46}$$

Since $\omega_\mu \omega^\mu$ is independent of λ, we still have the spin-invariant $\Gamma = S(S+1)$.

However, in the limit $b \to 0$, the algebra D goes into the product of the enveloping algebras of $SL(2, C)$ and the de Sitter group, so that we have the diagram of Fig. 1. The representations of the algebra D are obtained by finding the complete set of commuting observables and constructing the appropriate rigged Hilbert space,[21,22] as was done in the earlier work.[7]

Since, in the limit $b \to 0$, the algebra goes into $\epsilon(SL(2, C)) \times \epsilon(\text{de Sitter})$, we conjecture that the algebra D can be represented in a subspace of the rigged Hilbert space (RHS)

$$\phi_1 \hat{\otimes} \phi_2 \subset H_1 \overline{\otimes} H_2 \subset (\phi_1 \hat{\otimes} \phi_2)^x,$$

where $\phi_1 \subset H_1 \subset \phi_1^x$ is the RHS of $\epsilon(SL(2, C))$ labeled by eigenvalues (c, k_0) and $\phi_2 \subset H_2 \subset \phi_2^x$ is the RHS of $\epsilon(SO(4, 11))$[23] labeled by eigenvalues $(\alpha, \sigma = s)$. The notation (used above) and proofs that D is contained in the above RHS follow that given in Ref. 7 and need no further comment.

A generalized eigenvector in this space can be labeled by the eigenvalues

$$\mathbf{F}^2 - \mathbf{I}^2 = (1 + c^2 - k_0^2)\mathbf{1}, \quad \mathbf{F} \cdot \mathbf{I} = k_0 c \mathbf{1},$$
$$\Gamma = S(S+1)\mathbf{1}, \quad A = a\mathbf{1}. \tag{47}$$

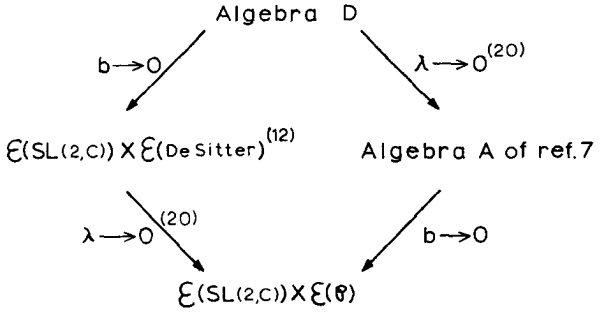


FIG. 1. Contraction scheme for algebra D. (12) and (20) refer to Refs. 12 and 20, respectively.

Then, taking the expectation value of A in the states $D_{a,\lambda(s,\alpha),(c,k_0)}$ gives

$$m^2 = a - \tfrac{9}{4}\lambda^2 + \lambda^2 S(S+1) - bI(I+1), \tag{48}$$

where λ is a universal constant, a is the parameter telling us which representation of the de Sitter group we are in, and b is fitted by experiment.

Although the mass formula derived above is of limited utility, the corresponding formula for $SL(3, C)$ case provides a unified mass formula for all mesons which is in excellent agreement with experiment.[14]

* Supported by NASA Traineeship, 1967–70.

[1] L. O'Raifeartaigh, Phys. Rev. Letters **14**, 332, 575 (1965).
[2] I. Segal, J. Functional Anal. **1**, 1 (1967).
[3] I. T. Grodsky and R. F. Streater, Phys. Rev. Letters **20**, 695 (1968).
[4] V. D. Kyakhovsky, Commun. Math. Phys. **11**, 131 (1968).
[5] G. C. Hegerfeldt and J. Hennig, Fortsch. Physik **16**, 491 (1968).
[6] V. Berzi, Commun. Math. Phys. **10**, 217 (1968).
[7] A. Böhm, J. Math. Phys. **8**, 1551 (1967).
[8] N. Mukunda, E. C. G. Sudarshan, and A. Böhm, Phys. Letters **24B**, 301 (1967).
[9] A. Böhm, Phys. Rev. **158**, 1408 (1967).
[10] J. Werle, ICTP, Trieste, Preprint IC/65/48, 1965.
[11] A. Böhm and E. C. G. Sudarshan, Phys. Rev. **178**, 2264 (1969); Nuovo Cimento **60**, 395 (1969); Syracuse University, Preprint SU-1206-188, 1969.
[12] A. Böhm, Phys. Rev. **145**, 1212 (1966). See also *Lectures in Theoretical Physics*, W. E. Brittin, A. O. Barut, and M. Guenin, Eds. (Gordon & Breach, 1966), Vol. 9B, p. 327.
[13] A. Böhm, Phys. Rev. **175**, 1767 (1968). In this paper, it is shown that the same mass formula can also be obtained within the framework of an algebraic model without the difficulties discussed in the text [Eq. (43) above and the following discussion]. The more interesting case of coupling two algebraic models, i.e., that of Refs. 12 and 13, will be the subject of another paper and provides an interesting basis for formulating particle physics.
[14] E. D. Dickens, Jr. and A. D. Levine, West Virginia University, Morgantown, W.Va., Preprint WV-EP-3, January, 1970.
[15] This is meant in the sense of Ref. 12, i.e., the de Sitter group generated by Poincaré generators via the coupling λ.
[16] This follows from the identity

$$V_\mu V^\mu = \omega_\mu \omega^\mu, \text{as } \lambda \to 0.$$

[17] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. **39**, 510 (1953).
[18] M. Levy-Nahas, J. Math. Phys. **8**, 1211 (1967).
[19] A. Sankaranarayanan, Douglas Advan. Res. Lab., Preprint, April, 1969.
[20] Contraction carried out such that Eqs. (45) and (46) are obeyed.
[21] A. Böhm, in *Lectures in Theoretical Physics*, W. E. Brittin, A. O. Barut, and M. Guenin, Eds. (Gordon and Breach, 1966), Vol. 9A, p. 255.
[22] J.-P. Antoine, J. Math. Phys. **10**, 53, 2276 (1969).
[23] The intimate relationship between $SO(4, 1)$, as employed here, and $SO(3, 1)$, employed for the Poincaré group with respect to the contraction and the definition of the de Sitter algebra, is discussed in Ref. 12.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Equivalence of Some Generalizations on the Spherical Model*

G. V. BETTONEY† AND R. M. MAZO
Institute of Theoretical Science and Department of Chemistry, University of Oregon, Eugene, Oregon 97403

(Received 6 February 1967)

We define n-spherical models as follows: (1) Divide a lattice of N points into n mutually exclusive subsets with $N_1, \cdots, N_n$ points, respectively; $\sum N_\alpha = N$. With each lattice point j, associate a variable ϵ_j, and impose the restrictions $\sum \epsilon_j^2 = N_\alpha$, $\alpha = 1, \cdots, n$; the sums are over the sites of the αth subset. (2) Adopt the energy expression for the Ising model with the ϵ's playing the role of the "scalar spins." We then prove that, in the thermodynamic limit ($N \to \infty$, $N_\alpha \to \infty$ for all α), the thermodynamic functions of the n-spherical model are equal to those of the ordinary spherical model.

I. INTRODUCTION

The spherical model of Berlin and Kac[1] was introduced as a model of an interacting many-body system, whose statistical thermodynamic properties can be computed exactly. Although the spherical model is only a caricature of the Ising model, which itself is a caricature of a real ferromagnet, it has the advantage that its partition function may be studied analytically. Since phase transitions in a statistical mechanics are intimately bound up with the analyticity (or lack of it) of the partition function, it is of great interest to be able to discuss the analyticity properties of the partition function without the necessity of drawing conclusions on the basis of approximate calculations, a procedure which is known to be fraught with pitfalls. For this reason, the spherical model has been studied by many authors.[2]

The energy expression for the spherical model is the same as that for the Ising model, namely,

$$E\{\epsilon\} = -J \sum_{\langle i,j \rangle} \epsilon_i \epsilon_j - \mu H \sum \epsilon_j . \tag{1}$$

ϵ_j is a scalar "spin" attached to the jth lattice site; the first sum is over nearest-neighbor pairs and the second is over all sites. J is the "exchange integral," positive for ferromagnetism, negative for antiferromagnetism, and μ is the magnetic moment of a "spin." The spherical model replaces the N constraints $\epsilon_j = \pm 1$, $j = 1, \cdots, N$, of the Ising model by a single constraint

$$\sum \epsilon_j^2 = N, \tag{2}$$

where N is the number of sites on the lattice. We must also mention the mean spherical model of Lewis and Wannier,[3] where the constraint is

$$\langle \sum \epsilon_j^2 \rangle = N, \tag{3}$$

the brackets meaning ensemble average.

In an attempt to suppress partially the fluctuations inherent in the spherical model, we examined what we called the bispherical model. The lattice was divided into two mutually interpenetrating sublattices, and two constraints were imposed, namely

$$\sum \epsilon_j^2 = \tfrac{1}{2}N, \tag{4}$$

on each sublattice separately. This procedure might be thought to be especially reasonable in the case of antiferromagnetism, where the two interpenetrating sublattices have physical significance. Much to our surprise, the thermodynamic properties of the bispherical model were identical to those of the spherical model. This led us to define a whole class of generalizations of the spherical model and to prove a theorem concerning them.

II. MAIN CONTENT

The class of models we wish to discuss is the following. Divide up the N lattice sites into n mutually exclusive groups of $N_1, \cdots, N_n$ sites, respectively. Retain the energy expression (1), but impose the n constraints

$$\sum \epsilon_j^2 = N_\alpha, \quad \alpha = 1, \cdots, n, \tag{5}$$

there being one such sum for each group of sites. We call this the "n-spherical model," or "n-model" for short. Clearly, the N-model is the Ising model, and the 1-model is the ordinary spherical model. We can also define mean-n-models by replacing the sums on the left-hand side of (5) by their mean values.

We now prove the following theorem:

Theorem: In the thermodynamic limit ($N \to \infty$, each $N_j \to \infty$), for any finite n, the thermodynamic functions of an n-spherical model are equal to those of the ordinary spherical model.

Proof: The proof begins by considering mean-n-models. The partition function of a mean-n-model is given by

$$\Xi(n) = \int \cdots \int d^N\epsilon \exp\left[-\beta E\{\epsilon\} - \sum_{\alpha=1}^{n} \gamma_\alpha \sum_{\alpha} \epsilon_j^2\right], \tag{6}$$

where the γ_α are determined by the n equations

$$\frac{\partial \ln \Xi(n)}{\partial \gamma_\alpha} = -N_\alpha, \quad \alpha = 1, \cdots, n. \tag{7}$$

The sums in the exponent in (6) are over the N_α sites corresponding to the αth constraint. The free energy $F(n)$ of the mean-n-model is given by

$$\bar{Q}(n) = \exp[-\beta F(n)] = \Xi(n) \exp(\sum \gamma_\alpha N_\alpha).$$

Now let us consider a mean-n-model and a mean-n-model with, say, $n > m$. Furthermore, let the constraints of the mean-m-model be expressible as sums of the constraints of the mean-n-model [e.g., as Eq. (2) is the sum of Eqs. (4) for the two sublattices]. Now the partition function for these models can be thought of as arising from maximizing the entropy expression subject to the given constraints, or alternatively, minimizing the free energy. Hence we have $\bar{Q}(n) \leq \bar{Q}(m)$ for this case, since the constraints of the mean-m-model are subsumed in those of the mean-n-model. Therefore, for any n, we have

$$\bar{Q}(N) \leq \bar{Q}(n) \leq \bar{Q}(1). \tag{8}$$

We now show that $\bar{Q}(N) = \bar{Q}(1)$, from which it follows that all mean spherical models have the same free energy. The proof of this is simple. Equations (7) for the mean-N-model become

$$\frac{\partial \ln \Xi(N)}{\partial \gamma_\alpha} = -1, \quad \alpha = 1, \cdots, N. \tag{9}$$

One set of roots of Eqs. (9) is $\gamma_1 = \gamma_2 = \cdots = \gamma_N = \gamma$, by symmetry. This set of roots gives a $\bar{Q}(N)$ which equals $\bar{Q}(1)$, where there is only one γ to start with. But, should any other set of roots exist, they must yield a $\bar{Q}(N) \leq \bar{Q}(1)$ and, hence, do not yield the minimum of the free energy.

Thus all mean spherical models yield the same free energy. We must now extend this result to spherical models proper, where the constraints are taken in the absolute, rather than mean, sense. The partition functions for n-spherical models are of the form

$$\begin{aligned} Q(n) &= \int \cdots \int d^N\epsilon \exp[-\beta E\{\epsilon\}] \prod_{\alpha=1}^{n} \delta(N_\alpha - \sum \epsilon_j^2) \\ &= (2\pi i)^{-n} \int \cdots \int d^N\epsilon \exp[-\beta E\{\epsilon\}] \\ &\quad \times \int_{a-i\infty}^{a+i\infty} \cdots \int d^n\gamma \exp\left[\sum_\alpha \gamma_\alpha (N_\alpha - \sum \epsilon_j^2)\right], \end{aligned} \tag{10}$$

where we have introduced integral representations of the delta functions in the usual way. We may write, then,

$$Q(n) = (2\pi i)^{-n} \int_{a-i\infty}^{a+i\infty} \cdots \int d^n\gamma \Xi(n) \exp\left[\sum_\alpha \gamma_\alpha N_\alpha\right]. \tag{11}$$

If all of the N_α are large, the integral in (11) may be evaluated by the method of steepest descents and

$$Q(n) = \Xi(n) \exp\left[\sum_\alpha \gamma_\alpha N_\alpha\right], \tag{12}$$

where the γ's are determined by the saddle-point conditions

$$\frac{\partial}{\partial \gamma_\alpha}\left(\sum_\beta \gamma_\beta N_\beta + \ln \Xi(n)\right) = 0, \quad \alpha = 1, \cdots, n. \tag{13}$$

But these are just the conditions given by Eqs. (7). Thus, when all the N_α are large, our conclusion about mean-n-models carries over to n-models.

Note, however, that we cannot say that n-spherical models are equivalent to ordinary spherical model for all n. This is because the steepest-descent argument used to establish the equivalence of spherical and mean spherical models breaks down unless all $N_\alpha \to \infty$. Such a conclusion would have been patently wrong anyway, because the N-spherical model is precisely the Ising model, which is known not to be equivalent to the spherical model.

III. DISCUSSION

There are two main results of this paper. The first is that all mean-n-models are equivalent to the ordinary spherical model (1-model). The second is that n-models, for any finite n (as $N \to \infty$), are equivalent to the 1-model. It might seem, at first, as though we are merely emphasizing a special case of a general experience in statistical thermodynamics, i.e., that adding additional constraints consistent with given constraints does not change the thermodynamic properties of the system under consideration. But this is not altogether correct, for when one introduces constraints reducing fluctuations in spatially distinct parts of a system, one also introduces a boundary energy. To be sure, this is usually negligible compared to the bulk energy in the thermodynamic limit, because the surface to volume ratio is small; that is why the thermodynamic properties are unchanged.

However, in our case, the sublattices can be intimately intermixed in space, so much so that all of the exchange energy is boundary energy, in a sense. Consider, for example, nearest-neighbor interactions

in the lattice $1\,2\,1\,2\,1\,2\cdots$ where 1 and 2 denote the sites of two sublattices (1-dimensional for typographical convenience only). In this case it is not at all clear that the usual arguments hold, although our proof shows that the usual result is valid.

Another consideration which indicates that our proof is actually required is the following. In the antiferromagnetic case, $J < 0$, when the magnetic field energy and exchange energy are of comparable strength, then one would expect one sublattice to contain a preponderance of the $\sum \epsilon_i^2$, when only one spherical constraint is imposed. In the intimate mixing case, a 2-model prevents this from happening, and the 1-model can have appreciably lower energies than the 2-model. Our result shows that this does not affect the thermodynamical properties. We conclude that our result actually tells us something about the spherical model and is not a special case of a general statistical mechanical dictum.

Our proof only holds for n finite. We conjecture that the theorem also holds if $n = o(N)$ as $N \to \infty$, based on the following ideas. In order to obtain the maximum effect from the multiple spherical constraints, all sublattices should have the same order of magnitude of number of sites, N_α. Otherwise, some subset of the sublattices will be asymptotically dominant, and one need only consider those. This means that $N_\alpha \to \infty$ as $N \to \infty$, and so a steepest-descent calculation on the n variables γ_α should go through. To make this argument rigorous, one must prove the first statement and also investigate the uniformity of the convergence of the complex integrals (11) as $n \to \infty$. We have not accomplished this.

After writing this paper, we learned that Montroll had introduced n-spherical models in 1949[4]; he did not investigate their properties. In view of our conjecture above, we feel that his suggestion that $n = N^{\frac{1}{2}}$ will produce different behavior is probably incorrect.

Finally, it should be pointed out that we have been using the term "thermodynamic properties" to mean properties derivable directly from the partition function, i.e., the free energy and its derivatives. It is possible that other quantities, such as long-range order (below the transition point), may vary with the model within the class of models we have considered.

* Supported in part by National Science Foundation Grant GP-6002.

† National Science Foundation Cooperative Fellow.

[1] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).

[2] There is a large literature on the subject. We mention here only a relatively small number of representative papers. T. H. Berlin and J. Thomsen, J. Chem. Phys. **20**, 1368 (1952); N. Uryu, Mem. Fac. Sci., Kyushu Univ., Ser B. **1**, 131 (1954); M. Lax, J. Chem. Phys. **20**, 1351 (1952), Phys. Rev. **97**, 1419 (1955); R. M. Mazo, J. Chem. Phys. **39**, 2196 (1963); H. A. Gersch, Phys. Fluids **6**, 599 (1963).

[3] H. W. Lewis and G. H. Wannier, Phys. Rev. **88**, 682 (1952).

[4] E. Montroll, Nuovo Cimento, Suppl., **6**, 1 (1949).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Integral-Transform Gaussian Functions for Heliumlike Systems

DAVID M. BISHOP
Department of Chemistry, University of Ottawa, Ottawa 2, Canada
AND
R. L. SOMORJAI
Division of Chemistry, National Research Council of Canada, Ottawa 2, Canada

(Received 23 June 1969)

The one-dimensional Laplace transform of the Gaussian $\exp(-r^2x)$,

$$g(r) = \mathfrak{L}\{G(x)\} = \int_0^\infty e^{-r^2x}G(x)\,dx,$$

was used to generate functions which could be useful as basis sets for atomic and molecular calculations. A particular choice of the weighting function $G(x)$ led to functions of the form $g(r) = (qr)^\nu K_\nu(qr)$, where $K_\nu(qr)$ are modified Bessel functions of the second kind. These functions were used as basis functions for the helium isoelectronic series and accounted for 98.98% (H^-), 99.89% (He), 99.96% (Li^+), 99.98% (Be^{++}), 99.996% (O^{6+}) of the Hartree–Fock energy.

I. INTRODUCTION

Recently a new class of approximate wavefunctions constructed from integral transforms has been described.[1] This general class was based on the one-dimensional Laplace transform

$$f(r) = \mathfrak{L}\{F(x)\} = \int_0^\infty e^{-rx}F(x)\,dx. \tag{1}$$

It can be shown that, for certain choices of $F(x)$, the $f(r)$ will be the same as some well-known approximate wavefunctions. For example, if $F(x)$ is a delta function, we obtain the simple 1s Slater orbital e^{-r}; if

$$\begin{aligned} F(x) &= 0, \quad 0 < x < \alpha, \\ &= 1, \quad \alpha \lesssim x \lesssim \beta, \\ &= 0, \quad x > \beta, \end{aligned}$$

we obtain the Hulthén function[2] $[\exp(-\alpha r) - \exp(-\beta r)]/r$, which is a linear combination of an infinite number of screened 1s orbitals with orbital exponents ranging continuously in the interval α to β, each orbital having the same weight in the combination. The virtue of this general class of functions [Eq. (1)] is that $F(x)$ may be chosen, in a variational manner, to appropriately weight the "orbital exponent space." Another advantage to this treatment is that extensive compilations[3] of Laplace transform pairs facilitate the generation of new types of basis functions, and that these can be compared and related to each other much more readily in terms of their inverse transforms $[F(x)]$. Furthermore, a number of useful theorems concerning Laplace transforms can be used. A computational advantage is that, in evaluation of matrix elements, integration over physical coordinates may be performed first, followed by integration over the parameter space. The authors have found that this is often a great deal easier than first integrating over the parameter space.

In a previous publication,[4] one of us has used Eq. (1) with $F(x) = x^{\mu-1}\exp(-Ax)/\Gamma(\mu)$ to generate the basis functions

$$f(r) = (r + A)^{-\mu},$$

and successfully applied these functions to the helium isoelectronic series. In this paper, we have applied the method to the function $\exp(-r^2x)$,

$$g(r) = \mathfrak{L}\{G(x)\} = \int_0^\infty e^{-r^2x}G(x)\,dx, \tag{2}$$

which is the Gaussian analog to Eq. (1). Currently, there are two commonly used basis functions: Slater and Gaussian type; as the transform procedure for the former has been developed, it was felt to be equally important to do so for the latter. The ubiquity of Gaussian wavefunctions is well known, mainly because integrals involving them are relatively easy to evaluate.

Having chosen functions of the form of $g(r)$, we next have to choose the weighting function $G(x)$. Criteria for doing this have been explored previously.[4] In the present case, we can see that, if $\exp(-r^2x)$ were the exact eigenfunction of the wave equation for some particular value of x, then in this limit $G(x)$ should be the delta function. However, in most cases $\exp(-r^2x)$ will not be an exact eigenfunction and we expect $G(x)$ to be a shape function. Further, it would be useful if $G(x)$ was such that $\lim g(r)$, as $r \to \infty$, was of exponential form. These criteria are met by choosing

$$G(x) = \exp(-q^2/4x)/x^{\nu+1}. \tag{3}$$

With this choice of weighting function, we obtain [see Eq. (55) of Ref. 5]

$$g(r) = 2^{\nu+1}k_\nu(qr)/q^{2\nu}. \tag{4}$$

The function $k_\nu(qr)$ is proportional to

$$r^{2\nu}\exp(-qr)\int_0^\infty \exp(-2qrt)[t(t+1)]^{\nu-\frac{1}{2}}\,dt;$$

if $\nu = \frac{1}{2}$, it is $\exp(-qr)$. Without loss of generality, we may therefore take as basis functions

$$\psi(r) = k_\nu(qr). \tag{5}$$

We will call $k_\nu(qr)$ the reduced modified Bessel functions of the second kind which are related to the normal modified Bessel functions of the second kind [$K_\nu(z)$; see Eq. (9.6.2) of Ref. 6] by the equality

$$k_\nu(z) = z^\nu K_\nu(z). \tag{6}$$

An additional advantage to the above choice of $G(x)$ is that there are a number of useful relations concerning the $k_\nu(z)$; some of these are given in Appendix A.

In this paper we present the results of using such a basis function to approximate the wavefunctions of a series of two-electron atomic systems. Each orbital is described by a function $\psi(r)$ and, for each system, ν and q are chosen such as to minimize the electronic energy. We are, therefore, essentially using two-parameter functions.

II. METHOD AND RESULTS

The Hamiltonian for a helium-type system is

$$\mathcal{H} = -\tfrac{1}{2}(\nabla_1^2 + \nabla_2^2) - \frac{z}{r_1} - \frac{z}{r_2} + \frac{1}{r_{12}}, \tag{7}$$

where z is the nuclear charge and r_1, r_2, and r_{12} are the nucleus–electron-1, nucleus–electron-2, and electron-1–electron-2 distances, respectively. If the two-electron wavefunction is given by

$$\Psi = \psi(r_1)\psi(r_2), \tag{8}$$

then the electronic energy E is

$$E = \frac{\iint \Psi^*\mathcal{H}\Psi\,d\tau_1\,d\tau_2}{\iint \Psi^*\Psi\,d\tau_1\,d\tau_2} = \frac{2S(T+V_n)+V_e}{S^2}, \tag{9}$$

where

$$S = \int \psi^*(r)\psi(r)\,d\tau, \tag{10}$$

$$T = \int \psi^*(r)(-\tfrac{1}{2}\nabla^2)\psi(r)\,d\tau, \tag{11}$$

$$V_n = \int \psi^*(r)\left(-\frac{z}{r}\right)\psi(r)\,d\tau, \tag{12}$$

and

$$V_e = \iint \psi^*(r_1)\psi(r_1)\left(\frac{1}{r_{12}}\right)\psi^*(r_2)\psi(r_2)\,d\tau_1\,d\tau_2. \tag{13}$$

Using Eqs. (5) and (6) and the equality[7]

$$\int_0^\infty x^{s-1}K_\nu^2(ax)\,dx = 2^{s-3}a^{-s}[\Gamma(\tfrac{1}{2}s)]^2 B(\tfrac{1}{2}s+\nu, \tfrac{1}{2}s-\nu),$$

$$\operatorname{Re} s > 2\,|\operatorname{Re}\nu|, \quad \operatorname{Re} a > 0, \tag{14}$$

where $B(x, y)$ is the β function, S, T, and V_n reduce to

$$S = 2^{2\nu+2}q^{-3}\pi[\Gamma(\nu+\tfrac{3}{2})]^2 B(2\nu+\tfrac{3}{2}, \tfrac{3}{2}), \tag{15}$$

$$T = 2^{2\nu+1}q^{-1}\pi[\Gamma(\nu+\tfrac{3}{2})]^2 B(2\nu+\tfrac{1}{2}, \tfrac{5}{2}), \tag{16}$$

$$V_n = -z2^{2\nu+1}q^{-2}\pi[\Gamma(\nu+1)]^2 B(2\nu+1, 1). \tag{17}$$

The two-electron integral V_e may be reduced (see Appendix B) to the form

$$\begin{aligned} V_e &= 16\pi^2 q^{-5}(2\nu+1)^{-1} \\ &\quad \times \int_0^\infty p^{4\nu+4}[K_\nu^2(p)K_{\nu+1}^2(p) - K_\nu^4(p)]\,dp \\ &= 16\pi^2 q^{-5}(2\nu+1)^{-1}I(\nu). \end{aligned} \tag{18}$$

An extensive search of the literature revealed no analytical form for the integral $I(\nu)$ if ν is unrestricted. $I(\nu)$ was therefore evaluated numerically for 40 different ν values equally spaced from $\nu = 0.025$ to 1.975 and the results used by the method of least squares to fit a 25-term linear combination of Chebyshev polynomials of the form

$$I(\nu) = \sum_{i=0}^{24} a_i T_i[(\nu - 1.0)/0.975]. \tag{19}$$

This series was then used to evaluate $I(\nu)$ for any ν value in the range 0 to 2.0. The details of this approach are given in Appendix B.

The electronic energy E is therefore finally a function of q, ν, and nuclear charge z. The values of q and ν which minimized E for $z = 1$, 2, 3, 4, and 8 were found by using Powell's minimization method[8] which does not use derivatives. All calculations were carried out in double precision on either an IBM 360/50 or 360/65 computer. In Table I we present the values of q, ν, and E which were obtained and also the exact Hartree–Fock results taken from Roothaan and Weiss.[9]

III. DISCUSSION

The energies presented in Table I show that integral transform Gaussian functions are capable of surprisingly high accuracy, considering the fact that they contain only *two* variable parameters. For the helium

TABLE I. Parameter values and energies for He and some He-like ions.

Atom or ion	z	ν	q	E^a	Hartree–Fock energy[b]	Optimized one-parameter Gaussian energy[c]
H^-	1	0.222364	0.551520	−0.48296	−0.48793	−0.35472
He	2	0.357521	1.521673	−2.85853	−2.86168	−2.30099
Li^+	3	0.404532	2.512546	−7.23366	−7.23641	−5.94491
Be^{++}	4	0.428263	3.508142	−13.60872	−13.61130	−11.28648
O^{6+}	8	0.464057	7.501736	−59.10880	−59.11114	−47.74648

[a] This work. [b] Ref. 9. [c] $-(1-8\frac{1}{2}z)^2/3\pi$.

series, they account for 98.98% (H^-), 99.89% (He), 99.96% (Li^+), 99.98% (Be^{++}), and 99.996% (O^{6+}) of the Hartree–Fock energy (which is the best result one can obtain with orbitals). Strictly, comparison should be made with normal Gaussian wavefunctions from which our basis functions are derived, and here the situation is even more dramatic—a two-term Gaussian wavefunction (having three variable parameters) for He produces only 95.99% of the Hartree–Fock energy,[10] and it is not until one reaches a five-term Gaussian wavefunction (nine variable parameters) that one produces an energy for He surpassing that of the integral transform Gaussian function. In Table I we also show the energies of the ions, using single optimized Gaussian orbitals; there do not appear to be, in the literature, any more complex Gaussian wavefunctions for these ions.

However, the calculation of near Hartree–Fock energies with integral transform functions is only one aspect of this work. The other is to investigate and set up criteria for the proper choice of weighting factors [$G(x)$ of Eq. (2)], and in Fig. 1 we show the optimized $G(x)$ of the form $\exp(-q^2/4x)/x^{\nu+1}$, normalized to $G(x)_{\max} = 1$, which were found in this work. It is pleasing to discover that, as suspected, they rise sharply and then fall off slowly and that, as the nuclear charge is increased, the higher exponent range is emphasized.

The sudden rise of $G(x)$ to its maximum value suggests that a weight function that is almost piecewise continuous rather than just continuous may be quite useful. First, a finite discontinuity in $G(x)$ necessarily implies that $g(r) = \exp(-\alpha r^2)f(r)$, where the discontinuity occurs at $x = \alpha$. Secondly, the final-value theorem of Laplace transforms[11] states that

$$\lim_{r\to 0} rg(r) = G(\infty),$$

and thus for our purposes $\lim G(x)$, as $x \to \infty$, should be finite. The simplest way to assure this is to make $G(x) \equiv 0$ for $x > \beta$, $\beta > \alpha$. [This also makes $g(r)$ an *entire* function.] Finally, the initial-value theorem states that

$$\lim_{r\to\infty} rg(r) = G(0+),$$

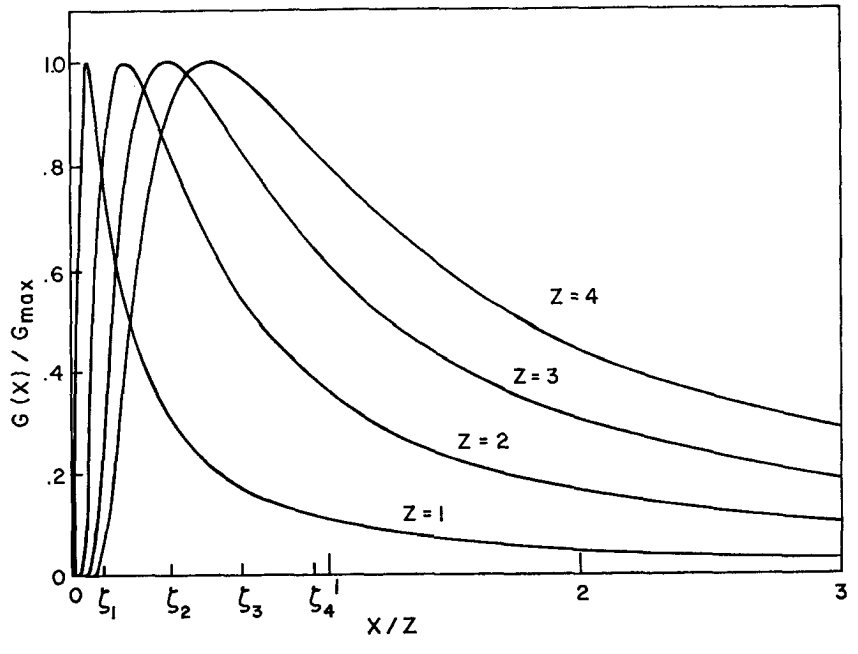


FIG. 1. Plots of normalized weight functions $G(x)/G_{\max}$ vs x/z. z is the nuclear charge.

$G_{\max} = G(x_{\max}) = [4(\nu+1)/q^2]^{\nu+1} \times \exp(-\nu-1)$,

$x_{\max} = q^2/4(\nu+1)$,

$\zeta_z = (1-8\frac{1}{2}z)^2/9\pi$ are the optimum orbital exponents of the one-parameter Gaussian orbitals $\exp(-\zeta_z r^2)$ for the He series.

so we know that $G(0)$ should vanish. In fact, the Hulthén function satisfies all three of these conditions, and this is perhaps one reason for its success. We have already found simple yet general weight functions that satisfy the above criteria:

$$\begin{aligned} G(x) &= 0, && 0 \lesssim x < \alpha, \\ &= (\beta - x)^N, && \alpha \lesssim x \lesssim \beta, \\ &= 0, && x > \beta, \end{aligned}$$

and calculations are in progress for $N = 0, 1, 2$. The results are promising and will be given in a later publication.

In Fig. 1 we have also indicated the position of the optimized exponent $[= (1 - 8\frac{1}{2}z)^2/9\pi]$ of a one-term Gaussian wavefunction for $z = 1, 2, 3$, and 4. The positions of these exponents clearly demonstrate that it is important to add in other Gaussian functions with exponents of smaller value. The energies for the one-term Gaussian wavefunctions are given in Table I.

Finally, it is worth mentioning that the reduced modified Bessel functions $[k_\nu(qr)]$, which in this work were the natural outcome of the transform method, can be considered in their own right as useful basis functions for atomic and molecular calculations. Attention was first drawn to this point as long ago as 1963 by Shavitt,[5] who saw them as a form of "fractional quantum number" orbital. The calculations presented above would seem to go some way towards justifying their usefulness.

ACKNOWLEDGMENT

One of us (D. M. B.) wishes to acknowledge the National Research Council of Canada for its financial support of this work.

APPENDIX A: USEFUL RELATIONS FOR $k_\nu(z)$

We have

$$k_\nu(z) = z^\nu K_\nu(z) = 2^{\nu-1}\int_0^\infty t^{\nu-1}\exp(-t - z^2/4t)\,dt,$$

$$\frac{dk_\nu(z)}{dz} = -zk_{\nu-1}(z),$$

$$k_{\nu+1}(z) = 2\nu k_\nu(z) + z^2 k_{\nu-1}(z),$$

$$k_{\frac{1}{2}}(z) = (\pi/2)^{\frac{1}{2}}\exp(-z),$$

$$k_{\frac{3}{2}}(z) = (\pi/2)^{\frac{1}{2}}(1 + z)\exp(-z),$$

$$k_{n+\frac{1}{2}}(z) = (\pi/2^{2n+1})^{\frac{1}{2}}\exp(-z) \times \sum_{i=0}^{n}(2n - i)!\,(2z)^i/i!\,(n - i)!.$$

APPENDIX B: EVALUATION OF V_e

The two-electron integral V_e is given by

$$V_e = \iint k_\nu^2(qr_1)(r_{12}^{-1})k_\nu^2(qr_2)\,d\tau_1\,d\tau_2.$$

Integration over all electronic coordinates except r_1 and r_2 reduces the above integral to

$$16\pi^2\int_0^\infty r_1 k_\nu^2(qr_1)\left[r_1\int_{r_1}^\infty r_2 k_\nu^2(qr_2)\,dr_2 + \int_0^{r_1} r_2^2 k_\nu^2(qr_2)\,dr_2\right]dr_1,$$

and by using the relation

$$\int_0^\infty\int_0^x f(x, y)\,dy\,dx = \int_0^\infty\int_y^\infty f(x, y)\,dx\,dy,$$

we obtain

$$32\pi^2\int_0^\infty r_1^2 k_\nu^2(qr_1)\left[\int_{r_1}^\infty r_2 k_\nu^2(qr_2)\,dr_2\right]dr_1.$$

The inner integral may be found from Eq. 11.3.31 of Ref. 6. We then obtain, after substituting $p = qr_1$,

$$V_e = 16\pi^2 q^{-5}(2\nu + 1)^{-1}I(\nu),$$

where

$$I(\nu) = \int_0^\infty p^{4\nu+4}[K_\nu^2(p)K_{\nu+1}^2(p) - K_\nu^4(p)]\,dp.$$

Though integrals over the product of four Bessel functions can be found analytically,[12] there does not appear to be any analytical form in the literature for integrals over the product of four reduced modified Bessel functions. It was therefore decided to evaluate $I(\nu)$ numerically, using a Gaussian–Laguerre quadrature formula

$$\int_0^\infty f(x)\,dx = \sum_{i=1}^{n} w_i \exp(x_i) f(x_i)$$

and taking $n = 32$. The abscissas (x_i) and weights (w_i) were taken from Krylov.[13] This necessitated the evaluation of the modified Bessel function $K_\nu(z)$ at a number of points. Since there is no series formula for $K_\nu(z)$, it was evaluated using an extended trapezoidal rule

$$\int_0^\infty f(x)\,dx = \tfrac{1}{2}f(0) + \sum_{r=1}^{\infty} f(rh),$$

which was applied to the known equality [Eq. (5) of Ref. 14]

$$K_\nu(z) = 2\pi^{\frac{1}{2}}\exp(-z)(2z)^{-\nu} \times \int_0^{+\infty}\exp(-x^2)x^{2\nu}(2z + x^2)^{\nu-\frac{1}{2}}\,dx/\Gamma(\nu + \tfrac{1}{2}) \quad \text{(B1)}$$

Though normally the trapezoidal rule is a crude method for integral evaluation, for this particular

type of integral it is extremely accurate.[15] For a given spacing h, the accuracy increases with increasing z and increasing ν. Because of the latter point it was decided to use recurrence relations (Eq. 9.6.26 of Ref. 6) and recur up ν values below 2.5. For $1.5 \lesssim \nu < 2.5$, we used

$$K_\nu(z) = K_{\nu+2}(z) - 2(\nu + 1)z^{-1}K_{\nu+1}(z) \quad \text{(B2)}$$

and, for $0 < \nu < 1.5$,

$$K_\nu(z) = -2(\nu + 1)z^{-1}K_{\nu+3}(z) + [1 + 4(\nu + 1)(\nu + 2)/z^2]K_{\nu+2}(z). \quad \text{(B3)}$$

For a spacing of $h = 0.01$ and using the above method, we could evaluate $K_\nu(z)$ to at least eight significant figures (by comparison with values in the *Handbook*[6]) for $0 < \nu < 2.5$.

TABLE II. Values of the coefficients a_i which appear in Eq. B4.

a_0 =	0.35394	35479	15791	(+02)
a_1 =	0.64008	20573	26086	(+02)
a_2 =	0.48186	32055	07485	(+02)
a_3 =	0.30841	37226	51208	(+02)
a_4 =	0.17153	11497	48487	(+02)
a_5 =	0.84366	87216	99388	(+01)
a_6 =	0.37248	63238	70390	(+01)
a_7 =	0.14938	42361	02270	(+01)
a_8 =	0.54965	14585	49135	(+00)
a_9 =	0.18705	71051	55576	(+00)
a_{10} =	0.59288	42934	38942	(−01)
a_{11} =	0.17602	60273	92932	(−01)
a_{12} =	0.49200	63307	90673	(−02)
a_{13} =	0.13001	92495	73327	(−02)
a_{14} =	0.32609	72634	96194	(−03)
a_{15} =	0.77880	90840	81385	(−04)
a_{16} =	0.17788	54897	33853	(−04)
a_{17} =	0.38978	60192	68732	(−05)
a_{18} =	0.84410	41927	65371	(−06)
a_{19} =	0.17967	05340	86563	(−06)
a_{20} =	0.57633	37185	27167	(−07)
a_{21} =	0.20665	99346	79070	(−07)
a_{22} =	0.21521	94156	22414	(−07)
a_{23} =	0.96540	90844	91849	(−08)
a_{24} =	0.69663	27958	93493	(−08)

The integral $I(\nu)$ could then be evaluated using the aforementioned 32-point Gaussian–Laguerre quadrature. The accuracy of this procedure is shown by the fact that $I(\frac{1}{2})$, $I(\frac{3}{2})$, and $I(\frac{5}{2})$, having half-integer values, can be found analytically to be 0.1927657109, 12.94662706, and 7014.70333, and the numerical method gives 0.192765714, 12.94662711, and 7014.703337, respectively.

A 25-term combination of Chebyshev polynomials, $T_i(x)$, was then least-squares fitted to 40 values of the integral for $\nu = 0.025(0.05)1.975$, i.e.,

$$I(\nu) = \sum_{i=0}^{24} a_i T_i[(\nu - 1.0)/0.975]. \quad \text{(B4)}$$

Agreement between $I(\nu)$ from Eq. (B4) and from direct calculation was better than eight significant figures for $0 < \nu < 2.0$. The coefficients a_i are given in Table II, and it was with these and Eq. (B4) that V_e was evaluated in the minimization procedure.

[1] R. L. Somorjai, Chem. Phys. Letters **2**, 399 (1968).
[2] L. Hulthén, Arkiv Fysik **28A**, 5 (1942). See also R. G. Parr and J. H. Weare, Progr. Theoret. Phys. (Kyoto) **36**, 854 (1966).
[3] G. E. Roberts and H. Kaufman, *Table of Laplace Transforms* (W. B. Saunders Co., Philadelphia, 1966).
[4] R. L. Somorjai, Chem. Phys. Letters **3**, 395 (1969).
[5] I. Shavitt, *Methods in Computational Physics*, B. Adler and S. Fernbach, Eds. (Academic Press Inc., New York, 1963), Vol. 2.
[6] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (Dover Publications, Inc., New York, 1965).
[7] A. Erdélyi, W. Magnus, F. Oberhettinger, and E. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill Book Co., Inc., New York, 1954), Vol. 1, p. 334.
[8] M. J. D. Powell, Comp. J. 7, 303 (1965). This program was made available through the Quantum Chemistry Program Exchange of the University of Indiana, Bloomington, Ind., as QCPE No. 60. It was converted to double precision by the present authors.
[9] C. C. J. Roothaan and A. W. Weiss, Rev. Mod. Phys. **32**, 194 (1960), Table V.
[10] S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).
[11] D. W. Widder, *The Laplace Transform* (Princeton Univ. Press, Princeton, N.J., 1946).
[12] Y. L. Luke, *Integrals of Bessel Functions* (McGraw-Hill Book Co., Inc., New York, 1962).
[13] V. I. Krylov, *Approximate Calculation of Integrals* (Macmillan Co., New York, 1962), p. 352.
[14] D. B. Hunter, Math. Comp. **18**, 123 (1964).
[15] H. E. Fettis, Math. Tables Aids Comp. **9**, 5 (1955).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Model-Independent Analysis of Nonrelativistic Multiparticle Reactions

RUSSELL E. WARREN AND WILLIAM H. KLINK
Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52240

(Received 6 June 1969)

The Clebsch–Gordan coefficients for the reduction of an n-fold tensor product of nonrelativistic free-particle states into one over-all state are calculated. The degeneracy parameters resulting from this decomposition are then cast into Galilean-invariant functions of the particle momenta. Finally, the reduction is applied to the scattering matrix for nonrelativistic reactions involving an arbitrary number of final particles.

I. INTRODUCTION

There has long been an interest in the quantum-mechanical applications of the physical representations of the Galilei group. Bargmann[1] first showed the need for constructing representations on a one-parameter extension of the Galilei group in order to obtain physically meaningful representations. Using Bargmann's analysis of the central extension, Lévy-Leblond[2] constructed physical representations of the Galilei group for 1- and 2-particle systems with the aid of little group theory and used Galilean invariance to establish many of the characteristics of nonrelativistic systems. Voisin[3] used Mackey's theory of induced representations[4] to construct physical representations of the Galilei group and to couple two single-particle states both in terms of helicity and spin-orbit coupling.

The purpose of the present work is to provide an extension of Voisin's treatment to the problem of coupling n single-particle states. The usual method of doing this relativistically involves a stepwise coupling of particle systems in several different frames.[5] The method of coupling to be used here effects a symmetric reduction of the n-fold system such that all single-particle states are reduced at once into the over-all state. This is convenient since the over-all state is then symmetric under particle interchange, thereby facilitating the construction of symmetric or antisymmetric states when some of the n particles are the same. It is also found that the set of degeneracy parameters that results can be written in the form of Galilean invariants having the same significance in any reference frame.

The second section of the paper reviews the use of the central extension of the Galilei group and treats the construction of induced unitary representations of the central extension defined on a single free-particle basis state in a manner very close to that used by Voisin. The third section is devoted to the calculation of the Clebsch–Gordan coefficients that relate the n single-particle states to one over-all state. Throughout this section the results of Mackey's theory of induced representations[4] are heavily employed; in particular, the method used in reducing the n-fold tensor product of single-particle states and in actually calculating the Clebsch–Gordan coefficients is taken from work of Klink and Smith,[6] which should be referred to for a more complete discussion of the relevant mathematics. In the fourth section, the degeneracy parameters resulting from the n-fold tensor product reduction are cast into the form of Galilean invariants involving quantities that are interpreted as the momenta of "particles" constructed in analogy to the reduced-mass "particle" of 2-body mechanics. In the fifth section, use is made of the Clebsch–Gordan reduction of the third section to write a partial wave analysis of the scattering amplitude for a nonrelativistic scattering problem involving two incoming and $(n-2)$ outgoing particles.

II. CONSTRUCTION OF PHYSICAL REPRESENTATIONS OF THE GALILEI GROUP

The Galilei group G consists of all transformations linking inertial reference frames; the most general transformation of $\mathbf{x}$ and t under G is

$$\mathbf{x}' = R\mathbf{x} + \mathbf{v}t + \mathbf{a},$$
$$t' = t + b. \tag{1}$$

This can be represented in matrix form as

$$\begin{pmatrix} \mathbf{x}' \\ t' \\ 1 \end{pmatrix} = \begin{pmatrix} R & \mathbf{v} & \mathbf{a} \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ t \\ 1 \end{pmatrix}, \tag{2}$$

where R represents the subgroup of proper rotations in three dimensions, $SO(3)$, $\mathbf{v}$ represents an arbitrary "pure" Galilei transformation, $\mathbf{a}$ represents an arbitrary space translation, and b represents an arbitrary time translation. While the matrix display of elements gives the correct combination law of elements of G by

simple matrix multiplication, it will be more useful to write the elements g of G in the form

$$g = [R, \mathbf{v}, \mathbf{a}, b]. \tag{3}$$

The combination law of g_1g_2 is

$$g_1g_2 = [R_1R_2, R_1\mathbf{v}_2 + \mathbf{v}_1, R_1\mathbf{a}_2 + \mathbf{v}_1b_2 + \mathbf{a}_1, b_1 + b_2]. \tag{4}$$

The goal in this section is the construction of irreducible representations of G that are unitary operators acting on momentum space wavefunctions of a free, nonrelativistic particle. It is well known[1] that there is a one-to-one mapping between physical states and rays in a Hilbert space where a ray $\bar{\varphi} \equiv \tau\varphi$ for $|\tau| = 1$. The vector φ is then called a representative of the ray $\bar{\varphi}$.

Therefore, the most general basis for representations allowed by quantum mechanics consists of rays and not vectors. This does not mean that rays must be used in constructing a representation consistent with quantum mechanics, only that their use is permitted. Indeed, it is often possible to fix the phase by setting $\tau = 1$ for all rays, thereby generating a vector representation of the group. An example of this is the Poincaré group in which vector representation can be interpreted physically.

However, when one attempts to set $\tau = 1$ for the Galilei group, it is found that the representations constructed are not consistent with the known transformation properties of momentum and energy under G. These properties are easily derived from Eq. (1) and the fact that $E = p^2/2M + v$ in any given reference frame (M denotes the mass of the particle and v denotes the constant internal energy of the particle). Differentiate Eq. (1) with respect to time, multiply by M, and substitute into the energy relation above to obtain

$$\begin{aligned} \mathbf{p}' &= R\mathbf{p} + M\mathbf{v}, \\ E' &= E + R\mathbf{p}\cdot\mathbf{v} + \tfrac{1}{2}Mv^2. \end{aligned} \tag{5}$$

However, it is found that vector representations constructed on G proper give only

$$\begin{aligned} \mathbf{p}' &= R\mathbf{p}, \\ E' &= E + R\mathbf{p}\cdot\mathbf{v} \end{aligned}$$

and so are too restrictive to be used as physical representations of the group.

The introduction of phase arbitrariness in the basis vectors of the representation allows one to construct representations that are consistent with Eq. (5), and thus provides a physical representation of G. By a representation of G one means that for each element of G there exists a homomorphic mapping that preserves the multiplication law of G by an operator $\bar{U}$. This operator ray is defined as $\bar{U} = \epsilon U$ for $|\epsilon| = 1$, where U is called an operator representative of $\bar{U}$ so that $\bar{\varphi}' = \bar{U}\bar{\varphi}$ and $\bar{U}(g_1)\bar{U}(g_2) = \bar{U}(g_1g_2)$ for all $g \in G$. For two elements of G, g_1 and g_2, one can then associate operator representatives, $U(g_1)$ and $U(g_2)$, with the product law

$$U(g_1)U(g_2) = \omega(g_1, g_2)U(g_1g_2), \quad |\omega(g_1, g_2)| = 1, \tag{6}$$

where $\omega(g_1, g_2)$ is a continuous function of g_1 and g_2 defined for some neighborhood of the identity of G. Bargmann[1] has shown that, by substituting $SU(2)$ for $SO(3)$ (i.e., by using the covering group of G), $\omega(g_1, g_2)$ can be defined continuously over the entire group manifold of G. Letting $\bar{U}(g_1) = \exp(i\theta_1)U(g_1)$ and $\bar{U}(g_2) = \exp(i\theta_2)U(g_2)$ then gives

$$U(g_1)U(g_2) = U(g_1g_2) = e^{i[\theta_1+\theta_2+\varphi(g_1,g_2)]}U(g_1g_2), \tag{7}$$

where $\omega(g_1, g_2) \equiv \exp[i\varphi(g_1, g_2)]$. $\varphi(g_1, g_2)$ is called the local exponent of G, or, in this case, simply the exponent of G, since $\omega(g_1, g_2)$ can be defined everywhere. The combination law of operator representatives led Bargmann to define the central extension of the Galilei group $\bar{G}$ with elements $\bar{g} = (\theta, g)$, where $g \in G$ and θ is any real number. The combination law of $\bar{G}$ is then

$$\begin{aligned} \bar{g}_1\bar{g}_2 &= [\theta_1, g_1][\theta_2, g_2] \\ &= [\theta_1 + \theta_2 + \varphi(g_1, g_2), g_1g_2]. \end{aligned} \tag{8}$$

The motivation for defining $\bar{G}$ is that one can work with vector representations of $\bar{G}$ that are equivalent to projective representations of G. This allows the use of ordinary representation theory. The ability to create representations of $\bar{G}$ consistent with Eq. (5) stems from the freedom available in the choice of $\varphi(g_1, g_2)$. It will be shown later that setting $\varphi(g_1, g_2) = \mathbf{v}_1 \cdot R_1\mathbf{a}_2 + \frac{1}{2}v_1^2b_2$ will yield such representations.

$\bar{G}$ can be decomposed into semidirect products of subgroups

$$\bar{G} = T \times [V \times SU(2)], \tag{9}$$

where T is defined as the invariant subgroup containing elements θ, $\mathbf{a}$, and b, V is defined as the subgroup of pure Galilei transformations $\mathbf{v}$, and $SU(2)$ is the simple covering group of rotations.

In order to use induced representation theory on G, it is first necessary to decompose $\bar{G}$ into a product of some subgroup H with right-coset elements g_c:

$$\bar{G} = \bigcup_c Hg_c. \tag{10}$$

The symbol $\bigcup$ means a union over all right cosets. Any subgroup of $\bar{G}$ can be used to induce a representation, but in order to obtain irreducible representations

of G, $H \equiv T \times L$ must be used. L denotes the little group restriction of $V \times SU(2)$ above.

When representations of H are known, induced representation theory provides the mathematical machinery for generating representations of $\bar{G}$. Physically, for the Galilei group, this means that H provides representations of $\bar{G}$ in the rest frame of the system. Induced representation theory then says what the representations look like in any Galilei frame generated from the rest frame by a pure Galilei transformation.

Since T consists only of translations plus the extensions $\{\theta\}$, it is Abelian and its unitary irreducible representations are all 1 dimensional. They can thus be written as

$$\mathfrak{L}(\mathbf{a}, b, \theta) = \exp i(M\theta + Eb - \mathbf{p}\cdot\mathbf{a}), \qquad (11)$$

where M, E, and $\mathbf{p}$ can be shown to be eigenvalues of the representations of the Lie-algebra elements of G, generating infinitesimal phase, time, and space translations. The minus sign in front of $\mathbf{p}\cdot\mathbf{a}$ is merely a convention suggested by the form of the Lorentz inner product. It is convenient to introduce a 5-dimensional scalar product with vectors

$$p = \begin{pmatrix} \mathbf{p} \\ E \\ M \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} \mathbf{a} \\ b \\ \theta \end{pmatrix}$$

so that

$$\mathfrak{L}^p(a) = e^{ip^T\beta a} \equiv e^{ip\cdot a}, \qquad (12)$$

where

$$\beta \equiv \begin{pmatrix} -I & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since T is normal, $\bar{g}T = T\bar{g}$ for all $\bar{g} \in \bar{G}$. This allows one to define an orbit as the set of representations of T generated by arbitrary inner automorphisms of T induced by $\bar{g} \in \bar{G}$ or as the set

$$\{\mathfrak{L}^{p'}(a) \text{ such that } \mathfrak{L}^{p'}(a) = \mathfrak{L}^p(\bar{g}a\bar{g}^{-1}) \text{ for all } \bar{g} \in \bar{G}\}.$$

To determine the relation of p' and p, calculate $\bar{g}a\bar{g}^{-1}$:

$$\begin{aligned}\bar{g}a'\bar{g}^{-1} &= [\theta, R, \mathbf{v}, \mathbf{a}, b][\theta', I, 0, \mathbf{a}', b'] \\ &\quad \times [-\theta + \mathbf{v}\cdot\mathbf{a} - \tfrac{1}{2}v^2 b, \\ &\qquad R^{-1}, -R^{-1}\mathbf{v}, R^{-1}\mathbf{v}b - R^{-1}\mathbf{a}, -b] \\ &= [\theta' + \mathbf{v}\cdot R\mathbf{a}' + \tfrac{1}{2}v^2 b', I, 0, R\mathbf{a}' + \mathbf{v}b', b'].\end{aligned}$$

Then,

$$\begin{aligned}\mathfrak{L}^p(\bar{g}a'\bar{g}^{-1}) &= \exp i[M(\theta' + \mathbf{v}\cdot R\mathbf{a}' + \tfrac{1}{2}v^2 b') \\ &\quad + Eb' - \mathbf{p}\cdot(R\mathbf{a} + \mathbf{v}b')] \\ &= \exp i[M\theta' + (E - \mathbf{p}\cdot\mathbf{v} + \tfrac{1}{2}Mv^2)b' \\ &\quad - (R^{-1}\mathbf{p} - MR^{-1}\mathbf{v})\cdot\mathbf{a}'] \\ &= \mathfrak{L}^{p'}(a') = \exp i(M'\theta' + E'b' - \mathbf{p}'\cdot\mathbf{a}').\end{aligned}$$

Comparing coefficients of θ', b', and $\mathbf{a}'$ gives

$$\begin{aligned} M &= M', \\ E &= E' + R\mathbf{p}'\cdot\mathbf{v} + \tfrac{1}{2}Mv^2, \\ \mathbf{p} &= R\mathbf{p}' + M\mathbf{v}. \end{aligned} \qquad (13)$$

Also, $E' - \mathbf{p}'^2/2M = E - \mathbf{p}^2/2M = v$. Thus, we are led back to Eq. (5) by having chosen $\varphi(g_1, g_2) = \mathbf{v}_1\cdot R_1\mathbf{a}_2 + \frac{1}{2}v_1^2 b_2$. In terms of the five vectors introduced previously, Eq. (13) can be written

$$p' = \Gamma^{-1}(R, \mathbf{v})p, \qquad (14)$$

where the matrix $\Gamma(R, \mathbf{v})$ and its inverse are defined as

$$\begin{aligned}\Gamma(R, \mathbf{v}) &\equiv \begin{pmatrix} R & 0 & \mathbf{v} \\ \mathbf{v}^T R & 1 & \frac{1}{2}v^2 \\ 0 & 0 & 1 \end{pmatrix}, \\ \Gamma^{-1}(R, \mathbf{v}) &\equiv \begin{pmatrix} R^{-1} & 0 & -R^{-1}\mathbf{v} \\ -\mathbf{v}^T & 1 & \frac{1}{2}v^2 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \qquad (15)$$

The orbit of $\mathbf{p}$ and E is thus seen to be all $\mathbf{p}'$ and E' consistent with constant M and v that can be reached through Galilei transformations of $\mathbf{p}$ and E.

L is defined as the subgroup of $V \times SU(2)$ that leaves at least one element $\hat{\mathbf{p}}$, $\hat{E}$ of the equivalence class, labeled by M and v, invariant. That is,

$$\hat{p} = \Gamma^{-1}(R, \mathbf{v})\hat{p}.$$

It is convenient to choose $\hat{\mathbf{p}} = 0$ and $\hat{E} = v$, which can be written

$$\hat{p} = \begin{pmatrix} 0 \\ v \\ M \end{pmatrix}.$$

With this choice, L consists of all rotations R and no elements of V. $R \in SU(2)$ has representations in the form of Wigner D functions, $D^J_{mm'}(R)$. The representations of $H = T \times L$ are

$$\begin{aligned}\mathfrak{K}^{(\hat{p},J)}(a, R) &= \exp(i\hat{p}^T\beta a)D^J_{mm'}(R) \\ &= \exp(i\hat{p}\cdot a)D^J_{mm'}(R). \end{aligned} \qquad (16)$$

A right-coset decomposition of $\bar{G}$ with respect to H is

$$[\theta, R, \mathbf{v}, \mathbf{a}, b] = [\theta, R, 0, \mathbf{a}, b][0, I, \mathbf{v}', 0, 0], \qquad (17)$$

where $\mathbf{v}' = R^{-1}\mathbf{v}$ is in V, the subgroup of pure Galilei transformations. As shown in Ref. 6, the representation induced by H acts on functions that may be considered to be functions over right-coset elements of G such that

$$U(g_0)f(g_c) = f(g_c g_0),$$

where

$$f(g) = f(hg_c) = \mathcal{K}(h)f(g_c). \tag{18}$$

Here the functions $f(g_c)$ are $f_m(\mathbf{v})$, that is, they are functions over velocity elements of $\bar{G}$ with m arising from the $(2J+1)$-dimensional representations of the inducing subgroup. The $f_m(\mathbf{v})$ are square integrable and form a Hilbert space with norm

$$\|f\|^2 = \sum_{m=-J}^{J} \int d\mu(\mathbf{v})\, |f_m(\mathbf{v})|^2 < \infty, \tag{19}$$

where $d\mu(\mathbf{v})$ is the invariant measure over the coset manifold of $\bar{G}$, $\bar{G}/H$.

The induced unitary irreducible representation of $\bar{G}$ is

$$\begin{aligned}
&U(g_0)f_m(\mathbf{v}) \\
&= f_m([0, I, \mathbf{v}, 0, 0][\theta_0, R_0, \mathbf{v}_0, \mathbf{a}_0, b_0]) \\
&= f_m[\theta_0 + \mathbf{v}\cdot\mathbf{a}_0 + \tfrac{1}{2}v^2 b_0, R_0, \mathbf{v} + \mathbf{v}_0, \mathbf{a}_0 + \mathbf{v}b_0, b_0] \\
&= f_m([\theta_0 + \mathbf{v}\cdot\mathbf{a}_0 + \tfrac{1}{2}v^2 b_0, R_0, 0, \mathbf{a}_0 + \mathbf{v}b_0, b_0] \\
&\quad \times [0, I, R_0^{-1}(\mathbf{v} + \mathbf{v}_0), 0, 0]) \\
&= \exp\,[i\hat{p}^T \Gamma^{-1T}(I, \mathbf{v})\beta a_0] \sum_{m'} D^J_{mm'}(R_0) f_{m'}(R_0^{-1}(\mathbf{v} + \mathbf{v}_0)) \\
&= \exp\,\{i[\Gamma^{-1}(I, \mathbf{v})\hat{p}]^T \beta a_0\} \sum_{m'} D^J_{mm'}(R_0) f_{m'}(R_0^{-1}(\mathbf{v} + \mathbf{v}_0)) \\
&= \exp\,[i\Gamma^{-1}(I, \mathbf{v})\hat{p}\cdot a_0] \sum_{m'} D^J_{mm'}(R_0) f_{m'}(R_0^{-1}(\mathbf{v} + \mathbf{v}_0)).
\end{aligned} \tag{20}$$

So far the induced representation $U(g)$ has been defined on vectors $f_m(\mathbf{v})$ in a rather abstract Hilbert space. To give the representation more physical meaning, let $f_m(\mathbf{v}) = \varphi_m(p)$ with $p = \Gamma^{-1}(I, \mathbf{v})\hat{p}$, where $\varphi_m(p)$ can be interpreted as the momentum space wavefunction satisfying the free-particle Schrödinger equation. This definition is possible since the transformation properties of p have been uniquely specified:

$$\begin{aligned}
U(g_0)\varphi_m(p) &\equiv U(g_0)f_m(\mathbf{v}) \\
&= \exp\,[i\Gamma^{-1}(I, \mathbf{v})\hat{p}\cdot a_0] \sum_{m'} D^J_{mm'}(R_0)\varphi_{m'}(p').
\end{aligned}$$

Since $p = \Gamma^{-1}(I, \mathbf{v})\hat{p}$, Eq. (20) implies that

$$p' = \Gamma^{-1}(I, R_0^{-1}(\mathbf{v} + \mathbf{v}_0))\hat{p} = \Gamma^{-1}(R_0, \mathbf{v}_0)p$$

and thus

$$U(g_0)\varphi_m(p) = e^{ip\cdot a_0} \sum_{m'} D^J_{mm'}(R_0)\varphi_{m'}(\Gamma^{-1}(R_0, \mathbf{v}_0)p). \tag{21}$$

Let $|[vMJ]\mathbf{p}m\rangle$ denote the state vector in Hilbert space corresponding to the wavefunction $\varphi_m(p)$. The labels v, M, and J denote the internal energy, mass, and spin, respectively. They constitute a complete set of irreducible representation labels of $\bar{G}$. The normalization of the states is taken to be

$$\langle [vMJ]\mathbf{p}m \mid [vMJ]\mathbf{p}'m'\rangle = \delta^3(\mathbf{p} - \mathbf{p}')\delta_{mm'}. \tag{22}$$

A general state $|\psi\rangle$ can be written

$$|\psi\rangle = \sum_m \int d\mu(\mathbf{p}, E)\varphi_m(p)\, |[vMJ]\mathbf{p}m\rangle, \tag{23}$$

where $d\mu(\mathbf{p}, E) \equiv d^3\mathbf{p}\, dE\delta(E - p^2/2M - v)$ is the invariant measure. Then,

$$\begin{aligned}
U(g_0)\,|\psi\rangle &= \sum_{m'} \int d\mu(\mathbf{p}, E)[U(g_0)\varphi_{m'}(p)]\, |[vMJ]\mathbf{p}m'\rangle \\
&= \sum_{m', m''} \int d\mu(\mathbf{p}, E) e^{ip\cdot a_0} D^J_{m'm''}(R_0) \\
&\quad \times \varphi_{m''}(\Gamma^{-1}(R_0, \mathbf{v}_0)p)\, |[vMJ]\mathbf{p}m'\rangle.
\end{aligned}$$

Let $\Gamma^{-1}(R_0, \mathbf{v}_0)p = p'$; also, with $d\mu(\mathbf{p}', E') = d\mu(\mathbf{p}, E)$,

$$\begin{aligned}
U(g_0)\,|\psi\rangle &= \sum_{m'm''} \int d\mu(\mathbf{p}', E') e^{i\Gamma(R_0, \mathbf{v}_0)p'\cdot a_0} D^J_{m'm''}(R_0) \\
&\quad \times \varphi_{m''}(p')\, |[vMJ](R_0\mathbf{p}' + M\mathbf{v}_0)m'\rangle, \\
U(g_0)\,|\psi\rangle &= \sum_{m'm} \int d\mu(\mathbf{p}, E) e^{i\Gamma(R_0, \mathbf{v}_0)p\cdot a_0} D^J_{m'm}(R_0) \\
&\quad \times \varphi_m(p)\, |[vMJ](R_0\mathbf{p} + M\mathbf{v}_0)m'\rangle,
\end{aligned}$$

but

$$U(g_0)\,|\psi\rangle = \sum_m \int d\mu(\mathbf{p}, E)\varphi_m(p) U(g_0)\, |[vMJ]\mathbf{p}m\rangle.$$

So, finally, the induced irreducible representation of the extended Galilei group can be written

$$\begin{aligned}
U(g_0)\,|[vMJ]\mathbf{p}m\rangle &= e^{i\Gamma(R_0, \mathbf{v}_0)p\cdot a_0} \sum_{m'} D^J_{m'm}(R_0) \\
&\quad \times |[vMJ](R_0\mathbf{p} + M\mathbf{v}_0), m'\rangle.
\end{aligned} \tag{24}$$

III. CLEBSCH–GORDAN COEFFICIENTS FOR COUPLING n SINGLE-PARTICLE STATES UNDER THE GALILEI GROUP

The method of finding induced representations defined on single-particle states is easily extended to the tensor product of n single-particle states. Define the outer product group $(\bar{G}_1, \bar{G}_2, \cdots, \bar{G}_n)$, consisting of an ordered set of elements $\bar{g}_i \in \bar{G}$. This product group is decomposed with respect to an outer product subgroup $(H_1, \cdots, H_n)$, where each H_i is chosen to induce irreducible representations of $\bar{G}$ and, hence, has representations of the form of Eq. (16). The right-coset elements are then elements of V, the subgroup of pure Galilei transformations. The decomposition is

$$(G_1, \cdots, G_n) = \bigcup_{\mathbf{v}_i} (H_1, \cdots, H_n)(\mathbf{v}_1, \cdots, \mathbf{v}_n). \tag{25}$$

$(H_1, \cdots, H_n)$ has representations

$$\begin{aligned}
&\mathcal{K}^{(\hat{p}_1 J_1, \cdots, \hat{p}_n J_n)}(a_1, R_1, \cdots, a_n, R_n) \\
&\qquad = \prod_{i=1}^{n} \exp\,(i\hat{p}_i\cdot a_i) D^{J_i}(R_i),
\end{aligned}$$

where

$$\hat{p}_i \equiv \begin{pmatrix} 0 \\ v_i \\ M_i \end{pmatrix}, \tag{26}$$

and induces a unitary representation defined on functions of velocity elements

$$f_{m_1 \cdots m_n}(\mathbf{v}_1, \cdots, \mathbf{v}_n).$$

These are square integrable and form a Hilbert space having norm

$$\|f\| = \sum_{m_1=J_1}^{J_1} \cdots \sum_{m_n=-J_n}^{J_n} \int d\mu(\mathbf{v}_1) \cdots d\mu(\mathbf{v}_n) \times |f_{m_1 \cdots m_n}(\mathbf{v}_1, \cdots, \mathbf{v}_n)|^2 < \infty. \tag{27}$$

They also have the property

$$f_{m_1 \cdots m_n}(g_1, \cdots, g_n) = \mathcal{K}(h_1, \cdots, h_n) \times f_{m_1 \cdots m_n}(\mathbf{v}_1, \cdots, \mathbf{v}_n),$$

where

$$h_i = (a_i, R_i). \tag{28}$$

In general, $U(g_0)$ acts on $f(g_{c_1}, \cdots, g_{c_n})$, where g_{c_i} are right-coset elements, to give

$$U(g_0) f(g_{c_1}, \cdots, g_{c_n}) = f(g_{c_1} g_0, g_{c_2} g_0, \cdots, g_{c_n} g_0). \tag{29}$$

The induced representation for the outer-product group is then

$$\begin{aligned}
&U(g_0) f_{m_1 \cdots m_n}(\mathbf{v}_1, \cdots, \mathbf{v}_n) \\
&= f_{m_1 \cdots m_n}([[0, \mathrm{I}, \mathbf{v}_1, 0, 0], \cdots, [0, I, \mathbf{v}_n, 0, 0]] \\
&\quad \times [[\theta_0, R_0, \mathbf{v}_0, \mathbf{a}_0, b_0], \cdots, [\theta_0, R_0, \mathbf{v}_0, \mathbf{a}_0, b_0]]) \\
&= f_{m_1 \cdots m_n}[[\theta_0 + \mathbf{v}_1 \cdot \mathbf{a}_0 + \tfrac{1}{2} v_1^2 b_0, R_0, \mathbf{v}_1 \\
&\quad + \mathbf{v}_0, \mathbf{a}_0 + \mathbf{v}_1 b_0, b_0], \cdots, \\
&\qquad [\theta_0 + \mathbf{v}_n \cdot \mathbf{a}_0 + \tfrac{1}{2} v_n^2 b_0, R_0, \mathbf{v}_n \\
&\quad + \mathbf{v}_0, \mathbf{a}_0 + \mathbf{v}_n b_0, b_0]] \\
&= f_{m_1 \cdots m_n}([[\theta_0 + \mathbf{v}_1 \cdot \mathbf{a}_0 + \tfrac{1}{2} v_1^2 b_0, R_0, 0, \mathbf{a}_0 \\
&\quad + \mathbf{v}_1 b_0, b_0], \cdots, \\
&\qquad [\theta_0 + \mathbf{v}_n \cdot \mathbf{a}_0 + \tfrac{1}{2} v_n^2 b_0, R_0, 0, \mathbf{a}_0 + \mathbf{v}_n b_0, b_0]] \\
&\quad \times [[0, I, R_0^{-1}(\mathbf{v}_1 + \mathbf{v}_0), 0, 0], \cdots, \\
&\qquad [0, I, R_0^{-1}(\mathbf{v}_n + \mathbf{v}_0), 0, 0]]) \\
&= \exp[i(\Gamma^{-1}(I, \mathbf{v}_1)\hat{p}_1)^T \beta a_0] \cdots \exp[i(\Gamma^{-1}(I, \mathbf{v}_n)\hat{p}_n)^T \beta a_0] \\
&\quad \times \sum_{m_1'} D^{J_1}_{m_1 m_1'}(R_0) \cdots \sum_{m_n'} D^{J_n}_{m_n m_n'}(R_0) \\
&\quad \times f_{m_1' \cdots m_n'}[R_0^{-1}(\mathbf{v}_1 + \mathbf{v}_0), \cdots, R_0^{-1}(\mathbf{v}_n + \mathbf{v}_0)] \\
&= \exp[i\Gamma^{-1}(I, \mathbf{v}_1)\hat{p}_1 \cdot a_0] \cdots \exp[i\Gamma^{-1}(I, \mathbf{v}_n)\hat{p}_n \cdot a_0] \\
&\quad \times \sum_{m_1'} D^{J_1}_{m_1 m_1'}(R_0) \cdots \sum_{m_n'} D^{J_n}_{m_n m_n'}(R_0) \\
&\quad \times f_{m_1' \cdots m_n'}[R_0^{-1}(\mathbf{v}_1 + \mathbf{v}_0), \cdots, R_0^{-1}(\mathbf{v}_n + \mathbf{v}_0)].
\end{aligned} \tag{30}$$

The representation as defined over $f_{m_1 \cdots m_n}(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is reducible, and the main object of this paper is to effect a reduction of this reducible representation through the use of double cosets that will allow calculation of the Clebsch–Gordan coefficients. Of primary significance in what follows is a result of Mackey[4] that induced representations defined on tensor product spaces are equivalent to induced representations defined on subspaces labeled by double cosets.

A double-coset decomposition of a group G is defined as

$$G = \bigcup_D H g_D H', \tag{31}$$

where H and H' are both subgroups of G. The g_D are elements of G labeling the double cosets. Using this decomposition, it is possible to construct induced representations on functions $F_D(g)$ for g in G/H_D which are square integrable in the subspace labeled by D. H_D is the subgroup inducing representations on this subspace and is defined as

$$H_D = g_D^{-1} H g_D \cap H'. \tag{32}$$

The direct integral decomposition of $f(g_c)$ is then

$$\|f\|^2 = \int dD \, \|F_D\|^2. \tag{33}$$

Applied to the outer product of the extended Galilei group,

$$\begin{aligned}
&(\bar{G}_1, \bar{G}_2, \cdots, \bar{G}_n) \\
&\quad = \bigcup_{D_1, D_2, \cdots, D_n} (H_1, H_2, \cdots, H_n) \\
&\qquad \times (g_{D_1}, g_{D_2}, \cdots, g_{D_n})(\bar{G}, \bar{G}, \cdots, \bar{G}),
\end{aligned} \tag{34}$$

where $(\bar{G}, \bar{G}, \cdots, \bar{G})$ is the diagonal subgroup of $(\bar{G}_1, \bar{G}_2, \cdots, \bar{G}_n)$ consisting of the same element $\bar{g} \in \bar{G}$ in all positions. It is clear that the double-coset elements are elements of V,

$$\begin{aligned}
&[[\theta_1, R_1', \mathbf{v}_1, \mathbf{a}_1, b_1], \cdots, [\theta_n, R_n', \mathbf{v}_n, \mathbf{a}_n b_n]] \\
&\quad = [[\theta_1, R_1, 0, \mathbf{a}_1, b_1], \cdots, [\theta_n, R_n, 0, \mathbf{a}_n, b_n]] \\
&\qquad \times [[0, I, \mathbf{v}_{D_1}, 0, 0], \cdots, [0, I, \mathbf{v}_{D_n}, 0, 0]] \\
&\qquad \times [[\theta, R, \mathbf{v}, \mathbf{a}, b], \cdots, [\theta, R, \mathbf{v}, \mathbf{a}, b]].
\end{aligned} \tag{35}$$

The $\mathbf{v}_{D_i}$ are chosen to be the smallest set consistent with a unique coverage of elements of the outer product. A convenient choice is

$$\begin{aligned}
&\mathbf{v}_{D_1} = 0, \\
&\mathbf{v}_{D_2} = (0, 0, v_{D_{2z}}), \\
&\mathbf{v}_{D_3} = (v_{D_{3x}}, 0, v_{D_{3z}}), \\
&\mathbf{v}_{D_i} \quad \text{arbitrary for} \quad i = 4, \cdots, n.
\end{aligned} \tag{36}$$

Then,

$$\begin{aligned}H_D &= (\mathbf{v}_{D_1}, \cdots, \mathbf{v}_{D_n})^{-1}(H_1, H_2, \cdots, H_n)\\ &\quad \times (\mathbf{v}_{D_1}, \cdots, \mathbf{v}_{D_n}) \cap (\vec{G}, \cdots, \vec{G})\\ &= [[0, I, -\mathbf{v}_{D_1}, 0, 0], \cdots, [0, I, -\mathbf{v}_{D_n}, 0, 0]]\\ &\quad \times [[\theta_1, R_1, 0, \mathbf{a}_1, b_1], \cdots, [\theta_n, R_n, 0, \mathbf{a}_n, b_n]]\\ &\quad \times [[0, I, \mathbf{v}_{D_1}, 0, 0], \cdots, [0, I, \mathbf{v}_{D_n}, 0, 0]]\\ &\quad \cap [[\theta, R, \mathbf{v}, \mathbf{a}, b], \cdots, [\theta, R, \mathbf{v}, \mathbf{a}, b]]\\ &= [[\theta_1 - \mathbf{v}_{D_1}\cdot\mathbf{a}_1 + \tfrac{1}{2}v_{D_1}^2 b_1, R_1,\\ &\qquad R_1\mathbf{v}_{D_1} - \mathbf{v}_{D_1}, \mathbf{a}_1 - \mathbf{v}_{D_1}b_1, b_1], \cdots,\\ &\qquad [\theta_n - \mathbf{v}_{D_n}\cdot\mathbf{a}_n + \tfrac{1}{2}v_{D_n}^2 b_n, R_n,\\ &\qquad R_n\mathbf{v}_{D_n} - \mathbf{v}_{D_n}, \mathbf{a}_n - \mathbf{v}_{D_n}b_n, b_n]]\\ &\quad \cap (\vec{G}, \vec{G}, \cdots, \vec{G}). \end{aligned} \tag{37}$$

Now, since $\mathbf{v}_{D_1} = 0$, $R_1\mathbf{v}_{D_1} - \mathbf{v}_{D_1} = 0$, and $0 \cap \mathbf{v} = 0$. Then, $0 \cap \mathbf{v} = R_2\mathbf{v}_{D_2} - \mathbf{v}_{D_2} \cap \mathbf{v} = \cdots = R_n\mathbf{v}_{D_n} - \mathbf{v}_{D_n} \cap \mathbf{v} = 0$. Since $\mathbf{v}$ is arbitrary, R_i can be at most a rotation about $\mathbf{v}_{D_i}$ for $i = 2, \cdots, n$. For $n = 2$, there is only one nonvanishing $\mathbf{v}_D$, so that R_1 can be taken to be $U(1)$ and $R_2 = R_1 = U(1)$. For $n \geq 3$, there are at least two different $\mathbf{v}_D$'s. Since $R_1 = R_2 = \cdots = R_n$, only $R_i = I$ will work. Because a number of authors have treated the problem of coupling two particles, the remainder of this paper will treat only the case of $n \geq 3$. Then,

$$H_D = [[\theta, I, 0, \mathbf{a}, b], \cdots, [\theta, I, 0, \mathbf{a}, b]]. \tag{38}$$

In order to have a unique association of elements on the left and right sides of the double-coset decomposition, the elements of H_D must be divided out of the diagonal subgroup. The result is

$$\begin{aligned}&[[\theta_1, R_1', \mathbf{v}_1, \mathbf{a}_1, b_1], \cdots, [\theta_n, R_n', \mathbf{v}_n, \mathbf{a}_n, b_n]]\\ &\quad = [[\theta_1, R_1, 0, \mathbf{a}_1, b_1], \cdots, [\theta_n, R_n, 0, \mathbf{a}_n, b_n]]\\ &\qquad \times [[0, I, \mathbf{v}_{D_1}, 0, 0], \cdots, [0, I, \mathbf{v}_{D_n}, 0, 0]]\\ &\qquad \times [[0, R, \mathbf{v}, 0, 0], \cdots, [0, R, \mathbf{v}, 0, 0]]\\ &\quad = [[\theta_1, R_1, 0, \mathbf{a}_1, b_1], \cdots, [\theta_n, R_n, 0, \mathbf{a}_n, b_n]]\\ &\qquad \times [[0, R, \mathbf{v}+\mathbf{v}_{D_1}, 0, 0], \cdots, [0, R, \mathbf{v}+\mathbf{v}_{D_n}, 0, 0]]. \end{aligned} \tag{39}$$

Now, one can define functions carrying representations induced by H_D as

$$\begin{aligned}F_{[m_i,\mathbf{v}_{D_i}]}(R, \mathbf{v}) &\equiv F_{m_1\cdots m_n}[(R, \mathbf{v}+\mathbf{v}_{D_1}),\\ &\qquad (R, \mathbf{v}+\mathbf{v}_{D_2}), \cdots, (R, \mathbf{v}+\mathbf{v}_{D_n})]. \end{aligned}$$

Under an arbitrary transformation $F_{[m_i,\mathbf{v}_{D_i}]}(R, \mathbf{v})$ transforms as

$$\begin{aligned}&U(g_0)F_{[m_1,\mathbf{v}_{D_i}]}(R, \mathbf{v})\\ &\equiv U(g_0)F_{m_1\cdots m_n}[(R, \mathbf{v}+\mathbf{v}_{D_1}), \cdots, (R, \mathbf{v}+\mathbf{v}_{D_n})]\\ &= F_{m_1\cdots m_n}([[0, R, \mathbf{v}+\mathbf{v}_{D_1}, 0, 0], \cdots,\\ &\qquad [0, R, \mathbf{v}+\mathbf{v}_{D_n}, 0, 0]]\\ &\quad \times [[\theta_0, R_0, \mathbf{v}_0, \mathbf{a}_0, b_0], \cdots, [\theta_0, R_0, \mathbf{v}_0, \mathbf{a}_0, b_0]])\\ &= F_{m_1\cdots m_n}[[\theta_0 + (\mathbf{v}+\mathbf{v}_{D_1})\cdot R\mathbf{a}_0 + \tfrac{1}{2}(\mathbf{v}+\mathbf{v}_{D_1})^2 b_0,\\ &\qquad RR_0, R\mathbf{v}_0 + \mathbf{v} + \mathbf{v}_{D_1},\\ &\qquad R\mathbf{a}_0 + (\mathbf{v}+\mathbf{v}_{D_1})b_0, b_0], \cdots,\\ &\qquad [\theta_0 + (\mathbf{v}+\mathbf{v}_{D_n})\cdot R\mathbf{a}_0 + \tfrac{1}{2}(\mathbf{v}+\mathbf{v}_{D_n})^2 b_0, RR_0,\\ &\qquad R\mathbf{v}_0 + \mathbf{v} + \mathbf{v}_{D_n}, R\mathbf{a}_0 + (\mathbf{v}+\mathbf{v}_{D_n})b_0, b_0]]\\ &= F_{m_1\cdots m_n}([[\theta_0 + (\mathbf{v}+\mathbf{v}_{D_1})\cdot R\mathbf{a}_0 + \tfrac{1}{2}(\mathbf{v}+\mathbf{v}_{D_1})^2 b_0,\\ &\qquad I, 0, R\mathbf{a}_0 + (\mathbf{v}+\mathbf{v}_{D_1})b_0, b_0], \cdots,\\ &\qquad [\theta_0 + (\mathbf{v}+\mathbf{v}_{D_n})\cdot R\mathbf{a}_0 + \tfrac{1}{2}(\mathbf{v}+\mathbf{v}_{D_n})^2 b_0, I, 0,\\ &\qquad R\mathbf{a}_0 + (\mathbf{v}+\mathbf{v}_{D_n})b_0, b_0]]\\ &\quad \times [[0, RR_0, R\mathbf{v}_0 + \mathbf{v} + \mathbf{v}_{D_1}, 0, 0], \cdots,\\ &\qquad [0, RR_0, R\mathbf{v}_0 + \mathbf{v} + \mathbf{v}_{D_n}, 0, 0]])\\ &= \exp[i\Gamma^{-1}(R, \mathbf{v}+\mathbf{v}_{D_1})\hat{p}_1\cdot a_0]\cdots\\ &\qquad \exp[i\Gamma^{-1}(R, \mathbf{v}+\mathbf{v}_{D_n})\hat{p}_n\cdot a_0]\\ &\quad \times F_{m_1\cdots m_n}[(RR_0, R\mathbf{v}_0 + \mathbf{v} + \mathbf{v}_{D_1}), \cdots,\\ &\qquad (RR_0, R\mathbf{v}_0 + \mathbf{v} + \mathbf{v}_{D_n})]\\ &= \exp[i\Gamma^{-1}(R, \mathbf{v})\tilde{p}\cdot a_0]F_{[m_i,v_{D_i}]}(RR_0, R\mathbf{v}_0 + \mathbf{v}), \end{aligned} \tag{40}$$

where $\tilde{p}$ means $\begin{pmatrix}\tilde{\mathbf{p}}\\ \tilde{E}\\ M\end{pmatrix}$ and is defined by

$$\begin{aligned}\Gamma^{-1}(R, \mathbf{v})\tilde{p} &\equiv \Gamma^{-1}(R, \mathbf{v}+\mathbf{v}_{D_1})\hat{p}_1\\ &\qquad + \cdots + \Gamma^{-1}(R, \mathbf{v}+\mathbf{v}_{D_n})\hat{p}_n \end{aligned}$$

or

$$\tilde{p} = \Gamma(I, -\mathbf{v}_{D_1})\hat{p}_1 + \cdots + \Gamma(I, -\mathbf{v}_{D_n})\hat{p}_n. \tag{41}$$

Since the subscripts $[m_i, \mathbf{v}_{D_i}]$ are left invariant under an arbitrary Galilei transformation, it is clear that they label degenerate subspaces of the invariant subspace of Hilbert-space vectors labeled by M and ν. Equation (41) yields the following equations for the components of $\tilde{p}$:

$$\begin{aligned}\tilde{\mathbf{p}} &= -M_1\mathbf{v}_{D_i} - \cdots - M_n\mathbf{v}_{D_n},\\ \tilde{E} &= \nu_1 + \tfrac{1}{2}M_1 v_{D_1}^2 + \cdots + \nu_n + \tfrac{1}{2}M_n v_{D_n}^2,\\ M &= M_1 + \cdots + M_n. \end{aligned} \tag{42}$$

The Clebsch–Gordan coefficients relate the n single-particle states to one over-all single-particle state

through the equation

$$|[\nu_1 M_1 J_1]\mathbf{p}_1 m_1\rangle\,|[\nu_2 M_2 J_2]p_2 m_2\rangle\cdots|[\nu_n M_n J_n]\mathbf{p}_n m_n\rangle$$
$$=\sum_{J=0}^{\infty}\sum_{m=-J}^{J}\sum_{\eta}\int d\mu(\mathbf{p})\times\langle[\nu MJ]\mathbf{p}m,\eta|\,[\nu_1 M_1 J_1]p_1 m_1,\cdots,[\nu_n M_n J_n]\mathbf{p}_n m_n\rangle\,|[\nu MJ]\mathbf{p}m,\eta\rangle, \quad (43)$$

where η denotes the set of degeneracy parameters distinguishing subspaces within the representation characterized by ν, M, and J. To obtain these Clebsch–Gordan coefficients, introduce the functions

$$D^{[\nu MJ]}_{\mathbf{p}'m'\mathbf{p}m}(R,\mathbf{v})$$

defined as

$$D^{[\nu MJ]}_{\mathbf{p}'m'\mathbf{p}m}(R,\mathbf{v})$$
$$\equiv\langle[\nu MJ]\mathbf{p}'m'|\,U(R,\mathbf{v})\,|[\nu MJ]\mathbf{p}m\rangle$$
$$=\sum_{m''}D^{J}_{m''m}(R)\,\langle[\nu MJ]\mathbf{p}'m'\,|\,[\nu MJ](R\mathbf{p}+M\mathbf{v})m''\rangle$$
$$=D^{J}_{m'm}(R)\delta^3(\mathbf{p}'-(R\mathbf{p}+M\mathbf{v})). \quad (44)$$

It is desired to leave $\mathbf{p}$ arbitrary and choose $\mathbf{p}'$ such that the D functions will transform like basis functions for the induced representation when the operator $U(R,\mathbf{v})$ acts to the left. Then, since $U(R,\mathbf{v})$ acting to the right causes the D functions to transform like the state vectors $|[\nu MJ]\mathbf{p}m\rangle$, one can regard them both as concrete realizations of $|[\nu MJ]\mathbf{p}m\rangle$ and as basis functions for the induced representation. This feature permits the Clebsch–Gordan coefficients to be expressed as an integral over $(R,\mathbf{v})$ because of the already completed double-coset reduction of basis functions for the n-fold tensor product.

Correct choices of D functions to represent the state vectors appearing in the Clebsch–Gordan coefficients must now be made. Choose

$$F_{[m_i',\mathbf{v}_{D_i}]}(R,\mathbf{v})\quad\text{to be}\quad D^{[\nu MJ]}_{\mathbf{p}k\mathbf{p}m}(R,\mathbf{v}),$$

where k is an additional degeneracy parameter. Likewise, choose

$$F'_{[m_i,\mathbf{v}_{D_i}]}(R,\mathbf{v})$$
$$=F'_{m_1'\cdots m_n'}[(R,\mathbf{v}+\mathbf{v}_{D_1}),\cdots,(R,\mathbf{v}+\mathbf{v}_{D_n})]$$

to be

$$\prod_{i=1}^{n}D^{[\nu_i M_i J_i]}_{\mathbf{p}_i m_i'\mathbf{p}_i m_i}(R,\mathbf{v}+\mathbf{v}_{D_i}).$$

Then,

$$\langle[\nu MJ]\mathbf{p}m,\,m_i'\mathbf{v}_{D_i}k\,|\,[\nu_1 M_1 J_1]\mathbf{p}_1 m_1,\cdots,[\nu_n M_n J_n]\mathbf{p}_n m_n\rangle$$
$$=N_J\int d(R,\mathbf{v})D^{[\nu MJ]^*}_{\tilde{\mathbf{p}}k\mathbf{p}m}(R,\mathbf{v})\times\prod_{i=1}^{n}D^{[\nu_i M_i J_i]}_{\hat{\mathbf{p}}_i m_i'\mathbf{p}_i m_i}(R,\mathbf{v}+\mathbf{v}_{D_i})$$
$$=N_J\int d(R,\mathbf{v})\delta^3(\tilde{\mathbf{p}}-(R\mathbf{p}+M\mathbf{v}))D^{J^*}_{km}(R)\times\prod_{i=1}^{n}\delta^3(\hat{\mathbf{p}}_i-[R\mathbf{p}_i+M_i(\mathbf{v}+\mathbf{v}_{D_i})])D^{J_i}_{m_i'm_i}(R), \quad (45)$$

where N_J is a normalization constant. The δ functions imply

$$\tilde{\mathbf{p}}=R\mathbf{p}+M\mathbf{v},$$
$$\hat{\mathbf{p}}_i=R\mathbf{p}_i+M_i(\mathbf{v}+\mathbf{v}_{D_i})=0,\quad i=1,\cdots,n. \quad (46)$$

Now Eq. (42) can be used with the above to obtain the expected conservation of momentum and energy for the over-all state:

$$\mathbf{p}=\sum_{i=1}^{n}\mathbf{p}_i,$$
$$E=\sum_{i=1}^{n}E_i. \quad (47)$$

When the integration over $(R,\mathbf{v})$ is performed, the Clebsch–Gordan coefficients are found to be

$$\langle[\nu MJ]\mathbf{p}m,\,m_i'k\mathbf{v}_{D_i}\,|\,[\nu_1 M_1 J_1]\mathbf{p}_1 m_1,\cdots,[\nu_n M_n J_n]\mathbf{p}_n m_n\rangle$$
$$=N_J\delta^3\Big(\mathbf{p}-\sum_i\mathbf{p}_i\Big)\delta\Big(E-\sum_i E_i\Big)\times\delta\Big(M-\sum_i M_i\Big)D^{J^*}_{km}(R)\prod_{i=1}^{n}D^{J_i}_{m_i'm_i}(R). \quad (48)$$

The significance of the rotation R appearing in the Wigner functions will be clarified in the following section.

IV. INTERPRETATION OF THE DEGENERACY PARAMETERS AND THE ROTATION R

At this point the set of degeneracy parameters is of the form $[m_i', k, \mathbf{v}_{D_i}]$. It is desirable to express the $\mathbf{v}_{D_i}$ in terms of quantities that have more physical significance. To do this, construct a coordinate system as shown, where $\mathbf{R}$ is defined as the position vector of the center of mass, $\mathbf{r}_i$ is defined as the position vector of the ith particle, and $\mathbf{r}_i'$ denotes the position vector of the ith particle relative to the center of mass (see Fig. 1). Let T represent the total kinetic energy of the

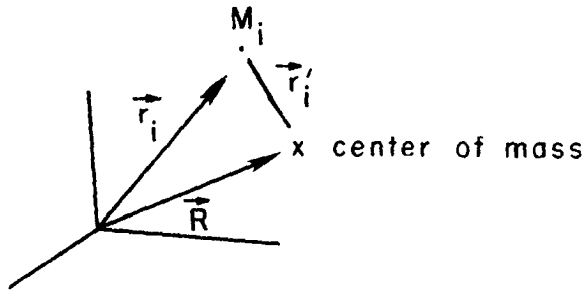


FIG. 1.

system:

$$T = \sum_i \tfrac{1}{2} M_i \dot{r}_i^2 = \sum \tfrac{1}{2} M_i (\dot{\mathbf{R}} + \dot{\mathbf{r}}_i')^2$$
$$= \tfrac{1}{2} M \dot{R}^2 + \sum_i \tfrac{1}{2} M_i \dot{r}_i'^2 = T_{\rm CM} + T_{\rm rel}, \quad (49)$$

since $\sum M_i \mathbf{r}_i' = 0$. Define $\mathbf{r}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$. Then,

$$\sum_i M_i (\mathbf{r}_j' + \mathbf{r}_{ij}) = 0,$$

such that

$$\dot{\mathbf{r}}_i' = \frac{1}{M} \sum_j M_j \mathbf{r}_{ij}, \quad T_{\rm rel} = \sum_i \tfrac{1}{2} M_i \sum_{k,l} \frac{M_k M_l}{M^2} \dot{\mathbf{r}}_{ik} \cdot \dot{\mathbf{r}}_{il}.$$

Then,

$$\dot{\mathbf{r}}_{kl} = \dot{\mathbf{r}}_k - \dot{\mathbf{r}}_l = (\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_l) - (\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_k) = \dot{\mathbf{r}}_{il} - \dot{\mathbf{r}}_{ik},$$
$$\dot{\mathbf{r}}_{kl}^2 = \dot{\mathbf{r}}_{il}^2 + \dot{\mathbf{r}}_{ik}^2 - 2\dot{\mathbf{r}}_{il} \cdot \dot{\mathbf{r}}_{ik},$$

such that

$$\dot{\mathbf{r}}_{ik} \cdot \dot{\mathbf{r}}_{il} = \tfrac{1}{2}(\dot{\mathbf{r}}_{il}^2 + \dot{\mathbf{r}}_{ik}^2 - \dot{\mathbf{r}}_{kl}^2),$$

$$T_{\rm rel} = \sum_{i,k,l} \frac{M_i M_k M_l}{2M^2} \tfrac{1}{2}(\dot{\mathbf{r}}_{il}^2 + \dot{\mathbf{r}}_{ik}^2 - \dot{\mathbf{r}}_{kl}^2)$$
$$= \sum_{i,l} \frac{M_i M_l}{4M} \dot{\mathbf{r}}_{il}^2 + \sum_{i,k} \frac{M_i M_k}{4M} \dot{\mathbf{r}}_{ik}^2 - \sum_{k,l} \frac{M_k M_l}{4M} \dot{\mathbf{r}}_{kl}^2.$$

Since all indices are dummy indices, the three sums are equal and

$$T_{\rm rel} = \tfrac{1}{2} \sum_i M_i \dot{\mathbf{r}}_i'^2 = \tfrac{1}{2} \sum_{i,j} \frac{M_i M_j}{2M} \dot{\mathbf{r}}_{ij}^2 = \sum_{i<j} \frac{M_i M_j}{2M} \dot{\mathbf{r}}_{ij}^2. \quad (50)$$

The $\mathbf{r}_{ij}$ can be regarded as canonical coordinates describing the relative positions of particles in the center-of-mass frame. Since no interactions have been introduced between particles, the Lagrangian $L = T_{\rm rel}$ and the momenta conjugate to $\mathbf{r}_{ij}$, defined as $\mathbf{p}_{ij}$, become

$$\mathbf{p}_{ij} = \frac{\partial L}{\partial \dot{\mathbf{r}}_{ij}} = \frac{M_i M_j}{M} \dot{\mathbf{r}}_{ij} = \frac{M_i M_j}{M} (\mathbf{v}_i - \mathbf{v}_j)$$
$$= \frac{M_j \mathbf{p}_i - M_i \mathbf{p}_j}{M}. \quad (51)$$

Letting $M_{ij} \equiv M_i M_j / M$, we have

$$T_{\rm rel} = \sum_{i<j} \tfrac{1}{2} M_{ij} \dot{\mathbf{r}}_{ij}^2 = \sum_{i<j} \frac{\mathbf{p}_{ij}^2}{2M_{ij}}. \quad (52)$$

Since $E = \sum_i E_i$,

$$\frac{p^2}{2M} + v = \sum \frac{p_i^2}{2M_i} + \sum v_i = \frac{p^2}{2M} + \sum_{i<j} \frac{\mathbf{p}_{ij}^2}{2M_{ij}} + \sum v_i,$$

such that

$$v = \sum v_i + \sum_{i<j} \frac{\mathbf{p}_{ij}^2}{2M_{ij}}. \quad (53)$$

The total internal energy v is seen to be the sum of the internal energies of the n single-particle states plus their relative kinetic energy about the center of mass. One can regard the $\mathbf{p}_{ij}$ as being the momenta of "particles" of mass M_{ij}, whose positions relative to the center of mass of the system are specified by $\mathbf{r}_{ij}$. This is a natural extension of the reduced-mass concept of 2-body mechanics. The reason for defining the $\mathbf{p}_{ij}$ is that it is now a simple matter to express the $\mathbf{v}_{D_i}$ in terms of $\mathbf{p}_{ij}$:

$$R\mathbf{p}_i = -M_i(\mathbf{v} + \mathbf{v}_{D_i}),$$
$$R\mathbf{p}_j = -M_j(\mathbf{v} + \mathbf{v}_{D_j}). \quad (46')$$

Therefore,

$$R(M_j \mathbf{p}_i - M_i \mathbf{p}_j) = M_i M_j (\mathbf{v}_{D_j} - \mathbf{v}_{D_i}),$$
$$\mathbf{v}_{D_j} - \mathbf{v}_{D_i} = R(M_j \mathbf{p}_i - M_i \mathbf{p}_j / M_i M_j)$$
$$= R\mathbf{p}_{ij} / M_{ij}. \quad (54)$$

There are $\frac{1}{2}n(n-1)$ independent $\mathbf{p}_{ij}$ and only $(3n-6)$ $\mathbf{v}_{D_i}$ components, so that a choice of $\mathbf{p}_{ij}$ can be made that will completely specify the $\mathbf{v}_{D_i}$, but this choice will not be unique. In particular, for the set of $\mathbf{v}_{D_i}$ defined previously, the following representation will cover the $\mathbf{v}_{D_i}$ components completely except for $(n-2)$ sign ambiguities. It is clear that this representation is invariant under an arbitrary Galilean transformation:

$$v_{D_{1x}} = 0, \quad v_{D_{2x}} = |\mathbf{p}_{12}| / M_{12},$$
$$v_{D_{ix}} = \frac{\mathbf{p}_{12} \cdot \mathbf{p}_{1i}}{|\mathbf{p}_{12}|\, M_{1i}}, \quad i \geq 3,$$
$$v_{D_{1z}} = 0, \quad v_{D_{2z}} = 0,$$
$$v_{D_{3x}} = \pm \left(\frac{p_{13}^2}{M_{13}^2} - \frac{\mathbf{p}_{12} \cdot \mathbf{p}_{13}}{|\mathbf{p}_{12}|\, M_{13}} \right)^{\frac{1}{2}},$$
$$v_{D_{ix}} = \pm \left(\frac{p_{13}^2}{M_{13}^2} - \frac{\mathbf{p}_{12} \cdot \mathbf{p}_{13}}{|\mathbf{p}_{12}|\, M_{13}} \right)^{-\frac{1}{2}} \left(\frac{\mathbf{p}_{13} \cdot \mathbf{p}_{1i}}{M_{13} M_{1i}} - v_{D_{3x}} v_{D_{iz}} \right), \quad i \geq 4,$$
$$v_{D_{1y}} = 0, \quad v_{D_{2y}} = 0, \quad v_{D_{3y}} = 0,$$
$$v_{D_{iy}} = \pm \left(\frac{p_{1i}^2}{M_{1i}^2} - v_{D_{ix}}^2 - v_{D_{iz}}^2 \right)^{\frac{1}{2}}, \quad i \geq 4. \quad (55)$$

To remove the $(n-2)$ sign ambiguities above, consider quantities of the form

$$\epsilon_{\alpha\beta\gamma} p_{ij}^{\alpha} p_{kl}^{\beta} p_{mn}^{\gamma}. \quad (56)$$

It is easy to show that these quantities are Galilean invariants. Now form

$$\epsilon_{\alpha\beta\gamma} p_{13}^{\alpha} p_{14}^{\beta} p_{15}^{\gamma} = M_{13} M_{14} M_{15} \epsilon_{\alpha\beta\gamma} v_{D_3}^{\alpha} v_{D_4}^{\beta} v_{D_5}^{\gamma},$$
$$\epsilon_{\alpha\beta\gamma} p_{12}^{\alpha} p_{13}^{\beta} p_{1i}^{\gamma} = M_{12} M_{13} M_{1i} \epsilon_{\alpha\beta\gamma} v_{D_2}^{\alpha} v_{D_3}^{\beta} v_{D_i}^{\gamma}$$
$$= M_{12} M_{13} M_{1i} v_{D_2}^{3} v_{D_3}^{1} v_{D_i}^{2}, \quad i = 4, \cdots, n, \quad (57)$$

where the last result follows from the choice of $\mathbf{v}_{D_i}$ made. These $(n-2)$ invariants are sufficient to remove the sign ambiguities in the $\mathbf{v}_{D_i}$ representation given.

The interpretation of R can be made most easily considering $\mathbf{p}_{12}$ and $\mathbf{p}_{13}$. In terms of the double-coset elements used here, they are

$$\mathbf{v}_{D_2} = (0, 0, v_{D_{2z}}) = R\mathbf{p}_{12}/M_{12},$$
$$\mathbf{v}_{D_3} = (v_{D_{3x}}, 0, v_{D_{3z}}) = R\mathbf{p}_{13}/M_{13}.$$

It is clear that R must be the rotation that takes $\mathbf{p}_{12}$ into the z axis while rotating $\mathbf{p}_{13}$ into the $x-z$ plane of the coordinate system in which the momenta are $\mathbf{p}_1, \cdots, \mathbf{p}_n$. In general, R must be specified by three parameters; a convenient choice is the set of Euler angles as defined by Goldstein,[7] $R = (\varphi, \theta, \psi)$. Using the rotation matrix given by Goldstein in terms of the Euler angles and the values of $\mathbf{v}_{D_2}$ and $\mathbf{v}_{D_3}$ given previously as functions of $\mathbf{p}_{12}$ and $\mathbf{p}_{13}$, one can specify R as a function of $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ and masses M_1, M_2, and M_3.

R can also be shown to rotate all momenta by the same angles in any Galilean reference frame. This means that the momenta can be thought of as defining a rigid body that rotates as a unit under R. Therefore, irrespective of the number of particle states coupled, the rotation R appearing in the Wigner functions is completely specified by the momenta of any three of the particles. The choice of these three is obviously irrelevant; the use of the first three here is due solely to the previous choice of double-coset elements. A different set of double cosets would lead to a different but physically equivalent interpretation of R.

V. PARTIAL WAVE ANALYSIS

As an application of the Clebsch–Gordan reduction calculated in Sec. III, consider the scattering matrix for the following nonrelativistic reaction:

$$1 + 2 \to 3 + 4 + \cdots + n. \tag{58}$$

An S-matrix element will have the form

$$\langle[\nu_3 M_3 J_3]\mathbf{p}_3 m_3| \cdots \langle[\nu_n M_n J_n]\mathbf{p}_n m_n|\, S\, |[\nu_1 M_1 J_1]\mathbf{p}_1 m_1\rangle \times |[\nu_2 M_2 J_2]\mathbf{p}_2 m_2\rangle. \tag{59}$$

In exactly the same manner used in Sec. III, the Clebsch–Gordan reduction, the two single-particle states can be shown to be

$$|[\nu_1 M_1 J_1]\mathbf{p}_1 m_1\rangle\, |[\nu_2 M_2 J_2]\mathbf{p}_2 m_2\rangle = \sum_{J=0}^{\infty} \sum_{m=-J}^{J} \sum_{i_1 i_2} \int d\mu(\mathbf{p})\, |[\nu MJ]\mathbf{p}m, i_1 i_2 \mathbf{v}_D\rangle \times \langle[\nu MJ]\mathbf{p}m, i_1 i_2 \mathbf{v}_D \,|\, [\nu_1 M_1 J_1]\mathbf{p}_1 m_1, [\nu_2 M_2 J_2]p_2 m_2\rangle \tag{60}$$

with

$$\langle[\nu MJ]\mathbf{p}m, i_1 i_2 \mathbf{v}_D \,|\, [\nu_1 M_1 J_1]\mathbf{p}_1 m_1 [\nu_2 M_2 J_2]\mathbf{p}_2 m_2\rangle = \delta^3(\mathbf{p} - \mathbf{p}_1 - \mathbf{p}_2)\delta(E - E_1 - E_2)\delta(M - M_1 - M_2) \times D^{J^*}_{i_1+i_2,m}(R') D^{J_1}_{i_1 m_1}(R') D^{J_2}_{i_2 m_2}(R'), \tag{61}$$

where $[i_1, i_2, \mathbf{v}_D]$ is the set of degeneracy parameters for the 2-particle case. Now let particles 1 and 2 define the z axis in the center-of-mass frame. Then, from the analysis of the last section, $R' = I$ and the Clebsch–Gordan coefficients become

$$\langle[\nu MJ]\mathbf{p}m, i_1 i_2 \mathbf{v}_D \,|\, [\nu_1 M_1 J_1]\mathbf{p}_1 m_1, [\nu_2 M_2 J_2]\mathbf{p}_2 m_2\rangle = \delta^3(\mathbf{p} - \mathbf{p}_1 - \mathbf{p}_2)\delta(E - E_1 - E_2)\delta(M - M_1 - M_2) \times \delta_{i_1+i_2,m}\delta_{i_1 m_1}\delta_{i_2 m_2}. \tag{62}$$

Then,

$$\begin{aligned}|[\nu_1 M_1 J_1]\mathbf{p}_1 m_1\rangle\, |[\nu_2 M_2 J_2]\mathbf{p}_2 m_2\rangle &= \sum_J \sum_m \sum_{i_1 i_2} N_J \int d\mu(\mathbf{p})\, |[\nu MJ]\mathbf{p}m, i_1 i_2 \mathbf{v}_D\rangle \\ &\quad \times \delta^3(\mathbf{p} - \mathbf{p}_1 - \mathbf{p}_2)\delta(E - E_1 - E_2) \\ &\quad \times \delta(M - M_1 - M_2)\delta_{i_1+i_2,m}\delta_{i_1 m_1}\delta_{i_2 m_2} \\ &= \sum_J N_J\, |[\nu(M_1 + M_2)J]\,(\mathbf{p}_1 + \mathbf{p}_2 = 0), \\ &\qquad m_1 + m_2, m_1 m_2 \mathbf{v}_D\rangle. \end{aligned} \tag{63}$$

Also,

$$\mathbf{v}_D = (0, 0, v_{D_z}) = \frac{\mathbf{p}_{12}}{M_{12}} = \frac{M_1 + M_2}{M_1 M_2}\,\mathbf{p}_1.$$

Since by Eqs. (48) and (43)

$$\begin{aligned}|[\nu_3 M_3 J_3]\mathbf{p}_3 m_3\rangle \cdots |[\nu_n M_n J_n]\mathbf{p}_n m_n\rangle &= \sum_{m_i', k} \sum_{J=0}^{\infty} \sum_{m=-J}^{J} N_J D^{J^*}_{km}(R) \\ &\quad \times \prod_{i=3}^{n} D^{J_i}_{m_i' m_i}(R)\, |[\nu MJ]\mathbf{p}m, m_i' k \mathbf{v}_{D_i}\rangle, \end{aligned} \tag{64}$$

the S-matrix element can be written

$$\sum_{m_i', k} \sum_J \sum_m \sum_{J'} N_J N_{J'} D^J_{km}(R) \prod_{i=3}^{n} D^{J_i^*}_{m_i' m_i}(R) \times \langle[\nu MJ]\mathbf{p}m, m_i' k \mathbf{v}_{D_i}|\, S\, |[\nu' M' J']\mathbf{p}' m', m_1 m_2 \mathbf{v}_D\rangle, \tag{65}$$

where $M' = M_1 + M_2$, $\mathbf{p}' = \mathbf{p}_1 + \mathbf{p}_2 = 0$, $m' = m_1 + m_2$. Now assume that the S matrix is invariant under arbitrary Galilean transformations. This allows one to write

$$\begin{aligned}\langle[\nu MJ]\mathbf{p}m, m_i' k \mathbf{v}_{D_i}|\, S\, |[\nu' M' J']\mathbf{p}' m', m_1 m_2 \mathbf{v}_D\rangle &= \delta^3(\mathbf{p} - \mathbf{p}')\delta(E - E')\delta(M - M')\delta_{JJ'}\delta_{m,m_1+m_2} \\ &\quad \times \langle[\nu MJ]\mathbf{p} = 0, m, m_i' k \mathbf{v}_{D_i}| \\ &\quad \times S\, |[\nu MJ]\mathbf{p} = 0, m, m_1 m_2 \mathbf{v}_D\rangle. \end{aligned} \tag{66}$$

Then,

$$\sum_{m_i'k}\sum_{J}\sum_{m}\sum_{J'} N_J^* N_{J'} D^J_{km}(R)\prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i}(R)$$
$$\times \delta^3(\mathbf{p}-\mathbf{p}')\delta(E-E')\delta(M-M')\delta_{JJ'}\delta_{m,m_1+m_2}$$
$$\times \langle [\nu MJ]\mathbf{p}=0, m, m_i' k \mathbf{v}_{D_i}|$$
$$\times S\,|[\nu MJ]\mathbf{p}=0, m, m_1 m_2 v_D\rangle$$
$$=\sum_{m_i',k}\sum_{J=m_1+m_2} N_J^2 D^J_{k,m_1+m_2}(R)$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i}(R) A_{(m_i',k,\mathbf{v}_{D_i},m_1,m_2,\mathbf{v}_D)}(\nu, M, J), \quad (67)$$

where

$$A_{(m_i',k,\mathbf{v}_{D_i},m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$
$$\equiv \langle [\nu MJ]\mathbf{p}=0, m_1+m_2, m_i', k, \mathbf{v}_{D_i}|$$
$$\times S\,|[\nu MJ]\mathbf{p}=0, m_1+m_2, m_1 m_2 \mathbf{v}_D\rangle. \quad (68)$$

The $\mathbf{v}_{D_i}$ can be expressed in terms of $\mathbf{p}_{ij}$ as shown previously, and $\mathbf{v}_D = [(M_1+M_2)/M_1M_2]\mathbf{p}_1$. Let μ denote the set of $\mathbf{v}_{D_i}$ in terms of $\mathbf{p}_{ij}$. Finally, the partial wave expansion becomes

$$\langle 3, 4, \cdots, n|\, S\,|1, 2\rangle$$
$$=\sum_{J=m_1+m_2}^{\infty}\sum_{m_i'k} N_J^2 D^J_{k,m_1+m_2}(R)$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i}(R) A_{(m_i'k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J), \quad (69)$$

where 3 refers to all labels associated with particle 3, etc.

As indicated previously, the choice of R has no absolute physical significance. If one wishes to express the partial wave expansion in terms of another rotation R', such that $R=\tilde{R}R'$, then the expansion becomes

$$\langle 3, 4, \cdots, n|\, S\,|1, 2\rangle$$
$$=\sum_{J}\sum_{m_i',k} N_J^2 D^J_{k,m_1+m_2}(R)$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i}(R) A_{(m_i',k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$
$$=\sum_{J}\sum_{m_i',k} N_J^2 D^J_{k,m_1+m_2}(\tilde{R}R')$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i}(\tilde{R}R') A_{(m_i',k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$
$$=\sum_{J}\sum_{m_i',k}\sum_{m',m_i''} N_J^2 D^J_{k,m'}(\tilde{R}) D^J_{m',m_1+m_2}(R')$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i''m_i}(R') D^{J_i^*}_{m_i''m_i}(R') A_{(m_i',k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$
$$=\sum_{J}\sum_{m',m_i''} N_J^2 D^J_{m',m_1+m_2}(R')\prod_{i=3}^{n} D^{J_i^*}_{m_i''m_i}(R')\sum_{k,m_i'} D^J_{k,m'}(\tilde{R})$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i''}(\tilde{R}) A_{(m_i',k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$
$$=\sum_{J}\sum_{m',m_i''} N_J^2 D^J_{m',m_1+m_2}(R')$$
$$\times \prod_{i=3}^{n} D^{J_i^*}_{m_i''m_i}(R')\tilde{A}_{(m_i'',m',\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J), \quad (70)$$

where

$$\tilde{A}_{(m_i'',m',\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$
$$=\sum_{k,m_i'} D^J_{km'}(\tilde{R})\prod_{i=3}^{n} D^{J_i^*}_{m_i'm_i''}(\tilde{R}) A_{(m_i',k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J). \quad (71)$$

VI. CONCLUSION

In this paper the Clebsch–Gordan coefficients that relate the n-fold tensor product of free-particle states to an over-all state have been calculated. Also, the continuous degeneracy parameters and the rotation that appears in the Clebsch–Gordan coefficients have been expressed in terms of invariants constructed from the momenta of the particles. Although the degeneracy parameters and the rotation constructed this way have no absolute significance in the sense that many other choices are possible, it is felt that the choice made here recommends itself through its easy identification with the momenta of the sytem.

Finally, the Clebsch–Gordan reduction has been applied to a scattering problem involving an arbitrary number of final particles. Using the assumption that the S matrix is a Galilean scalar, the maximum reduction of the scattering matrix allowed by Galilean invariance has been made. The quantities

$$A_{(m_i',k,\mu,m_1,m_2,\mathbf{v}_D)}(\nu, M, J)$$

contain the dynamics of the system and must be specified by experiment or a dynamical theory. It can be seen that the reduction derived here implies a clear distinction between quantities that can be specified on the basis of invariance under arbitrary Galilean transformations and quantities that must be determined by a particular dynamical model or experiment. It is to be noted that, because of the mass superselection rule, mass must be conserved in a multiparticle reaction. If mass differences caused by binding energy are neglected, the mass superselection rule follows from baryon number conservation, at least in the domain of low-energy nuclear physics. It is in this domain that the partial wave analysis [Eq. (67)] is relevant.

[1] V. Bargmann, Ann. Math. **59**, 1 (1954).
[2] J. M. Lévy-Leblond, J. Math. Phys. **4**, 776 (1961).
[3] J. Voisin, J. Math. Phys. **6**, 1519, 1822 (1965).
[4] G. W. Mackey, *The Theory of Group Representations* (University of Chicago Press, Chicago, 1955).
[5] A. J. MacFarlane, Rev. Mod. Phys. **34**, 41 (1962).
[6] W. H. Klink and G. J. Smith, Commun. Math. Phys. **10**, 231 (1968).
[7] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Co., Reading, Mass., 1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Evaluation of Certain Radial Coulomb Integrals Using Symmetry Properties of the Coulomb Field*

N. V. V. J. SWAMY AND R. G. KULKARNI
Oklahoma State University, Stillwater, Oklahoma 74074
AND
L. C. BIEDENHARN
Duke University, Durham, North Carolina 27706

(Received 2 May 1969)

The symmetry properties of the nonrelativistic Coulomb field problem allow one to construct an operator calculus for evaluating matrix elements of the multipole operator r^{-q}. By means of this operator calculus an explanation is given for the vanishing of certain radial integrals treated earlier by Pasternack and Sternheimer,as well as for the value of similar integrals occurring in Coulomb excitation.

I. INTRODUCTION

The $O(4)$ symmetry of the hydrogen atom has been of great interest as a model for high energy physics and the relevant literature is now quite large.[1] A knowledge of the dynamical symmetries of this quantum-mechanical problem, for which the exact solutions are known, has been of considerable importance in understanding the appearance of these symmetries in particle physics where the equations are not known.

It has been shown by Pasternack and Sternheimer[2] using direct evaluation that the following radial integrals vanish:

$$\int_0^\infty F_{N\,l+L}\frac{1}{r^q}F_{Nl}r^2\,dr \equiv \langle l+L|\frac{1}{r^q}|l\rangle = 0, \quad \text{for} \quad 2\le q\le L+1. \tag{1}$$

Here $F_{Nl}(r)$ is a nonrelativistic bound-state radial wavefunction of the hydrogen atom, normalized as usual:

$$\langle l+L|\frac{1}{r^q}|l\rangle = 1, \quad \text{when} \quad L=0=q. \tag{2}$$

It is the purpose of this note to point out that there is a physical interpretation of this result related to the $O(4)$ symmetry of the Coulomb field; in particular, the Pasternack–Sternheimer results are particular matrix elements of operator identities which are abstractly valid in general. The existence of this operator calculus is a direct consequence of the $O(4)$ symmetry. Our results allow of generalization to radial integrals wherein *continuum* functions $F_{\eta l}$ are used:

$$\langle l+L|\frac{1}{r^{L+1}}|l\rangle = \frac{2^L k_c^L\{\Gamma(L)\}^2}{\Gamma(2L)}\left|\frac{\Gamma(l+1+i\eta)}{\Gamma(l+1+L+i\eta)}\right|. \tag{3}$$

To understand these results most easily, it is convenient to work in the framework of a Pauli spin-$\frac{1}{2}$ particle in a Coulomb field. While this nonrelativistic spin (NRS) system has been elaborated elsewhere,[3] we shall briefly note down some of the key results needed for our operator derivation of the above equations. The spherical spinors (spin–angle functions), describing the motion of a Pauli spin-$\frac{1}{2}$ particle in a nonrelativistic Coulomb field, are defined as

$$\chi_x^\mu = \sum_\tau C(l(x),\tfrac{1}{2},j;\mu-\tau,\mu)Y_{l(x)}^{\mu-\tau}\chi_{\frac{1}{2}}^\tau, \tag{4}$$

and these satisfy the following equations:

$$(\boldsymbol{\sigma}\cdot\mathbf{L}+\mathbf{1})\chi_x^\mu \equiv \mathbf{K}_1\chi_x^\mu = -x\chi_x^\mu, \tag{5a}$$

$$\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\chi_x^\mu = -\chi_{-x}^\mu,$$

$$(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^2 = \mathbf{1}, \tag{5b}$$

$$(\chi_{x'}^{\mu'},\chi_x^\mu) = \delta_{xx'}\delta_{\mu\mu'}. \tag{5c}$$

Introducing the radial momentum operator

$$p_r = \tfrac{1}{2}(\hat{\mathbf{r}}\cdot\mathbf{p}+\mathbf{p}\cdot\hat{\mathbf{r}}) \leftrightarrow -i\left(\frac{\partial}{\partial r}+\frac{1}{r}\right), \tag{6a}$$

$$[r,p_r] = i, \tag{6b}$$

we write the nonrelativistic Hamiltonian as

$$\mathbf{H} = \frac{\mathbf{p}^2}{2m}-\frac{\alpha z}{r} \tag{7}$$

$$= \frac{p_r^2}{2m}+\frac{(\mathbf{K}_1^2-\mathbf{K}_1)}{2mr^2}-\frac{\alpha z}{r}. \tag{8}$$

Here α is the Sommerfeld fine-structure constant and we choose $\hbar = c = 1$. We readily notice

$$p_r = i^{-1}[r, m\mathbf{H}]; \quad [\mathbf{K}_1,\mathbf{H}] = 0. \tag{9}$$

The functions $\chi_x^\mu F_{Nl} = \Psi_{Nx\mu}$ are the bound-state eigenfunctions.

II. BOUND STATES

The vector-invariant characteristic of the $O(4)$ group, to which the nonrelativistic Coulomb field belongs, is the Runge–Lenz–Pauli vector $\mathcal{A}$.[3] Since the radial matrix elements discussed here pertain to a particular subspace of the Hilbert space corresponding to the principal quantum number N, we shall choose the invariant pseudoscalar "Coulomb helicity operator" as

$$\boldsymbol{\sigma}\cdot\mathcal{A} = k_b^{-2}\boldsymbol{\sigma}\cdot\mathbf{r}(\alpha z m + ip_r\mathbf{K}_1 - r^{-1}\mathbf{K}_1^2). \quad (10)$$

As discussed elsewhere,[3] it can be shown that

$$\boldsymbol{\sigma}\cdot\mathcal{A}\Psi_{Nx\mu} = -ik_b^{-1}\mathcal{A}_x\Psi_{N-x\mu}, \quad (11)$$

where we define

$$\begin{aligned}\Psi_{N-x\mu} &= F_{Nl-x}(i\chi^{\mu}_{-x}),\\ \mathcal{A}_x &\equiv |(N^2-x^2)^{\frac{1}{2}}|,\\ k_b &= \alpha z m/N.\end{aligned} \quad (12)$$

As a preliminary to understanding the Pasternack–Sternheimer result, we shall give some important operator relationships. A good number of other relations have also been of use, but we do not enumerate them here either because they are too obvious or they can easily be built from $\boldsymbol{\sigma}\cdot\mathcal{A}$, $\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}$, $\hat{\mathbf{z}}\cdot\boldsymbol{\sigma}\times\mathbf{L}$, and $\mathbf{K}_1$. With the help of the quantum condition $[r, p_r] = i$, the following are established by direct operator manipulation:

$$[r^{-q}, ip_r] = qr^{-(q+1)}, \quad q = \text{integer}, \quad (13)$$

$$r^{-q}ip_r = [2(q-1)]^{-1}[\boldsymbol{\beta}r^{-(q-1)}] + q(2r^{q+1})^{-1}, \quad q \neq 1, \quad (14)$$

where we have defined

$$\boldsymbol{\beta} \equiv 2m\mathbf{H} \quad (14')$$

and

$$[\boldsymbol{\sigma}\cdot\mathcal{A}, r^{-q}] = qk_b^{-2}\mathbf{K}_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-(q+1)}, \quad (15)$$

$$\begin{aligned}&\tfrac{1}{2}[r^{-q}[r, \tfrac{1}{2}\boldsymbol{\beta}], \boldsymbol{\beta}] + \tfrac{1}{4}q[r^{-(q+1)}, \boldsymbol{\beta}]\\ &= qr^{-(q+1)}\boldsymbol{\beta} + \alpha z m(2q+1)r^{-(q+2)}\\ &\quad + (q+1)\{\tfrac{1}{4}q(q+2) + (\mathbf{K}_1 - \mathbf{K}_1^2)\}r^{-(q+3)}.\end{aligned} \quad (16)$$

It is interesting to note that, by taking the expectation value of both sides of Eq. (16) with respect to the basis $\Psi_{Nx\mu}$, one gets the recursion relation

$$\begin{aligned}&q(l+\tfrac{1}{2}+\tfrac{1}{2}q)(l+\tfrac{1}{2}-\tfrac{1}{2}q)\langle l|\, r^{-(q+2)}\,|l\rangle\\ &= \alpha z m(2q-1)\langle l|\, r^{-(q+1)}\,|l\rangle - k_b^2(q-1)\langle l|\, r^{-q}\,|l\rangle,\\ &\qquad q = \text{integer or } 0.\end{aligned} \quad (17)$$

This was also shown by Pasternack[4] by a direct evaluation of the radial integrals. The invariant operator $\mathbf{K}_1$ anticommutes with $\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}$ and $\boldsymbol{\sigma}\cdot\mathcal{A}$. Using this property, we easily establish the following operator equations:

$$[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}] = 2\mathbf{K}_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-2}, \quad (18)$$

$$\begin{aligned}&\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}[2(q-1)]^{-1}[\boldsymbol{\beta}, \boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-(q-1)}]\\ &\quad = r^{-q}ip_r + (\mathbf{K}_1(q-1)^{-1} - \tfrac{1}{2}q)r^{-(q+1)}, \quad q \neq 1,\end{aligned} \quad (19)$$

$$\begin{aligned}&[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-(q-1)}ip_r, \boldsymbol{\beta}]\\ &\quad + [q(q-1) + 2\mathbf{K}_1](2q)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-q}, \boldsymbol{\beta}]\\ &\quad = \boldsymbol{\sigma}\cdot\hat{\mathbf{r}}[2(q-1)r^{-q}\boldsymbol{\beta} + 2\alpha z m(2q-1)r^{-(q+1)}\\ &\quad + (2q)^{-1}(q^2-1)(q^2-4\mathbf{K}_1^2)r^{-(q+2)}].\end{aligned} \quad (20)$$

Taking the matrix element ($\Psi_{N-x\mu}$ [Eq. (20)] $\Psi_{Nx\mu}$), we obtain the following recursion relation which plays an important role in the rest of our paper:

$$\begin{aligned}&q(q-2)(l+\tfrac{1}{2}+\tfrac{1}{2}q)(l+\tfrac{3}{2}-\tfrac{1}{2}q)\langle l+1|\, r^{-(q+1)}\,|l\rangle\\ &\quad - \alpha z m(q-1)(2q-3)\langle l+1|\, r^{-q}\,|l\rangle\\ &\quad + k_b^2(q-2)(q-1)\langle l+1|\, r^{-(q-1)}\,|l\rangle = 0,\\ &\qquad q = \text{positive integer or } 0.\end{aligned} \quad (21)$$

The angular operator $\hat{\mathbf{z}}\cdot\boldsymbol{\sigma}\times\mathbf{L} \equiv \Omega_0$ changes x to $-x-1$ without affecting the radial part of the function $\Psi_{Nx\mu}$, and it commutes with $\mathbf{H}$ and anticommutes with $(2\mathbf{K}_1 - 1)$. As has been shown elsewhere,[3] the operators $\boldsymbol{\sigma}\cdot\mathcal{A}\Omega_0$ and $\Omega_0\boldsymbol{\sigma}\cdot\mathcal{A}$ raise or lower the l value (more precisely the x value) of the wavefunctions $\Psi_{Nx\mu}$, respectively, whereas the angular operators $\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0$ and $\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}$ preserve the l value of the radial functions while changing x to $x+1$ and $x-1$, respectively. An alternative derivation of the Pasternack–Sternheimer result, using certain recursion relations between the radial function and a few other interesting numerical relationships between expectation values,etc., will be discussed in the Appendices A and B.

We shall now proceed with the proof for the vanishing of $\langle l+1|\, r^{-2}\,|l\rangle$. Consider the matrix element

$$(\Psi_{N-x-2\mu}, (\boldsymbol{\sigma}\cdot\mathcal{A}\Omega_0)^0(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})r^{-2}\boldsymbol{\sigma}\cdot\mathcal{A}\Psi_{Nx+1\mu}). \quad (22)$$

If we take x positive, then it is easy to see[3] that the above leads to

$$\begin{aligned}&[(l+2)(2l+3)]^{\frac{1}{2}}C(l+\tfrac{3}{2}, 1, l+\tfrac{1}{2}; \mu, 0, \mu)\\ &\qquad \times \mathcal{A}_{l+1}k_b^{-2}\langle l+1|\, r^{-2}\,|l\rangle.\end{aligned} \quad (23)$$

On the other hand, the operator

$$\begin{aligned}&\Omega_0 r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\boldsymbol{\sigma}\cdot\mathcal{A}\\ &\quad = k_b^{-2}\Omega_0 r^{-2}(\alpha z m + ip_r\mathbf{K}_1 - r^{-1}\mathbf{K}_1^2) \quad (24)\\ &\quad = r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\boldsymbol{\sigma}\cdot\mathcal{A}\Omega_0 + k_b^{-2}(r^{-2}ip_r - r^{-3})(1-2\mathbf{K}_1)\Omega_0, \quad (25)\end{aligned}$$

where we make use of the commutation relations between Ω_0 and $\mathbf{K}_1$, r^{-q} and ip_r. By means of Eqs. (14), (14′), and (20), we see that

$$r^{-2}ip_r - r^{-3} = \tfrac{1}{2}[\boldsymbol{\beta}, r^{-1}], \tag{26}$$

$$\begin{aligned}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-2} &= (2\alpha zm)^{-1}\{[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}ip_r, \boldsymbol{\beta}] + [\mathbf{K}_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-1}, \boldsymbol{\beta}]\}\\ &= (2\alpha zm)^{-1}[\tfrac{1}{2}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \beta], \boldsymbol{\beta}].\end{aligned} \tag{27}$$

The operator under consideration, therefore, is

$$\begin{aligned}&\left\{\frac{1}{2}\left[\boldsymbol{\beta}, \frac{1}{r}\right](1 - 2\mathbf{K}_1)\right.\\ &\quad\left. + \frac{1}{2\alpha zm}\left[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\left(ip_r - \frac{\mathbf{K}_1}{r}\right), \boldsymbol{\beta}\right]\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\right\}\Omega_0\\ &= \left\{\frac{1}{2}\left[\boldsymbol{\beta}, \frac{1}{r}\right](1 - 2\mathbf{K}_1) + \frac{1}{2\alpha zm}[\tfrac{1}{2}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}], \boldsymbol{\beta}]\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\right\}\Omega_0\end{aligned} \tag{28}$$

of the form of a commutator with the Hamiltonian. Hence, the matrix element vanishes between states of the same energy and we have

$$\langle l + 1|\, r^{-2}\,|l\rangle = 0.$$

We can now generalize this method by considering the operator

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^1(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^2 r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}; \tag{29}$$

this leads to the radial matrix element $\langle l + 2|\, r^{-q}\,|l\rangle$, when the matrix elements of the operator are computed in the same basis as before. Now let us consider

$$\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}. \tag{30}$$

Since Ω_0 commutes with the Hamiltonian, we can factor out Ω_0^2 by using the commutation relation

$$[\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}, \Omega_0^2] = 2\mathbf{K}_1\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}} \tag{31}$$

and also

$$\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} \equiv \tilde{\omega} = \Omega_0^2 + 2\mathbf{K}_1. \tag{32}$$

The operator can now be rewritten as

$$\tilde{\omega}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0 r^{-q}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}. \tag{33}$$

The procedure hereafter consists of the following general steps: We expand $\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}$ according to Eq. (10), take Ω_0 to the extreme right, replace $r^{-q}ip_r$ or $\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-q}ip_r$ by appropriate commutators with the Hamiltonian [Eqs. (14) and (19), or equivalents], and reduce the operator, after picking out commutators with $\mathbf{H}$, to a sum of terms like (const) r^{-q}. Following this general procedure, we see that by straightforward algebra the operator in Eq. (33) reduces to

$$\begin{aligned}&\frac{\tilde{\omega}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}}{k_b^2}\left\{\frac{1}{2(q-1)}\left[\boldsymbol{\beta}, \frac{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}}{r^{q-1}}\right](1 - \mathbf{K}_1)\right.\\ &\quad + \boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\left(\frac{\alpha zm}{r^q} + \tfrac{1}{2}[q-2]\right.\\ &\quad\left.\left.\times\left[1 - \frac{q-3}{q-1}\mathbf{K}_1 - \frac{2}{q-1}\mathbf{K}_1^2\right]\frac{1}{r^{q+1}}\right)\right\}\Omega_0.\end{aligned} \tag{34}$$

When $q = 3$, the matrix element becomes

$$\begin{aligned}&\alpha zm\,\langle l + 2|\, r^{-3}\,|l + 1\rangle\\ &\quad - \tfrac{1}{2}(l+1)(l+3)\,\langle l + 2|\, r^{-4}\,|l + 1\rangle,\end{aligned} \tag{34′}$$

which vanishes by virtue of Eq. (21). The operator in Eq. (34), for $q = 3$, finally becomes

$$\begin{aligned}&\tilde{\omega}\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}}{k_b^2}\left\{\frac{1}{4}\left[\boldsymbol{\beta}, \frac{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}}{r^2}\right](1 - \mathbf{K}_1) + \frac{1}{6}\left[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\frac{1}{r}ip_r, \boldsymbol{\beta}\right]\right.\\ &\quad + \frac{\mathbf{K}_1 + 1}{12}\left[\frac{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}}{r^2}, \boldsymbol{\beta}\right]\\ &\quad\left. - \frac{1}{6\alpha zm}\left([\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}ip_r, \boldsymbol{\beta}] + \mathbf{K}_1\left[\frac{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}}{r}, \boldsymbol{\beta}\right]\right)\boldsymbol{\beta}\right\}\Omega_0.\end{aligned} \tag{35}$$

We notice that the system of operator terms is rather complicated; but each one is a commutator with the Hamiltonian and the matrix element vanishes. When $q = 2$, we have a more trivial case wherein the operator degenerates into

$$\begin{aligned}&\frac{\tilde{\omega}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}}{2k_b^2}\left\{\left[\boldsymbol{\beta}, \frac{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}}{r}\right](1 - \mathbf{K}_1) + [\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}ip_r, \boldsymbol{\beta}]\right.\\ &\quad\left. + \mathbf{K}_1\left[\frac{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}}{r}, \boldsymbol{\beta}\right]\right\}\Omega_0.\end{aligned} \tag{36}$$

We thus see that $\langle l + 2|\, r^{-q}\,|l\rangle = 0$, when $q = 2, 3$. We shall give, without detailed proof, the results for $\langle l + 3|\, r^{-q}\,|l\rangle = 0$, when $q = 2,\ 3,\ 4$. This radial integral stems from the matrix element

$$(\Psi_{N\,-x-2\,\mu},\ (\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^2(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^3 r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Psi_{N\,x+1\,\mu}). \tag{37}$$

$(\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^2(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^3 r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}$ happens to be

$$\tilde{\omega}'\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\sum_i \mathcal{O}_{qi}\Omega_0,$$

where

$$\tilde{\omega}' = \boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\tilde{\omega}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}$$

and the $\mathcal{O}_{qi}$ are tabulated explicitly in Table I. Before discussing the generalization of this, we shall derive similar results for the continuum states.

TABLE I. The operators $\mathcal{O}_{qi}$.

q[a]	i	Operator
q	1	$-[k_b^2 2(q-1)]^{-1}\boldsymbol{\sigma}\cdot\mathcal{A}[r^{-(q-1)},\beta]\mathbf{K}_1$
q	2	$[k_b^2 2(q-1)]^{-1}\boldsymbol{\sigma}\cdot\mathcal{A}[r^{-(q-1)},\beta](2\mathbf{K}_1-1)$
q	3	$-[k_b^2 2(q-1)]^{-1}[r^{-(q-1)},\beta](2\mathbf{K}_1-1)\boldsymbol{\sigma}\cdot\mathcal{A}$
q	4	$-[k_b^4 2(q-1)]^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-(q-1)}ip_r,\beta](4\mathbf{K}_1^2-1)$
q	5	$\frac{1}{4}k_b^{-4}[r^{-q}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta](4\mathbf{K}_1^2-1)$
q	6	$-[k_b^4 2q(q-1)]^{-1}\mathbf{K}_1[r^{-q}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta](4\mathbf{K}_1^2-1)$
q	7	$[k_b^4 2(q-1)]^{-1}[r^{-(q-2)}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]\mathbf{K}_1(\alpha zm)$
q	8	$(k_b^4 2q)^{-1}[r^{-(q-1)}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]\mathbf{K}_1(q-2+\frac{1}{2}q\mathbf{K}_1-\mathbf{K}_1)$
2	9	$(\mathbf{K}_1^2-1)(6k_b^4)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-1}ip_r,\beta]$
2	10	$(1+\mathbf{K}_1)(\mathbf{K}_1^2-1)(12k_b^4)^{-1}[r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]$
2	11	$(\alpha zm)^2k_b^{-4}[(2\alpha zm)^{-1}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(ip_r-\mathbf{K}_1r^{-1}),\beta]$
2	12	$(1-\mathbf{K}_1^2)(3k_b^4)^{-1}[(2\alpha zm)^{-1}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(ip_r-r^{-1}\mathbf{K}_1),\beta]\beta$
3	9	$\alpha zm(6k_b^4)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-1}ip_r,\beta]$
3	10	$\alpha zm(12k_b^4)^{-1}(1+\mathbf{K}_1)[r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]$
4	9	$\frac{1}{30}(1-\mathbf{K}_1^2)[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-3}ip_r,\beta]$
4	10	$\frac{1}{120}(1-\mathbf{K}_1^2)(6+\mathbf{K}_1)[r^{-4}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]$
4	11	$\frac{1}{10}\alpha zm[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-2}ip_r,\beta]$
4	12	$\frac{1}{30}\alpha zm(3+\mathbf{K}_1)[r^{-3}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]$
4	13	$-\frac{1}{15}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-1}ip_r,\beta]\beta$
4	14	$-\frac{1}{30}(1+\mathbf{K}_1)[r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]\beta$
4	15	$(15\alpha zm)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}ip_r,\beta]\beta^2$
4	16	$(15\alpha zm)^{-1}\mathbf{K}_1[r^{-1}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}},\beta]\beta^2$
3	11	$-\alpha zm(3k_b^4)^{-1}[(2\alpha zm)^{-1}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(ip_r-r^{-1}\mathbf{K}_1),\beta]\beta$

[a] The first eight operators are common to the three cases with appropriate value of q in each case.

III. CONTINUUM STATES

The radial part of the continuum solutions of the Hamiltonian is derived from the bound-state function by analytic continuation, characterized by the quantum number N going over into $-i\eta$ where η is any positive number, not necessarily an integer. The continuum functions are normalized such that

$$\int_0^\infty F_{\eta l}F_{\eta' l}r^2\,dr=\delta(\eta-\eta'). \tag{38}$$

Introducing the wavenumber $k_c=\alpha zm/\eta$ and appropriately modifying the operator, we replace Eq. (11) by

$$\boldsymbol{\sigma}\cdot\mathcal{A}\Psi_{\eta\varkappa\mu}=-\frac{i}{k_c}|\varkappa+i\eta|\,\Psi_{\eta-\varkappa\mu}\equiv-\frac{i}{k_c}\mathcal{A}_\varkappa\Psi_{\eta-\varkappa\mu}. \tag{39}$$

It is important to note that, while the operator relationships derived in the earlier section are valid even here, some caution has to be exercised in taking matrix elements, since some of the radial integrals are likely to be singular.[5]

To begin with, let us consider the matrix element

$$(\Psi_{\eta-\varkappa-2\mu},(\boldsymbol{\sigma}\cdot\mathcal{A}\Omega_0)^0\{(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^1r^{-2}\boldsymbol{\sigma}\cdot\mathcal{A}-\text{adj}\}\Psi_{\eta\,\varkappa+1\,\mu}). \tag{40}$$

This matrix element is evaluated to be

$$i\mathcal{A}_{l+1}[(l+2)(2l+3)]^{\frac{1}{2}}C(l+\tfrac{3}{2},1,l+\tfrac{1}{2};\mu,0,\mu)\times\left(\frac{\mathcal{A}_{l+1}}{k_c}\langle l+1|\frac{1}{r^2}|l\rangle-\frac{\mathcal{A}_{l+2}}{k_c}\langle l+2|\frac{1}{r^2}|l+1\rangle\right) \tag{41}$$

$$=(\text{const})\left(\frac{\mathcal{A}_{l+1}}{k_c}\langle l+1|\frac{1}{r^2}|l\rangle-\frac{\mathcal{A}_{l+2}}{k_c}\langle l+2|\frac{1}{r^2}|l+1\rangle\right). \tag{42}$$

Using the commutation relations specified earlier, we can show that

$$\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-2}\boldsymbol{\sigma}\cdot\mathcal{A}-\boldsymbol{\sigma}\cdot\mathcal{A}r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0=(r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\boldsymbol{\sigma}\cdot\mathcal{A}-\boldsymbol{\sigma}\cdot\mathcal{A}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-2})\Omega_0-k_c^{-2}(r^{-2}ip_r-r^{-3})(2\mathbf{K}_1-1)\Omega_0 \tag{43}$$

$$=k_c^{-2}[r^{-1},\beta](\mathbf{K}_1-\tfrac{3}{2})\Omega_0 \tag{44}$$

$$=k_c^{-2}[r^{-1},\beta](\mathbf{K}_1-\tfrac{3}{2})\Omega_0. \tag{45}$$

The matrix element, therefore, vanishes and we get the difference equation

$$\frac{\mathcal{A}_{l+1}}{k_c}\langle l+1|\frac{1}{r^2}|l\rangle-\frac{\mathcal{A}_{l+2}}{k_c}\langle l+2|\frac{1}{r^2}|l+1\rangle=0, \tag{46}$$

which has a solution

$$\langle l+1|\frac{1}{r^2}|l\rangle=(\text{const})\frac{k_c}{|l+1+i\eta|}. \tag{47}$$

To evaluate $\langle l+2|r^{-3}|l\rangle$, we examine the matrix element

$$(\Psi_{\eta\,-\varkappa-2\,\mu},(\boldsymbol{\sigma}\cdot\mathcal{A}\Omega_0)^1[(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^2r^{-3}\boldsymbol{\sigma}\cdot\mathcal{A}-\text{adj}]\Psi_{\eta\,\varkappa+1\,\mu}),$$

which evaluates to

$$(\text{const})\left(\frac{\mathcal{A}_{l+2}\mathcal{A}_{l+1}}{k_c^2}\langle l+2|\frac{1}{r^3}|l\rangle-\frac{\mathcal{A}_{l+3}\mathcal{A}_{l+2}}{k_c^2}\langle l+3|\frac{1}{r^3}|l+1\rangle\right). \tag{48}$$

On the other hand, the operator reduces to a commutator with the Hamiltonian, as the following brief

steps indicate:

$$\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-3}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}} - \boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}r^{-3}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0 \tag{49}$$

$$= \tilde{\boldsymbol{\omega}}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-3}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}} - \boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0[\Omega_0\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-3} + k_c^{-2}(2r^{-4} - r^{-3}ip_r) \times \Omega_0(1 - 2\mathbf{K}_1)]\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0 \tag{50}$$

$$= \tilde{\boldsymbol{\omega}}\{\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}[r^{-3}, \boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}] + k_c^{-2}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}[2r^{-3}ip_r - r^{-4}(3 + 2\mathbf{K}_1)]\}\Omega_0 \tag{51}$$

$$= \tilde{\boldsymbol{\omega}}2k_c^{-2}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}[r^{-q}ip_r + \tfrac{1}{2}(\mathbf{K}_1 - 3)r^{-4}]\Omega_0 \tag{52}$$

$$= 2k_c^{-2}\tilde{\boldsymbol{\omega}}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}[\boldsymbol{\beta}, r^{-2}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}]\Omega_0. \tag{53}$$

In the above steps, Eqs. (16) and (19) have been used. The matrix element vanishes and we get, as before,

$$\langle l + 2|\frac{1}{r^3}|l\rangle = (\text{const})\frac{k_c^2}{|l + 1 + i\eta|\,|l + 2 + i\eta|}. \tag{54}$$

We shall give, without proof, the result for

$$\langle l + 3|\, r^{-4}\,|l\rangle.$$

The matrix element

$$(\Psi_{\eta\,-x-2\,\mu}(\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^2[(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^3r^{-4}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}} - \text{adj}]\Psi_{\eta\,x+1\,\mu}) \tag{55}$$

yields

$$(\text{const})\left\{\frac{\mathcal{A}_{l+1}\mathcal{A}_{l+2}\mathcal{A}_{l+3}}{k_c^3}\langle l + 3|\frac{1}{r^4}|l\rangle - \frac{\mathcal{A}_{l+4}\mathcal{A}_{l+3}\mathcal{A}_{l+2}}{k_c^3}\langle l + 4|\frac{1}{r^4}|l + 1\rangle\right\}. \tag{56}$$

It can be shown that the operator is equal to

$$\boldsymbol{\omega}'\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\sum_i \mathcal{O}_{qi}^c\Omega_0, \tag{57}$$

where the $\mathcal{O}_{qi}^c$'s are given in Table II.

TABLE II. The operators $\mathcal{O}_{qi}^c$.

q	i	Operator
4	1	$-k_c^{-2}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\frac{1}{6}[r^{-3}, \boldsymbol{\beta}]\mathbf{K}_1$
4	2	$k_c^{-2}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\frac{1}{6}[r^{-3}, \boldsymbol{\beta}](2\mathbf{K}_1 - 1)$
4	3	$k_c^{-2}\frac{1}{6}[r^{-3}, \boldsymbol{\beta}]\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}(2\mathbf{K}_1 + 1)$
4	4	$(6k_c^4)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-3}ip_r, \boldsymbol{\beta}](1 - 4\mathbf{K}_1^2)$
4	5	$(4k_c^4)^{-1}[r^{-4}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}](4\mathbf{K}_1^2 - 1)$
4	6	$(24k_c^4)^{-1}[r^{-4}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}](1 - 4\mathbf{K}_1^2)$
4	7	$(6k_c^2)^{-1}\mathbf{K}_1\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}[r^{-3}, \boldsymbol{\beta}]$
4	8	$(6k_c^2)^{-1}(2\mathbf{K}_1 - 3)\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}[r^{-3}, \boldsymbol{\beta}]$
4	9	$(6k_c^4)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-4}ip_r, \boldsymbol{\beta}](3 + 2\mathbf{K}_1)(2\mathbf{K}_1 + 1)$
4	10	$(24k_c^4)^{-1}(6 + \mathbf{K}_1)[r^{-4}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}](3 + 2\mathbf{K}_1)(2\mathbf{K}_1 + 1)$
4	11	$-3(4k_c^4)^{-1}[r^{-4}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}](2\mathbf{K}_1 + 3)(2\mathbf{K}_1 + 1)$
4	12	$(6k_c^4)^{-1}[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-3}ip_r, \boldsymbol{\beta}](2\mathbf{K}_1 + 1)$
4	13	$(6k_c^4)^{-1}\frac{1}{4}(\mathbf{K}_1 + 6)[r^{-4}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\beta}](2\mathbf{K}_1 + 1)$

Thus, we see that this combination of commutators with $\mathbf{H}$ renders the matrix element in Eq. (55) *zero*, and we evaluate as follows:

$$\langle l + 3|\frac{1}{r^4}|l\rangle = (\text{const})\frac{k_c^3}{|l + 1 + i\eta|\,|l + 2 + i\eta|\,|l + 3 + i\eta|}. \tag{58}$$

IV. DISCUSSION AND CONCLUSION

The general result for bound states arises from the matrix element

$$(\Psi_{N\,-x-2\,\mu}, (\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^{L-1}(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^L r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Psi_{N\,x+1\,\mu}), \tag{59}$$

and the parallel result for the continuum functions from the matrix element

$$(\Psi_{\eta\,-x-2\,\mu}, (\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^{L-1} \times [(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^L r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}} - \text{adj}]\Psi_{\eta\,x+1\,\mu}). \tag{60}$$

The extreme complexity of the resolution of the appropriate operators into a sum of commutators with the Hamiltonian noticed in $\langle l + 3|\, r^{-q}\,|l\rangle$, for bound as well as continuum states, shows that the general result would be too complicated to write down. However, it should be established in any particular case because the procedure is straightforward. For instance, we notice first that the invariant operator that multiplies the whole sum of commutators with Hamiltonian in the cases studied is seen to follow a systematic rule:

When the operator in the second parenthesis multiplying the invariant Ω_0 is	*The operator multiplying the entire sum is*
$(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\mathbf{1})$	$\mathbf{1} = (\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\mathbf{1})(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\mathbf{1})$
$(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})$	$\tilde{\boldsymbol{\omega}} \equiv \boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} = (\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})$
$\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^2$	$\tilde{\boldsymbol{\omega}}' \equiv \boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0\tilde{\boldsymbol{\omega}}\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} = [\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^2](\text{adj})$
.	.
.	.
.	.
$\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^{L-1}$	$[\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^{L-1}](\text{adj}).$ (61)

Secondly, Ω_0 happens to multiply the whole sum to the right. Thirdly, one of the important intermediate steps is to manipulate

$$\Omega_0\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-q} \quad \text{and} \quad \boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}r^{-q}\Omega_0 \tag{62}$$

such that Ω_0 can be moved to the right or left as required. And at the penultimate stage, one gets a linear combination of recursion relations mentioned in Eq. (21).

The essential point to be noted in our demonstration here is that the symmetry of the Coulomb field permits the construction of a variety of angular operators,[6] the matrix elements of which lead to interesting relationships among radial integrals. This is also reflected in the continuum where the "l degeneracy" is infinite. It has been proved elsewhere[5] that the zero-energy-loss limit (which means confining oneself to the subspace of Hilbert space belonging to a particular principal quantum number) simplifies the radial integrals considerably. Symmetry helps in the most convenient, though not the most general, evaluation of such integrals. The existence of simple recursion relations for monopole matrix elements, an exact quantum mechanical result in closed form for the total radiation loss in dipole bremsstrahlung, are specific instances in point. Furthermore, relativistic analogs of the operators $\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\Omega_0$ and $\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0$ exist in the case of the "symmetric Hamiltonian."[2] Thus, the vanishing of the matrix element

$$((\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}}\Omega_0)^{L-1}(\Omega_0\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})^L r^{-q}\boldsymbol{\sigma}\cdot\boldsymbol{\mathcal{A}})$$

ought to be expected even when the relativistic basis functions $\Psi_{Nx\mu}$ are used (this is why we departed slightly from the notation of Ref. 3 when referring to nonrelativistic spinors $|Nx\mu\rangle$).

ACKNOWLEDGMENT

One of us (R. G. K.) is thankful to the Department of Atomic Energy, Government of India, for financial support during the early phases of this work.

APPENDIX A

An alternative step-by-step derivation of the Pasternack–Sternheimer result is given here. This procedure exploits certain contiguous relations between radial functions F_{Nl} and an initial recurrence relation between matrix elements, which is established with the help of Eqs. (10) and (14) as follows:

$$F_{Nl+2} = \frac{-(l+2)\mathcal{A}_{l+1}}{(l+1)\mathcal{A}_{l+2}}F_{Nl} + \left(\frac{N(l+2)(2l+3)}{\alpha z m\mathcal{A}_{l+2}}\frac{1}{r} - \frac{N(2l+3)}{(l+1)\mathcal{A}_{l+2}}\right)F_{Nl+1}, \tag{A1}$$

$$F_{Nl+L} = -\frac{(l+L)\mathcal{A}_{l+L-1}}{(l+L-1)\mathcal{A}_{l+L}}F_{Nl+L-2} + \left(\frac{N(l+L)(2l+2L-1)}{\alpha z m\mathcal{A}_{l+L}}\frac{1}{r} - \frac{N(2l+2L-1)}{(l+L-1)\mathcal{A}_{l+L}}\right)F_{Nl+L-1}, \tag{A2}$$

$$\langle l+1|\frac{1}{r^q}|l\rangle = \frac{N}{\alpha z m}\frac{(l+1)}{\mathcal{A}_{l+1}}\left((l+1-\tfrac{1}{2}q)\left\langle\frac{1}{r^{q+1}}\right\rangle - \frac{\alpha z m}{l+1}\left\langle\frac{1}{r^q}\right\rangle\right); \tag{A3}$$

and this leads to

$$\langle l+L|r^{-q}|l\rangle = \sum_{i=1}^{q+1}\mathcal{C}_{qi}\langle r^{-q+1-i}\rangle,$$

$$\langle r^s\rangle \equiv \langle Nlm|r^s|Nlm\rangle \equiv \langle l|r^s|l\rangle. \tag{A4}$$

The expectation values on the right-hand side can be handled via Eq. (17). The values of the coefficients $\mathcal{C}_{qi}$ for the cases studied here are listed below:

$$\mathcal{C}_{11} = -\frac{N}{\mathcal{A}_{l+1}}, \quad \mathcal{C}_{12} = \frac{N(l+1)(l+1-\frac{1}{2}q)}{\alpha z m\mathcal{A}_{l+1}},$$

$$\mathcal{C}_{21} = \frac{N^2+(l+1)(l+2)}{\mathcal{A}_{l+1}\mathcal{A}_{l+2}},$$

$$\mathcal{C}_{22} = -\frac{N^2(2l+3)(2l+3-\frac{1}{2}q)}{\alpha z m\mathcal{A}_{l+1}\mathcal{A}_{l+2}},$$

$$\mathcal{C}_{23} = \frac{N^2(l+1)(l+2)(2l+3)(l+\frac{1}{2}-\frac{1}{2}q)}{(\alpha z m)^2\mathcal{A}_{l+1}\mathcal{A}_{l+2}},$$

$$\mathcal{C}_{31} = -\frac{1}{\mathcal{A}_{l+1}\mathcal{A}_{l+2}\mathcal{A}_{l+3}} \times\left[N^3\left(\frac{l+3}{l+1}\right)+N(3l^2+10l+9)\right],$$

$$\mathcal{C}_{32} = \frac{1}{\alpha z m\mathcal{A}_{l+1}\mathcal{A}_{l+2}\mathcal{A}_{l+3}} \times\{9N^3(l+2)^2+3N(l+1)(l+2)^2(l+3) - \tfrac{1}{2}q[3N^3(l+2)+N(l+1)(l+2)(l+3)]\},$$

$$\mathcal{C}_{33} = -\frac{N^3(l+2)(2l+3)(2l+5)(3l+4-q)}{(\alpha z m)^2\mathcal{A}_{l+1}\mathcal{A}_{l+2}\mathcal{A}_{l+3}},$$

$$\mathcal{C}_{34} = \frac{N^3(l+1)(l+2)(l+3)(2l+3)(2l+5)(l-\frac{1}{2}q)}{(\alpha z m)^3\mathcal{A}_{l+1}\mathcal{A}_{l+2}\mathcal{A}_{l+3}}.$$

Combining Eqs. (A1)–(A4), one can build the Pasternack–Sternheimer result for any higher case starting from the lowest result $\langle l+1|r^{-2}|l\rangle = 0$, because,as one goes to higher values of q,one can take advantage of all the earlier vanishing results.

APPENDIX B

A number of numerical relationships (for bound states) are listed here since they can be useful in practical applications; it is to be noted, however, that each relation has to be applied with caution only within its region of validity, since the existence of the angular-momentum quantum number l in the denominator makes certain matrix elements singular:

$$\left\langle \frac{1}{r^2} \right\rangle \equiv \langle l| \frac{1}{r^2} |l\rangle = \frac{k_b}{l+\frac{1}{2}} \left\langle \frac{1}{r} \right\rangle,$$

$$\left\langle \frac{1}{r^3} \right\rangle = \frac{\alpha z m}{l(l+1)} \left\langle \frac{1}{r^2} \right\rangle,$$

$$\left\langle \frac{1}{r^4} \right\rangle = \frac{k_b^2}{\alpha z m} \left\{ \frac{3N^2 - l(l+1)}{2(l-\frac{1}{2})(l+\frac{3}{2})} \right\} \left\langle \frac{1}{r^3} \right\rangle,$$

$$\left\langle \frac{1}{r^5} \right\rangle = \alpha z m \frac{5N^2 - 3l^2 - 3l + 1}{[3N^2 - l(l+1)](l+2)(l-1)} \left\langle \frac{1}{r^4} \right\rangle,$$

$$\left\langle \frac{1}{r^6} \right\rangle = \frac{k_b^2}{4\alpha z m} \left(\frac{35N^4 - 5N^2(6l^2 + 6l - 5) + 3(l-1)(l)(l+1)(l+2)}{(5N^2 - 3l^2 - 3l + 1)(l-\frac{3}{2})(l+\frac{5}{2})} \right) \left\langle \frac{1}{r^5} \right\rangle,$$

$$\langle l+1| \frac{1}{r} |l\rangle = \frac{k_b^2}{\alpha z m} \left(\frac{N-l-1}{N+l+1} \right)^{\frac{1}{2}}$$

$$= (l+\tfrac{1}{2})(l+1-N) \left\langle \frac{1}{r^2} \right\rangle,$$

$$\langle l+1| \frac{1}{r^3} |l\rangle = \frac{k_b \mathcal{A}_{l+1}}{(l+1)(2l+3)} \left\langle \frac{1}{r^2} \right\rangle$$

$$= \frac{k_b}{\alpha z m} \mathcal{A}_{l+1} \left(\frac{l}{2l+3} \right) \left\langle \frac{1}{r^3} \right\rangle,$$

$$\langle l+1| \frac{1}{r^4} |l\rangle = \frac{2\alpha z m}{l(l+2)} \langle l+1| \frac{1}{r^3} |l\rangle$$

$$= \frac{\alpha z m}{k_b} \mathcal{A}_{l+1} \left(\frac{2l-1}{l+2} \right) \frac{1}{[3N^2 - l(l+1)]} \left\langle \frac{1}{r^4} \right\rangle,$$

$$\langle l+1| \frac{1}{r^5} |l\rangle = \frac{3}{2} \left(\frac{k_b^2}{\alpha z m} \right) \left(\frac{5N^2 - l(l+2)}{(2l-1)(2l+5)} \right) \langle l+1| \frac{1}{r^4} |l\rangle.$$

* This work was supported in part by the Army Research Office (Durham) and the Research Foundation of Oklahoma State University.

[1] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters **17**, 148 (1965); E. C. G. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, *ibid.* **19**, 322 (1965); H. Bacry, Nuovo Cimento **41A**, 221 (1966); M. Y. Han, *ibid.* **42B**, 367 (1966); A. Böhm, *ibid.* **43A**, 665 (1966); A. O. Barut, P. Budini, and C. Fronsdal, Proc. Roy. Soc. **A291**, 106 (1966); R. Musto, Phys. Rev. **148**, 1274 (1966); R. H. Pratt and T. F. Jordan, *ibid.* 1276 (1966); M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966); A. Joseph, *ibid.* **39**, 829 (1967); C. A. Coulson and A. Joseph, *ibid.* 838 (1967); I. A. Malkin and V. I. Manko, P. N. Lebedev Physical Institute, Preprint N 145, 1967, Moscow, USSR; I. A. Malkin and V. I. Manko, P. N. Lebedev Physical Institute, Preprint N 9, 1968, Moscow, USSR.

[2] S. Pasternack and R. M. Sternheimer, J. Math. Phys. **3**, 1280 (1962).

[3] L. C. Biedenharn and N. V. V. J. Swamy, Phys. Rev. **133B**, 1353 (1964).

[4] S. Pasternack, U.S. Natl. Acad. Sci. **23**, 91 (1937).

[5] L. C. Biedenharn, J. L. Machale, and R. M. Thaler, Phys Rev. **100**, 376 (1955).

[6] L. C. Biedenharn and P. J. Brussard, *Coulomb Excitation* (Oxford University Press, London, 1964), Chap. III, Secs. 6 and 7; J. T. Reynolds, S. D. Onley, and L. C. Biedenharn, J. Math. Phys. **5**, 411 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Effective Permittivity of a Polycrystalline Dielectric

JOHN E. MOLYNEUX
Department of Mechanical and Aerospace Sciences, University of Rochester, Rochester, New York 14627

(Received 8 August 1969)

We use statistical variational principles to determine upper and lower bounds for the effective permittivity of a polycrystalline dielectric. We indicate how to derive bounds containing permittivity correlation functions of arbitrary order, and we obtain explicit expressions for bounds depending on one- and two-point correlation functions and for bounds containing one-, two-, and three-point correlation functions. We prove that for two classes of polycrystal, the effective permittivity may be exactly determined, and we use these exact expressions to show that we have obtained the best possible upper and lower bounds.

1. INTRODUCTION

In this paper we treat a polycrystalline dielectric as a statistically homogeneous and isotropic random medium. By this we mean that the permittivity tensor $\epsilon_{ij}(x)$ of the polycrystal is characterized by the sequence of correlation functions

$$C_{i_1 j_1 \cdots i_n j_n}(x_1, \cdots, x_n) = \langle \epsilon_{i_1 j_1}(x_1) \cdots \epsilon_{i_n j_n}(x_n) \rangle, \quad n = 1, 2, \cdots, \quad i_n, j_n = 1, 2, 3, \tag{1}$$

where brackets denote ensemble average and

$$C_{i_1 j_1 \cdots i_n j_n}(x_1 + y, \cdots, x_n + y) = C_{i_1 j_1 \cdots i_n j_n}(x_1, \cdots, x_n)$$

because of homogeneity. The electric field $E_i(x)$ and displacement vector $D_i(x)$ in the medium satisfy

$$\text{div } D = 0, \tag{2}$$

$$\text{curl } E = 0, \tag{3}$$

$$D_i(x) = \epsilon_{ij}(x) E_j(x), \tag{4}$$

and, as we shall show, are uniquely determined from these equations once the average field in the medium $\langle E_i \rangle$ is prescribed. To describe completely the electric field in a statistical sense we must, of course, find all the correlation functions

$$\langle E_{i_1}(x_1) \cdots E_{i_n}(x_n) \rangle \tag{5}$$

and

$$\langle E_{i_1}(x_1) \cdots E_{i_m}(x_m) \epsilon_{i_{m+1} j_{m+1}} \times (x_{m+1}) \cdots \epsilon_{i_{m+n} j_{m+n}}(x_{m+n}) \rangle. \tag{6}$$

This task is extremely difficult since the governing equations are nonlinear in stochastic variables. It is convenient, therefore, to consider a greatly reduced statistical description in which only the relationship between the prescribed average electric field and the average displacement field is sought. It may be shown that for a statistically homogeneous and isotropic medium this relationship must be of the form

$$\langle D_i \rangle = \epsilon_{\text{eff}} \langle E_i \rangle, \tag{7}$$

where ϵ_{eff}, the effective permittivity of the medium, is a constant whose value depends on the correlation functions given in (1). An exact determination of the effective permittivity still requires the solution of an infinite set of coupled partial differential equations. Thus, it is important from a practical point of view to obtain bounds on this quantity. It is desirable, in fact, to have a sequence of bounds depending on successively higher-order correlation functions, since it is reasonable to assume that the bounds will become increasingly accurate as more statistical information is included in them. In the following sections, we show that bounds of the type described may be readily determined by the use of suitable statistical variational principles. We also obtain two exact expressions for the effective permittivity for special classes of polycrystal and are able to demonstrate that the bounds we have derived, containing three-point correlations, are the best possible bounds.

2. STATISTICAL VARIATIONAL PRINCIPLES

We separate the average and fluctuating parts of the electric field and permittivity tensor and write

$$E_i(x) = \langle E_i \rangle + E_i'(x), \quad \epsilon_{ij}(x) = \langle \epsilon_{ij} \rangle + \epsilon_{ij}'(x),$$

where the bracketed quantities are constants because of homogeneity. We can then write (2), (3), and (4) in the form

$$\langle \epsilon_{ij} \rangle \frac{\partial E_j'}{\partial x_i} + \langle E_j \rangle \frac{\partial \epsilon_{ij}'}{\partial x_i} + \frac{\partial \epsilon_{ij}' E_j'}{\partial x_i} = 0, \tag{8}$$

$$\delta_{ijk} \frac{\partial E_k'}{\partial x_j} = 0, \tag{9}$$

where δ_{ijk} is the alternating tensor of Levi-Cività and sums are carried out over repeated indices. In media where $\epsilon_{ij}'(x)$ is not continuously differentiable in each

ensemble member (e.g., media composed of randomly oriented crystals), we can interpret the derivatives appearing in (8) and (9) in the mean-square sense[1] since we ultimately conceive of these equations as arising from the variational principles to be proved in this section. Alternately, we can first consider a fictitious ensemble in which the discontinuities have been smoothed, then use a limiting process to arrive at results for the actual ensemble. In either case, the validity of our final results depends on the behavior of the various permittivity correlation functions, and we will always assume that these functions are sufficiently smooth and well behaved for our formal manipulations to be valid. It can be shown that, for a statistically homogeneous medium, (8) and (9) determine the electric field uniquely if the average field is prescribed. Indeed, suppose there are two solutions $E_i^{(1)}(x)$ and $E_i^{(2)}(x)$, each with the same average value $\langle E_i \rangle$. The difference of these solutions $\Delta E_i = E_i^{(1)}(x) - E_i^{(2)}(x)$ must have zero average, vanishing curl, and satisfy

$$\frac{\partial \epsilon_{ij} \Delta E_i}{\partial x_i} = 0. \tag{10}$$

Since curl $\Delta E = 0$, $\Delta E = \nabla \psi$ for some scalar function ψ. Consider the correlation function

$$T_i(r) = \langle \epsilon_{ij}(x) \Delta E_j(x) \psi(x + r) \rangle.$$

Since $T_i(-r) = \langle \epsilon_{ij}(x + r) \Delta E_j(x + r) \psi(x) \rangle$, we find using (10) that

$$\begin{aligned} \frac{\partial T_i(r)}{\partial r_i} &= \left\langle \epsilon_{ij}(x) \Delta E_j(x) \frac{\partial \psi(x + r)}{\partial x_i} \right\rangle \\ &= - \frac{\partial T_i(-r)}{\partial r_i} \\ &= - \left\langle \psi(x) \frac{\partial \epsilon_{ij} \Delta E_j(x + r)}{\partial x_i} \right\rangle = 0. \end{aligned} \tag{11}$$

This relation, evaluated at $r = 0$, shows that

$$\langle \epsilon_{ij} \Delta E_i \Delta E_j \rangle = 0$$

and therefore, since ϵ_{ij} is a positive-definite tensor, that $\Delta E_i = 0$ almost everywhere.

We shall now formulate the two standard extremum principles of classical electrostatics as statistical variational principles. In doing so, we make use of the fact that, for a statistically homogeneous medium, the average value of any quantity is a constant. Therefore, average values can be considered as functionals. In particular, the average energy of a polycrystal may be considered either as a functional of the electric field or the displacement field.

Let us define the functional

$$J[A] = \langle \epsilon_{ij} A_i A_j \rangle \tag{12}$$

for all mean-square differentiable, statistically homogeneous, stochastic (DHS) vector fields $A = (A_1, A_2, A_3)$ which are defined on the polycrystalline medium and which satisfy the conditions

$$\langle A_i \rangle = \langle E_i \rangle, \tag{13}$$

$$\text{curl } A = 0, \tag{14}$$

and are such that the stochastic process $(A_i^{(1)}, A_i^{(2)}, \epsilon_{ij})$, where $A_i^{(k)}$ are any two A fields, is also statistically homogeneous. (In the future we shall assume that all correlations that arise are translation invariant and shall not mention conditions like this explicitly.) Then $J[A]$ is an absolute minimum when A also satisfies

$$\frac{\partial \epsilon_{ij} A_j}{\partial x_i} = 0, \tag{15}$$

i.e., when A is the electric field $E(x)$ in the polycrystalline medium. To prove this statement, let $A_i(x) = E_i(x) + a_i(x)$, where $a(x)$ has zero average and vanishing curl. Then

$$J[E + a] = J[E] + J[a] + 2\langle \epsilon_{ij} E_i a_j \rangle. \tag{16}$$

Since curl $a = 0$, we have $a = -\text{grad } \varphi$ and we can write the last term in (16) as

$$\langle \epsilon_{ij} E_i a_j \rangle = - \left\langle \epsilon_{ij} E_i \frac{\partial \varphi}{\partial x_j} \right\rangle = 0, \tag{17}$$

where this follows by manipulations similar to those used in obtaining (11). Noting that $J[a] \geq 0$, we have

$$J[E + a] \geq J[E], \tag{18}$$

where equality holds if and only if a vanishes almost everywhere.

We will assume now and throughout the remainder of the paper that the principal permittivities $\epsilon_i(x)$, $i = 1, 2, 3$, satisfy the inequalities

$$0 < \underline{\epsilon} \equiv \min_i \inf_x \epsilon_i(x) \leq \max_i \sup_x \epsilon_i(x) \equiv \bar{\epsilon} < \infty,$$

where $\underline{\epsilon}$ and $\bar{\epsilon}$ are taken to be constants, i.e., the same for each ensemble member. In this case, we can define the tensor $\theta_{ij}(x)$ by the equation

$$\theta_{ik}(x) \epsilon_{kj}(x) = \epsilon_{ik}(x) \theta_{kj}(x) = \delta_{ij}, \tag{19}$$

where δ_{ij} is the Kronecker δ function. Let the functional

$$I[B] = \langle \theta_{ij} B_i B_j \rangle \tag{20}$$

be defined for all DHS vector fields which are defined on the polycrystalline medium and which satisfy the

conditions

$$\langle B_i \rangle = \langle D_i \rangle, \tag{21}$$

$$\text{div } B = 0. \tag{22}$$

Then $I[B]$ is an absolute minimum when B also satisfies

$$\delta_{ijk} \frac{\partial \theta_{kl} B_l}{\partial x_j} = 0, \tag{23}$$

i.e., when B is the displacement field $D_i(x)$ in the polycrystal. The proof of this principle is quite similar to the first. We let $B = D + b$, where $\langle b \rangle$ and div $b = 0$. Then

$$I[D + b] = I[D] + I[b] + 2\langle \theta_{ij} b_i D_j \rangle, \tag{24}$$

and we can show that the last term on the right vanishes since b can be represented as the curl of another vector.

Using the fact that J and I take their minimum values, when evaluated for the actual electric and displacement fields, it may be shown that

$$J[E] = I[D] = \langle D_i \rangle \langle E_i \rangle. \tag{25}$$

Equation (25) together with (7) and the results obtained above show that the effective permittivity satisfies

$$\{I[B]\}^{-1} \langle D \rangle^2 \leq \epsilon_{\text{eff}} \leq \{\langle E \rangle^2\}^{-1} J[A], \tag{26}$$

where A is any trial field satisfying (13) and (14), and B is any trial field satisfying (21) and (22).

Let us now rewrite (26) in a form more convenient for our future work. For a fixed $A_i'(x)$ satisfying $\langle A_i' \rangle = 0$, curl $A' = 0$, we let $A_i(x) = \langle E_i \rangle + \lambda A_i'(x)$ and find the value of λ which makes $\phi(\lambda) = J[A]$ a minimum value. Using a similar procedure for $I[B]$, we find that

$$\begin{aligned} &\langle D \rangle^2 \{I[\langle D \rangle] - [I[B']]^{-1} [\langle D_i \rangle \langle \theta_{ij}' B_j' \rangle]^2\}^{-1} \\ &\quad \leq \epsilon_{\text{eff}} \\ &\quad \leq \{\langle E \rangle^2\}^{-1} \{J[\langle E \rangle] - [J[A']]^{-1} [\langle E_i \rangle \langle \epsilon_{ij}' A_j' \rangle]^2\} \end{aligned} \tag{27}$$

for arbitrary A' and B' satisfying $\langle A' \rangle = \langle B' \rangle = 0$, curl $A' = 0$, and div $B' = 0$ (θ_{ij}' is the fluctuating part of θ_{ij}). Since, by statistical isotropy, $\langle \theta_{ij} \rangle = \langle \theta \rangle \delta_{ij}$ and $\langle \epsilon_{ij} \rangle = \langle \epsilon \rangle \delta_{ij}$ where $\langle \theta \rangle$ and $\langle \epsilon \rangle$ are constants, (27), evaluated for $A' = B' = 0$, gives immediately the well-known bounds

$$\langle D \rangle^2 \{I[\langle D \rangle]\}^{-1} = \langle \theta \rangle < \epsilon_{\text{eff}} < \{\langle E \rangle^2\}^{-1} J[\langle E \rangle] = \langle \epsilon \rangle. \tag{28}$$

For a polycrystal composed of randomly oriented crystals all having the same principle permittivities $\epsilon_1, \epsilon_2, \epsilon_3$; $\langle \epsilon \rangle = \frac{1}{3}(\epsilon_1 + \epsilon_2 + \epsilon_3)$ and

$$\langle \theta \rangle = \tfrac{1}{3}(\epsilon_1^{-1} + \epsilon_2^{-1} + \epsilon_3^{-1}).$$

It is immediately obvious that any nonzero A' and B' give better bounds than (28).

In the next section, we use (27) to find bounds on ϵ_{eff} which involve one-, two-, and three-point correlation functions. To obtain a set of bounds involving only one- and two-point correlations it is necessary to derive another statistical variational principle which is closely related to the classical ones on which (27) is based, and is a generalization of a principle obtained by Hashin and Shtrikman.[2] We have already noted that $\langle \epsilon_{ij} E_i E_j \rangle = \langle \epsilon_{ij} E_j \rangle \langle E_i \rangle$. If we let α be an arbitrary positive constant and we add and subtract $\alpha \langle E_i E_i \rangle$ from the right-hand side of this equation, we obtain the identity

$$\langle \epsilon_{ij} E_i E_j \rangle = \alpha \langle E \rangle^2 + \langle C_i \rangle \langle E_i \rangle, \tag{29}$$

where $C_i \equiv (\epsilon_{ij} - \alpha \delta_{ij}) E_j$. If we restrict α so that it satisfies one of the inequalities

$$\alpha < \min_i \inf_x \epsilon_i(x) \quad \text{or} \quad \alpha > \max_i \sup_x \epsilon_i(x),$$

we can define the stochastic tensor $\beta_{ij}(x, \alpha)$ by

$$\beta_{ik}(\epsilon_{kj} - \alpha \delta_{kj}) = \delta_{ij}, \tag{30}$$

and since $\beta_{ij} C_j = E_i$, we see that the second term on the right-hand side of (29) is identically equal to $\langle \beta_{ij} C_i C_j \rangle - \langle C_i E_i' \rangle$. The correlation $\langle C_i E_i' \rangle$ can be expressed in terms of C_i only. In fact, from the definition of C_i it follows that

$$\nabla \cdot C = -\alpha \nabla \cdot E', \tag{31}$$

and since curl E' vanishes,

$$E' = \nabla \psi'. \tag{32}$$

From (31) and (32) we find

$$E_i'(x) = -\alpha^{-1} \int \frac{\partial G(x, \xi)}{\partial x_i} \frac{\partial C_j}{\partial \xi_j}(\xi)\, d\xi, \tag{33}$$

where $G(x, \xi) = -(4\pi |x - \xi|)^{-1}$ and the integral is interpreted to mean that

$$\langle E_i'(x) f'(y) \rangle = -\alpha^{-1} \int \frac{\partial G(x, \xi)}{\partial x_i} \left\langle f'(y) \frac{\partial C_j}{\partial \xi_j}(\xi) \right\rangle d\xi$$

for all sufficiently well-behaved stochastic fields $f'(y)$. Multiplying (33) by $C_i(x)$ and averaging, we obtain

$$\begin{aligned} \langle C_i E_i' \rangle &= -\alpha^{-1} \int \frac{\partial G(x, \xi)}{\partial x_i} \left\langle C_i(x) \frac{\partial C_j}{\partial \xi_j}(\xi) \right\rangle d\xi \\ &= -\alpha^{-1} \int G(x, \xi) \frac{\partial^2}{\partial \xi_i \partial \xi_j} \langle C_i(x) C_j(\xi) \rangle\, d\xi \\ &\equiv -\alpha^{-1} \mathcal{F}[C_i, C_j]. \end{aligned} \tag{34}$$

We have now succeeded in expressing $\langle \epsilon_{ij} E_i E_j \rangle$ as the functional

$$\langle \epsilon_{ij} E_i E_j \rangle = \mathfrak{J}[C] \equiv \alpha \langle E \rangle^2 + \langle \beta_{ij} C_i C_j \rangle + \alpha^{-1} \mathcal{F}[C_i, C_j]. \quad (35)$$

Let us consider $\mathfrak{J}[C]$ as a functional defined for all DHS vector fields which are defined on the polycrystalline medium and which satisfy the condition

$$\langle C_i \rangle = \langle D_i \rangle - \alpha \langle E_i \rangle = (\epsilon_{\text{eff}} - \alpha) \langle E_i \rangle. \quad (36)$$

Then $\mathfrak{J}[C]$ has an extremum for $C_i = (\epsilon_{ij} - \alpha \delta_{ij}) E_j \equiv \bar{C}_i$. This extremum is an absolute minimum for β_{ij}, positive definite, and an absolute maximum for $\beta_{ij} + \alpha^{-1}\delta_{ij}$, negative definite. To prove this, we let $C_i = \bar{C}_i + C_i'$ where C_i' has zero average. Then

$$\mathfrak{J}[\bar{C} + C'] = \mathfrak{J}[\bar{C}] + 2\langle E_i C_i' \rangle + 2\alpha^{-1} \mathcal{F}[C_i', \bar{C}_j] + \alpha^{-1} \mathcal{F}[C_i', C_j'] + \langle \beta_{ij} C_i' C_j' \rangle, \quad (37)$$

where statistical homogeneity has been invoked to equate $\mathcal{F}[C_i', \bar{C}_j]$ and $\mathcal{F}[\bar{C}_i, C_j']$. Since $\operatorname{div} \bar{C} = -\alpha \operatorname{div} E$ and $E = \operatorname{grad} \psi$,

$$\begin{aligned} \alpha^{-1} \mathcal{F}[C_i', \bar{C}_j] &= -\mathcal{F}[C_i' E_j] \\ &= -\int G(x, \xi) \frac{\partial^2}{\partial \xi_i \partial \xi_j} \left\langle C_i'(x) \frac{\partial \psi}{\partial \xi_j}(\xi) \right\rangle \\ &= -\int \frac{\partial^2 G(x, \xi)}{\partial \xi_i \partial \xi_j} \left\langle C_i'(x) \frac{\partial \psi}{\partial \xi_i}(\xi) \right\rangle \\ &= -\langle C_i' E_i \rangle. \end{aligned} \quad (38)$$

Employing Bochner's theorem,[1] we write

$$\langle C_i'(x) C_j'(x + r) \rangle = \int \Phi_{ij}(\kappa) e^{i\kappa \cdot r} \, d\kappa. \quad (39)$$

Then, changing variables of integration in (34) from ξ to $r = \xi - x$, we can write $\mathcal{F}[C_i', C_j']$ in the form

$$\begin{aligned} \mathcal{F}[C_i', C_j'] &= -\int G(|r|) \left[\int \Phi(|\kappa|) e^{i\kappa \cdot r} \, d\kappa \right] dr \\ &= 4\pi \int_0^\infty \Phi(|\kappa|) \, d\,|\kappa|, \end{aligned} \quad (40)$$

where $\Phi(|\kappa|) \equiv \Phi_{ij}(\kappa) \kappa_i \kappa_j$, and $\Phi(|\kappa|) > 0$ for all $\kappa \neq 0$ by well-known properties of the tensor $\Phi_{ij}(\kappa)$. Hence, $\mathcal{F}[C_i', C_j']$ is nonnegative, and

$$\mathfrak{J}[\bar{C} + C'] = \mathfrak{J}[\bar{C}] + \langle \beta_{ij} C_i' C_j' \rangle + \alpha^{-1} \mathcal{F}[C_i', C_j'] \geq \mathfrak{J}[\bar{C}] \quad (41)$$

for β_{ij} positive definite. To demonstrate the remainder of our assertion, we note that by (39) and (40)

$$\langle C_i' C_i' \rangle - \mathcal{F}[C_i' C_j'] = 4\pi \int_0^\infty [|\kappa|^2 \Phi_{ii}(\kappa) - \kappa_i \kappa_j \Phi_{ij}(\kappa)] \, d\,|\kappa| \geq 0, \quad (42)$$

and, therefore,

$$\begin{aligned} \mathfrak{J}[\bar{C} + C'] &= \mathfrak{J}[\bar{C}] + \langle (\beta_{ij} + \alpha^{-1}\delta_{ij}) C_i' C_j' \rangle \\ &\quad - \alpha^{-1} \{ \langle C_i' C_i' \rangle - \mathcal{F}[C_i' C_j'] \} \\ &\leq \mathfrak{J}[\bar{C}] \end{aligned} \quad (43)$$

for $\beta_{ij} + \alpha^{-1}\delta_{ij}$, negative definite. For fixed C_i' with $\langle C_j' \rangle = 0$, let us consider the function of the real variable λ, defined by

$$\begin{aligned} \Psi(\lambda; \alpha) &= \mathfrak{J}[(\epsilon_{\text{eff}} - \alpha) \langle E_i \rangle + \lambda C_i'] \\ &= \{ \langle \beta_{ij} C_i' C_j' \rangle + \alpha^{-1} \mathcal{F}[C_i' C_j'] \} \lambda^2 \\ &\quad + 2\lambda \{ (\epsilon_{\text{eff}} - \alpha) \langle E_i \rangle \langle \beta_{ij} C_j' \rangle \} \\ &\quad + \{ \alpha \langle E \rangle^2 + \langle \beta \rangle (\epsilon_{\text{eff}} - \alpha)^2 \langle E \rangle^2 \}, \end{aligned} \quad (44)$$

where $\langle \beta \rangle \delta_{ij} = \langle \beta_{ij} \rangle$. It follows, from the above considerations that at the value λ_0, defined by $\Psi'(\lambda_0; \alpha) = 0$, Ψ will have a minimum when β_{ij} is positive definite and a maximum when $\beta_{ij} + \alpha^{-1}\delta_{ij}$ is negative definite. Computing λ_0 and using the fact that

$$\mathfrak{J}[(\epsilon_{ij} - \alpha \delta_{ij}) E_j] = \epsilon_{\text{eff}} \langle E \rangle^2,$$

we find

$$\Psi_0(C_i'; \alpha_2) \leq \epsilon_{\text{eff}} \langle E \rangle^2 \leq \Psi_0(C_i'; \alpha_1), \quad (45)$$

where

$$\begin{aligned} &\Psi_0(C_i'; \alpha) \\ &= \alpha \langle E \rangle^2 + (\epsilon_{\text{eff}} - \alpha)^2 \{ \langle \beta \rangle \langle E \rangle^2 \\ &\quad - [\langle \beta_{ij} C_i' C_j' \rangle + \alpha^{-1} \mathcal{F}[C_i' C_j']]^{-1} [\langle E_i \rangle \langle \beta_{ij} C_j' \rangle]^2 \}, \end{aligned} \quad (46)$$

and α_1, α_2 are any values of α satisfying the inequalities

$$\alpha_1 < \min_{i=1,2,3} \inf_x \epsilon_i(x), \quad \max_{i=1,2,3} \sup_x \epsilon_i(x) < \alpha_2.$$

Letting $C_i' \equiv 0$ and taking the limits $\alpha_1 \to 0$, $\alpha_2 \to \infty$, we recover the bounds given by (28). Also, for fixed α_1 and α_2, it is clear that $\Psi_0(C_i'; \alpha_1) \leq \Psi_0(0; \alpha_1)$ and $\Psi_0(0; \alpha_2) \leq \Psi_0(C_i'; \alpha_2)$ so that any nonzero C_i' will give better bounds than (28).

3. BOUNDS ON THE EFFECTIVE PERMITTIVITY

We shall use a procedure similar to that developed by Beran[3] and Beran and Molyneux[4] to obtain bounds on ϵ_{eff} which depend on two- and three-point correlation functions of the permittivity tensor. We select trial functions A' and B' satisfying

$$\langle \epsilon \rangle \frac{\partial A_i'}{\partial x_i} + \langle E_j \rangle \frac{\partial \epsilon_{ij}'}{\partial x_j} = 0, \quad \delta_{ijk} \frac{\partial A_k'}{\partial x_j} = 0, \quad (47)$$

and

$$\langle \theta \rangle \delta_{ijk} \frac{\partial B_k'}{\partial x_j} + \langle D_l \rangle \delta_{ijk} \frac{\partial \theta_{kl}'}{\partial x_j} = 0, \quad \frac{\partial B_i'}{\partial x_i} = 0. \quad (48)$$

Equations (47) and (48) will be recognized as the first-order perturbation approximations to the actual problem (8) and (9). The bounds derived from them should, therefore, be quite accurate in the limit of small fluctuations. Indeed, it is shown that they coincide to the second order in small quantities and give the perturbation result for ϵ_{eff} derived in Molyneux.[5] We solve (47) and (48) by reducing them both to Poisson's equation (since curl $A' = 0$, $A' = \text{grad}\,\phi$, and since div $B' = 0$, $B' = \text{curl}\,C'$, where div $C' = 0$). The solutions are found to be

$$A_i'(x) = -\langle\epsilon\rangle^{-1}\langle E_j\rangle \int \frac{\partial G(x,\xi)}{\partial x_i}\frac{\partial \epsilon_{kj}'(\xi)}{\partial \xi_k}\,d\xi \quad (49)$$

and

$$B_i'(x) = -\langle\theta\rangle^{-1}\langle D_k\rangle\theta_{ik}'(x) + \frac{\partial^2}{\partial x_i \partial x_m}\langle\theta\rangle^{-1}\langle D_k\rangle \int G(x,\xi)\theta_{mk}'(\xi)\,d\xi, \quad (50)$$

where $G(x,\xi) = -(4\pi|x-\xi|)^{-1}$ and the integrals must be interpreted in the sense discussed after (33). We shall now use (49) and (50) to find the various moments which appear in the expressions for the upper and lower bounds on ϵ_{eff}.

To compute the upper bound, we first multiply (49) by $\langle E_j\rangle\epsilon_{ij}'(x)$ and average. This gives

$$\begin{aligned}\langle E_i\rangle\langle\epsilon_{ij}'A_j'\rangle &= -\langle\epsilon\rangle^{-1}\langle E_i\rangle\langle E_k\rangle\int\frac{\partial G(x,\xi)}{\partial x_j}\left\langle\epsilon_{ij}'(x)\frac{\partial\epsilon_{mk}'(\xi)}{\partial\xi_m}\right\rangle d\xi \\ &= -\langle\epsilon\rangle^{-1}\langle E_i\rangle\langle E_k\rangle\int G(|r|)\frac{\partial^2}{\partial r_j\partial r_m}C_{ijmk}(r)\,dr, \quad (51)\end{aligned}$$

where $C_{ijmk}(r) \equiv \langle\epsilon_{ij}'(x)\epsilon_{mk}'(x+r)\rangle$, and we have changed variables of integration from ξ to $r = \xi - x$. Since C_{ijkl} is a fourth-order isotropic tensor and satisfies the symmetry conditions $C_{ijkl}(r) = C_{jikl}(r) = C_{ijlk}(r) = C_{klij}(-r)$, it must have the form[6]

$$\begin{aligned}C_{ijkl}(r) = {} & R_1 r_i r_j r_k r_l + R_2(r_i r_j\delta_{kl} + r_k r_l\delta_{ij}) \\ & + R_3(r_i r_k\delta_{jl} + r_i r_l\delta_{jk} + r_j r_k\delta_{il} + r_j r_l\delta_{ik}) \\ & + R_4\delta_{ij}\delta_{kl} + R_5(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl}), \quad (52)\end{aligned}$$

where the R_n, $n = 1, 2, \cdots, 5$, are functions of $|r|^2$ only. Substituting (52) into (51), we find that

$$\langle E_i\rangle\langle\epsilon_{ij}'A_j'\rangle = -\langle\epsilon\rangle^{-1}\langle E\rangle^2[\tfrac{1}{9}\langle\epsilon_{ij}'\epsilon_{ij}'\rangle + K], \quad (53)$$

where

$$K = -\tfrac{2}{3}\int_0^\infty \rho[\rho^2 R_1(\rho) + 2R_2(\rho) + 5R_3(\rho)]\,d\rho.$$

The principal step in obtaining (53) involves the following result: If $T_{i_1\cdots i_{2n}}(r)$ is any suitably well-behaved isotropic tensor, then

$$\begin{aligned}\int G(|r|)\frac{\partial^2 T_{i_1\cdots j\cdots k\cdots i_{2n}}(r)}{\partial r_j\partial r_k}\,dr &= \tfrac{1}{3}T_{i_1\cdots k\cdots k\cdots i_{2n}}(0) \\ &\quad + (4\pi)^{-1}\int|r|^{-3}\,\Pi_{jk}(r)T_{i_1\cdots j\cdots k\cdots i_{2n}}(r)\,dr, \quad (54)\end{aligned}$$

where $\Pi_{jk}(r) = [\delta_{jk} - 3|r|^{-2}r_j r_k]$. Furthermore, the last triple integral on the right-hand side of (54) can always be reduced to a single integral of a suitable scalar function of $\rho = |r|$. The proof of this result is quite straightforward and is included in the Appendix. Equation (53) follows directly from (54) if we note that

$$C_{ijjk}(0) = 3^{-1}\delta_{ik}C_{mjjm}(0) = 3^{-1}\langle\epsilon_{mj}'\epsilon_{jm}'\rangle\delta_{ik},$$

$$(4\pi)^{-1}\int|r|^{-3}\,\Pi_{jm}(r)C_{ijmk}(r)\,dr = (12\pi)^{-1}\delta_{ik}\int|r|^{-3}\,\Pi_{jm}(r)C_{qjmq}(r)\,dr,$$

and that

$$\Pi_{jm}(r)C_{qjmq} = -2|r|^2[|r|^2 R_1 + 2R_2 + 5R_3].$$

Next, we multiply (49) by itself and obtain

$$\langle\epsilon\rangle\langle A_i'A_i'\rangle = \langle\epsilon\rangle^{-1}\langle E_j\rangle\langle E_k\rangle\iint G(x,\xi)G(x,\eta)\times\left\langle\frac{\partial^2\epsilon_{ij}'(\xi)}{\partial\xi_i\partial\xi_l}\frac{\partial^2\epsilon_{km}'(\eta)}{\partial\eta\partial\eta_m}\right\rangle d\xi\,d\eta. \quad (55)$$

Changing variables of integration from ξ, η to $r = \eta - \xi$, $s = \xi$, we find after some manipulation that

$$\begin{aligned}\langle\epsilon\rangle\langle A_i'A_i'\rangle &= \langle\epsilon\rangle^{-1}\langle E_j\rangle\langle E_k\rangle\int G(|r|)\frac{\partial^2}{\partial r_m\partial r_n}C_{jmkn}(r)\,dr \\ &= -\langle E_i\rangle\langle\epsilon_{ij}'A_j'\rangle. \quad (56)\end{aligned}$$

To find the remaining quantity necessary for computation of the upper bound, we multiply (49) by $\epsilon_{ij}'(x)A_j'(x)$ and average. This gives, after a change of variables of integration,

$$\langle\epsilon_{ij}'A_i'A_j'\rangle = \langle\epsilon\rangle^{-2}\langle E_k\rangle\langle E_l\rangle\iint G(|r|)G(|s|)\frac{\partial^4 D_{ijkmln}(r,s)}{\partial r_i\partial r_m\partial s_j\partial s_n}\,dr\,ds, \quad (57)$$

where

$$D_{ijkmln}(r,s) \equiv \langle\epsilon_{ij}'(x)\epsilon_{km}'(x+r)\epsilon_{ln}'(x+s)\rangle.$$

We can rewrite (57) in a slightly more convenient form if we make use of (54). Indeed, if we let

$$T_{ikml}(r) = \int G(|s|)\frac{\partial^2 D_{ijkmln}(r,s)}{\partial s_j\partial s_n}\,ds,$$

then (54) applies directly. The term $T_{mkml}(0)$ can be evaluated by a second use of (54), while the integrated term can be simplified by applying an obvious generalization of (54) to $T_{ikml}(r)$. These manipulations give

$$\langle \epsilon'_{ij} A'_i A'_j \rangle = \langle \epsilon \rangle^{-2} \langle E \rangle^2 [\tfrac{1}{27} \langle \epsilon'_{mj} \epsilon'_{jk} \epsilon'_{km} \rangle + L], \quad (58)$$

where

$$L = \tfrac{2}{9} \int_0^\infty |r|^{-1} [\delta_{im} - 3\,|r|^{-2} r_i r_m] D_{piqpqm}(0, r)\, d\,|r| + \frac{1}{48\pi^2} \iint |r|^{-3}\,|s|^{-3} [\delta_{im} - 3\,|r|^{-2} r_i r_m] \times [\delta_{jn} - 3\,|s|^{-2} s_j s_n] D_{ijqmqn}(r, s)\, dr\, ds.$$

The moments appearing in the lower bound are calculated in a similar fashion. If we let

$$F'_i(x) = -[B'_i(x) + \langle \theta \rangle^{-1} \langle D_k \rangle \theta'_{ik}(x)],$$

then (50) shows that F'_i is given by an expression of exactly the same form as A'_i. Thus, from (53) we can read directly the value of $\langle D_i \rangle \langle \theta'_{ij} F'_j \rangle$. Solving for $\langle D_i \rangle \langle \theta'_{ij} B'_j \rangle$, we obtain

$$\langle D_i \rangle \langle \theta'_{ij} B'_j \rangle = \langle \theta \rangle^{-1} \langle D \rangle^2 [-\tfrac{2}{9} \langle \theta'_{ij} \theta'_{ji} \rangle + P], \quad (59)$$

where

$$P = \tfrac{1}{3} \int_0^\infty |r|^{-1} [\delta_{jm} - 3\,|r|^{-2} r_j r_m] M_{jqqm}(r)\, d\,|r|,$$

and the tensor $M_{ijkl}(r) \equiv \langle \theta'_{ij}(x) \theta'_{kl}(x + r) \rangle$ has the same form as $C_{ijkl}(r)$ [see (34)]. From (56) we see that

$$\langle \theta \rangle \langle B'_i B'_i \rangle = -\langle D_i \rangle \langle \theta'_{ij} B'_j \rangle. \quad (60)$$

Finally, using (58), we can find the value of

$$\langle \theta'_{ij} F'_i F'_j \rangle = \langle \theta'_{ij} B'_i B'_j \rangle + 2 \langle \theta \rangle^{-1} \langle D_k \rangle \langle \theta'_{ij} \theta'_{ik} B'_k \rangle + 3^{-1} \langle \theta \rangle^{-2} \langle D \rangle^2 \langle \theta'_{ij} \theta'_{jk} \theta'_{kl} \rangle.$$

Hence to obtain $\langle \theta'_{ij} B'_i B'_j \rangle$ it is only necessary to calculate the moment $\langle \theta'_{li} \theta'_{lk} B'_i \rangle$. This is done by multiplying (50) by $\theta'_{li} \theta'_{lk}$ and averaging. The result is

$$\begin{aligned} &\langle \theta \rangle^{-1} \langle D_k \rangle \langle \theta'_{ji} \theta'_{jk} B'_i \rangle \\ &= -\langle \theta \rangle^{-2} \langle D_k \rangle \langle D_m \rangle \langle \theta'_{ji} \theta'_{jk} \theta'_{im} \rangle + \langle \theta \rangle^{-2} \langle D_k \rangle \langle D_m \rangle \\ &\quad \times \int \frac{\partial G(x, \xi)}{\partial x_i} \left\langle \theta'_{ji}(x) \theta'_{jk}(x) \frac{\partial \theta'_{mn}(\xi)}{\partial \xi_n} \right\rangle d\xi \\ &= \langle \theta \rangle^{-2} \langle D \rangle^2 \Big[-\tfrac{2}{9} \langle \theta'_{ij} \theta'_{jk} \theta'_{ki} \rangle \\ &\quad + \tfrac{1}{3} \int_0^\infty |r|^{-1} [\delta_{in} - 3\,|r|^{-2} r_i r_n] N_{ijjppn}(0, r)\, d\,|r| \Big], \end{aligned} \quad (61)$$

where

$$N_{ijklmn}(r, s) \equiv \langle \theta'_{ij}(x) \theta'_{kl}(x + r) \theta'_{mn}(x + s) \rangle.$$

From (61) together with (58), we find that

$$\langle \theta'_{ij} B'_i B'_j \rangle = \langle \theta \rangle^{-2} \langle D \rangle^2 [\tfrac{4}{27} \langle \theta'_{ij} \theta'_{jk} \theta'_{ki} \rangle + Q], \quad (62)$$

where

$$Q = -(\tfrac{4}{9}) \int_0^\infty |r|^{-1} [\delta_{im} - 3\,|r|^{-2} r_i r_m] N_{ijjkkm}(0, r)\, d\,|r| + \frac{1}{48\pi^2} \iint |r|^{-3}\,|s|^{-3} [\delta_{im} - 3\,|r|^{-2} r_i r_m] \times [\delta_{jn} - 3\,|s|^{-2} s_j s_n] N_{ijqmqn}(r, s)\, dr\, ds.$$

All moments appearing in (27) have now been calculated, and we have therefore determined a set of bounds on the effective permittivity. We find that, for the trial functions selected,

$$\begin{aligned} &\{\langle \theta \rangle - \langle \theta \rangle^{-1} [1 + \langle \theta \rangle^{-1} (\tfrac{2}{9} \langle \theta'_{ij} \theta'_{ji} \rangle - P)^{-1} \\ &\qquad \times (\tfrac{4}{27} \langle \theta'_{ij} \theta'_{jk} \theta'_{ki} \rangle + Q)]^{-1} (\tfrac{2}{9} \langle \theta'_{ij} \theta'_{ji} \rangle - P)\}^{-1} \\ &\quad \leq \epsilon_{\text{eff}} \leq \langle \epsilon \rangle - \langle \epsilon \rangle^{-1} [1 + \langle \epsilon \rangle^{-1} (\tfrac{1}{9} \langle \epsilon'_{ij} \epsilon'_{ji} \rangle + K)^{-1} \\ &\qquad \times (\tfrac{1}{27} \langle \epsilon'_{ij} \epsilon'_{jk} \epsilon'_{ki} \rangle + L)]^{-1} (\tfrac{1}{9} \langle \epsilon'_{ij} \epsilon'_{ji} \rangle + K). \end{aligned} \quad (63)$$

For polycrystalline media composed of randomly oriented crystals all having the same principal permittivities (ϵ_i; $i = 1, 2, 3$),

$$\langle \epsilon'_{ij} \epsilon'_{ji} \rangle = \langle \epsilon_{ij} \epsilon_{ji} \rangle - \langle \epsilon_{ij} \rangle \langle \epsilon_{ij} \rangle = \epsilon_{ij} \epsilon_{ij} - \langle \epsilon_{ij} \rangle \langle \epsilon_{ij} \rangle = \tfrac{1}{3} \sum_{i<j} (\epsilon_i - \epsilon_j)^2.$$

Here we use isotropy and the fact that $\epsilon_{ij} \epsilon_{ji}$ is an invariant. Similar arguments give

$$\langle \epsilon'_{ij} \epsilon'_{jk} \epsilon'_{ki} \rangle = (\epsilon_1^3 + \epsilon_2^3 + \epsilon_3^3) - (\epsilon_1 + \epsilon_2 + \epsilon_3)(\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2) + \tfrac{2}{9} (\epsilon_1 + \epsilon_2 + \epsilon_3)^3$$

and the corresponding moments for θ'_{ij} may be found by the substitution $\epsilon_i \to \epsilon_i^{-1}$ in these formulas.

In the limit of weak inhomogeneities, the bounds given by (63) coincide. In fact, from $\theta_{ik} \epsilon_{kj} = \delta_{ij}$ we find that

$$\langle \theta \rangle \cong \langle \epsilon \rangle^{-1} [1 - \tfrac{1}{3} \langle \epsilon \rangle^{-2} \langle \epsilon'_{ij} \epsilon'_{ji} \rangle], \quad \theta'_{ij} \cong -\langle \epsilon \rangle^{-2} \epsilon'_{ij},$$

and neglecting terms of order three and higher in primed quantities we obtain

$$\epsilon_{\text{eff}} \cong \langle \epsilon \rangle - \langle \epsilon \rangle^{-1} [\tfrac{1}{9} \langle \epsilon'_{ij} \epsilon'_{ji} \rangle + K]. \quad (64)$$

We observe that the bounds on ϵ_{eff} given by (63) contain one-, two-, and three-point correlation functions, while the bounds given by (28) involve only one-point moments. It is of interest to derive a set of bounds containing only one- and two-point moments, and we shall do this using (45).

First, let us use (54) to rewrite the functional $\mathcal{F}[C_i'C_j']$ in the form

$$\mathcal{F}[C_i'C_j'] = \tfrac{1}{3}\langle C_i'C_i'\rangle + F, \tag{65}$$

where

$$F = \int_0^\infty |r|^{-1}\,[\delta_{kl} - 3\,|r|^{-2}\,r_k r_l]\langle C_k'(x)C_l'(x+r)\rangle\,d\,|r|. \tag{66}$$

Substitution of (65) into (46) gives

$$\Psi_0(C_i';\alpha) = \alpha\langle E\rangle^2 + (\epsilon_{eff} - \alpha)^2\Big\{\langle\beta\rangle\langle E\rangle^2 - \left[\left\langle\left(\beta_{ij} + \frac{1}{3\alpha}\,\delta_{ij}\right)C_i'C_j'\right\rangle + \alpha^{-1}F\right]^{-1} \times (\langle E_i\rangle\langle\beta_{ij}C_j'\rangle)^2\Big\}. \tag{67}$$

We shall restrict our considerations to trial fields which can be expressed in the form

$$C_i' = \gamma_{ij}'\langle E_j\rangle, \tag{68}$$

where γ_{ij}' is a statistically homogeneous and isotropic stochastic tensor with zero average and γ_{ij}' may depend on α. With C_i' given by (68), the expression for Ψ_0 becomes

$$\Psi_0(\gamma';\alpha) = \langle E\rangle^2\{\alpha + (\epsilon_{\text{eff}} - \alpha)^2(\Sigma(\gamma';\alpha))^{-1} \times (\langle\beta\rangle\Sigma(\gamma',\alpha) - \tfrac{1}{3}\alpha\langle\beta_{ij}'\gamma_{ij}'\rangle^2)\}, \tag{69}$$

where

$$\Sigma(\gamma',\alpha) = \alpha\langle\gamma_{ij}'\beta_{jk}'\gamma_{ki}'\rangle + (\alpha\langle\beta\rangle + \tfrac{1}{3})\langle\gamma_{ij}'\gamma_{ij}'\rangle + I[\gamma'] \tag{70}$$

and

$$I[\gamma'] = \int_0^\infty |r|^{-1}\,\Pi_{ij}(r)\langle\gamma_{ik}'(0)\gamma_{kj}'(r)\rangle\,d\,|r|. \tag{71}$$

Since $(3\alpha)^{-1}\langle E\rangle^2\Sigma(\gamma',\alpha)$ is the denominator of the last term in (67), we see (from the proof of the variational principle) that $\Sigma(\gamma',\alpha_2) < 0 < \Sigma(\gamma',\alpha_1)$, for any α_1 and α_2, satisfy $\alpha_1 < \underline{\epsilon} < \bar{\epsilon} < \alpha_2$. Furthermore, the inequalities $\underline{\epsilon} \le \langle\theta\rangle^{-1} \le \epsilon_{\text{eff}} \le \langle\epsilon\rangle \le \bar{\epsilon}$ imply that $\alpha_1 < \epsilon_{\text{eff}} < \alpha_2$ for any admissible values of α_1 and α_2. Using (45) and (69) and the inequalities just discussed, we obtain the bounds

$$\Gamma(\alpha_1) \le \epsilon_{\text{eff}} \le \Gamma(\alpha_2), \tag{72}$$

where

$$\Gamma(\alpha) = [\langle\beta\rangle\Sigma(\gamma,\alpha) - \tfrac{1}{3}\alpha\langle\beta_{ij}'\gamma_{ji}'\rangle^2]^{-1} \times [(\alpha\langle\beta\rangle + 1)\Sigma(\gamma,\alpha) - \tfrac{1}{3}\alpha^2\langle\beta_{ij}'\gamma_{ji}'\rangle^2]. \tag{73}$$

These bounds are valid for arbitrary γ_{ij}'. We can be guided in our choice of γ_{ij}' by recourse to perturbation theory. Recall that the functional $\mathcal{J}[C]$ has its extremum for

$$C_i = (\epsilon_{ij} - \alpha\delta_{ij})E_j = (\langle\epsilon\rangle - \alpha)\langle E_i\rangle + \epsilon_{ij}'\langle E_j\rangle + (\langle\epsilon\rangle - \alpha)E_i + \epsilon_{ij}'E_j'.$$

If α is chosen so that $(\langle\epsilon\rangle - \alpha)$ is of order of the primed quantities, we see that to lowest order γ_{ij}' should be taken equal to ϵ_{ij}'. With this choice of γ_{ij}',

$$\langle\gamma_{ij}'\beta_{ji}'\rangle = \left\langle\sum_{i=1}^{3}(\epsilon_i(x) - \langle\epsilon\rangle)(\beta_i(x) - \langle\beta\rangle)\right\rangle = 3[\langle\beta\rangle(\alpha - \langle\epsilon\rangle) + 1],$$

$$\langle\gamma_{ij}'\beta_{jk}'\gamma_{ki}'\rangle = \left\langle\sum_{i=1}^{3}(\epsilon_i(x) - \langle\epsilon\rangle)^2(\beta_i(x) - \langle\beta\rangle)\right\rangle = (\alpha - \langle\epsilon\rangle)\langle\gamma_{ij}'\beta_{ji}'\rangle - \langle\beta\rangle\langle\epsilon_{ij}'\epsilon_{ji}'\rangle,$$

and $\Gamma(\alpha)$ becomes

$$\Gamma(\alpha) = \{3\alpha[\langle\beta\rangle(\langle\epsilon\rangle - \alpha) - 1] + \langle\beta\rangle[\tfrac{1}{3}\langle\epsilon_{ij}'\epsilon_{ji}'\rangle + I[\epsilon']]\}^{-1} \times \{3\alpha\langle\epsilon\rangle[\langle\beta\rangle(\langle\epsilon\rangle - \alpha) - 1] + (\alpha\langle\beta\rangle + 1)[\tfrac{1}{3}\langle\epsilon_{ij}'\epsilon_{ji}'\rangle + I[\epsilon']]\}, \tag{74}$$

with

$$I[\epsilon'] = \int_0^\infty |r|^{-1}\,\Pi_{ij}(r)\langle\epsilon_{ik}'(0)\epsilon_{kj}'(r)\rangle\,d\,|r| = 3K. \tag{75}$$

It is also of interest to consider another choice of γ_{ij}' which leads to a generalization of the bounds obtained by Hashin and Shtrikman.[2] We let γ_{ij}' be the fluctuating part of γ_{ij}, where γ_{ij} is taken to satisfy $\gamma_{ij}(3\alpha\beta_{jk} + \delta_{jk}) = \delta_{ik}$. In this case we find

$$\alpha\langle\gamma_{ij}'\beta_{ji}'\rangle = 1 - \langle\gamma\rangle(3\alpha\langle\beta\rangle + 1),$$

$$3\alpha\langle\gamma_{ij}'\beta_{jk}'\gamma_{ki}'\rangle = -3\alpha\langle\gamma\rangle\langle\gamma_{ij}'\beta_{ji}'\rangle - (3\alpha\langle\beta\rangle + 1)\langle\gamma_{ij}'\gamma_{ji}'\rangle,$$

$$\langle\gamma\rangle = (\tfrac{1}{3})\sum_{i=1}^{3}\langle(\epsilon_i(x) + 2\alpha)^{-1}(\epsilon_i(x) - \alpha)\rangle,$$

and $\Gamma(\alpha)$ becomes

$$\Gamma(\alpha) = \{\alpha^{-1}[\langle\gamma\rangle(3\alpha\langle\beta\rangle + 1) - 1](1 - \langle\gamma\rangle) + 3\langle\beta\rangle I[\gamma']\}^{-1} \times \{[\langle\gamma\rangle(3\alpha\langle\beta\rangle + 1) - 1](1 + 2\langle\gamma\rangle) + 3(\alpha\langle\beta\rangle + 1)I[\gamma']\}. \tag{76}$$

Equations (74) and (76) may now be used to obtain bounds on ϵ_{eff} after a suitable choice of values for α_1 and α_2 is made. In making this choice, we shall assume that one of the functions $\epsilon_i(x)$ actually takes on the value $\underline{\epsilon}$ on a set of positive probability measure (or on a set of positive volume in one ensemble member since we assume space and ensemble averages are equal) and also that the value $\bar{\epsilon}$ is achieved on a set of

positive measure. Then since

$$\langle\beta\rangle = \sum_{i=1}^{3} \langle(\epsilon_i(x) - \alpha)^{-1}\rangle,$$

$\langle\beta\rangle$ will approach $+\infty$ as $\alpha \to \underline{\epsilon} - 0$ and $-\infty$ as $\alpha \to \bar{\epsilon} + 0$. (These limits for α are suggested by perturbation theory.) Passage to the limit in (74) gives

$$\Gamma(\alpha) = \{3\alpha(\langle\epsilon\rangle - \alpha) + [\tfrac{1}{3}\langle\epsilon'_{ij}\epsilon'_{ji}\rangle + I[\epsilon']]\}^{-1} \times \{3\alpha\langle\epsilon\rangle(\langle\epsilon\rangle - \alpha) + \alpha[\tfrac{1}{3}\langle\epsilon'_{ij}\epsilon'_{ji}\rangle + I[\epsilon']]\}, \quad \alpha = \underline{\epsilon} \text{ or } \bar{\epsilon}. \quad (77)$$

A similar process in (76) yields

$$\Gamma(\alpha) = \alpha\{\langle\gamma\rangle(1 + 2\langle\gamma\rangle) + I[\gamma']\} \times \{\langle\gamma\rangle(1 - \langle\gamma\rangle) + I[\gamma']\}^{-1}, \quad (78)$$

where $\alpha = \underline{\epsilon}$ or $\bar{\epsilon}$, and $\langle\gamma\rangle$ and $I[\gamma']$ depend on α.

In summary, we note that the principal results of this section are contained in (63), (72), (77), and (78). These equations give three sets of upper and lower bounds on the effective permittivity. As we have already observed, (63) contains permittivity correlations up to and including the third order. Equation (77), on the other hand, gives a set of bounds which contains only one- and two-point permittivity correlations. Equation (78) also gives a set of bounds containing one- and two-point correlations, although these are not permittivity correlations. The bounds on ϵ_{eff} given in Ref. 2 are obtained from (78) by formally setting $I[\gamma'] = 0$. We have discussed these bounds elsewhere (see Molyneux[5]) and have shown that they are incorrect for arbitrary polycrystals.

4. AN EXACT EXPRESSION FOR ϵ_{eff}

In this section, we derive and discuss the perturbation series for the effective permittivity. We show that when certain integrals of the permittivity correlation tensors vanish, this perturbation series may be summed and an exact expression for ϵ_{eff} obtained.

The effective permittivity may be obtained directly from the correlation tensor $\langle\epsilon'_{ij}E'_j\rangle$ by the formula $\epsilon_{\text{eff}}\langle E_i\rangle = \langle\epsilon\rangle\langle E_i\rangle + \langle\epsilon'_{ij}E'_j\rangle$. Equations (8) and (9) give the formal integral equation for $E'_i(x)$,

$$E'_j(x) = -\langle\epsilon\rangle^{-1}\int G(|x - \xi|)\frac{\partial^2}{\partial\xi_j\partial\xi_k} \times [\epsilon'_{kl}(\xi)\langle E_l\rangle + \epsilon'_{kl}(\xi)E'_l(\xi)]\,d\xi, \quad (79)$$

and we shall find the perturbation series for $\langle\epsilon'_{ij}E'_j\rangle$ by iteration of this equation. We multiply (79) by $\epsilon'_{ij}(x)$ and average. Changing variables of integration from ξ to $r_1 = x - \xi$ and taking advantage of homogeneity to write $\langle\epsilon'_{ij}(x)\epsilon'_{kl}(\xi)\rangle = \langle\epsilon'_{kl}(0)\epsilon'_{ij}(r_1)\rangle$ and

$$\langle\epsilon'_{kl}(\xi)E'_l(\xi)\epsilon'_{ij}(x)\rangle = \langle\epsilon'_{kl}(0)E'_l(0)\epsilon'_{ij}(r_1)\rangle,$$

we find

$$\langle\epsilon'_{ij}E'_j\rangle = -\langle\epsilon\rangle^{-1}\int G_1 D_{i_1j_1} \times [\langle\epsilon_{li_1}(0)\epsilon_{j_1i}(r_1)\rangle\langle E_l\rangle + \langle\epsilon'_{li_1}(0)E'_l(0)\epsilon'_{j_1i}(r_1)\rangle]\,dr_1, \quad (80)$$

where we use the notation $G_1 = G(|r_1|)$, $D_{i_1j_1} = \partial^2/\partial r_{1i_1}\partial r_{1j_1}$. To evaluate the third-order correlation appearing in (80), we multiply (79) by $\epsilon'_{j_1i_1}(x)\epsilon'_{j_1i}(x + r_1)$ and average. Changing variables of integration from ξ to $r_2 = x - \xi$, using homogeneity, and adopting a notation similar to that used in (80), we obtain

$$\langle\epsilon'_{li_1}(0)E'_l(0)\epsilon'_{j_1i}(r_1)\rangle = -\langle\epsilon\rangle^{-1}\int G_2 D_{i_2j_2} \times [\langle\epsilon'_{li_2}(0)\epsilon'_{j_2i_1}(r_2)\epsilon'_{j_1i}(r_1 + r_2)\rangle\langle E_l\rangle + \langle\epsilon'_{li_2}(0)E'_l\epsilon'_{j_2i_1}(r_2)\epsilon'_{j_1i}(r_1 + r_2)\rangle]\,dr_2. \quad (81)$$

Combining (80) and (81), we find that to the fourth order

$$\langle\epsilon'_{ij}E'_j\rangle = \langle E_l\rangle\Big\{-\langle\epsilon\rangle^{-1}\int G_1 D_{i_1j_1}\langle\epsilon'_{li_1}(0)\epsilon'_{j_1i}(r_1)\rangle\,dr_1 + \langle\epsilon\rangle^{-2}\int G_1G_2 D_{i_1j_1}D_{i_2j_2} \times \langle\epsilon'_{li_2}(0)\epsilon'_{j_2i_1}(r_2)\epsilon'_{j_1i}(r_2 + r_1)\rangle\,dr_1\,dr_2\Big\} + \langle\epsilon\rangle^{-2}\int G_1G_2 D_{i_1j_1}D_{i_2j_2} \times \langle E'_l(0)\epsilon'_{li_2}(0)\epsilon'_{j_2i_1}(r_2)\epsilon'_{j_1i}(r_2 + r_1)\rangle\,dr_1\,dr_2. \quad (82)$$

A simple induction proof now shows that to order n,

$$\langle\epsilon'_{ij}E'_j\rangle = \langle E_l\rangle\Big\{\sum_{m=1}^{n}(-\langle\epsilon\rangle)^{-m} \times \int\prod_{k=1}^{m} G_k D_{i_kj_k}\langle\epsilon'_{li_m}(0)\epsilon'_{j_mi_{m-1}}(r_m)\cdots \epsilon'_{j_1i}(r_m + r_{m-1} + \cdots + r_1)\rangle\,dr_1\,dr_2\cdots dr_m\Big\} + (-\langle\epsilon\rangle)^{-n}\int\prod_{k=1}^{n} G_k D_{i_kj_k} \times \langle E'_l(0)\epsilon'_{li_n}(0)\epsilon'_{j_ni_{n-1}}(r_n)\cdots \epsilon'_{j_1i}(r_n + \cdots + r_1)\rangle\,dr_1\cdots dr_n. \quad (83)$$

We can partially integrate the first n terms in (83) by using (54). The result of applying (54) to the $m = 1$ and $m = 2$ terms is

$$\int G_1 D_{i_1j_1}\langle\epsilon'_{li_1}(0)\epsilon'_{j_1i}(r_1)\rangle\,dr_1 = \tfrac{1}{3}\langle\epsilon'_{lj_1}\epsilon'_{j_1i}\rangle + \int P_{i_1j_1}\langle\epsilon'_{li_1}(0)\epsilon'_{j_1i}(r_1)\rangle\,dr_1, \quad (84)$$

$$\int G_1 G_2 D_{i_1 j_1} D_{i_2 j_2} \langle \epsilon'_{l i_2}(0) \epsilon'_{j_2 i_1}(r_2) \epsilon'_{j_1 i}(r_2 + r_1) \rangle \, dr_1 \, dr_2$$
$$= \tfrac{1}{9} \langle \epsilon'_{l j_2} \epsilon'_{j_2 j_1} \epsilon'_{j_1 i} \rangle$$
$$+ \tfrac{1}{3} \int P_{i_1 j_1} \langle \epsilon'_{l j_2}(0) \epsilon'_{j_2 i_1}(0) \epsilon'_{j_1 i}(r_1) \rangle \, dr_1$$
$$+ \tfrac{1}{3} \int P_{i_2 j_2} \langle \epsilon'_{l i_2}(0) \epsilon'_{j_2 j_1}(r_2) \epsilon'_{j_1 i}(r_2) \rangle \, dr_2$$
$$+ \int P_{i_1 j_1} P_{i_2 j_2} \langle \epsilon'_{l i_2}(0) \epsilon'_{j_2 i_1}(r_2) \epsilon'_{j_1 i}(r_2 + r_1) \rangle \, dr_1 \, dr_2, \tag{85}$$

where

$$P_{i_1 j_1} = (4\pi)^{-1} |r_1|^{-3} (\delta_{i_1 j_1} - 3 |r_1|^{-2} r_{i_1} r_{j_1})$$

and $P_{i_2 j_2}$ is similarly defined. We cannot apply (54) directly to terms with $m \geq 3$. Consider, for example, the correlation function

$$\langle \epsilon'_{l i_3}(0) \epsilon'_{j_3 i_2}(r_3) \epsilon'_{j_2 i_1}(r_3 + r_2) \epsilon'_{j_1 i}(r_3 + r_2 + r_1) \rangle.$$

This approaches zero as $|r_3|$ and $|r_1|$ approach infinity, but as $|r_2|$ approaches infinity, it approaches

$$\langle \epsilon'_{l i_3}(0) \epsilon'_{j_3 j_2}(r_3) \rangle \langle \epsilon'_{j_2 i_1}(0) \epsilon'_{j_1 i}(r_1) \rangle.$$

Hence, we cannot apply (54) to the r_2 integration. To remedy this situation, we replace the correlation function appearing in the $m = 3$ term with the function

$$\langle \epsilon'_{l i_3}(0) \epsilon'_{j_3 i_2}(r_3) \epsilon'_{j_2 i_1}(r_3 + r_2) \epsilon'_{j_1 i}(r_3 + r_2 + r_1) \rangle$$
$$- \langle \epsilon'_{l i_3}(0) \epsilon'_{j_3 i_2}(r_3) \rangle \langle \epsilon'_{j_2 i_1}(0) \epsilon'_{j_1 i}(r_1) \rangle.$$

Since the subtracted term is independent of r_2, the value of the integral in question is unchanged, but (54) may now be applied to give

$$\int \prod_{k=1}^{3} G_k D_{i_k j_k} \langle \epsilon'_{l i_3}(0) \cdots \epsilon'_{j_1 i}(r_1 + r_2 + r_3) \rangle \, dr_1 \, dr_2 \, dr_3$$
$$= 3^{-3} [\langle \epsilon'_{l j_3} \epsilon'_{j_3 j_2} \epsilon'_{j_2 j_1} \epsilon'_{j_1 i} \rangle - \langle \epsilon'_{l j_3} \epsilon'_{j_3 j_2} \rangle \langle \epsilon'_{j_2 j_1} \epsilon'_{j_1 i} \rangle]$$
$$+ 3^{-2} \left\{ \int P_{i_1 j_1} [\quad] \, dr_1 + \int P_{i_2 j_2} [\quad] \, dr_2 \right.$$
$$\left. + \int P_{i_3 j_3} [\quad] \, dr_3 \right\}$$
$$+ 3^{-1} \left\{ \int P_{i_1 j_1} P_{i_2 j_2} [\quad] \, dr_1 \, dr_2 \right.$$
$$+ \int P_{i_1 j_1} P_{i_3 j_3} [\quad] \, dr_1 \, dr_3 + \left. \int P_{i_2 j_2} P_{i_3 j_3} [\quad] \, dr_2 \, dr_3 \right\}$$
$$+ \int P_{i_1 j_1} P_{i_2 j_2} P_{i_3 j_3} [\quad] \, dr_1 \, dr_2 \, dr_3, \tag{86}$$

where, for example, the first square bracket in the above sequence of integrals is

$$\langle \epsilon'_{l j_3}(0) \epsilon'_{j_3 j_2}(0) \epsilon'_{j_2 i_1}(0) \epsilon'_{j_1 i}(r_1) \rangle - \langle \epsilon'_{l j_3} \epsilon'_{j_3 j_2} \rangle \langle \epsilon'_{j_2 i_1}(0) \epsilon'_{j_1 i}(r_1) \rangle.$$

We have not written out the integrands in (86) in great detail since we are going to assume that the correlation functions of the permittivity tensor are such that they all vanish. Specifically, we assume that for all $m = 1, 2, \cdots$ the integrals which result from applying (54) to the term

$$\int \prod_{1}^{m} G_k D_{i_k j_k} \times \langle \epsilon'_{l i_m}(0) \cdots \epsilon'_{j_1 i}(r_m + r_{m-1} + \cdots + r_1) \rangle \, dr_1 \cdots dr_m$$

all vanish (for $m \geq 3$ we must first modify the integrand as outlined above). This assumption places an infinite number of conditions on the correlation functions of the random function $\epsilon'_{ij}(x)$. It is easy to show that these conditions are satisfied by the sequence of correlations

$$\langle \epsilon'_{i_0 j_0}(0) \epsilon'_{i_1 j_1}(r_1) \epsilon'_{i_2 j_2}(r_1 + r_2) \cdots \epsilon'_{i_n j_n}(r_1 + \cdots + r_n) \rangle$$
$$= \sum_{m=1}^{s} F^{(m)}(|r_1|, \cdots, |r_n|) T^{(m)}_{i_0 j_0 \cdots i_n j_n},$$

where $T^{(m)}_{i_0 j_0 \cdots i_n j_n}$ are constant isotropic tensors of the order $(2n + 2)$ and the sum goes over all possible such tensors. This sequence of correlations is fairly trivial, however, since each function depends only on the magnitudes of its arguments. On the other hand, there is some reason to believe all the integrals in question may be small in many cases. In fact, consider an integral like $\int P_{i_1 j_1} [\quad] \, dr_1$. The term in the square brackets will be a tensor function which is essentially zero for $|r_1|$, of order of a correlation length, l. If, for $|r_1| < l$, the term in the square brackets is weakly dependent on the angle variables, then the value of the integral should be small since the Cauchy principle value of $\int P_{i_1 j_1} \, dr_1$ is zero. In summary, the assumptions outlined above will yield an exact expression for $\langle \epsilon'_{ij} E'_j \rangle$ for any polycrystal whose permittivity correlation tensors satisfy a certain infinite set of integral relations, or will yield an approximate expression for $\langle \epsilon'_{ij} E'_j \rangle$, in case the correlation functions are weakly dependent on angle variables.

To find $\langle \epsilon'_{ij} E'_j \rangle$, we must now compute all terms in the series (83). We have shown how this is done up to the fourth order. The calculation of higher-order terms proceeds in exactly the same manner but is fairly lengthy. We therefore give only the result

$$\langle \epsilon'_{ij} E'_j \rangle = 3 \langle \epsilon \rangle \langle E_i \rangle$$
$$\times \left\{ \sum_{m=2}^{\infty} (-1)^{m-1} (3\langle \epsilon \rangle)^{-m} \left[\langle \epsilon'_{l j_1} \epsilon'_{j_1 j_2} \cdots \epsilon'_{j_{m-1} i} \rangle \right.\right.$$
$$+ \sum_{k=2}^{[\frac{1}{2}m]'} (-1)^{k-1} \langle \epsilon'_{l j_1} \epsilon'_{j_1 j_2} \cdots \epsilon'_{j_{n_1 - 1} j_{n_1}} \rangle$$
$$\times \langle \epsilon'_{j_{n_1} j_{n_1 + 1}} \cdots \epsilon'_{j_{n_1 + n_2 - 1} j_{n_1 + n_2}} \rangle \cdots$$
$$\left.\left. \langle \epsilon'_{j_{n_1 + n_2 + \cdots + n_{k-1}} j_{n_1 + n_2 + \cdots + n_{k-1} + 1}} \cdots \epsilon'_{j_{n_1 + n_2 + \cdots + n_k - 1} i} \rangle \right] \right\}, \tag{87}$$

where $[\frac{1}{2}m]$ indicates the integral part of $\frac{1}{2}m$ and the prime on the second summation means that the sum is defined as zero for $m = 2, 3$ and for $m \geq 4$ goes over all distinct sequences of integers $(n_1, n_2, \cdots, n_k)$ such that $\sum_{j=1}^{k} n_j = m$ and $n_j \geq 2; j = 1, 2, \cdots, k$.

We can write (87) in a much more compact form if we introduce some new notation. Let δ be the matrix with components $e_{ij}(x) = (3\langle\epsilon\rangle)^{-1}\epsilon'_{ij}(x)$ and denote the trace of a matrix $\mathcal{A}$ by Tr $\mathcal{A}$. Then it is easy to see that (87) can be written in the form

$$\langle\epsilon'_{ij}E'_j\rangle = -\langle\epsilon\rangle\langle E_i\rangle \operatorname{Tr}\left\{\sum_{m=2}^{\infty}(-1)^m \times\left[\langle\delta^m\rangle + \sum_{k=2}^{[\frac{1}{2}m]'}(-1)^{k-1}\langle\delta^{n_1}\rangle\langle\delta^{n_2}\rangle\cdots\langle\delta^{n_k}\rangle\right]\right\}, \tag{88}$$

where $\langle\delta^p\rangle$ is the matrix with components

$$\langle e_{ik_1}e_{k_1k_2}\cdots e_{k_{p-1}j}\rangle$$

and $\langle\delta^p\rangle\langle\delta^q\rangle$ denotes the matrix product of $\langle\delta^p\rangle$ and $\langle\delta^q\rangle$. To calculate the sum in (88) in closed form, we let t be real number, $|t| < 1$, and define

$$\phi(t) = \frac{t}{1+t} = \sum_{k=1}^{\infty}(-1)^{k+1}t^k. \tag{89}$$

Suppose y is a real random variable, $\langle y\rangle = 0$ and $|y| < \frac{1}{2}$. Then, if s is a real number $0 \leq s \leq 1$, it follows that $|sy(1+sy)^{-1}| < 1$ and, by expanding $[-sy(1+sy)^{-1}]$ in a power series and averaging, it is possible to show that

$$\langle[-sy(1+sy)^{-1}]\rangle^m = \sum_{k=2m}^{\infty}(-1)^k s^k \sum{}'' \langle y^{n_1}\rangle\langle y^{n_2}\rangle\cdots\langle y^{n_k}\rangle, \tag{90}$$

where $\sum''$ denotes summation over all sequences of integers $(n_1, n_2, \cdots, n_k)$ which satisfy $n_i \geq 2$ and $\sum_{j=1}^{k} n_j = m$. Now let $\psi(s) = \phi(\langle -sy(1+sy)^{-1}\rangle)$ and, using (89) and (90), write $\psi(s)$ as a power series in s. This power series may be shown to be of the form

$$\psi(s) = \phi(\langle -sy(1+sy)^{-1}\rangle) = \sum_{m=2}^{\infty} A_m s^m, \tag{91}$$

where

$$A_m = (-1)^m\left\{\langle y^m\rangle + \sum_{k=2}^{[\frac{1}{2}m]'}(-1)^{k-1}\langle y^{n_1}\rangle\langle y^{n_2}\rangle\cdots\langle y^{n_k}\rangle\right\}, \tag{92}$$

and the prime on the sum denotes the same restrictions given after (87). The principal values $e_i(x)$ of the matrix $\delta(x)$ satisfy

$$|e_i(x)| = |(3\langle\epsilon\rangle)^{-1}(\delta_i(x) - \langle\epsilon\rangle)| \leq (3\langle\epsilon\rangle)^{-1}(\bar{\epsilon} - \langle\epsilon\rangle) < \tfrac{1}{2}, \quad \text{for} \quad \bar{\epsilon} < \tfrac{5}{2}\langle\epsilon\rangle. \tag{93}$$

Hence, according to the theory of functions of matrices, the formal series in (88) converges for $\bar{\epsilon} < (\frac{5}{2})\langle\epsilon\rangle$ and represents the matrix $\phi(\langle -\delta(I+\delta)^{-1}\rangle)$ where I is the identity matrix, and $\mathcal{A}^{-1}$ denotes the inverse of matrix $\mathcal{A}$. Thus, (88) can be written in the form

$$\langle\epsilon'_{ij}E'_j\rangle = \langle\epsilon\rangle\langle E_i\rangle \operatorname{Tr}\{\langle\delta(I+\delta)^{-1}\rangle \times [I - \langle\delta(I+\delta)^{-1}\rangle]^{-1}\}. \tag{94}$$

We note that since

$$1 + \epsilon'_i(x)/3\langle\epsilon\rangle = 1 + (\epsilon_i(x) - \langle\epsilon\rangle)(3\langle\epsilon\rangle)^{-1} \geq \tfrac{2}{3},$$

the inverse of $I + \delta$ always exists. Also,

$$I - \langle\delta(I+\delta)^{-1}\rangle = \left\{3^{-1}\sum_{i=1}^{3}\langle[1 + (3\langle\epsilon\rangle)^{-1}\epsilon'_i(x)]^{-1}\rangle\right\}I,$$

so the inverse of this matrix always exists. Thus (94) may be used to represent $\langle\epsilon'_{ij}E'_j\rangle$ even when the condition $\bar{\epsilon} < (\frac{5}{2})\langle\epsilon\rangle$ is not satisfied.

To evaluate ϵ_{eff} from (94), we recall that $\epsilon_{\text{eff}}\langle E_i\rangle = \langle D_i\rangle = \langle\epsilon\rangle\langle E_i\rangle + \langle\epsilon'_{ij}E'_j\rangle$. Using this relation and calculating the various correlations appearing in (94), we find

$$\epsilon_{\text{eff}} = \langle\epsilon\rangle\left\{-2 + 9\left[\sum_{i=1}^{3}\langle[1 + (3\langle\epsilon\rangle)^{-1}\epsilon'_i(x)]^{-1}\rangle\right]^{-1}\right\}. \tag{95}$$

For a single phase polycrystal with principal permittivities ϵ_i, (95) becomes

$$\epsilon_{\text{eff}} = \langle\epsilon\rangle\left\{1 + \left[\sum_{i=1}^{3}\langle\epsilon\rangle(2\langle\epsilon\rangle + \epsilon_i)^{-1}\right]^{-1} \times \sum_{i=1}^{3}(\epsilon_i - \langle\epsilon\rangle)^{-1}(2\langle\epsilon\rangle + \epsilon_i)\right\}. \tag{96}$$

Since the effective permittivity satisfies the equation $\langle D_i\rangle = \epsilon_{\text{eff}}[\langle\theta\rangle\langle D_i\rangle + \langle\theta'_{ij}D'_j\rangle]$, where θ'_{ij} is the tensor defined in (19), we see that another expression for ϵ_{eff} may be found by deriving the perturbation series for the correlation $\langle\theta'_{ij}D'_j\rangle$. We will now derive this series, and by imposing certain requirements on the correlation functions of $\theta'_{ij}(x)$, we will be able to obtain a second exact expression for ϵ_{eff}.

From (2) and (3) we find that D' satisfies

$$\frac{\partial D'_i}{\partial x_i} = 0, \tag{97}$$

$$\delta_{ijk}\frac{\partial}{\partial x_j}[\langle\theta\rangle D'_k + \theta'_{km}\langle D_m\rangle + \theta'_{km}D'_m] = 0. \tag{98}$$

Equations (97) and (98) may be used to obtain a formal integral equation for the fluctuating part of the displacement field:

$$D'_i(x) = \langle\theta\rangle^{-1}\int G(x,\xi)\mathcal{A}_{im}(\xi) \times [\theta'_{mn}(\xi)\langle D_n\rangle + \theta'_{mn}(\xi)D'_n(\xi)]\,d\xi, \tag{99}$$

where $\mathcal{A}_{im}(\xi) = [\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}]\partial^2/\partial\xi_j\partial\xi_l$. Since (99) is similar in form to (79), we see from (83) that to any order n,

$$\langle\theta'_{ij}D'_j\rangle = \langle D_l\rangle\Big\{\sum_{m=1}^{n}\langle\theta\rangle^{-m}\int\prod_{k=1}^{m}G_k\mathcal{A}_{i_kj_k}(r_k) \times \langle\theta'_{li_m}(0)\theta'_{j_mi_{m-1}}(r_m)\cdots \theta'_{j_1i}(r_m + r_{m-1} + \cdots + r_1)\rangle\, dr_1\cdots dr_m\Big\} + \langle\theta\rangle^{-n}\int\prod_{i=1}^{n}G_k\mathcal{A}_{i_kj_k}(r_k)\langle D'_l(0)\theta'_{li_n}(0)\cdots \theta'_{j_1i}(r_n + r_{n-1} + \cdots + r_1)\rangle\, dr_1\cdots dr_n. \quad (100)$$

Just as in the calculation of $\langle\epsilon'_{ij}E'_j\rangle$, we may use (54) to rewrite the integrals which appear in (100). We thereby obtain a series of terms involving one-point, mth-order θ'_{ij} correlations plus a series of integrals containing contracted products of operators P_{ij} and θ'_{ij} correlation tensors. For example, the $m = 1$ term in (100) gives

$$\int G_1\mathcal{A}_{i_1j_1}(r_1)\langle\theta'_{li_1}(0)\epsilon'_{j_1i}(r_1)\rangle\, dr_1 = -(\tfrac{2}{3})\langle\theta'_{lj_1}\theta'_{j_1i}\rangle + \int P_{i_1j_1}\langle\theta'_{li_1}(0)\theta'_{j_1i}(r_1)\rangle\, dr_1. \quad (101)$$

We assume, as before, that the random function $\theta'_{ij}(x)$ is such that all integrals which arise from these manipulations vanish. Then, as with $\langle\epsilon'_{ij}E'_j\rangle$, we can sum the remaining series. Since the details of these calculations are rather lengthy, we give only the final result

$$\langle\theta'_{ij}D'_j\rangle = \tfrac{1}{2}\langle\theta\rangle\langle D_i\rangle\,\mathrm{Tr}\,\{\langle\Theta(I + \Theta)^{-1}\rangle \times [I - \langle\Theta(I + \Theta)^{-1}\rangle]^{-1}\}, \quad (102)$$

where Θ is the matrix with components $(\frac{2}{3})\langle\theta\rangle^{-1}\theta'_{ij}$. For a single phase polycrystal, we find using (102) and the definition of ϵ_{eff} that

$$\epsilon_{\mathrm{eff}} = \Big[\sum_{i=1}^{3}(\langle\theta\rangle + 2\theta_i)^{-1}\theta_i\Big]^{-1}\sum_{i=1}^{3}(\langle\theta\rangle + 2\theta_i)^{-1}. \quad (103)$$

5. COMPARISON OF BOUNDS

In this section, we will compare, for the case of a single-phase polycrystal, the three sets of bounds we have derived. If the principal permittivities are expressed in the form

$$\epsilon_1 = \epsilon, \quad \epsilon_2 = \epsilon(1 + s), \quad \epsilon_3 = \epsilon(1 + t); \quad 0 \le s \le t, \quad (104)$$

then we can write each lower bound, L_i, (where the subscript indicates the order in which the bounds were obtained) in the form

$$L_i(s, t) = \epsilon[F_i^{(2)}(s, t) + I_i^{(2)}(s, t)]^{-1}[F_i^{(1)}(s, t) + I_i^{(1)}(s, t)], \quad i = 1, 2, 3, \quad (105)$$

and each upper bound U_i in the form

$$U_i(s, t) = \epsilon[G_i^{(2)}(s, t) + J_i^{(2)}(s, t)]^{-1} \times [G_i^{(1)}(s, t) + J_i^{(1)}(s, t)]. \quad (106)$$

In (105) and (106), $F_i^{(k)}$ and $G_i^{(k)}$ are rational functions of s and t and arise from the various one-point correlations which appear in the bounds, while $I_i^{(k)}$ and $J_i^{(k)}$ involve integrals of correlation functions.

For the bounds L_1 and U_1 derived from the classical variational principles, we find

$$F_1^{(1)}(s, t) = s^2(1 + s)^{-3}(1 + \tfrac{4}{9}s) + t^2(1 + t)^{-3}(1 + \tfrac{4}{9}t) - st(1 + s)^{-1}(1 + t)^{-1} \times [1 - \tfrac{1}{3}s(1 + s)^{-1} - \tfrac{1}{3}t(1 + t)^{-1}], \quad (107)$$

$$F_1^{(2)}(s, t) = s^2(1 + s)^{-4}[1 + \tfrac{10}{9}s + \tfrac{4}{27}s^2] + t^2(1 + t)^{-4}[1 + \tfrac{10}{9}t + \tfrac{4}{27}t^2] - st(1 + s)^{-1}(1 + t)^{-1}[1 - \tfrac{1}{3}s(1 + s)^{-1} - \tfrac{1}{3}t(1 + t)^{-1} - \tfrac{10}{27}s^2(1 + s)^{-2} - \tfrac{10}{27}t^2(1 + t)^{-2} + \tfrac{2}{3}st(1 + s)^{-1}(1 + t)^{-1}], \quad (108)$$

$$G_1^{(1)}(s, t) = s^2[1 + \tfrac{7}{9}s + \tfrac{2}{27}s^2] + t^2[1 + \tfrac{7}{9}t + \tfrac{2}{27}t^2] - st[1 + \tfrac{1}{6}(s + t) - \tfrac{13}{54}(s^2 + t^2)], \quad (109)$$

$$G_1^{(2)}(s, t) = s^2(1 + \tfrac{4}{9}s) + t^2(1 + \tfrac{4}{9}t) - st[1 + \tfrac{1}{6}(s + t)], \quad (110)$$

$$I_1^{(1)}(s, t) = -\tfrac{27}{4}\epsilon^3[-\langle\theta\rangle P + Q], \quad (111)$$

$$I_1^{(2)}(s, t) = -\tfrac{27}{4}\epsilon^4[P(-\langle\theta\rangle^2 + \tfrac{2}{9}\langle\theta'_{ij}\theta'_{ij}\rangle - P) + \langle\theta\rangle Q], \quad (112)$$

$$J_1^{(1)}(s, t) = \tfrac{27}{2}\epsilon^{-4}[K(\langle\epsilon\rangle^2 - \tfrac{2}{9}\langle\epsilon'_{ij}\epsilon'_{ij}\rangle - K) + \langle\epsilon\rangle L], \quad (113)$$

$$J_1^{(2)}(s, t) = -\tfrac{27}{2}\epsilon^{-3}[\langle\epsilon\rangle K + L]. \quad (114)$$

As a special case of the above results, we consider a polycrystal whose principal permittivities are $\epsilon_1 = \epsilon_2 = \epsilon$ and $\epsilon_3 = \epsilon(1 + t)$. When the permittivity correlation tensors of such a medium satisfy the restrictions imposed in the previous section, the integrals K and L vanish and the upper bound becomes

$$U_1(0, t) = \epsilon(1 + \tfrac{4}{9}t)^{-1}(1 + \tfrac{2}{3}t)(1 + \tfrac{1}{9}t). \quad (115)$$

When the inverse of the permittivity tensor has correlations which satisfy the restrictions imposed in the previous section, the integrals P and Q vanish and the

lower bound becomes

$$L_1(0, t) = \epsilon(1 + \tfrac{10}{9}t + \tfrac{4}{27}t^2)^{-1}(1 + t)(1 + \tfrac{4}{9}t). \tag{116}$$

For the bounds L_2 and U_2, derived from the generalized Hashin variational principle with trial function $C_i' = \epsilon_{ij}'\langle E_j\rangle$, we find

$$F_2^{(1)}(s, t) = s(1 + \tfrac{5}{9}s) + t(1 + \tfrac{5}{9}t) + \tfrac{4}{9}st, \tag{117}$$

$$F_2^{(2)}(s, t) = s(1 + \tfrac{2}{9}s) + t(1 + \tfrac{2}{9}t) - \tfrac{2}{9}st, \tag{118}$$

$$I_2^{(1)}(s, t) = I_2^{(2)}(s, t) = 3\epsilon^{-2}K, \tag{119}$$

$$G_2^{(1)}(s, t) = (1 + t)[2t(1 + \tfrac{2}{9}t) - s(1 + \tfrac{5}{9}s) + \tfrac{5}{9}st], \tag{120}$$

$$G_2^{(2)}(s, t) = 2t(1 + \tfrac{8}{9}t) - s(1 + \tfrac{2}{9}s) - \tfrac{7}{9}st, \tag{121}$$

$$(1 + t)^{-1}J_2^{(1)}(s, t) = J_2^{(2)}(s, t) = -3\epsilon^{-2}K. \tag{122}$$

In the special cases considered in the previous paragraph, these bounds reduce to

$$U_2(0, t) = \epsilon(1 + \tfrac{8}{9}t)^{-1}(1 + t)(1 + \tfrac{2}{9}t) \tag{123}$$

and

$$L_2(0, t) = \epsilon(1 + \tfrac{2}{9}t)(1 + \tfrac{5}{9}t). \tag{124}$$

Finally, for the bounds L_3 and U_3 derived from the generalized Hashin principle with γ_{ij} satisfying

$$\gamma_{ij}(3\alpha\beta_{jk} + \delta_{jk}) = \delta_{ik},$$

we obtain

$$F_3^{(1)}(s, t) = 9 + 5s + 5t + \tfrac{7}{3}st, \tag{125}$$

$$F_3^{(2)}(s, t) = 9 + 2s + 2t + \tfrac{1}{3}st, \tag{126}$$

$$I_3^{(1)}(s, t) = I_3^{(2)}(s, t) = 3[s(3 + t) + t(3 + s)]^{-1} \times [(3 + s)(3 + t)]^2 I[\gamma_1'], \tag{127}$$

$$G_3^{(1)}(s, t) = (1 + t)[(3 + 2t)(9 + 2t) + s(15 + 8t)], \tag{128}$$

$$G_3^{(2)}(s, t) = (3 + 2t)(9 + 8t) + s(6 + 5t), \tag{129}$$

$$(1 + t)^{-1}J_3^{(1)}(s, t) = J_3^{(2)}(s, t) = -9[t(3 + 2t + s) + (t - s)(3 + 2t)]^{-1} \times [(3 + 2t)(3 + 2t + s)]^2 I[\gamma_3'], \tag{130}$$

where γ_1' and γ_3' denote the fluctuating part of γ_{ij} evaluated for $\alpha = \epsilon_1 = \epsilon$ and $\alpha = \epsilon_3 = \epsilon(1 + t)$, respectively. For the special cases considered above, we find

$$U_3(0, t) = \epsilon(1 + \tfrac{8}{9}t)^{-1}(1 + t)(1 + \tfrac{2}{9}t) \tag{131}$$

and

$$L_3(0, t) = \epsilon(1 + \tfrac{2}{9}t)^{-1}(1 + \tfrac{5}{9}t). \tag{132}$$

The exact expressions for ϵ_{eff}, derived in the previous section, may be written in terms of ϵ, s, and t. If we denote the two formulas for effective permittivity by $\epsilon_{\text{eff}}^{(i)}$, $i = 1, 2$, where superscript (1) denotes the first expression derived (for special permittivity-tensor correlations) and superscript (2) denotes the second expression derived (for special inverse permittivity-tensor correlations), then we can write

$$\epsilon_{\text{eff}}^{(i)} = \epsilon\Phi_i(s, t)/\Psi_i(s, t), \tag{133}$$

where

$$\Phi_1(s, t) = 4(s + t)(t^2 + 5st + s^2) + 3(20t^2 + 61st + 20s^2) + 243(s + t) + 243, \tag{134}$$

$$\Psi_1(s, t) = 3(8t^2 + 19ts + 8s^2 + 54(s + t) + 81), \tag{135}$$

$$\Phi_2(s, t) = 3(1 + s)(1 + t)[5s^2t^2 + 32(t^2s + s^2t) + 32(t^2 + s^2) + 130st + 108(s + t) + 81], \tag{136}$$

$$\Psi_2(s, t) = t^3s^3 + 18(t^3s^2 + s^3t^2) + 48(t^3s + s^3t) + 249t^2s^2 + 32(t^3 + s^3) + 504(t^2s + s^2t) + 276(t^2 + s^2) + 939ts + 486(t + s) + 243. \tag{137}$$

When $s = 0$,these formulas reduce to

$$\epsilon_{\text{eff}}^{(1)}(0, t) = \epsilon(1 + \tfrac{4}{9}t)^{-1}(1 + \tfrac{7}{9}t + \tfrac{2}{27}t^2) \tag{138}$$

and

$$\epsilon_{\text{eff}}^{(2)}(0, t) = \epsilon(1 + \tfrac{10}{9}t + \tfrac{4}{27}t^2)^{-1}(1 + t)(1 + \tfrac{4}{9}t). \tag{139}$$

It is easiest and most informative to compare these formulas for the two special media described in the previous section. Since all the bounds coincide when $s = t = 0$, we consider only the case when $s + t > 0$. We can then show that

$$U_1 < U_3 \leq U_2, \tag{140}$$

where the equality sign holds only for $s = 0$ or for $s = t$. We also have

$$L_2 \leq L_3 < L_1, \tag{141}$$

where the equality sign holds only for $s = 0$ or for $s = t$. Furthermore, if we compare (138) and (139)

with (115) and (116), we find that

$$U_1(0, t) = \epsilon^{(1)}(0, t); \quad L_1(0, t) = \epsilon^{(2)}(0, t). \quad (142)$$

Thus, U_1 and L_1 are the best possible upper and lower bounds on the effective permittivity. That is, if U and L are any other upper and lower bounds on ϵ_{eff} which hold for all possible stochastic tensors $\epsilon'_{ij}(x)$, then $U_1 \leq U$ and $L \leq L_1$.

APPENDIX

The derivation of (54) is quite straightforward. We let $s_R = \{r:|r| < R\}$, $s_\epsilon = \{r:|r| > \epsilon\}$, and $s_{R,\epsilon} = s_R - s_\epsilon$, and we consider the integral

$$I^{(R,\epsilon)}_{i_1 \cdots i_{k-1} i_{k+1} \cdots i_{l-1} i_{l+1} \cdots i_{2n}}$$

$$\equiv \int_{S_{R,\epsilon}} G(|r|) \frac{\partial^2 T_{i_1 \cdots i_k \cdots i_l \cdots i_{2n}}}{\partial r_{i_k} \partial r_{i_l}} \, dr$$

$$= \int_{S_{R,\epsilon}} \left\{ \frac{\partial}{\partial r_{i_k}} \left[G \frac{\partial T_{i_1 \cdots i_{2n}}}{\partial r_{i_2}} \right] - \frac{\partial}{\partial r_{i_l}} \left[\frac{\partial G}{\partial r_{i_k}} T_{i_1 \cdots i_{2n}} \right] + \frac{\partial^2 G}{\partial r_{i_k} \partial r_{i_l}} T_{i_1 \cdots i_{2n}} \right\} dr. \quad (A1)$$

Using the divergence theorem, we find

$$\int_{S_{R,\epsilon}} \frac{\partial}{\partial r_{i_k}} \left[G \frac{\partial T_{i \cdots i_{2n}}}{\partial r_{i_l}} \right] dr$$

$$= -\tfrac{1}{4}\pi \int_{\partial S_R} \frac{1}{R} \frac{\partial T_{i_1 \cdots i_{2n}}}{\partial r_{i_l}} n_{i_k} \, d\sigma + \tfrac{1}{4}\pi \int_{\partial S_\epsilon} \frac{1}{\epsilon} \frac{\partial T_{i_1 \cdots i_{2n}}}{\partial r_{i_l}} n_{i_k} \, d\sigma, \quad (A2)$$

$$\int_{S_{R,\epsilon}} \frac{\partial}{\partial r_{i_l}} \left[\frac{\partial G}{\partial r_{i_k}} T_{i_1 \cdots i_{2n}} \right] dr$$

$$= \tfrac{1}{4}\pi \int_{\partial S_R} \frac{1}{R^2} T_{i_1 \cdots i_{2n}} n_{i_k} n_{i_l} \, d\sigma - \tfrac{1}{4}\pi \int_{\partial S_\epsilon} \frac{1}{\epsilon^2} T_{i_1 \cdots i_{2n}} n_{i_k} n_{i_l} \, d\sigma, \quad (A3)$$

where n_k denotes $r_k/|r|$. Letting $R \to \infty$ and $\epsilon \to 0$, we see that (54) results, and we note that the integral involved in this equation is actually a Cauchy principle value. The fact that the triple integral

$$J_{i_1 \cdots i_{k-1} i_{k+1} \cdots i_{l-1} i_{l+1} \cdots i_{2n}} \equiv \tfrac{1}{4}\pi \int \frac{1}{|r|^3} \Pi_{i_k i_l}(r) T_{i_1 \cdots i_{2n}}(r) \, dr \quad (A4)$$

can always be written as an integral of a function of a single scalar variable also follows quite easily. Since $J_{i_1 \cdots i_{2n}}$ is a constant isotropic tensor of order $2n - 2$, it can be written as a sum of terms each of which is of the form $A \prod_{\nu=1}^{n-1} \delta_{i_{p_\nu} i_{q_\nu}}$, where $\{p_1, \cdots, p_{n-1}, q_1, \cdots, q_{n-1}\}$ is a permutation of $\{1, 2, \cdots, k-1, k+1, \cdots, l-1, l+1, \cdots, 2n\}$ and A is a combination of completely contracted forms of $J_{i_1 \cdots i_{2n}}$. It is obvious that each such contracted form of $J_{i_1 \cdots i_{2n}}$ can be written as an integral of a scalar function.

[1] M. S. Bartlett, *An Introduction to Stochastic Processes* (Cambridge University Press, New York, 1955).
[2] Z. Hashin and S. Shtrikman, Phys. Rev. **130**, 129 (1963).
[3] M. Beran, Nuovo Cimento **38**, 771 (1965).
[4] M. Beran and J. Molyneux, Quart. Appl. Math. **24**, 107 (1966).
[5] J. Molyneux, J. Math. Phys. **10**, 912 (1969).
[6] G. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge University Press, New York, 1953).

Spinors in a Weyl Geometry

RONALD J. ADLER
Department of Physics, Virginia Polytechnic Institute, Blacksburg, Virginia 24061

(Received 24 February 1969)

It was noted by London in 1927 that a mathematical connection exists between the unified field theory of Weyl and the old quantum theory of Bohr. We wish to demonstrate that a Dirac electron in an electromagnetic field described by the Weyl formalism appears to couple to the field via the usual minimal coupling recipe. It is generally acknowledged that the Weyl theory is unsatisfactory for describing macroscopic electromagnetic phenomena. The objections, however, may not be entirely applicable on a quantum-mechanical scale.

1. WEYL'S UNIFIED FIELD THEORY

Shortly after the successful introduction of general relativity theory to explain gravitational phenomena, Weyl attempted to imbed electromagnetic phenomena into a similar geometric framework by altering the notion of vector transplantation.[1,2] In this section we wish to call attention to the basic physical assumptions introduced by Weyl.

In normal general relativity theory the notion of parallel displacement is generalized so as to be applicable to more general coordinates than simple Cartesian systems. Specifically, one assumes the existence of an affine relation for the change in the components of a parallel displaced vector ξ^α,

$$d\xi^\alpha = \Gamma^\alpha_{\beta\gamma}\xi^\beta\, dx^\gamma, \tag{1.1}$$

where the coefficients of affine connection are symmetric in the lower indices[2]

$$\Gamma^\alpha_{\beta\gamma} = \Gamma^\alpha_{\gamma\beta}. \tag{1.2}$$

The demand that the length squared of a parallel displaced vector $\xi^\alpha\xi_\alpha$ not change then leads to a unique identification of the coefficients of affine connection with the Christoffel symbols[2]

$$\Gamma^\alpha_{\beta\gamma} = -\begin{Bmatrix}\alpha\\ \beta\gamma\end{Bmatrix} \equiv -\tfrac{1}{2}g^{\alpha\sigma}(g_{\gamma\sigma|\beta} + g_{\beta\sigma|\gamma} - g_{\beta\gamma|\sigma}). \tag{1.3}$$

From this result the familiar properties of conventional Riemann space follow in relatively straightforward fashion.

To expand this structure, Weyl generalized the Riemann space by allowing $\xi^2 = \xi^\alpha\xi_\alpha$ to change under parallel displacement. The change in ξ^2 is assumed to be given by a simple affine relation

$$d(\xi^2) = \phi_\beta\, dx^\beta\xi^2. \tag{1.4}$$

The four quantities ϕ_β thus introduced are intended to correspond to the electromagnetic vector potential and thereby provide a geometric framework for a description of the electromagnetic field. The conditions (1.1) and (1.4) lead to a set of coefficients of affine connection given by

$$\begin{aligned}\Gamma^\alpha_{\beta\gamma} &= -\begin{Bmatrix}\alpha\\ \beta\gamma\end{Bmatrix} + g^{\alpha\sigma}(g_{\gamma\sigma}\phi_\beta + g_{\beta\sigma}\phi_\gamma - g_{\beta\gamma}\phi_\sigma)\\ &= -\tfrac{1}{2}g^{\alpha\sigma}[(g_{\gamma\sigma|\beta} - 2g_{\gamma\sigma}\phi_\beta)\\ &\quad + (g_{\beta\sigma|\gamma} - 2g_{\beta\sigma}\phi_\gamma) - (g_{\beta\gamma|\sigma} - 2g_{\beta\gamma}\phi_\sigma)].\end{aligned} \tag{1.5}$$

We shall refer to a space with the above coefficients of affine connection as an extended Riemann space or, more succinctly, as a Weyl space.

The classical unified field theory which results from the above assumption is discussed in Refs. 1 and 2. It is not a particularly successful theory, although several rather elegant formal results do occur. We are not further concerned with the classical theory in the present paper but proceed directly to the possible relevance of Weyl space to quantum phenomena.

2. WEYL SPACE AND BOHR'S QUANTUM THEORY

It was pointed out by London in 1927[3] that Weyl's introduction of the affine displacement law (1.4) has a close mathematical connection with the idea of stationary phase inherent in the preliminary quantum theory of Bohr. We will illustrate this connection for the very simple case of circular orbits.[2]

Consider a single electron moving in the Coulomb field of a proton. We will assume that ϕ_β in Eq. (1.4) is proportional to the vector potential A_β of the Coulomb field

$$\phi_\beta = \lambda(e/r, 0, 0, 0) = \lambda A_\beta, \tag{2.1}$$

where r is the electron–proton separation. Then the affine relation (1.4) reads

$$dl = (\lambda e/r)lc\, dt, \tag{2.2}$$

where l is the length squared of any parallel displaced

vector associated with the electron. Since r is a constant, Eq. (2.2) yields, for a single orbit,

$$l = l_0 \exp(\lambda e T c/r), \tag{2.3}$$

where T is the period of the orbit. Elementary mechanics yields

$$T = 2\pi m^{\frac{1}{2}} r^{\frac{3}{2}}/e. \tag{2.4}$$

Thus, after one orbit, l has become

$$l = l_0 \exp[2\lambda\pi c(mr)^{\frac{1}{2}}]. \tag{2.5}$$

Certain orbits are clearly distinguished by this condition: those for which $l = l_0$ or for which we have stationary phase and a kind of length stability:

$$2\pi\lambda c(mr)^{\frac{1}{2}} = 2\pi i n, \quad r = -n^2/\lambda^2 c^2 m. \tag{2.6}$$

This may be compared with Bohr's expression for the allowed radii in hydrogen:

$$r = n^2\hbar^2/e^2 m. \tag{2.7}$$

The form of the expressions is the same, and they are identical if we choose

$$\lambda = \pm ie/\hbar c, \quad \phi_\beta = \pm i(e/\hbar c)A_\beta. \tag{2.8}$$

Thus, the Weyl geometry does lead in a relatively natural way to quantization conditions. Several interesting features of the above result should be noted. First, it is difficult to attach a simple physical meaning to the factor i in (2.8), since the 4-dimensional length of most vectors in classical physics is certainly real. We are not, however, attempting a classical interpretation, and have indeed given no specific meaning to the quantity l in (2.2). Secondly, in the classical limit of $h \to 0$, the change dl becomes infinite. If one may take the above result for ϕ_β seriously, this means that the classical limit does not make direct sense and the Weyl theory should be applied only on an atomic scale.

3. PROPERTIES OF SPINORS IN AFFINE SPACE

In this section we wish to review and discuss briefly the salient features of spinors and spinor analysis in affine space. The 4-component spinors of ordinary Dirac electron theory will be used, since the Dirac description is undeniably successful in the limit of special relativity. Our discussion uses the notation of Ref. 2 and follows Refs. 2 and 4 on basic concepts.

For a particle such as an electron, with an internal structure that is not amenable to a simple space–time interpretation, we introduce an internal or spin transformation S. This is a 4×4 matrix which is to operate on the 4-component spinors of the theory and, in general, may be a function $S(x)$ of position in space–time. In addition, we allow general coordinate transformations as in general relativity. *There is no a priori relation between the internal structure and space–time structure* so that the transformations are *independent of each other.*[5] Later, in Sec. 4, the internal and space–time structure is, however, related via the demand of a universal Clifford algebra.

Under a spin transformation a spinor is assumed to transform via

$$\psi'(x) = S(x)\psi(x). \tag{3.1}$$

A dual space may also be introduced wherein the dual spinors transform via[6]

$$\tilde{\psi}'(x) = \tilde{\psi}(x)S^{-1}(x). \tag{3.2}$$

Clearly, the inner product of a dual spinor and a spinor undergoes no change under a spin transformation, and may be interpreted as a "spin scalar."

In analogy with the vector transplantation law (1.1), we suppose that there exists a spinor transplantation law: Under a coordinate displacement dx^α the components of the spinor are assumed to change by

$$d\psi = -i\Gamma_\alpha \psi\, dx^\alpha \tag{3.3}$$

and the dual spinor components by

$$d\tilde{\psi} = i\tilde{\psi}\Gamma_\alpha\, dx^\alpha, \tag{3.3'}$$

where the Γ_α are a set of 4×4 matrices. The choice of the transplantation law (3.3') for the dual spinor guarantees that $\tilde{\psi}\psi$ be unchanged under spinor transplantation, and allows its continued interpretation as a spin scalar. The coefficients Γ_α of the affine spinor transplantation law are usually referred to as the Fock–Ivanenko coefficients.[7] It should be noted that the Γ_α need not be Hermitian; the use of Hermitian Γ_α and the identification of the dual spinor $\tilde{\psi}$ with the adjoint spinor $\psi^\dagger$ would be consistent, but is not necessary and is not used here.

The actual change in a spinor field $\psi(x)$ over dx^α is merely

$$\delta\psi(x) = \psi(x)_{|\alpha}\, dx^\alpha + O(dx)^2. \tag{3.4}$$

The change that would occur under spinor transplantation alone, the analog of parallel displacement, is given by (3.3). Thus, the difference of (3.4) and (3.3) may be viewed as the excess or "physical" change in ψ:

$$\delta\psi(x) - d\psi(x) = (\psi_{|\alpha} + i\Gamma_\alpha\psi)\, dx^\alpha + O(dx)^2. \tag{3.5}$$

As in ordinary vector analysis, the quantity in brackets is interpreted as a spin covariant derivative, which we denote by three slashes, i.e.,

$$\psi_{\|\alpha} = \psi_{|\alpha} + i\Gamma_\alpha\psi. \tag{3.6}$$

Similarly, the dual spinor has a spin covariant derivative defined as

$$\tilde{\psi}_{\|\alpha} = \tilde{\psi}_{|\alpha} - i\tilde{\psi}\Gamma_\alpha. \tag{3.6'}$$

This generalized derivative plays exactly the same role here as the covariant derivative does in general relativity theory.

The Fock–Ivanenko coefficients introduced in (3.3), in general, change under a spin transformation S; to obtain their transformation properties, we demand that a transplanted spinor remain a spinor. That is, ψ at $x + dx$ should still obey Eq. (3.1), or

$$S(x + dx)\psi(x + dx) = \psi'(x + dx). \tag{3.7}$$

If all the above quantities in (3.7) are expanded to first order in dx, a simple result is obtained for Γ'_α, the spin-transformed Fock–Ivanenko coefficient, namely,

$$\Gamma'_\alpha = S\Gamma_\alpha S^{-1} + iS_{|\alpha}S^{-1}. \tag{3.8}$$

One should note that, for the special case of a constant spin transformation $S_{|\alpha} = 0$, Eq. (3.8) is merely a similarity transformation. In order that (3.3) be invariant under a *pure coordinate transformation*, we assume that Γ_α behaves like a 4-vector. Thus,

$$\Gamma'_\alpha = \frac{\partial x^\beta}{\partial x'^\alpha}(S\Gamma_\beta S^{-1} + iS_{|\beta}S^{-1}). \tag{3.8'}$$

So far, we have dealt with only spinors and the Fock–Ivanenko coefficients under the spin transformations. More general objects also occur, such as various 4×4 matrices, which may or may not have a space–time index.[8] One of the simplest is a 4×4 object of the outer-product form

$$M = \psi\tilde{\phi}. \tag{3.9}$$

Under spinor transplantation,

$$dM = -i[\Gamma_\alpha, M]\,dx^\alpha, \tag{3.10}$$

so that a spin covariant derivative may be defined, analogous to Eq. (3.6), as

$$M_{\|\alpha} = M_{|\alpha} + i[\Gamma_\alpha, M]. \tag{3.11}$$

For objects such as 4×4 matrices with space–time indices, we can apply both coordinate and spin transformations simultaneously. Such an object is taken to satisfy, for the 2-index example, the transformation law

$$A'_{\mu\alpha} = \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\alpha} SA_{\nu\beta}S^{-1}, \tag{3.12}$$

which is merely a *combined* coordinate and similarity transformation. This transformation law guarantees, for example, that bilinear forms such as Dirac's

$$\tilde{\psi}\gamma_\mu\phi \tag{3.13}$$

transform as 4-vectors. We call objects such as $A_{\mu\alpha}$ "spin tensors."

The spin covariant derivative of a covariant spin vector may be inferred from (3.11) to be, plausibly,

$$A_{\mu\|\alpha} \equiv A_{\mu|\alpha} + i[\Gamma_\alpha, A_\mu]. \tag{3.14}$$

The use of the transformation laws (3.8') and (3.12) for Γ_α and A_μ allows one to verify that the spin covariant derivative in (3.14) is indeed a spin tensor, since it transforms as indicated in (3.12). The generalization of the spin covariant derivative to a spin tensor of any rank is straightforward:

$$A_{\mu\cdots\tau\|\alpha} = A_{\mu\cdots\tau|\alpha} + i[\Gamma_\alpha, A_{\mu\cdots\tau}]. \tag{3.15}$$

As in the tensor analysis of Riemann space, there occurs in the present spinor analysis a measure of the "distortion" of space. This analog of the Riemann tensor is called the spin curvature tensor. We will introduce it as a measure of the nonintegrability in the manner of Weyl. Consider a spinor transplanted along the vectors dx and $d\hat{x}$ of Fig. 1. After transplantation along path 1, dx followed by $d\hat{x}$, it has become

$$\psi(f)_1 = \psi - i\Gamma_\alpha\psi\,dx^\alpha - i\Gamma_\alpha\psi\,d\hat{x}^\alpha - i\Gamma_{\alpha|\beta}\psi\,dx^\beta\,d\hat{x}^\alpha - \Gamma_\alpha\Gamma_\beta\psi\,dx^\beta\,d\hat{x}^\alpha \tag{3.16}$$

at f. Alternatively, if it is transplanted along path 2, $d\hat{x}$ followed by dx, it has become at f

$$\psi(f)_2 = \psi - i\Gamma_\alpha\psi\,d\hat{x}^\alpha - i\Gamma_\alpha\psi\,dx^\alpha - i\Gamma_{\alpha|\beta}\psi\,d\hat{x}^\beta\,dx^\alpha - \Gamma_\alpha\Gamma_\beta\psi\,d\hat{x}^\beta\,dx^\alpha. \tag{3.17}$$

The difference of the transplanted spinors is, then,

$$\psi(f)_1 - \psi(f)_2 = d\psi = -i\hat{R}_{\alpha\beta}\psi\,dx^\beta\,d\hat{x}^\alpha,$$
$$\hat{R}_{\alpha\beta} = \Gamma_{\alpha|\beta} - \Gamma_{\beta|\alpha} + i[\Gamma_\beta, \Gamma_\alpha] = -\hat{R}_{\beta\alpha}. \tag{3.18}$$

This spin curvature tensor $\hat{R}_{\alpha\beta}$ is clearly a measure of the distortion or nonintegrability present in spin space; if it is zero, for example, a "constant" spinor field (one with zero-spin covariant derivative) can be established in the space by spinor transplantation. If it is not such, an attempt must meet with failure. $\hat{R}_{\alpha\beta}$ then plays much the same role as the Riemann curvature tensor in ordinary tensor analysis.[4]

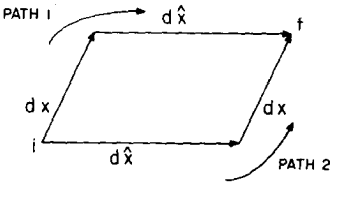


FIG. 1. Spinor transplantation along different paths.

In closing this section, we should note that our analysis has so far not depended on the number of components in the spinor fields considered, even though our motivation lies in the 4-component Dirac field. The detailed nature of the spinors only becomes relevant when some sort of coupling between the internal structure and the space–time structure of the fields is introduced in the next section.

4. SPACE–TIME COUPLING

So far, we have considered the internal or spin structure of the fields to be entirely independent of the space–time structure. In this section, we give a plausible coupling of the two structures. Our first basic postulate is that the Clifford algebra of ordinary Dirac theory in flat space remains true in the more general Weyl space:

$$\{\gamma_\mu, \gamma_\nu\} \equiv \gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2g_{\mu\nu}I. \tag{4.1}$$

The curly brackets here signify the anticommutator generally associated with fermion fields.

We have introduced the Clifford algebra in an ad hoc manner. It is well to remember its origins in Dirac theory.[8] The Dirac equation for a free electron in flat space is

$$(\gamma^\mu p_\mu - m)\psi = 0. \tag{4.2}$$

If both sides are now multiplied by the 4×4 matrix $\gamma^\mu p_\mu + m$ the result is

$$(\tfrac{1}{2}\{\gamma_\mu, \gamma_\nu\}p^\mu p^\nu - m^2)\psi = 0. \tag{4.3}$$

If one now *demands* that the above equation be a Klein–Gordon equation for *each* of the four components of ψ, then

$$(\tfrac{1}{2}\{\gamma_\mu, \gamma_\nu\}p^\mu p^\nu - m^2)\psi = (p^2 - m^2)\psi = 0, \tag{4.4}$$

so that

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}I. \tag{4.4'}$$

Thus, the demand that a free Dirac particle be on the mass shell $p^2 = m^2$ leads to the Clifford algebra. The obvious generalization of (4.1) is to allow any metric tensor on the right side instead of only the special case of the Lorentz metric. This generalization is, clearly, generally covariant and appears to be the simplest possible.

It may indeed be possible to motivate the use of a Clifford algebra more deeply or to consider other possibilities, but, in the spirit of simplicity, we use (4.1) as it stands.

Our next step is to check the consistency of the Clifford algebra when one takes the spin covariant derivative of both sides of (4.1). This is easily done if one first notes that the product rule for ordinary differentiation holds also for the spin covariant derivative

$$(AB)_{\|\alpha} = (A_{\|\alpha})B + A(B_{\|\alpha}). \tag{4.5}$$

In particular,

$$\{\gamma_\mu, \gamma_\nu\}_{\|\alpha} = \{\gamma_{\mu\|\alpha}, \gamma_\nu\} + \{\gamma_\mu, \gamma_{\nu\|\alpha}\}. \tag{4.6}$$

Applied to the Clifford algebra this yields

$$\begin{aligned}\{\gamma_{\mu\|\alpha}, \gamma_\nu\} + \{\gamma_\mu, \gamma_{\nu\|\alpha}\} - 2g_{\mu\nu\|\alpha} &= (\{\gamma_\mu, \gamma_\nu\} - 2g_{\mu\nu})_{\|\alpha} \\ &\quad + i\{[\Gamma_\alpha, \gamma_\mu], \gamma_\nu\} + i\{\gamma_\mu, [\Gamma_\alpha, \gamma_\nu]\}.\end{aligned} \tag{4.7}$$

The first term is identically zero by (4.1), while the other two terms cancel when expanded. We thus have the consistent relation

$$\{\gamma_\mu, \gamma_\nu\}_{\|\alpha} = \{\gamma_{\mu\|\alpha}, \gamma_\nu\} + \{\gamma_\mu, \gamma_{\nu\|\alpha}\} = 2g_{\mu\nu\|\alpha}, \tag{4.8}$$

so we may indeed postulate a Clifford algebra throughout the Weyl space.

In showing that (4.8) holds, we made no use of any properties of the Fock–Ivanenko coefficients Γ_α; clearly then, we cannot calculate them from a relation like (4.8). The usual procedure for the more restrictive case of a Riemann space is to make a slightly stronger demand at this point. For a Riemann space, one has the Ricci theorem[2]

$$g_{\mu\nu\|\alpha} = g_{\mu\nu\|\alpha} = 0 \tag{4.9}$$

so the right-hand side of (4.8) is zero. It is then natural to ask that the spin covariant derivatives of the γ_μ be zero also, which is stronger than, and guarantees, (4.8) for a Riemann space

$$\gamma_{\mu\|\alpha} = 0. \tag{4.10}$$

This must be modified for a Weyl space. A short calculation shows that the Ricci theorem becomes

$$g_{\mu\nu\|\alpha} = g_{\mu\nu\|\alpha} = 2g_{\mu\nu}\phi_\alpha, \tag{4.11}$$

so that the spin covariant derivatives of the Dirac matrices cannot be zero. The simplest generalization of (4.10) that is consistent with (4.8) involves making $\gamma_{\mu\|\alpha}$ linear in ϕ_α, or

$$\gamma_{\mu\|\alpha} = \gamma_\mu\phi_\alpha. \tag{4.12}$$

Using the definition of spin covariant derivative (3.14), we may write this as

$$[\gamma_\mu, \Gamma_\alpha] = -i\left(\gamma_{\mu|\alpha} - \begin{Bmatrix}\beta \\ \mu\alpha\end{Bmatrix}\gamma_\beta\right) - i(g_{\lambda\alpha}\phi_\mu - g_{\mu\alpha}\phi_\lambda)\gamma^\lambda. \tag{4.13}$$

This equation is treated in the Appendix. If the right side is given, subject to consistency requirements, the Γ_α can be determined up to a multiple of the identity matrix.

Before discussing Eq. (4.13) further, we should note the simplest case. In the flat Riemann space associated with special relativity, the right-hand side of (4.13) can be set equal to zero. The Γ_α are then simply multiples of the identity, as noted in Ref. 4,

$$\Gamma_\alpha = \xi_\alpha I, \tag{4.14}$$

where ξ_α is a set of four coefficients which may be identified with the electromagnetic field and minimal coupling. This is discussed further in Sec. 6.

5. THE FOCK–IVANENKO COEFFICIENTS IN FLAT WEYL SPACE

In this section, we wish to consider a space–time manifold free of gravitational effects and in which

$$\gamma_{\mu|\alpha} - \begin{Bmatrix} \beta \\ \mu\alpha \end{Bmatrix} \gamma_\beta = 0. \tag{5.1}$$

We call this a flat Weyl space. The only distortion remaining, then, is that due to a nonzero ϕ_α field. The Fock–Ivanenko coefficients are easily determined in this special case.

The basic equation for the Γ_α is (4.13), which becomes in the present case

$$[\gamma_\mu, \Gamma_\alpha] = -i(g_{\lambda\alpha}\phi_\mu - g_{\mu\alpha}\phi_\lambda)\gamma^\lambda. \tag{5.2}$$

This meets the consistency conditions of the Appendix and corresponds to

$$\begin{aligned} c_{\mu(\alpha)\lambda} &= -i(g_{\lambda\alpha}\phi_\mu - g_{\mu\alpha}\phi_\lambda) = -c_{\lambda(\alpha)\mu}, \\ \tau_{\mu(\alpha)} &= d^\lambda_{(\alpha)} = 0 \end{aligned} \tag{5.3}$$

in Eq. (A2). The solution for Γ_α is thus easily written down from (A10)

$$\begin{aligned} \Gamma_\alpha &= \xi_\alpha I + \tfrac{1}{4}(g_{\mu\alpha}\phi_\lambda - g_{\lambda\alpha}\phi_\mu)\sigma^{\mu\lambda} \\ &= \xi_\alpha I + \tfrac{1}{2}\sigma_{\alpha\lambda}\phi^\lambda, \end{aligned} \tag{5.4}$$

where ξ_α is an arbitrary set of four numbers. This clearly includes (4.14) as a special case.

6. DISPLACEMENT OPERATOR AND MINIMAL COUPLING

In quantum mechanics the displacement operator or momentum operator p_α is normally written in a space–time representation and in Cartesian coordinates as

$$p_\alpha = -i\hbar \frac{\partial}{\partial x^\alpha}. \tag{6.1}$$

In the present context, however, the derivative operator has no special significance, just as in ordinary tensor analysis it has no special significance. As in tensor analysis, we will replace it with an object that has well-defined transformation properties: in tensor analysis, the appropriate replacement is the covariant derivative and, in the present case, the replacement is the spin covariant derivative. Thus, instead of working with the operator relation (6.1) applied to spinors, we will use the *covariant* displacement operator

$$\begin{aligned} D_\alpha\psi &= -i\hbar\left(\frac{\partial}{\partial x^\alpha} + i\Gamma_\alpha\right)\psi \\ &= (p_\alpha + \hbar\Gamma_\alpha)\psi = -i\hbar\psi_{\|\alpha}. \end{aligned} \tag{6.2}$$

For the special case of a flat Riemann space the Fock–Ivanenko coefficients are multiples of the identity, as we have demonstrated in Sec. 4, Eq. (4.14). If ξ_α in Eq. (4.14) is chosen to be proportional to the electromagnetic field,

$$\Gamma_\alpha = -(e/\hbar c)A_\alpha I, \tag{6.3}$$

we recover the well-known principle of minimal coupling as pointed out by Pagels[4]:

$$p_\alpha \to p_\alpha - (e/c)A_\alpha. \tag{6.4}$$

For the case of a flat Weyl space we may instead use the result of Sec. 5, namely,

$$p_\alpha \to D_\alpha = p_\alpha + \hbar\xi_\alpha I + \tfrac{1}{2}\hbar\sigma_{\alpha\lambda}\phi^\lambda. \tag{6.5}$$

An important and basic change is that the last term in this expression is not Hermitian, so the eigenvalues are not, in general, real. Thus, the interpretation of the eigenvalues of D_α as physically measurable momenta is not valid. Moreover, one cannot consider the D_α as the generators of unitary transformations. In a Weyl space the simple connection between momentum and displacement operators familiar from elementary quantum mechanics must be abandoned.

The above unfamiliar properties of D_α should not be considered as overly unpleasant. If ϕ_α can indeed be considered as proportional to all or part of an electromagnetic field, then that field is intrinsically coupled to the spinor field, and a displacement operator that displaces the spinor field alone, and not the intimately associated electromagnetic field, should not be expected to be Hermitian.

Let us form the simplest generalization of the Dirac equation for a free electron, using D_α instead of p_α:

$$\gamma^\mu D_\mu\psi = m\psi = -i\hbar\gamma^\mu\psi_{\|\mu}. \tag{6.6}$$

A short calculation using the Clifford algebra reveals that

$$\gamma^\mu\sigma_{\mu\lambda} = i(g^\alpha_\alpha - 1)\gamma_\lambda = 3i\gamma_\lambda. \tag{6.7}$$

In a flat Weyl space we therefore have

$$\tfrac{1}{2}\gamma^{\mu}\sigma_{\mu\lambda}\phi^{\lambda} = \tfrac{3}{2}i\gamma_{\lambda}\phi^{\lambda}, \tag{6.8}$$

so that (6.6) becomes

$$\gamma^{\alpha}(p_{\alpha} + \hbar\xi_{\alpha} + \tfrac{3}{2}i\hbar\phi_{\alpha})\psi = m\psi. \tag{6.9}$$

This is indeed identical to the Dirac equation with minimal coupling to the fields ξ_{α} and ϕ_{α}. Indeed, if the proportionality indicated in Sec. 2 is assumed to hold, we obtain

$$\gamma^{\alpha}(p_{\alpha} + \hbar\xi_{\alpha} \pm \tfrac{3}{2}(e/c)A_{\alpha})\psi = m\psi. \tag{6.10}$$

Except for the factor of $\frac{3}{2}$, we therefore obtain a consistent coupling to the electromagnetic field, as identified in Sec. 2 with the Weyl field ϕ_{α}. Moreover, if ξ_{α} is arbitrarily taken to be also proportional to ϕ_{α}, which is, after all, the only 4-indexed quantity available, i.e.,

$$\xi_{\alpha} = -\tfrac{1}{2}i\phi_{\alpha}, \tag{6.11}$$

then we have precisely the Dirac equation with minimal coupling included

$$\gamma^{\alpha}(p_{\alpha} - (e/c)A_{\alpha}) = m\psi. \tag{6.12}$$

In conclusion, then, the replacement of p_{α} by D_{α} leads to a coupling with the Weyl field that is at least qualitatively the same as the minimal coupling of elementary quantum mechanics.

7. GRAVITATIONAL EFFECTS

Instead of assuming that

$$\gamma_{\mu|\alpha} - \begin{Bmatrix}\beta\\ \mu\alpha\end{Bmatrix}\gamma_{\beta} = 0, \tag{7.1}$$

as in Sec. 5, we may include gravitational effects by assuming an expansion of the rather general form

$$-i\left(\gamma_{\mu|\alpha} - \begin{Bmatrix}\beta\\ \mu\alpha\end{Bmatrix}\gamma_{\beta}\right) = \tilde{c}_{\mu(\alpha)\lambda}\gamma^{\lambda} + \tilde{d}^{\lambda}_{(\alpha)}\sigma_{\mu\lambda}, \tag{7.2}$$

where $\tilde{c}_{\mu(\alpha)\lambda}$ and $\tilde{d}^{\lambda}_{(\alpha)}$ are assumed to be small compared to the corresponding quantities associated with the electromagnetic field. The Fock–Ivanenko coefficients are then

$$\Gamma_{\alpha} = \xi_{\alpha}I + \tfrac{1}{2}\sigma_{\alpha\lambda}\phi^{\lambda} - \tfrac{1}{4}i\sigma_{\mu\lambda}\tilde{c}^{\mu}{}_{(\alpha)}{}^{\lambda} + \tfrac{1}{2}i\tilde{d}_{\lambda(\alpha)}\gamma^{\lambda}. \tag{7.3}$$

The $\tilde{c}^{\mu\lambda}_{(\alpha)}$ will only modify the electromagnetic coupling slightly, and can probably be safely ignored. The role of $\tilde{d}_{\lambda(\alpha)}$ is somewhat more interesting. Proceeding as before, we obtain an extra term in the Dirac equation, namely,

$$\tfrac{1}{2}i\tilde{d}_{\lambda(\alpha)}\gamma^{\alpha}\gamma^{\lambda}. \tag{7.4}$$

If $\tilde{d}_{\lambda(\alpha)}$ is taken to be proportional to the most obvious 2-indexed quantity, $g_{\lambda\alpha}$, we obtain an extra term of the form

$$g_{\lambda\alpha}\gamma^{\alpha}\gamma^{\lambda} = 4I. \tag{7.5}$$

That is, an extra scalar term appears in the Dirac equation that may be interpreted as a (presumably small) mass shift associated with an intrinsic coupling to the gravitational field. This is qualitatively consistent with the results of Pagels.[4]

8. CONCLUSION

We have attempted to show, using very plausible assumptions about the Clifford algebra of Dirac matrices, that the Fock–Ivanenko coefficients can be calculated for a Weyl space free of gravitational effects. The resulting generalized momentum operator is not Hermitian, which leads to interesting questions about its interpretation. If the generalized momentum operator is used, instead of $i\hbar$ times the simple derivative operator, the Dirac spinor field automatically couples to the Weyl field minimally, i.e., we obtain the usual Dirac equation for an electron in an electromagnetic field. The strength of the coupling is arbitrary, owing to the arbitrariness in the Fock–Ivanenko coefficients, so the result is qualitative.

APPENDIX

We wish to find solutions to matrix equations of the form

$$[\gamma_{\mu}, B_{(\alpha)}] = A_{\mu(\alpha)}, \quad \mu = 0, 1, 2, 3, \tag{A1}$$

where $A_{\mu(\alpha)}$ is given, γ_{μ} is a set of matrices obeying the Clifford algebra, and $B_{(\alpha)}$ is a set of 4×4 matrices to be determined. The parameter α may range over any number of indices, although we are concerned in this paper with the standard range 0 to 3.

Let us first suppose that $A_{\mu(\alpha)}$ can be expanded in terms of the usual Dirac scalar, vector, and tensor matrices,[8] I, γ^{μ}, and $\sigma^{\alpha\beta} = \frac{1}{2}i[\gamma^{\alpha}, \gamma^{\beta}]$, as

$$A_{\mu(\alpha)} = \tau_{\mu(\alpha)}I + c_{\mu(\alpha)\lambda}\gamma^{\lambda} + d^{\lambda}_{(\alpha)}\sigma_{\mu\lambda}. \tag{A2}$$

Then we expect a solution for $B_{(\alpha)}$ that can be similarly expanded in terms of coefficients to be determined:

$$B_{(\alpha)} = \xi_{(\alpha)}I + e_{(\alpha)\lambda}\gamma^{\lambda} + f^{\lambda\beta}_{(\alpha)}\sigma_{\lambda\beta}. \tag{A3}$$

If this trial solution is inserted in (A1), the result is

$$[\gamma_{\mu}, B_{(\alpha)}] = -2ie_{(\alpha)\lambda}\sigma^{\lambda}_{\mu} + \tfrac{1}{2}if^{\lambda\beta}_{(\alpha)}[\gamma_{\mu}, [\gamma_{\lambda}, \gamma_{\beta}]]. \tag{A4}$$

We may use the Clifford algebra and the identity

$$[x, yz] = \{x, y\}z - y\{x, z\} \tag{A5}$$

to simplify the second term in (A4) to obtain

$$[\gamma_\mu, [\gamma_\lambda, \gamma_\beta]] = 4(g_{\mu\lambda}\gamma_\beta - g_{\mu\beta}\gamma_\lambda). \quad \text{(A6)}$$

We thus have

$$[\gamma_\mu, B_{(\alpha)}] = -2ie^\lambda_{(\alpha)}\sigma_{\mu\lambda} + 2i(f_{(\alpha)\mu\beta} - f_{(\alpha)\beta\mu})\gamma\beta. \quad \text{(A7)}$$

If we compare this with the expansion (A2), we obtain the unknown coefficients appearing in $B_{(\alpha)}$, in terms of the known coefficients appearing in (A2). The result is

$$e^\lambda_{(\alpha)} = \tfrac{1}{2}id^\lambda_{(\alpha)}, \quad f_{(\alpha)\mu\lambda} = -\tfrac{1}{4}ic_{\mu(\alpha)\lambda}, \quad \xi_{(\alpha)} \text{ arbitrary.} \quad \text{(A8)}$$

In addition, we obtain consistency requirements on $A_{\mu(\alpha)}$, in order that a solution exists:

$$\tau_{\mu(\alpha)} = 0, \quad c_{\mu(\alpha)\lambda} = -c_{\lambda(\alpha)\mu}. \quad \text{(A9)}$$

In summary, if $A_{\mu(\alpha)}$ is given as in (A2) with the consistency conditions (A9) met, then $B_{(\alpha)}$ is

$$B_{(\alpha)} = \xi_{(\alpha)}I + \tfrac{1}{2}id_{\lambda(\alpha)}\gamma^\lambda - \tfrac{1}{4}ic_{\mu(\alpha)\lambda}\sigma^{\mu\lambda}. \quad \text{(A10)}$$

Lastly, we wish to note that the so-called pseudo-quantities containing $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ are disjoint from the above discussion in that a γ^5 term in $B_{(\alpha)}$ leads to a γ^5 term in $A_{\mu(\alpha)}$. Thus, if $A_{\mu(\alpha)}$ contains no γ^5 terms, we do not need any in $B_{(\alpha)}$. The same approach as used above may be applied using γ^5, $\gamma^5\gamma^\mu$, and $\gamma^5\sigma^{\alpha\beta}$, but this is unnecessary here and will not be carried out.

[1] H. Weyl, *Space, Time, Matter* (Dover Publications, New York, 1922).
[2] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill Book Co., Inc., New York, 1965), Chaps. 2 and 13. The notation of this reference is used throughout.
[3] F. London, Z. Physik **42**, 375 (1927).
[4] H. Pagels, Ann. Phys. (N.Y.) **31**, 64 (1965).
[5] This is a different viewpoint than is taken in most discussions of Dirac spinors in a Lorentz 4-space. Thus the $S(x)$ introduced here bears no simple relation to the spinor transformation used in Ref. 8.
[6] We differ from Pagels, Ref. 4, in that S need not be unitary.
[7] V. Fock and D. Ivanenko, Compt. Rend. **188**, 1470 (1929).
[8] For a discussion of the Dirac theory in ordinary Lorentz 4-space, see J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill Book Co., Inc., New York, 1964). It should be noted, however, that we deal with a more general space and treat the spin transformations entirely differently.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Matrix Elements of Relativistic Electrons in a Coulomb Field*

W. W. GARGARO AND D. S. ONLEY
Department of Physics, Ohio University, Athens, Ohio 45701

(Received 10 February 1969)

Matrix elements for a radiative interaction between states of a Dirac electron in the presence of a Coulomb field are reduced to a closed analytic form in the limit of zero electron mass; corrections for finite electron mass are indicated. The application of these to inelastic electron scattering and radiation problems is discussed.

I. INTRODUCTION

We wish to consider matrix elements of the electromagnetic interaction between states of a relativistic electron in the Coulomb field of a point charge. A closed form for these has been obtained previously[1] for an instantaneous Coulomb interaction. This form could be used also for a radiative interaction in the limit of the long-wavelength approximation for the photon with negligible energy loss for the electron. The integrals were employed in the analysis of electron scattering[2-4] for the contribution of high-angular-momentum components in the electron wavefunctions, which, since they have no appreciable amplitude near the origin, can be considered to be moving in the field of a point charge, even though the scattering center has finite extent. Later, because of the restriction on the energy loss of the electrons, the use of these closed forms was abandoned for inelastic electron scattering,[5] and the corresponding matrix elements were evaluated by numerical integration. The use of numerical integration has its limitations, however, because the integral over the electron radial coordinate has to be carried to infinity and approaches a limit only slowly and in an oscillatory fashion. In this paper we develop expressions for matrix elements between relativistic Coulomb wavefunctions of differing energies, which admit also the finite wavelength of the photon (or, equivalently, which admit retardation in the interaction with the nucleus). It is necessary to assume that the mass of the electron is negligible in comparison with its energy, in order to obtain a closed form; but this approximation is usual in high-energy electron scattering.

The Coulomb wavefunctions for the Dirac electron

to simplify the second term in (A4) to obtain

$$[\gamma_\mu, [\gamma_\lambda, \gamma_\beta]] = 4(g_{\mu\lambda}\gamma_\beta - g_{\mu\beta}\gamma_\lambda). \quad \text{(A6)}$$

We thus have

$$[\gamma_\mu, B_{(\alpha)}] = -2ie^\lambda_{(\alpha)}\sigma_{\mu\lambda} + 2i(f_{(\alpha)\mu\beta} - f_{(\alpha)\beta\mu})\gamma\beta. \quad \text{(A7)}$$

If we compare this with the expansion (A2), we obtain the unknown coefficients appearing in $B_{(\alpha)}$, in terms of the known coefficients appearing in (A2). The result is

$$e^\lambda_{(\alpha)} = \tfrac{1}{2}id^\lambda_{(\alpha)}, \quad f_{(\alpha)\mu\lambda} = -\tfrac{1}{4}ic_{\mu(\alpha)\lambda}, \quad \xi_{(\alpha)} \text{ arbitrary.} \quad \text{(A8)}$$

In addition, we obtain consistency requirements on $A_{\mu(\alpha)}$, in order that a solution exists:

$$\tau_{\mu(\alpha)} = 0, \quad c_{\mu(\alpha)\lambda} = -c_{\lambda(\alpha)\mu}. \quad \text{(A9)}$$

In summary, if $A_{\mu(\alpha)}$ is given as in (A2) with the consistency conditions (A9) met, then $B_{(\alpha)}$ is

$$B_{(\alpha)} = \xi_{(\alpha)}I + \tfrac{1}{2}id_{\lambda(\alpha)}\gamma^\lambda - \tfrac{1}{4}ic_{\mu(\alpha)\lambda}\sigma^{\mu\lambda}. \quad \text{(A10)}$$

Lastly, we wish to note that the so-called pseudo-quantities containing $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ are disjoint from the above discussion in that a γ^5 term in $B_{(\alpha)}$ leads to a γ^5 term in $A_{\mu(\alpha)}$. Thus, if $A_{\mu(\alpha)}$ contains no γ^5 terms, we do not need any in $B_{(\alpha)}$. The same approach as used above may be applied using γ^5, $\gamma^5\gamma^\mu$, and $\gamma^5\sigma^{\alpha\beta}$, but this is unnecessary here and will not be carried out.

[1] H. Weyl, *Space, Time, Matter* (Dover Publications, New York, 1922).
[2] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill Book Co., Inc., New York, 1965), Chaps. 2 and 13. The notation of this reference is used throughout.
[3] F. London, Z. Physik **42**, 375 (1927).
[4] H. Pagels, Ann. Phys. (N.Y.) **31**, 64 (1965).
[5] This is a different viewpoint than is taken in most discussions of Dirac spinors in a Lorentz 4-space. Thus the $S(x)$ introduced here bears no simple relation to the spinor transformation used in Ref. 8.
[6] We differ from Pagels, Ref. 4, in that S need not be unitary.
[7] V. Fock and D. Ivanenko, Compt. Rend. **188**, 1470 (1929).
[8] For a discussion of the Dirac theory in ordinary Lorentz 4-space, see J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill Book Co., Inc., New York, 1964). It should be noted, however, that we deal with a more general space and treat the spin transformations entirely differently.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Matrix Elements of Relativistic Electrons in a Coulomb Field*

W. W. GARGARO AND D. S. ONLEY
Department of Physics, Ohio University, Athens, Ohio 45701

(Received 10 February 1969)

Matrix elements for a radiative interaction between states of a Dirac electron in the presence of a Coulomb field are reduced to a closed analytic form in the limit of zero electron mass; corrections for finite electron mass are indicated. The application of these to inelastic electron scattering and radiation problems is discussed.

I. INTRODUCTION

We wish to consider matrix elements of the electromagnetic interaction between states of a relativistic electron in the Coulomb field of a point charge. A closed form for these has been obtained previously[1] for an instantaneous Coulomb interaction. This form could be used also for a radiative interaction in the limit of the long-wavelength approximation for the photon with negligible energy loss for the electron. The integrals were employed in the analysis of electron scattering[2-4] for the contribution of high-angular-momentum components in the electron wavefunctions, which, since they have no appreciable amplitude near the origin, can be considered to be moving in the field of a point charge, even though the scattering center has finite extent. Later, because of the restriction on the energy loss of the electrons, the use of these closed forms was abandoned for inelastic electron scattering,[5] and the corresponding matrix elements were evaluated by numerical integration. The use of numerical integration has its limitations, however, because the integral over the electron radial coordinate has to be carried to infinity and approaches a limit only slowly and in an oscillatory fashion. In this paper we develop expressions for matrix elements between relativistic Coulomb wavefunctions of differing energies, which admit also the finite wavelength of the photon (or, equivalently, which admit retardation in the interaction with the nucleus). It is necessary to assume that the mass of the electron is negligible in comparison with its energy, in order to obtain a closed form; but this approximation is usual in high-energy electron scattering.

The Coulomb wavefunctions for the Dirac electron

involve confluent hypergeometric functions, and the matrix elements consequently involve integrals over the product of two of these functions. We have been unable to find, in standard works, a complete account of the required integral, and we have therefore devoted Sec. II to its evaluation for various cases. We are also able to reproduce the result for the long-wavelength limit, mentioned above, which is a valuable check on our analysis. In Sec. III, we show how the results may be applied to inelastic electron scattering.

II. EVALUATION OF THE INTEGRAL

The integral which we wish to evaluate is the following:

$$I = \int_0^\infty dr r^{\alpha-1} e^{-k'r} {}_1F_1[a; b; k'r]\, {}_1F_1[\bar{a}; \bar{b}; kr]. \quad (1)$$

A value for this integral has been given by Slater [6] and Magnus and Oberhettinger[7] but appears to be incorrect unless one of the confluent hypergeometric series terminates. Also Reynolds, Onley, and Biedenharn[1] have considered the special case $k = k'$ (both purely imaginary). In the following we have listed all the cases in which we have been able to evaluate the integral and have shown the relationship to the previous results cited. Integral representations for the generalized hypergeometric functions and their analytic continuations have been used (except in Case IV); specifically,[8]

$${}_1F_1[a; b; z] = \Gamma\begin{bmatrix} b \\ a, b-a \end{bmatrix} \int_0^1 e^{zt} t^{a-1}(1-t)^{b-a-1}\, dt,$$
$$\mathrm{Re}\,(b) > \mathrm{Re}\,(a) > 0, \quad (2)$$

$${}_2F_1[a, b; c; z]$$
$$= \Gamma\begin{bmatrix} c \\ b, c-b \end{bmatrix} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\, dt,$$
$$\mathrm{Re}\,(c) > \mathrm{Re}\,(b) > 0, \quad (3)$$

$${}_3F_2[a_1, a_2, a_3; b_1, b_2; z]$$
$$= \Gamma\begin{bmatrix} b_1 \\ a_1, b_1 - a_1 \end{bmatrix} \int_0^1 t^{a_1-1}(1-t)^{b_1-a_1-1} \times {}_2F_1[a_2, a_3; b_2; zt]\, dt, \quad |z| < 1,$$
$$\mathrm{Re}\,(b_1) > \mathrm{Re}\,(a_1) > 0, \quad (4)$$

where we have adopted the notation

$$\Gamma\begin{bmatrix} a, b, \cdots \\ q, r, \cdots \end{bmatrix} = \frac{\Gamma(a)\Gamma(b)\cdots}{\Gamma(q)\Gamma(r)\cdots}.$$

Case (1): When $|k| > |k'|$, $\mathrm{Re}\,(k) < 0 < \mathrm{Re}\,(k')$, $\mathrm{Re}\,(b) > \mathrm{Re}\,(a) > 0$, $\mathrm{Re}\,(\alpha) > 0$, and $\mathrm{Re}\,(\bar{b}) > \mathrm{Re}\,(\bar{a}) > 0$, then

$$I = k^{-\alpha}\left\{\Gamma\begin{bmatrix} \alpha, \bar{b}, \bar{a}-\alpha \\ \bar{a}, b-\alpha \end{bmatrix} e^{i\pi\alpha} \times {}_3F_2\left[\alpha, b-a, 1+\alpha-\bar{b}; b, 1+\alpha-\bar{a}; \frac{k'}{k}\right] + \left(\frac{k'}{k}\right)^{\bar{a}-\alpha} \Gamma\begin{bmatrix} b, \bar{b}, \alpha-\bar{a}, b-a-\alpha+\bar{a} \\ b-a, b-\bar{a}, b-\alpha+\bar{a} \end{bmatrix} e^{i\pi\bar{a}} \times {}_3F_2\left[\bar{a}, 1+\bar{a}-b, b-a-\alpha+\bar{a}; 1+\bar{a}-\alpha, b+\bar{a}-\alpha; \frac{k'}{k}\right]\right\}. \quad (5)$$

By using Kummer's first theorem

$${}_1F_1[a; b; z] = e^z\, {}_1F_1[b-a; b; -z]$$

and the integral representation (2), we can put I in the form

$$I = \Gamma\begin{bmatrix} b, \bar{b} \\ a, \bar{a}, b-a, \bar{b}-\bar{a} \end{bmatrix} \int_0^1 du u^{b-a-1}(1-u)^{a-1} \times \int_0^1 dv v^{\bar{a}-1}(1-v)^{\bar{b}-\bar{a}-1} \int_0^\infty dr r^{\alpha-1} e^{-[k'u-kv]r}. \quad (6)$$

The integral over r can be carried out to yield

$$\int_0^\infty dr r^{\alpha-1} e^{-[k'u-kv]r} = \Gamma(\alpha)[k'u-kv]^{-\alpha}. \quad (7)$$

With $|k| > |k'|$ we write the result (7) as

$$\Gamma(\alpha)(k'u)^{-\alpha}[1-(k/k'u)v]^{-\alpha}. \quad (8)$$

Substituting this into expression (6), we have

$$I = \Gamma\begin{bmatrix} b, \bar{b}, \alpha \\ a, \bar{a}, b-a, \bar{b}-\bar{a} \end{bmatrix} (k')^{-\alpha} \times \int_0^1 du u^{b-a-\alpha-1}(1-u)^{a-1} \times \int_0^1 dv v^{\bar{a}-1}(1-v)^{b-\bar{a}-1}[1-(k/k'u)v]^{-\alpha}.$$

The integration over v is done, using formula (3), and yields

$$I = \Gamma\begin{bmatrix} b, \alpha \\ a, b-a \end{bmatrix} (k')^{-\alpha} \int_0^1 du u^{b-a-\alpha-1}(1-u)^{a-1} \times {}_2F_1[\alpha, \bar{a}; \bar{b}; k/k'u]. \quad (9)$$

This last expression is of the form of Eq. (4) except that the argument $|k/k'u| > 1$. However, by using the analytic continuation formula[9]

$${}_2F_1[a, b; c; z]$$
$$= \Gamma\begin{bmatrix} c, b-a \\ b, c-a \end{bmatrix} (-z)^{-a} \times {}_2F_1[a, 1+a-c; 1+a-b; z^{-1}] + \Gamma\begin{bmatrix} c, a-b \\ a, c-b \end{bmatrix} (-z)^{-b} \times {}_2F_1[b, 1+b-c; 1+b-a; z^{-1}],$$
$$\text{for } |\arg(-z)| < \pi, \quad (10)$$

we can express Eq. (9) in terms of two integrals which satisfy the conditions of Eq. (4):

$$I = \Gamma\begin{bmatrix} b, \alpha \\ a, b-a \end{bmatrix}(k')^{-\alpha} \times \left\{ \int_0^1 du u^{b-a-\alpha-1}(1-u)^{a-1}\Gamma\begin{bmatrix} \bar{b}, \bar{a}-\alpha \\ \bar{a}, \bar{b}-\alpha \end{bmatrix}\left(-\frac{k}{k'u}\right)^{-\alpha} \times {}_2F_1\left[\alpha, 1+\alpha-b; 1+\alpha-\bar{a}; \frac{k'}{k}u\right] + \int_0^1 du u^{b-a-\alpha-1}(1-u)^{a-1}\Gamma\begin{bmatrix} \bar{b}, \alpha-\bar{a} \\ \alpha, \bar{b}-\bar{a} \end{bmatrix}\left(-\frac{k}{k'u}\right)^{-\bar{a}} \times {}_2F_1\left[\bar{a}, 1+\bar{a}-\bar{b}; 1+\bar{a}-\alpha; \frac{k'}{k}u\right]\right\}. \quad (11)$$

Now, by applying formula (4), we obtain the result claimed in Eq. (5).

Case (2): When $|k'| > |k|$, $\mathrm{Re}\,(k) < 0 < \mathrm{Re}\,(k')$, $\mathrm{Re}\,(b) > \mathrm{Re}\,(a) > 0$, $\mathrm{Re}\,(\alpha) > 0$, and $\mathrm{Re}\,(\bar{b}) > \mathrm{Re}\,(\bar{a}) > 0$, then

$$I = (k')^{-\alpha}\Gamma\begin{bmatrix} \alpha, b, b-a-\alpha \\ b-a, b-\alpha \end{bmatrix} \times {}_3F_2\left[\alpha, \bar{a}, 1+\alpha-b; \bar{b}, 1+\alpha+a-b; \frac{k}{k'}\right] + k^{-\alpha}\left(\frac{k'}{k}\right)^{\alpha-b} \times e^{i\pi(\alpha+a-b)}\Gamma\begin{bmatrix} b, \bar{b}, \alpha-b+a, \bar{a}-\alpha+b-a \\ \bar{a}, a, \bar{b}-\alpha+b-a \end{bmatrix} \times {}_3F_2\left[1-a, b-a, \bar{a}-\alpha+b-a; 1+b-a-\alpha, \bar{b}+b-a-\alpha; \frac{k}{k'}\right]. \quad (12)$$

The proof of Case (1) can be followed through Eq. (7). Now, since $|k'| > |k|$, the result of Eq. (7) is expressed in the form

$$e^{i\pi\alpha}\Gamma(\alpha)(kv)^{-\alpha}[1-(k'/kv)u]^{-\alpha}$$

before integrating over u. The result is

$$I = \Gamma\begin{bmatrix} \bar{b}, \alpha \\ \bar{a}, \bar{b}-\bar{a} \end{bmatrix}e^{i\pi\alpha}(k)^{-\alpha}\int_0^1 dv v^{\bar{a}-\alpha-1}(1-v)^{\bar{b}-\bar{a}-1} \times {}_2F_1\left[\alpha, b-a; b; \frac{k'}{k}\frac{1}{v}\right]. \quad (13)$$

Using Eqs. (10) and (4), we obtain, in the same manner as before, the result (12).

Case (3): When $k = k'$, $\mathrm{Re}\,(k) = 0$, $\mathrm{Re}\,(b) > \mathrm{Re}\,(a) > 0$, for all values of $\alpha - b + 1$ except zero or a negative integer, and $\mathrm{Re}\,(\bar{b} - \bar{a} + a - \alpha) > 0$,

$$I = (k')^{-\alpha}\left\{e^{i\pi(a+\alpha-b)}\Gamma\begin{bmatrix} \alpha, b, 1+\alpha-b \\ a, 1+\alpha-a \end{bmatrix} \times {}_3F_2[\alpha, \bar{b}-\bar{a}, 1+\alpha-b; \bar{b}, 1+\alpha-a; 1] + e^{i\pi\alpha}\Gamma\begin{bmatrix} \alpha, b, 1+\alpha-b \\ b-a, 1+\alpha+a-b \end{bmatrix} \times {}_3F_2[\alpha, \bar{a}, 1+\alpha-b; \bar{b}, 1+\alpha+a-b; 1]\right\}. \quad (14)$$

In this case, owing to innumerable relations between ${}_3F_2$ functions of unit argument, the result can be expressed in many different forms. In the form (14), the result has been established independently by the methods of Ref. 1. We show here that our results for both Cases (1) and (2) reduce to Eq. (14) in the appropriate limit.

Consider the relation[10]

$$ {}_2F_1[a_2, a_3; 1+a_2+a_3-b_2; 1-z] = \Gamma\begin{bmatrix} 1+a_2+a_3-b_2, 1-b_2 \\ 1+a_2-b_2, 1+a_3-b_2 \end{bmatrix}{}_2F_1[a_2, a_3; b_2; z] + z^{1-b_2}\Gamma\begin{bmatrix} 1+a_2+a_3-b_2, b_2-1 \\ a_2, a_3 \end{bmatrix} \times {}_2F_1[1+a_2-b_2, 1+a_3-b_2; 2-b_2; z].$$

By inserting this into Eq. (4) we obtain a relationship between ${}_3F_2$ functions of unit argument as follows:

$$\Gamma\begin{bmatrix} a_1, 1+a_2+a_3-b_2, 1-b_2 \\ b_1, 1+a_2-b_2, 1+a_3-b_2 \end{bmatrix} \times {}_3F_2[a_1, a_2, a_3; b_1, b_2; 1] = \Gamma\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}{}_3F_2[b_1-a_1, a_2, a_3; b_1, 1+a_2+a_3-b_2; 1] - \Gamma\begin{bmatrix} 1+a_1-b_2, 1+a_2+a_3-b_2, b_2-1 \\ 1+b_1-b_2, a_2, a_3 \end{bmatrix} \times {}_3F_2[1+a_1-b_2, 1+a_2-b_2, 1+a_3-b_2; 1+b_1-b_2, 2-b_2; 1]. \quad (15)$$

Equation (4) can be extended to include $|z| = 1$ when the hypergeometric function can be expressed as a convergent series. Now setting $k = k'$ in Case (1) [Eq. (5)] yields

$$I = k^{-\alpha}\Gamma\begin{bmatrix} \alpha, \bar{b}, \bar{a}-\alpha \\ \bar{a}, \bar{b}-\alpha \end{bmatrix}e^{i\pi\alpha} \times {}_3F_2[\alpha, b-a, 1+\alpha-\bar{b}; b, 1+\alpha-\bar{a}; 1] + k^{-\alpha}e^{i\pi\bar{a}}\Gamma\begin{bmatrix} b, \bar{b}, \alpha-\bar{a}, b-a-\alpha+\bar{a} \\ b-a, \bar{b}-\bar{a}, b-\alpha+\bar{a} \end{bmatrix} \times {}_3F_2[\bar{a}, 1+\bar{a}-\bar{b}, b-a+\bar{a}-\alpha; 1+\bar{a}-\alpha, b+\bar{a}-\alpha; 1]. \quad (16)$$

Choosing $a_1 = b - a$, $a_2 = \alpha$, $a_3 = 1 + \alpha - \bar{b}$, $b_1 = b$, and $b_2 = 1 + \alpha - \bar{a}$, for identity (15), gives

$$\begin{aligned}\Gamma\begin{bmatrix} b, \alpha - \bar{a} \\ b - a, b + \bar{a} - \alpha \end{bmatrix} & \\ \times {}_3F_2[\bar{a}, 1 + \bar{a} - \bar{b}, b - a + \bar{a} - \alpha; & \\ 1 + \bar{a} - \alpha, b + \bar{a} - \alpha; 1] & \\ = \Gamma\begin{bmatrix} \alpha, 1 + \alpha - \bar{b} \\ b + \bar{a} - a - \alpha, \alpha + \bar{a} + 1 - \bar{b} \end{bmatrix} & \\ \times {}_3F_2[a, \alpha, 1 + \alpha - \bar{b}; b, 1 + \alpha + \bar{a} - \bar{b}; 1] & \\ - \Gamma\begin{bmatrix} \alpha, 1 + \alpha - \bar{b}, \bar{a} - \alpha \\ b + \bar{a} - a - \alpha, \bar{a}, 1 + \bar{a} - \bar{b} \end{bmatrix} & \\ \times {}_3F_2[\alpha, b - a, 1 + \alpha - \bar{b}; b, 1 + \alpha - \bar{a}; 1]. & \end{aligned} \tag{17}$$

The combination of Eqs. (16) and (17) yields the result (14) with the interchange of the sets of parameters (a, b) and $(\bar{a}, \bar{b})$. For $k = k'$, the original integral (1) is obviously symmetric in these sets of parameters. The same method reduces Case (2) [Eq. (12)] to Eq. (14) exactly. In this case, setting $k = k'$ gives us

$$\begin{aligned} I = k^{-\alpha}\Big\{ & \Gamma\begin{bmatrix} \alpha, b, b - a - \alpha \\ b - a, b - \alpha \end{bmatrix} \\ & \times {}_3F_2[\alpha, \bar{a}, 1 + \alpha - b; \bar{b}, 1 + \alpha + a - b; 1] \\ & + e^{i\pi(\alpha + a - b)}\Gamma\begin{bmatrix} \bar{b}, b, \alpha - b + a, \bar{a} - \alpha + b - a \\ \bar{a}, a, \bar{b} - \alpha + b - a \end{bmatrix} \\ & \times {}_3F_2[1 - a, b - a, \bar{a} - \alpha + b - a; \\ & \qquad 1 + b - a - \alpha, \bar{b} + b - a - \alpha; 1]\Big\}, \end{aligned} \tag{18}$$

and identity (15) must be used with the values $a_1 = \bar{a}$, $a_2 = \alpha$, $a_3 = 1 + \alpha - b$, $b_1 = \bar{b}$, and $b_2 = 1 + \alpha + a - b$. With these values, we have

$$\begin{aligned}\Gamma\begin{bmatrix} \bar{b}, b, \alpha - b + a, \bar{a} - \alpha + b - a \\ \bar{a}, a, \bar{b} - \alpha + b - a \end{bmatrix} & \\ \times {}_3F_2[1 - a, b - a, \bar{a} - \alpha + b - a; & \\ 1 + b - a - \alpha, \bar{b} + b - a - \alpha; 1] & \\ = \Gamma\begin{bmatrix} b, \alpha, 1 + \alpha - b \\ a, 1 + \alpha - a \end{bmatrix} & \\ \times {}_3F_2[\alpha, \bar{b} - \bar{a}, 1 + \alpha - b; \bar{b}, 1 + \alpha - a; 1] & \\ - \Gamma\begin{bmatrix} b, b - a - \alpha, \alpha, 1 + \alpha - b \\ a, b - a, 1 - a \end{bmatrix} & \\ \times {}_3F_2[\alpha, \bar{a}, 1 + \alpha - b; \bar{b}, 1 + \alpha + a - b; 1]. & \end{aligned} \tag{19}$$

Equations (18) and (19) combine to yield the result (14).

Case (4): When Re $(\alpha) > 0$, Re $(k') > 0$, Re $(\bar{b} - \bar{a} + a - \alpha) > 0$, and either $|k'| > |k|$ with the parameter a equal to a negative integer, or $|k'| \geq |k|$ with $\bar{a}$ equal to a negative integer, then

$$\begin{aligned} I = (k')^{-\alpha}\Gamma\begin{bmatrix} \alpha, b, b - a - \alpha \\ b - a, b - \alpha \end{bmatrix} & \\ \times {}_3F_2\left[\alpha, \bar{a}, 1 + \alpha - b; \bar{b}, 1 + \alpha + a - b; \frac{k}{k'}\right]. & \end{aligned} \tag{20}$$

The proof here follows the method of Slater.[6] The function ${}_1F_1[\bar{a}; \bar{b}; kr]$ in Eq. (1) is expanded as a series, and the integration is done term by term using

$$\begin{aligned} \int_0^\infty dt e^{-k't} t^{b-1} {}_1F_1[a; c; k't] & \\ = \Gamma(b)(k')^{-b}\, {}_2F_1[a, b; c; 1], & \end{aligned} \tag{21}$$

which is valid when the right-hand side converges, to yield

$$I = (k')^{-\alpha}\Gamma(\alpha)\sum_{m=0}^{\infty}\frac{(\bar{a})_m(\alpha)_m}{(\bar{b})_m m!}\left(\frac{k}{k'}\right)^m {}_2F_1[a, \alpha + m; b; 1]. \tag{22}$$

The hypergeometric functions of unit argument may be summed by Gauss' theorem if

$$\text{Re}\,[b - a - (\alpha + m)] > 0.$$

This can only be satisfied for all m if the sum terminates, i.e., if $\bar{a}$ is a negative integer. In this case the condition Re $(b - \alpha - a + \bar{a}) > 0$ assures that all the hypergeometric functions in Eq. (21) can be summed. If, on the other hand, the parameter a is a negative integer, then the hypergeometric series themselves terminate and may be summed by Vandermonde's theorem. In either case the result is

$${}_2F_1[a, \alpha + m; b; 1] = \frac{\Gamma(b)\Gamma(b - a - \alpha - m)}{\Gamma(b - a)\Gamma(b - \alpha - m)}. \tag{23}$$

Upon substituting this into Eq. (22) and performing minor rearrangements, we have the result (20). One might here note the somewhat curious result that the case of terminating series, Eq. (20), is just the first term of our answer in Case (2) [Eq. (12)].

III. APPLICATION TO INELASTIC ELECTRON SCATTERING

The occurrence of integrals of the type evaluated in Sec. II in election scattering derives from the form of the solution of the Dirac equation for an electron in the field of a point nucleus of charge Ze. These solutions

are well known and have the form[11]

$$\psi_\kappa^\mu = \begin{Bmatrix} g_\kappa(r)\chi_\kappa^\mu \\ if_\kappa(r)\chi_{-\kappa}^\mu \end{Bmatrix}, \tag{24}$$

where the representation called standard in Ref. 11 has been used for the Dirac matrices γ_ν. The spin-angle functions χ_κ^μ and $\chi_{-\kappa}^\mu$ are eigenfunctions of the operator $K = (\boldsymbol{\sigma}\cdot\mathbf{L} + 1)$ with eigenvalues $-\kappa$ and κ, respectively, and μ is the eigenvalue of the third component of total angular momentum. We use the radial functions with the normalization of Ref. 2,

$$\begin{Bmatrix} f_\kappa(r) \\ g_\kappa(r) \end{Bmatrix} = \begin{Bmatrix} -(E - m_e)/k \\ 1 \end{Bmatrix} \frac{(kr)^{\gamma-1}2^\gamma e^{\pi\eta/2}\,|\Gamma(\gamma + i\eta)|}{\Gamma(2\gamma + 1)} \times \begin{Bmatrix} \mathrm{Im} \\ \mathrm{Re} \end{Bmatrix} [\gamma + i\eta]e^{i\phi}e^{-ikr} \times {}_1F_1[\gamma + 1 + i\eta;\, 2\gamma + 1;\, 2ikr] \tag{25}$$

and

$$\gamma = [\kappa^2 - (\alpha Z)^2]^{\frac{1}{2}},$$
$$\eta = \alpha ZE/k, \tag{26}$$
$$e^{2i\phi} = e^{-i\pi}(\kappa - i\alpha Zm_e/k)/(\gamma + i\eta), \quad -\pi \leq \phi \leq 0,$$

where α is the fine-structure constant, E is the total energy, and k is the momentum.

When a nucleus is excited by an incident electron via one photon exchange, one may represent the process as the interaction between two charge and current distributions. The electron 4-current is

$$S_\nu(r_e, t) = -ie\psi_t^\dagger\gamma_4\gamma_\nu\psi_i, \tag{27}$$

where ψ_i and ψ_t are initial and final electron states. In the following we will only be interested in the transitions $(\kappa, \mu) \to (\kappa', \mu')$. With Ψ_i and Ψ_f the initial and final nuclear wavefunctions, the interaction matrix element becomes

$$\langle H_{\mathrm{int}}\rangle_{\mathrm{fi}} = \int \Psi_f^\dagger j_\nu^{(N)}[ieA_\nu(\kappa'\mu', \kappa\mu, \mathbf{r}_N)]\Psi_i\, d\mathbf{r}_N, \tag{28}$$

where A_ν is the 4-vector potential due to the electron and $j_\nu^{(N)}$ is the nuclear 4-vector current operator. The potential due to the electron 4-current, Eq. (27), is the retarded one which gives the result

$$\langle H_{\mathrm{int}}\rangle_{\mathrm{fi}} = \alpha \int d\mathbf{r}_N \int d\mathbf{r}_e \Psi_f^\dagger j_\nu^{(N)}\, \psi_{\kappa'}^{\mu'}\gamma_4\gamma_\nu \frac{e^{i\omega|\mathbf{r}_N - \mathbf{r}_e|}}{|\mathbf{r}_N - \mathbf{r}_e|}\, \psi_\kappa^\mu \Psi_i, \tag{29}$$

where ω is the excitation energy of the nucleus. The evaluation of this matrix element may be reduced to integrals over r_e and r_N, i.e., radial matrix elements. The procedure is simplified by recognizing that, for values of $|\kappa|$ greater than some minimum value $|\kappa|_{\min}$, the electron wavefunctions are sufficiently small in the neighborhood of the nucleus that they essentially do not penetrate the nuclear volume. However, owing to the long-range nature of electromagnetic forces, there is still significant interaction. Here we adopt the no-overlap approximation and consider $|\kappa| > |\kappa|_{\min}$. In this approximation several simplifications are possible. First, the integrations over nuclear and electron coordinates separate completely, because, as long as $r_e > r_N$, one has the expansion

$$e^{i\omega|\mathbf{r}_N - \mathbf{r}_e|}/|\mathbf{r}_N - \mathbf{r}_e| = 4\pi i\omega \sum_{L,M} j_L(\omega r_N)h_L^{(1)}(\omega r_e)Y_L^{M*}(\hat{\mathbf{r}}_N)Y_L^M(\hat{\mathbf{r}}_e).$$

Secondly, the scalar and longitudinal components of the interaction cancel,[12] and one need consider only transverse electric and transverse magnetic interactions. The contribution to the matrix element from transverse electric interactions of multipole order 2^L (projection M) is

$$E_{LM}(\kappa'\kappa) = 4\pi i\alpha\omega \int d\mathbf{r}_N \mathbf{j}_N \cdot \mathbf{N}_{LM}^{N*} \int d\mathbf{r}_e \psi_{\kappa'}^{\mu'}\boldsymbol{\alpha}\psi_\kappa^\mu \cdot \mathbf{N}_{LM}^e, \tag{30}$$

where $\mathbf{j}_N$ is the nuclear 3-current, $\boldsymbol{\alpha}$ is the Dirac operator for the electron 3-current, and N_{LM} is the transverse electric Hansen solution of the vector Helmholtz equation[13]

$$\mathbf{N}_{LM}^N = i\omega^{-1}[L(L+1)]^{-\frac{1}{2}}\nabla \times \mathbf{L}[j_L(\omega r_N)Y_L^M(\hat{\mathbf{r}}_N)],$$
$$\mathbf{N}_{LM}^e = i\omega^{-1}[L(L+1)]^{-\frac{1}{2}}\nabla \times \mathbf{L}[h_L^{(1)}(\omega r_e)Y_L^M(\hat{\mathbf{r}}_e)]. \tag{31}$$

The integral over nuclear coordinates in Eq. (30) will be denoted by $B_{LM}^{\mathcal{E}}$; this is the only way in which the structure of the nucleus enters the calculation. The integration over the electron angular coordinates is straightforward, so that

$$E_{LM} = \alpha B_{LM}^{\mathcal{E}}(-)^{j+\frac{1}{2}}\{4\pi(2j+1)/[L(L+1)]\}^{\frac{1}{2}} \times C(jj'L;\, -\tfrac{1}{2}, \tfrac{1}{2})C(jLj';\, \mu M\mu')R^{\mathcal{E}}(\kappa'L\kappa), \tag{32}$$

where j is the total angular momentum of the electron, and the radial matrix element is

$$\begin{aligned} R^{\mathcal{E}}(\kappa'L\kappa) &= L(L+1)\int_0^\infty r_e\, dr_e h_L^{(1)}(\omega r_e)(f_\kappa g_{\kappa'} - g_\kappa f_{\kappa'}) \\ &\quad - L(\kappa - \kappa')\int_0^\infty r_e\, dr_e h_L^{(1)}(\omega r_e)(f_\kappa g_{\kappa'} + g_\kappa f_{\kappa'}) \\ &\quad + \omega(\kappa - \kappa')\int_0^\infty r_e^2\, dr_e h_{L-1}^{(1)}(\omega r_e)(f_\kappa g_{\kappa'} + g_\kappa f_{\kappa'}). \end{aligned} \tag{33}$$

The contribution from transverse magnetic interactions can be developed similarly to the above, with the result

$$M_{LM} = \alpha B^{\mathcal{M}}_{LM}(-1)^{j-\frac{1}{2}}\{4\pi(2j+1)/[L(L+1)]\}^{\frac{1}{2}} \times C(jj'L; -\tfrac{1}{2}, \tfrac{1}{2})C(jLj'; \mu M\mu')R^{\mathcal{M}}(\kappa' L\kappa), \tag{34}$$

and the nuclear part is

$$B^{\mathcal{M}}_{LM} = \int d\mathbf{r}_N \mathbf{j}_N \cdot \{[L(L+1)]^{-\frac{1}{2}}\mathbf{L}[j_L(\omega r_N)Y^M_L(\hat{\mathbf{r}}_N)]\}^*. \tag{35}$$

Now the spherical Hankel function may be expanded as

$$h^{(1)}_\lambda(\omega r_e) = \sum_{n=1}^{\lambda+1}\Gamma\begin{bmatrix}\lambda+n\\ n, 2+\lambda-n\end{bmatrix}\frac{2^{1-n}i^{2+n-\lambda}e^{i\omega r_e}}{(\omega r_e)^n}. \tag{36}$$

Hence, the above integrations reduce to the type evaluated in Sec. II if one can write $\omega \simeq k_1 - k_2$, where $k_1(k_2)$ is the initial (final) electron momentum. This amounts to ignoring second-order terms in the electron mass as compared to k_1 and k_2, since

$$e^{i\omega r_e} = e^{i(k_1-k_2)r_e}[1 - (im_e^2/k_1k_2)r_e + \cdots]. \tag{37}$$

The first-order correction in the ratio m_e/k comes only from the coefficient $(E - m_e)/k$ in the definition of f_κ, Eq. (25), and can be included without difficulty. The lowest-order approximation (ignoring the electron mass altogether) results in

$$R^{\mathcal{E}}(\kappa' L\kappa) = \sum_{n=0}^{L+1}\Gamma\begin{bmatrix}L+n\\ n+1, 2+L-n\end{bmatrix}(2\omega)^{-n}i^{n-L} \times \left\{2nL(L+1)\int_0^\infty r^{1-n}\,dr e^{i(k_1-k_2)r}(g_\kappa f_{\kappa'} - f_\kappa g_{\kappa'}) + (\kappa-\kappa')[L(L+1)+n(n-1)] \times \int_0^\infty r^{1-n}\,dr e^{i(k_1-k_2)r}(f_\kappa g_{\kappa'} + g_\kappa f_{\kappa'})\right\}. \tag{38}$$

On substituting for f_κ, $f_{\kappa'}$, g_κ, and $g_{\kappa'}$ from Eq. (25) one has the above integrals cast into a common form, namely

$$I(\gamma_1\gamma_2\eta_1\eta_2k_1k_2lmn) = \int_0^\infty dr r^{(\gamma_1+\gamma_2-n)-1}e^{-2ik_2r} \times {}_1F_1[\gamma_1 + l + i\eta_1; 2\gamma_1+1; 2ik_1r] \times {}_1F_1[\gamma_2 + m + i\eta_2; 2\gamma_2+1; 2ik_2r], \tag{39}$$

where γ_1, η_1, $k_1(\gamma_2, \eta_2, k_2)$ are the initial (final) values of the parameters of Eq. (25) and Kummer's first theorem has been used to yield

$$\{e^{-ikr}{}_1F_1[\gamma+1+i\eta; 2\gamma+1; 2ikr]\}^* = e^{-ikr}{}_1F_1[\gamma+i\eta; 2\gamma+1; 2ikr]. \tag{40}$$

For brevity, integral (40) will be denoted by $I(l, m, n)$, and its value has already been found in Eq. (5):

$$I(l,m,n) = (2k_1)^{-\gamma_1-\gamma_2+n}\left\{e^{i\pi/2(\gamma_1+\gamma_2-n)}\Gamma\begin{bmatrix}\gamma_1+\gamma_2-n, 2\gamma_1+1, -\gamma_2+n+l+i\eta_1\\ \gamma_1+l+i\eta_1, \gamma_1-\gamma_2+1+n\end{bmatrix} \times {}_3F_2\left[\begin{matrix}\gamma_1+\gamma_2-n, \gamma_2+1-m-i\eta_2, \gamma_2-\gamma_1-n\\ 2\gamma_2+1, \gamma_2+1-n-l-i\eta_1\end{matrix}; \frac{k_2}{k_1}\right] + (k_2/k_1)^{-\gamma_2+n+l}e^{-\pi\eta_1+i[\frac{1}{2}\pi(\gamma_1-\gamma_2+n+2l)+\eta_1\log(k_2/k_1)]} \times \Gamma\begin{bmatrix}2\gamma_1+1, 2\gamma_2+1, \gamma_2-l-n-i\eta_1, 1+n+l-m+i(\eta_1-\eta_2)\\ \gamma_1+1-l-i\eta_1, \gamma_2+1-m-i\eta_2, \gamma_2+1+l+n+i\eta_1\end{bmatrix} \times {}_3F_2\left[\begin{matrix}\gamma_1+l+i\eta_1, -\gamma_1+l+i\eta_1, 1+n+l-m+i(\eta_1-\eta_2)\\ -\gamma_2+1+l+n+i\eta_1, \gamma_2+1+l+n+i\eta_1\end{matrix}; \frac{k_2}{k_1}\right]\right\}. \tag{41}$$

In terms of these functions we can now write

$$R^{\mathcal{E}}(\kappa' L\kappa) = \sum_{n=0}^{L+1}\Gamma\begin{bmatrix}L+n\\ n+1, 2+L-n\end{bmatrix}2^{-n-1}\omega^{-n}i^{n-L-1}A_{12} \times \{2nL(L+1)[(\gamma_1+i\eta_1)(\gamma_2-i\eta_2)e^{i(\phi_1-\phi_2)}I(1,0,n) - (\gamma_1-i\eta_1)(\gamma_2+i\eta_2)e^{i(\phi_2-\phi_1)}I(0,1,n)] + (\kappa-\kappa')[L(L+1)+n(n-1)] \times [(\gamma_1-i\eta_1)(\gamma_2-i\eta_2)e^{-i(\phi_1+\phi_2)}I(0,0,n) - (\gamma_1+i\eta_1)(\gamma_2+i\eta_2)e^{i(\phi_1+\phi_2)}I(1,1,n)]\}, \tag{42}$$

where

$$A_{12} = \frac{2^{\gamma_1+\gamma_2}k_1^{\gamma_1-1}k_2^{\gamma_2-1}e^{\frac{1}{2}\pi(\eta_1+\eta_2)}|\Gamma(\gamma_1+i\eta_1)\Gamma(\gamma_2+i\eta_2)|}{\Gamma(2\gamma_1+1)\Gamma(2\gamma_2+1)}.$$

The radial matrix element for the transverse magnetic interaction may be developed similarly, with the result

$$R^{\mathcal{M}}(\kappa' L\kappa) = \sum_{n=0}^{L}\Gamma\begin{bmatrix}L+n+1\\ n+1, 1+L-n\end{bmatrix}2^{-n-1}\omega^{-n}i^{n-L}A_{12} \times (\kappa+\kappa')[(\gamma_1+i\eta_1)(\gamma_2+i\eta_2)e^{i(\phi_1+\phi_2)}I(1,1,n) - (\gamma_1-i\eta_1)(\gamma_2-i\eta_2)e^{-i(\phi_1+\phi_2)}I(0,0,n)]. \tag{43}$$

IV. REMARKS AND CONCLUSIONS

The results given in Secs. II and III have been confirmed, as far as possible, by numerical methods. This not only serves as a check on the formal analysis, but also ensures that the expressions, although formally convergent, can be rendered in numerical form by a reasonably compact procedure.

The expressions given for the radials integrals (42) and (43) are for high-energy electron scattering with energy transferred to the nucleus. The real part of the same expressions would occur in the calculation of emission of radiation (Bremsstrahlung) by an electron moving in a Coulomb field. In either case, we are restricted to the "ultrarelativistic" limit, where $k \gg m_e$ except insofar as correction terms, in powers of m_e/k, can be added. Where $k \simeq m$, neither this analysis, nor the nonrelativistic analysis,[1,12] can be applied.

ACKNOWLEDGMENTS

We would like to thank Dr. Lucy J. Slater for helpful correspondence. During the early stages of this work, one of the authors (W. Gargaro) was supported by the National Aeronautics and Space Administration under its trainee program. Numerical work in connection with this paper was carried out on the IBM 360/44 computer at Ohio University.

* Work supported by the Army Research Office (Durham).

[1] J. T. Reynolds, D. S. Onley, and L. C. Biedenharn, J. Math. Phys. **5**, 411 (1964).

[2] J. A. Griffy, D. S. Onley, J. T. Reynolds, and L. C. Biedenharn, Phys. Rev. **128**, 833 (1962).

[3] D. S. Onley, T. A. Griffy, and J. T. Reynolds, Phys. Rev. **129**, 1689 (1962).

[4] D. S. Onley, J. T. Reynolds, and L. E. Wright, Phys. Rev. **134**, B945 (1964).

[5] S. T. Tuan, L. E. Wright, and D. S. Onley, Nucl. Instr. Meth. **60**, 70 (1968).

[6] L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge University Press, Cambridge, 1960), pp. 54–56.

[7] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), p. 322.

[8] E. D. Rainville, *Special Functions* (The Macmillan Co., New York, 1965).

[9] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge University Press, Cambridge, 1966), p. 36.

[10] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi, Eds., *Higher Transcendental Functions* (McGraw-Hill Book Co., Inc., New York, 1953), Vol. 1, p. 107.

[11] M. E. Rose, *Relativistic Electron Theory* (John Wiley & Sons, Inc., New York, 1961).

[12] L. C. Biedenharn and P. J. Brussaard, *Coulomb Excitation* (Clarendon Press, Oxford, 1965).

[13] W. W. Hansen, Phys. Rev. **47**, 139 (1935).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Reduced Matrix Elements of Tensor Operators*

JOHN S. BRIGGS
Argonne National Laboratory, Argonne, Illinois 60439

(Received 9 October 1969)

Reduced matrix elements of a tensorial product of n tensor operators in the basis of N-particle angular-momentum eigenstates are expanded into a sum of products of n 1-particle reduced matrix elements and a single $(N + n)$-particle recoupling coefficient. Application of the formula is illustrated by specific examples. The method leaves the coupling schemes of N-particle states undisturbed, which allows summation over intermediate states in a product of matrix elements to be made for any number of factors. A formula is given for the sum of products of two matrix elements and the extension to a greater number of matrix elements is illustrated by an example which is reduced to a form suitable for numerical evaluation. Such summations over products of matrix elements occur in the perturbation theory of configuration interaction and have hitherto been discussed in terms of effective operators. The connection of the method used here with the effective-operator approach is demonstrated.

1. INTRODUCTION

The study of the properties of tensor operators in the basis of eigenstates of angular momentum was initiated by Racah,[1] who gave an expression for the reduced matrix element of two coupled tensor operators between 2-particle states in terms of the reduced matrix elements of single tensor operators between 1-particle states. Subsequent extension to matrix elements of greater complexity proceeded by a sequence of recoupling transformations designed to bring the matrix element to the Racah form. The complicating features of such an approach are that each recoupling step introduces summations over intermediate quantum numbers and the procedure is generally different for each matrix element, being dependent upon the coupling schemes of the tensorial sets (operators or momentum eigenstates) involved. Here a formula is derived which is applicable to any matrix element of n coupled tensor operators between N-particle states. Furthermore, the formula permits any reduced matrix element to be immediately written as a sum of products of 1-particle reduced matrix elements with a single recoupling coefficient.

The method of derivation is a generalization of the method of Fano, Prats, and Goldschmidt[2] (hereafter referred to as FPG) who considered a product of two tensor operators acting between N-particle states. The method recognizes that both operators and momentum eigenstates are tensorial sets[3] and introduces mock particles whose eigenstates are tensorial sets of rank equal to that of a tensor operator. Fano[4] later used the same method to calculate matrix elements of the scalar product of Racah tensors between eigenstates of orbital angular momentum and called the mock particles "orbitons." In this general treatment the particles are called "momentons."

The general formula is derived in Sec. 2 and its application is illustrated by several examples in Sec. 3. The "momenton" method expresses the recoupling of momenta involved in any matrix element as a single recoupling coefficient. The recoupling coefficients occurring may be reduced to sums of products of 6-j and/or 9-j symbols for numerical evaluation. This reduction is readily achieved either by use of the diagrammatic techniques developed in the exhaustive treatment of Yutsis, Levinson, and Vanagas[5] (referred to as YLV) or by explicit expansion of the coefficient (Ref. 3, Chap. 9). The latter method is found to be more direct in this case and is used in the examples.

The momenton method has the further advantage that it does not disturb the coupling schemes of the N-particle states involved. This feature makes it eminently suitable for the evaluation of products of matrix elements where summation is made over the coupling schemes of intermediate states. Such products occur, for example, in the perturbation theory of configuration interaction in atoms.[6-8] The summation is carried out using the group property of recoupling coefficients and in Sec. 4 a formula is given for the product of two matrix elements summed over the states of the intermediate configuration of particles. The extension of the result to any number of matrix elements is straightforward and a worked example is given as illustration.

Previous treatments of configuration interaction[6-8] have investigated the structure of perturbation terms and showed that the sum of products of matrix elements may be replaced by "effective" operators acting between initial and final states only. The formulas given in Secs. 2 and 4 enable these effective operators to be derived by an expansion of the recoupling coefficients given by the momenton method. This connection with previous work is demonstrated in Sec. 5.

Shell-model calculations involve antisymmetrized matrix elements, which are expanded in terms of the corresponding unsymmetrized matrix elements.[1,4] The coefficients of the expansion are products of a phase factor, coefficients of fractional parentage and a numerical factor depending upon the occupation

numbers of the subshells. In the following, the unsymmetrized matrix elements are considered.

2. THE GENERAL CASE

The essential simplifying step in the treatment of FPG is to decouple the tensor operators so that they may be applied separately and independently to single-particle wavefunctions without disturbing the coupling scheme of the many-particle wavefunction. The general type of matrix element to be evaluated is a coupled product of n tensor operators acting between N-particle wavefunctions, i.e.,

$$([j_1 j_2 \cdots j_N], \alpha JM | [\mathbf{A}_r^{[a_r]}\mathbf{B}_s^{[b_s]}\mathbf{C}_t^{[c_t]} \cdots, \beta]_Q^{[K]} | [j_1' j_2' \cdots j_N'], \alpha' J' M'). \tag{2.1}$$

Here α, α', and β denote the coupling schemes and values of the intermediate quantum numbers of the two N-particle wavefunctions and the set of tensor operators, respectively. The subscripts r, s, t, $\cdots$ denote the 1-particle spaces in which the operators act; generally, any number of operators may act within each 1-particle space.

The coupled product of tensor operators in (2.1) is expanded as a linear combination of the direct products of tensor components, the expansion coefficients being generalized Clebsch–Gordan coefficients

$$[\mathbf{A}_r^{[a_r]}\mathbf{B}_s^{[b_s]}\mathbf{C}_t^{[c_t]} \cdots \beta]_Q^{[K]} = \sum_{x,y,z,\cdots} A_{r,x}^{[a_r]} B_{a,y}^{[b_s]} C_{t,z}^{[c_t]} \cdots (a_r x b_s y c_t z \cdots | (a_r b_s c_t \cdots)\beta, KQ), \tag{2.2}$$

where, for example, the subscript x denotes one of the $2a_r + 1$ components of the tensor $\mathbf{A}^{[a_r]}$. Momenton wavefunctions are now introduced and written in tensorial-set notation $u_q^{[k]}$. The orthogonality of the momenton wavefunctions

$$u_{q'}^{[k]*} u_q^{[k]} = \delta_{qq'}$$

enables (2.2) to be written

$$[\mathbf{A}_r^{[a_r]}\mathbf{B}_s^{[b_s]}\mathbf{C}_t^{[c_t]} \cdots \beta]_Q^{[K]} = \sum_{x,y,z,\cdots} \sum_{x',y',z',\cdots} (A_{r,x}^{[a_r]} B_{s,y}^{[b_s]} C_{t,z}^{[c_t]} \cdots) \times (u_x^{[a_r]*} u_y^{[b_s]*} u_z^{[c_t]*} \cdots)(u_{x'}^{[a_r]} u_{y'}^{[b_s]} u_{z'}^{[c_t]} \cdots) \times (a_r x' b_s y' c_t z' \cdots | (a_r b_s c_t \cdots)\beta, KQ). \tag{2.3}$$

Equation (2.3) is the generalization of Eq. (33) of FPG. The coupling scheme β has been transferred to the momenton wavefunctions, i.e., Eq. (2.3) may be written

$$[\mathbf{A}_r^{[a_r]}\mathbf{B}_s^{[b_s]}\mathbf{C}_t^{[c_t]} \cdots, \beta]_Q^{[K]} = (\mathbf{A}_r^{[a_r]}\mathbf{u}^{[a_r]*})(\mathbf{B}_s^{[b_s]}\mathbf{u}^{[b_s]*})(\mathbf{C}_t^{[c_t]}\mathbf{u}^{[c_t]*}) \cdots [\mathbf{u}^{[a_r]}\mathbf{u}^{[b_s]}\mathbf{u}^{[c_t]} \cdots, \beta]_Q^{[K]}. \tag{2.4}$$

The operators may now be applied independently to the 1-particle wavefunctions in the N-particle wavefunction

$$([j_1 j_2 j_3 \cdots j_N]\alpha, jm| = \sum_{m_1,m_2,m_3} ((j_1 j_2 j_3 \cdots), \alpha jm \,|\, j_1 m_1 j_2 m_2 j_3 m_3 \cdots) \times (j_1 m_1 \,|\, 1)(j_2 m_2 \,|\, 2)(j_3 m_3 \,|\, 3) \cdots. \tag{2.5}$$

It is important to recognize that the factors $\mathbf{A}_r^{[a_r]}\mathbf{u}^{[a_r]*}$ may be reordered to bring operators acting in the same 1-particle space together. For each 1-particle space the ordering of the operators must be the same as that in the original n-operator product. This is to be contrasted with the conventional approach which works with the coupled n-operator product so that the reordering into groups of operators acting in the same 1-particle space involves a recoupling transformation. Each group of operators acting within the same 1-particle space is now applied to the appropriate 1-particle wavefunction in (2.5). One such group of n_i operators $\mathbf{A}_i^{[a_i]}\mathbf{B}_i^{[b_i]} \cdots \mathbf{Z}_i^{[z_i]}$ acting in the ith 1-particle space will be considered. Operating to the left on the ith 1-particle wavefunction gives

$$(j_i m_i \,|\, i)(\mathbf{A}_i^{[a_i]}\mathbf{u}^{[a_i]*})(\mathbf{B}_i^{[b_i]}\mathbf{u}^{[b_i]*}) \cdots (\mathbf{Z}_i^{[z_i]}\mathbf{u}^{[z_i]*}) = \sum_{q,q',\cdots,q''} \sum_{j_\rho,j_\sigma,\cdots,j_{\omega+1}} \sum_{m_\rho,m_\sigma,\cdots,m_{\omega+1}} (j_i m_i \,|\, i) A_{i,q}^{[a_i]} (i \,|\, j_\rho m_\rho)(j_\rho m_\rho \,|\, i) B_{i,q'}^{[b_i]} \times (i \,|\, j_\sigma m_\sigma)(j_\sigma m_\sigma \,|\, i) \cdots (j_\omega m_\omega \,|\, i) Z_{i,q''}^{[z_i]} (i \,|\, j_{\omega+1} m_{\omega+1})(j_{\omega+1} m_{\omega+1} \,|\, i) u_q^{[a_i]*} u_{q'}^{[b_i]*} \cdots u_{q''}^{[z_i]*}. \tag{2.6}$$

Application of the Wigner–Eckart theorem to each matrix element of the tensor operators gives

$$\begin{aligned}(j_i m_i \,|\, i)(\mathbf{A}_i^{[a_i]}\mathbf{u}^{[a_i]*})(\mathbf{B}_i^{[b_i]}\mathbf{u}^{[b_i]*}) \cdots (\mathbf{Z}_i^{[z_i]}\mathbf{u}^{[z_i]*}) &= \sum_{q,q',\cdots,q''} \sum_{j_\rho,j_\sigma,\cdots,j_{\omega+1}} \sum_{m_\rho,m_\sigma,\cdots,m_{\omega+1}} ([j_i][j_\rho][j_\sigma] \cdots [j_\omega])^{-\frac{1}{2}} (j_i \| A_i^{[a_i]} \| j_\rho)(j_\rho \| B_i^{[b_i]} \| j_\sigma) \cdots (j_\omega \| Z_i^{[z_i]} \| j_{\omega+1}) \\ &\times (j_i m_i \,|\, j_\rho m_\rho a_i q)(j_\rho m_\rho \,|\, j_\sigma m_\sigma b_i q') \cdots (j_\omega m_\omega \,|\, j_{\omega+1} m_{\omega+1} z_i q'')(j_{\omega+1} m_{\omega+1} \,|\, i) u_q^{[a_i]*} u_{q'}^{[b_i]*} \cdots u_{q''}^{[z_i]*} \\ &= \sum_{j_\rho,j_\sigma,\cdots,j_{\omega+1}} [j_i]^{-\frac{1}{2}} (j_i \| A_i^{[a_i]} \| j_\rho)[j_\rho]^{-\frac{1}{2}} (j_\rho \| B_i^{[b_i]} \| j_\sigma) \cdots [j_\omega]^{-\frac{1}{2}} \\ &\times (j_\omega \| Z_i^{[z_i]} \| j_{\omega+1})[[[[\{j_{\omega+1}| \times \mathbf{u}^{[z_i]*}]^{[j_\omega]} \cdots]^{[j_\sigma]} \times \mathbf{u}^{[b_i]*}]^{[j_\rho]} \times \mathbf{u}^{[a_i]*}]_{m_i}^{[j_i]},\end{aligned} \tag{2.7}$$

where the usual abbreviation $[j] \equiv (2j+1)$ has been made.

Equation (2.7) gives the result of operating to the left on the ith 1-particle wavefunction with the n_i operators which are coupled with n_i momentons according to the transformation (2.4). This result is a product of n_i 1-particle reduced matrix elements, involving $n_i - 1$ intermediate summations over complete sets of 1-particle wavefunctions, together with a coupled product of particle i and n_i momentons. When the overlap is taken with the N-particle wavefunction on the rhs of the matrix element (2.1), the orthogonality of 1-particle wavefunctions gives $j_{\omega+1} = j_i'$ only. Therefore, the n_i momentons are coupled *sequentially* onto the momentum j_i' of particle i in the right-hand N-particle wavefunction to give a momentum j_i which is that of particle i in the left-hand N-particle wavefunction. As may be seen from (2.7), the coupling scheme of this group of wavefunctions is obtained by reading from right to left the sequence of momenta and ranks appearing in the product of 1-particle reduced matrix elements.

The intermediate momenta j_ρ, j_σ, etc., introduced by the summations over complete sets in (2.7) are local to the sequential coupling within each 1-particle group and do not enter into the coupling schemes α, α', or β. For the sake of clarity they will be suppressed henceforth except by indication under the summation sign. For instance, the product of 1-particle reduced matrix elements $(j_i\| A_i^{[a_i]} \| j_\rho) \cdots$, together with the "normalization" factors $[j]^{-\frac{1}{2}}$, will be indicated by the symbol $(j_i\| A_i^{[a_i]} B_i^{[b_i]} \cdots Z_i^{[z_i]} \| j_{\omega+1})$ inside the summation sign. Also, the sequentially coupled wavefunction $\{(([(j_{\omega+1} z_i) j_\omega \cdots] j_\sigma b_i) j_\rho a_i) j_i|$ of the particle and n_i momentons on the rhs of (2.7) will be denoted by $\{[j_{\omega+1}, \{r_i\}] j_i|$, where $\{r_i\}$ denotes the ordered group of momentons with rank $z_i \cdots a_i$. In this abbreviated notation, Eq. (2.7) is written, for the set of wavefunctions $\{j_i|$, as

$$\{j_i|\,(\mathbf{A}_i^{[a_i]}\mathbf{u}^{[a_i]^*})(\mathbf{B}_i^{[b_i]}\mathbf{u}^{[b_i]^*}) \cdots (\mathbf{Z}_i^{[z_i]}\mathbf{u}^{[z_i]^*}) = \sum_{j_\rho, j_\sigma, \cdots, j_{\omega+1}} (j_i\| A_i^{[a_i]} B_i^{[b_i]} \cdots Z_i^{[z_i]} \| j_{\omega+1}) \times \{[j_{\omega+1}, \{r_i\}] j_i|. \quad (2.8)$$

Each set of operators acting in the same 1-particle space gives rise to a factor like the rhs of (2.8). The final momenta j_i of each of the N groups, consisting of particle i and n_i momentons, are then coupled according to the scheme α of the left-hand N-particle wavefunction to give a coupled wavefunction of N particles and n momentons.

Upon substitution of the transformation (2.4), the rhs of the matrix element (2.1) becomes the (uncoupled) product of the n-momenton wavefunction

$$[\mathbf{u}^{[a_r]}\mathbf{u}^{[b_s]}\mathbf{u}^{[c_t]} \cdots, \beta]_Q^{[K]} \equiv |n, \beta K Q)$$

and the N-particle wavefunction

$$|[j_1' j_2' \cdots j_N'], \alpha' J' M') \equiv |N, \alpha' J' M').$$

The product is expanded as

$$|N, \alpha' J' M')\, |n, \beta K Q) = \sum_{jm} [|N, \alpha' J'\} \times |n, \beta K\}]_m^{[j]} (jm \mid J' M' K Q). \quad (2.9)$$

The orthogonality of $(N+n)$-particle wavefunctions gives only the term $j = J$, $m = M$ in the sum in (2.9) when the product is taken with the $(N+n)$-particle wavefunction arising from the n operators acting on the left-hand N-particle wavefunction. The scalar product of the two $(N+n)$-particle functions is a single recoupling coefficient.

The Wigner–Eckart theorem applied to (2.1) gives

$$(\alpha J M|\, \mathbf{O}_Q^{[K]}\, |\alpha' J' M') = [J]^{-\frac{1}{2}} (J M \mid J' M' K Q)(\alpha J\| O^{[K]} \|\alpha' J'), \quad (2.10)$$

where $\mathbf{O}_Q^{[K]}$ denotes the product of n tensor operators. Using the results (2.4), (2.8), (2.9), and (2.10), we can write the reduced matrix element of n tensor operators acting between N-particle momentum eigenstates as

$$([j_1 j_2 \cdots j_N], \alpha J\| [\mathbf{A}_r^{[a_r]}\mathbf{B}_s^{[b_s]}\mathbf{C}_t^{[c_t]} \cdots, \beta]^{[K]} \| [j_1' j' \cdots j_N'], \alpha' J') = \sum_{j_1,\rho,j_1,\sigma,\cdots,j_1,\omega} \sum_{j_2,\rho,j_2,\sigma,\cdots,j_2,\omega} \cdots \sum_{j_N,\rho,j_N,\sigma,\cdots,j_N,\omega} [J]^{\frac{1}{2}} (j_1\| A_1^{[a_1]} B_1^{[b_1]} \cdots Z_1^{[z_1]} \| j_1')(j_2\| A_2^{[a_2]} B_2^{[b_2]} \cdots Z_2^{[z_2]} \| j_2') \cdots (j_N\| A_N^{[a_N]} B_N^{[b_N]} \cdots Z_N^{[z_N]} \| j_N')([j_1', \{r_1\}] j_1, [j_2', \{r_2\}] j_2 \cdots [j_N', \{r_N\}] j_N, \alpha\, |[j_1' j_2' \cdots j_N', \alpha'] J', [a_r b_s c_t, \cdots, \beta] K)^{(J)}. \quad (2.11)$$

The formula (2.11) allows any reduced matrix element to be written as a sum of products of 1-particle reduced matrix elements and single recoupling coefficients. Aside from numerical factors of the type $[j]^{\frac{1}{2}}$, the recoupling coefficients are $3(N+n-1)$-j symbols or reduce to a product of lower-j symbols. The recoupling coefficient may be evaluated by the most direct method available. It is customary to expand the coefficient in products of 6-j and/or 9-j symbols. This reduction may be achieved either by the algebraic method used by Fano and Racah (Ref. 3, Chap. 9) or by the equivalent diagrammatic

technique of YLV. The present method enables the reduction to be used which gives the minimum number of intermediate summations, which is not always the case for the Racah method. Examples and specific cases are treated in the following section.

3. APPLICATIONS

A. Simple Examples

In the method of Racah,[1] the matrix elements of products of tensor operators are calculated by first expanding tensorial products (either operators or eigenstates) in terms of Clebsch–Gordan coefficients, extracting 1-particle reduced matrix elements, and subsequently recoupling. Fano and Racah[3] have circumvented the expansions in Clebsch–Gordan coefficients by direct recoupling of eigenstates. However, for a general matrix element of the type (2.1), several recoupling steps may be necessary although intermediate quantum numbers introduced by the recouplings may subsequently be summed over to give a single recoupling coefficient. The present method yields this single coefficient directly.

A simple example of the general type (2.1) is the matrix element of the product of two operators acting in different subspaces between 2-particle wavefunctions. In this case $(N + n - 1) = 3$ and the result is obtained from (2.11) in terms of a 9-j symbol, i.e.,

$$\begin{aligned}(j_1 j_2, j \| [\mathbf{A}_1^{[k_1]} \times \mathbf{B}_2^{[k_2]}]^{[k]} \| j_1' j_2', j') &= (j_1 \| A_1^{[k_1]} \| j_1')(j_2 \| B_2^{[k_2]} \| j_2')[j]^{\frac{1}{2}}[j_1]^{-\frac{1}{2}}[j_2]^{-\frac{1}{2}} \\ &\times ((j_1' k_1) j_1, (j_1' k_2) j_2 \,|\, (j_1' j_2') j', (k_1 k_2) k)^{(j)} \\ &= (j_1 \| A_1^{[k_1]} \| j_1')(j_2 \| B_2^{[k_2]} \| j_2')([j][k][j'])^{\frac{1}{2}} \\ &\times \begin{Bmatrix} j_1' & k_1 & j_1 \\ j_2' & k_2 & j_2 \\ j' & k & j \end{Bmatrix}. \end{aligned} \tag{3.1}$$

When derived by the Racah method, the expression differs from the rhs of (3.1) by a cyclic permutation of the columns of the 9-j symbol, which does not alter its value. Other well-known results involving 6-j symbols may also be obtained immediately from the general formula, for example, the case $N = 1$, $n = 2$,

$$\begin{aligned}(j_1 \| [\mathbf{A}_1^{[k_1]} \times \mathbf{B}_1^{[k_2]}]^{[k]} \| j_1') &= \sum_{j_1''} (j_1 \| A_1^{[k_1]} \| j_1'')(j_1'' \| B_1^{[k_2]} \| j_1')[j_1'']^{-\frac{1}{2}} \\ &\times ((j_1' k_2) j_1'', k_1 \,|\, j_1', (k_1 k_2) k)^{(j_1)} \\ &= \sum_{j_1''} (j_1 \| A_1^{[k_1]} \| j_1'')(j_1'' \| B_1^{[k_2]} \| j_1')[k]^{\frac{1}{2}}(-1)^{j_1 + j_1' + k} \\ &\times \begin{Bmatrix} k_1 & k_2 & k \\ j_1' & j_1 & j_1'' \end{Bmatrix} \end{aligned} \tag{3.2}$$

and the case $N = 2$, $n = 1$,[9]

$$\begin{aligned}(j_1 j_2, j \| \mathbf{A}_1^{[k]} \| j_1' j_2', j') &= (j_1 \| A_1^{[k]} \| j_1')[j]^{\frac{1}{2}}[j_1]^{-\frac{1}{2}}((j_1' k) j_1, j_2 \,|\, (j_1' j_2') j', k)^{(j)} \\ &= (j_1 \| A_1^{[k]} \| j_1')([j][j'])^{\frac{1}{2}}(-1)^{k + j_1 + j_2 + k'} \\ &\times \delta(j_2 j_2') \begin{Bmatrix} j_1 & j & j_2 \\ j' & j_1' & k \end{Bmatrix}. \end{aligned} \tag{3.3}$$

B. A Further Example

The general formula (2.11) is most useful in the case of more complicated matrix elements where it may avoid redundant intermediate summations. To illustrate the application of the present method, a specific example, treated in YLV (Sec. 34) by the conventional method, will be evaluated. According to Eq. (2.11), we have

$$\begin{aligned}(j_1 j_2, j \| [[\mathbf{T}_1^{[k_1]} \times \mathbf{U}_2^{[k_2]}]^{[k_{12}]} \times [\mathbf{V}_1^{[k_3]} \times \mathbf{W}_2^{[k_4]}]^{[k_{34}]}]^{[k]} \| j_1' j_2', j') &= \sum_{j_1'', j_2''} [j]^{\frac{1}{2}}(j_1 \| T_1^{[k_1]} V_1^{[k_3]} \| j_1')(j_2 \| U_2^{[k_2]} W_2^{[k_4]} \| j_2') \\ &\times ([j_1', \{r_1\}] j_1, [j_2', \{r_2\}] j_2 \,|\, (j_1' j_2') j', [(k_1 k_2) k_{12}, (k_3 k_4) k_{34}] k)^{(j)} \\ &= \sum_{j_1'', j_2''} [j]^{\frac{1}{2}}(j_1 \| T_1^{[k_1]} \| j_1'')(j_1'' \| V_1^{[k_3]} \| j_1')(j_2 \| U_2^{[k_2]} \| j_2'')(j_2'' \| W_2^{[k_4]} \| j_2') \\ &\times ([j_1][j_1''][j_2][j_2''])^{-\frac{1}{2}}([(j_1' k_3) j_1'' k_1] j_1, [(j_2' k_4) j_2'' k_2] j_2 \,|\, (j_1' j_2') j', [(k_1 k_2) k_{12}, (k_3 k_4) k_{34}] k)^{(j)}. \end{aligned} \tag{3.4}$$

The recoupling coefficient in (3.4) may be expanded in a way which involves a *single* intermediate summation to give an expression suitable for numerical evaluation

$$\begin{aligned}([(j_1' k_3) j_1'' k_1] j_1, [(j_2' k_4) j_2'' k_2] j_2 \,|\, (j_1' j_2') j', [(k_1 k_2) k_{12}, (k_3 k_4) k_{34}] k)^{(j)} &= \sum_l ((j_1'' k_1) j_1, (j_2'' k_2) j_2 \,|\, (j_1'' j_2'') l, (k_1 k_2) k_{12})^{(j)} \\ &\times ((j_1' k_3) j_1'', (j_2' k_4) j_2'' \,|\, (j_1' j_2') j', (k_3 k_4) k_{34})^{(l)} ((j' k_{34}) l, k_{12} \,|\, j', (k_{34} k_{12}) k)^{(j)} \times (-1)^{k_{34} + k_{12} - k} \\ &= \sum_l [l](-1)^{j + j' + k}([j_1][j_2][k_{12}][j_1''][j_2''][j'][k_{34}][k])^{\frac{1}{2}} \begin{Bmatrix} j' & k_{34} & l \\ k_{12} & j & k \end{Bmatrix} \begin{Bmatrix} j_1'' & k_1 & j_1 \\ j_2'' & k_2 & j_2 \\ l & k_{12} & j \end{Bmatrix} \begin{Bmatrix} j_1' & k_3 & j_1'' \\ j_2' & k_4 & j_2'' \\ j' & k_{34} & l \end{Bmatrix}. \end{aligned} \tag{3.5}$$

In this example, $N = 2$, $n = 4$, and the recoupling coefficient is proportional to a 15-j symbol. Utilizing the diagrammatic representation of YLV, we can show that the recoupling coefficient is proportional to a 15-j symbol of the third kind, namely,

$$\begin{Bmatrix} j_1 & j_1' & & j & j' & & j_2 & j_2' \\ & j_1'' & & & k & & & j_1'' \\ k_1 & & k_3 & k_{12} & & k_{34} & k_2 & & k_4 \end{Bmatrix}.$$

This is the result obtained in YLV (Sec. 34) by a procedure which involves a redundant intermediate summation. There is a standard formula given in YLV for the reduction of a 15-j symbol of the third kind which is equivalent to the expansion (3.5). The diagrammatic techniques are extremely useful for handling higher $3n$-j symbols, but in the present class of examples a direct expansion as in (3.5) may be preferable.

4. PRODUCTS OF MATRIX ELEMENTS

The examples of the previous section have dealt with situations where complex products of tensor operators act between states of simple structure (i.e., $N < n$). In shell-model calculations of many-particle systems, one usually encounters matrix elements between many-particle states of a complex structure. As noted in FPG, the momenton method then has the great advantage over conventional methods of leaving the coupling schemes of the many-particle states undisturbed. This property is even more useful in the case of products of matrix elements where summation is made over states of the intermediate configurations.

For a product of N matrix elements, application of Eq. (2.11) to each factor results in a product of N recoupling coefficients. Using the group property of recoupling coefficients, we can then make a summation over the coupling schemes of intermediate many-particle states. This summation gives a single recoupling coefficient which involves only the coupling schemes of the initial and final states in the original product of matrix elements. The case of a product of two matrix elements will be considered, since the extension to a product of many matrix elements is immediate. Consider the summation

$$\sum_{\alpha'} (j_1 \cdots j_N, \alpha J \| [\mathbf{A}_r^{[a_{r1}]} \mathbf{B}_s^{[b_{s1}]} \cdots \beta_1]^{[K_1]} \| j_1' \cdots j_N', \alpha' J') \times (j_1' \cdots j_N', \alpha' J' \| [\mathbf{A}_u^{[a_{u2}]} B_v^{[b_{v2}]} \cdots \beta_2]^{[K_2]} \| j_1'' \cdots j_N'', \alpha'' J''), \quad (4.1)$$

where the particles $r, s, \cdots, u, v, \cdots$ belong to the group of N particles whose momenta are coupled to form the states $(\alpha J|$, $(\alpha' J'|$, and $(\alpha'' J''|$. According to Eq. (2.11) each factor in (4.1) gives a sum over products of 1-particle reduced matrix elements with a recoupling coefficient. The 1-particle reduced matrix elements do not involve the coupling of angular momenta and will be omitted. The summation over α' in the product of recoupling coefficients arising from (4.1) follows from the group property of recoupling coefficients. The notation of (2.11) is used for the recoupling coefficients and the result is

$$\begin{aligned} \sum_{\alpha'} ([j_1', \{r_1^{(1)}\}] j_1 \cdots [j_N', \{r_N^{(1)}\}] j_N, \alpha \mid (j_1' \cdots j_N', \alpha') J', (a_{r1} b_{s1} \cdots, \beta_1) K_1)^{(J)} \\ \times ([j_1'', \{r_1^{(2)}\}] j_1' \cdots [j_N'', \{r_N^{(2)}\}] j_N', \alpha' \mid (j_1'' \cdots j_N'', \alpha'') J'', (a_{u2} b_{v2} \cdots, \beta_2) K_2)^{(J')} \\ = ([(j_1'', \{r_1^{(2)}\}) j_1', \{r_1^{(1)}\}] j_1 \cdots [(j_N'', \{r_N^{(2)}\}) j_N', \{r_N^{(1)}\}] j_N, \alpha \mid [(j_1'' \cdots j_N'', \alpha'') J'', \\ (a_{u2} b_{v2} \cdots, \beta_2) K_2] J', (a_{r1} b_{s1} \cdots, \beta_1) K_1)^{(J)}. \quad (4.2) \end{aligned}$$

From Eq. (4.2) it may be seen that the result of the summation over α' is simply to replace the momenta $j_i' \cdots j_N'$ on the lhs of the first recoupling coefficient by the coupled groups $[j_1'', \{r_1^{(2)}\}] j_1'$, etc., for the same particles on the lhs of the second recoupling coefficient. The rhs of the second coefficient has a coupling scheme $(J'' K_2) J'$ and replaces the N-particle state $|\alpha' J')$ on the rhs of the first coefficient. The resulting recoupling coefficient involves only the coupling schemes and α and α'' of the initial and final states. If only the first or only the second matrix element contains operators acting in the space of particle i, then a factor $\delta(j_i' j_i'')$ or $\delta(j_i j_i')$ is included, respectively. If neither matrix element contains operators acting in the space of particle i, then both factors are included.

By an extension of the procedure of (4.2), summation can be made over the states of intermediate configurations for any number of factors. The case of three factors involving two summations is illustrated by an example, i.e.,

$$\sum_{\alpha',\beta'}(j_rj_sj_t,\alpha J\|\,[\mathbf{A}_s^{[k_1]}\times\mathbf{B}_t^{[k_2]}]^{[k]}\,\|j_r'j_s'j_t',\alpha'J')(j_r'j_s'j_t',\alpha'J'\|\,\mathbf{C}_r^{[k_3]}\cdot\mathbf{D}_s^{[k_3]}\,\|l_r'l_s'l_t',\beta'L')(l_r'l_s'l_t',\beta'L'\|\,\mathbf{E}_t^{[k_4]}\,\|l_rl_sl_t,\beta L)$$
$$=(j_r'\|\,C_r^{[k_3]}\,\|l_r')(j_t\|\,B_t^{[k_2]}\,\|j_t')(l_t'\|\,E_t^{[k_4]}\,\|l_t)(j_s\|\,A_s^{[k_1]}\,\|j_s')(j_s'\|\,D_s^{[k_3]}\,\|l_s')$$
$$\times\Delta\times((l_rk_3)j_r,((l_sk_3)j_s',k_1)j_s,((l_tk_4)j_t'k_2)j_t,\alpha\,|\,((l_rl_sl_t)\beta L,k_4)J',(k_3k_3)0,(k_1k_2)k)^{(J)},\quad(4.3)$$

where Δ is the factor

$$\delta(J'L')\delta(l_r'l_r)\delta(j_r'j_r)\delta(l_s'l_s)\delta(l_t'j_t')([J][J'][L'][j_r'][j_t][l_t'][j_s][j_s'])^{-\frac{1}{2}}.$$

To evaluate the recoupling coefficient, the coupling schemes α and β must be given; we assume the schemes $((j_r,(j_sj_t)\bar{J})J|$ and $|((l_rl_s)\bar{L},l_t)L)$. In this case the recoupling in (4.3) reduces to a product of two 6-j symbols and a 12-j symbol and is evaluated via the expansion

$$((l_rk_3)j_r,[((l_sk_3)j_s',k_1)j_s,((l_tk_4)j_t'k_2)j_t]\bar{J}\,|\,(((l_rl_s)\bar{L},l_t)L,k_4)J',(k_3k_3)0,(k_1k_2)k)^{(J)}$$
$$=\sum_x((j_s'k_1)j_s,(j_t'k_2)j_t\,|\,(j_s'j_t')x,(k_1k_2)k)^{(\bar{J})}(j_r,(xk)\bar{J}\,|\,(j_rx)J',k)^{(J)}$$
$$\times(j_r,(j_s'j_t')x\,|\,(j_rj_s')\bar{L},j_t')^{(J')}((l_rk_3)j_r,(l_sk_3)j_s'\,|\,(l_rl_s)\bar{L},(k_3k_3)0)^{(L)}(\bar{L},(l_tk_4)j_t'\,|\,(\bar{L}l_t)L,k_4)^{(J')}$$
$$=P(-1)^{\phi}\begin{Bmatrix}\bar{L}&l_t&L\\k_4&J'&j_t'\end{Bmatrix}\begin{Bmatrix}j_r&l_r&k_3\\l_s&j_s'&\bar{L}\end{Bmatrix}\sum_x(-1)^x[x]\begin{Bmatrix}j_r&J'&x\\k&\bar{J}&J\end{Bmatrix}\begin{Bmatrix}j_r&J'&x\\j_t'&j_s'&\bar{L}\end{Bmatrix}\begin{Bmatrix}j_s'&j_t'&x\\k_1&k_2&k\\j_s&j_t&\bar{J}\end{Bmatrix},\quad(4.4)$$

where

$$P\equiv([k_3]^{-1}[j_s][j_t][k][\bar{J}][J'][\bar{L}][j_r][j_s'][j_t'][L])^{\frac{1}{2}}$$

and

$$\phi\equiv(2(\bar{L}+J'+j_s'+j_r)+J+k_3+k_4+k+l_r+l_t+j_t').$$

5. EFFECTIVE OPERATORS

The summations over intermediate states in a product of matrix elements which has been achieved in Sec. 4 by the momenton method, may also be performed by the introduction of unit tensor operators. This was the method introduced by Racah and Stein[7] in a study of the second-order perturbation of terms of the configuration l^N. The introduction of unit tensors allows 1-particle reduced matrix elements to be separated from the recoupling of momenta. When evaluated according to (2.11), the reduced matrix element of a product of unit tensor operators is simply a single recoupling coefficient, as illustrated by the following example:

$$(j_1j_2,j\|\,[\mathbf{A}_1^{[a]}\times\mathbf{B}_2^{[b]}]^{[k]}\,\|j_1'j_2',j')$$
$$=(j_1\|\,A_1^{[a]}\,\|j_1')(j_2\|\,B_2^{[b]}\,\|j_2')$$
$$\times(j_1j_2,j\|\,[\mathbf{a}_1^{[a]}\times\mathbf{b}_2^{[b]}]^{[k]}\,\|j_1'j_2',j')$$
$$=[j]^{\frac{1}{2}}([j_1][j_2])^{-\frac{1}{2}}(j_1\|\,A_1^{[a]}\,\|j_1')(j_2\|\,B_2^{[b]}\,\|j_2')$$
$$\times((j_1'a)j_1,(j_2'b)j_2\,|\,(j_1'j_2')j',(ab)k)^{(j)},\quad(5.1)$$

where

$$(j_1\|\,a_1^{[a]}\,\|j_1'')=\delta(j_1'j_1'')\quad\text{and}\quad(j_2\|\,b_2^{[b]}\,\|j_2'')=\delta(j_2'j_2'').$$

For each distinct 1-particle reduced matrix element, a different unit tensor operator has to be introduced. In the case of products of matrix elements such as were treated in Sec. 4, the corresponding product of matrix elements of unit operators may be summed over the states of intermediate configurations.[7] This gives a single matrix element of the product of all unit operators acting between the initial and final states. If this matrix element is evaluated by using Eq. (2.11), the result is the recoupling which would be obtained by performing the summation according to the procedure of Sec. 4. However, Racah and Stein recoupled the unit-operator product to bring together groups of operators acting in the same 1-particle space. The operators in each such group were then multiplied tensorially to form *effective* unit operators for each 1-particle space. The momenton method isolates the sequence of interactions within each 1-particle space on the lhs of the recoupling coefficient in Eq. (4.2). The effective unit operators can then be obtained by recoupling separately within each 1-particle space. This recoupling does not disturb the coupling schemes of N-particle states, and so it is not necessary to specify particular coupling schemes. The derivation of effective operators by the momenton method is illustrated by an example. In the perturbation of terms of the electronic configuration l^N by the configuration $l^{N-2}l'l''$, the following product of unsymmetrized matrix elements of Racah tensors occurs[6,7]:

$$\sum_{\alpha'}(l^{N-2},l^2|\,\mathbf{C}_i^{[k]}\cdot\mathbf{C}_j^{[k]}\,|l^{N-2}l'l'',\alpha')$$
$$\times(l^{N-2}l'l'',\alpha'|\,\mathbf{C}_i^{[k']}\cdot\mathbf{C}_j^{[k']}\,|l^{N-2},l^2).\quad(5.2)$$

Summation over α' may be made according to the method of Sec. 4 to give the result

$$(l\| C^{[k]} \|l')(l\| C^{[k]} \|l'')(l'\| C^{[k']} \|l)$$
$$\times (l''\| C^{[k']} \|l)[l]^{-1}([l'][l''])^{-\frac{1}{2}}$$
$$\times (((lk')l'k)l, ((lk')l''k)l \,|\, (ll)\bar{L}, (k'k')0, (kk)0)^{(\bar{L})}. \tag{5.3}$$

For simplicity in this case a coupling scheme $(l^2)\bar{L}$ has been specified, the configuration l^{N-2} being treated as "spectator" electrons. The recoupling coefficient is expanded

$$(((lk')l'k)l, ((lk')l''k)l \,|\, (ll)\bar{L}, (k'k')0, (kk)0)^{(\bar{L})}$$
$$= \sum_t ((lk')l', k \,|\, l, (k'k)t)^{(l)}((lk')l'', k \,|\, l, (k'k)t)^{(l)}$$
$$\times ((lt)l, (lt)l \,|\, (ll)\bar{L}, (tt)0)^{(\bar{L})}$$
$$= \sum_t [t]([l]^2[l'][l''])^{\frac{1}{2}} \begin{Bmatrix} k & k' & t \\ l & l & l' \end{Bmatrix} \begin{Bmatrix} k & k' & t \\ l & l & l'' \end{Bmatrix}$$
$$\times ((l^2)\bar{L}|\, \mathbf{c}_i^{[t]} \cdot \mathbf{c}_j^{[t]} \,|(l^2)\bar{L}), \tag{5.4}$$

where $(l\| c^{[t]} \,| j) = \delta(jl)$ and use has been made of the result (2.11). Finally, using (5.4), we have

$$\sum_{\alpha'} (l^N|\, \mathbf{C}_i^{[k]} \cdot \mathbf{C}_j^{[k]} \,|l^{N-2}l'l'', \alpha')$$
$$\times (l^{N-2}l'l'', \alpha'|\, \mathbf{C}_i^{[k']} \cdot \mathbf{C}_j^{[k']} \,|l^N)$$
$$= (l\| C^{[k]} \|l')(l\| C^{[k]} \|l'')(l'\| C^{[k']} \|l)(l''\| C^{[k']} \|l)$$
$$\times \sum_t [t] \begin{Bmatrix} k & k' & t \\ l & l & l' \end{Bmatrix} \begin{Bmatrix} k & k' & t \\ l & l & l'' \end{Bmatrix} (l^N|\, \mathbf{c}_i^{[t]} \cdot \mathbf{c}_j^{[t]} \,|l^N), \tag{5.5}$$

as obtained by Racah and Stein.

ACKNOWLEDGMENT

I would like to thank Professor U. Fano for helpful discussions.

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] G. Racah, Phys. Rev. **62**, 438 (1942); **63**, 367 (1943).
[2] U. Fano, F. Prats, and Z. Goldschmidt, Phys. Rev. **129**, 2643 (1963).
[3] U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic Press, Inc., New York, 1959).
[4] U. Fano, Phys. Rev. **140**, A64 (1965).
[5] A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, Jerusalem, 1962).
[6] K. Rajnak and B. G. Wybourne, Phys. Rev. **132**, 280 (1963).
[7] G. Racah and J. Stein, Phys. Rev. **156**, 58 (1967).
[8] B. G. Wybourne, J. Chem. Phys. **48**, 2596 (1968).

[9] For typographical convenience the notation $\delta(jj')$ is used for the Kronecker δ symbol.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

$O(5)$ Harmonics and Abnormal Solutions in the Bethe–Salpeter Equation

DEBABRATA BASU AND S. N. BISWAS*
Centre for Advanced Study in Physics, Department of Physics and Astrophysics, University of Delhi Delhi, India

(Received 23 May 1969)

Exact solutions of the covariant Bethe–Salpeter equation in the ladder approximation for two scalar particles bound by a massless particle have been obtained for all energies. By using Fock's stereographic projection, the Bethe–Salpeter equation is transformed on to the surface of a 5-dimensional Euclidean sphere and the solutions are then expressed as a series in $O(5)$ harmonics. The normalization condition has been imposed by requiring that the expectation value of appropriate components of the energy–momentum tensor with respect to the bound states is the total energy of the system; it is found that states corresponding to certain values of the quantum numbers do not satisfy the normalization requirement. These are the so-called abnormal asolutions.

1. INTRODUCTION

The study of the various properties of the covariant Bethe–Salpeter wave equation[1] has engaged the attention of many authors for a long time. Although the formulation of the BS equation has solved the problem of constructing a fully relativistic wave equation for a 2-body system in the framework of quantum mechanics, the appearance of many unfamiliar features in the equation has raised doubts in the minds of several authors. The presence of singularities in the interaction kernel led one to wonder whether any eigenvalue solution to the equation existed at all. This difficulty was resolved by Wick,[2] who transformed the equation by rotating the path

Summation over α' may be made according to the method of Sec. 4 to give the result

$$(l\| C^{[k]} \|l')(l\| C^{[k]} \|l'')(l'\| C^{[k']} \|l)$$
$$\times (l''\| C^{[k']} \|l)[l]^{-1}([l'][l''])^{-\frac{1}{2}}$$
$$\times (((lk')l'k)l, ((lk')l''k)l \,|\, (ll)\bar{L}, (k'k')0, (kk)0)^{(\bar{L})}. \tag{5.3}$$

For simplicity in this case a coupling scheme $(l^2)\bar{L}$ has been specified, the configuration l^{N-2} being treated as "spectator" electrons. The recoupling coefficient is expanded

$$(((lk')l'k)l, ((lk')l''k)l \,|\, (ll)\bar{L}, (k'k')0, (kk)0)^{(\bar{L})}$$
$$= \sum_t ((lk')l', k \,|\, l, (k'k)t)^{(l)}((lk')l'', k \,|\, l, (k'k)t)^{(l)}$$
$$\times ((lt)l, (lt)l \,|\, (ll)\bar{L}, (tt)0)^{(\bar{L})}$$
$$= \sum_t [t]([l]^2[l'][l''])^{\frac{1}{2}} \begin{Bmatrix} k & k' & t \\ l & l & l' \end{Bmatrix} \begin{Bmatrix} k & k' & t \\ l & l & l'' \end{Bmatrix}$$
$$\times ((l^2)\bar{L}|\, \mathbf{c}_i^{[t]} \cdot \mathbf{c}_j^{[t]} \,|(l^2)\bar{L}), \tag{5.4}$$

where $(l\| c^{[t]} \,| j) = \delta(jl)$ and use has been made of the result (2.11). Finally, using (5.4), we have

$$\sum_{\alpha'} (l^N|\, \mathbf{C}_i^{[k]} \cdot \mathbf{C}_j^{[k]} \,|l^{N-2}l'l'', \alpha')$$
$$\times (l^{N-2}l'l'', \alpha'|\, \mathbf{C}_i^{[k']} \cdot \mathbf{C}_j^{[k']} \,|l^N)$$
$$= (l\| C^{[k]} \|l')(l\| C^{[k]} \|l'')(l'\| C^{[k']} \|l)(l''\| C^{[k']} \|l)$$
$$\times \sum_t [t] \begin{Bmatrix} k & k' & t \\ l & l & l' \end{Bmatrix} \begin{Bmatrix} k & k' & t \\ l & l & l'' \end{Bmatrix} (l^N|\, \mathbf{c}_i^{[t]} \cdot \mathbf{c}_j^{[t]} \,|l^N), \tag{5.5}$$

as obtained by Racah and Stein.

ACKNOWLEDGMENT

I would like to thank Professor U. Fano for helpful discussions.

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] G. Racah, Phys. Rev. **62**, 438 (1942); **63**, 367 (1943).
[2] U. Fano, F. Prats, and Z. Goldschmidt, Phys. Rev. **129**, 2643 (1963).
[3] U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic Press, Inc., New York, 1959).
[4] U. Fano, Phys. Rev. **140**, A64 (1965).
[5] A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, Jerusalem, 1962).
[6] K. Rajnak and B. G. Wybourne, Phys. Rev. **132**, 280 (1963).
[7] G. Racah and J. Stein, Phys. Rev. **156**, 58 (1967).
[8] B. G. Wybourne, J. Chem. Phys. **48**, 2596 (1968).

[9] For typographical convenience the notation $\delta(jj')$ is used for the Kronecker δ symbol.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

$O(5)$ Harmonics and Abnormal Solutions in the Bethe–Salpeter Equation

DEBABRATA BASU AND S. N. BISWAS*
Centre for Advanced Study in Physics, Department of Physics and Astrophysics, University of Delhi Delhi, India

(Received 23 May 1969)

Exact solutions of the covariant Bethe–Salpeter equation in the ladder approximation for two scalar particles bound by a massless particle have been obtained for all energies. By using Fock's stereographic projection, the Bethe–Salpeter equation is transformed on to the surface of a 5-dimensional Euclidean sphere and the solutions are then expressed as a series in $O(5)$ harmonics. The normalization condition has been imposed by requiring that the expectation value of appropriate components of the energy–momentum tensor with respect to the bound states is the total energy of the system; it is found that states corresponding to certain values of the quantum numbers do not satisfy the normalization requirement. These are the so-called abnormal asolutions.

1. INTRODUCTION

The study of the various properties of the covariant Bethe–Salpeter wave equation[1] has engaged the attention of many authors for a long time. Although the formulation of the BS equation has solved the problem of constructing a fully relativistic wave equation for a 2-body system in the framework of quantum mechanics, the appearance of many unfamiliar features in the equation has raised doubts in the minds of several authors. The presence of singularities in the interaction kernel led one to wonder whether any eigenvalue solution to the equation existed at all. This difficulty was resolved by Wick,[2] who transformed the equation by rotating the path

of integration in the complex energy plane. The eigenvalue problem, in terms of the transformed equation, can be solved and the existence of the solutions can be shown to follow under fairly general assumptions that the wavefunction $\psi(p)$ is finite at small p and $p^2\psi(p)$ is finite at large p. It is also known that the BS equation may have too many solutions, a feature connected with the fact that the order of the equation is higher than that of the corresponding nonrelativistic equation. It has been suggested that the covariant bound-state equation must be supplemented by a normalization condition which, as for the nonrelativistic Schrödinger equation, serves to distinguish the physical solutions by their normalizability. Moreover, a knowledge of normalization integrals is always essential to the quantum mechanical probabilistic interpretation of the 2-particle wavefunctions given by the theory.

In the present paper, we investigate this particular aspect of the solutions of the BS equation and show the absence of the positive-definite norms for a class of wavefunctions satisfying the BS equation for two scalar particles bound by a massless scalar particle. These are the so-called "abnormal" solutions of the BS wave equation. The existence of such solutions in this particular case was also noted by Nakanishi,[3] who considered the solutions when the total energy of the 2-particle system is zero. The possibility that such solutions might exist for nonvanishing energy was also indicated by him. In particular, the presence of "abnormal" solution which do not possess a nonrelativistic limit was first noted by Wick[2] and Cutkosky.[4]

We first show that exact solutions of this particular BS equation can be obtained as a convergent series of $O(5)$ harmonics for arbitrary nonvanishing energies. This possibility stems from the fact that the BS equation at zero energy admits an exact $O(5)$ symmetry and thus has $O(5)$ polynomials as exact eigenfunctions. The existence of such a symmetry was pointed out by Cutkosky[4] and later by Salam *et al.*,[5] by applying the stereographic projection techniques of Fock.[6] The existence of this symmetry character in the BS equation has also been noted by one of us[7] to follow from the observation that the equation can be separated in a suitable system of coordinates. An extensive study of this dynamical symmetry in BS equation has recently been made by Kyriakopoulos.[8] In the present paper we adopt the method due to Fock[6] and Levy.[9] Our method of solution is a direct generalization of that used in our earlier paper[10] in solving the pion–nucleon BS equation in the instantaneous interaction approximation.

Normalizability criteria for the BS wavefunctions have then been used to eliminate the abnormal solutions. For this, we follow Nishijima and Singh[11] and adopt for normalization the requirement that the expectation value of the energy–momentum tensor operator with respect to the bound states be the total energy of the system. The equivalence of this criteria with other methods of normalization[12] of covariant wavefunctions has also been pointed out in Ref. 11.

In the next section, we briefly discuss the structure of our equation in Fock's 5-dimensional pseudosphere and obtain the solution in the form $\sum a_k P^{(3)}_{N+k,n}$ where $P^{(3)}_{N+k,n}$ are simply related to Gegenbauer polynomials and the a_k satisfy a difference equation of the form $La_{k+2} + M_k a_k + N_k a_{k-2} = 0$. We then solve this difference equation exactly and obtain the energy eigenvalue solution. In Sec. 3 we impose the normalization condition and point out the abnormal solutions of our equation.

2. THE BS EQUATION AND ITS SOLUTION IN $O(5)$ HARMONICS

We consider the following BS equation in the ladder approximation for two equal-mass scalar particles bound by a massless scalar meson:

$$[(p^2 + 1 - \epsilon^2)^2 + 4\epsilon^2 p_4^2]\psi(p) = \frac{\lambda}{\pi^2}\int d^4p' \frac{\psi(p')}{(p - p')^2}. \tag{1}$$

Here $\psi(p)$ is the Fourier transform of $\psi(x - y)$, the configuration-space wavefunction in relative coordinates. In obtaining Eq. (1) we have performed the Wick rotation in the relative time variable and further specialized to the rest frame $\frac{1}{2}P = (0, 0, 0, i\epsilon)$ where P denotes the total 4-momentum of the system. It should be mentioned, however, that because the mass of the exchanged scalar particle is zero ($\mu = 0$), the resulting BS equation admits of an extra symmetry. For example, when $\mu = 0$ and $\epsilon = 0$, the BS equation has exact $O(5)$ symmetry. The purpose of the present section is to show that, although for $\mu = 0$ and $\epsilon \neq 0$ the BS equation does not possess exact $O(5)$ symmetry, it still admits of an exact solution in terms of an infinite series in $O(5)$ harmonics. In this case, however, the equation has an exact $O(4)$ symmetry.

We first rewrite Eq. (1) in terms of a new variable q defined by

$$p_\mu = (1 - \epsilon^2)^{\frac{1}{2}} q_\mu \tag{2}$$

and further use

$$\begin{aligned} \rho_k &= 2q_k/(1 + q^2 + 2q_4), \\ \rho_4 &= (1 - q^2)/(1 + q^2 + 2q_4), \\ k &= 1, 2, 3, \end{aligned} \tag{3a}$$

so that

$$\rho^2 = (1 + q^2 - 2q_4)/(1 + q^2 + 2q_4). \quad (3b)$$

From Eqs. (3) it easily follows that

$$q_k = 2\rho_k/(1 + \rho^2 + 2\rho_4),$$
$$q_4 = (1 - \rho^2)/(1 + \rho^2 + 2\rho_4), \quad (3c)$$

and

$$q^2 = (1 + \rho^2 - 2\rho_4)/(1 + \rho^2 + 2\rho_4). \quad (3d)$$

The resulting integral equation is

$$[(1 + \rho^2)^2 - 4\epsilon^2\rho^2]\varphi(\rho) = \frac{\lambda}{\pi^2}\int \frac{\varphi(\rho')}{(\rho - \rho')^2}\, d^4\rho', \quad (4)$$

where φ is related to ψ by

$$\varphi = [2/(1 + \rho^2 + 2\rho_4)]^3\psi. \quad (5)$$

Equation (4) is manifestly $O(4)$ symmetric so that any solution can be written in the form

$$\varphi(\rho) = g(\rho)Y_{nlm}(\psi, \theta, \varphi) \quad (6)$$

where Y_{nlm} is a 4-dimensional spherical harmonic and is given by

$$Y_{nlm} = (N_{nl})^{\frac{1}{2}} \sin^l \psi C_{n-l}^{l+1}(\cos\psi) Y_l^m(\theta, \varphi),$$

N_{nl} being the normalization constant.

Separating the angular variables, we easily find that $g(\rho)$ satisfies the following 1-dimensional equation[13]:

$$[(1 + \rho^2)^2 - 4\epsilon^2\rho^2]g(\rho) = \frac{\lambda}{\pi^2}\int_0^\infty d\rho'\rho'^3 K_n(\rho, \rho')g(\rho'), \quad (7)$$

where

$$K_n(\rho, \rho') = 4\pi\int_{-1}^{1} \frac{P_{n,0}^{(2)}(x)(1 - x^2)^{\frac{1}{2}}}{\rho^2 + \rho'^2 - 2\rho\rho' x}\, dx.$$

We transform Eq. (7) onto Fock's 5-dimensional hyperspace through the substitution

$$\rho = \tan\tfrac{1}{2}\chi, \quad \rho' = \tan\tfrac{1}{2}\chi'. \quad (8)$$

Finally, we define

$$f(\chi) = \sec^6\tfrac{1}{2}\chi g(\tan\tfrac{1}{2}\chi). \quad (9)$$

Equation (7) can then be written as

$$(1 - \epsilon^2\sin^2\chi)f(\chi) = \frac{\lambda}{8\pi^2}\int \frac{f(\chi')P_{n,0}^{(2)}(\cos\psi')}{(1 - \cos\Theta)}\, d\Omega', \quad (10)$$

where $d\Omega'$ is the 5-dimensional solid angle and Θ is the angle between two 5-dimensional unit vectors of polar angles $(\chi, 0, 0, 0)$ and $(\chi', \psi', \theta', \varphi')$. We now attempt a solution in the form[14]

$$f(\chi) = \sum_{k=0}^{\infty} a_k P_{N+k,n}^{(3)}(\cos\chi). \quad (11)$$

Using the relation

$$\int d\Omega' \frac{P_{N,n}^{(3)}(\cos\chi')P_{n,0}^{(2)}(\cos\psi')}{(1 - \cos\Theta)} = \lambda_N P_{N,n}^{(3)}(\cos\chi), \quad (12)$$

where

$$\lambda_N = 8\pi^2/(N + 1)(N + 2), \quad (12')$$

we obtain (11) as a solution of (10), provided the coefficients a_k satisfy

$$L_k a_{k+2} + M_k a_k + N_k a_{k-2} = 0, \quad (13)$$

where

$$L_k = \epsilon^2 \frac{(N + k + 1)(N + k + 2)(N + k + n + 3)(N + k + n + 4)}{(2N + 2k + 5)(2N + 2k + 7)(N + k + 3)(N + k + 4)},$$

$$M_k = 1 - \frac{\lambda}{(N + k + 1)(N + k + 2)} + \epsilon^2\left(\frac{(N + k - n)(N + k + n + 2)}{(2N + 2k + 1)(2N + 2k + 3)} + \frac{(N + k - n + 1)(N + k + n + 3)}{(2N + 2k + 3)(2N + 2k + 5)} - 1\right),$$

$$N_k = \epsilon^2 \frac{(N + k + 1)(N + k + 2)(N + k - n - 1)(N + k - n)}{2(N + 2k + 1)(2N + 2k - 1)(N + k - 1)(N + k)}.$$

It is seen that the difference equation determines separately the even and the odd coefficients a_k. Using Perron's theorem,[15] we easily find the solutions for our energy-eigenvalue problem (see Appendix A), through the equation

$$M_0/L_0 = \cfrac{N_2/L_2}{M_2/L_2 - \cfrac{N_4/L_4}{\ddots}}, \quad (14)$$

and a similar equation is obtained which involves only the odd indexed L_k, etc. Both the equations, however, lead to the same eigenvalue equation, which, in a simplified way, can be expressed as

$$\lambda(\epsilon) = \sum_{m=0}^{\infty} X_m \epsilon^{2m}. \quad (15)$$

It is interesting to note that, if we put $\epsilon = 0$ in Eq. (14) or (15), we immediately recover the well-known eigenvalue condition[16]

$$\lambda(0) = (N + 1)(N + 2), \quad (16)$$

which was derived earlier in Refs. 4 and 7.

Before concluding this section, we add a few remarks regarding the convergence of our series (11). Using Eq. (A6), we find that the convergence criteria

$$\lim_{k\to\infty}\left|\frac{a_k}{a_{k-1}}\right|<1$$

is rigorously satisfied whenever $0<\epsilon<1$. The case $\epsilon=1$ needs a careful study. We must mention that, at this point, our series expansion fails, and a closer look at the original equation [see Eq. (7)] reveals that the integral equation for $g(\rho)$ can be easily solved by transforming it into a differential equation with proper boundary condition, whose solution is easily obtained in terms of a hypergeometric function. The eigenvalue equation is solved and it agrees with that given by Wick[2] and Cutkosky.[4] A short discussion is given in Appendix B. For a complete solution of the equation, we should also find a relation between odd and even a_k. This is achieved through the use of the normalization condition.

3. NORMALIZATION OF THE WAVEFUNCTION

The normalization condition is imposed by requiring that the total energy of the system be given by the expectation value of the energy–momentum tensor with respect to the given bound states. Thus, we have

$$\frac{1}{V}\langle P,\alpha|\int T_{4\nu}(x)\,d^3x\,|P,\beta\rangle=\frac{i}{V}P_\nu\delta_{\alpha\beta},$$

where $T_{\mu\nu}$ is the energy–momentum tensor operator and P_ν is the energy–momentum 4-vector. $|P,\beta\rangle$ denotes a bound state characterized by total energy–momentum P and another quantum number β.

Following Nishijima and Singh,[11] we note that the left-hand side of above can also be written as

$$\frac{iP_\nu}{2P_0V}\int d^4p\,d^4q\,\bar\psi_\alpha(p)\frac{\partial D(p,q,P)}{\partial P_4}\psi_\beta(q),$$

so that the normalization condition becomes

$$\int d^4p\,d^4q\,\bar\psi_\alpha(p)\frac{\partial D(p,q,P)}{\partial P_4}\psi_\beta(q)=-2iP_4\delta_{\alpha\beta}. \quad (17)$$

In Eq. (17), ψ_β is the bound-state wavefunction which satisfies the BS equation (1) and the conjugate wavefunction $\bar\psi_\alpha(p)$ is related to $\psi_\alpha(p)$ as follows[17]:

$$\bar\psi_\alpha(\mathbf{p},ip_4)=-\psi_\alpha^*(\mathbf{p},-ip_4).$$

Further, the D function is defined through the equation

$$\int d^4kD(k,p,P)K(k,p',P)=\delta(p-p').$$

The reciprocal of the kernel function D can be easily constructed, using the BS equation in the ladder approximation, and is given by

$$D(k,p,P)=-[(\tfrac{1}{2}P+k)^2+1][(\tfrac{1}{2}P-k)^2+1]\times\delta(p-k)-(i\lambda/\pi^2)(p-k)^{-2}. \quad (18)$$

To facilitate the evaluation of the normalization integral in (17), we use the transformation variables given in Sec. 2 [Eqs. (3)], and through the use of our solutions of the BS equation on the 5-dimensional pseudosphere, the normalization integral now takes the following simple form:

$$\int_0^\pi\sin^3\chi\,d\chi f^*(\pi-\chi)\sin^2\chi f(\chi)=1.$$

In the above equation, some kinematic factors have been absorbed in the wavefunctions $f(\chi)$.

We therefore conclude that only those solutions $f(\chi)$ are admissible for which

$$\int_0^\pi\sin^3\chi\,d\chi f^*(\pi-\chi)\sin^2\chi f(\chi)$$

is positive definite. Explicit evaluation of this integral using Eqs. (11) and (A5) shows that this is ensured if

$$|a_0|^2S_N\left(1-\frac{|a_1|^2}{|a_0|^2}\frac{S_{N+1}}{S_N}\right)(-1)^\kappa \quad (19)$$

is positive definite. In (19), S_N is a positive-definite quantity and $\kappa=N-n=0,1,2,\cdots$.

It is evident that, for states characterized by certain quantum numbers N and n, expression (19) cannot be made positive definite. Thus the wavefunctions $f(\chi)$ for which expression (19) is not positive definite are not admissible, and they constitute abnormal solutions. The possibility of the existence of such solutions was previously pointed out by Nakanishi[3] and also by Wick[2] and Cutkosky[4] from an approximate treatment of the BS equation at nonvanishing energies, but with zero mass for the exchanged particle. Our exact solution of the BS equation confirms this conclusion. In the next section, we give a brief summary of the nature of various types of solutions of the BS equation obtained so far.

4. CONCLUSION

The various solutions of the BS equation considered here can be classified as follows. When the total energy $E=0$ and the exchanged particle mass $\mu=0$, the equation admits an exact $O(5)$ symmetry; further, imposition of the normalization requirement shows that there are abnormal solutions in addition to the admissible ones. In the case $E=0$ and $\mu\neq0$, the

equation is $O(4)$ symmetric. It is interesting to point out that when $E = 0$, the BS equation becomes invariant under Lorentz transformations involving the relative 4-momentum p_μ, and the 4-dimensional angular momentum L is then a good quantum number. For a fixed value of L the ordinary angular momentum l is given by the relation $l = L - \kappa$, $\kappa = 0, 1, 2, \cdots$; when L as well as l are extended to Reggeized values, one obtains the so-called daughter trajectories.[18] This particular case of the BS equation has been discussed by various authors in connection with the existence of Lorentz poles.[19] When $E \neq 0$ and $\mu = 0$, the BS equation is $O(4)$ symmetric; however, the solutions can be obtained as a series of $O(5)$ polynomials. Imposition of the normalization condition again reveals that in the whole continuum of solutions there exist abnormal solutions in addition to the admissible ones. We have investigated this case here in detail. The point $\epsilon = \frac{1}{2}E \to 1$ deserves special mention; here our method of expansion of the wavefunction in $O(5)$ polynomials fails. Nevertheless, an exact solution exists and the eigenvalue problem is determinate. The case of $E \neq 0$ and $\mu \neq 0$ poses great difficulties. An attempt to solve the equation for this particular case was made by Okubo and Feldman,[20] by using an integral-representation method originally due to Wick. An exact analytical evaluation of the spectral function is extremely involved and no definite conclusions regarding the nature of the solutions can be made.

ACKNOWLEDGMENT

We wish to thank Dr. K. Datta for many interesting discussions.

APPENDIX A: DERIVATION OF THE EIGENVALUE EQUATION

The characteristic equation of the difference equation [Eq. (13)] possesses only two roots of distinct moduli; thus, according to Perron's theorem,[15] there is a fundamental system of solutions which falls into two classes. Denoting this fundamental system by $a_k^{(i)}$, $i = 1, 2$, we have

$$a_{k+2}^{(i)} + p_k a_k^{(i)} + q_k a_{k-2}^{(i)} = 0, \tag{A1}$$

where

$$p_k = M_k/L_k\,, \quad q_k = N_k/L_k\,.$$

Let us define

$$z_S = \frac{(k_S - k_{S+2})(k_{S+4} - k_{S+6})}{(k_S - k_{S+4})(k_{S+2} - k_{S+6})},$$
$$v_S = \frac{(k_{n+6} - k_S)(k_{S+2} - k_{S+4})}{(k_{n+6} - k_{S+2})(k_S - k_{S+4})}, \tag{A2}$$
$$S = 1, 2 \cdots n, \quad v_{n+2} = 1.$$

Then

$$v_S = 1 - z_S/v_{S+2}. \tag{A3}$$

We choose now

$$k_S = a_{k+S-6}^{(1)}/a_{k+S-6}^{(2)}.$$

It is then easy to show that

$$\frac{a_{k-1}^{(2)} - (a_{k+n}^{(2)}/a_{k+n}^{(1)})a_{k-1}^{(1)}}{a_{k-3}^{(2)} - (a_{k+n}^{(2)}/a_{k+n}^{(1)})a_{k-3}^{(1)}} = \cfrac{-q_{k-1}}{p_{k-1} - \cfrac{q_{k+1}}{p_{k+1} - \cfrac{q_{k+3}}{\ddots \cfrac{}{p_{k+n-4} - \cfrac{q_{k+n-2}}{p_{k+n-2}}}}}}. \tag{A4}$$

The right-hand member of this identity is obviously independent of the fundamental system chosen. Let us now choose, following Perron, the fundamental system in such a way that

$$\limsup_{n\to\infty} |a_{k+n}^{(1)}| = \alpha^{k+n},$$
$$\limsup_{n\to\infty} |a_{k+n}^{(2)}| = \beta^{k+n}.$$

Here α and β are the distinct moduli of the roots of the characteristic equation such that $\alpha > \beta$. Hence taking limit as $n \to \infty$ on both sides of Eq. (A4) and noting that

$$\lim_{n\to\infty} (a_{k+n}^{(2)} \mid a_{k+n}^{(1)}) = 0,$$

we obtain a solution of the difference equation in terms of an infinite continued fraction

$$\frac{a_{k+2}}{a_k} = -\cfrac{q_{k+2}}{p_{k+2} - \cfrac{q_{k+4}}{\ddots}}. \tag{A5}$$

Only this particular solution is acceptable, as it satisfies the convergence requirement of the solution of $f(\chi)$ given in terms of an infinite series in $O(5)$ harmonics; this is obvious when we note that this solution corresponds to

$$\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = \beta,$$
$$\beta < 1, \quad \alpha > 1, \quad 0 < \epsilon < 1. \tag{A6}$$

The eigenvalue Eq. (14) immediately follows as a compatibility requirement for the boundary conditions $a_{-1} = a_{-2} = \cdots = 0$.

APPENDIX B: THE SOLUTION OF EQ. (7) FOR $\epsilon = 1$

Applying the operator

$$\frac{1}{\rho^3}\frac{d}{d\rho}\left(\rho^3\frac{d}{d\rho}\right)$$

and setting $\epsilon = 1$, the integral Eq. (7) reduces to the following differential equation:

$$\frac{1}{\rho^3}\frac{d}{d\rho}\left(\rho^3\frac{du}{d\rho}\right) = \frac{n(n+2)u(\rho)}{\rho^2} - 4\lambda\frac{u(\rho)}{(1-\rho^2)^2}, \quad \text{(B1)}$$

where

$$u(\rho) = (1-\rho^2)^2 g(\rho)$$

and satisfies the boundary conditions that $u(\rho)$ is finite at small ρ and $\rho^2 u(\rho)$ is finite at large ρ.

A simple transformation

$$x = (1-\rho^2)/(1+\rho^2),$$
$$u(\rho) = (1-x^2)^{-\frac{1}{2}(n+2)}(1+x)G(x)$$

immediately transforms the equation into that obtained previously by Cutkosky, namely,

$$(1-x^2)G''(x) + 2nxG'(x) - n(n+1)G(x) + (\lambda/x^2)G(x) = 0. \quad \text{(B2)}$$

The solution of this equation which satisfies the boundary conditions at $x = \pm 1$ is given by

$$G(x) = (1-x^2)^{n+1}x^{\frac{1}{2}+\nu} \times {}_2F_1\{\tfrac{1}{4}(2n+5+2\nu), \tfrac{1}{4}(2n+3+2\nu); n+2; (1-x^2)\},$$
$$\nu = (\tfrac{1}{4}-\lambda)^{\frac{1}{2}}.$$

Following Wick, we find that the eigenvalue of λ for $\epsilon = 1$ is given by

$$\lambda(\epsilon = 1) = \tfrac{1}{4}.$$

* Present address: International Centre for Theoretical Physics, Trieste, Italy.

[1] M. Gell-Mann and F. E. Low, Phys. Rev. **84**, 350 (1951); E. E. Salpeter and H. A. Bethe, *ibid.* **84**, 1232 (1951).
[2] G.-C. Wick, Phys. Rev. **96**, 1124 (1954).
[3] N. Nakanishi, Phys. Rev. **138**, B1182 (1965).
[4] R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954).
[5] R. Delbourgo, Abdus Salam, and J. Strathdee, ICTP Preprint No. IC/66/60, 1966. See also A. O. Barut, P. Budini, and C. Fronsdal, ICTP Preprint No. IC/65/34, 1965.
[6] V. Fock, Z. Physik **98**, 145 (1935).
[7] S. N. Biswas, J. Math. Phys. **8**, 1109 (1967).
[8] E. Kyriakopoulos, Phys. Rev. **174**, 1846 (1968). For a general discussion on dynamical symmetries, see also N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965).
[9] M. Levy, Proc. Roy. Soc. (London) **A204**, 145 (1950).
[10] D. Basu and S. N. Biswas, J. Math. Phys. **10**, 2104 (1969).
[11] K. Nishijima and A. H. Singh, Phys. Rev. **162**, 1740 (1967).
[12] K. Nishijima, Progr. Theoret. Phys. (Kyoto) **10**, 549 (1953); **12**, 279 (1954); **13**, 305 (1955); A. Klein and C. Zemach, Phys. Rev. **108**, 126 (1957); S. Mandelstam, Proc. Roy. Soc. (London) **A233**, 248 (1955); G. R. Allcock, Phys. Rev. **104**, 1799 (1956); G. R. Allcock and D. Hooton, Nuovo Cimento **8**, 590 (1958); R. E. Cutkosky and M. Leon, Phys. Rev. **135**, B1445 (1964); I. Sato, J. Math. Phys. **4**, 24 (1963); D. Lurie, A. J. Macfarlane, and Y. Takahasi, Phys. Rev. **140**, B1091 (1965). See also E. Predazzi, Nuovo Cimento **40**, 913 (1965).
[13] In deriving this equation we have used Hecke's theorem. The hyperspherical harmonics appearing in this and in the subsequent equations are defined in accordance with our earlier paper (Ref. 10). It should be pointed out that, by modifying the transformation variables given by Eq. (3), the BS equation can also be reduced to a 1-dimensional integral equation analogous to (7), even when the masses of the external particles are unequal.
[14] The functions $P^{(3)}_{N,n}$ are simply related to the Gegenbauer polynomials

$$P^{(3)}_{N,n}(\cos\chi) = 2/[(N+1)(N+2)]^{-1}\sin^n\chi C^{n+\frac{3}{2}}_{N-n}(\cos\chi).$$

[15] L. M. Milne-Thomson, *Calculus of Finite Differences* (The Macmillan Co., London, 1933), Chap. 17. For a numerical evaluation of this type of continued fractions, see L. H. D. Reeves, Ph.D. thesis, Adelaide University, 1962.
[16] It can be easily shown that $\lambda(\epsilon)$ has a maximum at $\epsilon = 0$ which corresponds to maximal binding.
[17] In obtaining this relation we have made use of Mandelstam's prescription (Ref. 12). This relation also follows easily if we assume the spectral representation of the BS bound state due to Wick. For details see M. Ciafaloni and P. Menotti, Phys. Rev. **140**, B929 (1965). See also K. Nishijima in Ref. 12.
[18] D. Z. Freedman and J. M. Wang, Phys. Rev. **153**, 1596 (1967). See also the other references cited in our earlier paper (Ref. 10).
[19] See, for example, N. P. Chang and R. P. Saxena, Phys. Rev. **176**, 2101 (1968).
[20] S. Okubo and D. Feldman, Phys. Rev. **117**, 279 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Invariant Representation of All Analytic Petrov Type III Solutions to the Einstein Equations*

CARL H. BRANS
Department of Physics, Loyola University, New Orleans, Louisiana 70118

(Received 27 October 1969)

The results of the application of previously developed techniques to the analysis of the Petrov type III solutions to the vacuum Einstein equations are presented. The procedure involves the computer aided analysis of the Einstein–Petrov equations to the extent that the functions uniquely and invariantly generating all local analytic solutions are determined. For the case of type III it is shown that, relative to a given fixed point in the manifold, all local analytic solutions are uniquely and invariantly determined by six arbitrary analytic functions of one variable and six others of two variables. These functions, called generating functions, thus provide a representation of all such solutions and may be used for the study of the structure of the family of Einstein empty space metrics.

I. INTRODUCTION

It is an often noted paradox that, while Einstein's general relativity is a theory of remarkable elegance and logical simplicity, its gravitational field equations, even in empty space, are extremely complicated when explicitly spelled out. Consequently, only a comparatively few exact solutions are known, all of them involving the assumption of a fairly high degree of symmetry. Of course, linear and higher-order approximations have been thoroughly explored and exploited, for example, to derive equations of motions. Further, the detailed, local questions concerning the existence and uniqueness of solutions have been investigated thoroughly by Fourès-Bruhat.[1] A little later an invariant classification of solutions by means of the algebraic structure of the Ricci-null–Riemann tensor was worked out, with the name of Petrov associated with the three discrete types.[2]

This paper reports on another approach,[3] somewhat similar to that of Fourès-Bruhat. Here, however, the tangent frame of forms and local canonical coordinate system are completely determined, up to possible intrinsic symmetries, by the geometry, through the Petrov frame and its higher-order generalizations if needed. In addition, once the equations have been fully reduced as far as differential consistency is concerned, it is pointed out that *the local, analytic solutions are completely and uniquely determined by the "generating functions" (GF) which may be arbitrarily assigned.* Hence it is proposed that the GF's may be thought of as providing a unique and invariant representation of all local, analytic solutions. In this manner, the family of local, analytic vacuum Einstein metrics may be studied and structures imposed on it invariantly by studying and manipulating the family of analytic GF's. The question of relaxing the conditions of analyticity will be discussed in a later paper, but it seems reasonable that physically significant solutions will be at least piecewise analytic.

II. THE EINSTEIN-PETROV EQUATIONS

Much of the modern mathematical work in the field of differential geometry has been associated with a shift of emphasis from local coordinate patches to the larger structure of the tangent vector bundle. This is obviously quite natural since the content of the geometry of the space may be expressed essentially in terms of a relationship between vectors at neighboring points along curves (connection approach) or in terms of an inner product on the tangent vector space at each point (metric approach). In either case, the operations involve a linear vector space, with a global structure, rather than an open coordinate patch in a manifold. Hence, it is possible at each point to choose a vector space basis adapted to the intrinsic geometry, without concern for an underlying coordinate system. In particular, the vectors constituting the basis do not have to comprise an integrable system, that is, they need not be vectors tangent to the coordinate lines of a coordinate system.[4] Thus, the tangent frame may be chosen to be a "standard" one relative to the metric inner product. In the case of general relativity this means that the four base vectors will have inner products described by the standard Minkowski metric array $\eta_{ab} = \text{diag}\,(1, 1, 1, -1)$. This, of course, still leaves the freedom of a transformation of the homogeneous Lorentz group. This freedom may possibly be further reduced by means of the higher-order (in the differential sense) structure of the geometry. In approaching this problem we use both the work of Cartan and Petrov. The work of Petrov classifies solutions to the vacuum Einstein equations in terms of the algebraic structure of the Riemann tensor at each point. In the event of degeneracy in the algebraic

structure of the Riemann tensor itself, the structure of its covariant derivatives may be used to determine a canonical frame of forms at each point as strongly as possible, up to the intrinsic symmetries of the geometry.

In this formalism let ω^a represent four Lorentz orthonormal forms, so that the line element is

$$ds^2 = \eta_{ab}\omega^a\omega^b. \tag{1}$$

The condition that the frame of forms represent a canonical Petrov frame, and thus that the metric satisfy the Einstein equations, is expressed in terms of the Cartan structure equations

$$d\omega^a = \omega^a \wedge \omega^a{}_b, \quad \omega_{ab} + \omega_{ba} = 0, \tag{2}$$

$$d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b = (\tfrac{1}{2})R^a{}_{bcd}\omega^c \wedge \omega^d, \tag{3}$$

where the array R_{abcd} represents one of the three Petrov arrays.[5] The only case completely analyzed so far is that of Petrov type III, so in this paper we assume that R_{abcd} is of this type. In terms of the notational convention used here, this means that all R_{abcd} are zero except for the following elements and those obtained from them by known Riemann tensor symmetries:

$$R_{1424} = R_{2412} = R_{3431} = R_{3231} = 1. \tag{4}$$

In the other Petrov types the Riemann tensor itself does not always uniquely determine the frame ω^a. If this is so, further equations of the form

$$R_{abcd:e} = R_{abcd|e} - R_{nbcd}\gamma^n{}_{ae} - R_{ancd}\gamma^n{}_{be} - R_{abnd}\gamma^n{}_{ce} - R_{abcn}\gamma^n{}_{de} \tag{5}$$

must be added. In these, the array $R_{abcd;e}$ would represent one of the canonical forms discussed in the previous paper.[3] However, for the present, we need only consider (2) and (3), which, when expressed in terms of the canonical connection components $\gamma^a{}_{bc}$ defined by $\omega^a{}_b = \gamma^a{}_{bc}\omega^c$, become

$$d\omega^a = \gamma^a{}_{bc}\omega^b \wedge \omega^c, \tag{6}$$

$$\gamma^a{}_{bc|d} - \gamma^a{}_{bd|c} + \gamma^a{}_{bn}\gamma^n{}_{dc} - \gamma^a{}_{bn}\gamma^n{}_{cd} + \gamma^a{}_{nd}\gamma^n{}_{bc} - \gamma^n{}_{nc}\gamma^n{}_{bd} = R^a{}_{bdc}. \tag{7}$$

Here the vertical slash represents form differentiation with respect to the ω^a:

$$df = f_{|a}\omega^a. \tag{8}$$

The intrinsic invariance of the procedure now becomes evident. The functions $\gamma^a{}_{bc}$ refer to the canonical frame ω^a and are scalar fields which uniquely determine the geometry. These may not be arbitrarily assigned, however, but must satisfy the differential equations (6).

In Petrov types I and II, the canonical form for R_{abcd} will also contain nonconstant scalar fields and these must also be included. In any event, the Einstein equations are reduced to set of algebraic partial differential equations for a number of dependent scalar functions. The solutions to the equations then correspond in a unique and invariant way to the geometry satisfying the Einstein empty space equations.

III. INVARIANT COORDINATES

In order to describe the scalar fields effectively as functions on the space–time manifold, it is still necessary to introduce local coordinates. If these coordinates are chosen arbitrarily, the invariance gained by use of the canonical frames would be lost. Hence, it is necessary to define these coordinates in an invariant manner.

The first thought might be to use the invariant scalar fields $\gamma^a{}_{bc}$ to determine these coordinates. For example, these fields may be ordered in some standard way, the first nonconstant one chosen as x^1, the next functionally independent one as x^2, and so on until the intrinsic asymmetries of the geometry are exhausted. It turns out, however, that this procedure provides a very unwieldy condition on the equations and is probably completely impractical.

Consequently, another method was chosen which is more closely adapted to the canonical directions provided by the ω^a and to the structure equations themselves. These coordinates may be thought of as following as closely as possible along the canonical Petrov directions in the neighborhood of a fixed point. More precisely, they may be defined by means of the following result.

Theorem: Given four independent forms ω^a, defined over a neighborhood U of a point P, there exists a coordinate system x^i in a neighborhood V of P for which $V \subset U$, $x^i(P) = 0$, and the following conditions hold:

$$\begin{aligned} &\omega^a{}_4 = \delta^a{}_4, \\ &\omega^a{}_3 = \delta^a{}_3, \quad \text{for} \quad x^4 = 0, \\ &\omega^a{}_2 = \delta^a{}_2, \quad \text{for} \quad x^4 = x^3 = 0, \\ &\omega^a{}_1 = \delta^a{}_1, \quad \text{for} \quad x^4 = x^3 = x^2 = 0. \end{aligned} \tag{9}$$

Proof: Let y^i be some coordinate system in a neighborhood W of P, with $W \subset U$, in which the forms ω^a have components $\bar{\omega}^a{}_j(y^i)$ such that

$$\omega^a = \bar{\omega}^a{}_j(y^i)\, dy^j. \tag{10}$$

Further let $\bar{e}^i{}_a$ be components of a dual frame of tangent vectors so that

$$\bar{e}^i{}_a\bar{\omega}^a{}_j = \delta^i{}_j. \tag{11}$$

Consider the four sets of ordinary differential equations defined by

$$\frac{dy^i}{dx^j} = \bar{e}^i{}_j(y^k). \tag{12}$$

It is known that there exists, for sufficiently small x^j, unique solutions to each set of equations with given initial data. Let the solution for y^i to jth equation which assumes the values u^i at $x^j = 0$ be denoted $F^i{}_j(x^j, u^k)$. Thus,

$$\frac{\partial F^i{}_j}{\partial x^j} = \bar{e}^i{}_j(F^k{}_j), \quad F^i{}_j(0, u^k) = u^i. \tag{13}$$

The required coordinate system y^i is related to the x^j by

$$y^i = F^i{}_4(x^4, F^i{}_3(x^3, F^i{}_2(x^2, F^i{}_1(x^1, 0)))). \tag{14}$$

In fact, by denoting the components of ω^a with respect to the x^i system by $\omega^a{}_i$, it is easy to see that

$$\omega^a{}_4 = \frac{\partial y^i}{\partial x^4}\bar{\omega}^a{}_i = \bar{e}^i{}_4\bar{\omega}^a{}_i = \delta^a{}_4. \tag{15}$$

In a similar manner, it is easy to see that the remaining conditions in (9) are also satisfied by $\omega^a{}_i$.

A more intuitive definition of the coordinate system x^i for which the conditions (9) are satisfied can be given.[6] It is most easily explained in terms of the vector frame $\{e_b\}$ dual to the frame $\{\omega^a\}$. Each vector field of the frame $\{e_b\}$ has associated with it a family of curves to which it is tangent. Each curve has an intrinsically defined parameter such that the vector is differentiation with respect to this parameter. The point with canonical coordinates x^i is obtained by starting from P and going a parameter distance x^1 along the curve with tangent e_1, then proceeding from this point a parameter distance x^2 along the curve with tangent e_2, etc. It is easy to see that such a coordinate system has the properties listed in (9) and is unique.

IV. ANALYSIS OF THE EQUATIONS

The Einstein–Petrov equations in the form of (7) constitute a set of algebraic partial differential equations, but with form differentiation rather than ordinary variable differentiation. Assume that some coordinate system has been chosen, for example, the canonically defined system presented in the preceding section. The form differentiation of the fields $\gamma^a{}_{bc}$ in (7) will be replaced by ordinary coordinate differentiation by means of the identity

$$f_{|a}\omega^a{}_i = f_{,i}. \tag{16}$$

This, of course, entails the introduction of the sixteen variables $\omega^a{}_i$, which must satisfy (6), and the set of equations now must be enlarged to include this set. However, as is well known, the integrability conditions for (6) are nothing but one set of the algebraic identities satisfied automatically by any of the three Petrov forms for the Riemann tensor. Hence, assuming the satisfaction of (7) with R_{abcd} one of the Petrov arrays, integrability conditions need only be considered for (7) themselves, which may be retained in their form differentiation form.

At first glance it might seem that the integrability conditions for (7) are merely the Bianchi identities

$$R_{abcd;e} + R_{abde;c} + R_{abec;d} = 0 \tag{17}$$

for the covariant derivatives of the Riemann tensor. It is certainly true that this is the case if the R_{abcd} are assumed to be arbitrary functions subject only to the algebraic symmetries and cyclic identities. However, other conditions, such as the Einstein condition of vanishing Ricci tensor, and *a fortiori* the Petrov canonical form requirements, may impose new and independent conditions on the connection components which must then be checked against the differential conditions (7) on these fields, etc. Hence, in general it should not be expected that (17) are the full integrability conditions for (7), plus the requirement that R_{abcd} be in one of the Petrov forms, and more general techniques must be used.

The problem of adding all differential consistency relationships (generalized integrability conditions) to a set of partial differential equations has been fully explored. The approach used here is based on that given in the book by Ritt[7] and a brief summary will be useful.

First, an ordering is assigned to the dependent variables and their partial derivatives in such a way that any two derivatives are comparable and each has a finite number of predecessors. Note, however, that here we may be dealing with noncommutative differentiation. Hence, some standard order must be established for the taking of derivatives with the necessary commutations made by use of the identity

$$f_{|a|b} = f_{|b|a} + f_c(\gamma^c{}_{ab} - \gamma^c{}_{ba}). \tag{18}$$

Next, each equation is regarded as determining the value of the highest derivative in it in terms of the others. This determination of derivatives is used in the proof of the existence and uniqueness theorems for analytic solutions. The consistency, both algebraic

and differential, of these determinations may be checked by the following operations on the set of equations:

(A) *Algebraic consistency:* The set of equations is checked to see if any pair of equations determine precisely the same derivative of the same variable. If so, this quantity is eliminated algebraically from one of the equations and the other equation remains unchanged. Checks are then resumed with the new set, until no further action of this type need be taken. Clearly this will occur after a finite number of steps. For example, when applied to a set of linear equations this operation would produce an equivalent set in sequentially solvable triangular form.

(B) *Sequential derivative consistency:* In this check the set is tested for the possibility that one equation E_1 determines a derivative of a quantity determined by a second equation E_2. If so, E_2 is differentiated so as to provide the derivative occurring in E_1, and the result is a new equation E_3. The common quantity in E_3 is then algebraically eliminated between E_3 and E_1 and the result replaces E_1. The check is then resumed with the new set. Clearly, this iteration need be performed only a finite number of times. In fact, the result of one such cycle is to replace one equation in the set with another which determines a derivative lower than the original, and the ordering was required to be such that each derivative has only a finite number of lower derivatives.

(C) *Cross differentiation consistency:* This is the phase similar to the adding of the ordinary integrability conditions. That is, if a pair of equations determine different derivatives of the same variable, each is differentiated an appropriate number of times to provide the determination of the same derivative of the same variable. This quantity is then algebraically eliminated between these equations and the result is added as a new equation to the set. The original pair of equations is retained unchanged, but some record is kept that the consistency condition for them has been added, so that it will not be duplicated in future cycles. Again, it can be shown that, as a consequence of the properties of the ordering of the derivatives, this procedure will be repeated only a finite number of times.

These three checks (A)–(C) are applied in order stated and, if any change is made in the set, they are again repeated. It may happen that at some point an equation is produced which does not contain any of the dependent variables or the derivatives. If this equation is nontrivial, it cannot be satisfied for arbitrary values of the independent variables, so that it must be concluded that the original set of equations is inconsistent. If this does not occur, then a finite number of repetitions of (A)–(C) will produce a set for which all algebraic and differential consistency checks have been made. The final set may then be regarded as a completely integrable set of partial differential equations to which a theorem such as a generalized Cauchy–Kowalewski theorem may be applied.

V. COMPUTER TECHNIQUES

The application of these procedures to the Einstein–Petrov equations (7) is obviously a very complicated task involving not only the production of many large equations but also the application of many very involved algebraic and differential operations to them. Clearly, this is the sort of task for which digital computers could be put to very good use. The possibility of using computers to perform formal, nonnumerical, algebraic and differential operations on polynomials has been considered for some time.[8] In fact, most of the new computer languages contain internal capabilities for such operations. However, the techniques used in these approaches are most efficient for polynomials with a fixed, rather small number of variables and thus are not appropriate to the analysis of the Einstein–Petrov equations.

Consequently, a straightforward approach was developed for performing such operations within the framework of any standard computer language, e.g., Fortran. An early stage of this program was reported in a computer journal,[9] and a brief description will be given here.

Variables are represented by integers in some convenient coding. For example, the connection components with lower indices, γ_{abc}, can be represented by some integer-valued function of the indices a, b, and c, which takes into account the symmetries of γ_{abc}. In this particular case, the number of integers required is clearly less than 100 and hence requires only two decimal digit places.

The next step is to represent derivatives. The most convenient way would be to have this coding such that numerical inequalities would correspond to an ordering consistent with the requirements discussed in Sec. IV above. One simple way to do this is the following. Assume that undifferentiated variables are represented by integers less than 100. Then, if f is represented by n, $f_{|a}$ will be represented by $100a + n$. Higher derivatives are described similarly so that, in effect, the individual decimal digits read from left to right give the successive derivatives in order of most recent occurrence, while the last two digits indicate the dependent variable. Clearly, this technique is consistent with the requirements imposed on the ordering of variables described

in Sec. IV above. In fact, differentiation clearly increases the integer and since all of the representative integers are nonnegative, each number has only a finite number of predecessors. Also note that this takes into account the possible noncommutative property of form differentiation. Thus, if the function f is represented by the integer n, then

$$f_{|a_1|a_2} \rightleftarrows 1000a_2 + 100a_1 + n, \tag{19}$$

$$f_{|a_2|a_1} \rightleftarrows 1000a_1 + 100a_2 + n. \tag{20}$$

For convenience, the computer program was designed to produce all derivatives in standard order, with the digits corresponding to derivatives nonincreasing when read from right to left. To obtain such an arrangement, use is made of the identity (18), so that, if $a_2 < a_1$, $f_{|a_1|a_2}$ is replaced by the expression on the right-hand side of this equation.

Once an appropriate code has been chosen for the variables, polynomials involving them can be represented by arrays of integers describing the coefficients of each term and the variables occurring as factors in the term. Thus, a matrix of integers $m_{ij}(P)$, where $i = 1, \cdots, t, j = 1, \cdots, f+1$, represents the polynomial P of t terms with at most f factors in each term. The numerical coefficient of the ith term is m_{i1} while $m_{i2}, \cdots, m_{if+1}$ are nonnegative integers which give the numerical codes for the variables or derivatives which occur as factors in this term. If fewer than f factors occur, the corresponding excess m_{ij} are 0, i.e., $m_{ij} = 0$ can be interpreted as implying that the jth factor in the ith term is merely the trivial constant 1. For example, assume that the code for γ_{12i} is $10i + 1$, then the polynomial

$$P = 2\gamma_{121|2} - 3\gamma_{122|3|4} + \gamma_{121}\gamma_{123} - 4 \tag{21}$$

will be represented by an array with $t = 4, f = 2$, and

$$(m_{ij}) = \begin{pmatrix} 2 & 211 & 0 \\ -3 & 4321 & 0 \\ 1 & 11 & 31 \\ -4 & 0 & 0 \end{pmatrix}. \tag{22}$$

With this translation of algebraic quantities involving variables and their derivatives into numerical arrays, it is possible to use any computer language to perform operations of algebra and calculus on the corresponding polynomials. In fact, in the following we will use the terms "polynomial" and "matrix" interchangeably. For the purpose of the analysis of the differential equations the polynomials are regarded as representing the left-hand side of differential equations. Since in this analysis the equations determine the highest variable in them, it is useful to order each such polynomial in some standard manner. One basic subroutine ADD does this and further simplifies the polynomial. First, the factors in each term are ordered in nonincreasing manner so that

$$m_{ij} \geq m_{ik}, \quad \text{of} \quad k \geq j > 1. \tag{23}$$

Next, terms having the same sequence of variable factors are combined by adding their coefficients and terms with zero coefficients eliminated. Finally, the set of terms is ordered lexicographically within itself so that, if $i < k$, then the first nonzero term in the sequence $(m_{ij} - m_{kj})$, $j = 2, \cdots$, is positive. Such ordering thus eliminates all ambiguity associated with the representations of equivalent polynomials expressed in formally different manners.

Using this subroutine, we can easily perform algebraic and differential operations. For example, the addition of two polynomials P_1 and P_2, of t_1 and t_2, f_1 and f_2 terms and factors, respectively, is accomplished as follows. Assuming that $f_2 > f_1$, we find that the matrix corresponding to the sum $P_1 + P_2$ is then the result of the reduction by ADD of the matrix m_{ij} defined by

$$\begin{aligned} m_{ij} &= m_{ij}(P_1), \quad i \leq t_1, j \leq f_1 + 1, \\ m_{ij} &= 0, \qquad\quad i \leq t_1, j > f_1 + 1, \\ m_{ij} &= m_{ij}(P_2), \quad i > t_1. \end{aligned} \tag{24}$$

Similarly, for the product of two polynomials the result will be the reduction by ADD of $m_{ij}, i = 1, \cdots, t_1t_2, j = 1, \cdots, f_1 + f_2 + 1$:

$$\begin{aligned} m_{i1} &= m_{k1}(P_1)m_{n1}(P_2), \quad i = (k-1)t_1 + n, \\ m_{ij} &= m_{ji}(P_1), \qquad\quad 1 < j \leq f_1 + 1, \\ m_{ji} &= m_{ij-f_1}(P_2), \qquad j > f_1 + 1. \end{aligned} \tag{25}$$

Finally, operations of formal differentiation of polynomials can be performed according to obvious procedures. Here, however, care must be taken to order the differentiation in a standard manner. If necessary, (18) must be used. This, of course, entails some further technical difficulties in the writing of the program.

Given these subroutines, it is then clear how to proceed with the steps (A)–(C) described in Sec. IV. The highest variables in the equation $P = 0$, and thus the variable determined by this equation, is $m_{12}(P)$. Thus, to perform the checks in (A), it is only necessary to determine whether $m_{12}(P_1) = m_{12}(P_2)$, for P_1 and P_2 any distinct equations in the set. If so, the elimination of this common highest variable is accomplished by elementary algebraic manipulations involving multiplication and addition of polynomials, as

described above. The checks in(B)are made as follows. If $m_{12}(P_1) > m_{12}(P_2)$ and $10^n > m_{12}(P_2) \geq 10^{n-1}$, $n \geq 2$, then $m_{12}(P_1)$ is a derivative of $m_{12}(P_2)$ if and only if

$$m_{12}(P_2) - m_{12}(P_1) = 0, \quad \text{mod } (10^n). \tag{26}$$

If this is so, the operations involving both differentiation and algebraic eliminations can be carried out as described in(B).Finally, the cross differentiation check of (*C*) is satisfied if and only if

$$m_{12}(P_2) - m_{12}(P_1) = 0, \quad \text{mod } (100). \tag{27}$$

Thus the complete analysis of a set of algebraic partial differential equations and its reduction can be accomplished by this computer program. In fact, for this particular application, the input set of equations, the Einstein–Petrov equations plus Bianchi identities, is also made by the computer by means of another program whose input is a particular Petrov form for the Riemann tensor. Further, to save time and space, the cross differentiation checks of (C) in the structure equations (7) are not performed unless one of these equations is changed in a later step, since this would merely duplicate the Bianchi identities already added.

VI. REPRESENTATION OF SOLUTIONS

Given a set of equations which are complete in the sense that the steps (A)–(C) have been iterated until no changes are made, what can be said about the solutions? The usual questions concern first the existence and then the uniqueness of solutions. The first such theorem applicable to sets of analytic equations such as these, for which all differential consistency conditions are contained in the set, is the well-known Cauchy–Kowalewski theorem. This theorem essentially asserts the local existence of analytic solutions to such equations. The uniqueness part affirms that these solutions are completely determined by their "initial determinations," which consist of the evaluation at the origin of those derivatives of the solution which are not determined directly by the equations. *The generating functions are then those analytic functions obtained by taking the Taylor series consisting of the sums of the products of these initial determination with appropriate powers of the independent variables.* For example, if there are n independent variables and the variable u is such that only $\partial u/\partial x^1, \cdots, \partial u/\partial x^r$ (and derivatives of these) are determined by the set of equations then the initial determination for u consists of the value at the origin of all derivatives of it formed by combinations of members of the set $(\partial/\partial x^{r+1}, \cdots, \partial/\partial x^n)$. Hence, the generating function for u is any analytic function f of the variables $x^{r+1}, \cdots, x^n$, and the general analytic solution will have the form

$$u = f(x^{r+1}, \cdots, x^n) + \sum_{i=1}^{r} x^i F_i(x), \tag{28}$$

in which F is analytic and uniquely determined by f. *Thus, there is a* 1-*to*-1 *correspondence between arbitrary analytic functions f and local solutions for u to the set equations.* The GF's provide tools for investigation of the structure of the family of all local analytic solutions. For example, if f and g are the GF's of two solutions, then $f + g$ will be the GF for a third. In this way, a linear structure is established for the family of solutions. Similarly, other structures, such as algebraic, harmonic, etc., can be imposed on this family.

After the initial consideration of sets of partial differential equations under the assumptions of analyticity, it is soon realized that there are significant classification schemes for such equations which distinguish important characteristics of their solutions. Consider, for example, the set of equations

$$\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}, \quad \frac{\partial v}{\partial y} = \epsilon \frac{\partial u}{\partial x}. \tag{29}$$

Each of these sets (for $\epsilon = \pm 1$) constitute a completely integrable set of equations for the dependent variables u and v, their y derivative being determined in terms of lower derivatives. By the Cauchy–Kowalewski theorem, there exists a unique analytic solution to these equations for given initial values of the functions u and v on $y = 0$. In terms of the GF notation used above, the solutions are

$$u = f(x) + yF(x, y), \quad v = g(x) + yG(x, y), \tag{30}$$

where f and g are arbitrary analytic functions and F and G are analytic in both x and y and uniquely determined by f and g.

Two important questions now arise concerning the classification of the solutions in terms of the GF's f and g. First, must all solutions to (29) be analytic? Secondly, is the use of GF's f and g truly appropriate for the representation of the solutions? Investigating the first problem soon makes it apparent that the choice of positive or negative values for ϵ is crucial. If ϵ is negative, the set of equations is elliptic and, in fact, if $\epsilon = -1$, this set is the set of Cauchy conditions that u and v represent the real and imaginary parts of an analytic function of a complex variable. Thus, all solutions of (29) with $\epsilon = -1$ *must* be analytic in x and y. This result may be extended to a wider class of partial differential equations, in more than two independent variables, and a criterion for a set of such equations to be elliptic can be defined. It turns out that,

as in the simple case of (29) with $\epsilon = -1$, the solutions to generalized totally elliptic equations must be analytic functions. On the other hand, if $\epsilon = +1$, the set of equations is hyperbolic, and it is well known that solutions exist which are only C^1. Again, this notation can be generalized to sets of equations with more dependent and independent variables and it is possible to generalize the notion of hyperbolicity in such a way that hyperbolic equations have solutions whose order of differentiability is minimal to accommodate the equations.[10] At any rate, the solutions need not be analytic, although of course they may be. Finally, the notion of elliptic and hyperbolic classification is not fully extensive and some sets of equations may not fall into either class. The study of the properties of solutions of such equations, called mixed equations, has not yet been completed. However, the Cauchy–Kowalewski theorem is still applicable to mixed equations so that within the framework of analytic solutions, the above remarks concerning the representation of solutions by analytic GF's are valid. Thus, although the restriction to analyticity is probably stronger than needed, most physically significant information can be obtained by studying piecewise analytic solutions.

There yet remains the question of the appropriateness of the GF approach to the representation of solutions. In the case of hyperbolic equations, the use of generating functions to specify the solutions is merely a generalization of the use of Cauchy initial data. There is no dispute that this Cauchy problem is "well posed" in this case since the solutions depend "continuously" on the initial data in any reasonable definition of continuity. On the other hand, when elliptic equations are considered there is a widely held view that the Cauchy problem, i.e., the giving of the GF's, is not a "well-posed" problem at all.[11] On the contrary, for elliptic equations it is generally asserted that the only appropriate boundary-value problem consists of Dirichlet or Neumann conditions on a closed surface.

The argument that the Cauchy data are inappropriate for elliptic equations is based on two facts: (1) The solution to an elliptic equation must be analytic. Therefore, nonanalytic initial data may not generate a solution, whereas they will for a hyperbolic equation. (2) The solution may not depend continuously on the initial data for elliptic equations whereas it does for hyperbolic equations. These objections do not appear to be entirely cogent, however, for the following reasons.

In the first place, if the set is elliptic, so that solutions must be analytic, then clearly any Cauchy initial data must be also. The confusion arises from the fact that the Dirichlet or Neumann problems on analytically defined closed surfaces do have solutions for suitably restricted, but nonanalytic, boundary data in the interior region bounded by the given closed surface. However, if the boundary data are not analytic, obviously the solution may not be continued over an open region which includes the boundary since the restriction of an analytic function to an analytically defined surface must be analytic there. Thus, such boundary data do not correspond to solutions of the equations in a neighborhood of the boundary but rather indicate a solution for which the equations break down at the boundary. Thus, in order for the boundary-value problem to produce a solution to elliptic equations in a neighborhood of the boundary, it is both necessary and sufficient that the boundary data be analytic both in the Dirichlet and Neumann problems as well as in the Cauchy problems.

The second objection, that for elliptic equations the solutions may not depend continuously on the Cauchy initial data, is valid only if the topology of the function space is made to be artificially weak. Thus, the standard example is a Cauchy problem for (29) with $\epsilon = -1$ and with initial data of the form

$$u_n(x, 0) = \cos(nx)/n, \qquad v_n(x, 0) = 0. \tag{31}$$

The solution is

$$u_n(x, y) = \cos(nx)\cosh(ny)/n, \qquad v_n(x, y) = \sin(nx)\sinh(ny)/n. \tag{32}$$

If the function space is topologized by means of a norm such as

$$N(f) = \sup(|f(x, 0)|: |x| \leq b), \tag{33}$$

where b is some finite positive number, then clearly the initial data converges to zero, whereas the solutions (32) diverge to infinity off of the initial line. However, the function space that is appropriate for such problems is the space of analytic functions, as discussed above. This space is not complete, relative to N. For example, the sequence of analytic functions

$$f_n(x) = \sum_{k=1}^{n} e^{-k} \cos k^2 x \tag{34}$$

is convergent relative to N since

$$N(f_n - f_m) \to 0, \tag{35}$$

but f_n converges to a function which is nowhere analytic in any neighborhood of 0.

The difficulty with the norm N is, of course, due to the fact that it fails to take into account the behavior of

the derivatives. It is likely that other norms may be defined which include the behavior of all derivatives so as to provide a topology for the initial data sufficiently strong that the solution will depend continuously on them.

For these reasons, then, it is the author's opinion that the use of GF's does provide an appropriate representation of the solutions.

VII. PETROV TYPE III RESULTS

In this section, we discuss the results of an application of the techniques described above to the Einstein–Petrov type III equations. The first step is to apply the operations (A)–(C) described in Sec. IV above to Eq. (7). Since it is known, however, that the results of the first round will be merely the Bianchi identities, these were initially added by the program that produced the input equations in the computer coding described in Sec. V. Since the type III Riemann tensor consists of constants only, its covariant derivatives are all homogeneous linear combinations of the $\gamma^a{}_{bc}$ with constant coefficients. Since there are twelve independent conditions, they may be regarded as providing determinations of twelve of the $\gamma^a{}_{bc}$ in terms of the remaining twelve. In the coding used this means that $\gamma^a{}_{bp}$, $p = 3, 4$, are determined by $\gamma^a{}_{bu}$, $u = 1, 2$. Here and in the following, the indices u, v will only take on the values 1, 2 while p, q will assume only the values 3, 4. Procedures (A)–(C) must now be applied to this set of equations. For this case however the functions $\gamma^a{}_{bp}$ are determined linearly in terms of $\gamma^a{}_{bu}$ and it turns out to be more practical merely to substitute these forms for $\gamma^a{}_{bp}$ into the entire set of structure equations (7). This substitution then leaves those highest variables in (7) which are of the form $\gamma^a{}_{bd|c}$, $c > d$, unchanged except for those with $d = 3$ so that $c = 4$. The results of the substitution then give six new equations for $\gamma^a{}_{bu|3}$ which must be compared [process (A)] with the rest of (7). Somewhat surprisingly, it turns out that these new equations are merely linear combinations of equations already in (7), so that they can be eliminated and impose no independent conditions. What remains of (7) then are 30 equations which determine the derivatives $\gamma^a{}_{bu|c}$ with $c > u$. Further, this set is completely integrable.

The next step is to replace form differentiation by coordinate differentiation with respect to the canonically defined coordinates described in Sec. III, using the fields $\omega^a{}_i$ subject to (6). As discussed in Sec. IV, these latter equations for the $\omega^a{}_i$ are completely integrable if (7) is satisfied with a Petrov form Riemann tensor. Using the derivative coding and ordering scheme, we see that these equations determine the derivatives $\omega^a{}_{i,j}$, $j > i$. Thus, $\omega^a{}_4$ is undetermined by (6), but the choice of canonical coordinates fixes them to satisfy

$$\omega^a{}_4 = \delta^a{}_4. \tag{36}$$

Similarly, only the derivatives $\omega^a{}_{3,4}$ are determined by (6) so that the solution for $\omega^a{}_3$ will involve only GF's of x^1, x^2, x^3, at $x^4 = 0$. Again, however, the choice of canonical coordinates means that these functions are determined to be

$$\omega^a{}_3 = \delta^a{}_3, \quad \text{for} \quad x^4 = 0. \tag{37}$$

Similar considerations with $\omega^a{}_2$ and $\omega^a{}_1$ show that the definition (9) of the canonical coordinates provides a full determination of the arbitrary functions associated with the solution of (6). In other words, (9) and the structure equations (6) and (7) fully determine the $\omega^a{}_i$ in terms of the $\gamma^a{}_{bc}$.

The determination of the form derivatives $\gamma^a{}_{bu|c}$ from the remaining structure equations can then be translated into the determination of the corresponding ordinary derivatives (since $\omega^a{}_i = \delta^a{}_i$ at $x^i = 0$). Thus the completely integrable equations fully determine the $\omega^a{}_i$ and the derivatives $\gamma^a{}_{bu,c}$, $c > u$. From the Cauchy–Kowalewski theorem it then follows that the most general analytic solution in a neighborhood of $x^i = 0$ will be of the form

$$\gamma^a{}_{b1}(x^i) = {}_0\gamma^a{}_{b1}(x^1) + \sum_{k=2}^{4} x^k F^a{}_{bk}(x^i),$$

$$\gamma^a{}_{b2}(x^i) = {}_0\gamma^a{}_{b2}(x^1, x^2) + \sum_{k=3}^{4} x^k G^a{}_{bk}(x^i). \tag{38}$$

Here the ${}_0\gamma^a{}_{b1}$ are arbitrary analytic functions of x^1, and the ${}_0\gamma^a{}_{b2}$ are arbitrary analytic functions of x^1 and x^2. The functions F and G will then likewise be analytic and uniquely determined by the ${}_0\gamma^a{}_{bu}$ which are thus the generating functions referred to in the introduction. Finally, the actual metric (1) is determined by the $\omega^a{}_i$ which are obtained from the completely integrable equation (6) with initial determinations fully fixed by the coordinate conditions (9) as discussed above.

Thus, given the fixed point at which $x^i = 0$, that is, the origin of the canonical coordinates, there is a 1-*to*-1 *correspondence between the local analytic type III solutions and the arbitrary analytic functions* ${}_0\gamma^a{}_{bu}$.

VIII. APPLICATIONS AND DISCUSSION

This section discusses some of the applications and the problems to be considered in a subsequent paper. Before beginning any discussion of the use of the GF's to represent solutions, it should be pointed out that the necessity of keeping fixed the origin of the canonical coordinates gives rise to difficulty in the area of the uniqueness of these representations. In fact, it is clear

that the description of solutions by GF's depends on the choice of the fixed point so that two different GF's may not indeed give rise to essentially different solutions but rather to the same solution referred to different origins.

Also, the GF description is not readily adaptable to the more usual studies of solutions based on symmetries because the Killing vectors are not necessarily simply related to the Petrov directions. However, it is relatively easy to find GF's that produce solutions with no invariant varieties.

The most natural initial applications of the GF's are to produce the actual analytic form of the metric. Thus, the GF's would be defined by power-series expansions and successive terms in the $\omega^a{}_i$ obtained by algebraic operations. This approach would also be one in which a computer would be helpful. Thus the most general analytic type III metrics could be invariantly exhibited to as many power series terms as desired, in terms of the corresponding terms in the GF's ${}_0\gamma^a{}_{bu}$.

On a more theoretical plane, the correspondence between type III metrics and the arbitrary GF's can be used to investigate the space of local analytic type III metrics. Thus a linear and even algebraic structure can be imposed on the space of solutions simply by imposing it on the space of the GF's, which, because the GF's are arbitrary analytic functions, is closed under linear and algebraic operations. For example, solutions corresponding to two sets of generating functions may be "added" to give a third solution represented by the GF's which are the sums of the original pair.

Finally, another interesting problem is the possible relaxation of the condition of analyticity of the solutions. For mixed sets of equations which are neither totally elliptic nor hyperbolic, such as the completely reduced Einstein–Petrov equations appear to be, not all such questions have yet been settled.

ACKNOWLEDGMENT

The author would like to express his gratitude to E. Conway for helpful discussions concerning the mathematical structure of solutions to sets of partial differential equations.

* Work supported by a National Science Foundation Grant.

[1] For a good review see Y. Bruhat, in *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), Chap. IV.

[2] A. S. Petrow (Petrov), *Einstein Räume* (Akademie-Verlag, Berlin, 1964).

[3] Carl H. Brans, J. Math. Phys. **6**, 94 (1965), Lettere al Nuovo Cimento **2**, 699 (1969). The idea of expressing the Einstein equations in invariant form for scalar fields was proposed earlier by E. T. Newman and L. A. Tamburino [J. Math. Phys. **2**, 667 (1961)], while the problem of the freedom of specification for the initial data for an arbitrary Cauchy problem is examined in Petrov's book (Ref. 2), pp. 285–294.

[4] Other authors variously refer to such a basis as a "tetrad," "vierbein," or even as a "nonholonomic coordinate system."

[5] Note that the Latin indices $a, b, c, \cdots$, are raised and lowered by means of the constant η_{ab}.

[6] The author is indebted to R. Geroch for pointing this out to him.

[7] J. Ritt, *Algebraic Differential Equations* (American Mathematical Society Colloquium Publications, New York, 1932), Vol. XIV, especially Chaps. IX and X.

[8] W. Brown, Bell System Tech. J. **42**, 2081 (1963).

[9] Carl H. Brans, J. Assoc. Computing Machinery **14**, 45 (1967).

[10] A. Friedman, *Generalized Functions and Partial Differential Equations* (Prentice-Hall Inc., Englewood Cliffs, N.J. 1963).

[11] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II, pp. 226–231.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Reduction of the Direct Product of Representations of the Poincaré Group

H. E. Moses
Air Force Cambridge Research Laboratories, Bedford, Massachusetts 01730

(Received 12 September 1969)

We expand the direct product of two representations of the Poincaré group into representations of the Poincaré group in the general case that the factors of the direct product may have any mass, whether real, zero, or imaginary, and the total energy may be indefinite. The representations of the Poincaré group, which appear in the expansion of the direct product, have masses which run through a continuous spectrum of real and imaginary values and are irreducible in terms of the mass and sign of energy (for real mass), but are reducible in terms of the infinitesimal generators of the little groups. To obtain the expansion in terms of irreducible representations, one need only reduce the infinitesimal generators of the little groups. This reduction is carried out for the real mass components and, in principal at least, can be carried out for the infinitesimal generators of the little groups for the imaginary mass components. The factors of the direct product and the representations which appear in the expansion are expressed in terms of a particular momentum representation called "the standard helicity representation" which enables us to use a uniform notation for all masses, whether real, zero, or imaginary. The earlier portion of the present paper summarizes the properties of these representations.

1. INTRODUCTION

In Ref. 1, we obtained all the representations of the infinitesimal generators of the Poincaré group such that the generators were integrable and, in addition, that the infinitesimal generators corresponding to the energy and momentum were Hermitian (but not identically zero) and the generators corresponding to the angular momentum operators were Hermitian. These representations were in a form in which the helicity operator $\mathbf{P} \cdot \mathbf{J}$ played a particularly important role, and we used the term "standard helicity representations or realizations" to describe the form of the representation.

Each of these realizations acts on functions in a space which is the direct product of the space for realizations of the infinitesimal generators for scalar particles and of a space for the representation of the infinitesimal generators of the little group appropriate to the mass of the representation, that is, the rotation group, the Euclidean group in the plane, and the rotation group in pseudo-Euclidean space for positive, zero, and imaginary mass, respectively. Irreducible Hermitian representations of the infinitesimal generators of the Poincaré group are special cases of the above standard helicity representations.

In addition to the advantage of depending upon the helicity explicitly, the standard helicity realizations of the Poincaré group have a similar appearance as expressed in terms of the infinitesimal generators of the little group, for all masses, whether real, zero, or imaginary. It is not necessary, in many applications, to discuss separately the cases of real, zero, and imaginary masses.

It is the object of the present paper to exhibit the direct product of two such standard helicity representations and represent the direct product as a sum (actually an integral) of standard helicity representations. Working with helicity representations enables us to use a notation for the factors of the direct product largely independent of the value of the square of the mass or sign of energy of the factors. The square of the masses, which occurs in the sum, runs through a continuous spectrum of positive and negative values. Zero-mass representations are of zero measure and thus contribute nothing.

As part of the reduction, we give the little group associated with each representation in the sum. The little groups appear in unreduced form. In an appendix we reduce the little group, that is, the rotation group in a helicity representation, for the real mass representations. In principle, it is possible, though difficult, to reduce the little groups for imaginary mass representations, but we have not done so in the present paper. For some applications, however, such a reduction is not necessary.

Thus, the material of the present paper is a generalization over previous treatments of the reduction of the direct product of two representations of the Poincaré group (see, for example, Jacob and Wick,[2] Joos,[3] Kummer,[4] Lomont,[5] Macfarlane,[6] Moussa and Stora,[7] Werle,[8]) as follows: (1) in the present paper, one or both of the representations in the direct product may have imaginary mass; (2) the total energy may be indefinite so that the reduction of the direct product contains imaginary mass representations; (3) the representations in the direct product need not be

irreducible; and (4) while the infinitesimal generators corresponding to space–time rotations must be integrable and those corresponding to energy, components of linear momentum, and components of angular momentum must be integrable and Hermitian, the infinitesimal generators corresponding to space–time rotations need not be Hermitian.

To complete the reduction program, we also reduce the direct product of the representations of the Poincaré group, for which the energy and momentum operators are not identically zero, and a representation for which these operators are identically zero, that is, a representation of the homogeneous Lorentz group. The reduction contains representations of the Poincaré group in standard helicity representations.

Aside from the mathematical interest, our motivation for studying direct products for representations for more general situations than heretofore is our interest in reducing currents and interactions which appear in nonlinear systems of coupled wave equations, for example, the Dirac equation coupled to the electromagnetic vector potential. In Ref. 9 we showed how wavefunctions in configuration space constitute bases of representations of the Poincaré group. We proceeded to reduce these wavefunctions into irreducible representations of the Poincaré group for real nonzero mass in the Foldy–Shirokov realization. In Ref. 10 we extended the reduction process to include zero-mass representations. For the purpose of decomposing currents and interactions, the most general situation is obtained by expanding the wavefunction in configuration space into all possible representations of the Poincaré group. For the sake of simplicity we can use standard helicity representations for all masses, whether real, zero, or imaginary, for the case that the energy and momentum generators are not identically zero. These expansions can be carried out using the techniques discussed in Ref. 1. In the expansion of the wavefunction, we include the possibility that representations of the homogeneous group are also present, for the sake of the greatest generality. Then, having expanded the wavefunction as described above, we take the products of wavefunction as they appear in interactions and currents. We find that to reduce such products we are led immediately to the reduction problem considered in the present paper.

We shall give only the results for the sake of brevity. The derivation of the results, while lengthy, is a direct application of the reduction techniques discussed in Ref. 1. Furthermore, since we wish to use the results for calculations to be discussed in later papers, the reduction is given in very explicit form.

2. STANDARD HELICITY REPRESENTATIONS OF THE POINCARÉ GROUP

In the present section we shall give the standard helicity representations of the Poincaré group under the conditions that the infinitesimal generators are integrable, that the infinitesimal generators corresponding to energy, components of the linear momentum, and components of the angular momentum are Hermitian and that the infinitesimal generators corresponding to the energy and components of the linear momentum are not identically equal to zero.

Specifically, let P^α, $\alpha = 0, 1, 2, 3$, be the infinitesimal generators corresponding to the energy and components of linear momentum. We use the metric $g^{\alpha\beta} = g_{\alpha\beta} = 0$ if $\alpha \neq \beta$, $g^{00} = g_{00} = -g^{ii} = -g_{ii} = -1$. Hence, $P^0 = -P_0$ and $P^i = P_i$. We denote the components of the angular-momentum tensor by $J_{\alpha\beta} = -J_{\beta\alpha}$. Then we require the following commutation rules to be satisfied:

$$[P^\alpha, P^\beta] = 0, \quad [J_{\alpha\beta}, P_\gamma] = i(g_{\alpha\gamma}P_\beta - g_{\beta\gamma}P_\alpha),$$
$$[J_{\alpha\beta}, J_{\gamma\delta}] = i(g_{\alpha\gamma}J_{\beta\delta} - g_{\beta\gamma}J_{\alpha\delta} + g_{\alpha\delta}J_{\gamma\beta} - g_{\beta\delta}J_{\gamma\alpha}). \tag{1}$$

We shall write

$$H = P^0 = -P_0, \quad \mathbf{J} = (J_{23}, J_{31}, J_{12}),$$
$$\mathfrak{J} = (J_{01}, J_{02}, J_{03}). \tag{2}$$

The components of $\mathbf{J}$ are, of course, the components of the angular momentum, whereas those of $\mathfrak{J}$ correspond to infinitesimal space–time rotations.

Our integrability requirements can then be stated as follows: The operators $\exp[ia^\alpha P_\alpha]$, $\exp[i\boldsymbol{\theta}\cdot\mathbf{J}]$, and $\exp[i\boldsymbol{\beta}\cdot\mathfrak{J}]$ exist for all real numbers a^α, θ_i, and β_i, where θ_i and β_i are the components of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, respectively. The Hermiticity conditions are that the operators P^α and J_i are Hermitian.

We shall now introduce a notation for the representations of the little group. The three infinitesimal generators of the little group which we denote by T_1, T_2, and M satisfy the commutation relations

$$[T_1, M] = -iT_2, \quad [T_2, M] = iT_1,$$
$$[T_1, T_2] = iA(c)M, \tag{3a}$$

where c is a variable (which we shall later identify with the square of the mass) whose values include the entire real axis. The function $A(c)$ is given by

$$A(0) = 0, \quad A(c) = c/|c|, \quad \text{for} \quad c \neq 0. \tag{3b}$$

Thus, for positive values of c, the operators satisfy the commutation rules of the infinitesimal generators of

the rotation group, while, for $c = 0$ and $c < 0$, the commutation rules are those for the infinitesimal generators of the Euclidean group in the plane and the rotation group in pseudo-Euclidean space, respectively.

The integrability and Hermiticity requirements on J_i and $\mathcal{J}_i$ lead to the requirement on the infinitesimal generators of the little group that M be Hermitian and have only integers or half-odd integers as eigenvalues and that $\exp[i\mathbf{a}\cdot\mathbf{T}]$ exists for all real a_1 and a_2, where $\mathbf{a}\cdot\mathbf{T} = a_1T_1 + a_2T_2$.

A realization or representation of the operators T_i and M, which we shall call a realization of the first kind, is given in the following way. Each such realization has associated with it a Hilbert space $\{f(\lambda)\}$ of complex functions of a set of variables which we collectively denote by λ. The inner product in the Hilbert space is defined with the aid of a measure function $m(\lambda)$ such that the inner product of two functions is $\int f^{(1)*}(\lambda)f(\lambda)\,dm(\lambda)$. The operators T_1, T_2, and M are represented by linear operators in this space. We shall write the function resulting from applying the operators T_i and M on $f(\lambda)$ as $T_i^\lambda f(\lambda)$ and $M^\lambda f(\lambda)$, respectively, to emphasize that the operators act on the λ variables in the function. Any other operator A in this space will likewise be written A^λ to emphasize that it acts on the variable λ.

Realizations of the first kind are those customarily used for representing operators. However, we shall find it useful to introduce a slight generalization which we shall call realizations or representations of the second kind. Let W be any positive-definite Hermitian operator. For realizations of the second kind we replace the earlier inner product by $\int f^{(1)*}(\lambda)W^\lambda f(\lambda)\,dm(\lambda)$. For use in the representations of the Poincaré group, realizations of the generators of the little group of the second kind prove more useful than those of the first kind. We shall call W the weight operator associated with the representation.

Let us consider a second representation of the second kind of the operators T_i and M. The Hilbert space for this representation will be the space of complex functions $\{g(\mu)\}$ of a set of variables μ. The measure function is $M(\mu)$ and weight operator is V. We shall find the following definition useful:

Definition: The two realizations are unitarity equivalent if there is a one-to-one linear correspondence between the functions of the set $\{f(\lambda)\}$ and the set $\{g(\mu)\}$ and if the inner product is preserved in the correspondence, that is, $\int f^{(1)*}(\lambda)W^\lambda f(\lambda)\,dm(\lambda) = \int g^{(1)*}(\mu)V^\mu g(\mu)\,dM(\mu)$ where $f(\lambda)$ and $f^{(1)}(\lambda)$ correspond to $g(\mu)$ and $g^{(1)}(\mu)$, respectively.

The following theorem is then easily proved:

Theorem: Every realization is equivalent to one in which the weight operator is the identity.

We are now in a position to give the standard helicity representations for the infinitesimal generators of the Poincaré group under the very general conditions described in the Introduction. To give the Hilbert space of functions, we shall characterize the variables which appear in these functions.

Let c be a real variable which can take on all real values. This variable will later be identified as the square of the mass. Let the components of the vector $\mathbf{p} = (p_1, p_2, p_3)$ take on all values in the 3-dimensional space for the case that $c \geq 0$ and take on values outside the sphere $|\mathbf{p}| > -c$ for $c < 0$. The vector $\mathbf{p}$ will later be seen to be the momentum variable. Let ϵ take on the two values ± 1. The quantity ϵ will be the sign of energy.

For each value of $c \geq 0$ and each value of ϵ we assign a representation of the little group whose infinitesimal generators satisfy the commutation rules (3). Such an assignment provides a set of variables λ and a set of operators T_i^λ and M^λ for each value of c and ϵ. For $c < 0$ we also assign representations of the little group for each value of c which, however, are independent of ϵ. Thus, we have variables λ and operators T_i^λ and M^λ which depend on c but not on ϵ. (In Appendix A we extend the calculations of Ref. 1 to prove that λ, T_i^λ, and M^λ can always be chosen to be independent of ϵ for $c < 0$.)

The representations of the little groups which we require are of the second kind. We prescribe the weight operator $W^\lambda(c, \epsilon, p)$ with $p = |\mathbf{p}|$ as any function of the arguments. To complete the assignment of the little group and to enable us to define an inner product in the Hilbert space of functions upon which the infinitesimal generators of the Poincaré group act, we introduce three measure functions which are essentially measure functions in the product space of the c and λ variables. We denote these measure functions by $M^+(c, \epsilon, \lambda)$, $M^0(\epsilon, \lambda)$, and $M^-(c, \epsilon, \lambda)$ where $M^+ \equiv 0$ for $c < 0$ and $M^- \equiv 0$ for $c > 0$. We require that

$$M^-(c, -1, \lambda) = M^-(c, +1, \lambda).$$

Except for being consistent with the assignment with the little groups, $M^\pm$ are arbitrary as functions of c, and M^0 and M^+ are also arbitrary functions of ϵ.

Our Hilbert space for the representation of the infinitesimal generators of the Poincaré group is a space of complex functions $\{f(c, \epsilon, p, \lambda)\}$ with the

following inner product:

$$(f^{(1)},f) = \sum_{\epsilon}\iiint f^{(1)*}(c,\epsilon,\mathbf{p},\lambda)W^{\lambda}(c,\epsilon,p) \times f(c,\epsilon,\mathbf{p},\lambda)\frac{d\mathbf{p}}{\omega(c,p)}dM^{+}(c,\epsilon,\lambda) + \sum_{\epsilon}\iiint f^{(1)*}(0,\epsilon,\mathbf{p},\lambda)W^{\lambda}(0,\epsilon,p) \times f(0,\epsilon,\mathbf{p},\lambda)\frac{d\mathbf{p}}{\omega(c,p)}dM^{0}(\epsilon,\lambda) + \sum_{\epsilon}\iiint f^{(1)*}(c,\epsilon,\mathbf{p},\lambda)W^{\lambda}(c,\epsilon,p) \times f(c,\epsilon,\mathbf{p},\lambda)\frac{d\mathbf{p}}{\omega(c,p)}dM^{-}(c,\epsilon,\lambda), \quad (4)$$

where $\omega(c,p) = (c+p^2)^{\frac{1}{2}}$.

We shall now give the infinitesimal generators of the Poincaré group explicitly. First we note that, from the assignment of the little groups for all c and ϵ, we know the operators T_i^{λ} and M^{λ}, that is, we know the meaning of $T_i^{\lambda}f(c,\epsilon,\mathbf{p},\lambda)$ and $M^{\lambda}f(c,\epsilon,\mathbf{p},\lambda)$. Secondly, we define

$$p = |\mathbf{p}|, \quad \omega(c,p) = (c+p^2)^{\frac{1}{2}} \quad (5)$$

as before. Also

$$\mathbf{p}\cdot\mathbf{T}^{\lambda} = p_1T_1^{\lambda} + p_2T_2^{\lambda}, \quad B(c) = [|c|]^{\frac{1}{2}}, \quad \text{for} \quad c \neq 0, \quad B(0) = 1, \quad (6a)$$

and

$$\boldsymbol{\nabla} = (\nabla_1,\nabla_2,\nabla_3) = \left(\frac{\partial}{\partial p_1},\frac{\partial}{\partial p_2},\frac{\partial}{\partial p_3}\right). \quad (6b)$$

Then,

$$\begin{aligned}
&P_if(c,\epsilon,\mathbf{p},\lambda) = p_if(c,\epsilon,\mathbf{p},\lambda),\\
&Hf(c,\epsilon,\mathbf{p},\lambda) = P^0f(c,\epsilon,\mathbf{p},\lambda)\\
&\qquad = \epsilon\omega(c,p)f(c,\epsilon,\mathbf{p},\lambda),\\
&J_3f(c,\epsilon,\mathbf{p},\lambda) = [-i(\mathbf{p}\times\boldsymbol{\nabla})_3 + M^{\lambda}]f(c,\epsilon,\mathbf{p},\lambda),\\
&J_if(c,\epsilon,\mathbf{p},\lambda)\\
&\quad = \{-i(\mathbf{p}\times\boldsymbol{\nabla})_i + [p_i/(p+p_3)]M^{\lambda}\}f(c,\epsilon,\mathbf{p},\lambda), \quad i = 1,2,\\
&\mathcal{J}_3f(c,\epsilon,\mathbf{p},\lambda)\\
&\quad = \epsilon\{i\omega(c,p)\nabla_3 + [B(c)/p^2](\mathbf{p}\cdot\mathbf{T}^{\lambda})\}f(c,\epsilon,\mathbf{p},\lambda),\\
&\mathcal{J}_1f(c,\epsilon,\mathbf{p},\lambda)\\
&\quad = \epsilon\{i\omega(c,p)\nabla_1 + [p_2/p(p+p_3)]\omega(c,p)M^{\lambda}\\
&\qquad + \frac{B(c)}{p^2}[p_1\mathbf{p}\cdot\mathbf{T}^{\lambda}/(p+p_3) - pT_1^{\lambda}]\}f(c,\epsilon,\mathbf{p},\lambda),\\
&\mathcal{J}_2f(c,\epsilon,\mathbf{p},\lambda)\\
&\quad = \epsilon\{i\omega(c,p)\nabla_2 - [p_1/p(p+p_3)]\omega(c,p)M^{\lambda}\\
&\qquad + \frac{B(c)}{p^2}[p_2\mathbf{p}\cdot\mathbf{T}^{\lambda}/(p+p_3) - pT_2^{\lambda}]\}f(c,\epsilon,\mathbf{p},\lambda).
\end{aligned} \quad (7)$$

We note that $\mathbf{P}\cdot\mathbf{J}f(c,\epsilon,\mathbf{p},\lambda) = pM^{\lambda}f(c,\epsilon,\mathbf{p},\lambda)$; hence, the name "standard helicity representation."

We now give the integrated forms of the representation. Let us define p^0 by

$$p^0 = \epsilon\omega(c,p). \quad (8)$$

Then,

$$\exp[ia_{\alpha}P^{\alpha}]f(c,\epsilon,\mathbf{p},\lambda) = \exp[ia_{\alpha}p^{\alpha}]f(c,\epsilon,\mathbf{p},\lambda). \quad (9a)$$

In Eq. (9a) we use the summation convention where the a^{α} are four numbers which we take to be real for purposes of representation theory.

Let $\boldsymbol{\theta} = (\theta_1,\theta_2,\theta_3)$ be an arbitrary real vector and let

$$\theta = |\boldsymbol{\theta}|. \quad (9b)$$

Then

$$\exp[i\boldsymbol{\theta}\cdot\mathbf{J}]f(c,\epsilon,\mathbf{p},\lambda) = \exp[2i\Phi(\boldsymbol{\theta},\mathbf{p})M^{\lambda}]f(c,\epsilon,\mathbf{p}',\lambda), \quad (9c)$$

where Φ is defined as being the principal branch of

$$\tan\Phi(\boldsymbol{\theta},\mathbf{p}) = \frac{(\boldsymbol{\theta}\cdot\mathbf{p}+\theta_3p)\tan(\frac{1}{2}\theta)}{(p+p_3)+(\boldsymbol{\theta}\times\mathbf{p})_3\tan(\frac{1}{2}\theta)} \quad (9d)$$

and

$$\mathbf{p}' = \mathbf{p}\cos\theta + \frac{1-\cos\theta}{\theta^2}(\boldsymbol{\theta}\cdot\mathbf{p})\boldsymbol{\theta} + \frac{\sin\theta}{\theta}(\boldsymbol{\theta}\times\mathbf{p}). \quad (9e)$$

We note that Φ depends on the direction of the vector $\mathbf{p}$ but is independent of the magnitude of this vector.

To obtain $\exp[i\boldsymbol{\beta}\cdot\boldsymbol{\mathcal{J}}]$ for all real vectors $\boldsymbol{\beta} = (\beta_1,\beta_2,\beta_3)$, we use the identity

$$\exp[i\boldsymbol{\beta}\cdot\boldsymbol{\mathcal{J}}] = \exp[-i\boldsymbol{\xi}\cdot\mathbf{J}]\exp[i\beta\mathcal{J}_3]\exp[i\boldsymbol{\xi}\cdot\mathbf{J}], \quad (9f)$$

where the vector $\boldsymbol{\xi} = (\xi_1,\xi_2,\xi_3)$ is related to $\boldsymbol{\beta}$ through

$$\beta = |\boldsymbol{\beta}|, \quad \xi = |\boldsymbol{\xi}|,$$
$$\cos\xi = \frac{\beta_3}{\beta}, \quad \sin\xi\frac{\xi_2}{\xi} = \frac{\beta_1}{\beta}, \quad \sin\xi\frac{\xi_1}{\xi} = -\frac{\beta_2}{\beta}. \quad (9g)$$

We shall now give $\exp[i\beta\mathcal{J}_3]$. We define the functions $g(c,x)$ by

$$\begin{aligned}
&g(0,x) = x,\\
&g(c,x) = \arctan[(c)^{\frac{1}{2}}x], \quad \text{for} \quad c > 0,\\
&g(c,x) = \tfrac{1}{2}\log[|1+(-c)^{\frac{1}{2}}x|/|1-(-c)^{\frac{1}{2}}x|],\\
&\qquad\qquad \text{for} \quad c < 0.
\end{aligned} \quad (9h)$$

Also for every vector $\mathbf{p}$ and sign of energy ϵ we introduce the vector $\mathbf{p}'$ and sign of energy ϵ' through

the following Lorentz transformation:

$$p_1' = p_1, \quad p_2' = p_2,$$
$$p_3' = p_3 \cosh\beta - \epsilon\omega(c, p)\sinh\beta, \tag{9i}$$
$$\epsilon'\omega(c, p') = \epsilon\omega(c, p)\cosh\beta - p_3\sinh\beta, \quad p' = |\mathbf{p}'|.$$

We note that $\epsilon' = \pm 1$. Furthermore, when $c \geq 0$, we have $\epsilon' = \epsilon$. Then,

$$\exp[i\beta\mathfrak{J}_3]f(c, \epsilon, p, \lambda)$$
$$= \exp\left\{i\frac{\mathbf{p}\cdot\mathbf{T}^\lambda}{(p^2 - p_3^2)^{\frac{1}{2}}}\left[g\left(c, \frac{p_3}{(p^2 - p_3^2)^{\frac{1}{2}}\omega(c, p)}\right)\right.\right.$$
$$\left.\left. - g\left(c, \frac{p_3'}{(p^2 - p_3^2)^{\frac{1}{2}}\omega(c, p')}\right)\right]\right\}f(c, \epsilon', \mathbf{p}', \lambda). \tag{9j}$$

For $c \geq 0$, the above expressions can be simplified by noting that

$$(p^2 - p_3^2)^{-\frac{1}{2}}\left[g\left(0, \frac{p_3}{(p^2 - p_3^2)^{\frac{1}{2}}p}\right) - g\left(0, \frac{p_3'}{(p^2 - p_3^2)^{\frac{1}{2}}p'}\right)\right]$$
$$= \frac{\epsilon\tanh\beta}{p^2 - \epsilon p p_3\tanh\beta} \tag{9k}$$

and

$$g\left(c, \frac{p_3}{(p^2 - p_3^2)^{\frac{1}{2}}\omega(c, p)}\right) - g\left(c, \frac{p_3'}{(p^2 - p_3^2)^{\frac{1}{2}}\omega(c, p')}\right)$$
$$= \epsilon\arctan\left(\frac{[c(p^2 - p_3^2)]^{\frac{1}{2}}\tanh\beta}{p^2 - \epsilon p_3\omega(c, p)\tanh\beta}\right), \quad c > 0. \tag{9l}$$

We wish now to introduce the definition of equivalence of two helicity representations of the Poincaré group.

Let us assume that we have a second helicity realization defined in a space of functions $\{\hat{f}(c, \epsilon, \mathbf{p}, \mu)\}$, where μ is a variable analogous to λ of the previous representation and is used to describe the realization of the generators of the little group which in this second realization we designate by T_i^μ, M^μ. Furthermore, for this realization we have measures $\hat{M}^{\pm}(c, \epsilon, \mu)$, $\hat{M}^0(\epsilon, \mu)$ and weight operators $\hat{W}^\mu(c, \epsilon, p)$ for use in an inner product analogous to Eq. (4).

Definition: The two realizations are unitarily equivalent if there is a one-to-one linear correspondence between the functions of the two spaces such that the inner product is preserved.

The definition of unitary equivalence of two helicity representations leads to the following theorem:

Theorem: Two helicity representations are unitarily equivalent if and only if they contain the same values of c and ϵ and if the set of realizations of the corresponding little groups are unitarily equivalent.

We now wish to characterize irreducible helicity representations.

Definition: The set of infinitesimal generators of the little group (for all c and ϵ) are irreducible if there is no proper invariant subspace in the space of functions $\{f(c, \epsilon, \mathbf{p}, \lambda)\}$.

Theorem: The set of infinitesimal generators of the Poincaré group is irreducible if and only if all of the following conditions are satisfied:

(1) c takes on only one value,

(2) if $c \geq 0$, all of the functions $f(c, \epsilon, \mathbf{p}, \lambda)$ are identically zero for one value of ϵ, and

(3) the realization of the little group T_i^λ and M^λ is irreducible.

We note that, if the set of Poincaré infinitesimal generators is irreducible, then there is no integration over c and we can drop the argument c in places. For $c \geq 0$ we can similarly drop explicit dependence on ϵ. Then for $c \geq 0$ we write $f(\mathbf{p}, \lambda)$ for the function $f(c, \epsilon, \mathbf{p}, \lambda)$ and $W^\lambda(p)$ for the weight operator $W^\lambda(c, \epsilon, p)$. For $c < 0$ we write $f(\epsilon, \mathbf{p}, \lambda)$ for $f(c, \epsilon, \mathbf{p}, \lambda)$ and $W^\lambda(\epsilon, p)$ for $W^\lambda(c, \epsilon, p)$. For any value of c we replace the measure functions by the measure function $M(\lambda)$.

The inner product (4) is now written

$$(f^{(1)}, f)$$
$$= \iint f^{(1)*}(\mathbf{p}, \lambda)W^\lambda(p)f(\mathbf{p}, \lambda)\frac{d\mathbf{p}}{\omega(c, p)}\,dM(\lambda), \quad \text{for } c \geq 0,$$
$$= \sum_\epsilon \iint f^{(1)*}(\epsilon, \mathbf{p}, \lambda)W^\lambda(\epsilon, p)f(\epsilon, \mathbf{p}, \lambda)\frac{d\mathbf{p}}{\omega(c, p)}\,dM(\lambda), \quad \text{for } c < 0. \tag{10}$$

Up to now we have not required $\mathfrak{J}_i$ to be Hermitian. We wish now to examine the consequences of adding this requirement. We shall, however, not require that the representations be irreducible. The Hermiticity and irreducibility requirements can be combined in an obvious fashion to characterize irreducible Hermitian sets of infinitesimal generators, which are the realizations desired when discussing properties of particles.

Let us define the operators $\mathfrak{L}_i$, $i = 1, 2, 3$, as they act on our realizations by

$$\mathfrak{L}_i = i\epsilon\omega(c, p)\nabla_i. \tag{11}$$

The operators $\mathfrak{L}_i$ are the "orbital" parts of the operators $\mathfrak{J}_i$ [see Eq. (7)]. The following theorems hold:

Theorem: The operators $\mathfrak{L}_i$ are Hermitian if and only if the weight operators $W^\lambda(c, \epsilon, p)$ are independent of p for $c \geq 0$ and are independent of p and ϵ for $c < 0$.

Theorem: The operators $\mathfrak{L}_i$ are Hermitian if and only if the realization of the infinitesimal generators of the Poincaré group is equivalent to one in which the weight operators are the identity.

Theorem: The operators $\mathfrak{J}_i$ are Hermitian if and only if the operators $\mathfrak{L}_i$ and realizations of the little groups T_i^λ and M^λ are Hermitian.

Theorem: The operators $\mathfrak{J}_i$ are Hermitian if and only if the realization is unitarily equivalent to a realization in which the weight operators $W^\lambda(c, \epsilon, p)$ are the identity operator and in which the operators of realizations of the little group T_i^λ and M^λ are Hermitian.

From the last theorem we see that, within unitary equivalence, all Hermitian realizations of the Poincaré group can be constructed in the following way. One picks the values of c which one wishes to appear in the representation. The values of c can be in a continuous set, be a discrete set, or be a combination. Those values which appear can be incorporated into a measure function to be given shortly. For each value of $c > 0$ and for each value of ϵ one picks, a Hermitian realization of the rotation group T_i^λ and M^λ such that the corresponding weight operator W^λ is the identity. The values of λ which appear and the values of $c > 0$ which enter into the representation are characterized by the measure function $M^+(c, \epsilon, \lambda)$. The representations for $c = 0$ and $c < 0$ are treated in an analogous way except that for $c < 0$ the realizations of the little group are chosen to be independent of ϵ.

3. DIRECT PRODUCT OF HELICITY REPRESENTATIONS

We give the direct product of two helicity representations. In the next section we expand this direct product into standard helicity representations.

The infinitesimal generators of the factors of the direct product will be distinguished by a superscript (1) or (2). Thus, we have two sets of infinitesimal generators $P^{(1)\alpha}$, $J_i^{(1)}$, $\mathfrak{J}_i^{(1)}$, and $P^{(2)\alpha}$, $J_i^{(2)}$, $\mathfrak{J}_i^{(1)}$. The square of the mass, sign of energy, and realizations of the little group analogous to c, ϵ, T_i^λ, and M^λ, respectively, of the previous section are denoted by b, β, T_i^μ, and M^μ, respectively, where we now use the variable μ instead of λ for representations associated with the superscript (1). Moreover, the variable $\mathbf{p}$ which gives the spectrum of the linear-momentum operators in Sec. 2 is replaced by $\mathbf{y}$ for the superscript-(1) representations. Likewise for the superscript-(2) representations, the quantities analogous to c, ϵ, T_i^λ, M^λ, and λ are denoted by d, γ, T_i^ν, M^ν, and ν, respectively. The variable $\mathbf{p}$ is replaced by $\mathbf{z}$.

For simplicity we consider the direct product of two single-mass representations, that is, representations in which b and d take on only one value each, this value being any real number. In much of our notation we may thus suppress the appearance of b and d. The generalization to the case where b and d may run through a set of values will be obvious.

We also require that the realizations of the little groups are independent of the sign of the energy of the respective representations which appear in the product, not only for imaginary mass representations, as before, but also even when the mass is real. This requirement is automatically satisfied for representations obtained by expanding wavefunctions in configuration space in terms of the standard helicity representations, as will be shown in later papers. However, more generally, this requirement is not really restricting because for nonimaginary masses the realizations are reducible in terms of the signs of the energy, that is, for each realization the space of functions is a sum of orthogonal subspaces in each of which the energy has only one sign, that is, in each subspace the wavefunctions are identically zero when the sign of the energy is plus or minus, respectively. Then in each subspace the realization of the little group may be considered as being independent of the sign of the energy. This device enables us to use a uniform notation for reducing the direct product of representations of all masses, whether real or imaginary.

The space of functions upon which the infinitesimal generators involved in the inner product act is denoted by $\{f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\}$, where the variable $\mathbf{y}$ ranges over the entire space if $b \geq 0$ and over the space $y > -b$ if $b < 0$. (We use the notation $y = |\mathbf{y}|$.) Similarly, the range of $\mathbf{z}$ is the entire vector space if $d > 0$ and the space $z = |\mathbf{z}| > -d$ if $d < 0$.

To define the inner product, we introduce a measure function $m(\mu, \nu)$ and a weight operator $W^{\mu\nu}(\beta, y; \gamma, z)$, where the superscripts μ and ν indicate that the weight operators act on these variables in the functions f. Then the inner product is given by

$$(g, f) = \sum_{\beta,\gamma} \iiint g^*(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) W^{\mu\nu}(\beta, y; \gamma, z) \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) \frac{d\mathbf{y}}{\omega(b, y)} \frac{d\mathbf{z}}{\omega(d, z)} dm(\mu, \nu). \tag{12}$$

The way that the infinitesimal generators act on the functions $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ is now given. We define $B(c)$ as in Eq. (6). Furthermore,

$$y = |\mathbf{y}|, \quad z = |\mathbf{z}|, \quad \omega(b, y) = (b + y^2)^{\frac{1}{2}}, \quad \omega(d, z) = (d + z^2)^{\frac{1}{2}}, \tag{13}$$

as above. Also,

$$\nabla_y = (\nabla_{y1}, \nabla_{y2}, \nabla_{y3}), \quad \nabla_z = (\nabla_{z1}, \nabla_{z2}, \nabla_{z3}), \tag{14a}$$

with

$$\nabla_{yi} = \frac{\partial}{\partial y_i}, \quad \nabla_{zi} = \frac{\partial}{\partial z_i}. \tag{14b}$$

Finally,

$$\mathbf{y} \cdot \mathbf{T}^\mu = y_1 T_1^\mu + y_2 T_2^\mu, \quad \mathbf{z} \cdot \mathbf{T}^\nu = z_1 T_1^\nu + z_2 T_2^\nu. \tag{15}$$

In accordance with our assumption that the realizations of the little groups are independent of the sign of the energy, we take T_i^μ, M^μ, T_i^ν, and M^ν to be independent of β and γ. We have

$$\begin{aligned}
&P_i^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) = y_i f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&H^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad \equiv P^{(1)0} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \beta\omega(b, y) f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&J_3^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = [-i(\mathbf{y} \times \nabla_y)_3 + M^\mu] f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&J_i^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \{-i(\mathbf{y} \times \nabla_y)_i + [y_i/(y + y_3)]M^\mu\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu), \quad i = 1, 2,\\
&\mathfrak{J}_3^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \beta\{i\omega(b, y)\nabla_{y3} + [B(b)/y^2](\mathbf{y} \cdot \mathbf{T}^\mu)\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&\mathfrak{J}_1^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \beta\{i\omega(b, y)\nabla_{y1} + [y_2/y(y + y_3)]\omega(b, y)M^\mu\\
&\qquad + [B(b)/y^2][y_1\mathbf{y} \cdot \mathbf{T}^\mu/(y + y_3) - yT_1^\mu]\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&\mathfrak{J}_2^{(1)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \beta\{i\omega(b, y)\nabla_{y2} - [y_1/y(y + y_3)]\omega(b, y)M^\mu\\
&\qquad + [B(b)/y^2][y_2\mathbf{y} \cdot \mathbf{T}^\mu/(y + y_3) - yT_2^\mu]\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&P_i^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) = z_i f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&H^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad \equiv P^{(2)0} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \gamma\omega(d, z) f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&J_3^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = [-i(\mathbf{z} \times \nabla_z)_3 + M^\nu] f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&J_i^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \{-i(\mathbf{z} \times \nabla_z)_i + [z_i/(z + z_3)]M^\nu\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu), \quad i = 1, 2,\\
&\mathfrak{J}_3^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \gamma\{i\omega(d, z)\nabla_{z3} + [B(d)/z^2](\mathbf{z} \cdot \mathbf{T}^\nu)\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&\mathfrak{J}_1^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \gamma\{i\omega(d, z)\nabla_{z1} + [z_2/z(z + z_3)]\omega(d, z)M^\nu\\
&\qquad + [B(d)/z^2][z_1\mathbf{z} \cdot \mathbf{T}^\nu/(z + z_3) - zT_1^\nu]\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu),\\
&\mathfrak{J}_2^{(2)} f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\\
&\quad = \gamma\{i\omega(d, z)\nabla_{z2} - [z_1/z(z + z_3)]\omega(d, z)M^\nu\\
&\qquad + [B(d)/z^2][z_2\mathbf{z} \cdot \mathbf{T}^\nu/(z + z_3) - zT_2^\nu]\}\\
&\qquad \times f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu).
\end{aligned} \tag{16}$$

One can define unitary equivalence of two direct products of helicity representations in a manner analogous to the definition of unitary equivalence of two helicity representations as in Sec. 2. The following theorem is analogous to one in Sec. 2.

Theorem: $\mathfrak{J}_i^{(1)}$ and $\mathfrak{J}_i^{(2)}$ are both Hermitian if and only if the realizations of the little groups $T_i^\mu M^\mu$ and T_i^ν and M^ν are Hermitian and there is an unitarily equivalent direct product such that the weight operator corresponding to $W^{\mu\nu}$ is the identity.

Let us define the operators P^α, J_i, and $\mathfrak{J}_i$ by

$$\begin{aligned}
P^\alpha &= P^{(1)\alpha} + P^{(2)\alpha}, \quad \alpha = 0, 1, 2, 3,\\
J_i &= J_i^{(1)} + J_i^{(2)}, \quad i = 1, 2, 3,\\
\mathfrak{J}_i &= \mathfrak{J}_i^{(1)} + \mathfrak{J}_i^{(2)}, \quad i = 1, 2, 3.
\end{aligned} \tag{17}$$

In the next section we give a set of functions which are obtained through a one-to-one linear transformation on the functions $\{f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\}$ such that in terms of the new functions the operators act in a helicity representation as in Eq. (7).

4. FORMULAS FOR THE REDUCTION OF THE DIRECT PRODUCT

In the present section we give the explicit formulas for the reduction of the direct product. Specifically, we give a one-to-one linear transformation between the functions of the space $\{f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)\}$ and functions of a Hilbert space $\{F(c, \epsilon, \mathbf{p}, \lambda)\}$ such that, if A is any of the operators P^α, J_i, or $\mathfrak{J}_i$ as it operates on the functions $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ as in Sec. 3, then there is a corresponding operator $\hat{A}$ which operates in the standard helicity representation on the functions

$F(c, \epsilon, \mathbf{p}, \lambda)$ as in Eq. (7) in such a way that, if

$$f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) \leftrightarrow F(c, \epsilon, \mathbf{p}, \lambda),$$

then

$$Af(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) \leftrightarrow \hat{A}F(c, \epsilon, \mathbf{p}, \lambda). \quad (18)$$

We have to give the range of the variable c, identify the variable λ, and give the little group generators $\hat{T}_i$ and $\hat{M}$ as they act on λ in F. (By $\hat{T}_i$ and $\hat{M}$, we designate the generators of the little groups which appear in the standard helicity representations of the operators $\hat{J}_i$ and $\hat{\mathcal{J}}_i$.)

Without any loss in generality we take

$$b \geq d. \quad (19)$$

Then in the correspondence (18) we have the following results:

(i) The range of c is the entire negative real axis and the following portions of the positive real axis:

$$0 < c < \infty, \quad \text{if} \quad d \leq 0,$$
$$0 < c < (b^{\frac{1}{2}} - d^{\frac{1}{2}})^2, \quad (b^{\frac{1}{2}} + d^{\frac{1}{2}})^2 < c < \infty, \quad \text{if} \quad d \geq 0.$$

(ii) The contribution of $c = 0$ (that is, zero mass) is of zero measure and, hence, is not important in the reduction.

(iii) The nature of λ and the representations of the generators little groups $\hat{T}_i$ and $\hat{M}$ depend on the sign of c but are otherwise independent of the value of c.

It is convenient to divide the discussion of the character of λ and the relation of $F(c, \epsilon, \mathbf{p}, \lambda)$ to $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ into two parts corresponding to the sign of c.

A. $c > 0$

The variable λ consists of a set of variables which we denote by θ, φ, μ, and ν. The variables μ and ν are identical in range and character to the variables which appear in the functions $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$. The variables θ and φ are each real variables whose ranges are $0 < \theta < \pi$ and $0 < \varphi < 2\pi$, respectively, and are in the nature of the angular variables used in polar coordinates. For $c > 0$ we define $F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu)$ by

$$F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu) \equiv F(c, \epsilon, \mathbf{p}, \lambda). \quad (20)$$

We define the operators R_i by

$$R_1F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu) = \left[i\left(\sin\varphi \frac{\partial}{\partial\theta} + \cot\theta \cos\varphi \frac{\partial}{\partial\varphi}\right) + \cos\varphi(\tan\tfrac{1}{2}\theta \cdot M^\mu - \cot\tfrac{1}{2}\theta \cdot M^\nu)\right] \times F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu),$$

$$R_2F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu) = \left[-i\left(\cos\varphi \frac{\partial}{\partial\theta} - \cot\theta \sin\varphi \frac{\partial}{\partial\varphi}\right) + \sin\varphi(\tan\tfrac{1}{2}\theta \cdot M^\mu - \cot\tfrac{1}{2}\theta \cdot M^\nu)\right] \times F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu),$$

$$R_3F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu) = \left(-i\frac{\partial}{\partial\varphi} + M^\mu + M^\nu\right) \times F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu). \quad (21)$$

In Eq. (21) and later the operators T_i^μ, M^μ, T_i^ν, and M^ν are the same infinitesimal generators of the little groups associated with the direct product and used in Eq. (16). They act on the μ and ν variables as is indicated by the superscripts.

It is easily seen that the operators R_i satisfy the commutation rules for the infinitesimal generators of the rotation group. Then the infinitesimal generators of the little group $\hat{T}_i$ and $\hat{M}$ are given by

$$\hat{T}_1 = -R_2, \quad \hat{T}_2 = R_1, \quad \hat{M} = R_3. \quad (22)$$

The reduction of the direct product can be carried on still further by introducing a transformation in the λ variable such that $\hat{T}_i$ and $\hat{M}$ or, equivalently, the set of operators R_i are completely reduced. This reduction procedure is closely related to the procedure for reducing the set of angular-momentum operators in Eq. (7), which we shall call "a helicity representation" of the angular momentum. Both the reduction of the helicity representation of the angular momentum operators and the reduction of the operators R_i are discussed in Appendix B.

It will now be useful to introduce a "covariant" notation for some of the arguments appearing in the functions $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$. Let $y^i = y_i$, $i = 1, 2, 3$, be components of the vector $\mathbf{y}$. Similarly, we denote the components of $\mathbf{z}$ by $z^i = z_i$. We define $y^0 = -y_0$ and $z^0 = -z_0$ by

$$y^0 = -y_0 = \beta\omega(b, y), \quad z^0 = -z_0 = \gamma\omega(d, z). \quad (23)$$

Furthermore, we define the 4-vectors w^α and v^α by

$$w^\alpha = y^\alpha + z^\alpha, \quad v^\alpha = y^\alpha - z^\alpha, \quad \alpha = 0, 1, 2, 3. \quad (24a)$$

We define the vectors $\mathbf{w}$ and $\mathbf{v}$ as the space parts of the 4-vectors w^α and v^α respectively; that is,

$$\mathbf{w} = \mathbf{y} + \mathbf{z}, \quad \mathbf{v} = \mathbf{y} - \mathbf{z}, \quad w = |\mathbf{w}|, \quad v = |\mathbf{v}|. \quad (24b)$$

We can introduce a space for the set of independent variables $\mathbf{y}$, $\mathbf{z}$, β, γ which occur in the function

$f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$. We call this space S. That portion of the space S for which $w^\alpha w_\alpha < 0$ is called S_R where the subscript R stands for "real," since the function $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ determines F for real mass. Similarly, the part of S for which $w^\alpha w_\alpha > 0$ is denoted by S_I, where I stands for "imaginary," since for these values of the arguments, $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ determines F for imaginary mass.

For brevity, we denote the variables $\mathbf{y}$, $\mathbf{z}$, β, and γ collectively by Y. Thus, if $w^\alpha w_\alpha < 0$, then Y is in S_R. If $w^\alpha w_\alpha > 0$, then Y is in S_I.

We wish now to introduce a new set of variable which are functions of Y for the case Y is in S_R. We define the variables c, ϵ, $\mathbf{p}$, $\mathbf{r}$, θ, and φ by

$$c = -w^\alpha w_\alpha, \quad \epsilon = \operatorname{sgn} w^0, \quad \mathbf{p} = \mathbf{w},$$
$$r_i = v_i - w_i \mathbf{v}\cdot\mathbf{w}/w(w + w_3) - w_i v_3/(w + w_3), \quad i = 1, 2, \tag{25}$$
$$r_3 = (\operatorname{sgn} w^0)(-w^\alpha w_\alpha)^{-\frac{1}{2}} w^{-1}[w^0(d - b) - v^0 w^\alpha w_\alpha],$$
$$\theta = \arccos(r_3/r), \quad \varphi = \arctan(r_2/r_1), \quad r = |\mathbf{r}|.$$

In Eq. (25) we take the ranges of θ and φ to be $0 < \theta < \pi$ and $0 < \varphi < 2\pi$, respectively. The variable c is, of course, positive, and, in fact, has the same range as the argument c in $F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu)$. Each of the components of $\mathbf{p}$ covers the entire real axis.

We regard the set of variables c, ϵ, $\mathbf{p}$, θ, and φ, which we collectively denote by Q, as being the image of the point Y. As Y ranges through the space S_R, Q will go through the space S_R'.

Having given Q as a function of Y, we can give the inverse transformation, namely Y as a function of Q.

We first define the vector $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)$, the vector $\mathbf{k} = (k_1, k_2, k_3)$, and the functions s_1 and s_2 in terms of Q by

$$\boldsymbol{\tau} = \{[(c - b - d)^2 - 4bd]/c\}^{\frac{1}{2}} \times (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta),$$
$$\tau = |\boldsymbol{\tau}| = \{[(c - b - d)^2 - 4bd]/c\}^{\frac{1}{2}},$$
$$k_i = \left[\tau_i - \frac{p_i(p_1\tau_1 + p_2\tau_2)}{p(p + p_3)} + \frac{p_i}{p}\frac{\tau_3\omega(c, p)}{c^{\frac{1}{2}}} + p_i\frac{b - d}{c}\right], \quad i = 1, 2, \tag{26}$$
$$k_3 = -\frac{(p_1\tau_1 + p_2\tau_2)}{p} + \frac{p_3}{p}\frac{\tau_3\omega(c, p)}{c^{\frac{1}{2}}} + p_3\frac{b - d}{c},$$
$$s_1 = (c + b - d)\omega(c, p) + \tau_3 c^{\frac{1}{2}} p,$$
$$s_2 = (c - b + d)\omega(d, p) - \tau_3 c^{\frac{1}{2}} p.$$

The variable Y in terms of the variable Q is then as follows:

$$\mathbf{y} = \tfrac{1}{2}(\mathbf{p} + \mathbf{k}), \quad \mathbf{z} = \tfrac{1}{2}(\mathbf{p} - \mathbf{k}),$$
$$\beta = \epsilon \operatorname{sgn} s_1, \quad \gamma = \epsilon \operatorname{sgn} s_2. \tag{27}$$

We note that, if the value Q in $\boldsymbol{\tau}$ is replaced by its expression in terms of Y (Eq. 25) or if Y in the expressions for $\mathbf{r}$ in Eq. (25) is replaced by its expressions in terms of Q as in Eqs. (27), we have

$$\boldsymbol{\tau} = \mathbf{r}. \tag{28}$$

Similar replacements lead to

$$\mathbf{v} = \mathbf{k}. \tag{29}$$

From the Jacobians of the transformation (25) or its inverse (27) we can show

$$\frac{d\mathbf{y}}{\omega(b, y)}\frac{d\mathbf{z}}{\omega(d, z)} = \frac{1}{4}\frac{\tau}{c^{\frac{1}{2}}}dc\frac{d\mathbf{p}}{\omega(c, p)}\sin\theta\, d\theta\, d\varphi. \tag{30}$$

We now introduce certain operators which act on the μ and ν variables in $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ and $F_+(c, \epsilon, \mathbf{p}, \theta, \varphi, \mu, \nu)$.

First, let us define for any two vectors $\mathbf{a}$ and $\mathbf{b}$ the scalar function $f(\mathbf{a}, \mathbf{b})$:

$$f(\mathbf{a}, \mathbf{b}) = (a + a_3)(b + b_3) + (a_1 - ia_2)(b_1 + ib_2), \tag{31}$$

where

$$a = |\mathbf{a}|, \quad b = |\mathbf{b}|. \tag{31'}$$

We note

$$|f(\mathbf{a}, \mathbf{b})| = [2(a + a_3)(b + b_3)(ab + \mathbf{a}\cdot\mathbf{b})]^{\frac{1}{2}}. \tag{32}$$

Then we define

$$R_\mu(\mathbf{y}, \mathbf{w}) = [f(\mathbf{y}, \mathbf{w})/|f(\mathbf{y}, \mathbf{w})|]^{2M^\mu},$$
$$R_\nu(\mathbf{z}, \mathbf{w}) = [f(\mathbf{z}, \mathbf{w})/|f(\mathbf{z}, \mathbf{w})|]^{2M^\nu}. \tag{33}$$

Also, for simplicity,

$$M = (-w^\alpha w_\alpha)^{\frac{1}{2}}. \tag{34}$$

For $w^\alpha w_\alpha < 0$, we define

$$W_1(\mathbf{y}, \mathbf{z}) = \exp\left\{i\left(\frac{\mathbf{r}\cdot\mathbf{T}^\mu}{2}\frac{w}{|\mathbf{y}\times\mathbf{z}|}\left[g\left(b, \frac{\mathbf{y}\cdot\mathbf{w}}{|y^0[\mathbf{y}\times\mathbf{z}]|}\right) - g\left(b, \frac{Mwr_3}{|[M^2 + b - d][\mathbf{y}\times\mathbf{z}]|}\right)\right]\right)\right\},$$
$$W_2(\mathbf{y}, \mathbf{z}) = \exp\left\{-i\left(\frac{\mathbf{r}\cdot\mathbf{T}^\nu}{2}\frac{w}{|\mathbf{y}\times\mathbf{z}|}\left[g\left(d, \frac{\mathbf{z}\cdot\mathbf{w}}{|z^0[\mathbf{y}\times\mathbf{z}]|}\right) - g\left(d, -\frac{Mwr_3}{|[M^2 + d - b][\mathbf{y}\times\mathbf{z}]|}\right)\right]\right)\right\}, \tag{35}$$

where

$$\mathbf{r}\cdot\mathbf{T}^\mu = r_1 T_1^\mu + r_2 T_2^\mu, \quad \mathbf{r}\cdot\mathbf{T}^\nu = r_1 T_1^\nu + r_2 T_2^\nu, \tag{35'}$$

and the functions $g(c, x)$ are given in Eq. (9h).

Let us also define

$$S_1(\mathbf{p},\mathbf{k}) = \left[\frac{f(\mathbf{p}+\mathbf{k},\mathbf{p})}{|f(\mathbf{p}+\mathbf{k},\mathbf{p})|}\right]^{2M^\mu},$$

$$S_2(\mathbf{p},\mathbf{k}) = \left[\frac{f(\mathbf{p}-\mathbf{k},\mathbf{p})}{|f(\mathbf{p}-\mathbf{k},\mathbf{p})|}\right]^{2M^\nu}, \tag{36}$$

and

$$V_1(\mathbf{p},\mathbf{k}) = \exp\left\{i\left(\boldsymbol{\tau}\cdot\mathbf{T}^\mu \frac{p}{|\mathbf{p}\times\mathbf{k}|}\left[g\left(b,\frac{2c(p^2+\mathbf{p}\cdot\mathbf{k})}{|s_1[\mathbf{p}\times\mathbf{k}]|}\right) - g\left(b,\frac{2pc^{\frac{1}{2}}\tau_3}{|[c+b-d][\mathbf{p}\times\mathbf{k}]|}\right)\right]\right)\right\},$$

$$V_2(\mathbf{p},\mathbf{k}) = \exp\left\{-i\left(\boldsymbol{\tau}\cdot\mathbf{T}^\nu \frac{p}{|\mathbf{p}\times\mathbf{k}|}\left[g\left(d,\frac{2c(p^2-\mathbf{p}\cdot\mathbf{k})}{|s_2[\mathbf{p}\times\mathbf{k}]|}\right) - g\left(d,-\frac{2pc^{\frac{1}{2}}\tau_3}{|[c-b+d][\mathbf{p}\times\mathbf{k}]|}\right)\right]\right)\right\}, \tag{37}$$

where

$$\boldsymbol{\tau}\cdot\mathbf{T}^\mu = \tau_1 T_1^\mu + \tau_2 T_2^\mu, \quad \boldsymbol{\tau}\cdot\mathbf{T}^\nu = \tau_1 T_1^\nu + \tau_2 T_2^\nu. \tag{37'}$$

We note that in the above operators if the variables Q are replaced by the variables Y or Y by Q in accordance with the transformations (25) and (27), then

$$S_1(\mathbf{p},\mathbf{k}) = R_\mu(\mathbf{y},\mathbf{w}), \quad S_2(\mathbf{p},\mathbf{k}) = R_\nu(\mathbf{z},\mathbf{w}), \quad V_i(\mathbf{p},\mathbf{k}) = W_i(\mathbf{y},\mathbf{z}). \tag{38}$$

We can now give $F_+(c,\epsilon,\mathbf{p},\theta,\varphi,\mu,\nu)$ in terms of $f(\beta,\mathbf{y},\mu;\gamma,\mathbf{z},\nu)$ for Y in S_R:

$$F_+(c,\epsilon,\mathbf{p},\theta,\varphi,\mu,\nu) = V_1^{-1}(\mathbf{p},\mathbf{k})S_1^{-1}(\mathbf{p},\mathbf{k})V_2^{-1}(\mathbf{p},\mathbf{k})S_2^{-1}(\mathbf{p},\mathbf{k}) \times f(\epsilon\,\mathrm{sgn}\,s_1, \tfrac{1}{2}(\mathbf{p}+\mathbf{k}),\mu;\epsilon\,\mathrm{sgn}\,s_2,\tfrac{1}{2}(\mathbf{p}-\mathbf{k}),\nu). \tag{39}$$

The inverse of the relation [Eq. (39)] is

$$f(\beta,\mathbf{y},\mu;\gamma,\mathbf{z},\nu) = R_\mu(\mathbf{y},\mathbf{w})W_1(\mathbf{y},\mathbf{z})R_\nu(\mathbf{z},\mathbf{w})W_2(\mathbf{y},\mathbf{z}) \times F_+\left(M^2,\mathrm{sgn}\,w^0,\mathbf{w},\arccos\frac{r_3}{r},\arctan\frac{r_2}{r_1},\mu,\nu\right). \tag{40}$$

B. $c<0$

For $c<0$ the variable λ in $F(c,\epsilon,\mathbf{p},\lambda)$ consists of two continuous variables q_1 and q_2, whose range we give below; a discrete variable σ which takes on the values $+1$ or -1; and μ and ν whose range and character are the same as for the variables μ and ν in the function $f(\beta,\mathbf{y},\mu;\gamma,\mathbf{z},\nu)$. We shall use $\mathbf{q}$ to denote the 2-dimensional vector $\mathbf{q}=(q_1,q_2)$. We shall now define $F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu)$ by

$$F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu) \equiv F(c,\epsilon,\mathbf{p},\lambda), \quad \text{for} \quad c<0. \tag{41}$$

We shall now give the range of $\mathbf{q}$. We define K by

$$K = \frac{4bd-(c-b-d)^2}{c}. \tag{42}$$

Then, for values of c such that $K>0$, the components of $\mathbf{q}$ cover the entire (q_1,q_2) plane. For values of c such that $K<0$, the components of $\mathbf{q}$ cover the entire (q_1,q_2) plane such that $q>-K$ where

$$q = |\mathbf{q}| = (q_1^2+q_2^2)^{\frac{1}{2}}. \tag{43}$$

We shall now give $\hat{T}_i$ and $\hat{M}$. Let us define

$$K_+ = -(c+b-d)^2/c, \quad K_- = -(c+d-b)^2/c. \tag{44}$$

Also

$$\kappa = (-c)^{\frac{1}{2}}. \tag{45}$$

We then define the operators R_i, $i=1,2,3$, by

$$R_1F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu) = \sigma\Bigg[i\omega(K,q)\frac{\partial}{\partial q_1} + \frac{\kappa q_2\omega(K,q)}{\omega(K_+,q)[\kappa\omega(K_+,q)-c-b+d]}M^\mu + \frac{\kappa q_2\omega(K,q)}{\omega(K_-,q)[\kappa\omega(K_-,q)-c-d+b]}M^\nu + 2\frac{B(b)}{[\omega(K_+,q)]^2}\left(\frac{\kappa q_1}{[\kappa\omega(K_+,q)-c-b+d]}\mathbf{q}\cdot\mathbf{T}^\mu - \omega(K_+,q)T_1^\mu\right) - 2\frac{B(d)}{[\omega(K_-,q)]^2}\left(\frac{\kappa q_1}{\kappa\omega(K_-,q)-c-d+b}\mathbf{q}\cdot\mathbf{T}^\nu - \omega(K_-,q)T_1^\nu\right)\Bigg]F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu),$$

$$R_2F(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu) = \sigma\Bigg[i\omega(K,q)\frac{\partial}{\partial q_2} - \frac{\kappa q_1\omega(K,q)}{\omega(K_+,q)[\kappa\omega(K_+,q)-c-b+d]}M^\mu - \frac{\kappa q_1\omega(K,q)}{\omega(K_-,q)[\kappa\omega(K_-,q)-c-d+b]}M^\nu + 2\frac{B(b)}{[\omega(K_+,q)]^2}\left(\frac{\kappa q_2}{\kappa\omega(K_+,q)-c-b+d}\mathbf{q}\cdot\mathbf{T}^\mu - \omega(K_+,q)T_2^\mu\right) - 2\frac{B(d)}{[\omega(K_-,q)]^2}\left(\frac{\kappa q_2}{\kappa\omega(K_-,q)-c-d+b}\mathbf{q}\cdot\mathbf{T}^\nu - \omega(K_-,q)T_2^\nu\right)\Bigg]F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu),$$

$$R_3F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu) = \left[-i\left(q_1\frac{\partial}{\partial q_2} - q_2\frac{\partial}{\partial q_1}\right) + M^\mu + M^\nu\right]F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu), \tag{46}$$

where

$$\mathbf{q}\cdot\mathbf{T}^{\mu} = q_1T_1^{\mu} + q_2T_2^{\mu},$$
$$\mathbf{q}\cdot\mathbf{T}^{\nu} = q_1T_1^{\nu} + q_2T_2^{\nu}. \tag{46'}$$

Then

$$\hat{T}_1 = -\epsilon R_1, \quad \hat{T}_2 = -\epsilon R_2, \quad \hat{M} = R_3. \tag{47}$$

In Eq. (47) it is seen that $\hat{T}_i$ depends on ϵ, whereas in Sec. 2 it was stated that these operators could be chosen to be independent of the sign of the energy. In Appendix A it is shown that by changing the phase of F_-, the operators are indeed independent of ϵ in the new representation.

For reasons similar to those used for the introduction of the variables Q when Y is in S_R, we shall introduce a set of variables, which we shall also call Q, when Y is in S_I. As Y ranges through the space S_I, Q will range through an image space S_-'. Specifically, the variables summarized by the symbol Q, for Y in S_I, are c, ϵ, $\mathbf{p}$, $\mathbf{q}$, σ, where

$$c = -w^{\alpha}w_{\alpha}, \quad \mathbf{p} = \mathbf{w}, \quad \epsilon = \operatorname{sgn} w^0, \quad q_i = r_i, \quad i = 1, 2,$$
$$\sigma = \operatorname{sgn} S, \tag{48}$$

where the functions r_i are given in Eq. (25) and S is closely related to r_3 of Eq. (25):

$$S = [v^0w^{\alpha}w_{\alpha} - w^0(d - b)]. \tag{48'}$$

The ranges of the variables c, $\mathbf{p}$, ϵ, $\mathbf{q} = (q_1, q_2)$, σ are identical to the variables of the same name in $F_-(c, \epsilon, \mathbf{p}, \mathbf{q}, \sigma, \mu, \nu)$.

We can also find Y from Q. Let us define

$$d_1 = p\omega(K, q) - \epsilon\sigma[(c + b - d)/(-c)^{\frac{1}{2}}]\omega(c, p),$$
$$d_2 = p\omega(K, q) + \epsilon\sigma[(c + d - b)/(-c)^{\frac{1}{2}}]\omega(c, p), \tag{49}$$

and the vector $\mathbf{k} = (k_1, k_2, k_3)$ by

$$k_i = q_i - p_i\frac{\mathbf{p}\cdot\mathbf{q}}{p(p + p_3)} - p_i\frac{d - b}{c} + \epsilon\sigma\frac{p_i}{p}\frac{\omega(K, q)\omega(c, p)}{(-c)^{\frac{1}{2}}}, \quad i = 1, 2,$$
$$k_3 = -\frac{\mathbf{p}\cdot\mathbf{q}}{p} - p_3\frac{d - b}{c} + \epsilon\sigma\frac{p_3}{p}\frac{\omega(K, q)\omega(c, p)}{(-c)^{\frac{1}{2}}}, \tag{50}$$

where

$$\mathbf{p}\cdot\mathbf{q} = p_1q_1 + p_2q_2. \tag{50'}$$

Then the expression for Y in terms of Q is

$$\mathbf{y} = \tfrac{1}{2}(\mathbf{p} + \mathbf{k}), \quad z = \tfrac{1}{2}(\mathbf{p} - \mathbf{k}),$$
$$\beta = \sigma \operatorname{sgn} d_1, \quad \gamma = -\sigma \operatorname{sgn} d_2. \tag{51}$$

We note that, when Q is expressed in terms of Y or Y in terms of Q,

$$\mathbf{v} = \mathbf{k}, \quad |S| = (-c)^{\frac{1}{2}}p\omega(K, q), \quad q_i = r_i, \quad i = 1, 2. \tag{52}$$

Using the Jacobian of the transformations, we obtain

$$\frac{dy}{\omega(b, y)}\frac{dz}{\omega(d, z)} = \frac{1}{4}\frac{dq_1\,dq_2\,d\mathbf{p}\,dc}{\kappa\omega(K, q)\omega(c, p)}. \tag{53}$$

We now introduce operators in the variable μ, ν which enables us to relate $F_-(c, \epsilon, \mathbf{p}, \mathbf{q}, \sigma, \mu, \nu)$ to $f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu)$ when Y is in S_I.

First we define $R_{\mu}(\mathbf{y}, \mathbf{w})$, $R_{\nu}(\mathbf{z}, \mathbf{w})$, and $S_i(\mathbf{p}, \mathbf{k})$ for $i = 1, 2$, as in Eq. (33) and Eq. (36), using, however, Eq. (50) for $\mathbf{k}$ instead of Eq. (26). We also define W_i and V_i, $i = 1, 2$, by

$$W_1(\mathbf{y}, \mathbf{z}) = \exp\left\{i\left(\frac{\mathbf{r}\cdot\mathbf{T}^{\mu}}{2}\frac{w}{|\mathbf{y}\times\mathbf{z}|}\left[g\left(b, \frac{\mathbf{y}\cdot\mathbf{w}}{|y^0[\mathbf{y}\times\mathbf{z}]|}\right) - g\left(b, -\frac{w^2[c + b - d]}{2\,|S[\mathbf{y}\times\mathbf{z}]|}\right)\right]\right)\right\},$$
$$W_2(\mathbf{y}, \mathbf{z}) = \exp\left\{-i\left(\frac{\mathbf{r}\cdot\mathbf{T}^{\nu}}{2}\frac{w}{|\mathbf{y}\times\mathbf{z}|}\left[g\left(d, \frac{\mathbf{z}\cdot\mathbf{w}}{|z^0[\mathbf{y}\times\mathbf{z}]|}\right) - g\left(d, -\frac{w^2[c + d - b]}{2\,|S[\mathbf{y}\times\mathbf{z}]|}\right)\right]\right)\right\},$$
$$V_1(\mathbf{p}, \mathbf{k}) = \exp\left\{i\left(\mathbf{q}\cdot\mathbf{T}^{\mu}\frac{p}{|\mathbf{p}\times\mathbf{k}|}\left[g\left(b, \frac{2\kappa[p^2 + \mathbf{p}\cdot\mathbf{k}]}{|d_1[\mathbf{p}\times\mathbf{k}]|}\right) - g\left(b, -\frac{2p[c + b - d]}{|\kappa\omega(K, q)[\mathbf{p}\times\mathbf{k}]|}\right)\right]\right)\right\}, \tag{54}$$
$$V_2(\mathbf{p}, \mathbf{k}) = \exp\left\{-i\left(\mathbf{q}\cdot\mathbf{T}^{\nu}\frac{p}{|\mathbf{p}\times\mathbf{k}|}\left[g\left(d, \frac{2\kappa[p^2 - \mathbf{p}\cdot\mathbf{k}]}{|d_2[\mathbf{p}\times\mathbf{k}]|}\right) - g\left(d, -\frac{2p[c + d - b]}{|\kappa\omega(K, q)[\mathbf{p}\times\mathbf{k}]|}\right)\right]\right)\right\},$$

where $\mathbf{r}\cdot\mathbf{T}^{\mu}$, $\mathbf{r}\cdot\mathbf{T}^{\nu}$, $\mathbf{q}\cdot\mathbf{T}^{\mu}$, and $\mathbf{q}\cdot\mathbf{T}^{\nu}$ are defined in Eq. (35') and Eq. (46').

We note that when Q is expressed in terms of Y or Y in terms of Q, then

$$S_1(\mathbf{p}, \mathbf{k}) = R_{\mu}(\mathbf{y}, \mathbf{w}), \quad S_2(\mathbf{p}, \mathbf{k}) = R_{\nu}(\mathbf{z}, \mathbf{w}), \quad V_i(\mathbf{p}, \mathbf{k}) = W_i(\mathbf{y}, \mathbf{z}). \tag{55}$$

Then the formulas analogous to Eq. (39) and Eq. (40) which give the relations between F_- and f when Y is in S_I and Q is in S_I' are

$$F_-(c, \epsilon, \mathbf{p}, \mathbf{q}, \sigma, \mu, \nu) = V_1^{-1}(\mathbf{p}, \mathbf{k})S_1^{-1}(\mathbf{p}, \mathbf{k})V_2^{-1}(\mathbf{p}, \mathbf{k})S_2^{-1}(\mathbf{p}, \mathbf{k}) \times f(\sigma \operatorname{sgn} d_1, \tfrac{1}{2}(\mathbf{p} + \mathbf{k}), \mu; -\sigma \operatorname{sgn} d_2, \tfrac{1}{2}(\mathbf{p} - \mathbf{k}), \nu), \tag{56}$$

$$f(\beta, \mathbf{y}, \mu; \gamma, \mathbf{z}, \nu) = R_{\mu}(\mathbf{y}, \mathbf{w})W_1(\mathbf{y}, \mathbf{z})R_{\nu}(\mathbf{z}, \mathbf{w})W_2(\mathbf{y}, \mathbf{z}) \times F_-(-w^{\alpha}w_{\alpha}, \operatorname{sgn} w^0, \mathbf{w}, \hat{\mathbf{r}}, \operatorname{sgn} S, \mu, \nu), \tag{57}$$

where $\hat{\mathbf{r}}$ is the 2-component vector

$$\hat{\mathbf{r}} = (r_1, r_2), \tag{57'}$$

r_1 and r_2 being given by Eq. (25).

C. The Inner Product

The inner product is given by

$$\begin{aligned}(f^{(1)},f) = &\sum_{\beta,\gamma}\iiint f^{(1)*}(\beta,\mathbf{y},\mu;\gamma,\mathbf{z},\nu)\\ &\times W^{\mu\nu}(\beta,y;\gamma,z)f(\beta,\mathbf{y},\mu;\gamma,\mathbf{z},\nu)\\ &\times\frac{d\mathbf{y}}{\omega(b,y)}\frac{d\mathbf{z}}{\omega(c,z)}\,dm(\mu,\nu)\\ =&\frac{1}{4}\sum_{\epsilon}\int_{R_+}dc\,\frac{\tau}{c^{\frac{1}{2}}}\int\frac{d\mathbf{p}}{\omega(c,p)}\int_0^{\pi}\sin\theta\,d\theta\int_0^{2\pi}d\varphi\\ &\times\int dm(\mu,\nu)F_+^{(1)*}(c,\epsilon,\mathbf{p},\theta,\varphi,\mu,\nu)\\ &\times V_1^{\dagger}(\mathbf{p},\mathbf{k})S_1^{-1}(\mathbf{p},\mathbf{k})V_2^{\dagger}(\mathbf{p},\mathbf{k})S_2^{-1}(\mathbf{p},\mathbf{k})\\ &\times W^{\mu\nu}(\epsilon\,\mathrm{sgn}\,s_1,\tfrac{1}{2}(|\mathbf{p}+\mathbf{k}|);\\ &\quad\epsilon\,\mathrm{sgn}\,s_2,\tfrac{1}{2}(|\mathbf{p}-\mathbf{k}|))S_1(\mathbf{p},\mathbf{k})V_1(\mathbf{p},\mathbf{k})\\ &\times S_2(\mathbf{p},\mathbf{k})V_2(\mathbf{p},\mathbf{k})F_+(c,\epsilon,\mathbf{p},\theta,\varphi,\mu,\nu)\\ &+\frac{1}{4}\sum_{\epsilon,\sigma}\int_{-\infty}^{0}\frac{dc}{\kappa}\int\frac{d\mathbf{p}}{\omega(c,p)}\int\frac{d\mathbf{q}}{\omega(K,q)}\\ &\times\int dm(\mu,\nu)F_-^{(1)*}(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu)\\ &\times V_1^{\dagger}(\mathbf{p},\mathbf{k})S_1^{-1}(\mathbf{p},\mathbf{k})V_2^{\dagger}(\mathbf{p},\mathbf{k})\\ &\times S_2^{-1}(\mathbf{p},\mathbf{k})W^{\mu\nu}(\sigma\,\mathrm{sgn}\,d_1,\tfrac{1}{2}(|\mathbf{p}+\mathbf{k}|);\\ &-\sigma\,\mathrm{sgn}\,d_2,\tfrac{1}{2}(|\mathbf{p}-\mathbf{k}|))S_1(\mathbf{p},\mathbf{k})V_1(\mathbf{p},\mathbf{k})\\ &\times S_2(\mathbf{p},\mathbf{k})_2V(\mathbf{p},\mathbf{k})F_-(c,\epsilon,\mathbf{p},\mathbf{q},\sigma,\mu,\nu),\end{aligned} \tag{58}$$

where the symbol † means Hermitian adjoint, the range R_+ refers to the range of c on the positive real axis given earlier, and $d\mathbf{q} = dq_1\,dq_2$, where the ranges of q_1 and q_2 were given earlier. Also the operators V_i are given by Eq. (37) for $c > 0$ and by Eq. (54) for $c < 0$.

D. A Special Case

All of the cases treated in detail in the earlier papers on the reduction of the direct product of the representations of the Poincaré group consider only the direct products of representations of nonimaginary mass and positive energy. This important special case is, of course, contained in our more general reduction algorithm. It corresponds to the case $b \geq d \geq 0$ such that $f(\beta,\mathbf{y},\mu;\gamma,\mathbf{z},\nu)$ is identically zero for β and/or $\gamma = -1$.

In this case it is readily seen that F_- is identically zero. Other important special cases can also be discussed. We refrain for the sake of brevity.

5. DIRECT PRODUCT OF A HELICITY REPRESENTATION OF THE POINCARÉ GROUP AND A REPRESENTATION OF THE HOMOGENEOUS LORENTZ GROUP

In the present section we shall reduce the direct product of two representations of the Poincaré group, one of which has operators P^α which are not identically zero and is in the standard helicity representations of Sec. 2 and the second of which has operators P^α which are identically zero and hence is a representation of the proper, orthochronous, homogeneous Lorentz group.

Let us first discuss the representations of the homogeneous Lorentz group. These infinitesimal generators satisfy the commutation relations [Eq. (1)]. However,

$$P^\alpha \equiv 0, \quad \alpha = 0, 1, 2, 3. \tag{59}$$

We require the operators J_i to be Hermitian and integrable. However, the operators $\mathfrak{J}_i$ need be only integrable. The finite-dimensional irreducible representations of these operators are discussed, for example, by Lomont.[11] The operators $\mathfrak{J}_i$ are not Hermitian for such representations. The irreducible Hermitian representations of the set of operators J_i and $\mathfrak{J}_i$ are infinite dimensional and are discussed by Naimark.[12]

Let us set up a notation which includes representations of the homogeneous group for all cases of interest. We introduce a Hilbert space of complex functions $\{f(\nu)\}$. The inner product is defined with the aid of a positive-definite weight operator W^ν and a measure function $m(\nu)$ in analogy to the inner product for the standard helicity representations

$$(f^{(1)},f) = \int f^{(1)*}(\nu)W^\nu f(\nu)\,dm(\nu). \tag{60}$$

The superscript on W^ν is used to emphasize that this operator acts on the argument of the function $f(\nu)$. We shall also write

$$J_i f(\nu) = J_i^\nu f(\nu), \quad \mathfrak{J}_i f(\nu) = \mathfrak{J}_i^\nu f(\nu) \tag{61}$$

to indicate that J_i and $\mathfrak{J}_i$ operate with respect to the variable ν in this representation.

We shall now introduce the function space of the direct product which we shall denote by $\{f(\beta,\mathbf{y},\mu;\nu)\}$ in analogy to the function space introduced in Sec. 3. The square of the mass of the standard helicity representation is b which can have any real value. The ranges of the variables β, $\mathbf{y}$, and μ are the same as those

for the variables in the functions of Sec. 3. Furthermore, we shall assume, as in Sec. 3, that the representation of the generators of the little group T_i^μ and M^μ are independent of β. The operators P^α, J_i, and $\mathfrak{J}_i$ acting on the function $f(\beta, \mathbf{y}, \mu; \nu)$ are defined by

$$\begin{aligned} H &= H^{(1)} + H^{(2)}, \quad P_i = P_i^{(1)} + P_i^{(2)}, \\ J_i &= J_i^{(1)} + J_i^{(2)}, \\ \mathfrak{J}_i &= \mathfrak{J}_i^{(1)} + \mathfrak{J}_i^{(2)}, \end{aligned} \tag{62}$$

where $H^{(1)}$, $P_i^{(1)}$, and $\mathfrak{J}_i^{(1)}$ are defined as operators with respect to β, $\mathbf{y}$, and μ as in Eq. (16) and where

$$\begin{aligned} H^{(2)} &\equiv 0, \quad P_i^{(2)} \equiv 0, \\ J_i^{(2)} f(\beta, \mathbf{y}, \mu; \nu) &\equiv J_i^\nu f(\beta, \mathbf{y}, \mu; \nu), \\ \mathfrak{J}_i^{(2)} f(\beta, \mathbf{y}, \mu; \nu) &\equiv \mathfrak{J}_i^\nu f(\beta, \mathbf{y}, \mu; \nu), \end{aligned} \tag{63}$$

where J_i^ν and $\mathfrak{J}_i^\nu$ are operators corresponding to a representation of the homogeneous Lorentz group as described earlier in the present section.

The inner product is described with the aid of a weight operator $W^{\mu\nu}(\beta, y)$ and a measure function $m(\mu, \nu)$:

$$(f^{(1)}, f) = \iint f^{(1)*}(\beta, \mathbf{y}, \mu; \nu) W^{\mu\nu}(\beta, y) \times f(\beta, \mathbf{y}, \mu; \nu) \frac{d\mathbf{y}}{\omega(b, y)} dm(\mu, \nu). \tag{64}$$

We express $f(\beta, \mathbf{y}, \mu; \nu)$ in terms of a function $F(\beta, \mathbf{y}, \mu, \nu)$ such that, if A is any of the operators H, P_i, J_i, $\mathfrak{J}_i$ acting on $f(\beta, \mathbf{y}, \mu, \nu)$, then $Af(\beta, \mathbf{y}, \mu, \nu)$ becomes $\hat{A}F(\beta, \mathbf{y}, \mu, \nu)$ where $\hat{A}$ is in the standard helicity representation of Sec. 2 with the square of the mass given by b, the sign of energy by β, the momentum variable by $\mathbf{y}$, and the little-group variable λ given by the pair of variables (μ, ν).

Explicitly,

$$f(\beta, \mathbf{y}, \mu; \nu) = \{\exp [i\mathbf{\Theta} \cdot \mathbf{J}^\nu] \exp [i\rho \mathfrak{J}_3^\nu]\} F(\beta, \mathbf{y}, \mu, \nu) \tag{65a}$$

and, conversely,

$$F(\beta, \mathbf{y}, \mu, \nu) = \{\exp [-i\rho \mathfrak{J}_3^\nu] \exp [-i\mathbf{\Theta} \cdot \mathbf{J}^\nu]\} \times f(\beta, \mathbf{y}, \mu, \nu), \tag{65b}$$

where ρ and the vector $\mathbf{\Theta}$ are given in terms of y by

$$\begin{aligned} &y = (b)^{\frac{1}{2}} \sinh |\rho|, \quad \beta = \operatorname{sgn} \rho, \quad \text{for} \quad b > 0, \\ &y = e^{\beta\rho}, \quad \text{for} \quad b = 0, \\ &y = (-b)^{\frac{1}{2}} \cosh \rho, \quad \beta = \operatorname{sgn} \rho, \quad \text{for} \quad b < 0, \\ &y_1 = -y(\Theta_2/\Theta) \sin \Theta, \quad y_2 = y(\Theta_1/\Theta) \sin \Theta, \\ &y_3 = y \cos \Theta, \quad \text{with} \quad \mathbf{\Theta} = (\Theta_1, \Theta_2, 0), \\ &\Theta = |\mathbf{\Theta}|. \end{aligned} \tag{65c}$$

We now give the little group generators $\hat{T}_i$ and $\hat{M}$ which act on the μ and ν variables in $F(\beta, \mathbf{y}, \mu; \nu)$ which constitute the variable λ of the standard helicity representation.

Let us define the following operators acting on the variable ν in F:

$$\begin{aligned} &M^\nu = J_3^\nu, \quad \text{for all values of } b, \\ &T_1^\nu = -J_2^\nu, \quad T_2^\nu = J_1^\nu, \quad \text{for} \quad b > 0, \\ &T_1^\nu = \beta \mathfrak{J}_1^\nu - J_2^\nu, \quad T_2^\nu = -\beta \mathfrak{J}_2^\nu + J_1^\nu, \quad \text{for} \quad b = 0, \\ &T_1^\nu = -\beta \mathfrak{J}_i^\nu, \quad \text{for} \quad b < 0. \end{aligned} \tag{66}$$

Then,

$$\hat{M} = M^\mu + M^\nu, \quad \hat{T}_i = T_i^\mu + T_i^\nu. \tag{67}$$

The inner product is given by

$$\begin{aligned} &\iint f^{(1)*}(\beta, \mathbf{y}, \mu; \nu) W^{\mu\nu}(\beta, y) \\ &\quad \times f(\beta, \mathbf{y}, \mu; \nu) \frac{d\mathbf{y}}{\omega(b, y)} dm(\mu, \nu) \\ &= \iint F^{(1)*}(\beta, \mathbf{y}, \mu, \nu) \{\exp [-i\rho \mathfrak{J}_3^{\nu\dagger}] \\ &\quad \times \exp [-i\mathbf{\Theta} \cdot \mathbf{J}^\nu] W^{\mu\nu}(\beta, y) \exp [i\mathbf{\Theta} \cdot \mathbf{J}^\nu] \exp [i\rho \mathfrak{J}_3^\nu]\} \\ &\quad \times F(\beta, \mathbf{y}, \mu, \nu) \frac{d\mathbf{y}}{\omega(b, y)} dm(\mu, \nu). \end{aligned} \tag{68}$$

If $b > 0$, the direct-product representation can be completely reduced by reducing the little group, which is the rotation group. We refrain from details which are almost obvious.

APPENDIX A: THE INDEPENDENCE OF THE REPRESENTATIONS OF THE INFINITESIMAL GENERATORS OF THE LITTLE GROUP ON THE SIGN OF ENERGY FOR IMAGINARY MASS REPRESENTATIONS OF THE POINCARÉ GROUP

In Ref. 1 it is shown that for $c < 0$ one can choose the standard helicity representations such that the operators T_i^λ and M^λ of Eq. (7) are of the form

$$T_i^\lambda = -\epsilon R_i^\lambda, \tag{A1}$$

where the operators R_i^λ, $i = 1, 2$, and M^λ act on the λ variable, are independent of ϵ, and satisfy the commutation rules for the little group generators (Eq. 3).

Let us now choose a new standard helicity representation which consists of a space of functions $\{\hat{f}(c, \epsilon, \mathbf{p}, \lambda)\}$ related to the space of functions $\{f(c, \epsilon, \mathbf{p}, \lambda)\}$ of the original representation for which Eq.

(A1) holds, as follows:

$$\hat{f}(c, +1, \mathbf{p}, \lambda) = f(c, +1, \mathbf{p}, \lambda),$$

$$\hat{f}(c, -1, \mathbf{p}, \lambda) = \exp[i\pi M^{\lambda}] f(c, -1, \mathbf{p}, \lambda). \quad \text{(A2)}$$

Then, it is clear that the infinitesimal generators $\hat{T}_i^{\lambda}$ and $\hat{M}^{\lambda}$, which are the operators acting on $\hat{f}$, are given by

$$\hat{T}_i^{\lambda} = -R_i^{\lambda}, \quad \hat{M}^{\lambda} = M^{\lambda}, \quad \text{(A3)}$$

and are thus independent of ϵ.

APPENDIX B: REDUCTION OF THE HELICITY REPRESENTATIONS OF THE ROTATION GROUP AND ITS APPLICATION TO THE REDUCTION OF THE LITTLE GROUP FOR REAL MASS REPRESENTATIONS APPEARING IN THE DIRECT PRODUCT

The main purpose of the present appendix is to show how one may reduce the operators R_i of Eq. (21) and, hence, how one may reduce the representation of the generators of the little group $\hat{T}_i$ and $\hat{M}$ as they act on F_+. Such a reduction would thus enable one to reduce the real mass components of the direct product completely. But in order to carry out this reduction, it is convenient to give the reduction of the helicity representations of the rotation group.

Let α be any integer or half-odd integer of either sign. Let us introduce the helicity representation of the infinitesimal generators of the rotation group belonging to α in the following way: We first introduce a Hilbert space of complex functions $\{f(\theta, \varphi)\}$ of the continuous variables θ and φ whose ranges are $0 < \theta < \pi$ and $0 < \varphi < 2\pi$, respectively, such that the inner product in the Hilbert space is given by

$$(f^{(1)}, f) = \int_0^{2\pi} \int_0^{\pi} f^{(1)*}(\theta, \varphi) f(\theta, \varphi) \sin\theta \, d\theta \, d\varphi. \quad \text{(B1)}$$

Then the infinitesimal generators of the helicity representation J_i^{α} are given by

$$J_1^{\alpha} f(\theta, \varphi) = \left[i\left(\sin\varphi \frac{\partial}{\partial\theta} + \cot\theta \cos\varphi \frac{\partial}{\partial\varphi}\right) + \cos\varphi \tan\tfrac{1}{2}\theta \cdot \alpha \right] f(\theta, \varphi),$$

$$J_2^{\alpha} f(\theta, \varphi) = \left[-i\left(\cos\varphi \frac{\partial}{\partial\theta} - \cot\theta \sin\varphi \frac{\partial}{\partial\varphi}\right) + \sin\varphi \tan\tfrac{1}{2}\theta \cdot \alpha \right] f(\theta, \varphi), \quad \text{(B2)}$$

$$J_3^{\alpha} f(\theta, \varphi) = \left[-i \frac{\partial}{\partial\varphi} + \alpha \right] f(\theta, \varphi).$$

It is clear that the operators J_i^{α} are Hermitian [with the inner product (B1)] and satisfy the proper commutation relations for the infinitesimal generators of the rotation group for any real α. The requirement that α be an integer or half-odd integer is a necessary and sufficient condition that the integrated infinitesimal generators $\exp[i\boldsymbol{\mu} \cdot \mathbf{J}^{\alpha}]$, where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ is a vector with real components, constitute a ray representation of the rotation group when parameterized in terms of the axis and angle of rotation.

We note that the angular-momentum operators J_i of the standard helicity representation (7) take the form (B2) if we use polar coordinates for $\mathbf{p}$ and choose a variable λ such that M^{λ} is diagonal. Since M^{λ} is Hermitian, such a choice can always be made.

Let us define the unit vector $\boldsymbol{\eta}$ as the vector given by the polar coordinates θ and φ, that is,

$$\boldsymbol{\eta} = (\sin\theta \cos\varphi, \sin\theta \sin\varphi, \cos\theta). \quad \text{(B3)}$$

Let us denote by $R(\boldsymbol{\mu})$ a rotation matrix in terms of a parameterization such that the axis of rotation is given by the direction of the vector $\boldsymbol{\mu}$ and that the angle of rotation is given by $\mu = |\boldsymbol{\mu}|$:

$$R(\boldsymbol{\mu}) = \{R_{ij}(\boldsymbol{\mu})\}, \quad \text{(B4)}$$

where

$$R_{ij}(\boldsymbol{\mu}) = \delta_{ij} \cos\mu + \frac{\mu_i \mu_j}{\mu^2}(1 - \cos\mu) + \sum_k \epsilon_{ijk}\mu_k \frac{\sin\mu}{\mu}. \quad \text{(B4')}$$

In Eq. (B4'), ϵ_{ijk} is the usual antisymmetric 3-index symbol.

Let us define the unit vector $\boldsymbol{\eta}'$ by

$$\boldsymbol{\eta}' = R(-\boldsymbol{\mu})\boldsymbol{\eta} = \boldsymbol{\eta} \cos\mu + [(1 - \cos\mu)/\mu^2](\boldsymbol{\mu} \cdot \boldsymbol{\eta}) + [(\sin\mu)/\mu](\boldsymbol{\mu} \times \boldsymbol{\eta}). \quad \text{(B5)}$$

Furthermore, the angles θ' and φ' are the polar angles of $\boldsymbol{\eta}'$:

$$\boldsymbol{\eta}' = (\sin\theta' \cos\varphi', \sin\theta' \sin\varphi', \cos\theta'). \quad \text{(B6)}$$

We can now give the integrated form of the helicity representation of the rotation group as

$$\exp[i\boldsymbol{\mu} \cdot \mathbf{J}^{\alpha}] f(\theta, \varphi) = \exp[2i\Phi(\boldsymbol{\mu}, \boldsymbol{\eta})\alpha] f(\theta', \varphi'), \quad \text{(B7)}$$

where Φ is given by Eq. (9d). [Equation (B7) is, of course, essentially the same as Eq. (9c).]

It is now our intent to expand the functions $f(\theta, \varphi)$ into modes which transform under the irreducible

representations of the rotation group when the functions $f(\theta, \varphi)$ transforms as in Eq. (B7). This reduction has already been carried out by Moses.[13] [See Eq. (4.2) of that reference.] We reformulate these results for the purposes of the present paper.

Let us introduce the usual form for the irreducible representations of the infinitesimal generators of the rotation group. We denote the matrices corresponding to the irreducible representation characterized by the quantum number j, by $(j, m|\, J_i \,|j, m')$. Then,

$$(j, m|\, J_3 \,|j, m') = m\delta_{m,m'},$$

$$(j, m|\, J_1 \,|j, m') \pm i(j, m|\, J_2 \,|j, m') = [(j \mp m')(j \pm m' + 1)]^{\frac{1}{2}}\delta_{m,m'\pm 1}. \quad \text{(B8)}$$

In Ref. 13 we introduced certain functions $Y_j^{mn}(\theta, \varphi)$ which are generalizations of the usual surface harmonics. We called these functions generalized surface harmonics. We refer to Ref. 13 for their properties. Let us define the coefficients $C(\alpha; j, m)$ by

$$C(\alpha; j, m) = \int_0^{2\pi}\int_0^{\pi} Y_j^{m\alpha *}(\theta, \varphi) f(\theta, \varphi) \sin\theta \, d\theta \, d\varphi. \quad \text{(B9)}$$

Then from the properties of the generalized surface harmonics we can expand $f(\theta, \varphi)$ as follows:

$$f(\theta, \varphi) = \sum_{j=|\alpha|}^{\infty} \sum_{m=-j}^{j} Y_j^{m\alpha}(\theta, \varphi) C(\alpha; j, m). \quad \text{(B10)}$$

Expansion (81) is such that

$$J_i^a f(\theta, \varphi) = \sum_{j=|\alpha|}^{\infty} \sum_{m=-j}^{j} Y_j^{m\alpha}(\theta, \varphi) \times \left(\sum_{m'=-j}^{j} (j, m|\, J_i \,|j, m') C(\alpha; j, m') \right). \quad \text{(B11)}$$

The modes $C(\alpha; j, m)$ thus transform under the irreducible representations of the rotation group.

For the sake of completeness we note

$$(f^{(1)}, f) = \sum_{j=|\alpha|}^{\infty} \sum_{m=-j}^{j} C^{(1)*}(\alpha; j, m) C(\alpha; j, m). \quad \text{(B12)}$$

We shall now apply these results for the purpose of reducing the little group operators $\hat{T}_i$ and $\hat{M}$ of Eqs. (22) and (21). We shall proceed by reducing the operators R_i. Accordingly, we introduce a Hilbert space of complex functions $\{f(\theta, \varphi, \mu, \nu)\}$ where the ranges and character of the variables θ, φ, μ, and ν are the same as those for the same variables in the functions F_+ of Eq. (21). We introduce the inner product

$$(f^{(1)}, f) = \int\int_0^{2\pi}\int_0^{\pi} f^{(1)*}(\theta, \varphi, \mu, \nu) \times f(\theta, \varphi, \mu, \nu) \sin\theta \, d\theta \, d\varphi \, dm(\mu, \nu), \quad \text{(B13)}$$

where the measure function $m(\mu, \nu)$ is the same one used in the inner products (12) and (48).

We define the operators R_i acting on the functions of this Hilbert space as having the same form as the operators R_i acting on the variables θ, φ, μ, ν of F_+ in Eq. (21). It is readily seen that the operators R_i constitute a Hermitian representation of the infinitesimal generators of the rotation group.

For simplicity, it is convenient to suppress the variables μ and ν in the function $f(\theta, \varphi, \mu, \nu)$. Accordingly, we write

$$f(\theta, \varphi) \equiv f(\theta, \varphi, \mu, \nu). \quad \text{(B14)}$$

We can integrate the infinitesimal generators R_i. We obtain

$$\exp[i\boldsymbol{\mu} \cdot \mathbf{R}] f(\theta, \varphi) = \exp[2i\Phi(\boldsymbol{\mu}, \boldsymbol{\eta}) M^{\mu} + 2i\Phi(\boldsymbol{\mu}, -\boldsymbol{\eta}) M^{\nu}] f(\theta', \varphi'), \quad \text{(B15)}$$

where $\boldsymbol{\eta}$ is defined by Eq. (B3) and $\boldsymbol{\eta}'$, θ', φ' are defined by Eqs. (B5) and (B6) and $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ is an arbitrary vector with real components.

Let us now introduce a new Hilbert space of functions $\{g(\theta, \varphi)\}$ by means of a unitary transformation on the previous one:

$$g(\theta, \varphi) = \exp[2iM^{\nu}\varphi] f(\theta, \varphi). \quad \text{(B16)}$$

The operators R_i map into operators, which we denote by $\hat{R}_i$, and are thus defined by

$$\exp[i\boldsymbol{\mu} \cdot \hat{\mathbf{R}}] g(\theta, \varphi) = \exp[2iM^{\nu}\varphi] \exp[i\boldsymbol{\mu} \cdot \mathbf{R}] f(\theta, \varphi). \quad \text{(B17)}$$

On using Eq. (B15) and

$$\varphi - \varphi' = -[\Phi(\boldsymbol{\mu}, \boldsymbol{\eta}) + \Phi(\boldsymbol{\mu}, -\boldsymbol{\eta})], \quad \text{(B18)}$$

which can be proved in a straightforward manner, we find

$$\exp[i\boldsymbol{\mu} \cdot \hat{\mathbf{R}}] g(\theta, \varphi) = \exp[2i\Phi(\boldsymbol{\mu}, \boldsymbol{\eta})(M^{\mu} - M^{\nu})] g(\theta', \varphi'). \quad \text{(B19)}$$

Since M^{μ} and M^{ν} are Hermitian, they can be diagonalized. We assume that a transformation in the μ and ν variables has been made so that the diagonalization has been accomplished. Then, since the values

of M^μ and M^ν must be an integer or half-odd integer, we see that α defined by

$$\alpha = M^\mu - M^\nu \tag{B20}$$

is an integer or half-odd integer. Furthermore, from Eqs. (B19) and (B7) it is seen that $\hat{R}_i$ is identical to J_i^α of Eq. (B2). Hence, the reduction process for J_i^α can now be used to complete the reduction of R_i.

[1] H. E. Moses, J. Math. Phys. **9**, 2039 (1968).
[2] M. Jacob and G.-C. Wick, Ann. Phys. (N.Y.) **7**, 404 (1959).
[3] H. Joos, Fortschr. Physik **10**, 65 (1962).
[4] M. Kummer, J. Math. Phys. **7**, 997 (1966).
[5] J. S. Lomont, J. Math. Phys. **1**, 237 (1960).
[6] A. J. Macfarlane, Rev. Mod. Phys. **34**, 41 (1962).
[7] P. Moussa and R. Stora, *Lectures in Theoretical Physics, Vol. 7* (University of Colorado Press, Boulder, Colo., 1964), Part A, p. 37.
[8] J. Werle, Nucl. Phys. **44**, 579 (1963).
[9] H. E. Moses, J. Math. Phys. **8**, 837 (1967).
[10] H. E. Moses, J. Math. Phys. **9**, 16 (1968).
[11] J. S. Lomont, *Applications of Finite Groups* (Academic Press, New York, 1959).
[12] M. A. Naimark, Am. Math. Soc. Trans., Ser (2) **6**, 379 (1957); see the references to earlier work of I. M. Gel'fand and M. A. Naimark included therein.
[13] H. E. Moses, Ann. Phys. (N.Y.) **41**, 166 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Closed-Form Solution of the Differential Equation $\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0$ by Normal-Ordering Exponential Operators*

RALPH WILCOX
Department of Electrical Engineering, University of Southern California, Los Angeles, California 90024

(Received 23 June 1969)

A closed-form solution to Lambropoulos' partial differential equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P(x, y, t) = 0,$$

subject to the initial condition $P(x, y, 0) = \Phi(x, y)$, is presented. The applicability of the normal-ordering method to a class of partial differential equations is briefly discussed.

1. INTRODUCTION

The partial differential equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P(x, y, t) = 0, \tag{1}$$

subject to the initial condition

$$P(x, y, 0) = \Phi(x, y), \tag{1'}$$

was introduced and solved by Lambropoulos.[1] Here a, b, and c are constants, while $\Phi(x, y)$ is a prescribed function. In a recent paper,[2] Neuringer pointed out that Lambropoulos' solution is quite cumbersome, and has instead obtained a closed-form solution in terms of the 2-dimensional Fourier transform of the function $e^{-\alpha xy}\Phi(x, y)$, where α is a suitably chosen constant.

In the present paper, we obtain a different closed-form solution of (1) under the assumption that

$$R(x, y, t) \equiv \left[\exp\left(-\beta\frac{\partial^2}{\partial x \partial y}\right)\right]\Phi(x, y) \tag{2}$$

exists, where β, as defined by Eqs. (20) and (16)–(18), is a certain function of a, b, c, and t. The solution of (1) is then given by

$$P(x, y, t) = e^{-\delta - \omega xy}R(x/\varphi, y/\rho, t), \tag{3}$$

where δ, ω, φ, and ρ are also certain functions of a, b, c, and t defined by Eqs. (16)–(18), (20), and (23). The reader can most easily verify this for the case where the initial condition is

$$\Phi_{kq}(x, y) \equiv e^{ikx+iqy}, \tag{4}$$

where the solution to (1) is given by

$$P_{kq}(x, y, t) = \exp\left[\beta kq - \delta - \omega xy + i(kx/\varphi) + i(qy/\rho)\right]. \tag{5}$$

In more general cases, the initial condition may be expressed as a linear superposition of exponentials of the form (4) with coefficients depending upon k and q. Consequently, the solution is given by the same linear superposition. This includes the cases of Fourier and Laplace series and transforms. For example, if the initial condition $\Phi(x, y)$ is represented as a Fourier transform (FT)

$$\Phi(x, y) = \int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} dq \Psi(k, q)\Phi_{kq}(x, y), \tag{6}$$

then

$$P(x, y, t) = \int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} dq \Psi(k, q)P_{kq}(x, y, t) \tag{7}$$

is the corresponding solution.

Neuringer's solution has the same formal structure as (5) and (7), but with different expressions for the quantities β, δ, ω, φ, and ρ, since the $\Psi(k, q)$ there denotes the FT of $e^{-\alpha xy}\Phi(x, y)$ rather than the FT of $\Phi(x, y)$ itself. The advantage of our solution is apparent, since the FT of $\Phi(x, y)$ is often simpler than the FT of $e^{-\alpha xy}\Phi(x, y)$, so that the integrations in (7) are easier to perform than in Neuringer's case. Indeed, it is not difficult to conceive of cases where [although $\Phi(x, y)$ has an FT] the FT of $e^{-\alpha xy}\Phi(x, y)$ does not exist since the function $e^{-\alpha xy}$ diverges at large distances in two quadrants of the (x, y) plane.

It is also noted that, if $\Phi(x, y)$ is a polynomial in x or y, the $R(x, y, t)$ defined by (2) is also such a polynomial, so that a solution is simply given. This and other examples are discussed in Sec. 3. In Sec. 2, the solution is derived by the method of normal ordering an exponential operator,[3] while in Sec. 4 some other possible applications of the method are discussed.

2. DERIVATION

We rewrite (1) in the form

$$\left(H + \frac{\partial}{\partial t}\right) P(x, y, t) = 0, \tag{8}$$

where H denotes the operator

$$H \equiv XY + axX + byY + cxy. \tag{8'}$$

Here X and Y denote the operators $\partial/\partial x$ and $\partial/\partial y$, respectively. Since H is independent of t, a formal solution to Eq. (8) [and hence of Eq. (1)] is given by

$$P(x, y, t) = e^{-Ht}\Phi(x, y). \tag{9}$$

Clearly, the operators X, Y, x, and y satisfy the commutation relations

$$\begin{aligned} &[X, x] = I = [Y, y], \\ &[X, Y] = [x, y] = [X, y] = [x, Y] = 0, \end{aligned} \tag{10}$$

where I is the identity operator. Thus, H is a linear combination of the five operators XY, xX, yY, xy, and I which comprise a 5-dimensional Lie algebra with the commutation relations

$$\begin{aligned} &[XY, xX] = XY = [XY, yY], \\ &[XY, xy] = xX + yY + I, \\ &[xX, xy] = xy = [yY, xy]. \end{aligned} \tag{11}$$

All other commutators vanish.

Thus e^{-Ht} can be expressed in normal-ordered form as

$$e^{-Ht} = N \exp(-\beta XY - \mu xX - \nu yY - \omega xy - \delta), \tag{12}$$

where N is the normal-ordering superoperator which orders the derivative operators X and Y to the right of the coordinate operators x and y, *as if* they commuted,[3] while β, μ, ν, ω, and δ are scalar functions of a, b, c, and t which vanish at $t = 0$. To find the differential equations which they must satisfy, we differentiate (12) with respect to t, substitute the left-hand side of (12) into the result, and operate from the right by e^{Ht}. We thus find that

$$H = \dot{\beta}X(t)Y(t) + \dot{\mu}xX(t) + \dot{\nu}yY(t) + \dot{\omega}xy + \dot{\delta}, \tag{13}$$

where the dot is used to denote differentiation with respect to τ and

$$X(t) \equiv e^{-Ht}Xe^{Ht}, \tag{14a}$$

$$Y(t) \equiv e^{-Ht}Ye^{Ht}. \tag{14b}$$

The Lie-algebraic similarity transformations (14) satisfying the commutation relations (11) are easily obtained by a standard method.[4] One finds that

$$X(t) = \varphi X + \gamma y, \tag{15a}$$

$$Y(t) = \rho Y + \sigma x, \tag{15b}$$

where the functions φ, γ, ρ, and σ are defined in terms of the constants a, b, c, and

$$g \equiv [(a + b)^2 - 4c]^{\frac{1}{2}} \tag{16}$$

as follows:

$$\begin{aligned} e^{-\frac{1}{2}(a-b)t}\varphi &= e^{\frac{1}{2}(a-b)t}\rho \\ &= \cosh(\tfrac{1}{2}gt) + (a + b)g^{-1}\sinh(\tfrac{1}{2}gt), \end{aligned} \tag{17}$$

$$\begin{aligned} e^{-\frac{1}{2}(a-b)t}\gamma &= e^{\frac{1}{2}(a-b)t}\sigma \\ &= 2cg^{-1}\sinh(\tfrac{1}{2}gt). \end{aligned} \tag{18}$$

Substituting (15) and (8') into (13) and equating coefficients of XY, xX, yY, xy, and I, respectively, one obtains the following equations:

$$\begin{aligned} 1 &= \dot{\beta}\varphi\rho, \\ a &= \dot{\beta}\varphi\sigma + \dot{\mu}\rho, \\ b &= \dot{\beta}\gamma\rho + \dot{\nu}\rho, \\ c &= \dot{\beta}\gamma\sigma + \dot{\mu}\gamma + \dot{\nu}\sigma + \dot{\omega}, \\ 0 &= \dot{\beta}\varphi\sigma + \dot{\delta}. \end{aligned} \tag{19}$$

The solution to these equations which vanishes when $t = 0$ is found to be

$$\omega = c\beta = \gamma\varphi^{-1}, \tag{20}$$

$$\mu = 1 - \varphi^{-1}, \tag{21}$$

$$\nu = 1 - \rho^{-1}, \tag{22}$$

$$\delta = \ln(\varphi) - at. \tag{23}$$

Now, because of the way the normal-ordering superoperator N is defined, (12) may also be written as

$$e^{-Ht} = e^{-\delta - \omega xy}(Ne^{-\mu xX - \nu yY})e^{-\beta XY}. \tag{24}$$

Operating upon $\Phi(x, y)$ as in (10), we find that the rightmost exponential gives the definition (2) for $R(x, y, t)$, while the middle factor is then easily shown to give

$$R(x - x\mu, y - y\nu, t). \tag{25}$$

Using (21) and (22), one obtains the form given in (3), which completes the derivation.

3. EXAMPLES

In the following examples, we calculate only $R(x, y, t)$, since the prescription (3) for obtaining $P(x, y, t)$ is straightforward.

Consider, first, the case where $\Phi(x, y)$ is a function of x only, i.e.,

$$\Phi(x, y) = F(x). \tag{26}$$

Then, from (2), we have

$$R(x, y, t) = F(x). \tag{27}$$

More generally, if

$$\Phi(x, y) = y^n f(x), \tag{28}$$

then

$$R(x, y, t) = \sum_{j=0}^{\infty} (-\beta)^j \frac{\partial^{2j}}{\partial x^j \partial y^j} [y^n f(x)]$$

$$= \sum_{j=0}^{n} \frac{n!\,(-\beta)^j}{(n-j)!}\, y^{n-j} f^{(j)}(x), \tag{29}$$

where

$$f^{(0)}(x) \equiv f(x), \quad f^{(j+1)}(x) \equiv \frac{df^{(j)}(x)}{dx}.$$

For the special case

$$\Phi(x, y) = x^m y^n, \quad m \geq n,$$

we have

$$R(x, y, t) = \sum_{j=0}^{n} \frac{m!\, n!\, x^{m-j} y^{n-j}}{(m-j)!\,(n-j)!}. \tag{30}$$

Of course, more general functions may be handled by the method of linear superposition at the expense of increasing the number of terms.

As another example, consider the case where $P(x, y, t)$ is initially a 2-dimensional Gaussian, i.e.,

$$\Phi(x, y) = e^{-A(x^2+y^2)}, \tag{31}$$

where A is a positive constant. Then the FT $\Psi(k, q)$, defined by (6), is given by

$$(4\pi A)^{-1} e^{-(k^2+q^2)/4A}. \tag{32}$$

Substituting (32) into (7), using (5), and changing integration variables to u and v defined by $k = u + v$ and $q = u - v$, one obtains

$$(1 - 4\beta^2 A^2)^{\frac{1}{2}} R(x, y, t)$$

$$= \exp\left[-\tfrac{1}{2}(x+y)^2 (A^{-1} - 2\beta)^{-1} - \tfrac{1}{2}(x-y)^2 (A^{-1} + 2\beta)^{-1}\right] \tag{33}$$

upon performing the integrations. Note that for some values of the parameters a, b, c, and A it can happen, for large enough t, that $4\beta^2 A^2 \geq 1$, in which case $R(x, y, t)$—and hence a solution $P(x, y, t)$ satisfying Eqs. (1) and (1′)—no longer exists.

4. DISCUSSION

We remark that, by an obvious modification of the method given here, the form of the solution given by Neuringer can also be obtained. Neuringer obtained his solution by a more conventional method which involves transforming both independent and dependent variables. We have not investigated whether our solution can be obtained by conventional means, but we suspect that it can.

The method which we have used should (at least in principle) enable one to obtain closed-form solutions to other partial differential equations of the form (8) where the operator H is a second-degree polynomial in the coordinate and partial derivative operators, since the Lie algebra for such cases is finite dimensional. An analogous statement for the creation and annihilation operators describing a many-boson system has been made by Marburger.[5]

A further generalization can be made to the case where the coefficients multiplying the operators in H depend upon t. Although (10) is no longer valid, one can still write

$$P(x, y, t) = e^{-G(t)} \Phi(x, y),$$

where $G(t)$ is a t-dependent linear combination of the operators comprising the same finite-dimensional Lie algebra.[6] Since the normal-ordered form of $e^{-G(t)}$ is also an exponential of this same Lie algebra, a closed-form expression may (at least in principle) be obtained.

If (as commonly occurs in more practical problems) the Lie algebra of H is infinite dimensional, it still may be possible to approximate H by elements of a finite-dimensional Lie algebra by using techniques known for approximating exponential operators. Whether such an approach can compete with conventional methods for solving partial differential equations remains to be seen.

In conclusion, we observe that the procedure employed in this paper is just converse to that usually followed in quantum physics. There, one converts an abstract operator differential equation into an ordinary partial differential equation which one then proceeds to solve by conventional means. Here, on the other hand, we have solved a partial differential equation by taking explicit account of its abstract Lie-algebraical properties.

* This paper was supported by the National Science Foundation under Grant No. GP-8960 and the Atomic Energy Commission under Contract No. AT(11-1)-113, Project No. 19.

[1] P. Lambropoulos, J. Math. Phys. **8**, 2167 (1967).

[2] J. L. Neuringer, J. Math. Phys. **10**, 250 (1969).

[3] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967), Sec. 10; further references on normal ordering are given here.

[4] See Ref. 3, Sec. 6.

[5] J. H. Marburger, J. Math. Phys. **7**, 829 (1966).

[6] See Ref. 3, Sec. 8.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Analysis of the Newman–Unti Integration Procedure for Asymptotically Flat Space-Times

W. G. DIXON
Churchill College, Cambridge University, Cambridge, England

(Received 20 June 1969)

An analysis is given of the procedure of Newman and Unti for solving the vacuum gravitational field equations for all space-times in which $\Psi'_0 = O(r^{-5})$. It was found empirically by Newman and Unti that when the nonradial equations and three of the u-derivative equations have been satisfied to their lowest nontrivial order in r^{-1}, they are then found to be identically satisfied to all orders. A general proof of this result is given which avoids the need for direct verification.

1. INTRODUCTION

The behavior of asymptotically flat empty space-times has been the subject of much study in recent years. A considerable proportion of this work is based on the spin-coefficient formalism of Newman and Penrose.[1] Using their results, Newman and Unti[2] were able to give a systematic method for integrating the vacuum field equations for all space-times in which one of the Weyl spinor components, usually denoted by Ψ'_0, is $O(r^{-5})$ for large r, where r is a suitably defined radial coordinate on the hypersurfaces on which the retarded time u is constant. This condition may be interpreted[3] as requiring that the space-time be asymptotically flat and that any incoming gravitational radiation die off sufficiently rapidly as time goes on. Further developments of this work have been the study of the asymptotic multipole structure of the field by Janis Newman,[4] the extension of this by Newman and Unti,[5] and the discovery of some conservation laws of a new type by Newman and Penrose.[3,6]

The work of Newman and Unti,[2] however, contained a logical gap, in that they did not show that they had extracted all the information contained in the field equations. Work of other authors,[7,8] using somewhat different approaches, strongly suggested that they had done so, but no general proof of this was given. In view of the importance of their results, the present author feels that this gap should be filled, and it is the purpose of this paper to do so. In the course of the proof given here, insight is gained into the underlying structure of the NU procedure which would not be gained if the completeness of their results was verified by direct calculation.

The notation of NP is summarized in Sec. 2, together with those equations which will be needed. As the proof also needs certain details of the NU integration procedure which were not given in NU, a fairly detailed account of the NU procedure is given in Sec. 3. The proof of completeness is then given in Sec. 4, and its results are discussed in Sec. 5.

2. THE NP FORMALISM

The notation of NP is based on a null tetrad field $(l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$, which may be given arbitrarily at every point of the space-time. Here, l^μ and n^μ are real, m^μ is complex, and the overbar denotes complex conjugation. The only nonzero scalar products between these vectors are

$$l^\mu n_\mu = -m^\mu \bar{m}_\mu = 1, \tag{2.1}$$

so that the metric tensor of the space-time, taken as having signature $- - - +$, is given by

$$g^{\mu\nu} = l^\mu n^\nu + l^\nu n^\mu - m^\mu \bar{m}^\nu - m^\nu \bar{m}^\mu. \tag{2.2}$$

The affine connection may then be described by twelve complex *spin coefficients*, defined by

$$\begin{aligned}
&\kappa = l_{\mu;\nu} m^\mu l^\nu, \quad \pi = -n_{\mu;\nu} \bar{m}^\mu l^\nu, \quad \rho = l_{\mu;\nu} m^\mu \bar{m}^\nu,\\
&\epsilon = \tfrac{1}{2}(l_{\mu;\nu} n^\mu l^\nu - m_{\mu;\nu} \bar{m}^\mu l^\nu), \quad \lambda = -n_{\mu;\nu} \bar{m}^\mu \bar{m}^\nu,\\
&\alpha = \tfrac{1}{2}(l_{\mu;\nu} n^\mu \bar{m}^\nu - m_{\mu;\nu} \bar{m}^\mu \bar{m}^\nu), \quad \sigma = l_{\mu;\nu} m^\mu m^\nu,\\
&\beta = \tfrac{1}{2}(l_{\mu;\nu} n^\mu m^\nu - m_{\mu;\nu} \bar{m}^\mu m^\nu), \quad \nu = -n_{\mu;\nu} \bar{m}^\mu n^\nu,\\
&\gamma = \tfrac{1}{2}(l_{\mu;\nu} n^\mu n^\nu - m_{\mu;\nu} \bar{m}^\mu n^\nu), \quad \mu = -n_{\mu;\nu} \bar{m}^\mu m^\nu,\\
&\qquad \tau = l_{\mu;\nu} m^\mu n^\nu.
\end{aligned} \tag{2.3}$$

If we label the vectors l^μ, n^μ, m^μ, and $\bar{m}^\mu$ as 1, 2, 3, and 4 in that order, the tetrad components of any tensor may be numerically labeled in the usual manner—for example, $C_{1232} = C_{\kappa\lambda\mu\nu} l^\kappa n^\lambda m^\mu n^\nu$. The following complex functions are then introduced to describe the irreducible parts of the curvature tensor: The Weyl tensor is described by the five functions

$$\begin{aligned}
&\Psi_0 = -C_{1313}, \quad \Psi_1 = -C_{1213},\\
&\Psi_2 = \tfrac{1}{2}(C_{1234} - C_{1212}),\\
&\Psi_3 = C_{1224}, \quad \Psi_4 = -C_{2424};
\end{aligned} \tag{2.4}$$

the tracefree part of the Ricci tensor $R_{\mu\nu}$ by a 3×3 Hermitian matrix Φ_{AB}, $A, B = 0, 1, 2$, so that $\bar{\Phi}_{AB} = \Phi_{BA}$, defined by

$$\begin{aligned} \Phi_{00} &= -\tfrac{1}{2}R_{11}, \quad \Phi_{01} = -\tfrac{1}{2}R_{13}, \\ \Phi_{02} &= -\tfrac{1}{2}R_{33}, \quad \Phi_{11} = -\tfrac{1}{4}(R_{12} + R_{34}), \\ \Phi_{12} &= -\tfrac{1}{2}R_{23}, \quad \Phi_{22} = -\tfrac{1}{2}R_{22}; \end{aligned} \tag{2.5}$$

and finally the curvature scalar is expressed in terms of $\Lambda = \frac{1}{24}R$. This notation, while seeming very much ad hoc as thus defined, appears very natural when expressed in the notation of 2-component spinors, which is the form in which it was originated in NP. The conventions adopted for the curvature tensor are given in the Appendix.

A special coordinate system, together with an associated null tetrad field, is next introduced as follows. The coordinate $x^1 = u$ labels a family of null hypersurfaces, so that

$$g^{\mu\nu}u_{,\mu}u_{,\nu} = 0. \tag{2.6}$$

Take $l^\mu = g^{\mu\nu}u_{,\nu}$. Then[9] l^μ is tangent to a null geodesic congruence, and the coordinate $x^2 = r$ is chosen to be an affine parameter along each geodesic of the congruence. The coordinates x^3 and x^4 are left arbitrary, as are the other vectors of the tetrad, subject only to having the appropriate scalar products. Their components with respect to this coordinate system then have the form

$$\begin{aligned} l^\mu &= \delta^\mu_2, \quad m^\mu = \omega\delta^\mu_2 + \xi^k\delta^\mu_k, \\ n^\mu &= \delta^\mu_1 + U\delta^\mu_2 + X^k\delta^\mu_k, \end{aligned} \tag{2.7}$$

where $k = 3, 4$, thus defining the *metric coefficients* U, X^k, ω, and ξ^k, of which U and X^k are real and ω and ξ^k complex. From (2.2) the contravariant metric components are then

$$\begin{aligned} g^{11} &= g^{1k} = 0, \quad g^{12} = 1, \\ g^{22} &= 2(U - \omega\bar{\omega}), \\ g^{2k} &= X^k - (\xi^k\bar{\omega} + \bar{\xi}^k\omega), \\ g^{kl} &= -(\xi^k\bar{\xi}^l + \bar{\xi}^k\xi^l), \end{aligned} \tag{2.8}$$

where $k, l = 3, 4$.

As l^μ is tangent to a hypersurface-orthogonal null-geodesic congruence, with r an affine parameter along its rays, the spin coefficients satisfy

$$\kappa = 0, \quad \epsilon + \bar{\epsilon} = 0, \quad \rho = \bar{\rho}, \quad \tau = \bar{\alpha} + \beta. \tag{2.9}$$

The remaining spin coefficients are then determined by the *metric equations*

$$D\xi^k = (\rho + \epsilon - \bar{\epsilon})\xi^k + \sigma\bar{\xi}^k, \tag{2.10a}$$
$$D\omega = (\rho + \epsilon - \bar{\epsilon})\omega + \sigma\bar{\omega} - \tau + \bar{\pi}, \tag{2.10b}$$
$$DX^k = (\bar{\tau} + \pi)\xi^k + (\tau + \bar{\pi})\bar{\xi}^k, \tag{2.10c}$$
$$DU = (\bar{\tau} + \pi)\omega + (\tau + \bar{\pi})\bar{\omega} - (\gamma + \bar{\gamma}), \tag{2.10d}$$
$$\delta X^k - \Delta\xi^k = (\mu + \bar{\gamma} - \gamma)\xi^k + \bar{\lambda}\bar{\xi}^k, \tag{2.11a}$$
$$\delta U - \Delta\omega = (\mu + \bar{\gamma} - \gamma)\omega + \bar{\lambda}\bar{\omega} - \bar{\nu}, \tag{2.11b}$$
$$\delta\bar{\xi}^k - \bar{\delta}\xi^k = (\bar{\beta} - \alpha)\xi^k + (\bar{\alpha} - \beta)\bar{\xi}^k, \tag{2.11c}$$
$$\delta\bar{\omega} - \bar{\delta}\omega = (\bar{\beta} - \alpha)\omega + (\bar{\alpha} - \beta)\bar{\omega} + (\mu - \bar{\mu}), \tag{2.11d}$$

where the derivative operators D, δ, and Δ are defined by

$$\begin{aligned} D &= l^\mu\partial_\mu = \frac{\partial}{\partial r}, \\ \delta &= m^\mu\partial_\mu = \omega\frac{\partial}{\partial r} + \xi^k\frac{\partial}{\partial x^k}, \\ \Delta &= n^\mu\partial_\mu = U\frac{\partial}{\partial r} + \frac{\partial}{\partial u} + X^k\frac{\partial}{\partial x^k}. \end{aligned} \tag{2.12}$$

By applying the Ricci identity to the vectors l^μ, m^μ, and n^μ and writing the results in the above notation, we obtain the following 18 equations which form a redundant set for the evaluation of Ψ_A, Φ_{AB}, and Λ:

$$D\rho = \rho^2 + \sigma\bar{\sigma} + \Phi_{00}, \tag{2.13a}$$
$$D\sigma = (2\rho + 4\epsilon)\sigma + \Psi_0, \tag{2.13b}$$
$$D\tau = (\tau + \bar{\pi})\rho + (\bar{\tau} + \pi)\sigma + 2\epsilon\tau + \Psi_1 + \Phi_{01}, \tag{2.13c}$$
$$D\alpha - \bar{\delta}\epsilon = (\rho - 3\epsilon)\alpha + \beta\bar{\sigma} - \bar{\beta}\epsilon + (\rho + \epsilon)\pi + \Phi_{10}, \tag{2.13d}$$
$$D\beta - \delta\epsilon = (\alpha + \pi)\sigma + (\rho + \epsilon)\beta + (\bar{\pi} - \bar{\alpha})\epsilon + \Psi_1, \tag{2.13e}$$
$$D\gamma - \Delta\epsilon = (\tau + \bar{\pi})\alpha + (\bar{\tau} + \pi)\beta - (\gamma + \bar{\gamma})\epsilon + \tau\pi + \Psi_2 - \Lambda + \Phi_{11}, \tag{2.13f}$$
$$D\lambda - \bar{\delta}\pi = (\rho - 4\epsilon)\lambda + \bar{\sigma}\mu + (\pi + \alpha - \bar{\beta})\pi + \Phi_{20}, \tag{2.13g}$$
$$D\mu - \delta\pi = \rho\mu + \sigma\lambda + \pi\bar{\pi} - (\bar{\alpha} - \beta)\pi + \Psi_2 + 2\Lambda, \tag{2.13h}$$
$$D\nu - \Delta\pi = (\pi + \bar{\tau})\mu + (\bar{\pi} + \tau)\lambda + (\gamma - \bar{\gamma})\pi - 2\epsilon\nu + \Psi_3 + \Phi_{21}, \tag{2.13i}$$

$$\Delta\lambda - \bar{\delta}\nu = (\bar{\gamma} - 3\gamma - \mu - \bar{\mu})\lambda + (2\alpha + \pi)\nu - \Psi'_4, \tag{2.14a}$$

$$\delta\rho - \bar{\delta}\sigma = \rho\tau - \sigma(3\alpha - \bar{\beta}) - \Psi'_1 + \Phi_{01}, \tag{2.14b}$$

$$\delta\alpha - \bar{\delta}\beta = \mu\rho - \lambda\sigma + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta + \epsilon(\mu - \bar{\mu}) - \Psi'_2 + \Lambda + \Phi_{11}, \tag{2.14c}$$

$$\delta\lambda - \bar{\delta}\mu = (\mu - \bar{\mu})\pi + \mu\bar{\tau} + \lambda(\bar{\alpha} - 3\beta) - \Psi'_3 + \Phi_{21}, \tag{2.14d}$$

$$\delta\nu - \Delta\mu = \lambda\bar{\lambda} + (\mu + \gamma + \bar{\gamma})\mu - \bar{\nu}\pi - 2\beta\nu + \Phi_{22}, \tag{2.14e}$$

$$\delta\gamma - \Delta\beta = \mu\tau - \sigma\nu - \epsilon\bar{\nu} - \beta(\gamma - \bar{\gamma} - \mu) + \alpha\bar{\lambda} + \Phi_{12}, \tag{2.14f}$$

$$\delta\tau - \Delta\sigma = (\mu + \bar{\gamma} - 3\gamma)\sigma + \bar{\lambda}\rho + 2\beta\tau + \Phi_{02}, \tag{2.14g}$$

$$\Delta\rho - \bar{\delta}\tau = (\gamma + \bar{\gamma} - \bar{\mu})\rho - \sigma\lambda - 2\alpha\tau - \Psi'_2 - 2\Lambda, \tag{2.14h}$$

$$\Delta\alpha - \bar{\delta}\gamma = (\rho + \epsilon)\nu - (\tau + \beta)\lambda + (\bar{\gamma} - \gamma - \bar{\mu})\alpha - \Psi'_3. \tag{2.14i}$$

The redundancy corresponds to the cyclic identity $R_{\kappa[\lambda\mu\nu]} = 0$ of the curvature tensor. A particular instance is that, in view of (2.9), any two of the equations (2.13c), (2.13d), and (2.13e) imply the third.

We shall also need the Bianchi identities. These are given in NP only for a vacuum and for an Einstein–Maxwell field, but the full set has been calculated independently by Szekeres and by McLenaghan, and is given by Pirani.[10] With the simplifications induced by (2.9), they are

$$D(\Phi_{01} - \Psi'_1) - \delta\Phi_{00} + \bar{\delta}\Psi'_0 = (4\alpha - \pi)\Psi'_0 - (4\rho + 2\epsilon)\Psi'_1 + (\bar{\pi} - 2\tau)\Phi_{00} + 2(\rho + \epsilon)\Phi_{01} + 2\sigma\Phi_{10}, \tag{2.15a}$$

$$\begin{aligned} D(2\Phi_{11} - 3\Psi'_2) - 2\delta\Phi_{10} + \bar{\delta}(3\Psi'_1 + \Phi_{01}) - \Delta\Phi_{00} \\ = 3\lambda\Psi'_0 + 6(\alpha - \pi)\Psi'_1 - 9\rho\Psi'_2 + (\bar{\mu} - 2\mu - 2\gamma - 2\bar{\gamma})\Phi_{00} + 2(\alpha + \pi + \bar{\tau})\Phi_{01} \\ + 2(\tau - 2\bar{\alpha} + \bar{\pi})\Phi_{10} + 2\rho\Phi_{11} + 2\sigma\Phi_{20} - \bar{\sigma}\Phi_{02}, \end{aligned} \tag{2.15b}$$

$$\begin{aligned} D(\Phi_{21} - 3\Psi'_3) - \delta\Phi_{20} + \bar{\delta}(3\Psi'_2 + 2\Phi_{11}) - 2\Delta\Phi_{10} \\ = 6\lambda\Psi'_1 - 9\pi\Psi'_2 + 6(\epsilon - \rho)\Psi'_3 - 2\nu\Phi_{00} + 2\lambda\Phi_{01} + 2(\bar{\mu} - \mu - 2\bar{\gamma})\Phi_{10} + (2\pi + 4\bar{\tau})\Phi_{11} \\ + (4\beta + \bar{\pi})\Phi_{20} - 2\bar{\sigma}\Phi_{12} - 2\epsilon\Phi_{21}, \end{aligned} \tag{2.15c}$$

$$\begin{aligned} D\Psi'_4 - \bar{\delta}(\Phi_{21} + \Psi'_3) + \Delta\Phi_{20} \\ = -3\lambda\Psi'_2 + 2(\alpha + 2\pi)\Psi'_3 + (\rho - 4\epsilon)\Psi'_4 + 2\nu\Phi_{10} - 2\lambda\Phi_{11} + (2\bar{\gamma} - 2\gamma - \bar{\mu})\Phi_{20} \\ - 2\bar{\beta}\Phi_{21} + \bar{\sigma}\Phi_{22}, \end{aligned} \tag{2.15d}$$

$$\begin{aligned} \Delta\Psi'_0 - \delta(\Phi_{01} + \Psi'_1) + D\Phi_{02} \\ = (4\gamma - \mu)\Psi'_0 - 2(2\tau + \beta)\Psi'_1 + 3\sigma\Psi'_2 - \bar{\lambda}\Phi_{00} + 2(\bar{\pi} - \beta)\Phi_{01} + 2\sigma\Phi_{11} + (\rho + 4\epsilon)\Phi_{02}, \end{aligned} \tag{2.16a}$$

$$\begin{aligned} \Delta(3\Psi'_1 - \Phi_{01}) - \delta(2\Phi_{11} + 3\Psi'_2) + \bar{\delta}\Phi_{02} + 2D\Phi_{12} \\ = 3\nu\Psi'_0 + 6(\gamma - \mu)\Psi'_1 - 9\tau\Psi'_2 + 6\sigma\Psi'_3 - \bar{\nu}\Phi_{00} + 2(\bar{\mu} - \mu - \gamma)\Phi_{01} - 2\bar{\lambda}\Phi_{10} + 2(\tau + 2\bar{\pi})\Phi_{11} \\ + (3\alpha - \bar{\beta} + 2\pi)\Phi_{02} + 4\epsilon\Phi_{12} + 2\sigma\Phi_{21}, \end{aligned} \tag{2.16b}$$

$$\begin{aligned} \Delta(3\Psi'_2 - 2\Phi_{11}) - \delta(\Phi_{21} + 3\Psi'_3) + 2\bar{\delta}\Phi_{12} + D\Phi_{22} \\ = 6\nu\Psi'_1 - 9\mu\Psi'_2 - 6\bar{\alpha}\Psi'_3 + 3\sigma\Psi'_4 - 2\nu\Phi_{01} - 2\bar{\nu}\Phi_{10} + 2(2\bar{\mu} - \mu)\Phi_{11} + 2\lambda\Phi_{02} - \bar{\lambda}\Phi_{20} \\ + 2(\pi + \bar{\tau} - 2\bar{\beta})\Phi_{12} + 2(\beta + \tau + \bar{\pi})\Phi_{21} - \rho\Phi_{22}, \end{aligned} \tag{2.16c}$$

$$\begin{aligned} \Delta(\Psi'_3 - \Phi_{21}) - \delta\Psi'_4 + \bar{\delta}\Phi_{22} \\ = 3\nu\Psi'_2 - 2(\gamma + 2\mu)\Psi'_3 + (4\beta - \tau)\Psi'_4 - 2\nu\Phi_{11} - \bar{\nu}\Phi_{20} + 2\lambda\Phi_{12} + 2(\gamma + \bar{\mu})\Phi_{21} - \bar{\tau}\Phi_{22}, \end{aligned} \tag{2.16d}$$

$$\begin{aligned} D(\Phi_{11} + 3\Lambda) - \delta\Phi_{10} - \bar{\delta}\Phi_{01} + \Delta\Phi_{00} \\ = (2\gamma + 2\bar{\gamma} - \mu - \bar{\mu})\Phi_{00} + (\pi - 2\alpha - 2\bar{\tau})\Phi_{01} + (\bar{\pi} - 2\bar{\alpha} - 2\tau)\Phi_{10} + 4\rho\Phi_{11} + \bar{\sigma}\Phi_{02} + \sigma\Phi_{20}, \end{aligned} \tag{2.17a}$$

$$\begin{aligned} D\Phi_{12} - \delta(\Phi_{11} - 3\Lambda) - \bar{\delta}\Phi_{02} + \Delta\Phi_{01} \\ = \bar{\nu}\Phi_{00} + (2\gamma - \mu - 2\bar{\mu})\Phi_{01} - \bar{\lambda}\Phi_{10} + 2(\bar{\pi} - \tau)\Phi_{11} + (\pi + \bar{\beta} - 3\alpha)\Phi_{02} + (3\rho + 2\epsilon)\Phi_{12} \\ + \sigma\Phi_{21}, \end{aligned} \tag{2.17b}$$

$$\begin{aligned} D\Phi_{22} - \delta\Phi_{21} - \bar{\delta}\Phi_{12} + \Delta(\Phi_{11} + 3\Lambda) \\ = \nu\Phi_{01} + \bar{\nu}\Phi_{10} - 2(\mu + \bar{\mu})\Phi_{11} - \lambda\Phi_{02} - \bar{\lambda}\Phi_{20} + (2\pi - \bar{\tau} + 2\bar{\beta})\Phi_{12} + (2\bar{\pi} - \tau + 2\beta)\Phi_{21} \\ + 2\rho\Phi_{22}. \end{aligned} \tag{2.17c}$$

3. THE NU INTEGRATION PROCEDURE

Provided that

$$\xi^3\bar{\xi}^4 - \bar{\xi}^3\xi^4 \neq 0, \tag{3.1}$$

whatever values are chosen for the real functions U and X^k, and complex functions ω and ξ^k, the metric (2.8) will be nonsingular and have the correct signature $- - - +$ for a space-time, and the tetrad vectors determined by (2.7) will have the correct scalar products (2.1). Moreover, the corresponding spin coefficients will satisfy (2.9). But without altering either the metric or the coordinate system, one can perform any Lorentz transformation of the tetrad frames which keeps l^μ fixed, thus altering the metric coefficients U, X^k, ω, and ξ^k, while leaving the functional form (2.8) of the components of $g^{\mu\nu}$ unaltered. Using this freedom, one may additionally impose that the tetrad frame should be parallelly transported along the rays of the l^μ congruence. The spin coefficients will then satisfy the stronger restrictions

$$\kappa = \epsilon = \pi = 0, \quad \rho = \bar{\rho}, \quad \tau = \bar{\alpha} + \beta, \tag{3.2}$$

with corresponding differential equations restricting the metric coefficients.

In an empty space-time one also has the vacuum field equations

$$\Lambda = 0, \quad \Phi_{AB} = 0, \quad A, B = 0, 1, 2. \tag{3.3}$$

Under the specializations (3.2) and (3.3), the above equations (2.10), (2.11), and (2.13)–(2.16) simplify considerably, while (2.17) becomes trivial. Let us denote this simplified form of an equation by prefixing the equation number with an asterisk, e.g., (*2.13a). The NU integration procedure aims to integrate these simplified equations for all empty space-times in which $\Psi_0 = O(r^{-5})$ as $r \to \infty$. It is necessary to assume also that the hypersurfaces $u =$ const are asymptotically neither plane nor cylindrical. The precise statement of this condition is given in NP, where, under these assumptions, the order of magnitude of the asymptotic r-dependence of all the metric and spin coefficients and components Ψ_A of the Weyl tensor is determined. These results form the starting point of NU.

In NU, a slightly more restrictive assumption is placed on Ψ_0 than just $\Psi_0 = O(r^{-5})$, in order to be able to consider the time dependence of the solution. They assume that

$$\Psi_0 = \Psi_0^0 r^{-5} + O(r^{-6}), \tag{3.4}$$

where Ψ_0^0 is independent of r. They further assume that this may be formally differentiated with respect to r once,

$$D\Psi_0 = -5\Psi_0^0 r^{-6} + O(r^{-7}), \tag{3.5}$$

and that up to four or three derivatives of (3.4) and (3.5), respectively, with respect to x^3 or x^4 may be performed without altering this r-dependence of the remainder term. This latter assumption is described as "uniform smoothness." In subsequent work,[3] still more terms have been assumed in this expansion of Ψ_0, and so we shall adopt the more general assumption

$$\Psi_0 = \sum_{n=0}^{N} \Psi_0^n r^{-n-5} + O(r^{-N-6}), \tag{3.6}$$

again with the same assumptions about derivatives.

The NU procedure solves simultaneously for the metric and spin coefficients, and Ψ_1 through Ψ_4, in terms of the given Ψ_0. Its first step is to integrate the "radial equations" (*2.10), (*2.13), and (*2.15), there being exactly one equation of this set giving the radial derivative of each of the unknown quantities. One arbitrary function of u and x^k occurs in the integration corresponding to each unknown quantity, and this is denoted by the same symbol as the unknown but with a superscript zero. We shall call these the *initial functions*. By using the freedom remaining in the choice of tetrad frame and coordinate system, without altering the family of hypersurfaces $u =$ const on which the coordinates are based, it is possible to make

$$\rho^0 = \tau^0 = 0, \quad \xi^{04} = i\xi^{03} \tag{3.7}$$

and to make ξ^{03} real. We write $\xi^{03} = P$.

To perform these integrations, the equations are divided into sets which are integrated sequentially, all members of a set being integrated simultaneously. These sets, together with the functions determined by them, are given in Table I. The group letters in this table are for future reference. The simplifications (3.7) are made as soon as that particular initial function

TABLE I. Integration of radial equations.

Group	Set number	Equations (*2.10)	(*2.13)	(*2.15)	Functions determined
A	1		a, b		ρ, σ
	2	a			ξ^k
B	3	b	d, e	a	$\omega, \alpha, \beta, \tau, \Psi_1$
	4	c			X^k
C	5		g, h	b	λ, μ, Ψ_2
	6				γ
	7	d			U
D	8			c	Ψ_3
	9		i		ν
	10			d	Ψ_4

appears. The leading terms in the resulting expressions are given below, to show the coefficients in which the initial functions first appear. More terms in the expansions are given in NU, but we do not need them here. We find

$$
\begin{aligned}
\Psi'_1 &= \Psi'^0_1 r^{-4} + \cdots, & \Psi'_2 &= \Psi'^0_2 r^{-3} + \cdots,\\
\Psi'_3 &= \Psi'^0_3 r^{-2} + \cdots, & \Psi'_4 &= \Psi'^0_4 r^{-1} + \cdots,\\
\rho &= -r^{-1} - \sigma^0\bar{\sigma}^0 r^{-3} + \cdots, & \sigma &= \sigma^0 r^{-2} + \cdots,\\
\alpha &= \alpha^0 r^{-1} + \bar{\sigma}^0\bar{\alpha}^0 r^{-2} + \cdots, & \beta &= -\bar{\alpha}^0 r^{-1} - \sigma^0\alpha^0 r^{-2} + \cdots,\\
\tau &= -\tfrac{1}{2}\Psi'^0_1 r^{-3} + \cdots, & \lambda &= \lambda^0 r^{-1} + \cdots,\\
\mu &= \mu^0 r^{-1} + \cdots, & \gamma &= \gamma^0 - \tfrac{1}{2}\Psi'^0_2 r^{-2} + \cdots,\\
\nu &= \nu^0 - \Psi'^0_3 r^{-1} + \cdots, & \omega &= \omega^0 r^{-1} + \cdots,\\
U &= -(\gamma^0 + \bar{\gamma}^0) r + U^0 + \cdots, & \xi^k &= \xi^{0k} r^{-1} - \sigma^0 \bar{\xi}^{0k} r^{-2} + \cdots,\\
& X^k = X^{0k} + \tfrac{1}{6}(\Psi'^0_1 \bar{\xi}^{0k} + \bar{\Psi}'^0_1 \xi^{0k}) r^{-3} + \cdots.
\end{aligned}
\tag{3.8}
$$

With the assumption (3.6), this procedure will give the Ψ'_A up to and including the terms in r^{-N-5} with a remainder term $O(r^{-N-6})$, the spin coefficients X^k and ξ^k up to the terms in r^{-N-4} with a remainder term $O(r^{-N-5})$, and U and ω up to the terms in r^{-N-3} with a remainder term $O(r^{-N-4})$. It can be shown that all the formal derivatives of these expressions that we need are also valid.

The next step is to use the "nonradial equations" (*2.11) and (*2.14) to evaluate most of the initial functions. By substituting the above expressions into the nonradial equations, they take the form of series in powers of r^{-1} with a remainder term whose power of r^{-1} increases with N. The coefficients of every power of r^{-1} less than that of the remainder term must separately vanish if the resulting metric is to satisfy the vacuum field equations. If we equate to zero only the first nontrivial coefficient in each equation, we obtain the results given in Table II, where the order of the term used is also given. The operators ∇ and $\eth$ used in these results are defined in the Appendix. The equations (*2.14e), (*2.14f), and (*2.14i) are not listed because their first nontrivial terms are identically satisfied in consequence of the results obtained from the other equations. A special feature arises when (*2.11a) is considered. Its r^{-1} terms are first used to deduce $\nabla(X^{03} + iX^{04}) = 0$, from which it follows that $X^{03} + iX^{04}$ is an analytic function of $x^3 + ix^4$. In consequence[2,9] it is possible to use a coordinate transformation to set $X^{03} = X^{04} = 0$ without loss of generality. The r^{-1} terms of (*2.11a) are then used again to evaluate γ^0. The group letters in the table indicate that the determination of the initial functions of that group depend only on the radial integrations in the corresponding or earlier groups, e.g., ω^0 may be evaluated after integrating only the first four sets of radial equations.

TABLE II. Integration of nonradial equations.

Group	Equation	Term used	Result
B	(*2.11c)	r^{-2}	$\alpha^0 = \tfrac{1}{2}\bar{\nabla}P$
	(*2.14b)	r^{-3}	$\omega^0 = -\bar{\eth}\sigma^0$
	(*2.11a)	r^{-1}	$X^{03} = X^{04} = 0$ (see text); $\gamma^0 = -\tfrac{1}{2}(\log P)_{,1}$
	(*2.14g)	r^{-2}	$\lambda^0 = \bar{\sigma}^0_{,1} + 2\gamma^0\bar{\sigma}^0$
C	(*2.14c)	r^{-2}	$\mu^0 = -\eth\bar{\eth}\log P$
	(*2.11d)	r^{-1}	$\mu^0 = \bar{\mu}^0$ (trivial)
		r^{-2}	$\Psi'^0_2 - \bar{\Psi}'^0_2 = \eth\bar{\omega}^0 - \bar{\eth}\omega^0 + \bar{\sigma}^0\lambda^0 - \sigma^0\bar{\lambda}^0$
	(*2.14h)	r^{-2}	$U^0 = \mu^0$
	(*2.14d)	r^{-2}	$\Psi'^0_3 = \eth\lambda^0 - \bar{\eth}\mu^0$
D	(*2.11b)	r^0	$\nu^0 = -\bar{\eth}(\gamma^0 + \bar{\gamma}^0)$
	(*2.14a)	r^{-1}	$\Psi'^0_4 = -\bar{\eth}\nu^0 - \lambda^0_{,1} - 4\gamma^0\lambda^0$

At this stage the remaining undetermined data are the Ψ'^n_0 and the initial functions P, σ^0, Ψ'^0_1, and $\Psi'^0_2 + \bar{\Psi}'^0_2$, all as functions of u and x^k. Finally, the "u-derivative equations" (*2.16) are invoked. Of these, (*2.16a) may be considered after groups C of the radial and nonradial equations, the others after groups D. Equation (*2.16a) determines $\partial\Psi'^n_0/\partial u$ for all n, showing that Ψ'_0 need only be given on one null hypersurface $u = \text{const}$, its propagation off this hypersurface being then completely determined. The other three u-derivative equations are treated similarly to the nonradial equations, only the first nontrivial coefficient of each being equated to zero, to obtain the u-dependence of Ψ'^0_1 and Ψ'^0_2. These results are given in Table III. The first nontrivial coefficient in (*2.16d) yields no new information.

After this stage of the calculations, Newman and Unti found that all the higher coefficients of the nonradial and u-derivative equations which they evaluated were identically satisfied in consequence of relations

TABLE III. Integration of u-derivative equations.

Group	Equation	Term used	Result
C	(*2.16a)	r^{-5}	$P^3\frac{\partial}{\partial u}(P^{-3}\Psi^0_0) = -\eth\Psi^0_1 + 3\sigma^0\Psi^0_2$
		r^{-5-n}	Equation for $\frac{\partial}{\partial u}\Psi^n_0$
D	(*2.16b)	r^{-4}	$P^3\frac{\partial}{\partial u}(P^{-3}\Psi^0_1) = -\eth\Psi^0_2 + 2\sigma^0\Psi^0_3$
	(*2.16c)	r^{-3}	$P^3\frac{\partial}{\partial u}(P^{-3}\Psi^0_2) = -\eth\Psi^0_3 + \sigma^0\Psi^0_4$

already derived, and they concluded that they had extracted all the information contained in the field equations. For any fixed value of N it would be possible, but very laborious, to verify this directly, but Newman and Unti did not even carry the case $N = 0$ to completion in this way, and for $N > 0$ the labor required would be enormous. The purpose of the present paper is to give a general proof that, for arbitrary N, the remaining coefficients vanish identically, and this is done in the following section.

4. PROOF OF COMPLETENESS

To begin this discussion it is simplest to assume that the term in (3.6) of $O(r^{-N-6})$ is explicitly known (although not necessarily as a power series in r^{-1}), and that the integration of the radial equations and of the first u-derivative equation (*2.16a) has been carried out exactly. The data needed to perform the NU integration are then three functions of three variables, namely $P(u, x^k)$ and $\sigma^0(u, x^k)$ for all u and Ψ^0_0 on a null hypersurface $u = u_0$, say, and two functions of two variables, namely Ψ^0_1 and $\Psi^0_2 + \bar{\Psi}^0_2$ for $u = u_0$. However, as yet we have not shown that these data may be specified arbitrarily; there may still be interrelations between them contained in those higher coefficients of the nonradial and u-derivative equations which have not yet been evaluated.

To prove that no such extra interrelations actually exist, we start from the observation at the beginning of Sec. 3: that whatever values are chosen for the metric coefficients, provided (3.1) is satisfied, they do define a space-time and a null tetrad system throughout it. For this, as any other space-time, Eqs. (2.10), (2.11), and (2.13)–(2.17) form a consistent set of equations. If some of these are used to evaluate the spin coefficients and the Ψ_A, Φ_{AB}, and Λ, then the resulting quantities identically satisfy the remaining equations. In particular, if we can deduce from these equations that

$$\epsilon = \pi = 0, \quad \Phi_{AB} = 0, \quad \Lambda = 0, \tag{4.1}$$

then all the equations used in the NU procedure are identically satisfied. So our task is as follows: assuming the expressions for the metric coefficients obtained by the NU procedure, we must deduce (4.1). Completeness of the NU procedure is then assured.

Now, in practice, the NU integration is not usually carried out exactly, so that the remainder terms whose evaluation requires knowledge of the $O(r^{-N-6})$ terms in Ψ^0_0 are not known. As remarked above, these remainders are $O(r^{-N-5})$ in X^k and ξ^k and $O(r^{-N-4})$ in U and ω. It also is useful to know to what accuracy (4.1) is satisfied if the metric coefficients are only known to this order of magnitude. For this reason, we assume only that the NU procedure has been carried as far as possible without explicit knowledge of the $O(r^{-N-6})$ terms of Ψ^0_0. For example, (*2.13a) and (*2.13b) are only assumed satisfied with an error term $O(r^{-N-6})$. The proof for the exact case is obtained by formally setting $N = \infty$, as it does not depend on Ψ^0_0 being given in the form of a power series, but only on the accuracy to which it is known.

During the proof it is necessary to consider simultaneously the values of quantities calculated in the NU procedure and the values which actually follow from the assumed form of the metric coefficients. To distinguish between them, all quantities obtained in the NU procedure are indicated by a caret. Thus the integration of (*2.13a), (*2.13b), and (*2.14b) implies that

$$D\hat{\rho} = \hat{\rho}^2 + \hat{\sigma}\hat{\bar{\sigma}} + O(r^{-N-6}), \tag{4.2a}$$

$$D\hat{\sigma} = 2\hat{\rho}\hat{\sigma} + \hat{\Psi}^0_0 + O(r^{-N-6}), \tag{4.2b}$$

and

$$\hat{\delta}\hat{\rho} - \hat{\bar{\delta}}\hat{\sigma} = \hat{\rho}\hat{\tau} - \hat{\sigma}(3\hat{\alpha} - \hat{\bar{\beta}}) - \hat{\Psi}^0_1 + O(r^{-4}), \tag{4.3}$$

whereas the unaccented ρ and σ satisfy (2.13a), (2.13b), and (2.14b) in exactly the form given in Sec. 2.

The proof is most easily carried out by considering the situation at the end of each of the groups A, B, C, and D of integrations indicated in Tables I–III. After group A, the only known NU metric coefficients are the $\hat{\xi}^k$. We thus take

$$\xi^k = \hat{\xi}^k + O(r^{-N-5}) \tag{4.4}$$

and assume also that three radial derivatives of the remainder term may be formally taken, as well as as many derivatives in the x^k directions as necessary. This can be proved to hold for the remainder term in $\hat{\xi}^k$. Then (*2.10a) and (2.10a) give, respectively,

$$D\hat{\xi}^k = \hat{\rho}\hat{\xi}^k + \hat{\sigma}\hat{\bar{\xi}}^k + O(r^{-N-6}) \tag{4.5a}$$

and

$$D\xi^k = (\rho + \epsilon - \bar{\epsilon})\xi^k + \sigma\bar{\xi}^k. \tag{4.5b}$$

But by hypothesis $D(\xi^k - \hat{\xi}^k) = O(r^{-N-6})$; hence, because $\hat{\rho} = O(r^{-1})$ and $\hat{\sigma} = O(r^{-2})$, we get

$$(\rho - \hat{\rho} + \epsilon - \bar{\epsilon})\xi^k + (\sigma - \hat{\sigma})\bar{\xi}^k = O(r^{-N-6}). \tag{4.6}$$

Now $\xi^k = O(r^{-1})$ and $\xi^3\bar{\xi}^4 - \bar{\xi}^3\xi^4 = -2iP^2r^{-2} + O(r^{-3})$, so that Eqs. (4.6) may be solved to give

$$\rho - \hat{\rho} + \epsilon - \bar{\epsilon} = O(r^{-N-5}) \tag{4.7}$$

and

$$\sigma - \hat{\sigma} = O(r^{-N-5}). \tag{4.8}$$

But by (2.9) ρ and $\hat{\rho}$ are real and ϵ is purely imaginary. Equation (4.7) thus separates into its real and imaginary parts to give

$$\rho - \hat{\rho} = O(r^{-N-5}) \quad \text{and} \quad \epsilon = O(r^{-N-5}). \tag{4.9}$$

Two formal radial derivatives of (4.8) and (4.9) are also valid, which is also true for all similar orders of magnitude which will be obtained for spin coefficients. If the NU integrations were performed exactly, (4.8) and (4.9) would show that the form obtained for ξ^k after group A would already ensure that $\epsilon = 0$ and that the forms obtained for ρ and σ correctly followed from this value for ξ^k. Subtraction of (4.2a) and (4.2b) from (2.13a) and (2.13b), respectively, now also gives

$$\Phi_{00} = O(r^{-N-6}) \quad \text{and} \quad \Psi'_0 - \hat{\Psi}'_0 = O(r^{-N-6}), \tag{4.10}$$

showing in the exact version that at this stage one field equation is already satisfied and that Ψ'_0 is already determined and has the form initially assumed. One formal radial derivative of (4.10) is also valid, as is also true for all similar orders of magnitude which will be obtained for curvature tensor components.

This uses up all the information in group A. In group B we additionally calculate $\hat{\omega}$ and $\hat{X}^k$, so we now also take

$$\omega = \hat{\omega} + O(r^{-N-4}) \quad \text{and} \quad X^k = \hat{X}^k + O(r^{-N-5}), \tag{4.11}$$

again with three formal radial derivatives and as many x^k derivatives as necessary. Then, as above, a comparison of (*2.10b) and (*2.10c) with (2.10b) and (2.10c), respectively, implies

$$\tau - \hat{\tau} = O(r^{-N-5}) \quad \text{and} \quad \pi = O(r^{-N-5}). \tag{4.12}$$

Equations (2.13c) and (*2.13c) now give

$$(\Psi'_1 - \hat{\Psi}'_1) + \Phi_{01} = O(r^{-N-6}). \tag{4.13}$$

From now on the argument becomes more tortuous. Use of (2.12) and (3.8) shows that $\delta\bar{\xi}^k = O(r^{-2})$, with the help of which (2.11c) implies $\bar{\alpha} - \beta = O(r^{-1})$. But $\tau = \bar{\alpha} + \beta$; thus, on using (4.12) and (3.8), we now get separately $\alpha, \beta = O(r^{-1})$. Together with (4.9) and (4.12) this shows that the terms in (2.13d) involving ϵ and π are $O(r^{-N-6})$, which is the accuracy to which (*2.13d) is integrated. Again by using $\tau = \bar{\alpha} + \beta$ with (4.12), a comparison of (2.13d) with (*2.13d) now yields

$$D(\alpha - \hat{\alpha}) = \rho(\alpha - \hat{\alpha}) - \bar{\sigma}(\bar{\alpha} - \hat{\bar{\alpha}}) + \Phi_{10} + O(r^{-N-6}). \tag{4.14}$$

Now treat similarly the final radial equation (2.15a) used in group B. First observe that (2.14b) implies $\Psi'_1 - \Phi_{01} = O(r^{-3})$, which, together with (4.13) and (3.8), gives $\Psi'_1, \Phi_{01} = O(r^{-3})$ separately. Equation (2.15a) thus simplifies to

$$D(\Phi_{01} - \Psi'_1) + \delta\Psi'_0 = 4\alpha\Psi'_0 - 4\rho\Psi'_1 + 2\rho\Phi_{01} + 2\sigma\Phi_{10} + O(r^{-N-7}). \tag{4.15}$$

Subtract (*2.15a) from this and use (4.13) to give

$$D\Phi_{01} = 3\rho\Phi_{01} + \sigma\Phi_{10} + 2(\alpha - \hat{\alpha})\Psi'_0 + O(r^{-N-7}). \tag{4.16}$$

Using the technique of NU for integrating radial equations, we may integrate (4.14) and (4.16) simultaneously for Φ_{01} and $(\alpha - \hat{\alpha})$. Note that the other quantities ρ, σ, and Ψ'_0 occurring in the equations are now known to the required accuracy. This integration shows that, if it is known that

$$\alpha - \hat{\alpha} = O(r^{-2}) \tag{4.17}$$

and

$$\Phi_{01} = O(r^{-4}), \tag{4.18}$$

then it follows that

$$\alpha - \hat{\alpha} = O(r^{-N-5}) \quad \text{and} \quad \Phi_{01} = O(r^{-N-6}). \tag{4.19}$$

Now group B includes satisfying the r^{-2} terms of (*2.11c), which, with (2.11c) and (4.12), gives

$$(\alpha - \hat{\alpha})\xi^k - (\bar{\alpha} - \hat{\bar{\alpha}})\bar{\xi}^k = O(r^{-3}),$$

from which (4.17) follows. Also the r^{-3} terms of (*2.14b) are satisfied, which, with (2.14b) and (4.17), gives

$$(\Psi'_1 - \hat{\Psi}'_1) - \Phi_{01} = O(r^{-4}).$$

Together with (4.13) this implies (4.18). Equations (4.19) thus follow, which, with (4.12) and (4.13), also give

$$\beta - \hat{\beta} = O(r^{-N-5}), \quad \Psi'_1 - \hat{\Psi}'_1 = O(r^{-N-6}). \tag{4.20}$$

We now have α, β, ρ, σ, τ, Ψ_0', and Ψ_1' correctly given by the NU procedure, with Φ_{00} and Φ_{01} both of $O(r^{-N-6})$. This completes group *B*.

Group *C* is the most complicated one to treat. We here add in

$$U = \hat{U} + O(r^{-N-4}), \tag{4.21}$$

again with three radial derivatives and as many x^k derivatives as are needed. Equations (2.10d) and (*2.10d) immediately give

$$\mathrm{Re}\,(\gamma - \hat{\gamma}) = O(r^{-N-5}); \tag{4.22}$$

then, on remembering that ϵ is purely imaginary, comparison of the real parts of (2.13f) and (*2.13f) gives

$$\mathrm{Re}\,(\Psi_2' - \hat{\Psi}_2') - \Lambda + \Phi_{11} = O(r^{-N-6}). \tag{4.23}$$

The entire equations (2.13f) and (*2.13f) then give

$$D(\gamma - \hat{\gamma}) = \mathrm{Im}\,(\Psi_2' - \hat{\Psi}_2') + O(r^{-N-5}), \tag{4.24}$$

where $\mathrm{Re}\ \varphi = \frac{1}{2}(\varphi + \bar{\varphi})$, $\mathrm{Im}\ \varphi = \frac{1}{2}(\varphi - \bar{\varphi})$, for any quantity φ. Now satisfying the r^{-1} and r^{-2} terms of (*2.11d) gives

$$\mathrm{Im}\,(\mu - \hat{\mu}) = O(r^{-3}), \tag{4.25}$$

whilst satisfying the r^{-1} terms of (*2.11a) gives

$$\mu - \hat{\mu} - 2\,\mathrm{Im}\,(\gamma - \hat{\gamma}) = O(r^{-1}) \tag{4.26}$$

and $\lambda = O(r^{-1})$. Noting that $\hat{\lambda}$, $\hat{\mu}$, $\mathrm{Im}\,\hat{\gamma} = O(r^{-1})$, we see that Eqs. (4.25) and (4.26) together give μ, $\mathrm{Im}\,\gamma = O(r^{-1})$, and thus also $\gamma - \hat{\gamma} = O(r^{-1})$ on using (4.22). Comparison of the equations (2.13g) and (2.13h) with (*2.13g) and (*2.13h) now yields

$$D(\lambda - \hat{\lambda}) = \rho(\lambda - \hat{\lambda}) + \bar{\sigma}(\mu - \hat{\mu}) + \Phi_{20} + O(r^{-N-6}) \tag{4.27}$$

and

$$D(\mu - \hat{\mu}) = \rho(\mu - \hat{\mu}) + \sigma(\lambda - \hat{\lambda}) + \Psi_2' - \hat{\Psi}_2' + 2\Lambda + O(r^{-N-6}). \tag{4.28}$$

Turning next to the nonradial equations, by satisfying the r^{-2} terms of (*2.14g) we ensure that

$$\rho(\lambda - \hat{\lambda}) + \Phi_{20} = O(r^{-3}). \tag{4.29}$$

With (4.27) this implies $D(\lambda - \hat{\lambda}) = O(r^{-3})$, which integrates to give

$$\lambda - \hat{\lambda} = O(r^{-2}). \tag{4.30}$$

With (4.29), this implies

$$\Phi_{02} = O(r^{-3}). \tag{4.31}$$

Next, satisfy the r^{-2} terms of (*2.14h). The r^{-3} terms are then identically satisfied in virtue of the values of γ^0 and λ^0 already obtained from (*2.11a) and (*2.14g); hence, on using (4.30), we obtain

$$\rho(\bar{\mu} - \hat{\bar{\mu}}) + \Psi_2' - \hat{\Psi}_2' + 2\Lambda = O(r^{-4}). \tag{4.32}$$

Together with (4.28) this implies

$$D(\mu - \hat{\mu}) = 2\rho\,\mathrm{Im}\,(\mu - \hat{\mu}) + O(r^{-4}),$$

which, on using (4.25), integrates to give

$$\mu - \hat{\mu} = O(r^{-3}). \tag{4.33}$$

Put this back into (4.32) and use (4.23) to obtain

$$\mathrm{Im}\,(\Psi_2' - \hat{\Psi}_2') = O(r^{-4}) \quad \text{and} \quad \Phi_{11} - 3\Lambda = O(r^{-4}). \tag{4.34}$$

The r^{-2} terms of (*2.14c) are next satisfied, giving

$$(\Psi_2' - \hat{\Psi}_2') - \Lambda - \Phi_{11} = O(r^{-3}),$$

which, with (4.23) and (4.34), shows that Φ_{11}, $\Lambda = O(r^{-3})$.

The next step is to derive radial equations for Λ, Φ_{11}, Φ_{02}, and $\mathrm{Im}\,(\Psi_2' - \hat{\Psi}_2')$. Comparison of the imaginary parts of (2.15b) and (*2.15b) gives

$$D\,\mathrm{Im}\,(\Psi_2' - \hat{\Psi}_2') = 3\rho\,\mathrm{Im}\,(\Psi_2' - \hat{\Psi}_2') - \mathrm{Im}\,[(\lambda - \hat{\lambda})\Psi_0' + \sigma\Phi_{20}] + O(r^{-N-7}). \tag{4.35}$$

Their real parts, together with (2.17a), give the two equations

$$D(\Phi_{11} - \Lambda) = 2\rho(\Phi_{11} - \Lambda) + \tfrac{2}{3}\,\mathrm{Re}\,[(\lambda - \hat{\lambda})\Psi_0'] + O(r^{-N-6}) \tag{4.36}$$

and

$$D(\Phi_{11} + 3\Lambda) = \rho(\Phi_{11} + 3\Lambda) + 3\rho(\Phi_{11} - \Lambda) + 2\,\mathrm{Re}\,(\sigma\Phi_{20}) + O(r^{-N-6}). \tag{4.37}$$

Finally, satisfaction of the u-derivative equation (*2.16a) with a remainder $O(r^{-N-6})$ gives

$$D\Phi_{02} = \rho\Phi_{02} + 4(\gamma - \hat{\gamma})\Psi_0' - (\mu - \hat{\mu})\Psi_0' + 3\sigma(\Psi_2' - \hat{\Psi}_2') + 2\sigma\Phi_{11} + O(r^{-N-6}). \tag{4.38}$$

In deriving these last four equations it is necessary to use many of the orders of magnitude previously calculated. We now have a system of seven radial equations, (4.36), (4.38), (4.37), (4.35), (4.24), (4.27), and (4.28), which may be simultaneously integrated by iterating them sequentially in this order. Starting from the initial conditions

$$\Phi_{11}, \Phi_{02}, \Lambda = O(r^{-3}), \quad \mathrm{Im}\,(\Psi_2' - \hat{\Psi}_2') = O(r^{-4}), \quad \gamma - \hat{\gamma} = O(r^{-1}), \quad \lambda - \hat{\lambda} = O(r^{-2}), \quad \mu - \hat{\mu} = O(r^{-3}), \tag{4.39}$$

which have already been derived, and using (4.23) during the iteration, we then obtain

$$\Phi_{11}, \Phi_{02}, \Lambda, \Psi_2 - \hat{\Psi}_2 = O(r^{-N-5}),$$
$$\gamma - \hat{\gamma}, \lambda - \hat{\lambda}, \mu - \hat{\mu} = O(r^{-N-4}). \quad (4.40)$$

A further application of (4.35), and then also of the imaginary part of (4.28), yields the slightly more stringent results

$$\Psi_2 - \hat{\Psi}_2 - \Lambda + \Phi_{11} = O(r^{-N-6})$$

and

$$\mathrm{Im}\,(\mu - \hat{\mu}) = O(r^{-N-5}), \quad (4.41)$$

but these cannot be extended to improve the results (4.40) themselves.

Finally, we come to group D. First deduce $\nu = O(1)$ from (2.11b) and then $\Phi_{12} = O(r^{-1})$ from (2.14f). With these, (2.17b) simplifies to

$$D\Phi_{12} = 3\rho\Phi_{12} + \sigma\Phi_{21} + O(r^{-N-6}). \quad (4.42)$$

Now compare (2.15c) with (*2.15c) and use (4.42) to put the result in the form

$$D(\Psi_3 - \hat{\Psi}_3 - \Phi_{21}) = 2\rho(\Psi_3 - \hat{\Psi}_3 - \Phi_{21}) + O(r^{-N-6}). \quad (4.43)$$

But in group D we satisfy the r^{-2} terms of (*2.14d), which gives

$$\Psi_3 - \hat{\Psi}_3 - \Phi_{21} = O(r^{-3}).$$

Using this, we integrate (4.43) to give

$$\Psi_3 - \hat{\Psi}_3 - \Phi_{21} = O(r^{-N-5}). \quad (4.44)$$

We also satisfy the r^{-4} terms of (*2.16b), from which

$$D\Phi_{12} = 3\sigma(\Psi_3 - \hat{\Psi}_3) + \sigma\Phi_{21} + O(r^{-5}) = 4\sigma\Phi_{21} + O(r^{-5})$$

by (4.44). Using $\Phi_{12} = O(r^{-1})$, we integrate to give $\Phi_{12} = O(r^{-4})$, which may be used as the initial condition in (4.42). On using (4.44) again, we obtain

$$\Phi_{12} = O(r^{-N-5}) \quad \text{and} \quad \Psi_3 - \hat{\Psi}_3 = O(r^{-N-5}). \quad (4.45)$$

Together with (2.13i) and (*2.13i), these now give $D(\nu - \hat{\nu}) = O(r^{-N-5})$. But satisfying the r^0 terms of (*2.11b) gives $\nu - \hat{\nu} = O(r^{-1})$, so that this integrates to give

$$\nu - \hat{\nu} = O(r^{-N-4}). \quad (4.46)$$

Next use the vanishing of the r^{-1} terms of (*2.14a) to obtain

$$\Psi_4 - \hat{\Psi}_4 = O(r^{-2}). \quad (4.47)$$

With this, the vanishing of the r^{-3} terms of (*2.16c) gives

$$D\Phi_{22} = -\rho\Phi_{22} + O(r^{-4}). \quad (4.48)$$

But (2.14e) gives $\Phi_{22} = O(r^{-1})$, so that (4.48) integrates to give $\Phi_{22} = O(r^{-3})$. Now (2.17c) reduces to

$$D\Phi_{22} = 2\rho\Phi_{22} + O(r^{-N-5}).$$

By using $\Phi_{22} = O(r^{-3})$ as initial condition, this implies

$$\Phi_{22} = O(r^{-N-4}). \quad (4.49)$$

Finally, comparison of (2.15d) with (*2.15d) gives

$$D(\Psi_4 - \hat{\Psi}_4) = \rho(\Psi_4 - \hat{\Psi}_4) + O(r^{-N-5}),$$

which, on using (4.47), integrates to give

$$\Psi_4 - \hat{\Psi}_4 = O(r^{-N-4}). \quad (4.50)$$

This completes the proof. We have now shown that, if the NU integration is performed exactly, then the metric coefficients obtained generate an empty space–time, together with a null tetrad satisfying (3.2), and that the correct values are obtained for the spin coefficients and Weyl tensor components of this metric. If the remainder terms are not explicitly calculated, then we have discovered the accuracy to which the field equations are satisfied and to which the spin coefficients and Weyl tensor components are obtained. A summary of these results is given in Table IV. Within each group of NU integrations, the corresponding completeness proof divides naturally into two parts. Some results are obtained by direct comparison of starred and unstarred radial equations, while the others require integration of a radial equation using initial conditions which are derived from the nonradial and u-derivative equations of NU. Table IV also shows the group in which each result is derived and whether it follows by direct comparison or needs such a radial integration.

5. DISCUSSION

The above proof makes clear the reason why some NU equations have to be satisfied exactly, while for others only the first nontrivial term need be satisfied. Those equations which are satisfied exactly either immediately ensure that some spin coefficient or Weyl tensor component is correctly given by the NU procedure, or else give a radial equation for the error in such a quantity. These radial equations may be integrated to prove that such an error is actually zero, provided that as an initial condition we know that that error vanishes at least as rapidly as some power of r^{-1}, which power is determined by the radial equation. It is these initial conditions which are provided by those NU equations which need only be satisfied to their lowest order.

TABLE IV. Summary of results.

Group	Derivation	$O(r^{-N-4})$	$O(r^{-N-5})$	$O(r^{-N-6})$	Exact
	geometry				$\rho = \bar{\rho}$, $\kappa = 0$, $\epsilon + \hat{\epsilon} = 0$, $\tau = \bar{\alpha} + \beta$,
A	assumption		$\xi^k - \hat{\xi}^k$		
	comparison		$\rho - \hat{\rho}$, $\sigma - \hat{\sigma}$, ϵ	Φ_{00}, $\Psi_0 - \hat{\Psi}_0$	
B	assumption	$\omega - \hat{\omega}$	$X^k - \hat{X}^k$		
	comparison		$\tau - \hat{\tau}$, π		
	integration		$\alpha - \hat{\alpha}$, $\beta - \hat{\beta}$	Φ_{01}, $\Psi_1 - \hat{\Psi}_1$	
C	assumption	$U - \hat{U}$			
	comparison		$\mathrm{Re}\,(\gamma - \hat{\gamma})$	$\mathrm{Re}\,(\Psi_2 - \hat{\Psi}_2) - \Lambda + \Phi_{11}$	
	integration	$\gamma - \hat{\gamma}$, $\lambda - \hat{\lambda}$, $\mu - \hat{\mu}$	Φ_{11}, Φ_{02}, Λ, $\Psi_2 - \hat{\Psi}_2$, $\mathrm{Im}\,(\mu - \hat{\mu})$	$\mathrm{Im}\,(\Psi_2 - \hat{\Psi}_2)$	
D	integration	$\nu - \hat{\nu}$, $\Psi_4 - \hat{\Psi}_4$, Φ_{22}	$\Psi_3 - \hat{\Psi}_3$, Φ_{12}		

For completeness, it should be noted that the values of the initial functions given in Table II can be simplified by a suitable choice of the hypersurfaces $u =$ const. It is shown in NU that this choice can be made so that $\partial P/\partial u = 0$. In particular, this gives $\gamma^0 = \nu^0 = 0$, and thus by (3.8) makes $\hat{\gamma} = O(r^{-2})$ and $\hat{\nu} = O(r^{-1})$. However, it would not have simplified the above proof if we had used this further restriction on the coordinate system, and so we did not introduce it earlier.

APPENDIX: NOTATION AND CONVENTIONS

The signature of space–time is taken as $- - - +$. The curvature tensor is taken so that the Ricci identity for a covariant vector A_α is

$$\nabla_\alpha \nabla_\beta A_\gamma - \nabla_\beta \nabla_\alpha A_\gamma = R_{\alpha\beta\gamma}^{\cdots\ \ \ \delta} A_\delta .$$

The Ricci tensor $R_{\beta\gamma}$, curvature scalar R, and Weyl tensor $C_{\alpha\beta\gamma\delta}$ are then defined by

$$R_{\beta\gamma} = R_{\alpha\beta\gamma}^{\cdots\ \ \ \alpha}, \quad R = g^{\beta\gamma} R_{\beta\gamma},$$

$$C_{\alpha\beta\gamma\delta} = R_{\alpha\beta\gamma\delta} + \tfrac{1}{2}(g_{\alpha\gamma} R_{\beta\delta} - g_{\alpha\delta} R_{\beta\gamma} + g_{\beta\delta} R_{\alpha\gamma} - g_{\beta\gamma} R_{\alpha\delta}) + \tfrac{1}{6} R (g_{\alpha\delta} g_{\beta\gamma} - g_{\alpha\gamma} g_{\beta\delta}).$$

In Tables II and III the operators ∇ and $\eth$ are used. These are defined in the context of the coordinate system introduced in Sec. 2, when the further specialization (3.7) has been made, but before the reality of ξ^{03} is imposed. At this stage the freedom remaining in the choice of tetrad system is merely that of a rotation in the $(m, \bar{m})$ plane:

$$l^\mu \to l^\mu, \quad n^\mu \to n^\mu, \quad m^\mu \to e^{i\psi} m^\mu, \quad \bar{m}^\mu \to e^{-i\psi} \bar{m}^\mu, \tag{A1}$$

where ψ is an arbitrary real function of u and x^k. A quantity η, referred to this tetrad system, is said to have *spin weight* s if, under the transformation (A1),

$$\eta \to e^{is\psi} \eta. \tag{A2}$$

Suppose $s \geq 0$. Then the real quantity defined by

$$\eta_{\alpha_1 \cdots \alpha_s} = \eta \bar{m}_{\alpha_1} \cdots \bar{m}_{\alpha_s} + \bar{\eta} m_{\alpha_1} \cdots m_{\alpha_s} \tag{A3}$$

is independent of ψ. It is thus a well-defined tensor of rank s, from which η may be reconstructed:

$$\eta = \eta_{\alpha_1 \cdots \alpha_s} m^{\alpha_1} \cdots m^{\alpha_s}. \tag{A4}$$

On using (2.3) and (2.12), we get from (A3) that

$$(m^\beta \nabla_\beta \eta_{\alpha_1 \cdots \alpha_s}) m^{\alpha_1} \cdots m^{\alpha_s} = \eth \eta + s(\bar{\alpha} - \beta)\eta \tag{A5}$$

and

$$(\bar{m}^\beta \nabla_\beta \eta_{\alpha_1 \cdots \alpha_s}) m^{\alpha_1} \cdots m^{\alpha_s} = \bar{\eth} \eta - s(\alpha - \bar{\beta})\eta. \tag{A6}$$

We also observe that the left-hand sides, and thus also the right-hand sides, of (A5) and (A6) are quantities of spin weights $(s + 1)$ and $(s - 1)$, respectively. Hence, if their values are known for one particular choice of orientation of $(m, \bar{m})$, they can be obtained for any other orientation by using (A2).

Now suppose that η is independent of r. This is well defined, as the transformation (A2) only allows

functions ψ which are themselves independent of r. We can then use (3.8), together with the value of α^0 given in Table II, to evaluate the right-hand sides of (A5) and (A6) for the orientation of m and $\bar{m}$ which makes $\xi^{03} = P$ real. This is the choice originally made in Sec. 3. We obtain

$$\delta\eta + s(\bar{\alpha} - \beta)\eta = -r^{-1}\eth\eta + r^{-2}\sigma^0\bar{\eth}\eta + O(r^{-3}) \quad \text{(A7)}$$

and

$$\bar{\delta}\eta - s(\alpha - \bar{\beta})\eta = -r^{-1}\bar{\eth}\eta + r^{-2}\bar{\sigma}^0\eth\eta + O(r^{-3}), \quad \text{(A8)}$$

where $\eth$ and $\bar{\eth}$ are defined by

$$\eth\eta = -P^{1-s}\nabla(P^s\eta) \quad \text{and} \quad \bar{\eth}\eta = -P^{1+s}\bar{\nabla}(P^{-s}\eta), \quad \text{(A9)}$$

with $\nabla = (\partial/\partial x^3) + i(\partial/\partial x^4)$ and $\bar{\nabla}$ being the complex conjugate of ∇.

The above discussion only holds if $s \geq 0$. But if $s < 0$, we may apply (A7) through (A9) to $\bar{\eta}$, which has positive spin weight $-s$. We then define

$$\eth\eta = (\bar{\eth}\bar{\eta})^* \quad \text{and} \quad \bar{\eth}\eta = (\eth\bar{\eta})^*, \quad \text{(A10)}$$

where, to avoid undue complexity in printing, * and $^-$ are both used to denote complex conjugation. With this definition we see that (A7) through (A10) now hold for all s.

If a different choice is made for ψ, then $\eth\eta$ and $\bar{\eth}\eta$ are taken as defined by the expansions (A7) and (A8). As each term in these expansions must have the same spin weight, this shows that $\eth\eta$ and $\bar{\eth}\eta$ also have spin weights $(s + 1)$ and $(s - 1)$, respectively. The operators $\eth$ and $\bar{\eth}$ are only defined when acting on quantities which are independent of r and which have well-defined spin weights. They were introduced by Newman and Penrose[11] and have been more extensively studied by Goldberg *et al.*[12] The results of NU are all expressed in terms of ∇, but many of them take a much simpler form when $\eth$ is used. Not all the spin coefficients have well-defined spin weights. Those that do are σ:2; κ, τ:1; $\rho, \mu, \epsilon + \bar{\epsilon}, \gamma + \bar{\gamma}$:0; π, ν:-1; λ:-2. Ψ_A and Φ_{AB} have spin weights $2 - A$ and $B - A$, respectively. The initial functions σ^0, etc., have the same spin weight as the corresponding spin coefficient, being defined in terms of the asymptotic power series of that spin coefficient.

[1] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962), hereafter referred to as NP.
[2] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962), hereafter referred to as NU.
[3] E. T. Newman and R. Penrose, Proc. Roy. Soc. (London) **A305**, 175 (1968).
[4] A. I. Janis and E. T. Newman, J. Math. Phys. **6**, 902 (1965).
[5] E. T. Newman and T. W. J. Unti, J. Math. Phys. **6**, 1806 (1965).
[6] E. T. Newman and R. Penrose, Phys. Rev. Letters **15**, 231 (1965).
[7] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[8] R. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962).
[9] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962).
[10] A. Trautman, F. A. E. Pirani, and H. Bondi, *Lectures on General Relativity, at the 1964 Brandeis Summer Institute in Theoretical Physics* (Prentice-Hall Inc., Englewood Cliffs, N.J., 1965).
[11] E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[12] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Representations of Noncompact Semisimple Lie Groups*

DAVID W. JOSEPH
Behlen Laboratory of Physics, University of Nebraska, Lincoln, Nebraska

(Received 10 February 1969)

A method is presented for the construction of unitary representations of semisimple Lie groups (or, more precisely, of the corresponding algebras), proceeding directly from the commutation relations among the canonical generators $e_{\pm\alpha}$ and h_α. In the case of the orthogonal groups, the correspondence between the canonical generators and the more usual tensor generators is written down explicitly. It is shown how this method can be used to construct a certain class of representations (bounded above or below) of the Lie algebras of the noncompact groups $SU(p, q)$ and $SO(p, q)$, which arise in the dynamical group treatment of certain physical systems.

1. INTRODUCTION

Noncompact groups and their algebras have arisen in connection with models for hadron spectra and transition amplitudes,[1,2] strong-coupling models,[3] the properties of Regge trajectories at $t = 0$,[4] and the Bethe–Salpeter equations,[5] among other structures.[6] It appears likely that further applications will arise as we become more familiar with such groups. Although properties of unitary representations of compact groups (and finite-dimensional representations in general) have been studied in terms of the commutation relations among the canonical generators $e_{+\alpha}$, $e_{-\alpha}$, and h_α,[7] it seems that little use has so far been made of these commutation relations for the explicit construction of unitary representations of noncompact groups (or their algebras). The purpose of the present article is to describe rules for constructing certain unitary representations of noncompact groups in this way, proceeding directly from the canonical commutation relations.

Section 2 presents a brief review of the general structure of semisimple Lie algebras. Section 3 shows how this structure can be used to construct representations of the algebra A_1. Section 4 derives a method of constructing bounded representations of more general semisimple Lie algebras. Sections 5 and 6 carry this out in more detail for the representations of A_n [corresponding to the groups $SU(p, q)$], and of B_n and D_n [corresponding to the groups $SO(p, q)$], respectively.[8] In particular, in Sec. 6 the canonical generators are related to the usual tensor generators, and the hermiticity properties of the former are derived for unitary representations. Section 7 illustrates these methods by using them to obtain some previously known results.[5,9] Finally, the Appendices review the representations of the rank-one groups (as opposed to the algebras); this material is needed for some applications of larger groups.[2]

2. THE STRUCTURE OF SEMISIMPLE LIE ALGEBRAS

The Lie algebra arises from expressing Lie group elements in exponential form:

$$a = e^{\sum_j \alpha_j X_j}. \tag{2.1}$$

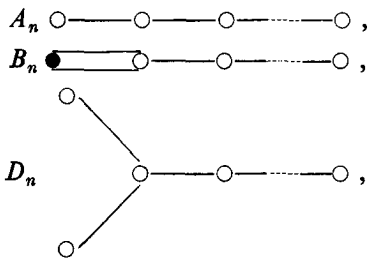
The α_j are the parameters of the group, and the generators X_j constitute a basis for the corresponding Lie algebra, which consists of all elements of the form of the exponent. As the product of two elements must be preserved in any representation of the group, so, correspondingly, the commutation relations among generators must be preserved in the representation of the algebra; these define the structure of the Lie algebra. The general classification of Lie algebras allows the parameters in (2.1) to be arbitrary complex numbers; imposing appropriate reality conditions on these parameters leads to groups such as the compact group $SU(p)$ or the noncompact groups $SU(p, q)$.

The general classification of simple Lie algebras[7] is into four sequences, called A_n, B_n, C_n, D_n, and the five "exceptional" algebras E_5, E_6, E_7, F_4, and G_2. We are particularly interested in the sequences A_n, B_n, and D_n, which correspond to the unitary groups, the odd-dimensional rotation groups, and the even-dimensional rotation groups, respectively, and to their noncompact versions:

$$\begin{aligned} A_n &\sim SU(n + 1 - q, q), \\ B_n &\sim SO(2n + 1 - q, q), \\ D_n &\sim SO(2n - q, q), \end{aligned} \tag{2.2}$$

where n is a positive integer ($n \geq 3$ for D_n) and where q, $n + 1 - q$, $2n + 1 - q$, and $2n - q$ are all nonnegative integers. The structure of these algebras can be summarized by the following Dynkin diagrams,

the significance of which will be explained below:



(2.3)

where n is equal to the total number of filled or open dots. The dots will be labeled α, β, γ, etc. To each dot (filled or open) there corresponds a triplet of generators, say $e_{\pm\alpha}$ and h_α, which satisfy the commutation relations

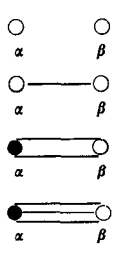
$$[e_{+\alpha}, e_{-\alpha}] = h_\alpha, \tag{2.4}$$

$$[h_\alpha, e_{\pm\alpha}] = \pm 2e_{\pm\alpha}. \tag{2.5}$$

We note that for $SU(2) \sim A_1$, which corresponds to just one dot, these are essentially angular momentum operators; $e_{\pm\alpha} = \sqrt{2}\, j_\pm$ and $h_\alpha = 2j_3$. Any two operators corresponding to two dots not connected by a line commute. However, the commutator of two generators corresponding to connected dots is generally not zero and in some cases defines a new generator.

A general semisimple Lie algebra is a direct sum of simple Lie algebras, with the generators corresponding to different summands commuting. The Dynkin diagram for a general semisimple Lie algebra is thus a set of connected diagrams, such as those in (2.3).

In order to specify the commutation relations completely, it is convenient to introduce the Cartan scalar product (α, β) defined on the $\alpha, \beta, \gamma, \cdots$ (called the simple positive roots), considered as basis vectors in a vector space of their own. Up to an over-all factor which is irrelevant, this scalar product is defined in terms of the kind of connection as follows:



$$(\alpha, \beta) = 0,$$

$$(\alpha, \beta) = -\tfrac{1}{2}(\alpha, \alpha) = -\tfrac{1}{2}(\beta, \beta),$$

$$(\alpha, \beta) = -(\alpha, \alpha) = -\tfrac{1}{2}(\beta, \beta),$$

$$(\alpha, \beta) = -\tfrac{3}{2}(\alpha, \alpha) = -\tfrac{1}{2}(\beta, \beta). \tag{2.6}$$

(The last kind of connection is listed for completeness; it occurs only for G_2, and so is not of interest here.) The set of all positive roots is defined to include the simple positive roots $\alpha, \beta, \gamma, \cdots$ and certain other expressions of the form

$$x = j\alpha + k\beta + l\gamma + \cdots, \tag{2.7}$$

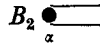
where $j, k, l, \cdots$ are nonnegative integers, not all zero. The expressions (2.7) which occur can be found from the rule that, if x is a positive root, then

$$x + j\alpha \tag{2.8}$$

is a positive root for integral j such that

$$0 \leq j \leq -\frac{2(x, \alpha)}{(\alpha, \alpha)}. \tag{2.9}$$

When constructing the positive roots according to this rule, it is sufficient to use only positive roots x such that $x - \alpha$ is not a positive root. As an example, we note that (2.6)–(2.9) applied to



(2.10)

yield the positive roots

$$\alpha, \quad \beta, \quad \alpha + \beta, \quad 2\alpha + \beta. \tag{2.11}$$

A complete set of generators (i.e., a basis for the Lie algebra) is given by the set of all h_α and $e_{\pm\alpha}$, where α is any simple positive root, x is any positive root, and the $e_{\pm x}$ for x not a simple root are defined by expressions such as

$$e_{\pm(\alpha+\beta+\gamma)} \equiv N'_{\alpha\beta\gamma}[[e_{\pm\alpha}, e_{\pm\beta}], e_{\pm\gamma}]. \tag{2.12}$$

Note that a definite order must be chosen for the commutators in definitions such as (2.12). The complete set of commutation relations can then be put in the form

$$[e_u, e_v] = \begin{cases} N_{uv} e_{u+v} & \text{with } N_{uv} \neq 0 \\ & \text{if } u + v \text{ is a root,} \\ 0 & \text{if } u + v \text{ is not a root} \\ & \text{and } u + v \neq 0, \end{cases} \tag{2.13a}$$

$$[e_\alpha, e_{-\alpha}] = h_\alpha, \tag{2.13b}$$

$$[h_\alpha, e_u] = \frac{2(\alpha, u)}{(\alpha, \alpha)} e_u, \tag{2.13c}$$

$$[h_\alpha, h_\beta] = 0, \tag{2.13d}$$

where u and v are roots (that is, expressions of the form $\pm x$, where x is a positive root) and α and β are simple positive roots.

3. THE ALGEBRA A_1 AND ITS REPRESENTATIONS

The algebra A_1, along with its associated groups, deserves special discussion before proceeding to the other algebras—not only because it provides the simplest example of the general method, but also because the content of its commutation relations can be taken over directly to the larger algebras.

The group $SL(2)$ is defined by the set of all 2×2 matrices with determinant $+1$. If we write a typical element in exponential form,

$$a = e^{ib}, \tag{3.1}$$

then ib is a 2×2 matrix with trace zero. (The i has been inserted for convenience later.) The set of all such 2×2 traceless matrices forms the Lie algebra A_1. The Pauli matrices constitute a basis for A_1; that is,

$$b = \sum_{i=1}^{\xi} \xi_i \sigma_i, \tag{3.2}$$

where the ξ_i are complex numbers. The definitions

$$e_\pm \equiv \tfrac{1}{2}(\sigma_1 \pm i\sigma_2) \quad \text{and} \quad h \equiv \sigma_3 \tag{3.3}$$

lead to the commutation relations

$$[h, e_\pm] = \pm 2e_\pm \quad \text{and} \quad [e_+, e_-] = h, \tag{3.4}$$

in agreement with (2.4) and (2.5), or (2.13b) and (2.13c). The expansion in terms of this basis will be written as

$$b = \xi_+ e_+ + \xi_- e_- + \xi_3 h, \tag{3.5}$$

where

$$\xi_\pm = \xi_1 \mp i\xi_2 . \tag{3.6}$$

The group $SL(2)$ contains several interesting subgroups. One is the subgroup which preserves the scalar product

$$x^\dagger y \equiv x_1^* y_1 + x_2^* y_2; \tag{3.7}$$

this is the subgroup $SU(2)$ of unitary 2×2 matrices with determinant equal to $+1$. From (3.1), the condition $a^\dagger a = I$ implies that b is Hermitian, i.e., $b^\dagger = b$. Thus the ξ's in (3.2) must be real, or, from (3.6),

$$\xi_\pm^* = \xi_\mp \quad \text{and} \quad \xi_3^* = \xi_3 \tag{3.8}$$

for this subgroup, $SU(2)$.

Another subgroup is the one which preserves the scalar product

$$x^\dagger \sigma_3 y = x_1^* y_1 - x_2^* y_2; \tag{3.9}$$

this is the pseudounitary group $SU(1, 1)$. Preservation of (3.9) evidently requires that

$$a^\dagger \sigma_3 a = \sigma_3; \tag{3.10}$$

that is,

$$\sigma_3 = e^{-ib^\dagger} \sigma_3 e^{ib} = \sigma_3 e^{-i\hat{b}} e^{ib}, \tag{3.11}$$

where

$$\hat{b} \equiv \sigma_3 b^\dagger \sigma_3 = -\xi_1^* \sigma_1 - \xi_2^* \sigma_2 + \xi_3^* \sigma_3 . \tag{3.12}$$

Multiplication from the left by σ_3 shows that (3.11) can be satisfied in general only if $\hat{b} = b$; from (3.12), this means that ξ_1 and ξ_2 must be imaginary and ξ_3 must be real. From (3.6),

$$\xi_\pm^* = -\xi_\mp \quad \text{and} \quad \xi_3^* = \xi_3 \tag{3.13}$$

for this subgroup, $SU(1, 1)$.

Finally, there is the subgroup $SL(2, R)$ of real 2×2 matrices with determinant $+1$, which could act on a space of two real coordinates. In this case, the elements of the matrix b in (3.1) must be imaginary. If we make the permutation $\sigma_1 \to \sigma_2 \to \sigma_3 \to \sigma_1$ on the usual representation of the Pauli matrices, which preserves the commutation relations, then ξ_1 and ξ_2 must be imaginary and ξ_3 real, as for $SU(1, 1)$. Thus $SL(2, R)$ is isomorphic to $SU(1, 1)$.

The groups $SU(2)$ and $SU(1, 1) \cong SL(2, R)$ are called real forms of $SL(2)$, since in each case a choice of generators can be made such that the parameters are real. Conversely, $SL(2)$ is called the complexification (or complex extension) of any of the others since it can be obtained from any one of them by making the parameters complex. Of course, $SU(2)$ is compact, while the others are noncompact.

The set of all 2×2 nonsingular matrices and the subsets which are unitary or pseudounitary are called the fundamental representations of the abstract groups $SL(2)$, $SU(2)$, and $SU(1, 1)$, respectively. To find general representations of $SL(2)$, $SU(2)$, or $SU(1, 1)$, we must find (finite- or infinite-dimensional) matrices $E_\pm$ and H such that

$$[H, E_\pm] = \pm 2E_\pm \tag{3.14a}$$

and

$$[E_+, E_-] = H. \tag{3.14b}$$

The matrices representing group elements are then of the form

$$A = e^{iB}, \tag{3.15}$$

where

$$B = \xi_+ E_+ + \xi_- E_- + \xi_3 H, \tag{3.16}$$

with

$$\xi_3^* = \xi_3 \quad \text{and} \quad \begin{cases} \xi_\pm^* = \xi_\mp & \text{for } SU(2), \\ \xi_\pm^* = -\xi_\mp & \text{for } SU(1, 1). \end{cases} \tag{3.17}$$

There are no restrictions on the ξ's for $SL(2)$. For unitary representations we must also have

$$B^\dagger = B; \tag{3.18}$$

this means that[10]

$$H^\dagger = H \quad \text{and} \quad \begin{cases} E_\pm^\dagger = E_\mp & \text{for } SU(2), \\ E_\pm^\dagger = -E_\mp & \text{for } SU(1, 1). \end{cases} \tag{3.19}$$

It is evidently impossible to satisfy (3.18) for $SL(2)$ by any condition on the generators. However, this impasse can be broken by splitting the parameters into real and imaginary parts, writing

$$B = \sum_{i=1}^{3} (\xi_i \sigma_i + i\eta_i \sigma_i). \tag{3.20}$$

If we take the σ_i and $i\sigma_i$ as generators, then the condition on the parameters is that they must all be

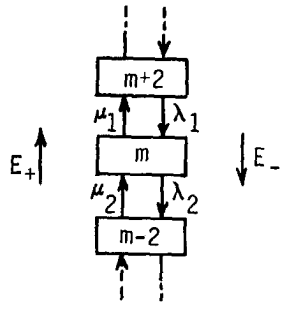


FIG. 1. Part of a weight diagram for A_1.

real; for unitary representations, the matrices corresponding to these generators must then be Hermitian. A direct check of the commutation relations shows that we can make the identifications

$$\begin{aligned} \tfrac{1}{2}\sigma_i &= L_{jk}, \quad i,j,k \text{ cyclic}, \\ \tfrac{1}{2}(i\sigma_i) &= iL_{4i}, \end{aligned} \tag{3.21}$$

where the L_{ij} satisfy the commutation relations for orthogonal transformations. From the reality of the parameters, it follows that this is just the algebra corresponding to the pseudo-orthogonal group $SO(3, 1)$.

Now we turn to explicit determination of the structure of the unitary representations of $SU(2)$ and $SU(1, 1)$. Equation (3.14a) states that $E_\pm$ raises/lowers the eigenvalues of H by two units:

$$H(E_\pm |m\rangle) = E_\pm H |m\rangle \pm 2E_\pm |m\rangle = (m \pm 2)(E_\pm |m\rangle), \tag{3.22}$$

where $H |m\rangle = m |m\rangle$. But normalization is not necessarily preserved; in fact, $E_\pm |m\rangle$ could be zero. Therefore, we set

$$\begin{aligned} E_+ |m\rangle &= \mu_1 |m+2\rangle, \quad E_- |m+2\rangle = \lambda_1 |m\rangle, \\ E_- |m\rangle &= \lambda_2 |m-2\rangle, \quad E_+ |m-2\rangle = \mu_2 |m\rangle, \end{aligned} \tag{3.23}$$

with the conditions $\langle m \mid m\rangle = \langle m+2 \mid m+2\rangle = \langle m-2 \mid m-2\rangle = 1$. This can be diagrammed as in Fig. 1. Equation (3.14b) yields the relationship

$$\lambda_2\mu_2 = \lambda_1\mu_1 + m. \tag{3.24}$$

Next we consider the implications of the unitary conditions (3.19). First we see that the eigenvalues of H must be real:

$$m^* = m. \tag{3.25}$$

For the stepping operators, we find

$$\langle m| E_+ = (\pm)(E_- |m\rangle)^\dagger = (\pm)\lambda_2^* \langle m-2|, \tag{3.26}$$

where the $(\pm)$ depends on whether the group is $SU(2)$ or $SU(1, 1)$. Thus

$$\langle m| E_+E_- |m\rangle = (\pm)\lambda_2^*\lambda_2\langle m-2 \mid m-2\rangle = (\pm)\lambda_2^*\lambda_2, \tag{3.27}$$

assuming the states to be normalized. But also,

$$\begin{aligned} \langle m| E_+E_- |m\rangle &= \lambda_2 \langle m| E_+ |m-2\rangle \\ &= \lambda_2\mu_2\langle m \mid m\rangle = \lambda_2\mu_2. \end{aligned} \tag{3.28}$$

Thus

$$\mu_2 = (\pm)\lambda_2^*, \quad \text{if} \quad \lambda_2 \neq 0. \tag{3.29}$$

Similarly,

$$\lambda_1 = (\pm)\mu_1^*, \quad \text{if} \quad \mu_1 \neq 0. \tag{3.30}$$

Substitution of these equations into (3.24) yields

$$|\lambda_2|^2 = |\lambda_1|^2 + (\pm)m. \tag{3.31}$$

We shall choose the phases of the states to make the λ_i real and positive; Eqs. (3.29), (3.30), and (3.31) then serve to determine all the normalization factors when any one is known.

Taking first the case of $SU(2)$, for which

$$\mu_i = \lambda_i \quad \text{and} \quad |\lambda_{i+1}|^2 = |\lambda_i|^2 + m_i, \tag{3.32}$$

we see that a chain such as indicated in Fig. 1 cannot go down forever, since the m_i eventually become negative. Thus (3.32) must lead to a λ_j which is zero for some j. Neither can the chain go up forever, since the m_i eventually become positive; thus μ_k must vanish for some k. Consideration of (3.32) also shows that the normalization factors are symmetrical about the middle; in particular, if the top vector is $|m\rangle$, the bottom vector must be $|-m\rangle$.

For $SU(1, 1)$, for which

$$\mu_i = -\lambda_i \quad \text{and} \quad |\lambda_{i+1}|^2 = |\lambda_i|^2 - m_i, \tag{3.33}$$

the limitation is that the chain can have a lower end only at some positive m_j or an upper end only at some negative m_k. It cannot have two ends; it may have none.

The above results can also be obtained by consideration of the invariant operator

$$\Sigma^2 \equiv \Sigma_1^2 + \Sigma_2^2 + \Sigma_3^2 \begin{cases} = 4E_+E_- + H(H-2), \\ = 4E_-E_+ + H(H+2). \end{cases} \tag{3.34}$$

(The invariance of this operator follows directly from the commutation relations.) Equations (3.28) and (3.29) yield

$$\langle m| E_+E_- |m\rangle = (\pm)\,|\lambda_2|^2 \begin{cases} > 0 & \text{for } SU(2), \\ < 0 & \text{for } SU(1,1). \end{cases} \tag{3.35}$$

That is, the eigenvalues of E_+E_- are positive for $SU(2)$ and negative for $SU(1, 1)$. If the chain extends arbitrarily far up or down, the eigenvalues of $H(H-2)$ become arbitrarily large and positive, while Σ^2 remains unchanged. This leads to a contradiction of (3.34) for $SU(2)$, but not for $SU(1, 1)$. However, for the latter there can be a contradiction

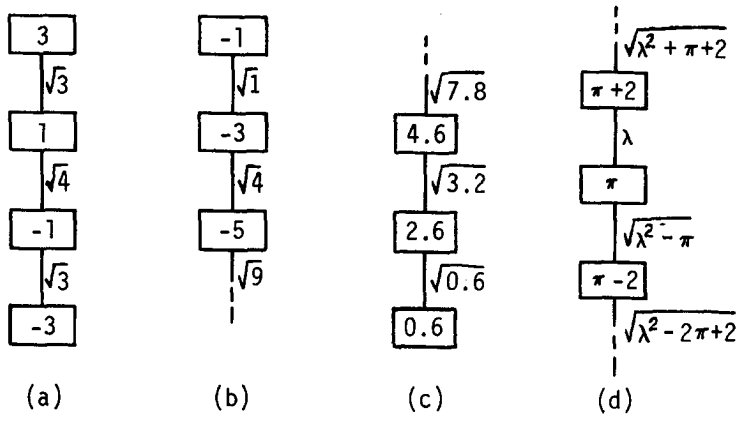


FIG. 2. Some representations of A_1: (a) is unitary for SU (2); (b) is unitary for $SU(1, 1)$; (c) and (d) are unitary for the covering group of $SU(1, 1)$. The square roots are values of $\lambda_i = (\pm)\mu_i$.

for small eigenvalues of $H(H-2)$, if Σ^2 is small enough. That is, unitary representations of $SU(2)$ must be bounded at top and bottom, while those of $SU(1, 1)$ may be bounded at top or bottom, but not both.

Some typical representations of A_1 are shown in Fig. 2. These representations are, of course, well known; that in Fig. 2(a) is just one of the usual spin representations (with states labeled by $2j_3$), while the other kinds have been discussed, for example, by Barut.[11] Incidentally, they illustrate the general property that unitary representations are finite-dimensional for compact groups such as $SU(2)$ and infinite-dimensional for noncompact groups such as $SU(1, 1)$. The evaluation of group elements in such representations is discussed in the Appendices.

4. CONSTRUCTION OF BOUNDED REPRESENTATIONS

We now proceed to the construction of bounded representations of larger semisimple Lie algebras, that is, representations having either a highest or a lowest weight. This construction will be further restricted to representations for which $E_\alpha^\dagger = \pm E_{-\alpha}$ when the matrices of the fundamental representation satisfy the corresponding conditions $e_\beta^\dagger = \pm e_{-\beta}$ (the choice between plus and minus for a given operator may be different in the two representations).[12] As before, Greek subscripts will be used for simple positive roots, x and y for roots which are positive but not necessarily simple, and u, v, and w for general positive or negative roots.

The construction of such representations will be based on the following relations, which refer only to simple operators, i.e., those labeled by simple roots:

$$[E_\alpha, E_{-\beta}] = \delta_{\alpha\beta} H_\alpha, \tag{4.1a}$$

$$[H_\alpha, E_{\pm\beta}] = \pm \frac{2(\alpha, \beta)}{(\alpha, \alpha)} E_{\pm\beta}, \tag{4.1b}$$

$$[H_\alpha, H_\beta] = 0. \tag{4.1c}$$

It will first be shown that these relations imply the complete set (2.13) for the matrices E_u and H_α in representations satisfying the above conditions.

Equations (4.1) do not specify the commutators $[E_\alpha, E_\beta]$, $[[E_\alpha, E_\beta], E_\gamma]$, $\cdots$. In part, such commutators serve to define new matrix operators $E_{\alpha+\beta}$, $E_{\alpha+\beta+\gamma}$, $\cdots$, in those cases where $\alpha + \beta$, $\alpha + \beta + \gamma$, $\cdots$ are positive roots. So we define the operators $E_{\pm x}$ recursively by[12]

$$E_{x+\alpha} \equiv \frac{1}{N_{x\alpha}} [E_x, E_\alpha], \tag{4.2a}$$

$$E_{-(x+\alpha)} \equiv \frac{1}{N_{-x,-\alpha}} [E_{-x}, E_{-\alpha}] \tag{4.2b}$$

for all positive roots $x + \alpha$, where the $N_{\pm x,\pm\alpha}$ are the coefficients which occur in (2.13a). (Note that $N_{\pm x,\pm\alpha} \neq 0$ if $x + \alpha$ is a positive root.) In case $x + \alpha = y + \beta$, with y also a positive root, we choose just one of the possible commutators to define $E_{x+\alpha} = E_{y+\beta}$, or $E_{-x-\alpha} = E_{-y-\beta}$. *A priori*, there is the possibility that some of the operators (4.2) vanish; however, we do not have to concern ourselves directly with this problem, since it will be shown below that these operators have all the correct commutation relations.

Use of the Jacobi identity

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \tag{4.3}$$

yields

$$\begin{aligned}
[H_\alpha, E_{u+\beta}] &= \frac{1}{N_{u\beta}} [H_\alpha, [E_u, E_\beta]] \\
&= -\frac{1}{N_{u\beta}} ([E_u, [E_\beta, H_\alpha]] + [E_\beta, [H_\alpha, E_u]]) \\
&= \frac{1}{N_{u\beta}} \left(\frac{2(\alpha, \beta)}{(\alpha, \alpha)} [E_u, E_\beta] + \frac{2(\alpha, u)}{(\alpha, \alpha)} [E_u, E_\beta] \right) \\
&= \frac{2(\alpha, u + \beta)}{(\alpha, \alpha)} E_{u+\beta},
\end{aligned} \tag{4.4}$$

provided H_α and E_u satisfy

$$[H_\alpha, E_u] = \frac{2(\alpha, u)}{(\alpha, \alpha)} E_u. \tag{4.5}$$

Thus an induction on the order of E_u [the total number of commutants in the definition (4.2a) or (4.2b)] proves (4.5) for operators E_u of any order. Equation (4.5) is the analog of (2.13c) for matrix operators.

Since (4.1a) includes all relations of the form (2.13b), it remains only to show that the matrix operators satisfy

$$[E_u, E_v] = N_{uv} E_{u+v} \tag{4.6}$$

with the same coefficients N_{uv} that occur in (2.13a). We start by showing that

$$[E_u, E_{\pm\alpha}] = N_{u,\pm\alpha} E_{u\pm\alpha} \tag{4.7}$$

(where u is any positive or negative root) by an induction on the order n of E_u.

Equation (4.7) is satisfied for $n = 1$; this follows from (4.1a), the definition (4.2), the fact that $[A, B] = -[B, A]$ for both matrices and abstract operators, and the equation

$$[E_{\pm\alpha}, E_{\pm\beta}] = 0 \quad \text{if} \quad \alpha + \beta \text{ is not a root}, \tag{4.8}$$

which will now be proven. Use of (4.3), (4.1a), and (4.1b) yields

$$\begin{aligned}[[E_\alpha, E_\beta], E_{-\zeta}] &= -[[E_\beta, E_{-\zeta}], E_\alpha] - [[E_{-\zeta}, E_\alpha], E_\beta] \\ &= -\delta_{\beta\zeta}[H_\beta, E_\alpha] + \delta_{\alpha\zeta}[H_\alpha, E_\beta] \\ &= 2(\alpha, \beta)\left(-\frac{E_\alpha}{(\beta, \beta)} + \frac{E_\beta}{(\alpha, \alpha)}\right). \end{aligned} \tag{4.9}$$

This evidently vanishes for $\alpha = \beta$. For $\alpha \neq \beta$, $(\alpha, \beta) \leq 0$; hence (2.9) requires that it vanish, since $\alpha + \beta$ is not a root. Thus the commutator (4.9) vanishes for all ζ; it will be shown below that this implies

$$[E_\alpha, E_\beta] = 0 \tag{4.10}$$

in representations which are bounded below. Since $E_{-\alpha} = \pm E_\alpha^\dagger$ for the representations in which we are interested, the other half of (4.8) follows immediately. In the case of representations bounded above, the roles of raising and lowering operators are just interchanged. Thus (4.7) is valid for $n = 1$ for bounded unitary representations.

Assume now that (4.7) is valid when E_u is of order $\leq n$. Consider first commutators of the form

$$[E_x, E_{-\alpha}] = \frac{1}{N_{y\beta}} [[E_y, E_\beta], E_{-\alpha}] \tag{4.11}$$

with E_x of order $n + 1$. Use of (4.3), (4.1a), (4.5), and (4.7) for order $\leq n$ yields

$$\begin{aligned}&[E_x, E_{-\alpha}] \\ &\quad = -\frac{1}{N_{y\beta}} ([[E_\beta, E_{-\alpha}], E_y] + [[E_{-\alpha}, E_y], E_\beta]) \\ &\quad = -\frac{1}{N_{y\beta}} (\delta_{\alpha\beta}[H_\beta, E_y] + N_{-\alpha,y}[E_{y-\alpha}, E_\beta]) \\ &\quad = -\frac{1}{N_{y\beta}} \left(\frac{2\delta_{\alpha\beta}(\beta, y)}{(\beta, \beta)} E_y + N_{-\alpha,y} N_{y-\alpha,\beta} E_{y-\alpha+\beta}\right) \\ &\quad = -\frac{1}{N_{y\beta}} \left(\frac{2\delta_{\alpha\beta}(\beta, y)}{(\beta, \beta)} + N_{-\alpha,y} N_{y-\alpha,\beta}\right) E_{x-\alpha}. \end{aligned} \tag{4.12}$$

Note that $E_{y-\alpha}$ is of order $n - 1$, so that a separate evaluation must be carried out for the case $n = 1$; but it is straightforward, and is left to the reader. Should it happen that $y - \alpha$ or $y - \alpha + \beta$ is not a root, we define $N_{-\alpha,y} \equiv 0$ or $N_{y-\alpha,\beta} \equiv 0$, respectively; the definition of $E_{y-\alpha}$ or $E_{y-\alpha+\beta}$ is then immaterial. An evaluation analogous to (4.12) can also be carried out in the abstract algebra, with the result that

$$[e_x, e_{-\alpha}] = -\frac{1}{N_{y\beta}} \left(\frac{2\delta_{\alpha\beta}(\beta, y)}{(\beta, \beta)} + N_{-\alpha,y} N_{y-\alpha,\beta}\right) e_{x-\alpha}. \tag{4.13}$$

But we also know that

$$[e_x, e_{-\alpha}] = N_{x,-\alpha} e_{x-\alpha}. \tag{4.14}$$

Now either $x - \alpha$ is a positive root—in which case $e_{x-\alpha}$ is well defined and the coefficients in (4.13) and (4.14) are equal—or it is not, in which case both coefficients vanish. Thus (4.12) can be rewritten

$$[E_x, E_{-\alpha}] = N_{x,-\alpha} E_{x-\alpha} \quad (= 0 \text{ if } x - \alpha \text{ is not a root}). \tag{4.15}$$

A similar proof holds for

$$[E_{-x}, E_\alpha] = N_{-x,\alpha} E_{-x+\alpha}. \tag{4.16}$$

This derivation of (4.15) and (4.16) could be summarized as follows: The commutator in (4.15) or (4.16) can be evaluated in terms of commutation relations already known to have the correct form; since this evaluation is the same in the matrix algebra as in the abstract algebra, we must obtain the correct answer, namely, (4.15) or (4.16).

Next consider commutators of the form

$$[E_x, E_\alpha], \tag{4.17}$$

where E_x is of order $n + 1$. Use of (4.3), (4.1a), (4.15) for order $n + 1$, (4.5), and (4.7) for order n yields

$$[[E_x, E_\alpha], E_{-\zeta}] = -\left(\frac{2\delta_{\alpha\zeta}(\alpha, x)}{(\alpha, \alpha)} + N_{-\zeta,x} N_{x-\zeta,\alpha}\right) E_{x+\alpha-\zeta}. \tag{4.18}$$

The same operations can be carried out in the abstract algebra, with the analogous result; also,

$$[[e_x, e_\alpha], e_{-\zeta}] = N_{x\alpha}[e_{x+\alpha}, e_{-\zeta}] = N_{x\alpha}N_{x+\alpha,-\zeta}e_{x+\alpha-\zeta}. \tag{4.19}$$

Thus the coefficients in (4.18) and (4.19) must be equal, so that

$$[[E_x, E_\alpha], E_{-\zeta}] = N_{x\alpha}N_{x+\alpha,-\zeta}E_{x+\alpha-\zeta}. \tag{4.20}$$

On the other hand, if the definitions (4.2a) prescribe

$$E_{x+\alpha} = \frac{1}{N_{y+\alpha,\eta}}[E_{y+\alpha}, E_\eta] \tag{4.21}$$

(note that y need not be a root, though $y + \alpha$ necessarily is), then

$$\begin{aligned}
&[E_{x+\alpha}, E_{-\zeta}] \\
&= -\frac{1}{N_{y+\alpha,\eta}}([[E_\eta, E_{-\zeta}], E_{y+\alpha}] + [[E_{-\zeta}, E_{y+\alpha}], E_\eta]) \\
&= -\frac{1}{N_{y+\alpha,\eta}}\left(2\delta_{\eta\zeta}\frac{(\eta, y+\alpha)}{(\eta,\eta)} + N_{-\zeta,y+\alpha}N_{y+\alpha-\zeta,\eta}\right)E_{x+\alpha-\zeta} \\
&= N_{x+\alpha,-\zeta}E_{x+\alpha-\zeta},
\end{aligned} \tag{4.22}$$

where the last expression follows, as before, from comparison with the result of carrying out similar operations in the abstract algebra. From (4.20) and (4.22) we have

$$[[E_x, E_\alpha] - N_{x\alpha}E_{x+\alpha}, E_{-\zeta}] = 0 \tag{4.23}$$

for all ζ. As will be shown below, this is sufficient to guarantee the vanishing of the first commutant for representations bounded below; i.e.,

$$[E_x, E_\alpha] = N_{x\alpha}E_{x+\alpha}. \tag{4.24}$$

Since $E_{-\beta} = \pm E_\beta^\dagger$ and $e_{-\beta} = \pm e_\beta^\dagger$ for the unitary and fundamental representations in which we shall be interested (though the choice between plus or minus for a given operator may be different in the two representations), it also follows that

$$[E_{-x}, E_{-\alpha}] = N_{-x,-\alpha}E_{-x-\alpha}. \tag{4.25}$$

To show this, one need only express (4.24) and its analog in the fundamental representation in terms of commutators of simple operators; the Hermitian conjugates of these two equations can be rearranged into the form (4.25), with the same coefficients. In the case of representations bounded above, the same arguments evidently go through with positive and negative roots interchanged. Thus (4.7) has been proven for order $n + 1$ and, hence, by induction for all orders of E_u.

Finally, we can prove (4.6) by an induction on the order m of E_v. For $m = 1$, it reduces to (4.7). If

$$E_v \equiv \frac{1}{N_{w\zeta}}[E_w, E_{\pm\zeta}] \tag{4.26}$$

is of order $m + 1$, then use of (4.3) and (4.6) for order of the second commutant less than or equal to m yields

$$\begin{aligned}
[E_u, E_v] &= \frac{1}{N_{w\zeta}}[E_u, [E_w, E_{\pm\zeta}]] \\
&= \frac{1}{N_{w\zeta}}(-[[E_u, E_{\pm\zeta}], E_w] + [[E_u, E_w], E_{\pm\zeta}]) \\
&= \frac{1}{N_{w\zeta}}(-N_{u\pm\zeta,w}N_{u,\pm\zeta} + N_{u+w,\pm\zeta}N_{u,w})E_{u+v}.
\end{aligned} \tag{4.27}$$

Consideration of the corresponding expression in the abstract algebra shows the coefficient of E_{u+v} to be just N_{uv}.

To complete these proofs, it must be shown that in a representation bounded below

$$[Z_+, E_{-\zeta}] = 0 \tag{4.28}$$

for all ζ implies

$$Z_+ = 0, \tag{4.29}$$

where Z_+ stands for $E_{+\alpha}$ or a commutator of such operators. First, note that in an irreducible representation every basis vector except the lowest can be lowered by some $E_{-\alpha}$. For example, suppose that some vector $|a\rangle$ cannot be lowered. Then, since this is an irreducible representation, there must be an operator $E_{+\alpha}$ which raises it, say to $|b\rangle$ (see Fig. 3). If it were possible to lower $|b\rangle$ by some operator $E_{-\beta}$ with $\beta \neq \alpha$, then (4.1a) would be violated; and lowering $|b\rangle$ by $E_{-\alpha}$ just leads back to $|a\rangle$, by (4.1a). Thus all of the states connected to $|a\rangle$ must lie higher than $|a\rangle$; that is, $|a\rangle$ is the lowest vector of the representation. Now suppose that there exists a raising operator Z_+ which commutes with all the $E_{-\zeta}$, but yet yields a nonzero result when applied to some vector $|a\rangle$. Since $Z_+|a\rangle \equiv |b\rangle$ is not the lowest vector, there is some operator $E_{-\alpha}$ which lowers it. Thus we must have the condition depicted in Fig. 4. That is, there must exist a vector $|d\rangle$ lower than $|a\rangle$, such that Z_+ applied to it yields a nonzero result. But this will ultimately lead to a contradiction for a representation bounded below. Thus $Z_+ = 0$.

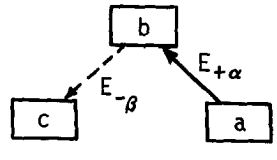


FIG. 3. Demonstration that $E_{-\alpha}|a\rangle = 0$ for all α implies that $|a\rangle$ is the lowest vector.

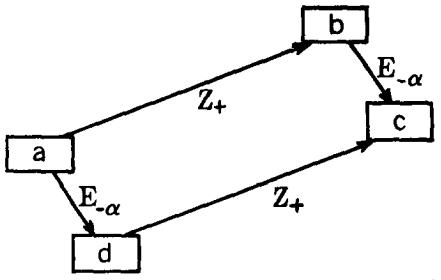


FIG. 4. Demonstration that Z_+ must vanish identically.

A similar argument shows that in a representation bounded above

$$[Z_-, E_\zeta] = 0 \tag{4.30}$$

for all ζ implies

$$Z_- = 0, \tag{4.31}$$

where Z_- stands for $E_{-\alpha}$ or a commutator of such operators.

Thus it has been shown that Eqs. (4.1) imply the rest of the required commutation relations for representations which are bounded either above or below and satisfy the conditions $E_\alpha^\dagger = \pm E_{-\alpha}$ when the matrices of the fundamental representation are chosen to satisfy $e_\beta^\dagger = \pm e_{-\beta}$. The general method of using (4.1) will now be indicated; specific examples are given in the following sections.

Equations (4.1c) will be satisfied by choosing basis states for which the operators H_α are diagonal. Relations (4.1b) state that the $E_{\pm\beta}$'s are raising and lowering operators for the eigenvalues of the H_α's. Actually, only two kinds of connection occur for the algebras considered here, leading to the following specializations of (4.1b):

$$\begin{matrix} \circ\!\!-\!\!-\!\!\circ \\ \alpha \quad \beta \end{matrix} \quad \begin{aligned} &[H_\alpha, E_{\pm\alpha}] = \pm 2E_{\pm\alpha}, \quad [H_\alpha, E_{\pm\beta}] = \mp E_{\pm\beta}, \\ &[H_\beta, E_{\pm\beta}] = \pm 2E_{\pm\beta}, \quad [H_\beta, E_{\pm\alpha}] = \mp E_{\pm\alpha}, \end{aligned} \tag{4.32}$$

$$\begin{matrix} \bullet\!\!-\!\!\circ \\ \alpha \quad \beta \end{matrix} \quad \begin{aligned} &[H_\alpha, E_{\pm\alpha}] = \pm 2E_{\pm\alpha}, \quad [H_\alpha, E_{\pm\beta}] = \mp 2E_{\pm\beta}, \\ &[H_\beta, E_{\pm\beta}] = \pm 2E_{\pm\beta}, \quad [H_\beta, E_{\pm\alpha}] = \mp E_{\pm\alpha}. \end{aligned} \tag{4.33}$$

That is, $E_{\pm\xi}$ always raises/lowers H_ξ by two units; it lowers/raises H_ζ by one unit if ζ is connected to ξ, except when $\xi = \beta$ and $\zeta = \alpha$ for the case of the double connection (4.33). The actions of the lowering operators, for example, can be diagrammed as in Fig. 5. A set of eigenvalues $[\alpha\ \beta\ \gamma \cdots]$ labeling states in such a diagram is called a weight. The normalization factors to be associated with the raising and lowering operators in order to produce normalized states follow from (4.1a) for $\alpha = \beta$, just as in Sec. 3. Also, these equations can often be used to show that $E_{-\alpha}$ and $E_{+\alpha}$ applied in succession to a state yield that state back again, as indicated in Fig. 5 by the dashed arrows. Note that two boxes have been drawn around the weight $[\alpha\ \ \beta-1\ \ \gamma+1 \cdots]$ to indicate that more than one vector may correspond to this weight; whether or not this happens is determined by (4.1a) with $\alpha \neq \beta$. If there is only one state with weight $[\alpha\ \ \beta-1\ \ \gamma+1 \cdots]$, then the product of the normalization factors corresponding to the lower path $E_{+\alpha}E_{-\beta}$ must equal the product of the two normalization factors corresponding to the upper path $E_{-\beta}E_{+\alpha}$. If the two products are unequal, this indicates either that there are two (nonorthogonal) states with the weight $[\alpha\ \ \beta-1\ \ \gamma+1 \cdots]$ or that the representation is impossible, depending on the direction of the inequality.

5. THE ALGEBRAS A_n

The group $SL(p)$ has as fundamental representation the group of all unimodular $p \times p$ matrices. We take the exponential form (3.1) for each group element a and expand b as follows:

$$b = \sum_{i=1}^{p-1} \xi_i h_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{p} \xi_{ij} e_{ij}, \tag{5.1}$$

where h_i is defined to have $+1$ in the iith position, -1 in the $(i+1)(i+1)$th position, and zeros elsewhere, while e_{ij} $(i \neq j)$ has $+1$ in the ijth position and zeros elsewhere. This form automatically satisfies the condition $\mathrm{Tr}\, b = 0$, or $\det a = 1$. We note that

$$[e_{ij}, e_{kl}] = \begin{cases} \delta_{jk} e_{il} & \text{if } i \neq l, \\ -\delta_{il} e_{kj} & \text{if } j \neq k. \end{cases} \tag{5.2}$$

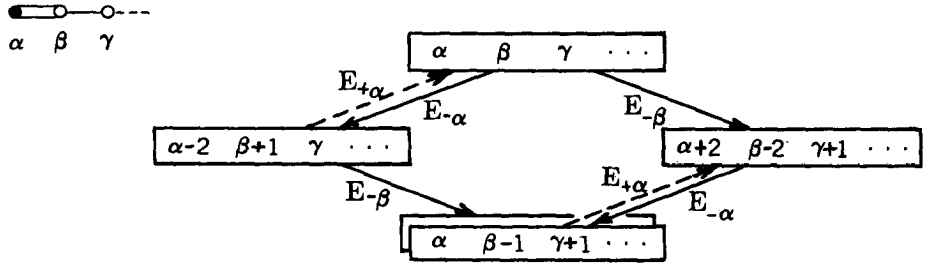


FIG. 5. Part of a weight diagram corresponding to B_n; $\alpha, \beta, \gamma, \cdots$ denote eigenvalues of the operators $H_\alpha, H_\beta, H_\gamma, \cdots$.

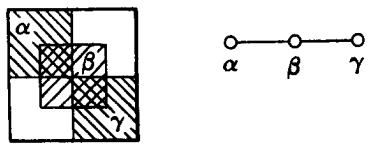


FIG. 6. Generators of $SL(4)$ in the fundamental representation. Shading indicates nonzero matrix elements.

This means that e's with indices differing by more than one unit can be constructed from those with indices differing by ± 1; e.g., $e_{13} = [e_{12}, e_{23}]$. So the set

$$\begin{aligned} &e_{+\alpha} \equiv e_{12}, \quad e_{-\alpha} \equiv e_{21}, \quad h_\alpha \equiv h_1, \\ &e_{+\beta} \equiv e_{23}, \quad e_{-\beta} \equiv e_{32}, \quad h_\beta \equiv h_2, \quad \text{etc.}, \end{aligned} \tag{5.3}$$

constitutes a subset of generators from which the rest can be obtained by commutation. Thus (5.1) can be rewritten

$$b = \sum_\rho \xi_\rho h_\rho + \sum_{\pm,\rho} \xi_{\pm\rho} e_{\pm\rho} + \sum_{\pm,\rho,\sigma} \xi_{\pm(\rho+\sigma)} [e_{\pm\rho}, e_{\pm\sigma}] + \cdots, \tag{5.4}$$

where the ξ's are complex numbers, the dots denote terms involving higher-order commutators, and $\rho, \sigma, \cdots$ range over $\alpha, \beta, \cdots$. The generators (5.3) satisfy commutation relations of the form (2.13). Explicitly,

$$\begin{aligned} &[e_{\pm\rho}, e_{\pm\sigma}] \equiv e_{\pm(\rho+\sigma)} \neq 0 \quad \text{if } \rho, \sigma \text{ are adjacent}, \\ &[e_{+\rho}, e_{-\sigma}] = \delta_{\rho\sigma} h_\rho, \\ &[h_\rho, e_{\pm\rho}] = \pm 2 e_{\pm\rho}, \\ &[h_\rho, e_{\pm\sigma}] = \mp e_{\pm\sigma} \qquad \text{if } \rho, \sigma \text{ are adjacent}, \\ &[h_\rho, h_\sigma] = 0. \end{aligned} \tag{5.5}$$

Furthermore, comparison with (2.13), (2.6), and (2.3) shows that these commutation relations are those of A_n with $n = p - 1$. For example, the generators (5.3) for $SL(4)$ take the form shown in Fig. 6, with the indicated correspondence to the Dynkin diagram for A_3. We note that the third sum in (5.4) can be restricted to values of ρ and σ such that each sum $\alpha + \beta, \beta + \gamma, \cdots$ occurs precisely once; combinations such as $\alpha + \gamma$ are not needed, since e's with nonadjacent indices commute, according to the definitions (5.3). That is, this sum is really over just the positive roots of the form $\rho + \sigma$ of A_n. Similar remarks apply to the subsequent sums in (5.4).

The pseudounitary group $SU(p-q, q)$ is that subgroup of $SL(p)$ which preserves the scalar product

$$\begin{aligned} x^\dagger g y \equiv x_1^* y_1 + \cdots + x_{p-q}^* y_{p-q} \\ - x_{p-q+1}^* y_{p-q+1} - \cdots - x_p^* y_p, \end{aligned} \tag{5.6}$$

where g is a diagonal matrix with $+1$ in the first $p - q$ positions and -1 thereafter. We note that $SU(p, 0)$ is just $SU(p)$. Preservation of (5.6) under the transformation

$$x \to ax, \quad y \to ay \tag{5.7}$$

requires

$$a^\dagger g a = g \tag{5.8}$$

or

$$g a^\dagger g = a^{-1}, \tag{5.9}$$

since $g^2 = I$. Referring to the exponential expression (3.1), we see that this requires

$$g b^\dagger g = b \tag{5.10}$$

for the matrix b of (5.4), in the fundamental representation. Now we note the following relations, which follow from the definitions (5.3):

$$\begin{aligned} h_\rho^\dagger &= h_\rho, \\ e_{\pm\rho}^\dagger &= e_{\mp\rho}, \\ g h_\rho g &= h_\rho, \\ g e_{\pm\rho_0} g &= -e_{\pm\rho_0}, \\ g e_{\pm\rho} g &= +e_{\pm\rho} \quad \text{for} \quad \rho \neq \rho_0, \end{aligned} \tag{5.11}$$

where $e_{\pm\rho_0}$ are the operators which span the change in metric; for example, if g has the diagonal elements $(1, 1, 1, 1, -1, -1)$, then e_{ρ_0} and $e_{-\rho_0}$ are the matrices e_{45} and e_{54}, respectively. Substitution of (5.4) and (5.11) into (5.10) yields

$$\begin{aligned} &\sum_\rho \xi_\rho^* h_\rho + \sum_{\pm,\rho} \epsilon_\rho \xi_{\pm\rho}^* e_{\mp\rho} - \sum_{\pm,\rho,\sigma} \epsilon_\rho \epsilon_\sigma \xi_{\pm(\rho+\sigma)}^* [e_{\mp\rho}, e_{\mp\sigma}] + \cdots \\ &= \sum_\rho \xi_\rho h_\rho + \sum_{\pm,\rho} \xi_{\pm\rho} e_{\pm\rho} + \sum_{\pm,\rho,\sigma} \xi_{\pm(\rho+\sigma)} [e_{\pm\rho}, e_{\pm\sigma}] + \cdots, \end{aligned} \tag{5.12}$$

where

$$\begin{aligned} \epsilon_{\rho_0} &\equiv -1, \\ \epsilon_\rho &\equiv +1 \quad \text{for} \quad \rho \neq \rho_0. \end{aligned} \tag{5.13}$$

Thus

$$\begin{aligned} \xi_\rho^* &= \xi_\rho, \\ \xi_{\pm\rho}^* &= \epsilon_\rho \xi_{\mp\rho}, \\ \xi_{\pm(\rho+\sigma)}^* &= -\epsilon_\rho \epsilon_\sigma \xi_{\mp(\rho+\sigma)}. \end{aligned} \tag{5.14}$$

In a unitary representation, on the other hand, we must have

$$B^\dagger = B. \tag{5.15}$$

That is,

$$\begin{aligned} &\sum_\rho \xi_\rho^* H_\rho^\dagger + \sum_{\pm,\rho} \xi_{\pm\rho}^* E_{\pm\rho}^\dagger - \sum_{\pm,\rho,\sigma} \xi_{\pm(\rho+\sigma)}^* [E_{\pm\rho}^\dagger, E_{\pm\sigma}^\dagger] + \cdots \\ &= \sum_\rho \xi_\rho H_\rho + \sum_{\pm,\rho} \xi_{\pm\rho} E_{\pm\rho} + \sum_{\pm,\rho,\sigma} \xi_{\pm(\rho+\sigma)} [E_{\pm\rho}, E_{\pm\sigma}] + \cdots, \end{aligned} \tag{5.16}$$

where the parameters are the same as in the fundamental representation; in particular, they satisfy (5.14). Thus

$$\begin{aligned} H_\rho^\dagger &= H_\rho, \\ E_{\pm\rho_0}^\dagger &= -E_{\mp\rho_0}, \\ E_{\pm\rho}^\dagger &= +E_{\mp\rho} \quad \text{for} \quad \rho \neq \rho_0. \end{aligned} \tag{5.17}$$

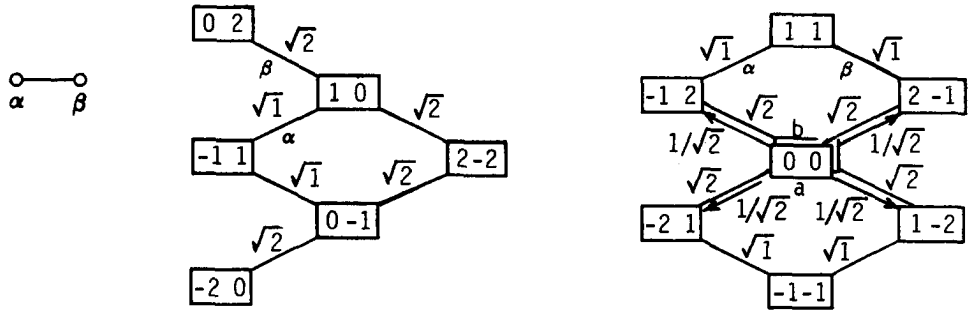


FIG. 7. Two unitary representations of $SU(3)$. The actions of $E_{\pm\alpha}$ are indicated by lines sloping downward toward the left, and the actions of $E_{\pm\beta}$, by lines sloping downward toward the right. Arrowheads indicate paths to be traversed in one direction only; note that $|0\,0, b\rangle$ must be replaced by a linear combination of $|0\,0, a\rangle$ and $|0\,0, b\rangle$ orthogonal to $|0\,0, a\rangle$ in order that $E_{\pm\alpha}$ and $E_{\pm\beta}$ satisfy the unitarity relations (5.17).

These are the unitarity conditions for representations of $SU(p-q, q)$. Together with the commutation relations among simple triplets, which are the same as for $SU(2)$ and $SU(1, 1)$, relations (5.17) yield

$$|\lambda_{\rho,i+1}|^2 = |\lambda_{\rho,i}|^2 + \epsilon_\rho m_{\rho,i} \quad \text{and} \quad \mu_{\rho,i} = \epsilon_\rho \lambda_{\rho,i}, \tag{5.18}$$

in complete analogy to (3.32) and (3.33).

The commutation relations in all representations are the same as in the fundamental representation. Namely, within each triplet,

$$\begin{aligned} [H_\alpha, E_{\pm\alpha}] &= \pm 2E_{\pm\alpha}, \\ [E_{+\alpha}, E_{-\alpha}] &= H_\alpha; \end{aligned} \tag{5.19}$$

and, between neighboring triplets,

$$\begin{aligned} [H_\alpha, E_{\pm\beta}] &= \mp E_{\pm\beta}, \\ [E_{\pm\alpha}, E_{\pm\beta}] &\equiv E_{\pm(\alpha+\beta)} \neq 0, \\ [E_{+\alpha}, E_{-\beta}] &= 0, \\ [H_\alpha, H_\beta] &= 0. \end{aligned} \tag{5.20}$$

The first equations of (5.19) and (5.20) show that while $E_{\pm\alpha}$ still raises/lowers the eigenvalues of H_α by two units, it also lowers/raises the eigenvalues of any neighboring H by one unit.

As an example, two possible representations of $SU(3)$ are shown in Fig. 7. In this case, where there is no ρ_0, the unitarity relations (5.18) tell us that both α and β chains must be finite and symmetrical with respect to the eigenvalues of H_α and H_β, respectively, as for $SU(2)$. Note that the new commutation relation $[E_{+\alpha}, E_{-\beta}] = 0$ precludes situations such as those depicted in Fig. 8; it also serves to determine the multiplicities and some of the normalization factors for weights such as [0 0] in the second representation of Fig. 7. This latter representation also illustrates a complication which arises with degenerate weights. The state $|0\,0, a\rangle = E_{-\alpha}|2\,-1\rangle$ is not orthogonal to $|0\,0, b\rangle = E_{-\beta}|-1\,2\rangle$, and the portions of $E_{\pm\alpha}$ and $E_{\pm\beta}$ (as defined by Fig. 7) referring to these states do not satisfy the unitarity relations (5.17). This is easily remedied by replacing $|0\,0, b\rangle$ by the "isospin zero" state

$$|0\,0, \lambda\rangle \equiv -\sqrt{\tfrac{1}{3}}\,|0\,0, a\rangle + \sqrt{\tfrac{4}{3}}\,|0\,0, b\rangle. \tag{5.21}$$

The new normalization coefficients, easily found from Fig. 7, satisfy (5.17).

As a second example we take $SU(2, 1)$, with the second triplet chosen to be the noncompact one so that $\rho_0 = \beta$. Relations (5.18) tell us that α chains are finite and symmetrical, while each β chain is infinite and can have only a "positive" lower end, a "negative" upper end, or no ends at all. Figure 9 shows the general form of unitary representations having an α singlet at the upper end of the leading β chain. The convention has been adopted that noncompact generators act horizontally. Representations with α doublets, α triplets, etc., at the upper end can be obtained from Fig. 9 by setting $\omega = 0, -1$, etc., and taking the portion to the right of the vanishing normalization factors; there are also other representations starting with α doublets, α triplets, etc., which have degenerate weights. We note that Fig. 9 is a representation of the group $SU(2, 1)$ only for integral ω; since $\exp(2\pi i h_\beta)$ is the identity in the fundamental representation, $\exp(2\pi i H_\beta)$ must be the identity in any other true representation of the group.

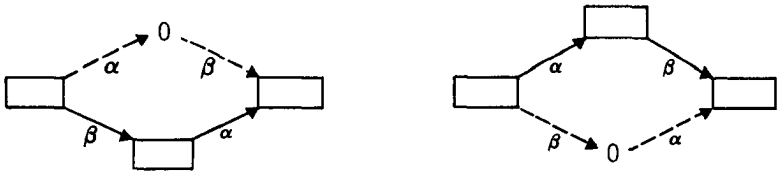


FIG. 8. Structures prohibited by $[E_{+\alpha}, E_{-\beta}] = 0$.

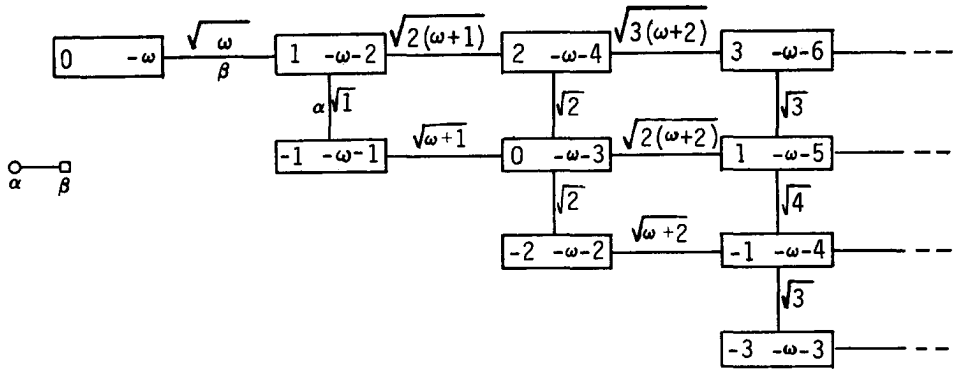


FIG. 9. A class of unitary representations of the algebra of $SU(2, 1)$, $\omega > 0$. (Normalization factors for $E_{+\beta}$ are the negatives of those shown for $E_{-\beta}$.)

Finally, Fig. 10 shows the classic example of a representation of $SU(3, 1)$ which yields the spectrum of the simple harmonic oscillator in three dimensions; this can be generalized in an obvious way to the representation of $SU(N, 1)$ which yields the spectrum of the simple harmonic oscillator in N dimensions.

6. THE ALGEBRAS B_n AND D_n

The proper complex orthogonal group in m dimensions, $SO(m, C)$, has as fundamental representation the group of all unimodular $m \times m$ matrices which preserve the bilinear form

$$\tilde{x}y \equiv x_1y_1 + x_2y_2 + \cdots + x_my_m. \tag{6.1}$$

We again take the exponential form

$$a = e^{ib}, \tag{6.2}$$

so that preservation of (6.1) is equivalent to

$$\tilde{b} = -b. \tag{6.3}$$

This condition is satisfied by setting

$$b = \sum_{\substack{i,j \\ i<j}} \xi_{ij} l_{ij}, \tag{6.4}$$

where each l_{ij} is defined to have $+1$ in the ijth position, -1 in the jith position, and zeros elsewhere. Tracelessness in this case is an automatic consequence of the antisymmetry of b. We note that the parameters need not be real, since we are defining the complex orthogonal groups. The commutation relations among the above generators are

$$[l_{ij}, l_{kl}] = \delta_{jk}l_{il} - \delta_{jl}l_{ik} + \delta_{il}l_{jk} - \delta_{ik}l_{jl}. \tag{6.5}$$

For the case that m is odd we make the definitions

$$\begin{aligned} l_{\pm\alpha} &\equiv l_{12} \mp il_{13}, \\ l_{\pm\beta} &\equiv l_{14} \mp il_{15}, \\ l_{\pm\gamma} &\equiv l_{16} \mp il_{17}, \quad \text{etc.} \end{aligned} \tag{6.6}$$

We note that any l_{ij}, with $i \neq 1$ and $j \neq 1$, can be constructed as the commutator

$$l_{ij} = -[l_{1i}, l_{1j}], \tag{6.7}$$

and thus as a linear combination of commutators of

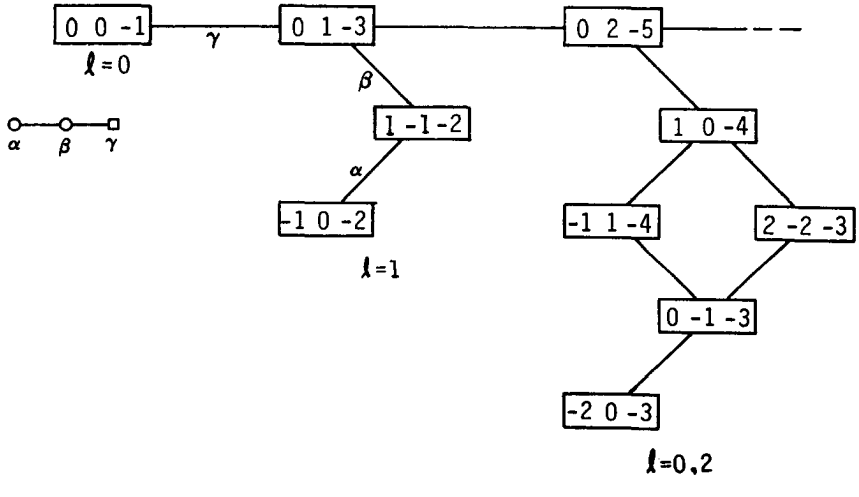


FIG. 10. A unitary representation of $SU(3, 1) \sim A_3$ corresponding to the energy levels of the simple harmonic oscillator; $l_\pm = e_{\pm\alpha} + e_{\pm\beta}$, $l_3 = h_\alpha + h_\beta$. Most of the γ paths have been omitted.

the $l_{\pm\alpha}$, $l_{\pm\beta}$, etc. Next we define, for $SO(2n+1, C)$,

$$\begin{aligned} e_{\pm\alpha} &\equiv \pm l_{\pm\alpha} = \pm l_{12} - i l_{13}, \\ h_\alpha &\equiv [e_{+\alpha}, e_{-\alpha}] = 2i l_{23}, \\ e_{\pm\beta} &\equiv \pm\tfrac{1}{2}[l_{\pm\beta}, l_{\mp\alpha}] = \tfrac{1}{2}\{\pm(l_{24} + l_{35}) - i(l_{25} - l_{34})\}, \\ h_\beta &\equiv [e_{+\beta}, e_{-\beta}] = i(l_{45} - l_{23}), \\ e_{\pm\gamma} &\equiv \pm\tfrac{1}{2}[l_{\pm\gamma}, l_{\mp\beta}] = \tfrac{1}{2}\{\pm(l_{46} + l_{57}) - i(l_{47} - l_{56})\}, \\ h_\gamma &\equiv [e_{+\gamma}, e_{-\gamma}] = i(l_{67} - l_{45}), \quad \text{etc.}, \end{aligned} \tag{6.8}$$

where the final expressions follow from the commutation relations (6.5). The α operators are a special case, but thereafter it is clear that the scheme will continue in the fashion of the β and γ operators. The l's defined in (6.6) can be expressed in terms of commutators of the e's, since

$$\begin{aligned} l_{\pm\alpha} &= \pm e_{\pm\alpha}, \\ l_{\pm\beta} &= \pm[l_{\pm\alpha}, e_{\pm\beta}], \\ l_{\pm\gamma} &= \pm[l_{\pm\beta}, e_{\pm\gamma}], \quad \text{etc.} \end{aligned} \tag{6.9}$$

Thus (6.4) can be rewritten in terms of the $e_{\pm\rho}$ and their commutators:

$$b = \sum_\rho \xi_\rho h_\rho + \sum_{\pm,\rho} \xi_{\pm\rho} e_{\pm\rho} + \sum_{\pm,\rho,\sigma} \xi_{\pm(\rho+\sigma)}[e_{\pm\rho}, e_{\pm\sigma}] + \cdots. \tag{6.10}$$

The commutation relations below have been used to eliminate certain terms from this expansion.

The virtue of the operators defined in (6.8) is that they satisfy commutation relations of the general form (2.13). Explicitly,

$$[e_{\pm\rho}, e_{\pm\sigma}] \equiv e_{\pm(\rho+\sigma)} \neq 0 \quad \text{if } \rho, \sigma \text{ are adjacent,} \tag{6.11a}$$

$$[e_{+\rho}, e_{-\sigma}] = \delta_{\rho\sigma} h_\rho, \tag{6.11b}$$

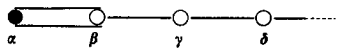
$$[h_\rho, e_{\pm\rho}] = \pm 2 e_{\pm\rho}, \tag{6.11c}$$

$$[h_\rho, e_{\pm\sigma}] = \mp e_{\pm\sigma} \quad \text{if } \rho, \sigma \text{ are adjacent,} \quad \rho \neq \alpha, \tag{6.11d}$$

$$[h_\alpha, e_{\pm\beta}] = \mp 2 e_{\pm\beta}, \tag{6.11e}$$

$$[h_\rho, h_\sigma] = 0. \tag{6.11f}$$

The only relation above which differs from the $SL(p)$ case is (6.11e); comparison with (2.13) shows that it corresponds to the double connection of (2.6). There are, of course, other commutators involving $e_{\pm(\rho+\sigma)}$, etc., and operators with different nonadjacent indices commute. The Dynkin diagram and a summary of the actions of the raising operators are given below (we note that the commutation relations are those of B_n, where $m = 2n+1$), for $SO(2n+1, C) \sim B_n$:



$$\begin{aligned} e_{+\alpha}&: [\ 2 \quad -1 \quad 0 \quad 0 \quad \cdots] \\ e_{+\beta}&: [-2 \quad 2 \quad -1 \quad 0 \quad \cdots] \\ e_{+\gamma}&: [\ 0 \quad -1 \quad 2 \quad -1 \quad \cdots] \\ &\text{etc.} \end{aligned} \tag{6.12}$$

For the case that m is even, we make the definitions

$$\begin{aligned} l_{\pm\xi} &\equiv l_{03} \pm i l_{02}, \\ l_{\pm\eta} &\equiv l_{12} \mp i l_{13}, \\ l_{\pm\beta} &\equiv l_{14} \mp i l_{15}, \quad \text{etc.}, \end{aligned} \tag{6.13}$$

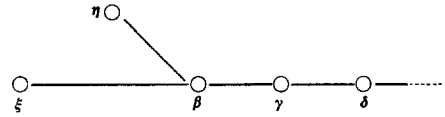
replacing (6.6). Instead of (6.8), we set, for $SO(2n, C)$,

$$\begin{aligned} e_{\pm\xi} &\equiv \pm\tfrac{1}{2}(l_{\pm\xi} + l_{\pm\eta}) = \tfrac{1}{2}\{\pm(l_{12} + l_{03}) - i(l_{13} - l_{02})\}, \\ e_{\pm\eta} &\equiv \pm\tfrac{1}{2}(-l_{\pm\xi} + l_{\pm\eta}) \\ &\quad = \tfrac{1}{2}\{\pm(l_{12} - l_{03}) - i(l_{13} + l_{02})\}, \\ h_\xi &\equiv i(l_{23} + l_{01}), \\ h_\eta &\equiv i(l_{23} - l_{01}), \\ e_{\pm\beta} &\equiv \pm\tfrac{1}{2}[l_{\pm\beta}, l_{\mp\eta}] = \tfrac{1}{2}\{\pm(l_{24} + l_{35}) - i(l_{25} - l_{34})\}, \\ h_\beta &\equiv i(l_{45} - l_{23}), \quad \text{etc.} \end{aligned} \tag{6.14}$$

Subsequent expressions are exactly the same as (6.8), for $SO(2n+1, C)$. These operators satisfy the commutation relations

$$\begin{aligned} &[e_{+\xi}, e_{-\xi}] = h_\xi, && [h_\xi, e_{\pm\xi}] = \pm 2 e_{\pm\xi}, \\ &[e_{+\eta}, e_{-\eta}] = h_\eta, && [h_\eta, e_{\pm\eta}] = \pm 2 e_{\pm\eta}, \\ &[e_{\pm\xi}, e_{+\eta}] = [e_{\pm\xi}, e_{-\eta}] = 0, \\ &[h_\xi, e_{\pm\eta}] = [h_\eta, e_{\pm\xi}] = 0, \\ &[h_\beta, e_{\pm\xi}] = \mp e_{\pm\xi}, && [h_\xi, e_{\pm\beta}] = \mp e_{\pm\beta}, \\ &[h_\beta, e_{\pm\eta}] = \mp e_{\pm\eta}, && [h_\eta, e_{\pm\beta}] = \mp e_{\pm\beta}, \\ &\qquad\qquad \text{etc.} \end{aligned} \tag{6.15}$$

These commutation relations and all subsequent ones are of the form (6.11), omitting (6.11e), if ξ and η are both considered to be adjacent to β but not to each other. Thus the Dynkin diagram and the actions of the raising operators are as follows, for $SO(2n, C) \sim D_n$:



$$\begin{aligned} e_{+\xi}&: [\ 2 \quad 0 \quad -1 \quad 0 \quad 0 \quad \cdots] \\ e_{+\eta}&: [\ 0 \quad 2 \quad -1 \quad 0 \quad 0 \quad \cdots] \\ e_{+\beta}&: [-1 \quad -1 \quad 2 \quad -1 \quad 0 \quad \cdots] \\ e_{+\gamma}&: [\ 0 \quad 0 \quad -1 \quad 2 \quad -1 \quad \cdots] \\ &\text{etc.} \end{aligned} \tag{6.16}$$

The real pseudo-orthogonal group $SO(p,q)$ preserves the bilinear form

$$\tilde{x}'y' \equiv x_1'y_1' + \cdots + x_p'y_p' - x_{p+1}'y_{p+1}' - \cdots - x_{p+q}'y_{p+q}', \quad (6.17)$$

where the coordinates x_j' and y_j' are real; it can be viewed as the subgroup of $SO(p+q, C)$ which preserves the reality conditions of the subspace

$$\begin{aligned} x_j &= x_j' \text{(real)} && \text{for} \quad 1 \le j \le p, \\ x_j &= ix_j' \text{(imaginary)} && \text{for} \quad p+1 \le j \le p+q \end{aligned} \quad (6.18)$$

of the complex $(m = p+q)$-dimensional space of (6.1). Thus Eqs. (6.2)–(6.16) can be retained, but with certain reality conditions imposed on the parameters defined in (6.4) and (6.10). For the parameters of (6.4), these are evidently

$$\begin{aligned} \xi_{ij} &\text{ is real} && \text{if} \quad i \le p < j, \\ \xi_{ij} &\text{ is imaginary} && \text{if} \quad i < j \le p \quad \text{or} \quad p < i < j; \end{aligned} \quad (6.19)$$

that is, the ξ_{ij} are real if i, j span the change in metric, and are imaginary otherwise. The reality conditions on the parameters in (6.10) are determined by the requirement that the replacement of the h_ρ and $e_{\pm\rho}$ by the l_{ij} according to the definitions (6.8) or (6.14) must lead to an expression of the form (6.4) with the ξ_{ij} satisfying (6.19). Examination of (6.8) and (6.14) shows that, except when p and q are both odd, we can label the coordinates in such a way that only one pair of operators $e_{\pm\rho_0}$ are constructed from l_{ij}'s which span the change in metric. Thus we find the reality conditions

$$\begin{aligned} \xi_\rho^* &= +\xi_\rho, \\ \xi_{\pm\rho_0}^* &= -\xi_{\mp\rho_0}, \\ \xi_{\pm\rho}^* &= +\xi_{\mp\rho} && \text{if} \quad \rho \neq \rho_0, \\ \xi_{\pm(\rho+\sigma)}^* &= +\xi_{\mp(\rho+\sigma)} && \text{if} \quad \rho = \rho_0 \text{ or } \sigma = \rho_0, \\ \xi_{\pm(\rho+\sigma)}^* &= -\xi_{\mp(\rho+\sigma)} && \text{if} \quad \rho \neq \rho_0 \text{ and } \sigma \neq \rho_0, \quad \text{etc.}, \end{aligned} \quad (6.20)$$

provided that p and q are not both odd. (We note that $\xi_{\pm 2\rho_0}$ does not occur in the expansion since the corresponding commutator vanishes.)

In a unitary representation, with

$$A = e^{iB}, \quad (6.21)$$

we must have

$$B^\dagger = B, \quad (6.22)$$

where

$$B = \sum \xi_\rho H_\rho + \sum_{\pm\rho,\rho} \xi_{\pm\rho} E_{\pm\rho} + \sum_{\pm,\rho,\sigma} \xi_{\pm(\rho+\sigma)} [E_{\pm\rho}, E_{\pm\sigma}] + \cdots. \quad (6.23)$$

Equations (6.20), (6.22), and (6.23) are satisfied by the choice[10]

$$\begin{aligned} H_\rho^\dagger &= H_\rho, \\ E_{\pm\rho_0}^\dagger &= -E_{\mp\rho_0}, \\ E_{\pm\rho}^\dagger &= +E_{\mp\rho} \quad \text{for} \quad \rho \neq \rho_0, \end{aligned} \quad (6.24)$$

where $E_{\pm\rho_0}$ are the operators which span the change in metric. Thus, again, there is only one pair of noncompact simple generators (or stepping operators), provided p and q are not both odd; that is, the ρ_0 strings are infinite and subject to the normalization condition (3.33), while all others are finite and symmetrical and are subject to the normalization condition (3.32).

The algebras corresponding to $SO(p,q)$ with p and q both odd remain as a special case. The difficulty can be seen in the first such semisimple group, $SO(3,1) \sim SL(2, C)$. The generators in this case are just the first six (the ξ and η operators) of (6.14), and the expansion (6.10) is simply

$$b = \xi_\xi h_\xi + \xi_\eta h_\eta + \xi_{+\xi} e_{+\xi} + \xi_{-\xi} e_{-\xi} + \xi_{+\eta} e_{+\eta} + \xi_{-\eta} e_{-\eta}, \quad (6.25)$$

since $[e_{\pm\xi}, e_{\pm\eta}] = 0$. If the coordinates are labeled so that the metric takes the form $(+ - - -)$, then the coefficients in (6.25) must be chosen so that the coefficients of l_{0i} are real, while those of l_{ij} for $i, j \neq 0$ are imaginary. This requires

$$\xi_\xi^* = \xi_\eta \quad \text{and} \quad \xi_{\pm\xi}^* = \xi_{\mp\eta} \quad (6.26)$$

which suggests

$$H_\xi^\dagger = H_\eta \quad \text{and} \quad E_{\pm\xi}^\dagger = E_{\mp\eta} \quad (6.27)$$

for a unitary representation. Equations (6.27) are consistent with the commutation relations

$$\begin{aligned} &[H_\xi, E_{\pm\xi}] = \pm 2E_{\pm\xi}, \quad [H_\eta, E_{\pm\eta}] = \pm 2E_{\pm\eta}, \\ &[E_{+\xi}, E_{-\xi}] = H_\xi, \quad [E_{+\eta}, E_{-\eta}] = H_\eta, \\ &[E_{+\xi}, E_{\pm\eta}] = [E_{-\xi}, E_{\pm\eta}] = [H_\eta, E_{\pm\xi}] \\ &\qquad = [H_\xi, E_{\pm\eta}] = 0. \end{aligned} \quad (6.28)$$

However, diagonalization of H_ξ and H_η is evidently not possible, since the first of (6.27) requires the eigenvalues to be related by complex conjugation, while Eqs. (6.28) say that $E_{+\xi}$ raises the former eigenvalue by $+2$ but leaves the latter eigenvalue unchanged. The question of bounded representations for the algebras corresponding to $SO(p,q)$ with p and q both odd will not be pursued further here.

Figure 11 shows a representation of the algebra of $SO(3,2)$ constructed from the commutation relations (6.11b)–(6.11f) and the unitarity conditions (5.18).

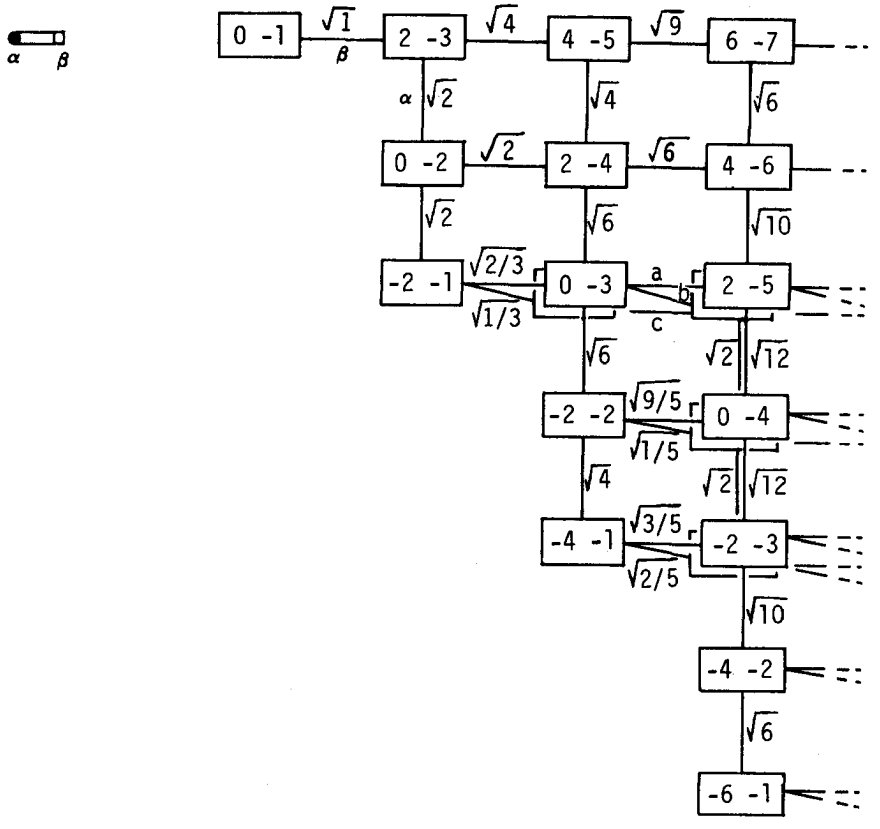


FIG. 11. A unitary representation of $SO(3, 2) \sim B_2$; $a \equiv (\frac{1}{6})^{\frac{1}{2}}$, $b \equiv (\frac{1}{10})^{\frac{1}{2}}$, $c \equiv (\frac{1}{3})^{\frac{1}{2}}$. The basis states corresponding to degenerate weights have been chosen to be eigenstates of the $SU(2)_\alpha$ Casimir operator. (Normalization factors for $E_{+\beta}$ are the negatives of those shown for $E_{-\beta}$.)

Note that some weights are degenerate in this representation, with a resulting complexity in the normalization factors.

7. APPLICATIONS TO DYNAMICAL GROUPS

The methods described above will now be illustrated by application to two cases of possible physical interest. The results are not new; the purpose of this section is to show how the present methods yield a rather pictorial derivation of some known results, proceeding directly from the commutation relations for the groups in question.

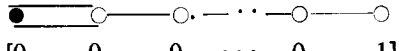
A. Spherical Harmonics in N Dimensions

The dynamical groups of a certain class of physically interesting equations have been shown to be $SO(N, 2)$.[5] This demonstration proceeded from the fact that the solutions are essentially spherical harmonics in N dimensions. The connection between these spherical harmonics and $SO(N, 2)$ is readily shown as follows.

We first note that a basis for the N-dimensional spherical harmonics Y_k^t of degree t can be obtained from the tth-order monomials

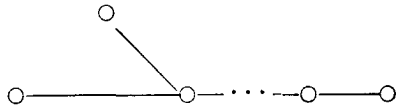
$$z_{i_1} z_{i_2} \cdots z_{i_t} \tag{7.1}$$

in the coordinates z_i of an N-dimensional space by forming a maximal set of polynomials irreducible under $SO(N)$ and evaluating them on the unit sphere. The coordinates z_i transform like the (N-dimensional) vector representation of $SO(N)$. The highest weight of this representation for $SO(2n + 1) \sim B_n$ with $N = 2n + 1 \geq 5$ is



$$[0 \quad 0 \quad 0 \quad \cdots \quad 0 \quad 1] \tag{7.2a}$$

and for $SO(2n) \sim D_n$ with $N = 2n \geq 6$ is



$$[0 \quad 0 \quad 0 \quad \cdots \quad 0 \quad 1] \tag{7.2b}$$

The monomials (7.1) are necessarily symmetric; that is, they belong to the t-fold symmetric direct product of the appropriate representation (7.2) with itself. Since the weight of a direct product of vectors is just the sum of the weights of the vectors, the highest weight occurring in the product representation is just

$$[0 \quad 0 \quad 0 \quad \cdots \quad 0 \quad t] \tag{7.3}$$

for N either odd or even. Thus the Y_k^t form a basis for

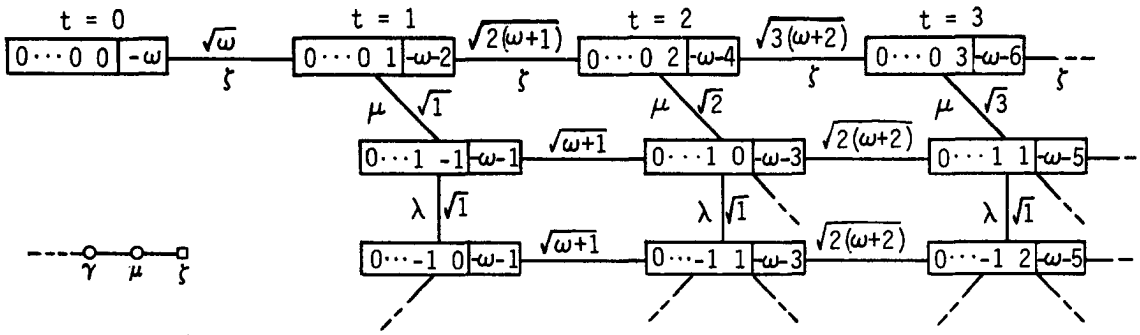
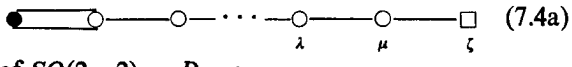


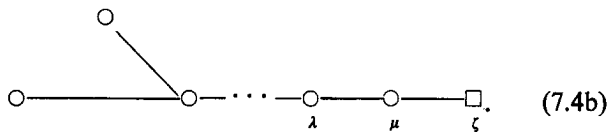
FIG. 12. A unitary representation of $SO(N, 2)$ for $N \geq 5$, which is spanned by the N-dimensional spherical harmonics Y_k^t for all t. The relation $[E_{+\lambda}, E_{-\zeta}] = 0$ requires $\omega = 1$, by Eq. (7.8). (Normalization factors for $E_{+\zeta}$ are the negatives of those shown for $E_{-\zeta}$.)

the representation with greatest weight (7.3). Representations of $SO(N, 2)$ of the form indicated in Fig. 12 appear to yield the complete set of representations (7.3) for all t, where the last eigenvalue in each weight is that of H_ζ, defined by the scheme of $SO(2n + 1, 2) \sim B_{n+1}$:



(7.4a)

or of $SO(2n, 2) \sim D_{n+1}$:



(7.4b)

There is, however, a further condition on ω arising from the commutation relation $[E_{+\mu}, E_{-\zeta}] = 0$. This relation is satisfied for the states shown in Fig. 12, and trivially for the states lying below these, until we come to the last two states of the $t = 1$ multiplet and the states of the $t = 2$ multiplet connected to them. These states are shown in Fig. 13. At first sight, the situation appears to be complicated by the fact that the weight $[0 \cdots 0\ 0 \mid -\omega - 2]$ is degenerate. However, only the state $(1/\sqrt{2})E_{-\mu} |0 \cdots -1\ 2 \mid -\omega - 3\rangle$ connects with $|0 \cdots 0\ -1 \mid -\omega\rangle$; for example, $[E_{+\zeta}, E_{-\lambda}] = 0$ yields

$$E_{+\zeta}E_{-\lambda} |0 \cdots 2\ -1 \mid -\omega - 2\rangle = E_{-\lambda}E_{+\zeta} |0 \cdots 2\ -1 \mid -\omega - 2\rangle = 0, \quad (7.5)$$

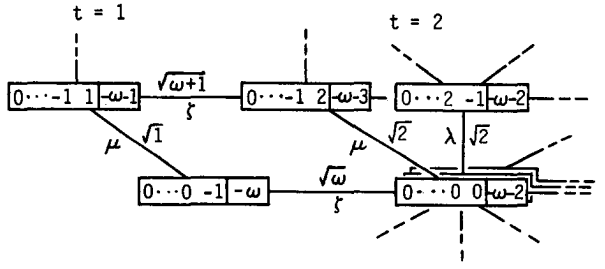


FIG. 13. The portion of the representation of Fig. 12 which contains the bottom of the $t = 1$ multiplet and the states of the $t = 2$ multiplet which connect to it. The weight $[0 \cdots 0\ 0 \mid -\omega - 2]$ has multiplicity n for N odd and $n - 1$ for N even.

since no state of the $t = 1$ multiplet is connected to $|0 \cdots 2\ -1 \mid -\omega - 2\rangle$. Thus the relation

$$0 = [E_{+\mu}, E_{-\zeta}] |0 \cdots 0\ -1 \mid -\omega\rangle \quad (7.6)$$

yields just

$$0 = (2\omega)^{\frac{1}{2}} - (\omega + 1)^{\frac{1}{2}} \quad (7.7)$$

or

$$\omega = 1. \quad (7.8)$$

The cases $N = 3$ and $N = 4$ remain to be treated, but they go through in much the same way. Note that (7.3) is replaced by $[2t]$ for $N = 3$ and by $[t\ \ t]$ for $N = 4$; $\omega = \frac{1}{2}$ for $N = 3$.

The observation[5] that these representations remain irreducible under the subgroup $SO(N, 1)$ can be made as follows: From (6.8) or (6.14) we see that the restriction of $SO(N, 2)$ to this subgroup is equivalent to dropping the stepping operators $E_{\pm\zeta}$ and substituting just the one operator

$$N \text{ odd: } L_{1,N+1} = \tfrac{1}{2}[\cdots[[E_{+\alpha} - E_{-\alpha}, E_{+\beta} - E_{-\beta}], E_{+\gamma} - E_{-\gamma}], \cdots, E_{+\zeta} - E_{-\zeta}] \quad (7.9a)$$

or

$$N \text{ even: } L_{1,N} = \tfrac{1}{2}[\cdots[[E_{+\xi} - E_{-\xi} + E_{+\eta} - E_{-\eta}, E_{+\beta} - E_{-\beta}], E_{+\gamma} - E_{-\gamma}], \cdots, E_{+\zeta} - E_{-\zeta}]. \quad (7.9b)$$

Applied to the $t = 0$ state of Fig. 12, it evidently yields a state of the $t = 1$ multiplet. [Note that only one ordering of the commutants in (7.9a) or (7.9b) contributes, and only the lowering operator from each pair.] Since the stepping operators of $SO(N)$ are still included, we get all of the $t = 1$ multiplet. In general, the operator (7.9a) or (7.9b) applied to the highest state of multiplet t yields a linear combination of a state of multiplet $t + 1$ and a state of multiplet $t - 1$. Thus, if we progress from left to right across Fig. 12, we see that the irreducible representation of $SO(N, 1)$ which contains the $t = 0$ state must also contain all the other states of that $SO(N, 2)$ representation.

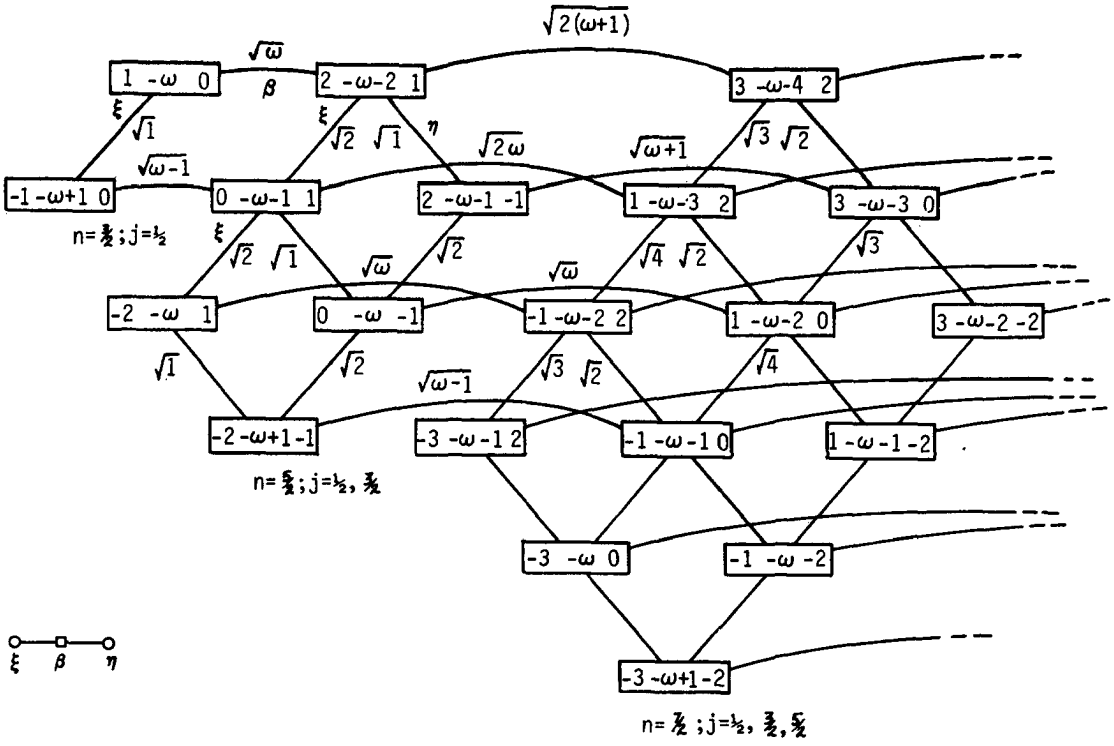


FIG. 14. The baryon representation (see Ref. 9). $[E_{+\xi}, E_{-\beta}] = 0$ requires $\omega \geq 2$ in general, and $\omega = 2$ if the weights are to be nondegenerate. For $\omega = 2$, n is the eigenvalue of $-\frac{1}{2}H$. Normalization factors for $E_{+\beta}$ are the negatives of those shown for $E_{-\beta}$. For $\omega \neq 2$, two of the $n = \frac{5}{2}$ weights and six of the $n = \frac{7}{2}$ weights become degenerate, with a corresponding increase in the number of connecting paths and associated normalization factors. [A similar case of degenerate weights for $SO(3, 2)$ is shown in Fig. 11.] $\sqrt{2}J_{\pm} = E_{\pm\xi} + E_{\pm\eta}$, $2J_z = H_\xi + H_\eta$.

B. The Dynamical Group Model of the Baryons

In the dynamical group model of the baryons,[9] the $S = 0$, $T = \frac{1}{2}$, $T_3 = +\frac{1}{2}$ baryons are associated with the representation of $SO(4, 2) \sim SU(2, 2)$ shown in Fig. 14, for $\omega = 2$. More precisely, these baryons are assigned to linear combinations of the states of this representation and those of the similar representation with greatest weight $[0\ -\omega\ 1]$ which have positive parity; the other combinations are presumably to be associated with a sequence of negative-parity baryon resonances. These representations are singled out by the requirement that they have a spin doublet at the top. The corresponding antibaryon states are then, of course, to be associated with the representations having lowest weights $[-1\ \omega\ 0]$ and $[0\ \omega\ -\ 1]$. The choice[9] of $\omega = 2$ is favored by the present experimental situation, since it corresponds to nondegenerate weights and, hence, to the least number of low-spin baryons.

One would like to be able to identify the proton, for example, with the $n = \frac{3}{2}$ doublet of Fig. 14. However, in order to obtain the correct characteristic mass $\approx m_\omega$ (rather than $2m_p$) in the form factors, it is necessary to apply the tilt operator $(1/N)\exp(i\theta L_{04})$ $[(1/N)\exp(i\theta L_{45})$ in the notation of Ref. 9] to these states; this has the effect that the proton is now a linear combination of all the $j = \frac{1}{2}$ multiplets in Fig. 14. Alternatively, of course, one can view this instead as a complication of the simple form of the current and boost operators which enter the model.[9]

The calculation of the proton magnetic form factor can now be easily traced out; that for the electric form factor is similar, except that parity eigenstates must be used throughout to avoid spurious terms. We have[9]

$$G_M^p(t) = \frac{1}{\sinh\frac{1}{2}\zeta}\langle \mathbf{p}' = 0, j_z = -\tfrac{1}{2}|\,\mathcal{J}^2\,|\mathbf{p}, j_z = +\tfrac{1}{2}\rangle$$
$$= \sum_i \frac{1}{N_p^2 \sinh\frac{1}{2}\zeta}$$
$$\times \langle -1 \quad -\omega+1 \quad 0|\, e^{-i\theta L_{04}}\mathcal{J}^2 e^{i\theta L_{04}}\,|i\rangle$$
$$\times \langle i|\, e^{-i\theta L_{04}} e^{i\zeta L_{14}} e^{i\theta L_{04}}\,|1 \quad -\omega \quad 0\rangle, \quad (7.10)$$

where $\mathcal{J}^2$ is the x_2 component of the current operator, the sum is over a complete set of intermediate states, and $\exp(i\zeta L_{14})$ is the boost which transforms a state from rest to momentum $\mathbf{p}$ along the x_1 axis. The space coordinates have been taken to be $x_1 = z$, $x_2 = x$, and $x_3 = y$, so that $J_z = J_1 = \frac{1}{2}(H_\xi + H_\eta)$ is diagonal; $x_4 = t$ is the time coordinate, while x_0 and x_5 are the extra spacelike and timelike coordinates, respectively.

Also,

$$\zeta \equiv \sinh^{-1}(p/m) \quad \text{and} \quad t = 2m(m - E). \tag{7.11}$$

The first matrix elements after the summation sign in (7.10) are easily evaluated, since the current operator in this model is chosen to be a sum of terms,

$$\mathcal{J}^\mu = \sum a_i \mathcal{J}_i^\mu, \tag{7.12}$$

at most linear in momenta and in $SO(4, 2)$ generators; hence, each has simple transformation properties. In particular, let us consider the term made purely from $SO(4, 2)$ generators:

$$\mathcal{J}_1^\mu \equiv (L_{15}, L_{25}, L_{35}, L_{45}). \tag{7.13}$$

Since [see Eqs. (6.14)]

$$\mathcal{J}_1^2 = L_{25} = \tfrac{1}{2}i(E_{+\beta} + E_{-\beta} + [E_{+\xi}, [E_{+\beta}, E_{+\eta}]] + [E_{-\xi}, [E_{-\beta}, E_{-\eta}]]) \tag{7.14}$$

commutes with L_{04}, the matrix elements in question are easily evaluated from Fig. 14; the nonzero matrix elements involving $\langle -1 \quad -\omega+1 \quad 0|$ are, for $\omega = 2$,

$$\langle -1 \quad -\omega+1 \quad 0|\, \mathcal{J}_1^2 \,|0 \quad -\omega-1 \quad 1\rangle = -\tfrac{1}{2}i,$$
$$\langle -1 \quad -\omega+1 \quad 0|\, \mathcal{J}_1^2 \,|-2 \quad -\omega+1 \quad -1\rangle = i/\sqrt{2}. \tag{7.15}$$

Evaluation of the other matrix elements in (7.10) is facilitated by writing the operator in the form[13]:

$$e^{-i\theta L_{04}} e^{i\zeta L_{14}} e^{i\theta L_{04}} = e^{\alpha L_{10}} e^{i\beta L_{14}} e^{\gamma L_{10}}. \tag{7.16}$$

This can be viewed as a sort of Euler-angle representation in 014 space. The parameters α, β, and γ in (7.16) can be evaluated by replacing the L_{ij} by $\frac{1}{2}i\sigma_k$; this yields

$$e^{-\frac{1}{2}i(\alpha+\gamma)} = \frac{\cosh \tfrac{1}{2}\zeta + i \sinh\theta \sinh \tfrac{1}{2}\zeta}{\cosh \tfrac{1}{2}\beta},$$
$$e^{\frac{1}{2}i(\alpha-\gamma)} = 1, \tag{7.17}$$
$$\sinh \tfrac{1}{2}\beta = \cosh\theta \sinh \tfrac{1}{2}\zeta.$$

Matrix elements of the operator on the right-hand side of (7.16) are readily evaluated, since L_{10} is diagonal with respect to the states of Fig. 14:

$$L_{10} = \tfrac{1}{2}i(H_\xi - H_\eta), \tag{7.18}$$

from (6.14). This leaves only the evaluation of matrix elements of $\exp(i\beta L_{14})$. For this purpose we decompose Fig. 14 into representations of an appropriate triple of operators satisfying the canonical commutation relations

$$[E_+, E_-] = H, \quad [H, E_\pm] = \pm 2E_\pm, \tag{7.19}$$

with $E_+ + E_-$ proportional to L_{14}; $\exp(i\beta L_{14})$ can

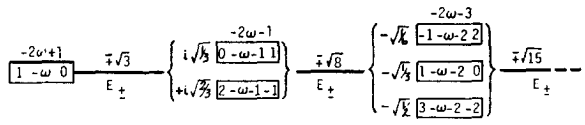


FIG. 15. A subspace of Fig. 14 which is a representation of $E_\pm$ and H, for $\omega = 2$. For $\omega \neq 2$, the second and third states become linear combinations of three and six states, respectively. The numbers above the boxes are eigenvalues of H.

then be evaluated as in Appendix A. From (6.14), we see that a suitable choice is

$$E_\pm \equiv -i[E_{\pm\beta}, E_{\pm\xi} + E_{\pm\eta}] = iL_{14} \pm L_{15},$$
$$H \equiv H_\xi + 2H_\beta + H_\eta = 2iL_{45}. \tag{7.20}$$

The representation of (7.20) which starts with the weight $[1 \quad -\omega \quad 0]$ can easily be extracted from Fig. 14; it is shown in Fig. 15. Since $2iL_{14} = E_+ + E_-$, Eqs. (A22) and (A23) together with Fig. 15 yield, for $\omega = 2$,

$$e^{i\beta L_{14}}|1 \quad -\omega \quad 0\rangle = (\cosh \tfrac{1}{2}\beta)^{-3}[|1 \quad -\omega \quad 0\rangle + i\sqrt{3}\tanh(\tfrac{1}{2}\beta)(\sqrt{\tfrac{1}{3}}\,|0 \quad -\omega-1 \quad 1\rangle + \sqrt{\tfrac{2}{3}}\,|2 \quad -\omega-1 \quad -1\rangle) + \cdots], \tag{7.21}$$

so that the matrix elements between $|1 \quad -\omega \quad 0\rangle$ and the kets of (7.15) are

$$\langle 0 \quad -\omega-1 \quad 1|\, e^{\alpha L_{10}} e^{i\beta L_{14}} e^{\gamma L_{10}} |1 \quad -\omega \quad 0\rangle = i\frac{\sinh \tfrac{1}{2}\beta}{(\cosh \tfrac{1}{2}\beta)^4} e^{\frac{1}{2}i(\gamma-\alpha)},$$
$$\langle -2 \quad -\omega+1 \quad -1|\, e^{\alpha L_{10}} e^{i\beta L_{14}} e^{\gamma L_{10}} |1 \quad -\omega \quad 0\rangle = 0. \tag{7.22}$$

Equations (7.10), (7.15), (7.17), and (7.22) yield $G_{M1}^p(t)$ for $\omega = 2$. The form of the result for $\omega \neq 2$ is obtained by noting that the power of $\cosh \frac{1}{2}\beta$ in (7.21) is always $-2\omega + 1$, the eigenvalue of H on the state $|1 \quad -\omega \quad 0\rangle$. Thus

$$G_{M1}^p(t) = \frac{f(\omega)}{2N_p^2} \frac{\cosh\theta}{(\cosh \tfrac{1}{2}\beta)^{2\omega}}, \tag{7.23}$$

where $f(\omega) = 1$ for $\omega = 2$. From (7.11) and the last of (7.17),

$$\cosh^2 \tfrac{1}{2}\beta = 1 - \frac{\cosh^2\theta}{4m^2}\, t. \tag{7.24}$$

The contributions to the magnetic form factor from the other terms in the current (7.12) used in Ref. 9 can be calculated in the same way. These contributions also have the t dependence $(\cosh \frac{1}{2}\beta)^{-2\omega}$; thus the choice[9] of $\omega = 2$ not only yields the smallest number of low-spin resonances, but also leads to a second-order pole in $G_M^p(t)$, in agreement with

experiment.[9] (Of course, θ is chosen so that the pole occurs at the correct place, namely, $2m/\cosh\theta \approx m_\omega$.)

ACKNOWLEDGMENTS

The author would like to express his appreciation to C. J. Hieggelke for help in working through portions of the manuscript, to Professor W. M. Fairbairn for pointing out several errors in the original manuscript, and to Professors B. Hamprecht and H. Kleinert for providing references.

APPENDIX A: EVALUATION OF GROUP ELEMENTS

Matrix elements of noninfinitesimal group elements are often needed in the applications of noncompact groups. Except for the simplest representations, it is impractical to obtain these directly from the exponential expression (2.1). For completeness we describe here, in some detail and in terms of the present notation, a method of calculating such matrix elements for the rank-one subgroups $SU(2)$ and $SU(1,1)$.[14] As was seen in Sec. 7B, this permits the calculation of matrix elements for larger groups also.

The differential operators

$$\mathcal{E}_+ \equiv x\partial_y, \quad \mathcal{E}_- \equiv y\partial_x, \quad \mathcal{K} \equiv x\partial_x - y\partial_y \tag{A1}$$

satisfy the commutation relations

$$[\mathcal{K}, \mathcal{E}_\pm] = \pm 2\mathcal{E}_\pm, \quad [\mathcal{E}_+, \mathcal{E}_-] = \mathcal{K} \tag{A2}$$

and hence are a realization of the generators for $SL(2)$ or its subgroups $SU(2)$ or $SU(1,1)$. A general transformation of $SL(2)$ is then written in the form

$$\mathcal{W} = e^{\alpha\mathcal{E}_+ + \beta\mathcal{E}_- + \gamma\mathcal{K}}. \tag{A3}$$

Since the operators (A1) are differential operators, a general analytic function $f(x,y)$ transforms according to

$$\mathcal{W}f(x,y) = f[\mathcal{W}(x), \mathcal{W}(y)]. \tag{A4}$$

For example,

$$e^{\alpha\mathcal{E}_+} = e^{\alpha x\partial_y} = 1 + (\alpha x)\partial_y + \tfrac{1}{2}(\alpha x)^2\partial_y^2 + \cdots, \tag{A5}$$

so that

$$e^{\alpha\mathcal{E}_+}f(x,y) = f(x, y + \alpha x), \tag{A6}$$

as can be seen by comparison with a Taylor series expansion of f in the second variable about the point y. A similar result holds for the subgroup generated by $\mathcal{E}_- = y\partial_x$. Finally, a consideration of the action of the differential operators on the monomials in a power series expansion shows that

$$e^{\gamma\mathcal{K}}f(x,y) = f(e^\gamma x, e^{-\gamma}y). \tag{A7}$$

The actions of these three subgroups on the coordinate functions are thus

$$e^{\alpha\mathcal{E}_+}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}x\\ \alpha x + y\end{bmatrix}, \quad e^{\beta\mathcal{E}_-}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}x + \beta y\\ y\end{bmatrix},$$
$$e^{\gamma\mathcal{K}}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}e^\gamma x\\ e^{-\gamma}y\end{bmatrix}. \tag{A8}$$

The transformations (A3) or (A8) on the coordinates can also be written in the form

$$\mathbf{x}' = \tilde{W}\mathbf{x}, \tag{A9}$$

where $\mathbf{x}$ is the column formed from x and y, and W is the matrix operator

$$W = e^{\alpha e_+ + \beta e_- + \gamma h} \tag{A10}$$

in the fundamental representation, with

$$e_+ = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, \quad e_- = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}, \quad h = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{A11}$$

We review first the familiar case of $SU(2)$, for which

$$\tilde{W} = \begin{bmatrix}u & v\\ -v^* & u^*\end{bmatrix} \quad \text{with} \quad |u|^2 + |v|^2 = 1, \tag{A12}$$

so that $\tilde{W}^\dagger\tilde{W} = I$ and

$$\mathbf{x}^\dagger\mathbf{x} = |x|^2 + |y|^2 \equiv a^2 \tag{A13}$$

is an invariant. These transformations can be visualized as moving points around on "spheres" in a complex two-dimensional space; in particular, they induce transformations on sets of functions defined on the unit sphere. Such a set of functions must, of course, transform according to some representation of $SU(2)$; generally, this will be a larger representation than the two-dimensional one which acts on the $x - y$ space. Using the definition (A1), it is easily shown that the monomials[15]

$$N_k x^{-k-\omega}y^k, \quad k \text{ integral}, \tag{A14}$$

belong to a representation of the form shown in Fig. 16. But we know that such a representation of $SU(2)$ is unitary if and only if it is bounded at both ends. Some typical unitary representations are shown in Fig. 17. We note that for unitary representations ω is a nonpositive integer and $0 \le k \le |\omega|$. With these restrictions, the monomials (A14) are analytic everywhere (except at infinity) and hence are a suitable

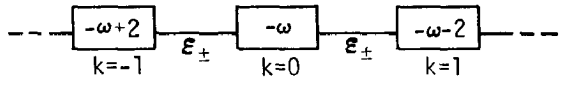


FIG. 16. The form of the monomial representations. The numbers in the boxes are eigenvalues of $\mathcal{K}$.

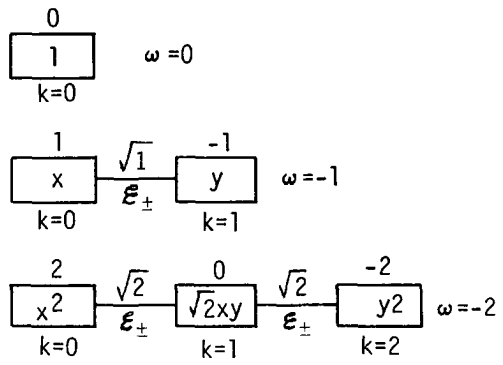


FIG. 17. Unitary representations of $SU(2)$. The eigenvalues of $\mathcal{K}$ are written above the boxes.

domain for the operators (A1). The explicit evaluation of matrix elements for $SU(2)$ will not be carried out here, since it is both well known and analogous to that for $SU(1,1)$, which will be treated next.

For $SU(1,1)$,

$$\tilde{W} = \begin{bmatrix} u & v \\ v^* & u^* \end{bmatrix} \quad \text{with} \quad |u|^2 - |v|^2 = 1, \tag{A15}$$

so that $\tilde{W}^\dagger \sigma_3 \tilde{W} = \sigma_3$ and

$$\mathbf{x}^\dagger \sigma_3 \mathbf{x} = |x|^2 - |y|^2 \equiv a^2 \tag{A16}$$

is an invariant. These transformations move points around on "hyperbolas" in the complex two-dimensional space; in particular, functions on the unit hyperbola ($a = 1$) transform according to representations of $SU(1,1)$. Again, the monomials (A14) belong to a representation of the form of Fig. 16. However, the conditions for unitarity are changed; for example, there is the class of unitary representations bounded above shown in Fig. 18. In this case the functions are singular at $x = 0$; but this point is not on the unit hyperbola, so that the generators (A1) are still well defined.

We next consider the problem of obtaining matrix elements corresponding to the transformation (A15) for $SU(1,1)$. Let

$$z \equiv y/x, \quad w \equiv 1/x, \tag{A17}$$

so that the basis functions (A14) take the form

$$N_k z^k w^\omega, \quad k \text{ integral}, \quad \omega > 0, \tag{A18}$$

and the unit hyperbola $|x|^2 - |y|^2 = 1$ becomes the sphere

$$|w|^2 + |z|^2 = 1. \tag{A19}$$

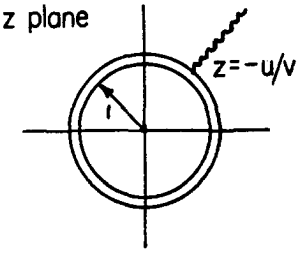


FIG. 19. The circle of convergence of Eq. (A20) (outer circle) and the range of z on the manifold (A19) (inner circle). The branch cut exists unless ω is integral.

When the transformation (A15) is applied to $\mathbf{x}$, the basis functions transform to

$$\begin{aligned} &N_k z'^k w'^\omega \\ &= N_k \frac{(v^*x + u^*y)^k}{(ux + vy)^{k+\omega}} = N_k w^\omega \frac{(u^*z + v^*)^k}{(vz + u)^{k+\omega}} \\ &= N_k u^{-k-\omega} w^\omega (u^*z + v^*)^k [1 - (k+\omega)(v/u)z \\ &\quad + \tfrac{1}{2}(k+\omega)(k+\omega+1)(v/u)^2 z^2 - \cdots]. \end{aligned} \tag{A20}$$

The functions above have only the one singularity at $z = -u/v$, so this expansion converges within the circle $|z| = |u/v|$ (see Fig. 19). Since $|u/v| > 1$ according to (A15), this region of convergence includes the entire manifold (A19). The expansion (A20) can evidently be put into the form

$$N_k z'^k w'^\omega = \sum_l A_l N_l z^l w^\omega. \tag{A21}$$

The coefficients A_l are then the desired matrix elements of $\mathcal{W}$. For example, the operators of interest in Sec. 7B are of the form

$$W_\alpha = e^{\alpha(e_+ + e_-)} = \begin{bmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{bmatrix}, \quad \alpha \text{ real}; \tag{A22}$$

that is, $u^* = u = \cosh\alpha$ and $v^* = v = \sinh\alpha$. Labeling states by the eigenvalues of $\mathcal{K}$ ($= -2k - \omega$), we find

$$\begin{aligned} \mathcal{W}_\alpha |-\omega\rangle &= w'^\omega \\ &= c^{-\omega} w^\omega [1 - \omega t z + \tfrac{1}{2}\omega(\omega+1) t^2 z^2 - \cdots] \\ &= c^{-\omega} [|-\omega\rangle + (\sqrt{\omega}) t \, |-\omega - 2\rangle \\ &\quad + [\tfrac{1}{2}\omega(\omega+1)]^{\frac{1}{2}} t^2 \, |-\omega - 4\rangle + \cdots], \end{aligned} \tag{A23}$$

where $c \equiv \cosh\alpha$ and $t \equiv \tanh\alpha$. The following

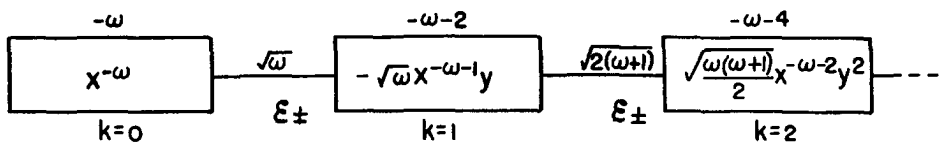


FIG. 18. Unitary representations of the algebra of $SU(1,1)$ bounded above; $\omega > 0$. Eigenvalues of $\mathcal{K}$ are written above the boxes. The normalization factors given are for $\mathcal{E}_-$; those for $\mathcal{E}_+$ differ by a minus sign.

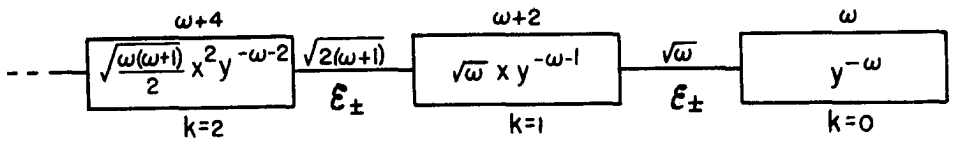


FIG. 20. Unitary representations of the algebra $SU(1, 1)$, bounded below; $\omega > 0$.

matrix elements can now be read off immediately:

$$\begin{aligned}
&\langle -\omega|\, \mathcal{W}_\alpha \,|-\omega\rangle = c^{-\omega},\\
&\langle -\omega - 2|\, \mathcal{W}_\alpha \,|-\omega\rangle = (\sqrt{\omega})tc^{-\omega}, \qquad \text{(A24)}\\
&\langle -\omega - 4|\, \mathcal{W}_\alpha \,|-\omega\rangle = [\tfrac{1}{2}\omega(\omega+1)]^{\frac{1}{2}}t^2c^{-\omega}, \quad \text{etc.}
\end{aligned}$$

These matrix elements are, of course, independent of the fact that we have used a specific realization of the algebra in terms of differential operators; they are determined by the commutation relations.

A similar calculation can be carried through for representations which are bounded below. In this case the monomials take the form

$$N_k x^k y^{-\omega-k}, \qquad \text{(A25)}$$

and

$$z \equiv x/y, \quad w \equiv 1/y. \qquad \text{(A26)}$$

The representations then have the form shown in Fig. 20; Eq. (A20) is replaced by

$$\begin{aligned}
&N_k z'^k w'^\omega\\
&= N_k u^{*-k-\omega} w^\omega (uz + v)^k (1 + v^* z/u^*)^{-k-\omega}\\
&= N_k u^{*-k-\omega} w^\omega (uz + v)^k [1 - (k + \omega)(v^*/u^*)z\\
&\quad + \tfrac{1}{2}(k + \omega)(k + \omega + 1)(v^*/u^*)^2 z^2 - \cdots], \qquad \text{(A27)}
\end{aligned}$$

and Eq. (A24) is replaced by

$$\begin{aligned}
&\langle \omega|\, \mathcal{W}_\alpha \,|\omega\rangle = c^{-\omega},\\
&\langle \omega + 2|\, \mathcal{W}_\alpha \,|\omega\rangle = (-\sqrt{\omega})tc^{-\omega}, \qquad \text{(A28)}\\
&\langle \omega + 4|\, \mathcal{W}_\alpha \,|\omega\rangle = [\tfrac{1}{2}\omega(\omega+1)]^{\frac{1}{2}}t^2c^{-\omega}, \quad \text{etc.}
\end{aligned}$$

General expressions could be written for matrix elements such as those in (A24) and (A28); but for small values of k it is probably easier to evaluate the special cases directly. It should be noted that the representations discussed above are true representations of the *group* $SU(1, 1)$ only when ω is integral, since only in that case is $\exp(2\pi i\mathcal{K})$ equal to the identity, as it is in the fundamental representation and hence in any other true representation. The significance of the other representations of the algebra and the question of multiple-valuedness are discussed in Appendix B.

APPENDIX B: THE COVERING GROUP OF $SO(2, 1)$

The Lie algebra determines much, but not all, of the structure of unitary representations of the corresponding groups. For example, to the algebra A_2 there correspond the two compact groups $SU(3)$ (with quarks) and $SU(3)/Z_3$ (without). The choice among noncompact groups belonging to a particular algebra is much wider; for example, to A_1 there correspond the noncompact groups $SO(2, 1)$, $SO(2, 1) \times Z_2 \cong SU(1, 1)$, $SO(2, 1) \times Z_3$, etc. For true representations of these groups there are the restrictions that $\omega = 2m$, m, $2m/3$, etc., respectively, where m is an integer. The purpose of this appendix is to clarify the status of these representations of the algebra with nonintegral values of ω.

A natural set of generators for $SO(2, 1)$ is

$$\begin{aligned}
&\mathcal{L}_{ij} \equiv x_i\partial_j - x_j\partial_i, \quad \text{for } i, j = 1, 2, 3;\\
&\mathcal{E}_\pm \equiv i\mathcal{L}_{23} \pm \mathcal{L}_{31}, \quad \mathcal{K} \equiv 2i\mathcal{L}_{12}. \qquad \text{(B1)}
\end{aligned}$$

These are essentially the usual angular momentum operators. Setting

$$y \equiv x_1 - ix_2, \quad \bar{y} \equiv x_1 + ix_2, \quad z = -ix_3, \qquad \text{(B2)}$$

we find

$$\begin{aligned}
&\mathcal{K}y = 2y, \quad \mathcal{K}z = 0, \quad \mathcal{K}\bar{y} = -2\bar{y},\\
&\mathcal{E}_+ y = 0, \quad \mathcal{E}_+ z = iy, \quad \mathcal{E}_+ \bar{y} = 2iz, \qquad \text{(B3)}\\
&\mathcal{E}_- y = -2iz, \quad \mathcal{E}_- z = -i\bar{y}, \quad \mathcal{E}_- \bar{y} = 0.
\end{aligned}$$

From these equations, one can easily obtain the familiar finite-dimensional representations that are unitary for $SO(3)$, starting with y^l and ending with $\bar{y}^{-l}$. Of more interest here are representations which are unitary for $SO(2, 1)$; one such class is shown in Fig. 21. In the last function listed, use has been made of the restriction $x_1^2 + x_2^2 + x_3^2 = x_1^2 + x_2^2 - z^2 = 1$, which defines a manifold preserved by the group $SO(3, C)$ with complex parameters. A suitable choice of reality conditions on the parameters leads to the subgroup $SO(2, 1)$, which preserves the real

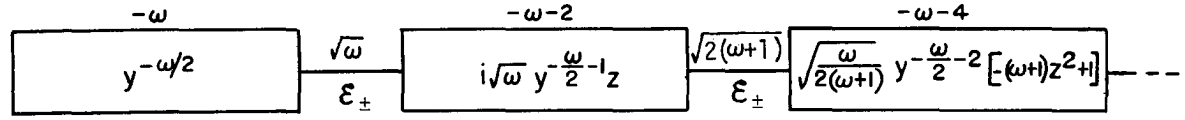


FIG. 21. Unitary representations of the algebra of $SO(2, 1)$ bounded above; $\omega > 0$. The eigenvalues of $\mathcal{K}$ are written above the boxes.

manifold

$$x_1^2 + x_2^2 - z^2 = 1, \quad \text{with } x_1, x_2, \text{ and } z \text{ real.} \tag{B4}$$

The analog of the transformation (A22) is

$$\mathfrak{W}(\alpha) \equiv e^{\alpha(\mathcal{E}_+ + \mathcal{E}_-)} = e^{2i\alpha\mathcal{L}_{23}}. \tag{B5}$$

The action of this transformation on the x_i can be deduced from the series expansion of the exponential, with the result

$$\mathfrak{W}(\alpha)\begin{bmatrix} y \\ \bar{y} \\ z \end{bmatrix} = \begin{bmatrix} c^2 & -s^2 & -2ics \\ -s^2 & c^2 & 2ics \\ ics & -ics & c^2 + s^2 \end{bmatrix}\begin{bmatrix} y \\ \bar{y} \\ z \end{bmatrix}, \tag{B6}$$

where $c \equiv \cosh\alpha$ and $s \equiv \sinh\alpha$. In particular, this yields

$$\begin{aligned} \mathfrak{W}(\alpha)|-\omega\rangle &= y'^{-\frac{1}{2}\omega} \\ &= c^{-\omega}y^{-\frac{1}{2}\omega}[1 - 2itz/y - t^2(1+z^2)/y^2]^{-\frac{1}{2}\omega} \\ &= c^{-\omega}y^{-\frac{1}{2}\omega}[1 + i\omega tz/y \\ &\quad + \tfrac{1}{2}\omega t^2(-\omega z^2 - z^2 + 1)/y^2 + \cdots] \\ &= c^{-\omega}[|-\omega\rangle + (\sqrt{\omega})t\,|-\omega-2\rangle \\ &\quad + [\tfrac{1}{2}\omega(\omega+1)]^{\frac{1}{2}}t^2\,|-\omega-4\rangle + \cdots], \end{aligned} \tag{B7}$$

where $t \equiv \tanh\alpha$. The last expression in (B7) is identical to that in (A23), as indeed it must be. The series expansion in (B7) is valid over the entire real manifold (B4), since it amounts to expansion of the function

$$[1 - 2itzw - t^2(1+z^2)w^2]^{-\frac{1}{2}\omega} \tag{B8}$$

in powers of w; and the singular points $w_\pm = t(iz \pm 1)$ of (B8) lie outside the real manifold (B4), on which $|w|^2 = |y|^{-2} = (z^2+1)^{-1}$.

We introduce an "Euler-angle" parametrization of $SO(2,1)$:

$$\mathfrak{W}(\phi,\alpha,\psi) \equiv \mathfrak{W}(\phi)\mathfrak{W}(\alpha)\mathfrak{W}(\psi) \equiv e^{\varphi\mathcal{L}_{12}}e^{2i\alpha\mathcal{L}_{23}}e^{\psi\mathcal{L}_{12}}, \tag{B9}$$

where the matrix corresponding to $\mathfrak{W}(\alpha)$ has already been written down in (B6). Equations (B1) and (B3) yield

$$\mathfrak{W}(\phi)\begin{bmatrix} y \\ \bar{y} \\ z \end{bmatrix} = \begin{bmatrix} e^{-i\varphi} & 0 & 0 \\ 0 & e^{i\varphi} & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y \\ \bar{y} \\ z \end{bmatrix}, \tag{B10}$$

and a corresponding expression for $\mathfrak{W}(\psi)$. The angles in (B9) must evidently cover the range $0 \le \phi, \psi \le 2\pi$, as can be checked by actually carrying out the matrix multiplication corresponding to (B9).

The corresponding parametrization for $SU(1,1)$ is

$$\begin{aligned} \mathfrak{W}'(\phi,\alpha,\psi) &= \mathfrak{W}'(\phi)\mathfrak{W}'(\alpha)\mathfrak{W}'(\psi) \\ &= e^{\varphi\mathcal{L}'_{12}}e^{2i\alpha\mathcal{L}'_{23}}e^{\psi\mathcal{L}'_{12}} \end{aligned} \tag{B11}$$

with

$$\mathcal{L}'_{12} \equiv \mathcal{K}'/2i \quad \text{and} \quad \mathcal{L}'_{23} \equiv (\mathcal{E}'_+ + \mathcal{E}'_-)/2i, \tag{B12}$$

where $\mathcal{E}'^{\,}_{\pm}$ and $\mathcal{K}'$ are now the operators defined in (A1). As in (A22),

$$\mathfrak{W}'(\alpha)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c & s \\ s & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{B13}$$

and, from (A8),

$$\mathfrak{W}'(\phi)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e^{-\frac{1}{2}i\varphi} & 0 \\ 0 & e^{\frac{1}{2}i\varphi} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}. \tag{B14}$$

Remembering that matrices have to be multiplied in the opposite order from their corresponding differential operators, we find

$$\mathfrak{W}'(\phi,\alpha,\psi)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ce^{-\frac{1}{2}i(\varphi+\psi)} & se^{\frac{1}{2}i(\varphi-\psi)} \\ se^{-\frac{1}{2}i(\varphi-\psi)} & ce^{\frac{1}{2}i(\varphi+\psi)} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}. \tag{B15}$$

On the other hand, the group $SU(1,1)$ is defined by (A15); Fig. 22 indicates some ranges of the angles ϕ and ψ for which (B15) covers each of the matrices (A15) just once.

The range $0 \le \phi, \psi \le 2\pi$ of the parameters for $SO(2,1)$ corresponds to the shaded square in the lower left of Fig. 22. That is, the elements of $SO(2,1)$ can be put into one-to-one correspondence with "half" of the elements of $SU(1,1)$; the other half of $SU(1,1)$ can be obtained from this half by multiplication with the element

$$\mathfrak{I}_1 \equiv \mathfrak{W}'(\phi = 2\pi) = -I \tag{B16}$$

in $SU(1,1)$. That is,

$$SU(1,1) \cong SO(2,1) \times Z_2, \tag{B17}$$

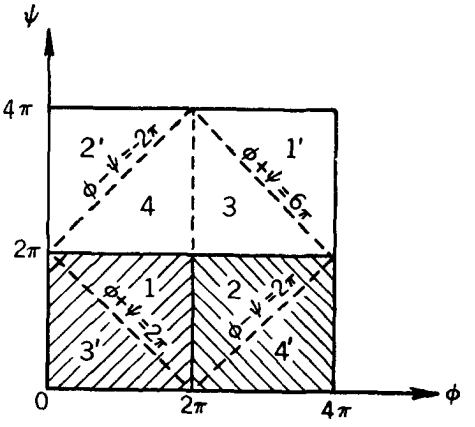


FIG. 22. Ranges of the parameters ϕ and ψ. Triangles 1′, 2′, 3′, and 4′ are equivalent to 1, 2, 3, and 4, respectively. The elements of $SU(1,1)$ are covered just once when ϕ and ψ range over either the dashed square or the two shaded squares.

where

$$Z_2 \equiv \{I, -I\}. \tag{B18}$$

The product in (B17) is only semidirect, since $SO(2, 1)$ is not a normal subgroup of $SU(1, 1)$ (it is not even a subgroup). Alternatively, if $SU(1, 1)$ is partitioned into cosets with respect to Z_2, the two shaded squares of Fig. 22 become equivalent; that is,

$$SU(1, 1)/Z_2 \cong SO(2, 1). \tag{B19}$$

Since $\mathfrak{W}(\phi = 2\pi) = \exp(-\pi i\mathcal{K})$ is the identity in $SO(2, 1)$, we see from Fig. 21 that ω must be an even integer for true representations of $SO(2, 1)$. The restriction is somewhat less stringent for $SU(1, 1)$, where the identity first occurs for $\mathfrak{W}'(\phi = 4\pi) = \exp(-2\pi i\mathcal{K})$; there it is only necessary that ω be an integer.

In the case of nonintegral ω, the basis functions in Fig. 21 are multiple-valued, because of the factor $y^{-\frac{1}{2}\omega}$. Thus the manifold (B4) has to be taken to be a multi-sheeted hyperboloid; transformations will generally carry some points from one sheet to another. This suggests that we introduce the new coordinate space (θ, z), where the old coordinates y and $\bar{y}$ are given by the functions

$$\begin{aligned} y &= +(z^2 + 1)^{\frac{1}{2}} e^{-i\theta}, \\ \bar{y} &= +(z^2 + 1)^{\frac{1}{2}} e^{i\theta}. \end{aligned} \tag{B20}$$

Equation (B10) is then replaced by

$$\mathfrak{W}(\phi)\begin{bmatrix}\theta \\ z\end{bmatrix} = \begin{bmatrix}\theta + \phi \\ z\end{bmatrix}, \tag{B21}$$

and similarly for $\mathfrak{W}(\psi)$. Equations (B6) and (B20) yield the nonlinear relations

$$\mathfrak{W}(\alpha)\begin{bmatrix}\theta \\ z\end{bmatrix} = \begin{bmatrix}\tan^{-1}[(c^2 + s^2)\tan\theta \\ \qquad + 2csz/(z^2 + 1)^{\frac{1}{2}}\cos\theta] \\ (c^2 + s^2)z + 2cs(z^2 + 1)^{\frac{1}{2}}\sin\theta\end{bmatrix}, \tag{B22}$$

where θ must evidently remain on the same branch of the tangent function, since $\mathfrak{W}(\alpha = 0)$ is the identity. At $\alpha = 0$, (B22) yields

$$\begin{aligned} \frac{d\theta}{d\alpha} &= \frac{2z\cos\theta}{(1 + z^2)^{\frac{1}{2}}}, \\ \frac{dz}{d\alpha} &= 2(1 + z^2)^{\frac{1}{2}}\sin\theta, \end{aligned} \tag{B23}$$

which tell us that $\mathfrak{W}(\alpha)$ moves points along the "flow lines" sketched in Fig. 23. The transformation (B21) is, of course, just a translation along the θ axis; and a general "boost," $\mathfrak{W}(-\phi)\mathfrak{W}(\alpha)\mathfrak{W}(\phi)$,

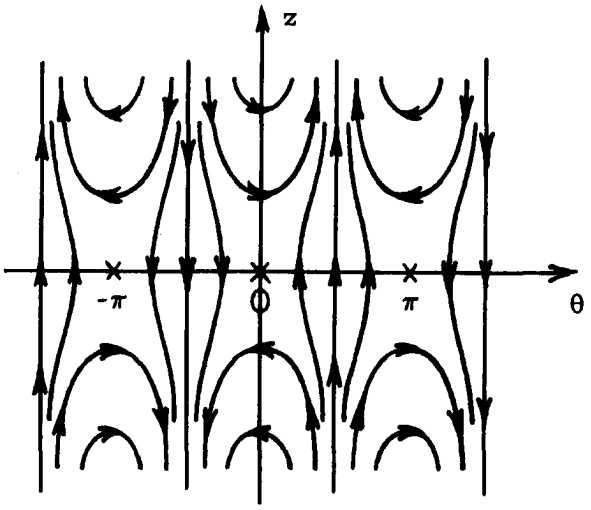


FIG. 23. Qualitative representation of the transformation $\mathfrak{W}(\alpha)$ on the (θ, z) plane.

corresponds to a flow pattern obtained from Fig. 23 by shifting it a distance ϕ along the θ axis.

From this last comment, we see that

$$\mathfrak{I}_n \equiv \mathfrak{W}(\phi = 2\pi n), \quad n \text{ integral}, \tag{B24}$$

commutes with $\mathfrak{W}(\alpha)$. Thus for arbitrary ϕ, α, and ψ we have

$$\mathfrak{W}(\phi)\mathfrak{W}(\alpha)\mathfrak{W}(\psi) = \mathfrak{I}_n\mathfrak{W}(\bar{\phi})\mathfrak{W}(\alpha)\mathfrak{W}(\bar{\psi}) \quad \text{with} \quad 0 \leq \bar{\phi}, \bar{\psi} \leq 2\pi, \tag{B25}$$

for a suitable choice of n. Furthermore, any product of such transformations can also be written in this form. Since the set of all $\mathfrak{I}_n$ forms a group isomorphic to the additive group of integers Z, and since the elements $\mathfrak{W}(\bar{\phi})\mathfrak{W}(\alpha)\mathfrak{W}(\bar{\psi})$ can be put into one-to-one correspondence with the elements of $SO(2, 1)$, this is just the statement that

$$C \cong SO(2, 1) \leftthreetimes Z, \tag{B26}$$

where C is the set of all transformations of the form (B25).

The basis functions of Fig. 21 take the form

$$N_k e^{i(\frac{1}{2}\omega + k)\theta} F(z), \tag{B27}$$

where $F(z)$ is a well-defined function of z, provided we always take the branch of $(z^2 + 1)^{-\frac{1}{2}\omega - \frac{1}{2}k}$ which goes to 1 at $z = 0$. Thus all representations of the form of Fig. 21 are representations of the group C. We can also construct the representations of the algebra which are bounded below or unbounded in terms of similar sets of functions; and these also are representations of C. Thus C is the covering group for the algebra $SO(2, 1)$.

For representations with $\omega = 2n/m$, $\theta + 2\pi m$ is evidently equivalent to θ, from (B27); and so $\mathfrak{I}_m$ is the

identity on these representations. That is, such representations are also true representations of

$$SO(2,1) \supset\!\!\times Z_m, \tag{B28}$$

where Z_m is the additive group of integers modulo m. Representations with irrational ω evidently belong only to C.

Finally, we note that the expressions obtained for matrix elements of $\mathfrak{W}(\alpha)$ in Eqs. (A24), (A28), and (B7) are valid in any case; we are evidently to take the branch of $(\cosh \alpha)^{-\omega}$ which goes to unity at $\alpha = 0$, since $\mathfrak{W}(\alpha = 0)$ is the identity.

* Supported in part by a grant from the National Science Foundation.

[1] E. Gotsman and Y. Ne'eman, in *Symmetry Principles at High Energy III*, A. Perlmutter, J. Wojtaszek, G. Sudarshan and B. Kurşunoğlu, Eds. (W. H. Freeman and Co., San Francisco, 1966), p. 137; C. Fronsdal, *ibid.*,p. 48; Y. Ne'eman, *Algebraic Theory of Particle Physics* (W. A. Benjamin, New York, 1967), Chap. 10; R. Delbourgo and A. Salam, Phys. Rev. **186**, 1516 (1969).

[2] H. Kleinert, Phys. Rev. **163**, 1807 (1967); A. O. Barut, D. Corrigan, and H. Kleinert, Phys. Rev. **167**, 1527 (1968); A. O. Barut, *Lectures in Theoretical Physics*, A. O. Barut and W. E. Brittin, Eds. (Gordon and Breach, Science Publ., New York, 1968), Vol. XB, p. 377; K. C. Tripathy, Phys. Rev. **170**, 1626 (1968); D. Corrigan, Ph.D. thesis, University of Colorado, 1968 (unpublished).

[3] C. Goebel, *Noncompact Groups in Particle Physics*, Y. Chow, Ed. (W. A. Benjamin, Inc., New York, 1966), p. 143; T. Cook and B. Sakita, J. Math. Phys. **8**, 708 (1967).

[4] G. Domokos and G. L. Tindle, Phys. Rev. **165**, 1906 (1968); G. Domokos and S. Kövesi-Domokos, Nuovo Cimento **60A**, 437 (1969).

[5] E. Kyriakopoulos, Phys. Rev. **174**, 1846 (1968).

[6] The above list is certainly incomplete; some references to further work can be found in Refs. 1–5 above.

[7] E. Dynkin, Amer. Math. Soc. Transl., Ser. 2, **6**, 319 (1957); M. Mehta, J. Math. Phys. **7**, 1824 (1966); M. Mehta and P. Srivastava, J. Math. Phys. **7**, 1833 (1966); an exposition of the methods has been given by P. A. Rowlatt, *Group Theory and Elementary Particles* (American Elsevier Publ. Co., Inc., New York, 1966).

[8] It is shown in the following paper [D. Joseph and C. Hieggelke, J. Math. Phys. **11**, 1272 (1970)] that bounded representations exist for $SO(p, q)$, with p and q not both odd, only if p or $q = 0$ or 2.

[9] See Refs. 2, especially the second.

[10] Since $\xi_\pm$ are not independent, the necessity of $E_\pm{}^\dagger = (\pm)E_\mp$ is not immediate. However, $[H, E_\mp{}^\dagger] = -[H, E_\mp]^\dagger = \pm 2E_\mp{}^\dagger$; this implies (Ref. 7) that $E_\mp{}^\dagger$ are proportional to $E_\pm$, which leads to the desired result.

[11] A. O. Barut, *Lectures in Theoretical Physics, Vol. IXA*, W. E. Brittin, A. O. Barut, and M. Guenin, Eds. (Gordon and Breach, Science Publ., New York, 1967), p. 125.

[12] Note that E_α and $N_{a,b}$ are synonymous with $E_{+\alpha}$ and N_{ab}, respectively.

[13] H. Kleinert, Ref. 2, and K. C. Tripathy, Ref. 2.

[14] This method is not new; it is essentially that used by Barut, Ref. 11.

[15] Note that the relative sizes of the N_k are determined by the unitarity conditions (3.32) or (3.33).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Noncompact Algebras with Additively Labeled Multiplets*

D. W. JOSEPH AND C. J. HIEGGELKE
Behlen Laboratory of Physics, University of Nebraska, Lincoln, Nebraska 68508

(Received 27 May 1969)

A large class of noncompact simple Lie algebras is examined to determine those that admit unitary irreducible representations in which the multiplets associated with the compact subalgebra are labeled by the eigenvalues of a generator, with only a finite number of multiplets corresponding to each eigenvalue. It is shown that such representations are precisely those which are bounded above or below, so that the eigenvalues of the labeling generator are bounded above or below. It is found that many of the algebras examined do not have such representations; for example, among the pseudo-orthogonal groups $SO(p, q)$ with p and/or q even, only those of the form $SO(p, 2)$ do admit representations of this kind.

1. INTRODUCTION

In a number of applications of noncompact groups,[1] the multiplets corresponding to the maximum compact subgroup (henceforth called compact multiplets) are labeled by an additive quantum number, which often has a physical significance. In the classical examples of the hydrogen atom and the simple harmonic oscillator, for instance, this is the principal quantum number, specifying the principal part of the energy difference between states. It is also true in these applications that only a finite number of multiplets correspond to each value of this quantum number.

With the expectation that this feature will be of importance in future applications, we examine here a large class of algebras to determine which admit unitary representations with the compact multiplets labeled in this way. As will be shown in Sec. 2, these representations are precisely those which are bounded above or below; this means that the spectrum of the multiplet-labeling quantum number is, in fact, bounded above or below. The algebras to be examined include, among others, those for $SU(p, q)$ and for $SO(p, q)$ with p and q not both odd.

The formalism[2,3] to be used will now be briefly reviewed. The form

$$g = e^{iB} \tag{1.1}$$

will be assumed for group elements, with

$$B = \sum_{\rho} (\xi_{\rho} H_{\rho} + \xi_{+\rho} E_{+\rho} + \xi_{-\rho} E_{-\rho}) + \sum_{\rho,\sigma} (\xi_{+\rho+\sigma}[E_{+\rho}, E_{+\sigma}] + \xi_{-\rho-\sigma}[E_{-\rho}, E_{-\sigma}]) + \cdots, \tag{1.2}$$

where the ξ's are complex numbers, the dots denote terms involving higher-order commutators, and $\rho, \sigma, \cdots$ range over the simple positive roots. The number of simple positive roots is equal to the rank of the algebra; the H's are a maximal set of commuting generators. The following commutation relations will be needed:

$$[E_u, E_v] = N_{uv} E_{u+v}, \text{ with } N_{uv} \neq 0 \text{ if } u + v \text{ is a root,}$$
$$= 0, \quad \text{if } u + v \text{ is not a root and } u + v \neq 0, \tag{1.3a}$$

$$[E_{+\rho}, E_{-\sigma}] = \delta_{\rho\sigma} H_{\rho}, \tag{1.3b}$$

$$[H_{\rho}, E_u] = [2(\rho, u)/(\rho, \rho)] E_u, \tag{1.3c}$$

where ρ, σ denote simple positive roots and u, v denote general roots. The definition of (u, v) and further discussion can be found in Refs. 2 and 3; the above commutation relations correspond, for example, to Eqs. (2.13) of Ref. 3.

The algebras to be considered here are those for which

$$\xi_{\rho}^* = \xi_{\rho}, \quad \xi_{\pm\rho}^* = \epsilon_{\rho} \xi_{\mp\rho},$$
$$\xi_{\pm(\rho+\sigma)}^* = -\epsilon_{\rho}\epsilon_{\sigma} \xi_{\mp(\rho+\sigma)}, \quad \text{etc.,} \tag{1.4}$$

$$\epsilon_{\rho} \equiv +1, \quad \text{for} \quad \rho \neq \rho_0,$$
$$\epsilon_{\rho_0} \equiv -1. \tag{1.5}$$

(These algebras differ from the corresponding compact algebras only by the presence of ϵ_{ρ_0}.) For a unitary representation, Eqs. (1.1), (1.2), (1.4), and (1.5) imply

$$H_{\rho}^{\dagger} = H_{\rho}, \tag{1.6a}$$

$$E_{\pm\rho}^{\dagger} = \epsilon_{\rho} E_{\mp\rho}. \tag{1.6b}$$

If $E_{+\rho} |a\rangle = 0$, Eqs. (1.3b) and (1.6b) yield

$$\omega_{\rho}(a) \equiv \langle a| H_{\rho} |a\rangle = \langle a| E_{+\rho} E_{-\rho} |a\rangle = \epsilon_{\rho} (E_{-\rho} |a\rangle)^{\dagger} E_{-\rho} |a\rangle. \tag{1.7}$$

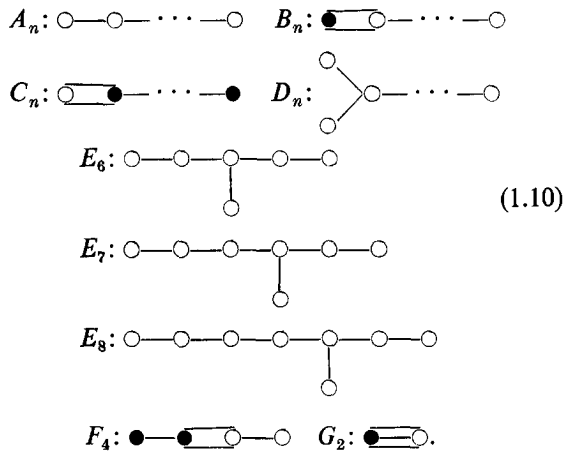
Thus

$$E_{+\rho} |a\rangle = 0 \Rightarrow \omega_{\rho}(a) \geq 0, \text{ for } \rho \neq \rho_0,$$
$$\Rightarrow \omega(a) \equiv \omega_{\rho_0}(a) \leq 0, \text{ for } \rho = \rho_0, \tag{1.8}$$

in a unitary representation; and, similarly,

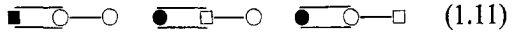
$$E_{-\rho} |a\rangle = 0 \Rightarrow \omega_{\rho}(a) \leq 0 \text{ for } \rho \neq \rho_0,$$
$$\Rightarrow \omega(a) \equiv \omega_{\rho_0}(a) \geq 0 \text{ for } \rho = \rho_0. \tag{1.9}$$

The complex simple Lie algebras are conventionally classified in terms of the Dynkin (or Schouten) diagrams[2]:



(1.10)

Each dot corresponds to a simple root; the type of connection and the coloring determine the Cartan scalar product (ρ, σ) which enters, e.g., in (1.3c). The noncompact algebras we are considering here can be described by replacing one circle by a square to indicate the location of ρ_0. For example, corresponding to B_3 there are the three real noncompact algebras



(1.11)

belonging to $SO(1,6)$, $SO(3,4)$, and $SO(5,2)$, respectively.[3]

The elementary stepping operators $E_{\pm\rho}$, with ρ a simple positive root, connect the basis states of an irreducible representation; and, in general,

$$|b\rangle = \sum \left(\prod_{\pm\rho} E_{\pm\rho} \right) |a\rangle, \tag{1.12}$$

where $|a\rangle$ and $|b\rangle$ are any two states of an irreducible representation, and the symbol $\sum$ without a subscript has been used to denote a linear combination of the terms following. The generators $E_{\pm\rho_0}$ step between different levels of the representation; within each level, the operators $E_{\pm\rho}$ with $\rho \neq \rho_0$ connect basis states belonging to the same compact multiplet (i.e., to the same irreducible representation of the maximum compact subalgebra). In general, there will be more than one compact multiplet to a level.

The commuting operators H_ρ will be chosen to be diagonal, so that basis states are (partially) labeled by their eigenvalues; the set of eigenvalues corresponding to a given state is called the weight of that state. From the commutation relations (1.3c), each application of the raising operator $E_{+\rho}$ (or lowering operator $E_{-\rho}$) changes the weight of a state by a prescribed amount, which will be denoted by $+[\rho]$ (or $-[\rho]$). If $|a\rangle$ and $|b\rangle$ in (1.12) have the definite weights $[a]$ and $[b]$, then

$$[b] = [a] + \sum_{\pm\rho} \pm[\rho], \tag{1.13}$$

where the sum is over the same set of simple roots $\pm\rho$ as the product in any term of (1.12). [Note that a given simple root may occur more than once in the sum of (1.13), and in the products of (1.12).] If after canceling equal positive and negative terms, the sum in (1.13) reduces to a sum over only positive (negative) roots, then $[b]$ will be said to lie above (below) $[a]$, and $|b\rangle$ above (below) $|a\rangle$. A weight will be called maximal (minimal) if no weights lie above (below) it. In general, a representation may contain no maximal or minimal weights, or it may contain several, with several states corresponding to each. However, an irreducible representation of any semisimple Lie algebra[4] contains at most one maximal and one minimal weight; when such highest and lowest weights exist, they are nondegenerate [see statements (I) in Sec. 2]. Since a compact multiplet is an irreducible representation of the compact subalgebra and is finite dimensional, it has a highest weight and corresponding highest state; these completely specify the structure of the multiplet as a representation of the compact subalgebra. If an irreducible representation of the whole algebra has a highest (or lowest) weight and state, it will be said to be bounded above (or below); its structure is then completely specified by this highest (or lowest) weight. Such a semibounded representation[5] has a highest (or lowest) level, where the levels are ordered with respect to the number of times $\pm[\rho_0]$ occurs in (1.13); in fact, the highest (or lowest) level of an irreducible representation contains just one compact multiplet. [This follows from statements (I) in Sec. 2.]

It is always possible to define an operator of the form

$$N \equiv \sum_{\rho} k_\rho H_\rho \tag{1.14}$$

which labels the levels of an irreducible representation. Thus our objective is to determine which algebras admit unitary irreducible representations with only a finite number of compact multiplets at each level. We show in the next section that, for the algebras defined by (1.4) and (1.5), these representations are precisely those that are bounded above or below; and in Sec. 3 we determine which of these algebras can have such representations.

2. EQUIVALENCE OF SEMIBOUNDEDNESS AND HAVING A FINITE NUMBER OF MULTIPLETS AT EACH LEVEL

We first show that a semibounded irreducible representation (i.e., one bounded above or below) can have only a finite number of multiplets at each level. Let $|a\rangle$ and $|b\rangle$ be basis states of an irreducible representation, and express $|b\rangle$ in the form (1.12). From the commutation relation (1.3b), the raising operators $E_{+\rho}$ can be commuted toward the right (without changing their order) at the expense of introducing additional terms containing the operators H_ρ. But the latter operators are diagonal; hence

$$|b\rangle = \sum \left(\prod_\rho E_{-\rho}\right)\left(\prod_{\rho'} E_{+\rho'}\right)|a\rangle, \tag{2.1}$$

where $\sum$ again denotes a linear combination of the terms following. If the representation is bounded above, let $|a\rangle$ be the highest state; then (2.1) reduces to

$$|b\rangle = \sum \left(\prod_\rho E_{-\rho}\right)|a\rangle. \tag{2.2}$$

By (1.3a), any $E_{-\rho_0}$'s which occur in (2.2) can be commuted toward the left with the addition of similar terms involving operators E_{-x}, where x denotes a general positive root containing ρ_0. The latter operators can also be commuted to the left, with the final result

$$|b\rangle = \sum \left(\prod_x E_{-x}\right)\left(\prod_{\rho\neq\rho_0} E_{-\rho}\right)|a\rangle. \tag{2.3}$$

Since there are only a finite number of linearly independent states in a compact multiplet, there is only a finite number of linearly independent states of the form $(\prod_{\rho\neq\rho_0} E_{-\rho})\,|a\rangle$. Furthermore, there is only a finite number of positive roots x corresponding to any given algebra, and the total number of times ρ_0 occurs in the roots x in any term of (2.3) is just equal to the number of levels by which $|b\rangle$ lies lower than $|a\rangle$; hence there is only a finite number of possible factors $\prod_x E_{-x}$. Thus there is only a finite number of linearly independent states (2.3) (or, equivalently, of compact multiplets) to a given level. An analogous proof holds, of course, for an irreducible representation bounded below.

The rest of this section is devoted to proving the converse statement that any irreducible representation with a finite number of compact multiplets (or of states) at each level is necessarily bounded above or below. In fact, it is necessary to know only that one level has a finite number of compact multiplets; this can be seen as follows. Let $|a\rangle$ and $|b\rangle$ belong to levels l and m, respectively, and assume $l > m$ (i.e., l higher than m). Then $|b\rangle$ can be expressed in the form (2.1), where the difference between the number of $E_{-\rho_0}$'s and the number of $E_{+\rho_0}$'s is equal to $l - m$. Now regroup the operators in (2.1) (without changing their order),

$$|b\rangle = \sum \left(\prod_{\rho''} E_{-\rho''}\right)\left(\prod_{\rho'''} E_{-\rho'''}\right)\left(\prod_{\rho'} E_{+\rho'}\right)|a\rangle$$

$$\equiv \sum \left(\prod_{\rho''} E_{-\rho''}\right)|a'\rangle, \tag{2.4}$$

so that $E_{-\rho_0}$ occurs in $\prod_{\rho''} E_{-\rho''}$ the same number of times as $E_{+\rho_0}$ occurs in $\prod_{\rho'} E_{+\rho'}$; that is, so that $|a'\rangle$ belongs to the same level as $|a\rangle$. If we know that level l contains only a finite number of linearly independent states $|a'\rangle$, then an argument similar to that following (2.2) applied to the last expression in (2.4) shows the same to be true of level m. Conversely, if there is only a finite number of independent states in the lower level m, it can be shown in a similar way that there is also only a finite number of independent states in the higher level l.

The following statements will be needed for the main proof of this section:

(Ia) If

$$E_{+\rho}\,|a\rangle = 0 \tag{2.5a}$$

for all simple positive roots ρ, then $|a\rangle$ is the highest state of the irreducible representation to which it belongs; in particular, if $|a\rangle$ belongs to level l, there are no states in levels higher than l.

(Ib) If

$$E_{-\rho}\,|a\rangle = 0 \tag{2.5b}$$

for all simple positive roots ρ, then $|a\rangle$ is the lowest state of the irreducible representation to which it belongs; in particular, if $|a\rangle$ belongs to level l, there are no states in levels lower than l.

Figure 1 gives a diagrammatic representation of these statements. To prove (Ia), we note that any other state $|c\rangle$ of the representation can be expressed in the form (2.1). If $|c\rangle$ did not lie lower than $|a\rangle$, then each of the terms in (2.1) would have to contain

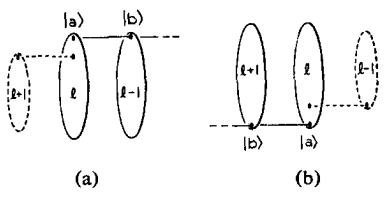


FIG. 1. Diagrammatic representations of level structures for the cases that (a) $E_{+\rho}\,|a\rangle = 0$ for all ρ, and (b) $E_{-\rho}\,|a\rangle = 0$ for all ρ. The ellipses represent levels, and the operators $E_{\pm\rho_0}$ act toward the left and right, respectively. The dashed lines indicate levels which cannot occur by statements (Ia) and (Ib), respectively. Also, $|a\rangle$, $|b\rangle$, $\cdots$ are the highest states of their respective multiplets in case (a) and the lowest states in case (b).

at least one raising operator; hence by (2.5a) it would vanish. The proof of (Ib) is analogous.

Also needed are the following statements concerning irreducible representations with finite-dimensional levels:

(IIa) If the states $|s\rangle \equiv (E_{+\rho_0})^s |a\rangle$ belong to weights which are maximal in their respective levels for $s = 0$, 1, 2, and 3, then they are the highest states of their respective levels for all s.

(IIb) If the states $|s\rangle \equiv (E_{-\rho_0})^s |a\rangle$ belong to weights which are minimal in their respective levels for $s = 0$, 1, 2, and 3, then they are the lowest states of their respective levels for all s.

Note that if either sequence of states terminates at $|s_0\rangle$, then so does the corresponding sequence of levels—that is, there are no levels beyond that containing $|s_0\rangle$, by statements (I). Statements (II) are diagramatically represented in Fig. 2.

To prove (IIa), we note that any other state $|b\rangle$ of the irreducible representation containing $|a\rangle$ can be expressed as in (2.1). Consideration of the weights shows that if $|b\rangle$ did not lie lower in its level than the corresponding state $(E_{+\rho_0})^s |a\rangle$, then each of the terms of (2.1) would take the form[6]

$$\left(\prod_{\rho} E_{-\rho}\right)\left(\prod_{\rho''} E_{+\rho''}\right) E_{+\sigma}(E_{+\rho_0})^r |a\rangle \qquad (2.6)$$

with $\sigma \neq \rho_0$. These terms can be put into more transparent form by use of the identity

$$AB^r = \sum_{m=0}^{r} \frac{r!}{(r-m)!\, m!} B^{r-m}[A, B]_m, \qquad (2.7)$$

where, for example,

$$[A, B]_3 \equiv [[[A, B], B], B], \qquad (2.8)$$

and $[A, B]_0 \equiv A$. The conditions of (IIa) imply that

$$[E_{+\sigma}, E_{+\rho_0}]_m |a\rangle = 0, \quad \text{for} \quad m = 0, 1, 2, 3, \qquad (2.9)$$

where σ is any simple positive root other than ρ_0. Furthermore, it follows from the commutation relations (1.3a) and the rules for the construction of roots [see, e.g., (2.6)–(2.9) of Ref. 3] that

$$[E_{+\sigma}, E_{+\rho_0}]_m = 0 \quad \text{for} \quad m \geq 4. \qquad (2.10)$$

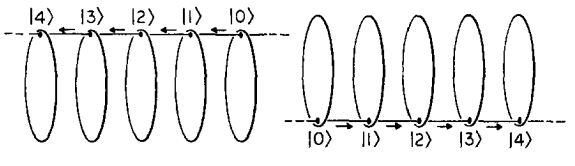


FIG. 2. Diagrammatic representation of statements (II). If $|0\rangle$, $|1\rangle$, $|2\rangle$, and $|3\rangle$ belong to maximal (minimal) weights of their respective levels, then $|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, \cdots$ are the highest (lowest) states of their respective levels until they vanish.

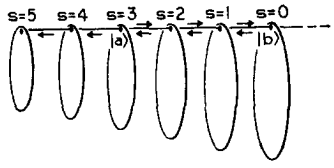


FIG. 3. A typical situation for the case that $(E_{-\rho_0})^3 |a\rangle \neq 0$.

[Actually, the third-order commutator can be nonzero only for G_2, and for A_n even the second-order commutator vanishes; cf. (3.20) below.] From (2.7), (2.9), and (2.10) it follows that

$$E_{+\sigma}(E_{+\rho_0})^r |a\rangle = 0 \qquad (2.11)$$

for all r, and so all the terms (2.6) vanish. A similar proof holds for statement (IIb).

We now proceed to the main proof of this section, the first part of which is summarized in Fig. 3. If a level contains only a finite number of compact multiplets, it must have at least one maximal weight with at least one corresponding state $|a\rangle$. If

$$|b\rangle \equiv (E_{-\rho_0})^3 |a\rangle \qquad (2.12)$$

does not vanish (the case when it does will be considered below), construct the series

$$|s\rangle \equiv (E_{+\rho_0})^s |b\rangle. \qquad (2.13)$$

Note that $|s = 3\rangle$ belongs to the same weight as $|a\rangle$, which is maximal in its level. If $|s = 2\rangle$ did not belong to a maximal weight of its level, there would be some maximal weight and a corresponding state $|2'\rangle$ lying above it. Then,if

$$|3'\rangle \equiv E_{+\rho_0} |2'\rangle \qquad (2.14)$$

vanished, $|2'\rangle$ would satisfy the conditions of (Ia), and the existence of $|a\rangle$ would be contradicted; if $|3'\rangle$ did not vanish, [3'] would be a weight lying above [a], in the same level. Hence $|s = 2\rangle$ belongs to a maximal weight of its level. In the same way, it follows that $|s = 1\rangle$ and $|s = 0\rangle$ also belong to maximal weights of their respective levels. Thus,by (IIa), the states (2.13) are all highest states of their respective levels, and hence of compact multiplets within these levels. However, each application of $E_{+\rho_0}$ lowers the eigenvalues of those H_ρ with $\rho \neq \rho_0$ for which ρ is connected to ρ_0,[7] whereas all these eigenvalues must be nonnegative for the highest state of a compact multiplet [from (1.8)]. Therefore the sequence (2.13) must terminate with some state $|s_0\rangle$, satisfying

$$E_{+\rho} |s_0\rangle = 0 \qquad (2.15)$$

for all ρ; and by (Ia), the representation is bounded above.

If, on the other hand, $(E_{-\rho_0})^3 |a\rangle$ vanishes, let $|b\rangle = (E_{-\rho_0})^t |a\rangle$ be the nonzero state such that

$$E_{-\rho_0} |b\rangle = 0. \tag{2.16}$$

(See Fig. 4. Actually, it always turns out that $|b\rangle = |a\rangle$, but this is not needed for our proof.) As was noted in (1.9), for a unitary representation (2.16) implies that

$$\omega(b) \geq 0, \tag{2.17}$$

where $H_{\rho_0} |b\rangle \equiv \omega(b) |b\rangle$. As above, $|b\rangle$ belongs to a maximal weight of its level, and must therefore be the highest state of some compact multiplet. Let $|c\rangle$ be the lowest state of this same multiplet. If the multiplet is the trivial one, $|c\rangle = |b\rangle$, then

$$E_{-\rho} |b\rangle = 0 \tag{2.18}$$

for all ρ, and by (Ib) the representation is bounded below. If the multiplet containing $|b\rangle$ and $|c\rangle$ is nontrivial, the eigenvalues ω_ρ with $\rho \neq \rho_0$ occurring in $[b]([c])$ must be nonnegative (nonpositive) integers, with at least one not equal to zero. We note that while $E_{-\rho}$ lowers ω_ρ, it raises ω_σ if $\sigma \neq \rho$ but is connected to it. From this it follows that any term in the expression

$$|c\rangle = \sum \left(\prod_{\rho \neq \rho_0} E_{-\rho} \right) |b\rangle \tag{2.19}$$

must contain at least one operator which raises $\omega \equiv \omega_{\rho_0}$; i.e.,

$$\omega(c) > \omega(b). \tag{2.20}$$

Let $[d]$ be the minimal weight of the level which lies below $[c]$ (actually, it can be shown that $[d] = [c]$); then

$$\omega(d) \geq \omega(c). \tag{2.21}$$

From (2.17), (2.20), and (2.21), we have

$$\omega(d) > 0. \tag{2.22}$$

Therefore

$$|e\rangle \equiv (E_{+\rho_0})^3 |d\rangle \neq 0 \tag{2.23}$$

by (1.8) and the fact that each application of $E_{+\rho_0}$ raises ω. It follows from (Ib) that the states

$$|s\rangle \equiv (E_{-\rho_0})^s |e\rangle \tag{2.24}$$

belong to minimal weights of their respective levels for $s = 3$, 2, 1, and 0, by reasoning analogous to the proof that the states (2.13) belong to maximal weights. Therefore, by (IIb), the states (2.24) must be lowest states of their respective levels. But the eigenvalues of H_ρ for $\rho \neq \rho_0$ must be nonpositive for the lowest state of a compact multiplet, while some of these eigenvalues are raised by each application of $E_{-\rho_0}$. Thus the sequence (2.24) must terminate; and so by (Ib), the representation is bounded below.

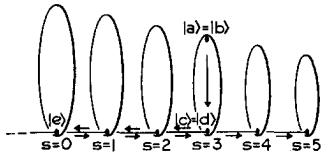


FIG. 4. A typical situation for the case that $E_{-\rho_0}|a\rangle = 0$.

Incidentally, Figs. 3 and 4 depict typical representations bounded above and below, respectively, with compact multiplets increasing in size (and usually also in the number per level) toward the right in the first case and toward the left in the second case. It can be shown from statements (I) and (II) that the sequence of states

$$(E_{-\rho_0})^n |h\rangle \quad \text{or} \quad (E_{+\rho_0})^n |l\rangle \tag{2.25}$$

formed from the highest state $|h\rangle$ or the lowest state $|l\rangle$ are all highest or lowest states of their respective levels.

3. ALGEBRAS WITH SEMIBOUNDED UNITARY REPRESENTATIONS

In this section it is shown that many of the noncompact algebras described in Sec. 1 do not possess unitary representations bounded above or below.

Let $|a\rangle$ belong to a minimal weight of the highest level of a unitary representation bounded above, so that

$$E_{-\rho} |a\rangle = 0, \quad \text{for} \quad \rho \neq \rho_0,$$
$$E_{+\rho_0} |a\rangle = 0. \tag{3.1}$$

If, as before, $\omega_\rho(a)$ is defined by $H_\rho |a\rangle \equiv \omega_\rho(a) |a\rangle$, then

$$\omega_\rho(a) \leq 0, \quad \text{for all } \rho, \tag{3.2}$$

by (1.8) and (1.9). In fact, we can assume that

$$\omega_\sigma(a) < 0 \tag{3.3}$$

for at least one simple positive root σ, since otherwise (1.3b) and (3.1) would yield

$$0 = \langle a| H_\rho |a\rangle = -\langle a| E_{-\rho} E_{+\rho} |a\rangle$$
$$= -(E_{+\rho} |a\rangle)^\dagger E_{+\rho} |a\rangle \tag{3.4}$$

for every $\rho \neq \rho_0$ and a similar equation for $\rho = \rho_0$. Combined with (3.1), these would imply

$$E_{\pm\rho} |a\rangle = 0 \tag{3.5}$$

for all ρ; thus $|a\rangle$ could only be the one state of the trivial one-dimensional representation of the whole algebra.

For many of the algebras described in Sec. 1, there is a further condition on the $\omega_\rho(a)$ which can be derived by consideration of the operators $E_{\pm m}$,

where m (the "maximum" positive root of the given algebra) is such that $m + \rho$ is not a root for any positive simple root ρ. From (1.3a),

$$[E_{-m}, E_{-\rho}] = 0. \tag{3.6}$$

If

$$m = \sigma + \mu + \cdots + \tau \tag{3.7}$$

in terms of simple positive roots $\sigma, \mu, \cdots, \tau$, we make the definitions

$$\begin{aligned} E_{+m} &\equiv [[\cdots [E_{+\sigma}, E_{+\mu}], \cdots], E_{+\tau}], \\ E_{-m} &\equiv [E_{-\tau}, [\cdots, [E_{-\mu}, E_{-\sigma}] \cdots]], \end{aligned} \tag{3.8}$$

in agreement with (1.3a). Note that the commutators defining E_{-m} have been written in the reverse order from those defining E_{+m}, so that (1.6b) yields

$$E_{+m}^{\dagger} = (-1)^{N_m} E_{-m}, \tag{3.9}$$

where N_m is the number of times that ρ_0 occurs in the expression (3.7).

It can be shown that

$$[E_{+m}, E_{-m}] = \sum_{\rho} f_m^{\rho} H_{\rho}, \tag{3.10}$$

with

$$f_m^{\rho} > 0, \quad \text{for all } \rho. \tag{3.11}$$

The general form (3.10) follows from the form of the commutation relations (1.3), by repeated use of the Jacobi identity. One way of demonstrating (3.11) is to note that in a *finite-dimensional* (nonunitary) representation[8,9] of the algebra we can choose

$$\mathcal{E}_{+\rho}^{\dagger} = \mathcal{E}_{-\rho} \tag{3.12}$$

for all ρ, so that

$$\mathcal{E}_{+m}^{\dagger} = \mathcal{E}_{-m}. \tag{3.13}$$

The operators $\mathcal{E}_{\pm m}$ are, of course, defined as in (3.8); and since commutation relations are the same in all representations,

$$[\mathcal{E}_{+m}, \mathcal{E}_{-m}] = \sum_{\rho} f_m^{\rho} \mathcal{H}_{\rho} \tag{3.14}$$

with the same coefficients f_m^{ρ} as in (3.10). Let $|a\rangle$ be the highest state of a nontrivial finite-dimensional representation; then (3.14) and (3.13) yield

$$\begin{aligned} \sum_{\rho} f_m^{\rho} \langle a| \mathcal{H}_{\rho} |a\rangle &= \langle a| \mathcal{E}_{+m} \mathcal{E}_{-m} |a\rangle \\ &= (\mathcal{E}_{-m} |a\rangle)^{\dagger} \mathcal{E}_{-m} |a\rangle > 0. \end{aligned} \tag{3.15}$$

Note that equality is not permitted in this last relation, for that would mean

$$\mathcal{E}_{-m} |a\rangle = 0; \tag{3.16}$$

since any other state $|\ell\rangle$ in this representation can be expressed [cf. (2.2)] as

$$|\ell\rangle = \sum \left(\prod_{\rho} \mathcal{E}_{-\rho} \right) |a\rangle, \tag{3.17}$$

Eqs. (3.6) and (3.16) would require $\mathcal{E}_{-m}$ to be identically zero, in conflict with (1.3a). Equation (3.15) must be true for any finite-dimensional representation; and the only condition on the set of numbers $\langle a| \mathcal{H}_{\rho} |a\rangle$ (the greatest weight of the representation) is that they be nonnegative integers.[9] Thus (3.15) implies conditions (3.11).

Since ρ_0 occurs at least once in m [see, e.g., (3.20) below] and $|a\rangle$ belongs to the highest level of the unitary representation,

$$E_{+m} |a\rangle = 0. \tag{3.18}$$

Therefore (3.10) and (3.9) yield

$$\begin{aligned} (-1)^{N_m} \sum_{\rho} f_m^{\rho} \langle a| H_{\rho} |a\rangle &= (-1)^{N_m} \langle a| E_{+m} E_{-m} |a\rangle \\ &= (E_{-m} |a\rangle)^{\dagger} E_{-m} |a\rangle \geq 0. \end{aligned} \tag{3.19}$$

This is clearly inconsistent with (3.2), (3.3), and (3.11) if N_m is even. That is, if the maximum root m of an algebra contains the noncompact root ρ_0 an even number of times, then the algebra does not possess any unitary representations bounded above; there is, of course, an analogous proof that it does not possess any unitary representations bounded below. (The trivial representation is naturally excluded.)

The positive roots corresponding to a given algebra are easily calculated according to the rules given in Refs. 2 and 3 [e.g., Eqs. (2.6)–(2.9) of Ref. 3]. The maximum positive roots m are found to take the following forms, in terms of simple positive roots $\alpha, \beta, \xi, \eta, \cdots, \tau, \zeta$:

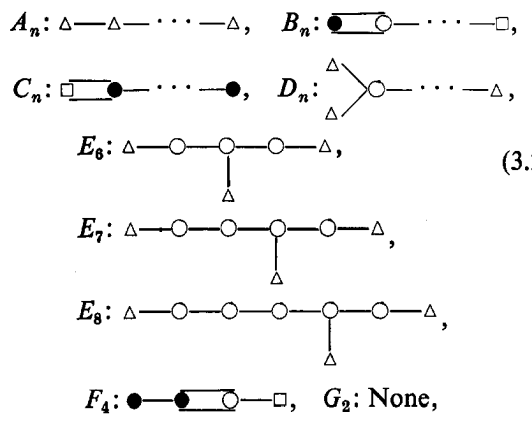
$$\begin{aligned} A_n&: m = \alpha + \beta + \cdots + \tau + \zeta, \\ B_n, F_4&: m = 2\alpha + 2\beta + \cdots + 2\tau + \zeta, \\ C_n&: m = \alpha + 2\beta + \cdots + 2\tau + 2\zeta, \\ D_n, E_6, E_7, E_8&: m = \xi + \eta + 2\beta + \cdots + 2\tau + \zeta, \\ G_2&: m = 3\alpha + 2\zeta, \quad l = 2\alpha + \zeta, \end{aligned} \tag{3.20}$$

where α and ζ denote roots at the left and right ends in (1.10), respectively, while ξ and η denote end roots in general. The additional root l has been included for G_2, since it yields an additional restriction of the same form as does m; we note that direct calculation yields

$$[E_{+l}, E_{-l}] = 8H_{\alpha} + 12H_{\zeta}, \tag{3.21}$$

so that (3.11) is satisfied also by the f_l^{ρ}.

The general result from (3.20) for the noncompact algebras described in Sec. 1 is that semibounded unitary representations can exist only in the following cases:



(3.22)

where the triangles represent possible alternative locations of ρ_0.

4. CONCLUSIONS

We have shown that among the noncompact algebras corresponding to Eqs. (1.4) and (1.5), or to replacing one circle by a square in (1.10), only those indicated in (3.22) can have unitary irreducible representations in which the compact multiplets are labeled by an additive quantum number (i.e., the eigenvalues of a generator) with only a finite number of multiplets corresponding to each eigenvalue. It has also been shown that these are precisely the unitary irreducible representations which are bounded above or below.

The algebras considered include those corresponding to a number of familiar groups.[10] The A_n-type algebras correspond to $SU(p, q)$; all of these can have unitary representations of the above type. The B_n-type algebras correspond to $SO(p, q)$ with $p + q$ odd; our result is that of these only the ones of the form $SO(p, 2)$ can have unitary representations of the above type. The D_n-type algebras (except with ρ_0 at the left ends) correspond to $SO(p, q)$ with p and q both even and $p + q \geq 6$; again, only those of the form $SO(p, 2)$ can have unitary representations of the above type.

ACKNOWLEDGMENTS

We would like to thank Professor C. D. Anderson and the Department of Physics at the California Institute of Technology for their hospitality during the period when this work was completed.

* Supported in part by a grant from the National Science Foundation and in part by a grant from the University of Nebraska Research Council.

[1] See A. O. Barut and H. Kleinert [Phys. Rev. **160**, 1149 (1967)] for a discussion of transition form factors in the hydrogen atom. For applications of $SO(3, 2)$ and $SO(4, 2)$ to particle physics, see A. O. Barut, D. Corrigan, and H. Kleinert, Phys. Rev. **167**, 1527 (1968); K. C. Tripathy, *ibid.* **170**, 1626 (1968); and A. O. Barut, *Lectures in Theoretical Physics, Vol. 10* (Gordon & Breach, Science Publishers, Inc., New York, 1968), Part B, p. 377. E. Kyriakopoulos [Phys. Rev. **174**, 1846 (1968); **177**, 2442 (1969)] discusses dynamical groups, the Bethe-Salpeter equation, and applications to bosons. Further references are to be found in all of the above.

[2] E. B. Dynkin, Am. Math. Soc. Transl., Ser. (2) **6**, 319 (1957); P. A. Rowlatt, *Group Theory and Elementary Particles* (American Elsevier Publ. Co., Inc., New York, 1966).

[3] D. W. Joseph, J. Math. Phys. **11**, 1249 (1970) (preceding article).

[4] The class of semisimple Lie algebras consists of the simple algebras, corresponding to the diagrams of (1.10), and of the algebras corresponding to sums of such diagrams.

[5] These were called bounded representations in Ref. 3.

[6] It is assumed that $|a\rangle$ is an eigenstate of $(E_{-\rho_0}E_{+\rho_0})$ so that any term of the form $(E_{-\rho_0})^t(E_{+\rho_0})^{s+t}|a\rangle = \Sigma\,(E_{+\rho_0})^s(E_{-\rho_0}E_{+\rho_0})^u|a\rangle$ is proportional to $(E_{+\rho_0})^s|a\rangle$. This is necessarily true if $[a]$ is nondegenerate. If $[a]$ were degenerate, the hermiticity of $(E_{-\rho_0}E_{+\rho_0})$ would allow us to find an eigenstate $|a'\rangle$ with weight $[a]$ which satisfied the conditions of (IIa); the conclusion of (IIa) would then show $|a'\rangle$ to be the (unique) highest state of its level, i.e., $|a'\rangle = |a\rangle$.

[7] This follows from the properties of the scalar product (u, v) which occurs in (1.3c) (see Refs. 2 and 3).

[8] As far as the generators are concerned, these are precisely the unitary representations of the corresponding compact algebra.

[9] See Ref. 2 for a discussion of finite-dimensional representations.

[10] See, for example, Secs. 5 and 6 of Ref. 3.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Statistical Average of a Product of Phase Sums Arising in the Study of Disordered Lattice. II

PRIYAMVADA SAH
Department of Physics, Banaras Hindu University, Varanasi, India

(Received 2 June 1969)

This recurrence formula, derived in a previous paper [J. Math. Phys. **10**, 2263 (1969)] to evaluate $\langle S(k_1)S(k_2)\cdots S(k_n)\rangle$, is extended here, for the case when $k_1 + k_2 + \cdots + k_n = 0$ but no other partial sum of the set $k_1, k_2, \cdots, k_n$ is zero. The average is of $O(N)$ as compared to the $O(1)$ when $k_1 + k_2 + \cdots + k_n \neq 0$. Formulas are then proposed for the situations when more than one partial sum of the set vanish. $\langle S(k)S(-k)\rangle$ is considered for a parabolic probability distribution $s(r)$ with a cutoff which shows striking similarity with x-ray diffraction pattern of liquids.

1. INTRODUCTION

A few definitions and equations from a previous paper[1] are collected for the sake of completeness:

$$S(k) = \sum_{j=1}^{N} e^{ikx_j}$$
$$= e^{ikr_1} + e^{ik(r_1+r_2)} + \cdots + e^{ik(r_1+r_2+\cdots+r_N)}, \quad (1)$$

$$S(k_1)S(k_2)\cdots S(k_n) = \sum_{K_1, K_2, \cdots, K_n} e^{i(K_1r_1+K_2r_2+\cdots+K_Nr_N)} \quad (2)$$

[Eqs. (I.1) and (I.2)], where $K_1 = k_1 + k_2 + \cdots + k_n$ and the sum is subjected to condition I.

$$\Xi(N) = S(k_1)S(k_2)\cdots S(k_n). \quad (3)$$

$\langle\Xi\rangle = \lim_{N\to\infty}\langle\Xi(N)\rangle$ was determined under the condition that no partial sum of $k_1, k_2, \cdots, k_n$ is zero. $\Xi(N)$ was expressed as

$$\Xi(N) = \Xi^{(1)}(N) + \Xi^{(2)}(N) + \cdots + \Xi^{(n)}(N). \quad (4)$$

It was shown by the method of induction that

$$\langle\Xi^{(j)}(N)\rangle = P_j(K_1, K_1^{(1)}, \cdots, K_1^{(j-1)}) + Q_j(K_1, K_1^{(1)}, \cdots, K_1^{(j-1)}, N), \quad (5)$$

where

$$P_j = \sum_{K_1, K_1^{(1)}, \cdots, K_1^{(j-1)}} \frac{f(K_1)}{1-f(K_1)} \times \frac{f(K_1^{(1)})}{1-f(K_1^{(1)})} \cdots \frac{f(K_1^{(j-1)})}{1-f(K_1^{(j-1)})} \quad (6)$$

and

$$Q_j = \sum_{K_1, \cdots, K_1^{(j-1)}} [C_0^{(j)} f^N(K_1) + C_1^{(j)} f^N(K_1^{(1)}) + \cdots + C_{j-1}^{(j)} f^N(K_1^{(j-1)})]. \quad (7)$$

$C_0^{(j)}, C_1^{(j)}, \cdots, C_{j-1}^{(j)}$ depend on $f(K_1), f(K_1^{(1)}), \cdots, f(K_1^{(j-1)})$ but do not depend on N. Thus Q_j involves the Nth powers of $f(K)$ which would tend to zero as $N \to \infty$ because $|f(K)| < 1$, for $K \neq 0$. In Sec. 2, we consider the Q_j in detail because the derivation of the formula for $\langle S(k_1)S(k_2)\cdots S(k_n)\rangle$ when $k_1 + k_2 + \cdots + k_n = 0$ is found to depend on the result for $\langle\Xi(N-1)\rangle, \langle\Xi(N-2)\rangle, \cdots, \langle\Xi(n-1)\rangle$; therefore, it is not enough to know just the limit of $\langle\Xi(N)\rangle$ as $N \to \infty$.

2. A RECURRENCE RELATION

$$\Xi(N) = \sum_{K_1, K_2, \cdots, K_n} e^{i(K_1r_1+K_2r_2+\cdots+K_Nr_N)}$$

as in I, but $K_1 = k_1 + k_2 + \cdots + k_n = 0$ in this paper. $K_2, K_3, \cdots, K_N$ are subjected to the condition I. The sum $\Xi(N)$ is here written as

$$\Xi(N) = \Xi_1(N) + \Xi_2(N) + \cdots + \Xi_N(N), \quad (8)$$

where

$$\Xi_j(N) = \Xi_j^{(1)}(N) + \Xi_j^{(2)}(N) + \cdots + \Xi_j^{(n)}(N), \quad \text{when } j \leq N-n+1,$$
$$= \Xi_j^{(1)}(N) + \Xi_j^{(2)}(N) + \cdots + \Xi_j^{N+1-j}(N), \quad \text{when } N \geq j > N-n+1. \quad (9)$$

$\Xi^{(\nu)}(N)$ is the sum of those terms of $\Xi(N)$ which have ν distinct K_n and $\Xi_j^{(\nu)}(N)$ is that subsum of $\Xi^{(\nu)}(N)$ in which the coefficient of $r_1, r_2, \cdots, r_j$ is K_1. In view of this, Eq. (9) defines $\Xi_j(N)$ as that subsum of $\Xi(N)$ in which every term has $K_1 = K_2 = \cdots = K_j$. The subdivision of $\Xi(N)$ in $\Xi_j(N)$, $j = 1, 2, \cdots, N$, is preferred to its subdivision in $\Xi^{(\nu)}(N)$, $\nu = 1, 2, \cdots, n$, as in Eq. (4), because $\Xi_j(N)$ has a factor $e^{iK_1(r_1+r_2+\cdots+r_j)}$ which is equal to unity and the second factor

$$\sum_{K_{j+1}, \cdots, K_N} e^{i(K_{j+1}r_{j+1}+\cdots+K_Nr_N)}$$

can be evaluated by use of the following equation [Eqs. (I.18), (I.8), and (I.10)]:

$$\begin{aligned}\langle \Xi(N)\rangle &= \sum_{K_1}\frac{f(K_1)}{1-f(K_1)} + \sum_{K_1,K_1^{(1)}}\frac{f(K_1)}{1-f(K_1)} \\ &\times \frac{f(K_1^{(1)})}{1-f(K_1^{(1)})} + \cdots + \sum_{K_1,\cdots,K_1^{(n-1)}}\frac{f(K_1)}{1-f(K_1)} \\ &\times \frac{f(K_1^{(1)})}{1-f(K_1^{(1)})}\cdots\frac{f(K_1^{(n-1)})}{1-f(K_1^{(n-1)})} \\ &+ \sum_{K_1} C_0^{(1)} f^N(K_1) \\ &+ \sum_{K_1,K_1^{(1)}}[C_0^{(2)} f^N(K_1) + C_1^{(1)} f^{(N)}(K_1^{(1)})] \\ &+ \cdots + \sum_{K_1,K_1^{(1)},\cdots,K_1^{(n-1)}}[C_0^{(n)} f^N(K_1) \\ &+ \cdots + C_{n-1}^{(n)} f^N(K_1^{(n-1)})]. \end{aligned} \tag{10}$$

Consider the subsums

$$\begin{aligned}&\Xi_j(N) \\ &= e^{iK_1(r_1+r_2+\cdots+r_j)} \sum_{K_{j+1},\cdots,K_N} e^{i(K_{j+1}r_{j+1}+\cdots+K_Nr_N)}, \\ &\qquad j = 1, 2, \cdots, N-1, \\ &\Xi_N(N) = e^{iK_1(r_1+r_2+\cdots+r_N)} = 1, \end{aligned} \tag{11}$$

where $K_{j+1} \neq K_1$ (the sum of the full set) but can be the sum of any subset including the empty set.

Considering again Ξ_j, $j \leq N-1$, one can take out the single term, in which K_{j+1} is the empty set, and write

$$\begin{aligned}\Xi_j(N) &= 1 + \sum_{K_{j+1},\cdots,K_N} e^{i(K_{j+1}r_{j+1}+\cdots+K_Nr_N)}, \\ &\qquad j = 1, 2, \cdots, (N-1). \end{aligned} \tag{12}$$

In the remaining sum all terms have a nonvanishing phase because $K_{j+1} \neq K_1$ and $K_{j+1} \neq 0$.

Hence, by Eqs. (8) and (12),

$$\begin{aligned}\Xi(N) &= \sum_{j=1}^{N}\Xi_j(N) \\ &= N + \sum_{j=1}^{N-1}\sum_{K_{j+1},\cdots,K_N} e^{i(K_{j+1}r_{j+1}+\cdots+K_Nr_N)}. \end{aligned} \tag{13}$$

Note that K_{j+1} can be equal to the sum of any of $2^n - 2$ subsets $(= {}^nc_1 + {}^nc_2 + \cdots + {}^nc_{n-1})$ of $k_1, k_2, \cdots, k_n$, which is the same for all j from 1 to $(N-1)$. The summation over j is now performed and the subscripts in the exponents are redefined. The labels attached to the variables r_i are immaterial (until two different r's are replaced by two different ones), and the condition I indicates that, in any sum

$$\sum_{K_{j+1},\cdots,K_N} e^{i(K_{j+1}r_{j+1}+\cdots+K_Nr_N)},$$

the permissible values of $K_{j+2}, \cdots, K_N$ are only dependent on K_{j+1}. Therefore, Eq. (13) can be written as

$$\begin{aligned}\Xi(N) = N &+ \sum_{K_2,\cdots,K_N} e^{i(K_2r_2+\cdots+K_Nr_N)} \\ &+ \sum_{K_2,\cdots,K_{N-1}} e^{i(K_2r_2+\cdots+K_{N-1}r_{N-1})} \\ &+ \cdots + \sum_{K_2} e^{iK_2r_2}. \end{aligned} \tag{14}$$

The second, third, $\cdots$, $(N-n+2)$th terms of Eq. (14) are averaged by the application of (10). Since the set of possible values of K_2 is identical with the set of values of $K_1^{(1)}$ in Paper I (because $K_2 \neq K_1$, $K_2 \neq 0$) and is the sum of a subset of $k_1, k_2, \cdots, k_n$, K_2 can be replaced by $K_1^{(1)}$. We then have

$$\begin{aligned}&\langle \Xi(N)\rangle \\ &= N + (N-n+1)\Bigg(\sum_{K_1^{(1)}}\frac{f(K_1^{(1)})}{1-f(K_1^{(1)})} + \cdots \\ &\quad + \sum_{K_1^{(1)},\cdots,K_1^{(n-1)}}\frac{f(K_1^{(1)})}{1-f(K_1^{(1)})}\cdots\frac{f(K_1^{(n-1)})}{1-f(K_1^{(n-1)})}\Bigg) \\ &\quad + \Bigg(\Bigg\langle \sum_{K_2,\cdots,K_{n-1}} e^{i(K_2r_2+\cdots+K_{n-1}r_{n-1})}\Bigg\rangle \\ &\quad + \cdots + \Bigg\langle \sum_{K_2} e^{iK_2r_2}\Bigg\rangle\Bigg) \\ &\quad + \sum_{\nu=n-1}^{N-1}\Bigg\{\sum_{K_1^{(1)}} C_0^{(1)} f^\nu(K_1^{(1)}) \\ &\quad + \sum_{K_1^{(1)},K_1^{(2)}}[C_0^{(2)} f^\nu(K_1^{(1)}) + C_1^{(2)} f^\nu(K_1^{(2)})] \\ &\quad + \cdots + \sum_{K_1^{(1)},\cdots,K_1^{(n-1)}}[C_0^{(n-1)} f^\nu(K_1^{(1)}) \\ &\quad + \cdots + C_{n-2}^{(n-1)} f^\nu(K_1^{(n-1)})]\Bigg\}. \end{aligned} \tag{15}$$

The values $K_1^{(1)}$ takes in the first sum are different from those it takes in the second sum. $K_1^{(1)}$ in the first sum is any proper subsum of $k_1, k_2, \cdots, k_n$, whereas $K_1^{(1)}$ in the second sum is equal to only those proper subsums which have more than one k_i in them. In the last term the sum over ν leads to geometric series, the common ratios being $f(K_1^{(1)})$, $f(K_1^{(2)}), \cdots, f(K_1^{(n-1)})$, the modulus of which are all less than unity. The sum of each geometric series is, therefore, finite. The average of the last $(n-2)$ terms of Eq. (14) is also finite. Since in the limit of large N only the terms of order N need be retained,

$$\begin{aligned}\langle \Xi\rangle &= \langle S(k_1)S(k_2)\cdots S(k_n)\rangle \\ &= N\Bigg(1 + \sum_{K_1^{(1)}}\frac{f(K^{(1)})}{1-f(K_1^{(1)})} + \cdots \\ &\quad + \sum_{K_1^{(1)},\cdots,K_1^{(n-1)}}\frac{f(K_1^{(1)})}{1-f(K_1^{(1)})}\cdots\frac{f(K_1^{(n-1)})}{1-f(K_1^{(n-1)})}\Bigg). \end{aligned} \tag{16}$$

This is of $O(N)$. We showed in I that $\langle \Xi\rangle$ is of $O(1)$ if $k_1 + k_2 + \cdots + k_n \neq 0$. Thus, there is a change of

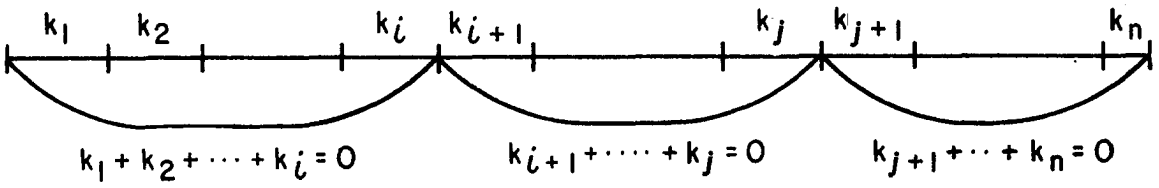


Diagram 1.

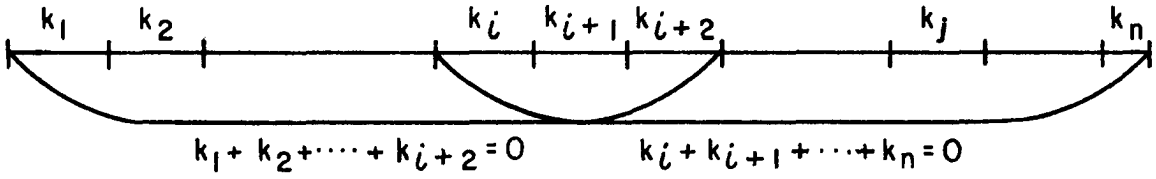


Diagram 2.

order from 1 to N as $k_1 + k_2 + \cdots + k_n$ takes a value of zero instead of one different from zero.

The δ-singularity of the operators (in the plane-wave representation) $(H_0 - zI)^{-1}U(H_0 - zI)^{-1}U \cdots U(H_0 - zI)^{-1}$ arises from this change of order. The diagonal elements correspond to $k_1 + k_2 + \cdots + k_n = 0$ and the nondiagonal elements to

$$k_1 + k_2 + \cdots + k_n \neq 0.$$

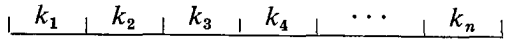
On regrouping the terms of Eq. (16) in the same was as the terms of Eq. (I.20) were regrouped to give the recurrence formula of Eq. (I.21), we obtain

$$\langle S(k_1)S(k_2)\cdots S(k_n)\rangle = N\Big(1 + \sum_{i=1}^{n}\langle S(k_i)\rangle + \cdots + \sum_{i,j,\cdots,l=1}^{n}{}'\langle S(k_i)S(k_j)\cdots S(k_l)\rangle\Big). \quad (17)$$

Since the lower products $\langle S(k_i)\cdots S(k_l)\rangle$ are such that $k_i + \cdots + k_l \neq 0$ and there is no subsum of $k_i, \cdots, k_l$ which is zero, the conditions of (I.21) are satisfied and can be evaluated by its application.

3. CASE OF SEVERAL VANISHING PARTIAL SUMS

We now generalize the recurrence formula of Eq. (17) when more than one partial sum vanishes. In order to distinguish between two possibilities that may arise, we proceed as follows: Let us consider a segment of the real axis and allot unit lengths to each of $k_1, k_2, \cdots, k_n$:



With every partial sum of $k_1, k_2, \cdots, k_n$ we shall associate a union of intervals. One possibility is that the unions for vanishing partial sums have a null intersection and the other is that they have a non-null intersection, as Diagrams 1 and 2 illustrate.

The main difference between Eqs. (I.21) and (17) is the replacement of the factor $f(K_1)/[1 - f(K_1)]$ by N. We conjecture that when a partial sum vanishes, a factor N is introduced in place of $f(K)/[1 - f(K)]$, where K stands for that partial sum. This arises from the fact that in the geometric series

$$r + r^2 + \cdots + r^N = \frac{r - r^{N+1}}{1 - r} \simeq \frac{r}{1 - r}$$

when $|r| < 1$ and is equal to N if $r = 1$. Suppose we write the expression for $\langle S(k_1)\cdots S(k_n)\rangle$ when $k_1, k_2, \cdots, k_n$ satisfy the conditions of Diagram 1, ignoring the fact that the expression is derived under the assumption that no other partial sum vanishes. Among the terms within the bracket of the rhs of Eq. (17) will occur a term which contains j S's.

$$\langle S(k_1)S(k_2)\cdots S(k_j)\rangle$$

has a factor

$$f(k_1 + k_2 + \cdots + k_j)/[1 - f(k_1 + k_2 + \cdots + k_j)]$$

which because of $k_1 + \cdots + k_j = 0$ should be replaced by N. Further, these are terms involving the product of this with

$$f(k_1 + k_2 + \cdots + k_i)/[1 - f(k_1 + k_2 + \cdots + k_i)]$$

which on account of $k_1 + k_2 + \cdots + k_i = 0$ introduces another N, giving an average of $O(N^3)$. There are two other sums of $O(N^3)$ arising from the expressions having $n - i$, $i + n - j$ S's. The rest of the terms are of smaller order and can be neglected. We now state without giving a proof that, if the j vanishing partial sums belong to unions of intervals which have a null intersection, the average is $O(N^j)$.

To arrive at a result when the vanishing partial sums belong to unions which have a nonnull intersection, note that the factors

$$\frac{f(k_1 + k_2 + \cdots + k_{i+1})}{1 - f(k_1 + k_2 + \cdots + k_{i+1})}$$

and

$$\frac{f(k_{i-1} + k_i + k_{i+1} + \cdots + k_j)}{1 - f(k_{i-1} + k_i + k_{i+1} + \cdots + k_j)}$$

never occur as a product in $\langle S(k_1)S(k_2)\cdots S(k_n)\rangle$ because the factors that multiply are always such that the set of k's in the partial sum of one contains or is contained in the set of k's occurring in the partial sum of the other. Therefore, $k_1, k_2, \cdots, k_n$ of Diagram 2 will give terms of $O(N^2)$ and not of $O(N^3)$. This should hold generally.

The formulas we have established can handle fully the problem of averaging of the perturbation terms, i.e., determining the average energy spectrum. When these ideas are applied to discuss the perturbation series elsewhere, we shall see that the case of null intersection is of greater interest than the case of nonnull intersection.

4. THE AVERAGE OF TWOFOLD PRODUCTS

The behavior of twofold products $\langle S(k_1)S(k_2)\rangle$ when $k_1 + k_2 = 0$ for a parabolic probability distribution $s(r)$ with a cutoff at a reasonable distance is studied. According to Eq. (17),

$$\langle S(k)S(-k)\rangle = N\left(1 + \sum_{i=1}^{2}\langle S(k_i)\rangle\right)$$
$$= N\left(1 + \frac{f(k)}{1 - f(k)} + \frac{f(-k)}{1 - f(-k)}\right),$$
$$f(-k) = f^*(k) \quad \text{from (I.9).} \quad (18)$$

Thus,

$$\langle S(k)S(-k)\rangle = N\left[1 + 2\mathrm{R}_e\frac{f(k)}{1 - f(k)}\right].$$

If

$$s(r) = -\frac{3}{4d^3}[(r - r_0)^2 - d^2],$$
$$\text{when} \quad -d \le r - r_0 \le d,$$
$$= 0, \quad \text{otherwise,} \quad (19)$$

then

$$f(k) = \int_{r_0-d}^{r_0+d} e^{ikr}s(r)\,dr$$
$$= \frac{3}{kd}j_1(kd)e^{ikr_0}, \quad (20)$$

$$\frac{1}{N}\langle S(kr_0)S(-kr_0)\rangle$$
$$= \frac{(kd)^2 - 9j_1^2(kd)}{(kd)^2 - 6(kd)\,j_1(kd)\cos kr_0 + 9j_1^2(kd)}.$$

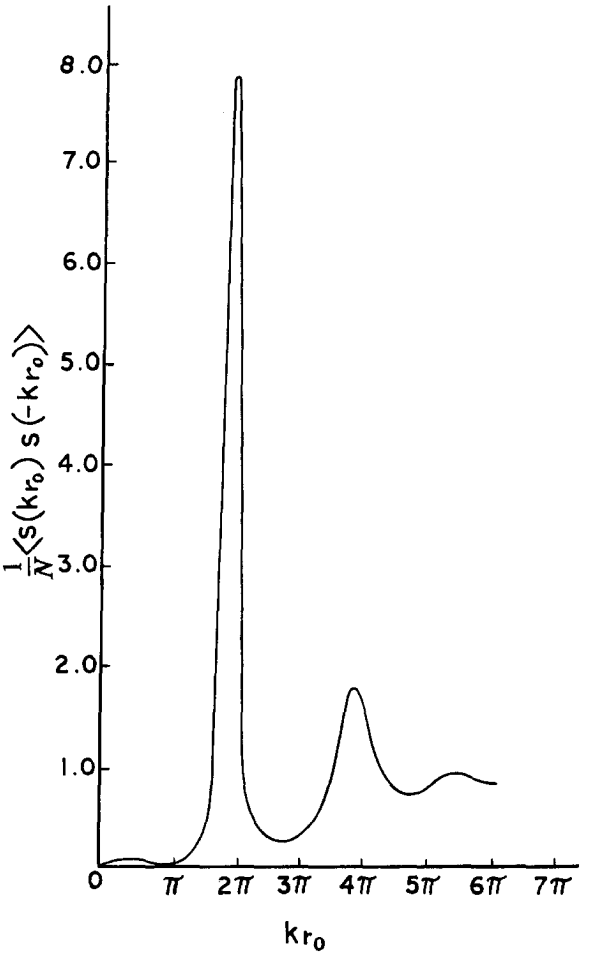


FIG. 1.

As $kr_0 \to \infty$, this function tends to 1; as $kr_0 \to 0$, it tends to $\frac{1}{5}(d/r_0)^2$. The behavior of this function for $r_0 = 4d$ is shown in Fig. 1. This is finite for all kr_0. This figure shows great similarity with the x-ray diffraction pattern of liquids, the characteristic of which is a strong maximum at $kr_0 = 2\pi$ with one or two subsidiary maxima.

A study of the averages of higher products and the connection of this work with the evaluation of perturbation series for energy of one-electron levels will be published elsewhere.

ACKNOWLEDGMENT

Thanks are due to the University Grants Commission for the award of a Senior Research fellowship and to I. I. T. Kanpur for extending library facilities.

[1] P. Sah, J. Math. Phys. **10**, 2263 (1969), to be referred to as I; equations therein will be denoted by (I.1), (I.2), etc.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Nonlinear Stochastic Differential Equations Containing Random Parameters with Small and Large Correlation Time*

N. G. F. SANCHO
Department of Mathematics, McGill University, Montreal 110, Quebec

(Received 3 November 1969)

This paper is an extension of a previous one [J. Math. Phys. **11**, 762 (1970)], and describes a technique for finding the moment equations of a nonlinear stochastic system containing random parameters with small and large correlation time. The nonlinear term can then be linearized by various methods. Finally, the moment inequalities are preserved using the technique described by Bellman and Richardson.

I. INTRODUCTION

In a recent paper, Sancho[1] described a technique for finding the moment equations of a nonlinear stochastic system of the form

$$\ddot{x}(t) + \alpha\dot{x}(t) + \psi(x(t)) = f(t), \tag{1}$$

where α is a constant, $\psi(x(t))$ is an arbitrary "smooth" function of $x(t)$, and f is a random forcing function which is not Gaussian white noise. Such a system was also considered by Morton and Corrsin,[2] who gave the assumptions needed to derive the Fokker–Planck equation for the random forcing function, and the experimental confirmation of the equation is shown for the moment equation in the steady state.

The purpose of this paper is to extend the results described by Sancho[1] for finding the moment equations of a nonlinear stochastic system containing random parameters with small and large correlation time. After the moment equations are found, the nonlinear terms can then be linearized by various "bootstrap" or "self-consistent" methods,[3,4] or even by a dishonest approach.[5,6] Finally the moment properties are preserved, e.g., $\langle x^2\rangle \geq \langle x\rangle^2$ using the technique of Bellman and Richardson.[7] The point, however, is that no matter how crude an approximation we take to linearize the nonlinear term in the moment equations, the technique of Bellman and Richardson[7] ensures the moment-preserving properties.

II. STATISTICAL ANALYSIS

A. Some Statistical Analysis

Let us consider now a simple nonlinear system of the form

$$\dot{x}(t) + b_0x^2(t) + b_1x(t) + n(t)x(t) = 0, \tag{2}$$

where b_0 and b_1 are known constants and $n(t)$ a random noise term that is Gaussian with small correlation time and zero mean. We wish to derive $\langle \Delta x \mid x\rangle$ and $\langle (\Delta x)^2 \mid x\rangle$, where Δ is a finite forward-increment operator over the time increment Δt and $\langle \cdots \mid x\rangle$ is the conditional expectation given x.

We assume that $n(t)$ has the same statistical properties for the random forcing function $f(t)$ of the system (1) described by Morton and Corrsin.[2] The results are summarized as follows:

(1) $n(t)$ is a stationary Gaussian random variable with zero average value and noninfinite integral scale;

(2) the statistical properties of $n(t)$ are independent of the system response;

(3) the largest statistically characteristic time of $n(t)$, $T_{\max}$, say, must be so much smaller than the smallest characteristic time of $x(t)$, δ_x, say, that there can exist a time θ which is very much larger than the former and very much smaller than the latter, i.e.,

$$\delta_x \gg \theta \gg T_{\max}.$$

In finding $\langle \Delta x \mid x\rangle$ by conventional means, we may have some form of stochastic integral, which may lead us to certain difficulty. The stochastic integral as defined by Ito (see Ref. 8) is defined by the limit

$$\int_a^b f(s)\,dz(s) = \lim \sum_{k=0}^{n-1} f(t_k)[z(t_{k+1}) - z(t_k)], \tag{3}$$

where z is the random term, $a = t_0 < t_1 < t_2 < t_3 < \cdots < t_{n-1} < t_n = b$, and $\tau = \max(t_{i+1} - t_i)$. Such a definition is quite consistent mathematically, for if $z(t)$ were a well-behaved function it would give the usual notion of an integral.

However, Doob[8] has shown that formal integrals differ quite often from stochastic integrals. For example, we have

$$\int_a^b [z(t) - z(a)]\,dz(t) = \tfrac{1}{2}[z(b) - z(a)]^2 - D(b - a), \tag{4}$$

where $(dz)^2 = 2D\,dt$. Hence, one can show that

$$\int F'(z)\,dz \neq F(z). \tag{5}$$

In Eq. (2) the $\dot{x}(t)$ term is often a mathematical approximation for a Gaussian process with a short correlation time. Hence, $x(t)$ is, therefore, not so obscure as to require a different form of integration. The engineer and physicist (see Ref. 9) would find it far more convenient to use a formal rather than a stochastic integral in the interpretation of Eq. (2).

It would be possible, on the other hand, to define an integral[10] in a more symmetric way than was done in (3). This gives a "symmetric" stochastic integral as the limit

$$\int_a^b f(s)\,dz(s) = \lim_{\tau\to 0}\sum_{k=0}^{n-1}\tfrac{1}{2}[f(t_{k+1})+f(t_k)][z(t_{k+1})-z(t_k)]. \quad (6)$$

Here t_k and τ are as in Eq. (3). Stratonovich's definition is also used by Wong and Zakai,[11] who also show that such a definition leads to an integral that is symmetric in time and identical to the formal integral. Therefore, differentials would retain their usual form

$$dF(z) = F'(z), \quad (7)$$

and integration would not yield the odd results. The mathematican, therefore, uses the stochastic integral when $n(t)$ is strictly white noise, but physicists use a formal integral since $n(t)$ is an approximation for a Gaussian process with a very small correlation time. We retain our identity of a formal integration by integrating Eq. (2) to obtain

$$\Delta x = x(t+\Delta t) - x(t) \quad (8)$$

$$= x(t)\left[\exp\left(-b_0x(t)\Delta t - b_1\Delta t - \int_t^{t+\Delta t} n(t)\,dt\right) - 1\right] \quad (9)$$

$$= -b_0x^2\Delta t - b_1x\Delta t - x\int_t^{t+\Delta t} n(t)\,dt + \tfrac{1}{2}x\left(\int_t^{t+\Delta t} n(t)\,dt\right)^2 + O(\Delta t)^2. \quad (10)$$

Therefore,

$$\langle \Delta x \mid x\rangle = -b_0x^2\Delta t - b_1x\Delta t + \tfrac{1}{2}x\left(\int_t^{t+\Delta t} n(t)\,dt\right)^2 + O(\Delta t)^2, \quad (11)$$

and

$$\langle (\Delta x)^2 \mid x\rangle = x^2\left[\int_t^{t+\Delta t} n(t)\,dt\right]^2 + O(\Delta t)^2, \quad (12)$$

since $n(t)$ has zero mean, i.e.,

$$\left\langle \int_t^{t+\Delta t} n(t)\,dt \right\rangle = \int_t^{t+\Delta t}\langle n\rangle\,dt = 0,$$

and its statistical properties are independent of system $x(t)$. It should be pointed out that if we had used Ito's definition of the stochastic integral, we would have obtained, for (11),

$$\langle \Delta x \mid x\rangle = -b_0x^2\Delta t - b_1x\Delta t + O(\Delta t)^2, \quad (13)$$

and Eq. (12) would have been the same since $n(t)$ has zero mean.

Let us now put

$$\Lambda = \int_t^{t+\Delta t}\int_t^{t+\Delta t}\langle n(t_1)n(t_2)\rangle\,dt_1\,dt_2, \quad (14)$$

and introduce the autocorrelation function

$$R_n(\tau) \equiv \langle n(t)n(t+\tau)\rangle/\langle n^2\rangle. \quad (15)$$

Equation (14) can be written as

$$\Lambda(\Delta t) = \langle n^2\rangle\int_0^{\Delta t}\int_0^{\Delta t} R_n(t_1-t_2)\,dt_1\,dt_2, \quad (16)$$

provided that $n(t)$ satisfies the conditions given earlier.

Let us now note that, as Δt becomes smaller, then

$$\Lambda \to \langle n^2\rangle(\Delta t)^2, \quad (17)$$

and let us put

$$D \equiv \Lambda(\theta)\theta^{-1} \simeq \lim_{\Delta t\to\infty}[\Lambda(\Delta t)/\Delta t].$$

Equations (11) and (12) can now be written

$$\langle \Delta x \mid x\rangle = -b_0x^2\Delta t - b_1x\Delta t + \tfrac{1}{2}Dx\Delta t + O(\Delta t)^2 \quad (18)$$

and

$$\langle (\Delta x)^2 \mid x\rangle = Dx^2\Delta t + O(\Delta t)^2. \quad (19)$$

B. Further Statistical Analysis

We consider again the system given by Eq. (2), but this time $n(t)$ satisfies the first two conditions of Morton and Corrsin[2] as given in Sec. II.A but does not satisfy the third, i.e., it has a large correlation time. The random noise term $n(t)$ is considered to be continuous, Gaussian, and Markovian; it is therefore governed by a stochastic differential equation of the form

$$dn(t) + an(t)\,dt = dw(t), \quad (20)$$

where a is a known constant and $w(t)$ is a Wiener process with the following incremental properties:

$$\langle dw(t)\rangle = 0,$$
$$\langle dw(t)dw(t)\rangle = 2C\,dt. \quad (21)$$

We now regard $n(t)$ as a state variable in Eq. (2). However, the dimension of (2) is increased by one, with a new type of nonlinearity, namely $n(t)x(t)$. In

any case, one type of nonlinearity is as bad as another, so we have little to lose if $n(t)$ is taken as a new state variable. However, if Eq. (2) were linear, then it would have now become nonlinear by the introduction of $n(t)$ as a new state variable. We, therefore, have

$$\langle \Delta x \mid x, n\rangle = -b_0x^2\Delta t - b_1x\Delta t - nx\Delta t + O(\Delta t)^2, \tag{22}$$

$$\langle (\Delta x)^2 \mid x, n\rangle = (b_0^2x^4 + b_1^2x^2 + n^2x^2)(\Delta t)^2 + O(\Delta t)^2, \tag{23}$$

$$\langle \Delta n \mid x, n\rangle = -an\Delta t + O(\Delta t)^2, \tag{24}$$

and

$$\langle (\Delta n)^2 \mid x, n\rangle = 2C\Delta t + a^2n^2(\Delta t)^2 + O(\Delta t)^2. \tag{25}$$

Equations (22)–(25) and also (18) and (19) are needed in using the Fokker–Planck equation for the probability density function. It also is used in the next section to find the differential equations governing the moments. It should be pointed out that the same results are obtained for (22)–(25); no matter what type of definition, Ito[8] or Stratonovich,[10] for stochastic integral is used.

It is noted that systems described by differential equations involving δ-function-correlated random parameters have been discussed by Leibowitz.[12]

III. MOMENT EQUATIONS

A. Moment Equations of Stochastic System

In order to find the moment equations of (2), where $n(t)$ is a random noise term with a small correlation time, we follow the same procedure given by Sancho[1] and also by Cumming.[13] We consider now $B(x; t)$, an arbitrary function of $x(t)$, whose partial derivatives B_x and B_{xx} are continuous and bounded on an interval of $x(t)$. We derive a differential equation for the expected value $\langle B\rangle$ of B. Using Taylor's series, we have

$$\Delta B = B_x\Delta x + \tfrac{1}{2}B_{xx}(\Delta x)^2 + O(\Delta x)^2. \tag{26}$$

Taking the conditional expectation of (26), given x, and using (18) and (19), we have

$$\langle \Delta B \mid x\rangle = -B_x(b_0x^2 + b_1x - \tfrac{1}{2}Dx)\Delta t + \tfrac{1}{2}B_{xx}Dx^2\Delta t + O(\Delta t)^2. \tag{27}$$

Now taking the expected value of (27), we have as $\langle\langle \Delta B \mid x\rangle\rangle = \langle \Delta B\rangle$,

$$\langle \Delta B\rangle = -\langle B_x(b_0x^2 + b_1x - \tfrac{1}{2}Dx)\rangle\Delta t + \tfrac{1}{2}D\langle B_{xx}x^2\rangle\Delta t + O(\Delta t)^2. \tag{28}$$

Dividing through by Δt and taking limit as $\Delta t \to 0$ and interchanging $\langle -\rangle$ and d operator on the left-hand side, we obtain the ordinary differential

$$\frac{d\langle B\rangle}{dt} = -\langle B_x(b_0x^2 + b_1x - \tfrac{1}{2}Dx)\rangle + \tfrac{1}{2}D\langle B_{xx}x^2\rangle. \tag{29}$$

The moment equations of the system are found by substituting $B = x$, x^2, x^3, etc., and we have

$$\frac{d\langle x\rangle}{dt} = -b_0\langle x^2\rangle - b_1\langle x\rangle + \tfrac{1}{2}D\langle x\rangle, \tag{30}$$

$$\frac{d\langle x^2\rangle}{dt} = -2b_0\langle x^3\rangle - 2b_1\langle x^2\rangle + D\langle x^2\rangle + D\langle x^2\rangle, \tag{31}$$

$$\frac{d\langle x^3\rangle}{dt} = -3b_0\langle x^4\rangle - 3b_1\langle x^3\rangle + \tfrac{3}{2}D\langle x^3\rangle + 3D\langle x^3\rangle, \tag{32}$$

and so on. If we continue in this manner, an infinite system of differential equations is obtained.

B. Further Moment Equations

We consider again the system (2), with $n(t)$ as a random noise term with large correlation time. As in Sec. II.B we regard $n(t)$ as a new state variable governed by Eq. (20). Consider again $B(x, n; t)$ an arbitrary function of $x(t)$ and $n(t)$ whose first and second partial derivatives are bounded. We have used Taylor's series

$$\Delta B = B_x\Delta x + B_n\Delta n + \tfrac{1}{2}B_{xx}(\Delta x)^2 + \tfrac{1}{2}B_{nn}(\Delta n)^2 + O(\Delta x, \Delta n). \tag{33}$$

Following the same procedure in Sec. III.A and using Eqs. (22), (23), (24), and (25), we arrive at the ordinary differential equation

$$\frac{d\langle B\rangle}{dt} = -b_0\langle B_xx^2\rangle - b_1\langle B_xx\rangle - \langle B_xnx\rangle - a\langle B_nn\rangle + C\langle B_{nn}\rangle. \tag{34}$$

The moment equations are obtained by substituting the appropriate value for B. Therefore, we have

$$\frac{d\langle x\rangle}{dt} = -b_0\langle x^2\rangle - b_1\langle x\rangle - \langle nx\rangle, \tag{35}$$

$$\frac{d\langle x^2\rangle}{dt} = -2b_0\langle x^3\rangle - 2b_1\langle x^2\rangle - 2\langle nx^2\rangle, \tag{36}$$

$$\frac{d\langle nx\rangle}{dt} = -b_0\langle nx^2\rangle - b_1\langle nx\rangle - \langle n^2x\rangle - a\langle nx\rangle, \tag{37}$$

$$\frac{d\langle nx^2\rangle}{dt} = -2b_0\langle nx^3\rangle - 2b_1\langle nx^2\rangle - 2\langle n^2x^2\rangle - a\langle nx^2\rangle, \tag{38}$$

$$\frac{d\langle n^2x\rangle}{dt} = -b_0\langle n^2x^2\rangle - b_0\langle n^2x\rangle - 2a\langle n^2x\rangle + 2C\langle x\rangle. \tag{39}$$

Continuing in this manner, we again obtain an infinite system of differential equations.

C. Moment Equations of Higher-Order Systems

A differential equation governing the moment equations is found similarly for an nth-order stochastic differential equation whose coefficients contain random parameters with small and large correlation time. If x_i is a state variable (or x is an n-dimensional state vector) of an nth-order system, then we again have, using Taylor's series,

$$\Delta B = \sum_{i}^{n} B_{x_i} \Delta x_i + \tfrac{1}{2} \sum_{i,j}^{n} B_{x_i x_j} \Delta x_i \Delta x_j + O(\Delta x \Delta x^T). \tag{40}$$

Taking the conditional expectation of (40), given x, we have

$$\langle \Delta B \mid x \rangle = \sum_{i}^{n} B_{x_i} \langle \Delta x_i \mid x \rangle + \tfrac{1}{2} \sum_{i,j}^{n} B_{x_i x_j} \langle \Delta x_i \Delta x_j \mid x \rangle + O(\Delta t)^2. \tag{41}$$

We then find $\langle \Delta x_i \mid x \rangle$ and $\langle \Delta x_i \Delta x_j \mid x \rangle$ by the methods described in Secs. II.A and II.B. The differential equation governing the moment equations can then be found by taking the expected value of (41), dividing by Δt, and taking limit as $\Delta t \to 0$.

It should be pointed out that in systems containing random parameters with large correlation time a new state variable representing random term always has to be introduced into system.

IV. LINEARIZATION

So far, we have obtained an infinite system of differential equations governing the moment equations. In order to obtain either an analytic or computational solution, we employ a closure method which replaces the infinite system by a finite system.

The simplest one is to use the Poincaré–Lyapunov stability theorem.[14] Consider a vector–matrix differential equation

$$\frac{dz}{dt} = Az + g(z), \quad z(0) = c, \tag{42}$$

where A is a constant matrix whose characteristic roots have negative real parts and $g(z)$ is a vector all of whose components are power series in the component of z lacking constant and first-degree terms. The classical stability theorem of Poincaré and Lyapunov then asserts that, if $\|c\|$ is sufficiently small, the solution of (42) is bounded and tends to zero as $t \to \infty$. Bellman and Richardson[3] have shown that an infinite system of differential equations can be replaced by a finite one, where z_i is now the order of moment of system, by ignoring the higher-degree terms (or nonlinear terms) provided they satisfy the conditions of the Poincaré–Lyapunov stability theorem. Moreover, the larger the order of the system (or moments), the better the approximation. For example, the moment equations described by Eqs. (30)–(32) can be approximated to a finite one to third order by omitting the term $\langle x^4 \rangle$ in (32), providing the conditions for the Poincaré–Lyapunov theorem are satisfied.

If the conditions for the Poincaré–Lyapunov theorem are not satisfied, then a "dishonest" approach can be used[5,6] to linearize the nonlinear term. For example, in Eqs. (35)–(37) for moments up to second order we put $\langle nx^2 \rangle \simeq \langle n \rangle \langle x^2 \rangle = 0$, since n has zero mean, and $\langle n^2 x \rangle \simeq \langle n^2 \rangle \langle x \rangle$; and $\langle n^2 \rangle$ is a constant if n is stationary. If we wish to approximate up to third-order moments, we use Eqs. (38) and (39), together with (35)–(37), and put $\langle nx^3 \rangle \simeq \langle n \rangle \langle x^3 \rangle = 0$,

$$\langle n^2 x^2 \rangle \simeq \langle n^2 \rangle \langle x^2 \rangle,$$

and $\langle n^3 x \rangle \simeq \langle n^3 \rangle \langle x \rangle = 0$, since we can show that the moments of n raised to odd powers are zero in stationary case by using (34).

Another form of linearization[3,15] is to replace the nonlinear term by a constant times a linear term and use a mean-square norm as a measure of approximation to find the constant term. This method of approximation is probably more cumbersome than the others previously described.

V. PRESERVATION OF MOMENT PROPERTIES

In approximating moment equations, we have not shown that moment properties are preserved. For example, we must have $\langle x^2 \rangle \geq \langle x^2 \rangle$. The method of Bellman and Richardson[4] ensures these moment properties, no matter what type of approximation is used to form a finite closure. The technique to be described is also given by Sancho.[1] It consists of using the linear matrix differential equation

$$Y' = BY + YB^{\mathrm{T}}, \quad Y(0) = K, \tag{43}$$

whose solution is given by

$$Y(t) = e^{Bt} K e^{B^{\mathrm{T}} t}. \tag{44}$$

Now, $Y(t)$ is positive definite if K is positive definite for $t > 0$. If moments up to second order are required, we replace Y' by the linearized solution of moment equations (up to second order in the example shown

below) in a symmetric form, i.e., Y' is replaced by

$$\begin{pmatrix} \langle x^0 \rangle & \langle x \rangle \\ \langle x \rangle & \langle x^2 \rangle \end{pmatrix}', \tag{45}$$

where $\langle x^0 \rangle = 1$ and the linearized form of $\langle x \rangle'$ and $\langle x^2 \rangle'$ has been substituted into (45). We let

$$B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}, \tag{46}$$

where b_1, b_2, b_3, and b_4 are real parameters to be determined in such a manner that

$$\left\| \begin{pmatrix} \langle x^0 \rangle' & \langle x \rangle' \\ \langle x \rangle' & \langle x^2 \rangle \end{pmatrix} - B \begin{pmatrix} 1 & \langle x \rangle \\ \langle x \rangle & \langle x^2 \rangle \end{pmatrix} - \begin{pmatrix} 1 & \langle x \rangle \\ \langle x \rangle & \langle x^2 \rangle \end{pmatrix} B^{\mathrm{T}} \right\|^2 \tag{47}$$

is a minimum. We always obtain $b_1 = b_2 = 0$ if $\langle x^0 \rangle = 1$ and $b_3 = 0$ if $\langle x \rangle \to 0$ as $t \to \infty$. Hence we set

$$\|Y\|^2 = \int_0^\infty \mathrm{Tr}\,(YY^{\mathrm{T}})\,dt = \int_0^\infty \left(\sum_{i,j} y_{ij}^2 \right) dt. \tag{48}$$

The resulting equations for the first and second moments after (47) is solved for b_4 are

$$\langle x \rangle = k_1 e^{b_4 t} \quad \text{and} \quad \langle x^2 \rangle = k_2 e^{2b_4 t}, \tag{49}$$

where $k_1 = \langle x(0) \rangle$ and $k_2 = \langle x^2(0) \rangle$. Since $\langle x \rangle^2 / \langle x^2 \rangle = k_1^2/k_2$, the moment inequality is clearly preserved.

For higher-order moments we choose

$$B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & b_5 & b_6 \\ 0 & b_8 & b_9 \end{pmatrix}, \tag{50}$$

since $\langle x^0 \rangle$ is always taken to be unity. The nine parameters are reduced to four and we proceed as above.

CONCLUSION

We have studied the solution of moment equations of a nonlinear stochastic system containing random parameters with small and large correlation time. For systems containing random parameters with large correlation time, we assume that the random term $n(t)$ as a new state variable. It is, therefore, Gaussian and Markovian and governed by a stochastic differential equation (20). In either case, an infinite system of moment equations is always obtained, and we have described three different ways by which these moment equations can be closed to a finite one. The method of Bellman and Richardson[7] is then used to preserve certain moment inequalities; for example, we have $\langle x^2 \rangle \geq \langle x \rangle^2$, a consequence of the Schwarz inequality. The point, however, is that, no matter what type of approximation we use to close infinite system into finite one, the moment properties are always preserved.

* This research was supported in part by the National Research Council of Canada under contract No. A-4051.

[1] N. G. F. Sancho, J. Math. Phys. **11**, 762 (1970).
[2] J. B. Morton and S. Corrsin, J. Math. Phys. **10**, 361 (1969).
[3] R. Bellman and J. M. Richardson, Quart. Appl. Math. **20**, 333 (1963).
[4] R. Bellman and J. M. Richardson, Proc. Natl. Acad. Sci. U.S. **47**, 1191 (1961).
[5] J. M. Richardson, *Proceedings of Symposia in Applied Mathematics, Vol. 16* (American Mathematical Society, Providence, R.I., 1964).
[6] W. E. Boyce, SIAM J. Appl. Math. **15**, 143 (1967).
[7] R. Bellman and J. M. Richardson, J. Math. Anal. Appl. **23**, 639 (1968).
[8] J. L. Doob, *Stochastic Processes* (John Wiley & Sons, Inc., New York, 1953).
[9] A. H. Gray and T. K. Caughey, J. Math. & Phys. **44**, 288 (1965).
[10] R. L. Stratonovich, SIAM J. Control **4**, 362 (1966).
[11] E. Wong and M. Zakai, Third International Federation for Automatic Control Congress, London, **3B** (1966).
[12] M. A. Leibowitz, J. Franklin Inst. **287**, 223 (1969).
[13] I. G. Cumming, Intern. J. Control **5**, 85 (1967).
[14] R. Bellman, *Stability Theory of Differential Equations* (McGraw-Hill Book Company, New York (1959).
[15] N. G. F. Sancho, J. Math. Anal. Appl. (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Quasiparticle Formalism and Atomic Shell Theory. II. Mixed Configurations*

M. J. CUNNINGHAM AND B. G. WYBOURNE
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 17 July 1969)

The quasiparticle technique and formalism of an earlier paper is extended to describe general mixed configurations. The main features of the equivalent electron case are preserved, viz., a loss of spin quantum number and particle number, compensated by a rich classificatory scheme, and general operator-matrix-element evaluation without the use of coefficients of fractional parentage.

I. INTRODUCTION

Recently, there has been some interest among atomic spectroscopists in exploring the possibilities of applying the quasiparticle technique, well known in nuclear and superconducting physics, to problems in atomic spectroscopy[1–3] following an initial paper by Armstrong and Judd.[2]

Shudeman[4] has shown that the states of an atomic shell may be classified by considering the symmetrical product of a spin-up (all $m_s = +\frac{1}{2}$) and a spin-down (all $m_s = -\frac{1}{2}$) space. Group-theoretically, in the case of the l shell, this amounts to factorizing the group $U_{2^{4l+2}}$, by which all 2^{4l+2} states of the l shell transform, into the chain of product groups

$$\begin{aligned} R_{8l+5} &\to R^{\uparrow}_{4l+3} \times R^{\downarrow}_{4l+3} \to (R^{\uparrow}_{2} \times R^{\uparrow}_{2l+1}) \times (R^{\downarrow}_{2} \times R^{\downarrow}_{2l+1}) \\ &\to (R^{\uparrow}_{2} \times R^{\uparrow}_{3}) \times (R^{\downarrow}_{2} \times R^{\downarrow}_{3}) \to R_2 \times R_3, \end{aligned} \tag{1}$$

where the arrow denotes whether the spin is "up" or "down" and the representations of $R^{\uparrow}_2$, $R^{\downarrow}_2$, and R_2 label the states according to the quantum numbers $M^{\uparrow}_S$, $M^{\downarrow}_S$, and $M_S = M^{\uparrow}_S + M^{\downarrow}_S$. The quasiparticle method has the distinguishing feature of an alternative factorization of each of the groups R_{4l+3} into the product of two R_{2l+1} groups. This factorization is achieved first by defining operators

$$\begin{aligned} \lambda^{(l)}_q &= 2^{-\frac{1}{2}}[a^{+}_{\frac{1}{2}q} + (-1)^{l-q} a_{\frac{1}{2}-q}], \\ \mu^{(l)}_q &= 2^{-\frac{1}{2}}[a^{+}_{\frac{1}{2}q} - (-1)^{l-q} a_{\frac{1}{2}-q}], \\ \nu^{(l)}_q &= 2^{-\frac{1}{2}}[a^{+}_{-\frac{1}{2}q} + (-1)^{l-q} a_{-\frac{1}{2}-q}], \\ \xi^{(l)}_q &= 2^{-\frac{1}{2}}[a^{+}_{-\frac{1}{2}q} - (-1)^{l-q} a_{-\frac{1}{2}-q}], \end{aligned} \tag{2}$$

where the subscripts to the annihilation and creation operators a and a^+ specify m_s and m_l for an electron, and then by showing that coupled tensor products of the form $\frac{1}{2}(\lambda^{(l)}\lambda^{(l)})^{(K)}_Q$ for *odd* K form the generators of an R_{2l+1} group. The relevant group chain for an atomic l shell was shown in our recent paper[1] (hereafter referred to as I) to be

$$\begin{aligned} U_{2^{4l+2}} &\to R_{8l+5} \to R^{\uparrow}_{4l+3} \times R^{\downarrow}_{4l+3} \to R^{\uparrow}_{4l+2} \times R^{\downarrow}_{4l+2} \\ &\to (R^{\lambda}_{2l+1} \times R^{\mu}_{2l+1})^{\uparrow} \times (R^{\xi}_{2l+1} \times R^{\nu}_{2l+1})^{\downarrow} \\ &\to (R^{\lambda}_{3} \times R^{\mu}_{3})^{\downarrow} \to (R^{\xi}_{3} \times R^{\nu}_{3})^{\downarrow} \to R^{\lambda\mu\uparrow}_{3} \times R^{\xi\nu\downarrow}_{3} \\ &\to R_3 \to R_2. \end{aligned} \tag{3}$$

In I we define fermion quasiparticle creation and annihilation operators, directly related to the operators of Eq. (2), and suitable quasiparticle vacuum states, and are able to show that an antisymmetric quasiparticle state transforms according to the basic spin representation Δ of R_{4l+3} and as one of the conjugate spin representations

$$\Delta_1 = [\tfrac{1}{2}\ \tfrac{1}{2} \cdots \tfrac{1}{2}]$$

or

$$\Delta_2 = [\tfrac{1}{2}\ \tfrac{1}{2} \cdots -\tfrac{1}{2}]$$

of R_{4l+2}. Although the particle number is lost in this system, we were able to show that Δ_1 is associated with an *even* number of particles (real particles or quasiparticles) while Δ_2 is associated with an *odd* number of particles. The states of the l shell transform according to the $\{1\}$ vector representation[5] of $U_{2^{4l+2}}$ which decomposes under the group chain of Eq. (3) as (the R_{4l+3} groups are now suppressed)

$$\begin{aligned} \{1\} &\to \Delta \to (\Delta_1 + \Delta_2)^{\uparrow} \times (\Delta_1 + \Delta_2)^{\downarrow} \\ &\to (\Delta_\lambda \times \Delta_\mu)^{\uparrow} \times (\Delta_\xi \times \Delta_\nu)^{\downarrow} \\ &\to (L_\lambda \times L_\mu)^{\uparrow} \times (L_\xi \times L_\nu)^{\downarrow} \\ &\to L^{\uparrow}_{\lambda\mu} \times L^{\downarrow}_{\xi\nu} \to L \to M_L. \end{aligned} \tag{4}$$

We note that the spin quantum number S has been lost.

The quasiparticle scheme has the weakness of yielding eigenfunctions associated with neither a well-defined number of particles nor spin quantum numbers S and M_S. We have in compensation, first, a remarkably rich classification scheme, since the spin representation of R_{2l+1} decomposes without duplication upon restriction of R_3 for all $l < 9$, and, second, a method for calculating matrix elements without recourse to the usual fractional-parentage methods, since any state can be regarded as the vector coupling of four $R_3 - R_2$ states and, consequently, can be handled by the usual vector coupling methods.[6,7] Details of these methods have been discussed in I for the case of an atomic shell of equivalent l orbitals.

To date, the application of quasiparticle methods to mixed configurations has been treated by only one author, Feneuille,[8] who considered the case of $(s + d)^N$. In this paper we extend the methods and formalism of I to general mixed configurations. The relevant group structure becomes

$$\begin{aligned} U_{2^{2\rho}} &\to R_{4\rho+1} \to R_{2\rho}^{\uparrow} \times R_{2\rho}^{\downarrow} \\ &\to (R_\rho^\lambda \times R_\rho^\mu)^{\uparrow} \times (R_\rho^\xi \times R_\rho^\nu)^{\downarrow} \\ &\to (R_3^\lambda \times R_3^\mu)^{\uparrow} \times (R_3^\xi \times R_3^\nu)^{\downarrow} \to R_3^{\lambda\mu\uparrow} \times R_3^{\xi\nu\downarrow} \\ &\to R_3 \to R_2, \end{aligned} \tag{5}$$

where $\rho = \sum_i (2l_i + 1)$ and the summation is taken over all orbitals characterizing the mixed configurations under study. Suitable linear combinations of fermion quasiparticle states are found forming bases for the spin representations Δ_1 or Δ_2 of $R_{2\rho}$ and Δ or one of Δ_1 or Δ_2 of R_ρ, as ρ is *odd* or *even*. This linear combination allows us to find the decomposition of the vector representation $\{1\}$ of $U_{2^{2\rho}}$ under the group structure of Eq. (5) and also to separate, in a natural manner, certain multiplicities that arise at the R_3 level if three or more orbitals are under consideration. Finally, the $\mathbf{W}^{(\kappa k)}(ll')$ tensor operators are expressed in the quasiparticle scheme and their matrix elements, which are shown to be generally both real and imaginary, are evaluated, as in I, without the necessity of fractional-parentage coefficients.

II. THE GROUP STRUCTURE

In this section we establish the group structure specified in Eq. (5). This is achieved by forming quasiparticle operators $\lambda_q^{(l)}$, etc., as in Eq. (2), where now the l quantum number can range over all the orbitals of the mixed configuration under consideration. The first two groups of the chain $(U_{2^{2\rho}} \to R_{4\rho+1})$ exist as a result of a self-evident extension of the results of Judd[5] and Feneuille[9] on the ordinary creation and annihilation operators which may easily be retrieved from Eq. (2); e.g.,

$$a_{\frac{1}{2}q}^{(l)+} = 2^{-\frac{1}{2}}(\lambda_q^{(l)} + \mu_q^{(l)}), \quad \text{etc. [cf. Eq. (8) of I].}$$

To prove the existence of the lower groups, we must consider coupled products of the form $\frac{1}{2}(A^{l_i}B^{l_j})_Q^{(K)}$, where A and B are one of λ, μ, ξ, and ν. We note that, if $A \equiv B$ and $l_i \equiv l_j$, this coupled product exists only for *odd* k and $k = 0$. These coupled products satisfy the commutation relations (all l's assumed integral)

$$\begin{aligned} &[\tfrac{1}{2}(A^{l_1}B^{l_2})_{Q_1}^{(K_1)}, \tfrac{1}{2}(C^{l_3}D^{l_4})_{Q_2}^{(K_2)}] \\ &= \tfrac{1}{2}a_{BC}\delta(l_2, l_3)\delta(B, C)\sum_{K_3,Q_3}(-1)^{l_1+l_4+K_1+K_2+K_3-Q_3}[K_1, K_2, K_3]^{\frac{1}{2}}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_4 & l_1 & l_2 \end{Bmatrix}\begin{pmatrix} K_1 & K_2 & K_3 \\ Q_1 & Q_2 & -Q_3 \end{pmatrix}\tfrac{1}{2}(A^{l_1}D^{l_4})_{Q_3}^{(K_3)} \\ &\quad - \tfrac{1}{2}a_{AD}\delta(l_1, l_4)\delta(A, D)\sum_{K_3,Q_3}(-1)^{l_2+l_3-Q_3}[K_1, K_2, K_3]^{\frac{1}{2}}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_3 & l_2 & l_1 \end{Bmatrix}\begin{pmatrix} K_1 & K_2 & K_3 \\ Q_1 & Q_2 & -Q_3 \end{pmatrix}\tfrac{1}{2}(C^{l_3}D^{l_4})_{Q_3}^{(K_3)} \\ &\quad - \tfrac{1}{2}a_{DB}\delta(l_4, l_2)\delta(D, B)\sum_{K_3,Q_3}(-1)^{l_1+l_2+K_1+K_3-Q_3}[K_1, K_2, K_3]^{\frac{1}{2}}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_3 & l_1 & l_2 \end{Bmatrix}\begin{pmatrix} K_1 & K_2 & K_3 \\ Q_1 & Q_2 & -Q_3 \end{pmatrix}\tfrac{1}{2}(A^{l_1}C^{l_3})_{Q_3}^{(K_3)} \\ &\quad + \tfrac{1}{2}a_{AC}\delta(l_1, l_3)\delta(A, C)\sum_{K_3,Q_3}(-1)^{l_2+l_3+K_2-Q_3}[K_1, K_2, K_3]^{\frac{1}{2}}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_4 & l_2 & l_1 \end{Bmatrix}\begin{pmatrix} K_1 & K_2 & K_3 \\ Q_1 & Q_2 & -Q_3 \end{pmatrix}\tfrac{1}{2}(D^{l_4}B^{l_2})_{Q_3}^{(K_3)}, \end{aligned} \tag{6}$$

where a_{XY} is $+1$, if $X = Y = \lambda$ or ξ, or -1, if $X = Y = \mu$ or ν.

We now consider the operators like

$$(-1)^{\frac{1}{2}(a+b)}\tfrac{1}{2}(A^{l_1}B^{l_2})_Q^{(K)} - (-1)^{l_1-l_2+K}(\tfrac{1}{2}B^{l_2}A^{l_1})_Q^{(K)},$$

where $a = 2$, if $A = \lambda$ or ξ, or $a = +1$, if $A = \mu$ or ν, and similarly between b and B; thus, we find the following commutation relation [we note that $(-1)^{\frac{1}{2}(a+b)}$ can be imaginary]:

$$
\begin{aligned}
&[(-1)^{\frac{1}{2}(a+b)}\{\tfrac{1}{2}(A^{l_1}B^{l_2})^{(K_1)}_{Q_1} - (-1)^{l_A-l_B+K_1}\tfrac{1}{2}(B^{l_2}A^{l_1})^{(K_1)}_{Q_1}\}, (-1)^{\frac{1}{2}(c+d)}\{\tfrac{1}{2}(C^{l_3}D^{l_4})^{(K_2)}_{Q_2} - (-1)^{l_C-l_D+K_2}\tfrac{1}{2}(D^{l_4}C^{l_3})^{(K_2)}_{Q_2}\}] \\
&\quad = \sum_{K_3Q_3}(-1)^{K_3-Q_3+l_1+l_4}[K_1, K_2, K_3]^{\frac{1}{2}}\begin{pmatrix} K_1 & K_2 & K_3 \\ Q_1 & Q_2 & -Q_3 \end{pmatrix} \\
&\qquad \times \Big(\delta(l_2l_3)\delta(BC)(-1)^{K_1+K_2}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_4 & l_1 & l_2 \end{Bmatrix}(-1)^{\frac{1}{2}(a+d)}\{\tfrac{1}{2}(A^{l_1}D^{l_4})^{(K_3)}_{Q_3} - (-1)^{l_1-l_4+K_3}\tfrac{1}{2}(D^{l_4}A^{l_1})^{(K_3)}_{Q_3}\} \\
&\qquad + \delta(l_1l_4)\delta(AD)\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_3 & l_2 & l_1 \end{Bmatrix}(-1)^{\frac{1}{2}(b+c)}\{\tfrac{1}{2}(B^{l_2}C^{l_3})^{(K_3)}_{Q_3} - (-1)^{l_2-l_3+K_3}\tfrac{1}{2}(C^{l_3}B^{l_2})^{(K_3)}_{Q_3}\} \\
&\qquad - \delta(l_2l_4)\delta(BD)(-1)^{K_1}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_3 & l_1 & l_2 \end{Bmatrix}(-1)^{\frac{1}{2}(a+c)}\{\tfrac{1}{2}(A^{l_1}C^{l_3})^{(K_3)}_{Q_3} - (-1)^{l_1-l_3+K_3}\tfrac{1}{2}(C^{l_3}A^{l_1})^{(K_3)}_{Q_3}\} \\
&\qquad - \delta(l_1l_3)\delta(AC)(-1)^{(K_2)}_{Q_2}\begin{Bmatrix} K_1 & K_2 & K_3 \\ l_4 & l_2 & l_1 \end{Bmatrix}(-1)^{\frac{1}{2}(b+d)}\{\tfrac{1}{2}(B^{l_2}D^{l_4})^{(K_3)}_{Q_3} - (-1)^{l_2-l_4+K_3}\tfrac{1}{2}(D^{l_4}B^{l_2})^{(K_3)}_{Q_3}\}\Big). \qquad (7)
\end{aligned}
$$

This is, however, precisely the commutation relation of the

$$\bar{V}^{(k)}_q(l_1, l_2) = V^{(k)}_q(l_1, l_2) - (-1)^{l_1-l_2+k}V^{(k)}_q(l_2, l_1)$$

operators of Butler and Wybourne.[10] Moreover, they have shown that the set of operators $\bar{V}^{(k)}_q(l_i, l_i)$, with *odd* allowed k and $\bar{V}^{(k)}_q(l_i, l_j)$ for all l_i and l_j, $i < j$, of a given configuration with all allowed k, form the infinitesimal operators for the group R_ρ. Hence, the set of operators

$$(-1)^a[\tfrac{1}{2}(A^{l_i}A^{l_j})^{(K)}_Q - (-1)^{l_i-l_j+K}\tfrac{1}{2}(A^{l_j}A^{l_i})^{(K)}_Q], \qquad l_i \le l_j,$$

form the infinitesimal operators for the group R^A_ρ (taking K *odd* if $l_i = l_j$), while this set, together with

$$(-1)^b[\tfrac{1}{2}(B^{l_i}B^{l_j})^{(K)}_Q - (-1)^{l_i-l_j+K}\tfrac{1}{2}(B^{l_j}B^{l_i})^{(K)}_Q], \qquad l_i \le l_j,$$

where B is the other of the pair (λ, μ) or (ξ, ν), and

$$(-1)^{\frac{1}{2}(a+b)}[\tfrac{1}{2}(A^{l_i}B^{l_j})^{(K)}_Q - (-1)^{l_i-l_j+K}\tfrac{1}{2}(B^{l_j}A^{l_i})^{(K)}_Q]$$

form the infinitesimal operators of $R^{\uparrow}_{2\rho}$ if the pair is (λ, μ) or $R^{\downarrow}_{2\rho}$ if the pair is (ξ, ν). We note that this construction of R_ρ differs from Feneuille's treatment,[8] for he constructs this group in the two orbital case essentially from the coupled products $(\lambda^{l_1}\lambda^{l_1})^{(K)}_Q$, $(\mu^{l_2}\mu^{l_2})^{(K)}_Q$, and $(\lambda^{l_1}\mu^{l_2})^{(K)}_Q$. The R_3 part of the group chain follows immediately by limiting the tensor ranks to one and zero.

III. THE WEYL SELF-COMMUTING OPERATORS

At this stage we define quasiparticle creation and annihilation operators of the same form as in I, viz.,

$$
\begin{aligned}
\beta^{(l)+}_{\pm\frac{1}{2}q} &= 2^{-\frac{1}{2}}(a^+_{\pm\frac{1}{2}q} + (-1)^{l-q}a_{\pm\frac{1}{2}-q}) \\
&\equiv \lambda^{(l)}_q \quad \text{or} \quad \xi^{(l)}_q, \\
(\beta^{(l)+}_{\pm\frac{1}{2}q})^+ &= \beta^{(l)}_{\pm\frac{1}{2}q} = 2^{-\frac{1}{2}}(a_{\pm\frac{1}{2}q} + (-1)^{l-q}a_{\pm\frac{1}{2}-q}) \\
&\equiv (-1)^{l-q}\lambda^{(l)}_q \quad \text{or} \quad (-1)^{l-q}\xi^{(l)}_q, \\
\gamma^{(l)+}_{\pm\frac{1}{2}q} &= 2^{-\frac{1}{2}}(a_{\pm\frac{1}{2}q} - (-1)^{l-q}a^+_{\pm\frac{1}{2}-q}) \\
&\equiv -(-1)^{l-q}\mu^{(l)}_{-q} \quad \text{or} \quad -(-1)^{l-q}\nu^{(l)}_{-q}, \\
(\gamma^{(l)+}_{\pm\frac{1}{2}q})^+ &= \gamma^{(l)}_{\pm\frac{1}{2}q} = 2^{-\frac{1}{2}}(a^+_{\pm\frac{1}{2}q} - (-1)^{l-q}a_{\pm\frac{1}{2}-q}) \\
&\equiv \mu^{(l)}_q \quad \text{or} \quad \nu^{(l)}_q,
\end{aligned}
$$

all with $q > 0$. In analogy with I, the quasiparticle vacuum becomes

$$N\prod_l\prod_{q>0}\beta^{(l)}_q\prod_l\prod_{q>0}\gamma^{(l)}_q|0\rangle = |\bar{0}\rangle,$$

where N is a normalization constant suitably chosen such that $\langle\bar{0}\,|\,\bar{0}\rangle = 1$. This definition of the vacuum satisfies $\beta^{(l)}_q|\bar{0}\rangle$ for all l and q, as required.

In order to find the representation by which a quasiparticle state transforms, we must first form the Weyl commuting operators,[11] hereafter referred to as the H operators, for the groups $R_{2\rho}$ and R_ρ. Let us designate a configuration by $(l_1 + l_2 + \cdots + l_{2n})^N$ for an *even* number of orbitals, or by $(l_1 + l_2 + \cdots + l_{2n+1})^N$ for an *odd* number of orbitals, where l_i and l_j may be identical if they are associated with different principal quantum numbers. Butler and Wybourne[10] have shown, in terms of this notion, that, if $W^{l_il_j}_{ab}$ is defined as

$$
\begin{aligned}
W^{l_il_j}_{ab} = \sum_{k,q}(-1)^{l_i-a}[k]^{\frac{1}{2}}&\begin{pmatrix} l_i & k & l_j \\ -a & q & b \end{pmatrix} \\
&\times [V^{(k)}_q(l_i, l_j) - (-1)^{l_i-l_j+k}V^{(k)}_q(l_j, l_i)], \qquad (8)
\end{aligned}
$$

then the operators required for R_ρ are of the forms $W^{l_il_i}_{aa}$ for all l_i, characterizing the configuration with all permissible a, and

$$(-1)^{\frac{1}{2}(l_{2i-1}+l_{2i}+1)}W^{l_{2i-1}l_{2i}}_{00},$$

where $i = 1, \cdots, n$.

In the quasiparticle scheme we find for R_ρ^λ

$$W_{aa}^{ll} = \tfrac{1}{2}(-1)^{l-a}[\lambda_a^{(l)}, \lambda_{-a}^{(l)}] = \tfrac{1}{2}[\beta_a^{(l)+}, \beta_a^{(l)}]$$

and

$$(-1)^{\frac{1}{2}(l_x+l_y+1)}W_{00}^{l_xl_y} = \tfrac{1}{2}(-1)^{\frac{1}{2}(l_x+3l_y+1)}[\lambda_0^{(l_x)}, \lambda_0^{(l_y)}]$$
$$= \tfrac{1}{2}(-1)^{\frac{1}{2}(l_x-l_y+1)}[\beta_0^{(l_x)}, \beta_0^{(l_y)}],$$

while for R_ρ^μ

$$W_{aa}^{ll} = -\tfrac{1}{2}(-1)^{l-a}[\mu_a^{(l)}, \mu_{-a}^{(l)}] = -\tfrac{1}{2}[\gamma_a^{(l)+}, \gamma_a^{(l)}]$$

and

$$(-1)^{\frac{1}{2}(l_x+l_y+1)}W_{00}^{l_xl_y} = -\tfrac{1}{2}(-1)^{\frac{1}{2}(l_x+3l_y+1)}[\mu_0^{(l_x)}, \mu_0^{(l_y)}]$$
$$= -\tfrac{1}{2}(-1)^{\frac{1}{2}(l_x-l_y+1)}[\gamma_0^{(l_x)}, \gamma_0^{(l_y)}],$$

where $x = 2i - 1$ and $y = 2i$ in each case.

The H operators for the group $R_{2\rho}^{\uparrow}$ are identical to those for R_ρ^λ and R_ρ^μ except for the case of an odd number of orbitals when we must add the operator

$$H^{\lambda\mu} = \tfrac{1}{2}[\lambda_0^{(l)}, \mu_0^{(l)}] = \tfrac{1}{2}(-1)^l[\beta_0^{(l)}, \gamma_0^{(l)}],$$

where $l \equiv l_{2n+1}$, i.e., l is the single arbitrarily chosen unpaired orbital of the configuration. It now remains to find the eigenfunctions of the H operators.

IV. REPRESENTATIONS

A general eigenfunction for the H operators of R_ρ^λ takes the form

$$N\{\cdots\}\prod_{i=1}^{n}\{[(-1)^{\frac{1}{2}(1-l_x)}\beta_0^{(l_x)} \pm_x (-1)^{\frac{1}{2}l_y}\beta_0^{(l_y)}]$$
$$\pm_y [-2^{-\frac{1}{2}} \pm_x (-1)^{\frac{1}{2}(1-l_x+l_y)}2^{\frac{1}{2}}\beta_0^{(l_x)}\beta_0^{(l_y)}]\}\,|\tilde{0}\rangle$$
$$= \{\cdots\}\prod_{i=1}^{n} M_{\pm x,\pm y}^{\lambda}, \qquad (9)$$

where $\{\cdots\}$ denotes an arbitrary product of β quasiparticle creation operators with nonzero subscripts, apart from possibly $\beta_0^{l_{2n+1}}$ if the number of orbitals is *odd* (see I), i is defined as in Sec. III, i.e., the product is being taken over pairs of orbitals, and the subscript to the $\pm$ signs means that signs with different subscripts are independent; as before, N is a suitable normalizing constant and $|\tilde{0}\rangle$ is the vacuum for the R_ρ^λ group and is

$$\propto \prod_i \prod_{q>0} \beta_q^{(i)}\,|0\rangle.$$

The operation of W_{aa}^{ll} on the eigenfunctions of Eq. (9) yields eigenvalues of $+\frac{1}{2}$ or $-\frac{1}{2}$, depending on whether $\beta_0^{(l)+}$ is or is not present in $\{\cdots\}$ (see I), while operating with

$$(-1)^{\frac{1}{2}(l_x+l_y+1)}W_{00}^{l_xl_y}$$

gives eigenvalues $\pm_x\frac{1}{2}$. The corresponding eigenfunction for R_ρ^μ is

$$N\{\cdots\}\prod_{i=1}^{n}\{[(-1)^{\frac{1}{2}(1-l_x)}\gamma_0^{(l_y)} \pm_x (-1)^{\frac{1}{2}l_y}\gamma_0^{(l_y)}]$$
$$\pm_y [2^{-\frac{1}{2}} \mp_x (-1)^{\frac{1}{2}(1-l_x+l_y)}2^{\frac{1}{2}}\gamma_0^{(l_x)}\gamma_0^{(l_y)}]\}\,|\tilde{0}\rangle$$
$$= \{\cdots\}\prod_{i=1}^{n} M_{\pm x,\pm y}^{\mu}, \qquad (10)$$

where now W_{aa}^{ll} yields eigenvalues $-\frac{1}{2}$ or $+\frac{1}{2}$ depending on whether $\gamma_0^{(l)+}$ is or is not present in $\{\cdots\}$, while

$$(-1)^{\frac{1}{2}(l_x+l_y+1)}W_{00}^{l_xl_y}$$

gives eigenvalues $\pm_x(-1)^{l_x+l_y}\frac{1}{2}$.

It would appear that the quasiparticle states do indeed form a basis for a spin representation. However, if ρ is *odd*, the dimension of the basic spin representation Δ of R_ρ is

$$2^{\sum_i^{2n+1}l_i+n},$$

with n defined as in Sec. III, while the dimension of the conjugate spin representations of R_ρ for ρ *even* is

$$2^{\sum_i^{2n}l_i+n-1}.$$

It is clear from Eqs. (9) and (10) that the number of eigenstates for R_ρ is

$$2^{\sum_i^{2n}l_i+2n},$$

(for ρ *even*) or

$$2^{\sum_i^{2n+1}l_i+2n}$$

(for ρ *odd*), i.e., 2^n greater than the dimension of the corresponding spin representations. This apparent discrepancy arises since the weight vector can be obtained in any of 2^n ways simply by allowing all combinations in the choice of sign labeled by y. To resolve this problem, we consider the group $R_{2\rho}$.

A general eigenfunction for this group is of the form

$$N\{\cdots\}\prod_{i=1}^{n} M_{(\pm x,\pm y)}^{\lambda}\prod_{i=1}^{n} M_{(\pm x,\pm y)}^{\mu}\,|\tilde{0}\rangle,$$

where $\{\cdots\}$ is now a product of β and γ quasiparticle creation operators, with the possible inclusion of $\beta_0^{(l_{2n+1})}$ and $\gamma_0^{(l_{2n+1})}$. This is equal to

$$N\{\cdots\}\prod_{i=0}^{n} M_{(\pm x,\pm y)}^{\lambda}M_{(\pm x,\pm y)}^{\mu}\,|\tilde{0}\rangle$$

to within a phase since any two nonequivalent M's anticommute. We now look at a typical term $M_{(\pm x,\pm y)}^{\lambda}$,

$M^{\mu}_{(\pm x,\pm y)}$ which, expanded in full, is

$$
\begin{aligned}
&[-\tfrac{1}{2} \pm_{x_\mu} (-1)^{\frac{1}{2}(1-l_x+l_y)} \beta_0^{(l_x)} \beta_0^{(l_y)} \\
&\qquad \pm_{x_\lambda} (-1)^{\frac{1}{2}(1-l_x+l_y)} \beta_0^{(l_y)} \beta_0^{(l_x)} \pm_{x_\lambda} \pm_{x_\mu} \tfrac{1}{2}] \\
&\qquad \pm_{y_\mu} 2^{-\frac{1}{2}} [(-1)^{\frac{1}{2}(1-l_x)} \beta_0^{(l_x)} \mp_{x_\mu} (-1)^{\frac{1}{2}(2+l_y)} \beta_0^{(l_y)} \\
&\qquad \pm_{x_\lambda} (-1)^{\frac{1}{2}l_y} \beta_0^{(l_y)} \pm_{x_\lambda} \pm_{x_\mu} (-1)^{\frac{1}{2}(1-l_x)} \beta_0^{(l_x)}] \\
&\qquad \pm_{y_\lambda} 2^{-\frac{1}{2}} [(-1)^{\frac{1}{2}(1-l_x)} \beta_0^{(l_x)} \pm_{x_\lambda} (-1)^{\frac{1}{2}l_y} \beta_0^{(l_y)} \\
&\qquad \mp_{x_\mu} (-1)^{\frac{1}{2}(2+l_y)} \beta_0^{(l_y)} \pm_{x_\lambda} \pm_{x_\mu} (-1)^{\frac{1}{2}(1-l_x)} \beta_0^{(l_x)}] \\
&\qquad \pm_{y_\lambda} \pm_{y_\mu} [\tfrac{1}{2} \mp_{x_\mu} (-1)^{\frac{1}{2}(1-l_x+l_y)} \beta_0^{(l_x)} \beta_0^{(l_y)} \\
&\qquad \pm_{x_\lambda} (-1)^{\frac{1}{2}(1-l_x+l_y)} \beta_0^{(l_x)} \beta_0^{(l_y)} \pm_{x_\lambda} \pm_{x_\mu} \tfrac{1}{2}],
\end{aligned}
$$

where, for instance, $\pm_{x_\lambda} \pm_{x_\mu}$ is plus if x_λ and x_μ have the same sign and minus if their sign is different. We justify replacement of any $\gamma_0^{(l)}$ by $\beta_0^{(l)}$ by noting that

$$
\begin{aligned}
\gamma_0^{(l)} |\bar{0}\rangle &= \left(N \prod_l \prod_{q>0} \beta_q^{(l)} \prod_l \prod_{q>0} \gamma_q^{(l)} \right) \gamma_0^{(l)} |0\rangle \\
&= \left(N \prod_l \prod_{q>0} \beta_q^{(l)} \prod_l \prod_{q>0} \gamma_q^{(l)} \right) 2^{-\frac{1}{2}} a_0^{(l)+} |0\rangle \\
&= \left(N \prod_l \prod_{q>0} \beta_q^{(l)} \prod_l \prod_{q>0} \gamma_q^{(l)} \right) \beta_0^{(l)} |0\rangle = \beta_0^{(l)} |\bar{0}\rangle.
\end{aligned}
$$

Clearly, the above expression vanishes for certain sign choices, e.g., $x_\lambda = +1, y_\lambda = +1, x_\mu = +1$, and $y_\mu = -1$. In fact, if we choose $x_\lambda = +1$ and $y_\lambda = +1$, then the only μ sign choices compatible (i.e., giving nonvanishing states) are $x_\mu = +1$, $y_\mu = +1$ and $x_\mu = -1, y_\mu = -1$, which are in turn compatible only with the original choice, and $x_\lambda = -1, y_\lambda = -1$. If we choose to start with a different sign choice, our compatible set of sign choices, although different, will give the same quasiparticle states, to within a phase. As a consequence, we see that at the R_ρ level we cannot make random choices in the sign labeled y_λ at the risk of throwing up a vanishing state at the $R_{2\rho}$ level. Restricting ourselves then to some compatible set of sign choices, we see that we have

$$2^{2\sum_i^{2n} l_i + 2n} \quad \text{or} \quad 2^{2\sum_i^{2n+1} l_i + 2n+1}$$

states for $R_{2\rho}$ (as ρ *even* or *odd*), which is just the sum of the dimensions of Δ_1 and Δ_2 of $R_{2\rho}$, and

$$2^{\sum_i^{2n} l_i + n}$$

states (ρ *even*) or

$$2^{\sum_i^{2n+1} l_i + n}$$

(ρ *odd*) states at the R_ρ level, as required.

Since the group operators, being coupled products of quasiparticle operators, connect only states differing by two or zero particles, we see by considering the action of the H operators on a typical eigenstate that the set of eigenstates in the λ space (and, similarly, for the μ space) form a basis for the spin representation Δ^λ or one of Δ_1^λ, Δ_2^λ for the group R_ρ^λ according as ρ is *odd* or *even*, i.e., whether there is an odd or even number of orbitals. Moreover, they also form a basis for either the $\Delta_1^\uparrow$ or $\Delta_2^\uparrow$ spin representations of $R_{2\rho}^\uparrow$, where the parity of the particle number associated with $\Delta_1^\uparrow$ is the parity of n defined above and vice versa for $\Delta_2^\uparrow$. For instance, this means that, for the one orbital case ($n = 0$), $\Delta_1^\uparrow$ is associated with an even number of particles and $\Delta_2^\uparrow$ with an odd number, in agreement with I. The opposite is the case for two orbitals.

V. BRANCHING RULES

We now show that the spin representation Δ of $R_{4\rho+1}$ decomposes as $(\Delta_1 + \Delta_2)^\uparrow \times (\Delta_1 + \Delta_2)^\downarrow$ under restriction to $R_{2\rho}^\uparrow \times R_{2\rho}^\downarrow$ and that, upon restriction to $R_\rho^\lambda \times R_\rho^\mu$, $\Delta_1^\uparrow$ and $\Delta_2^\uparrow$ decompose to $\Delta^\lambda \times \Delta^\mu$ if ρ is odd (number of orbitals odd) or as $\Delta_m^\uparrow \to \Delta_m^\lambda \times \Delta_2^\mu + \Delta_2^\lambda \times \Delta_{m'}^\mu$ for ρ even where $m' = 2$ or 1 as $m = 1$ or 2. Further, upon restricting R_ρ to R_3, we show that we get a minimum multiplicity of 2^{n-1} or 2^n, depending on whether n is even or odd. These duplicated R_3 representations can be separated in a natural manner.

The first branching rule is a direct consequence of the analysis of Sec. IV, i.e., that the representations Δ_1 and Δ_2 (spin up or down) are associated as follows: one with all possible states containing an even number of particles and one with all possible states containing an odd number of particles. Restricting the H operators to those of R_ρ clearly results in $\Delta_1^\uparrow$ and $\Delta_2^\uparrow$ decomposing to $\Delta^\lambda \times \Delta^\mu$ for ρ odd, but for ρ even a closer scrutiny is probably necessary. To this end, we consider a set of eigenfunctions

$$\psi_{\Delta_m^{a \text{ or } b}} \quad \text{of} \quad R_\rho^\uparrow,$$

where the subscript $\Delta_m^{a \text{ or } b}$ defines the representation by which the ψ's transform, Δ_1 if $m = 1$, and Δ_2 if $m = 2$. Here $\psi_{\Delta_m{}^a}$ is defined as being that subset of ψ_{Δ_m} that has an even number of $-\frac{1}{2}$ eigenvalues under the H operators of R_ρ^λ, i.e., a restricted set of the H operators of $R_{2\rho}^\uparrow$. Similarly, $\psi_{\Delta_m{}^b}$ is that subset having an odd number of $-\frac{1}{2}$ eigenvalues under the H operators of R_ρ^λ. However, since the H operators of $R_{2\rho}^\uparrow$ for ρ even are just the H operators of R_ρ^λ and R_ρ^μ (Sec. III), then the set of H operators of R_ρ^μ must give an even (odd) number of $-\frac{1}{2}$'s if $m = 1$ ($m = 2$) when acting on the set $\psi_{\Delta_m{}^a}$ and an odd (even) number of $-\frac{1}{2}$'s if $m = 1$ ($m = 2$) when acting on the set

$\psi_{\Delta_m{}^b}$. Thus, $\psi_{\Delta_m{}^a}$ must transform, by definition, as $\psi_{\Delta_1{}^\lambda}$ (the subscript to ψ now refers to the R_ρ^λ group) and as $\psi_{\Delta_m{}^\mu}$ under R_ρ^μ, i.e., as $\psi_{\Delta_1{}^\lambda}\psi_{\Delta_m{}^\mu}$; $\psi_{\Delta_m{}^b}$ must transform as $\psi_{\Delta_2{}^\lambda}$, by definition, and as $\psi_{\Delta_{m'}{}^\mu}$, i.e., $\psi_{\Delta_m{}^b}$ transforms as $\psi_{\Delta_2{}^\lambda}\psi_{\Delta_{m'}{}^\mu}$. Hence, the set of functions ψ_{Δ_m}, being by definition $\psi_{\Delta_m{}^a} \cup \psi_{\Delta_m{}^b}$, must transform as

$$\psi_{\Delta_1} \times \psi_{\Delta_m{}^\mu} + \psi_{\Delta_2{}^\lambda}\psi_{\Delta_{m'}{}^\mu};$$

i.e., we have established the branching rule under $R_{2\rho}^\uparrow \to R_\rho^\lambda \times R_\rho^\mu$ as $\Delta_m^\uparrow \to \Delta_1^\lambda \times \Delta_m^\mu + \Delta_2^\lambda \times \Delta_{m'}^\mu$.

It is interesting to note that for ρ even the $R_{2\rho}^\uparrow$ spin-representation label becomes redundant in that this branching rule shows that the specification of the R_ρ^λ and R_ρ^μ label determines it. This is implicit in Feneuille's work.[8]

VI. DUPLICATED R_3 REPRESENTATIONS

Upon finally restricting R_ρ to R_3, we note that, since $\mathbf{L} \propto \sum_l \mathbf{W}^{(01)}(ll)$, we are restricting the set of transformations in such a manner that only the H's which do not mix the l's (i.e., the ${}^\lambda H_{aa}^{ll}$) remain (cf. I and Sec. V) since

$$L_0 = \sum_l \sum_{a>0}^{l} a^\lambda H_{aa}^{ll} \quad \text{for } R_\rho^\lambda.$$

We now consider that subset of the eigenfunctions forming a basis for Δ_1 (or Δ_2) for R_ρ^λ, ρ even, such that any wavefunction belonging to the set has $+\frac{1}{2}$ eigenvalues for H_{aa}^{ll}. On restriction to R_3, the eigenvalue of L_0 when acting upon any of these eigenfunctions will take its maximum value equal to the maximum $L = L_M$ in the $R_\rho \to R_3$ branching rule. In fact,

$$L_M = \sum_l \sum_{a>0}^{l} \tfrac{1}{2}a = \tfrac{1}{4}\sum_l l(l+1).$$

However, since we have n H's that mix the l's the eigenvalues of which, when acting on the above subset of wavefunctions, are unrestricted, we can see that this subset contains 2^{n-1} eigenfunctions. Consequently, under $R_\rho \to R_3$ we have $\Delta_1 \to 2^{n-1}L_M + \cdots$. A similar analysis for ρ odd shows that

$$\Delta \to 2^n L_M + \cdots.$$

Thus, the multiplicity of the L_M representation of R_3 is 2^{n-1}, ρ even, or 2^n, ρ odd, and the multiplicity for any smaller allowed L will be an integral multiple of 2^{n-1} or 2^n; e.g., for $(s+p+d)^N$, under $R_9 \to R_3$ we have $[\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}] \to 2P + 2D$.

Fortunately, in this case, we have available a natural separation of duplicated R_3 representations amounting essentially to specifying, in addition to the R_ρ spin-representation label, an R_ρ weight label. A state will then be labeled by an $R_{2\rho}$ and R_ρ spin-representation label, an R_3 label, and an n-vector Γ specifying the eigenvalues of the state under the H's that mix the l's. Thus, for instance, the two D states of $(s+p+d)^N$ would be labeled as $|\Delta_a^\uparrow \Gamma = \pm\frac{1}{2} D\rangle$ (there being no R_ρ label needed here since it adds no new information).

VII. OPERATORS AND MATRIX ELEMENTS

As for equivalent electrons, any interaction can be written as sums of products of the tensor operators

$$\mathbf{W}^{(\kappa k)}(l_A l_B) = \sum_{i=1}^{N} \mathbf{w}_i^{(\kappa k)}(l_A l_B),$$

where

$$\langle slA \| w^{(\kappa k)}(l_B l_C) \| s' l D\rangle = [\kappa, k]^{\frac{1}{2}} \delta(s,s')\delta(l_A l_B)\delta(l_B l_C).$$

Two-particle interactions will involve terms of the form

$$\sum_{i\neq j} [\mathbf{w}_i^{(\kappa_1 k_1)}(l_A, l_B)\mathbf{w}_j^{(\kappa_2 k_2)}(l_C, l_D)]_Q^{(\kappa k)\bar{\kappa}}$$
$$= [\mathbf{W}^{(\kappa_1 k_1)}\mathbf{W}^{(\kappa_2 k_2)}]_Q^{(\kappa k)\bar{\kappa}}$$
$$- \delta(l_B, l_C)(-1)^{k+\kappa+l_A+l_D+2s}[\kappa_1, \kappa_2, k_1, k_2]^{\frac{1}{2}}$$
$$\times \begin{Bmatrix} k_2 & k & k_1 \\ l_A & l_B & l_D \end{Bmatrix} \begin{Bmatrix} \kappa_2 & \kappa & \kappa_1 \\ s & s & s \end{Bmatrix} (\mathbf{W}^{(\kappa k)}(l_A l_D))_Q^{\bar{\kappa}}.$$

Morrison[12] has shown that $W_{\pi q}^{(\kappa k)}(l_A l_B)$ can be written in terms of coupled products of annihilation and creation operators as

$$W_{\pi q}^{(\kappa k)}(l_A l_B) = -(a^{l_A\dagger}\tilde{a}^{l_B})_{\pi q}^{(\kappa k)},$$

where

$$\tilde{a}^{l_B}_{m_s m_l} = (-1)^{s+l_B-m_s-m_l} a_{-m_s-m_l}.$$

We may then express $W_{\pi q}^{(\kappa k)}(l_A l_B)$ in terms of our quasiparticle operators, as was done in I, giving

$$\begin{aligned} W_{\pi q}^{(\kappa k)}(l_A l_B) &= -\tfrac{1}{2}\{\langle s\tfrac{1}{2}s\tfrac{1}{2} \mid \kappa\pi\rangle\{(\lambda^{l_A}\nu^{l_B})_q^{(k)} - (\mu^{l_A}\nu^{l_B})_q^{(k)} \\ &\quad + (\mu^{l_A}\nu^{l_B})_q^{(k)} - (\lambda^{l_A}\xi^{l_B})_q^{(k)} + \delta(k,0)\delta(l_A, l_B)[l_A]^{\frac{1}{2}}\} \\ &\quad + \langle s - \tfrac{1}{2}s\tfrac{1}{2} \mid \kappa\pi\rangle\{(\nu^{l_A}\nu^{l_B})_q^{(k)} - (\xi^{l_A}\xi^{l_B})_q^{(k)} \\ &\quad + (\xi^{l_A}\nu^{l_B})_q^{(k)} - (\nu^{l_A}\xi^{l_B})_q^{(k)} + \delta(k,0)\delta(l_A, l_B)[l_A]^{\frac{1}{2}}\} \\ &\quad - \langle s\tfrac{1}{2}s - \tfrac{1}{2} \mid \kappa\pi\rangle\{(\lambda^{l_A}\lambda^{l_B})_q^{(k)} - (\mu^{l_A}\mu^{l_B})_q^{(k)} \\ &\quad + (\mu^{l_A}\lambda^{l_B})_q^{(k)} - (\lambda^{l_A}\mu^{l_B})_q^{(k)} + \delta(k,0)\delta(l_A l_B)[l_A]^{\frac{1}{2}}\} \\ &\quad - \langle s - \tfrac{1}{2}s - \tfrac{1}{2} \mid \kappa\pi\rangle(\nu^{l_A}\lambda^{l_B})_q^{(k)} - (\xi^{l_A}\mu^{l_B})_q^{(k)} \\ &\quad - (\xi^{l_A}\lambda^{l_B})_q^{(k)} - (\nu^{l_A}\mu^{l_B})_q^{(k)} + \delta(k,0)\delta(l_A l_B)[l_A]^{\frac{1}{2}}\}). \end{aligned} \tag{11}$$

For $l_A \equiv l_B$, this formula splits into two disjoint parts for k even and k odd, as in I.

It is clear that, as in I, any quasiparticle matrix element can be evaluated by standard methods of tensor operators and angular momentum recoupling,[6,7] requiring us to know only the reduced matrix elements

of the quasiparticle operators. These can be evaluated by expanding a particular component of the operator in question, say λ_0^l, and the bra and ket in terms of quasiparticle or ordinary-particle creation and annihilation operators and then applying the Wigner–Eckart theorem.

If the general expression for the eigenfunctions of H is written in its modified form (Sec. IV), viz.,

$$H\{\cdots\}\prod_{i=1}^{n} M^{\lambda}_{(\pm x\pm y)}, M^{\mu}_{\pm x\pm y}|\bar{0}\rangle,$$

then it is clear that $A_0^{l_p}$, $p \neq 2n+1$, $A = \lambda, \mu, \xi, \nu$, acts only on that $M^{\lambda}M^{\mu}$ product where x or y equals p. We write $(\Gamma_A)_i$ as

$$(-1)^{r+1+\frac{1}{2}[1-(-1)^a](l_y+l_x)}\tfrac{1}{2},$$

i.e., $\pm\frac{1}{2}$ or $\pm(-1)^{l_y+l_x}\frac{1}{2}$ if $A = \mu, \nu$ and r is an arbitrary integer, and similarly for $(\Gamma_B)_i$, with t replacing r (A, B are a pair as defined in Sec. II); then we find

$$A_0^{l_p}|(\Gamma_A)_i(\Gamma_B)_i\rangle = (-1)^{\frac{1}{2}\{r[1-(-1)^p]+p+l_p-[1-(-1)^a](l_p-r-t-1)\}} \times 2^{-\frac{1}{2}}|-(\Gamma_A)_i(\Gamma_B)_i\rangle, \qquad p \neq 2n+1,$$

if the states are properly normalized. Other quantum numbers have been suppressed as unimportant. Since the rest of the Γ vector has remained unchanged, we have the selection rule that $A_0^{l_p}$ acts only between $|\Delta_a^{\uparrow(\downarrow)}\Delta_b^A\rangle$ and $|\Delta_{a'}^{\uparrow(\downarrow)}\Delta_{b'}^A\rangle$, which in turn implies that coupled products of the form $\frac{1}{2}(A^lA^{l'})_Q^{(K)}$ can act only between $|\Delta_a^{\uparrow(\downarrow)}\Delta_b^A\Delta_c^B\rangle$ and $|\Delta_a^{\uparrow(\downarrow)}\Delta_b^A\Delta_c^B\rangle$ while products of the form $\frac{1}{2}(A^lB^{l'})_Q^{(K)}$ can act only between $|\Delta_a^{\uparrow(\downarrow)}\Delta_b^A\Delta_c^B\rangle$ and $|\Delta_a^{\uparrow(\downarrow)}\Delta_b^A\Delta_{c'}^B\rangle$.

As in I we can now write a general expression for the reduced matrix element for the particular case of maximum $L = L_M$ by noting under our separation of the duplicated L_M that

$$|(\Delta_a^{\uparrow})(\Delta_b^{\lambda}\Gamma^{\lambda}L_M - L_M), (\Delta_c^{\mu}\Gamma^{\mu}L_ML_M)\rangle = N\prod_{i}^{n} M^{\lambda}M^{\mu}|\bar{0}\rangle,$$

where the Δ and Γ are determined by the particular form of M. This gives us

$$\langle\Delta_{m'}^{\uparrow(\downarrow)}\Delta_{n'}^A - (\Gamma)_i^AL_M\| A^{l_x}\|\Delta_m^{\uparrow(\downarrow)}\Delta_n^A(\Gamma)_i^AL_M\rangle = P(-1)^{\frac{1}{2}(p+l_p)}\frac{[\frac{1}{2}(2L_M-l_p)!\,(2L_M+l_p+1)!]^{\frac{1}{2}}}{(2L_M)!}, \qquad p \neq 2n+1, \quad (12)$$

where $P = (-1)^{\frac{1}{2}\{n[1-(-1)^p]+m[1-(-1)^a]\}}$ and m and n are used as before. For ρ odd the Δ^A label is not present. In this case we find, as in I,

$$\langle\Delta_m^{\uparrow(\downarrow)}L_M\| A^{l_{2n+1}}\|\Delta_{m'}\rangle = (-1)^{\frac{1}{2}\{[1+(-1)^a]l_{2n+1}+[1+(-1)^m](l_{2n+1}+a)\}} \times \frac{[\frac{1}{2}(2L_M-l_{2n+1})!\,(2L_M+l_{2n+1}+1)!]^{\frac{1}{2}}}{(2L_M)!}. \quad (13)$$

We note that these matrix elements can be both real and imaginary so that the matrix element of a general operator will, in general, be complex. This perhaps unsatisfactory state of affairs is a natural consequence of the possible imaginary coefficients we have to take in order to form eigenfunctions of the H operators of the various groups. We consider two quasiparticle states $|QP\rangle = \sum_i a_i\,|i\rangle$ and $|QP\rangle' = \sum_k b_k\,|k\rangle$, where the $|i\rangle$ and $|k\rangle$ are ordinary states and the a_i and b_k can be both real and imaginary; thus, if W is any operator and

$$W\,|i\rangle = \sum_j W_{ij}\,|j\rangle,$$

where W_{ij} is real, then

$$\langle QP|\,W\,|QP\rangle = \sum_{ijk} b_k^*a_iW_{ij}\langle k\,|\,j\rangle = \sum_{ij} b_j^*a_iW_{ij},$$

which is, in general, complex. It is not hard to see that, if W is Hermitian and $|QP\rangle = |QP\rangle'$, then the expectation value of W is real. We find that $W_{\pi q}^{(\kappa k)}(l_il_j)$ has real matrix elements if $(-1)^{i+j} = (-1)^{l_i+l_j}$.

To conclude this section, we shall evaluate the diagonal matrix element of $W_{00}^{(02)}(sd)$ of $(s+d)^N$ with the state $|(\Delta_1^{\uparrow})(\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}\lambda})(\Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}\mu})L = 3, M_L = 3\rangle$. The Γ quantum numbers are not used as they are unnecessary for two orbitals. The only part of $W_{00}^{(02)}(sd)$ that we are interested in is $2^{-\frac{3}{2}}[(\lambda^0\lambda^2)_0^{(2)} - (\mu^0\mu^2)_0^{(2)}]$, since the rest vanishes by the selection rules above. Since

$$\begin{aligned}
\langle\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}}\|\,\lambda^0\,\|\Delta_2\Delta_2^{\lambda}\rangle &= i\sqrt{2},\\
\langle\Delta_2^{\uparrow}\Delta_{2\frac{1}{2}}^{\lambda\frac{3}{2}}\|\,\lambda^2\,\|\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}}\rangle &= \sqrt{10},\\
\langle\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}}\|\,\mu^0\,\|\Delta_2\Delta_2^{\mu}\rangle &= i\sqrt{2},\\
\langle\Delta_2^{\uparrow}\Delta_{2\frac{1}{2}}^{\mu\frac{3}{2}}\|\,\mu^2\,\|\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}}\rangle &= -\sqrt{10},
\end{aligned} \quad (14)$$

then

$$\begin{aligned}
&\langle(\Delta_1)^{\uparrow}(\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}\lambda}, \Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}\mu})L = 3, M_L = 3|\;W_{00}^{(02)}(sd)\\
&\quad\times|(\Delta_1^{\uparrow})(\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}\lambda}, \Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}\mu})L = 3, M_L = 3\rangle\\
&= +\frac{1}{2\sqrt{2}}\begin{pmatrix}3 & 2 & 3\\ -3 & 0 & 3\end{pmatrix}7(\sqrt{5})\begin{Bmatrix}\frac{3}{2} & \frac{3}{2} & 0\\ \frac{3}{2} & \frac{3}{2} & 2\\ 3 & 3 & 2\end{Bmatrix}\\
&\quad\times\begin{bmatrix}\langle\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}}\|\,\lambda^0\,\|\Delta_2^{\uparrow}\Delta_{2\frac{1}{2}}^{\lambda\frac{3}{2}}\rangle\langle\Delta_2^{\uparrow}\Delta_{2\frac{1}{2}}^{\lambda\frac{3}{2}}\|\,\lambda^2\,\|\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\lambda\frac{3}{2}}\rangle\\ -\langle\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}}\|\,\mu^0\,\|\Delta_2^{\uparrow}\Delta_{2\frac{1}{2}}^{\mu\frac{3}{2}}\rangle\langle\Delta_2^{\uparrow}\Delta_{2\frac{1}{2}}^{\mu\frac{3}{2}}\|\,\mu^2\,\|\Delta_1^{\uparrow}\Delta_{1\frac{1}{2}}^{\mu\frac{3}{2}}\rangle\end{bmatrix}\\
&= i\sqrt{\tfrac{5}{14}}.
\end{aligned}$$

VIII. CONCLUSIONS

In this paper, we have examined the application of the quasiparticle formalism to mixed configurations. As would have been anticipated from the results of our earlier work,[1] we obtain a remarkably rich classification scheme that permits the matrix elements of interactions to be calculated without the usual recourse to coefficients of fractional parentage but rather by the more common vector coupling methods. Compared with the previously treated case of equivalent electron orbitals, there has been, however, a not unexpected increase in complexity. We cite, for instance, the care that must be taken over the precise form of the eigenfunctions of the Weyl operators (Sec. IV) and the complexity of the matrix elements (Sec. VI). Where many orbitals are involved, the problem of handling the duplications arising in the decomposition of the spin representations under $R_\rho \to R_3$ will become increasingly severe, though, of course, the same holds for the conventional method.

The quasiformalism gives further insight into the mathematical structure of atomic shell theory, and in this respect its study is valuable. The fact that the quasiparticle eigenfunctions involve linear combinations of eigenfunctions defined on different numbers of particles constitutes the fundamental weakness of the scheme from a physical viewpoint. In the case of the *angular* parts of the wavefunction, this is of no moment,but, in terms of the *radial* parts, we certainly do not wish to assume their invariance with respect to the number of particles. This objection could, of course, be overcome by projecting out wavefunctions defined on a definite number of particles, but only with an ensuing increase in complexity.

ACKNOWLEDGMENT

We would like to thank P. Butler for some helpful discussions.

* Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force under AFOSR Grant No. 1275–67.

[1] M. J. Cunningham and B. G. Wybourne, J. Math. Phys. **10**, 2149 (1969).
[2] L. Armstrong, Jr., and B. R. Judd, Proc. Roy. Soc. (London), "Quasiparticles in Atomic Shell Theory," (to be published) (1970).
[3] L. Armstrong, Jr., and B. R. Judd, Proc. Roy. Soc. (London), "Atomic Structure Calculations in a Factored Shell," (to be published , 1970).
[4] C. L. B. Shudeman, J. Franklin Inst. **224**, 301 (1937).
[5] B. R. Judd, "Group Theory in Atomic Spectroscopy," in *Group Theory and its Applications*, E. M. Loebel, Ed. (Academic Press, New York, 1968).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, New York, 1963).
[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1960).
[8] S. Feneuille, J. Phys. Radium, "Traitement des configurations $(d + s)^N$ dans le formalisme des quasi-particles " (in press).
[9] S. Feneuille, J. Phys. Radium **28**, 61 (1967).
[10] P. H. Butler and B. G. Wybourne (unpublished data).
[11] H. Weyl, Math. Z. **23**, 271 (1925); **24**, 328, 377 (1925).
[12] J. C. Morrison, J. Math. Phys. **10**, 1431 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Relativistic Fields Due to a Particle in a Grounded Cylindrical Box

RICHARD L. LIBOFF
Department of Applied Physics, School of Electrical Engineering, Cornell University, Ithaca, New York 14850

(Received 5 September 1969)

The exact relativistic fields inside a perfectly conducting, closed, grounded cylindrical box of finite length, which are due to a charged particle moving down the axis in vacuum, are obtained. This permits construction of the solution for a line segment propagating down the axis and the return surface current. The fields are shown to be causal so that no field is present before the light front. The technique of solution employs superposition of Lorentz-transformed fields of the positive and negative images. The initial condition that there be no field in the box at time $t = 0$ is satisfied by adding a superposition of the cavity modes to the inhomogeneous solution. The fields and energy in the wake are also found.

1. INTRODUCTION

In this paper we obtain the exact relativistic fields inside of a perfectly conducting, closed, grounded cylindrical box, which are due to a charged point particle moving down its axis in vacuum. These are the Green's electromagnetic fields which immediately enable one to obtain the solution for a charged line segment propagating down the axis.

The solution is found to reduce to one for a particle in a semi-infinite cylinder, in the limit that the end of the cylinder is extended to infinity. In all cases the return current is obtained from the value of the magnetic field B_ϕ at the surface of the cylinder.

The solutions obtained are found to be causal so that no field is present in the region $z > ct$, where z is the axis coordinate, t is time, and c is the speed of

VIII. CONCLUSIONS

In this paper, we have examined the application of the quasiparticle formalism to mixed configurations. As would have been anticipated from the results of our earlier work,[1] we obtain a remarkably rich classification scheme that permits the matrix elements of interactions to be calculated without the usual recourse to coefficients of fractional parentage but rather by the more common vector coupling methods. Compared with the previously treated case of equivalent electron orbitals, there has been, however, a not unexpected increase in complexity. We cite, for instance, the care that must be taken over the precise form of the eigenfunctions of the Weyl operators (Sec. IV) and the complexity of the matrix elements (Sec. VI). Where many orbitals are involved, the problem of handling the duplications arising in the decomposition of the spin representations under $R_\rho \to R_3$ will become increasingly severe, though, of course, the same holds for the conventional method.

The quasiformalism gives further insight into the mathematical structure of atomic shell theory, and in this respect its study is valuable. The fact that the quasiparticle eigenfunctions involve linear combinations of eigenfunctions defined on different numbers of particles constitutes the fundamental weakness of the scheme from a physical viewpoint. In the case of the *angular* parts of the wavefunction, this is of no moment,but, in terms of the *radial* parts, we certainly do not wish to assume their invariance with respect to the number of particles. This objection could, of course, be overcome by projecting out wavefunctions defined on a definite number of particles, but only with an ensuing increase in complexity.

ACKNOWLEDGMENT

We would like to thank P. Butler for some helpful discussions.

* Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force under AFOSR Grant No. 1275–67.

[1] M. J. Cunningham and B. G. Wybourne, J. Math. Phys. **10**, 2149 (1969).
[2] L. Armstrong, Jr., and B. R. Judd, Proc. Roy. Soc. (London), "Quasiparticles in Atomic Shell Theory," (to be published) (1970).
[3] L. Armstrong, Jr., and B. R. Judd, Proc. Roy. Soc. (London), "Atomic Structure Calculations in a Factored Shell," (to be published , 1970).
[4] C. L. B. Shudeman, J. Franklin Inst. **224**, 301 (1937).
[5] B. R. Judd, "Group Theory in Atomic Spectroscopy," in *Group Theory and its Applications*, E. M. Loebel, Ed. (Academic Press, New York, 1968).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, New York, 1963).
[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1960).
[8] S. Feneuille, J. Phys. Radium, "Traitement des configurations $(d + s)^N$ dans le formalisme des quasi-particles " (in press).
[9] S. Feneuille, J. Phys. Radium **28**, 61 (1967).
[10] P. H. Butler and B. G. Wybourne (unpublished data).
[11] H. Weyl, Math. Z. **23**, 271 (1925); **24**, 328, 377 (1925).
[12] J. C. Morrison, J. Math. Phys. **10**, 1431 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Relativistic Fields Due to a Particle in a Grounded Cylindrical Box

RICHARD L. LIBOFF
Department of Applied Physics, School of Electrical Engineering, Cornell University, Ithaca, New York 14850

(Received 5 September 1969)

The exact relativistic fields inside a perfectly conducting, closed, grounded cylindrical box of finite length, which are due to a charged particle moving down the axis in vacuum, are obtained. This permits construction of the solution for a line segment propagating down the axis and the return surface current. The fields are shown to be causal so that no field is present before the light front. The technique of solution employs superposition of Lorentz-transformed fields of the positive and negative images. The initial condition that there be no field in the box at time $t = 0$ is satisfied by adding a superposition of the cavity modes to the inhomogeneous solution. The fields and energy in the wake are also found.

1. INTRODUCTION

In this paper we obtain the exact relativistic fields inside of a perfectly conducting, closed, grounded cylindrical box, which are due to a charged point particle moving down its axis in vacuum. These are the Green's electromagnetic fields which immediately enable one to obtain the solution for a charged line segment propagating down the axis.

The solution is found to reduce to one for a particle in a semi-infinite cylinder, in the limit that the end of the cylinder is extended to infinity. In all cases the return current is obtained from the value of the magnetic field B_ϕ at the surface of the cylinder.

The solutions obtained are found to be causal so that no field is present in the region $z > ct$, where z is the axis coordinate, t is time, and c is the speed of

light. At $t = 0$, the particle is at $z = 0$. The particle is assumed to move in free flight, $z = vt$, $0 \leq z \leq L$.

The wake fields present in the cavity after the particle leaves are calculated. The energy in these modes gives a divergent sum. This singularity stems from the approximation that the particle leaves (as well as enters) the cavity through a point hole. When the particle reaches the far wall, there is an abrupt cancellation of charge. This gives rise to an infinite pulse of radiation. This pulse is trapped in the box. To obviate this fictitious result, the sum over radial k-vectors is cut off at some appropriate "hole"-wavelength.

Similar problems have been considered in relation to the calculation of fields in accelerators.[1–3] These investigations usually are concerned with pipes which are infinitely long, albeit both the motion of the particle and nature of the walls are more complicated than in the present work. Our investigation is experimentally motivated by the Cornell relativistic beam. In this experiment a pulse of current $\sim$ 100,000 A, at voltages $\sim$ 300 KeV ($\beta \sim 0.78$), is propagated down a drift tube of diameter $\sim 5\frac{1}{2}$ in. and length from 1 ft to 60 ft. The pulse is of duration $\sim$ 50 nsec. Fields are measured ahead of the beam in one of a number of techniques.[4] It is hoped that the beam width can be made small enough to approximate the configuration studies below. An externally supported B_z field is contemplated to maintain the structure of the beam in vacuum. We hope to report on such measurements at a later time.

2. OUTLINE

The simplest technique for obtaining the fields of a relativistic particle is to Lorentz-transform the static Coulomb field in the frame of the particle to the "lab" frame.[5] When boundaries are present, such a technique is still possible if the static problem is soluble using the method of images. This idea was suggested by Ott and Shmoys[6] in an investigation of a particle impinging on a dielectric half space.

The static solution for a particle in a cylindrical box is solved with an infinite array of positive and negative images[7] (owing to the infinite number of reflections in the two end plates). This, together with one other observation, permits solution in terms of Lorentz transformations of the positive and negative fields. This second observation is that when the positive source charge (in the box) moves to the right, all of the positive images rigidly move with it, while all the negative images rigidly move in the opposite direction. First we calculate the static field of the positive charges and then Lorentz-transform this field to the lab (box) frame. Then we calculate the static field of the negative charges (these are all image charges) and Lorentz-transform it to the lab frame. Addition of these two sets of fields gives the inhomogeneous part of the solution to the problem. It contains the singularity at the particle position $z = vt$ and $r = 0$, and satisfies the boundary condition that the tangential component of **E** vanishes at the walls. It does not satisfy the initial condition that the fields are zero in the whole box at $t = 0$. (At $t = 0$, the positive and negative images cross. This current generates a magnetic field.) To insert this piece of data, we construct a homogeneous solution which cancels the inhomogeneous solution at $t = 0$. The homogeneous solution is a superposition of the TM modes of the cavity. (The inhomogeneous solution is a TM wave.)

The addition of the homogeneous solution also guarantees that the whole solution is causal so that all fields vanish for $z > ct$. Furthermore, for times $ct < L$ (the box is of length L) the fields are those of a semi-infinitely long ($0 \leq z$) cylinder. This must be the case, since the fields are not influenced by the far wall before $t = L/c$.

The chief assumption in the analysis is that the particle moves in free flight. This assumption is used in the analysis to obviate a singularity in the energy of the wake fields. For an ultrarelativistic electron, it is found that free flight is maintained, provided the electron leaves the cavity through a hole of diameter greater than 10^{-13} cm (1 F).

3. ANALYSIS

A. The Inhomogeneous Component

The static potential due to a point charge q on the axis at $z = 0$ interior to an infinitely long grounded cylinder, with perfectly reflecting walls, is

$$\Phi(r, z) = \frac{2q}{r_0^2} \sum_{j=1}^{\infty} \frac{e^{-k_j|z|} J_0(k_j r)}{k_j [J_1(k_j r_0)]^2} . \tag{1}$$

The radius of the tube is r_0. The zeroth- and first-order Bessel functions are J_0 and J_1, respectively. The above expression contains the proper singularity at $z = 0$, through the δ-function representation

$$\delta(r) = \frac{1}{\pi} \sum_j \frac{J_0(k_j r)}{r_0^2 [J_1(k_j r_0)]^2} ,$$

where

$$\int_0^{r_0} \delta(r) 2\pi r \, dr = 1, \quad J_0(k_j r_0) = 0.$$

It follows that

$$2\pi q \delta(r) = \left(\frac{\partial \phi_>}{\partial z} - \frac{\partial \phi_<}{\partial z} \right)_{z=0}$$

where $\gtrless$ denotes $z \gtrless 0$, respectively.

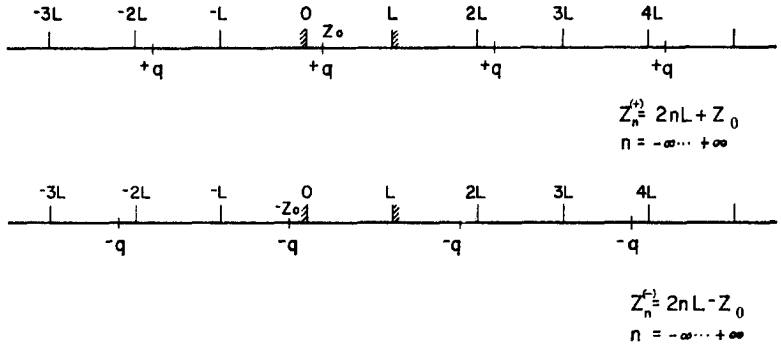


FIG. 1. Image locations for a particle in a cylindrical box.

For a closed cylindrical box with end plates at $z = 0$ and $z = L$ and a test charge q at $z = z_0$, an infinite array of image charges are induced exterior to the box along the axis. Their relative positions are depicted in Fig. 1. The static solution to this problem is obtained by replacing the exponential z-dependent part of the solution in (1) with the corresponding potential due to all the images. There are positive images at $z_n = 2nL + z_0$ and negative images at $z_n = 2nL - z_0$, where n is an integer running from $-\infty$ to $+\infty$.

For the relativistic problem we must separate the field into that due to the positive images (plus source) and that due to the negative images. The z-dependent part of the solution due to the positive charges appears as

$$\psi_j^{(+)} = \sum_{n=1}^{\infty} e^{-k_j(2nL-z_0+z)} + \sum_{n=1}^{\infty} e^{-k_j(2nL+z_0-z)} + e^{-k_j|z-z_0|} = \frac{\cosh k_j(L - |z - z_0|)}{\sinh k_j L}.$$

Similarly, the z-dependent part of the solution due to the negative charges appears as

$$\psi_j^{(-)} = \frac{\cosh k_j(L - |z + z_0|)}{\sinh k_j L}.$$

Superposition of the total solutions generated by $\psi^{(+)}$ and $\psi^{(-)}$ (as given by these latter two expressions) reproduces the static solution.[7]

If the test charge is moving on the axis inside the box, all of the positive images move with it. Let the frame where these charges are at rest be O'. At $t = 0$ the source charge is at the origin of O' and, furthermore, at this time the O' frame and the lab frame O are coincident.

The potential due to all the positive charges in O' (where the source is *always* at the origin) is

$$\Phi^{(+)\prime}(r'z') = \frac{2q}{r_0^2} \sum_j \frac{\cosh k_j(L' - |z'|)}{\sinh k_j L'} \frac{J_0(k_j r)}{k_j J_1^2}.$$

In this latter formula and expressions to follow, J_1^2 is written for $J_1^2(k_j r_0)$.

The corresponding components of the electric field are

$$E_z^{(+)\prime} = \frac{2q}{r_0^2} \sum_j \frac{\sinh k_j(L' - |z'|)}{\sinh k_j L'} \frac{J_0(k_j r)}{k_j J_1^2} \operatorname{sgn} z',$$

$$E_r^{(+)\prime} = \frac{2q}{r_0^2} \sum_j \frac{\cosh k_j(L' - |z'|)}{\sinh k_j L'} \frac{J_1(k_j r)}{J_1^2}.$$

Next we must write the rhs of the above expressions in terms of the coordinates of the lab frame O. The transformation for the positive chain is $z' = \gamma(z - vt)$ and $z' = \gamma(z + vt)$ for the negative chain (see Fig. 2). The (proper) length between image charges in O' is L' while the length between these same charges in O is L, so that $L' = \gamma L$. There results

$$E_z^{(+)\prime} = \frac{2q}{r_0^2} \sum \frac{\sinh \gamma k_j(L - |z - vt|)}{\sinh \gamma k_j L} \frac{J_0(k_j r)}{k_j J_1^2} \times \operatorname{sgn}(z - vt), \quad (2)$$

$$E_r^{(+)\prime} = \frac{2q}{r_0^2} \sum \frac{\cosh \gamma k_j(L - |z - vt|)}{\sinh \gamma k_j L} \frac{J_1(k_j r)}{J_1^2},$$

$$\gamma \equiv (1 - \beta^2)^{-\frac{1}{2}}, \quad \beta \equiv v/c.$$

When observed in the lab frame, these fields become

$$E_z^{(+)} = E_z^{(+)\prime}, \quad E_r^{(+)} = \gamma E_r^{(+)\prime}, \quad B_\phi^{(+)} = \gamma\beta E_r^{(+)\prime}. \quad (3)$$

In similar manner, we obtain for the fields of the negative charges, expressed in the coordinates of O,

$$E_z^{(-)\prime} = -\frac{2q}{r_0^2} \sum_j \frac{\sinh \gamma k_j(L - |z + vt|)}{\sinh \gamma k_j L} \frac{J_0(k_j r)}{J_1^2} \times \operatorname{sgn}(z - vt),$$

$$E_r^{(-)\prime} = -\frac{2q}{r_0^2} \sum_j \frac{\cosh \gamma k_j(L - |z + vt|)}{\sinh \gamma k_j L} \frac{J_1(k_j r)}{J_1^2}. \quad (4)$$

Transforming these fields to the lab frame gives

$$E_z^{(-)} = E_z^{(-)\prime}, \quad E_r^{(-)} = \gamma E_r^{(-)\prime}, \quad B_\phi^{(-)} = -\gamma\beta E_r^{(-)\prime}. \quad (5)$$

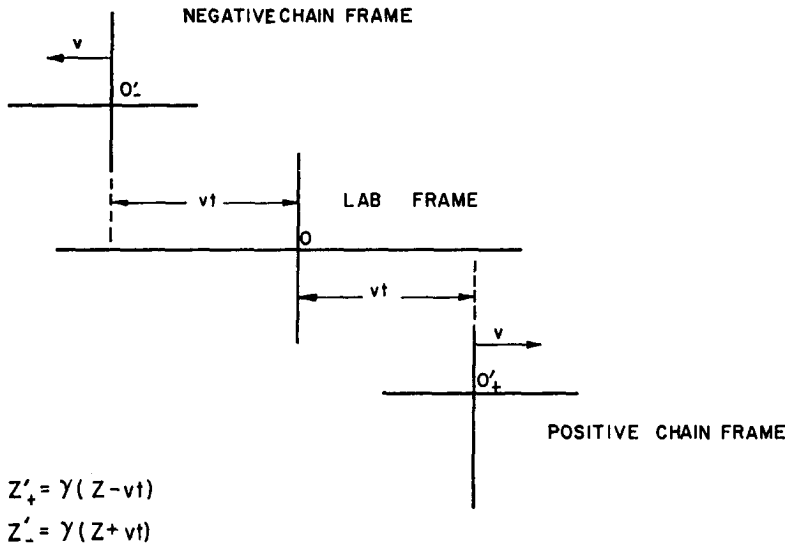


FIG. 2. Positive and negative image frames.

Superposing these fields with those of the positive chain, Eq. (3) gives the inhomogeneous component of the total solution

$$E_z^{\mathrm{i}} = \frac{4q}{r_0^2}\sum_j \frac{J_0(k_j r)}{J_1^2}\,\frac{\sinh \Omega_j t \cosh K_j(1 - z/L)}{\sinh K_j},$$

$$E_r^{\mathrm{i}} = \frac{4q\gamma}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2}\,\frac{\sinh \Omega_j t \sinh K_j(1 - z/L)}{\sinh K_j}, \quad (6a)$$

$$B_\phi^{\mathrm{i}} = \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2}\,\frac{\cosh \Omega_j t \cosh K_j(1 - z/L)}{\sinh K_j},$$

for $z > vt$. (The superscript "i" denotes inhomogeneous.) "Behind" the particle for $z < vt$, these fields become

$$E_z^{\mathrm{i}} = -\frac{4q}{r_0^2}\sum_j \frac{J_0(k_j r)}{J_1^2}\,\frac{\sinh (K_j z/L) \cosh (K_j - \Omega_j t)}{\sinh K_j},$$

$$E_r^{\mathrm{i}} = \frac{4q\gamma}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2}\,\frac{\sinh (K_j z/L) \sinh (K_j - \Omega_j t)}{\sinh K_j}, \quad (6b)$$

$$B_\phi^{\mathrm{i}} = \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2}\,\frac{\cosh (K_j z/L) \cosh (K_j - \Omega_j t)}{\sinh K_j}.$$

This solution is a segment of the periodic field present in the "extended lab frame." (See Fig. 3.) At any fixed position, the time period is $2L/v$, while the length period is $2L$. At the instant when the images cover one

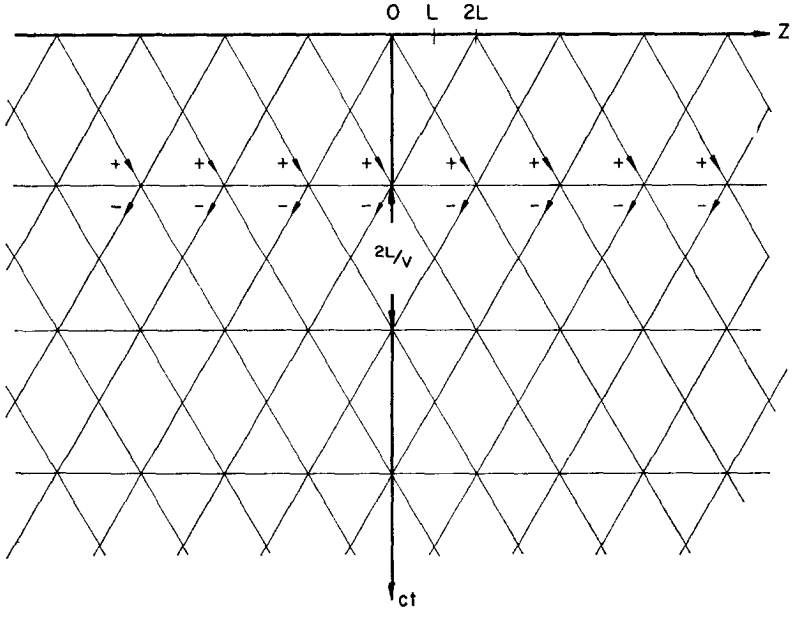


FIG. 3. The extended lab frame.

another ($vt = nL$) the electric field vanishes while the magnetic field is minimum midplane between images and maximum in the image planes.

In Eq. (6), we have written

$$\Omega_j \equiv \gamma k_j v, \quad K_j \equiv \gamma k_j L.$$

Since **B** does not vanish at $t = 0$, it follows that the solution constructed from images as outlined above, while satisfying boundary and singularity conditions, does not satisfy the initial data that $\mathbf{E} = \mathbf{B} = 0$ for $t \leq 0$ everywhere in the box. Insofar as the solution so constructed incorporates the presence of the source, it is a particular solution (viz., to the wave equations). The total solution to our initial-value problem is obtained by adding to this particular integral a solution to the homogeneous wave equations. This final form gives $\mathbf{E} = \mathbf{B} = 0$ at $t = 0$ and is singular on $z = vt$; $E_{\|}$ vanishes at the walls and is causal. It follows that it is the correct Green's solution to the stated problem.

B. The Homogeneous and Total Green's Function

Insofar as $B_z = 0$ in the inhomogeneous solution (6), it represents a TM wave. It follows that a superposition of TM waves must be added thereto to give the desired null effects. These are TM_{0jp} modes, j referring to wavenumber k_j and p to z-harmonic dependence. The zero relates to Bessel-function order. The eigenfrequencies ω_{jp} which accompany these modes are

$$\omega_{jp}^2 = c^2[k_j^2 + (p\pi/L)^2].$$

Superposition of these modes give the fields

$$\begin{aligned}
\mathfrak{E}_z &= \sum_p \sum_j \epsilon_{pj} J_0(k_j r) \cos\left(\frac{p\pi z}{L}\right) e^{-i\omega_{pj}t}, \\
\mathfrak{E}_r &= \sum_p \sum_j \epsilon_{pj} \left(\frac{p\pi}{Lk_j}\right) J_1(k_j r) \sin\left(\frac{p\pi z}{L}\right) e^{-i\omega_{pj}t}, \qquad (7) \\
\mathfrak{B}_\phi &= \sum_p \sum_j \epsilon_{pj} \left(\frac{i\omega_{pj}}{ck_j}\right) J_1(k_j r) \cos\left(\frac{p\pi z}{L}\right) e^{-i\omega_{pj}t}.
\end{aligned}$$

The coefficient ϵ_{pj} is to be determined. Both the real and imaginary components of these fields, respectively, are solutions to the homogeneous wave equations. At $t = 0$, Im $\mathfrak{B}_\phi \equiv B_\phi^{\mathrm{h}}$ ("h" denotes homogeneous) appears as

$$B_\phi^{\mathrm{h}} = -\sum_p \sum_j \epsilon_{pj} \left(\frac{\omega_{pj}}{ck_j}\right) J_1(k_j r) \cos\left(\frac{p\pi z}{L}\right).$$

Comparison with Eq. (6) indicates that ϵ_{pj} must be chosen so that

$$\frac{4q\gamma\beta}{J_1^2 r_0^2} \frac{\cosh K_j(1 - z/L)}{\sinh K_j} = \sum_p \epsilon_{pj} \frac{\omega_{pj}}{ck_j} \cos\frac{p\pi z}{L}.$$

To solve for ϵ_{pj}, we employ the Fourier decomposition

$$\frac{\cosh K(1 - z/L)}{K \sinh K} = \sum_{p=-\infty}^{\infty} \frac{\cos p\pi z/L}{(\pi p)^2 + K^2}.$$

To validate this representation, we rewrite the summation as an integral in the complex p plane:

$$\sum_{p=-\infty}^{\infty} \frac{\cosh(\pi p z/L)}{(\pi p)^2 + K^2} = \frac{1}{2\pi i} \int_C \frac{\cos(\pi p z/L) \cot \pi p \, dp}{(\pi p)^2 + K^2} \equiv I.$$

The curve C encircles the real p axis as shown in Fig. 4(a). Using the cosine addition law, we obtain

$$I = \frac{1}{2\pi i} \int_C \frac{\cos \pi p(1 - z/L) \, dp}{[(\pi p)^2 + K^2] \sin \pi p} - \frac{1}{2\pi i} \int_C \frac{\sin(\pi p z/L) \, dp}{(\pi p)^2 + K^2}.$$

Since the second integrand is an analytic function along the whole real axis, its integral vanishes. To evaluate the first integral, we distort the contour C into C_1 and C_2 as shown in Fig. 4(b). This gives the two residues from the poles at $\pi p = \pm iK$ which add to yield the desired result.

For ϵ_{pj} we then have

$$\frac{\omega_{pj}\epsilon_{pj}}{ck_j} = \frac{4q\gamma\beta K_j}{J_1^2 r_0^2[(\pi p)^2 + K_j^2]},$$

where $-\infty \leq p \leq +\infty$.

Substituting this value for ϵ_{pj} into Eq. (7), taking the imaginary part thereof, and adding the resultant fields to the inhomogeneous solution, Eq. (6), gives

$$\begin{aligned}
E_z &= \frac{4q}{r_0^2} \sum_{j=0}^{\infty} \frac{J_0(k_j r)}{J_1^2} \left(\frac{\sinh \Omega_j t \cosh K_j(1 - z/L)}{\sinh K_j} - \sum_{p=-\infty}^{\infty} \frac{K_j \Omega_j \sin \omega_{pj} t \cos \pi p z/L}{\omega_{pj}[(\pi p)^2 + K_j^2]} \right) \\
&\equiv \frac{4q}{r_0^2} \sum_j \frac{J_0(k_j r)}{J_1^2} U_j^{>}(z, t, v, L), \\
E_r &= \frac{4q\gamma}{r_0^2} \sum_{j=1}^{\infty} \frac{J_1(k_j r)}{J_1^2} \left(\frac{\sinh \Omega_j t \sinh K_j(1 - z/L)}{\sinh K_j} - \sum_{p=-\infty}^{+\infty} \frac{\pi p \Omega_j \sin \omega_{pj} t \sin \pi p z/L}{\omega_{pj}[(\pi p)^2 + K_j^2]} \right) \qquad (8a) \\
&\equiv \frac{4q\gamma}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2} V_j^{>}(z, t, v, L), \\
B_\phi &= \frac{4q\gamma\beta}{r_0^2} \sum_{j=0}^{\infty} \frac{J_1(k_j r)}{J_1^2} \left(\frac{\cosh \Omega_j t \cosh K_j(1 - z/L)}{\sinh K_j} - \sum_{p=-\infty}^{\infty} \frac{K_j \cos \omega_{pj} t \cos \pi p z/L}{[(\pi p)^2 + K_j^2]} \right) \\
&\equiv \frac{4q\gamma\beta}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2} W_j^{>}(z, t, v, L).
\end{aligned}$$

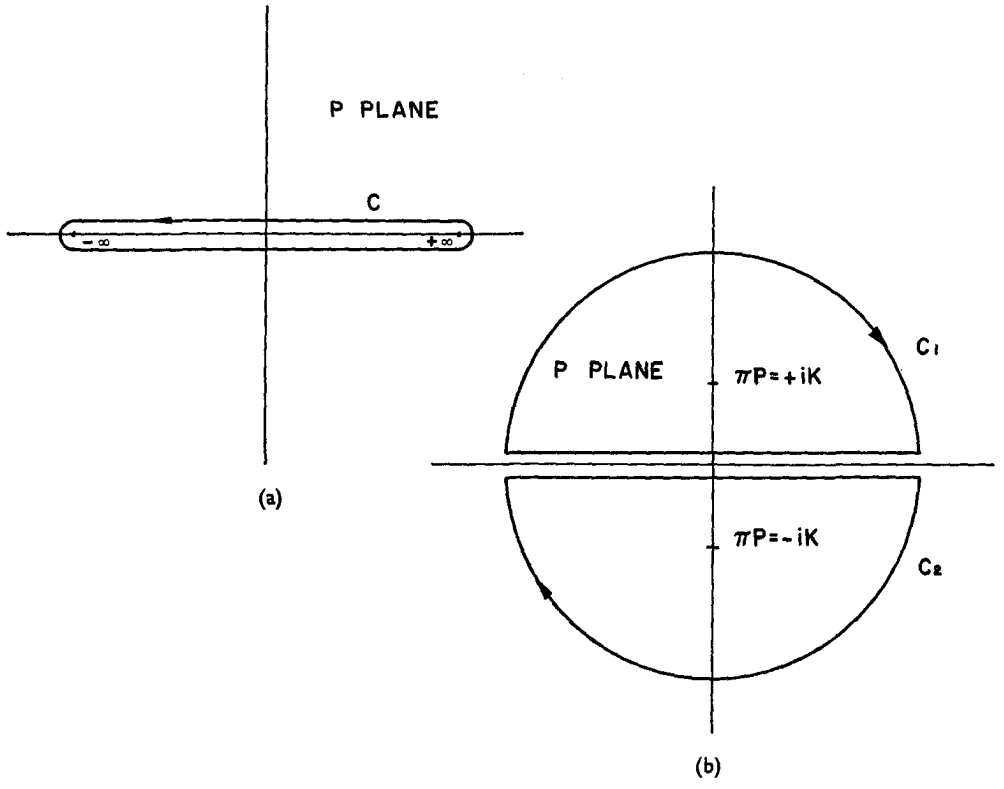


FIG. 4. The contours C, C_1, and C_2.

The second identification in each case above serves to define the (z, t) dependent forms, U, V, and W. These are the fields for $z > vt$.

For $z < vt$, merely substitute the inhomogeneous terms in the above expressions with those given in Eq. (6b). The homogeneous component of the total solution (for $z \leq L$) is then continuous across the plane $z = vt$,while the inhomogeneous component includes the singularity at the particle position. For $z < vt$, there results

$$E_z = -\frac{4q}{r_0^2}\sum_j \frac{J_0(k_j r)}{J_1^2}\left(\frac{\sinh(K_j z/L)\cosh(K_j - \Omega_j t)}{\sinh K_j} + \sum_p \frac{K_j \Omega_j \sin \omega_{pj} t \cos \pi p z/L}{\omega_{pj}[(\pi p)^2 + K_j^2]}\right) \equiv -\frac{4q}{r_0^2}\sum_j \frac{J_0(k_j r)}{J_1^2} U^<,$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2}\left(\frac{\sinh(K_j z/L)\sinh(K_j - \Omega_j t)}{\sinh K_j} - \sum_p \frac{\pi p \omega_j \sin \omega_{pj} t \sin \pi p z/L}{\omega_{pj}[(\pi p)^2 + K_j^2]}\right) \equiv \frac{4q\gamma}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} V^<,$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2}\left(\frac{\cosh(K_j z/L)\cosh(K_j - \Omega_j t)}{\sinh K_j} - \sum_p \frac{K_j \cos \omega_{pj} t \cos \pi p z/L}{[(\pi p)^2 + K_j^2]}\right) \equiv \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} W^<. \tag{8b}$$

Equations (8) with $q = 1$ are the components of the relativistic Green's tensor field $G_{\mu\nu}(\mathbf{x}, t)$ (this notation will be used below). The components of G are the values of the fields at $\mathbf{x}$, t due to a point unit charge moving with v along the axis of a cylindrical grounded box, which was at $z = 0$ at $t = 0$.

C. Causality

To show that the components of G as given by Eq. (8) are all causal, the summation over p is converted to a contour integration in the p plane (after multiplying by cot πp). The following addition formulas are used:

$$\cos \pi p\left(1 - \frac{z}{L}\right) = \cos \frac{\pi p z}{L} \cos \pi p + \sin \frac{\pi p z}{L} \sin \pi p,$$

$$\sin \pi p\left(1 - \frac{z}{L}\right) = -\sin \frac{\pi p z}{L} \cos \pi p + \cos \frac{\pi p z}{L} \sin \pi p.$$

In each case it is found that U, W, and V vanish for $z > ct$. For $z \leq ct$, in the general case a simplifying contour distortion is not evident and we must work with the summations over p in their generic form.

To illustrate the causal property of the solution (8), we consider B_ϕ. We wish to show that

$$W = 0, \quad z > ct,$$

or equivalently (deleting the j index)

$$\frac{\cosh \Omega t \cosh K(1 - z/L)}{K \sinh K} = \sum_{p=-\infty}^{+\infty} \frac{\cos \omega_p t \cos \pi pz/L}{(\pi p)^2 + K^2} = \frac{1}{2\pi i} \int_C dp \frac{\cos \omega_p t \cos (\pi pz/L) \cot \pi p}{(\pi p)^2 + K^2} \equiv \Lambda,$$

where the contour C is depicted in Fig. 4(a). From the cosine law the latter integral is decomposed into

$$\Lambda = \frac{1}{2\pi i} \int_C \frac{\cos \omega_p t \cot \pi p(1 - z/L)\, dp}{(\pi p)^2 + K^2} - \frac{1}{2\pi i} \int_C \frac{\cos \omega_p t \sin \pi pz/L}{(\pi p)^2 + K^2}.$$

The second integrand is regular in the domain enclosed by C so that only the first integral contributes to Λ. Expanding the integrand of this first integral gives

$$\Lambda = \frac{1}{4\pi i} \int_C \frac{\cos [\omega_p t + \pi p(1 - z/L)] + \cos [\omega_p t - \pi p(1 - z/L)]}{[(\pi p)^2 + K^2] \sin \pi p} dp$$
$$\equiv \frac{1}{4\pi i} \int_C \tilde{\Lambda}\, dp.$$

In the limit that $\operatorname{Im} p \equiv p' \to \pm\infty$, we have that $\omega_p \to c\pi p'/L$ and that

$$|[(\pi p)^2 + K^2]\, \tilde{\Lambda}| \to 2\, |e^{-|\pi p|}\{\exp [-(\pi p'/L)(z - ct) + \pi p'] + \exp [(\pi p'/L)(z - ct) - \pi p'] + \exp [(\pi p'/L)(z + ct) - \pi p'] + \exp [-(\pi p'/L)(z + ct) + \pi p']\}|.$$

For $p' \to \pm\infty$ the rhs of this latter expression goes to zero providing $ct < z$ and $2L > z + ct$. Both inequalities are satisfied for $ct < z < L$. It follows that for these values of z and t the curve C may be distorted into C_1 and C_2 as depicted in Fig. 4(b). This picks up the two residues at $K = \pm i\pi p$ to yield

$$\Lambda = \frac{1}{2\pi i} \int_C \frac{\cos \omega_p t \cos \pi p(1 - z/L)}{[(\pi p)^2 + K^2] \sinh \pi p}$$
$$= \frac{1}{2\pi i} \left(\int_{C_1} + \int_{C_2} \right)$$
$$= \frac{\cosh \Omega t \cosh K(1 - z/L)}{K \sinh K}.$$

It follows that $W = 0$ for $z > ct$. Similar constructions hold for U and V.

D. The Semi-Infinite and Completely Infinite Pipe

The solution to the above problem as given by Eq. (8) reduces to a very simple form in the interval $t < L/c$. In this interval the pulse is not influenced by the forward wall at $z = L$ and must reduce to that owing to a moving charged particle in a semi-infinite tube with walls at $z = 0$ and $z = \infty$.

In this limit as $L \to \infty$, the inhomogeneous contributions in U, V, and W become

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix}_i \to e^{-\gamma kz} \begin{pmatrix} \sinh \Omega t \\ \sinh \Omega t \\ \cosh \Omega t \end{pmatrix}.$$

In the homogeneous p-summations, the discrete variable p becomes the continuous variable ξ through the transformation

$$\pi p/L \to \xi, \quad \pi/L \to d\xi.$$

There results

$$\frac{\pi v}{\Omega} \begin{pmatrix} U \\ V \\ W \end{pmatrix}_h \to \begin{pmatrix} \Omega \int_{-\infty}^{\infty} \frac{d\xi \sin \omega t \cos \xi z}{\omega(\xi^2 + \gamma^2 k^2)} \\ v \int_{-\infty}^{\infty} \frac{\xi\, d\xi \sin \omega t \sin \xi z}{\omega(\xi^2 + \gamma^2 k^2)} \\ \int_{-\infty}^{\infty} \frac{d\xi \cos \omega t \cos \xi z}{\xi^2 + \gamma^2 k^2} \end{pmatrix}.$$

Rewriting the trigonometric products in exponential form and closing the integration along the real p axis with the upper or lower semicircle, depending on which of these the integrand vanishes, gives the following:

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix}_h = -\begin{pmatrix} U \\ V \\ W \end{pmatrix}_i, \quad z > ct,$$

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix}_h = 0, \qquad z \leq ct.$$

Only the inhomogeneous solution survives in the limit $L \to \infty$, and we obtain for $vt < z \leq ct$ (ahead of

the particle)

$$E_z = \frac{4q}{r_0^2}\sum \frac{J_0(k_j r)}{J_1^2} e^{-\gamma k_j z} \sinh \Omega_j t,$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum \frac{J_1(k_j r)}{J_1^2} e^{-\gamma k_j z} \sinh \Omega_j t, \qquad (9a)$$

$$B_\phi = \frac{4q\gamma\beta}{r_2^0}\sum \frac{J_1(k_j r)}{J_1^2} e^{-\gamma k_j z} \cosh \Omega_j t.$$

For $z > ct$ the fields vanish. Behind the particle ($0 \leq z < vt$), we find

$$E_z = -\frac{4q}{r_0^2}\sum_j \frac{J_0(k_j r)}{J_1^2} e^{-\Omega_j t} \sinh \gamma k_j z,$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} e^{-\Omega_j t} \sinh \gamma k_j z, \qquad (9b)$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} e^{-\Omega_j t} \cosh \gamma k_j z.$$

When the particle is very far removed from the after wall of the tube (which in the present case is the *only* wall of the tube) i.e., in the limit $\Omega t \gg 1$ and $kz \gg 1$, one obtains the sharply peaked wave

$$E_z = \frac{4q}{r_0^2}\sum_j \frac{J_0(k_j r)}{J_1^2} e^{-\gamma k_j |z - vt|} \operatorname{sgn}(z - vt),$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} e^{-\gamma k_j |z - vt|}, \qquad (10)$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} e^{-\gamma k_j |z - vt|}.$$

These are the exact fields of a particle moving in a pipe which stretches from $z = -\infty$ to $z = +\infty$. They are very simply obtained through a single Lorentz transformation of a particle in an infinite cylinder.

4. APPLICATIONS

A. The Wake Fields

At $vt = L$, the particle leaves the cylindrical box. To obtain the fields in the box for $vt > L$, we do the following. The fields "behind" the particle (for $vt \leq L$) are given by Eqs. (8b). Symbolically, we write these fields as

$$G = G^{\mathrm{i}} + G^{\mathrm{h}},$$

where, as before, "i" denotes inhomogeneous and "h" denotes homogeneous. Let us construct a field $\tilde{G}$, which satisfies the following initial-value problem:

$$\Box^2 \tilde{G} = 0, \quad \text{for} \quad vt > L,$$

and

$$\tilde{G}(r, vt, z) = G^{\mathrm{i}}(r, vt, z), \quad \text{at} \quad vt = L.$$

It then follows that

$$\Box^2[\tilde{G} + G^{\mathrm{h}}] = 0, \quad \text{for} \quad vt > L$$

(since G^{h} is a solution to the homogeneous equation for all time) and that

$$[\tilde{G} + G^{\mathrm{h}}]_{vt=L} = [G^{\mathrm{i}} + G^{\mathrm{h}}]_{vt=L}.$$

We may conclude that the field $\tilde{G} + G^{\mathrm{h}}$, so constructed, satisfies the homogeneous wave equation and the "initial" conditions at $vt = L$, whence it is the desired solution for the wake domain.

The field G^{i} is given by Eq. (6b) with $vt = L$. Only B_ϕ^{i} survives:

$$B_\phi^{\mathrm{i}} = \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} \frac{\cosh K_j z/L}{\sinh K_j}.$$

From our previous Fourier decomposition, we have

$$\frac{\cosh K}{K \sinh K} = \sum \frac{\cos \pi p z/L}{(\pi p)^2 + L^2}.$$

It follows that

$$\frac{\cosh Kz/L}{K \sinh K} = \sum \frac{\cos \pi p(1 - z/L)}{(\pi p)^2 + K^2}$$

$$= \sum \frac{(-)^b \cos \pi p z/L}{(\pi p)^2 + K^2}.$$

We need only recall the following to obtain the desired result. If $\Phi(x, t)$ is a solution to the homogeneous wave equation, so is $\Phi(x, t + a)$, where a is an arbitrary constant. It follows that, if Eq. (7) represents a homogeneous solution, then so do these same expressions with t replaced by $(t - L/v)$. The coefficients ϵ_{pj} which enter follow from the above Fourier decomposition:

$$-\frac{\omega_{pj}\epsilon_{pj}}{ck_j} = \frac{4q\gamma\beta(-)^b K_j}{J_1^2 r_0^2[(\pi p)^2 + K_j^2]}.$$

Combining these fields with G^{h} [the second terms in each of Eq. (8a)] gives the closed fields in the wake domain:

$$E_z = \frac{4q}{r_0^2}\sum_j \sum_p \frac{K_j \Omega_j \cos \pi p z/L}{\omega_{pj}[(\pi p)^2 + K_j^2]} \frac{J_0(k_j r)}{J_1^2}$$
$$\times [-\sin \omega_{pj} t + (-1)^b \sin \omega_{pj}(t - L/v)],$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum_j \sum_p \frac{p\pi \Omega_j \sin \pi p z/L}{\omega_{pj}[(\pi p)^2 + K_j^2]} \frac{J_1(k_j r)}{J_1^2}$$
$$\times [-\sin \omega_{pj} t + (-1)^p \sin \omega_{pj}(t - L/v)], \qquad (11)$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2}\sum_j \sum_p \frac{K_j \cos \pi p z/L}{[(\pi p)^2 + K_j^2]} \frac{J_1(k_j r)}{J_1^2}$$
$$\times [-\cos \omega_{pj} t + (-)^p \cos \omega_{pj}(t - L/v)]$$
$$\equiv \frac{4q\gamma\beta}{r_0^2}\sum_j \frac{J_1(k_j r)}{J_1^2} \overline{W}_j.$$

This latter identity serves to define the function $\overline{W}_j$, to be used below. These fields given in Eq. (11) represent the fields in the cavity at time t, due to a particle which was at the far wall at $t = L/v$.

B. Wake Energy

In this section we calculate the energy in the wake fields (11). Insofar as the energy in the cavity is conserved, the time average of the energy is the actual energy. If we call the time factors of E_z and E_r, S, and that of B_ϕ, C, then

$$\langle S^2\rangle = \langle C^2\rangle = 1 - (-1)^p \cos\frac{L\omega_{pj}}{v}.$$

Furthermore, the energy in the electric field equals the energy in the magnetic field:

$$\tfrac{1}{2}U = \tfrac{1}{4}\iint (E_r^2 + E_z^2)\, dz\, r\, dr = \tfrac{1}{4}\iint B_\phi^2\, dz\, r\, dr$$
$$= L\left(\frac{q\gamma\beta}{r_0}\right)^2 \sum_{p=-\infty}^{+\infty}\sum_{j=1}^{+\infty} \frac{[1-(-1)^p\cos(\omega_{pj}L/v)]K_j^2}{J_1^2(k_j r_0)[(\pi p)^2 + K_j^2]^2} > 0. \tag{12}$$

Since the cos $(\omega L/v)$ term does not exceed unity and enters in an oscillating series, it follows that the sum over the first term (unity) is greater than the sum over the cos term, whence

$$U < 4L\left(\frac{q\gamma\beta}{r_0}\right)^2 \sum_p \sum_j \frac{K_j^2}{J_1^2[(\pi p)^2 + K_j^2]^2}. \tag{13}$$

The p summation may be evaluated as follows. Define

$$M \equiv \sum_{p=-\infty}^{+\infty} \frac{1}{[(\pi p)^2 + K^2]^2} = \frac{1}{2\pi i}\int_C \frac{\cot \pi p\, dp}{[(\pi p)^2 + K^2]^2},$$

where the contour C is depicted in Fig. 4. Since

$$\lim_{p\to\pm i\infty} |\cot \pi p| = 1,$$

the contour C may be distorted into C_1 and C_2 to yield

$$M = \frac{2}{(2K)^2\pi \sinh^2 K}\left[1 + \frac{\sinh K \cosh K}{K}\right].$$

It follows that

$$U < \frac{2L}{\pi}\left(\frac{q\gamma\beta}{r_0}\right)^2 \sum_j \frac{1}{J_1^2 \sinh^2 K_j}\left(1 + \frac{\sinh K_j \cosh K_j}{K_j}\right). \tag{14}$$

In the limit as $j \to \infty$ the second term in the summand goes to a constant and gives a divergent sum. This singularity [evident from Eq. (12)] stems from idealization that the point particle q leaves (as well as enters) through a point hole. When the particle reaches the far wall, it coalesces with its (nearest) image and stops. The resulting singular pulse is trapped in the box. This conclusion is consistent with Ott's[8] calculation for the transition radiation problem[9] in which a particle is incident on a grounded plane. After the particle passes through the plane, a hemispherical wave propagates away from the wall carrying zero field behind it and the previous field in front of it. The fields at the wave surface are singular. In the similar problems with a hole in the plane[10,11] this singularity is obviated. Similarly, in the problem considered herein, if the series above, written in the form $U < \sum_j U_j$, is cut off at $r_0 k_j \simeq 2\pi r_0/d$, then the sum is finite. This would be, roughly, the energy deposited in a finite cylindrical box with holes of diameter d in its end plates.

The large-order zeros of J_0 go as $r_0 k_j \simeq \pi j$ so that there are $j \sim 2r_0/d$ terms in the cut-off series of Eq. (12). It follows that an upper estimate of this series is given by

$$U < jU_{j_{\max}} = j\frac{2\gamma q^2\beta^2}{\pi r_0} = \frac{2\gamma q^2\beta^2}{\pi d}.$$

Our assumption that the particle does not lose too much of its energy to the stimulated wake fields will be valid if

$$\frac{U}{(\gamma - 1)m_0c^2} \ll 1.$$

For an electron with $\beta \simeq 1$ one obtains

$$r_0 \gg d \gg 10^{-13}\text{ cm},$$

which is easily satisfied in most practical cases. The left inequality insures that the hole is, at most, a small perturbation in the included analysis.

C. The Charged Line Segment

In this section we consider a line charge of length $b < L$ and charge q. There are four relevant epochs. (See Fig. 5.) In epoch (1), the segment is partially in the cavity ($vt < b$). In epoch (2), the segment is completely in the cavity ($b \le vt \le L$). In epoch (3), the segment is leaving the cavity ($L - b < vt - b < L$). In epoch (4), the pulse has completely left the cavity ($vt - b \ge L$).

In the first two epochs we only need the Green's fields given by Eq. (8). We recall that these fields (with $q = 1$) are those at $(\mathbf{x}, t)$ due to a point unit charge which entered the cavity at $t = 0$. In epoch (1),

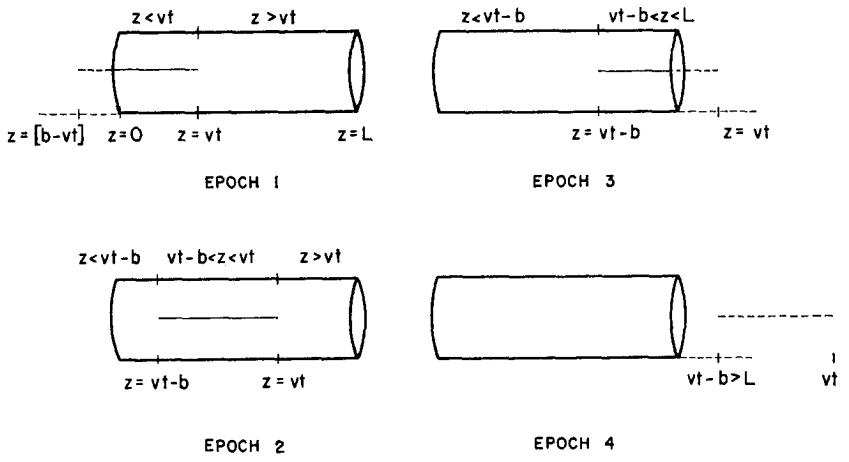


FIG. 5. Relevant time intervals for the line charge problem.

one obtains

$$F_{\mu\nu} = \frac{qv}{b}\int_0^t G_{\mu\nu}(x, t-\eta)\,d\eta$$
$$\equiv \frac{qv}{b}\int_0^t G_{\mu\nu}(x,\tau)\,d\tau, \quad t \le b/v,$$
$$\tau \equiv t - \eta.$$

The Lagrangian variable η labels the time of entry into the can of distinct particles in the beam. More explicitly, the B field is given by

$$\mathfrak{B}^{(1)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\int_0^t W_j^{>}\,d\tau, \quad vt < z, \tag{15a}$$

ahead of the leading edge of the pulse, and

$$\mathfrak{B}^{(1)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\left(\int_0^{T_z} W_j^{>}\,d\tau + \int_{T_z}^t W_j^{<}\,d\tau\right), \quad z < vt, \tag{15b}$$

behind the leading edge of the pulse. The factors W are given in Eq. (8), while $\mathfrak{B}$ and T_z are defined through

$$B_\phi \equiv (4qv\gamma\beta/br_0^2)\mathfrak{B},$$
$$T_z \equiv z/v.$$

In the second epoch there are three relevant domains: $z \ge vt$, ahead of the leading edge, $z \le vt - b$, behind the leading edge, and $vt - b < z < vt$. There results

$$\mathfrak{B}^{(2)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\int_{t-T_b}^t W_j^{>}\,d\tau, \quad z \ge vt, \tag{16a}$$

$$\mathfrak{B}^{(2)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\left(\int_{t-T_b}^t W_j^{<}\,d\tau\right), \quad z \le vt - b, \tag{16b}$$

$$\mathfrak{B}^{(2)} = \sum_j \frac{J_1(v_j r)}{J_1^2}\left(\int_{t-T_b}^{T_z} W_j^{>}\,d\tau + \int_{T_z}^t W_j^{<}\,d\tau\right),$$
$$vt - b < z < vt, \tag{16c}$$
$$T_b \equiv b/v.$$

In the third epoch the pulse is leaving the cavity, i.e., $L - b < vt - b < L$; there are two relevant domains. In the domain ahead of the after edge of the pulse,

$$\mathfrak{B}^{(3)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\left(\int_{t-T_b}^{T_z} W_j^{>}\,d\tau + \int_{T_z}^{T_L} W_j^{<}\,d\tau + \int_{T_L}^t \overline{W}_j(\tau)\,d\tau\right), \tag{17a}$$
$$T_L \equiv L/v, \quad T_L = K/\Omega.$$

The wake factor $\overline{W}_j(t)$ is given by Eq. (11).

In the domain behind the after edge of the pulse,

$$\mathfrak{B}^{(3)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\left(\int_{t-T_b}^{T_L} W_j^{<}\,d\tau + \int_{T_L}^t \overline{W}_j(\tau)\,d\tau\right), \quad z < vt - b. \tag{17b}$$

In the fourth epoch, $vt - b \ge L$, the solution is a superposition of homogeneous wake fields

$$\mathfrak{B}^{(4)} = \sum_j \frac{J_1(k_j r)}{J_1^2}\int_0^{T_b} \overline{W}_j(t-\eta)\,d\eta, \tag{18}$$

and all modes are harmonic.

In addition to purely harmonic behavior, the time dependence of the fields in the first three epochs includes the hyperbolic components

$$\int \begin{pmatrix}\cosh \Omega\tau \\ \sinh \Omega\tau\end{pmatrix}\Omega\,d\tau = \begin{pmatrix}\sinh \Omega\tau \\ \cosh \Omega\tau\end{pmatrix}.$$

More generally, all of the time integrations in Eqs. (11)–(15) are simply performed.

In the extremely relativistic limit much of the segment may enter the cavity before the fields are influenced by the far wall. (See Fig. 7.) The fields are then most simply obtained by integrating the asymptotic forms (9). Writing only the B fields, we have for $vt < z < ct$ (and $vt < b$), in front of the leading edge

of the pulse,

$$B_\phi = \frac{4q\gamma v\beta}{br_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2} e^{-\gamma k_j z} \frac{\sinh \Omega_j t}{\Omega_j},$$

and, behind the leading edge of the pulse,

$$B_\phi = \frac{4q\gamma v\beta}{br_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2 \Omega_j} (1 - e^{-\Omega_j t} \cosh \gamma k_j z).$$

For any of these pulse problems, if the total charge of the segment is q and I_A is the beam current in amperes, then

$$q = (10/c) I_A T_b \text{ esu.}$$

D. Return Current Point and Line Charge Segment

The surface current at the wall of the perfectly conducting cylinder is given by

$$S_z = (c/4\pi) B_\phi(r_0) \text{ (statamp/cm).}$$

From Eq. (8) one obtains, for the propagating point charge,

$$S_z^{\gtrless} = \frac{c}{\pi} \frac{q\gamma\beta}{r_0^2} \sum_j \frac{W_j^{\gtrless}(z, t; v, L)}{J_1(k_j r_0)},$$

where, as before, $\gtrless$ denotes $z \gtrless vt$.

After the particle leaves the box, the surface current is obtained from the wake solution, Eq. (11). For the line pulse of charge, there are four distinct epochs ($\alpha = 1, \cdots, 4$). The surface current during these intervals is given by

$$S_z^{(\alpha)} = (q\gamma c^2\beta^2/\pi b r_0^2) \mathcal{B}^{(\alpha)}$$

with $\mathcal{B}^{(\alpha)}$ given in Eqs. (15)–(18). The time behavior of S_z follows $\mathcal{B}$.

5. CONCLUSIONS

In this analysis we have studied the fields induced in a finite, closed, cylindrical cavity, with perfectly conducting walls, by a relativistically moving charged point particle. Referring to Fig. 6, we see that the fields conveniently divide into four distinct domains. In the region bounded by the triangle OAB the fields are those of a point charge moving in a semi-infinite tube. All events in this domain are not influenced by the forward wall (whose world-line is $z = L$). In the domain bounded by the triangle OAD all fields vanish since all events in this domain are not influenced by the particle. Above the line $ct = -z + 2L$, i.e., above the triangle OAB, the full solution, Eq. (8), comes into play. The wake fields, Eq. (11), come into play above the line $t = T_L$, again, excluding the points in OAB.

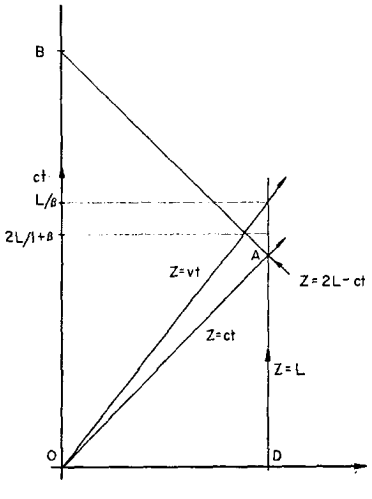


FIG. 6. Space-time diagram for the point charge.

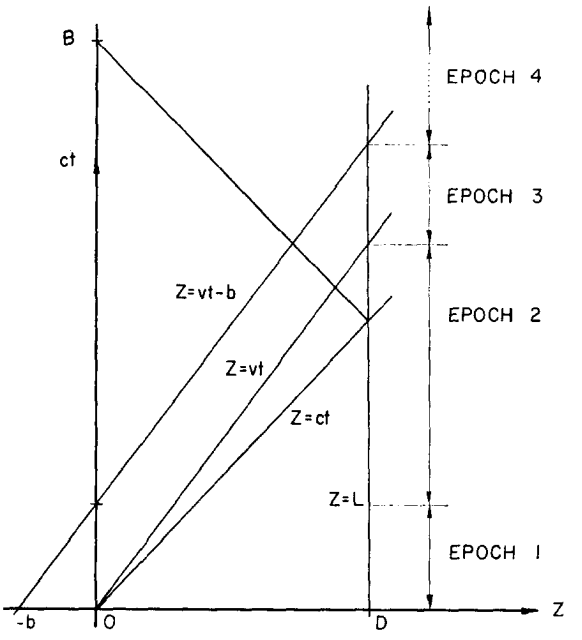


FIG. 7. Space-time diagram for the line charge.

In calculating the energy in the wake fields, a singularity enters owing to the idealization in our model that the point charge enters and leaves the cavity through point holes. This in turn necessitates that the point charge coalesce with its image and vanish in zero time. The infinite deceleration launches a singular pulse back into the cavity. Any finite hole obviates this singularity. For an electron, it was found that a hole of diameter exceeding one fermi insures that the initial energy of the particle is large compared to the energy excited in the wake fields.

The point charge solution was used to obtain the fields of a line segment of charge. Here it was found that the time domain relevant to the problem divides into four epochs as depicted in Fig. 5. The space-time diagram for this problem is shown in Fig. 7. In the triangle OAB the fields are those of a finite segment propagating down a semi-infinite cylinder. Above the

line $ct = -z + 2L$ the total solution as given by Eqs. (15)–(18) comes into play.

The formalism introduced herein for obtaining the relativistic Green's solution may be easily extended to a variety of problems, provided one is able to formulate the static solution in terms of images. Such problems include motion in a cylinder of arbitrary cross section, the motion of any charge configuration which lies in a plane of constant z, and motion in a dielectric medium. The limitation of the theory is that it does not account for interaction between charges in a given configuration.

ACKNOWLEDGMENTS

Discussions on these topics with Professor E. Ott proved invaluable. My sincere appreciation and gratitude are extended to him. Special thanks are also due Professor N. Rostoker, Professor P. McIsaac, and Dr. J. Hieronymus for their very worthwhile suggestions relating to this paper. I am indebted to Dr. E. D. Courant for a critical reading of the final manuscript.

This research was supported, in part, under Contract Nonr 401-(60), with the Physics Branch of the Office of Naval Research.

[1] P. L. Morton, V. K. Neil, and A. M. Sessler, J. Appl. Phys. **37**, 3875 (1966).
[2] E. D. Courant, IEEE Trans. Nucl. Sci. **12**, 550 (1965).
[3] S. Weinberg, J. Math. Phys. **5**, 1371 (1964).
[4] J. A. Nation, Bull. Am. Phys. Soc. **13**, 1572 (1968).
[5] J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1963).
[6] E. Ott and J. Shmoys, Quart. J. Appl. Math. **25**, 378 (1968); **26**, 188 (1968).
[7] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Co., New York, 1968), 3rd ed.
[8] E. Ott, thesis, Polytechnic Institute of Brooklyn, New York, 1967.
[9] F. G. Bass and V. M. Yakovenko, Usp. Fiz. Nauk **86**, 189 (1965) [Sov. Phys. Usp. **8**, 420 (1965)].
[10] Y. N. Dnestrovskii and D. P. Kostomarov, Radiotekhnika Elektron. **4**, 1644 (1959).
[11] R. W. P. King and T. T. Wu, *The Scattering and Diffraction of Waves* (Harvard University Press, Cambridge, Mass., 1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Expected Number of Distinct Sites Visited by a Random Walk with an Infinite Variance

JOSEPH E. GILLIS AND GEORGE H. WEISS*
Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel

(Received 6 October 1969)

Consider a random walk of n steps on an infinite, simple cubic lattice. Let $p(r)$ be the (symmetric) probability of a vector jump r, and let S_n be the expected number of distinct lattice points visited in the course of the random walk. In the present paper we calculate asymptotic values for S_n for the particular choice of jump probabilities $p(r) = p(-r) = Ar^{-(1+\alpha)}$, where $2 \geq \alpha > 0$, and $p(r_1, r_2) = Ar^{-\beta}$, where $r^2 = r_1^2 + r_2^2$, $2 \geq \beta > 1$, and A denotes the normalizing constant. The results are, in 1D, (1) $S_n \sim An$, $1 > \alpha > 0$, (2) $S_n \sim Bn/\ln n$, $\alpha = 1$, (3) $S_n \sim Cn^{1/\alpha}$, $2 > \alpha > 1$, (4) $S_n \sim D(n \ln n)^{\frac{1}{2}}$, $\alpha = 2$, where A, B, C, and D are calculable constants, and, in 2D, (1) $S_n \sim An$, $2 > \beta > 1$, (2) $S_n \sim Bn/\ln \ln n$, $\beta = 2$.

1. INTRODUCTION

Dvoretzky and Erdos[1] were the first to study the statistics of the distinct number of sites visited in an n-step random walk on a lattice. Their results were subsequently rederived by Vineyard[2] and later extended by Montroll and Weiss[3] to include random walks whose jump probabilities had a finite variance (cf. also Spitzer[4]). The results of Montroll and Weiss were derived through the use of Karamata's Tauberian theorem, a technique first introduced into the study of random walks by Darling and Kac[5] and Kac.[6] Although all of the known results pertain to jump probabilities with finite variance, the use of Tauberian theorems allows one to extend the study to jump probabilities having infinite variance. If $p(n)$ is the probability of changing position by a vector n in a single step, then the variance (in one dimension) is defined by

$$\sigma^2 = \sum_{n=-\infty}^{\infty} n^2 p_n, \tag{1.1}$$

with analogous definitions in higher dimensions. In the present paper, we derive results for certain specific random walks in one and two dimensions for which the variance associated with jump probabilities is infinite. In three and higher dimensions, the problem is of lesser interest since, even in the case of a finite variance, the expected number of distinct sites visited in an n-step random walk is asymptotically An, where A is a calculable constant. Since the result can be at most n, the order of the asymptotic dependence on n cannot depend on whether the variance is finite or infinite, though the value of A, as well as correction terms, will depend on the variance.

For random walks with finite variance, it is known that the expected number of distinct sites visited during an n-step random walk is as follows:

$$\begin{aligned} S_n &= a_1\sqrt{n}, && \text{in 1D,} \\ S_n &= a_2 n/\ln n, && \text{in 2D,} \\ S_n &= a_3 n, && \text{in 3D,} \end{aligned} \tag{1.2}$$

where a_1, a_2, and a_3 are constants. The change of variance from finite to infinite will be shown to lead to an increase in the order of magnitude of S_n in one and two dimensions for the specific models

$$\begin{aligned} p(n) &= An^{-(1+\alpha)}, && \text{in 1D,} \\ p(n_1, n_2) &= B(n_1^2 + n_2^2)^{-\beta}, && \text{in 2D,} \end{aligned} \tag{1.3}$$

and certain extensions of these random walks.

2. A 1-DIMENSIONAL SET OF TRANSITION PROBABILITIES

Let $p(n)$ denote the probability of a jump of n sites at a single step, when $n = (n)$ in one dimension and $n = (n_1, n_2)$ in two dimensions. We define the structure factor of the random walk by

$$\lambda(\boldsymbol{\theta}) = \sum_{\mathbf{n}} p(\mathbf{n}) \exp(i\mathbf{n} \cdot \boldsymbol{\theta}), \tag{2.1}$$

where $\boldsymbol{\theta} = (\theta)$ or $\boldsymbol{\theta} = (\theta_1, \theta_2)$. In k dimensions, we define the integral

$$P(z) = \frac{1}{\pi^k} \int_0^{\pi} \cdots \int \frac{d^k\boldsymbol{\theta}}{1 - z\lambda(\boldsymbol{\theta})}. \tag{2.2}$$

Then it is known[3] that the asymptotic value of S_n, the expected number of distinct sites visited on an n-step walk, is given by

$$S_n \sim n/P(1 - n^{-1}), \tag{2.3}$$

provided that $P(1 - n^{-1})$ has the asymptotic form

$$P(1 - n^{-1}) \sim n^{\beta} L(n), \tag{2.4}$$

where $L(n)$ is a slowly varying function, i.e.,

$$\lim L(cn)/L(n) = 1,$$

for all $c > 0$, as $n \to \infty$.

In what follows we assume that $p(\mathbf{n})$ is symmetric in its indices. The specific form of $p(\mathbf{n})$ to be studied in one dimension will be

$$p(n) = p(-n) = B(n^{-(1+\alpha)} + \epsilon_n), \tag{2.5}$$

where it is assumed that ϵ_n is such that

$$\sum_{n=1}^{\infty} n^2 \epsilon_n < \infty \tag{2.6}$$

and B is a normalization constant. The condition in Eq. (2.6) can be weakened somewhat, but we shall not pursue this generalization.

Since we are interested in the case of infinite variance, α must satisfy

$$2 \geq \alpha > 0. \tag{2.7}$$

It follows from Eq. (2.2) that the singular behavior of $P(z)$ near $z = 1$ must come from the behavior of the integrand in the neighborhood of the root of $\lambda(\theta) = 1$. For the cases of present interest, $\theta = 0$ will be the only such root.[7] In order to calculate the behavior of $P(1 - n^{-1})$ for large n, we must find an expansion for $\lambda(\theta)$ valid for small $|\theta|$. This requires analysis of the function

$$G_{\alpha}(\theta) = \sum_{n=1}^{\infty} \frac{1 - \cos n\theta}{n^{1+\alpha}} \tag{2.8}$$

for small θ.

We substitute

$$\frac{1}{n^{1+\alpha}} = \frac{1}{\Gamma(1+\alpha)} \int_0^{\infty} t^{\alpha} e^{-nt}\, dt \tag{2.9}$$

in (2.8) and perform the resulting summation. This leads to the exact representation

$$G_{\alpha}(\theta) = \frac{1}{\Gamma(1+\alpha)} \times \int_0^{\infty} \frac{t^{\alpha} e^{-t}(1 + e^{-t})(1 - \cos\theta)\, dt}{(1 - e^{-t})[(1 - e^{-t})^2 + 2e^{-t}(1 - \cos\theta)]}. \tag{2.10}$$

To study the behavior of this integral near $\theta = 0$, we approximate it by

$$g_{\alpha}(\theta) \sim \frac{\theta^2}{2\Gamma(1+\alpha)} \int_0^{\infty} \frac{t^{\alpha} e^{-t}(1 + e^{-t})\, dt}{(1 - e^{-t})[(1 - e^{-t})^2 + \theta^2 e^{-t}]}. \tag{2.11}$$

It is easily verified directly that $|G_{\alpha}(\theta) - g_{\alpha}(\theta)|$ is of lower order than the terms retained.

It is clear from (2.11) that, if θ is set equal to zero in the integrand, the resulting integral

$$\int_0^{\infty} \frac{t^{\alpha} e^{-t}(1 + e^{-t})\, dt}{(1 - e^{-t})^3}$$

diverges because of the singularity at $t = 0$. It follows that the behavior of the integral for small θ depends only on the behavior of the integrand in a neighborhood of $t = 0$. It is shown in Appendix A that the limiting behavior of $g_{\alpha}(\theta)$ can be found from the integral

$$g_{\alpha}(\theta) \sim h_{\alpha}(\theta) = \frac{\theta^2}{\Gamma(1+\alpha)} \int_0^{\infty} \frac{t^{\alpha-1} e^{-t}\, dt}{(t^2 + \theta^2)} = \frac{\theta^{\alpha}}{\Gamma(1+\alpha)} \int_0^{\infty} \frac{x^{\alpha-1} e^{-\theta x}\, dx}{1 + x^2}. \tag{2.12}$$

We can now evaluate the limiting behavior of $g_{\alpha}(\theta)$ for small θ. For $\alpha < 2$,

$$\frac{h_{\alpha}(\theta)}{\theta^{\alpha}} \sim \frac{1}{\Gamma(1+\alpha)} \int_0^{\infty} \frac{x^{\alpha-1}\, dx}{1 + x^2} = \frac{\pi \csc \frac{1}{2}\pi\alpha}{2\Gamma(1+\alpha)}. \tag{2.13}$$

When $\alpha = 2$, we must include the exponential term to insure convergence at $t = \infty$. It follows that

$$\frac{2h_2(\theta)}{\theta^2} \sim \int_0^{\infty} \frac{x e^{-\theta x}\, dx}{1 + x^2}. \tag{2.14}$$

We see that, for small θ,

$$\frac{d}{d\theta}\left(\frac{h_2(\theta)}{\theta^2}\right) \sim -\frac{1}{2\theta} + \tfrac{1}{4}\pi + O(1) \tag{2.15}$$

so that

$$h_2(\theta) \sim -\tfrac{1}{2}\theta^2 \ln\theta = \tfrac{1}{2}\theta^2 \ln^{-1}\theta. \tag{2.16}$$

We now return to the evaluation of $P(z)$ for z close to 1. For $0 < \alpha < 1$, we find

$$P(z) \sim \frac{1}{\pi} \int_0^{\pi} \frac{d\theta}{1 - z + B z \theta^{\alpha}}, \tag{2.17}$$

where B is the normalizing constant in Eq. (2.5). Hence, $P(1)$ is finite and it follows from Eq. (2.2) that

$$S_n \sim n/P(1). \tag{2.18}$$

For $\alpha = 1$, we have

$$P(z) \sim \frac{1}{\pi} \int_0^{\pi} \frac{d\theta}{1 - z + \frac{1}{2}B\theta} \sim -\frac{2}{B} \ln(1 - z) + O(1) \tag{2.19}$$

so that

$$S_n \sim \frac{Bn}{2 \ln n}. \tag{2.20}$$

It is interesting to notice that this resembles the result for a finite variance random walk in two dimensions.[3] When $2 > \alpha > 1$, the integral is of the form shown in (2.17), but $P(1)$ is no longer finite. For that case, we set $\theta^\alpha = (1 - z)x$ so that

$$P(z) \sim \frac{1}{\pi(1-z)^{1-1/\alpha}} \int_0^{z/(1-z)^{1/\alpha}} \frac{dx}{(1 + Bx^\alpha)x^{1-1/\alpha}} \sim \frac{1}{\pi(1-z)^{1-1/\alpha}} \int_0^\infty \frac{dx}{(1 + Bx^\alpha)x^{1-1/\alpha}}. \tag{2.21}$$

This result implies that, for $2 > \alpha > 1$,

$$S_n \sim cn^{1/\alpha} \tag{2.22}$$

where c is found from Eq. (2.21).

Finally, when $\alpha = 2$, we must find the asymptotic behavior of the integral

$$P(z) \sim \frac{1}{\pi} \int_0^a \frac{d\theta}{1 - z + Bz\theta^2 \ln \theta^{-1}}, \tag{2.23}$$

where a can be any fixed constant which we choose less than 1 [so that the singularity at $\theta = 1$, which is due to our choice of an approximate form for $\lambda(\theta)$, does not have to be discussed]. For simplicity, let us define the integral

$$F(\epsilon) = \int_0^a \frac{d\theta}{\epsilon + \theta^2 \ln \theta^{-1}} \tag{2.24}$$

so that

$$\lim_{z \to 1} P(z) = \frac{1}{\pi B} \lim_{z \to 1} F\left(\frac{1-z}{B}\right). \tag{2.25}$$

In this integral change variables by the transformation

$$\theta^2 \ln \theta^{-1} = y, \tag{2.26}$$

with a solution represented by

$$\theta = f(y). \tag{2.27}$$

Then $F(\epsilon)$ can be formally written

$$F(\epsilon) = \int_0^b \frac{f'(y)}{y + \epsilon} dy, \tag{2.28}$$

where $b = a^2 \ln a^{-1}$. The integral for $F(\epsilon)$ clearly diverges at $\epsilon = 0$ because of the singularity at $y = 0$ in the integrand. The nature of the singularity can be determined by giving an accurate representation for $f'(y)$ in the neighborhood of $y = 0$. For this purpose set $\theta = \sqrt{y}\, u(y)$ in Eq. (2.26) so that $u(y)$ is the solution to

$$u^2(y)\{\tfrac{1}{2} \ln y^{-1} + \ln [u(y)]^{-1}\} = 1. \tag{2.29}$$

Let us replace this equation by the iterative scheme

$$u_{n+1}^2(y) = \frac{2}{\ln y^{-1} + 2 \ln [u_n(y)]^{-1}}, \quad n = 1, 2, \cdots,$$
$$u_0(y) = 1. \tag{2.30}$$

The first approximation to a solution is

$$u_1(y) = (2/\ln y^{-1})^{\frac{1}{2}} \tag{2.31}$$

and the second approximation is

$$u_2(y) = [2/(\ln y^{-1} + \ln \ln y^{-1} - \ln 2)]^{\frac{1}{2}}. \tag{2.32}$$

This suggests that the general solution to Eq. (2.29) can be expressed in the form

$$u_n(y) = \{2/[\ln y^{-1} + \eta_n(y)]\}^{\frac{1}{2}}, \tag{2.33}$$

where

$$\lim_{y \to 0} \frac{\eta_n(y)}{\ln y^{-1}} = 0. \tag{2.34}$$

An inductive argument serves to establish the validity of Eq. (2.33) and also the approximation $\eta_n(y) \sim \ln \ln y^{-1}$.

With these results we have, finally, that for y in a neighborhood of zero

$$f(y) = \left(\frac{2y}{\ln y^{-1}}\right)^{\frac{1}{2}} \left[1 + O\left(\frac{\ln \ln y^{-1}}{\ln y^{-1}}\right)\right] \tag{2.35}$$

so that $F(\epsilon)$ behaves asymptotically as

$$F(\epsilon) \sim \frac{1}{\sqrt{2}} \int_0^{b'} \frac{dy}{y + \epsilon} \left(\frac{1}{(y \ln y^{-1})^{\frac{1}{2}}} + \frac{1}{(y \ln^3 y^{-1})^{\frac{1}{2}}}\right), \tag{2.36}$$

with the singular behavior of $F(\epsilon)$ still determined by the behavior of the integrand at the origin. The second term in the brackets, $(y \ln^3 y^{-1})^{-\frac{1}{2}}$, can be neglected in comparison with $(y \ln y^{-1})^{-\frac{1}{2}}$, as $y \to 0$, so that

$$F(\epsilon) \sim \frac{1}{\sqrt{2}} \int_0^{b'} \frac{dy}{y + \epsilon} \frac{1}{y \ln y^{-1}}. \tag{2.37}$$

In this integral we make the substitution $y = \epsilon x$ which leads to the representation

$$F(\epsilon) \sim \frac{1}{(2\epsilon)^{\frac{1}{2}}} \int_0^\infty \frac{dx}{1 + x} \frac{1}{[x(\ln x^{-1} + \ln \epsilon^{-1})]^{\frac{1}{2}}}, \tag{2.38}$$

where the upper limit of integration has been replaced by ∞ since the resulting integral is convergent. At this point we split the range of integration $(0, \infty)$ into $(0, A\epsilon)$ and $(A\epsilon, \infty)$, where A is chosen so that $A \gg 1$ and $A\epsilon \ll 1$. The first integral can be bounded as

follows:

$$\int_0^{A\epsilon} \frac{dx}{(1+x)[x(\ln x^{-1} + \ln \epsilon^{-1})]^{\frac{1}{2}}} < \frac{1}{(\ln \epsilon^{-1})^{\frac{1}{2}}} \int_0^x \frac{dx}{\sqrt{x}} = 2\left(\frac{A\epsilon}{\ln \epsilon^{-1}}\right)^{\frac{1}{2}}. \quad (2.39)$$

In the second range of integration, we can estimate

$$\int_{A\epsilon}^{\infty} \frac{dx}{(1+x)[x(\ln x^{-1} + \ln \epsilon^{-1})]^{\frac{1}{2}}} = \frac{1}{(\ln \epsilon^{-1})^{\frac{1}{2}}} \int_{A\epsilon}^{\infty} \frac{dx}{(1+x)[x(1 + \ln x^{-1}/\ln \epsilon^{-1})]^{\frac{1}{2}}} \sim \frac{1}{(\ln \epsilon^{-1})^{\frac{1}{2}}} \int_0^{\infty} \frac{dx}{(1+x)\sqrt{x}} + O\left(\frac{1}{\ln \epsilon^{-1}}\right) = \frac{\pi}{(\ln \epsilon^{-1})^{\frac{1}{2}}}\left[1 + O\left(\frac{1}{(\ln \epsilon^{-1})^{\frac{1}{2}}}\right)\right]. \quad (2.40)$$

We therefore see, by comparing the last two equations, that the dominant contribution comes from the second range of integration and that

$$F(\epsilon) \sim \pi/(2\epsilon \ln \epsilon^{-1})^{\frac{1}{2}}. \quad (2.41)$$

But, by (2.3), this implies that, for $\alpha = 2$,

$$S_n \sim (2Bn \ln n)^{\frac{1}{2}}. \quad (2.42)$$

When $\alpha > 2$, results in earlier references imply that S_n is asymptotically proportional to $\sqrt{n}$.

3. A 2-DIMENSIONAL SET OF TRANSITION PROBABILITIES

We next consider a 2-dimensional generalization of the jump probabilities given in Eq. (2.5),

$$p(n, m) = B[(n^2 + m^2 + D^2)^{-\beta} + \epsilon_{nm}], \quad (3.1)$$

where we will assume that the ϵ_{nm} satisfy

$$\sum_{n,m} n^2\epsilon_{nm}, \quad \sum_{n,m} nm\epsilon_{nm}, \quad \sum_{n,m} m^2\epsilon_{nm} < \infty. \quad (3.2)$$

The constant β is chosen to satisfy

$$2 \geq \beta > 1, \quad (3.3)$$

which implies that the covariances associated with $p(n, m)$ are infinite. We notice that the inserted constant D^2 implies that $p(0, 0) \neq 0$. This is done to simplify the algebra. No loss of generality follows from the particular form of Eq. (3.1) since the asymptotic dependence on n does not depend on D, although the coefficients may be functions of this parameter.

In order to calculate the function $P(1 - n^{-1})$, we must study the behavior of $\lambda(\theta)$ defined in Eq. (2.1), noting that $\lambda(0) = 1$. To do so, we invoke the 2-dimensional form of the Poisson summation formula

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n, m) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \iint_{-\infty}^{\infty} f(x, y) \exp[2\pi i(rx + sy)]\, dx\, dy, \quad (3.4)$$

where it is assumed that both sides of this equation exist. The contribution to $\lambda(\theta)$ that determines the asymptotic form of $P(1 - n^{-1})$ comes from the sum

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \frac{\exp[i(n\theta_1 + m\theta_2)]}{(n^2 + m^2 + D^2)^{\beta}}, \quad (3.5)$$

evaluated near $\theta = 0$. Let us therefore analyze the behavior of the function

$$\frac{\Delta\lambda(\boldsymbol{\theta})}{B} = \sum_{n=-\infty}^{\infty} \sum_{-\infty}^{\infty} \frac{1 - \exp[i(n\theta_1 + m\theta_2)]}{(n^2 + m^2 + D^2)^{\beta}} \quad (3.6)$$

by applying the transformation (3.4). It is found that[8]

$$\frac{\Delta\lambda(\boldsymbol{\theta})}{B} = \frac{2}{2^{\beta-1}\Gamma(\beta)} \frac{1}{D^{\beta-1}} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \Bigg\{ [2\pi(r^2 + s^2)^{\frac{1}{2}}]^{\beta-1} \times K_{\beta-1}(2\pi D(r^2 + s^2)^{\frac{1}{2}}) - \left\{2\pi\left[\left(r - \frac{\theta_1}{2\pi}\right)^2 + \left(s - \frac{\theta_2}{2\pi}\right)^2\right]^{\frac{1}{2}}\right\}^{\beta-1} \times K_{\beta-1}\left(2\pi D\left[\left(r - \frac{\theta_1}{2\pi}\right)^2 + \left(s - \frac{\theta_2}{2\pi}\right)^2\right]^{\frac{1}{2}}\right)\Bigg\}, \quad (3.7)$$

where $K_{\beta-1}(x)$ is a modified Bessel function of the second kind. If $\beta < 2$, we see that a possible branch-point singularity occurs in the term $r = s = 0$. The exact nature of such a singularity can be determined from the identity[8]

$$K_{\nu}(x) = (\pi/\sin \pi\nu)[I_{-\nu}(x) - I_{\nu}(x)] \quad (3.8)$$

for noninteger ν, where $I_{\nu}(x)$ is a modified Bessel function of the first kind defined by the series

$$I_{\nu}(x) = \sum_{k=0}^{\infty} \frac{1}{k!\,\Gamma(\nu + k + 1)} (\tfrac{1}{2}x)^{\nu+2k}. \quad (3.9)$$

The relation (3.8) requires that ν not be equal to an integer. It follows from (3.8) and (3.9) that

$$\lim_{\epsilon \to 0} \left(\frac{\epsilon}{D}\right)^{\beta-1} K_{\beta-1}(\epsilon) = \frac{2^{\beta-1}}{D^{\beta-1}\Gamma(2 - \beta)}. \quad (3.10)$$

Hence, by considering the term $r = s = 0$ on the right-hand side of Eq. (3.7), we find that

$$\Delta\lambda(\boldsymbol{\theta}) \sim a(\theta_1^2 + \theta_2^2)^{\beta-1}, \quad (3.11)$$

where a is a constant. The asymptotic behavior of $P(z)$ near $z = 1$, therefore, depends on the integral

$$P(z) \sim \frac{1}{\pi^2} \iint_0^\pi \frac{d\theta_1\, d\theta_2}{1 - z + a(\theta_1^2 + \theta_2^2)^{\beta-1}}. \quad (3.12)$$

Transforming to polar coordinates and setting $z = 1$, we see that the integral converges for $\beta < 2$, so that

$$\lim_{n\to\infty} P(1 - n^{-1}) = P(1) \quad (3.13)$$

and, therefore, by Eq. (2.2), when $2 > \beta > 1$, it follows that for large n

$$S_n \sim n/P(1), \quad (3.14)$$

where $P(1)$ must be evaluated numerically.

The case $\beta = 2$ is more difficult. For this case we note that, for small[8] ϵ,

$$K_1(\epsilon) \sim \tfrac{1}{2}\epsilon \ln \epsilon. \quad (3.15)$$

Again we see from Eq. (3.7) that the singular behavior of $\Delta\lambda(\theta)$ near the origin is determined by the term $r = s = 0$. If we use Eq. (3.15) to determine this behavior, we find that

$$\Delta\lambda(\boldsymbol{\theta})/B \sim \tfrac{1}{4}\pi(\theta_1^2 + \theta_2^2) \ln (\theta_1^2 + \theta_2^2)^{-1}. \quad (3.16)$$

Thus, we must study the singular behavior of

$$P(z) \sim \frac{1}{\pi^2} \iint_C \frac{d\theta_1\, d\theta_2}{1 - z + b(\theta_1^2 + \theta_2^2) \ln (\theta_1^2 + \theta_2^2)^{-1}}, \quad (3.17)$$

in which $b = \frac{1}{4}\pi B$ and in which C can be chosen to be a circle in the (θ_1, θ_2) plane without changing the nature of the singularity at $z = 1$ (since only the behavior at $\theta = 0$ determines the singularity). The form of the integral suggests a transformation to polar coordinates in which the angular integration is immediate. We are thus led to consider the integral

$$G(\epsilon) = \int_0^R \frac{r\, dr}{\epsilon + r^2 \ln r^{-1}} = \int_0^{R^2} \frac{dv}{2\epsilon + v \ln v^{-1}}, \quad (3.18)$$

in terms of which we have

$$\lim_{z\to 1} P(z) = \frac{2}{\pi b} \lim_{z\to 1} G\left(\frac{1 - z}{b}\right) \quad (3.19)$$

and in which R^2 is chosen less than 1 but is otherwise arbitrary. To determine the limiting behavior of $G(\epsilon)$, we use the same device as for the analysis of the integral appearing in Eq. (2.24). Making the substitution

$$v \ln v^{-1} = x, \quad (3.20)$$

we find that, for x small,

$$v \sim \frac{x}{\ln x^{-1}} \quad (3.21)$$

so that

$$G(\epsilon) \sim \int_0^{R_1} \frac{dx}{2\epsilon + x} \frac{1}{\ln x^{-1}}. \quad (3.22)$$

But we can write

$$\frac{1}{\ln x^{-1}} = \frac{1}{\ln (x + 2\epsilon)^{-1}} + \frac{\ln x(x + 2\epsilon)^{-1}}{\ln x^{-1} \ln (x + 2\epsilon)^{-1}}, \quad (3.23)$$

which, when substituted into Eq. (3.22), yields the estimate

$$G(\epsilon) \sim \ln \ln \epsilon^{-1} + R(\epsilon), \quad (3.24)$$

where

$$R(\epsilon) = \int_0^{R_1} \frac{dx}{2\epsilon + x} \frac{\ln x(x + 2\epsilon)^{-1}}{\ln x^{-1} \ln (x + 2\epsilon)^{-1}}. \quad (3.25)$$

It is shown in Appendix B that $R(\epsilon) = O(1)$ as $\epsilon \to 0$ so that asymptotically $G(\epsilon) \sim \ln \ln \epsilon^{-1}$, which implies that the asymptotic expected number of distinct sites visited is, for $\beta = 2$,

$$S_n \sim \tfrac{1}{8}\pi^2 Bn/\ln \ln n. \quad (3.26)$$

For $\beta > 2$, S_n is asymptotic to $n/\ln n$, as is shown in Ref. 3.

Finally, we note that it has been shown that the asymptotic expected number of points visited exactly once in an n-step random walk is

$$V_n \sim n/P^2(1 - n^{-1}). \quad (3.27)$$

Since we have calculated the function $P(1 - n^{-1})$ for a number of cases, we only list the results. In one dimension, for the jump probabilities given by Eq. (2.5), we have

$$\begin{aligned} &V_n \sim n/P^2(1), && 1 > \alpha > 0, \\ &V_n \sim B^2 n/(4 \ln^2 n), && \alpha = 1, \\ &V_n \sim C^2 n^{(2/\alpha)-1}, && 2 > \alpha > 1, \\ &V_n \sim 2B \ln n, && \alpha = 2. \end{aligned} \quad (3.28)$$

In two dimensions, for jump probabilities given by Eq. (3.1), we have

$$\begin{aligned} &V_n \sim n/P^2(1), && 2 > \beta > 1, \\ &V_n \sim \pi^2 b^2 n/(4 \ln \ln^2 n), && \beta = 2. \end{aligned} \quad (3.29)$$

For $\alpha > 2$ in one dimension and for $\beta > 2$ in two dimensions, results for V_n are given in Ref. 3.

In concluding, we note that the methods of the present paper allow us to analyze results for the 2-dimensional jump probabilities

$$p(n_1, n_2) = A(|n_1| + |n_2|)^{-\alpha},$$

where A and α are constants. It is quite possible that results can also be obtained for jump probabilities of the form that we have considered, but multiplied by a slowly varying function of its parameters. Such investigations would involve analyses of the kind found in Zygmund[9] and appear to be beyond the scope of any method based on Karamata's theorem.

APPENDIX A: LIMITING PROPERTY OF $g_\alpha(\theta)$

Let us write

$$g_\alpha(\theta) = h_\alpha(\theta) + [g_\alpha(\theta) - h_\alpha(\theta)]. \tag{A1}$$

We will show that

$$\lim_{\theta\to 0} \frac{g_\alpha(\theta) - h_\alpha(\theta)}{h_\alpha(\theta)} = 0, \tag{A2}$$

for $2 \geq \alpha > 0$. The difference $g_\alpha(\theta) - h_\alpha(\theta)$ can be written explicitly as, say,

$$\begin{aligned} &g_\alpha(\theta) - h_\alpha(\theta) \\ &= \frac{\theta^2}{2\Gamma(1+\alpha)} \int_0^\infty t^\alpha e^{-t} \\ &\quad \times \left(\frac{t(t^2+\theta^2)(1+e^{-t}) - 2(1-e^{-t})[(1-e^{-t})^2 + \theta^2 e^{-t}]}{t(1-e^{-t})(t^2+\theta^2)[(1-e^{-t})^2+\theta^2 e^{-t}]} \right) dt \\ &= \frac{\theta^2}{2\Gamma(1+\alpha)} I(\theta). \end{aligned} \tag{A3}$$

It is clear that, for any θ, the integrand is such that there is no trouble with convergence at the upper limit. The only possible difficulty can arise from the lower limit. If we approximate e^{-t} by $1 - t$ near the origin, then the integrand is approximately $2/(t^2 + \theta^2)$ in the neighborhood of the origin. Hence, if $I(\theta)$ diverges, the divergent behavior will be that of the integral

$$J(\theta) = \int_0^\infty \frac{t^\alpha e^{-t}}{t^2+\theta^2}\, dt = \theta^{\alpha-1} \int_0^\infty \frac{x^\alpha e^{-\theta x}}{x^2+1}\, dx. \tag{A4}$$

For $2 \geq \alpha > 1$,

$$J(\theta) \to J(0) = \int_0^\infty t^{\alpha-2} e^{-t}\, dt = \text{const}, \tag{A5}$$

so that $g_\alpha(\theta) - h_\alpha(\theta) = O(\theta^2)$ for $\theta \to 0$ and

$$\frac{g_\alpha(\theta) - h_\alpha(\theta)}{h_\alpha(\theta)} = O(\theta^{2-\alpha}). \tag{A6}$$

For $\alpha = 1$,

$$J(\theta) = \int_0^\infty \frac{x e^{-\theta x}}{1+x^2}\, dx \sim \ln \theta^{-1} + O(1) \tag{A7}$$

by Eq. (A4) so that since $h_1(\theta) \to$ const, the ratio in Eq. (A2) tends to 0 as $\theta^2 \ln(\theta^{-1})$. When $\alpha < 1$, $J(\theta) \sim a\theta^{\alpha-1}$, where a is a constant such that

$$g_\alpha(\theta) - h_\alpha(\theta) \sim b\theta^{1+\alpha}, \tag{A8}$$

where b is a constant. Since $h_\alpha(\theta)$ is $O(\theta^\alpha)$, the ratio tends to 0 as $\theta \to 0$.

APPENDIX B: PROOF THAT $R(\epsilon)$ IS BOUNDED AS $\epsilon \to 0$

The upper limit R_1 appearing in the definition of R_1 in Eq. (3.25) is strictly less than 1. Hence, ϵ can always be chosen small enough to ensure that $R_1 + 2\epsilon < 1$. Thus, we can write

$$\begin{aligned} |R(\epsilon)| &= \left| \int_0^{R_1} \frac{dx}{2\epsilon + x} \frac{\ln x(x+2\epsilon)^{-1}}{\ln x^{-1} \ln (x+\epsilon)^{-1}} \right| \\ &\leq \frac{1}{\ln R_1^{-1} \ln (R_1+\epsilon)^{-1}} \left| \int_0^{R_1} \frac{dx}{2\epsilon + x} \ln \frac{x}{x+2\epsilon} \right| \\ &\leq \frac{1}{\ln R_1^{-1} \ln (R_1+\epsilon)^{-1}} \left| \int_0^\infty \frac{dy}{y+1} \ln \frac{y}{y+1} \right|. \end{aligned} \tag{B1}$$

The indicated integral converges, so that $|R(\epsilon)|$ is bounded as $\epsilon \to 0$. Indeed, a more careful analysis reveals that $\lim R(\epsilon) = 0$ as $\epsilon \to 0$.

* Permanent address: Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Department of Health, Education and Welfare, Bethesda, Maryland 20014.

[1] A. Dvoretzky and P. Erdos, in *Proceedings of the Second Berkeley Symposium on Mathematics and Statistics* (University of California Press, Berkeley, 1951), p. 33.
[2] G. H. Vineyard, J. Math. Phys. **4**, 1191 (1963).
[3] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
[4] F. Spitzer, *Principles of Random Walk* (D. Van Nostrand, Princeton, N.J., 1964).
[5] D. A. Darling and M. Kac, Trans. Am. Math. Soc. **84**, 444 (1957).
[6] M. Kac, Trans. Am. Math. Soc. **84**, 459 (1957).
[7] This follows immediately if we require that $p(n) > 0$ for all n.
[8] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic Press, New York, 1965).
[9] A. Zygmund, *Trigonometric Series* (Cambridge University Press, Cambridge, 1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Natural Spin-Orbitals and Generalized Overlap Amplitudes*

O. GOSCINSKI AND P. LINDNER
Quantum Chemistry Group for Research in Atomic, Molecular, and Solid-State Theory
Uppsala, Sweden

(Received 31 March 1969)

The generalized overlap amplitudes for "hole" states which appear in the one-particle Green's function are, in general, not linearly independent. We show that canonical orthonormalization of them yields the natural spin-orbitals (eigenfunctions of the first-order reduced density matrix).

INTRODUCTION

The eigenfunctions of the first-order density matrix, introduced by Löwdin as natural spin-orbitals,[1] are an essential tool for treating many-electron systems in stationary states. Their optimal convergence properties have been repeatedly emphasized[1,2] and exploited.[3] They are the most suitable set of 1-particle functions to use in discussing a quantum system, even though their direct determination, without previous knowledge of the wavefunction, is hindered by the N-representability problem.[2]

The first-order density matrix can be considered as an initial value for the 1-particle Green's function of field theory, which is becoming common, even for the treatment of finite electron-systems.[4] A certain set of 1-particle functions,[5] the generalized overlap amplitudes, appears naturally in the spectral weight function of the Green's functions. These overlap amplitudes are characterized by

$$\begin{aligned}\varphi_s(\mathbf{x}) = g_s(\mathbf{x}) &= \langle N, 0|\, \psi^\dagger(\mathbf{x})\, |N-1, s\rangle, \quad \epsilon_s < \mu, \\ &= f_s(\mathbf{x}) = \langle N, 0|\, \psi(\mathbf{x})\, |N+1, s\rangle, \quad \epsilon_s > \mu,\end{aligned} \tag{1}$$

and are associated with the elementary excitations

$$\begin{aligned}\epsilon_s &= E(N, 0) - E(N-1, s), \quad \epsilon_s < \mu, \\ &= E(N+1, s) - E(N, 0), \quad \epsilon_s > \mu,\end{aligned} \tag{2}$$

where $|N, 0\rangle$ denotes the ground state of the N-particle system, $|N+1, s\rangle$ the sth excited state of the $(N+1)$-particle system, $E(N, 0)$ and $E(N+1, s)$ the associated energies, μ the chemical potential, and $\psi(\mathbf{x})$ the field operator in the Heisenberg representation. The spectral weight function, given by

$$A(\mathbf{x}, \mathbf{x}', \epsilon) = \sum_s \varphi_s(\mathbf{x})\varphi_s^*(\mathbf{x}')\delta(\epsilon - \epsilon_s), \tag{3}$$

is important in determining 1-electron properties.

The set $\{g_s(\mathbf{x})\}$, for reasons that we will discuss later, induces the temptation of identifying it with the natural spin-orbitals.[6] The g's have been assumed to be orthonormal,[7] but there is every indication that they are linearly dependent,[8] whereas the natural spin orbitals are orthonormal.[1,2] It has been stated that, except for independent particle models, there is no 1-to-1 correspondence between those many-particle states for which the $\varphi_s(\mathbf{x})$ are different from zero and an orthonormal set.[9]

To make clear the need for a clarification of the relation between the natural orbitals and the generalized overlap amplitudes, we refer to the theory of "capture" and "ionization" processes, where the problem arises naturally and where attempts to establish a connection between the functions we are discussing fail.[10]

In this paper, we establish the connection and derive interrelationships between the $g_s(\mathbf{x})$, the $f_s(\mathbf{x})$, and the natural spin-orbitals, which we will denote by $\chi_i(x)$. In the process, we obtain a derivation of a well-known bound for the occupation numbers of the first-order density matrix. The tool employed is Löwdin's canonical orthonormalization procedure, which is specially devised to cope with linear dependences in a set[11] and which is of importance in the non-orthogonality problem.[12] We briefly review the properties of canonical orthogonalization, density matrices, and field-theoretical Green's functions, then establish the desired connections and discuss the results.

CANONICAL ORTHONORMALIZATION: AN APPLICATION

In order to describe the method, we consider first a basis set, $\boldsymbol{\phi} = \{\phi_k\}$, of n functions which spans a subspace V_p of order p, with $n > p$. The problem is to find a linearly independent basis $\boldsymbol{\varphi}$ for this subspace. The fact that $\boldsymbol{\phi}$ is a linearly dependent set implies that the metric matrix

$$\boldsymbol{\Delta} = \boldsymbol{\phi}^\dagger\boldsymbol{\phi} \tag{4}$$

has p nonvanishing eigenvalues, which are positive since

$$\boldsymbol{\Delta} \geq 0. \tag{5}$$

Let U be the unitary transformation that diagonalizes Δ:

$$U^\dagger \Delta U = d = \begin{pmatrix} \mu & 0 \\ 0 & 0 \end{pmatrix}. \tag{6}$$

The diagonal matrix d has p nonvanishing eigenvalues which can be considered to form a diagonal submatrix μ. From (6), it follows that

$$\Delta U = Ud \tag{7}$$

and

$$\Delta V = V\mu, \quad \Delta A = 0, \tag{8}$$

where V is a partition of U:

$$U = (VA). \tag{9}$$

The matrix U is of order $n \times n$, V is a rectangular submatrix of order $n \times p$, and A of order $n \times q$ with $q = n - p$. It easily follows[12] that

$$V^\dagger V = 1_p, \quad V^\dagger \Delta V = \mu, \quad \Delta = V\mu V^\dagger, \tag{10}$$

where 1_p is a unit matrix of order p. We introduce the set

$$\eta = \phi V \mu^{-\frac{1}{2}}, \tag{11}$$

which has the property

$$\eta^\dagger \eta = \mu^{-\frac{1}{2}} V^\dagger \Delta V \mu^{-\frac{1}{2}} = 1_p. \tag{12}$$

This is the canonical orthonormalization procedure[11,12]; it means that not only are (11) and (12) fulfilled, but also that in V_p we have

$$1 = \eta\eta^\dagger = \sum_{i=1}^{p} |\eta_i\rangle\langle\eta_i|, \tag{13}$$

i.e., the resolution of the identity. We also have

$$\phi = \eta \mu^{\frac{1}{2}} V^\dagger. \tag{14}$$

The transformations (11) and (14) are between two sets of different dimensionality, via rectangular matrices. It can be verified that $\mu^{\frac{1}{2}} V^\dagger$ is the explicit "generalized inverse" of $V\mu^{-\frac{1}{2}}$, which has no ordinary inverse since it is rectangular.[13]

We apply now the preceding results to a seemingly artificial problem in linear spaces which will turn out to be of importance in the next sections. Consider a linear space of dimension p. Assume that there are two sets g and f, each of them of dimension larger than p. We do not assume that either of them is linearly independent. In fact, the only assumptions made are that g contains a linear independent subset of order $r \leq p$ and, what is crucial, that the relation

$$1 = gg^\dagger + ff^\dagger \tag{15}$$

is satisfied. It looks like the resolution of the identity (13), but notice that the analogy is superficial. The g's and the f's form a set with linear dependencies and, of course, with no orthogonality conditions whatsoever imposed on it. Yet, (15) leads to interesting consequences. Canonical orthonormalization of g leads to

$$\chi = gV\mu^{-\frac{1}{2}} \tag{16}$$

and, hence,

$$gg^\dagger = \chi\mu\chi^\dagger. \tag{17}$$

The set χ is orthonormal and of dimension $r \leq p$. Therefore, the resolution of the identity in the space considered is

$$1 = \chi\chi^\dagger + \omega\omega^\dagger, \tag{18}$$

where ω is in the orthogonal complement to χ,

$$\omega^\dagger \chi = \chi^\dagger \omega = 0. \tag{19}$$

If $p = r$, (18) becomes simply $1 = \chi\chi^\dagger$. From (15), (17), and (18) it follows that

$$ff^\dagger = \chi(1_r - \mu)\chi^\dagger + \omega\omega^\dagger. \tag{20}$$

Furthermore, on account of (19),

$$\chi^\dagger f f^\dagger = (1_r - \mu)\chi^\dagger, \tag{21}$$

which means that the relation (15) imposes a bound on the nonvanishing eigenvalues of $g^\dagger g$:

$$\mu_i \leq 1, \quad i = 1, 2, \cdots, r. \tag{22}$$

Trivially,

$$g^\dagger \chi = V\mu^{\frac{1}{2}} \tag{23}$$

and

$$\chi^\dagger g g^\dagger \chi = \mu. \tag{24}$$

Also, from (20), we conclude that, if $\mu = 1_p$, then

$$ff^\dagger = \omega\omega^\dagger \tag{25}$$

and conversely. μ is idempotent in such a case.

FIRST-ORDER DENSITY MATRIX AND NATURAL ORBITALS

The first-order density matrix is defined in terms of the ground-state wavefunction of the N-particle system:

$$\gamma(1 \mid 1') = N \int \Psi(1, 2, \cdots, N) \times \Psi^*(1', 2, \cdots, N)\, dv_2 \cdots dv_N, \tag{26}$$

where 1 denotes the spin-space coordinates of one particle.[1] The kernel $\gamma(1 \mid 1')$ is associated with a nonnegative Hermitian operator γ_{op} of finite trace and, hence, is diagonalizable[1,2,14,15]:

$$\gamma(1 \mid 1') = \sum_i n_i \chi_i(1) \chi_i^*(1'). \tag{27}$$

The χ_i are the natural spin-orbitals[1] and the eigenvalues n_i are the usually called occupation numbers.

Since $\mathrm{Tr}\,\gamma(1 \mid 1') = N$, it follows that

$$\sum_i n_i = N. \tag{28}$$

The Hilbert–Schmidt theory of integral equations[15] implies that the spectrum of n_i consists of a non-increasing sequence of positive numbers which can be finite or infinite. In the latter case, it tends to zero and the χ_i form a complete set in L_2 with optimal convergence properties in the expansion of Ψ.[1,2]

An independent particle model is characterized by $n_1 = n_2 = \cdots = n_N$, in which case γ_{op} is idempotent. Upper and lower bounds for the n_i have been given by several authors. It is well known that[1,2,16]

$$n_i \leq 1. \tag{29}$$

This result has been proven using an expansion of the wavefunction in terms of ordered configurations (Slater determinants)[1] or by using the Schmidt theory[14] of integral equations.[2] Lower bounds have been given,[17] but they involve restricting assumptions on the system.

GREEN'S FUNCTIONS AND GENERALIZED OVERLAP AMPLITUDES

The Green's function[5] is defined by

$$G(1, t; 1', t') = -i\langle T\{\psi(1, t)\psi^\dagger(1', t')\}\rangle. \tag{30}$$

T is the time-ordering operator and the brackets indicate an expectation value with respect to the exact ground state of the interacting system. Atomic units are employed. From (30) and the second quantization form of (26), one gets

$$\gamma(1 \mid 1') = \langle\psi^\dagger(1)\psi(1')\rangle, \tag{31}$$

and it follows that[5,6,8,18]

$$\gamma(1 \mid 1') = -i \lim_{\delta \to 0+} G(1, 0; 1', -\delta). \tag{32}$$

From (30) one can obtain the spectral resolution

$$\begin{aligned} G(1, t; 1', t') &= -i \sum_s f_s(1) f_s^*(1') e^{-i\epsilon_s(t-t')}, \quad t > t', \\ &= i \sum_s g_s(1) g_s^*(1') e^{-i\epsilon_s(t'-t)}, \quad t' > t, \end{aligned} \tag{33}$$

where the f_s, g_s, and ϵ_s are defined in (1) and (2). One gets (33) from (30) by inserting a resolution of the identity between the field operators, where appropriate; this involves states of the $(N + 1)$- or the $(N - 1)$-particle system.

From (32) and (33) it readily follows that[8]

$$\gamma(1 \mid 1') = \sum_s g_s(1) g_s^*(1'). \tag{34}$$

This last equation should be compared with (27). They are both diagonal sums and the temptation of identifying the χ_i with the g_s is great.[6,10] If the g's were orthogonal, their norms would be the occupation numbers. Yet we know that the g's are linearly dependent.[8]

On the other hand, we know that, from the commutation relations of the field operators

$$\psi(1)\psi^\dagger(1') + \psi^\dagger(1')\psi(1) = \delta(1 - 1'), \tag{35}$$

one can obtain a completeness relation for the overlap amplitudes:

$$\sum_s f_s(1) f_s^*(1') + \sum_s g_s(1) g_s^*(1') = \delta(1 - 1'). \tag{36}$$

It is clear from (27) and (34) that we cannot identify the g_s with the χ_i, but there is on the other hand a clue in (36): We have two sets of functions, the f_s and the g_s, neither of which are necessarily linearly independent, but both of which satisfy the completeness relation (36).

NATURAL ORBITALS AND GENERALIZED OVERLAP AMPLITUDES

We notice that the density matrix is the kernel of a completely continuous transformation[19] and that the appearance of a continuum in the expansion (34) does not cause any formal trouble. In fact, this is an ordinary situation: If one formally expresses $\gamma(1 \mid 1')$ in terms of the eigenfunctions of a hydrogen-like operator with a continuum, diagonalization will lead to (27). This is not done in practice, just a conceptual possibility which can be justified.[18] The novelty is not the appearance of the continuum in (34), but the linear dependences.

At this stage we can employ the results which we derived previously with canonical orthonormalization. We assumed before that we dealt with a metric space of order p. Now we deal with an infinite-dimensional space where the completeness relation (36) holds. The set $\mathbf{g} = \{g_1, g_2, \cdots\}$ is, of course, infinite and not discrete, but since

$$\mathrm{Tr}\,\gamma(1 \mid 1') = N, \tag{37}$$

from (34) we conclude that we can form the metric matrix associated to $\mathbf{g}$, and that it has a finite trace, equal to N.

Canonical orthonormalization of the g_s leads to an orthonormal set, and with the preceding assumptions we see that (36) is an extension of (15) and that (17) expresses the interrelationship between (34) and (27). The natural orbitals are thus the canonically orthonormalized overlap amplitudes g_s. The eigenvalues

of the metric matrix associated with the g_s are the occupation numbers of the first-order density matrix.

There are a number of consequences of this connection. From (24) it follows that

$$\sum_s \langle \chi_i \mid g_s \rangle \langle g_s \mid \chi_j \rangle = n_i \delta_{ij}. \tag{38}$$

From (21) we have

$$\sum_s \langle \chi_i \mid f_s \rangle \langle f_s \mid \chi_j \rangle = (1 - n_i)\delta_{ij}. \tag{39}$$

It should be noted that (22) implies (29) and that this derivation of a bound for the occupation numbers required the following prerequisites:

(i) the anticommutation relation (35) in its form (36);

(ii) the formal definition of the Green's function (30) and its connection with the density matrix (32);

(iii) canonical orthonormalization of the set.

The independent-particle model is fully characterized by (25). If the density matrix is idempotent, then $\mathbf{g}^\dagger \mathbf{f} = \mathbf{0}$, i.e., the f's are orthogonal to the g's. Furthermore, the f's are associated with the zero eigenvalue of the density operator and thus represent "empty" orbitals; the set $\{g_s\}$ spans the N-dimensional subspace of the Hilbert space associated with the "occupied" orbitals in an independent-particle model.

We can gain further insight into the problem by examining in more detail the elements of the metric matrix $\mathbf{\Delta} = \mathbf{g}^\dagger \mathbf{g}$:

$$\Delta_{st} = \langle g_s \mid g_t \rangle. \tag{40}$$

Using the definition (1) and a first-quantization representation, we obtain

$$\begin{aligned}\Delta_{st} &= N \int \Psi_t^{N-1*}(2, \cdots, N) \Psi_0^N(1, \cdots, N) \\ &\quad \times \Psi_0^N(1, 2', \cdots, N') \Psi_s^{N-1}(2', \cdots, N')\, dv\, dv' \\ &= \int \Psi_t^{N-1*}(2, \cdots, N) \Gamma_{N,0}^{(N-1)}(2, \cdots, N \mid 2', \cdots, N') \\ &\quad \times \Psi_s(2', \cdots, N') \frac{dv\, dv'}{dv_1\, dv_{1'}}. \end{aligned} \tag{41}$$

Here Ψ_s^{N-1} is the wavefunction of the sth excited state of the $(N-1)$-particle system and $\Gamma_{N,0}^{(N-1)}$ the $(N-1)$th-order reduced density matrix of the ground-state wavefunction of the N-particle system. Therefore, the diagonalization of $\mathbf{\Delta}$ is nothing but the diagonalization of $\Gamma_{N,0}^{(N-1)}$ when the latter is expressed in terms of the eigenfunctions of the $(N-1)$-particle Hamiltonian. Equation (41) confirms that the nonvanishing eigenvalues of $\mathbf{\Delta}$ are the occupation numbers of $\gamma(1 \mid 1')$. This follows from the fact that the nonvanishing eigenvalues of $\Gamma_{N,0}^{(N-1)}$ and of $\Gamma_{N,0}^{(1)} \equiv \gamma$ are identical according to the Carlson–Keller theorem.[2,20] Canonical orthogonalization is required since the eigenvalue zero may occur in both $\Gamma_{N,0}^{(N-1)}$ and $\Gamma_{N,0}^{(1)}$ and, what is worse, with different multiplicity in each case.[20]

Except for a few, if any, bound states, the set f has a norm in the δ-function sense, and the procedure which was used to obtain the natural orbitals from the g's cannot be used. This is most easily seen by considering the quantity analogous to (40); i.e., we study the metric of the f's (we examine the diagonal elements only):

$$\begin{aligned}&\langle f_s \mid f_s \rangle \\ &\quad = \int f_s^*(1) f_s(1)\, dv_1 \\ &\quad = \int \Psi_0^{N*}(1, \cdots, N) \Gamma_{N+1,s}^{(N)}(1, \cdots, N \mid 1', \cdots, N') \\ &\qquad \times \Psi_0(1', \cdots, N')\, dv\, dv'. \end{aligned} \tag{42}$$

We see that the diagonal elements of the metric matrix associated with the f's are expectation values, with respect to the ground state of the N-particle system, of the Nth-order reduced density matrices of the sth excited states of the $(N+1)$-particle system. Since we are dealing with scattering states, the Schmidt theory does not apply in this case.

DISCUSSION

We have shown that the natural spin-orbitals[1] which diagonalize the first-order reduced density matrix are related to the generalized overlap amplitudes of Green's function theory[5–10] by the canonical orthonormalization procedure of Löwdin.[11,12]

This relation, which aroused interest in several contexts,[6,7–10] is established and it leads to orthogonality properties like (38) and (39) and other consequences which follow from (15)–(25).

The natural orbitals provide a basis for expanding the Green's function, for $t < t'$, in a natural way. For $t > t'$, we have no such simple property and a discussion of the zero eigenvalue of $\gamma(1 \mid 1')$ should precede any statement on this.

The results obtained in this paper exploit a connection between conceptual tools sometimes used in a mutually exclusive fashion. An immediate application is to the theory of "capture" and "ionization" processes.[10] Applications to atomic and molecular physics are under way. In treating finite systems, the modifications of external fields (e.g., nuclear

framework, center of mass, etc.) upon changes in the number of particles must be properly treated.

ACKNOWLEDGMENT

We are grateful to Professor P.-O. Löwdin for generous support and stimulating interest.

* Sponsored in part by the Swedish National Research Council and in part by the Air Force Office of Scientific Research (OSR) through the European Office of Aerospace Research (OAR), U.S. Air Force, under Grant EOOAR-69-0043.

[1] P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955); in *Advances in Chemical Physics*, I. Prigogine, Ed. (Interscience Publishers, New York, 1959), Vol. 2, p. 207.

[2] For a review, see A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963). For updated references and discussions, see *Reduced Density Matrices with Applications to Physical and Chemical Systems* (Queen's Papers on Pure and Applied Mathematics, No. 11), A. J. Coleman and R. M. Erdahl, Eds. (Queen's University, Kingston, Ontario, 1968).

[3] P.-O. Löwdin and H. Shull, Phys. Rev. **101**, A30 (1956). (This is the first "natural analysis" of a wavefunction and it has been followed by many others. See Ref. 2.)

[4] For a review, see J. Linderberg, Magy. Fiz. Folyóirat **16**, 5 (1968). See also J. Linderberg and Y. Öhrn, J. Chem. Phys. **49**, 716 (1968), and references therein. See also L. Hedin, Phys. Rev. **139**, A796 (1965).

[5] They can be found in any textbook on the Green's function method, e.g., A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).

[6] L. Hedin and S. Lundqvist, Uppsala University, Uppsala, Sweden, Quantum Chemistry Group, TIII, 1960 (unpublished).

[7] G. W. Pratt, Rev. Mod. Phys. **35**, 502 (1963).

[8] L. Hedin, University of Uppsala, Uppsala, Sweden, Quantum Chemistry Group, TN 84, 1962 (unpublished).

[9] L. Hedin, A. Johansson, B. I. Lundqvist, S. Lundqvist, and V. Samathiyakanit, Arkiv Fysik **39**, 97 (1967).

[10] T. Berggren, Nucl. Phys. **72**, 337 (1965); C. F. Clement, Phys. Letters **28B**, 395, 398 (1969).

[11] P.-O. Löwdin, Advan. Phys. **5**, 1 (1956). See also P.-O. Löwdin, Ann. Rev. Phys. Chem. **11**, 107 (1960).

[12] P.-O. Löwdin, University of Uppsala, Uppsala, Sweden, Quantum Chemistry Group, TN 213, 1968 (to be published); P.-O. Löwdin, Intern. J. Quantum Chem. **1S**, 811 (1967).

[13] E. H. Moore, *General Analysis* (American Philosophical Society, Philadelphia, 1935), Vol. 1; R. Penrose, Proc. Cambridge Phil. Soc. **51**, 406 (1955).

[14] E. Schmidt, Math. Ann. **63**, 433 (1907).

[15] F. G. Tricomi, *Integral Equations* (Interscience Publishers, New York, 1957).

[16] F. Sasaki, Phys. Rev. **138**, B1338 (1965).

[17] W. Kutzelnigg and V. H. Smith, J. Chem. Phys. **42**, 2791 (1965).

[18] M. H. Stone, *Linear Transformations in Hilbert Space* (American Mathematical Society Colloquium Publications, New York, 1932), Vol. XV, Chap. VII.

[19] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (F. Ungar Publishing Co., New York, 1955), p. 206.

[20] B. C. Carlson and J. M. Keller, Phys. Rev. **121**, 659 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Some Simple Observations on Griffiths' Theorems for the Classical Heisenberg Model*

ROBERT H. T. YEH
Department of Physics and Astronomy, State University of New York Buffalo, New York 14214

(Received 6 November 1969)

We show that Griffiths' theorems are valid for any classical ferromagnetic Heisenberg model in the weak interaction limit. They are also valid for certain chain-type and ring-type models, regardless of interaction strength.

Griffiths[1] has shown that, for an Ising ferromagnet in zero field, with arbitrary crystal structure and range of interaction, the spin correlation obeys two theorems. First, it is nonnegative (Griffiths' first theorem). Furthermore, it is a monotonic increasing function of the interactions (Griffiths' second theorem). These theorems are quite useful and have been generalized in various ways.[2] Recently, it was pointed out[3] that Griffiths' second theorem does not hold for the (quantum) ferromagnetic Heisenberg model because of the existence of a counterexample. It is natural to ask, what is the reason for this difference between the Ising model and the Heisenberg model? Is it due to the quantum nature of spin operators which give rise to some uncertainties in the orientation of each spin vector? Or is it due to the 1-dimensional nature of spin vectors of the Ising model which plays an important role in previous proofs of Griffiths' second theorem? We do not have a complete answer to these questions. However, we show that Griffiths' theorems are valid for the classical ferromagnetic Heisenberg model under certain conditions.

Let

$$\beta H = -\left(\sum_{i>j} J_{ij}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j + \sum_i B\sigma_{iz}\right),$$

where J_{ij} is the interaction parameter and B is the applied magnetic field. Since temperature is kept

framework, center of mass, etc.) upon changes in the number of particles must be properly treated.

ACKNOWLEDGMENT

We are grateful to Professor P.-O. Löwdin for generous support and stimulating interest.

* Sponsored in part by the Swedish National Research Council and in part by the Air Force Office of Scientific Research (OSR) through the European Office of Aerospace Research (OAR), U.S. Air Force, under Grant EOOAR-69-0043.

[1] P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955); in *Advances in Chemical Physics*, I. Prigogine, Ed. (Interscience Publishers, New York, 1959), Vol. 2, p. 207.

[2] For a review, see A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963). For updated references and discussions, see *Reduced Density Matrices with Applications to Physical and Chemical Systems* (Queen's Papers on Pure and Applied Mathematics, No. 11), A. J. Coleman and R. M. Erdahl, Eds. (Queen's University, Kingston, Ontario, 1968).

[3] P.-O. Löwdin and H. Shull, Phys. Rev. **101**, A30 (1956). (This is the first "natural analysis" of a wavefunction and it has been followed by many others. See Ref. 2.)

[4] For a review, see J. Linderberg, Magy. Fiz. Folyóirat **16**, 5 (1968). See also J. Linderberg and Y. Öhrn, J. Chem. Phys. **49**, 716 (1968), and references therein. See also L. Hedin, Phys. Rev. **139**, A796 (1965).

[5] They can be found in any textbook on the Green's function method, e.g., A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).

[6] L. Hedin and S. Lundqvist, Uppsala University, Uppsala, Sweden, Quantum Chemistry Group, TIII, 1960 (unpublished).

[7] G. W. Pratt, Rev. Mod. Phys. **35**, 502 (1963).

[8] L. Hedin, University of Uppsala, Uppsala, Sweden, Quantum Chemistry Group, TN 84, 1962 (unpublished).

[9] L. Hedin, A. Johansson, B. I. Lundqvist, S. Lundqvist, and V. Samathiyakanit, Arkiv Fysik **39**, 97 (1967).

[10] T. Berggren, Nucl. Phys. **72**, 337 (1965); C. F. Clement, Phys. Letters **28B**, 395, 398 (1969).

[11] P.-O. Löwdin, Advan. Phys. **5**, 1 (1956). See also P.-O. Löwdin, Ann. Rev. Phys. Chem. **11**, 107 (1960).

[12] P.-O. Löwdin, University of Uppsala, Uppsala, Sweden, Quantum Chemistry Group, TN 213, 1968 (to be published); P.-O. Löwdin, Intern. J. Quantum Chem. **1S**, 811 (1967).

[13] E. H. Moore, *General Analysis* (American Philosophical Society, Philadelphia, 1935), Vol. 1; R. Penrose, Proc. Cambridge Phil. Soc. **51**, 406 (1955).

[14] E. Schmidt, Math. Ann. **63**, 433 (1907).

[15] F. G. Tricomi, *Integral Equations* (Interscience Publishers, New York, 1957).

[16] F. Sasaki, Phys. Rev. **138**, B1338 (1965).

[17] W. Kutzelnigg and V. H. Smith, J. Chem. Phys. **42**, 2791 (1965).

[18] M. H. Stone, *Linear Transformations in Hilbert Space* (American Mathematical Society Colloquium Publications, New York, 1932), Vol. XV, Chap. VII.

[19] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (F. Ungar Publishing Co., New York, 1955), p. 206.

[20] B. C. Carlson and J. M. Keller, Phys. Rev. **121**, 659 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Some Simple Observations on Griffiths' Theorems for the Classical Heisenberg Model*

ROBERT H. T. YEH
Department of Physics and Astronomy, State University of New York Buffalo, New York 14214

(Received 6 November 1969)

We show that Griffiths' theorems are valid for any classical ferromagnetic Heisenberg model in the weak interaction limit. They are also valid for certain chain-type and ring-type models, regardless of interaction strength.

Griffiths[1] has shown that, for an Ising ferromagnet in zero field, with arbitrary crystal structure and range of interaction, the spin correlation obeys two theorems. First, it is nonnegative (Griffiths' first theorem). Furthermore, it is a monotonic increasing function of the interactions (Griffiths' second theorem). These theorems are quite useful and have been generalized in various ways.[2] Recently, it was pointed out[3] that Griffiths' second theorem does not hold for the (quantum) ferromagnetic Heisenberg model because of the existence of a counterexample. It is natural to ask, what is the reason for this difference between the Ising model and the Heisenberg model? Is it due to the quantum nature of spin operators which give rise to some uncertainties in the orientation of each spin vector? Or is it due to the 1-dimensional nature of spin vectors of the Ising model which plays an important role in previous proofs of Griffiths' second theorem? We do not have a complete answer to these questions. However, we show that Griffiths' theorems are valid for the classical ferromagnetic Heisenberg model under certain conditions.

Let

$$\beta H = -\left(\sum_{i>j} J_{ij}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j + \sum_i B\sigma_{iz}\right),$$

where J_{ij} is the interaction parameter and B is the applied magnetic field. Since temperature is kept

constant in our calculation, we absorb the temperature factor β into J_{ij} and B. We note that, if all the interaction parameters are identical or $J_{ij} = J \geq 0$, then the canonical ensemble average of $\sum_{i>j} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j$ is a monotonic increasing function of J. This holds for arbitrary crystal structure and range of interaction, and $\boldsymbol{\sigma}$ may be either a classical spin vector or a quantum operator. Also, $\boldsymbol{\sigma}$ may be of any dimension or magnitude.

The proof is as follows. It is obvious that $Z = \operatorname{Tr} e^{-\beta H} \geq 1$ and

$$\operatorname{Tr}\left[\left(\sum_{i>j}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right)^2 e^{-\beta H}\right] \geq \left\{\operatorname{Tr}\left[\left(\sum_{i>j}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right) e^{-\beta H}\right]\right\}^2,$$

so that

$$\begin{aligned}\frac{\partial}{\partial J}\left\langle \sum_{i>j}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \right\rangle &= \frac{\partial^2}{\partial J^2}\log Z \\ &= Z^{-2}\left\{Z \operatorname{Tr}\left[\left(\sum_{i>j}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right)^2 e^{-\beta H}\right]\right. \\ &\quad \left. - \left(\operatorname{Tr}\left[\left(\sum_{i>j}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right) e^{-\beta H}\right]\right)^2\right\} \geq 0. \end{aligned} \tag{1}$$

From now on, we assume that $B = 0$. Let us consider a classical Heisenberg model of arbitrary crystal structure and range of interaction. The partition function is

$$Z = \operatorname{Tr}\left(\exp \sum_{i>j} J_{ij}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right), \quad J_{ij} \geq 0.$$

Tr indicates integration over all possible directions of $\boldsymbol{\sigma}$'s. Let us classify all crystal structures by connecting all $J_{ij} \neq 0$ bonds. If there is at least one ring formed by these connected bonds, we call the model a ring type; otherwise, we call it a chain type. For the chain type, we can change integration variables so that all the J_{ij} are decoupled and $Z = \prod_{i>j} f(J_{ij})$. [For example, if we have a linear chain of N spins with nearest-neighbor interactions, then

$$Z = (2\pi)^N I_0(J_{12}) I_0(J_{23}) \cdots I_0(J_{N-1,N})$$

for 2-dimensional spins and

$$Z = (4\pi)^N f(J_{12}) f(J_{23}) \cdots f(J_{N-1,N})$$

for 3-dimensional spins. I_0 is the modified Bessel function and $f(J) = (\sinh J)/J$.] It is clear that $\partial^2 \log Z/\partial J_a \partial J_b = 0$ if $J_a \neq J_b$. This is true, independent of the dimension of the spin vector. Furthermore, for 2- and 3-dimensional spins, it is easy to see that

$$\langle \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \rangle = \frac{\partial \log Z}{\partial J_{ij}} \geq 0.$$

So both theorems of Griffiths are valid for chain-type models.

Ring-type models are more complicated and we restrict our discussions to the weak-coupling limit (i.e., all the $J_{ij} \to 0$). In this limit, we can expand $\log (Z/Z_0)$ in powers of J, where Z_0 is the value of Z when all the $J_{ij} = 0$. For 2- or 3-dimensional spins, we find that

$$\log (Z/Z_0) = L_1 + L_2 + L_3 + \cdots, \tag{2}$$

where L_i is of the order J^i and

$$\begin{aligned} L_1 &= 0, \\ L_2 &= \sum_{i>j} C_{ij} J_{ij}^2, \qquad C_{ij} \geq 0, \\ L_3 &= \sum C_{ijk} J_{ij} J_{jk} J_{ki}, \quad C_{ijk} \geq 0. \end{aligned} \tag{3}$$

It is clear that $\partial L_2/\partial J_a \geq 0$ and $\partial^2 L_3/\partial J_a \partial J_b \geq 0$; hence, both theorems of Griffiths are valid. If the smallest ring in the structure contains n bonds, then the first nonvanishing term of $\sum_i \partial^2 L_i/\partial J_a \partial J_b$ has $i = n$, and it is nonnegative. (For $i < n$, the system is essentially like chain type; hence, $\partial^2 L_i/\partial J_a \partial J_b = 0$ for $J_a \neq J_b$.)

To illustrate these results, let us consider a concrete example of three 2-dimensional spins $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_3$. They form a ring by the bonds J_{12}, J_{23}, J_{31}. The partition function Z of the system is

$$\begin{aligned} Z = \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \int_0^{2\pi} d\theta_3 \\ \times \exp\left[J_{12}\cos(\theta_1 - \theta_2) + J_{23}\cos(\theta_2 - \theta_3)\right. \\ \left. + J_{31}\cos(\theta_3 - \theta_1)\right]. \end{aligned} \tag{4}$$

In the limit of small J's, we easily find that

$$\log (Z/Z_0) = \tfrac{1}{4}(J_{12}^2 + J_{23}^2 + J_{31}^2) + \tfrac{1}{4} J_{12} J_{23} J_{31}, \tag{5}$$

where $Z_0 = (2\pi)^3$.

Actually, Griffiths' second theorem is valid for the above model and some other models, regardless of interaction strength. Let us consider a lattice structure of 2-dimensional spins. We assume that there are three spins $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_3$, in which $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_3$ interact only with $\boldsymbol{\sigma}_2$ and with each other, but $\boldsymbol{\sigma}_2$ may interact with any number of spins. We also assume that $J_{12} = J_{23}$ $(= J)$, and we let $J_{31} = J'$. Then Griffiths' second theorem is valid for the three spins in this ring; that is, $\partial^2 \log Z/\partial J \partial J' \geq 0$.

To prove this, we first make a variable transformation from $(\theta_1, \theta_2, \theta_3)$ to $(\theta_{12}, \theta_2, \theta_{23})$ in the partition function Z, where $\theta_{12} = \theta_1 - \theta_2$ and $\theta_{23} = \theta_2 - \theta_3$. Then $Z = AZ_R$, where Z_R denotes the ring contribution to the partition function and A is independent of J

and J':

$$Z_R = \int_0^{2\pi} d\theta_{12} \int_0^{2\pi} d\theta_{23} \times \exp\left[J(\cos\theta_{12} + \cos\theta_{23}) + J'\cos(\theta_{12} + \theta_{23})\right] = 8\pi \int_0^{\frac{1}{2}\pi} d\alpha e^{J'\cos 2\alpha} I_0(2J\cos\alpha). \tag{6}$$

We note that

$$\langle \sigma_1 \cdot \sigma_2 \rangle = \langle \sigma_2 \cdot \sigma_3 \rangle = \frac{1}{2}\frac{\partial \log Z}{\partial J} = \frac{1}{2}\frac{\partial \log Z_R}{\partial J} \geq 0,$$

$$\langle \sigma_1 \cdot \sigma_3 \rangle = \frac{\partial \log Z}{\partial J'} = \frac{\partial \log Z_R}{\partial J'} \geq 0. \tag{7}$$

From Eq. (6), we obtain

$$\frac{\partial^2 \log Z}{\partial J' \partial J} = \left(\int_0^{\frac{1}{2}\pi} d\gamma e^{J'\cos 2\gamma} I_0(2J\cos\gamma)\right)^{-2} \times \int_0^{\frac{1}{2}\pi} d\alpha \int_0^{\frac{1}{2}\pi} d\beta e^{J'(\cos 2\alpha + \cos 2\beta)} \times I_0(2J\cos\alpha) I_0(2J\cos\beta)(\cos^2\alpha - \cos^2\beta) \times \left(\cos\alpha \frac{I_1(2J\cos\alpha)}{I_0(2J\cos\alpha)} - \cos\beta \frac{I_1(2J\cos\beta)}{I_0(2J\cos\beta)}\right). \tag{8}$$

It is clear that $\partial^2 \log Z/\partial J' \partial J$ is nonnegative provided that

$$\frac{I_1(x)}{I_0(x)} = \frac{d}{dx}\log I_0(x)$$

is a monotonic increasing function of x, i.e.,

$$\frac{d^2}{dx^2}\log I_0(x) \geq 0.$$

This is indeed the case, since

$$\frac{d^2}{dx^2}\log I_0(x) = 2[I_0(x)]^{-2} \times \int_0^{\pi} d\theta \int_0^{\pi} d\theta' e^{x(\cos\theta + \cos\theta')}(\cos\theta - \cos\theta')^2 \geq 0. \tag{9}$$

From this we conclude that both theorems of Griffiths' are valid for all J_{ij}, for chain-type models of 2-dimensional spins with three-spin rings of the above kind decorated at the tips.

It is of interest to note that Griffiths' second theorem is also valid for a similar quantum Heisenberg ring model. Consider again a simple ring of three spins, with interactions $J_{12} = J_{23} = J$ and $J_{31} = J'$. The correlation function is obtained from a modified partition function[3]

$$Z^* = \mathrm{Tr}\,[\exp 2(JP_{12} + JP_{23} + J'P_{31})],$$

where $P_{kl} = \frac{1}{2}(1 + \sigma_k \cdot \sigma_l)$ and Z^* differs from Z only by a constant factor

$$\frac{\partial \log Z}{\partial J_{kl}} = \frac{\partial \log Z^*}{\partial J_{kl}} - 1.$$

Eigenvalues of $J(P_{12} + P_{23}) + J'P_{31}$ are easily found to be $2J + J'$, $2J + J'$, $J - J'$, and $J' - J$: each is doubly degenerate. So

$$Z^* = 4[e^{2(2J+J')} + \cosh 2(J - J')]. \tag{10}$$

From this we obtain

$$\frac{\partial^2 \log Z}{\partial J \partial J'} = 4[e^{4J+2J'} + \cosh 2(J - J')]^{-2}(e^{6J} - 1) \geq 0. \tag{11}$$

Hence, Griffiths' second theorem is valid for this model for all values of positive J and J'.

* This work was supported by a faculty research fellowship from the Research Foundation of State University of New York.

[1] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).

[2] J. Ginibre, Phys. Rev. Letters **23**, 828 (1969), and references cited therein; R. B. Griffiths, "Rigorous Results for Ising Ferromagnets of Arbitrary Spin" (we wish to thank Professor Griffiths for sending us this preprint before publication); D. C. Mattis and M. Plischke, J. Math. Phys. **10**, 1107 (1969); R. Yeh, Phys. Letters **30A**, 297 (1969).

[3] C. A. Hurst and S. Sherman, Phys. Rev. Letters **23**, 1357 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

The Adiabatic Invariant of the Linear or Nonlinear Oscillator*

K. R. SYMON
Department of Physics, University of Wisconsin, Madison, Wisconsin 53706

(Received 4 June 1969)

H. R. Lewis's treatment of the time-dependent harmonic oscillator is extended to the general time-dependent nonlinear oscillator. A prescription is given for obtaining an adiabatic invariant as a power series in the amplitude and to any order in the time variation of the coefficients. An exact invariant of the same algebraic form as the adiabatic invariant is also derived. The exact invariant can be so chosen that, at any particular time, it is equal to the adiabatic invariant. Comparison of the adiabatic and exact invariants permits a calculation of the nonadiabatic changes in the adiabatic invariant in the course of time. Applications to a linear and a nonlinear example are presented and the statistical distribution of the long time changes in the adiabatic invariants are determined in the two cases.

I. INTRODUCTION

In this paper, we propose to extend Lewis's treatment[1] of the slowly varying harmonic oscillator to the general case of a slowly varying Hamiltonian in one degree of freedom, which we suppose can be expanded in powers of the coordinate and momentum

$$H(x, p, t) = H_2(x, p, t) + H_3(x, p, t) + \cdots, \quad (1)$$

where

$$H_2(x, p, t) = \tfrac{1}{2}a(t)p^2 + b(t)xp + \tfrac{1}{2}c(t)x^2 \quad (2)$$

is a homogeneous quadratic function of the variables x and p with slowly varying coefficients, and where H_3, H_4, etc., are likewise polynomials of degrees 3, 4, etc. It turns out that Lewis's treatment, when so extended, yields convenient practical procedures for calculating adiabatic invariants to at least the fourth power in the amplitudes of oscillation and to fairly high orders in the time derivatives of the coefficients in $H(x, p, t)$, and, furthermore, yields practical methods for calculating the statistical distribution of the nonadiabatic changes for the adiabatic invariants over a period of time for any particular time variation of $H(x, p, t)$. The method enables us to treat separately the nonlinearities and the time variations of the coefficients. It also allows us to distinguish transient variations in the lower-order adiabatic invariants, which can be eliminated by calculating corrections to higher order in the time derivatives, from genuinely nonadiabatic changes, which can be calculated by comparing an exact invariant with the adiabatic invariant.

Throughout this paper, the term "exact invariant" means a function $J(x, p, t)$ which is either exactly constant, in the linear case, in virtue of the equations of motion derived from the Hamiltonian (2), or, in the nonlinear case, a function which is constant to some order in the amplitude of the motion and independently of the time rates of change of the coefficients in the Hamiltonian (1). By "adiabatic invariant," we mean a function $J(x, p, t)$ which is approximately constant when the coefficients in the Hamiltonian are slowly varying; more precisely, a function whose time derivative contains only terms higher than some particular order in the time derivatives of the coefficients of the Hamiltonian. In addition, we require the adiabatic invariant $J(x, p, t)$ at any time to be expressible in terms of the coefficients in the Hamiltonian $H(x, p, t)$, and possibly their time derivatives, evaluated at the time t; this is in contrast with the exact invariant which in general depends on the entire history of the oscillator, that is, it involves integrals over t.

We have omitted a linear term H_1, so that $x = p = 0$ is a solution of the equations of motion. In the general case when a linear term is present, we may choose any solution of the equations of motion, which we may call the *equilibrium solution* or *reference solution*, and the variables x, p in Eq. (1) are measured relative to the reference solution. A sufficiently small linear term H_1 may, however, be treated as a perturbation, and introduced after the linear problem arising from H_2 has been solved.

We are interested only in the case in which the solutions given by the Hamiltonian (1) are oscillatory when the coefficients a, b, c, etc., are held constant. We further assume that the amplitudes are small enough so that expansions in powers of the variables are appropriate, and, in fact, we assume that successively higher-order terms are successively smaller in magnitude, except for an occasional term which may vanish.

In Sec. II, we consider the solution of the linear problem given by the Hamiltonian H_2. Lewis[1] has considered the special case in which the Hamiltonian has the form

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2(t)x^2. \quad (3)$$

He discovered an exact invariant of the equations of

motion which is very closely related to the adiabatic invariant for the oscillator (3). In fact, Lewis's method also gives a very convenient way of calculating the adiabatic invariant to any desired order. We derive essentially the same result in Sec. II by a somewhat different procedure for the more general Hamiltonian (2). We obtain the invariant by making a linear transformation to a set of variables X and P, such that the new Hamiltonian takes the form

$$K_2(X, P, t) = \tfrac{1}{2}\Omega(t)(X^2 + P^2). \tag{4}$$

The new equations of motion now yield the exact invariant

$$J = \tfrac{1}{2}(X^2 + P^2) = K_2/\Omega. \tag{5}$$

Note that the action J is also the adiabatic invariant for the Hamiltonian K_2; we shall see that it is closely related to the adiabatic invariant for the Hamiltonian H_2, of which J, when expressed in terms of the variables x and p, is an exact invariant. We may also transform to polar coordinates J, γ in the X, P plane by means of a canonical transformation.

The linear transformation of x, p to X, P or J, γ is also made in the higher-order terms H_3, etc. In Sec. III, we show how to construct a series of canonical transformations which transform away successively higher terms H_3, H_4, etc., except for certain terms which depend only upon the action J. Thus, to any order in x and p, we are able to reduce the Hamiltonian (1) to the form

$$K(J, t) = \Omega(t)J + \tfrac{1}{2}k_2(t)J^2 + \cdots. \tag{6}$$

The action J is, therefore, an exact invariant of the equations of motion to this order. The coefficients in the power series which expresses J in terms of x and p are functions which can be so chosen that at any particular time t_0, J is exactly equal to the adiabatic invariant at that time t_0. The transformation from x and p to J and γ generally depends upon the history of the system between t_0 and the later time t, that is, it depends upon the coefficients in the Hamiltonian during that time period. However, an approximate transformation can be derived to any order in x and p, and also to any order in the rates of change of the coefficients, which depends only upon the coefficients and their derivatives at a particular time t at which J is to be evaluated. The resulting J is the adiabatic invariant and is approximately equal to the exact invariant if the coefficients vary sufficiently slowly. Since the exact and adiabatic invariants are equal at t_0, the difference between them at any later time t is a measure of the change in the adiabatic invariant. Examples are given in Sec. IV.

We note, in passing, that the frequency of the nonlinear oscillator at any particular time is given as a function of the amplitude J by the equation

$$\frac{d\gamma}{dt} = \frac{\partial K}{\partial J} = \Omega + k_2 J + \cdots. \tag{7}$$

In constructing an invariant for the linear problem given by H_2, we make use of the phase–amplitude form of the solution of the equations of motion. The phase–amplitude solution was used in a similar way by Courant and Snyder[2] to construct an invariant for the case when the coefficients in Eq. (2) are periodic functions of the time. Our method of attacking the nonlinear problem by a sequence of successive canonical transformations is analogous to the method used by Birkhoff[3] in treating the corresponding dynamical problem when the coefficients are periodic functions of the time. Birkhoff's method was later extended by Moser[4] to the case when there is a resonance between the oscillator frequency and the periodicity of the coefficients; this case, of course, does not occur in our problem, where the coefficients are all taken to be slowly varying functions of the time. The method could, however, be easily extended to the case of periodic coefficients with slowly varying amplitudes and phases.

II. THE LINEAR PROBLEM

The linear equations of motion which follow from the Hamiltonian H_2 are

$$\begin{aligned} \dot{x} &= ap + bx, \\ \dot{p} &= -bp - cx. \end{aligned} \tag{8}$$

We assume that periodic or approximately periodic solutions of these equations exist and that a particular complex solution $\mathbf{x}$, $\mathbf{p}$ is given which we write in the form

$$\mathbf{x} = we^{i(\phi+\chi_1)}, \quad \mathbf{p} = ize^{i(\phi+\chi_2)}, \tag{9}$$

where w, z, χ_1, χ_2, and ϕ are real functions, where $w > 0$ and $z > 0$, and we assume that the solution (9) is genuinely complex; that is, not both χ_1 and χ_2 are constant. The complex-conjugate solution $\mathbf{x}^*$ and $\mathbf{p}^*$ is then an independent solution of Eqs. (8). The phase ϕ is to be given by

$$\phi = \int \Omega(t)\, dt, \tag{10}$$

where $\Omega(t)$ is a frequency which we may choose to suit our convenience; we intend that $\Omega(t)$ at least approximate the (varying) oscillator frequency so that w, z, χ_1, and χ_2 are slowly varying under adiabatic or nearly adiabatic conditions. We agree to normalize the

solution (9) by setting the Wronskian of that solution with its complex conjugate equal to $2i$:

$$(\mathbf{x}^*\mathbf{p} - \mathbf{x}\mathbf{p}^*)/2i = wz\cos(\chi_1 - \chi_2) = 1. \quad (11)$$

The Wronskian is constant, as may readily be verified by direct substitution in the equations of motion (8).

By substituting the solution (9) in the equations of motion and using the normalization (11), we can obtain the following differential equations for the functions w, z, and ϕ:

$$\dot{w} - bw = az\sin(\chi_1 - \chi_2), \quad (12)$$

$$\dot{z} + bz = -cw\sin(\chi_1 - \chi_2), \quad (13)$$

$$\dot{\phi} = \Omega = a/2w^2 + c/2z^2 - \dot{\chi}_1 - \dot{\chi}_2. \quad (14)$$

By squaring Eq. (12) and using Eq. (11), we can express z as a function of w and its first derivative:

$$z^2 = w^{-2} + (\dot{w} - bw)^2a^{-2}. \quad (15)$$

If we insert Eq. (15) in the left-hand side of Eq. (13) and use Eq. (12) to eliminate $\sin(\chi_1 - \chi_2)$, we obtain

$$\ddot{w} - a^{-1}\dot{a}\dot{w} + (ac - b^2 - \dot{b} + a^{-1}\dot{a}b)w = a^2w^{-3}. \quad (16)$$

The term involving $\dot{w}$ can be eliminated by defining a new variable $a^{-\frac{1}{2}}w$ (see Appendix). We assume that an exact or suitable approximate solution of Eq. (16) is available for the amplitude function w, and we use this function to construct the invariant J. In the case of the Hamiltonian (3), for which $a = 1$, $b = 0$, and $c = \omega^2$, Eq. (16) reduces to Lewis's Eq. (4)[1] which is the same as Eq. (3.2) of Courant and Snyder.[2]

If we multiply solution (9) by an arbitrary complex constant $Ae^{i\theta}$ and take the imaginary part, we obtain the most general real solution of Eq. (8), which we write in the form

$$x = Aw\cos\chi_1\sin(\phi + \theta) + Aw\sin\chi_1\cos(\phi + \theta),$$

$$p = Az\sin\chi_2\sin(\phi + \theta) + Az\cos\chi_2\cos(\phi + \theta). \quad (17)$$

We now make a linear transformation to variables X and P:

$$x = Xw\cos\chi_1 + Pw\sin\chi_1,$$
$$p = Xz\sin\chi_2 + Pz\cos\chi_2. \quad (18)$$

These equations may be solved for X and P, making use of Eq. (11):

$$X = xz\cos\chi_2 - pw\sin\chi_1,$$
$$P = -xz\sin\chi_2 + pw\cos\chi_1. \quad (19)$$

Since the Poisson bracket $(x, p)_{XP}$ is 1 by Eq. (11), the transformation (18) is canonical. We see from Eqs. (17) and (18) that

$$X = A\sin(\phi + \theta), \quad P = A\cos(\phi + \theta), \quad (20)$$

which is the general solution of the equations of motion which follow from the Hamiltonian (4). It is not difficult to construct a generating function for the transformation (18) and to verify that it transforms the Hamiltonian H_2 [Eq. (2)] into the Hamiltonian K_2 [Eq. (4)].

We immediately see from Eq. (20) that the following quantity is a constant of the motion:

$$J = \tfrac{1}{2}(X^2 + P^2) = \tfrac{1}{2}z^2x^2 - zwxp\sin(\chi_1 - \chi_2) + \tfrac{1}{2}w^2p^2. \quad (21)$$

By using Eqs. (11) and (15), we can express J in terms of the single function w and its derivative

$$J = \tfrac{1}{2}w^{-2}x^2 + \tfrac{1}{2}[wp - (\dot{w} - bw)a^{-1}x]^2. \quad (22)$$

In case $a = 1$, $b = 0$, this is the invariant obtained by Lewis.[1] If the coefficients a, b, and c are periodic functions of t, and we choose for w the periodic solution of Eq. (16), then J is the familiar invariant of the Hill equation.[2]

For any time-dependent coefficients a, b, and c, if w is a solution of Eq. (16), then J as given by Eq. (22) is a constant of the motion. If a, b, and c, are constant, Eq. (16) has the solution

$$w = \left(\frac{a}{\omega}\right)^{\frac{1}{2}}, \quad (23)$$

where

$$\omega = (ac - b^2)^{\frac{1}{2}}. \quad (24)$$

When the coefficients are slowly varying, Eq. (23) gives a solution for w which satisfies Eq. (16) to the zeroth order in the time derivatives. If this solution for w is substituted in Eq. (22) and $\dot{w}$ neglected, we obtain the familiar zero-order expression for the adiabatic invariant,[5] $J = H_2/\omega$. Equation (16) can be solved in a straightforward way for w to any desired order in the time derivatives by successive approximations starting with Eq. (23). If the resulting w is substituted in Eq. (22), we obtain a formula for the adiabatic invariant J to the corresponding order in the time derivatives. If at any time t_0, we start with initial values w_0 and $\dot{w}_0$ given by Eq. (23) or by some higher-order adiabatic solution of Eq. (16), and then integrate Eq. (16) to find w at any later time t, then this w, when substituted in Eq. (22), gives an exact invariant J which is initially equal to the adiabatic invariant. The difference between the two at time t then gives the change in the adiabatic invariant.

We note that if the coefficients a, b, and c are constant, we can readily write down the general solution of Eqs. (8), and hence we can write down the general solution of the nonlinear Eq. (16) for w, which turns out to be

$$w^2 = (a^2\omega^{-2} + A^2 + B^2)^{\frac{1}{2}} + A\cos 2\omega t + B\sin 2\omega t, \quad (25)$$

where ω is given by Eq. (24) and A and B are arbitrary constants. If the constants A and B are zero, Eq. (25) yields the adiabatic solution (23). Thus, if we consider a case where the coefficients a, b, and c are constant, both initially and finally, then the solution w is given initially by Eq. (23) and finally by Eq. (25), where the constants A and B characterize the deviation from adiabaticity. We use this result in Sec. IV to study the nonadiabatic change in the adiabatic invariant for the linear oscillator. Equation (25), when A and B are allowed to be slowly varying quantities, is also a convenient starting point for developing approximate nonadiabatic solutions to Eq. (16). (See Appendix.)

III. THE NONLINEAR PROBLEM

If we make the canonical transformation (18), the nonlinear Hamiltonian (1) becomes

$$K = \tfrac{1}{2}\Omega(X^2 + P^2) + K_3(X, P, t) + \cdots, \quad (26)$$

where K_3, K_4, etc., are obtained from H_3, H_4, etc., by simply substituting from Eq. (18).

We may introduce polar coordinates J, γ in the phase plane for which X, P are rectangular coordinates by means of the following canonical transformation:

$$X = (2J)^{\frac{1}{2}}\sin\gamma, \quad P = (2J)^{\frac{1}{2}}\cos\gamma. \quad (27)$$

It is readily verified that this transformation satisfies the Poisson-bracket condition $(X, P)_{J\gamma} = 1$. The Hamiltonian $K(J, \gamma)$ is obtained from the Hamiltonian $K(X, P)$ by a simple substitution from Eq. (27) and we therefore use the same symbol K for the Hamiltonian in either case. If K is independent of γ, then J is a constant of the motion and J and γ are called action-angle variables.[5]

If we make the substitution (27) in the Hamiltonian (26), we obtain

$$K = \Omega J + K_3(J, \gamma, t) + \cdots. \quad (28)$$

Since K_n is a polynomial of degree n in X and P, it is evident that with a little trigonometric rearrangement we may write

$$K_n(J, \gamma, t) = \sum_l K_{nl}(t)(2J)^{\frac{1}{2}n}\cos[l\gamma + \beta_{nl}(t)], \quad (29)$$

where the sum is over all integers l less than or equal to n and having the same parity (even or odd) as n. We now show how to construct a sequence of transformations which successively eliminate the γ dependence from the terms K_3, K_4, etc. At any stage in this process, the transformed variables to nth order $J_{(n)}$, $\gamma_{(n)}$ may be thought of as polar coordinates in a phase space whose rectangular coordinates $X_{(n)}$, $P_{(n)}$ are given by Eq. (27).

Suppose that we have succeeded in transforming away the γ dependence of all terms through K_{n-1}, and that the resulting canonical variables are $J_{(n-1)}$ and $\gamma_{(n-1)}$. The Hamiltonian then has the form

$$K_{(n-1)} = \Omega J_{(n-1)} + (\text{terms of order} < n) + \sum_l K_{nl(n-1)}(t)(2J_{(n-1)})^{\frac{1}{2}n} \times \cos[l\gamma_{(n-1)} + \beta_{nl(n-1)}(t)] + (\text{terms of order} > n). \quad (30)$$

We now transform to nth-order variables by means of the generating function

$$S = J_{(n)}\gamma_{(n-1)} + \sum_l S_{nl}(t)(2J_{(n)})^{\frac{1}{2}n}e^{il\gamma_{(n-1)}} + \text{c.c.}, \quad (31)$$

which yields the transformation

$$J_{(n-1)} = \frac{\partial S}{\partial\gamma_{(n-1)}} = J_{(n)} + \sum_l ilS_{nl}(2J_{(n)})^{\frac{1}{2}n}e^{il\gamma_{(n-1)}} + \text{c.c.}, \quad (32)$$

$$\gamma_{(n)} = \frac{\partial S}{\partial J_{(n)}} = \gamma_{(n-1)} + \sum_l nS_{nl}(2J_{(n)})^{\frac{1}{2}n-1}e^{il\gamma_{(n-1)}} + \text{c.c.}, \quad (33)$$

$$K_{(n)} = K_{(n-1)} + \frac{\partial S}{\partial t} = K_{(n-1)} + \sum_l \dot{S}_{nl}(2J_{(n)})^{\frac{1}{2}n}e^{il\gamma_{(n-1)}} + \text{c.c.}, \quad (34)$$

where "c.c." stands for "complex conjugate." If we substitute from Eqs. (30) and (32) in Eq. (34), we can see that we obtain

$$K_{(n)} = \Omega J_{(n)} + (\text{terms of order} < n) + \sum_l (\dot{S}_{nl} + il\Omega S_{nl} + \tfrac{1}{2}K_{nl(n-1)}e^{i\beta_{nl(n-1)}})(2J_{(n)})^{\frac{1}{2}n}e^{il\gamma_{(n-1)}} + \text{c.c.} + (\text{terms of order} > n). \quad (35)$$

The terms of order less than n in Eq. (35) are of precisely the same form as in Eq. (30) with $J_{(n)}$ substituted for $J_{(n-1)}$. The terms of order greater than n are in general modified by the substitution (32). If l

is not zero, we can eliminate the corresponding term from the sum in Eq. (35) by requiring that S_{nl} satisfy the differential equation

$$\dot{S}_{nl} + il\Omega S_{nl} = -\tfrac{1}{2}K_{nl(n-1)}(t)e^{i\beta_{nl(n-1)}(t)} \quad (36)$$

We could also eliminate the terms with $l = 0$, but the resulting transformation would contain secular terms increasing linearly with the time, even when the Hamiltonian is constant or slowly varying, a situation we prefer to avoid. Since l and n have the same parity, we see that the transformation (31) eliminates all terms of order n when n is odd, and eliminates all terms except that with $l = 0$ when n is even. To complete the transformation, Eq. (33) must be solved for $\gamma_{(n-1)}$ by successive approximations to whatever order we wish to carry out the procedure. The result is then to be substituted in the terms of order n and higher to obtain the Hamiltonian $K_{(n)}$ in terms of nth order variables. The result is clearly of the form

$$K_{(n)} = \Omega J_{(n)} + \sum_{n'} \frac{1}{n'} k_{n'}(t) J_{(n)}^{n'} + (\text{terms of order} > n), \quad (37)$$

where the sum is over integers n', for which $2 \leq n' \leq \frac{1}{2}n$,

$$k_{n'} = \tfrac{1}{2}n' K_{2n',0,(2n'-1)},$$

and we may take $\beta_{n0} = 0$. To any desired order n in the amplitude of the motion, we have now succeeded in reducing the Hamiltonian to the form (6) by means of a succession of canonical transformations. The final nth-order variables $J_{(n)}$, $\gamma_{(n)}$ are related to the original variables by a transformation of order n in the amplitude. The final action variable $J_{(n)}$ is a constant of the motion to this order.

The solution of Eq. (36) is

$$S_{nl} = -\tfrac{1}{2}e^{-il\phi}\int K_{nl(n-1)}e^{il\phi + i\beta_{nl(n-1)}}\,dt, \quad (38)$$

where $\phi = \int \Omega\, dt$.

If $Ke^{i\beta}$ is slowly varying, then Eq. (36) can be solved in the adiabatic approximation by successive approximations as

$$S_{nl}^{(0)} = -K_{nl(n-1)}e^{i\beta_{nl(n-1)}}/2il\Omega,$$
$$S_{nl}^{(m+1)} = S_{nl}^{(0)} - \dot{S}_{nl}^{(m)}/il\Omega. \quad (39)$$

If $S_{nl}^{(m)}$ is substituted in the transformation equations (32) and (33), the resulting action variable $J_{(n)}^{(m)}$, when expressed in terms of the original variables, gives a formula for the adiabatic invariant constant to order n in the amplitudes of the motion and to order m in the time derivatives of the coefficients in the Hamiltonian. Note that two sequences of successive approximations are involved in calculating the adiabatic invariant, in powers of the amplitude, and in successive orders in the time derivatives of the coefficients. If we use the exact solution (38) for S, we can treat the time dependence exactly, to obtain an invariant exact to nth order in the amplitude. By properly choosing the constant of integration in Eq. (38), we can make the exact solution correspond initially to the adiabatic solution (39). Any exact or nonadiabatic approximation to the integral in Eq. (38) (e.g., a saddle-point approximation) then allows us to calculate or estimate the nonadiabatic changes in the value of the adiabatic invariant.

Although the procedure we have outlined is straightforward in principle, in practice it is difficult to carry the procedure to an order n larger than about four without a prohibitive amount of labor, unless lower-order terms in the Hamiltonian happen to vanish. One can, however, use the method to estimate the effect of the time variation of any particular higher-order term in the Hamiltonian upon either the form of the adiabatic invariant or the nonadiabatic change in the adiabatic invariant. The successive orders of adiabatic approximation in the procedure indicated in Eq. (39) involve considerably less algebraic labor and can therefore be carried several orders beyond the lowest before the formulas become too unwieldy.

A difficulty may arise in carrying the above procedure to higher-order terms when the exact solution (38) is used in an earlier transformation. The exact solution (38) of Eq. (36) differs in general from the adiabatic solution (39) by terms which oscillate with frequency $l\Omega$. These transient terms can be present even during a period of time when K and β are constant or very slowly varying; they record the effects of past nonadiabatic behavior of the coefficients K and β. During a period when the Hamiltonian is constant, the phase points move on curves of constant Hamiltonian. However, if nonadiabatic changes have occurred previously, then a group of phase points which initially lay on a curve of constant Hamiltonian no longer lie on such a curve, and consequently the curve on which they lie moves periodically in the phase space. The role of the terms of period $l\Omega$ in S is to describe this behavior. If, now, the exact solution of Eq. (36) is used in the transformation (34), the new Hamiltonian contains higher-order terms which oscillate with frequencies $l\Omega$ and higher harmonics. Such terms may lead to resonances in the solution of Eq. (36) for higher-order terms. These resonances lead to secular increases in the amplitude of the higher-order transformation coefficients, even during periods when the original Hamiltonian was constant. This difficulty is

presumably associated with the fact that, for a nonlinear oscillator, the phase points traverse the phase trajectories with frequencies which depend upon the amplitude. Hence, even a simple closed curve which does not coincide with a phase trajectory becomes progressively more complicated as its various parts traverse phase trajectories with various frequencies. This difficulty is not serious when the above method is used for the purpose of obtaining practical estimates of the nonadiabatic changes of the adiabatic invariants. If the calculation is carried only to the first order in which such changes occur, the difficulty does not arise. Even in higher orders, the secular increase is very slow, so long as the nonadiabatic changes have been small; they therefore do not interfere with estimates of small nonadiabatic changes produced by variations in the Hamiltonian during a finite period of time.

It should be noted that nonlinearities may not only modify the oscillator frequency and the shape of the phase trajectories, but that they may also change the topological character of those trajectories at large amplitudes. There are in general separatrices in the phase space which separate the regions in which the motion has a different topological character. Our treatment here is applicable only to small-amplitude motions. Successively higher-order terms in the amplitude do not become successively smaller when the amplitude reaches the neighborhood of the smallest separatrix. Since the frequency of oscillation vanishes on a separatrix, formula (7) can be used to estimate the amplitude at which the separatrix occurs, although this is often obvious from inspection of the function H. If the motion beyond the separatrix is also periodic, it can often be treated approximately by methods similar to those outlined above. When a phase point crosses a separatrix due to the time variation of H, the adiabatic approximation is not valid. However, if the motion is periodic on both sides of the separatrix, and if the adiabatic approximation is valid everywhere except near the separatrix, then the transition across the separatrix is quasiadiabatic; that is, we can identify the value of the new adiabatic invariant corresponding to any given value of the original invariant for all except a very small set of phase points. We do not discuss this question further here.[6]

IV. APPLICATIONS

A. The Linear Oscillator

To avoid inessential algebraic complications, we consider in this example a linear oscillator described by a Hamiltonian of the form (3). Our discussion can be extended in a straightforward way to the more general case in which the Hamiltonian is given in the form (2). This problem has been studied by many authors.[1,7-11] We therefore restrict our attention to results which are new or which illustrate particular features of the present method of attacking the problem. We consider the case when the frequency ω has initially the constant value ω_0, then changes in some fashion for a period of time, and finally becomes again constant equal to a final value ω_1.

The invariant J is given by Eq. (22) which in the present case reduces to

$$J = \tfrac{1}{2}w^{-2}x^2 + \tfrac{1}{2}(wp - \dot{w}x)^2, \tag{40}$$

where w is a solution of Eq. (16):

$$\ddot{w} + \omega^2 w = w^{-3}. \tag{41}$$

The adiabatic approximations to the solution of this equation are, to second order in the time derivatives,

$$\begin{aligned} w^{(0)} &= w^{(1)} = \omega^{-\frac{1}{2}}, \\ w^{(2)} &= \omega^{-\frac{1}{2}} + \tfrac{1}{8}\ddot{\omega}\omega^{-\frac{7}{2}} - \tfrac{3}{16}\dot{\omega}^2\omega^{-\frac{9}{2}}. \end{aligned} \tag{42}$$

If these approximations are substituted in Eq. (40), we obtain expressions for the adiabatic invariant, which are

$$\begin{aligned} J^{(0)} &= \tfrac{1}{2}\omega x^2 + \tfrac{1}{2}\omega^{-1}p^2, \\ J^{(1)} &= \tfrac{1}{2}\omega x^2 + \tfrac{1}{2}\omega^{-1}(p + \tfrac{1}{2}\dot{\omega}\omega^{-1}x)^2, \end{aligned} \tag{43}$$

to zeroth and first orders, respectively. During periods when ω is constant, all approximations to the adiabatic invariant reduce to $J^{(0)}$. Thus, initially the adiabatic invariant to any order is

$$J_0 = \tfrac{1}{2}\omega_0 x^2 + \tfrac{1}{2}\omega_0^{-1}p^2. \tag{44}$$

Since we choose the exact solution of Eq. (41) which begins with the initial value $w = \omega_0^{-\frac{1}{2}}$, Eq. (44) also represents the exact invariant during the initial period.

Finally, when $\omega = \omega_1$, the adiabatic invariant is given to any order by

$$J = \tfrac{1}{2}\omega_1 x^2 + \tfrac{1}{2}\omega_1^{-1}p^2. \tag{45}$$

During the final period, the exact solution of Eq. (41) has the form (25):

$$w^2 = (\omega_1^{-2} + C^2)^{\frac{1}{2}} + C\cos(2\omega_1 t + \theta), \tag{46}$$

where the constants C and θ depend upon the history of $\omega(t)$ during the time in which it was changing. We substitute in Eq. (40) to obtain the exact invariant, which of course still has its initial value J_0. Consider now a collection of oscillators whose phase points initially move on an ellipse (44) for some fixed value of J_0. Finally, these oscillators have phase points moving on an ellipse (40) for $J = J_0$. We see that the final ellipse changes its orientation and shape with time,

according to Eq. (46), twice per period. Each individual phase point moves on a fixed ellipse (45), intersected by the moving ellipse (40). In order to find the distribution of final values of the adiabatic invariant J, let us choose a particular instant t_1 for which $2\omega_1 t_1 + \theta = 2n\pi$, so that $\dot{w}(t_1) = 0$. The exact invariant (40) at t_1 then has its principal axes along x and p,

$$J_0 = \tfrac{1}{2}\omega_1 Q^{-2}x^2 + \tfrac{1}{2}\omega_1^{-1}Q^2p^2, \tag{47}$$

where

$$Q^2 = (1 + \omega_1^2C^2)^{\frac{1}{2}} + \omega_1 C. \tag{48}$$

Each phase point (x, p) on the ellipse (47) lies also on an ellipse (45). We wish to determine the distribution of the values of J. In order to determine the statistical distribution of phase points around the ellipse (47), let us make the following transformation of variables:

$$X = \omega_1^{\frac{1}{2}}Q^{-1}x, \quad P = \omega_1^{-\frac{1}{2}}Qp. \tag{49}$$

The invariant then defines a circle in the new phase plane,

$$J_0 = \tfrac{1}{2}(X^2 + P^2). \tag{50}$$

It is readily verified that the transformation (49) is canonical, so that the area of the circle (50) equals the area of the ellipse (47). Let us now imagine a set of phase points uniformly distributed in the annular area between two initial ellipses (J_0 and $J_0 + dJ_0$) given by Eq. (44). These phase points are then uniformly distributed in the annulus between two circles (50) corresponding to the same values of J_0. It is now evident that all phases around the circle (50) are equally likely, so that if we let

$$X = (2J_0)^{\frac{1}{2}} \sin\Phi, \quad P = (2J_0)^{\frac{1}{2}} \cos\Phi, \tag{51}$$

then the phase variable Φ is uniformly distributed.

We make the substitutions (49) and (51) in Eq. (45) to obtain the final values of the adiabatic invariant:

$$\begin{aligned} J &= J_0(Q^2 \sin^2\Phi + Q^{-2}\cos^2\Phi) \\ &= J_0[(1 + \omega_1^2C^2)^{\frac{1}{2}} - \omega_1 C \cos 2\Phi]. \end{aligned} \tag{52}$$

Thus the distribution in final values of J is entirely determined by the constant $\omega_1 C$, which is to be obtained by integrating Eq. (41). In particular, the mean and variance of J are given by

$$\langle J\rangle = J_0(1 + \omega_1^2C^2)^{\frac{1}{2}}, \tag{53}$$

$$\langle(\Delta J)^2\rangle = \langle J^2\rangle - \langle J\rangle^2 = \tfrac{1}{2}J_0^2\omega_1^2C^2. \tag{54}$$

Note that these results are exact, independent of the nature of the variation of $\omega(t)$, provided only that it is constant initially and finally. The maximum and minimum of J can also be read from Eq. (52):

$$J_{\min}^{\max} = J_0[(1 + \omega_1^2C^2)^{\frac{1}{2}} \pm \omega_1 C]. \tag{55}$$

If either $J_{\max}$ or $J_{\min}$ is known, the constant $\omega_1 C$ may be determined, and the entire distribution in J is then determined.

We show in the Appendix how to obtain an approximate solution to Eq. (41), when $\omega_1 C \ll 1$. The constant C is given in this approximation, according to Eq. (A23) by

$$\omega_1 Ce^{i\theta} = \int_{-\infty}^{\infty} \dot{\omega}\omega^{-1}e^{2i\phi}\,dt, \tag{56}$$

where

$$\phi = \int^t \omega(t)\,dt. \tag{57}$$

[Actually, in order for θ in Eq. (56) to correspond to θ in Eq. (46), $\omega_1 t$ should be replaced by ϕ.]

The change in the adiabatic invariant J has been calculated for a particular function $\omega(t)$ by Howard[11] using the saddle-point method to evaluate an integral given by Vandervoort.[10] Vandervoort obtains the approximate result

$$J = J_0 \exp\,[-\omega_1 C \cos\theta + \tfrac{1}{2}(\omega_1 C \cos\theta)^2], \tag{58}$$

where $\omega_1 Ce^{i\theta}$ is given by formula (56). Note that Eq. (58) agrees to second order with our exact formula (52) if we identify $\theta = 2\Phi$. Formula (58), however, contains the phase θ, which, in turn, depends upon the phase ϕ_0 of the oscillator at a particular time $t = 0$ during the period when ω is changing. Since the probability distribution of the phase ϕ_0 is not obvious, the probability distribution of J is not easy to determine, although we can read off its maximum and minimum excursions. Howard finds agreement with the maximum and minimum values of $J - J_0$ from numerical solutions of Eq. (3) within about 10%. Our formulas above give the complete distribution in J in terms of the amplitude C.

B. A Nonlinear Case

As a simple example of the application of the methods of Sec. III, we consider a nonlinear oscillator whose equation of motion is

$$\ddot{x} + \omega^2 x + a(t)x^2 = 0. \tag{59}$$

The corresponding Hamiltonian is

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2x^2 + \tfrac{1}{3}a(t)x^3. \tag{60}$$

We assume, to simplify the problem, that ω is constant, and we seek the effect of the time variation of the coefficient a. We further assume, to be specific, that the coefficient $a(t)$ is zero both initially and finally, so

that we have a simple harmonic oscillator subject for a period of time to a nonlinear perturbing term.

The transformation (18) is now trivial, and we have, after making the transformation (27),

$$X = \omega^{\frac{1}{2}}x = (2J_{(2)})^{\frac{1}{2}} \sin \gamma_{(2)},$$
$$P = \omega^{-\frac{1}{2}}p = (2J_{(2)})^{\frac{1}{2}} \cos \gamma_{(2)}, \tag{61}$$
$$K = \omega J_{(2)} + \tfrac{1}{4}a\omega^{-\frac{3}{2}}(2J_{(2)})^{\frac{3}{2}}(\sin \gamma_{(2)} - \tfrac{1}{3} \sin 3\gamma_{(2)}). \tag{62}$$

We now make the third-order transformation (31), which in our case is

$$S = J_{(3)}\gamma_{(2)} + 2S'_{31}(2J_{(3)})^{\frac{3}{2}} \cos (\gamma_{(2)} + \sigma_{31}) + 2S'_{33}(2J_{(3)})^{\frac{3}{2}} \cos (3\gamma_{(2)} + \sigma_{33}), \tag{63}$$

where

$$S_{31} = S'_{31}e^{i\sigma_{31}}, \quad S_{33} = S'_{33}e^{i\sigma_{33}}, \tag{64}$$

and the coefficients are solutions of the equations

$$\dot{S}_{31} + i\omega S_{31} = \tfrac{1}{8}ia\omega^{-\frac{3}{2}},$$
$$\dot{S}_{33} + 3i\omega S_{33} = -\tfrac{1}{24}ia\omega^{-\frac{3}{2}}. \tag{65}$$

The transformation

$$J_{(2)} = \frac{\partial S}{\partial \gamma_{(2)}}$$

leads to the following expression for the invariant, to third order in the amplitude:

$$J_{(3)} = J_{(2)} + 2S'_{31}(2J_{(2)})^{\frac{3}{2}} \sin (\gamma_{(2)} + \sigma_{31}) + 6S'_{33}(2J_{(2)})^{\frac{3}{2}} \sin (3\gamma_{(2)} + \sigma_{33}). \tag{66}$$

The variables $J_{(2)}$ and $\gamma_{(2)}$ can be replaced by x and p in a straightforward way by using Eq. (61).

The adiabatic solutions to Eqs. (65) are easily written down to any desired order in the time derivatives according to the scheme (39). The first two orders are

$$S_{31}^{(0)} = \tfrac{1}{8}a\omega^{-\frac{5}{2}}, \quad S_{33}^{(0)} = -\tfrac{1}{72}a\omega^{-\frac{5}{2}}, \tag{67}$$
$$S_{31}^{(1)} = \tfrac{1}{8}a\omega^{-\frac{5}{2}} + \tfrac{1}{8}i\dot{a}\omega^{-\frac{7}{2}},$$
$$S_{33}^{(1)} = -\tfrac{1}{72}a\omega^{-\frac{5}{2}} - \tfrac{1}{216}i\dot{a}\omega^{-\frac{7}{2}}. \tag{68}$$

If Eq. (67) is substituted in Eq. (66), we obtain a formula for the adiabatic invariant correct to third order in the amplitude and to lowest order in the time derivatives:

$$J_{(3)}^{(0)} = \tfrac{1}{2}\omega x^2 + \tfrac{1}{2}\omega^{-1}p^2 + \tfrac{1}{6}a\omega^{-1}x^3. \tag{69}$$

Initially and finally, the coefficient a is zero and the adiabatic invariant to any order is just the invariant for the linear oscillator:

$$J = \tfrac{1}{2}\omega x^2 + \tfrac{1}{2}\omega^{-1}p^2. \tag{70}$$

The initial and final values of J, however, are generally different because of the nonlinear term which appeared during the intervening time. Initially, we take $S_{31} = S_{33} = 0$ as the solutions of Eqs. (65) so that the exact invariant is also given by Eq. (70). Let us assume that the initial value is $J = J_0$. The final solution for S is to be obtained by integrating Eqs. (65) through the intervening time. During the final time when $a = 0$, the solutions of Eqs. (65) are given by

$$S_{31} = A_1e^{-i(\omega t-\theta_1)}, \quad S_{33} = A_3e^{-3i(\omega t-\theta_3)}, \tag{71}$$

where the constants are given by Eq. (37),

$$A_1e^{i\theta_1} = \tfrac{1}{8}i\omega^{-\frac{3}{2}}\int_{-\infty}^{\infty} a(t)e^{i\omega t}\,dt,$$
$$A_3e^{3i\theta_3} = -\tfrac{1}{24}i\omega^{-\frac{3}{2}}\int_{-\infty}^{\infty} a(t)e^{3i\omega t}\,dt. \tag{72}$$

To be specific, let us take the coefficient $a(t)$ to be of the form

$$a(t) = a_0\tau^2(t^2 + \tau^2)^{-1}. \tag{73}$$

The integrals in Eq. (72) are then readily evaluated and we find

$$A_1 = \tfrac{1}{8}\pi a_0\tau\omega^{-\frac{3}{2}}e^{-\omega\tau}, \quad A_3 = \tfrac{1}{24}\pi a_0\tau\omega^{-\frac{3}{2}}e^{-3\omega\tau},$$
$$\theta_1 = \theta_3 = \tfrac{1}{2}\pi. \tag{74}$$

We substitute these results into Eq. (66) to obtain, for the exact invariant, to third order

$$J_0 = (\tfrac{1}{2}\omega x^2 + \tfrac{1}{2}\omega^{-1}p^2)[1 + (2A_1 + 18A_3) \times (\omega^{\frac{1}{2}}x \sin \omega t + \omega^{-\frac{1}{2}}p \cos \omega t)] - 24A_3(\omega^{\frac{1}{2}}x \sin \omega t + \omega^{-\frac{1}{2}}p \cos \omega t)^3. \tag{75}$$

We see that the final exact invariant (75) is a cubic polynomial, in spite of the fact that the oscillator finally becomes again a simple harmonic oscillator. The phase points corresponding to a fixed value of J_0 lie on the cubic curve (75), which changes periodically with time as each phase point on that curve traverses an ellipse (70). If we choose a particular time at which $\sin \omega t = 1$, the invariant curve takes the simpler form

$$J_0 = \tfrac{1}{2}\omega x^2 + \tfrac{1}{2}\omega^{-1}p^2 + \tfrac{1}{8}\pi a_0\tau(e^{-\omega\tau} - 5e^{-3\omega\tau})x^3. \tag{76}$$

The probability distribution of the final adiabatic invariant J is most easily obtained directly from Eq. (66). First, we recognize that $J_{(2)}$ is the adiabatic invariant J, and to third order $J_{(3)} = J_0$. We also note that, since to third order the γ dependence has been transformed away in the Hamiltonian, the phase points move on circles of constant $J_{(3)}$. In Eq. (66), again to third order, we may let $\gamma_{(2)} = \gamma_{(3)}$. By the same argument we gave in the last example, the variable $\gamma_{(3)}$ is uniformly distributed in probability. The

probability distribution of J is therefore determined by making these replacements in Eq. (66), so that

$$J = J_0 - 2A_1(2J_0)^{\frac{3}{2}} \sin \Phi - 6A_3(2J_0)^{\frac{3}{2}} \sin 3\Phi, \quad (77)$$

where $\Phi = \gamma_{(3)} - \omega t - \frac{1}{2}\pi$ is uniformly distributed at any instant t. We may readily read off the mean and variance of J:

$$\langle J \rangle = J_0,$$
$$\langle (J - J_0)^2 \rangle = (16A_1^2 + 48A_3^2)J_0^3$$
$$= \tfrac{1}{4}\pi^2 a_0^2 \tau^2 \omega^{-3}(e^{-2\omega\tau} + \tfrac{1}{3}e^{-6\omega\tau})J_0^3. \quad (78)$$

It is evident by inspection of Eq. (60) that there is a separatrix with a singular point at $p = 0$ and $x = -a/\omega^2$, beyond which the motion in this example is no longer oscillatory. The transformation (63) changes its character at amplitudes near and beyond the separatrix, where the third-order terms dominate the lowest-order term.

We note finally that the transformation (63) which eliminates the cubic terms from the Hamiltonian (62) leaves a higher-order residue

$$\Delta K_{(3)} = a\omega^{-\frac{3}{2}}[(2J_{(2)})^{\frac{3}{2}} - (2J_{(3)})^{\frac{3}{2}}] \times (\tfrac{1}{4} \sin \gamma_{(2)} - \tfrac{1}{12} \sin 3\gamma_{(2)}). \quad (79)$$

The fourth-order terms may be readily calculated from the above expression. The γ-dependent fourth-order terms could then be transformed away by the method given in Sec. III and an expression for the invariant to fourth order in the amplitude could be obtained. We content ourselves here with noting that the remaining γ-independent fourth-order term in the Hamiltonian is

$$K_{40}(2J_{(4)})^2 = (\tfrac{3}{2})a\omega^{-\frac{3}{2}}(2J_{(4)})^2 \times (S'_{33} \cos \sigma_{33} - S'_{31} \cos \sigma_{31}). \quad (80)$$

The Hamiltonian in fourth order therefore reduces to the form (6). If we substitute the lowest-order adiabatic expressions (67), the coefficient k_2 takes the simple form

$$k_2 = -\tfrac{5}{6}a^2\omega^{-4}, \quad (81)$$

which, according to Eq. (7), gives the amplitude dependence of the frequency. This result is not new, at least when a is constant, and can be obtained by almost any method of solving the nonlinear equation (59). Our treatment shows that the result gives correctly, to second order in the amplitude and to lowest order in the time variation, the time rate of change of the phase as given by Eq. (7), even when the coefficient a is changing with time. Higher-order accuracy in the time variation may of course be obtained by substituting more accurate expressions for S in Eq. (80).

ACKNOWLEDGMENTS

The author wishes to thank the Los Alamos Scientific Laboratory for support during the time in which the present work was conceived. I also wish to thank H. R. Lewis, whose discovery of the exact invariant (40) called my attention to the possibility of using the foregoing methods in treating the adiabatic problem. I am indebted to J. E. Howard and K. W. Li for checking the foregoing calculations. I am also indebted to Howard for the private communication of his results which stimulated the application of the above methods to the calculations in Sec. 4A.

APPENDIX

We wish to solve Eq. (16) in the general case when the coefficients a, b, and c are functions of time by using a method of variation of the constants A and B in the solution (25). To simplify the form of Eq. (16), we let

$$W = a^{-\frac{1}{2}}w, \quad (A1)$$

$$\Omega^2 = ac - b^2 - \dot{b} + \dot{a}a^{-1}b + \tfrac{1}{2}\ddot{a}a^{-1} - \tfrac{3}{4}\dot{a}^2a^{-2}. \quad (A2)$$

Equation (A2) gives a convenient choice for the frequency Ω. The equation to be solved is then

$$\ddot{W} + \Omega^2 W = W^{-3}. \quad (A3)$$

We try a solution in the form[12]

$$\Omega W^2 = Q + A \cos 2\phi + B \sin 2\phi, \quad (A4)$$

where

$$\phi = \int \Omega \, dt, \quad (A5)$$

$$Q = (1 + A^2 + B^2)^{\frac{1}{2}}. \quad (A6)$$

If A, B, and Ω are constant, then Eq. (A4) is a solution to Eq. (A3). We differentiate Eq. (A4) to obtain, after omitting a factor 2Ω,

$$W\dot{W} = -A \sin 2\phi + B \cos 2\phi + \tfrac{1}{2}\Omega^{-1}(\dot{A} \cos 2\phi + \dot{B} \sin 2\phi + \dot{Q} - \dot{\Omega}W^2). \quad (A7)$$

We require that the quantity in parenthesis vanish. Differentiating again, we obtain

$$\dot{W}^2 + W\ddot{W} = -2\Omega A \cos 2\phi - 2\Omega B \sin 2\phi - (\dot{A} \sin 2\phi - \dot{B} \cos 2\phi). \quad (A8)$$

We again require that the quantity in parenthesis vanish. It is then a matter of straightforward algebra to verify that, when Eqs. (A7) and (A8) are substituted in Eq. (A3), the latter equation is satisfied. We have then two first-order differential equations to solve for

the quantities A and B, which we write in the form

$$\dot{A} = (\dot{\Omega}W^2 - \dot{Q})\cos 2\phi,$$
$$\dot{B} = (\dot{\Omega}W^2 - \dot{Q})\sin 2\phi. \tag{A9}$$

We now introduce the complex variable

$$Z = A + iB. \tag{A10}$$

If we use as independent variable the phase ϕ defined by Eq. (A5), which is a monotonic function of the original independent variable t, Eqs. (A9) can be written in the simple form

$$Z' = Re^{2i\phi}, \tag{A11}$$

where a prime denotes differentiation with respect to ϕ, and

$$R = -Q' + \Omega'\Omega^{-1}(Q + A\cos 2\phi + B\sin 2\phi). \tag{A12}$$

The coefficient R depends on Z as well as on ϕ, but it depends on Z only through the combination ZZ^* and its derivative, since we may rewrite Eq. (A6) in the form

$$Q = (1 + ZZ^*)^{\frac{1}{2}}, \tag{A13}$$

and Eqs. (A11) and (A12) then imply that R is a solution of the equation

$$R = \Omega'\Omega^{-1}Q - Q' + \tfrac{1}{2}\Omega'\Omega^{-1}R^{-1}(ZZ^*)'. \tag{A14}$$

We seek an approximate nonadiabatic solution of Eq. (A11), valid at least in the case when Ω is a slowly varying function of t (and therefore of ϕ). If we write $\Omega(\epsilon\phi)$ for the functional dependence of Ω on ϕ, then we expect the over-all change in the coefficients A and B during a long time, which represent the nonadiabatic changes in the solution of Eq. (A13), to go to zero faster than any power of ϵ as ϵ approaches zero. During times when Ω is changing, we see from Eqs. (A11) and (A12) that the coefficients A and B contain transient parts of order ϵ which oscillate with frequency 2Ω. Therefore, it is important, in deriving an expression for the long-time nonadiabatic changes, to make sure that terms which are neglected do not contribute more to the final result than terms which are retained.[13]

We first derive an adiabatic solution of Eq. (A11). We integrate both sides of Eq. (A11) with respect to ϕ and carry out $(n + 2)$ integrations by parts, where n is an even number, to obtain

$$Z = \int Re^{2i\phi}\,d\phi = \left[\frac{K}{2i} - \frac{K'}{(2i)^2}\right]e^{2i\phi} + \frac{1}{(2i)^{n+2}}\int e^{2i\phi}\frac{d^{n+2}R}{d\phi^{n+2}}\,d\phi, \tag{A15}$$

where

$$K = R + \frac{R''}{(2i)^2} + \cdots + \frac{1}{(2i)^n}\frac{d^nR}{d\phi^n}. \tag{A16}$$

We note that since R is real and only even powers of $2i$ appear in Eq. (A16), the quantity K is therefore also real. Since the quantity R turns out to be slowly varying, each of the indicated derivatives in Eqs. (A15) and (A16) represents an additional power in the quantity ϵ, the smallness parameter for the time variation of Ω. The first term in Eq. (A15) is the transient adiabatic solution to order $(n + 2)$ in ϵ; as we see, it oscillates with frequency 2Ω. The last term contains higher-order parts of the transient solution together with any long term secular changes in the coefficients A and B. We are interested in a solution which corresponds initially to the adiabatic solution of Eq. (A4). If Ω is initially constant, then $Z = 0$ initially. If Ω is not initially constant, then Z should be taken initially to be the adiabatic solution, which to any order $(n + 2)$ is given by neglecting the last term in Eq. (A15). The adiabatic solution at any later time to order $n + 2$ is then again given by the first term in Eq. (A15). We have, therefore, for the adiabatic solution to order $(n + 2)$,

$$Z = (\tfrac{1}{2}K\sin 2\phi + \tfrac{1}{4}K'\cos 2\phi) + i(-\tfrac{1}{2}K\cos 2\phi + \tfrac{1}{4}K'\sin 2\phi). \tag{A17}$$

If we substitute this result in Eqs. (A13) and (A14), we obtain the approximate values

$$Q = (1 + \tfrac{1}{4}K^2 + \tfrac{1}{16}K'^2)^{\frac{1}{2}}, \tag{A18}$$

$$R = \Omega'\Omega^{-1}Q - Q' + \tfrac{1}{4}\Omega'\Omega^{-1}K'. \tag{A19}$$

To first order in ϵ we have, therefore,

$$R^{(1)} = \Omega'/\Omega. \tag{A20}$$

By an iterative process involving repeated substitutions of R into Eqs. (A16), (A18), and (A19), we may obtain successive approximations $R^{(3)}$, $R^{(5)}$, etc., to any desired order in ϵ. The final result $R^{(n)}$ to any order n is then given as an explicit function of Ω and its derivatives. We may then write an approximate solution for Z in the form

$$Z = \int R^{(n)}e^{2i\phi}\,d\phi. \tag{A21}$$

The integral in Eq. (A21), evaluated between the initial and final times, gives the final values of the coefficients A and B, which represent the nonadiabatic effects due to changes in Ω during the intervening time. If the time variation of Ω is slow, the integral in Eq. (A21) can be evaluated to a good approximation

by the saddle-point method. If at the upper limit of integration $\Omega' = 0$, the result is of the form

$$Z_{\text{final}} = T^{(n)} e^{-k/\epsilon}, \tag{A22}$$

where the coefficient $T^{(n)}$ is a polynomial of order n in ϵ which depends upon the coefficient $R^{(n)}$ and upon the functional form of Ω, and where the coefficient k in the exponent depends only on the form of the function $\Omega(\epsilon\phi)$, and not on the coefficient $R^{(n)}$. We see, therefore, that higher-order terms in R do not give rise to terms in the final value of Z which are larger than those contributed by lower-order terms. For most purposes an entirely adequate approximation for the nonadiabatic effects is obtained by using only the first-order expression (A20) for R. If we do so and return to the independent variable t, we have the final result

$$Z = \int \dot{\Omega}\Omega^{-1} e^{2i\phi}\, dt. \tag{A23}$$

* This work supported by the U.S. Atomic Energy Commission.

[1] H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968); Phys. Rev. Letters **18**, 510 (1967); also an erratum: Phys. Rev. Letters **18**, 636 (1967).

[2] E. D. Courant and H. S. Snyder, Ann. Phys. (N.Y.) **3**, 1 (1958).

[3] G. D. Birkhoff, *Dynamical Systems* (Am. Math. Soc. Colloquium Publ., New York, 1927), Chap. 3.

[4] J. Moser, Nachr. Akad. Wiss. Goettingen, Math. Physik. Kl. IIa **6**, 87 (1955).

[5] In order to avoid a profusion of π's we have chosen to define the action variable J so that it is the phase area of the corresponding ellipse divided by 2π. Our J, therefore, differs by a factor of 2π from the more customary definition. Likewise, the angle variable γ is a true angle, increasing by 2π in one revolution.

[6] A case of this type is discussed by R. W. B. Best, Physica **40**, 182 (1968).

[7] R. M. Kulsrud, Phys. Rev. **106**, 205 (1957).

[8] F. Hertweck and A. Schlüter, Z. Naturforsch. **12a**, 844 (1957).

[9] G. Backus, A. Lenard, and R. Kulsrud, Z. Naturforsch. **15a**, 1007 (1960).

[10] P. O. Vandervoort, Ann. Phys. (N.Y.) **13**, 436 (1961).

[11] J. E. Howard, "The Non-Adiabatic Harmonic Oscillator," submitted to Phys. Fluids.

[12] The quantities A and B in Eq. (A4) differ by a factor Ω/a from those in Eq. (25).

[13] I am indebted to J. B. Taylor for pointing out this flaw in previous derivations of ΔJ, including my own first draft of this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

High Symmetry Fields and the Homogeneous Field in General Relativity*

P. C. AICHELBURG
International Atomic Energy Agency, International Centre for Theoretical Physics
Trieste, Italy

(Received 22 July 1969)

The usual definition of a homogeneous field in general relativity implies a space with $R_{iklm} = 0$, thus admitting a group of motions isomorphic to the Poincaré group. After discussing the symmetry group of the homogeneous field in Newtonian space, we point out that there exists no space with $R_{ik} = 0$, which is a "true" field, i.e., $R_{iklm} \neq 0$, and which admits an analogous relativistic group. We then study fields, solutions of $R_{ik} = 0$, which define spaces that admit a 4-parameter group of motions locally isomorphic to the groups $T_1 \otimes [T_2 \otimes_s O(2)]$ and $T_1 \otimes [T_2 \otimes_s O(1, 1)]$. We compare the motion of a test particle in these fields with the motion in the usual homogeneous field.

I. INTRODUCTION

If, by definition, a homogeneous gravitational field in general relativity (GR) is a field that can be transformed away not only locally but over all space, it defines a space with $R_{iklm} = 0$ and is only an apparent field. [In this paper, Latin indices take the values 0, 1, 2, 3 and Greek indices, 0, 1, 2, or as specified. A comma means ordinary differentiation and a semicolon, covariant differentiation. Coordinates are also written as $x^0 = t$, $x^1 = x$, $x^2 = y$, and $x^3 = z$. The signature of g_{ik} is $(+ - - -)$, G_r stands for an r-parameter group and T_r for an Abelian group, and $\otimes_s$ means semidirect product.]

So far, interest in a so-called homogeneous field in GR has been mainly in connection with the twin paradox and the equivalence principle; see, e.g., Refs. 1 and 2. These fields can always be obtained by a suitable coordinate transformation from a pseudo-Euclidean metric. Besides the condition $R_{iklm} = 0$, they have to fulfill supplementary conditions in order to be static, have the right nonrelativistic limit, etc. Since for such fields space is flat, they admit a ten-parameter group of motions isomorphic to the Poincaré group. Just as in Newtonian theory, however, the symmetry of the field-free space is reduced in GR by the presence of "true" fields (with $R_{iklm} \neq 0$).

Thus, our aim is to find a field which is an exact solution of Einstein's empty space equation $R_{ik} = 0$, but with $R_{iklm} \neq 0$ and with a relativistic symmetry

by the saddle-point method. If at the upper limit of integration $\Omega' = 0$, the result is of the form

$$Z_{\text{final}} = T^{(n)} e^{-k/\epsilon}, \tag{A22}$$

where the coefficient $T^{(n)}$ is a polynomial of order n in ϵ which depends upon the coefficient $R^{(n)}$ and upon the functional form of Ω, and where the coefficient k in the exponent depends only on the form of the function $\Omega(\epsilon\phi)$, and not on the coefficient $R^{(n)}$. We see, therefore, that higher-order terms in R do not give rise to terms in the final value of Z which are larger than those contributed by lower-order terms. For most purposes an entirely adequate approximation for the nonadiabatic effects is obtained by using only the first-order expression (A20) for R. If we do so and return to the independent variable t, we have the final result

$$Z = \int \dot{\Omega}\Omega^{-1} e^{2i\phi}\, dt. \tag{A23}$$

* This work supported by the U.S. Atomic Energy Commission.

[1] H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968); Phys. Rev. Letters **18**, 510 (1967); also an erratum: Phys. Rev. Letters **18**, 636 (1967).

[2] E. D. Courant and H. S. Snyder, Ann. Phys. (N.Y.) **3**, 1 (1958).

[3] G. D. Birkhoff, *Dynamical Systems* (Am. Math. Soc. Colloquium Publ., New York, 1927), Chap. 3.

[4] J. Moser, Nachr. Akad. Wiss. Goettingen, Math. Physik. Kl. IIa **6**, 87 (1955).

[5] In order to avoid a profusion of π's we have chosen to define the action variable J so that it is the phase area of the corresponding ellipse divided by 2π. Our J, therefore, differs by a factor of 2π from the more customary definition. Likewise, the angle variable γ is a true angle, increasing by 2π in one revolution.

[6] A case of this type is discussed by R. W. B. Best, Physica **40**, 182 (1968).

[7] R. M. Kulsrud, Phys. Rev. **106**, 205 (1957).

[8] F. Hertweck and A. Schlüter, Z. Naturforsch. **12a**, 844 (1957).

[9] G. Backus, A. Lenard, and R. Kulsrud, Z. Naturforsch. **15a**, 1007 (1960).

[10] P. O. Vandervoort, Ann. Phys. (N.Y.) **13**, 436 (1961).

[11] J. E. Howard, "The Non-Adiabatic Harmonic Oscillator," submitted to Phys. Fluids.

[12] The quantities A and B in Eq. (A4) differ by a factor Ω/a from those in Eq. (25).

[13] I am indebted to J. B. Taylor for pointing out this flaw in previous derivations of ΔJ, including my own first draft of this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

High Symmetry Fields and the Homogeneous Field in General Relativity*

P. C. AICHELBURG
International Atomic Energy Agency, International Centre for Theoretical Physics
Trieste, Italy

(Received 22 July 1969)

The usual definition of a homogeneous field in general relativity implies a space with $R_{iklm} = 0$, thus admitting a group of motions isomorphic to the Poincaré group. After discussing the symmetry group of the homogeneous field in Newtonian space, we point out that there exists no space with $R_{ik} = 0$, which is a "true" field, i.e., $R_{iklm} \neq 0$, and which admits an analogous relativistic group. We then study fields, solutions of $R_{ik} = 0$, which define spaces that admit a 4-parameter group of motions locally isomorphic to the groups $T_1 \otimes [T_2 \otimes_s O(2)]$ and $T_1 \otimes [T_2 \otimes_s O(1, 1)]$. We compare the motion of a test particle in these fields with the motion in the usual homogeneous field.

I. INTRODUCTION

If, by definition, a homogeneous gravitational field in general relativity (GR) is a field that can be transformed away not only locally but over all space, it defines a space with $R_{iklm} = 0$ and is only an apparent field. [In this paper, Latin indices take the values 0, 1, 2, 3 and Greek indices, 0, 1, 2, or as specified. A comma means ordinary differentiation and a semicolon, covariant differentiation. Coordinates are also written as $x^0 = t$, $x^1 = x$, $x^2 = y$, and $x^3 = z$. The signature of g_{ik} is $(+ - - -)$, G_r stands for an r-parameter group and T_r for an Abelian group, and $\otimes_s$ means semidirect product.]

So far, interest in a so-called homogeneous field in GR has been mainly in connection with the twin paradox and the equivalence principle; see, e.g., Refs. 1 and 2. These fields can always be obtained by a suitable coordinate transformation from a pseudo-Euclidean metric. Besides the condition $R_{iklm} = 0$, they have to fulfill supplementary conditions in order to be static, have the right nonrelativistic limit, etc. Since for such fields space is flat, they admit a ten-parameter group of motions isomorphic to the Poincaré group. Just as in Newtonian theory, however, the symmetry of the field-free space is reduced in GR by the presence of "true" fields (with $R_{iklm} \neq 0$).

Thus, our aim is to find a field which is an exact solution of Einstein's empty space equation $R_{ik} = 0$, but with $R_{iklm} \neq 0$ and with a relativistic symmetry

analogous to Newtonian space with a homogeneous field.

In Sec. II, we summarize the symmetry of Newtonian space–time and recall that the ten-parameter Galilean group of the field free space is reduced to a 6-parameter subgroup when there is a homogeneous field present. Then we turn to Einstein's space–time. As with the Newtonian homogeneous field, we require that the space should admit a subgroup isomorphic to a Poincaré subgroup by the restriction that a certain spacelike coordinate is invariant, i.e., a group of motions locally isomorphic to the group $T_3 \otimes_s O(2, 1)$.[3] Fields defined by space admitting this group of motions would be analogous to the homogeneous field of Newtonian space–time. However, no space exists which is a true-field solution of the empty-space equation and which has the required 6-parameter symmetry group. Since the above symmetry group gives only flat-space solutions, we require from physical considerations that our space should admit a 4-parameter group acting transitively on a timelike hypersurface and having a 3-parameter Abelian subgroup. There are two nonisomorphic groups with the structures $T_1 \otimes [T_2 \otimes_s O(2)]$ and $T_1 \otimes [T_2 \otimes_s O(1, 1)]$.

In Sec. III, this leads us directly to 1-dimensional static gravitational fields. Corresponding to each group there is a solution of the field equations. The Riemann tensor for these solutions has an essentially singular timelike hypersurface.

In Sec. IV, we investigate the nonrelativistic limit of these fields and compare it to the field given by Møller.[1] It turns out that in this limit the gravitational potential is $\phi \sim pgz$, with the value of the constant p determined by the solutions. We thus have a homogeneous field giving a constant acceleration for a geodesically moving particle in the nonrelativistic limit.

In the last section, we discuss the trajectory of a freely moving particle. We show that the motion in a field which admits a symmetry group of the local structure $T_1 \otimes [T_2 \otimes_s O(1, 1)]$ is similar to that given in Ref. 1.

II. HIGH SYMMETRY FIELDS

Newtonian space–time can be described by a 4-dimensional differential manifold with a flat symmetric affine connection. The group of motions admitted by the manifold is obtained from the objects defining it and is the ten-parameter inhomogeneous Galilean group.[4,5]

In the classical theory the gravitational field is introduced as a new absolute object. Since with the presence of a field there are more objects defining the space, its symmetry is reduced by the further condition $\bar{\delta}\phi = 0$, where ϕ is the gravitational potential and $\bar{\delta}$ denotes the Lie derivative.

If we take for ϕ a homogeneous field which in a suitable coordinate system is defined as

$$\phi = a + bz,$$

with z a space coordinate, then the symmetry group is reduced to the 6-parameter subgroup of the Galilean group which leaves z invariant. This group contains the 3-parameter Abelian group of translations along x, y and t, two pure Galilean transformations leaving z unchanged, and the rotation about z.

From this we see that Newtonian space–time with a homogeneous field which everywhere satisfies the empty–space equation $\Delta\phi = 0$ admits a 6-parameter group of motions acting transitively on the hypersurfaces $z = \text{const}$.

In GR the space–time manifold is Riemannian, described by the affine connection Γ^i_{kl} and the metric tensor g_{ik}. To see the possible symmetries of the manifold, we must first look not only to empty space but also to a space with no true gravitational field present, i.e., with $R^i_{klm} = 0$. It then follows that Γ^i_{kl} is the flat affinity and the symmetry group is obtained from

$$\bar{\delta}\Gamma^i_{kl} = 0, \quad \bar{\delta}g_{ik} = 0, \tag{1}$$

where the second condition gives the Killing equation

$$\bar{\delta}g_{ik} = g_{il}\xi^l_{,k} + g_{lk}\xi^l_{,i} + g_{ik,l}\xi^l = 0. \tag{2}$$

It is well known that for the above case the conditions (2) lead to a ten-parameter group with the structure of the Poincaré group.

We now turn to $R_{iklm} \neq 0$ with $R_{ik} = 0$, i.e., empty but not field-free space. We have seen that the Galilean symmetry group is reduced to a 6-parameter subgroup if a homogeneous field is present. Similarly, we require that the empty space should allow a symmetry group isomorphic to a Poincaré subgroup. The subgroup analogous to the reduced Galilean group which leaves a spacelike coordinate invariant has a structure isomorphic to $T_3 \otimes_s O(2, 1)$, which in terms of the infinitesimal generators is given by

$$\begin{aligned}
&[X_\alpha, X_\beta] = 0, && \alpha, \beta = 1, 2, 3,\\
&[X_1, X_4] = 0, && [X_2, X_4] = X_3,\\
&[X_1, X_5] = X_2, && [X_2, X_5] = X_1,\\
&[X_1, X_6] = X_3, && [X_2, X_6] = 0,\\
&[X_3, X_4] = -X_2, && [X_4, X_5] = -X_6,\\
&[X_3, X_5] = 0, && [X_4, X_6] = X_5,\\
&[X_3, X_6] = -X_1, && [X_5, X_6] = X_4.
\end{aligned} \tag{3}$$

We have so far not chosen a coordinate system and have not given physical meaning to the generators. We only demand that the space should admit a group of motions with the structure given above and that the generators act on a timelike hypersurface.

Since the group is multiply transitive on 3-dimensional submanifold, it contains a nondiscrete isotropy group. From this it follows by a theorem of Ehlers and Kundt,[6] that the Riemann tensor of the manifold must be of degenerate Petrov type,[7] i.e., D or N.

N-type terms are pure radiation fields and can be excluded because the group is required to act on a timelike hypersurface. So only D-type is possible.

Because a G_4 is the maximal group for a D-type space, it cannot admit a G_6. We conclude that in GR there is no space with a symmetry group analogous to the group of Newtonian space with a homogeneous field.

We proceed to investigate D-type spaces which admit a G_4 and contain a T_3. This group has to be a subgroup of $T_3 \otimes_s O(2, 1)$. Although exact vacuum solutions of this type are the best known ones,[8] we start, for reasons to become clear in Sec. IV, with a space admitting a T_3. As in the Newtonian case, this group should act transitively on a timelike hypersurface. It follows then that these surfaces are geodesically parallel, and it is possible to introduce a coordinate system such that the equation for the hypersurface is $z = \text{const}$.[9] The metric takes the form

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta - dz^2, \quad \alpha, \beta = 0, 1, 2.$$

The rank of the matrix formed by the Killing vectors $({}_A\xi)$, where A denotes independent vectors, is 3, and we can choose the vectors as

$${}_1\xi^i = \delta^i_0, \quad {}_2\xi^i = \delta^i_1, \quad {}_3\xi^i = \delta^i_2.$$

From Eq. (2) we have for the g_{jk}

$$g_{\alpha\beta} = g_{\alpha\beta}(z), \quad \alpha, \beta = 0, 1, 2.$$

If we now require a 4-parameter group by adding one generator from (3) to the Abelian group, we get groups with the structures, $\alpha, \beta = 1, 2, 3$:

$$\begin{aligned} &[X_\alpha, X_\beta] = 0, \qquad [X_1, X_4] = 0, \\ &[X_2, X_4] = -X_3, \quad [X_3, X_4] = X_2, \end{aligned} \tag{4a}$$

$$\begin{aligned} &[X_\alpha, X_\beta] = 0, \qquad [X_1, X_4] = 0, \\ &[X_2, X_4] = X_3, \quad [X_3, X_4] = X_2. \end{aligned} \tag{4b}$$

Group (4a) is isomorphic to the group $T_1 \otimes [T_2 \otimes_s O(2)]$, while (4b) has the structure of $T_1 \otimes [T_2 \otimes_s O(1, 1)]$. For each group the fourth independent Killing vector satisfying the commutation relations (4a) and (4b), respectively, is given by

$${}_4\xi^i = \delta^i_1 x^2 - \delta^i_2 x^1, \quad {}_4\xi^i = \delta^i_0 x^1 + \delta^i_1 x^0.$$

Using the Killing Eq. (2) with these vectors, we reduce the metric to diagonal form: for (4a),

$$ds^2 = g_{00}(z)\, dt^2 + g_{11}(z)(dx^2 + dy^2) - dz^2 \tag{5a}$$

and, for (4b),

$$ds^2 = g_{00}(z)(dt^2 - dx^2) + g_{22}(z)\, dy^2 - dz^2, \tag{5b}$$

or x and y interchanged.

If we try to enhance the groups (4) to the 6-parameter group given in (3), we find, from the Killing equation, the metric

$$ds^2 = g_{00}(z)(dz^2 - dx^2 - dy^2) - dz^2,$$

which is only a solution of $R_{ik} = 0$ if $g_{00}(z) \sim \text{const}$ so that $R^i_{klm} = 0$. From the above, we see that requiring the space to admit a 3-parameter Abelian group leads to 1-dimensional fields, and the groups given by (4) reduce these to special types of static fields.

III. THE KASNER SOLUTION

A special solution admitting a 3-parameter Abelian group of motions was found by Kasner,[10] with the line element given by

$$ds^2 = z^{2p}\, dt^2 - z^{2q}\, dx^2 - z^{2r}\, dy^2 - dz^2, \tag{6}$$

where p, q, and r satisfy the relations

$$p + q + r = 1, \quad p^2 + q^2 + r^2 = 1. \tag{7}$$

Since the hypersurfaces $z = \text{const}$ have indefinite metrics, other solutions with the same symmetry group also exist, but only the above solution will admit by further specification a 4-parameter group with the required structures [Eqs. (4)].

From a theorem given by Taub[11] it is clear that we cannot expect solutions with the above group structure to have everywhere a finite Riemann tensor. The metric (6) has singularities for $z = 0$ and $z \to \pm\infty$. It is shown in Ref. 12 that physical singularities are those occurring in the canonical form of the Riemann tensor (as given by Petrov). For the Kasner solution, the space has an essentially singular hypersurface $z = 0$ (which can be transformed to any $z = \text{const}$, but not completely away), showing that there the space is strongly curved, presumably corresponding to the presence of matter. This is similar to the exterior Schwarzschild solution, which has an essential singularity for $r = 0$.

From the metric (6) we get the following $\Gamma \neq 0$:

$$\begin{aligned} &\Gamma^3_{00} = pz^{2p-1}, \quad \Gamma^3_{11} = -qz^{2q-1}, \quad \Gamma^3_{22} = -rz^{2r-1}, \\ &\Gamma^0_{03} = p/z, \qquad \Gamma^1_{13} = q/z, \qquad \Gamma^2_{23} = r/z; \end{aligned} \tag{8}$$

TABLE I. How space and group structure varies with the parameters p, q, and r. We use the symbols D for the degenerate type I with two equal eigenvalues and O for three equal eigenvalues.

Values of parameters p, q, r	$p \neq q \neq r$	$p = -\frac{1}{3}, q = r = \frac{2}{3}$	$q = -\frac{1}{3}, p = r = \frac{2}{3}$ or $r = -\frac{1}{3}, p = q = \frac{2}{3}$	$(p, q, r) = (1, 0, 0)$
Type of space	I	D	D	O
G_r	G_3	G_4	G_4	G_{10}
Structure	T_3	$T_1 \otimes [T_2 \otimes_s O(2)]$	$T_1 \otimes [T_2 \otimes_s O(1,1)]$	Poincaré

and from this the nonvanishing components of the Riemann tensor:

$$R_{1212} = +qrz^{-2p}, \quad R_{1313} = z^{2(q-1)}q(q-1),$$
$$R_{1010} = -pqz^{-2r}, \quad R_{2323} = z^{2(r-1)}r(r-1), \quad (9)$$
$$R_{2020} = -prz^{-2q}, \quad R_{0303} = -z^{2(p-1)}p(p-1).$$

For the metric (6) to take one of the forms (5), it is necessary that two of the parameters (p, q, r) be equal; e.g., for $p \neq q = r$, one gets (5a). There are only two possible values of the triplet (p, q, r), which are found from Eq. (7) to be $(1, 0, 0)$ and $(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ and their permutations. But from Eqs. (9) we see that for $(1, 0, 0)$ all components of the Riemann tensor vanish, so that space is flat.

From the eigenvalue equation of the Riemann tensor in the induced 6-dimensional bivector space (see, for example, Ref. 6)

$$|R_{ab} - \lambda g_{ab}| = 0, \quad a, b = 1, 2, \cdots, 6$$

with $R_{iklm} \to R_{ab}$ and $g_{ik}g_{lm} - g_{il}g_{km} \to g_{ab}$, one obtains the type classification for the Kasner solution. The results are given in Table I (see Refs. 13 and 14).

In what follows we discuss some physical aspects of the D-type spaces, but actually we shall work with the Kasner solution (6), keep p, q, and r arbitrary, and remember that for $(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ and its permutations the space has the right symmetry, while for $(1, 0, 0)$ the space is flat.

IV. THE NONRELATIVISTIC LIMIT

In discussing the clock paradox, Møller has defined a homogeneous field by the metric

$$ds^2 = (1 + gz)^2\, dt^2 - dx^2 - dy^2 - dz^2 \quad (10)$$

with the following properties: (a) $R_{iklm} = 0$, space is flat and the field is due to the choice of the coordinate system; (b) it is static; (c) it has the right nonrelativistic limit $g_{00} \simeq 1 + 2\phi_{NR} = 1 + 2gz$; (d) it leads to hyperbolic motion of the $z = \text{const}$ surfaces as seen from a geodesically moving observer (note that this has no absolute meaning because the space admits a G_{10} acting on the whole 4-dimensional manifold).

We should like to compare our solutions with the metric (10). But before we go on to this, we must say something about the physical meaning of the coordinate system as given by Eq. (6).

On the hypersurfaces $z = \text{const}$ we have introduced Gaussian coordinates so that the z lines are geodesics. Because the group G_4 acts only on these hypersurfaces, this has a covariant meaning. Second, our space is static and, therefore, precisely one G_1 exists with timelike hypersurface orthogonal congruence.[6] If we choose the t lines along the congruence and Fermi-transport the spacelike coordinates along them, we have a preferred coordinate system (t lines are not geodesic). The metric then satisfies the conditions

$$g_{01} = g_{02} = g_{03} = g_{13} = g_{23} = 0, \quad g_{33} = -1,$$

and the coordinate system is fixed up to transformations of the form

$$\bar{t} = at + b, \quad \bar{x} = f_1(x, y),$$
$$\bar{z} = z + d, \quad \bar{y} = f_2(x, y).$$

If we now make use of this freedom for the coordinates and transform according to

$$t = \bar{t}g^p, \quad x = \bar{x}g^p, \quad y = \bar{y}g^r,$$
$$z = (1 + g\bar{z})/g, \quad g = \text{const}, \quad (11)$$

the line element (6) becomes

$$ds^2 = (1 + g\bar{z})^{2p}\, d\bar{t}^2 - (1 + g\bar{z})^{2q}\, d\bar{x}^2 - (1 + g\bar{z})^{2r}\, d\bar{y}^2 - d\bar{z}^2; \quad (12)$$

for $p = 1$ and $q = r = 0$, we have exactly the metric (10).

The unit tangent vector to the static congruence of (12) is

$$u^i = (1 + g\bar{z})^{-p}\delta^i_0.$$

Then, the 4-force exerted on a particle with unit mass is Fermi-propagated along the t lines and is given by

$$\frac{\delta u^i}{\delta s} = \frac{pg}{1 + g\bar{z}}\,\delta^i_3.$$

Since $\delta u^i/\delta s$ is invariant under the groups G_4 and $(\delta/\delta s)(u_{i;k} - u_{k;i}) = 0$, one can define a scalar potential ϕ by

$$\frac{\delta u_i}{\delta s} = -\phi_{,i}, \quad \text{so that} \quad \phi = p \ln(1 + g\bar{z}). \tag{13}$$

Over regions where $g\bar{z} \ll 1$ the nonrelativistic limit leads to

$$\phi_{\mathrm{NR}} = pg\bar{z},$$

which is the classical homogeneous field.

The 3-acceleration of a freely moving particle in this limit is obtained from

$$\frac{d^2\bar{x}^\alpha}{dt^2} \simeq \Gamma^\alpha_{00} = -pg(1 + g\bar{z})^{2p-1}\delta^\alpha_3$$
$$\simeq -pg\delta^\alpha_3 = -\phi_{\mathrm{NR},\alpha}, \quad \alpha = 1, 2, 3. \tag{14}$$

Only g_{00} contributes to ϕ_{NR}, although the derivations from the Euclidean values of the spatial components are of the same order. In this limit the acceleration of a freely moving particle, as well as the force acting on a particle at rest, is constant. The difference between the fields given in (12) for $p = 1$ and $p \neq 1$ appears in higher orders of $g\bar{z}$ and, of course, in the constant p. For the solution with $p = -\frac{1}{3}$, the field has the opposite direction and the motion of the particle is reversed (see Sec. V).

V. TIMELIKE GEODESICS

We now proceed to investigate how a freely falling, i.e., a geodesically moving, particle is seen by an observer at rest in a coordinate system where the singular hypersurface is at rest. We solve the geodesic equation in the system given by the metric (6) and this solution can then be transformed to (12). From the Γ's given in (8) the equations for the geodesics are found to be

$$\ddot{t} + \frac{2p}{z}\dot{t}\dot{z} = 0,$$
$$\ddot{x} + \frac{2q}{z}\dot{x}\dot{z} = 0,$$
$$\ddot{y} + \frac{2r}{z}\dot{y}\dot{z} = 0, \tag{15}$$
$$\ddot{z} + pz^{2p-1}\dot{t}^2 - qz^{2q-1}\dot{x}^2 - rz^{2r-1}\dot{y}^2 = 0,$$

with

$$\dot{x}^i = \frac{d}{ds}x^i(s).$$

We are only interested in timelike geodesics and to simplify the solution we take the initial values such that the particle is at rest when $s = 0$; i.e., for the 4-velocity we have

$$\dot{x}^i(0) = \delta^i_0 z^{-p} \quad \text{satisfying} \quad \dot{x}^i g_{ik}\dot{x}^k = 1.$$

From the geodesic equations it then follows that $\ddot{x}(0) = 0$ and $\ddot{y}(0) = 0$, so that these components of the velocity will remain zero if they were zero initially. We need then only the z-t dependence of Eqs. (15). One immediately gets an integral of the first equation

$$\dot{t}(s) = [z(0)/z(s)^2]^p, \tag{16}$$

where the integration constant is chosen so as to fulfill the initial condition. Introducing this into the last Eq. (15), with only the t-z dependence, gives

$$\ddot{z}(s) + pz(0)^{2p}z(s)^{-(2p+1)} = 0$$

and a first integral

$$\dot{z}(s) = \pm\{[z(0)/z(s)]^{2p} + C\}^{\frac{1}{2}}, \tag{17}$$

where C must be equal to -1 and the over-all sign must be chosen from physical considerations. For $\dot{z}(s)$ to be real,

$$[z(0)/z(s)]^{2p} \geq 1.$$

It follows that

$$z(0) \gtrless z(s) \quad \text{for} \quad p \gtrless 0,$$

so that in Eq. (17) we have

$$\mp \quad \text{for} \quad p \gtrless 0.$$

The 3-velocity can be calculated from (16) and (17):

$$\frac{dz}{dt} = \mp z^p(s)\left[1 - \left(\frac{z(s)}{z(0)}\right)^{2p}\right]^{\frac{1}{2}}, \quad \text{for} \quad p \gtrless 0.$$

If we plot dz/dt against z, we can roughly deduce from this the trajectory of the freely moving particle (see Fig. 1). Here one should remember that, although these trajectories are not generally covariant, they

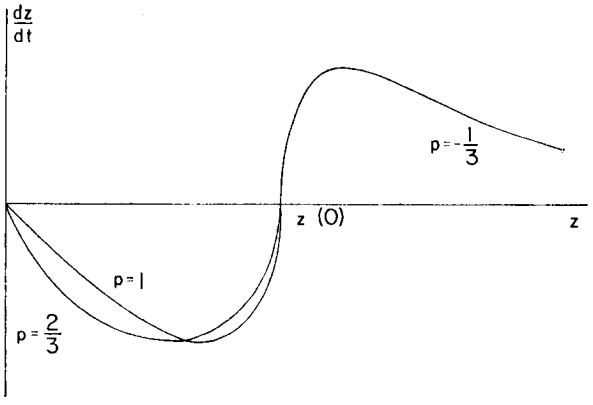


FIG. 1. Schematic velocity-space dependence for a geodesically moving particle in fields with $p = 1$, $p = \frac{2}{3}$, and $p = -\frac{1}{3}$.

are given in the preferred coordinate system owing to the symmetry of the manifold (Sec. IV) and the z lines are geodesics.

$p = 1$: For $s = 0$, the particle is at rest at the point $z = z(0)$. It is then accelerated towards the singular plane $z = 0$. After a certain time its velocity decreases and, finally, the particle comes to rest at $z = 0$ for $t \to +\infty$.

$0 < p < 1$: Here again the particle is accelerated towards the $z = 0$ plane but reaches this after a finite time; the velocity has then decreased to zero.

$p < 0$: For this value of p the motion is reversed. The particle is accelerated from the initial point $z = z(0)$ in the opposite direction, i.e., from the singular plane away. Again, after a certain time the velocity decreases and will tend to zero for $z \to +\infty$.

The transformation (11) to the coordinate system defined by (12) does not change the main features of the trajectory, but the singular surface is shifted to $z = -1/g$. We see from this that the field with $p = \frac{2}{3}$, i.e., a space with group structure (4b), has features similar to the field for $p = 1$.

VI. CONCLUSION

We have seen that in GR, from the point of view of symmetries, a homogeneous field as in the Newtonian theory does not exist. On the other hand, we were led only by symmetry requirements to static fields with nonvanishing Riemann tensor having the right nonrelativistic limit. Moreover, the motion of a free particle in a field admitting a group of the structure (4b) reflects properties of the motion in the usual defined homogeneous field, for which space is actually flat.

Because our spaces have essentially singular timelike hypersurfaces, one could try to find a complete solution of the Einstein field equations with a corresponding distribution of matter as source term, so that the Riemann tensor is regular everywhere. Physically meaningful solutions can only be expected for spaces with $p > 0$, since $p < 0$ would have the effect of antigravitation.

ACKNOWLEDGMENTS

The author is indebted to Professor W. Thirring for his hospitality at the Institute for Theoretical Physics in Vienna. He also thanks Professor A. Salam and Professor P. Budini, as well as the International Atomic Energy Agency, for the Fellowship kindly extended to him at the International Center for Theoretical Physics, Trieste, where the manuscript was prepared. Further, he thanks Dr. W. Graf for many valuable discussions and for drawing his attention to the Kasner solution.

* Work supported in part by the Österreichischer Forschungsrat.

[1] C. Møller, Kgl. Danske Videnskab **20**, 19 (1943).
[2] F. Rohrlich, Ann. Phys. (N.Y.) **22**, 169 (1963).
[3] Throughout this paper, groups of motions are also called symmetry groups and are defined only locally by their corresponding Lie algebra. We always mean real groups with real structures.
[4] J. L. Anderson, *Principles of Relativity Physics* (Academic Press, New York, 1967), Chap. 4.
[5] For further details, see also P. C. Aichelburg, University of Vienna, Preprint, 1968.
[6] J. Ehlers and W. Kundt, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (Wiley & Sons, New York, 1962), p. 49.
[7] A. Petrov, *Einstein-Räume* (Akademie-Verlag, Berlin, 1966), Chap. IV.
[8] P. Jordan, J. Ehlers, and W. Kundt, Akad. Wiss. Lit (Mainz) Abhand. Math-Nat. Kl. **2**, 21 (1960).
[9] P. F. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, N.J., 1966), sixth printing.
[10] E. Kasner, Trans. Amer. Math. Soc. **27**, 249 (1925).
[11] A. H. Taub, Ann. Math. **53**, 472 (1951).
[12] V. Joseph, Proc. Cambridge Phil. Soc. **53**, 836 (1957).
[13] F. A. E. Pirani, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (Wiley & Sons, New York, 1962), p. 199.
[14] R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Stability of Time-Dependent Particlelike Solutions in Nonlinear Field Theories. I

D. L. T. ANDERSON* AND G. H. DERRICK
Department of Theoretical Physics, School of Physical Sciences, University of St. Andrews, Fife, Scotland

(Received 25 April 1969)

The stability of time-dependent particlelike solutions of the form $\psi = \varphi(r)e^{-i\omega t}$ is examined for the nonlinear field $\nabla^2\psi - c^{-2}\partial^2\psi/\partial t^2 = \kappa^2\psi - \mu^2\psi\psi^*\psi$. It is found that such solutions are unstable for all ω.

1. INTRODUCTION

In this paper, we investigate the stability of time-dependent particlelike solutions of the nonlinear field equation

$$\nabla^2\psi - c^{-2}\frac{\partial^2\psi}{\partial t^2} = \kappa^2\psi - \mu^2\psi\psi^*\psi, \tag{1}$$

where κ and μ are real parameters. We are interested in a nonlinear field, since with such a field it may be possible to describe an elementary particle without some of the difficulties of conventional field theory. In a nonlinear field, one does not have to represent particles as point singularities, but can consider them as local concentrations of some suitable quantity, usually the energy density. Once the field has been specified, one can determine the interaction between particles and what happens to a particle when it is disturbed. Does its energy remain localized or does it dissipate? The question of particle interaction has been studied by a large number of authors,[1] beginning with Born and Infeld. In particular, Rosen and Rosenstock[1] have shown that the interaction between two particles at large distances in the above field is governed by the Yukawa potential. Work has also been done on the stability of 1-particle solutions. Hobart and one of us (G. H. D.)[2] have shown independently that time-independent solutions to a large class of nonlinear Lorentz-invariant field equations are unstable, and Rosen[3] has investigated a solvable time-independent field. Little, so far, has been done on the stability of time-dependent particlelike solutions, which forms the subject of the present investigation. The treatment given here is purely classical.

2. PARTICLELIKE SOLUTIONS OF (1)

Equation (1) may be derived from the variation principle $\delta\int \mathfrak{L}\, d^3\mathbf{r}\, dt = 0$, with the Lagrangian density[4]

$$\mathfrak{L} = c^{-2}\left|\frac{\partial\psi}{\partial t}\right|^2 - |\nabla\psi|^2 - \kappa^2|\psi|^2 + \tfrac{1}{2}\mu^2|\psi|^4. \tag{2}$$

The energy density corresponding to (2) is

$$\mathfrak{E} = c^{-2}\left|\frac{\partial\psi}{\partial t}\right|^2 + |\nabla\psi|^2 + \kappa^2|\psi|^2 - \tfrac{1}{2}\mu^2|\psi|^4, \tag{3}$$

while the charge and current densities

$$\rho = -i\sigma\left(\psi\frac{\partial\psi^*}{\partial t} - \psi^*\frac{\partial\psi}{\partial t}\right), \tag{4}$$

$$\mathbf{j} = -i\sigma c^2(\psi^*\nabla\psi - \psi\nabla\psi^*) \tag{5}$$

satisfy the continuity equation $\nabla\cdot\mathbf{j} + \partial\rho/\partial t = 0$. The constant σ is arbitrary but is fixed when we specify how the field ψ interacts with the electromagnetic field A^ν. Adopting the standard prescription,

$$\nabla \to \nabla - \frac{ie}{\hbar c}\mathbf{A} \quad \text{and} \quad \frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \frac{ie}{\hbar}A^0,$$

yields $\sigma = e/\hbar c^2$.

We shall restrict ourselves to time-dependent solutions of the form

$$\psi = \varphi(r)e^{-i\omega t}, \tag{6}$$

where ω is a real parameter and $\varphi(r)$ is real and spherically symmetric. Then, (1) reduces to

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{\partial\varphi}{\partial r} = \left(\kappa^2 - \frac{\omega^2}{c^2}\right)\varphi - \mu^2\varphi^3. \tag{7}$$

We seek solutions with finite energy, and for which φ and all its Cartesian-space derivatives exist everywhere. This implies $(d\varphi/dr)_{r=0} = 0$, $\varphi \to 0$ as $r \to \infty$,

$$\kappa^2 - \omega^2/c^2 > 0.$$

Making the transformations

$$r' = r(\kappa^2 - \omega^2/c^2)^{\frac{1}{2}},$$
$$\varphi' = \mu\varphi(\kappa^2 - \omega^2/c^2)^{-\frac{1}{2}}, \tag{8}$$

φ' is found to satisfy

$$\frac{d^2\varphi'}{dr'^2} + \frac{2}{r'}\frac{d\varphi'}{dr'} = \varphi' - \varphi'^3. \tag{9}$$

This equation has been extensively studied.[5] There is an infinity of solutions, the simplest having no nodes, the next simplest one node, etc. The three solutions of lowest order are shown in Fig. 1. The energy $E = \int \mathfrak{E}\, d^3\mathbf{r}$ can be reduced to

$$E = 2\kappa\mu^{-2}(1 - \omega'^2)^{-\frac{1}{2}}I, \tag{10}$$

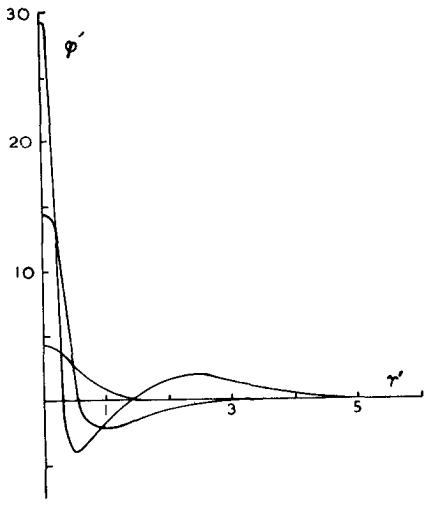


FIG. 1. The three lowest-order solutions to (9). φ' is plotted against r' for the solutions of (9) having 0, 1, 2 nodes, respectively.

where $\omega' = \omega/\kappa c$, $I = 4\pi \int \varphi'^2 r'^2\, dr'$. Since ω' is an arbitrary parameter in the range $-1 < \omega' < 1$, this gives a continuous spectrum of allowed energy values. The values of the integral $\int \varphi'^2 r'^2\, dr'$ for the three lowest-order solutions are, respectively, 1.502, 9.63, 29.18. Using (6) and (8), we can reduce (3) to

$$\mathcal{E} = \frac{\kappa^4}{\mu^2}(1-\omega'^2)^2\left[\left(\frac{1+\omega'^2}{1-\omega'^2}\right)\varphi'^2 + \left(\frac{\partial\varphi'}{\partial r'}\right)^2 - \tfrac{1}{2}\varphi'^4\right].$$

The reduced energy density is defined as $\mathcal{E}' = \mathcal{E}\mu^2/\kappa^4$. When $\omega' = 0$,

$$\mathcal{E}' = \varphi'^2 + (\partial\varphi'/\partial r')^2 - \tfrac{1}{2}\varphi'^4.$$

In Fig. 2 the lowest-order solution to (9) is shown in more detail than in Fig. 1, and in Fig. 3 the reduced energy density is plotted for this solution when $\omega' = 0$. The field is perhaps unsatisfactory in one respect: the energy density is negative in a certain region of space. It would be preferable to have $\mathcal{E}$ everywhere positive, especially if one wants to think of a particle as a local concentration of energy. However, the total integrated energy is, in fact, positive for solutions of type (6), for which we have

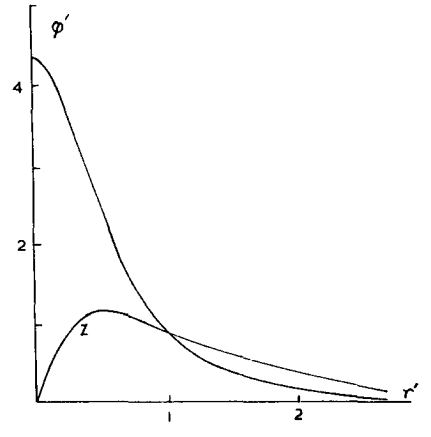


FIG. 2. Graph of φ' and $Z = r'\varphi'$ for the lowest-order solution of (9). Z can be seen to have a maximum value of 1.2.

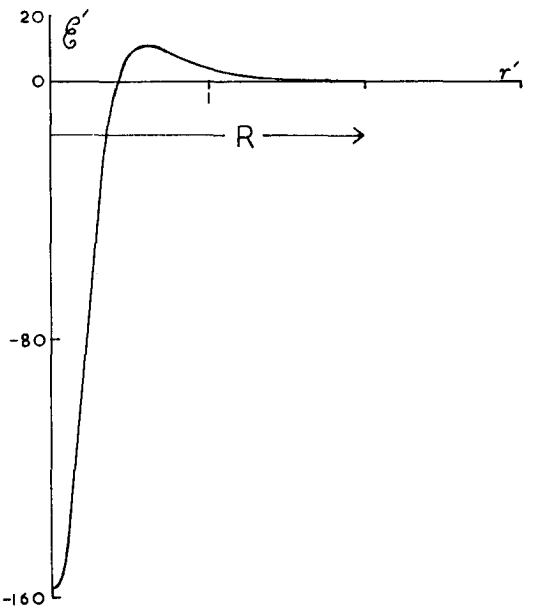


FIG. 3. Reduced energy density $\mathcal{E}'$ for the lowest-order solution to (9) when $\omega' = 0$. The particle radius R is marked. For $r' > R$, $\mathcal{E}'$ is negligible. For $r' < 0.4$, $\mathcal{E}'$ is negative.

$$\mathcal{E} - \nabla\cdot(\varphi\nabla\varphi) = (\omega^2/c^2 + \kappa^2)\varphi^2 - \varphi\nabla^2\varphi - \tfrac{1}{2}\mu^2\varphi^4$$
$$= 2\omega^2 c^{-2}\varphi^2 + \tfrac{1}{2}\mu^2\varphi^4 > 0.$$

On integrating over all space, the divergence on the left-hand side disappears, yielding a positive total energy. The size of a "particle" is not a well-defined quantity, but could be taken as the distance denoted by R in Fig. 3. For $r' > R$, $\mathcal{E}'$ is negligible. R is ≈ 2 units of r'. From Fig. 2 we see that φ' is also small for $r' > 2$.

3. STABILITY BY FIRST-ORDER PERTURBATION THEORY

A. The First-Order Perturbation Equations

Let us denote by $\psi_0 = \varphi_0(r)e^{-i\omega t}$ the unperturbed state and by $\psi_1(r, t)$ the disturbance, assumed small, at least initially. We shall concentrate on the case where $\varphi_0(r)$ is the lowest-order solution of (7). Putting $\psi = \psi_0 + \psi_1$ in (1) and keeping only up to first order in ψ_1 gives

$$\nabla^2\psi_1 - c^{-2}\frac{\partial^2\psi_1}{\partial t^2} = (\kappa^2 - 2\mu^2\psi_0^*\psi_0)\psi_1 - \mu^2\psi_0^2\psi_1^*,$$

$$\nabla^2\psi_1^* - c^{-2}\frac{\partial^2\psi_1^*}{\partial t^2} = (\kappa^2 - 2\mu^2\psi_0^*\psi_0)\psi_1^* - \mu^2\psi_0^{*2}\psi_1. \tag{11}$$

Let us try a solution to (11) of the form

$$\psi_1 = \eta(\mathbf{r})e^{-i(\Omega+\omega)t} + \chi^*(\mathbf{r})e^{i(\Omega^*-\omega)t}, \tag{12}$$

where $\Omega = \Omega_r + i\Omega_i$ is a complex constant. Then, defining $\Omega' = \Omega_r' + i\Omega_i' = \Omega/\kappa c$ and using (7) and (8), we can reduce (11) to

$$\left(\nabla'^2 + \frac{(\Omega' + \omega')^2 - 1}{(1 - \omega'^2)} + 2(\varphi_0')^2\right)\eta = -(\varphi_0')^2\chi,$$

$$\left(\nabla'^2 + \frac{(\Omega' - \omega')^2 - 1}{(1 - \omega'^2)} + 2(\varphi_0')^2\right)\chi = -(\varphi_0')^2\eta,$$

$$-1 < \omega' < 1, \quad (13)$$

where φ_0' is the lowest-order solution to (9) (Fig. 2) and the boundary conditions are η, χ, both finite at $r' = 0$ and η, $\chi \to 0$ as $r' \to \infty$. If $(\omega', \Omega', \eta, \chi)$ is a solution to (13), then so are $(\omega', -\Omega', \chi, \eta)$, $(\omega', \Omega'^*, \eta^*, \chi^*)$, and $(-\omega', -\Omega', \eta, \chi)$. Hence, there is no loss of generality in taking $0 < \omega' < 1$, $\Omega_r' > 0$, $\Omega_i' > 0$. If Ω is not purely real, then $e^{-i\Omega t} = e^{-i\Omega_r t}e^{\Omega_i t}$ will build up with time and destroy the solution ψ_0. Thus, Ω_i^{-1} gives a measure of the lifetime of ψ_0.

B. Numerical Solutions of (13)

Equation (13) is a type of eigenvalue problem. For a given ω', it is necessary to find eigenvalues Ω' and eigenfunctions (η, χ) which satisfy (13). If Ω' is real, the particle will be stable, while, if Ω' has an imaginary part, it will be unstable. An intensive search for solutions to (13) has been made numerically. $\varphi_0'(r')$ is the lowest-order solution to (9) which is not known analytically. Thus, it is necessary to feed into (13) the numerical solution φ_0', derived from (9), and solve (13) numerically.

Let us examine first the case where η, χ are spherically symmetric. Instead of the quantities η, χ, it is more convenient numerically to deal with $g = r'\eta$, $f = r'\chi$. Then, the boundary condition at $r' = 0$ is $g = f = 0$. The asymptotic form of (13) can easily be found, and our numerical method was based on it. Knowing the asymptotic form for large r', we integrated in to $r' = 0$, where we evaluated g and f. If g and f were not both 0 at $r' = 0$, we corrected the eigenvalues and tried again until the boundary conditions were satisfied. No solutions existed if Ω_r' was not zero. For $\Omega_r' = 0$, we found only one value of Ω_i' for a given ω', and Fig. 4 shows this eigenvalue Ω_i' as a function of ω'. (We disregard the trivial solution $\Omega' = 0$, $\eta = -\chi$, corresponding to no perturbation.)

It has previously been stated that we expected to find eigenvalues Ω' for a given ω', and this is indeed confirmed. But there remains the possibility that additional special solutions to (13) may exist for certain particular ω'. An intensive search for such solutions was made, but none was found. Figure 4 thus shows all the known eigenvalues for the spherically symmetric solutions to (13).

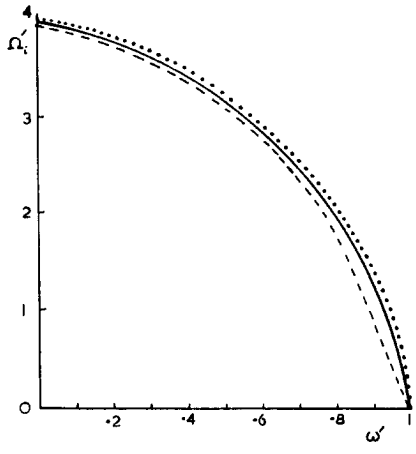


FIG. 4. Eigenvalues of (13). The eigenvalue Ω_i' is shown plotted against ω' for the nontrivial solutions to (13). The dashed curve indicates results obtained by the numerical method (Sec. 3B), and the solid curve results by the variational method (Sec. 3C). The dotted curve limits the region within which the eigenvalues must lie (Appendix A). Note that this limiting curve gives a good upper bound to the values of Ω_i' derived from Secs. 3B and 3C.

We now turn to nonspherically symmetric solutions to (13). The equation can be separated in spherical polar coordinates by assuming that the angular dependence of η and χ is given by the spherical harmonics Y_l^m. One solution of this type is known for $l = 1$ corresponding to the translational invariance of (1), viz.,

$$\Omega' = 0, \quad \eta = \chi = \frac{\partial\varphi_0'}{\partial r'} Y_1^m.$$

We sought this solution numerically and found it, but could find no other nonspherically symmetric solutions to (13). In Appendix A, it is proved that no unstable solutions can exist for $l > 1$.

C. Variational Solutions of (13)

If (13) is rewritten in the form

$$\left[\begin{pmatrix} \nabla'^2 - 1 + 3\varphi_0'^2 & 0 \\ 0 & \nabla'^2 - 1 + \varphi_0'^2 \end{pmatrix} + \frac{2\Omega'\omega'}{(1 - \omega'^2)}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{\Omega'^2}{(1 - \omega'^2)}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} \eta + \chi \\ \eta - \chi \end{pmatrix} = 0 \quad (14)$$

and ω_B, H, B, and ξ are defined as follows,

$$\omega_B = \frac{\Omega'}{(1 - \omega'^2)^{\frac{1}{2}}},$$

$$H = \begin{pmatrix} -\nabla^2 + 1 - 3\varphi_0'^2 & 0 \\ 0 & -\nabla^2 + 1 - \varphi_0'^2 \end{pmatrix}, \quad (15)$$

$$B = \frac{-\omega'}{(1 - \omega'^2)^{\frac{1}{2}}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \xi = \begin{pmatrix} \eta + \chi \\ \eta - \chi \end{pmatrix},$$

then (14) can be rewritten

$$(\omega_B^2 I - 2\omega_B B - H)\xi = 0, \tag{16}$$

which is now in a form examined by Barston.[6] H and B are Hermitian, but, since ω_B can be complex, $\Pi = (\omega_B^2 I - 2\omega_B B - H)$ is not. However, by doubling the dimensions of the vector space we can construct the Hermitian operator

$$L = \begin{pmatrix} 0 & \Pi \\ \Pi^\dagger & 0 \end{pmatrix}.$$

Then, it is sensible to ask for the real eigenvalues of L, i.e., solutions of the problem $L\zeta = \lambda\zeta$:

$$\begin{pmatrix} 0 & \Pi \\ \Pi^\dagger & 0 \end{pmatrix}\begin{pmatrix} v \\ u \end{pmatrix} = \lambda\begin{pmatrix} v \\ u \end{pmatrix},$$

where

$$\zeta = \begin{pmatrix} v \\ u \end{pmatrix}$$

and v and u are 2-component vectors. Since $\Pi^\dagger = \Pi^*$, there is no loss in generality in taking $v = u^*$. Thus, we have

$$\begin{pmatrix} 0 & \Pi \\ \Pi^* & 0 \end{pmatrix}\begin{pmatrix} u^* \\ u \end{pmatrix} = \lambda\begin{pmatrix} u^* \\ u \end{pmatrix},$$

yielding the variation principle $\delta[\lambda] = 0$ for arbitrary variations δu, where

$$[\lambda] = \frac{\int (u, u^*)\begin{pmatrix} 0 & \Pi \\ \Pi^* & 0 \end{pmatrix}\begin{pmatrix} u^* \\ u \end{pmatrix} \cdot d^3\mathbf{r}}{\int (u, u^*)\begin{pmatrix} u^* \\ u \end{pmatrix} d^3\mathbf{r}} = \frac{\mathrm{Re}\left[\int (u\Pi u)\, d^3\mathbf{r}\right]}{\int uu^*\, d^3\mathbf{r}}. \tag{17}$$

The eigenvalue λ is a function of ω_B and ω_B^*, and the problem is to find those values of ω_B for which λ vanishes. A suitable trial wavefunction is

$$u = \begin{pmatrix} \gamma e^{-\alpha r} \\ \beta e^{-\alpha r} \end{pmatrix}, \tag{18}$$

where α, β, and γ are complex parameters $\alpha = \alpha_r + i\alpha_i$, etc. Because the system is homogeneous, any one of γ_r, γ_i, β_r, and β_i can be normalized to unity. If one seeks the case $\eta = -\chi$, $\Omega' = 0$ for any ω', then one will obviously normalize β_r or β_i, while if one seeks solutions more closely resembling $\eta = +\chi$, either γ_r or γ_i would be normalized. (The case $\eta = +\chi$ occurs when $\omega' = 0$.) It must still be decided what would be a suitable choice for φ_0'. As previously mentioned, no analytic solution for φ_0' is known, but variational solutions to (9) have been found by Betts *et al.*[5] The simplest, and in some ways most suitable, form for φ_0' is $\varphi_0' = 4 \times 2^{\frac{1}{2}} \exp(-3^{\frac{1}{2}}r')$. With the trial function (18) and with φ_0' of the above form, $[\lambda]$ was evaluated. We then sought solutions to the problem

$$[\lambda] = 0, \quad \frac{\partial[\lambda]}{\partial\mu_A} = 0,$$

(a) when $\mu_A = \gamma_i, \beta_r, \beta_i, \alpha_r, \alpha_i$ ($\gamma_r = 1$), for fixed (ω', Ω_r'), in terms of the six unknowns $\Omega_i', \gamma_i, \beta_r, \beta_i, \alpha_r, \alpha_i$,

(b) when $\mu_A = \gamma_r, \gamma_i, \beta_r, \alpha_r, \alpha_i$ ($\beta_i = 1$), for fixed (ω', Ω_r'), in terms of the six unknowns $\Omega_i', \gamma_r, \gamma_i, \beta_r, \alpha_r, \alpha_i$.

In both (a) and (b), solutions were found only if Ω_r' was zero. In (a), only one Ω_i' was found for a given ω', the values of which are given in Table I. In (b), we found the expected trivial case $\Omega' = 0$, $\eta = -\chi =$ const $\times\ \varphi_0'$. In addition to (a) and (b), we tried other types of solution. As pointed out in Sec. 3B, special solutions might exist for certain discrete ω'. Thus, ω' must be allowed to vary in order to let it converge on such a solution. A random search was made for solutions to the set of equations

$$[\lambda] = 0, \quad \frac{\partial[\lambda]}{\partial\mu_A} = 0,$$

(c) when $\mu_A = \gamma_i, \beta_r, \beta_i, \alpha_r, \alpha_i$ ($\gamma_r = 1$), solving for the unknowns $\Omega_r', \Omega_i', \omega', \gamma_i, \beta_r, \beta_i, \alpha_r, \alpha_i$,

(d) when $\mu_A = \gamma_r, \gamma_i, \beta_r, \alpha_r, \alpha_i$ ($\beta_i = 1$), solving for the unknowns $\Omega_r', \Omega_i', \omega', \gamma_r, \gamma_i, \beta_r, \alpha_r, \alpha_i$.

Solutions to (c) were only found for $\Omega_r' = 0$ and (Ω_i', ω') lying on the curve obtained from (a). Solutions to (d) were only found for $\Omega_r' = 0$, and $\Omega_i' = 0$, i.e., the trivial case.

TABLE I. Values of Ω_i' calculated from numerical and variational methods.

ω'	Ω_i' (numerical)	Ω_i' (variational)
0	3.95	3.96
0.1	3.93	3.94
0.2	3.85	3.86
0.3	3.73	3.74
0.4	3.54	3.56
0.5	3.29	3.32
0.6	2.97	2.99
0.7	2.51	2.57
0.8	1.89	2.00
0.9	0.92	1.18

D. Comparison of Results

For the trivial solution, the variational results give $\Omega' = 0$, $\eta = -\chi =$ const $\times \exp(-3^{\frac{1}{2}}r')$, i.e., apart

from the arbitrary normalizing constant, we get the variational solution for φ_0', with Ω' equal to zero. In the numerical case, we do not expect to get $\Omega' = 0$ exactly nor to get $\eta = -\chi \propto \varphi_0'$ exactly, since there will always be some numerical error involved. What is found in practice is that Ω' is very small and $\eta = -\chi \approx \text{const} \times \varphi_0'$. In neither method do we find solutions for nonzero Ω_r'. In both methods we find solutions for $\Omega_r' = 0$, $\Omega_i' \neq 0$. The values of Ω_i' for the two methods are given in Table I and graphs of Ω_i' v ω' for the two methods shown in Fig. 4. It is interesting to note the good agreement obtained. Figure 4 also shows the limiting curve[7] $\bar{\Omega}'(\omega') = 3.95\,(1 - \omega'^2)^{\frac{1}{2}}$ inside which the eigenvalues Ω_i' must lie. Note that it gives a good upper bound to Ω_i'.

4. STABILITY BY DIRECT-PERTURBATION METHOD

Section 3 gives a useful indication of what might happen to ψ_0 when it is disturbed. In particular, it gives a measure of how long the state ψ_0 will live after being disturbed. But the main method we used to investigate the stability of ψ_0 was to disturb it and follow its time development from (1). The "particle" is deemed stable if $\mathcal{E}$ remains everywhere finite and localized, but unstable if its energy is dissipated. Let $\boldsymbol{\rho} = \kappa\mathbf{r}$, $\Psi = \mu\kappa^{-1}\psi$, $\tau = \kappa ct$. Then (1) reduces to

$$\nabla_\rho^2\Psi - \frac{\partial^2\Psi}{\partial\tau^2} = \Psi - \Psi\Psi^*\Psi. \tag{19}$$

The reduced energy density $\mathcal{E}' = \mathcal{E}\mu^2/\kappa^4$ can be found from (3) to be

$$\mathcal{E}' = \left|\frac{\partial\Psi}{\partial\tau}\right|^2 + |\nabla_\rho\Psi|^2 + |\Psi|^2 - \tfrac{1}{2}|\Psi|^4. \tag{20}$$

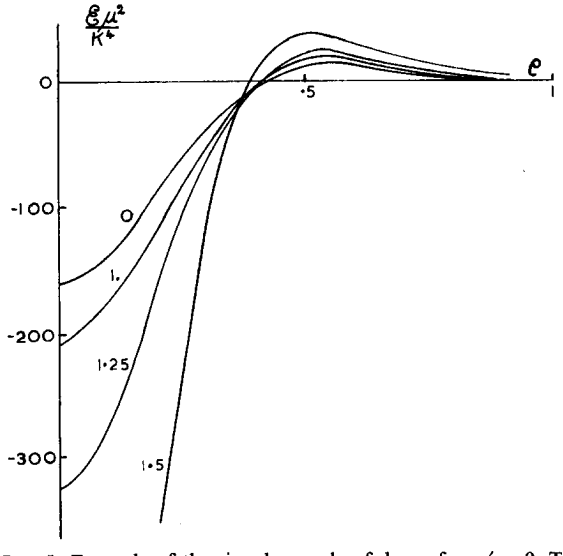


FIG. 5. Example of the singular mode of decay for $\omega' = 0$. The reduced energy density $\mathcal{E}'$ is plotted against the reduced radial distance ρ at a sequence or reduced times $\tau = 0, 1, 1.25, 1.5$. One sees how rapidly the energy density at the origin becomes large and negative. For $\tau \approx 1.5$, the energy density goes singular. When this occurs, we no longer consider $\mathcal{E}'$ to be particlelike.

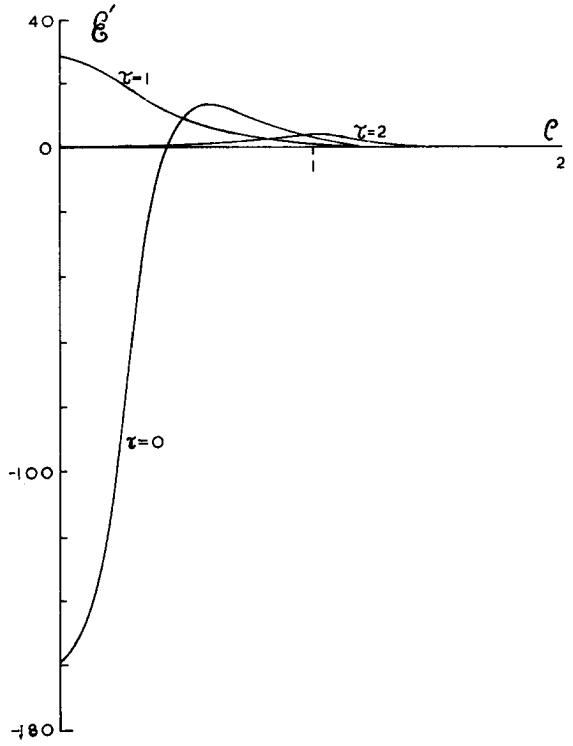


FIG. 6. Example of the dissipative mode of decay for $\omega' = 0$. The reduced energy density $\mathcal{E}'$ is plotted against ρ at a sequence of reduced τ values 0, 1, 2, 3. This shows how $\mathcal{E}'$ rapidly changes in shape and how the energy is pushed away from the origin. For $\tau = 3$, $\mathcal{E}'$ is not visibly different from zero anywhere.

Let us denote by subscript 0 the undisturbed state, i.e.,

$$\Psi_0 = \mu\kappa^{-1}\psi_0 = \mu\kappa^{-1}\varphi_0 e^{-i\omega t} = (1 - \omega'^2)^{\frac{1}{2}}\varphi_0' e^{-i\omega'\tau},$$
$$\rho = r'(1 - \omega'^2)^{-\frac{1}{2}}. \tag{21}$$

As in Sec. 3, φ_0' is taken to be the zero-node solution of (9) (see Fig. 2). We disturb Ψ_0 to Ψ and follow its time development from (19). We shall examine initially the case $\omega' = 0$, when $\Psi_0 = \varphi_0'$ and $\rho = r'$.

The undistrubed state Ψ_0 has a reduced energy density shown in Fig. 3. When Ψ_0 is disturbed, it is found that it decays in one of two distinct ways, which may be designated the singular or dissipative mode. Only spherically symmetric disturbances were used. Examples of the singular and dissipative modes are shown in Figs. 5 and 6, respectively. $(\Psi_0)_{\tau=0}$ and $(\partial\Psi_0/\partial\tau)_{\tau=0}$ were disturbed by a random-number generator operating in the ranges -0.02 to $+0.02$, and -0.4 to $+0.4$, respectively.

One drawback of a field which permits $\mathcal{E}$ to be negative is that singular decays are possible. In this

mode of decay the particle draws more and more negative energy into a decreasing region around the origin and compensates by increasing the positive bump. This decay is probably not physically sensible, but the dissipative one seems reasonable. Although it is easy to classify the decay modes, it is not so easy to say what the decay time is. When shall we consider the particle to have ceased to exist? If we look at Fig. 6, we see that, after 3 units of τ, the energy density corresponding to the state Ψ is not visibly different from zero anywhere; so it would be reasonable to take this as some measure of the decay time. Examination of Fig. 5 reveals that after ≈ 1.5 units of τ, $\mathcal{E}'$ is becoming rapidly more singular. Indeed, the plot of $\mathcal{E}'_{\min}\nu\tau$ in Fig. 7 shows how marked this effect is ($\mathcal{E}'_{\min}$ is the minimum value of $\mathcal{E}'$). In the singular mode of decay it would thus be reasonable to take a lifetime of around 1.5 units of τ. It is stressed, however, that these definitions are arbitrary and are given only as a measure of the decay period. Let us take for the sake of comparison an average figure of ≈ 2 units of τ for the decay time. This compares quite favorably with the estimate from first-order perturbation theory in Sec. 3, which gave $\Omega_i' = 3.95$ and, hence, an estimated lifetime of $(4\kappa c)^{-1}$ units of t, or $\frac{1}{4}$ units of τ (i.e., the same to within an order of magnitude). Figures 8–10 show further examples of decay for the $\omega' = 0$ case. When one puts $\omega' = 0$ in (12) and (13), one obtains

$$\psi_1 = 2\eta(r)\exp(\Omega_i'\tau),$$

where η is a solution to

$$\left[\frac{d^2}{dr'^2} + \frac{2}{r'}\frac{d}{dr'} - \Omega_i'^2 - 1 + 3\varphi_0'^2\right]\eta = 0. \quad (22)$$

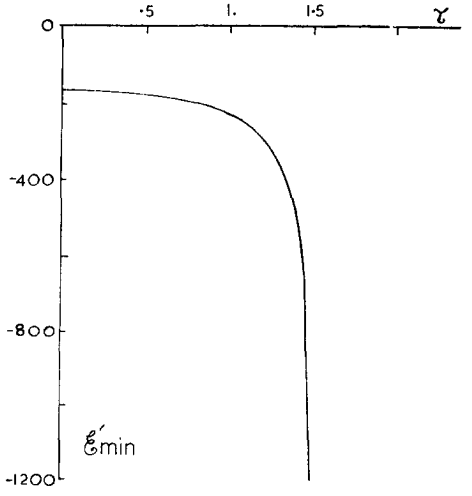


FIG. 7. Plot of $\mathcal{E}'_{\min}\nu\tau$ for the disturbance shown in Fig. 5. This shows how rapidly $\mathcal{E}'$ goes singular at $\tau = 1.5$.

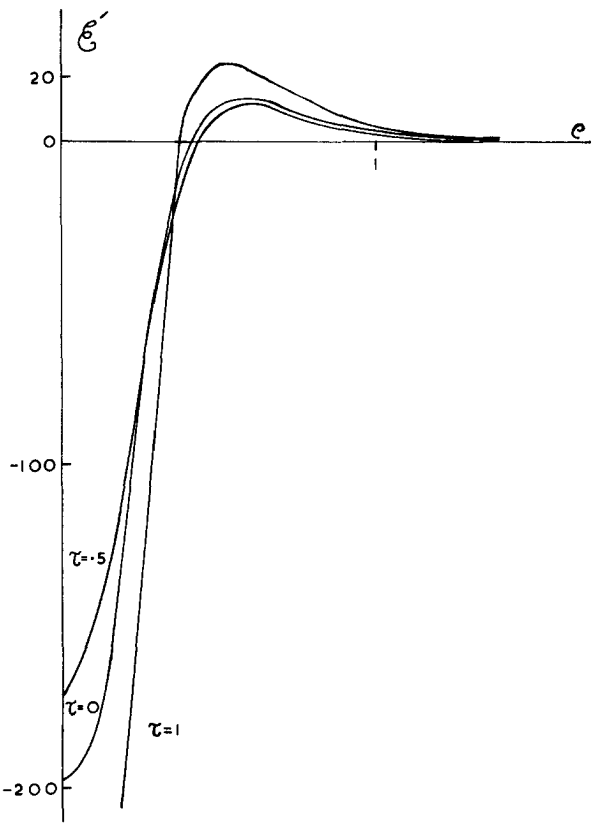


FIG. 8. Singular mode for $\omega' = 0$. The disturbance (23) with $\alpha = 3.5$ is applied at $\tau = 0$. We see that at $\tau = 0.5$, $\mathcal{E}'$ is trying to return to the undisturbed state (Fig. 3), i.e., Ψ_1 is still decaying. However, by $\tau = 1$, Ψ_1 has ceased its exponential decay and $\mathcal{E}'$ is rapidly going singular. Compare with Fig. 5.

This equation has solution for $\Omega_i' = \pm\epsilon$, where ϵ was found numerically to be 3.95 in Sec. 3B (see Table I). Let us consider the disturbance $\psi_1]_{\tau=0} = \text{const} \times \eta(r)$ and $(\partial\psi_1/\partial\tau)_{\tau=0} = -\text{const} \times \epsilon\eta(r)$. Then, ψ_0 should be stable to such a disturbance ψ_1, since ψ_1 decays exponentially with time. Because ψ_1 starts small and gets smaller with increasing time, the results of first-order perturbation theory should hold good and be consistent with direct-perturbation methods. Unfortunately, one can never hope to get the decaying solution exactly in numerical direct methods based on (19), since there will always be some numerical error and the effect of this is to couple in some random disturbance which will build up and wreck the solution. Although it is not possible to obtain ψ_1 exactly, it is still possible to partially verify the results. Let us consider the disturbance

$$\Psi_1]_{\tau=0} = a\xi(\rho), \quad \left.\frac{\partial\Psi_1}{\partial\tau}\right]_{\tau=0} = -a\alpha\xi(\rho), \quad (23)$$

where a is a constant, arbitrary other than that Ψ_1 should be small, $\xi(\rho)$ is the solution $\eta(r')$ of (22) appropriately transformed (according to $\rho = \kappa r$,

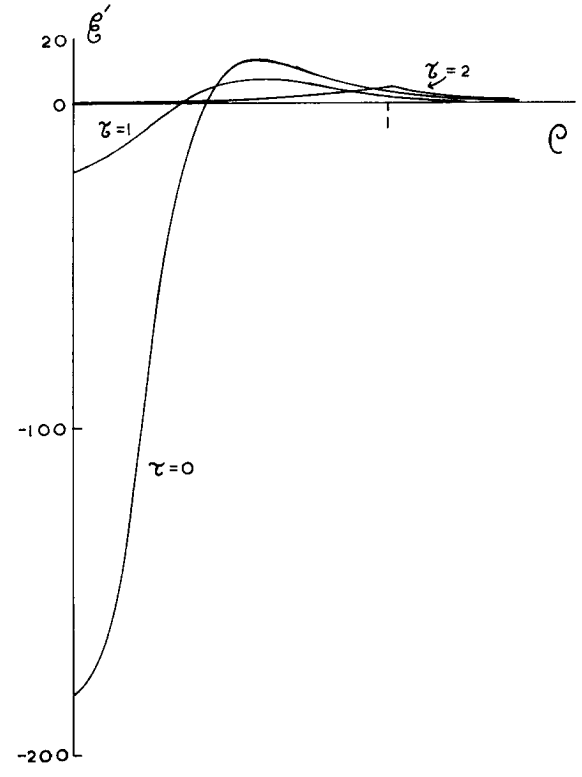


FIG. 9. Further example of dissipative mode for $\omega' = 0$. The applied disturbance is (23) with $\alpha = 5$. Compare with Fig. 6.

$\Psi_1 = \mu\psi_1/\kappa$), and α is a real constant. If $\xi(\rho)$ and α were exactly correct ($\alpha = \epsilon$) and (19) could be solved exactly, then Ψ_0 would be stable to (23). In Figs. 8 and 9, we show the destruction of Ψ_0 by the disturbance (23) for $\alpha = 3.5$ and 5.0, respectively.

We find that, as we apply the disturbance (23) with α approaching ϵ from above and below, the time of decay increases as expected. We also find that for $\alpha < \epsilon$, decay is always by the singular mode, while for $\alpha > \epsilon$ it is always by the dissipative mode, where for this method we find ϵ to be around 4.1 compared with the value 3.95 from Sec. 3. It would thus appear that (23), for $\alpha = \epsilon$, plays a more fundamental role than being merely a stable disturbance. It is, in some sense, the separating curve between the two types of decay. This result cannot be predicted from first-order perturbation theory.

Figure 10 shows a time sequence of events with the following disturbance:

$$\Psi_1]_{\tau=0} = 0, \quad \frac{\partial\Psi_1}{\partial\tau}\Big]_{\tau=0} = 0, \quad \text{for} \quad \rho < 3.5,$$

$$\Psi_1]_{\tau=0} = 2.5 \times 10^8 \times \frac{(\rho - 3.5)}{\rho} e^{-5\rho},$$

$$\frac{\partial\Psi_1}{\partial\tau}\Big]_{\tau=0} = 1 \times 10^8 \times \frac{(\rho - 3.5)}{\rho} e^{-5\rho}, \quad \text{for} \quad \rho > 3.5.$$

In Fig. 10 we plot the positive section of the reduced energy density versus the reduced radial distance ρ. Curve a of Fig. 10 shows $\mathcal{E}'$ immediately after the disturbance has been applied. Note that the disturbance is well outside the particle radius R. Curve b of Fig. 10 shows the situation 0.5τ units later. The disturbance is moving in on the particle. Curves b, c, and d show oscillations forming on the disturbance and travelling inwards. Outgoing oscillations do also occur, but are very small and are not visible on this scale. So far the particle has not been greatly affected, most of the action taking place outside its radius. At e the disturbance is beginning to enter the particle radius and oscillations are building up. Curves f, g, h of Fig. 10 show the violent interaction when the disturbance really hits the particle. In less than 0.2

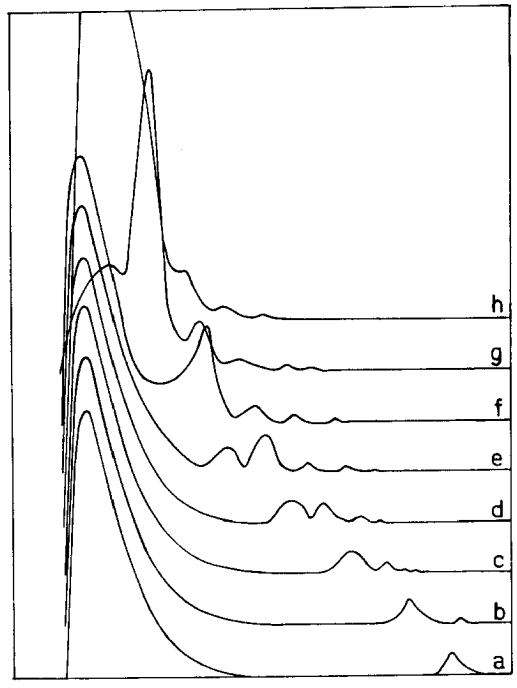


FIG. 10. Time sequence of events leading to a singular decay for $\omega' = 0$. This series of diagrams shows the time development of a disturbance applied to the particlelike solution well outside the particle radius (curve a). The positive part of $\mathcal{E}'$ is plotted against ρ for the series of reduced times $\tau = 0$, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5. It can be seen how the disturbance propagates in a wavelike motion. Until the disturbance crosses the particle radius, the interaction is gentle, but once the disturbance starts interacting within R (curve e), the reaction becomes more drastic as can be seen in curves e, f, g, h. Indeed, in h the initial configuration is unrecognizable. Within 0.2 units after h the particle has decayed by the singular mode (adopting the definition of singular decay given in Sec. 4). If we do not consider interaction with the particle to start until the disturbance is within a distance R, then the time of decay is about 1.7τ units. Note that the disturbance appears to travel with velocity c, travelling from $\rho = 3.5$ to $\rho = 0.0$ in a time $\tau = 3.5$. The scale is the same for all the curves (a–h), viz.,

$$\mathcal{E}': 0 \rightarrow 11, \quad \rho: 0 \rightarrow 4.$$

The time interval is 0.5τ units. The horizontal axis is displaced vertically for different times fo the sake of clarity.

units of τ after h, the particle decays by the singular mode (using the definition of singular given previously). In this type of interaction we should not start measuring decay time until the disturbance starts crossing the particle radius, since it is not until this time that the particle really feels the disturbance. Adopting this convention, the decay time of the above disturbance is about 1.7 units of τ. A dissipative sequence would be similar to the above, but decay this time would occur by the dissipative mode rather than by the singular. Figures 5–10 show the effects of various types of disturbance on the particlelike solution when $\omega' = 0$. For all the disturbances shown in this paper, the decay time for a given mode is more or less the same: 1.5 units of τ for a singular decay, 3 units of τ for a dissipative decay. Indeed, for all disturbances we have applied, we find that singular decay times lie in the range 1 to 2.5 units of τ and dissipative ones between 2 and 4 units of τ, except if we apply the disturbance (23) with $\alpha \to \epsilon$ when it takes longer for decay to set in.

We now examine the time-dependent case ($\omega' \neq 0$). The same two decay modes were found to exist, but no others. In general, Ψ decays more slowly as ω' increases, in agreement with results from Sec. 3, where Ω_i' decreases as ω' increases. Figures 11 and 12 show singular and dissipative decays for the case $\omega' = 0.8$,

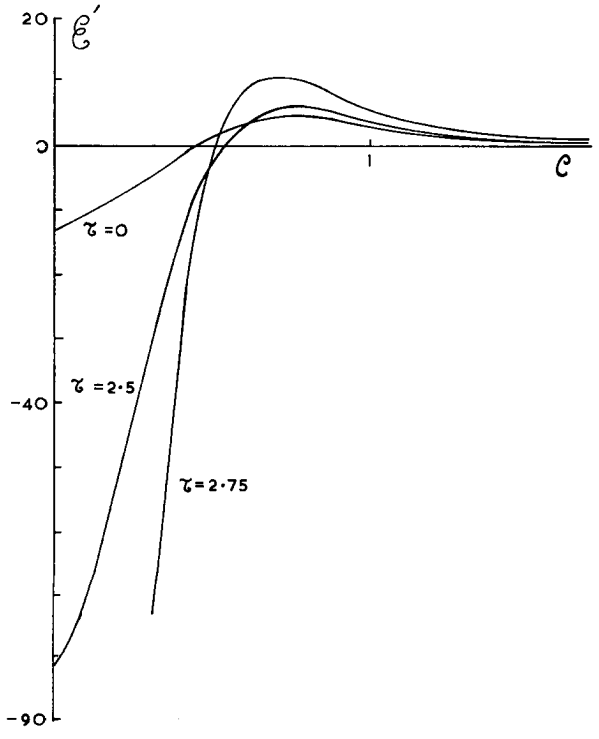


FIG. 11. Singular decay for $\omega' = 0.8$. The behavior is similar to that of the $\omega' = 0$ case (Figs. 5, 8) but the decay-time, about 2.8τ, is longer. Notice that $\mathcal{E}'$ is less negative in the higher ω' case.

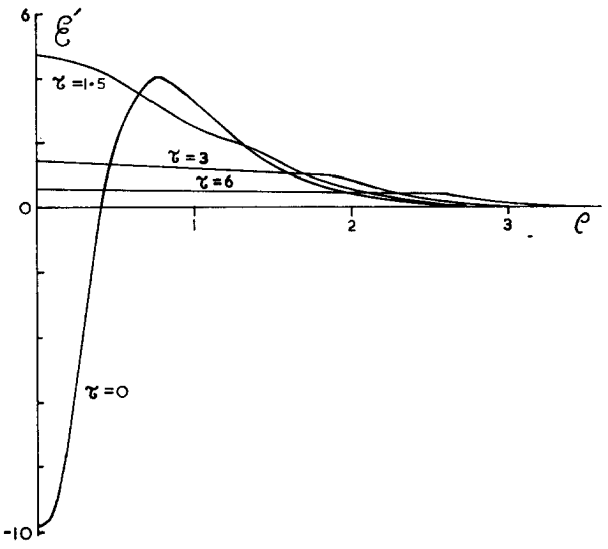


FIG. 12. Dissipative decay for $\omega' = 0.8$. The same form of decay as for the $\omega' = 0$ case is apparent, but the time of decay is longer. It is not until $\tau = 7$ that $\mathcal{E}'$ is not visibly different from zero.

which we take as being typical. The disturbances are applied by a random-number generator similar to that used for the $\omega' = 0$ case. The singular decay of Fig. 11 is very similar to those shown for the $\omega' = 0$ case, but the decay time is longer. Figure 13, which shows $\mathcal{E}'_{\min}\nu\tau$ for this disturbance, closely resembles Fig. 7 for the $\omega' = 0$ case, but the value of τ at which $\mathcal{E}'$ goes singular is different (2.8 compared with 1.5). Figure 12 shows a dissipative decay. Again note the close resemblance to Fig. 6, but this time it is not

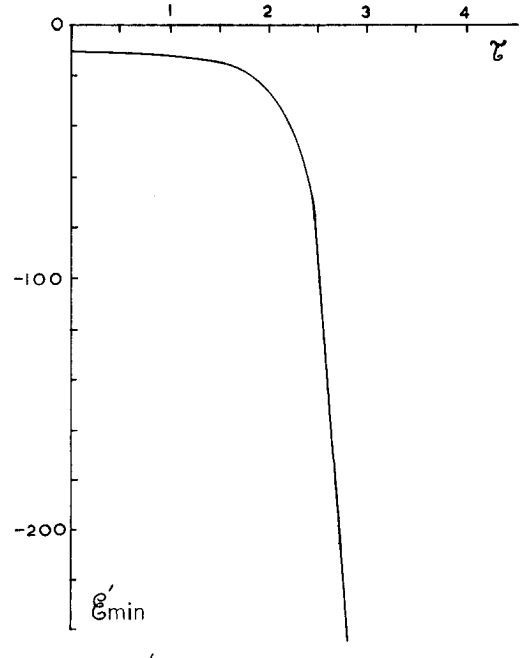


FIG. 13. $\mathcal{E}'_{\min}\nu\tau$ for the singular decay of Fig. 11.

until $\tau = 7$ that $\mathcal{E}'$ is not visibly different from 0, compared with $\tau = 3$ for Fig. 6.

Conclusion: Both time-dependent and time-independent solutions of the form (6) are unstable for the field investigated.

5. ASSIGNMENT OF PARAMETERS

Energy, charge, lifetime, and size are four of the physical quantities which can be given some significance in terms of this theory. The energy and charge are respectively,

$$E = \frac{2\kappa}{\mu^2(1-\omega'^2)^{\frac{1}{2}}} I, \quad Q = \frac{2e\omega' I}{\hbar c\mu^2(1-\omega'^2)^{\frac{1}{2}}}.$$

For a neutral particle, $\omega' = 0$ and E reduces to $2\kappa\mu^{-2}I$. The values of I for the three lowest-order particlelike solutions are 1.5, 9.6, and 29, so that the theory predicts neutral particles with masses in the ratios $1.5:9.6:29:\cdots$. We consider only the particle of lowest energy. The lifetime in the $\omega' = 0$ case is about $2(\kappa c)^{-1}$ sec if one takes the estimates of Sec. 4 and about $(4\kappa c)^{-1}$ sec if one takes the perturbation estimates of Sec. 3. From Fig. 3 the size of the particle R is of the order of 2 units of r'. This means that the ratio (lifetime/size) is of the order of $1/c$, a constant independent of the field parameters κ and μ. The value of this ratio ($\approx 10^{-9}$ sec/cm) is far removed from that of the metastable mesons, but of the correct order of magnitude for the highly unstable mesons if we assume the size of such particles to be of the order of 1 Fermi. In the time-dependent case, a good approximation to the values of Ω_i' is given by Barston's limiting curve (see Appendix A) for which

$$\Omega_i' \propto (1-\omega'^2)^{\frac{1}{2}}.$$

The size of the particle R is about 2 units of r', but since $r' = r(1-\omega'^2)^{+\frac{1}{2}}$, then R is proportional to $(1-\omega'^2)^{-\frac{1}{2}}$. Thus, the lifetime/size ratio remains approximately $1/c$, independent of ω'.

So far no attempt has been made to impose charge quantization. If we postulate $Q = \pm e$, 0, then we obtain the relation

$$\frac{2\omega' I}{\hbar c\mu^2(1-\omega'^2)^{\frac{1}{2}}} = \pm 1, 0,$$

which determines ω' to be

$$\omega' = \pm[1 + (2I/\hbar c\mu^2)^2]^{-\frac{1}{2}} \quad \text{or} \quad 0,$$

yielding

$$E = [\hbar c\kappa)^2 + (2\kappa I/\mu^2)^2]^{\frac{1}{2}} \quad \text{or} \quad (2\kappa I/\mu^2).$$

The mass of the charged particle is thus always greater than that of the neutral particle, but approaches the neutral mass in the limit $(2\kappa I/\mu^2)^2 \gg (\hbar c\kappa)^2$. This is consistent with experiment, when it is found that the mass of the charged particles in an isotriplet is greater than the neutral member, characteristically by a few per cent.

Although the field chosen is too simple to be considered as a serious model of a meson, it does exhibit some attractive features, and at least does not discourage further study of such fields.

ACKNOWLEDGMENT

D. L. T. Anderson would like to express his thanks to the Carnegie Trust for the Universities of Scotland for a research scholarship.

APPENDIX A: BOUNDS ON AND NONEXISTENCE OF EIGENVALUES TO (13)

Equation (13) can be recast into the form (14), (15), and (16), the latter equation being of a form examined by Barston.[6] Applying his results, we find that, provided the imaginary part of ω_B is not zero,

$$|\omega_B|^2 \leq -[\text{lowest eigenvalue of } H]$$
$$= -[\text{lowest eigenvalue of } (-\nabla'^2 + 1 - 3\varphi_0'^2)],$$
$$[\text{Re}\,\omega_B]^2 = (\xi, B\xi)^2/(\xi,\xi)^2 \leq (\xi,\xi)(B\xi, B\xi)/(\xi,\xi)^2$$
$$= (B\xi, B\xi)/(\xi,\xi) = \omega'^2/(1-\omega'^2).$$

The latter yields immediately

$$(\Omega_r')^2 \leq \omega'^2. \tag{A1}$$

Let us now consider spherically symmetric solutions. Then

$$|\omega_B|^2 \leq -\text{lowest eigenvalue of the operator}$$
$$\left(-\frac{d^2}{dr'^2} - \frac{2}{r'}\frac{d}{dr'} + 1 - 3\varphi_0'^2\right).$$

This operator was numerically found to have a lowest eigenvalue -15.6. Thus,

$$(\Omega_r'^2 + \Omega_i'^2)/(1-\omega'^2) \leq 15.6. \tag{A2}$$

Because there is no loss in generality in taking Ω_r', Ω_i' both positive, (A1) and (A2) bound the real and imaginary parts of the eigenvalue Ω' as follows:

$$\Omega_i' \leq 3.95(1-\omega'^2)^{\frac{1}{2}}, \quad \Omega_r' \leq 3.95(1-\omega'^2)^{\frac{1}{2}},$$
$$\Omega_r' \leq \omega', \quad \text{provided } \Omega_i' \neq 0.$$

We can thus restrict the Ω_i' space to that lying between the axes and the dotted curve in Fig. 4, the equation of this curve being given by $\bar{\Omega}_i' = 3.95(1-\omega'^2)^{\frac{1}{2}}$.

We now want to examine nonspherically symmetric solutions to (13). The most general solution to (13) is an expansion of η, χ in terms of the spherical harmonics Y_l^m. If we so expand η, χ, then there is no coupling between different l values, the "radial" wavefunctions $\eta_l(r')$ and $\chi_l(r')$ being determined by $l = 0, 1, 2, \cdots$:

$$\left[\frac{d^2}{dr'^2} + \frac{2}{r'}\frac{d}{dr'} - \frac{l(l+1)}{r'^2} + \frac{(\Omega' + \omega')^2 - 1}{(1-\omega'^2)} + 2\varphi_0'^2\right]\eta_l = -\varphi_0'^2\chi_l,$$

$$\left[\frac{d^2}{dr'^2} + \frac{2}{r'}\frac{d}{dr'} - \frac{l(l+1)}{r'^2} + \frac{(\Omega' - \omega')^2 - 1}{(1-\omega'^2)} + 2\varphi_0'^2\right]\chi_l = -\varphi_0'^2\eta_l, \quad \text{(A3)}$$

$$\eta_l, \chi_l \to 0 \quad \text{as} \quad r' \to \infty,$$
$$\eta_l, \chi_l \propto r'^l \quad \text{as} \quad r' \to 0.$$

Barston's inequality now becomes

$$\frac{\Omega_{il}'^2 + \Omega_{rl}'^2}{(1-\omega'^2)} \leq -\left[\text{lowest eigenvalue of } \left(-\frac{d'^2}{dr'^2} - \frac{2}{r'}\frac{d}{dr'} + \frac{l(l+1)}{r'^2} + 1 - 3\varphi_0'^2\right)\right],$$

where the dependence of Ω_r', Ω_i' on l is indicated by Ω_{rl}', Ω_{il}'. We will show that the operator

$$\Theta \equiv \left(-\frac{d'^2}{dr'^2} - \frac{2}{r'}\frac{d}{dr'} + \frac{l(l+1)}{r'^2} + 1 - 3\varphi_0'^2\right)$$

can have (square integrable) eigenfunctions only for $l = 0$, $l = 1$. We have already discussed the $l = 0$ case when Θ has the lowest eigenvalue, -15.6. We want to solve $\Theta z = \Lambda z$ when l is not zero. Λ must be real since Θ is Hermitian, and there is no loss in generality in taking z real. Writing $y = r'z$, we must solve

$$\left(\frac{d^2}{dr'^2} - \frac{l(l+1)}{r'^2} - 1 + \Lambda + 3\varphi_0'^2\right)y = 0,$$
$$y = 0 \quad \text{at} \quad r' = 0,$$
$$y \to 0 \quad \text{as} \quad r' \to \infty. \quad \text{(A4)}$$

It will now be shown that (A4) can have solutions only for $l = 0$, 1. For large r', (A4) reduces to

$$\frac{d^2y}{dr'^2} - \gamma^2 y = 0, \quad \text{where} \quad \gamma^2 = 1 - \Lambda.$$

[It is necessary that $(1 - \Lambda)$ be positive for square-integrable solutions.] y has the asymptotic form $y = e^{-\gamma r'}$ and must also have the value zero at $r' = 0$.

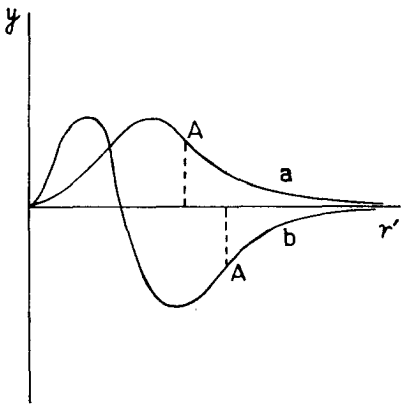


FIG. 14. The two simplest forms of solution to (A4), with no nodes and one node, curves a and b, respectively. There must be at least one point of inflexion (at A) at which the function is nonzero.

The simplest form of solution for y is shown in Fig. 14, curve a. This solution must have at least one point of inflexion at A at which $d^2y/dr'^2 = 0$, $y \neq 0$. Thus, we require

$$\left(-\frac{l(l+1)}{r'^2} - \gamma^2 + 3\varphi_0'^2\right) = 0 \quad \text{at } A.$$

This can be rewritten as

$$\gamma^2 = [3Z^2 - l(l+1)]/r'^2.$$

Now γ^2 must be ≥ 0. Hence, at a point A we must have

$$3Z^2 - l(l+1) \geq 0. \quad \text{(A5)}$$

A plot of Zvr' is shown in Fig. 2, from which it can be seen that Z has a maximum value of 1.2. The condition (A5) can, therefore, be satisfied for $l = 0$ or 1, but not for $l = 2$ or higher. Thus, a solution of the form of curve a of Fig. 14 cannot exist for $l > 1$. The next simplest form of solution (with one node) is shown in curve b of Fig. 14. It must have at least one point of inflexion satisfying the same conditions as for the nodeless solution. Similarly, for the next simplest (2 nodes), and so on. Thus, any solution for y must have at least one point of inflexion at which y is not zero. The condition for such a point is (A5) which can be satisfied only for $l = 0$, 1. Hence, (A4) can have no solutions for $l > 1$, implying that (13) can have no nontrivial solutions for $l > 1$. The only solution found to (13) for $l = 1$ was the solution $\Omega' = 0$, $\eta = \chi = \text{const} \times (\partial\varphi_0'/\partial r')Y_1^m$. The solutions for $l = 0$ have been discussed in the text.

* Present address: Commonwealth Meteorology Research Centre, P.O. Box 5089 AA, Melbourne.

[1] M. Born, Nature **132**, 282 (1933); M. Born and L. Infeld, *ibid.*, 1004 (1933); Proc. Roy. Soc. (London) **A143**, 410 (1934); **A144**, 425 (1934); **147**, 522 (1934); **150**, 141 (1935); M. Born, Proc. Indian Acad. Sci. **3**, 8, 85 (1936); J. Frenkel, Proc. Roy. Soc. (London) **A146**, 930 (1934); E. Feenberg, Phys. Rev. **47**, 148 (1935); M. H. L. Pryce, Proc. Roy. Soc. (London) **A155**, 597 (1936); B. Hoffman and L. Infeld, Phys. Rev. **51**, 765 (1937); N. Rosen, Phys. Rev. **55**, 94

(1939); A. C. Menuis and N. Rosen, *ibid.* **62**, 436 (1942); N. Rosen and H. Rosenstock, *ibid.* **85**, 257 (1952); H. Rosenstock, *ibid.* **93**, 331 (1954); H. Schiff, Proc. Roy. Soc. (London) **A269**, 277 (1962); G. Rosen, J. Math. Phys. **8**, 573 (1967); G. H. Derrick and K-K. Wan, *ibid.* **9**, 232 (1968); G. Pinski, *ibid.* **9**, 1323 (1968).

[2] R. H. Hobart, Proc. Phys. Soc. (London) **82**, 201 (1963); **85**, 610 (1965); G. H. Derrick, J. Math. Phys. **5**, 1252 (1964).

[3] G. Rosen, J. Math. Phys. **6**, 1269 (1965).

[4] No generality is gained by inserting a multiplicative constant into (2), since such a constant can be absorbed into ψ, with a corresponding redefinition of μ.

[5] R. Finkelstein, R. Le Levier, and M. Ruderman, Phys. Rev. **83**, 326 (1951); N. Rosen and H. Rosenstock, *ibid.*, **85**, 257 (1952); N. V. Mitskevich, Zh. Eksp. Teor. Fiz. **29**, 354 (1955) [Sov. Phys.—JETP **2**, 197 (1956)]; V. B. Glasko, F. Leriust, I. P. Terletskii, and S. F. Shushurin, *ibid.*, **35**, 452 (1958) [Sov. Phys.—JETP **8**, 312 (1959)]; D. D. Betts, H. Schiff, and W. B. Strickfaden, J. Math. Phys. **4**, 334 (1963); Z. Nehari, Proc. Roy. Irish Acad. **62**, 117 (1963). A field similar to (1), but with μ^2 negative [leading to a change in sign of the cubic term in (9)], has been studied by the following authors: L. I. Schiff, Phys. Rev. **84**, 1 (1951); **92**, 766 (1953); A. Salam, *ibid.*, **82**, 217 (1951); B. J. Malenka, *ibid.*, **85**, 686(L) (1952); **86**, 68 (1952); R. O. Fornaguera, Nuovo Cimento **1**, 132 (1955); W. E. Thirring, Z. Naturforsch. **7a**, 63 (1952); and C. S. Morawetz, Proc. Roy. Soc. (London) **A306**, 291 (1968). However, Mitskevich showed that if μ^2 is negative,then (1) has no regular solutions of type (6), while Morawetz proved that any wavepacket of finite energy must necessarily dissipate in such a field. Hence, a positive value of μ^2 would seem to be appropriate, if one seeks a classical model of elementary particles. On the quantum approach, little has been done to the μ^2 positive case, though L. I. Schiff has studied the $\mu^2 - ve$ case.

[6] E. M. Barston, J. Math. Phys. **8**, 523 (1967).

[7] This result is derived in Appendix A.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Exact State and Fugacity Equations for the Ideal Quantum Gases

MICHAEL MARTIN NIETO
The Niels Bohr Institute, University of Copenhagen, Denmark

(Received 30 June 1969)

The fully relativistic expressions for the density, pressure, and energy density of the ideal quantum gases are solved exactly for the fugacity. This allows the exact, fully relativistic equations of state to be obtained for the ideal Fermi and Bose gases. All these expressions are studied in detail, including a discussion of physical implications and various limiting cases.

I. INTRODUCTION

The most familiar results of quantum-statistical mechanics are the expressions for the density n, pressure P, and energy density E of the ideal Fermi and Bose gases in terms of the chemical potential μ, or, equivalently, the fugacity

$$z = e^{\mu/kT}. \tag{1}$$

They are

$$n = \frac{g}{2\pi^2\hbar^3}\int_0^\infty \frac{p^2\,dp}{\exp[(\mathcal{E}-\mu)/kT] \pm 1} + \begin{Bmatrix} +0 \\ +V^{-1}z(1-z)^{-1} \end{Bmatrix}, \tag{2}$$

$$P = \frac{g}{6\pi^2\hbar^3}\int_0^\infty \frac{\partial \mathcal{E}}{\partial p}\frac{p^3\,dp}{\exp[(\mathcal{E}-\mu)/kT] \pm 1} + \begin{Bmatrix} +0 \\ -kTV^{-1}\ln(1-z) \end{Bmatrix}, \tag{3}$$

$$E = \frac{g}{2\pi^2\hbar^3}\int_0^\infty \frac{\mathcal{E}p^2\,dp}{\exp[(\mathcal{E}-\mu)/kT] \pm 1}, \tag{4}$$

where the upper (lower) sign is for Fermi (Bose) particles, g is the statistical weight ($2S+1$ for massive particles), and $\mathcal{E}$ is the kinetic energy of a particle:

$$\mathcal{E} = (p^2c^2 + m^2c^4)^{\frac{1}{2}} - mc^2. \tag{5}$$

For Bose particles, the last terms on the right-hand side in Eqs. (2) and (3) are important only if a finite fraction of the total number of particles (N) is in the $\mathbf{p} = 0$ state.[1] This would represent a Bose-Einstein condensation into the ground state.

Over many decades, it has been thought desirable to solve Eqs. (2)–(4) for z, and hence to be able to find exact equations of state for the ideal gases. Unfortunately, although equations of state were long ago found in limiting cases,[2] until recently no exact equations of state were ever obtained.

The key to the solution was found by Leonard,[3] who very cleverly succeeded in inverting the nonrelativistic limit of Eq. (2) with complex variable theory. Using a modification of Leonard's method, we recently inverted Eqs. (2)–(4) and, by combining the results, obtained the *exact* relativistic equations of state for the ideal Fermi and Bose gases for the first time.[4]

In this paper, we give a more detailed discussion of our earlier note.[4] In the next section, the Fermi equations (2)–(4) are inverted, yielding the exact

(1939); A. C. Menuis and N. Rosen, *ibid.* **62**, 436 (1942); N. Rosen and H. Rosenstock, *ibid.* **85**, 257 (1952); H. Rosenstock, *ibid.* **93**, 331 (1954); H. Schiff, Proc. Roy. Soc. (London) **A269**, 277 (1962); G. Rosen, J. Math. Phys. **8**, 573 (1967); G. H. Derrick and K-K. Wan, *ibid.* **9**, 232 (1968); G. Pinski, *ibid.* **9**, 1323 (1968).

[2] R. H. Hobart, Proc. Phys. Soc. (London) **82**, 201 (1963); **85**, 610 (1965); G. H. Derrick, J. Math. Phys. **5**, 1252 (1964).

[3] G. Rosen, J. Math. Phys. **6**, 1269 (1965).

[4] No generality is gained by inserting a multiplicative constant into (2), since such a constant can be absorbed into ψ, with a corresponding redefinition of μ.

[5] R. Finkelstein, R. Le Levier, and M. Ruderman, Phys. Rev. **83**, 326 (1951); N. Rosen and H. Rosenstock, *ibid.*, **85**, 257 (1952); N. V. Mitskevich, Zh. Eksp. Teor. Fiz. **29**, 354 (1955) [Sov. Phys.—JETP **2**, 197 (1956)]; V. B. Glasko, F. Leriust, I. P. Terletskii, and S. F. Shushurin, *ibid.*, **35**, 452 (1958) [Sov. Phys.—JETP **8**, 312 (1959)]; D. D. Betts, H. Schiff, and W. B. Strickfaden, J. Math. Phys. **4**, 334 (1963); Z. Nehari, Proc. Roy. Irish Acad. **62**, 117 (1963). A field similar to (1), but with μ^2 negative [leading to a change in sign of the cubic term in (9)], has been studied by the following authors: L. I. Schiff, Phys. Rev. **84**, 1 (1951); **92**, 766 (1953); A. Salam, *ibid.*, **82**, 217 (1951); B. J. Malenka, *ibid.*, **85**, 686(L) (1952); **86**, 68 (1952); R. O. Fornaguera, Nuovo Cimento **1**, 132 (1955); W. E. Thirring, Z. Naturforsch. **7a**, 63 (1952); and C. S. Morawetz, Proc. Roy. Soc. (London) **A306**, 291 (1968). However, Mitskevich showed that if μ^2 is negative,then (1) has no regular solutions of type (6), while Morawetz proved that any wavepacket of finite energy must necessarily dissipate in such a field. Hence, a positive value of μ^2 would seem to be appropriate, if one seeks a classical model of elementary particles. On the quantum approach, little has been done to the μ^2 positive case, though L. I. Schiff has studied the $\mu^2 - ve$ case.

[6] E. M. Barston, J. Math. Phys. **8**, 523 (1967).

[7] This result is derived in Appendix A.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Exact State and Fugacity Equations for the Ideal Quantum Gases

MICHAEL MARTIN NIETO
The Niels Bohr Institute, University of Copenhagen, Denmark

(Received 30 June 1969)

The fully relativistic expressions for the density, pressure, and energy density of the ideal quantum gases are solved exactly for the fugacity. This allows the exact, fully relativistic equations of state to be obtained for the ideal Fermi and Bose gases. All these expressions are studied in detail, including a discussion of physical implications and various limiting cases.

I. INTRODUCTION

The most familiar results of quantum-statistical mechanics are the expressions for the density n, pressure P, and energy density E of the ideal Fermi and Bose gases in terms of the chemical potential μ, or, equivalently, the fugacity

$$z = e^{\mu/kT}. \tag{1}$$

They are

$$n = \frac{g}{2\pi^2\hbar^3}\int_0^\infty \frac{p^2\,dp}{\exp[(\mathcal{E}-\mu)/kT] \pm 1} + \begin{Bmatrix} +0 \\ +V^{-1}z(1-z)^{-1} \end{Bmatrix}, \tag{2}$$

$$P = \frac{g}{6\pi^2\hbar^3}\int_0^\infty \frac{\partial \mathcal{E}}{\partial p}\frac{p^3\,dp}{\exp[(\mathcal{E}-\mu)/kT] \pm 1} + \begin{Bmatrix} +0 \\ -kTV^{-1}\ln(1-z) \end{Bmatrix}, \tag{3}$$

$$E = \frac{g}{2\pi^2\hbar^3}\int_0^\infty \frac{\mathcal{E}p^2\,dp}{\exp[(\mathcal{E}-\mu)/kT] \pm 1}, \tag{4}$$

where the upper (lower) sign is for Fermi (Bose) particles, g is the statistical weight ($2S+1$ for massive particles), and $\mathcal{E}$ is the kinetic energy of a particle:

$$\mathcal{E} = (p^2c^2 + m^2c^4)^{\frac{1}{2}} - mc^2. \tag{5}$$

For Bose particles, the last terms on the right-hand side in Eqs. (2) and (3) are important only if a finite fraction of the total number of particles (N) is in the $\mathbf{p} = 0$ state.[1] This would represent a Bose-Einstein condensation into the ground state.

Over many decades, it has been thought desirable to solve Eqs. (2)–(4) for z, and hence to be able to find exact equations of state for the ideal gases. Unfortunately, although equations of state were long ago found in limiting cases,[2] until recently no exact equations of state were ever obtained.

The key to the solution was found by Leonard,[3] who very cleverly succeeded in inverting the nonrelativistic limit of Eq. (2) with complex variable theory. Using a modification of Leonard's method, we recently inverted Eqs. (2)–(4) and, by combining the results, obtained the *exact* relativistic equations of state for the ideal Fermi and Bose gases for the first time.[4]

In this paper, we give a more detailed discussion of our earlier note.[4] In the next section, the Fermi equations (2)–(4) are inverted, yielding the exact

equation of state for Fermi particles. This is followed in Sec. III by a similar discussion of Bose particles. The fourth section presents a discussion that includes an investigation of various limiting properties.

II. IDEAL FERMI GAS

By changing variables to $y = \exp(\mathcal{E}/kT)$, the Fermi equation (1) can be written as

$$\begin{aligned} n &\equiv -YF_n(j) \\ &\equiv -jY\int_1^\infty \frac{dy(\ln y + M)[\ln y(\ln y + 2M)]^{\frac{1}{2}}}{y(y-j)} \\ &\equiv -jY\int_1^\infty \frac{dyf_n(y)}{y(y-j)}, \end{aligned} \tag{6}$$

where

$$g\Lambda^{-3} \equiv Y \equiv \frac{g}{2\pi^2}\left(\frac{kT}{\hbar c}\right)^3, \quad M \equiv \left(\frac{mc^2}{kT}\right), \quad j \equiv -z. \tag{7}$$

"Λ" is what we will call the *optical wavelength*, since it is equal to the wavelength of a photon with energy $(4\pi)^{\frac{1}{2}}kT$. Note also that M is the ratio of the rest-mass energy to the thermal energy kT.

In addition to the functions $f_n(t)$ and $F_n(t)$ defined in (6), we also define

$$f_P(t) = [\ln t(\ln t + 2M)]^{\frac{1}{2}}, \quad F_P(t) = t\int_1^\infty \frac{dyf_P(y)}{y(y-t)}, \tag{8}$$

$$\begin{aligned} f_E(t) &= (\ln t)^{\frac{1}{2}}(\ln t + M)(\ln t + 2M)^{\frac{1}{2}}, \\ F_E(t) &= t\int_1^\infty \frac{dyf_E(y)}{y(y-t)}. \end{aligned} \tag{9}$$

Note that, as t goes from 1 to ∞, $f_n(t)$, $f_P(t)$, and $f_E(t)$ are monotonically increasing functions that go from 0 to ∞, and the principal values of $F_n(t)$, $F_P(t)$, and $F_E(t)$ are monotonically decreasing functions that go from a positive constant to $-\infty$.

Now consider the function $\mathcal{N}$ of a complex variable s, defined by

$$\mathcal{N}(s) = 1 + YF_n(s)/n. \tag{10}$$

$\mathcal{N}(s)$ has a cut on the real axis $[1, \infty)$. Just by the definition of $F_n(s)$ in (6), it also has a zero at $s = j$. This is the only zero or pole of $\mathcal{N}$ since n is a real monotonic function of a real variable j. (This point can also be seen more rigorously by using the principle of the argument,[5] as Leonard did and which we do in Appendix A.)

By using[6]

$$\frac{1}{x - a \pm i\epsilon} = \mathbf{P}\left\{\frac{1}{x-a}\right\} \mp i\pi\delta(a), \tag{11}$$

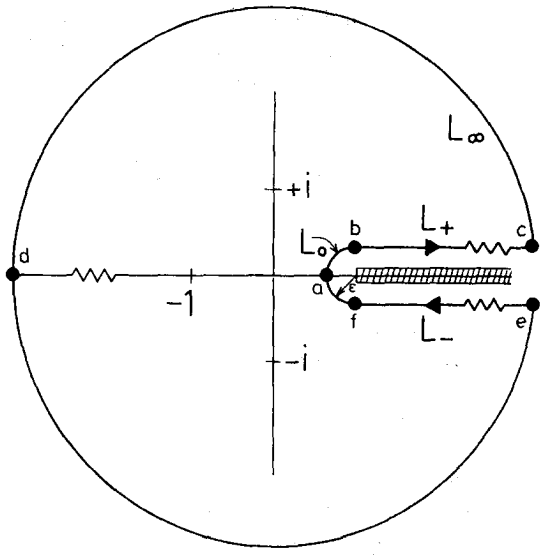


FIG. 1. The complex s plane, with the cut $[1, \infty)$. The contour L is divided into four parts L_∞, L_-, L_0, and L_+. L_0 travels from the point f through the point a to point b, on a half-circle of radius ϵ centered at $s = 1$.

we see that the limiting values of $\mathcal{N}$ above and below the cut (see Fig. 1) are[7]

$$\mathcal{N}^\pm(s) = 1 + (Y/n)\mathbf{P}\{F_n(s)\} \pm i(\pi/n)Yf_n(s), \tag{12}$$

so that

$$\mathcal{N}^+(s)/\mathcal{N}^-(s) = \exp[i2\theta_n(s)], \tag{13}$$

$$\theta_n(s) = \tan^{-1}\left(\frac{\pi Yf_n(s)/n}{YF_n(s)/n + 1}\right). \tag{14}$$

As s goes from 1 to ∞, the argument of $\tan^{-1}$ in (14) goes from 0 to $-\infty$, is discontinuous to $+\infty$, and then goes to 0. By taking the branch of $\tan^{-1}$ such that $\theta_n(\infty) = 0$, we have

$$\begin{aligned} \lim_{s\to 1+} \theta_n(s) &= -\pi_+, \\ \lim_{s\to\infty-} \theta_n(s) &= 0_-. \end{aligned} \tag{15}$$

This means that $\arg[\mathcal{N}(s)]$ increases by π as $\mathcal{N}$ traverses L_+, which agrees with Appendix A.

By now defining

$$\mathcal{N}_0(s) = \mathcal{N}(s)/(s-j), \tag{16}$$

we have a function which is analytic everywhere except on the cut and which also satisfies the same limiting ratio (13). Finding the solution for $j = -z$ of a function of the form of (16) is the Hilbert problem.[8] This solution is obtained by first realizing that the use of Eq. (11) implies that the function

$$G(s) = \exp\Gamma(s) = \exp\left(\frac{s}{\pi}\int_1^\infty \frac{dt\theta_n(t)}{t(t-s)}\right) \tag{17}$$

satisfies the ratio condition (13) on the cut. However, $\Gamma(s)$ has a logarithmic singularity at $s = 1$, which can be exhibited explicitly by writing

$$\Gamma(s) = \frac{s}{\pi}\int_1^\infty \frac{[\theta_n(t)+\pi]\,dt}{t(t-s)} - \frac{s}{\pi}\int_1^\infty \frac{\pi\,dt}{t(t-s)}$$
$$= \frac{s}{\pi}\int_1^\infty \frac{[\theta_n(t)+\pi]\,dt}{t(t-s)} + \ln(1-s). \quad (18)$$

This means $G(s)$ has a zero at $s = 1$ and has no other zeros or poles. Thus, we can write

$$\mathcal{N}(s)/(s-j) = KG(s)/(1-s), \quad (19)$$

where K is a constant. This is so because both sides of (19) satisfy the ratio condition (13) on the cut and have no other singularities.

By evaluating both sides of (19) at $s = 0$, we get $K = -1/j = 1/z$. Then, taking the derivative with respect to s of (19) and evaluating at $s = 0$, we find that

$$\frac{YI_n}{n} \equiv \frac{Y}{n}\int_1^\infty \frac{dyf_n(y)}{y^2} = \frac{1}{z_F} + 1 + \frac{1}{\pi}\int_1^\infty \frac{dt\theta_n(t)}{t^2}. \quad (20)$$

The integral $I_n(M)$ on the left-hand side of (20) can be done exactly in terms of $K_n(M)$, the modified Bessel functions of the second kind.[9] (This is done in Appendix B.) Inserting Eq. (B13) of Appendix B into (20), we end with the result

$$\frac{1}{z_F} = \frac{Y}{n}M^2e^M[K_2(M)] - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_n(t)}{t^2}. \quad (21)$$

Similarly, we obtain

$$\frac{1}{z_F} = \frac{YkT}{P}M^2e^M[K_2(M)] - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_P(t)}{t^2}, \quad (22)$$

$$\frac{1}{z_F} = \frac{YkT}{E}M^2e^M[MK_1(M) + (3-M)K_2(M)] - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_E(t)}{t^2}, \quad (23)$$

$$\theta_P(t) = \tan^{-1}\left(\frac{\pi YkTf_P(t)/3P}{YkTF_P(t)/3P + 1}\right), \quad (24)$$

$$\theta_E(t) = \tan^{-1}\left(\frac{\pi YkTf_E(t)/E}{YkTF_E(t)/E + 1}\right). \quad (25)$$

The $\theta_{P,E}(t)$ have the same limits at $t = 1, \infty$ as $\theta_n(t)$.

Now, combining (21) and (22), we have the exact Fermi equation of state

$$YM^2e^MK_2(M)\left(\frac{1}{n} - \frac{kT}{P}\right) = \frac{1}{\pi}\int_1^\infty \frac{dt[\theta_n(t) - \theta_P(t)]}{t^2}. \quad (26a)$$

If one prefers, one can obtain the T-P-E Fermi equation of state by combining (22) and (23):

$$kTYM^2e^MK_2(M)\left[\frac{1}{P} - \frac{1}{E}\left((3-M) + \frac{MK_1(M)}{K_2(M)}\right)\right] = \frac{1}{\pi}\int_1^\infty \frac{dt[\theta_P(t) - \theta_E(t)]}{t^2}. \quad (26b)$$

III. IDEAL BOSE GAS

For the Bose case we start in a similar manner, except that, since we have Bose integrals[10] to begin with, we do not have to use the intermediate step of taking $j = -z$. To invert the Bose equation (2), we consider the function $\mathcal{N}_B(s)$ defined by

$$\mathcal{N}_B(s) = 1 - \frac{YF_n(s)}{n} - \frac{1}{nV}\frac{s}{1-s}. \quad (27)$$

As before, $\mathcal{N}_B(s)$ has a zero at $s = z$ and a branch cut on the real axis $[1, \infty)$. However, from the last term, this time $\mathcal{N}_B(s)$ also has a pole at $s = 1$, and so finding this solution presents a new complication. The limits of $\mathcal{N}_B(s)$ above and below the cut are[11]

$$\mathcal{N}_B^{\pm}(s) = 1 - \frac{Y}{n}\mathbf{P}\{F_n(s)\} - \frac{1}{nV}\frac{s}{1-s} \mp i\frac{\pi}{n}Yf_n(s), \quad (28)$$

so that

$$\mathcal{N}_B^+(s)/\mathcal{N}_B^-(s) = \exp[i2\Phi_n(s)], \quad (29)$$

$$\Phi_n(s) = \tan^{-1}\left(\frac{\pi Yf_n(s)/n}{YF_n(s)/n - (1/nV)s/(s-1) - 1}\right). \quad (30)$$

At $s = 1$ and $s = \infty$, the limits of Φ_n are

$$\Phi_n(1_+) = \Phi_n(\infty_-) = 0_-. \quad (31a)$$

For $1 < s < \infty$, Φ_n first decreases to some angle ϕ, $-\pi < \phi < 0$, and then returns to zero. Also, there is the limit

$$\lim_{\substack{V\to\infty \\ \delta\to 0}} \Phi_n(1+\delta) = -\pi_+, \quad (31b)$$

which means that Φ_n behaves similarly to θ_n when Bose–Einstein condensation is unimportant. (The detailed behavior of Φ_n is discussed in Appendix A.)

The solution to this problem is obtained by considering

$$\bar{\mathcal{N}}_0(s) = [(1-s)/(s-z)]\mathcal{N}_B(s) \tag{32}$$

and

$$K_B G_B(s) \equiv K_B \exp\left(\frac{s}{\pi}\int_1^\infty \frac{dt\Phi_n(t)}{t(t-s)}\right). \tag{33}$$

Since both Eqs. (32) and (33) satisfy the ratio condition (29) and have no singularities other than the cut,

$$[(1-s)/(s-z)]\mathcal{N}_B(s) = K_B G_B(s) \tag{34}$$

is our solution equation.

Note that the appearance of the Bose–Einstein pole $1/(1-s)$ in $\mathcal{N}_B(s)$ caused the disappearance of the $(1-s)$ zero in $G_B(s)$, but that in the limit $V \to \infty$ the situation would revert to the same singularity structure as in the Fermi case. In either event, the solution equation (34) is the same form as the Fermi solution equation (19).

We now can complete the solution as in the Fermi case. Matching both sides of Eq. (33) at $s = 0$ determines $K_B = 1/z$. Matching the derivatives of (33) at $s = 0$ and performing the necessary Bessel function integrations gives us the final solution

$$\frac{1}{z_B} = \frac{Y}{n} M^2 e^M [K_2(M)] + 1 + \frac{1}{nV} + \frac{1}{\pi}\int_1^\infty \frac{dt\Phi_n(t)}{t^2}. \tag{35}$$

Going now to the Bose equation (3), we see that there is also a complication here due to a condensation term. This condensation term does not imply an extra zero or pole, but rather a second cut. However, the branch cut of $\ln(1-z)$ is also $[1, \infty)$, so that, conveniently, this only implies a second contribution to our first cut.

The solution then proceeds as in the Fermi case. We consider

$$\mathfrak{F}(s) = 1 - \frac{YkTF_P(s)}{3P} + \frac{kT}{PV}\ln(1-s), \tag{36}$$

which has a single zero at $s = z$. The limits of $\mathfrak{F}(s)$ on the cut are

$$\mathfrak{F}^{\pm}(s) = 1 - \frac{YkT}{3P}\mathbf{P}\{F_P(s)\} \mp i\frac{\pi YkT}{3P} f_P(s) + \frac{kT}{PV}\ln|s-1| - i\frac{kT}{PV}\arg(s-1), \tag{37}$$

so that

$$\mathfrak{F}^+(s)/\mathfrak{F}^-(s) = \exp[i2\Phi_P(s)], \tag{38}$$

$$\Phi_P(s) = \tan^{-1}\left(\frac{\pi YkTf_P(s)/3P + kT\arg(s-1)/PV}{YkTF_P(s)/3P - 1 - kT\ln|s-1|/PV}\right). \tag{39}$$

Even with the condensation term, the limits of $\Phi_P(s)$ are the same as in the Fermi case (see Appendix A), so that

$$\lim_{s\to 1+}\Phi_P(s) = -\pi_+, \qquad \lim_{s\to\infty-}\Phi_P(s) = 0_-. \tag{40}$$

Due to the limit $\Phi_P(1) = -\pi$, our matching function $G_P(s)$ will have a zero at $s = 1$, as in the Fermi case. The solution obviously then proceeds in the same manner, yielding

$$\frac{1}{z_B} = \frac{YkT}{P} M^2 e^M [K_2(M)] + 1 + \frac{kT}{PV} + \frac{1}{\pi}\int_1^\infty \frac{dt\Phi_P(t)}{t^2}. \tag{41}$$

Finally, we observe that, since the Bose equation (4) has no condensation term, it is inverted the same as in the Fermi case (except, of course, for not using $j = -z$). The result is

$$\frac{1}{z_B} = \frac{YkT}{E} M^2 e^M [MK_1(M) + (3-M)K_2(M)] + 1 + \frac{1}{\pi}\int_1^\infty \frac{dt\Phi_E(t)}{t^2}, \tag{42}$$

$$\Phi_E(t) = \tan^{-1}\left(\frac{YkTf_E(t)/E}{YkTF_E(t)/E - 1}\right), \tag{43}$$

where the same limits exist on $\Phi_E(t)$:

$$\Phi_E(1_+) = -\pi_+, \quad \Phi_E(\infty_-) = 0_-. \tag{44}$$

Combining (35) and (41), we finally obtain the exact relativistic equation of state for an ideal Bose gas:

$$YM^2 e^M K_2(M)\left(\frac{1}{n} - \frac{kT}{P}\right) = -\left(\frac{1}{nV} - \frac{kT}{PV}\right) - \frac{1}{\pi}\int_1^\infty \frac{dt[\Phi_n(t) - \Phi_P(t)]}{t^2}. \tag{45a}$$

Again, if one prefers, the T-P-E Bose equation of state can be obtained, from Eqs. (41) and (42):

$$kTYM^2 e^M K_2(M)\left[\frac{1}{P} - \frac{1}{E}\left((3-M) + \frac{MK_1(M)}{K_2(M)}\right)\right] = -\frac{kT}{PV} - \frac{1}{\pi}\int_1^\infty \frac{dt[\Phi_P(t) - \Phi_E(t)]}{t^2}. \tag{45b}$$

IV. DISCUSSION

In this section we discuss some of the clearest physical implications and limiting cases of our results. With the equations we have obtained, these points are seen from a new and general viewpoint.

A. Effect of Statistics

Our solutions clearly show where the differences in behavior between Fermi and Bose gases manifest themselves. First, of course, there is the obvious effect caused by the Bose–Einstein condensation terms. But in addition to this, the solutions for the fugacity show that there are three other dissimilarities. Referring, for example, to Eqs. (23) and (42), the differences between the Fermi and Bose fugacities are found in three places on the right-hand sides: (a) the ± 1; (b) the $\mp \int d t \psi(t)/t^2$, where ψ is, respectively, θ_E or Φ_E; and (c) within the definitions of the above angles as being of the form $\tan^{-1} b(t)$, the functions $b(t)$ have a change of ± 1 in the denominator.

These three ($\pm$) signs and the condensation terms are the exact manifestations of the difference in fugacities (and the equations of state) of the ideal Fermi and Bose gases.

B. Ranges of Chemical Potential

One also trivially sees the ranges of chemical potential for the two gases. Obviously, from Eqs. (21) and (35) we have

$$\lim_{n\to 0} \frac{1}{z_F} = \lim_{n\to 0} \frac{1}{z_B} = \infty. \tag{46}$$

But, from (14) and (30),

$$\lim_{n\to\infty} \theta_n = -\pi, \quad \lim_{n\to\infty} \Phi_n = 0, \tag{47}$$

so that

$$\lim_{n\to\infty} \frac{1}{z_F} = 0, \quad \lim_{n\to\infty} \frac{1}{z_B} = 1. \tag{48}$$

Since n and z are monotonic functions of each other, this yields

$$0 \le z_F \le \infty, \quad 0 \le z_B \le 1, \tag{49}$$

$$-\infty \le \mu_F \le \infty, \quad -\infty \le \mu_B \le 0. \tag{50}$$

C. Perfect Gas Law

In the low-density limit, we easily obtain the perfect gas law (Boyle's law) for both Fermi and Bose gases. Referring to the equations of state (26a) and (45a), we see that the angle terms on the right are bounded by ± 1. When n becomes small, the first term on the left-hand side is large. Then, to first order, it is balanced by the second term on the left, so that

$$P = nkT \quad \text{(Boyle's law)}. \tag{51}$$

From this discussion we also see that

$$D_F = \frac{1}{\pi}\int_1^\infty \frac{dt(\theta_n - \theta_P)}{t^2},$$

$$D_B = -\frac{1}{\pi}\int_1^\infty \frac{dt(\Phi_n - \Phi_P)}{t^2} \tag{52}$$

are expressions which give exact measures of the deviations of the ideal quantum gases from the perfect-gas law.

D. Nonrelativistic and Relativistic Limits

We now look at the relationship between the pressure and the energy density in the nonrelativistic and relativistic limits. Using Appendix B and Eq. (21), we have for a Fermi gas in the nonrelativistic limit ($M \gg 1$)

$$\lim_{M\to\infty} \frac{1}{z_F} = \frac{gkT}{\lambda^3 P} - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_P^{NR}(t)}{t^2}, \tag{53}$$

$$\theta_P^{NR}(t) = \tan^{-1}\left(\frac{4\pi^{\frac{1}{2}} gkT(\ln t)^{\frac{3}{2}}/3P\lambda^3}{(4gkT/3\pi^{\frac{1}{2}}P\lambda^3)t\int_1^\infty dy(\ln y)^{\frac{3}{2}}/y(y-t)+1}\right), \tag{54}$$

$$\lambda \equiv (\pi\hbar^2/mkT)^{\frac{1}{2}}. \tag{55}$$

But by taking the same limit for the E equation (22), we get exactly the same result as if each time there were a P in (54) and in (55) we substituted

$$P = \tfrac{2}{3}E \quad \text{(nonrelativistic limit, } M \gg 1). \tag{56}$$

Thus, this is the relation of the pressure to the energy density of an ideal Fermi gas in the nonrelativistic limit.

The expression λ is interesting in itself. It is called the *thermal wavelength.*[12] λ is the same size as the de Broglie wavelength of a nonrelativistic particle with kinetic energy πkT. In the relativistic limit ($M \ll 1$), Eq. (21) becomes

$$\lim_{M\to 0} \frac{1}{z_F} = \frac{2YkT}{P} - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_P^{R}(t)}{t^2}, \tag{57}$$

$$\theta_P^{R}(t) = \tan^{-1}\left(\frac{\pi YkT(\ln t)^3/3P}{(YkT/3P)t\int_1^\infty dy(\ln y)^3/y(y-t)+1}\right). \tag{58}$$

Again we get the same form from Eq. (22) if

$$P = \tfrac{1}{3}E \quad \text{(relativistic limit, } M \ll 1\text{)}. \tag{59}$$

This is the relativistic pressure energy density relation. If we can ignore Bose–Einstein condensation effects, the Bose and Fermi equations are the same except for consistent changes in $(\pm)$ signs. It is obvious that then Eqs. (56) and (59) would hold for a Bose gas, also.

For completeness, we also give the Fermi-density limiting cases

$$\lim_{M\to\infty}\frac{1}{z_F} = \frac{g}{n}\frac{1}{\lambda^3} - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_n^{NR}(t)}{t^2}, \tag{60}$$

$$\theta_n^{NR}(t) = \tan^{-1}\left(\frac{2\pi^{\frac{1}{2}}g(\ln t)^{\frac{1}{2}}/n\lambda^3}{(2g/\pi^{\frac{1}{2}}n\lambda^3)t\int_1^\infty dy(\ln y)^{\frac{1}{2}}/y(y-t)+1}\right), \tag{61}$$

which is Leonard's result,[3] and

$$\lim_{M\to 0}\frac{1}{z_F} = \frac{2Y}{n} - 1 - \frac{1}{\pi}\int_1^\infty \frac{dt\theta_n^R(t)}{t^2}, \tag{62}$$

$$\theta_n^R(t) = \tan^{-1}\left(\frac{\pi Y(\ln t)^2/n}{(Yt/n)\int_1^\infty dy(\ln y)^2/y(y-t)+1}\right). \tag{63}$$

All of the above limits can trivially be converted to the Bose limits by changing the three appropriate $(\pm)$ signs and, if desired, inserting the condensation terms.

E. Bose–Einstein Condensation

Following the discussion of Huang,[1] the Bose equation (2) can be written as

$$\frac{1}{V}\frac{z}{1-z} \equiv \frac{\langle n_0\rangle}{V} = n - YF_n(z). \tag{64}$$

The condition for the existence of Bose condensation can be stated as being that the left-hand side of Eq. (64) is greater than zero, or alternatively,

$$n > YF_n(1). \tag{65}$$

$F_n(1)$ is evaluated in Appendix B, yielding the condensation condition

$$n > n_c \equiv YM^3\sum_{t=1}^\infty e^{tM}\frac{K_2(tM)}{tM} \equiv g\Lambda_m^{-3}\sum_{t=1}^\infty e^{tM}\frac{K_2(tM)}{tM}, \tag{66}$$

$$\Lambda_m = hc/(4\pi)^{\frac{1}{3}}mc^2. \tag{67}$$

We call Λ_m the *mass wavelength*, it being the wavelength of a photon with energy $(4\pi)^{\frac{1}{3}}mc^2$. The limiting cases of condition (66) are (see Appendix B)

$$\lim_{M\to\infty} n_c = \frac{\zeta(\frac{3}{2})g}{\lambda^3} = \frac{(2.612)g}{\lambda^3}, \tag{68}$$

$$\lim_{M\to 0} n_c = 2Y\zeta(3) = 2g\Lambda^{-3}\zeta(3) = (2.404)g\Lambda^{-3}. \tag{69}$$

The solution of (66) for T, of course, implies that for a given density n, there exists a critical temperature T_c, below which condensation occurs.

We know from (64) that, above the critical density, the fraction of particles in the ground state is $(N \equiv nV)$

$$\langle n_0\rangle/N = 1 - n_c/n, \tag{70}$$

since $z = 1$.

As we near the critical point, it is of interest to have an expression for the fugacity. This can be obtained[3] by putting (27) in the left-hand side of (34) and taking the limit $s \to 1$. With a little algebra, and using the value of $K_B = -1/z$, one obtains

$$\frac{1}{z_B} = 1 + \frac{1}{nV}\exp\left(-\frac{1}{\pi}\int_1^\infty \frac{\Phi_n(t)\,dt}{t(t-1)}\right). \tag{71}$$

The largest contribution to the integral is possible when $t \approx 1$, at which point

$$\Phi_n(t \to 1) \approx \tan^{-1}\left(\frac{2\pi^{\frac{1}{2}}g(\ln t)^{\frac{1}{2}}/n\lambda^3}{(n_c/n) - (1/nV)/(t-1) - 1}\right). \tag{72}$$

Now consider the case $n < n_c$, but $n \approx n_c$. Since the numerator in (72) is logarithmically small, the condition for Φ_n to be $-\pi$ and not 0 is that the denominator be $\geqslant 0$. In other words,

$$\lim_{t\to 1}(t-1) \geq \left[nV\left(\frac{n_c}{n} - 1\right)\right]^{-1}. \tag{73}$$

At $\Phi_n = -\pi$, putting this into the integral in (71) produces a logarithmic divergence in (nV) as $N \to \infty$. However, it is canceled by the other (nV) in (70) and leaves

$$\lim_{z\to 1}\frac{1}{z_B} \simeq 1 + C\left(\frac{n_c}{n} - 1\right), \quad n \lesssim n_c, \tag{74}$$

where C is a constant. Leonard[3] has considered other limiting functional forms near the critical point, and the interested reader is referred there.

V. CONCLUSION

By inverting the expressions for the density, pressure, and energy density, we have obtained exact, fully relativistic expressions for the fugacities

of the ideal quantum gases. This allowed us to find the exact equations of state for the ideal Fermi and Bose gases. All these expressions were studied for physical content, and various cases, such as the nonrelativistic and relativistic limits, were discussed.

Because our solutions involve double integrals (the angles themselves are in terms of an integral), the solutions we have found are not necessarily always easy to use in computer calculations. Be that as it may, the value of our results is, first, that they are *exact* solutions for the fugacity and *exact* equations of state and, second, that they can test the exactness of approximations in particular cases, especially since the inner integrals in the angles are all well-known Bose integrals.[10] Furthermore, these exact equations clearly unify and show the common origin of special limiting cases, and provide a new perspective from which to understand the physics involved.

APPENDIX A

1. The Principle of the Argument and the Angle θ_n

In this appendix we will first use the principle of the argument[5] to verify that $\mathcal{N}(s)$ has only one pole. The principle of the argument states that, as a function goes around a closed contour, the argument of the function increases by 2π times the number of zeros minus the number of poles. To apply this principle to $\mathcal{N}(s)$, we use the contour L in Fig. 1.

First consider L_∞. The dominant term in $\mathcal{N}(s)$ as $|s| \to \infty$ is the integral $YF_n(s)/n$. By changing variables to $x = \ln(y)$ and replacing s by $x_0 = \ln(-s)$ (this last step is analytic on the contour L_∞), we have

$$\mathcal{N}(s)\big|_{L_\infty} = -\frac{Y}{n}\int_0^\infty \frac{dx(x+M)[x(x+2M)]^{\frac{1}{2}}}{\exp(x-x_0)+1}. \quad \text{(A1)}$$

However, for $|s| \to \infty$,

$$\lim_{|s|\to\infty} x_0 = \ln|s| + i\arg(-s) \simeq \ln|s|. \quad \text{(A2)}$$

Then the integral in (A2) can be evaluated by the Sommerfeld technique,[13] which states that, for x_0 large,

$$I = \int_0^\infty \frac{dg(x)}{dx}\frac{dx}{\exp(x-x_0)+1} \simeq g(x_0). \quad \text{(A3)}$$

In our case,

$$g(x) = \tfrac{1}{8}[x(x+2M)]^2, \quad \text{(A4)}$$

so that

$$\mathcal{N}(s)\big|_{L_\infty} \simeq -(Y/3n)[\ln s(\ln s + 2M)]^{\frac{3}{2}} \quad \text{(A5)}$$

$$\to -(Y/3n)[\ln|s|\,(\ln|s| + 2M)]^{\frac{3}{2}}. \quad \text{(A6)}$$

Equation (A6) means that $\arg[\mathcal{N}(s)]$ does not increase as $\mathcal{N}$ traverses L_∞.

Next, consider $\mathcal{N}(s)$ as it traverses L_+ from point b to point c. Since $F_n(s)$ goes from $F_n(1) > 0$ to $-\infty$, and $f_n(s)$ goes from 0 to a positive logarithmic infinity such that

$$\lim_{s\to\infty}\left(\frac{f_n(s)}{F_n(s)}\right) = 0, \quad \text{(A7)}$$

we have that $\arg[\mathcal{N}^+(s)] = \theta_n(s)$ increases by π on L_+. Taking the branch cut of the tan as we did in Eq. (14), this means

$$\theta_n(1) = -\pi, \quad \theta_n(\infty) = 0. \quad \text{(A8)}$$

Similarly, one sees that as $\mathcal{N}(s)$ traverses L_- from point e to point f, $\arg[\mathcal{N}(s)]$ increases by π again. {With our choice of the branch of tan, $\arg[\mathcal{N}(s)]$ goes from -2π to $-\pi$ on L_-.} On L_0, in the limit $\epsilon \to 0$, $\mathrm{Re}\,[\mathcal{N}(s)]$ is positively infinite while $\mathrm{Im}\,[\mathcal{N}(s)]$ is finite, so that there is no increase in $\arg[\mathcal{N}(s)]$ on L_0. Thus, the total increase in $\arg[\mathcal{N}(s)]$ as $\mathcal{N}(s)$ traverses L is 2π, so there exists only the one pole, at $s = j = -z$.

2. The Angle Φ_n

From Eq. (30), the angle Φ_n is defined by[11]

$$\Phi_n(s) = \tan^{-1}\left(\frac{\pi Y f_n(s)/n}{YF_n(s)/n - (1/nV)s/(s-1) - 1}\right) \quad \text{(A9)}$$

$$= \arg[\mathcal{N}_B(s)]. \quad \text{(A10)}$$

As $s \to 1_+$, the condensation term's negative infinity dominates over the logarithmic positive infinity of F_n in the argument of $\tan^{-1}$. Thus,[11] $\Phi_n(1_+) = 0_-$. However, as $s \to \infty_-$, the terms $f_n(s)$ and $F_n(s)$ dominate the $\tan^{-1}$ argument in the same manner as in Eq. (A7), so that $\Phi_n(\infty_-) = 0_-$. For $1 < s < \infty$, the numerator of the tan increases from 0 to ∞, as s is going from 1 to ∞. At the same time, in the denominator, $F_n(s)$ is going from $F_n(1) > 0$ to $-\infty$ and the condensation term is going from $-\infty$ to 0. The denominator may or may not go through 0, depending on the size of V. In either case, the denominator eventually stops increasing and returns to $-\infty$, since the $F_n(s)$ term eventually dominates.

This can all be summed up by saying that, for $1 < s < \infty$, $\Phi_n(s)$ first decreases to some angle ϕ ($-\frac{1}{2}\pi < \phi < 0$ if the numerator of the argument of $\tan^{-1}$ does not go through 0, $-\pi < \phi \leq -\frac{1}{2}\pi$ if it does), and then as s increases further, $\Phi_n(s)$ turns around and returns to 0. However, as V becomes larger, the $F_n(s)$ term dominates the numerator of

the tan for smaller and smaller s. The larger V becomes, the more quickly (δ smaller) and closer $\Phi_n(1+\delta)$ approaches $-\pi$. This means that, when Bose–Einstein condensation becomes unimportant ($V \to \infty$), the limiting behavior of $\Phi_n(s)$ is similar to $\theta_n(s)$. In fact,

$$\lim_{\substack{V\to\infty \\ \delta\to 0}} \Phi_n(1+\delta) = -\pi_+. \tag{A11}$$

This last step is true since the $F_n(s)$ term becomes $[n(1)/n(z<1)] > 1$ by the physical values of z_B (see Sec. IVB).

3. The Angle Φ_P

From Eq. (39) we have

$$\Phi_P(s) = \tan^{-1}\left(\frac{\pi YkTf_P(s)/3P + kT \arg(s-1)/PV}{YkTF_P(s)/3P - 1 - kT \ln|s-1|/PV}\right). \tag{A12}$$

For V finite, at $s=1$, the condensation terms dominate both the numerator and the denominator of $\tan^{-1}$. The denominator term is a positive infinity, which implies that, as s goes from 1 to a positive number, the argument goes from 0 to a positive number. However, because of the limiting properties of the f and F functions $\{f_P \to 0$, and the F_P term is $[P(z=1)/P(z<1)] > 1\}$, this is true even when $V \to \infty$, and the F_P term dominates.

As s increases, the denominator of the argument goes through zero, as before, and then heads to a negative infinity. Therefore, Φ_P is going from $-\pi$ to 0. Thus, with or without the condensation term, the limits on $\Phi_P(s)$ are the same as in the Fermi case, $\theta_P(s)$:

$$\lim_{s\to 1+} \Phi_P(s) = -\pi_+,$$

$$\lim_{s\to\infty-} \Phi_P(s) = 0_-. \tag{A13}$$

APPENDIX B

1. Modified Bessel Function of the Second Kind

The modified Bessel function of the second kind can be defined by[9]

$$K_\nu(z) = \tfrac{1}{2}\pi[I_{-\nu}(z) - I_\nu(z)]/\sin\nu\pi, \tag{B1}$$

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{(\frac{1}{2}z)^{\nu+2m}}{m!\,\Gamma(\nu+m+1)}. \tag{B2}$$

They satisfy the following relations[14]:

$$K_{\nu-1}(z) - K_{\nu+1}(z) = -(2\nu/z)K_\nu(z), \tag{B3}$$

$$K_{\nu-1}(z) + K_{\nu+1}(z) = -2K'_\nu(z), \tag{B4}$$

$$zK'_\nu(z) + \nu K_\nu(z) = -zK_{\nu-1}(z), \tag{B5}$$

$$zK'_\nu(z) - \nu K_\nu(z) = -zK_{\nu+1}(z). \tag{B6}$$

For ν an integer, they have the following limiting properties as z approaches 0 and $+\infty$ on the real axis[15]:

$$\lim_{M\to\infty} K_n(M) = \left(\frac{\pi}{2M}\right)^{\frac{1}{2}} e^{-M} \times \left(1 + \frac{4n^2-1^2}{(1!)8M} + \frac{(4n^2-1^2)(4n^2-3^2)}{(2!)(8M)^2} + \cdots\right), \tag{B7}$$

$$\lim_{M\to 0} K_n(M) = \frac{(n-1)!}{2}\left(\frac{2}{M}\right)^n, \quad n \geq 1, \tag{B8a}$$

$$= \ln\left(\frac{2}{M}\right), \quad n = 0. \tag{B8b}$$

2. The Integrals $I_n(M)$, $I_P(M)$, and $I_E(M)$

To obtain the solutions in the final form, it is necessary to evaluate the three integrals

$$I_{n,P,E}(M) \equiv \int_1^\infty \frac{dy f_{n,P,E}(y)}{y^2}. \tag{B9}$$

Using

$$x = (\ln y)/M + 1, \tag{B10}$$

I_n can be written as

$$I_n = M^3 e^M \int_1^\infty dx e^{-xM}(x^2-1)^{\frac{1}{2}}x = -M^3 e^M \frac{\partial}{\partial M}\left(\int_1^\infty dx e^{-xM}(x^2-1)\right), \tag{B11}$$

which can be integrated to give[16]

$$I_n = -M^3 e^M \frac{\partial}{\partial M}[M^{-1}K_1(M)]. \tag{B12}$$

Now use of (B6) yields

$$I_n = M^2 e^M [K_2(M)]. \tag{B13}$$

Similarly, letting $x = (\ln y)/M$, we obtain[17]

$$I_P = M^4 \int_0^\infty dx e^{-xM}[x(x+2)]^{\frac{3}{2}} = M^4[3M^{-2}e^M K_2(M)] = 3M^2 e^M [K_2(M)]. \tag{B14}$$

Finally, using (B10),

$$I_E = M^4 e^M \int_1^\infty dx e^{-Mx}(x^2-1)^{\frac{1}{2}}(x^2-x)$$
$$= M^4 e^M \left(\frac{\partial^2}{\partial M^2} + \frac{\partial}{\partial M}\right)\int_1^\infty dx e^{-Mx}(x^2-1)^{\frac{1}{2}}$$
$$= M^4 e^M \left(\frac{\partial^2}{\partial M^2} + \frac{\partial}{\partial M}\right)[M^{-1}K_1(M)]. \quad \text{(B15)}$$

With the help of the recursion relations above, this gives

$$I_E = M^2 e^M [MK_1(M) + (3-M)K_2(M)]. \quad \text{(B16)}$$

From the limits of the K_n given above, we also have

$$\lim_{M\to 0} I_n = \tfrac{1}{3}\lim_{M\to 0} I_P = \tfrac{1}{3}\lim_{M\to 0} I_E = 2, \quad \text{(B17)}$$

$$\lim_{M\to\infty} I_n = \tfrac{1}{3}\lim_{M\to\infty} I_P = \tfrac{2}{3}\lim_{M\to\infty} I_E = M^{\frac{3}{2}}(\tfrac{1}{2}\pi)^{\frac{1}{2}}. \quad \text{(B18)}$$

3. The Integrals $F_n(1)$ and $F_P(1)$

The integral $F_n(1)$ is necessary to determine the Bose–Einstein condensation condition. Starting from the definition of F_n in Eq. (6), $F_n(1)$ can be evaluated by changing variables to $x_0 = \ln(y)$, writing

$$\frac{1}{e^{x_0}-1} = \sum_{t=1}^\infty e^{-tx_0}, \quad \text{(B19)}$$

taking the summation sign outside the integral and changing variables again to $x = (x_0/M + 1)$. One then has

$$F_n(1) = M^3 \sum_{t=1}^\infty e^{tM}\int_1^\infty dx e^{-tMx}(x^2-1)^{\frac{1}{2}}x. \quad \text{(B20)}$$

But this is the same type of integral as in (B11), and use of the same evaluation procedure yields

$$F_n(1) = M^3 \sum_{t=1}^\infty e^{tM}\frac{K_2(tM)}{tM}. \quad \text{(B21)}$$

The limiting cases are

$$\lim_{M\to\infty} F_n(1) = M^{\frac{3}{2}}(\tfrac{1}{2}\pi)^{\frac{1}{2}}\zeta(\tfrac{3}{2}) = (2.612)(\tfrac{1}{2}\pi)^{\frac{1}{2}}M^{\frac{3}{2}}, \quad \text{(B22)}$$

$$\lim_{M\to 0} F_n(1) = 2\zeta(3) = 2(1.202), \quad \text{(B23)}$$

where $\zeta(d)$ is the Riemann ζ function:

$$\zeta(d) = \sum_{t=1}^\infty t^{-d}. \quad \text{(B24)}$$

In the same manner,

$$F_P(1) = 3M^4 \sum_{t=1}^\infty e^{tM}\frac{K_2(tM)}{(tM)^2}, \quad \text{(B25)}$$

$$\lim_{M\to\infty} F_P(1) = 3\zeta(\tfrac{5}{2})(\tfrac{1}{2}\pi)^{\frac{1}{2}}M^{\frac{3}{2}} = 3(1.341)(\tfrac{1}{2}\pi)^{\frac{1}{2}}M^{\frac{3}{2}}, \quad \text{(B26)}$$

$$\lim_{M\to 0} F_P(1) = 6\zeta(4) = \tfrac{1}{15}\pi^4. \quad \text{(B27)}$$

[1] K. Huang, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1963), Secs. 9.6 and 12.3.

[2] See, for example, the discussion in H. Y. Chiu, *Stellar Physics* (Blaisdell Publ. Co., Waltham, Mass., 1968), Vol. I, Chap. 3.

[3] A. Leonard, Phys. Rev. **175**, 221 (1968).

[4] M. M. Nieto, Nuovo Cimento Letters **1**, 677 (1969).

[5] E. T. Copson, *Theory of Functions of a Complex Variable* (Oxford University Press, London, 1935), p. 119.

[6] A. Messiah, *Quantum Mechanics* (North-Holland Publ. Co., Amsterdam, 1961), Vol. I, p. 469.

[7] P, of course signifies "principal value." In what follows we will often omit the P sign where it is obviously understood.

[8] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Ltd., Groningen, The Netherlands, 1953), Chap. 5.

[9] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1944), 2nd ed., Sec. 3.7.

[10] Actually, by defining our variables oppositely, all our integrals could have been Fermi-type integrals, but with the argument of the angles defined along the cut $(-\infty, -1]$.

[11] In fact, at this point both Eqs. (28) and (30) should have an additional term like $\lim_{\epsilon\to 0}$ (Re s)/[(Re $s-1 \mp i\epsilon$)nV]. To understand this, one must realize that on L_+, just at the point $s = 1 + i\epsilon$, there is a contribution to the angle Φ_n even for $\epsilon \to 0$. This is because, before $\epsilon \to 0$, Φ_n increases from $-3\pi/2$ at point f, to π at point a, to $-\pi/2$ at point b. The increase is caused by the condensation pole at $s = 1$, and even when $\epsilon \to 0$ it produces a step function in arg $[\mathcal{N}_E^+(s)]$ so that $\Phi_n(1) = -\frac{1}{2}\pi$. In the limit $\epsilon \to 0$, the effect is only at the point $s = 1$, which is a set of measure zero on L_+. However, for our matching equation (34) the zero is eliminated, so that the angle Φ_n given by Eq. (30) is the correct one to use for our solution.

[12] Ref. 1, p. 197

[13] See, for example, the Appendix of Ref. 2.

[14] Ref. 9, p. 79.

[15] Equation (B7) is found on p. 202 of Ref. 9. Equation (B8) is found in N. N. Lebedev, *Special Function and Their Applications* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1965), p. 111. See also Eq. (14) on p. 80 of Ref. 9.

[16] W. Gröbner and N. Hofreiter, *Integraltafel, Part 2, Bestimmte Integrale* (Springer-Verlag, Vienna, 1950), line (313.23).

[17] The second line of Eq. (B14) is from Ref. (16), line (312.8).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Lorentz-Covariant Matrices for Elementary-Particle Theories as Polynomials in the Spin Matrices*

T. J. NELSON†
Lawrence Radiation Laboratory, University of California, Berkeley, California 94720
AND
R. H. GOOD, JR.‡
Stanford Linear Accelerator Center, Stanford, California 94305

(Received 25 September 1969)

The problem of expressing the $(2s+1)\times(2s+1)$ covariantly defined matrices $S_{\mu\nu\rho}\ldots$ (used in theories of particles with spin) in terms of the angular-momentum matrices s is shown to be equivalent to the problem of finding a certain type of polynomial. Explicit expressions for the polynomials, recursion formulas, and differentiation properties are given.

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

In discussing particles with spin s, free or interacting, there is an approach which uses $(0,s)\oplus(s,0)$ wavefunctions[1–3] for massive particles, $(0,s)$ or $(s,0)$ wavefunctions[4,5] for massless particles. This approach is useful because it permits many of the well-known ideas about Dirac particles and 2-component neutrinos to be carried over directly into the higher-spin cases. Central in all the discussions are matrices which are generalizations of the Dirac matrices and Pauli matrices to the higher-spin problems.[6] In some of the discussions, explicit formulas for the matrices are needed. The purpose of this paper is to provide expressions for these matrices in terms of the angular-momentum matrices.

The problem leads to new considerations of the Lorentz group which may be of mathematical interest. It leads to the definition of a set of polynomials, implied by the fact that R_3 is a subgroup of the Lorentz group, which may have value in other Lorentz-group applications.

One is led to the definition of the matrices from consideration of complete sets of matrices. For any integer or half-integer s, one can form a complete set of $(2s+1)\times(2s+1)$ matrices from the three spin matrices s_i. One procedure is to form tensors from products of the matrices, irreducible with respect to R_3, up to tensor rank $2s$. For example, for $s=\frac{1}{2}$ the complete set is $\{1, s_i\}$, and for $s=1$ the complete set is $\{1, s_i, s_is_j+s_js_i-\frac{4}{3}\delta_{ij}\}$. However, an alternate procedure exists in which the $(2s+1)^2$ linearly independent matrices occur as components of a traceless symmetric Lorentz tensor of rank $2s$. The main subject of this paper is the organization based on Lorentz symmetry.

Consider the continuous Lorentz group. A general transformation of the coordinate 4-vector is written

$$x'_\mu = a_{\mu\nu}x_\nu, \tag{1}$$

where x_4 is it, $a_{\mu\nu}a_{\mu\rho}$ is $\delta_{\nu\rho}$, and a_{ij}, ia_{4j}, ia_{i4}, and a_{44} are real. The $(s,0)$ and $(0,s)$ representations are $e^{i\tau^*\cdot s}$ and $e^{i\tau\cdot s}$, respectively, where τ_i are three complex parameters that serve to label the Lorentz transformations. Let $\{S_\Gamma\}$ be any complete set of $(2s+1)\times(2s+1)$ matrices, the index Γ ranging over $(2s+1)^2$ values. Since $\{S_\Gamma\}$ is a complete set, one can expand

$$(e^{i\tau\cdot s})^\dagger S_\Gamma (e^{i\tau\cdot s})$$

in terms of members of the set,

$$(e^{i\tau\cdot s})^\dagger S_\Gamma (e^{i\tau\cdot s}) = \sum_\Delta A_{\Gamma\Delta}S_\Delta. \tag{2}$$

Here $A_{\Gamma\Delta}$ are the $(2s+1)^4$ expansion coefficients, which are functions of τ. (The dagger indicates the Hermitian conjugate. It does not coincide with the inverse owing to the nonunitary nature of the representations.) Evidently, $A_{\Gamma\Delta}$ is another representation of the continuous Lorentz group; in general, it is (s,s). Therefore, it is possible to make a special choice of the matrices S_Γ in which they are labeled like the components of a Lorentz tensor of rank $2s$, symmetric with respect to interchange of any index pair and with all contractions zero. Let these matrices be $S_{\mu\nu\rho}\ldots$, with $2s$ subscripts, where

$$S_{\mu\nu\rho}\ldots = S_{\nu\mu\rho}\ldots = S_{\mu\rho\nu}\ldots = \cdots, \tag{3}$$

$$S_{\mu\mu\rho}\ldots = 0. \tag{4}$$

Equation (2) for these special covariantly defined matrices reads

$$(e^{i\tau\cdot s})^\dagger S_{\mu\nu\rho}\ldots(e^{i\tau\cdot s}) = a_{\mu\alpha}a_{\nu\beta}a_{\rho\gamma}\cdots S_{\alpha\beta\gamma}\ldots. \tag{5}$$

The problem is how to construct the matrices $S_{\mu\nu\rho}\ldots$ from the spin matrices s_i.

For $s=\frac{1}{2}$, the result is well known:

$$(e^{i\tau\cdot s})^\dagger S_\mu (e^{i\tau\cdot s}) = a_{\mu\nu}S_\nu, \tag{6}$$

where S_i are the Pauli matrices σ_i and S_4 is i. For

$s = 1$ the matrices are[2]

$$S_{ij} = s_i s_j + s_j s_i - \delta_{ij}, \tag{7a}$$

$$S_{i4} = S_{4i} = i s_i, \tag{7b}$$

$$S_{44} = -1, \tag{7c}$$

and they satisfy

$$(e^{i\tau\cdot\mathbf{s}})^\dagger S_{\mu\nu}(e^{i\tau\cdot\mathbf{s}}) = a_{\mu\alpha}a_{\nu\beta}S_{\alpha\beta}. \tag{8}$$

The results for $s = \frac{3}{2}$ have also been worked out.[7]

The $S_{\mu\nu\rho}\ldots$ are a special case of the matrices introduced by Barut, Muzinich, and Williams.[6] One process for developing them in terms of angular-momentum matrices has already been given by Weinberg.[2] In the present paper an alternative approach to the problem is given and some new properties of the matrices are found.

The matrices provide especially concise and convenient descriptions of elementary particles. For a massless particle and antiparticle of helicity $-s$ and $+s$, the wave equation is[5]

$$S^\dagger_{\mu\nu\rho}\ldots\partial\psi/\partial x_\mu = 0. \tag{9}$$

That is, this equation has just two solutions of definite momentum $\mathbf{q}$ and energy $E = q$: one for the particle with helicity $-s$ and one for the antiparticle with helicity $+s$. For a particle and antiparticle of finite mass m and spin s, described by a $(0, s) \oplus (s, 0)$ wavefunction,[1-3] the wavefunction satisfies the Schrödinger–Klein–Gordon equation

$$\partial^2\psi/\partial x_\alpha\partial x_\alpha = m^2\psi \tag{10}$$

and Weinberg's equation

$$\gamma_{\mu\nu\rho}\ldots\partial^{2s}\psi/\partial x_\mu\partial x_\nu\partial x_\rho\ldots = (-m)^{2s}\psi. \tag{11}$$

We use units in which $\hbar = c = 1$ so that mass has the dimensions of (length)$^{-1}$. Here

$$\gamma_{\mu\nu\rho}\ldots = \begin{pmatrix} 0 & -S^\dagger_{\mu\nu\rho}\ldots \\ -S_{\mu\nu\rho}\ldots & 0 \end{pmatrix} \tag{12}$$

are the generalized Dirac matrices.

In this paper the problem of finding the matrices $S_{\mu\nu\rho}\ldots$ in terms of the spin matrices s_i is converted to the problem of calculating certain polynomials $P^s_N(z)$. Explicit formulas for the polynomials are given and some differentiation and recursion relations between them are reported.

II. THE EQUIVALENT POLYNOMIAL PROBLEM

As shown in Ref. 5, Eq. (5) is sufficient to determine the $S_{\mu\nu\rho}\ldots$ in terms of $\mathbf{s}$ except for an over-all numerical factor. That is, for an infinitesimal rotation,

$$a_{ij} = \delta_{ij} + \epsilon_{ijk}\theta_k, \tag{13a}$$

$$a_{i4} = a_{4i} = 0, \tag{13b}$$

$$a_{44} = 1, \tag{13c}$$

$$\tau_i = \theta_i \tag{13d}$$

[given $a_{\mu\nu}$, τ can be readily determined from Eq. (6) and the properties of the Pauli matrices], and Eq. (5) implies

$$[s_j, S_{4\ldots4}]_- = 0, \tag{14}$$

$$[s_j, S_{4\ldots4i_1\ldots i_N}]_- = i\epsilon_{ji_1k}S_{4\ldots4ki_2\ldots i_N} + \cdots(N \text{ terms}). \tag{15}$$

Here N can be any number from 1 to $2s$, $S_{4\ldots4}$ is the matrix with all subscripts 4, and $S_{4\ldots4i_1\ldots i_N}$ is the matrix with $(2s - N)$ subscripts 4, the rest being $i_1, i_2, \cdots, i_N$. The notation $+\cdots(N$ terms) indicates a sum in which each i index is treated the same way. That is, the right side of Eq. (15) is

$$i\epsilon_{ji_1k}S_{4\ldots4ki_2\ldots i_N} + i\epsilon_{ji_2k}S_{4\ldots4i_1ki_3\ldots i_N}\cdots + i\epsilon_{ji_Nk}S_{4\ldots4i_1\ldots i_{N-1}k}.$$

Also, for an infinitesimal pure Lorentz transformation,

$$a_{ij} = \delta_{ij}, \tag{16a}$$

$$a_{i4} = -a_{4i} = iv_i, \tag{16b}$$

$$a_{44} = 1, \tag{16c}$$

$$\tau_i = iv_i, \tag{16d}$$

Eq. (5) implies

$$[s_j, S_{4\ldots4}]_+ = 2isS_{4\ldots4j}, \tag{17}$$

$$[s_j, S_{4\ldots4i_1\ldots i_N}]_+ = i(2s - N)S_{4\ldots4ji_1\ldots i_N} - i\delta_{ji_1}S_{4\ldots4i_2\ldots i_N} + \cdots(N \text{ terms}), \tag{18}$$

for $N = 1, 2, \cdots, (2s - 1)$. The point is that Eq. (14) shows that $S_{4\ldots4}$ is a multiple of the identity, Eq. (17) fixes $S_{4\ldots4j}$ once $S_{4\ldots4}$ is decided on, and Eq. (18) determines $S_{4\ldots4ji_1\ldots i_N}$ once $S_{4\ldots4i_1\ldots i_N}$ and $S_{4\ldots4i_2\ldots i_N}$ are known. Thus, the $S_{\mu\nu\rho}\ldots$ are specified up to an over-all numerical factor.

In agreement with the previous work,[5] let the factor be chosen so that

$$S_{4\ldots4} = (i)^{2s}. \tag{19}$$

For $s = \frac{1}{2}$ and 1, this agrees with the usual conventions $S_4 = i$ and $S_{44} = -1$. It differs by a factor of i from the $s = \frac{3}{2}$ convention in Ref. 7. This choice is appropriate because it yields the results

$$S^\dagger_{\mu\nu\rho}\ldots = \pm S_{\mu\nu\rho}\ldots, \tag{20a}$$

$$C_s S_{\mu\nu\rho}\ldots C_s^{-1} = (-1)^{2s}S^*_{\mu\nu\rho}\ldots. \tag{20b}$$

The sign in Eq. (20a) is $+1$ (-1) when there is an even (odd) number of 4's in the indices. The asterisk denotes the complex conjugate and C_s is the unitary matrix such that

$$C_s \mathbf{s} C_s^{-1} = -\mathbf{s}^*.$$

Equations (20) are easily verified by using Eqs. (17)–(19). An important consequence of Eq. (20b) is the complement of Eq. (5):

$$(e^{i\tau^*\cdot\mathbf{s}})^\dagger S^\dagger_{\mu\nu\rho\ldots}(e^{i\tau^*\cdot\mathbf{s}}) = a_{\mu\alpha}a_{\nu\beta}a_{\rho\gamma}\cdots S^\dagger_{\alpha\beta\gamma\ldots}. \quad (21)$$

This can be derived by operating on Eq. (5) from the left by C_s, from the right by C_s^{-1}, and then taking the transpose.

The problem of expressing $S_{\mu\nu\rho\ldots}$ in terms of $\mathbf{s}$ is equivalent to the problem of determining a certain set of polynomials, as will now be shown. First of all, Eq. (17) gives

$$S_{4\ldots4j} = (i)^{2s-1}s_j/s. \quad (22)$$

Next, one observes from Eq. (18) that $S_{4\ldots4i_1\cdots i_N}$ is a sum of terms built from numerical factors, Kronecker δ symbols, and angular-momentum matrices. For even (odd) N, only terms of even (odd) degree in $\mathbf{s}$ occur, and the highest degree is N. Each term is symmetric in the indices $i_1 i_2 \cdots i_N$. Consequently, for even $N \geq 2$ one can write

$$\begin{aligned} S_{4\cdots4i_1\cdots i_N} &= c_0\langle\delta_{i_1i_2}\cdots\delta_{i_{N-1}i_N}\rangle \\ &\quad + c_2\langle\delta_{i_1i_2}\cdots\delta_{i_{N-3}i_{N-2}}s_{i_{N-1}}s_{i_N}\rangle \\ &\quad + \cdots + c_N\langle s_{i_1}s_{i_2}\cdots s_{i_N}\rangle, \end{aligned} \quad (23a)$$

and for odd $N \geq 3$,

$$\begin{aligned} S_{4\cdots4i_1\cdots i_N} &= c_1\langle\delta_{i_1i_2}\cdots\delta_{i_{N-2}i_{N-1}}s_{i_N}\rangle \\ &\quad + c_3\langle\delta_{i_1i_2}\cdots s_{i_{N-2}}s_{i_{N-1}}s_{i_N}\rangle \\ &\quad + \cdots + c_N\langle s_{i_1}s_{i_2}\cdots s_{i_N}\rangle. \end{aligned} \quad (23b)$$

Here the c's are numerical coefficients to be determined and the angular brackets indicate $1/N!$ times a sum over all permutations of the indices so that, for example,

$$\begin{aligned} \langle s_is_j\rangle &= \tfrac{1}{2}(s_is_j + s_js_i), \\ \langle\delta_{ij}\rangle &= \delta_{ij}, \\ \langle\delta_{ij}s_k\rangle &= \tfrac{1}{3}(\delta_{ij}s_k + \delta_{ik}s_j + \delta_{jk}s_i). \end{aligned}$$

The c coefficients define, for $N \geq 2$, an even or an odd polynomial, say

$$P^s_N(z) = (i)^{N-2s}\frac{(2s)!}{(2s-N)!\,(N)!}\sum c_r z^r, \quad (24)$$

where the sum ranges over $r = N, N-2, N-4, \cdots$, 1 or 0. The factor multiplying the sum is put in to make later formulas concerning the polynomials simpler. On contracting each i index of $S_{4\ldots4i_1\cdots i_N}$ with that of an arbitrary unit vector $\hat{p}_i$, the same polynomial is reached:

$$\begin{aligned} S_{4\cdots4i_1\cdots i_N}\hat{p}_{i_1}\cdots\hat{p}_{i_N} &= \sum c_r(s_i\hat{p}_i)^r \\ &= (i)^{2s-N}\frac{(2s-N)!\,(N)!}{(2s)!}P^s_N(s_i\hat{p}_i). \end{aligned} \quad (25)$$

It is appropriate to let this equation define the polynomials for $N = 0$ and 1, $N = 0$ meaning $S_{4\ldots4}$ on the left. Equations (19) and (22) then give the first two polynomials as

$$P^s_0(z) = 1, \quad (26)$$

$$P^s_1(z) = 2z. \quad (27)$$

By contracting each i index in Eq. (18) with that of $\hat{p}_i$, a recursion formula is obtained:

$$(N+1)P^s_{N+1}(z) = 2zP^s_N(z) - (2s+1-N)P^s_{N-1}(z). \quad (28)$$

This formula is valid for $N = 1, 2, \cdots, (2s-1)$, and thus it serves to determine all the polynomials up to $P^s_{2s}(z)$. As a matter of convenience, this equation is taken as defining $P^s_N(z)$ for $N > 2s$ also, although those higher degree polynomials are not needed for the covariantly defined matrices. The polynomials up to $N = 7$ are displayed in Table I. Once the polynomials are found, the c coefficients are known and $S_{\mu\nu\rho\ldots}$ are given by Eqs. (23).

III. THE POLYNOMIALS

In the recursion formula, Eq. (28), the coefficient of $P^s_{N-1}(z)$ goes negative for $N > 2s + 1$. It follows[8] that we are not dealing with a classical set of orthogonal polynomials. Nevertheless, it is easy to demonstrate the following general formula for the polynomials:

$$\begin{aligned} P^s_N(z) = \sum_{n=0 \text{ or } \frac{1}{2}}^{\frac{1}{2}N} &\frac{(-1)^{\frac{1}{2}N-n}2^{2n}}{(2n)!\,(\frac{1}{2}N-n)!} \\ &\times \frac{\Gamma(s-n+\frac{1}{2})\Gamma(z+n+\frac{1}{2})}{\Gamma(s-\frac{1}{2}N+\frac{1}{2})\Gamma(z-n+\frac{1}{2})}. \end{aligned} \quad (29)$$

TABLE I. The polynomials $P^s_N(z)$ for $N = 0$ to 7. When expressed as

$$P^s_N(z) = (i)^{N-2s}\frac{(2s)!}{(2s-N)!\,(N)!}\Sigma\, c_r z^r,$$

the coefficients c_r give $S_{\mu\nu\rho\ldots}$ according to Eqs. (23).

$P^s_0(z) = 1$
$P^s_1(z) = 2z$
$P^s_2(z) = 2z^2 - s$
$P^s_3(z) = \frac{2}{3}[2z^3 + (1-3s)z]$
$P^s_4(z) = \frac{1}{6}[4z^4 + 4(2-3s)z^2 + 3s(s-1)]$
$P^s_5(z) = \frac{1}{15}[4z^5 + 20(1-s)z^3 + (15s^2 - 25s + 6)z]$
$P^s_6(z) = \frac{1}{90}[8z^6 + 20(4-3s)z^4 + 2(45s^2 - 105s + 46)z^2 - 15s(s-1)(s-2)]$
$P^s_7(z) = \frac{1}{315}[8z^7 + 28(5-3s)z^5 + 7(30s^2 - 90s + 56)z^3 - 3(35s^3 - 140s^2 + 147s - 30)z]$

In the sum, n ranges in integer steps from 0 to $\frac{1}{2}N$ when N is even, from $\frac{1}{2}$ to $\frac{1}{2}N$ when N is odd. To prove the validity of this formula, one first derives, from Eq. (28), a recursion formula that links even polynomials or odd polynomials:

$$(N+1)(N+2)P^s_{N+2} = (4z^2 + 2N^2 - 4Ns - 2s)P^s_N - (2s+1-N)(2s+2-N)P^s_{N-2}. \quad (30)$$

It is clear that this relation, together with the polynomials for $N = 0, 1, 2, 3$, determines all the polynomials. One can easily verify that Eq. (29) does give the correct polynomials for $N = 0$–3 and that it satisfies Eq. (30).

The polynomials P^s_{2s+1} have special values. Because of the $\Gamma(s - \frac{1}{2}N + \frac{1}{2})$ factor in the denominator of Eq. (29), where $N = 2s + 1$ only the $n = s + \frac{1}{2}$ term in the sum contributes and

$$P^s_{2s+1}(z) = \frac{2^{2s+1}}{(2s+1)!}\frac{\Gamma(z+s+1)}{\Gamma(z-s)} = [2^{2s+1}/(2s+1)!](z+s) \times (z+s-1)\cdots(z-s+1)(z-s). \quad (31)$$

These polynomials are thus factors times the spin-matrix polynomials[9,10] whose roots are the eigenvalues of $s_i\hat{p}_i$. Furthermore, the recursion relation, Eq. (28), implies that

$$P^s_{2s+2}(z) = (2z/2s+2)P^s_{2s+1}(z) \quad (32)$$

and, thus, that all polynomials P^s_N with $N \geq 2s + 1$ contain the sth spin-matrix polynomial as a factor. It follows that

$$P^s_N(s_i\hat{p}_i) = 0, \quad \text{if} \quad N \geq 2s+1. \quad (33)$$

Referring to Eq. (25) one sees that, in the formulas for $S_{4\cdots 4i_1\cdots i_N}$ as functions of angular-momentum matrices $\mathbf{s}$, if you use matrices for representation s with $2s \leq N - 1$, the $S_{4\cdots 4i_1\cdots i_N}$ are identically zero.

One can set up recursion formulas on the s index of the polynomials also. Relating polynomials with s indices differing by $\frac{1}{2}$, the result is

$$P^{s+\frac{1}{2}}_N = \sum_r \frac{(2r)!}{2^{2r}(1-2r)(r!)^2} P^s_{N-2r}, \quad (34)$$

where in the sum $r = 0, 1, 2, \cdots, \frac{1}{2}N$ for N even, $r = 0, 1, 2, \cdots, \frac{1}{2}(N-1)$ for N odd. This can be verified by substitution in Eq. (28). We first write the recursion formula with s replaced by $s + \frac{1}{2}$, then substitute using Eq. (34), and finally use Eq. (28) again to eliminate polynomials multiplied by z. We could use $n = N/2 - r$ as the dummy index in Eq. (34). The sum then goes from $0(\frac{1}{2})$ to $N/2$ as in Eq. (29). However, this way of expressing the result is particularly simple. It is remarkable that the coefficients on the right are independent of N and s. The recursion formula relating polynomials with s indices differing by 1 is even simpler:

$$P^{s+1}_N = P^s_N - P^s_{N-2}. \quad (35)$$

This formula can be easily verified by substitution using the explicit form of Eq. (29) for the polynomials.

Another question is how the derivatives of the polynomials are expressible in terms of the polynomials. The answer is that

$$(d/dz)P^s_N(z) = \sum_r [2/(2r+1)]P^s_{N-2r-1}(z), \quad (36)$$

where $r = 0, 1, 2, \cdots, \frac{1}{2}(N-2)$ for N even, $r = 0, 1, 2, \cdots, \frac{1}{2}(N-1)$ for N odd. This result can be verified in essentially the same way as Eq. (34). Again, the coefficients are independent of N and s. In principle the polynomials could be determined by successive integration of these equations; that is, starting from P^s_0 known, one could integrate the $N = 1$ equation to get P^s_1, and so on. This program requires knowledge of the values of the polynomials at $z = 0$ to fix the integration constants at each step:

$$P^s_N(0) = 0, \quad \text{for } N \text{ odd}, \quad (37a)$$

$$P^s_N(0) = (-1)^{\frac{1}{2}N}\Gamma(s+1)/(\tfrac{1}{2}N)!\,\Gamma(s-\tfrac{1}{2}N+1), \quad \text{for } N \text{ even}. \quad (37b)$$

Thus, Eqs. (36) and (37) could be used as the definition of the polynomials. Equation (37b) can be verified by showing it satisfies the specialized form of Eq. (28) that results when $z = 0$.

IV. DISCUSSION

The matrices $S_{\mu\nu\rho\ldots}$ discussed here are covariantly defined using the representation $e^{i\tau\cdot s}$ and its Hermitian conjugate as in Eq. (5). Alternatively, one can set up covariantly defined matrices using the representation $e^{i\tau\cdot s}$ and its inverse; that is, for any complete set of $(2s+1)\times(2s+1)$ matrices S_Γ, instead of Eq. (2), one can consider

$$(e^{i\tau\cdot s})^{-1}S_\Gamma(e^{i\tau\cdot s}) = \sum_\Delta B_{\Gamma\Delta}S_\Delta. \quad (38)$$

Evidently, $B_{\Gamma\Delta}$ also is a representation of the continuous Lorentz group. Matrices covariantly defined in the sense of Eq. (38) can in fact be built from products of $S_{\mu\nu\rho\ldots}$ and $S^\dagger_{\mu\nu\rho\ldots}$. By combining Eqs. (5) and (21), one finds that

$$(e^{i\tau\cdot s})^{-1}S^\dagger_\mu\ldots S_\nu\ldots(e^{i\tau\cdot s}) = a_{\mu\alpha}\cdots a_{\nu\beta}\cdots S^\dagger_\alpha\ldots S_\beta\ldots. \quad (39)$$

The representation produced by $S_\mu^\dagger \cdots S_\nu \cdots$, in this way, is reducible and some of the parts provide a complete set of $(2s+1)\times(2s+1)$ matrices. For spin 1, for example, from $S_{\mu\nu}^\dagger S_{\rho\sigma}$ one needs the parts

$$T = \tfrac{1}{12} S_{\mu\nu}^\dagger S_{\mu\nu}, \tag{40a}$$

$$U_{\mu\nu} = -\tfrac{1}{6} i (S_{\mu\rho}^\dagger S_{\nu\rho} - S_{\nu\rho}^\dagger S_{\mu\rho}), \tag{40b}$$

$$V_{\mu\rho,\nu\sigma} = -\tfrac{1}{3}(S_{\mu\nu}^\dagger S_{\rho\sigma} + S_{\rho\sigma}^\dagger S_{\mu\nu} - S_{\rho\nu}^\dagger S_{\mu\sigma} - S_{\mu\sigma}^\dagger S_{\rho\nu} + 2\delta_{\mu\nu}\delta_{\rho\sigma} - 2\delta_{\rho\nu}\delta_{\mu\sigma}). \tag{40c}$$

The nonzero components of these parts are given by

$$T = 1, \tag{41a}$$

$$U_{i4} = \tfrac{1}{2}\epsilon_{ijk} U_{jk} = s_i, \tag{41b}$$

$$V_{i4,j4} = \tfrac{1}{2}\epsilon_{ikl} V_{kl,j4} = \tfrac{1}{2} V_{i4,kl}\epsilon_{jkl} = \tfrac{1}{4}\epsilon_{ikl} V_{kl,mn}\epsilon_{jmn} = s_i s_j + s_j s_i - \tfrac{4}{3}\delta_{ij}. \tag{41c}$$

Thus T, U, and V have 1, 3, and 5 independent components and they form the complete set of 3×3 matrices. From spin-1 functions φ and χ [(1, 0) and (0, 1)] transforming as

$$\varphi'(x') = e^{i\tau^*\cdot s}\varphi(x),$$
$$\chi'(x') = e^{i\tau\cdot s}\chi(x),$$

one can therefore make the following bilinear covariants: $\varphi^\dagger\chi$, $\varphi^\dagger U_{\mu\nu}\chi$, $\varphi^\dagger V_{\mu\rho,\nu\sigma}\chi$. The types of covariants that can occur are evident from a group-theoretical point of view since

$$(0,1)\otimes(0,1) = (0,0)\oplus(0,1)\oplus(0,2).$$

The quantities $S_{4\cdots 4i_1\cdots i_N}$ are reducible with respect to R_3; they are symmetric among the i indices but contracting on a pair of them does not give zero. However, one can build up linear combinations that are irreducible. Suppose $O_{i_1\cdots i_N}$ is irreducible, symmetric among all indices, has zero contractions, and is built up from $S_{4\cdots 4i_1\cdots i_r}$ with $r \le N$. Thus, for even N,

$$O_{i_1\cdots i_N} = a_0\langle\delta_{i_1i_2}\cdots\delta_{i_{N-1}i_N}\rangle + a_2\langle\delta_{i_1i_2}\cdots\delta_{i_{N-2}i_{N-1}}S_{4\cdots 4i_Ni_N}\rangle + \cdots + a_N\langle S_{4\cdots 4i_1\cdots i_N}\rangle \tag{42a}$$

and, for odd N,

$$O_{i_1\cdots i_N} = a_1\langle\delta_{i_1i_2}\cdots\delta_{i_{N-2}i_{N-1}}S_{4\cdots 4i_Ni_N}\rangle + a_3\langle\delta_{i_1i_2}\cdots S_{4\cdots 4i_{N-2}i_{N-1}i_Ni_N}\rangle + \cdots + a_N\langle S_{4\cdots 4i_1\cdots i_N}\rangle. \tag{42b}$$

On contracting each i index with an arbitrary vector p_i, one finds

$$O_{i_1i_2\cdots i_N}p_{i_1}p_{i_2}\cdots p_{i_N} = \sum_r a_r p^{N-r} S_{4\cdots 4i_1\cdots i_r}p_{i_1}\cdots p_{i_r},$$

where the sum ranges over $r = N, N-2, N-4, \cdots$, 1 or 0. Here one can operate with $\partial^2/\partial p_a\partial p_a$. The result is zero on the left because it is a contraction of two of the O indices. In the terms involving $S_{4\cdots 4aai_3\cdots i_r}$ on the right, this quantity can be replaced by $-S_{4\cdots 4i_3\cdots i_r}$ and the sum index revised accordingly. After simplification, the recursion formula

$$(r+1)(r+2)a_{r+2} = (N-r)(N+r+1)a_r$$

is obtained. This implies that

$$a_r = \frac{2^r\Gamma(\frac{1}{2}N + \frac{1}{2}r + \frac{1}{2})}{(\frac{1}{2}N - \frac{1}{2}r)!\, r!}\, C, \tag{43}$$

where C is some normalization constant independent of r. Equations (42) and (43) solve the problem of finding irreducible combinations of the $S_{4\cdots 4i_1\cdots i_N}$.

The polynomials can also be expressed as hypergeometric symbols:

$$P_N^s(z) = \frac{(-1)^N\Gamma(N-2s)}{N!\,\Gamma(-2s)}\, {}_2F_1(-N, -s-\tfrac{1}{2}+z, -2s; 2). \tag{44}$$

If this formula defines $P_N^s(z)$ for s negative and z pure imaginary, then $e^{-iN\pi/2}P_N^{-\lambda}(iX)$ is a member of one of Pollaczek's systems of polynomials.[11]

ACKNOWLEDGMENTS

R. Good was generously supported by the Ames Laboratory and by Iowa State University through the Faculty Improvement Leave program.

* Research supported by the U.S. Atomic Energy Commission.

† A.E.C. Postdoctoral Fellow. Address after January, 1970: Bell Telephone Laboratories, Murray Hill, N.J.

‡ On leave 1968–9 from Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Ia. 50010.

[1] H. Joos, Forschr. Physik **10**, 65 (1962).

[2] S. Weinberg, Phys. Rev. **133**, B1318 (1964).

[3] D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135**, B241 (1964).

[4] C. L. Hammer and R. H. Good, Jr., Phys. Rev. **111**, 342 (1958).

[5] T. J. Nelson and R. H. Good, Jr., Phys. Rev. **179**, 1445 (1969).

[6] A. O. Barut, I. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963).

[7] D. Shay, H. S. Song, and R. H. Good, Jr., Nuovo Cimento Suppl. **3**, 455 (1965).

[8] *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., Inc., New York, 1953), Sec. 10.3.

[9] T. A. Weber and S. A. Williams, J. Math. Phys. **6**, 1980 (1965).

[10] S. A. Williams, J. P. Draayer, and T. A. Weber, Phys. Rev. **152**, 1207 (1966).

[11] Ref. 8, Sec. 10.21.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

$SO^+(n, 1)$-Invariant Finite-Component Wave Equations and Reflection Invariance

P. M. MATHEWS*
International Atomic Energy Agency, International Centre for Theoretical Physics, Miramare-Trieste, Italy

(Received 5 August 1969)

First-order finite-dimensional equations of the generalized Bhabha type in a space with n spacelike and one timelike dimensions are considered. It is observed that invariance under the connected part $SO^+(n, 1)$ of the pseudo-orthogonal group in this space requires the use of a wavefunction transforming according to a representation of the larger group $SO^+(n + 1, 1)$. It is shown that though the *wavefunction* [and each part of it obtained by reduction with respect to $SO^+(n, 1)$] transforms into itself under the strong reflection R also, nevertheless the *wave equation* in an odd-dimensional space ($n + 1 = 2k + 1$) may or may not be invariant under R, when the wavefunction employed transforms irreducibly under the connected group $SO^+(n + 1, 1)$. Thus, contrary to the suggestion contained in the recent literature, where specific equations are considered, the mere fact of oddness of the dimension of the space does not force the use of representations reducible with respect to the latter group (the matter is representation dependent). It is shown, however, that,irrespective of the representation used and of the dimension of the space, the equation is invariant under a transformation Θ, defined as a combination of an improper transformation which reverses the sign of time and another *improper* transformation of reflection in any hyperplane in the spacelike variables. It is suggested that *TCP* in an arbitrary space be identified with Θ rather than with R, which simply reverses the sign of all coordinates.

I. INTRODUCTION

The discovery of ever-increasing numbers of elementary particles in recent years has served to direct attention to noncompact groups containing the Lorentz group as a subgroup. On one hand, such groups have been employed as "noninvariance" groups to generate mass spectra with infinitely many levels and, recently, also to make predictions regarding other quantities of interest like form factors.[1] On the other hand, wave equations which are invariant under groups like the inhomogeneous de Sitter group and similar higher-dimensional groups have been systematically explored, partly with the hope of incorporating internal symmetries alongside the space–time symmetry and partly as a way of presenting in unified form systems of Lorentz-invariant wave equations.[2] The use of invariant wave equations in odd-dimensional spaces [e.g., the (4 + 1)-dimensional de Sitter space] has in turn raised questions regarding the role of improper transformations in such spaces and, in particular, regarding the *TCP* invariance of these wave equations. Considering a space with $2k$ spacelike and one timelike dimensions, Rosen[3] and Fushchich[4] define the parity operation through a reversal of signs of all space coordinates, which is a *proper* transformation in view of the evenness of the number of space dimensions. [Here we give the notation of the coordinates x^μ ($\mu = 0, 1, 2, \cdots, n$). The first, x^0, is considered timelike and the x^i ($i = 1, 2, \cdots, n$) is spacelike. We use a time-favored metric: $g^{00} = +1$, $g^{11} = g^{22} = \cdots = g^{nn} = -1$. The group defined by matrices orthogonal with respect to this metric is denoted by $O(n, 1)$, its subgroup consisting of those matrices with determinant $+1$ by $SO(n, 1)$, and the subgroup of matrices continuous with the identity (called proper, orthochronous) by $SO^+(n, 1)$.] They conclude therefrom that *TCP* invariance really boils down to *TC* invariance and then proceed to show that the equation

$$(i\Gamma^\mu\partial_\mu + m)\psi = 0, \tag{1}$$

with the Γ^μ specifically chosen to be the lowest (2^n)-dimensional representation of a generalized Dirac algebra

$$\Gamma^\mu\Gamma^\nu + \Gamma^\nu\Gamma^\mu = 2g^{\mu\nu}, \tag{2}$$

does not possess *TCP* invariance, even though it is invariant under the proper orthochronous group $SO^+(2k, 1)$.

In this paper, we wish to point out that there exists a transformation, in any space of odd or even dimension, under which an $SO^+(n, 1)$-invariant equation would be automatically invariant. We prefer to call this operation *TCP* and use the symbol Θ for it. The operation considered in Refs. 3 and 4 is denoted here by R ("strong reflection"). We discuss first, in Sec. II, the irreducible representations (IR) of $SO^+(n, 1) \times R$, and show that for any n, fields transforming irreducibly under $SO^+(n, 1)$ provide, rather surprisingly, a representation space also for the operation R. The additional requirements for the existence of an $SO^+(n, 1)$-invariant *equation* of the general form (1) are then considered. It is shown that there is plenty of scope for choosing the Γ^μ to be used in (1) in such a way that the equation is also R invariant. The choice contained in (2) is a very special one and the lack of R invariance in this case (if the Γ^μ are taken to be

irreducible) is verified, and the kind of reducible representation of Γ^μ needed to ensure R invariance is specified. We then proceed to introduce the Θ operation in Sec. IV and discuss the invariance under this operation. We conclude with a few remarks on TC and P invariance.

II. FIELDS TRANSFORMING IRREDUCIBLY UNDER $SO^+(n, 1) \times R$

The basis of the expectation (and realization) of TCP invariance in Lorentz-invariant theories may be stated in broad terms as follows. The operation TCP, which in this case is the same as the "strong reflection" R, takes $x \to -x$, $t \to -t$, and $\psi(x, t) \to \psi(-x, -t)$ to within a linear transformation on the components of ψ. (Since we are considering only the invariance of free-field equations in which ψ occurs linearly, the more recondite questions associated with TCP invariance of systems of interacting quantum fields do not arise here. For this reason it makes no practical difference whether ψ is a quantum field or not and it may conveniently be visualized as a c-number field.) It effects a transformation of determinant $+1$ on the coordinates and, though it does not belong to $SO^+(3, 1)$, it does belong to the associated compact group $SO(4)$. Now, every irreducible representation D of $SO^+(3, 1)$ gives, through Weyl's "unitary trick," an irreducible representation D' of $SO(4)/Z_2$, where Z_2 is the Abelian invariant subgroup consisting of the elements I (identity) and R of $SO(4)$. The direct product of D' and of a 1-dimensional IR of Z_2 gives an IR of $SO(4)$ which, when continued back by the unitary trick, gives an IR of $SO^+(3, 1) \times R$, of the same dimension as the original representation D. Thus, the carrier space of any IR of the group $SO^+(3, 1)$ serves as one for $SO^+(3, 1) \times R$ also. In the context of our problem, this means that any field ψ which transforms according to an IR of $SO^+(3, 1)$ transforms into itself under R. For example, the wavefunction ψ in the Dirac equation transforms according to the direct sum of the IR's $D(\frac{1}{2}, 0)$ and $D(0, \frac{1}{2})$, of $SO^+(3, 1)$; the R (or TCP) transformation takes ψ into $\gamma^5\psi$, and here, γ^5 does not connect the two irreducible parts, but it carries $D(\frac{1}{2}, 0)$ and $D(0, \frac{1}{2})$ each into itself. It is then not surprising that invariance under the proper Lorentz group should lead to invariance under TCP. Clearly, the above arguments go through for any pseudo-orthogonal space of *even* dimension.

When odd-dimensional spaces are considered, then the operation

$$R: x^\mu \to -x^\mu, \quad \mu = 0, 1, \cdots, 2k, \quad \psi(x) \to \psi(-x),$$

is really TC together with a rotation, as noted in the introduction. The determinant of the transformation on the coordinates being -1, it is an *improper* transformation which does not belong to $SO(2k + 1)$. Therefore, an IR of $SO^+(2k, 1) \times R$ would be continued, by the unitary trick, to one of $SO(2k + 1) \times R \sim O(2k + 1)$, i.e., the full orthogonal group containing improper transformations also. But now a curious thing happens: The IR's of $O(2k + 1)$ are also IR's[5] of $SO(2k + 1)$, so that on continuing back to $SO^+(2k, 1) \times R$ one finds again that the carrier space of an IR of $SO^+(2k, 1)$ admits the operation $R = TC$ also and no enlargement of the space is necessary.

The conclusion so far is that the operation R can be represented on *any* irreducible representation space of $SO^+(n, 1)$ irrespective of whether n be even or odd. This statement would seem to be in contradiction with the findings of Rosen[3] and Fushchich,[4] but this contradiction is only apparent. What we have shown is merely that, given an $SO^+(n, 1)$-invariant equation involving a field transforming reducibly under this group, each irreducible part goes into itself even under the reflection R. This, however, is *not enough*, in general, to ensure form invariance of a given field equation. This kind of situation is not new. Indeed, already in the case of proper transformations which, by definition, carry an IR into itself, one does not have the freedom to insist on a wavefunction having only one irreducible part, *if* it is required to satisfy a particular wave equation. It is well known, for example, that the vector operator Γ^μ needed for setting up an equation invariant under $SO^+(3, 1)$ in the general form (1) can be constructed only in a *reducible* representation of this group.

The implication for the case of strong reflection is now clear. Despite the fact that the field ψ in any equation invariant under the proper group $SO^+(n, 1)$ automatically provides a representation space for R also, an operator Γ^μ with the correct behavior *may* nevertheless *not* exist within such a representation space. It would then be necessary to enlarge the space suitably and use a wavefunction which transform reducibly under $SO^+(n, 1) \times R$. We now examine in more detail the conditions under which this becomes necessary.

III. INVARIANCE OF WAVE EQUATIONS UNDER $SO^+(n, 1) \times R$

A general equation of the form (1) is invariant under the proper group if and only if the Γ^μ transform like a vector, which means that they must have the following commutation relations with the generators $I^{\mu\nu}$ of

the group:

$$[I^{\mu\nu}, \Gamma^{\rho}] = -ig^{\mu\rho}\Gamma^{\nu} + ig^{\nu\rho}\Gamma^{\mu}. \tag{3}$$

The generators themselves have the Lie algebra

$$[I^{\mu\nu}, I^{\rho\sigma}] = ig^{\nu\rho}I^{\mu\sigma} - ig^{\mu\rho}I^{\nu\sigma} + ig^{\mu\sigma}I^{\nu\rho} - ig^{\nu\sigma}I^{\mu\rho}. \tag{4}$$

Under R,

$$I^{\mu\nu} \to I^{\mu\nu}, \tag{5}$$

and for invariance of (1) we must further have

$$\Gamma^{\mu} \to -\Gamma^{\mu}. \tag{6}$$

Equations (3) and (4) are invariant under these transformations so that under R any representation of (3) and (4) would go into some representation, but whether these two are equivalent depends on the nature of the algebraic relations other than (3) and (4) which hold among the $I^{\mu\nu}$ and Γ^{μ} in specified representations. To consider a special example, in the case of $SO^{+}(4, 1)$, we can take

$$\Gamma^{0} = \epsilon\gamma^{0}, \quad \Gamma^{1} = \epsilon\gamma^{1}, \quad \Gamma^{2} = \epsilon\gamma^{2}, \quad \Gamma^{3} = \epsilon\gamma^{3}, \quad \Gamma^{4} = \epsilon\gamma^{5} \tag{7}$$

(where the γ's are the familiar 4×4 Dirac matrices with $\gamma^{5} = \gamma^{0}\gamma^{1}\gamma^{2}\gamma^{3}$ and $\epsilon = \pm 1$). The choices of $\epsilon = \pm 1$ in (7) both satisfy (2), and together with

$$I^{\mu\nu} = -i(\Gamma^{\mu}\Gamma^{\nu} - \Gamma^{\nu}\Gamma^{\mu}) \tag{8}$$

also satisfy (3) and (4). [Here, in (8), a factor $\frac{1}{4}$ on the rhs has been omitted for later convenience. This corresponds to normalizing the Γ^{μ} by multiplication by a factor $\frac{1}{2}$ and would bring an extra factor $\frac{1}{2}$ on the rhs of (7) and a factor $\frac{1}{4}$ in Eq. (2), but these factors are not relevant for our purposes.] But the two choices, $\epsilon = \pm 1$, provide *inequivalent* IR's of (3), (4), and (8), because the matrix $\Gamma^{0}\Gamma^{1}\Gamma^{2}\Gamma^{3}\Gamma^{4}$ which commutes with all the Γ^{μ} and $I^{\mu\nu}$ is a *different* multiple, $+1$ or -1, of the unit matrix in the two cases. The transformation R, which by virtue of (6) leads to the interchange of $\epsilon = +1$ and $\epsilon = -1$, would interchange these two inequivalent IR's of the set of matrices $(I^{\mu\nu}, \Gamma^{\rho})$ and so an $SO^{+}(4, 1)$-invariant equation which involves only one of these cannot be invariant under R.

We are now in a position to consider the general case of Eq. (1) in an $(n + 1)$-dimensional pseudo-orthogonal space. We do *not* require the Γ^{μ} to obey Eq. (2), but it is assumed that they obey the $SO^{+}(n, 1)$-invariance conditions (3) and that the Γ^{μ} together with the $I^{\mu\nu}$ form a finite-dimensional Lie algebra through the relation (8) which enables the algebra to be closed. Thus, the equations considered are a generalization of the Bhabha equation[6] to higher-dimensional spaces.

The essential point to note now is that Eqs. (3), (4), and (8) together define the Lie algebra associated with the group $SO^{+}(n + 1, 1)$ in a space of one-higher dimension than what we first started.[7] It has the generators

$$I^{\mu\nu} \quad \text{and} \quad I^{n+1,\mu} = -I^{\mu,n+1} = \Gamma^{\mu}. \tag{9}$$

Under the reflection R in the $(n + 1)$-dimensional space,

$$I^{\mu\nu} \to I^{\mu\nu} \quad \text{and} \quad I^{n+1,\mu} \to -I^{n+1,\mu}, \tag{10}$$

which corresponds to the transformation R':

$$x^{\mu} \to -x^{\mu}, \quad x^{n+1} \to x^{n+1} \tag{11}$$

in the $(n + 2)$-dimensional space.

It is clear from (9) that the invariance of Eq. (1) under the proper group $SO^{+}(n, 1)$ requires that the representation of the Γ matrices employed should belong to an IR (or of course, to a direct sum of IR's) of the larger group $SO^{+}(n + 1, 1)$. *The question now is whether or not every such equation would also be invariant under R.* The answer is in the affirmative only if the IR considered is also an IR of the group obtained by adjoining the element R' defined by (11) to $SO^{+}(n + 1, 1)$. We see below that the IR's of this enlarged group may contain more than one IR of $SO^{+}(n + 1, 1)$, in which case such higher-dimensional *reducible* representations of the latter group would have to be used in (1) to ensure invariance under R also. For a detailed analysis of this point, we have to consider the even- and odd-dimensional cases separately.

(a) *Equation* (1) *in even-dimensional spaces:* $n + 1 = 2k$. In this case the transformation (11) has determinant $+1$ and is therefore an element of the compact proper group $SO(n + 2)$ associated with $SO^{+}(n + 1, 1)$. It follows then from the line of argument used in Sec. II that *any* irreducible representation space of $SO^{+}(n + 1, 1)$ also accommodates the transformation R', and therefore, no enlarging of the dimension of Eq. (1) is necessitated by the introduction of strong reflection invariance. An interesting special case is the spinor representation which, in view of the fact that $n + 1 = 2k$, is of dimension 2^{k}. The matrices $\Gamma^{\mu} = I^{n+1,\mu}$ of this representation satisfy (2). This representation reduces into two inequivalent IR's of the original group $SO^{+}(n, 1)$ with which we started. The simplest example is the Dirac equation in $(3 + 1)$ dimensions, where the 4-dimensional wavefunction ψ transforms irreducibly under $SO^{+}(4, 1)$ but reduces into the two IR's, $D(\frac{1}{2}, 0)$ and $D(0, \frac{1}{2})$, of $SO^{+}(3, 1)$ as well as $SO^{+}(3, 1) \times R$.

(b) *Equation* (1) *in odd-dimensional spaces:* $n = 2k$. Now the transformation (11) has determinant -1.

So the compact group corresponding to $SO^+(n+1, 1)$ enlarged by this operation is not $SO(n+2)$ but $O(n+2)$. The IR's of $O(n+2)$, $n+2=2k+2$ being even, are of two types.[8] One type is characterized by nonselfassociated Young diagrams. These IR's are irreducible also under $SO(n+2)$ and lead, via the unitary trick, to IR's of $SO^+(n+1, 1)$ in the representation spaces of which the transformation R' is also defined. In other words, if the $I^{n+1,\mu}$ of this kind of IR of $SO^+(n+1, 1)$ are taken as the Γ^μ in (1), *no increase in dimension is necessary* to incorporate strong-reflection invariance. Thus, the doubling of the representation space, found necessary in the context of odd-dimensional pseudo-orthogonal spaces in Refs. 3 and 4, is a special feature of the representation of the Γ matrices employed there and *is by no means a general property associated with the dimensionality of the space.*

The second type of IR's of $O(n+2)$, characterized by selfassociated Young diagrams, is reducible into two inequivalent IR's of equal dimension of $SO(n+2)$. From this, one concludes that strong reflection interchanges two inequivalent IR's of $SO^+(n+1, 1)$, so that an equation of the form (1), employing one of these IR's alone, cannot be invariant under

$$SO^+(n, 1) \times R$$

though it is invariant under $SO^+(n, 1)$. In such cases, a doubling of the dimension of the equation invariant under the proper group, so as to accommodate both the above *inequivalent* IR's, is necessary if strong-reflection invariance is required.

Besides the above single-valued representations, one also has the fundamental spinor representation of dimension 2^{k+1} (where $2k=n$) of $O(n+2)$. It splits into two inequivalent IR's each of dimension 2^k under $SO(n+2)$. In both these IR's of $SO(n+2)$ the generators $I^{n+1,\mu} \equiv \Gamma^\mu$ obey Eq. (2) and, in addition, one of these sets satisfies

$$\Gamma^0\Gamma^1 \cdots \Gamma^n = (i)^{\frac{1}{2}n} I, \tag{12a}$$

while the other has

$$\Gamma^0\Gamma^1 \cdots \Gamma^n = -(i)^{\frac{1}{2}n} I. \tag{12b}$$

Either of these can be used in (1) to produce an $SO^+(n, 1)$-invariant equation, but a direct sum of these must be used if R invariance is also desired, since R interchanges (12a) and (12b) by virtue of its effect on Γ^μ as in (6). It is this very special case which is the basis of the remarks in Refs. 3 and 4 regarding R invariance in odd-dimensional spaces. Santhanam[9] suggests that, since R invariance cannot be had within the 2^k-dimensional IR of the Dirac–Clifford algebra C_{2k+1} [defined by (2)] which satisfies (12a), a reducible representation of these equations be used. However, it is clear from our consideration above that *no* representation (even a reducible one) satisfying both (2) and (12a) can ensure R invariance. The important point which is missed is that a pairing of a representation (reducible or irreducible) of (2) and (12a) with the corresponding one of (2) and (12b) is essential. Such a representation would, of course, be a reducible representation of (2) alone, but a special one, which forms $2k+1$ of the elements of the Dirac–Clifford algebra C_{2k+3}.

The considerations of the last paragraph may be illustrated by the example of the $SO^+(2, 1)$-invariant equation which has either the set of Pauli matrices $(\sigma_1, \sigma_2, \sigma_3)$ or the set $(-\sigma_1, -\sigma_2, -\sigma_3)$ as the Γ^μ. The equations formed from these two possibilities go over into one another under the R operation and can be combined into one R-invariant equation which has three of the Dirac γ matrices, say $\gamma^1, \gamma^2, \gamma^3$ for the Γ^μ.

IV. THE *TCP* OPERATION

The preceding discussion concerned the invariance of (1) under the strong reflection R which is identical with the usual *TCP* in even-dimensional spaces, but is just *TC* in the odd-dimensional case. We recall here that in the former case, the fact of R being a transformation with determinant $+1$ led to the expectation that it could be built into every IR of $SO^+(2k+1, 1)$ by passing to $SO^+(2k+1, 1) \times R$, with the help of the unitary trick, from $SO(2k+2)$ to which R belongs. It follows that when R is a transformation of determinant -1, as in the case of odd-dimensional spaces, there is no longer any such strong reason to expect R invariance. But it follows equally well that one *should* expect invariance of the theory in *any* dimension under any transformation which has determinant $+1$; for example,

$$x^0 \to -x^0, \quad x^1 \to -x^1, \quad x^i \to x^i, \quad i = 2, 3, \cdots, n. \tag{13}$$

This is a combination of $TC(x^0 \to -x^0)$ with the *improper* transformation of *space-reflection*,[10] i.e., a reflection of space in a hyperplane which passes through the origin. (The special choice of this hyperplane which singles out the 1-direction is really irrelevant; any other would do as well and is related to the one given, by a proper transformation.) We define this to be the *TCP* operation Θ in a space of arbitrary dimension and verify that *any* equation of the form (1) is automatically invariant under this operation. It does not require any special consideration in the even-dimensional case since it differs from R only by a

rotation. In the case of odd dimensions, the point to note is that $SO^+(n, 1)$ together with Θ—a semidirect product, since Θ does not commute with $SO^+(n, 1)$, unlike R—goes over into $SO(n + 1)$ through the unitary trick [and not into $O(n + 1)$ as when we had R instead of Θ]. It is easy then to see, by arguments similar to those used repeatedly above, that any IR of the group obtained from $SO^+(n, 1)$ by adjoining Θ has the same dimension as that of $SO(n, 1)$ by itself. Nothing new comes up when we consider the invariance of Eq. (1) either, for we have to consider for that purpose the group $SO^+(n + 1, 1)$ to which the transformation

$$x^0 \to -x^0, \quad x^1 \to -x^1, \quad x^i \to x^i, \quad i = 2, 3, \cdots, n + 1, \quad (14)$$

a transformation with determinant $+1$, is to be adjoined instead of R' as defined by (11). It follows then that every IR of this group is irreducible under the proper subgroup $SO^+(n + 1, 1)$ and therefore, any equation based on an IR of $SO^+(n + 1, 1)$ [and hence invariant under $SO^+(n, 1)$] is invariant under Θ also.

In the case of the special representation of the Γ^μ defined by (2), for example, it is obvious that there is a matrix, namely $\Gamma^0\Gamma^1$, which induces the transformation

$$\Gamma^0 \to -\Gamma^0, \quad \Gamma^1 \to -\Gamma^1, \quad \Gamma^i \to \Gamma^i, \quad i = 2, \cdots, n, \quad (15)$$

necessary for invariance of (1) under the Θ operation (13). This statement is independent of whether n is odd or even.

V. THE OPERATIONS *TC*, *P*, *T*, AND *C*

The question of *TC* invariance in odd-dimensional spaces has been treated as R invariance in Sec. III and, in view of the discussion in the last section, the requirements for invariance under P (improper transformation of space *reflection*) are the same as for *TC* invariance.

In even-dimensional spaces, the question of *TC* invariance leads one to IR's of $O(n + 2)$, where $n + 1 = 2k$; but since these are IR's of $SO(n + 2)$ also, as mentioned in Sec. II, one finds by previous arguments that no increase in dimension of an $SO^+(n, 1)$-invariant equation is necessary to incorporate *TC* invariance.

On the question of separate *T* or *C* invariance, some interesting observations have been made by Fushchich[4] in the context of the special spinor representation of the Γ matrices. A general discussion has to be based on the relation between the relevant IR's and their complex-conjugate IR's and is not attempted here.

ACKNOWLEDGMENTS

The author would like to acknowledge discussions with Dr. M. M. Bakri which have been useful in clarifying some ideas. He is grateful to Professor A. Salam and Professor P. Budini and to the International Atomic Energy Agency for hospitality at the International Centre for Theoretical Physics, Trieste, which has made this work possible.

* Present Address: Division of Theoretical Physics, University of Madras, Madras-25, India.

[1] Y. Nambu, Progr. Theoret. Phys. (Kyoto), Suppl. No. **37–38**, 368 (1966); J. G. Kuriyan and E. C. G. Sudarshan, Phys. Rev. **162**, 1650 (1967); *Lectures in Theoretical Physics*, A. O. Barut and W E. Brittin, Eds. (Gordon and Breach, Sci. Publishers, Inc., New York, 1968), Vol. XB, p. 377; A. O. Barut, Nucl. Phys. **B9** 194 (1969). In these applications, infinite-dimensional representations of the relevant groups are employed. In the present work, we are concerned with finite-component wave equations having noncompact invariance groups.

[2] See, for example, M. M. Bakri, J. Math. Phys. **10**, 289 (1969); V. I. Fushchich and I. Yu. Krivsky, Nucl. Phys. **B7**, 79 (1968).

[3] S. P. Rosen, J. Math. Phys. **9**, 1593 (1968).

[4] V. I. Fushchich, Akad. Nauk Ukr. S.S.R, Kiev, Preprint No. ИТФ-69-I.

[5] F. D. Murnaghan, *The Theory of Group Representations* (Dover Publications, Inc., New York, 1963), Chs. IX and X.

[6] H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945).

[7] See Ref. 6, for case when $n = 3$.

[8] For the results used here regarding the irreducible representations of the orthogonal groups $O(n)$ and the reducibility of these with respect to the proper subgroup $SO(n)$, see Murnaghan, Ref. 5, or H. Weyl, *The Classical Groups, their Invariants and Representations* (Princeton University Press, Princeton, N.J., 1946).

[9] T. S. Santhanam, International Centre for Theoretical Physics, Preprint IC/69/46 (1969).

[10] The suggestion here is that the essential aspect of the parity operation, in the familiar case when the number of spacelike dimensions is odd, is not that it changes the sign of all these coordinates, but rather that it is an *improper* transformation, one which is not contained in the rotation group and which is not physically realizable. These same properties are characteristic of space reflection as defined above, and when the inversion itself becomes a proper transformation as when the number of spacelike dimensions is even, its essential role must be taken over by space reflection. It is an inconvenience that it does not express itself symmetrically with respect to all the coordinates, but one cannot ignore the existence of such transformations for this reason.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Asymptotic Evaluation of Multidimensional Integrals*

LAWRENCE SIROVICH
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

(Received 8 May 1969; Revised Manuscript Received 24 November 1969)

The asymptotic evaluation of a wide class of multidimensional integrals occurring in mathematical physics is considered. In this class are included integrals of the form

$$\frac{1}{(2\pi)^N}\int_{-\infty}^{\infty}\rho(\mathbf{k})\exp[i\mathbf{k}\cdot\mathbf{x}-\sigma(\mathbf{k})t]\,d\mathbf{k}.$$

A semiconstructive method is proven and certain classes of integrals are asymptotically evaluated. Examples involving problems in partial differential equations and a transport equation are given.

1. INTRODUCTION

It is often the case in mathematical physics that the resolution of a problem reduces itself to the evaluation of integrals. This is especially true in the case of linear problems. In spite of this, formidable problems still usually remain. Often the integrals one encounters do not have representations in terms of familiar or, for that matter, tabulated functions. In such cases one tries to take advantage of the presence of large parameters in the integrand.[1] Techniques for exploiting the presence of a single large parameter occurring in 1-dimensional integrands have been considered exhaustively in the literature.[2–4] These classical techniques have also proven successful in a number of cases involving multidimensional integrals,[5–9] but progress there has not been as great. Generally speaking, these methods represent the asymptotic evaluation in terms of an evaluation at the stationary point of a function. The location of the stationary point of this exponent is, of course, not part of the classical methods, and this part of the calculation usually proves impossible except when only elementary functions are involved.

In this paper, we develop a method for the asymptotic evaluation of integrals which avoids these restrictions and difficulties. We consider integrals over an arbitrary number of dimensions, containing a number of parameters. In order to do this, we naturally have to give up a certain amount of generality. We do this by focusing on integrals which are typical of a large class that occur in mathematical physics. As the reader will see, the restrictions placed on the integrand of the integrals under study are typically the case in physical problems containing a dissipative mechanism.

The method discussed in this paper has already proven successful in a number of problems in gas dynamics,[10–13] magnetohydrodynamics[14,15] and kinetic theory.[16,17] A general discussion for integrals over one dimension has already been given.[10]

2. STATEMENT OF THE MAIN RESULT

To begin with, we consider integrals of the form

$$I=\frac{1}{(2\pi)^N}\int_{-\infty}^{\infty}e^{i\mathbf{k}\cdot\mathbf{x}-\sigma(\mathbf{k})t}\rho(\mathbf{k})\,d\mathbf{k}. \tag{1}$$

[This is generalized below by Eq. (21).] Both $\mathbf{k}$ and $\mathbf{x}$ denote N-dimensional vectors and $d\mathbf{k}$ represents the N-space volume element. The integration may extend over any part of N space. The infinite limits of integration are indicated only for simplicity; other limits can be included in the support of ρ. The sole restriction in this regard is that the region of integration include the origin.

Without loss of generality we may take

$$\sigma(\mathbf{k})=0,\quad \mathbf{k}=0. \tag{2}$$

$\sigma(\mathbf{k})$ is said to be admissible if it satisfies the following five conditions:

(i) $\operatorname{Re}\sigma=\sigma_r\geq 0$,

(ii) $\sigma_r=0$, only if $\mathbf{k}=0$,

(iii) $\sigma\in C$,

(iv) in the neighborhood of the origin

$$\sigma=if(\mathbf{k})+g(\mathbf{k})+O(\mathbf{k}^3),$$

where f and g are real, continuous, and homogeneous degree one and two, respectively,

(v) $g=0$, only if $\mathbf{k}=0$.

Condition (i), which demands that $-\sigma_r$ have a global maximum, states that the system in question is stable, and condition (ii) then adds that it be dissipative. Condition (iv) is obtained if $\sigma\in C^3$, and is therefore somewhat weaker. That the first order is pure imaginary and the second pure real is often a direct consequence of the transformation properties of the equations governing the system.

For most purposes it suffices to place the following

weak restriction on the function $\rho(\mathbf{k})$:

(vi) $|\rho|$, $\int_{-\infty}^{\infty} |\rho|\, d\mathbf{k} < M < \infty$.

[Actually, as will be clear, (vi) is stronger than necessary, but we avoid such mathematical niceties.]

Main Result

If σ is admissible and ρ satisfies (vi), then I, as given by (1), can be written as

$$I = I_0 + O^*(t^{-\frac{1}{2}(N+1)}),$$

where

$$I_0 = \frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x} - if(\mathbf{k})t - g(\mathbf{k})t} \rho(\mathbf{k})\, d\mathbf{k} \tag{3}$$

and $O^*(t^{-p})$ represents a quantity such that

$$\lim_{t\to\infty} t^{p-\delta} O^*(t^{-p}) = 0$$

for any small $\delta > 0$.

With the additional condition at the origin

(vii) $\rho = \rho_0 + O(\mathbf{k})$,

we obtain

$$I = \frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x} - if(\mathbf{k})t - g(\mathbf{k})t} d\mathbf{k}\rho_0 + O^*(t^{-\frac{1}{2}(N+1)})$$

or

$$I = I^0\rho_0 + O^*(t^{-\frac{1}{2}(N+1)}), \tag{4}$$

where I^0 is defined through (4). We prove (3) and (4), and an extension (21) and (22) in Sec. 3. In the remainder of this section we comment on certain aspects of the calculations involved in (3) and (4).

Before going further, it should be noted that the main result is, in a sense, only semiconstructive. The integral appearing in (3), and even the one in (4), cannot generally be carried out in terms of elementary functions. Even after taking into consideration the homogeneity requirements on f and g [see (iv)], we are left with an integral which cannot, in general, be carried out. In Sec. 4, however, we carry out the full integration in several important special cases.

So far, nothing has been said of the parameters $\mathbf{x}$. In fact, the error estimates in (3) and (4) are completely independent of $\mathbf{x}$. This is another aspect of the semiconstructive nature of the calculation. The region of validity in $\mathbf{x}$ space of the calculation is the restriction to those $\mathbf{x}$ such that the integral of (3) or (4) is large compared with the error estimate. Often this region becomes apparent only after the completion of the integration of the integrals in (3) or (4). We give explicit examples of this in Sec. 4.

In this same vein, we point out that it is conceivable that the integral terms in (3) or (4) are less than, or of the same order as, the error estimate for all values of $\mathbf{x}$. In such a case, the calculation as it stands only represents an estimate for the integral I.

At this point, we mention an essential difference between (3) and (4). In general, the modulus of the error estimates in (3) and (4) are quite different. The first form (3) only involves an expansion in terms of the scales of the underlying operator leading to (1), while (4) involves in addition an expansion of the data of the problem. In other words, (3) leads to a sharper result and, hence, may be used for significantly shorter times. As an illustration, in gas dynamics,[10] (3) is valid for times large compared to the mean time between molecular collision, while (4), in addition, requires that the time be large compared with the time taken by a sound wave to traverse the initial disturbance.

In the remainder of this section, we indicate how one can typically obtain the functions $f(\mathbf{k})$ and $g(\mathbf{k})$. For many problems of mathematical physics, this usually presents a simple calculation. The following remarks are only meant to be formal.

Let us consider a problem which may be considered as an initial-value problem. Consider

$$\frac{\partial v}{\partial t} = Lv. \tag{5}$$

L is a linear operator and v belongs, say, to a Hilbert space (perhaps finite). The problem, then, is to solve (5), subject to specified initial data

$$v\,(t = 0) = v^0. \tag{6}$$

Further, let us assume that the problem has already been Fourier transformed, i.e., L, v, and v^0 are to be regarded as functions of $\mathbf{k}$. Using formal manipulations and inverting the transformations leads to the following representation for the solution:

$$v = \frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x}} e^{tL} v^0\, d\mathbf{k}. \tag{7}$$

The representation of e^{tL} itself involves a number of problems, but, generally speaking, it can be represented in terms of the spectrum of L. Therefore, a typical term which arises out of the point spectrum of L has the form

$$\frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x} + \lambda(\mathbf{k})t} \rho(\mathbf{k})\, d\mathbf{k}. \tag{8}$$

The function $\rho(\mathbf{k})$ is partly due to the operator and partly due to the initial data. The function $\lambda(\mathbf{k})$ is an

eigenvalue of L, i.e., there exists a q such that

$$Lq = \lambda q.$$

The above integral is, of course, of the form I in Eq. (1). To employ the main result it is, of course, necessary to prove λ admissible. Aside from this, it is important to note the way in which $\lambda(\mathbf{k})$ arises. Now, although λ may be quite difficult to obtain, its expansion is in practice much simpler to obtain. Formally, one writes

$$L = L_0 + kL_1 + k^2L_2 + \cdots$$

and

$$\lambda = \lambda_0 + k\lambda_1 + k^2\lambda_2 + \cdots,$$

where the L_i (known) and the λ_i are homogeneous of degree zero in $\mathbf{k}$. A number of results and methods for such perturbation series for L are discussed in the literature.[18,19]

Finally, although it is not our intention here to consider the initial-value problem in any detail, one further point is worth mentioning. This has to do with the solution to (5) and (6), say, in the form (7). Suppose there exists a discrete eigenvalue of L, λ, such that its real part for $\mathbf{k} = 0$ is greater than any other part of the spectrum of L. Clearly, then, for $t \to \infty$ the major contribution to (7) is given by (8), and by our main result this has the form (3) or (4).

3. PROOF OF THE MAIN RESULT AND ITS GENERALIZATION

We require the following lemma in our proof.

Lemma: For $\sigma(\mathbf{k})$ admissible there exists a $g_0 > 0$ and an $\epsilon_1 > 0$, such that

$$\sigma_r - \hat{\epsilon}^2 g_0 \geq 0 \tag{9}$$

for all $|\mathbf{k}| > \hat{\epsilon}$ and any $\hat{\epsilon} > 0$ such that $\hat{\epsilon} \leq \epsilon_1$.

Proof: Since g is homogeneous of degree two

$$g(\mathbf{k}) = k^2 g(\mathbf{e}),$$

with

$$\mathbf{e} = \mathbf{k}/k. \tag{10}$$

From the continuity of g and condition (v) we can also write

$$g_M \geq g(\mathbf{e}) \geq g_m > 0,$$

with g_M and g_m the maximum and minimum, respectively.

From condition (iv),

$$\sigma_r - \tfrac{1}{2}k^2 g_m > \tfrac{1}{2}k^2 g_m + O(k^3).$$

Hence, there exists an $\epsilon_0 > 0$ such that

$$\sigma_r - \tfrac{1}{2}k^2 g_m > 0, \quad |k| < \epsilon_0. \tag{11}$$

In fact, let ϵ_0 be the maximum such value.

Next from the continuity of σ_r and the dissipative condition (ii), we have that σ_r is bounded away from zero if $|\mathbf{k}|$ is bounded away from zero. Therefore, for all $k_0 > 0$, we have

$$\sigma_r \geq G(k_0) = \inf_{|\mathbf{k}| > k_0} \sigma_r(\mathbf{k}) > 0. \tag{12}$$

Then there exists an $\hat{\epsilon}$ such that

$$0 < \hat{\epsilon} < \epsilon_0$$

and

$$G(k_0) \geq \tfrac{1}{2}\hat{\epsilon}^2.$$

For, if this were not true, there would exist a point set $\{k_i\}$ such that

$$\sigma(\mathbf{k}_i) < \tfrac{1}{2}g_m \epsilon_i^2,$$

where $\{\epsilon_i\}$ is a sequence converging to zero. From (11), $|k_i| \geq \epsilon_0$ for all i. But then this contradicts (12). Denote the largest such $\hat{\epsilon} \leq \epsilon_0$ by ϵ_1. Then from (11) we have

$$\sigma_r - \tfrac{1}{2}\hat{\epsilon}^2 g_m \geq 0$$

for all $|\mathbf{k}| > \hat{\epsilon}$ and $\hat{\epsilon} \leq \epsilon_1$. Setting $g_0 = \tfrac{1}{2}g_m$, we have proven the lemma.

Proof of the Main Result

From condition (iv) we have

$$\lim_{|\mathbf{k}| \to 0} \left| \frac{\sigma - if - g}{k^3} \right| = c' < \infty,$$

which may be zero. In any case, we set

$$c = 1 + c'.$$

There exists an $\epsilon_2 > 0$ such that

$$|\sigma - if - g| \leq c\,|\mathbf{k}|^3 \tag{13}$$

for

$$|\mathbf{k}| \leq \epsilon_2.$$

Next, we choose

$$\epsilon_3 = \min(\epsilon_1, \epsilon_2),$$

where ϵ_1 is the same as that of the lemma. Then, for $\epsilon < \epsilon_3$ we decompose the integral (1) as follows:

$$\begin{aligned}(2\pi)^N I = &\int_{|\mathbf{k}| \geq \epsilon} e^{i\mathbf{k}\cdot\mathbf{x} - \sigma(\mathbf{k})t} \rho(\mathbf{k})\, d\mathbf{k} \\ &+ \int_{|\mathbf{k}| < \epsilon} (e^{i\mathbf{k}\cdot\mathbf{x} - \sigma(\mathbf{k})t} - e^{i\mathbf{k}\cdot\mathbf{x} - if t - gt}) \rho(\mathbf{k})\, d\mathbf{k} \\ &+ \int_{|\mathbf{k}| < \epsilon} e^{i\mathbf{k}\cdot\mathbf{x} - if(\mathbf{k})t - g(\mathbf{k})t} \rho(\mathbf{k})\, d\mathbf{k} \\ &= I_1 + I_2 + I_3.\end{aligned}$$

Then, since

$$|(2\pi)^N I - I_3| \leq |I_1| + |I_2|,$$

we need only estimate the first two integrals. Using (9) of the lemma, we easily have

$$|I_1| < \int_{|\mathbf{k}|\geq\epsilon} e^{-\sigma_r t}\,|\rho|\,d\mathbf{k}$$
$$= e^{-g_0\epsilon^2 t}\int_{|\mathbf{k}|\geq\epsilon} e^{-(\sigma_r - g_0\epsilon^2)t}\,|\rho|\,d\mathbf{k} \leq Me^{-g_0\epsilon^2 t}. \quad (14)$$

Next, writing

$$I_2 = \int_{|\mathbf{k}|<\epsilon} e^{i\mathbf{k}\cdot\mathbf{x} - ift - gt}(e^{(g+if-\sigma)t} - 1)\rho\,d\mathbf{k}$$

and taking absolute values, we have

$$|I_2| < M\int_{|\mathbf{k}|<\epsilon} |e^{(g+if-\sigma)t} - 1|\,d\mathbf{k}.$$

Using a well-known inequality and Eq. (13), we obtain

$$|I_2| \leq Mc\Omega_N t\int_0^\epsilon e^{ck^3 t}k^{N+2}\,dk$$
$$= \frac{e^{c\epsilon^3 t}Mc\Omega_N t\epsilon^{N+3}}{N+3} = K\epsilon^{N+3}te^{c\epsilon^3 t}, \quad (15)$$

where Ω_N is the surface area of the unit sphere in N-dimensions.[20]

We now set

$$\epsilon = t^{-\frac{1}{2}(1-\delta)}, \quad (16)$$

where $\delta > 0$ is small and

$$t > \epsilon_0^{-2/(1-\delta)}. \quad (17)$$

With this choice of ϵ, Eq. (13) becomes

$$|I_1| < Me^{-g_0 t^\delta}$$

and (15) becomes[21]

$$|I_2| \leq Ke^{c/(t^{\frac{1}{2}} - \frac{3}{2}\delta)}/t^{\frac{1}{2}(N+1)-\frac{1}{2}\delta(N+1)} = O^*(t^{-\frac{1}{2}(N+1)}). \quad (18)$$

This proves the main result (3), since the extension of the limits of integration to ∞ in I adds an asymptotically small contribution.

It is clear from the above proof that condition (vii) immediately leads to (4). In fact, it seems that a many-term expansion of $\rho(k)$ leads to an asymptotic expansion. There would be no value in this, since, if $\rho = O(\mathbf{k})$, a simple estimate on I shows that $I_1 = O^*(t^{-\frac{1}{2}(N+1)})$, i.e., it is of the same order as already neglected terms. Therefore, if $\rho(\mathbf{k})$ satisfies (vii), Eq. (4) is obtained, i.e.,

$$I = I^0\rho_0 + O^*(t^{-\frac{1}{2}(N+1)}),$$

where I^0 is the same as defined through (4):

$$I^0(\mathbf{x}, t) = \frac{1}{(2\pi)^N}\int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x} - ift - gt}\,d\mathbf{k}. \quad (19)$$

For reasons which are discussed in the next section, it is sometimes best not to use the expansion of $\rho(\mathbf{k})$ even if ρ satisfies (vii). In these cases, we can write, instead of (3),

$$I = I^0 * \rho(\mathbf{x}) + O^*(t^{-\frac{1}{2}(N+1)}), \quad (20)$$

where[22]

$$\rho(\mathbf{x}) = \int_{-\infty}^{\infty} e^{-i\mathbf{k}\cdot\mathbf{x}}\rho(\mathbf{k})\,d\mathbf{k}.$$

The asterisk in the first term of (20) denotes the N-dimensional spatial convolution product.

An examination of the proof of the main result given in this section shows that it depends in no essential way on the form $e^{i\mathbf{k}\cdot\mathbf{x}}$, in which the vector $\mathbf{x}$ appears. In fact, if this exponential is replaced by any function $F(\mathbf{x}, \mathbf{k})$ which is uniformly bounded, no alteration in the proof is necessary. Hence, writing

(viii) $|F(\mathbf{x}, \mathbf{k})| < \infty$ uniformly,

we extend our main result.

Extension of the Main Result

Consider the integral

$$I' = \frac{1}{(2\pi)^N}\int_{-\infty}^{\infty} F(\mathbf{x}, \mathbf{k})e^{-\sigma(\mathbf{k})t}\rho(\mathbf{k})\,d\mathbf{k}, \quad (21)$$

with σ admissible, ρ satisfying (vi), and F satisfying (viii). Then, for large t, we have

$$I' = \int_{-\infty}^{\infty} F(\mathbf{x}, \mathbf{k})e^{-if(\mathbf{k})t - g(\mathbf{k})t}\rho(\mathbf{k})\,d\mathbf{k} + O^*(t^{-\frac{1}{2}(N+1)}). \quad (22)$$

If ρ satisfies (vii), an expression similar to (4) may also be written.

It is clear that the above generalization is quite extensive, and we have chosen to focus attention on Fourier transform type integrals (1) only because of their natural importance.

4. SPECIAL CASES AND EXTENSIONS

Case (1):

$$\sigma \sim i\boldsymbol{\alpha}\cdot\mathbf{k} + \sum_{i,j=1}^{N}\beta_{ij}k_i k_j$$

in the neighborhood of the origin: If, in addition to being admissible, we have also that σ has two continuous derivatives at the origin, then we may

conclude that

$$f = \boldsymbol{\alpha} \cdot \mathbf{k},$$
$$g = \sum_{i,j=1}^{N} \beta_{ij} k_i k_j,$$

where $\boldsymbol{\alpha}$ is a real constant vector and $\boldsymbol{\beta}$ is a real, symmetric, positive-definite matrix of order N. In this case, I^0 takes the form

$$I^0 = \frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot(\mathbf{x}-\boldsymbol{\alpha}t)-\mathbf{k}\cdot\boldsymbol{\beta}\cdot\mathbf{k}t}\, d\mathbf{k}, \tag{23}$$

which may be integrated immediately and gives

$$I^0 = (4\pi t)^{-\frac{1}{2}N} e^{-(\mathbf{x}-\boldsymbol{\alpha}t)\cdot\boldsymbol{\beta}^{-1}\cdot(\mathbf{x}-\boldsymbol{\alpha}t)/4t} / (\det \boldsymbol{\beta})^{\frac{1}{2}}. \tag{24}$$

In the above forms, the dot product denotes the inner product in Euclidean N space.

Now, having the form (24) for I^0, we can, in this case, give a precise characterization to the region in $\mathbf{x}$ space for which the asymptotic approximation is valid. Writing, for example,

$$I = I^0 \rho_0 + O^*(t^{-\frac{1}{2}(N+1)}),$$

we clearly have that $\mathbf{x}$ must be such that

$$(\mathbf{x} - \boldsymbol{\alpha}t) \cdot \boldsymbol{\beta}^{-1} \cdot (\mathbf{x} - \boldsymbol{\alpha}t) = o(t \ln t). \tag{25}$$

Using the properties of $\boldsymbol{\beta}$, a cruder estimate is that $|\mathbf{x} - \boldsymbol{\alpha}t| = o((t \ln t)^{\frac{1}{2}})$. Outside these regions we have the estimate that $I = O^*(t^{-\frac{1}{2}(N+1)})$.

Case (2): 1-Dimensional Integrals: For $N = 1$, $\mathbf{k} = k$ and the admissibility condition (iv) is clearly equivalent to σ having two derivatives at the origin. In this case I^0 [Eq. (24)]has the form

$$I^0\,(N = 1) = e^{-(x-\alpha t)^2/4\beta t} / (4\pi\beta t)^{\frac{1}{2}}. \tag{26}$$

The range of validity is still given by (25). In terms of the integral[22]

$$\mathfrak{I} = \frac{1}{2\pi} \int e^{-\sigma(k)t + ikx}\, dk,$$

we can write

$$\mathfrak{I} = I^0 + O^*(t^{-1}). \tag{27}$$

We now consider the next term in the asymptotic development of $\mathfrak{I}$ or I. To accomplish this, we assume that $\sigma(k)$ satisfies

$$\text{(iv')}: \sigma = i\alpha k + \beta k^2 - i\gamma k^3 + O(k^4)$$

instead of (iv). ($\beta > 0$ and α and γ real.) The condition (13) is now replaced by

$$|\sigma - i\alpha k - \beta k^2 + i\gamma k^3| < ck^4, \quad |k| < \epsilon_2. \tag{28}$$

Also, instead of condition (vii), we now let $\rho(k)$ be such that

$$\text{(vii')}: \rho = \rho_0 + \rho_1 k + O(k^2).$$

Then, using (iv') and repeating an argument analogous to that given in the previous section, we can directly prove the following:

$$I = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-\alpha t) - \beta k^2 t + i\gamma k^3 t} \rho(k)\, dk + O^*(t^{-\frac{3}{2}}). \tag{29}$$

Equivalently, instead of (27) we can write

$$\mathfrak{I} = \mathfrak{J}^0 + O^*(t^{-\frac{3}{2}})$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-\alpha t) - \beta k^2 t + i\gamma k^3 t}\, dk + O^*(t^{-\frac{3}{2}}). \tag{30}$$

Finally, with the additional requirement (vii') on $\rho(k)$, we can write

$$I = \left(\rho_0 - \frac{i}{\alpha}\rho_1 \frac{\partial}{\partial x}\right)\mathfrak{J}^0 + O^*(t^{-\frac{3}{2}}), \tag{31}$$

where $\mathfrak{J}^0$ is defined through (30). On setting

$$k = \eta - i\beta/3\gamma$$

in $\mathfrak{J}^0$, we can reduce it to a standard representation of the Airy function Ai (x), and we obtain[23]

$$\mathfrak{J}^0 = \frac{e^{\beta(x-\alpha t)/3\gamma + 2\beta^3 t/27\gamma^2}}{\pi(3\gamma t)^{\frac{1}{3}}} \operatorname{Ai}\left(\frac{x - \alpha t}{(3\gamma t)^{\frac{1}{3}}} + \frac{\beta^2 t^{\frac{2}{3}}}{(3\gamma)^{\frac{4}{3}}}\right). \tag{32}$$

This, in turn, may be expanded for t large and we obtain[24]

$$\mathfrak{J}^0 = \frac{e^{-(x-\alpha t)^2/4\beta t}}{(4\pi\beta t)^{\frac{1}{2}}}\left[1 - \frac{3(x - \alpha t)}{4\beta^2 t} + \frac{(x - \alpha t)^3}{8\beta^3 t^2} + \cdots\right]. \tag{33}$$

It is also clear that the range of validity is only marginally extended. That is, the form in (33) holds in the basically parabolic region

$$(x - \alpha t)^2 = o(t \ln t),$$

and outside this region we have the estimate $O^*(t^{-\frac{3}{2}})$.

The further development for, say, $\mathfrak{I}$ may be continued in this way. Further differentiability conditions on σ at the origin have to be assumed, and their series development substituted. It is clear that the exact evaluation of $\mathfrak{J}^0$ given by (32) was fortuitous and that the integrals in the general case cannot be expected to have known forms. However, as (33) already indicates, such an evaluation is not really necessary and a direct (second) asymptotic analysis of $\mathfrak{J}^0$ for $t \to \infty$ could have to be performed. This is also true in the general case, although we do not pursue this line of study.

The same remarks are also valid in the N-dimensional case. In general, the estimate $O^*(t^{-\frac{1}{2}(N+1)})$ may be improved upon by assuming further differentiability conditions on σ, and the development of I may be obtained. Since this is straightforward and, perhaps, of only limited value, we do not pursue it further.

Finally, we remark on the distinction between using the development of $\rho(k)$ directly in I and, on the other hand, leaving this intact and only developing $\mathfrak{I}$ in

$$I = \mathfrak{I}(x, t) * \rho(x).$$

This distinction is important and even crucial in certain problems. Referring back to the formal problem posed by (5)–(8) in Sec. 2, we recognize that, in expanding in small k, two distinct expansions are in play. There is, of course, the expansion of the underlying operator, but there is also the expansion of the data of the problem. In general, this involves two entirely different time scales. For example, in problems involving gas dynamics, the expansion of the operator is tantamount to considering times large compared to the time between molecular collisions (which is extremely small under ordinary conditions). If the initial data is also expanded, the circumstances become more involved and the time it takes a sound-wave to traverse the data comes into play. This latter quantity can be quite large, and the utility of the resulting asymptotic development becomes quite limited. These remarks manifest themselves in the modulus of the error term $O^*(t^{-\frac{1}{2}(N+1)})$. The constant that is implicit in this symbol can be radically different under the two different expansions. This is already clear in (31), where the presence of ρ_1 can signal that t must be extremely large for the development to be valid. As a practical rule, one may say that only that portion of $\rho(k)$ arising from the underlying operator should be expanded and that expanding the remaining portion can badly inhibit the usefulness of the asymptotic development.

The above remarks are applicable without modification to the N-space case.

Case (3): $\sigma = \sigma(|\mathbf{k}|)$. In a number of applications (see, e.g., Refs. 12 and 13), due to the isotropy of the underlying equations, an admissible σ is a function of only $k = |\mathbf{k}|$. Although σ is not differentiable in this case, the admissibility condition immediately leads to

$$\sigma = i\alpha k + \beta k^2 + O(k^3)$$

with α real and $\beta > 0$. As shown in the previous discussion, when $N = 1$, the calculation is straightforward. This important case, however, in more than one dimension is far from trivial, and we now consider the case $N \geq 2$ in some detail.

We first note that the estimate of the error term may be greatly improved. To accomplish this, we can start with I itself, (1) or, alternatively, we may consider

$$\mathfrak{I} = \frac{1}{(2\pi)^N} \int e^{-\sigma(k)t + i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k}, \tag{34}$$

since

$$I = \mathfrak{I}(\mathbf{x}, t) * \rho(\mathbf{x}).$$

The limits of integration in (34) have been purposely left out since, if convergence problems appear with infinite limits, we may take the limits of integration in (34) to be finite without loss of generality. [We have already demonstrated through (14) that the contribution from outside the neighborhood of the origin is exponentially small in time, uniformly in $\mathbf{x}$.[22]]

The integration of (34) can be carried out most easily by introducing spherical coordinates in N-space. Integrating over all angles but the polar angle yields

$$\mathfrak{I} = \frac{\Omega_{N-1}}{(2\pi)^N} \int_0^\pi d\theta e^{ikr\cos\theta} \sin^{N-2}\theta\, d\theta \int_{k\geq 0} e^{-\sigma(k)t} k^{N-1}\, dk,$$

where

$$\Omega_k = 2\pi^{\frac{1}{2}k}/\Gamma(\tfrac{1}{2}k) \tag{35}$$

is the area of unit sphere in k space. The remaining angular integration can be carried in terms of Bessel functions and yields,[25]

$$\mathfrak{I} = \frac{r^{1-\frac{1}{2}N}}{(2\pi)^{\frac{1}{2}N}} \int_{k\geq 0} J_{\frac{1}{2}N-1}(kr) k^{\frac{1}{2}N} e^{-\sigma(k)t}\, dk.$$

We now focus attention on

$$\mathfrak{J} = r^{\frac{1}{2}N-1}\mathfrak{I} = (2\pi)^{-\frac{1}{2}N} \int_{k\leq 0} J_{\frac{1}{2}N-1}(kr) k^{\frac{1}{2}N} e^{-\sigma(k)t}\, dk. \tag{36}$$

As mentioned before,[22] the upper limit of integration may be taken to be finite if convergence difficulties appear with infinite limits of integration.

Since

$$|J_\nu(x)| \leq 1, \quad \nu \geq 0,$$

the integral $\mathfrak{J}$ clearly falls under the hypothesis (viii) of the extension of the main result and we may apply (22). The modification due to the presence of $k^{\frac{1}{2}N}$ is, of course, of no consequence. Therefore, writing

$$\mathfrak{J}_0 = (2\pi)^{-\frac{1}{2}N} \int_0^\infty J_{\frac{1}{2}N-1}(kr) e^{-i\alpha kt - \beta k^2 t} k^{\frac{1}{2}N}\, dk \tag{37}$$

and using the arguments leading to (18), we can easily show that

$$\mathfrak{J} = \mathfrak{J}_0 + O^*(t^{-\frac{1}{2}N-1}). \tag{38}$$

As preparation for the evaluation of $\mathfrak{J}_0$, we first express the Hankel expansion of the Bessel function[26]:

$$J_\nu(x) = \left(\frac{2}{\pi x}\right)^{\frac{1}{2}}\Bigg[\cos(x - \tfrac{1}{4}(2\nu+1)\pi) \times \left(\sum_{m=0}^{P}(-1)^m(\nu, 2m)(2x)^{-2m} + O(|x|^{-2P-2})\right) + \sin(x - \tfrac{1}{4}(2\nu+1)\pi) \times \left(\sum_{m=0}^{Q}(-1)^m(\nu, 2m+1)(2x)^{-2m-1} + O(|x|^{-2Q-3})\right)\Bigg]. \quad (39)$$

If ν is of half-odd-integer order ($N = 3, 5, \cdots$), these series are known to terminate with

$$P = [\tfrac{1}{4}(2\nu - 1)] \geq 0,$$
$$Q = [\tfrac{1}{4}(2\nu - 3)] \geq 0, \quad (40)$$

i.e., with the limits (40), the finite expansions in (39) are exact. In this case, N odd, the integration may be carried out explicitly and, in fact, if $N = 3$, then

$$\mathfrak{J}_0(N=3) = [16r^{\frac{1}{2}}(\beta t\pi)^{\frac{3}{2}}]^{-1} \times [(r+\alpha t)e^{-(r+\alpha t)^2/4\beta t}\operatorname{erfc}(i(r+\alpha t)/2(\beta t)^{\frac{1}{2}}) + (r-\alpha t)e^{-(r-\alpha t)^2/4\beta t}\operatorname{erfc}(i(\alpha t - r)/2(\beta t)^{\frac{1}{2}})]. \quad (41)$$

The argument in the first expression of the bracket is large, and, on performing the required asymptotic expansion, we find

$$\mathfrak{J}_0(N=3) = \frac{r-\alpha t}{16r^{\frac{1}{2}}(\beta t\pi)^{\frac{3}{2}}}e^{-(r-\alpha t)^2/4\beta t}\operatorname{erfc}\left(\frac{i(\alpha t - r)}{2(\beta t)^{\frac{1}{2}}}\right) + O(t^{-\frac{5}{2}}). \quad (42)$$

[In this last expression, r should be regarded as being $r \geq O(t)$. For r small the entire expression (36) will be shown below to be of negligible order.]

The general case for N odd may be obtained, but we do not give it, since it is tedious to express and, as we will shortly see, it carries already neglected orders. For N even, no explicit integration seems to be available.[27] At this point of the analysis, we abandon the search for an explicit calculation of (37), and perform a second asymptotic analysis. As will be seen, this is at no expense to the $O^*(t^{-\frac{1}{4}N-1})$ estimate, and we find an explicit calculation independently of the dimension N.

The asymptotic analysis of $\mathfrak{J}_0$ (37), under the condition $r = o(t)$, is fairly straightforward and we merely quote the result:

$$\mathfrak{J} \sim \frac{(-i)^N\Gamma(N-1)}{\Gamma(\frac{1}{2}N)\pi^{\frac{1}{2}N}2^{N-1}\alpha^N}\frac{r^{\frac{1}{2}N-1}}{t^N}, \quad r = o(t).$$

It is, therefore, clear that, for all $N \geq 2$, this is already small, compared with neglected terms. In what follows, therefore, we may restrict attention to

$$r \geq O(t).$$

Using this and returning to $\mathfrak{J}$ in (34) and (36), we see the superiority of the estimate (38) over the estimate given by (4).

As a first step in our evaluation of $\mathfrak{J}_0$, we demonstrate that

$$\hat{\mathfrak{J}}_0 = \frac{1}{(2\pi)^{\frac{1}{2}N}}\int_0^{O(t^{-p})} J_{\frac{1}{2}N-1}(kr)k^{\frac{1}{2}N}e^{-i\alpha tk-\beta k^2t}\,dk,$$

with p such that $O(t^{-p}) = O^*(t^{-1})$ is of an already neglected order. To avoid carrying unimportant constants in our estimates, we consider instead $O(\hat{\mathfrak{J}}_0)$. Then, clearly,

$$O(\hat{\mathfrak{J}}_0) \leq \int_0^{t^{-p}} J_{\frac{1}{2}N-1}(kr)k^{\frac{1}{2}N}\,dk.$$

The integral on the right may be explicitly evaluated[28]:

$$O(\hat{\mathfrak{J}}_0) \leq r^{-\frac{1}{2}N-1}(r/t^p)^{\frac{1}{2}N}J_{\frac{1}{2}N}(r/t^p).$$

In view of the fact that $r \geq O(t)$ and $p < 1$, we can asymptotically evaluate the Bessel function and find

$$O(\hat{\mathfrak{J}}_0) \leq (r^{1+\frac{1}{2}}t^{\frac{1}{2}p(N-1)})^{-1},$$

which is, clearly, of an already neglected order. We next consider

$$\mathfrak{J}_0 - \hat{\mathfrak{J}}_0 = \frac{1}{(2\pi)^{\frac{1}{2}N}}\int_{O(t^{-p})}^{\infty} J_{\frac{1}{2}N-1}(kr)k^{\frac{1}{2}N}e^{-i\alpha tk-\beta k^2t}\,dk.$$

From the limits of integration and the condition on r, kr is large, and we therefore write

$$J_{\frac{1}{2}N-1}(kr) = (2/\pi kr)^{\frac{1}{2}}\cos(kr - \tfrac{1}{4}(N-1)\pi) + O(|kr|^{-\frac{3}{2}}).$$

Hence, we consider

$$A = \frac{1}{(2\pi)^{\frac{1}{2}N}}\int_{O(t^{-p})}^{\infty}\left[J_{\frac{1}{2}N-1}(kr) - \left(\frac{2}{\pi kr}\right)^{\frac{1}{2}}\cos(kr - \tfrac{1}{4}(N-1)\pi)\right]k^{\frac{1}{2}N}e^{-i\alpha tk-\beta k^2t}\,dk.$$

Proceeding as before, we have

$$O(A) \leq \frac{1}{r^{\frac{3}{2}}}\int_{t^{-p}}^{\infty} k^{\frac{1}{2}(N-3)}e^{-\beta tk^2}\,dk \leq \frac{1}{r^{\frac{3}{2}}t^{\frac{1}{4}(N-1)}} \leq O(t^{-\frac{1}{4}N-\frac{5}{4}}),$$

which is also of a neglected order. Finally, it only

remains for us to consider

$$B = \frac{1}{(2\pi)^{\frac{1}{2}N}} \int_0^{O(t^{-p})} k^{\frac{1}{2}N} \left(\frac{2}{\pi kr}\right)^{\frac{1}{2}} \times \cos(kr - \tfrac{1}{4}(N-1)\pi) e^{-iakt - \beta k^2 t}\, dk,$$

from which we directly obtain

$$O(B) \leq [r^{\frac{1}{2}} t^{\frac{1}{2}p(N+1)}]^{-1}$$

and which again is of negligible order. We have, therefore, demonstrated that

$$\mathfrak{J}_0 = \frac{2}{r^{\frac{1}{2}}(2\pi)^{\frac{1}{2}(N+1)}} \int_0^{\infty} k^{\frac{1}{2}(N-1)} e^{-\beta k^2 t - iakt} \times \cos(kr - \tfrac{1}{4}(N-1)\pi)\, dk + O^*(t^{-\frac{1}{4}N-1}). \quad (43)$$

The resulting integral may now be carried out in terms of confluent hypergeometric functions[29] and, due to their special form in our case, these may in turn be written as parabolic cylinder functions.[30] Choosing these latter forms, we find that

$$\mathfrak{J}_0 = \frac{\Gamma(\frac{1}{4}(N+1))\Gamma(\frac{1}{4}(N+3))2^{\frac{1}{4}(N+1)}}{r^{\frac{1}{2}}(2\pi)^{\frac{1}{2}(N+1)}2(\beta t)^{\frac{1}{4}(N+1)}\Gamma(\frac{1}{2})} \times \left[e^{-\frac{1}{4}i(N-1)\pi - (r-\alpha t)^2/8\beta t} D_{-\frac{1}{2}(N+1)}\left(\frac{i(\alpha t - r)}{(2\beta t)^{\frac{1}{2}}}\right) + e^{\frac{1}{4}i(N-1)\pi - (r+\alpha t)^2/8\beta t} D_{-\frac{1}{2}(N+1)}\left(\frac{i(r+\alpha t)}{(2\beta t)^{\frac{1}{2}}}\right)\right] + O^*(t^{-\frac{1}{4}N-1}). \quad (44)$$

If N is set equal to 3 in (44), we get our previous result (42). Noting that the argument of the second term is large, and using the asymptotic estimate[31]

$$D_\nu(Z) = e^{-\frac{1}{4}Z^2} Z^\nu (1 + O(Z^{-2})), \quad |\arg Z| \leq \tfrac{3}{4}\pi,$$

we conclude that this term is of negligible order. Therefore, we finally have

$$\mathfrak{J} = (r^{\frac{1}{2}N-1})\mathfrak{J} = \frac{\Gamma(\frac{1}{4}(N+1))\Gamma(\frac{1}{4}(N+3))e^{-\frac{1}{4}i(N-1)\pi - (r-\alpha t)^2/8\beta t}}{2\Gamma(\frac{1}{2})r^{\frac{1}{2}}(2\pi^2\beta t)^{\frac{1}{4}(N+1)}} \times D_{-\frac{1}{2}(N+1)}\left(\frac{i(\alpha t - r)}{(2\beta t)^{\frac{1}{2}}}\right) + O^*(t^{-\frac{1}{4}N-1}), \quad (45)$$

where we have used (38) to replace $\mathfrak{J}_0$ by $\mathfrak{J}$. In the interest of completeness, we note that, if N is odd, $\frac{1}{2}(N+1)$ is integer and the following formula is of the value[32] (n integer)

$$D_{-n-1}(Z) = (\tfrac{1}{2}\pi)^{\frac{1}{2}} \frac{(-1)^n}{n!} e^{-\frac{1}{4}Z^2} \frac{d^n}{dZ^n}\left[e^{\frac{1}{2}Z^2} \operatorname{erfc}\left(\frac{Z}{\sqrt{2}}\right)\right].$$

For N even, the following may be useful[32]:

$$D_{-\frac{3}{2}}(Z) = (2\pi)^{-\frac{1}{2}} Z^{\frac{3}{2}} [K_{\frac{3}{4}}(\tfrac{1}{4}Z^2) - K_{\frac{1}{4}}(\tfrac{1}{4}Z^2)]$$

and

$$\nu D_{\nu-1}(Z) = e^{-\frac{1}{4}Z^2} \frac{d}{dZ}[e^{\frac{1}{4}Z^2} D_\nu(Z)],$$

where the K_μ refer to the modified Bessel functions.

We once again note from (45) that the estimate (38) is superior to (3) and (4) (except in the case $N = 2$, when it gives the same result). Also, note that the range of validity may be obtained from (45). Without going into details, we further note that, since $D_\nu = O(1)$ in the neighborhood of the origin, (45) is valid for at least

$$|r - \alpha t| = O(t^{\frac{1}{2}}).$$

Before ending this section, we add a cautionary example. Consider the following integral:

$$f(t) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} e^{-t(1-e^{ik})}\, dk.$$

In this case, $\sigma = 1 - e^{ik}$ in the interval $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ certainly satisfies all the admissibility conditions. Then, applying the main result, we obtain

$$f(t) = (2\pi/t)^{\frac{1}{2}} e^{-\frac{1}{2}t} + O^*(t^{-1}).$$

Hence, the result of the asymptotic analysis is less than the error estimate. This signals the failure of the main result for this integral, as it should, since standard methods show that $f(t) \sim 2\pi i e^{-t}$.

5. APPLICATIONS

We consider three applications in the following. These have been chosen to demonstrate the range of the main result, rather than for their physical importance. Applications to a number of specific physical problems have already been cited.[10–17]

Although in each problem below a mathematically rigorous analysis may be given, our discussion is only meant to be formal.

Problem 1: Consider the following initial-value problem[33]:

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 - \mu\nabla^2 \frac{\partial}{\partial t}\right) s(\mathbf{x}, t) = 0,$$

$$s(t=0) = \delta(\mathbf{x}), \quad \frac{\partial s(t=0)}{\partial t} = \mu\nabla^2\delta(\mathbf{x}), \quad (46)$$

with the constant $\mu > 0$. The partial differential equation (46) is probably the simplest one demonstrating wave propagation and diffusion. Introducing the Fourier transform

$$s(\mathbf{k}, t) = \int_{-\infty}^{\infty} e^{-i\mathbf{k}\cdot\mathbf{x}} s(\mathbf{x}, t)\, d\mathbf{x},$$

we see that the solution to (46) is

$$s(\mathbf{x}, t) = \frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x}}(e^{\sigma^+ t} + e^{\sigma^- t})\, dk,$$

where

$$\sigma^{\pm} = \tfrac{1}{2}[-\mu k^2 \pm (\mu^2 k^4 - 4k^2)^{\frac{1}{2}}].$$

Both σ^+ and σ^- satisfy the conditions of admissibility and, at the origin,

$$\sigma^{\pm} = \pm ik - \tfrac{1}{2}\mu k^2 + O(k^3).$$

The main result is, therefore, applicable. In particular, the results of Case (2) of the previous section apply. On using (45), we have, for $N \geq 2$,

$$s(\mathbf{x}, t) = \frac{\Gamma(\frac{1}{4}(N+1))\Gamma(\frac{1}{4}(N+3))e^{-\frac{1}{4}i(N-1)\pi - (r-t)^2/4\mu t}}{2r^{\frac{1}{2}(N+1)}\Gamma(\frac{1}{2})(\pi^2\mu t)^{\frac{1}{4}(N+1)}} \times D_{-\frac{1}{2}(N+1)}\left(\frac{i(t-r)}{(\mu t)^{\frac{1}{2}}}\right) + O^*(t^{-\frac{3}{4}N}),$$

and from (26), for $N = 1$,

$$s(x, t) = e^{-(x-t)^2/2\mu t}/(2\pi\mu t)^{\frac{1}{2}} + e^{-(x+t)^2/2\mu t}/(2\pi\mu t)^{\frac{1}{2}} + O^*(t^{-1}).$$

Problem 2: Consider the following transport equation[34,35]:

$$\left(\frac{\partial}{\partial t} + \boldsymbol{\xi}\cdot\nabla + \nu(\xi)\right) f = \nu(\xi)\left(\int (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}\xi^2} f\nu\, d\xi\right) \Big/ \left(\int (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}\xi^2} \nu\, d\xi\right)$$

$$= (\nu/\nu_1)\hat{\rho}(\mathbf{x}, t) = Kf. \tag{47}$$

The collision frequency ν is a positive, monotonically increasing function of the magnitude of the molecular velocity $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$. We attempt to solve (47) in an unbounded domain and subject to the initial data

$$f(t = 0) = \delta(\mathbf{x}). \tag{48}$$

We first consider the Fourier-transformed problem

$$\left(\frac{\partial}{\partial t} + i\mathbf{k}\cdot\boldsymbol{\xi} + \nu - K\right) f = 0, \qquad f(t = 0) = 1. \tag{49}$$

Next, writing

$$L = -i\mathbf{k}\cdot\boldsymbol{\xi} - \nu + K,$$

we write the Eq. (49) as

$$\frac{df}{dt} = Lf.$$

Following the formalism given in Sec. 2, the solution to Eqs. (49) is

$$f(\mathbf{k}, \boldsymbol{\xi}, t) = e^{tL}. \tag{50}$$

The operator $\nu^{-1}K$ is clearly a projector and, hence, $\nu - K$ is nonpositive. Further, we may prove that L has just one eigenvalue $\lambda(\mathbf{k})$ and that it satisfies all the admissibility conditions. (For $\mathbf{k}$ sufficiently large this eigenvalue may disappear.) In addition, the operator L has a continuous spectrum which covers a 2-dimensional region to the left of

$$\operatorname{Re}\sigma = \nu(0)$$

in the complex σ plane.[35] (For ν constant, this region degenerates to a single line $\operatorname{Re}\sigma = -\nu$.) Denoting this region by $C(k)$ and an element of area in the complex σ plane by ds, we can write (50) in the form

$$f(\mathbf{k}, \boldsymbol{\xi}, t) = e^{\lambda(\mathbf{k})t} g_0(\xi, k) + \int_{C(\mathbf{k})} e^{\sigma t} g(\sigma, \boldsymbol{\xi}, \mathbf{k})\, ds, \tag{51}$$

where the eigenfunction g_0 and the "improper eigenfunction" g are still to be determined. The eigenvalue determination leads to

$$(\lambda + i\mathbf{k}\cdot\boldsymbol{\xi} + \nu) g_0 = Kg_0.$$

On using the perturbation analysis outlined in Sec. 2, we easily find

$$\lambda = -\alpha k^2 + O(k^4),$$

with

$$\alpha = \frac{1}{3}\left(\int e^{-\frac{1}{2}\xi^2} \nu^{-2}\, d\boldsymbol{\xi}\right) \Big/ \left(\int e^{-\frac{1}{2}\xi^2} \nu^{-1}\, d\boldsymbol{\xi}\right)$$

and

$$g_0 = \beta + O(k), \quad \beta = \text{const.}$$

Since $C(k)$ lies to the left of $\operatorname{Re}\sigma = -\nu(0)$, the contribution from the continuous spectrum is asymptotically small when compared with the point spectrum contribution. Therefore, for large times, we may neglect the integral term on the right of (51). On inverting the Fourier transforms and making use of the main result, we find

$$f(\mathbf{x}, \boldsymbol{\xi}, t) \sim \frac{\beta}{(2\pi)^3} \int_{-\infty}^{\infty} e^{-\alpha k^2 t + i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k}$$

and, from the evaluation given in (24),

$$f(\mathbf{x}, \boldsymbol{\xi}, t) \sim \beta e^{-r^2/4\alpha t}/(4\alpha\pi t)^{\frac{3}{2}}.$$

Finally, to calculate the constant β, we note that (47) leads to the continuity equation and, hence, the total number of particles at the initial instant is conserved. Carrying out the required integration, we find that

$\beta = 1$ and, therefore,

$$f(\mathbf{x}, \boldsymbol{\xi}, t) \sim e^{-r^2/4\alpha t}/(4\alpha\pi t)^{\frac{3}{2}}.$$

Problem 3: Consider the following 3 problems:

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2}, \quad w(t=0) = \delta(x), \tag{52a}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right) w = \delta(t)\delta(x), \tag{52b}$$

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} + \frac{\partial w}{\partial t} = 0, \quad \frac{\partial w}{\partial t}(t=0) = \delta(x), \quad w(t=0) = 0. \tag{52c}$$

Using transform techniques, we can easily analyze each of these and make them fall under the hypothesis leading to the main result. In fact, for $t \to \infty$, each problem leads to the same asymptotic result:

$$w = e^{-x^2/4t}/(4\pi t)^{\frac{1}{2}} + O^*(t^{-1}).$$

[Problem (52b), of course, should be considered in the complete (x, t) plane; for $t < 0$, however, the solution is exponentially small.]

Each of the problems (52) can, of course, be exactly solved; however, this is not the point. Equations (52) represent the three basic types of partial differential equations of second order. It is, of course, amusing that all three have the same asymptotic solution, but of more importance is the fact that the main result and the methods associated with it can be used independently of the type of partial differential equation.

ACKNOWLEDGMENT

The final version of the manuscript was prepared while the author was a Visiting Professor at the Institue Henri Poincaré of the University of Paris. He wishes to take this opportunity to express his gratitude for the invitation to spend the year at the Institute and for the gracious hospitality received during that year.

* This work was supported by the U.S. Office of Naval Research under Contract 562(39).

[1] We use the word "parameters" generically to refer both to parameters in the usual sense as well as to independent variables.

[2] A. Erdélyi, *Asymptotic Expansions* (Dover Publications, New York, 1956).

[3] E. T. Copson, *Asymptotic Expansions* (Cambridge University Press, London, 1965).

[4] H. Jeffrys, *Asymptotic Approximations* (Oxford University Press, London, 1962).

[5] D. S. Jones and M. Kline, J. Math. Phys. **37**, (1958).

[6] N. G. van Kampen, Physica **14**, 575 (1949).

[7] N. Bleistein and R. Handelsman, J. Math. Anal. Appl. **27**, 434 (1969).

[8] J. Focke, Berichte Sach. Akad. Wiss. Leipzig **101**, Heft 3 (1954).

[9] N. Chako, J. Inst. Math. Appl. **1**, 1965.

[10] L. Sirovich, Phys. Fluids **10**, 24 (1967).

[11] L. Sirovich, Phys. Fluids **11**, 1424 (1968).

[12] T. H. Chong and L. Sirovich, Phys. Fluids (to be published).

[13] L. Halabisky and L. Sirovich, Quart. Appl. Math. (to be published).

[14] E. P. Salathe and L. Sirovich, Phys. Fluids **10**, 1477 (1967).

[15] E. P. Salathe and L. Sirovich, J. Fluid Mech. **33**, 361 (1968).

[16] L. Sirovich, Phys. Fluids **6**, 10, 218, 1428 (1963).

[17] L. Sirovich and J. K. Thurber, Quart. Appl. Math. **25**, 175 (1967).

[18] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).

[19] K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (American Mathematical Society, Providence, R.I., 1965).

[20] See Eq. (35) in Sec. 4. K just represents a constant.

[21] Henceforth, $\delta > 0$ represents a small quantity, whose definition may change from equation to equation.

[22] The presence of infinite limits is of no importance, since we have already demonstrated that the contribution exterior to the neighborhood of the origin is asymptotically small. We can, in fact, consider ρ as having compact support at this point.

[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Government Printing Office, Washington, D.C., 1964), p. 447 #10.4.32.

[24] Reference 23, p. 448 #10.4.59.

[25] Reference 23, p. 376 #9.6.18.

[26] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), p. 139.

[27] An integration for N arbitrary is available, but is in infinite series form. This, however, does not display the essential character of $\mathfrak{J}$. See Y. L. Luke, *Integrals of Bessel Functions* (McGraw-Hill Book Co., New York, 1962), p. 313.

[28] Reference 23, p. 484 #11.3.20.

[29] Reference 26, pp. 403 and 408.

[30] Reference 26, p. 324.

[31] Reference 26, p. 331.

[32] Reference 26, p. 326.

[33] The δ-function data in this form is chosen only for the simplicity to which it leads. However, as is well known, the ensuing solution can be used to generate the solution to an arbitrary initial value problem.

[34] L. Sirovich and J. Thurber, J. Math. Phys. **8**, 888 (1967) (see Sec. 4).

[35] L. Sirovich and J. Thurber, J. Math. Phys. **10**, 239 (1969).

Generalized Hypergeometric Function of Unit Argument*

BARRY M. MINTON
Physics Department, University of Calgary, Calgary, Alberta, Canada

(Received 29 September 1969)

Two summation theorems are given for the terminating generalized hypergeometric function ${}_pF_{p-1}$, for arbitrary p, with certain restrictions on the parameters.

The generalized hypergeometric function of unit argument may be defined by the series

$$ {}_pF_q[a_1, a_2, \cdots, a_p; b_1, b_2, \cdots, b_q; 1] = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k (1)_k}, \quad (1) $$

where as usual

$$ (a)_k = \frac{\Gamma[a+k]}{\Gamma[a]} = \Gamma\begin{bmatrix} a+k \\ a \end{bmatrix}. \quad (2) $$

We are concerned with the case in which $q = p - 1$ when the series converges for

$$ \mathrm{Re}\left(\sum_{i=1}^{p} a_i\right) < \mathrm{Re}\left(\sum_{j=1}^{q} b_j\right) \quad (3) $$

or terminates when $a_i = -n$ for $i = 1, 2, \cdots, p$ and $n = 0, 1, \cdots$. It is assumed that $b_j \neq -n$ for $j = 1, 2, \cdots, q$ and $n = 0, 1, \cdots$, so that the function is defined.

This type of function often appears in solutions to problems in theoretical physics (for example, Racah coefficients are of this type with $p = 4$) and it would be of some interest if it were possible to sum the series in closed form in terms of say Γ-function products, thus simplifying the physical solution.

For $p = 2$, the series may be summed by Gauss' theorem for all values of the parameters, but even for $p = 3$ it has been summed only for parameters satisfying certain conditions, as in, say, the Saalschutz theorem.[1] There are a number of other special summation theorems for particular values of $p > 3$,[2] again involving restrictive conditions on the parameters, but as far as we know there have been no theorems given for arbitrary p, even with such restrictions on the parameters. We have so far been unable to obtain a generalization of the Saalschutz theorem for any p, which would be most useful for the physical problem, but give two other summation theorems for arbitrary p, one of which is very similar to the Saalschutzian form.

We first give the relation

$$ {}_{p+1}F_p[a_1, a_2, \cdots, a_{p+1}; b_1, b_2, \cdots, b_p; x] = \sum_{q=0}^{q=-a_p} \frac{(a_1)_q \cdots (a_p)_q (b_p - a_{p+1})_q}{(b_1)_q \cdots (b_p)_q (1)_q} (-x)^q \times {}_pF_{p-1}[a_1 + q, \cdots, a_p + q; b_1 + q, \cdots, b_{p-1} + q; x]. \quad (4) $$

We have assumed that a_p is a negative integer so that both sides terminate and the relation is defined for all x. The proof of Eq. (4) may be performed by expanding the ${}_pF_{p-1}$ function as a series in x, changing summation indices, and using the Gauss summation theorem, i.e.,

$$ {}_2F_1[a_1, a_2; b_1; 1] = \Gamma\begin{bmatrix} b_1, b_1 - a_1 - a_2 \\ b_1 - a_1, b_1 - a_2 \end{bmatrix}. \quad (5) $$

From Eq. (4) we may derive directly two relations for the case $p = 2$ and $x = 1$. First,

$$ {}_3F_2[a_1, a_2, a_3; b_1, b_2; 1] = \Gamma\begin{bmatrix} b_1, b_1 - a_1 - a_2 \\ b_1 - a_1, b_1 - a_2 \end{bmatrix} {}_3F_2[a_1, a_2, b_2 - a_3; b_2, a_1 + a_2 + 1 - b_1; 1] \quad (6) $$

which gives the Saalschutz theorem when $b_2 - a_3 = a_1 + a_2 + 1 - b_1$, i.e., for

$$ a_1 + a_2 + a_3 + 1 = b_1 + b_2, \quad (7) $$

$$ {}_3F_2[a_1, a_2, a_3; b_1, b_2; 1] = \Gamma\begin{bmatrix} b_1, b_1 - a_1 - a_2, b_2, b_2 - a_1 - a_2 \\ b_1 - a_1, b_1 - a_2, b_2 - a_1, b_2 - a_2 \end{bmatrix}. \quad (8) $$

Secondly,

$$ {}_3F_2[a_1, a_2, a_3; b_1, b_2; 1] = \Gamma\begin{bmatrix} b_1, b_2, s \\ a_3, s + a_1, s + a_2 \end{bmatrix} {}_3F_2[b_2 - a_3, b_1 - a_3, s; s - a_1, s - a_2; 1], \quad (9) $$

where $s = b_1 + b_2 - a_1 - a_2 - a_3$. This is the generalization of Dixon's theorem and, in particular, if

$b_2 - a_3$ is a negative integer so that the right-hand side terminates, we may, remembering that a_2 is also a negative integer, put $s = 0$ to give

$$ {}_3F_2[a_1, a_2, a_3; b_1, b_2; 1] = \Gamma\begin{bmatrix} 1 - a_1, 1 - a_2, 1 - a_3 \\ 1 - b_1, 1 - b_2 \end{bmatrix} \quad (10) $$

for the case where

$$ a_1 + a_2 + a_3 = b_1 + b_2. \quad (11) $$

The first theorem for arbitrary p is that, for a_0 negative integral and $a_0 < \sum_{i=1}^{p-1} b_i - a_i$, if $b_i - a_i$ is negative integral, $i = 1, 2, \cdots, p - 1$, then

$$ {}_{p+1}F_p[a_0, b, a_1, a_2, \cdots, a_{p-1}; b + 1, b_1, \cdots, b_{p-1}; 1] = \Gamma\begin{bmatrix} b + 1, 1 - a_0 \\ b + 1 - a_0 \end{bmatrix} \frac{(a_1)_{-b}(a_2)_{-b} \cdots (a_{p-1})_{-b}}{(b_1)_{-b}(b_2)_{-b} \cdots (b_{p-1})_{-b}}. \quad (12) $$

The proof is by induction using Eq. (4) with $x = 1$. On substituting from Eq. (12) into Eq. (4) and rearranging terms, we finally arrive at

$$ {}_{p+2}F_{p+1}[a_0, b, a_1, a_2, \cdots, a_{p-1}, a_p; b + 1, b_1, \cdots, b_{p-1}, b_p; 1] = \frac{(a_1)_{-b} \cdots (a_{p-1})_{-b}}{(b_1)_{-b} \cdots (b_{p-1})_{-b}} \Gamma\begin{bmatrix} b + 1, 1 - a_0 \\ b + 1 - a_0 \end{bmatrix} \times {}_2F_1[b, b_p - a_p; b_p; 1], $$

where we now require $b_p - a_p$ negative integral and $a_0 < \sum_{i=1}^{p} b_i - a_i$. The ${}_2F_1$ function then terminates for $b_p - a_p$ negative integral, and is summable by Eq. (5) to give

$$ {}_{p+2}F_{p+1}[a_0, b, a_1, a_2, \cdots, a_p; b + 1, b_1, \cdots, b_p; 1] = \Gamma\begin{bmatrix} b + 1, 1 - a_0 \\ b + 1 - a_0 \end{bmatrix} \frac{(a_1)_{-b} \cdots (a_p)_{-b}}{(b_1)_{-b} \cdots (b_p)_{-b}}. \quad (13) $$

This is now in the same form as Eq. (12) with the extra conditions on the parameters a_p and b_p. We have thus shown that, if the theorem is true for the ${}_{p+1}F_p[1]$ function, it is also true for the ${}_{p+2}F_{p+1}[1]$ function. In particular, for $p = 1$ we have, from Eq. (12),

$$ {}_2F_1[a_0, b; b + 1; 1] = \Gamma\begin{bmatrix} b + 1, 1 - a_0 \\ b + 1 - a_0 \end{bmatrix}, $$

which is simply the result given by Gauss' theorem. Knowing that the theorem is true for $p = 1$, we may then say that, from Eqs. (12) and (13), it is true by induction for all p.

The second theorem may be obtained from the first. Thus, writing Eq. (13), for example, in more symmetric form, with the previous restrictions on the parameters understood, we have

$$ {}_{p+2}F_{p+1}[a_0, b, a_1, \cdots, a_p; b + 1, b_1, \cdots, b_p; 1] = \frac{(1)_b(1 - b_1)_b(1 - b_2)_b \cdots (1 - b_p)_b}{(1 - a_0)_b(1 - a_1)_b \cdots (1 - a_p)_b}. $$

We may let $b \to \infty$ to give the second theorem:

$$ {}_{p+1}F_p[a_0, a_1, \cdots, a_p; b_1, \cdots, b_p; 1] = \lim_{b \to \infty} \frac{(1)_b(1 - b_1)_b \cdots (1 - b_p)_b}{(1 - a_0)_b(1 - a_1)_b \cdots (1 - a_p)_b} = \Gamma\begin{bmatrix} 1 - a_0, 1 - a_1, \cdots, 1 - a_p \\ 1 - b_1, 1 - b_2, \cdots, 1 - b_p \end{bmatrix}, \quad (14) $$

for

$$ a_0 + \sum_{i=1}^{p} a_i = \sum_{i=1}^{p} b_i. \quad (15) $$

When $p = 2$, this reduces to the already proved result (10) and (11). The result (14) is very similar to a generalization of the Saalschutz theorem, while Eq. (15) differs from the Saalschutzian condition by unity on the left-hand side.

* Supported in part by the National Research Council of Canada.

[1] E. D. Rainville, *Special Functions* (The Macmillan Co., New York, 1960), p. 87.

[2] L. J. Slater, *Generalised Hypergeometric Functions* (Cambridge University Press, Cambridge, England, 1966), Chap. 2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Strong Coupling Solution to Schrödinger Equation: The Mixing of States

L. M. GARRIDO AND L. NAVARRO
Instituto de Física Teórica and G.I.F.T., Barcelona, Spain

(Received 6 October 1969)

In this paper we give some ideas that can be useful to solve Schrödinger equations in the case when the Hamiltonian contains a large term. We obtain an expansion of the solution in reciprocal powers of the large coupling constant. The procedure followed consists in considering that the small part of the Hamiltonian engenders a motion adiabatic to the motion generated by the large part of the same.

1. INTRODUCTION

We define strong coupling, stating that the perturbation expansion is not valid. Then we pretend to obtain the solutions to the equations of motion in negative powers of the coupling constant g. Such an expansion has not been achieved so far and constitutes a crucial problem of elementary-particles dynamics. It is assumed nowadays that perturbation expansion is not valid; however, most of the calculation performed at the present time to discover the symmetry of elementary particles assume the validity of perturbation theory, an assumption that probably is not true.

We cannot state without discrimination that in the case of strong coupling we have to expand the solutions of the equation of motion into negative powers of the coupling constant. The coupling constant, being large, may be multiplied in the solution by factors that are small, which yields a small product. Thus, for strong coupling the general solution has to be expanded into negative powers of the coupling constant when multiplied by small factors. Most likely there is an intermediate region in which the solution has to be evaluated by variational methods, as it corresponds to intermediate effective coupling.

As is well known, it is hoped that, for the strong coupling approximation, the eigenstates of the large part of the Hamiltonian play the most important role. What we pretend in this paper is to find out, in powers of the reciprocal of the coupling constant, the mixing that the small part of the Hamiltonian produces among the eigenstates of the large part. The procedure followed consists in considering that the small part of the Hamiltonian engenders a motion adiabatic[1] to the motion generated by the large part of the same. This paper is no more than an introduction to the abovementioned problem.

Indeed, we do not solve the problem completely, since we do not obtain completely a series of negative powers of the coupling constant; but this paper is a first approximation to the desired solution.

The Schrödinger equation which is studied is

$$i\hbar \frac{\partial}{\partial t} |t\rangle = (H_0 + gH_1) |t\rangle,$$

where H_0 and H_1 are time independent, while the coupling constant g is large.

The solution to this equation is exactly

$$|t\rangle = \exp\left[\frac{t}{i\hbar}(H_0 + gH_1)\right] |0\rangle,$$

but, written in this way, we do not see how the eigenstates of H_1 are mixed amongst themselves. In what follows, we pretend to solve such a problem.

2. STRONG COUPLING SOLUTIONS TO SCHRÖDINGER EQUATION

Essentially, we want to solve the equation

$$i\hbar \frac{\partial}{\partial t} U(t) = (H_0 + gH_1)U(t)$$

for the unitary operator $U(t)$, since the time evolution of the state vector $|t\rangle$ at instant t is given by

$$|t\rangle = U(t) |0\rangle$$

when we suppose that $t = 0$ is the time origin. Correspondingly, we impose on $U(t)$ the following boundary condition:

$$U(0) = I.$$

The total Hamiltonian $H = H_0 + gH_1$ contains the dimensionless coupling constant g, which we suppose to be large. This implies that gH_1 gives a large contribution to the time derivative of the evolution operator $U(t)$, i.e., such an operator generates a fast time dependence of $U(t)$. Perturbation expansion in powers of g is not valid, since g is large.

We assume also, for the sake of concreteness, that H_0 and H_1 do not have an explicit time dependence. The study of the cases when H_0 and H_1 may have explicit time dependence yields much more complicated solutions and no better insight of the method used in this paper is gained. Besides, we should remember

that for most physical applications H_0 and H_1 do not depend on time explicitly, though they may have a dynamical time dependence; this is the case of quantum field theory.

Since the contribution of gH_1 to $U(t)$ is very large, it is sensible to suppose that the eigenvectors of H_1 play an important role in the expansion of $U(t)$ in powers of g^{-1}. This evident fact means that we will have difficulties in interpreting the physical meaning of the solution found, since,usually, free or undressed particles are the eigenvalues of the Hamiltonian H_0, which in this case has a much smaller contribution to the motion than gH_1.

Therefore, we use the eigenvectors of H_1 as the basis for the representation of the solution. So we assume that the following equation has been solved,

$$H_1 |E\rangle = E |E\rangle,$$

and that the eigenvectors $|E\rangle$ form a complete set in the sense that the identity operator is given by

$$I = \sum_E |E\rangle \langle E|.$$

Let A be any time-dependent operator. We can split it into a diagonal part A^D and a nondiagonal part A^N, in relation to the basis of our representation, the eigenvectors $|E\rangle$. So we have

$$A = A^D + A^N,$$

where

$$A^D \equiv \sum_E |E\rangle \langle E| A |E\rangle \langle E|,$$

$$A^N \equiv \sum_{E,E'} \{|E\rangle \langle E| A |E'\rangle \langle E'| (-\delta_{E,E'} + 1)\}.$$

It is very easy to check that diagonal parts of any two operators commute, while the corresponding nondiagonal parts generally do not commute; the product of any two diagonal parts of operators is a diagonal operator, while the product of two nondiagonal parts of operators may have diagonal and nondiagonal parts. We also have from the definitions above

$$\langle E| A^D |E\rangle = \langle E| A |E\rangle.$$

The diagonal part of the Hamiltonian H is

$$H^D = H_0^D + gH_1,$$

and its nondiagonal part is

$$H^N = H_0^N.$$

We solve the time equation for $U(t)$ exactly for H^D and treat H^N as the perturbation that mixes the eigenvectors of H_1, as we have already done in preceding papers.[2-4] Accordingly, we write

$$U(t) = S_D(t)S(t), \tag{1}$$

where $S_D(t)$ is a diagonal operator in our chosen representation, i.e., an exact solution of the following equation,

$$i\hbar \frac{\partial}{\partial t} S_D(t) = (H_0^D + gH_1)S_D(t),$$

and $S(t)$ is an operator that contains diagonal and nondiagonal parts in general and that we have to expand in negative powers of g.

We have to impose the conditions $S_D(0) = I$ and $S(0) = I$. We have

$$S_D(t) = \exp\left\{\frac{t}{i\hbar}(H_0^D + gH_1)\right\}, \tag{2}$$

the exact solution of the equation for $S_D(t)$. The equation for $S(t)$ is

$$i\hbar \frac{\partial}{\partial t} S(t) = H_0^N[t]S(t),$$

when, as usually occurs, the effective-perturbation Hamiltonian is defined by

$$H_0^N[t] = S_D^{-1}(t)H_0^N S_D(t),$$

which is well known since $S_D(t)$ has been obtained exactly and H_0^N can be evaluated. Then

$$S(t) = I + \frac{1}{i\hbar}\int_0^t dt' H_0^N[t'] + \frac{1}{(i\hbar)^2}\int_0^t dt' H_0^N[t']\int_0^{t'} dt'' H_0^N[t''] + \cdots, \tag{3}$$

as is well known. We have to show that, indeed, the terms of the expansion (3) as a series in powers of H_0^N contain negative powers of g only, when g is large.

We define

$$\begin{aligned}\bar{A}(t) &\equiv \frac{1}{i\hbar}\int_0^t dt' S_D^{-1}(t')A(t')S_D(t') \\ &= \frac{1}{i\hbar}\int_0^t dt' S_D^{-1}(t')A^N(t')S_D(t') \\ &\quad + \frac{1}{i\hbar}\int_0^t dt' S_D^{-1}(t')A^D(t')S_D(t') \\ &= S_D^{-1}(t)\,{}_\delta[A^N]S_D(t) + [A^D]_\delta,\end{aligned}$$

where

$${}_\delta[A^N] \equiv S_D(t)\left\{\frac{1}{i\hbar}\sum_{E,E'}\int_0^t dt'\left(\exp\frac{t}{i\hbar}[\langle E| H_0^D + gH_1 |E\rangle - \langle E'| H_0^D + gH_1 |E'\rangle]\right) \times |E\rangle \langle E| A^N(t') |E'\rangle \langle E'|\right\} S_D^{-1}(t) \tag{4}$$

and

$$[A^D]_\delta = \frac{1}{i\hbar}\int_0^t dt' A^D(t').$$

Integrating (4) by parts, we obtain

$$
\begin{aligned}
{}_{\delta}[A^N] = \sum_{E,E'} & (\langle E|\, H_0^D + gH_1\,|E\rangle \\
& - \langle E'|\, H_0^D + gH_1\,|E'\rangle)^{-1} \\
& \times \Bigg\{ S_D(t)\Bigg[\Bigg(\exp\frac{t}{i\hbar}\,[\langle E|\, H_0^D + gH_1\,|E\rangle \\
& - \langle E'|\, H_0^D + gH_1\,|E'\rangle]\Bigg) \\
& \times |E\rangle\langle E|\, A^N(t)\,|E'\rangle\langle E'|\Bigg]_0^t S_D^{-1}(t) \\
& - S_D(t)\Bigg[\int_0^t dt'\Bigg(\exp\frac{t'}{i\hbar}\,[\langle E|\, H_0^D + gH_1\,|E\rangle \\
& - \langle E'|\, H_0^D + gH_1\,|E'\rangle]\Bigg) \\
& \times |E\rangle\langle E|\,\frac{d}{dt'}\,A^N(t')\,|E'\rangle\langle E'|\Bigg] S_D^{-1}(t)\Bigg\} \\
= {} & O(1/g), \qquad (5)
\end{aligned}
$$

because we suppose, as usual, that $A(t)$ and its derivatives are bounded operators.

Now we can evaluate every term in (3):

$$
\begin{aligned}
I_1 &= \frac{1}{i\hbar}\int_0^t dt' H_0^N[t'] = \frac{1}{i\hbar}\int_0^t dt' S_D^{-1}(t') H_0^N S_D(t') \\
&= S_D^{-1}(t)\,{}_{\delta}[H_0^N] S_D(t) = O(1/g), \\
I_2 &= \frac{1}{(i\hbar)^2}\int_0^t dt' H_0^N[t']\int_0^{t'} dt'' H_0^N[t''] \\
&= \frac{1}{i\hbar}\int_0^t dt' S_D^{-1}(t') H_0^N\,{}_{\delta}[H_0^N] S_D(t') \\
&= S_D^{-1}(t)\,{}_{\delta}[H_0^N\,{}_{\delta}[H_0^N]] S_D(t) + \frac{1}{i\hbar}\int_0^t dt'[H_0^N\,{}_{\delta}[H_0^N]]^D.
\end{aligned}
$$

The first term in the expression of I_2 is of the order of g^{-2} and the second is of the order of g^{-3} because

$$\tau/h \simeq 1/g^2$$

by definition of the coupling constant g, where τ is approximately the time during which the interaction takes place. Similarly,

$$
\begin{aligned}
I_3 &= \frac{1}{i\hbar}\int_0^t H_0^N[t'] I_2(t') \\
&= S_D^{-1}(t)\,{}_{\delta}[H_0^N\,{}_{\delta}[H_0^N\,{}_{\delta}[H_0^N]]] S_D(t) \\
&\quad + \frac{1}{i\hbar}\int_0^t dt'[H_0^N\,{}_{\delta}[H_0^N\,{}_{\delta}[H_0^N]]]^D.
\end{aligned}
$$

Therefore, the degree of approximation of every term can be simply obtained by adding to the number of δ twice the number of integrals. The terms in I_3 are

$$O(1/g^3), \quad O(1/g^4).$$

Therefore, the unitary evolution operator in the second-order approximation, for instance, is

$$U(t) = \{I + {}_{\delta}[H_0^N] + {}_{\delta}[H_0^N\,{}_{\delta}[H_0^N]] + O(1/g^3)\} S_D(t). \qquad (6)$$

Evidently, it is quite easy to continue with the calculation of further terms in the expansion of $S(t)$. We see that successive terms contain higher and higher negative powers of the dimensionless coupling constant g.

A very interesting case is that in which both H_0 and H_1 can be simultaneously diagonalized. Then our procedure results are considerably simplified (the same occurs in the ordinary perturbation theory). The expression for $U(t)$ can be reduced to

$$U(t) = S_D(t).$$

Then,in this special case, it is not possible to obtain the evolution operator as an expansion in reciprocal powers of the large coupling constant using this procedure. But when this occurs, the problem is easier than usual and frequently can be solved, as in Example I in the next section.

We should remark that it is essential for the validity of this expansion that the difference

$$\langle E'|\, H_0^D + gH_1\,|E'\rangle - \langle E|\, H_0^D + gH_1\,|E\rangle$$

be large for any two values of E and E'. If it were small, the method would not be acceptable, since the denominators in ${}_{\delta}[A]$ would not be large.

The time dependence of $U(t)$ is very fast, since the exponent of $S_D(t)$ contains a constant without dimensions which is large. But $S(t)$ will not necessarily vary rapidly with time. We have here two time scales: the fast one generated by H^D and the slow one generated by H^N. We separate the two time dependences, fast and slow, to solve the problem considering the slow time dependence as producing a small perturbation compared to the fast time variation. Both motions are adiabatic to each other. The method is similar to those used by us in preceding papers to treat adiabatic motions.

3. EXAMPLES

A. Example I

Consider two identical linear oscillators with spring constant k and an interaction potential given by gx_1x_2, where x_1 and x_2 are the oscillator variables, and g is large. The total Hamiltonian is

$$\mathcal{H} = \mathcal{H}_0 + g\mathcal{H}_1,$$

where

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right) + \frac{k}{2}(x_1^2 + x_2^2)$$

and

$$\mathcal{H}_1 = x_1 x_2 .$$

The energy levels of $\mathcal{H}_0$ are thus given by

$$E_0 = (n_1 + \tfrac{1}{2})\hbar\omega + (n_2 + \tfrac{1}{2})\hbar\omega,$$

where

$$\omega^2 = k/m.$$

As we really have a two-dimensional problem, it is advantageous to use a $\mathcal{N} \otimes \mathcal{N}$ representation, $\mathcal{N}$ being the occupation number of each oscillator. Then the unperturbated eigenfunctions are

$$|n_1 n_2\rangle = n_1 |1.0\rangle + n_2 |0.1\rangle.$$

The Hamiltonian matrix of the unperturbed motion is

$$\mathcal{H}_0 = \tfrac{3}{2}\hbar\omega \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the Hamiltonian matrix of the perturbation, taken with respect to the unperturbed states, is

$$\mathcal{H}_1 = \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix},$$

where

$$a = \langle 1.0| \mathcal{H}_1 |0.1\rangle = g \langle 0| x_1 |1\rangle \langle 1| x_2 |0\rangle = \hbar/2m\omega,$$

as may be deduced from the quantum-mechanical version of the virial theorem.

In order to apply the approximation given by (6), it is necessary to make a transformation of state vector bases to achieve a new representation in which $\mathcal{H}_1$ is diagonal. This matrix transformation is

$$A = \frac{1}{i\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then

$$H_0 = A^{-1}\mathcal{H}_0 A = \tfrac{3}{2}\hbar\omega \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$H_1 = A^{-1}\mathcal{H}_1 A = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix},$$

and

$$_{\delta}[H_0] = S_D(t)\left[\int_0^t dt' S_D^{-1}(t') H_0^N S_D(t')\right] S_D^{-1}(t) = 0.$$

Formula (5) gives

$$U(t) = S_D(t) = \exp\left[-\frac{i}{\hbar} t(H_0 + gH_1)\right],$$

or, in the original representation,

$$U(t) = \exp\left[-\frac{i}{\hbar} t(\mathcal{H}_0 + g\mathcal{H}_1)\right].$$

In this case we have not achieved an expansion in powers of g^{-1}. But, owing to the simultaneous diagonalization of $\mathcal{H}_0$ and $\mathcal{H}_1$, the problem can be easily solved. Defining new variables u and v by

$$x_1 = (u + v)/2, \quad x_2 = (u - v)/2,$$

we express the Hamiltonian as

$$H = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right) + \tfrac{1}{2}(k + g)u^2 + \tfrac{1}{2}(k - g)v^2.$$

The exact energy levels are thus given by

$$E = n_1\hbar\omega_1 + n_2\hbar\omega_2 + (\hbar/2)(\omega_1 + \omega_2),$$

where n_1 and n_2 are positive integers and

$$\omega_1^2 = (k - g)/m, \quad \omega_2^2 = (k + g)/m,$$

and the eigenfunctions are

$$|n_1 n_2\rangle = F_{n_1}(u) F_{n_2}(v),$$

where the F's are simple harmonic oscillator wavefunctions.

Actually, this problem has nothing to do with the procedure that we introduced in Sec. 2. It is easy enough to be solved with a simple change of the variables.

We have included it for two reasons: First, it serves to introduce the notation that we are going to employ with some further complications in the next example, which is the one that demonstrates the validity of our method. Secondly, it represents the special case in which $\mathcal{H}_0$ and $\mathcal{H}_1$ can be diagonalized simultaneously. In this case, our procedure is not available, but then we have shown that the problem is easy to solve, as we anticipated in Sec. 2.

B. Example II

In the foregoing example, it has not been possible to check the validity of the approximation suggested by formula (6). In order to verify it, we consider two different oscillators strongly coupled. In this case,

$$\mathcal{H}_0 = \tfrac{3}{2}\hbar \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix},$$

$$\mathcal{H}_1 = \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix},$$

where

$$a = \langle 1.0| \mathcal{H}_1 |0.1\rangle = \hbar(2m)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}.$$

The diagonalization of $\mathcal{H}_1$ as in the preceding example can be done by transforming both Hamiltonians with the matrix

$$A = \frac{1}{i\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The new Hamiltonians are

$$H_0 = A^{-1}\mathcal{H}_0 A = \tfrac{3}{4}\hbar\begin{pmatrix}\omega_1+\omega_2 & \omega_2-\omega_1\\ \omega_2-\omega_1 & \omega_1+\omega_2\end{pmatrix},$$

$$H_1 = A^{-1}\mathcal{H}_1 A = \begin{pmatrix}a & 0\\ 0 & -a\end{pmatrix} = \hbar(2m)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

$$H = \frac{\hbar}{2}\begin{pmatrix}\frac{3}{2}(\omega_1+\omega_2)+g(m)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}} & \frac{3}{2}(\omega_2-\omega_1)\\ \frac{3}{2}(\omega_2-\omega_1) & \frac{3}{2}(\omega_1+\omega_2)-g(m)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}\end{pmatrix}.$$

The first-order approximation of formula (6) is given by

$$U(t) = (1 + {}_\delta[H_0^N])S_D(t),$$

where

$$S_D(t) = \exp(t/i\hbar)(H_0^D + gH_1)$$

and

$$H_0^D = \tfrac{3}{4}\hbar\begin{pmatrix}\omega_1+\omega_2 & 0\\ 0 & \omega_1+\omega_2\end{pmatrix}.$$

The validity of the method can be checked by evaluating H_0^N and verifying that it is of the order of g^{-1}. Now we may easily calculate

$$S_D^{-1}(t)H_0^N S_D(t) = \tfrac{3}{4}\hbar(\omega_2-\omega_1)\begin{pmatrix}0 & \exp[-gt(im)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}]\\ \exp[gt(im)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}] & 0\end{pmatrix}.$$

Integration yields

$$\begin{aligned}{}_\delta[H_0^N] &= S_D(t)\left[\int_0^t dt' S_D^{-1}(t')H_0^N S_D(t')\right]S_D^{-1}(t)\\ &= \tfrac{3}{4}\hbar(\omega_2-\omega_1)\frac{im\sqrt{\omega_1\omega_2}}{g}\begin{pmatrix}0 & 1-\exp[gt(im)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}]\\ 1-\exp[-gt(im)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}] & 0\end{pmatrix},\end{aligned}$$

in accordance with (5).

The above expression is of the order of $1/g$ because $\omega_2 \neq \omega_1$. Thus,

$$U(t) = \left\{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + \tfrac{3}{4}\hbar(\omega_2-\omega_1)\frac{im\sqrt{\omega_1\omega_2}}{g}\begin{pmatrix}0 & 1-\exp[gt(im)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}]\\ 1-\exp[-gt(im)^{-1}(\omega_1\omega_2)^{-\frac{1}{2}}] & 0\end{pmatrix}\right\}S_D. \tag{7}$$

Expression (7) allows us to solve this problem in the first order and prove our assertion about the non-diagonal character of ${}_\delta[H_0^N]$.

4. CONCLUSION

We have found how the eigenvalues of H_1 mix. We have found an expression of $S(t)$ that, expanded in powers of g^{-1}, is the term that indicates how the eigenvalues of H_1 mix.

The process developed in this paper will be applied in a future paper to evaluate the mass renormalization of a nucleon coupled to a cloud of mesons when the coupling constant is large.

[1] A. Lenard, Ann. Phys. **6**, 261 (1959).
[2] L. M. Garrido, J. Math. Phys. **5**, 335 (1964).
[3] L. Navarro and L. M. Garrido, Proc. Phys. Soc., Ser. 2,**1**, 326 (1968).
[4] L. Navarro and L. M. Garrido, Physica **39**, 361 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Spherical Symmetry and Mass-Energy in General Relativity. I. General Theory

MICHAEL E. CAHILL AND G. C. MCVITTIE
University of Illinois Observatory, Urbana, Illinois 61801

(Received 10 July 1969)

The mass-energy of spherically symmetric distributions of material is studied according to general relativity. An arbitrary orthogonal coordinate system is used whenever invariant properties are discussed. The Bianchi identity is shown to imply that the Misner-Sharp-Hernandez mass function is an integral of two combinations of Einstein's equations for any energy-momentum tensor and that mass-energy flow is conservative. The two mass equations thus found and the mass function provide a technique for casting Einstein's field equations into alternative forms. This mass-function technique is applied to the general problem of the motion of a perfect fluid and especially to the examination of negative-mass shells and their relation to singular behavior. The technique is then specialized to the study of a known class of solutions of Einstein's equations for a perfect fluid and to a brief treatment of uniform model universes and the charged point-mass solution.

1. INTRODUCTION

The physical situation to be dealt with in this paper is the motion of a self-gravitating spherically symmetric distribution of material. By spherical symmetry is meant that an observer at the center of the distribution will discern at any instant the same physical picture regardless of his orientation, which makes two of the coordinates, θ and φ, cyclic in character.

Much of the treatment is devised so that forces other than gravitation may be present, but particular attention is paid to the motion of a perfect fluid. The material may have discontinuities, in which case the external region is assumed to be empty. Einstein's general theory of relativity is assumed to provide an accurate description of the given physical situation.

Several problems in this paper are examined in an arbitrary orthogonal spherically symmetric coordinate system. It is possible to find an integral of Einstein's field equations which represents the total quantity of mass-energy enclosed by the 2-space of equivalent points passing through an arbitrary point in space-time. This mass function is shown to be a scalar invariant. The Bianchi identity is used to demonstrate that both the space and time derivatives of the mass function are related to components of the energy-momentum tensor and these equations are called the mass equations. The mass function may be used to construct a vector which represents the flow of mass-energy, and this vector has vanishing divergence. The mass equations and the mass function allow Einstein's equations to be put in an alternative and sometimes more transparent form. The case of a perfect fluid is considered with the aid of the mass function, as is that of fitting a solution to a Schwarzschild exterior.

Another class of problems explored involves the motion of a perfect fluid as described by a comoving coordinate system. Negative-mass shells, which are regions that produce a negative contribution to the total mass-energy content of the material, are examined and are shown to imply singular behavior under a certain restriction. The mass function technique is used to elucidate, simplify, and generalize the solutions of McVittie.[1]

The last series of problems involves well-known solutions. The mass function approach is applied to the uniform model universe solutions and negative-mass shells are found in the closed solutions. The charged point-mass solution is then considered, and the problem of a zero-pressure charged fluid is briefly treated.

This work is related to that of several other investigators. Hernandez and Misner[2] have found the expression for the mass-energy of a perfect fluid in spherical symmetry by a different technique. They have also dealt with the problem of the occurrence of gravitational-collapse-type singularities when negative-mass shells arise in imploding perfect fluids. Misner and Sharp[3] have formulated and examined Einstein's field equations of a perfect fluid in a comoving coordinate system and have fitted an arbitrary solution of them to an empty exterior. May and White[4] have integrated the field equations by computer for a perfect fluid undergoing adiabatic collapse using formulas similar to those of Misner and Sharp. They also encounter situations in which the mass-energy does not increase monotonically with the comoving radial coordinate and show that where this happens a negative-mass shell occurs. McVittie[1] has developed a class of solutions of Einstein's equations. Thompson and Whitrow[5] have generalized McVittie's class of solutions by an alternative method and have investigated the uniform density case. Lastly, Taub[6] has examined this class of

solutions, and paid particular attention to their thermodynamic properties.

2. BASIC FORMULAS FOR SPHERICAL SYMMETRY

In this section we present important formulas for an arbitrary spherically symmetric space-time. The metric is

$$ds^2 = e^{2\gamma}(dx^4)^2 - c^{-2}[e^{2\alpha}(dx^1)^2 + r^2\, d\Omega^2], \quad (2.1)$$

where γ, α, and r are functions of x^1 and x^4, and

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\varphi^2. \quad (2.2)$$

The center of symmetry is defined by

$$x^1 = 0, \quad (2.3)$$

and so

$$r(0, x^4) = 0. \quad (2.4)$$

It is also assumed that

$$r(x^1, x^4) > 0, \quad (2.5)$$

for all $x^1 \neq 0$, that

$$e^{2\alpha} > 0, \quad e^{2\gamma} > 0, \quad (2.6)$$

for all x^1 and x^4, and that α and γ are bounded unless otherwise stated.

Throughout this paper, numerical subscripts on lower case letters refer to partial or ordinary derivatives as the context requires, while a subscript on a capital letter is a covariant index.

The nonvanishing Christoffel symbols are

$$\begin{aligned}
&\Gamma^1_{11} = \alpha_1, && \Gamma^2_{12} = r_1/r, && \Gamma^3_{13} = r_1/r, && \Gamma^4_{11} = e^{2(\alpha-\gamma)}\alpha_4/c^2,\\
&\Gamma^1_{14} = \alpha_4, && \Gamma^2_{24} = r_4/r, && \Gamma^3_{23} = \cot\theta, && \Gamma^4_{14} = \gamma_1,\\
&\Gamma^1_{22} = -e^{-2\alpha}rr_1, && \Gamma^2_{33} = -\sin\theta\cos\theta, && \Gamma^3_{34} = r_4/r, && \Gamma^4_{22} = e^{-2\gamma}rr_4/c^2,\\
&\Gamma^1_{33} = \sin^2\theta\Gamma^1_{22}, && && && \Gamma^4_{33} = \sin^2\theta\Gamma^4_{22},\\
&\Gamma^1_{44} = c^2e^{2(\gamma-\alpha)}\gamma_1, && && && \Gamma^4_{44} = \gamma_4.
\end{aligned} \quad (2.7)$$

The Riemann–Christoffel tensor, defined by

$$R^\sigma_{\lambda\mu\nu} = \frac{\partial\Gamma^\sigma_{\lambda\nu}}{\partial x^\mu} - \frac{\partial\Gamma^\sigma_{\lambda\mu}}{\partial x^\nu} + \Gamma^\tau_{\lambda\nu}\Gamma^\sigma_{\mu\tau} - \Gamma^\tau_{\lambda\mu}\Gamma^\sigma_{\tau\nu},$$

has the following nonzero components:

$$\begin{aligned}
R_{2323} &= -r^2\sin^2\theta(1 + e^{-2\gamma}r_4^2/c^2 - e^{-2\alpha}r_1^2)/c^2,\\
R_{1212} &= R_{3131}/\sin^2\theta = r(r_{11} - \alpha_1 r_1 - e^{2(\alpha-\gamma)}\alpha_4 r_4/c^2)/c^2,\\
R_{1224} &= R_{1334}/\sin^2\theta = -r(r_{14} - \alpha_4 r_1 - \gamma_1 r_4)/c^2,\\
R_{1414} &= e^{2\alpha}(\alpha_{44} + \alpha_4^2 - \alpha_4\gamma_4)/c^2 - e^{2\gamma}(\gamma_{11} + \gamma_1^2 - \alpha_1\gamma_1),\\
R_{2424} &= R_{3434}/\sin^2\theta = r(r_{44} - \gamma_4 r_4 - c^2e^{2(\gamma-\alpha)}\gamma_1 r_1)/c^2.
\end{aligned} \quad (2.8)$$

The components of the Einstein tensor may now be found by

$$G_{\mu\nu} = R^\sigma_{\mu\nu\sigma} - \tfrac{1}{2}g_{\mu\nu}g^{\alpha\beta}R^\sigma_{\alpha\beta\sigma}.$$

Thus,

$$\begin{aligned}
G^1_1 &= -g^{22}(R^3_{223} + 2R^4_{224}),\\
G^2_2 = G^3_3 &= -g^{22}(R^1_{221} + R^4_{224}) - g^{11}g^{44}R_{4114},\\
G^4_4 &= -g^{22}(R^3_{223} + 2R^1_{221}),\\
G^1_4 &= g^{11}g_{44}G^4_1 = 2g^{22}R^1_{224}.
\end{aligned} \quad (2.9)$$

Upon evaluation, Eqs. (2.9) become

$$\begin{aligned}
G^1_1 &= -e^{-2\gamma}\left(2\frac{r_{44}}{r} + \frac{r_4^2}{r^2} - 2\gamma_4\frac{r_4}{r}\right)\\
&\quad - \frac{c^2}{r^2} + c^2e^{-2\alpha}\left(\frac{r_1^2}{r^2} + 2\gamma_1\frac{r_1}{r}\right),\\
G^2_2 = G^3_3 &= -e^{-2\gamma}\left(\alpha_{44} + \alpha_4^2 - \alpha_4\gamma_4\right.\\
&\quad \left. + \frac{r_{44}}{r} + \alpha_4\frac{r_4}{r} - \gamma_4\frac{r_4}{r}\right)\\
&\quad + c^2e^{-2\alpha}\left(\gamma_{11} + \gamma_1^2 - \alpha_1\gamma_1\right.\\
&\quad \left. + \frac{r_{11}}{r} + \gamma_1\frac{r_1}{r} - \alpha_1\frac{r_1}{r}\right),\\
G^4_4 &= -e^{-2\gamma}\left(\frac{r_4^2}{r^2} + 2\alpha_4\frac{r_4}{r}\right) - \frac{c^2}{r^2}\\
&\quad + c^2e^{-2\alpha}\left(2\frac{r_{11}}{r} + \frac{r_1^2}{r^2} - 2\alpha_1\frac{r_1}{r}\right),
\end{aligned} \quad (2.10)$$

$$e^{2\gamma}G^4_1 = -e^{2\alpha}G^1_4/c^2 = 2(r_{14} - \alpha_4 r_1 - \gamma_1 r_4)/r.$$

The Einstein tensor has vanishing vectorial divergence, and hence obeys

$$G^\mu_{\nu,\mu} = 0, \quad (2.11)$$

and this equation gives rise to two nontrivial equations which are

$$\frac{\partial G_1^1}{\partial x^1} + \frac{\partial G_1^4}{\partial x^4} + \left(\gamma_4 + \alpha_4 + 2\frac{r_4}{r}\right)G_1^4 = \gamma_1(G_4^4 - G_1^1) + 2(G_2^2 - G_1^1)\frac{r_1}{r}, \quad (2.12)$$

$$\frac{\partial G_4^1}{\partial x^1} + \frac{\partial G_4^4}{\partial x^4} + \left(\alpha_1 + \gamma_1 + 2\frac{r_1}{r}\right)G_4^1 = \alpha_4(G_1^1 - G_4^4) + 2(G_2^2 - G_4^4)\frac{r_4}{r}. \quad (2.13)$$

Finally, Einstein's field equations are

$$-8\pi\mathfrak{G}T_\nu^\mu = G_\nu^\mu. \quad (2.14)$$

Formulas (2.7), (2.8), and (2.10) are slightly modified versions of those developed by Synge.[7]

3. THE MASS FUNCTION

In this section the meaning and properties of R^3_{232} are studied. This function, by Eq. (2.8), is

$$R^3_{232} = (1 + e^{-2\gamma}r_4^2/c^2 - e^{-2\alpha}r_1^2). \quad (3.1)$$

This component of the Riemann–Christoffel tensor is of interest because it involves only first derivatives and because it is invariant under transformations of the form $\bar{x}^1 = \bar{x}^1(x^1, x^4)$, $\bar{x}^4 = \bar{x}^4(x^1, x^4)$. Examination of (3.1) reveals that if r is used as a coordinate and t as a conjugate orthogonal coordinate to r, then in this coordinate system the metric is

$$ds^2 = e^{2\psi}\,dt^2 - \frac{1}{c^2}\left[\frac{dr^2}{(1 - R^3_{232})} + r^2\,d\Omega^2\right]. \quad (3.2)$$

Because r is the curvature of the 2-space of symmetry passing through $(x^1, \theta, \varphi, x^4)$, it is called the curvature coordinate. It is spacelike in character when R^3_{232} is less than unity, but timelike when R^3_{232} exceeds unity. Thus, in this coordinate system, $e^{2\alpha}$ could diverge. If curvature coordinates are used in an empty region of spherically symmetric space–time, the Schwarzschild metric results wherein

$$ds^2 = (1 - 2M/r)\,dt^2 - \frac{1}{c^2}\left[\frac{dr^2}{(1 - 2M/r)} + r^2\,d\Omega^2\right], \quad (3.3)$$

and M is a constant.

Consider next an arbitrary spherical distribution of material which is surrounded by empty space. In such a case, the boundary of the material is a surface of discontinuity. Israel[8] has shown that the theory of surfaces of discontinuity due to Lichnerowicz[9] implies that in curvature coordinates the metric tensor must be continuous across a surface of discontinuity. Comparison of (3.2) with (3.3) then reveals that

$$(R^3_{232})_b = 2M/r_b,$$

where subscript "b" means evaluation at the boundary. This suggests that a function $m(x^1, x^4)$ be defined by

$$m = \tfrac{1}{2}rR^3_{232} = \tfrac{1}{2}r(1 + e^{-2\gamma}r_4^2/c^2 - e^{-2\alpha}r_1^2), \quad (3.4)$$

and because of the invariance of R^3_{232} and r, m is an invariant also. The function m may be tentatively defined as the total amount of mass–energy entrapped between the center of the distribution and the 2-space of symmetry passing through the point $(x^1, \theta, \varphi, x^4)$. Equation (3.4) was originally developed by Hernandez and Misner,[2] who also proved the invariance of m by a different method. The mass function possesses the important property that the only derivatives present in it are those of the first order and they occur to the second power. In the following, it is shown that the above tentative identification is a reasonable one.

The mass function is an integral of Einstein's equations. This may be demonstrated by examination of the Bianchi identity

$$R^3_{232,\sigma} + R^3_{22\sigma,3} + R^3_{2\sigma3,2} = 0, \quad (3.5)$$

where σ is 1 or 4, and where a comma indicates covariant differentiation. Equations (2.7) and (2.9) may be used to evaluate the terms in (3.5) and, with the identity

$$\Gamma^2_{2\lambda}R^\lambda_{2\sigma2} = \Gamma^\lambda_{22}R^3_{\lambda\sigma3},$$

Eq. (3.5) becomes

$$m_\sigma = \tfrac{1}{2}(R^3_{232}r_\sigma + 2\Gamma^2_{2\lambda}rR^\lambda_{2\sigma2}). \quad (3.6)$$

In view of Eqs. (2.7) and (2.9), the last equation is expressible as

$$m_1 = r^2(G_1^4r_4 - G_4^4r_1)/2c^2, \quad (3.7)$$

$$m_4 = r^2(G_4^1r_1 - G_1^1r_4)/2c^2. \quad (3.8)$$

The field equations (2.14) then imply that

$$m_1 = 4\pi\mathfrak{G}r^2(T_4^4r_1 - T_1^4r_4)/c^2, \quad (3.9)$$

$$m_4 = 4\pi\mathfrak{G}r^2(T_1^1r_4 - T_4^1r_1)/c^2. \quad (3.10)$$

Equations (3.9) and (3.10) are the mass equations, and imply that the mass function is an integral of two combinations of Einstein's equations. Moreover, they are equivalent to two of Einstein's equations. Also, the form of these equations clearly indicates that m is a constant in empty space. Finally, (3.9) may be regarded as a source equation for mass–energy.

The mass function may also be used to determine

another equation which gives the second partial derivative of r with respect to x^4. If the right-hand side of (3.4) is differentiated with respect to x^4, the result must equal the right-hand side of (3.10). Hence,

$$4\pi\mathfrak{G}r^2(T_1^1 r_4 - T_4^1 r_1)/c^2 = mr_4/r + re^{-2\alpha}(\alpha_4 r_1^2 - r_1 r_{14}) + re^{-\gamma}r_4(e^{-\gamma}r_4)_4/c^2.$$

With the aid of Eq. (2.10) and the field equations, r_{14} may be eliminated with the result that

$$e^{-\gamma}(e^{-\gamma}r_4)_4 = 4\pi\mathfrak{G}rT_1^1 - mc^2/r^2 + c^2e^{-2\alpha}r_1\gamma_1. \quad (3.11)$$

This equation is the relativistic analog of Newton's law of gravitation. The left-hand side corresponds to the acceleration, while the right-hand side contains the familiar inverse square attraction, a stress term and a term which involves γ_1. Equation (3.11) was originally derived for the case of a perfect fluid, in the form given below as (4.13), by Misner and Sharp.[3]

Mass–energy is conserved in spherically symmetric space–times. This may be shown by defining a mass flow vector J^σ by

$$J^\sigma = \sin\theta(-m_4, 0, 0, m_1)/[4\pi(-g)^{\frac{1}{2}}]. \quad (3.12)$$

Its vector character follows from the tensor character of the Levi-Cività density $\epsilon^{\alpha\beta\gamma\delta} = \delta^{\alpha\beta\gamma\delta}_{1234}(-g)^{-\frac{1}{2}}$, where $\delta^{\alpha\beta\gamma\delta}_{1234}$ is the generalized Kronecker δ function, and of

$$\Theta_{\mu\nu} = \sin\theta(\delta^3_\mu\delta^2_\nu - \delta^2_\mu\delta^3_\nu),$$

which was introduced by Regge and Wheeler.[10] Thus, it is seen that

$$J^\sigma = \epsilon^{\sigma\alpha\beta\gamma}\Theta_{\alpha\beta}m_\gamma/8\pi. \quad (3.13)$$

The reason why J^2 and J^3 vanish is that no mass–energy may flow in these directions. The radial component J^1 gives the rate at which mass–energy is flowing out of the coordinate sphere $x^1 = \text{const}$ at any fixed x^4. The J^4 component describes the distribution of mass–energy in the radial direction. The conservation of mass–energy follows immediately from (3.12) for

$$J^\sigma_{,\sigma} = 0. \quad (3.14)$$

The mass-flow vector is orthogonal to m_σ, because

$$J^\sigma m_\sigma = 0. \quad (3.15)$$

This is to be expected, for the direction of mass–energy flow should be tangential to the surface $m = \text{const}$. Thorne[11] has also developed Eqs. (3.12) and (3.14) by an alternative method. The Bianchi identity, which ensures that there is local conservation of mass–energy, has thus been shown to be involved in the construction of the mass function, which represents a globally conserved quantity.

4. SPHERICALLY SYMMETRIC PERFECT FLUID

In this section some of the familiar formulas for the motion of a perfect fluid are developed from the mass function approach. The energy–momentum tensor for such a fluid is

$$T^\mu_\nu = (\rho + p/c^2)U^\mu U_\nu - \delta^\mu_\nu p/c^2, \quad (4.1)$$

where ρ and p are the internal energy density and pressure as measured by a local comoving observer and U^μ is the 4-velocity of the fluid which must be a unit vector and hence satisfy

$$U^\mu U_\mu = 1. \quad (4.2)$$

Einstein's field equations (2.14) and the form of G^μ_ν in (2.10) imply that U^2 and U^3 vanish. The field equations and (2.11) imply that

$$T^\mu_{\nu,\mu} = 0, \quad (4.3)$$

which expresses the local conservation of energy and momentum. All the components of the energy–momentum tensor are not independent and (4.1) may be used to show that

$$T_4^1 T_1^4 = (T_2^2 - T_1^1)(T_2^2 - T_4^4), \quad (4.4)$$

and so

$$G_4^1 G_1^4 = (G_2^2 - G_1^1)(G_2^2 - G_4^4), \quad (4.5)$$

which is called a consistency relation by McVittie.[12] In order to specify a particular coordinate system, the four coordinates (x^1, x^2, x^3, x^4) must be defined. Two definitions are $x^2 = \theta$ and $x^3 = \varphi$, and a third is implied by $g_{14} = 0$. There is, therefore, one further specification available. Thus, in an arbitrary spherically symmetric coordinate system, the functions α, γ, r, m, ρ, p, U^1, and U^4 are determined by four statements defining the coordinates, one of which is arbitrary, by the mass function equation (3.4), by the two mass equations for the derivatives of m, Eqs. (3.9) and (3.10), by the dynamical equation (3.11), by the condition that U^μ be a unit vector (4.2), by the consistency equation (4.5) or some equation equivalent to it, and lastly, by an equation of state or some condition that permits the integration of the entire set of equations.

If the motion of the fluid is such that a comoving coordinate system may be used, then T_4^1, which is proportional to U^1, must vanish. Thus, by means of the field equations, the nonvanishing components of the energy–momentum tensor are

$$T_1^1 = T_2^2 = T_3^3 = -p/c^2, \quad T_4^4 = \rho. \quad (4.6)$$

Equation (4.2) is then $U^4U_4 = 1$. Because the vanishing of T_4^1 is equivalent to the vanishing of U^1, the last

specification for defining the coordinate system is

$$r_{14} = \gamma_1 r_4 + \alpha_4 r_1. \tag{4.7}$$

The mass function equation is still (3.4), namely,

$$m = \tfrac{1}{2} r(1 + e^{-2\gamma} r_4^2/c^2 - e^{-2\alpha} r_1^2), \tag{4.8}$$

while the mass equations (3.9) and (3.10) now become

$$m_1 = 4\pi \mathfrak{G} \rho r^2 r_1/c^2, \tag{4.9}$$

$$m_4 = -4\pi \mathfrak{G} p r^2 r_4/c^4. \tag{4.10}$$

The consistency relation is $G_1^1 = G_2^2$ but this is equivalent to either of the nontrivial local conservation equations (4.3). By means of the field equations (2.12) and (2.13) and the form of T_ν^μ, they are

$$p_1/c^2 = -\gamma_1(\rho + p/c^2), \tag{4.11}$$

$$\rho_4 = -(2r_4/r + \alpha_4)(\rho + p/c^2). \tag{4.12}$$

The dynamical equation (3.11) takes the form

$$e^{-\gamma}(e^{-\gamma} r_4)_4 = -(4\pi \mathfrak{G} r p/c^2 + mc^2/r^2) + c^2 e^{-2\alpha} r_1 \gamma_1. \tag{4.13}$$

Equations (4.8)–(4.10) and (4.13) were discovered by Misner and Sharp[3] and Bardeen.[13] Equation (4.13) was also examined by May and White.[4] Since it is merely necessary to determine $(\alpha, \gamma, r, m, \rho, p)$ in comoving coordinates, only five of the equations, (4.7)–(4.13), are independent.

The mass function may be used to show that the boundary pressure of a perfect fluid surrounded by empty space must vanish. If the equation of the boundary is

$$f(x^1, x^4) = 0,$$

then

$$\left(\frac{dm}{dx^4}\right)_b = \left(-m_1 \frac{f_4}{f_1} + m_4\right)_b \tag{4.14}$$

and

$$(U^1 f_1 + U^4 f_4)_b = 0. \tag{4.15}$$

By means of Eqs. (3.9), (3.10), and (4.1), Eq. (4.14) becomes

$$\left(\frac{dm}{dx^4}\right)_b = \frac{4\pi \mathfrak{G} r_b^2}{c^2 (f_1)_b} [\rho(U_1 r_4 - U_4 r_1)(U^1 f_1 + U^4 f_4) + (p/c^2)(U^1 r_1 + U^4 r_4)(U_1 f_4 - U_4 f_1)]_b. \tag{4.16}$$

Because the mass function is an invariant and depends only on first derivatives, the Lichnerowicz continuity conditions require that it be continuous across the boundary of the material. Thus $(dm/dx^4)_b$ must vanish. The first term on the right-hand side of (4.16) vanishes because of (4.15). The second term can vanish only if

$$p_b = 0. \tag{4.17}$$

5. NEGATIVE-MASS SHELLS

Negative-mass shells were first encountered by May and White.[4] A negative-mass shell occurs in an orthogonal comoving coordinate system if m_1 becomes negative during the motion of the matter distribution. Equation (4.9) shows that this condition is equivalent to r_1 becoming negative. Hernandez and Misner[2] have shown that, for a situation in which r_4 is always negative, gravitational collapse must occur if r_1 ever becomes negative. They proved this by making use of the theorem of Penrose,[14] which gives very general sufficient conditions under which a solution of Einstein's equations either has a singularity or possesses no initial Cauchy hypersurface. The purpose of this section is to investigate negative-mass shells without restrictions on r_4.

In the following it is assumed that all derivatives employed exist and are bounded and that the pressure and density are nonnegative. Furthermore it is supposed that a negative-mass shell exists in an orthogonal comoving coordinate system for all values of x^4 on an unbounded interval of x^4. Because r_1 must be positive at the center of the material, it follows that, at each value of x^4 on the interval, there is a value of x^1 at which r_1 vanishes, while for all smaller values of x^1, r_1 is positive. If there is a set of such points in the (x^1, x^4) plane given by

$$x^1 = f(\sigma), \quad x^4 = g(\sigma), \tag{5.1}$$

where f and g are continuous functions of a continuously varying parameter σ, and g takes on all values of x^4 on the unbounded interval, then this set of points forms the inner boundary of a permanent negative-mass shell. Clearly,

$$r_1(f(\sigma), g(\sigma)) = 0. \tag{5.2}$$

The path given by (5.1) is denoted by C. Thus, on C the dynamical equation (4.13) becomes

$$e^{-\gamma}(e^{-\gamma} r_4)_4 = -(4\pi \mathfrak{G} r p/c^2 + mc^2/r^2),$$

which may be integrated along C to give

$$\int_{x_a^4}^{x_b^4} (e^{-\gamma} r_4)_4 \, dx^4 = -\int_{x_a^4}^{x_b^4} \left(\frac{4\pi \mathfrak{G} r p}{c^2} + \frac{mc^2}{r^2}\right) e^{\gamma} \, dx^4, \tag{5.3}$$

where x_a^4 and x_b^4 are the initial and final values of x^4. Now

$$\int_{x_a^4}^{x_b^4} [(e^{-\gamma} r_4)_4 \, dx^4 + (e^{-\gamma} r_4)_1 \, dx^1] = [e^{-\gamma} r_4(f(x^4), x^4)]_{x_a^4}^{x_b^4}, \tag{5.4}$$

but, by Eq. (4.7),

$$(e^{-\gamma}r_4)_1 = e^{-\gamma}(r_{14} - \gamma_1 r_4) = e^{-\gamma}\alpha_4 r_1,$$

which vanishes on C. Thus (5.3) becomes

$$[e^{-\gamma}r_4(f(x^4), x^4)]_{x_a^4}^{x_b^4} = -c^2\int_{x_a^4}^{x_b^4}\left(\frac{4\pi\mathcal{G}p}{c^4} + \frac{m}{r^2}\right)e^{\gamma}\,dx^4. \tag{5.5}$$

Since $r_1(x^1, x^4)$ is positive for x^1 less than $f(x^4)$, Eq. (4.9) shows that m is positive along C. It is assumed that either p or m/r^2 has positive lower bounds along C. As the interval of integration in (5.5) becomes arbitrarily large, so, too, does the right-hand side as long as e^{γ} has a positive lower bound on C. In this case, $|r_4|$ must become arbitrarily large. If there is no positive lower bound for e^{γ}, then it must vanish for some x^4. In either case, the physical situation must be one in which a singularity is occurring. This shows that permanent negative-mass shells arise only in physical situations in which there is a singularity. Moreover, if a fluid distribution has a negative Schwarzschild mass due to such a negative-mass shell, it must exhibit singular behavior. The question of whether a finite number of temporary negative-mass shells could exist in a spherically symmetric distribution of material which does not exhibit singular behavior is left open, as is the question of whether an infinite number of temporary negative-mass shells could occur in a spherically symmetric fluid distribution which does not exhibit singular behavior but does possess a negative Schwarzschild mass. In Sec. 7, it is shown that permanent negative-mass shells occur in the closed uniform model universe solutions. Also in Sec. 7, the issue of whether negative-mass shells occur in charged distributions is discussed.

6. A CLASS OF EXACT SOLUTIONS OF EINSTEIN'S EQUATIONS

In this section, a class of solutions of Einstein's equations for a perfect fluid is developed by imposing a restriction on the form of the mass function. The material is assumed to occupy the interior of a sphere at any time and to be surrounded by empty space. Moreover, a comoving coordinate system is employed so that all the equations of Sec. 4 may be used.

The class of solutions to be discussed has been investigated by a number of authors. McVittie[1] examined special cases of the class by placing restrictions on the form of the metric. Thompson and Whitrow[5] and Taub[6] derived the class by imposing Eq. (6.5) as a condition on the metric. Taub[6] has shown that the metric restrictions of McVittie[1] are necessary if p is to depend on ρ alone. The class of solutions is developed here by the introduction of a function $\psi(x^1, x^4)$ connected with m by

$$m = 4\pi\mathcal{G}\rho r^3/3c^2 + \psi. \tag{6.1}$$

Because of Eq. (4.9), it follows that

$$4\pi\mathcal{G}r^3\rho_1/3c^2 = -\psi_1. \tag{6.2}$$

The time derivative of ψ may be found by differentiating (6.1) with respect to x^4 and employing (4.10) for m_4 and (4.12) for ρ_4. This results in

$$\frac{4\pi\mathcal{G}}{3c^2}\left(\rho + \frac{p}{c^2}\right)r^3\left(\alpha_4 - \frac{r_4}{r}\right) = \psi_4. \tag{6.3}$$

The restriction that will be placed on the mass function is

$$\psi = \psi(x^1). \tag{6.4}$$

This is the simplest class of functions ψ that gives a nonzero spatial gradient for ρ, by Eq. (6.2). The resulting class of solutions is called simple density gradient (SDG) solutions. Equation (6.3) then implies that

$$\alpha_4 = r_4/r, \tag{6.5}$$

which may be integrated to give

$$e^{\alpha} = ru_1/u, \tag{6.6}$$

where u is a function of x^1 alone. Therefore dl^2, the spatial part of the metric (2.1),becomes

$$dl^2 = (r^2/u^2)(du^2 + u^2\,d\Omega^2). \tag{6.7}$$

Henceforth, u will be used as the radial coordinate. Whenever the spatial part of a spherically symmetric metric is some function times the metric of a Euclidean 3-dimensional space, it is said to be in isotropic form. Thus, in a comoving coordinate system, the restrictions (6.4) or (6.5) or the condition that the spatial part of the coordinate system be isotropic are equivalent and result in the same class of solutions. Equation (4.7), with the aid of Eq. (6.5), becomes

$$(\ln r_4)_1 = (\gamma + \ln r)_1,$$

which may be integrated to give

$$e^{\gamma} = Hr_4/r, \tag{6.8}$$

where H is a function of x^4 alone. The mass function (4.8), by means of (6.6) and (6.8), takes the form

$$m = \tfrac{1}{2}r[1 + (r/Hc)^2 - (ur_u/r)^2]. \tag{6.9}$$

If the material is not bounded by empty space,then (2.1), (6.8), and (6.7) yield the metric

$$ds^2 = \left(\frac{r_4 H}{r}\right)^2(dx^4)^2 - \frac{r^2}{c^2u^2}(du^2 + u^2\,d\Omega^2). \tag{6.10}$$

It depends on the function r alone and, therefore, all the information of physical interest is determined once r is known. An equation which determines r may be found by differentiation of (6.1) with respect to x^1 and use of (6.2) to eliminate ρ_1. The result is

$$u^2 r r_{uu} - 2u^2 r_u^2 + u r r_u + r^2 = 3\psi r. \quad (6.11)$$

Since the mass function is known from (6.9), the density, by means of (6.1), is

$$4\pi\mathfrak{G}\rho/3c^2 = (m - \psi)/r^3. \quad (6.12)$$

The pressure may be found by use of (6.1) in (4.10), which yields

$$p/c^2 = -\rho - \tfrac{1}{3} r\rho_4/r_4. \quad (6.13)$$

In the event that the material is bounded by empty space, examination of the metric (6.10) reveals that no generality is lost if the boundary of the material is given by

$$u = 1.$$

Because the Schwarzschild solution is the external solution, the boundary value of the mass function is M. Hence, by (6.9), H is given by

$$H^{-2} = \left\{\frac{c^2}{r^2}\left[\frac{2M}{r} - 1 + \left(\frac{r_u}{r}\right)^2\right]\right\}_b.$$

Therefore, the mass function is

$$m = M\left(\frac{r}{r_b}\right)^3 + \tfrac{1}{2} r\left[1 - \left(\frac{r}{r_b}\right)^2 + r^2\left(\frac{r_u}{r^2}\right)_b^2 - u^2\left(\frac{r_u}{r}\right)^2\right]. \quad (6.14)$$

The form of the metric which follows from the above expression for H and (6.10) is

$$ds^2 = \left(\frac{r_4 H}{r}\right)^2 (dx^4)^2 - \frac{r^2}{c^2 u^2}(du^2 + u^2\, d\Omega^2),$$

$$H^{-2} = \left\{\frac{c^2}{r^2}\left[\frac{2M}{r} - 1 + \left(\frac{r_u}{r}\right)^2\right]\right\}_b. \quad (6.15)$$

It is possible to determine alternative expressions for ρ and p. Equation (6.2) may be integrated to give

$$\frac{4\pi\mathfrak{G}}{3c^2}(\rho - \rho_b) = \int_u^1 \frac{\psi_u}{r^3}\, du,$$

which with the aid of Eq. (6.12) becomes

$$\frac{4\pi\mathfrak{G}\rho}{3c^2} = \frac{M - \psi_b}{r_b^3} + \int_u^1 \frac{\psi_u}{r^3}\, du. \quad (6.16)$$

Equation (4.11) may be integrated and with the aid of (6.8) the expression for p becomes

$$\frac{p}{c^2} = e^{-\gamma}\int_u^1 \rho(e^{\gamma})_u\, du, \quad (6.17)$$

where use has been made of the vanishing of the boundary pressure which follows from the fact that $m_b = M$ is a constant. Finally, p may also be found by use of the dynamical equation (4.13).

Equation (6.11) may be put into other useful forms. First let

$$y = 1/(ru), \quad x = \ln u, \quad (6.18)$$

so that (6.11) becomes

$$y_{xx} + 2y_x = -3\psi e^x y^2. \quad (6.19)$$

In the event that

$$-3\psi e^x = Bx^n, \quad (6.20)$$

Eq. (6.19) becomes a special case of the generalized Lane–Emden equation which is mentioned by Kamke.[15] As an alternative, let the independent variable be

$$z = 16u^{-2} = 16e^{-2x}; \quad (6.21)$$

then Eq. (6.19) becomes

$$y_{zz} = -3\psi z^{-\frac{5}{2}} y^2. \quad (6.22)$$

In the event that

$$\psi = -2k^2 z^{\frac{1}{2}}, \quad (6.23)$$

where k is some constant, then,with the aid of Kamke,[15]

$$y = \wp(kx + A(x^4)), \quad (6.24)$$

where $\wp$ is the Weierstrass $\wp$ function with invariants

$$g_2 = 0, \quad g_3 = B(x^4). \quad (6.25)$$

Thus, A and B are functions of integration. This solution is noticed but it will not be examined further here. It is also possible to integrate (6.22),when

$$\psi = 0, \quad (6.26)$$

to obtain

$$y = z/16R + K/R,$$

where R and K are functions of x^4 alone. Thus, r is given by

$$r = Ru/(1 + Ku^2). \quad (6.27)$$

Because of Eq. (6.2), it follows that

$$\rho = \rho(x^4) \quad (6.28)$$

is equivalent to (6.26) and so the general uniform density solution has been recovered in comoving coordinates. In Sec. 7, the uniform model universe solutions are shown to be a special case of this solution. The physical properties of the uniform-density solutions have been extensively investigated. Bonnor and Faulkes[16] have constructed oscillating uniform-density solutions that have no singularities. Thompson and Whitrow[5] have shown that all uniform-density

solutions that correspond to inward motions which have a negative density-gradient undergo gravitational collapse. McVittie and Stabell[17] have investigated some uniform-density solutions and shown that they must undergo collapse even though they have infinite central pressure. Taub[6] has shown that the outer boundary in an oscillating uniform-density solution which is free of singularities does not fall through the Schwarzschild radius of the configuration. Bondi[18] finds conditions under which initial inward motion in an uniform-density solution will be halted. Particular non-uniform-density SDG solutions have also been examined by McVittie[1] and by Nariai.[19]

The regularity conditions which ensure that an SDG solution be free of singularities will now be developed. Because the material at $u = 0$ is assumed to be at the center of the distribution,

$$r(0, x^4) = 0. \tag{6.29}$$

In order that the coefficient $(r/u)^2$, which occurs in the spatial section of the metric (6.15), be finite at the origin, it must be the case that

$$\lim_{u\to 0} (r/u) = r_u(0, x^4) > 0. \tag{6.30}$$

The radius of curvature by (2.5) should vanish only at the center and therefore

$$r(u, x^4) > 0, \tag{6.31}$$

for all u on $(0, 1]$. The coefficient of $(dx^4)^2$ in metric (6.10) must never vanish and so, without any loss of generality,

$$r_4/[2m/r - 1 + (r_u/r)^2]_b > 0. \tag{6.32}$$

Because the density is finite in the vicinity of the center, it follows from (4.9) that

$$\lim_{u\to 0} m/r^3 = 4\pi\mathfrak{G}\rho(0, x^4)/3c^2. \tag{6.33}$$

This result, by virtue of Eq. (6.12), is equivalent to ψ obeying

$$\lim_{u\to 0} \psi/u^3 = 0. \tag{6.34}$$

Equation (6.34) moreover implies that m/r^3 does approach a finite limit. The right-hand side of Eq. (6.11) is zero to order u^3 at least, by virtue of Eq. (6.34). The left-hand side of (6.11), when expanded to order u^3, depends on the behavior of r to order u^3. Thus the solution of (6.11) is the same as the uniform-density solution to order u^3. Now in the uniform-density solution (6.29) and (6.30) hold and also

$$r_{uu}(0, x^4) = 0, \tag{6.35}$$

and so these equations are satisfied for any solution of (6.11) so long as (6.34) is satisfied. When Eq. (6.11) is examined at $u = 0$ with the aid of (6.34) and (6.35), it turns out that

$$\left| \lim_{u\to 0} \left(-2\frac{r_u^2}{ru} + \frac{r_u}{u^2} + \frac{r}{u^3} \right) \right| < \infty.$$

But if this expression is rewritten and (6.30) is employed, we find that

$$\left| \lim_{u\to 0} \frac{r}{u}\left(-2\frac{r_u u}{r}\frac{r_u}{ru} + \frac{r_u}{ru} + \frac{1}{u^2} \right) \right| < \infty$$

implies

$$\left| \lim_{u\to 0} \left(\frac{1 - ur_u/r}{u^2} \right) \right| < \infty. \tag{6.36}$$

Equation (6.36) implies that ρ is finite at the center of the material for, by Eq. (6.14),

$$\lim_{u\to 0} \frac{m}{r^3} = \frac{M}{r_b^3} - \frac{[1 - (r_u/r)_b^2]}{2r_b^2} + \tfrac{1}{2} \lim_{u\to 0} \frac{u^2}{r^2}\left(\frac{1 - (ur_u/r)^2}{u^2} \right).$$

By means of (6.30) the last term becomes

$$\frac{1}{2r_u^2(0, x^4)} \lim_{u\to 0} \left(1 + \frac{ur_u}{r} \right) \lim_{u\to 0} \left(\frac{1 - ur_u/r}{u^2} \right) = \frac{1}{r_u^2(0, x^4)} \lim_{u\to 0} \left(\frac{1 - ur_u/r}{u^2} \right),$$

which has been proven to be finite. Thus the conclusions are that if Eq. (6.34) is satisfied, then (6.29) and (6.30) both hold and also that the central density is always finite. Equation (6.34) is the necessary and sufficient condition for the spatial part of the metric, and the density, to be well behaved at $u = 0$. Equation (6.17) then implies that $p(0, x^4)$ is also well behaved provided $e^{2\gamma}$ is positive and bounded. If in addition, Eq. (6.31) holds and ρ and p are positive, the solution is physically acceptable and free from any singularities.

7. APPLICATIONS OF THE MASS FUNCTION

In this section the mass function is used to examine certain properties of the uniform model universe solutions, the charged point-mass solution, and distributions of charged fluid at zero pressure.

In the following it is shown that the uniform model universe solutions are the subclass of the uniform-density solution discussed in Sec. 6 in which p is a function of x^4 alone and Λ, the cosmical constant, is taken to be zero. For, if both p and ρ depend on x^4 alone, Eq. (6.13) implies that r_4/r depends on x^4 alone. In the uniform-density solution r is given by (6.27)

and therefore

$$\frac{r_4}{r} = \frac{R_4}{R} - \frac{K_4 u^2}{1 + Ku^2}.$$

Thus, if r_4/r depends on x^4 alone, K_4 must be zero, which in turn means that K is a constant. Moreover, the form of the spatial part of metric (6.10) implies that K is a scale factor and so without loss of generality we take

$$K = k/4, \quad k = \{1, 0, -1\}. \tag{7.1}$$

Because g_{44} in the metric (6.10) depends on time alone, a new time coordinate t may be introduced by

$$dt = r_4 H\, dx^4/r. \tag{7.2}$$

Thus the metric (6.10) with the aid of (6.27), (7.1), and (7.2) becomes

$$ds^2 = dt^2 - \frac{R^2}{c^2(1 + \frac{1}{4}ku^2)^2}(du^2 + u^2\, d\Omega^2), \tag{7.3}$$

which is one of the standard forms of the metric of a uniform model universe. In order to examine these solutions by means of the mass function technique, it is preferable to use another form of (7.3), namely,

$$ds^2 = dt^2 - (R^2/c^2)[d\omega^2 + S_k(\omega)\, d\Omega^2], \tag{7.4}$$

where $R(t)$ is the scale function, ω is the radial comoving coordinate, and $S_k(\omega)$ is given by

$$\begin{aligned} S_k(\omega) &= \sin\omega, \quad \text{if } k = +1, \\ &= \omega, \quad \text{if } k = 0, \\ &= \sinh\omega, \quad \text{if } k = -1. \end{aligned} \tag{7.5}$$

Equation (2.1) implies that the radius of curvature, r, is given by

$$r = RS_k(\omega). \tag{7.6}$$

Comparison of (2.1) with (7.4) indicates that $t = x^4$, $\omega = x^1$. Hence, the mass function (3.4) is

$$m = \tfrac{1}{2}r[1 + \dot{R}^2 S^2(\omega)/c^2 - C_k^2(\omega)],$$

where $\dot{R} = dR/dt$ and $C_k = dS_k/d\omega$. Thus, with the aid of (7.6) and $C_k^2 + kS_k^2 = 1$, the mass function becomes

$$m = \tfrac{1}{2}r^3(k + \dot{R}^2/c^2)/R^2. \tag{7.7}$$

Because (6.26) must hold in any uniform-density solution, Eq. (6.1) becomes

$$m = 4\pi\mathfrak{G}\rho r^3/3c^2. \tag{7.8}$$

Comparison of (7.7) and (7.8) immediately gives ρ as

$$\frac{8\pi\mathfrak{G}\rho}{c^2} = 3\frac{(k + \dot{R}^2/c^2)}{R^2}. \tag{7.9}$$

A striking feature of the spherical universe, which occurs when $k = +1$, is the existence of a permanent negative-mass shell on the ω interval $(\frac{1}{2}\pi, \pi)$. This is because r is $R \sin\omega$ by (7.6) and r increases with increasing ω to a maximum value of R when ω is $\frac{1}{2}\pi$, and then falls off to zero when ω is π, at the antipole. In fact, Eq. (7.8) implies that the total amount of mass–energy between the origin $\omega = 0$ and its antipole $\omega = \pi$ is zero, which means that

$$m(\pi, t) = 0. \tag{7.10}$$

Since $r_1 = R\cos\omega$, it follows that r_1 is negative in the ω interval $(\frac{1}{2}\pi, \pi)$ and thus the condition for negative-mass shells, stated at the beginning of Sec. 5, is satisfied.

A second application of the mass function to a specific case is found in the solution for the external field of a charged spherically symmetric distribution of material due to Nordström[20] and Reissner.[21] Adler, Bazin, and Schiffer[22] give the metric for the external field as

$$\begin{aligned} ds^2 = &\left[1 - \frac{2}{r}\left(M - \frac{2\pi\mathfrak{G}q^2}{c^4 r}\right)\right] dt^2 \\ &- \frac{1}{c^2}\left\{\left[1 - \frac{2}{r}\left(M - \frac{2\pi\mathfrak{G}q^2}{c^4 r}\right)\right]^{-1} dr^2 + r^2\, d\Omega^2\right\}, \end{aligned} \tag{7.11}$$

where M and q are constants. Their expression for the electric field, which is radial, is

$$E = q/r^2, \tag{7.12}$$

where q is interpreted as the charge of the distribution and Heaviside–Lorentz units are used. Their expression for the energy density is

$$T_4^4 = E^2/2c^2. \tag{7.13}$$

Because the radius of curvature r is the radial coordinate in (7.11) and $r_4 = \partial r/\partial t$ is zero, Eqs. (3.2) and (3.4) imply that the mass function is

$$m = M - 2\pi\mathfrak{G}q^2/c^4 r. \tag{7.14}$$

Equation (7.14) implies that

$$m(\infty) = M, \tag{7.15}$$

which means that M is the total amount of mass-energy between the center of the material and infinity calculated from the external field. Equation (7.14) also shows that the mass function decreases as r does. The reason for this may be established as follows. When (7.14) is differentiated with respect to r and (7.12) is also used, it is found that

$$\frac{dm}{dr} = \frac{2\pi \mathfrak{G} q^2}{c^4 r^2} = \frac{\mathfrak{G}}{2c^4} E^2 4\pi r^2.$$

If this is integrated, an equivalent form of (7.14) is obtained, namely,

$$m(r) = M - \frac{\mathfrak{G}}{c^2} \int_r^\infty \left(\frac{E^2}{2c^2}\right) 4\pi r^2 \, dr, \qquad (7.16)$$

a result that would also be derived from (3.9) and (7.13). Therefore, $m(r)$ is the total mass-energy minus the mass equivalent of the energy stored in the electric field between radius r and infinity.

Another aspect of (7.14) is that m becomes negative for sufficiently small r. This means that a small particle close to a point charge will be repelled. If the metric (7.11) is that of the field exterior to a charged, static, and finite distribution of matter at zero pressure, then two alternatives arise. The first is that, because of the presence of the electric field, the distribution has a positive m_0. The second alternative is that the boundary value of m may be negative. In this case, for a static incoherent charged fluid Adler, Bazin, and Schiffer[23] give T_4^4 as

$$T_4^4 = \rho + E^2/2c^2. \qquad (7.17)$$

Equation (3.9) yields

$$m = \frac{4\pi \mathfrak{G}}{c^2} \int_0^{x^1} T_4^4 r^2 r_1 \, dx^1. \qquad (7.18)$$

This shows, since T_4^4 is necessarily positive, that the only way in which the above material could have a negative mass at the boundary would be for a negative-mass shell to occur.

[1] G. C. McVittie, Ann. Inst. Henri Poincaré **6**, 1 (1967).
[2] W. C. Hernandez and C. W. Misner, Astrophys. J. **143**, 452 (1966).
[3] C. W. Misner and D. H. Sharp, Phys. Rev. **136**, B571 (1964).
[4] M. M. May and R. H. White, Phys. Rev. **141**, 1232 (1966).
[5] I. H. Thompson and G. J. Whitrow, Monthly Notices Roy. Astron. Soc. **136**, 207 (1967).
[6] A. H. Taub, Ann. Inst. Henri Poincaré **9**, 153 (1968).
[7] J. L. Synge, *Relativity: The General Theory* (North-Holland Publ. Co., Amsterdam, 1964), pp. 270–272.
[8] W. Israel, Proc. Roy. Soc. (London) **A248**, 404 (1958).
[9] A. Lichnerowicz, *Theories relativistes de la gravitation et de l'electromagnetisme* (Masson et Cie., Paris, 1955).
[10] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[11] K. S. Thorne, thesis, Princeton University, 1965.
[12] G. C. McVittie, *General Relativity and Cosmology* (The University of Illinois Press, Urbana, Ill., 1965), p. 66.
[13] J. M. Bardeen, *Second Texas Symposium on Relativistic Astrophysics*, 1964.
[14] R. Penrose, Phys. Rev. Letters **14**, 57 (1965).
[15] E. Kamke, *Differentialgleichungen* (Akademische Verlagsgellschaft Geest und Portig K.G., Leipzig, 1959), pp. 561, 542.
[16] W. B. Bonnor and M. C. Faulkes, Monthly Notices Roy. Astron. Soc. **137**, 239 (1967).
[17] G. C. McVittie and R. Stabell, Ann. Inst. Henri Poincaré **7**, 103 (1967).
[18] H. Bondi, Monthly Notices Roy. Astron. Soc. **142**, 333 (1969).
[19] H. Nariai, Progr. Theoret. Phys. (Kyoto) **38**, 740 (1967).
[20] G. Nordström, Verhandel. Koninkl. Ned. Akad. Wetenschap., Afdel Natuurk. Sec. II. **26**, 1201 (1918).
[21] H. Reissner, Ann. Physik **50**, 106 (1916).
[22] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill Book Co., New York, 1965), Sec. (13.1).
[23] Ref. 22, Eqs. (9.70) and (13.13).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Spherical Symmetry and Mass–Energy in General Relativity. II. Particular Cases

MICHAEL E. CAHILL AND G. C. MCVITTIE
University of Illinois Observatory, Urbana, Illinois 61801

(Received 10 July 1969)

A solution of Einstein's field equations for the motion of a spherically symmetric distribution of perfect fluid is investigated in an isotropic comoving coordinate system. It is shown that this solution includes all solutions discussed by McVittie in which the density depends on both the radial and time coordinates. Necessary and sufficient conditions for the solution to be singularity-free and for the density and pressure to be nonnegative and monotonically decreasing from the center of the material outwards to an outer boundary are found. The material is surrounded by empty space. Examples of both oscillating and "bouncing" solutions are produced. It is shown that the outer boundary of the material never penetrates the Schwarzschild radius in all singularity-free solutions.

1. INTRODUCTION

The radial motion of a spherically symmetric mass of perfect fluid may be treated exactly in general relativity by solving Einstein's equations. Classes of analytical solutions of these equations are attainable under certain circumstances. One such class was developed by McVittie.[1] Thompson and Whitrow[2] showed that McVittie's solutions were a subclass of a still more extensive group and Cahill and McVittie[3] (hereinafter referred to as Paper I) have discussed some properties of this group.

There are three classes of solutions treated by McVittie[1] and it is shown in Sec. 2 of this paper that in class (iii) the density depends on the time coordinate alone. This class of solutions, therefore, is a special case of the general uniform-density solution which is briefly treated in Paper I,where references to several more complete descriptions of the general uniform-density solution are given. However, it should be noted that Taub[4] has shown that the outer boundary of any singularity-free uniform-density solution cannot penetrate the Schwarzschild radius. It is also shown in Sec. 2 that the remaining two classes of solutions, classes (i) and (ii), may be treated together and that the isotropic coordinate system employed in Paper I may be used to develop these classes in a simple form. The purpose of this paper is to find members of these classes which are free of singularities and in which the density and pressure are nonnegative and monotonically decreasing from the center of the material to an outer boundary. This is accomplished by use of the general results obtained in Paper I. Lastly, we point out that a special case of these classes due to Nariai[5] has already appeared in the literature.

2. McVITTIE'S SUBCLASS OF SIMPLE DENSITY GRADIENT SOLUTIONS

The class of solutions of Einstein's equations developed by McVittie[1] has the metric

$$ds^2 = y^2(dx^4)^2 - R_0^2S^2e^{\eta}c^{-2}[(dx^1)^2 + f^2\,d\Omega^2],$$
$$y = 1 - \tfrac{1}{2}\eta_z, \quad e^z = Q/S, \qquad (2.1)$$
$$d\Omega^2 = d\theta^2 + \sin^2\theta\,d\varphi^2,$$

where x^1 and x^4 are the radial and time coordinates, respectively, η is a function of z, Q and f are functions of x^1, S is a function of x^4, and R_0 is a constant. In Paper I the more general metric

$$ds^2 = e^{2\gamma}(dx^4)^2 - c^{-2}[e^{2\alpha}(dx^1)^2 + r^2\,d\Omega^2] \qquad (2.2)$$

is employed, where α, γ, and r are functions of x^1 and x^4. This is Eq. (I.2.1) [that is, Eq. (2.1) of Paper I]. Hence (2.1) is the special case in which

$$e^{2\gamma} = y^2, \qquad (2.3a)$$
$$e^{2\alpha} = R_0^2S^2e^{\eta}, \qquad (2.3b)$$
$$r = R_0Se^{\eta/2}f. \qquad (2.3c)$$

If the energy–momentum tensor is that of a perfect fluid, then the coordinates are comoving by virtue of the previous formulas.

The condition that T_1^1 and T_2^2 are identical leads McVittie to the three equations

$$Q_{11}/Q - Q_1f_1/fQ = a(Q_1/Q)^2, \qquad (2.4a)$$
$$f_{11}/f - f_1^2/f^2 + 1/f^2 = b(Q_1/Q)^2, \qquad (2.4b)$$
$$y_{zz} + (a - 3 + y)y_z + y[a + b - 2 - (a - 3)y - y^2] = 0, \qquad (2.4c)$$

where the suffix 1 means a derivative with respect to x^1 and $y_z = dy/dz$, and a and b are two constants

which may be specified arbitrarily. There are three known first integrals of Eq. (2.4c), namely,

$$\text{case (i):} \qquad \frac{dy}{dz} = \tfrac{1}{5}(a-3)y + \tfrac{1}{2}y^2, \tag{2.5}$$

provided that

$$b = -\tfrac{1}{25}(6a^2 - 11a + 4);$$

$$\text{case (ii):} \qquad \frac{dy}{dz} = -\tfrac{1}{2}(b+1) + \tfrac{1}{2}y^2, \tag{2.6}$$

provided that $a = 3$; and the particular integral

$$\text{case (iii):} \qquad \frac{dy}{dz} = (a+b-2) - (a-3)y - y^2. \tag{2.7}$$

The density turns out to be

$$\tfrac{8}{3}\pi\mathcal{G}\rho = \left(\frac{\dot{S}}{S}\right)^2 + \frac{c^2 e^{-\eta}}{R_0^2 S^2}\left\{\frac{(1-f_1^2)}{f^2} - 2(1-y)\frac{Q_1 f_1}{Qf} - \tfrac{1}{3}[2b - 2y_z + (1-y)(2a-1-y)] \times \left[\frac{Q_1}{Q}\right]^2\right\}, \tag{2.8}$$

where $\dot{S} = dS/dx^4$.

The theory of the mass function will now be applied to McVittie's metrics. The mass function was given by Eq. (I.3.4) as

$$m = \tfrac{1}{2}r(1 + e^{-2\gamma}r_4^2/c^2 - e^{-2\alpha}r_1^2), \tag{2.9}$$

where r_4 and r_1 are the partial derivatives of r with respect to x^4 and x^1, respectively. Its calculation requires the derivatives of r, where r is defined by Eq. (2.3c). With the aid of (2.1) also, the derivatives are

$$r_1 = R_0 S e^{\frac{1}{2}\eta}[f_1 + (1-y)Q_1 f/Q],$$
$$r_4 = R_0 \dot{S} e^{\frac{1}{2}\eta} f y = r y(\dot{S}/S). \tag{2.10}$$

The mass function is then, by (2.9), (2.3), and (2.10),

$$m = \tfrac{1}{2}r\{1 + (r\dot{S}/cS)^2 - [f_1 + (1-y)fQ_1/Q]^2\},$$

which with the aid of Eq. (2.8) becomes

$$m = 4\pi\mathcal{G}\rho r^3/3c^2 + [a + b - 2 - (a-3)y - y^2 - y_z]\tfrac{1}{3}r(Q_1 f/Q)^2.$$

Thus, the mass functions corresponding to the three integrals of Eq. (2.4c), corresponding to cases (i)–(iii), respectively, are

$$m = 4\pi\mathcal{G}\rho r^3/3c^2 - \tfrac{1}{2}[y + \tfrac{2}{5}(a-3)]^2 r(Q_1 f/Q)^2, \tag{2.11a}$$

$$m = 4\pi\mathcal{G}\rho r^3/3c^2 + \tfrac{1}{2}(b + 1 - y^2)r(Q_1 f/Q)^2, \tag{2.11b}$$

$$m = 4\pi\mathcal{G}\rho r^3/3c^2. \tag{2.11c}$$

In Paper I, Einstein's equations for a perfect fluid, in which the mass function was written in the form given in (I.6.1), are discussed. This equation is

$$m = (4\pi\mathcal{G}\rho r^3/3c^2) + \psi. \tag{2.12}$$

The class of solutions which occur when ψ depends on x^1 alone was first considered by Thompson and Whitrow.[2] As in Paper I, this class is referred to as the class of simple density gradient (SDG) solutions. Equations (2.11) and (2.12) may be used to evaluate ψ in each of the three cases with the results

$$\psi = -\tfrac{1}{2}[y + \tfrac{2}{5}(a-3)]^2 r(Q_1 f/Q)^2, \tag{2.13a}$$

$$\psi = \tfrac{1}{2}(b + 1 - y^2)r(Q_1 f/Q)^2, \tag{2.13b}$$

$$\psi = 0, \tag{2.13c}$$

for cases (i)–(iii), respectively. Because ψ is zero in case (iii), all solutions of this type are uniform-density solutions which, as mentioned in Sec. 1, are treated elsewhere. In cases (i) and (ii), ψ will be shown to depend on x^1 alone by completing the solution of the problem in the following way.

In case (i): let $\mu = \frac{1}{5}(a-3)$. Then (2.5) may be integrated and the definition $y = 1 - \frac{1}{2}\eta_z$ used to give

$$y = -2\mu e^{\mu z}/(\Gamma + e^{\mu z}), \quad e^{\frac{1}{2}\eta} = e^z(\Gamma + e^{\mu z})^2, \tag{2.14}$$

where Γ is a constant of integration and a multiplicative constant of integration in the determination of η has been equated to unity. Equation (2.3c), the relation $e^z = Q/S$, and (2.14) imply that

$$r = R_0 Q f(\Gamma + (Q/S)^\mu)^2. \tag{2.15}$$

Equation (2.13a) for case (i), with the aid of Eqs. (2.14) and (2.15), yields

$$\psi = -2R_0(\mu\Gamma)^2 f^3 Q_1^2/Q. \tag{2.16}$$

This clearly shows that ψ is a function of x^1 alone and concludes the treatment of case (i).

In case (ii), let $\nu^2 = b + 1$, $b \neq -1$. Equation (2.6) may then be integrated and the definition $y = 1 - \frac{1}{2}\eta_z$ used to give

$$y = \nu(1 - \Gamma e^{\nu z})/(1 + \Gamma e^{\nu z}),$$
$$e^\eta = (1 + \Gamma e^{\nu z})^4 e^{2(1-\nu)z}, \tag{2.17}$$

where Γ is a constant of integration and a multiplicative constant of integration in the determination of η has been equated to unity. Equation (2.3c) implies that

$$r = R_0 S f(1 + \Gamma e^{\nu z})^2 e^{(1-\nu)z}. \tag{2.18}$$

Equation (2.13b) for case (ii) with the aid of Eqs.

(2.17) and (2.18) requires that

$$\psi = \tfrac{1}{2}\nu^2\left(1 - \frac{(1-\Gamma e^{\nu z})^2}{(1+\Gamma e^{\nu z})^2}\right) R_0 S f (1+\Gamma e^{\nu z})^2 e^{(1-\nu)z}\left(\frac{Q_1 f}{Q}\right)^2$$
$$= 2R_0\nu^2\Gamma S e^z f^3 \frac{Q_1^2}{Q^2}.$$

Because $e^z = Q/S$, it is seen that

$$\psi = 2R_0\nu^2\Gamma f^3 Q_1^2/Q, \tag{2.19}$$

which shows that ψ again depends on x^1 alone. Equation (2.18) may be put into a simpler form if the quantities φ and ϵ are introduced by

$$\varphi = \ln|\Gamma| + z = \ln|\Gamma| + \ln Q - \ln S, \quad \epsilon = |\Gamma|/\Gamma, \tag{2.20}$$

for it thereby reads

$$r = (R_0 f Q/|\Gamma|)(e^{-\frac{1}{2}\varphi} + \epsilon e^{+\frac{1}{2}\varphi})^2$$
$$= (R_0 f Q/|\Gamma|)(e^{\varphi} + e^{-\varphi} + 2\epsilon),$$

whence

$$r = (2R_0 f Q/|\Gamma|)(\cosh\varphi + \epsilon). \tag{2.21}$$

The form of r in the last equation is

$$r = Dn[\cosh(h+G) + \epsilon], \tag{2.22}$$

where n and h are functions of x^1, G is a function of x^4, ϵ is ± 1, and D is an arbitrary constant length. Thus Eq. (2.22) gives the form of r in case (ii). Moreover, if

$$\epsilon = -1, \quad n = \bar{n}/\delta^2, \quad h = \bar{h}\delta, \quad G = \bar{G}\delta, \tag{2.23}$$

where $\bar{n}$ and $\bar{h}$ are functions of x^1, $\bar{G}$ is a function of x^4, and δ is a constant, then

$$\lim_{\delta\to 0} D\bar{n}[\cosh\delta(\bar{h}+\bar{G}) + \epsilon]/\delta^2 = D\bar{n}(\bar{h}+\bar{G})^2, \tag{2.24}$$

which is the form of r in Eq. (2.15) for case (i). Thus it is seen that the form of r, given in Eq. (2.22), covers both cases (i) and (ii). In Paper I it is shown that for all spherically symmetric perfect fluid solutions, in which ψ depends on x^1 alone, the metric may be expressed as (I.6.10) which is

$$ds^2 = \left(\frac{r_4 H}{r}\right)^2 (dx^4)^2 - \frac{r^2}{c^2 u^2}(du^2 + u^2\, d\Omega^2), \tag{2.25}$$

where x^4 is the time coordinate, u is the radial comoving coordinate, r is the radius of curvature, H is an arbitrary function of x^4, and r_4 is the partial derivative of r with respect to x^4. The functions n and h in Eq. (2.22) may therefore be regarded as functions of u. It is also shown in Paper I that r must be a solution of Eq. (I.6.11) which is

$$u^2 r r_{uu} - 2u^2 r_u^2 + u r r_u + r^2 = 3\psi r, \tag{2.26}$$

where the subscript of u refers to partial differentiation with respect to u. In Sec. 3 of this paper the most general forms of $n(u)$ and $h(u)$ consistent with (2.26) are found, the function $G(x^4)$ being left arbitrary.

3. McVITTIE'S CLASS OF NONUNIFORM-DENSITY SDG SOLUTIONS

The plan of this section is to use the generalized form of r for McVittie's SDG solutions of cases (i) and (ii) given by Eq. (2.22) and substitute it into the differential equation (2.26) in order to determine $n(u)$ and $h(u)$. It will be supposed that $G(x^4)$ is arbitrary since (2.26) is a differential equation which involves derivatives with respect to u alone and coefficients depending only on u and on r itself. Nariai[5] has investigated McVittie's case (i) solutions with a restriction on the constants of integration. Thus, the solutions to be developed will include the Nariai solutions as a special case. The metric of any SDG solution which is bounded by empty space has the form of (2.25) but H is no longer arbitrary because of the condition that m shall be constant at the boundary. A quantity with the subscript "b" is evaluated at the boundary and thus $m_b = M$, the Schwarzschild mass. Without loss of generality the boundary can be defined by $u = 1$. Equation (I.6.15) for the metric in this situation is

$$ds^2 = \left(\frac{r_4 H}{r}\right)^2 (dx^4)^2 - \frac{1}{c^2}\frac{r^2}{u^2}(du^2 + u^2\, d\Omega^2),$$
$$H^{-2} = \left\{\frac{c^2}{r^2}\left[\frac{2M}{r} - 1 + \left(\frac{r_u}{r}\right)^2\right]\right\}_b. \tag{3.1}$$

Conditions under which e^{α} and e^{ν} are bounded and positive and under which r is positive, except at the center of the material, will be developed and discussed. Expressions for the density and pressure will be found and the conditions under which both quantities are nonnegative will be determined.

Since

$$r = Dn[\cosh(h+G) + \epsilon], \tag{3.2}$$

its logarithmic derivative with respect to x^4 is

$$\frac{r_4}{r} = \frac{\sinh(h+G)\dot{G}}{\cosh(h+G) + \epsilon}, \tag{3.3}$$

where the dot denotes differentiation with respect to x^4. Metric (3.1) thereby becomes

$$ds^2 = \left(\frac{\sinh(h+G)\dot{G}H}{\cosh(h+G)+\epsilon}\right)^2 (dx^4)^2$$
$$- \frac{D^2 n^2}{c^2 u^2}[\cosh(h+G) + \epsilon]^2[du^2 + u^2\, d\Omega^2], \tag{3.4}$$

where

$$H^{-2} = \{(c^2/r^2)[2M/r - 1 + (r_u/r)^2]\}_b. \quad (3.5)$$

The values of the metric coefficients are unchanged if $h(u) + G(x^4)$ is replaced by $-h(u) - G(x^4)$. Moreover, since g_{44} may not vanish, $h + G$ may never equal zero and it is therefore possible to impose the condition that

$$h + G > 0, \quad (3.6)$$

for all (x^1, x^4) without any loss of generality. The remaining condition which ensures that g_{44} does not vanish is

$$\dot{G}H > 0. \quad (3.7)$$

The determination of $n(u)$ and $h(u)$ is now undertaken. The first and second derivatives of r with respect to u are

$$\frac{r_u}{D} = \frac{n_u r}{nD} + nh_u \sinh(h + G),$$

$$\frac{r_{uu}}{D} = \left(\frac{n_{uu}}{n} + h_u^2\right)\frac{r}{D} + (2n_u h_u + nh_{uu}) \sinh(h + G) - h_u^2 n\epsilon. \quad (3.8)$$

When the above equations are used in Eq. (2.26), it becomes an equation of the form

$$\zeta_1 n[\cosh(h + G) + \epsilon] + \zeta_2 \sinh(h + G) + \zeta_3 = 0, \quad (3.9)$$

where

$$\zeta_1 = n_{uu}/n - h_u^2 - 2n_u^2/n^2 + n_u/un + 1/u^2,$$
$$\zeta_2 = nh_{uu} - 2n_u h_u + nh_u/u,$$
$$\zeta_3 = (3/u^2)(h_u^2 nu^2\epsilon - \psi/D).$$

If this equation is to hold for an arbitrary function G, then ζ_1, ζ_2, and ζ_3 must be identically zero. Thus, the following three equations are obtained:

$$n_{uu}/n - h_u^2 - 2n_u^2/n^2 + n_u/un + 1/u^2 = 0, \quad (3.10)$$

$$nh_{uu} - 2n_u h_u + nh_u/u = 0, \quad (3.11)$$

$$\psi = Dh_u^2 nu^2\epsilon. \quad (3.12)$$

Equations (3.10), (3.11), and (3.12) are the conditions that n, h, and ψ must satisfy in order that r be of the form given in Eq. (3.2). Equation (3.11) is a first-order linear equation in h_u, the integral of which is

$$h_u = \frac{c_0}{c_2}\frac{n^2}{u}, \quad (3.13)$$

where the constant of integration has been written as c_0/c_2. This equation may be used to eliminate h_u in Eq. (3.10) which, when multiplied by u^2/n^3, becomes

$$\frac{u^2 n_{uu}}{n^4} - \frac{2u^2 n_u^2}{n^5} + \frac{un_u}{n^4} = \left(\frac{c_0}{c_2}\right)^2 n - \frac{1}{n^3}.$$

But if un_u/n^2 is factored from the left-hand side then

$$\left(\frac{un_u}{n^2}\right)\left[\frac{u\,dn_u/dn}{n^2} - \frac{2un_u}{n^3} + \frac{1}{n^2}\right] = \left(\frac{c_0}{c_2}\right)^2 n - \frac{1}{n^3}$$

results. The bracketed term is $d(un_u/n^2)/dn$ and so

$$\frac{d}{dn}\left(\frac{un_u}{n^2}\right)^2 = 2\left(\frac{c_0}{c_2}\right)^2 n - \frac{2}{n^3},$$

which may be integrated to give

$$\left(\frac{un_u}{n^2}\right)^2 = \left(\frac{c_0}{c_2}\right)^2 n^2 + \frac{1}{n^2} + \frac{2c_1}{c_2}, \quad (3.14)$$

where the constant of integration has been written as c_1/c_2. The substitution $n^2 = 1/x$ reduces this equation to

$$\frac{dx}{[(c_0/c_2)^2 + 2(c_1/c_2)x + x^2]^{\frac{1}{2}}} = -2\frac{du}{u}.$$

Hence,

$$\ln\left\{x + \frac{c_1}{c_2} + \left[\left(\frac{c_0}{c_2}\right)^2 + 2\left(\frac{c_1}{c_2}\right)x + x^2\right]^{\frac{1}{2}}\right\} = -\ln(c_2 u^2),$$

where c_2 is the constant of integration. Thus,

$$\left(\frac{c_0}{c_2}\right)^2 + 2\left(\frac{c_1}{c_2}\right)x + x^2 = \left(\frac{1}{c_2 u^2} - \frac{c_1}{c_2} - x\right)^2.$$

This equation may be solved for x and, because $x = 1/n^2$, the result is

$$n^2 = 2c_2 u^2/[(1 - c_1 u^2)^2 - c_0^2 u^4]. \quad (3.15)$$

The function n^2 must be positive for all u on $(0, 1]$. This implies that

$$c_2 > 0. \quad (3.16)$$

Moreover, the denominator in (3.15) must be positive also, whence

$$(1 - c_1 u^2)^2 > c_0^2 u^4, \quad (3.17)$$

for all u on $[0, 1]$. The inequality is equivalent to

$$|1 - c_1 u^2| > |c_0|\, u^2,$$

and because the left-hand side may not vanish for all u on $[0, 1]$, $1 - c_1 u^2$ is positive. The inequality becomes

$$1 > (c_1 + |c_0|)u^2,$$

which is true for all u if it holds at $u = 1$. Thus (3.17) is satisfied if and only if

$$1 > c_1 \pm c_0. \quad (3.18)$$

The constant c_2 is not present in the function h because of Eqs. (3.13) for h_u and (3.15) for n^2. It therefore enters the metric (3.4) through n alone. Because the expression for n^2 only depends multiplicatively on c_2 and n in turn is always multiplied by D in the metric, c_2 is a scale factor and may be taken to be unity without loss of generality. Thus,

$$n^2 = 2u^2/[(1 - c_1u^2)^2 - c_0^2u^4], \tag{3.19}$$

and c_1 and c_0 must satisfy (3.18) in order that n^2 be positive. The derivative of n is needed for later use and is given by

$$\frac{n_u}{n} = \frac{1 + (c_0^2 - c_1^2)u^4}{[(1 - c_1u^2)^2 - c_0^2u^4]u}. \tag{3.20}$$

By means of Eq. (3.13) with $c_2 = 1$, Eq. (3.12) for ψ becomes

$$\psi = D\epsilon c_0^2 n^5. \tag{3.21}$$

Equation (3.13) with c_2 taken to be unity may be integrated with the aid of Eq. (3.19), which leads to

$$h = c_3 + \tfrac{1}{2} \ln \{1 + 2c_0u^2/[1 - (c_1 + c_0)u^2]\}.$$

Inequality (3.18) implies that $1 - (c_1 + c_0)u^2$ is always positive for all values of u; moreover, Eq. (3.13) implies that h is a monotonically increasing function if c_0 is positive and monotonically decreasing if c_0 is negative. The argument of the logarithm in the expression for h is always positive because

$$1 + \frac{2c_0u^2}{1 - (c_1 + c_0)u^2} = \frac{1 + (c_0 - c_1)u^2}{1 - (c_1 + c_0)u^2},$$

and inequality (3.18) implies that both numerator and denominator are positive. Finally, the function h enters the metric (3.4) additively with G. Thus, c_3 in the expression for h may be taken to be zero without loss of generality so that

$$e^{2h} = [1 - (c_1 - c_0)u^2]/[1 - (c_1 + c_0)u^2], \tag{3.22}$$

and its derivative is given by Eq. (3.13), with $c_2 = 1$, as

$$h_u = c_0n^2/u = 2c_0u/[(1 - c_1u^2)^2 - c_0^2u^4]. \tag{3.23}$$

This concludes the integration of Eqs. (3.10) and (3.11) whose integrals are (3.19) and (3.22). The generalizations of McVittie's cases (i) and (ii) are thus seen to be SDG solutions in which ψ is given by (3.21).

If the material is surrounded by empty space it follows that Eqs. (I.6.14), (I.6.16), and (I.6.17), for the mass function, the density, and the pressure, respectively, must hold. They are

$$m = M\left(\frac{r}{r_b}\right)^3 + \tfrac{1}{2}r\left[1 - \left(\frac{r}{r_b}\right)^2 + r^2\left(\frac{r_u}{r^2}\right)_b^2 - u^2\left(\frac{r_u}{r}\right)^2\right], \tag{3.24}$$

$$\frac{4\pi\mathfrak{G}\rho}{3c^2} = \frac{M - \psi_b}{r_b^3} + \int_u^1 \frac{\psi_u}{r^3}\, du, \tag{3.25}$$

$$\frac{p}{c^2} = e^{-\gamma}\int_u^1 \rho(e^{\gamma})_u\, du. \tag{3.26}$$

Thus, Eq. (3.25) for the density may be evaluated by means of Eq. (3.21) for ψ and (3.2) for r. Differentiation of (3.21) yields

$$\psi_u = 5D\epsilon c_0^2 n^4 n_u, \tag{3.27}$$

and thus (3.25) becomes

$$\frac{4\pi\mathfrak{G}}{15c^2} D^2(\rho - \rho_b) = \epsilon c_0^2 \int_u^1 \frac{nn_u\, du}{[\cosh(h + G) + \epsilon]^3}. \tag{3.28}$$

Equations (2.12), (3.2), and (3.21) furnish the form of ρ_b as

$$\frac{4\pi\mathfrak{G}}{3c^2} D^2\rho_b = \frac{M/D - \epsilon c_0^2 n_b^5}{n_b^3[\cosh(h_b + G) + \epsilon]^3}. \tag{3.29}$$

If the density is to be a monotonically decreasing function of u for any x^4, n_u must be positive when $\epsilon = +1$ and negative when $\epsilon = -1$. In Eq. (3.20) for n_u the denominator is always positive by condition (3.18). Hence the sign of n_u is that of the numerator, $1 + (c_0^2 - c_1^2)u^4$. This cannot be negative for any u in the $\epsilon = -1$ case and therefore the density increases monotonically from the center of the material outwards for some, and perhaps all, u. In the $\epsilon = +1$ case, however, if $1 + (c_0^2 - c_1^2)u^2$ is positive at the boundary,then n_u is positive for all u and the condition becomes

$$1 + c_0^2 - c_1^2 > 0, \quad \epsilon = +1. \tag{3.30}$$

If the density falls off monotonically from the center of the material toward the outer boundary at any x^4, then the condition for the density to be nonnegative is that the boundary density be nonnegative. The boundary density ρ_b given by Eq. (3.29) is nonnegative if

$$M/D \geq \epsilon c_0^2 n_b^5. \tag{3.31}$$

In the event that ρ is not a monotonically decreasing function of u, there is no simple requirement like (3.31) which ensures nonnegative density, although (3.31) is a necessary condition. When the equality holds in the $\epsilon = +1$ case,the boundary density is zero.

Equation (3.26) for the pressure may be evaluated once e^{γ} is known. Comparison of Eqs. (3.4) and

(2.2) shows that

$$e^{\gamma} = \sinh(h + G)\dot{G}H/\{\cosh(h + G) + \epsilon\},$$

and hence, since $\epsilon = \pm 1$,

$$\gamma_u = \epsilon h_u/\sinh(h + G).$$

Thus Eq. (3.26) becomes

$$\frac{p}{c^2} = \frac{\cosh(h + G) + \epsilon}{\sinh(h + G)} \int_u^1 \frac{\rho\epsilon h_u \, du}{\cosh(h + G) + \epsilon}, \quad (3.32)$$

and with the aid of the properties of the hyperbolic functions we have

$$p/c^2 = \tfrac{1}{2}\coth\tfrac{1}{2}(h + G)\int_u^1 \rho h_u \operatorname{sech}^2\tfrac{1}{2}(h + G)\, du, \quad \epsilon = +1,$$

$$= -\tfrac{1}{2}\tanh\tfrac{1}{2}(h + G)\int_u^1 \rho h_u \operatorname{csch}^2\tfrac{1}{2}(h + G)\, du, \quad \epsilon = -1. \quad (3.33)$$

Condition (3.6) which requires that $h + G$ be positive, coupled with the properties of n and H, implies that both the density and pressure have no divergences. Furthermore, p must be nonnegative in order that it be physically acceptable, and examination of Eqs. (3.33) shows that this is true if and only if ϵh_u is positive. The expression (3.23) for h_u implies that the sign of h_u is that of c_0. Therefore, the condition that p be nonnegative is equivalent to

$$c_0 \geq 0, \quad \epsilon = +1, \quad c_0 \leq 0, \quad \epsilon = -1. \quad (3.34)$$

When $c_0 = 0$, which may occur for either value of ϵ, the pressure is uniformly zero throughout the material. Moreover, when ϵc_0 is positive, it follows from Eq. (3.33) that the pressure falls off monotonically from the center of the material to zero at the boundary. Inequality (3.6) may now be written in a simpler form. Equation (3.23) implies that in the $\epsilon = +1$ case h is either a monotonically increasing function of u or a constant, while in the $\epsilon = -1$ case h is either a monotonically decreasing function of u or a constant. Thus, the minimum value of $h(u)$ when $\epsilon = +1$ is $h(0)$, which is zero by Eq. (3.22); but if $\epsilon = -1$ the minimum value of $h(u)$ is $h(1)$. Hence inequality (3.6) becomes

$$G(x^4) > 0, \quad \epsilon = +1; \quad G(x^4) > h_b, \quad \epsilon = -1, \quad (3.35)$$

for all x^4.

Because of inequality (3.7), $\dot{G}H$ is required to be a positive bounded dimensionless function $W(x^4)$, say, of x^4. Equation (3.5) for H may therefore be used to give

$$\dot{G}^2 = W^2(x^4)\left(\frac{c}{r_b}\right)^2\left[\frac{2M}{r_b} - 1 + \frac{1}{r_b^2}\left(\frac{\partial r}{\partial u}\right)_b^2\right].$$

If a new dimensionless time coordinate τ is introduced by

$$d\tau = \frac{Wc}{r_b}\, dx^4, \quad (3.36)$$

the equation for G in terms of τ becomes

$$\left(\frac{dG}{d\tau}\right)^2 = \frac{2M}{r_b} - 1 + \frac{1}{r_b^2}\left(\frac{\partial r}{\partial u}\right)_b^2. \quad (3.37)$$

Therefore, the metric (3.4) becomes

$$ds^2 = \left(\frac{Dn_b}{c}\sinh[h(u) + G(\tau)] \times \frac{\cosh[h_b + G(\tau)] + \epsilon}{\cosh[h(u) + G(\tau)] + \epsilon}\right)^2 d\tau^2 - \frac{D^2n^2}{c^2u^2}\{\cosh[h(u) + G(\tau)] + \epsilon\}^2 \times (du^2 + u^2\, d\Omega^2). \quad (3.38)$$

An important consequence of Eq. (3.37) is that in this solution of Einstein's equations the boundary may never pass inside the Schwarzschild radius, for if $2M/r_b > 1$ occurred for some G, then, as G became smaller, $\dot{G}^2$ in (3.37) would always be positive. Thus condition (3.35) could not be satisfied, for either G was negative at a previous time, corresponding to the positive root of (3.37) for $\dot{G}$, or G becomes negative at a later time, corresponding to the negative root of (3.37) for $\dot{G}$. Taub[4] has proven that the boundary of the uniform-density solution also has this property.

The nature of the function $G(\tau)$ determined by Eq. (3.37) will now be discussed. Because of the peculiar behavior of the density in the $\epsilon = -1$ case, only the $\epsilon = +1$ case will be considered. Inequality (3.35) then implies that the only acceptable solutions of (3.37) are those in which G is positive for all values of τ. If, in addition, G is bounded from above, the solution is oscillatory, while, if G may become arbitrarily large, a "bounce" type solution occurs. A "bounce" is a solution that refers to a case in which the material, at a moment in the infinitely remote past, was spread out throughout all space. It then contracted to a sphere of finite radius and thereafter will expand again to a state in which it fills all space at a moment in the infinitely remote future. The mathematical characteristic of a "bounce" solution is that the continuous function r possesses the properties that

$$\lim_{\tau \to \pm\infty} r(u, \tau) = \infty,$$

while $r(u, \tau)$ is finite for finite τ. In such a solution the pressure overwhelms the attraction of gravity.

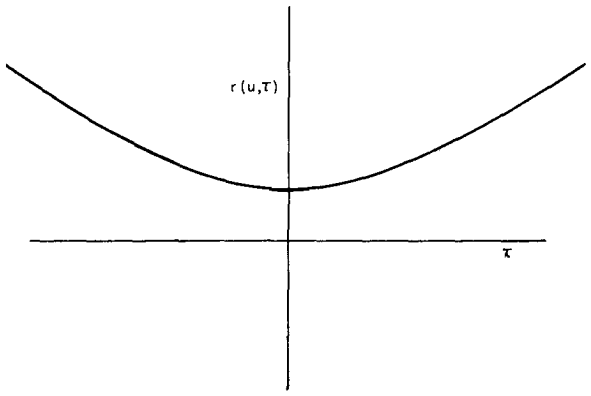


FIG. 1. Typical behavior of r in a "bounce."

The typical behavior of $r(u, \tau)$ for a fixed u is illustrated in Fig. 1.

By means of Eqs. (3.2) and (3.8), Eq. (3.37) becomes

$$\left(\frac{dG}{d\tau}\right)^2 = \frac{2M/D}{n_b[\cosh(h_b + G) + 1]} - 1 + \left[\left(\frac{n_u}{n}\right) + \frac{h_u \sinh(h + G)}{\cosh(h + G) + 1}\right]_b^2,$$

which may be expressed with the aid of the hyperbolic half-argument formulas as

$$\left(\frac{dG}{d\tau}\right)^2 = \left[\frac{M}{Dn} + \left(\frac{n_u}{n}\right)^2 - 1 + 2\left(\frac{h_u n_u}{n}\right)\tanh\tfrac{1}{2}(h + G) + \left((h_u)^2 - \frac{M}{Dn}\right)\tanh^2\tfrac{1}{2}(h + G)\right]_b.$$

This differential equation may be simplified with the aid of the definitions

$$A = \frac{M}{Dn_b} - (h_u^2)_b, \quad B = \left(\frac{h_u n_u}{n}\right)_b,$$
$$C = 1 - \frac{M}{Dn_b} - \left(\frac{n_u}{n}\right)_b^2. \tag{3.39}$$

The constant A defined in Eq. (3.39) may be written as

$$A = (M/D - c_0^2 n_b^5)/n_b,$$

and is therefore positive by virtue of inequality (3.31), provided that ρ_b does not vanish. The constant B is also positive because in the $\epsilon = +1$ case inequality (3.34) implies that h_u as given in (3.23) is positive and (3.30) implies that n_u is positive. The constant C will turn out to be positive in situations of physical interest because of the condition that G always be positive, which will be examined shortly. The differential equation for G now becomes

$$\left(\frac{dG}{d\tau}\right)^2 = -A\tanh^2\tfrac{1}{2}(h_b + G) + 2B\tanh\tfrac{1}{2}(h_b + G) - C. \tag{3.40}$$

This equation may be integrated with the aid of the substitution $x = \tanh\frac{1}{2}(h_b + G)$, but it is merely desired to show that G is positive here. The roots of $(dG/d\tau)^2$ will determine the bounds on G and these roots are

$$\tanh\tfrac{1}{2}(h_b + G_\pm) = [B \pm (B^2 - AC)^{\frac{1}{2}}]/A, \tag{3.41}$$

where

$$0 < G_- \leq G \leq G_+ \leq \infty \tag{3.42}$$

and $A \neq 0$. In order for the roots to be real,

$$B^2 \geq AC. \tag{3.43}$$

When the equality holds in (3.43) both roots are equal and the oscillations have zero amplitude, and hence the solution is a static one. It is also interesting to note that oscillations cannot occur if c_0 vanishes, for then B would vanish also. But Eqs. (3.33) and (3.23) show that c_0 vanishes if and only if the pressure is zero for all x^1 and x^4. Thus, a pressure is essential for the existence of oscillations.

If G_- is to be positive, then the condition

$$\tanh\tfrac{1}{2}h_b < [B - (B^2 - AC)^{\frac{1}{2}}]/A, \tag{3.44}$$

where

$$[B - (B^2 - AC)^{\frac{1}{2}}]/A < 1 \tag{3.45}$$

must be imposed because the hyperbolic tangent of a finite positive quantity is less than unity. If G_+ is to be finite, then

$$[B + (B^2 - AC)^{\frac{1}{2}}]/A < 1. \tag{3.46}$$

Thus, for the $\epsilon = +1$ case any set of constants c_0, c_1, and M/D, which satisfy inequalities (3.18), (3.30), (3.31), (3.34), (3.43), (3.44), and (3.45), determine a solution of Einstein's equations for which the metric coefficients g_{44} and g_{11} never vanish while r is positive and vanishes at the center only and for which the density and pressure are nonnegative and monotonically decreasing from the center of the material outwards. Also, if inequality (3.46) is satisfied, the solution is oscillatory and g_{11}, g_{44} are bounded, while if it is not satisfied the solution is of the "bounce" type.

4. EXAMPLES OF McVITTIE'S CLASS OF NONUNIFORM-DENSITY SDG SOLUTIONS

The purpose of this section is to demonstrate that values of c_0, c_1, and M/D, which satisfy the inequalities derived in Sec. 3 for $\epsilon = +1$, do indeed exist. A "bounce" and an oscillatory solution are developed. The method of attack consists of solving all the inequalities in Sec. 3 which involve c_0 and c_1 alone. The solution consists of a region in the (c_1, c_0) plane of values of c_0 and c_1 for which there may be values of

M/D such that the remaining inequalities are satisfied. Two values of c_1 and c_0 are selected and a range of values of M/D is then found so that the remaining inequalities are satisfied.

In either a "bouncing" or oscillating solution where $\epsilon = +1$, the inequalities which depend on c_1 and c_0 alone are (3.18),

$$1 > c_1 \pm c_0$$

(the condition for n and h to be real and bounded), (3.30),

$$1 > c_1^2 - c_0^2$$

(the condition that the density be a monotonically decreasing function of u at any τ), and (3.34),

$$c_0 > 0,$$

which is used in the strong form to ensure that the pressure is always positive in the interior of the material. Each of these inequalities defines a region of inadmissible values of c_0 and c_1 in the(c_1, c_0)plane and the boundaries are given in Fig. 2. The only permissible region lies within *GFID*. That this region of possible acceptable values of c_0 and c_1 is divided into possible bounces and possible oscillations by the line *OR* will be demonstrated.

The inequalities that c_0, c, and M/D must satisfy and which are common to both "bouncing" and oscillating solutions are: (3.31),

$$M/D \geq c_0^2 n_b^5$$

(which ensures that the boundary density is non-negative for all values of τ), (3.43),

$$B^2 \geq AC$$

(which, with the definitions (3.39), ensures that the roots of $(dG/d\tau)^2$ are real and distinct if $B^2 > AC$), and (3.44) and (3.45),

$$\tanh \tfrac{1}{2}h_b < [B - (B^2 - AC)^{\frac{1}{2}}]/A,$$
$$[B - (B^2 - AC)^{\frac{1}{2}}]/A < 1$$

(which together imply that G_- is positive and finite). If the inequality (3.46),

$$[B + (B^2 - AC)^{\frac{1}{2}}]/A < 1,$$

is satisfied, then G_+ is finite and the solution is oscillatory while, if this inequality is not satisfied, there is no finite second root of $(dG/d\tau)^2$, the solution G is therefore unbounded, and $\lim G = \infty$ as $\tau \to \pm\infty$. The solution is a "bounce" when G is unbounded at $\tau = \pm\infty$ because the continuous function $r(u, \tau)$ satisfies both $\lim r(u, \tau) = \infty$, as $\tau \to \pm\infty$, and $r(u, \tau \neq \infty) \neq \infty$, and hence the material was spread out over all space in the distant past, contracted to a finite size, and then will expand outward to occupy all space in the distant future. If an oscillation occurs, then Eq. (3.46) implies that

$$2B < A + C,$$

which, by means of Eq. (3.39), becomes

$$(h_u + n_u/n)_b^2 < 1.$$

Because h_u and n_u/n are positive by inequalities (3.34) and (3.30), respectively, this inequality becomes

$$(h_u + n_u/n)_b < 1,$$

and, with the aid of equations (3.23) for $(h_u)_b$ and (3.20) for $(n_u/n)_b$, this becomes

$$2c_0 + 1 + c_0^2 - c_1^2 < (1 - c_1)^2 - c_0^2,$$

and thus

$$0 < (c_1 + c_0)(-1 + c_1 - c_0). \qquad (4.1)$$

If the solution "bounces," then (4.1) does not hold and so the inequality (4.1) divides the (c_1, c_0) plane into regions of points for which either "bounces" or oscillations, but not both, may occur. These regions and their boundaries are illustrated in Fig. 3. All the boundary points do not satisfy (3.46) and they are therefore "bounce" possibilities. The boundary line $c_1 = -c_0$ thus divides the region of points which

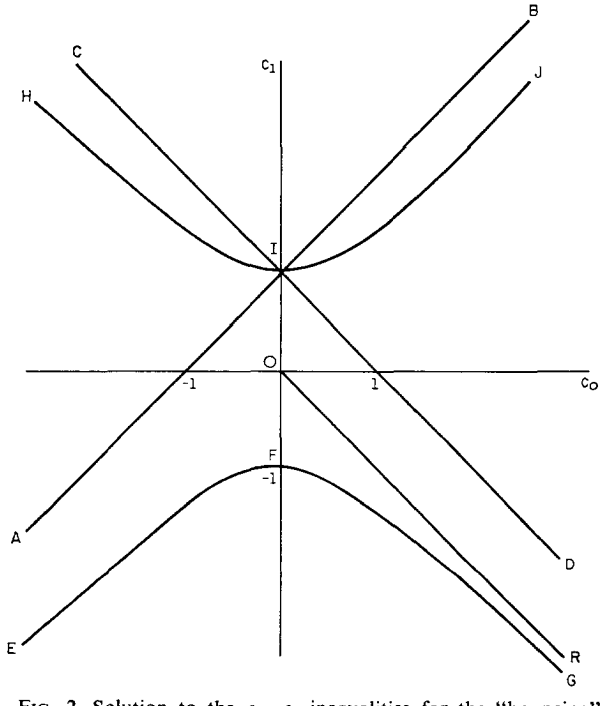


FIG. 2. Solution to the c_0, c_1 inequalities for the "bouncing" solutions and oscillating solutions. *AB* is the line $1 = c_1 - c_0$; *CD* is the line $1 = c_1 + c_0$; *EFG* and *HIJ* are the branches of the hyperbola $1 = c_1^2 - c_0^2$; *OR* is the line $c_1 = -c_0$.

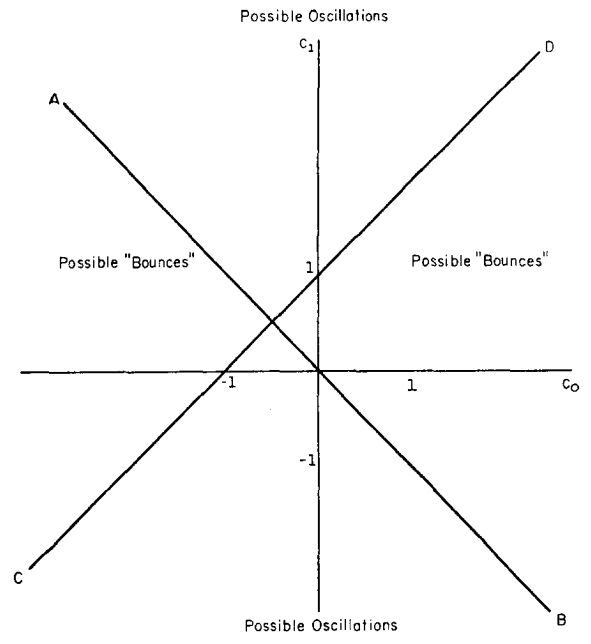


FIG. 3. Solution of the inequality (3.46). AB is the line $c_1 = -c_0$ and CD is the line $c_1 - c_0 = 1$.

satisfy inequalities (3.18), (3.30), and (3.34) illustrated in Fig. 2 into "bounce" possibilities which include $c_0 = -c_1$ (region ROID), and oscillation possibilities (region GFOR).

A set of "bouncing" solutions will now be constructed. In this $\epsilon = +1$ solution, the values of c_0 and c_1 are given by

$$c_0 = \tfrac{1}{2}, \quad c_1 = -\tfrac{1}{2}. \tag{4.2}$$

Values of M/D which satisfy inequalities (3.31) and (3.43)–(3.45) but not (3.46) will be found. Equations (4.2) and (3.19) give, at the boundary $u = 1$,

$$n_b = 1, \tag{4.3}$$

while Eq. (3.20) gives

$$(n_u/n)_b = \tfrac{1}{2}. \tag{4.4}$$

Equation (3.22) for h yields

$$h_b = \tfrac{1}{2} \ln 2, \tag{4.5}$$

and Eq. (3.23) for h_u becomes

$$(h_u)_b = \tfrac{1}{2}. \tag{4.6}$$

Equation (4.5) may be used to evaluate $\tanh \frac{1}{2}h_b$ as

$$\tanh \tfrac{1}{2}h_b = (\sqrt{2} - 1)/(\sqrt{2} + 1). \tag{4.7}$$

Equations (4.3), (4.4), and (4.6) may be used to evaluate A, B, and C defined by Eq. (3.39) which thereby become

$$A = \frac{M}{D} - \frac{1}{4}, \quad B = \frac{1}{4}, \quad C = \frac{3}{4} - \frac{M}{D}. \tag{4.8}$$

Inequality (3.31) becomes

$$M/D \geq \tfrac{1}{4}, \tag{4.9}$$

and inequality (3.43) becomes

$$\left(\frac{1}{4}\right)^2 \geq \left(\frac{M}{D} - \frac{1}{4}\right)\left(\frac{3}{4} - \frac{M}{D}\right) = \left[\left(\frac{M}{D} - \frac{1}{2}\right) + \frac{1}{4}\right]\left[\frac{1}{4} - \left(\frac{M}{D} - \frac{1}{2}\right)\right],$$

which is always true. Inequality (3.45) with the aid of (4.8) becomes

$$\frac{\frac{1}{4} - [(\frac{1}{4})^2 - (M/D - \frac{1}{4})(\frac{3}{4} - M/D)]^{\frac{1}{2}}}{M/D - \frac{1}{4}} < 1,$$

but because the bracketed term is $(M/D - \frac{1}{2})^2$ the inequality reduces to

$$M/D > \tfrac{1}{2}. \tag{4.10}$$

Inequality (3.44) with the aid of Eqs. (4.7) and (4.8) becomes

$$\frac{\sqrt{2} - 1}{\sqrt{2} + 1} < \frac{\frac{3}{4} - M/D}{M/D - \frac{1}{4}},$$

and thus

$$M/D < \tfrac{1}{2} + \sqrt{2}/8. \tag{4.11}$$

Inequalities (4.9)–(4.11) are satisfied if

$$\tfrac{1}{2} < M/D < \tfrac{1}{2} + \sqrt{2}/8. \tag{4.12}$$

It remains to show that inequality (3.46) does not hold so that G is unbounded and a "bounce" occurs. This is true because $[B^2 + (B^2 - AC)^{\frac{1}{2}}]/A = 1$. This completes the demonstration that "bounce" solutions having all the properties cited at the end of the last section do exist.

The fact that oscillating solutions exist will now be demonstrated. In Fig. 2 it may be seen that the point (c_0, c_1) for the previous bounce solution lies on the line separating the possible "bounces" from the possible oscillations. Thus, the values

$$c_0 = \tfrac{1}{2} - \delta, \quad c_1 = -\tfrac{1}{2}, \tag{4.13}$$

where δ is chosen so that

$$1 \gg \delta > 0, \tag{4.14}$$

are in the possible oscillation region of Fig. 2. Moreover, (4.12) gives a finite range of values of M/D such that inequalities (3.31) and (3.43)–(3.45) are satisfied. All expressions resulting in these inequalities are bounded and continuous for values of c_0 and c_1 in the vicinity of $c_0 = \frac{1}{2}, c_1 = -\frac{1}{2}$. Therefore, there must also be a finite set of values of M/D satisfying (3.31) and (3.43)–(3.45) for c_0 and c_1 given by (4.13). The new

region differs from the one given by (4.12) in that the endpoints of the new region are displaced from the endpoints of the old region by amounts proportional to δ. The endpoints of the new region have been calculated to first order in δ, but these results are not presented here. This completes the demonstration that both oscillating and "bouncing" cases occur in McVittie's class of nonuniform-density SDG solutions and that they are solutions in which the pressure and density are bounded positive functions which are monotonically decreasing in u. Furthermore, in these solutions g_{44} and g_{11} are always positive while r vanishes only at the origin.

APPENDIX

In this appendix the functions $\bar{n}$ and $\bar{h}$ as defined in (2.23) for McVittie's case (i) are found. For the special case in which Eq. (2.23) holds, r is given by Eq. (2.24) when δ approaches zero and thus

$$r = D\bar{n}(\bar{h} + \bar{G})^2.$$

If the transformation (2.23) is applied to Eq. (3.9), the conditions that ζ_1, $\delta\zeta_2$, and ζ_3 vanish as δ approaches zero give rise to the three equations

$$\frac{\bar{n}_{uu}}{\bar{n}} - \frac{2\bar{n}_u^2}{\bar{n}^2} + \frac{\bar{n}_u}{u\bar{n}} + \frac{1}{u^2} = 0, \tag{A1}$$

$$\bar{n}\bar{h}_{uu} - 2\bar{n}_u\bar{h}_u + \bar{n}\bar{h}_u/u = 0, \tag{A2}$$

$$\psi = -D\bar{h}_u^2\bar{n}u^2. \tag{A3}$$

The solution of Eq. (A1) may be obtained from the solution of (3.10) because these two equations differ only in that the latter contains a $-(h_u)^2$ term. Because of the first member of Eq. (3.23) this term gives rise to the c_0^2 term in Eq. (3.19). The solution of equation (A1) thus results when the c_0^2 term of Eq. (3.19) is omitted. Hence

$$\bar{n}^2 = 2u^2/(1 - c_1u^2)^2. \tag{A4}$$

When Eq. (A4) is substituted into the first term of Eq. (3.23), the result may be integrated to give

$$\bar{h} = c_0u^2/(1 - c_1u^2), \tag{A5}$$

where, as before, in the derivation of h, an additive constant of integration has been equated to zero. Clearly both $\bar{n}$ and $\bar{h}$ are bounded monotonic functions if and only if

$$1 > c_1. \tag{A6}$$

Equation (3.12) with the aid of the first term of Eq. (3.23) becomes

$$\psi = -Dc_0^2\bar{n}^5.$$

[1] G. C. McVittie, Ann. Inst. Henri Poincaré **6**, 1 (1967).
[2] I. H. Thompson and G. J. Whitrow, Monthly Notices Roy. Astron. Soc. **136**, 207 (1967).
[3] M. E. Cahill and G. C. McVittie, J. Math. Phys. **11**, 1382 (1970).
[4] A. H. Taub, Ann. Inst. Henri Poincaré **9**, 153 (1968).
[5] H. Nariai, Progr. Theoret. Phys. (Kyoto) **38**, 740 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Coulomb Scattering. I. Single Channel*

DENIS MULHERIN
Department of Physics, Manhattan College, Bronx, New York
AND
I. I. ZINNES
Department of Physics, Fordham University, Bronx, New York

(Received 29 May 1968; Revised Manuscript Received 30 December 1968)

With the aid of a Coulomb "free propagator" $U_c^{(\pm)}(t)$ constructed from the asymptotic form of the conventional time-independent solution to the scattering problem in a pure Coulomb field, Møller wave operators are shown to exist for general Coulomb-like (Coulomb + short-range) potentials. It is emphasized that when scattering occurs in such long-range fields, one should distinguish three classes of states: the prepared and detected states, the scattering states, and certain asymptotic states which are peculiar to long-range potentials. The existence of the third class as a class distinct from the first suggests a connection between $U_c^{(\pm)}(t)$ and the quantum theory of measurements.

I. INTRODUCTION

The time-dependent theory of potential scattering, first formulated in a mathematically satisfactory form by Jauch,[1,2] has been the subject of considerable study during the past decade. A review of this work emphasizing the mathematical problems can be found in a recent book by Putnam.[3] Following Møller[4] and Cook,[5] Jauch defines a single-channel scattering system as one whose essential property is that one can associate with the given scattering system the two wave or Møller operators $\Omega^{(\pm)}$, which are defined by the relations

$$\Omega^{(\pm)} = \lim_{t\to\pm\infty} e^{iHt}e^{-iH_0t}. \tag{1}$$

Here $H = H_0 + V$ and $H_0 = -\Delta/2m$ (we take units with $\hbar = 1$) are the full and the free Hamiltonians respectively, and the limits are taken in the strong topology of linear operators on a Hilbert space. When these operators exist, they can be applied to an arbitrary[6–9] wavepacket or state Φ in the space of square-integrable functions $L^2(E_3)$ to yield

$$\Psi^{(\pm)} = \Omega^{(\pm)}\Phi. \tag{2}$$

The significance of the states $\Psi^{(\pm)}$ and Φ can be obtained by combining (1) and (2), which imply that

$$\|e^{-iHt}\Psi^{(\pm)} - e^{-iH_0t}\Phi\| \to 0, \quad \text{as} \quad t\to\pm\infty. \tag{3}$$

Thus, $\Psi^{(\pm)}$ can be interpreted as the scattering state

$$\Psi^{(\pm)}(t) = e^{-iHt}\Psi^{(\pm)}$$

at time $t = 0$, which at long times in the past $(-)$ [future $(+)$] becomes indistinguishable from the state $\Phi(t) = e^{-iH_0t}\Phi$. The latter state is developing in time like a conventional free system for which $V = 0$. Therefore, it seems clear that, because of the zero limit in (3), one should identify $\Phi(t)$ at large negative (positive) times with the beam of projectiles which one has prepared for the scattering (or which one will detect after the scattering). This identification of $\Phi(t)$ with the prepared (detected) beam is consistent with the theorem[10] that, if $H_0 = -\Delta/2m$, then at large times the probability of finding a projectile in any finite region of space when it is in the state $\Phi(t) = e^{-iH_0t}\Phi$ approaches zero as $t\to\pm\infty$. Thus, $\Phi(t)$ describes a freely moving system which at large enough times is far from any finite region, just where the projectile originates (or where it is detected) in an idealization of a real scattering experiment. As a result of this identification of $\Phi(t)$ with the prepared (detected) state, relation (3) asserts the existence of scattering states $\Psi^{(\pm)}$ at $t = 0$ which arose from $(-)$ [will give rise to $(+)$] the freely moving prepared (detected) asymptotic state $\Phi(t)$.

According to (2), one may define the wave operator $\Omega^{(\pm)}$ as that operator which converts the prepared (detected) state into the corresponding scattering state. Useful sufficient conditions[5,8,9] for the existence of $\Omega^{(\pm)}$ are that $V(\mathbf{r})$ be either (1) square integrable, or (2) locally square integrable and $O(r^{-\beta})$, $\beta > 1$, as $r\to\infty$. Such potentials will be called short-range potentials.

These conditions clearly exclude the Coulomb potential α/r. In fact, it can be shown[11] that there is no wave operator which satisfies (1), and consequently (3), when the Coulomb potential is involved. This conclusion is reasonable if it is recalled, from the conventional time-independent scattering theory, that the scattering state corresponding to a pure Coulomb interaction is asymptotic (in space) not to a plane wave, but to a distorted plane wave. [See Eq. (14) below.] Thus, in the distant past, for example, one

must expect the scattering state to approximate an asymptotic state which is *not* identical with that of the beam prepared (and monitored) in the absence of the scatterer. And it is the state of the latter beam at $t = 0$, say, which contains the initial data for the scattering. Thus, for scattering carried out in Coulomb or Coulomb-like (Coulomb + short-range potential) fields one becomes involved with three states[12]:

(a) $\Phi_0(t)$, the state of the projectile prepared in the absence of the scatterer (we shall refer to this state as the prepared state);

(b) $\Psi^{(-)}(t)$, the resulting scattering state;

(c) $\Phi(t)$, the asymptotic state to which $\Psi^{(-)}(t)$ converges as $t \to -\infty$, and which is indistinguishable from $\Phi_0(t)$ in the short-range case.

The scattering problem now consists in relating $\Psi^{(-)}$ (and also $\Psi^{(+)}$) with a given Φ_0. It can also be formulated by the relations (1) and (3) in which the conventional free propagator e^{-iH_0t} is replaced by a more general "free" propagator, one which either converts Φ_0 into Φ and propagates the asymptotic state in a suitable manner, or propagates Φ_0 as a conventional free particle and converts it at any time t into a suitable asymptotic state. In either case, we must replace Eqs. (1)–(3) with the relations

$$\Omega^{(\pm)} = \lim_{t \to \pm\infty} e^{iHt}U(t), \tag{4}$$

$$\Psi^{(\pm)} = \Omega^{(\pm)}\Phi_0, \tag{5}$$

$$\lim_{t \to \pm\infty} \|e^{-iHt}\Psi^{(\pm)} - U(t)\Phi_0\| = 0, \tag{6}$$

where $U(t)$ is a suitable generalized "free" propagator. These relations become identical with the earlier ones if, for the short-range potentials, we define $U(t) = e^{-iH_0t}$.

In the more general case, when we are concerned with the Coulomb or Coulomb-like potentials, the relations (4)–(6) pose an ambiguous problem: Given Φ_0, an experimentally determined prepared state, and the full Hamiltonian H for the interacting system, find $\Psi^{(\pm)}$ and $U(t)$ satisfying (6). Clearly, the problem is indeterminate, since any $U(t)$ for which the limits (4) exist determines some $\Psi^{(\pm)}$, while any $\Psi^{(\pm)}$ determines the asymptotic behavior of $U(t)$. Because of this ambiguity, we are no longer guaranteed, as we are in the short-range case, that a solution of (6) will yield the scattering states appropriate to the prepared state Φ_0.

In this paper, we present a solution to the Coulomb problem assuming that the relation between Φ_0 and $\Psi^{(\pm)}$ is known. We then prove that a particular $U(t)$, whose structure is physically plausible, satisfies (6) which then gives rise to the wave operator and the relation (5).

The assumed relation between Φ_0 and $\Psi^{(\pm)}$ is a generalization of a result proved by Ikebe[13] for a special class of very smooth short-range potentials, in particular, those potentials which vanish at infinity at least as fast as r^{-2-h}, $h > 0$. This result asserts that, if $f \in L^2$ and if $\hat{f}$ is the Fourier transform of f, then

$$(\Omega^{(\pm)}f)(\mathbf{r}) = \text{l.i.m.}\,(2\pi)^{-\frac{3}{2}}\int d\mathbf{k}\psi^{(\pm)}(\mathbf{k},\mathbf{r})\hat{f}(\mathbf{k}), \tag{7}$$

where the $\psi^{(\pm)}(\mathbf{k},\mathbf{r})$ are the conventional stationary scattering states which satisfy the spatial asymptotic condition

$$\psi^{(-)}(\mathbf{k},\mathbf{r}) \approx e^{i\mathbf{k}\cdot\mathbf{r}} + g(\vartheta,\psi)e^{ikr}/r, \tag{8}$$

$$\psi^{(+)}(\mathbf{k},\mathbf{r}) = \psi^{(-)*}(-\mathbf{k},\mathbf{r}) \tag{8'}$$

and the $\Omega^{(+)}$ are the short-range wave operators defined in (1). Following this structure, which seems physically plausible, we shall assume for the pure Coulomb case that, if $\hat{f}(\mathbf{k}) \in L^2$ and is the Fourier transform of the prepared state Φ_{0f}, i.e., if

$$\Phi_{0f}(\mathbf{r}) = (2\pi)^{-\frac{3}{2}}\int d\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{r}}\hat{f}(\mathbf{k}), \tag{9}$$

then

$$\Psi_{cf}^{(\pm)}(\mathbf{r}) = (2\pi)^{-\frac{3}{2}}\int d\mathbf{k}\psi_c^{(\pm)}(\mathbf{k},\mathbf{r})\hat{f}(\mathbf{k}) \tag{10}$$

are the physically associated scattering states, where

$$\psi_c^{(-)}(\mathbf{k},\mathbf{r}) = \Gamma(1 + i\lambda/k)e^{-\pi\lambda/2k}e^{i\mathbf{k}\cdot\mathbf{r}} \times {}_1F_1\,(-i\lambda/k|\,1\,|i(kr - \mathbf{k}\cdot\mathbf{r})),$$

$$\psi_c^{(+)}(\mathbf{k},\mathbf{r}) = \psi_c^{(-)*}(-\mathbf{k},\mathbf{r}) \tag{11}$$

are the conventional stationary Coulomb[14] scattering states. Here $\lambda = m\alpha$, where m is the mass; $V(\mathbf{r}) = \alpha/r$ is the Coulomb potential energy; and ${}_1F_1$ is Kummer's form of the confluent hypergeometric function. Therefore, the states that play the role of asymptotic states may be expected to be constructed from the asymptotic forms of $\psi_c^{(\pm)}(\mathbf{k},\mathbf{r})$. Thus, associated with the states in (9) and (10), we define the asymptotic states

$$\Phi_{cf}^{(\pm)}(\mathbf{r}) = (2\pi)^{-\frac{3}{2}}\int d\mathbf{k}\phi_c^{(\pm)}(\mathbf{k},\mathbf{r})\hat{f}(\mathbf{k}), \tag{12}$$

where

$$\phi_c^{(\pm)}(\mathbf{k},\mathbf{r}) = \exp i[\mathbf{k}\cdot\mathbf{r} \mp (\lambda/k)\ln(kr \pm \mathbf{k}\cdot\mathbf{r})] \tag{13}$$

and

$$\psi_c^{(-)}(\mathbf{k},\mathbf{r}) \approx \phi_c^{(-)}(\mathbf{k},\mathbf{r}) + f_c(\vartheta)\{\exp i[kr - (\lambda/k)\ln 2kr]\}/r. \tag{14}$$

Referring to (9) and (10), we note that, since the Fourier and Coulomb transforms are isometric on L^2, we have,[11] for $f \in L^2$,

$$\|\Phi_{0f}\| = \|f\|, \tag{15a}$$

$$\|\Psi_{cf}^{(\pm)}\| = \|f\|. \tag{15b}$$

On the other hand, for $f \in \mathcal{M} \subset L^2$, where $\mathcal{M}$ is a space whose elements are dense in L^2 and is defined in the next section, it can be shown that, as a result of several judicious integrations by parts,

$$\|\Phi_{cf}^{(\pm)}\| < \infty, \quad f \in \mathcal{M}. \tag{16}$$

[We conjecture that the norm in (16) is, in fact, equal to $\|f\|$ with $f \in L^2$. We shall discuss this point further in Sec. III.]

With these definitions, we shall prove the following in the next section:

Theorem 1: The operators $\Omega_c^{(\pm)}$ which map the prepared (or detected) state Φ_{0f} onto the Coulomb scattering states $\Psi_{cf}^{(\pm)}$, i.e., the operators $\Omega_c^{(\pm)}$ defined by the relation

$$\Omega_c^{(\pm)}\Phi_{0f} = \Psi_{cf}^{(\pm)}, \quad f \in L^2(E_3), \tag{17}$$

are the unique continuous extensions of the operators $\hat{\Omega}_c^{(\pm)}$. The latter operators, in turn, are defined on $\mathcal{M}$ by the following relations. If

$$H = -\Delta/2m + \alpha/r, \quad \alpha = \text{const}, \tag{18}$$

and the "free Coulomb propagator" $U_c^{(\pm)}(t)$ is defined for $f \in \mathcal{M}$ by

$$U_c^{(\pm)}(t)\Phi_{0f}(\mathbf{r}) = (2\pi)^{-\frac{3}{2}} \int d\mathbf{k}\, \phi_c^{(\pm)}(\mathbf{k}, \mathbf{r}) e^{-ik^2t/2m} \hat{f}(\mathbf{k}), \tag{19}$$

then

$$\hat{\Omega}_c^{(\pm)} = \lim_{t \to \pm\infty} e^{iHt} U_c^{(\pm)}(t), \quad \text{on } \mathcal{M}, \tag{20}$$

and

$$\lim_{t \to \pm\infty} \|e^{-iHt}\Psi_{cf}^{(\pm)} - U_c^{(\pm)}(t)\Phi_{0f}\| = 0, \quad f \in \mathcal{M}. \tag{21}$$

Recently, motivated by considerations other than ours, Dollard[11,15] proved (21) for $f \in L^2$ using the following "free" propagator in place of our $U_c^{(\pm)}(t)$:

$$U_D(t) = e^{-iH_D(t)}, \tag{22}$$

where

$$H_D(t) = -\frac{\Delta}{2m}t + \frac{\epsilon(t) m e_1 e_2}{(-\Delta)^{\frac{1}{2}}} \ln\left(-\frac{2|t|\Delta}{m}\right)$$

and $\epsilon(t) = \pm 1$ when $t \gtrless 0$. Thus Dollard has proven the existence of wave operators

$$\Omega_D^{(\pm)} = \lim_{t \to \pm\infty}{}_S\, e^{iHt} U_D(t), \tag{23}$$

which satisfy (17). Assuming the validity of Theorem 1, we have the result

$$\Omega_D^{(\pm)} = \Omega_c^{(\pm)}. \tag{24}$$

In addition, he has proven the existence of the limit in (23) when the potential in the full Hamiltonian is Coulomb-like. We have occasion to use both of these results to facilitate the proof of Theorem 1 and its extension to Coulomb-like potentials.

II. PROOF OF THEOREM 1

The proof of Theorem 1 will be carried out with the aid of several lemmas. In order to be able to integrate by parts, and to be able to use the asymptotic expansion of the confluent hypergeometric functions which arise in the Coulomb problem, it is convenient to introduce a linear space of testing functions $\mathcal{M}$ defined in the following way: $f(\mathbf{r}) \in \mathcal{M}'(\hat{n})$ if its Fourier transform $\hat{f}(\mathbf{k})$ satisfies the following conditions in momentum ($\mathbf{k}$) space. (See Fig. 1.)

(a) $\hat{f}(\mathbf{k}) = 0$ inside some sphere of radius $\rho > 0$ with center at the origin.

(b) $\hat{f}(\mathbf{k}) = 0$ outside a region $\mathcal{R}_2'$ bounded by an infinitely long right circular cone, whose apex is at the origin, and whose generators make an angle $\Theta' \leq \frac{1}{6}\pi - \delta$, $\delta > 0$, with the central axis of the cone, whose direction $\hat{n}$ is arbitrary.

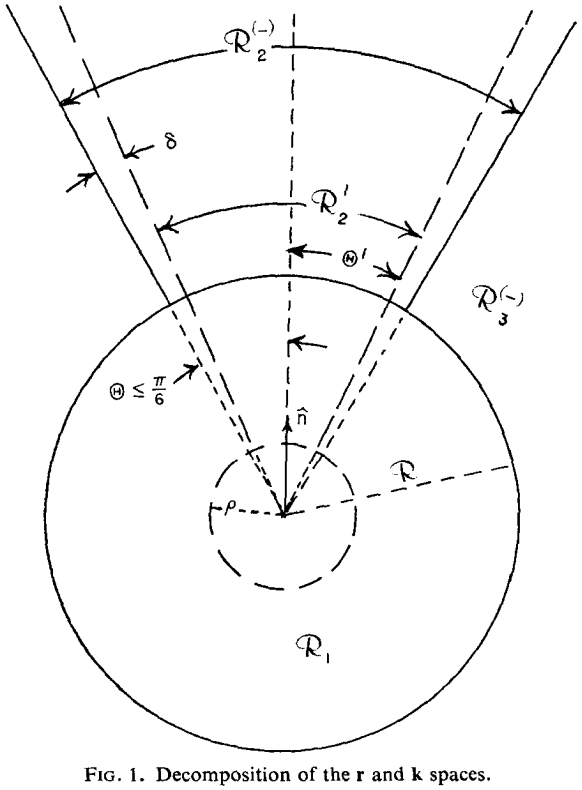


FIG. 1. Decomposition of the r and k spaces.

(c) $\hat{f}(\mathbf{k})$ has continuous derivatives of at least third order in the components k_1, k_2, k_3 of $\mathbf{k}$.

(d) $\hat{f}(\mathbf{k})$ and all its derivatives to third order are functions of rapid decrease,[16] i.e., functions which vanish when $k \to \infty$ faster than any positive power of $1/k$, i.e.,

$$|\mathbf{k}|^n \left| \frac{\partial^{m_1+m_2+m_3}}{\partial k_1^{m_1} \partial k_2^{m_2} \partial k_3^{m_3}} \hat{f}(\mathbf{k}) \right| \quad \text{is bounded}$$

for arbitrary integers n and for integers m_1, m_2, and m_3 such that $0 \leq m_1 + m_2 + m_3 \leq 3$. We now define $\mathcal{M}$ as the space of functions which are finite sums of functions in the spaces $\{\mathcal{M}'(\hat{n})\}$, each space defined for an arbitrary direction $\hat{n}$, i.e., $\mathcal{M}$ is the span of the union of the family $\{\mathcal{M}'(\hat{n})\}$. Using the properties of the Schwartz space[16] $\mathcal{S}$ of testing functions, in particular, that $\mathcal{S}$ is dense in L^2, it can be shown that our space of functions $\mathcal{M}$ is not empty and is, in fact, dense in L^2. Using a standard argument from the theory of Hilbert space, we are able to assert the existence of an operator on the whole of the Hilbert space $\mathcal{H}$ of L^2-functions as the unique continuous extension of an operator shown to exist and to be bounded on $\mathcal{M}$.

The strategy to be used in proving Theorem 1 consists in decomposing configuration space into three disjoint regions $\mathcal{R}_i^{(\pm)}$, $i = 1, 2, 3$, and proving that, for arbitrary $\Phi_{0f} \in \mathcal{M}'(\hat{n})$,

$$\lim_{t \to \pm\infty} \|\mathcal{F}_i^{(\pm)}[e^{-iHt}\Psi_{cf}^{(\pm)} - U_c^{(\pm)}(t)\Phi_{0f}]\| = 0, \quad (25)$$

$i = 1, 2, 3$, where $\mathcal{F}_i^{(\pm)}$ is a projection into region $\mathcal{R}_i^{(\pm)}$. The region $\mathcal{R}_i^{(\pm)} = \mathcal{R}_1$ is bounded by a sphere of radius R centered at the origin of configuration space. (Since the configuration and momentum spaces are isomorphic, we consider them superimposed in Fig. 1.) The remaining two regions depend on the special subspace $\mathcal{M}'(\hat{n})$ being considered, in particular, on the values of Θ and $\hat{n}$ describing the geometry of $\mathcal{M}'(\hat{n})$. Thus, for a given $\mathcal{M}'(\hat{n})$, $\mathcal{R}_2^{(-)}$ is bounded by $\mathcal{R}_1$ and a cone which is coaxial with the one defining $\mathcal{R}_2'$ and whose generators make an angle $\Theta = \Theta' + \delta$ with the common axis $\hat{n}$. Region $\mathcal{R}_3^{(-)}$ is the complement of $\mathcal{R}_1 \cup \mathcal{R}_2^{(-)}$. The decomposition just described will be useful when the limit $t \to -\infty$ is being considered. On the other hand, when the limit $t \to +\infty$ is being considered, it will be convenient to decompose configuration space into the regions $\mathcal{R}_i^{(+)}$, $i = 1, 2, 3$, which are defined as the reflections in the origin of the corresponding regions $\mathcal{R}_i^{(-)}$.

Clearly, (25) is equivalent to (21) when (21) is restricted to the elements of $\mathcal{M}'(\hat{n}) \subset \mathcal{M}$. One can immediately remove this restriction from (21) by recalling that $\mathcal{M}$ consists of finite linear combinations of the functions in $\{\mathcal{M}'(\hat{n})\}$.

We now prove several lemmas on the evanescence of the prepared, the asymptotic, and the scattered states in a Coulomb field.

Lemma 1: For $f \in L^2$ and for $\mathcal{F}_1 \equiv \mathcal{F}_1^{(-)} \equiv \mathcal{F}_1^{(+)}$,

$$\lim_{t \to \pm\infty} \|\mathcal{F}_1 e^{-iH_0 t}\Phi_{0f}\| = 0, \quad (26a)$$

$$\lim_{t \to \pm\infty} \|\mathcal{F}_1 e^{-iHt}\Psi_{cf}^{(\pm)}\| = 0, \quad (26b)$$

and, for $f \in \mathcal{M}$,

$$\lim_{t \to \pm\infty} \|\mathcal{F}_1 U_c^{(\pm)}(t)\Phi_{0f}\| = 0. \quad (26c)$$

Proof: Since R is arbitrary but finite, Eqs. (26a) and (26b) state that a particle moving as a free particle ($H_0 = -\Delta/2m$) or as a particle scattered in a Coulomb field eventually disappears from any finite region of space. In addition, (26c) makes a similar statement about a particle whose propagator is given by $U_c^{(\pm)}(t)$ and whose state is in $\mathcal{M}$. The result (26a) has been proved elsewhere,[10] and is included in the lemma for the sake of completeness.

The proof of (26b) can most readily be carried out with the aid of two results from Ref. 15, where it is shown that, for $f \in L^2$,

$$\lim_{t \to \pm\infty} \left\| U_D(t)\Phi_{0f}(\mathbf{r}) - \left(\frac{m}{it}\right)^{\frac{3}{2}} \chi_c(\mathbf{r}, t)\hat{f}\left(\frac{m\mathbf{r}}{t}\right) \right\| = 0, \quad (27a)$$

$$\lim_{t \to \pm\infty} \|e^{-iHt}\Psi_{cf}^{(\pm)} - U_D(t)\Phi_{0f}\| = 0, \quad (27b)$$

where χ_c is a function of absolute value 1 and $U_D(t)$ is the unitary operator defined in (22). By combining these two results, we obtain the relation

$$\lim_{t \to \pm\infty} \left\| \mathcal{F}\left[e^{-iHt}\Psi_{cf}^{(\pm)}(\mathbf{r}) - \left(\frac{m}{it}\right)^{\frac{3}{2}} \chi_c(\mathbf{r}, t)\hat{f}\left(\frac{m\mathbf{r}}{t}\right) \right] \right\| = 0, \quad (28)$$

where $\mathcal{F}$ is an arbitrary projection into a measurable region of configuration space. With $\mathcal{F} = \mathcal{F}_1$ and $f \in \mathcal{M}$, the result (28) is equivalent to (26b), since one can always choose $|t|$ large enough so that the argument of $\hat{f}$ in (28) is outside the support of $\hat{f}$.

To complete the proof of this part of the lemma, we must show that (26b) holds for arbitrary $f \in L^2$. But this follows immediately from the fact that $\mathcal{M}$ is dense in L^2, and from the boundedness of the operators $\mathcal{F}_1$ and e^{-iHt}. For, given any $f \in L^2$ and $\epsilon > 0$, because $\mathcal{M}$ is dense in L^2, and because of (15), there is a $g \in \mathcal{M}$ such that $\|\Psi_{cf}^{(\pm)} - \Psi_{cg}^{(\pm)}\| < \epsilon$. Furthermore, because $g \in \mathcal{M}$, there is a T such that, for $|t| > T$,

$$\|A_t \Psi_g^{(\pm)}\| < \epsilon,$$

where $A_t = \mathfrak{I}_1 e^{-iHt}$ and $\|A_t\| \leq 1$. Therefore,

$$\|A_t\Psi_{cf}^{(\pm)}\| = \|A_t(\Psi_{cf}^{(\pm)} - \Psi_{cg}^{(\pm)}) + A_t\Psi_{cg}^{(\pm)}\| \leq \|\Psi_{cf}^{(\pm)} - \Psi_{cg}^{(\pm)}\| + \|A_t\Psi_{cg}^{(\pm)}\| \leq 2\epsilon.$$

QED

The proof of (26c) follows from an estimate of the integral in (19) with $f \in \mathcal{M}'(\hat{n})$. Integrating by parts[17] with respect to $k = |\mathbf{k}|$ and making use of the support properties of the functions in $\mathcal{M}'(\hat{n})$ lead one quite readily to the estimate[18] that, for $f \in \mathcal{M}'(\hat{n})$ and, therefore, also for $f \in \mathcal{M}$,

$$|U_c^{(\pm)}(t)\Phi_{0f}(\mathbf{r})| \leq (ar + b|\ln r| + c)/|t|,$$

where a, b, and c are constants independent of t. Integration of the square of the rhs over $\mathcal{R}_1$ in configuration space leads immediately to (26c).

We now prove an extension of what may be called the local evanescence theorems of Lemma 1, to distinguish them from the following long-range evanescence theorems.

Lemma 2: For $f \in \mathcal{M}'(\hat{n})$,

$$\lim_{t\to\pm\infty} \|\mathfrak{I}_2^{(\pm)} e^{-iHt}\Psi_{cf}\| = 0, \tag{29a}$$

$$\lim_{t\to\pm\infty} \|\mathfrak{I}_2^{(\pm)} U_c^{(\pm)}(t)\Phi_{0f}\| = 0. \tag{29b}$$

Proof: The proof of (29a) follows from (28) with $\mathfrak{I} = \mathfrak{I}_2^{(\pm)}$, since, when $\mathbf{r} \in \mathcal{R}_2^{(\pm)}$, $t \gtrsim 0$, and $f \in \mathcal{M}'(\hat{n})$, mr/t, the argument of $\hat{f}$, falls outside the support of $\hat{f}$.

The proof of (29b) is straightforward, but tedious, and is, therefore, only outlined here.[18] As before, choose $f \in \mathcal{M}'(\hat{n})$, insert its transform $\hat{f}$ in (19), and integrate twice by parts with respect to k. Using the support properties of $\hat{f}$ and its rapid decrease at infinity, it is not difficult to obtain the following estimate:

$$|\mathfrak{I}_2^{(\pm)} U_c^{(\pm)}(t)\Phi_{0f}(\mathbf{r})| \leq \frac{a\ln r + b(\ln r)^2}{(r \pm 4\rho t)^2} + \frac{c|t|\ln r}{(r \pm 4\rho t)^3}, \tag{30}$$

where a, b, and c are constants independent of t and r. Integration of the square of the rhs of (30) over $\mathcal{R}_2^{(\pm)}$ $[0 \ll R \leq r \leq \infty]$ yields an expression which is $O(|t|^{-\frac{1}{2}})$ as $|t| \to \infty$. QED

In the final lemma, we consider regions $\mathcal{R}_3^{(\pm)}$ in configuration space. With $\mathbf{r} \in \mathcal{R}_3^{(\pm)}$ and $f \in \mathcal{M}'(\hat{n})$ so that the support of the test function $\hat{f}(\mathbf{k}) \in \mathcal{R}_2'$, the argument $kr \pm \mathbf{k}\cdot\mathbf{r}$ of the Kummer function appearing in the Coulomb scattering state in (11) cannot vanish. In fact, the argument of the Kummer function can be made arbitrarily large by choosing the value R of $\mathcal{R}_1$ arbitrarily large. This permits use of the asymptotic expansion of the Kummer function. Since $\phi_c^{(\pm)}$ in (13) differs from the asymptotic form of $\psi_c^{(\pm)}$ in (14) by the conventional scattered wave, it is reasonable to expect that in $\mathcal{R}_3^{(\pm)}$ the scattering states would approach the asymptotic states as $t \to \pm\infty$. In fact, we can prove

Lemma 3: For $f \in \mathcal{M}'(\hat{n})$,

$$\lim_{t\to\pm\infty} \|\mathfrak{I}_3^{(\pm)}[e^{-iHt}\Psi_{cf}^{(\pm)} - U_c^{(\pm)}(t)\Phi_{0f}]\| = 0. \tag{31}$$

We first find a suitable estimate of

$$I^{(\pm)}(\mathbf{r}, t) = e^{-iHt}\Psi_{cf}^{(\pm)} - U_c^{(\pm)}(t)\Phi_{0f}(\mathbf{r}) \tag{32}$$

whose norm, taken over $\mathcal{R}_3^{(\pm)}$, is required in (31). For the proof, we exploit the properties of $\mathcal{M}'(\hat{n})$ and $\mathcal{R}_3^{(\pm)}$. The Kummer function appearing in the Coulomb wavefunction $\psi_c^{(\pm)}$ [see Eq. (11)] is a function of $kr(1 \pm \cos\vartheta)$ where ϑ is the angle made by $\mathbf{k}$ with $\mathbf{r}$, which may be taken as the z axis of a spherical coordinate system. The definitions of $\mathcal{M}'(\hat{n})$ and $\mathcal{R}_3^{(\pm)}$ imply that, for $\mathbf{r} \in \mathcal{R}_3^{(\pm)}$ and $\mathbf{k} \in \mathcal{R}_2'$, we have

$$0 < (1 - \cos\delta) \leq (1 \pm \cos\vartheta) \leq 2$$

(see Fig. 1). Therefore, for $k > \rho$ [in $\mathcal{M}'(\hat{n})$, $\hat{f} = 0$ for $k < \rho$] we can choose r large enough and make $kr(1 \pm \cos\vartheta)$ as large as desired. This value of r determines the radius of the spherical region $\mathcal{R}_1$, and also makes it possible to use the asymptotic representation of the Kummer function[19,20]:

$$\begin{aligned} {}_1F_1(\alpha|\,\gamma\,|\mp is) &= [\Gamma(\gamma)/\Gamma(\gamma-\alpha)]e^{\mp\alpha\pi i}[(\mp is)^{-\alpha} + r_0(\alpha|\,\gamma\,|\mp is)] \\ &\quad + [\Gamma(\gamma)/\Gamma(\alpha)]e^{\mp(\alpha-\gamma)\pi i} \\ &\quad \times e^{\mp is}[(\pm is)^{-(\gamma-\alpha)} + r_0(\gamma-\alpha|\,\gamma\,|\pm is)], \end{aligned} \tag{33}$$

where s is real and positive, and

$$|r_0(\alpha|\,\gamma\,|\mp is)| \leq \left|\frac{1+\alpha-\gamma}{\Gamma(\alpha)}\right| \frac{e^{\mp\frac{1}{2}\pi\,\mathrm{Im}(\alpha)}e^{\frac{1}{2}\pi|\mathrm{Im}(\gamma-\alpha)|}\Gamma(1+\mathrm{Re}(\alpha))}{S^{[1+\mathrm{Re}(\alpha)]}}.$$

Using this expression for the Kummer function appearing in (11) and carrying out several integrations by parts,[18] we obtain the following estimate for $I^{(\pm)}$ in (32). For $\mathbf{r} \in \mathcal{R}_3^{(\pm)}$,

$$|I^{(\pm)}(\mathbf{r}, t)| < \frac{a_1 + a_2\ln r}{r(r \pm 2\rho t)} + \frac{a_3}{r^2|t|}, \tag{34}$$

where the a's are constants independent of $\mathbf{r}$ and t. From (34), it follows that

$$\lim_{t\to\pm\infty} \int_{\mathbf{r}\in\mathcal{R}_3^{(\pm)}} |I^{(\pm)}(\mathbf{r}, t)|^2\, d\mathbf{r} = 0.$$

QED

We may now combine Lemmas 1, 2, and 3 and prove Theorem 1 for $f \in \mathcal{M}'(\hat{n})$, and thus, for $f \in \mathcal{M}$. For $f \in \mathcal{M}'(\hat{n})$ we have

$$\begin{aligned}
&\|e^{-iHt}\Psi_{cf}^{(\pm)} - U_c^{(\pm)}(t)\Phi_{0f}\| \\
&\quad = \left\| \sum_{i=1}^{3} \mathfrak{I}_i^{(\pm)}(e^{-iHt}\Psi_{cf}^{(\pm)} - U_c^{(\pm)}(t)\Phi_{0f}) \right\| \\
&\quad \leq \sum_{i=1}^{2} \|\mathfrak{I}_i^{(\pm)} e^{-iHt}\Psi_{cf}^{(\pm)}\| + \sum_{i=1}^{2} \|\mathfrak{I}_i^{(\pm)} U_c^{(\pm)}(t)\Phi_{0f}\| \\
&\qquad + \|\mathfrak{I}_3^{(\pm)} I^{(\pm)}(t)\| \to 0, \quad \text{when} \quad t \to \pm\infty, \quad (35)
\end{aligned}$$

according to the lemmas. This proves the existence of the wave operator $\hat{\Omega}_c^{(\pm)}$, as defined in (20), on the elements of $\mathcal{M}$, and supports the identification implied in (21) of $\Psi_{cf}^{(\pm)}$ as the scattering states associated with the prepared (or detected) state Φ_{0f}. Further, since $\hat{\Omega}_c^{(\pm)}$ and $\Omega_D^{(\pm)}$ both satisfy (17) for elements $f \in \mathcal{M}$, we have the relation (24) valid in $\mathcal{M}$.

To complete the proof of Theorem 1, we should show that $\hat{\Omega}_c^{(\pm)}$ is a bounded operator on $\mathcal{M}$ and, therefore, has a unique continuous extension onto $\mathcal{H}$. The quickest way to show this is to exploit the equality (24) of $\Omega_D^{(\pm)}$ and $\hat{\Omega}_c^{(\pm)}$ on $\mathcal{M}$. It is clear from (22) that $U_D(t)$ is unitary and that, since the domain of $\Omega_D^{(\pm)}$ is the entire Hilbert space, $\|\Omega_D^{(\pm)}\| = 1$. Therefore, $\hat{\Omega}_c^{(\pm)}$ is also bounded on $\mathcal{M}$ and has a unique extension $\Omega_c^{(\pm)}$ onto the whole of $\mathcal{H}$ with unit norm. This proves (24) and completes the proof of Theorem 1.

We can now exploit the relation (24) to prove an extension of Theorem 1.

Theorem 2: The operators $\Omega'^{(\pm)}$ exist as the unique continuous extension onto the whole of L^2 of $\hat{\Omega}'^{(\pm)}$, which is defined on $\mathcal{M}$ by the relation

$$\hat{\Omega}'^{(\pm)} = \lim_{t \to \pm\infty} e^{iHt} U_c^{(\pm)}(t), \quad (36)$$

where $H = -\Delta/2m + V$, V is Coulomb-like, and where $U_c^{(\pm)}(t)$ is defined in (19), just as in the pure Coulomb case. In addition, $\Omega'^{(\pm)}$ are isometries.

Proof: To carry out the proof, we first note that

$$\lim_{t \to \pm\infty}{}_S\, U_D^{\dagger}(t) U_c^{(\pm)}(t) = 1 \quad \text{on } \mathcal{M}. \quad (37)$$

This follows from (24), (20), and (23), since, for $f \in \mathcal{M}$ and $V = \alpha/r$,

$$\begin{aligned}
&\|[1 - U_D^{\dagger}(t) U_c^{(\pm)}(t)]f\| \\
&\quad = \|[U_D(t) - U_c^{(\pm)}(t)]f\| \\
&\quad = \|e^{iHt}U_D(t)f - \Omega_D^{(\pm)}f - e^{iHt}U_c^{(\pm)}(t)f + \Omega_c^{(\pm)}f\| \\
&\quad \leq \|e^{iHt}U_D(t)f - \Omega_D^{(\pm)}f\| \\
&\qquad + \|e^{iHt}U_c^{(\pm)}(t)f - \Omega_c^{(\pm)}f\| \to 0 \quad \text{when} \quad t \to \pm\infty.
\end{aligned}$$

The remainder of the proof rests on the theorem proved by Dollard[11] that $\Omega_D^{(\pm)}$ defined in (23) for pure Coulomb potentials also exists for Coulomb-like potentials and is an isometry on L^2. Applying this theorem and (37) to

$$\begin{aligned}
e^{iHt}U_c^{(\pm)}(t) &= e^{iHt}U_D(t)U_D^{\dagger}(t)U_c^{(\pm)}(t) \\
&\to \Omega_D^{(\pm)} \quad \text{on } \mathcal{M} \text{ when } \quad t \to \pm\infty,
\end{aligned}$$

we obtain the result that $\hat{\Omega}'^{(\pm)}$ exists on $\mathcal{M}$, and, on $\mathcal{M}$, we have

$$\hat{\Omega}'^{(\pm)} = \Omega_D^{(\pm)}. \quad (38)$$

Therefore, as before, $\hat{\Omega}'^{(\pm)}$ has a unique continuous extension onto all of L^2. QED

III. DISCUSSION

Theorem 2 guarantees the existence of a wave operator, constructed from the pure Coulomb "free propagator" $U_c^{(\pm)}(t)$, for a wide class of potentials. When the charge vanishes [$\lambda = 0$ in (13)], $U_c^{(\pm)}(t)$ in (19) becomes just the conventional free propagator $\exp(-iH_0t)$, from which the wave operator $\Omega^{(\pm)}$ appropriate to short-range potentials is constructed, as can be seen in (1). Therefore, since $\Omega'^{(\pm)} = \Omega^{(\pm)}$ for short-range potentials, we may use the simpler notation $\Omega^{(\pm)}$ for the wave operator in all cases considered in this paper, and we may define the wave operator by (20), with H containing any Coulomb-like potential.

Theorem 2 is an extension of that part of Theorem 1 which relates to the existence of a wave operator. However, unlike Theorem 1, Theorem 2 does *not* relate the scattering state $\Psi'^{(\pm)} = \Omega^{(\pm)}\Phi_0$ to the conventional time-independent scattering states $\psi^{(\pm)}$ in the manner shown in Eqs. (9) and (10). However, this relationship is guaranteed by Theorem 1 for the pure Coulomb case, and by Ikebe's work[13] for those cases involving a special class of short-range potentials. These results strongly suggest the validity of the conjecture that this relation is true in general. Moreover, the general applicability of $U_c^{(\pm)}(t)$ as the appropriate "free" propagator, at least for scattering involving Coulomb-like potentials, suggests that the (spatially asymptotic) boundary condition required to define the conventional steady-state solution is just that employed in the pure Coulomb problem.

Finally, in addition to the evanescence properties of the various states considered in the first two lemmas, there is another interesting result which pertains to the theory of measurements in quantum mechanics. The structure of the "free" propagator $U_c^{(\pm)}(t)$ exhibited in (19), together with the comments in the introduction concerning the relation between the prepared, scattering, and asymptotic states, suggests that it may be

useful to look upon $U_c^{(\pm)}(t)$ as a product of two operators:

$$U_c^{(\pm)}(t) = K^{(\pm)}(\lambda)U_t, \tag{39}$$

where $U_t = e^{-iH_0t}$ is the free propagator for short-range potentials; $K^{(\pm)}(\lambda)$ is defined implicitly by (39) and (19); and λ is proportional to the charge. [See (13).] Since $K^{(\pm)}(0) = 1$, we see that, for $\lambda = 0$, $U_c^{(\pm)}(t)$ becomes just the short-range propagator of the prepared as well as the asymptotic states. However, in the presence of Coulomb-like potentials, $\lambda \neq 0$. The operator $K^{(\pm)}(\lambda)$ is no longer trivial, and converts the prepared state

$$\Phi_{0f}(\mathbf{r}, t) = (2\pi)^{-\frac{3}{2}} \int d\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r} - ik^2t/2m} \hat{f}(\mathbf{k})$$

into the asymptotic states

$$\Phi_{cf}(\mathbf{r}, t) = (2\pi)^{-\frac{3}{2}} \int d\mathbf{k} \phi_c^{(\pm)}(\mathbf{k}, \mathbf{r}) e^{-ik^2t/2m} \hat{f}(\mathbf{k}).$$

Thus, $K^{(\pm)}(\lambda)$ describes the effect of the scatterer on the source and sink of the projectile.

Whether one can give the asymptotic states produced by $K^{(\pm)}(\lambda)$ objective significance depends to some extent on the validity of the conjecture which we made in connection with (16). But in any case, $K^{(\pm)}(\lambda)$ effectively alters the initial and final conditions from the free-field conditions to those appropriate for long-range interactions.

* Based on a thesis submitted by one of the authors (D. M.) to the Department of Physics of Fordham University in partial fulfillment of the requirements for the doctorate.

[1] J. M. Jauch, Helv. Phys. Acta **31**, 127 (1958).
[2] J. M. Jauch, Helv. Phys. Acta **31**, 661 (1958).
[3] C. R. Putnam, *Commutation Properties of Hilbert Space Operators and Related Topics* (Springer-Verlag, Berlin, 1967), see Chap. V.
[4] C. Møller, Kgl. Danske. Videnskab. Selskab. Math.-Fys. Medd. **23**, No. 1 (1945).
[5] J. M. Cook, J. Math. Phys. **36**, 82 (1957).
[6] We are here assuming that H_0 is the conventional free Hamiltonian $-\Delta/2m$. More generally, one must restrict oneself to that part of L^2 which is the absolutely continuous (or continuum) part of H_0 (see Ref. 7).
[7] S. T. Kuroda, Nuovo Cimento **12**, 431 (1959).
[8] M. N. Hack, Nuovo Cimento **9**, 731 (1958).
[9] J. M. Jauch and I. I. Zinnes, Nuovo Cimento **11**, 553 (1959).
[10] I. I. Zinnes, Nuovo Cimento Suppl. **12**, 87 (1959).
[11] J. D. Dollard, thesis, Princeton University, 1963.
[12] We are here emphasizing the limit $t \to -\infty$. A similar discussion can be carried out for the limit $t \to +\infty$. In addition, we use the notation Φ without explicit time dependence to represent a Schrödinger state at $t = 0$, i.e., $\Phi = \Phi(0)$.
[13] T. Ikebe, Archiv Ratl. Mech. Anal. **5**, 1 (1960).
[14] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Co., New York, 1955), p. 116.
[15] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[16] L. Schwartz, *Théorie des distributions* (Herman Cie., Paris, 1959), Vol. II.
[17] Here and in the sequel, whenever multiple integrals are replaced by iterated integrals so that integration by parts can be carried out, the replacement can be justified by an application of Fubini's theorem.
[18] Here, as well as in similar situations below, tedious details are omitted. The interested reader may refer to the thesis of one of the authors (D. M.) for these details.
[19] L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge University Press, Cambridge, 1960). See especially formula (4.1.6).
[20] N. N. Lebedev, *Special Functions and their Applications* (Prentice-Hall, Englewood Cliffs, N.J., 1965). On page 269, there is a useful form for the remainder which permits one to estimate the contribution of k which appears in more than the principal argument of the Kummer function [see Eq. (11)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Off-Shell T-Matrix Elements in Potential Scattering

J. UNDERHILL
Department of Applied Mathematics, The University of Liverpool, Liverpool, England

(Received 23 July 1968; Revised Manuscript Received 13 August 1969)

We give derivations of integral equations satisfied by the off-shell extension of the scattering amplitude (in nonrelativistic potential scattering) defined by the Lippmann–Schwinger equation. The input information in these equations consists of phase shifts. The derivation is based on the work of R. G. Newton on the inverse scattering problem in potential scattering.

We are concerned in this paper with the determination of 2-body T-matrix elements corresponding to nonrelativistic scattering by a local, central potential. In applications to the 3-body problem, for example, one needs to know the fully off-shell matrix element $t(\mathbf{k}', \mathbf{k}; z)$, defined as the solution of the Lippmann–Schwinger equation

$$t(\mathbf{k}', \mathbf{k}; z) = -\frac{1}{4\pi} v(\mathbf{k}' - \mathbf{k}) - \frac{1}{(2\pi)^3}\int v(\mathbf{k}' - \mathbf{k}'')(k''^2 - z)^{-1} t(\mathbf{k}'', \mathbf{k}; z)\, d\mathbf{k}'', \quad (1)$$

where $v(\mathbf{k}) = \int V(r) \exp(-i\mathbf{k}\cdot\mathbf{r})\, d\mathbf{r}$. If $V(r)$ is known, then $t(\mathbf{k}', \mathbf{k}; z)$ may be determined from (1) by integrating numerically [or by assuming a separable rather than a local form for the potential, in which case the solution of (1) becomes a straightforward algebraic problem]. We consider here the possibility of determining $t(\mathbf{k}', \mathbf{k}; z)$ from knowledge of scattering phase shifts.

It is known that $t(\mathbf{k}', \mathbf{k}; z)$ is essentially determined once

$$t(\mathbf{k}', \mathbf{k}) = \lim_{\epsilon \to 0+} t(\mathbf{k}', \mathbf{k}; k^2 + i\epsilon)$$

has been found.[1] We prove that $t(\mathbf{k}', \mathbf{k})$ may be derived from a set of integral equations in which the input information consists of the scattering amplitude $f(k, \theta)$ at energy k^2 rather than the potential. [$f(k, \theta)$ is the on-shell matrix element $t(\mathbf{k}', \mathbf{k})$, $k'^2 = k^2$, $\mathbf{k}\cdot\mathbf{k}' = k^2 \cos\theta$.] The derivation is based on Newton's solution of the inverse scattering problem at fixed energy,[2-4] the relevant part of which we summarize briefly in Sec. 1. The following sections are devoted to the derivation of the required equations.

1. THE INVERSE SCATTERING PROBLEM

Our notation is essentially that of Newton.[4] Let the $\varphi_l(k, r)$, $l = 0, 1, 2, \cdots$, satisfy the coupled linear equations

$$\varphi_l(k, r) = u_l(kr) - \sum_{l'=0}^{\infty} L_{ll'}(kr) c_{l'} \varphi_{l'}(k, r), \quad (2)$$

where

$$u_l(x) = (\tfrac{1}{2}\pi x)^{\frac{1}{2}} J_{l+\frac{1}{2}}(x), \quad L_{ll'}(x) = \int_0^x \frac{dy}{y^2}\, u_l(y) u_{l'}(y)$$

and the numbers c_l, $l = 0, 1, 2, \cdots$, are arbitrary real coefficients. It can be shown that the functions $\varphi_l(k, r)$ satisfy equations of the form

$$\frac{\partial^2}{\partial r^2}\varphi_l(k, r) + \left(k^2 - V_1(r) - \frac{l(l+1)}{r^2}\right)\varphi_l(k, r) = 0,$$

with $V_1(r)$ a potential which can be determined from the coefficients c_l. In fact, if

$$K(r, r') = \sum_{l=0}^{\infty} c_l \varphi_l(k, r) u_l(kr'), \quad (3)$$

then

$$V_1(r) = -\frac{2}{kr}\frac{d}{dr}\left(\frac{K(r, r)}{r}\right). \quad (4)$$

Also from (2), $\varphi_l(k, 0) = 0$, $l = 0, 1, 2, \cdots$.

Suppose now that the c_l are determined from the condition that the $\varphi_l(k, r)$ have the asymptotic behavior

$$\varphi_l(k, r) \sim A_l \sin(kr - \tfrac{1}{2}l\pi + \delta_l), \quad l = 0, 1, 2, \cdots; \quad (5)$$

then the corresponding potential $V_1(r) = V(r)$ has the property that the phase shifts arising from it have the given values δ_l.

In order to determine the c_l, one takes the limit $r \to \infty$ in (2), using

$$L_{ll'}(kr) \to L^{\infty}_{ll'} = i^{l'-l-1} M_{ll'}, \quad l' \neq l,$$
$$= \pi/2(2l+1), \quad l' = l,$$

where

$$M_{ll'} = 0, \quad l - l' \text{ even},$$
$$= [(l' - l)(l' + l + 1)]^{-1}, \quad l - l' \text{ odd}.$$

Then (2) becomes

$$A_l \sin(kr - \tfrac{1}{2}l\pi + \delta_l) = \sin(kr - \tfrac{1}{2}l\pi) - \sum_{l'=0}^{\infty} L^{\infty}_{ll'} c_{l'} A_{l'} \sin(kr - \tfrac{1}{2}l'\pi + \delta_{l'}).$$

From this identity one obtains

$$\frac{1}{c_l} = -\frac{\pi}{2(2l+1)} + \frac{\sum_{l'} M_{ll'}a_{l'}}{a_l \tan\delta_l},$$

where a_l is obtained from

$$\sum_{l'} M_{ll'}a_{l'} + \sum_{l'} \tan\delta_l M_{ll'} \tan\delta_{l'} a_{l'} = \tan\delta_l. \quad (6)$$

Because the matrix $\mathbf{M} = (M_{ll'})$ does not have a unique inverse, Eq. (6) does not have a unique solution a_l, $l = 0, 1, 2, \cdots$. In fact it possesses a 1-parameter set of solutions, associated with the existence of a vector $\mathbf{v}$, which is annihilated by $\mathbf{M}$. It follows that a 1-parameter family of potentials will all give rise to the same phase shifts δ_l at energy k^2. Sabatier[3] has proved that if $\delta_l = O(l^{-3-\epsilon})$ as $l \to \infty$, for some $\epsilon > 0$, then exactly one of the phase equivalent potentials determined by this method decreases as $r \to \infty$ faster than $r^{-\frac{3}{2}}$. This short-range potential decreases as $r \to \infty$ faster than $r^{-2+\epsilon}$ for every $\epsilon > 0$, and is presumably the only one of practical interest, since every other potential possesses an oscillatory tail of the form $r^{-\frac{3}{2}} \cos(2kr - \pi/4)$.

A class of phase-equivalent potentials larger than that found by Newton's procedure has been found by Sabatier.[5] However, we shall consider here only the amplitude $t(\mathbf{k}', \mathbf{k})$ arising from the unique "short-range" potential described above. In fact, at two points in the following derivation, we find it necessary to assume a stronger condition on the potential $V(r)$ than the $r^{-2+\epsilon}$ behavior, namely, we assume that

$$\int_0^\infty r\,|V(r)|\,dr < \infty.$$

Since our aim is to derive equations which make no mention of the potential, it is unfortunate that we have to make this assumption, and it is likely that a more delicate analysis would show it to be unnecessary. On the other hand, the condition is not very stringent and is satisfied provided that the potential is of constant sign for all sufficiently large r. To show this we observe that in this case $\int_0^\infty r\,|V(r)|\,dr$ converges provided $\int_0^\infty rV(r)\,dr$ converges. [That $rV(r)$ is finite at $r = 0$ has been proved by Sabatier.[6]] But

$$\int_0^\infty rV(r)\,dr = -\frac{2}{k}\int_0^\infty \frac{d}{dr}\left(\frac{K(r,r)}{r}\right)dr = -\frac{2}{k}\lim_{r\to\infty}\left(\frac{K(r,r)}{r}\right).$$

Sabatier shows[3] that $K(r,r)/r \to \alpha$, a constant, as $r \to \infty$, so the result follows.

Finally, we shall need the following estimate of the asymptotic behavior of the product $c_l A_l$ for the short-range potential, derived in Ref. 3 under the assumption $\delta_l = O(l^{-3-\epsilon})$:

$$c_l A_l = 4\alpha/\pi + O(l^{-\frac{5}{3}-\epsilon}). \quad (7)$$

2. PROPERTIES OF $t(\mathbf{k}', \mathbf{k})$

Let $\psi_l(k, r)$ be the regular solution of

$$\frac{\partial^2}{\partial r^2}\psi_l(k,r) + \left(k^2 - V(r) - \frac{l(l+1)}{r^2}\right)\psi_l(k,r) = 0, \quad (8)$$

with the asymptotic behavior

$$\psi_l(k,r) \sim e^{i\delta_l}\sin(kr - \tfrac{1}{2}l\pi + \delta_l). \quad (9)$$

In (8), $V(r)$ is the short-range potential corresponding to the phase shifts δ_l at energy k^2, where $\delta_l = O(l^{-3-\epsilon})$ for some $\epsilon > 0$.

Then

$$t(\mathbf{k}', \mathbf{k}) = \sum_{l=0}^{\infty}(2l+1)t_l(k', k)P_l(\mathbf{k}\cdot\mathbf{k}'/kk'), \quad (10)$$

where

$$t_l(k', k) = -\frac{1}{kk'}\int_0^\infty u_l(k'r)V(r)\psi_l(k,r)\,dr. \quad (11)$$

If we substitute for $V(r)\psi_l(k,r)$ from the differential equation (8), integrate by parts, and then use

$$\varphi_l(k,r) = A_l e^{-i\delta_l}\psi_l(k,r) \quad (12)$$

[as is clear from (5) and (9)], we can express $t_l(k', k)$ in terms of $\varphi_l(k,r)$. Elimination of the wavefunctions $\varphi_l(k,r)$ between this relation and Eq. (2) leads to a set of coupled equations for the $t_l(k', k)$. This method, while straightforward in principle, seems to be rather more difficult to justify than the indirect approach which we shall adopt.

Define

$$\Theta_{ll'}(k, k', k'') = 2\int_0^\infty dr\,\frac{u_l(kr)u_l(k'r)}{r}\frac{d}{dr}\left(\frac{u_{l'}(kr)u_{l'}(k''r)}{r}\right). \quad (13)$$

We prove in Appendix A that

$$\lim_{\epsilon\to 0+}\int_0^\infty dk''\,\frac{u_{l'}(k''r)\Theta_{ll'}(k,k',k'')}{k''^2 - k^2 - i\epsilon} = \pi\int_0^\infty ds\,\frac{G_{l'}(r,s)u_{l'}(ks)}{s}\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right) \quad (14)$$

and

$$\lim_{\epsilon\to 0+}\int_0^\infty dk''\,\frac{u_{l'}(k''r)\Theta_{ll'}(k,k',k'')}{k''^2 - k'^2 - i\epsilon} = \tfrac{1}{2}\pi k(L_{ll'}(kr)u_l(k'r) - i^{l-l'}L_{ll'}^{\infty}u_{l'}(k'r)). \quad (15)$$

In (14), $G_l(r, s)$ is the Green's function for Eq. (8), with boundary condition (9):

$$\psi_l(k, r) = u_l(kr) + \int_0^\infty G_l(r, s)V(s)\psi_l(k, s)\, ds. \quad (16)$$

We wish to evaluate

$$P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - K^2}\, dk''$$

in the two cases $K = k$ and $K = k'$.

We use the usual prescription

$$P((k''^2 - K^2)^{-1}) = (k''^2 - K^2 - i\epsilon)^{-1} - i\pi\delta(k''^2 - K^2)$$

(justified for our case in Appendix B) to obtain

$$\lim_{\epsilon\to 0+}\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - K^2 - i\epsilon}\, dk'' - \frac{i\pi}{2} t_{l'}(K, k)\Theta_{ll'}(k, k', K).$$

The integral is

$$-\frac{1}{k}\lim_{\epsilon\to 0+}\int_0^\infty dk''\, \frac{\Theta_{ll'}(k, k', k'')}{k''^2 - K^2 - i\epsilon} \times \int_0^\infty dr u_{l'}(k''r)V(r)\psi_{l'}(k, r).$$

The integrand is bounded for all k'', r and less in modulus than $C\,|k''^2 - K^2 - i\epsilon|^{-1} r^{-\frac{3}{2}}$, so the integral is absolutely convergent. The order of integration may be changed to give

$$-\frac{1}{k}\lim_{\epsilon\to 0+}\int_0^\infty drV(r)\psi_{l'}(k, r) \times \int_0^\infty dk''\, \frac{\Theta_{ll'}(k, k', k'')u_{l'}(k''r)}{k''^2 - K^2 - i\epsilon}.$$

The integral over k'' can be found explicitly (Appendix A), and this makes it easy to justify taking the $\lim_{\epsilon\to+0}$ inside the integration over r (Appendix B). In the two cases $K = k$ and $K = k'$, respectively, we get

$$P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - K^2}\, dk''$$

$$= -\tfrac{1}{2}i\pi t_{l'}(k, k)\Theta_{ll'}(k, k', k) - \frac{\pi}{k}\int_0^\infty drV(r)\psi_{l'}(k, r)\int_0^\infty ds\, \frac{G_{l'}(r, s)u_{l'}(ks)}{s} \times \frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right), \quad K = k, \quad (17)$$

$$= -\tfrac{1}{2}i\pi t_{l'}(k', k)\Theta_{ll'}(k, k', k') - \frac{\pi}{2}\int_0^\infty drV(r)\psi_{l'}(k, r) \times [L_{ll'}(kr)u_l(k'r) - i^{l-l'}L^\infty_{ll'}u_{l'}(k'r)], \quad K = k'. \quad (18)$$

Again, the integral on the right of (17) is absolutely convergent, since the integrand is bounded and less in modulus than $Cs^{-2}r^{-\frac{3}{2}}$, so the order of integration may be changed. Using (16) and the well-known result

$$t_{l'}(k, k) = k^{-1}e^{i\delta_{l'}}\sin\delta_{l'},$$

we find

$$P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - k^2}\, dk''$$

$$= -\frac{i\pi}{2k}e^{i\delta_{l'}}\sin\delta_{l'}\Theta_{ll'}(k, k', k) - \frac{\pi}{k}\int_0^\infty ds\, \frac{u_{l'}(ks)}{s}(\psi_{l'}(k, s) - u_{l'}(ks)) \times \frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right)$$

$$= -\frac{\pi}{2k}e^{i\delta_{l'}}\cos\delta_{l'}\Theta_{ll'}(k, k', k) - \frac{\pi}{k}\int_0^\infty ds\, \frac{u_{l'}(ks)\psi_{l'}(k, s)}{s}\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right). \quad (19)$$

Similarly,

$$P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - k'^2}\, dk'' = -\frac{\pi^2}{4(2l+1)}kk't_l(k', k)\delta_{ll'} - \frac{\pi}{2}\int_0^\infty drV(r)u_l(k'r)L_{ll'}(kr)\psi_{l'}(k, r), \quad (20)$$

where we have used (11) and the result (proved in Appendix A)

$$\Theta_{ll'}(k, k', k') = kk'M_{ll'}.$$

If we multiply both sides of (19) and (20) by $c_{l'}A_{l'}e^{-i\delta_{l'}}$, we replace the $\psi_{l'}(k, s)$ and $\psi_{l'}(k, r)$ on the right-hand side of these equations by $c_{l'}\varphi_{l'}(k, s)$ and $c_{l'}\varphi_{l'}(k, r)$. Now sum over l', taking the summation inside the integrals over s and r, respectively, on the right-hand side. [N.B.: It is only in the justification of this step, carried out in Appendix B, that we use the assumption

$$\int_0^\infty r\,|V(r)|\, dr < \infty.$$

If these interchanges of summation and integration can be justified, then our final equations hold—not only without this assumption, but even for the long-range oscillatory potentials phase-equivalent to $V(r)$.]

Using (3) and (4), we obtain

$$\sum_{l'=0}^{\infty} c_{l'}A_{l'}e^{-i\delta_{l'}}P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - k^2}\,dk''$$
$$= -\frac{\pi}{2k}\sum_{l'=0}^{\infty} c_{l'}A_{l'}\cos\delta_{l'}\Theta_{ll'}(k, k', k)$$
$$-\frac{\pi}{k}\int_0^\infty \frac{ds}{s}K(s, s)\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right)$$
$$= -\frac{\pi}{2k}\sum_{l'=0}^{\infty} c_{l'}A_{l'}\cos\delta_{l'}\Theta_{ll'}(k, k', k)$$
$$-\frac{\pi}{2}\int_0^\infty ds V(s)u_l(ks)u_l(k's). \qquad (21)$$

[In the integration by parts which leads to the final integral on the right-hand side, the contributions from the end points vanish since $K(s, s)/s$ is bounded.]

Similarly, using (2), (11), and (12), we obtain

$$\sum_{l'=0}^{\infty} c_{l'}A_{l'}e^{-i\delta_{l'}}P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - k'^2}\,dk''$$
$$= -\frac{\pi^2}{4(2l+1)}kk't_l(k', k)c_lA_le^{-i\delta_l}$$
$$-\frac{\pi}{2}\int_0^\infty V(r)u_l(k'r)(u_l(kr) - \varphi_l(k, r))\,dr$$
$$= -\frac{\pi}{2}A_le^{-i\delta_l}\left(1 + \frac{\pi c_l}{2(2l+1)}\right)kk't_l(k', k)$$
$$-\frac{\pi}{2}\int_0^\infty V(r)u_l(kr)u_l(k'r)\,dr. \qquad (22)$$

Subtracting (22) from (21) gives the required coupled integral equations for the amplitudes $t_l(k', k)$:

$$t_l(k', k)\left(1 + \frac{\pi c_l}{2(2l+1)}\right)$$
$$= \frac{e^{i\delta_l}}{A_lk'^2k^2}\sum_{l'=0}^{\infty} c_{l'}A_{l'}\cos\delta_{l'}\Theta_{ll'}(k, k', k)$$
$$+\frac{2e^{i\delta_l}}{\pi A_l}\frac{(k^2 - k'^2)}{kk'}\sum_{l'=0}^{\infty} c_{l'}A_{l'}e^{-i\delta_{l'}}$$
$$\times P\int_0^\infty dk''\frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{(k''^2 - k^2)(k''^2 - k'^2)}. \qquad (23)$$

Finally, we list some further problems suggested by the set of equations (23), with which a certain amount of progress has been made: First, do they determine a unique set of amplitudes $t_l(k', k)$ once the coefficients c_l, A_l corresponding to the short-range potential with given phase shifts have been evaluated? Secondly, are they of any value in numerical calculations of the $t_l(k', k)$?

The answers might be clearer if we could write the equations as a single integral equation for the 3-dimensional matrix element $t(\mathbf{k}', \mathbf{k})$. This does not seem to be possible in any simple way, but it still appears that there may be a 3-dimensional form of the equations. These questions may be the subjects of a further paper.

APPENDIX A: PROPERTIES OF $\Theta_{ll'}(k, k', k'')$

We outline the proof of the following properties:

$$\Theta_{ll'}(k, k', k'') = \frac{k^2k'k''}{2}\int_{-1}^{1} ds$$
$$\times P_l(s)\int_{-1}^{1} dt\frac{P_{l'}(t)}{k'^2 - k''^2 - 2kk's + 2kk''t} \qquad (A1)$$

(the integral over t is a principal-value integral if the denominator vanishes for $-1 < t < 1$);

$$\Theta_{ll'}(k, k', k') = kk'M_{ll'}, \qquad (A2)$$

$$\lim_{\epsilon\to 0+}\int_0^\infty dk''\frac{u_{l'}(k''r)\Theta_{ll'}(k, k', k'')}{k''^2 - K^2 - i\epsilon}$$
$$= -\frac{i\pi}{K}u_{l'}(Kr)\int_r^\infty ds\frac{u_{l'}(ks)w_{l'}(Ks)}{s}\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right)$$
$$-\frac{i\pi}{K}w_{l'}(Kr)\int_0^r ds\frac{u_{l'}(ks)u_{l'}(Ks)}{s}\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right), \qquad (A3)$$

where

$$w_l(x) = (\tfrac{1}{2}\pi x)^{\frac{1}{2}}H^{(1)}_{l+\frac{1}{2}}(x).$$

We assume the following results:

$$\lim_{R\to\infty}\int_a^b f(x)\frac{[1 - \cos R(x - y)]}{x - y}\,dx = \int_a^b \frac{f(x)}{x - y}\,dx \qquad (A4)$$

[where the second integral is a principal value if $a < y < b$; the result certainly holds if $f(x)$ is a polynomial, the only case we shall use] and

$$\frac{u_l(kr)u_l(k'r)}{r} = \tfrac{1}{2}\int_{|k-k'|}^{k+k'} \sin prP_l\left(\frac{k^2 + k'^2 - p^2}{2kk'}\right)dp. \qquad (A5)$$

For the proof of (A1), let

$$A(R) = 2\int_0^R dr\frac{u_l(kr)u_l(k'r)}{r}\frac{d}{dr}\left(\frac{u_{l'}(kr)u_{l'}(k''r)}{r}\right).$$

From (A5)

$$A(R) = \tfrac{1}{2}\int_0^R dr\int_{|k-k'|}^{k+k'} dp\sin prP_l\left(\frac{k^2 + k'^2 - p^2}{2kk'}\right)$$
$$\times\int_{|k-k''|}^{k+k''} p'\cos p'rP_{l'}\left(\frac{k^2 + k''^2 - p'^2}{2kk''}\right)dp'$$
$$= \tfrac{1}{4}\int_{|k-k'|}^{k+k'} dp\int_{|k-k''|}^{k+k''} dp'p'P_l\left(\frac{k^2 + k'^2 - p^2}{2kk'}\right)$$
$$\times P_{l'}\left(\frac{k^2 + k''^2 - p'^2}{2kk''}\right)$$
$$\times\left(\frac{1 - \cos R(p + p')}{p + p'} + \frac{1 - \cos R(p - p')}{p - p'}\right).$$

We wish to find $\lim A(R)$, as $R \to \infty$. We can take the limit, as $R \to \infty$, inside the integral over p provided the integrand possesses an integrable bound, independent of R. Let

$$\int_{|k-k''|}^{k+k''} p'\,dp' P_{l'}\left(\frac{k^2+k''^2-p'^2}{2kk''}\right)\left(\frac{1-\cos R(p+p')}{p+p'}+\frac{1-\cos R(p-p')}{p-p'}\right) = J+K+L,$$

where

$$J = \int_{|k-k''|}^{k+k''} p'\,dp' P_{l'}\left(\frac{k^2+k''^2-p'^2}{2kk''}\right)\left(\frac{1-\cos R(p+p')}{p+p'}\right),$$

$$K = \int_{|k-k''|}^{k+k''} dp' \left(\frac{p'P_{l'}((k^2+k''^2-p'^2)/2kk'') - pP_{l'}((k^2+k''^2-p^2)/2kk'')}{p-p'}\right)[1-\cos R(p-p')],$$

$$L = pP_{l'}\left(\frac{k^2+k''^2-p^2}{2kk''}\right)\int_{|k-k''|}^{k+k''} dp'\,\frac{1-\cos R(p-p')}{p-p'}.$$

It is easily shown that

$$|J| \le 2(k+k''-|k-k''|) - 2p\log\left(\frac{p+k+k''}{p+|k-k''|}\right),$$

$$|K| \le 2(k+k''-|k-k''|)\max_{\substack{|k-k''|<p'<k+k'' \\ |k-k'|<p<k+k'}}\left|\frac{p'P_{l'}((k^2+k''^2-p'^2)/2kk'') - pP_{l'}((k^2+k''^2-p^2)/2kk'')}{p-p'}\right|,$$

$$|L| \le 2p\log\left|\frac{p-|k-k''|}{p-k-k''}\right|\max_{|k-k'|<p<k+k'}\left|P_l\left(\frac{k^2+k''^2-p^2}{2kk''}\right)\right|.$$

Hence, using (A4),

$$\Theta_{ll'}(k,k',k'') = \lim_{R\to\infty} A(R)$$
$$= \tfrac{1}{2}\int_{|k-k'|}^{k+k'} p\,dp P_l\left(\frac{k^2+k'^2-p^2}{2kk'}\right)$$
$$\times \int_{|k-k''|}^{k+k''} \frac{p'\,dp'}{p^2-p'^2} P_{l'}\left(\frac{k^2+k''^2-p'^2}{2kk''}\right).$$

Changing the variables to s and t given by

$$p^2 = k^2+k'^2-2kk's,\quad p'^2 = k^2+k''^2-2kk''t$$

gives the result stated in (A1).

In order to prove (A2), we have from (A1):

$$\Theta_{ll'}(k,k',k') = \frac{kk'}{4}\int_{-1}^{1} ds P_l(s) P\int_{-1}^{1} dt\,\frac{P_{l'}(t)}{t-s},$$
$$= -\frac{kk'}{2}\int_{-1}^{1} ds P_l(s)Q_{l'}(s)$$
$$= -\frac{kk'}{2}\,\frac{[1-(-1)^{l-l'}]}{(l-l')(l+l'+1)} = kk'M_{ll'}.$$

In order to prove (A3), for z not real and positive we have

$$\int_0^\infty dk''\,\frac{u_{l'}(k''r)}{k''^2-z}\,\Theta_{ll'}(k,k',k'')$$
$$= -2\int_0^\infty dk''\,\frac{u_{l'}(k''r)}{k''^2-z}$$
$$\times \int_0^\infty ds\,\frac{u_{l'}(k''s)u_{l'}(ks)}{s}\,\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right).$$

The integrand is bounded for all k'' and s, and less in modulus than $Ck''^{-2}s^{-2}$. The integral therefore converges absolutely and the order of integration may be changed:

$$\int_0^\infty dk''\,\frac{u_{l'}(k''r)}{k''^2-z}\,\Theta_{ll'}(k,k',k'')$$
$$= -2\int_0^\infty ds\,\frac{u_{l'}(ks)}{s}\,\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right)$$
$$\times \int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2-z}. \quad \text{(A6)}$$

Now

$$\int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2-z}$$
$$= \frac{\pi}{2}(rs)^{\frac{1}{2}} I_{l'+\frac{1}{2}}(r\sqrt{-z})K_{l'+\frac{1}{2}}(s\sqrt{-z}),\quad r \le s,$$
$$= \frac{\pi}{2}(rs)^{\frac{1}{2}} I_{l'+\frac{1}{2}}(s\sqrt{-z})K_{l'+\frac{1}{2}}(r\sqrt{-z}),\quad r \ge s. \quad \text{(A7)}$$

The square root is defined so that if $z = \rho e^{i\theta}$, $0 < \theta \le \pi$, then $\sqrt{-z} = -i(\sqrt{\rho})e^{i\theta/2}$, so that

$$I_{l'+\frac{1}{2}}(r\sqrt{-z}) = e^{-i\pi(l'/2+\frac{1}{4})}J_{l'+\frac{1}{2}}[r(\sqrt{\rho})e^{i\theta/2}],$$
$$K_{l'+\frac{1}{2}}(r\sqrt{-z}) = \tfrac{1}{2}i\pi e^{i\pi(l'/2+\frac{1}{4})}H^{(1)}_{l'+\frac{1}{2}}(r(\sqrt{\rho})e^{i\theta/2}).$$

If $\rho \to K^2$, $\theta \to 0+$,

$$I_{l'+\frac{1}{2}}(r-z)K_{l'+\frac{1}{2}}(s-z) \to \tfrac{1}{2}i\pi J_{l'+\frac{1}{2}}(Kr)H^{(1)}_{l'+\frac{1}{2}}(Ks)$$

and

$$\lim_{\epsilon\to 0+}\int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2-K^2-i\epsilon}=\frac{i\pi}{2K}u_{l'}(Kr)w_{l'}(Ks),\quad r\le s,$$
$$=\frac{i\pi}{2K}u_{l'}(Ks)w_{l'}(Kr),\quad r\ge s.$$

(A3) now follows provided we are justified in taking the $\lim_{z\to K^2+i\epsilon}$ inside the integral over s on the right-hand side of (A6). This will be the case if there exists an integrable function $\Psi(s)$, independent of ϵ, such that,for sufficiently small ϵ,

$$\left|\frac{u_{l'}(ks)}{s}\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right)\int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2-K^2-i\epsilon}\right|<\Psi(s).$$

For $\Psi(s)$ one may take the left-hand side with the integral over k'' replaced by a suitable constant, since the integral is bounded, for sufficiently small ϵ, as a function of s, ϵ. To show this one uses the explicit expressions for the Bessel functions $J_{l'+\frac{1}{2}}[r(K^2+i\epsilon)^{\frac{1}{2}}]$ and $H^{(1)}_{l'+\frac{1}{2}}[r(K^2+i\epsilon)^{\frac{1}{2}}]$ and finds that,if $\epsilon<K^2$,

$$\left|\int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2-K^2-i\epsilon}\right|\le 2re^{2Kr}[K_{l'+\frac{1}{2}}(Kr)]^2,\quad r\le s,$$
$$\le\frac{\pi r}{2}e^{Kr}I_{l'+\frac{1}{2}}(Kr2^{\frac{1}{4}})K_{l'+\frac{1}{2}}(Kr),\quad r>s.$$

Observe that if $K=k$, since the kernel $G_l(r,s)$ defined in (16) is

$$G_l(r,s)=-(i/k)w_l(kr)u_l(ks),\quad r\ge s,$$
$$=-(i/k)w_l(ks)u_l(kr),\quad r\le s,$$

the result (A3) may be written as

$$\lim_{\epsilon\to 0+}\int_0^\infty dk''\,\frac{u_{l'}(k''r)\Theta_{ll'}(k,k',k'')}{k''^2-k^2-i\epsilon}=\pi\int_0^\infty ds\,\frac{G_{l'}(r,s)u_{l'}(ks)}{s}\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right),$$

which is Eq. (14).

Even when $K\neq k$, the result of (A3) implies that

$$y=\lim_{\epsilon\to 0+}\int_0^\infty dk''\,\frac{u_{l'}(k''r)\Theta_{ll'}(k,k',k'')}{k''^2-K^2-i\epsilon}$$

satisfies

$$\frac{d^2y}{dr^2}+\left(K^2-\frac{l'(l'+1)}{r^2}\right)y=\frac{\pi}{r}u_{l'}(kr)\frac{d}{dr}\left(\frac{u_l(kr)u_l(k'r)}{r}\right),\quad\text{(A8)}$$

with the boundary conditions

$$y\underset{r\to\infty}{\sim}\frac{\pi}{2K}\Theta_{ll'}(k,k',K)e^{i(Kr-\frac{1}{2}l'\pi)},\quad y(0)=0.\quad\text{(A9)}$$

When $K=k'$, one verifies explicitly that

$$\frac{k\pi}{2}L_{ll'}(kr)u_l(k'r)$$

satisfies (A8), so that (A9) implies

$$y=\frac{k\pi}{2}L_{ll'}(kr)u_l(k'r)+Au_{l'}(k'r),$$

where

$$A\sin(k'r-\tfrac{1}{2}l'\pi)+\frac{k\pi}{2}L^\infty_{ll'}\sin\left(k'r-\frac{l\pi}{2}\right)=\frac{\pi}{2k'}\Theta_{ll'}(k,k',k')e^{i(k'r-\frac{1}{2}l'\pi)}.$$

Using the result (A2), we easily find

$$A=-\frac{\pi k}{2}e^{\frac{1}{2}i(l-l')\pi}L^\infty_{ll'}.$$

This completes the proof of Eq. (15).

APPENDIX B

We justify here some of the operations carried out in the derivation of Eqs. (23), namely: (i) the use of the prescription

$$P\left(\frac{1}{k^2-K^2}\right)=\frac{1}{k^2-K^2-i\epsilon}-i\pi\delta(k^2-K^2);$$

(ii) the interchange of $\lim_{\epsilon\to 0+}$ and the integration over r in the equation preceding (17); (iii) the interchange of summation and integration which took us from (19) to (21); (iv) the similar operation which took us from (20) to (22).

Proof of (i)*:* It can be shown that,for each $y>0$,

$$\lim_{\epsilon\to 0+}\int_0^\infty\frac{f(x)}{x-y-i\epsilon}\,dx=P\int_0^\infty\frac{f(x)}{x-y}\,dx+i\pi f(y),$$

provided that (a) $\int_c^\infty\frac{f(x)}{x}\,dx$ converges for some

$c > 0$, and (b) $f(x)$ satisfies a Lipschitz condition of positive order for each $x \geq 0$. We consider

$$P\int_0^\infty \frac{t_{l'}(k'', k)k''\Theta_{ll'}(k, k', k'')}{k''^2 - K^2}\,dk'' = \tfrac{1}{2}P\int_0^\infty \frac{t_{l'}(\sqrt{p}, k)\Theta_{ll'}(k, k', \sqrt{p})}{p - K^2}\,dp.$$

Condition (a) is satisfied, since

$$|t_{l'}(\sqrt{p}, k)\Theta_{ll'}(k, k', \sqrt{p})| < C/\sqrt{p}.$$

Next, we observe that $t_{l'}(\sqrt{p}, k)\Theta_{ll'}(k, k', \sqrt{p})$ depends on p through a term of the form $u_{l'}(r\sqrt{p})u_{l'}(s\sqrt{p})/\sqrt{p}$, which, using (A5), is easily shown to satisfy

$$\left|\frac{u_{l'}(r\sqrt{p})u_{l'}(s\sqrt{p})}{\sqrt{p}} - \frac{u_{l'}(r\sqrt{q})u_{l'}(s\sqrt{q})}{\sqrt{q}}\right| < \tfrac{4}{3}|p - q|^{\frac{1}{4}}((r + s)^{\frac{3}{2}} - |r - s|^{\frac{3}{2}}), < C|p - q|^{\frac{1}{4}} r\sqrt{s}, \quad r \leq s,$$

with a similar inequality if $r \geq s$. Hence, from the definitions (11) and (13) of $t_{l'}(k', k)$ and $\Theta_{ll'}(k, k', k'')$, we can show that

$$|t_{l'}(\sqrt{p}, k)\Theta_{ll'}(k, k', \sqrt{p}) - t_{l'}(\sqrt{q}, k)\Theta_{ll'}(k, k', \sqrt{q})| < A|p - q|^{\frac{1}{4}},$$

where

$$A < \int_0^\infty ds\,|V(s)|\,|\psi_{l'}(k, s)|\,(A_1 + A_2 \log s)\sqrt{s}.$$

Here A_1 and A_2 depend only on k and k', and the integral converges if $V(s)$ is the short-range potential. Hence (b) is satisfied.

Proof of (ii): Since

$$\int_0^\infty dr\,|V(r)\psi_{l'}(k, r)|$$

converges, we have only to show that

$$\int_0^\infty dk''\,\frac{\Theta_{ll'}(k, k', k'')u_{l'}(k''r)}{k''^2 - K^2 - i\epsilon}$$

possesses a bound independent of r and ϵ. This can be done using estimates of the kind obtained in the proof of (A3). There we showed that

$$\int_0^\infty dk''\,\frac{\Theta_{ll'}(k, k', k'')u_{l'}(k''r)}{k''^2 - K^2 - i\epsilon} = -2\int_0^\infty ds\,\frac{u_{l'}(ks)}{s}\,\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right) \times \int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2 - K^2 - i\epsilon}.$$

We estimate the last integral on the right-hand side using the explicit expressions for the Bessel functions $J_{l'+\frac{1}{2}}(Kr)$ and $H^{(1)}_{l'+\frac{1}{2}}(Kr)$, and we obtain, introducing a fixed length $a > 0$,

$$\left|\int_0^\infty dk''\,\frac{u_{l'}(k''r)u_{l'}(k''s)}{k''^2 - K^2 - i\epsilon}\right| \leq 2ae^{2Ka}[K_{l'+\frac{1}{2}}(Ka)]^2, \quad r > a, s > a,$$

$$\leq \frac{\pi a}{2}\,e^{Ka}K_{l'+\frac{1}{2}}(Ka)I_{l'+\frac{1}{2}}(Ka2^{\frac{1}{4}}), \quad r > a, s < a \text{ or } r < a, s > a,$$

$$\leq \max_{0\leq r\leq a}\left[\frac{\pi r}{2}\,e^{Kr}K_{l'+\frac{1}{2}}(Kr)I_{l'+\frac{1}{2}}(Kr2^{\frac{1}{4}})\right], \quad r < a, s < a,$$

provided $\epsilon < K^2$. If the largest of the expressions on the right is M,

$$\left|\int_0^\infty dk''\,\frac{\Theta_{ll'}(k, k', k'')u_{l'}(k''r)}{k''^2 - K^2 - i\epsilon}\right| \leq M\int_0^\infty ds\left|\frac{u_{l'}(ks)}{s}\,\frac{d}{ds}\left(\frac{u_l(ks)u_l(k's)}{s}\right)\right|.$$

Proof of (iii): We have to show that

$$\sum_{l'=0}^\infty d_{l'}\int_0^\infty dr\,\frac{u_{l'}(kr)\psi_{l'}(k, r)}{r}\,\frac{d}{dr}\left(\frac{u_l(kr)u_l(k'r)}{r}\right) = \int_0^\infty \frac{dr}{r}\,\frac{d}{dr}\left(\frac{u_l(kr)u_l(k'r)}{r}\right)\sum_{l'=0}^\infty d_{l'}u_{l'}(kr)\psi_{l'}(k, r),$$

where $d_{l'} = c_{l'}A_{l'}e^{-i\delta_{l'}}$. We prove the result in two steps: (a) with $\psi_{l'}(k, r)$ replaced by $\psi_{l'}(k, r) - u_{l'}(kr)$; (b) with $\psi_{l'}(k, r)$ replaced by $u_{l'}(kr)$.

(a) Since the justification is trivial for a finite sum, there is no loss of generality in replacing $\sum_{l'=0}^\infty$ by $\sum_{l'=N}^\infty$, where N is sufficiently large that

$$[(N + \tfrac{1}{4})(N + \tfrac{3}{4})]^{\frac{1}{4}} > \pi\int_0^\infty x\,|V(x)|\,dx.$$

We then have the following estimate, given by de Alfaro and Regge [7]:

$$|\psi_{l'}(k, r)| \leq |u_{l'}(kr)| + \frac{\pi}{2}(\sqrt{r})I_{l'}[(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}}, \tag{B1}$$

where

$$I_{l'} \leq \frac{\int_0^\infty (\sqrt{s})\,|u_{l'}(ks)|\,|V(s)|\,ds}{1 - \tfrac{1}{2}\pi[(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}}\int_0^\infty x\,|V(x)|\,dx} \leq 2\int_0^\infty (\sqrt{s})\,|u_{l'}(ks)|\,|V(s)|\,ds, \quad l' \geq N. \tag{B2}$$

The proof of (B1) depends on the following result (Ref. 7, p. 190):

$$|w_{l'}(kr)u_{l'}(ks)| \leq \frac{\pi}{2} k(rs)^{\frac{1}{2}}[(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}}, \quad r \geq s. \tag{B3}$$

Other estimates of Bessel functions which we need are as follows:

$$|u_{l'}(kr)| \leq Cl'^{\frac{1}{3}}, \quad l' > 0, \tag{B4}$$

$$|u_{l'}(kr)w_{l'}(kr)| \leq C(kr)^{\frac{1}{2}}(l' + 1)^{\frac{1}{3}}, \tag{B5}$$

where the constants are independent of r and l'. (B4) is the weakest of the estimates given by Sabatier (Ref. 3, p. 1528). If we write

$$\begin{aligned} u_{l'}(kr)w_{l'}(kr) &= \frac{\pi kr}{2}\{[J_{l'+\frac{1}{2}}(kr)]^2 + iJ_{l'+\frac{1}{2}}(kr)Y_{l'+\frac{1}{2}}(kr)\} \\ &= kr\left[\int_0^{\pi/2} J_{2l'+1}(2kr\cos\theta)\,d\theta + i\int_0^{\infty} J_{2l'+1}(2kr\cosh t)\,dt\right] \end{aligned}$$

and use (B4) to estimate the integrals, we obtain (B5).

Next, we require the following estimates: If $n^a b_n$ is bounded, where $-\frac{1}{2} < \alpha < 1$, then

$$\sum_{n=0}^{\infty} b_n[u_n(x)]^2 < Cx^{1-\alpha}. \tag{B6}$$

The case $\alpha = 1$ is slightly different: If nb_n is bounded,

$$\sum_{n=0}^{\infty} b_n[u_n(x)]^2 < C + C'\,|\log x|. \tag{B7}$$

Then (B6) and (B7) are proved by noting that

$$b_n\Gamma(n + \tfrac{3}{2} + \alpha/2)/(2n + 1)\Gamma(n + \tfrac{1}{2} - \alpha/2)$$

in the first case and $b_n n(n + 1)/(2n + 1)$ in the second are both bounded, so that the sums on the left-hand side of (B6) and (B7) are bounded by multiples of

$$\begin{aligned} \sum_{n=0}^{\infty}(2n + 1)\frac{\Gamma(n + \frac{1}{2} - \alpha/2)}{\Gamma(n + \frac{3}{2} + \alpha/2)}(u_n(x))^2 &= \frac{\Gamma[(1 - \alpha)/2]}{\Gamma[(1 + \alpha)/2]}\frac{x^{1-\alpha}}{2^{\alpha}}\int_0^{2x}\frac{\sin p}{p^{1-\alpha}}\,dp \\ &< Cx^{1-\alpha} \end{aligned}$$

(since $\int_0^{\infty}\frac{\sin p}{p^{1-\alpha}}\,dp$ converges) and

$$\begin{aligned} \sum_{n=1}^{\infty}\frac{(2n + 1)}{n(n + 1)}(u_n(x))^2 &= \int_0^{2x}\frac{1 - \cos p}{p}\,dp - \sin^2 x \\ &< C + C'\,|\log x|, \end{aligned}$$

respectively.

Observe that (B6) and (B7) remain true if $\sum_{n=0}^{\infty}$ is replaced by $\sum_{n=N}^{\infty}$.

We shall prove that

$$\sum_{l'=N}^{\infty}|d_{l'}|\,|\psi_{l'}(k, r) - u_{l'}(kr)|\,|u_{l'}(kr)| < Cr^{\frac{1}{2}+\epsilon},$$

for $\epsilon > 0$ arbitrarily small. Step (a) will then follow, since

$$\int_0^{\infty}\frac{dr}{r}\left|\frac{d}{dr}\left(\frac{u_l(kr)u_l(k'r)}{r}\right)\right| r^{\frac{1}{2}+\epsilon} < \infty, \quad \epsilon < \tfrac{1}{2}.$$

Writing out $G_{l'}(r, s)$ explicitly, we have

$$\begin{aligned} \psi_{l'}(k, r) - u_{l'}(kr) &= -\frac{i}{k}\int_0^r w_{l'}(kr)u_{l'}(ks)V(s)\psi_{l'}(k, s)\,ds \\ &\quad - \frac{i}{k}\int_r^{\infty} w_{l'}(ks)u_{l'}(kr)V(s)\psi_{l'}(k, s)\,ds \\ &= ie^{i\delta_{l'}}\sin\delta_{l'}w_{l'}(kr) \\ &\quad + \frac{i}{k}\int_r^{\infty} w_{l'}(kr)u_{l'}(ks)V(s)\psi_{l'}(k, s)\,ds \\ &\quad - \frac{i}{k}\int_r^{\infty} w_{l'}(ks)u_{l'}(kr)V(s)\psi_{l'}(k, s)\,ds, \end{aligned}$$

where we have again used

$$\begin{aligned} t_{l'}(k, k) &= -\frac{1}{k^2}\int_0^{\infty} u_{l'}(kr)V(r)\psi_{l'}(k, r)\,dr \\ &= \frac{1}{k}e^{i\delta_{l'}}\sin\delta_{l'}. \end{aligned}$$

It follows from (B1) that

$$\begin{aligned} &|u_{l'}(kr)(\psi_{l'}(k, r) - u_{l'}(kr))| \\ &\quad \leq |\sin\delta_{l'}|\,|u_{l'}(kr)w_{l'}(kr)| \\ &\qquad + \frac{1}{k}|u_{l'}(kr)w_{l'}(kr)|\int_r^{\infty}|V(s)|\,(u_{l'}(ks))^2\,ds \\ &\qquad + \frac{1}{k}|u_{l'}(kr)w_{l'}(kr)|\frac{\pi}{2}I_{l'}[(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} \\ &\qquad \times\int_r^{\infty}|V(s)|\sqrt{s}\,|u_{l'}(ks)|\,ds \\ &\qquad + \frac{1}{k}|u_{l'}(kr)|\int_r^{\infty}|u_{l'}(kr)w_{l'}(ks)|\,|V(s)|\,|u_{l'}(ks)|\,ds \\ &\qquad + \frac{1}{k}|u_{l'}(kr)|\frac{\pi}{2}I_{l'}[(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} \\ &\qquad \times\int_r^{\infty}|u_{l'}(kr)w_{l'}(ks)|\,|V(s)|\,(\sqrt{s})\,ds, \quad l' \geq N, \end{aligned}$$

and so, from (B3) and (B5),

$$
\begin{aligned}
&|u_{l'}(kr)(\psi_{l'}(k,r)-u_{l'}(kr))| \\
&\quad \le C\delta_{l'}(l'+1)^{\frac{1}{3}}(kr)^{\frac{1}{2}} \\
&\qquad + \frac{\pi}{2} r[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{4}} \int_r^\infty |V(s)|\,(u_{l'}(ks))^2\,ds \\
&\qquad + \left(\frac{\pi}{2}\right)^2 r[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} I_{l'} \\
&\qquad \times \int_r^\infty |V(s)|\,|u_{l'}(ks)|\,(\sqrt{s})\,ds \\
&\qquad + \frac{\pi}{2}\sqrt{r}\,|u_{l'}(kr)|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{4}} \\
&\qquad \times \int_r^\infty |V(s)|\,|u_{l'}(ks)|\,(\sqrt{s})\,ds \\
&\qquad + \left(\frac{\pi}{2}\right)^2 \sqrt{r}\,|u_{l'}(kr)|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} I_{l'} \\
&\qquad \times \int_r^\infty s\,|V(s)|\,ds, \quad l' \ge N.
\end{aligned}
$$

We deal with these terms separately, treating three typical cases. Clearly, since $d_{l'}$ is bounded,

$$
\begin{aligned}
\sum_{l'=N}^{\infty} |d_{l'}|\,C\delta_{l'}(l'+1)^{\frac{1}{3}}(kr)^{\frac{1}{2}} &< C'\sum_{l'=N}^{\infty} l_1'^{-3}(l'+1)^{\frac{1}{3}}(kr)^{\frac{1}{2}} \\
&< C''\sqrt{r}.
\end{aligned}
$$

From (B6)

$$
\sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{4}}(u_{l'}(ks))^2 < C(ks)^{\frac{1}{2}}.
$$

So

$$
\begin{aligned}
\int_r^\infty |V(s)| \sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{4}}(u_{l'}(ks))^2\,ds & \\
< C(\sqrt{k})\int_r^\infty |V(s)|\,(\sqrt{s})\,ds < C'\int_r^\infty \frac{ds}{s^{\frac{3}{2}-\epsilon}} & \\
< C'' r^{-\frac{1}{2}+\epsilon} < \infty &
\end{aligned}
$$

and, hence,

$$
\begin{aligned}
&\sum_{l'=N}^{\infty} |d_{l'}|\,\frac{\pi}{2} r[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{4}} \int_r^\infty |V(s)|\,(u_{l'}(ks))^2\,ds \\
&\quad = \frac{\pi}{2} r \int_r^\infty |V(s)| \sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{4}}(u_{l'}(ks))^2 \\
&\quad < Cr^{\frac{1}{2}+\epsilon}.
\end{aligned}
$$

Next,

$$
\begin{aligned}
&\sum_{l'=N}^{\infty} |d_{l'}|\,r[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} I_{l'} \int_r^\infty |V(s)|\,|u_{l'}(ks)|\,(\sqrt{s})ds \\
&\quad = r\int_r^\infty (\sqrt{s})\,|V(s)| \sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} I_{l'} \\
&\qquad \times |u_{l'}(ks)|\,ds,
\end{aligned}
$$

provided the right-hand side exists, and

$$
\begin{aligned}
&\sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} I_{l'}\,|u_{l'}(ks)| \\
&\qquad \le 2\sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}}\,|u_{l'}(ks)| \\
&\qquad\quad \times \int_0^\infty (\sqrt{x})\,|u_{l'}(kx)|\,|V(x)|\,dx \\
&\qquad = 2\int_0^\infty (\sqrt{x})\,|V(x)| \sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} \\
&\qquad\quad \times |u_{l'}(ks)|\,|u_{l'}(kx)|\,dx,
\end{aligned}
$$

provided the right-hand side exists. Now

$$
\begin{aligned}
&\left\{\sum_{l'=N}^{\infty} |d_{l'}|\,[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}}\,|u_{l'}(ks)|\,|u_{l'}(kx)|\right\}^2 \\
&\qquad \le C\sum_{l'=N}^{\infty} [(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}}(u_{l'}(ks))^2 \\
&\qquad\quad \times \sum_{l'=N}^{\infty} [(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}}(u_{l'}(kx))^2 \\
&\qquad < (C + C'\,|\log s|)(C + C'\,|\log x|).
\end{aligned}
$$

Since

$$
\int_0^\infty (\sqrt{x})\,|V(x)|\,\sqrt{(C + C'\,|\log x|)}\,dx < \infty
$$

and

$$
\int_r^\infty (\sqrt{s})\,|V(s)|\,\sqrt{(C + C'\,|\log s|)}\,ds < \infty,
$$

both interchanges are justified, and

$$
\begin{aligned}
&\sum_{l'=N}^{\infty} |d_{l'}|\,r[(l'+\tfrac{1}{4})(l'+\tfrac{3}{4})]^{-\frac{1}{2}} I_{l'} \\
&\qquad \times \int_r^\infty |V(s)|\,|u_{l'}(ks)|\,(\sqrt{s})\,ds \\
&\qquad < Cr\int_r^\infty ds\,|V(s)|\,(\sqrt{s})[\sqrt{(C + C\,|\log s|)}] \\
&\qquad < Cr\int_r^\infty ds\,\frac{s^{\frac{1}{2}}\cdot s^{\epsilon}}{s^{2-\epsilon}} \\
&\qquad < Cr^{\frac{1}{2}+2\epsilon},
\end{aligned}
$$

where ϵ is arbitrarily small.

The remaining terms are similarly dealt with and give the same result. Summing up, we obtain

$$
\sum_{l'=N}^{\infty} |d_{l'}|\,|\psi_{l'}(k,r) - u_{l'}(kr)|\,|u_{l'}(kr)| < Cr^{\frac{1}{2}+\epsilon},
$$

and (a) is proved.

For (b) we have to prove

$$\sum_{l'=0}^{\infty} d_{l'} \int_0^{\infty} dr \, \frac{u_{l'}(kr)u_{l'}(kr)}{r} \frac{d}{dr}\left(\frac{u_l(kr)u_l(k'r)}{r}\right)$$
$$= \int_0^{\infty} dr \sum_{l'=0}^{\infty} d_{l'} \frac{(u_{l'}(kr))^2}{r} \frac{d}{dr}\left(\frac{u_l(kr)u_l(k'r)}{r}\right). \quad \text{(B8)}$$

It is sufficient to prove the result with $d_{l'}$ replaced by 1. Equation (7) shows that $d_{l'} = f_{l'} + 4\alpha/\pi$ where $f_{l'} = O(l^{-\frac{5}{3}-\epsilon})$. Since, as is shown in Ref. 3,

$$\sum_{l'=0}^{\infty} f_{l'}(u_{l'}(kr))^2$$

is a bounded function of r, the integral on the right-hand side of (B8), with $d_{l'}$ replaced by $f_{l'}$, is absolutely convergent. Hence (B8) is valid when $d_{l'}$ is replaced by $f_{l'}$. Hence we have to show that

$$\sum_{l'=0}^{\infty} \int_0^{\infty} dr \, \frac{d}{dr}\left(\frac{(u_{l'}(kr))^2}{r}\right) \frac{u_l(kr)u_l(k'r)}{r}$$
$$= \int_0^{\infty} dr \sum_{l'=0}^{\infty} \frac{d}{dr}\left(\frac{(u_{l'}(kr))^2}{r}\right) \frac{u_l(kr)u_l(k'r)}{r}.$$

[This differs from (B8) in that we have integrated by parts.] Now

$$\frac{d}{dr}\left(\frac{(u_{l'}(kr))^2}{r}\right) = \frac{u_{l'}(kr)}{r}(u_{l'-1}(kr) - u_{l'+1}(kr)),$$

so that

$$\sum_{l'=0}^{M} \int_0^{\infty} dr \, \frac{d}{dr}\left(\frac{(u_{l'}(kr))^2}{r}\right) \frac{u_l(kr)u_l(k'r)}{r}$$
$$= \int_0^{\infty} \frac{dr}{r} \sum_{l'=0}^{M} (u_{l'}(kr)u_{l'-1}(kr) - u_{l'+1}(kr)u_{l'}(kr)) \times \frac{u_l(kr)u_l(k'r)}{r}$$
$$= \int_0^{\infty} \frac{dr}{r} (u_0(kr)u_{-1}(kr) - u_{M+1}(kr)u_M(kr)) \times \frac{u_l(kr)u_l(k'r)}{r}$$
$$= \int_0^{\infty} dr \sum_{l'=0}^{\infty} \frac{d}{dr}\left(\frac{(u_{l'}(kr))^2}{r}\right) \frac{u_l(kr)u_l(k'r)}{r}$$
$$- \int_0^{\infty} \frac{dr}{r} u_{M+1}(kr)u_M(kr) \frac{u_l(kr)u_l(k'r)}{r}.$$

The result now follows if we can prove that

$$\lim_{M\to\infty} \int_0^{\infty} u_{M+1}(kr)u_M(kr)u_l(kr)u_l(k'r) \frac{dr}{r^2} = 0.$$

Using the estimates of the $u_M(kr)$ given in Ref. 3, we have

$$\left| \int_0^{\infty} u_{M+1}(kr)u_M(kr)u_l(kr)u_l(k'r) \frac{dr}{r^2} \right|$$
$$< C_1 \int_0^{\sqrt{M}} dr \left(\frac{4M}{e^2}\right)^{-M/2} + C_2 \int_{\sqrt{M}}^{M-M^{\frac{1}{3}}} \frac{M^{-\frac{1}{6}}}{r^2} dr$$
$$+ C_3 \int_{M-M^{\frac{1}{3}}}^{M} \frac{M^{\frac{1}{3}}}{r^2} dr + C_4 \int_M^{M^2} \frac{M^{\frac{2}{3}}}{r^2} dr$$
$$+ C_5 \int_{M^2}^{\infty} \frac{dr}{r^2} \to 0, \quad \text{as} \quad M \to \infty.$$

This completes the proof of (iii).

Proof of (iv): To prove that

$$\sum_{l'=0}^{\infty} d_{l'} \int_0^{\infty} V(r)u_l(k'r)L_{ll'}(kr)\psi_{l'}(k, r)\, dr$$
$$= \int_0^{\infty} V(r)u_l(k'r) \sum_{l'=0}^{\infty} d_{l'} L_{ll'}(kr)\psi_{l'}(k, r)\, dr,$$

it is sufficient to prove that

$$\sum_{l'=0}^{\infty} |d_{l'}|\,|L_{ll'}(kr)|\,|\psi_{l'}(k, r)| < Cr^{\frac{1}{2}}, \quad \text{(B9)}$$

since $u_l(k'r)$ is bounded and $\int_0^{\infty} r^{\frac{1}{2}}|V(r)|\, dr < \infty$. Also, since $L_{ll'}(kr)\psi_{l'}(k, r)$ is bounded, it is enough to verify (B9) with $\sum_{l'=0}^{\infty}$ replaced by $\sum_{l'=N}^{\infty}$. We shall suppose that $N > l$ and

$$[(N + \tfrac{1}{4})(N + \tfrac{3}{4})]^{\frac{1}{4}} > \pi \int_0^{\infty} x\,|V(x)|\, dx.$$

Then, using (B4),

$$L_{ll'}(kr) = \int_0^{kr} u_l(x)u_{l'}(x) \frac{dx}{x^2}$$
$$= \frac{u_{l'}(kr)u_{l+1}(kr) - u_{l'+1}(kr)u_l(kr)}{(l' - l)(l' + l + 1)} + \frac{u_{l'}(kr)u_l(kr)}{kr(l' + l + 1)}$$
$$< Cl'^{-\frac{4}{3}} + \frac{C'l'^{-\frac{1}{3}}}{r},$$

so that, from (B1),

$$\sum_{l'=N}^{\infty} |d_{l'}|\,|L_{ll'}(kr)|\,|\psi_{l'}(k, r)|$$
$$< \sum_{l'=N}^{\infty} \left(Cl'^{-\frac{4}{3}} + \frac{C'l'^{-\frac{1}{3}}}{r}\right) \times \left(|u_{l'}(kr)| + \frac{\pi}{2}\sqrt{r}[(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} I_{l'}\right).$$

Now

$$\left(\sum_{l'=N}^{\infty} l'^{-\frac{4}{3}} |u_{l'}(kr)|\right)^2 \leq \sum_{l'=N}^{\infty} l'^{-\frac{8}{3}} \sum_{l'=N}^{\infty} (u_{l'}(kr))^2 < Cr$$

and

$$\left(\sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} |u_{l'}(kr)|\right)^2 \leq \sum_{l'=N}^{\infty} l'^{-\frac{5}{3}} \sum_{l'=N}^{\infty} l'(u_{l'}(kr))^2 < Cr^2,$$

so

$$\frac{1}{r} \sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} |u_{l'}(kr)| < C.$$

Also

$$\sum_{l'=N}^{\infty} l'^{-\frac{4}{3}} [(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} I_{l'}$$

and

$$\sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} [(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} I_{l'}$$

both converge, e.g.,

$$\left(\sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} [(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} |u_{l'}(ks)|\right)^2 \leq C \sum_{l'=N}^{\infty} (u_{l'}(ks))^2 < C's$$

and

$$\int_0^{\infty} (\sqrt{s})\, |V(s)| \sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} [(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} |u_{l'}(ks)|\, ds < C \int_0^{\infty} s\, |V(s)|\, ds < \infty,$$

so that

$$\sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} [(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} I_{l'} < 2 \int_0^{\infty} ds(\sqrt{s})\, |V(s)| \times \sum_{l'=N}^{\infty} l'^{-\frac{1}{3}} [(l' + \tfrac{1}{4})(l' + \tfrac{3}{4})]^{-\frac{1}{4}} |u_{l'}(ks)| < C \int_0^{\infty} s\, |V(s)|\, ds < \infty.$$

It follows that

$$\sum_{l'=N}^{\infty} |d_{l'}|\, |L_{ll'}(kr)|\, |\psi_{l'}(k, r)| < C\sqrt{r},$$

which proves (iv).

[1] L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering Theory* (I.P.S.T., 1965), p. 25.
[2] R. G. Newton, J. Math. Phys. **3**, 75 (1962).
[3] P. C. Sabatier, J. Math. Phys. **7**, 1515 (1966).
[4] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Book Co., New York, 1966), pp. 623–633.
[5] P. C. Sabatier, J. Math. Phys. **8**, 1957 (1967).
[6] P. C. Sabatier, J. Math. Phys. **7**, 2079 (1966).
[7] V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland Publ. Co., 1965), p. 84.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Relaxation to Quantum Statistical Equilibrium of the Wigner–Weisskopf Atom in a One-Dimensional Radiation Field. II. Finite Systems

RUSSELL DAVIDSON
Center for Statistical Mechanics and Thermodynamics, The University of Texas at Austin
Austin, Texas 78712

AND

JOHN J. KOZAK
Department of Chemistry, The University of Notre Dame, Notre Dame, Indiana 46556

(Received 7 October 1969)

In this paper, a model is studied in which it was possible to obtain an expression for the exact solution of the master equation for the problem of spontaneous emission in a finite system. The model chosen for discussion is that of the Wigner–Weisskopf atom in interaction with a massless boson field, here 1-dimensional. The solution takes the form of a constant term plus a time-dependent one, expressed as the sum of residues at a series of poles along the real axis of a Laplace transform variable. Numerical calculations were performed on various aspects of the solution, and although these were too delicate to be quite certain around sensitive values of the time, the general picture is clear: After an initial decay to a value near zero, the series gave rise at fairly regular intervals to rapid and large fluctuations, the size of which never quite attains the initial value, but may nonetheless be large even after very long times. This result seems to be in agreement with the observation made by Zwanzig, namely, that for finite systems the master equation might demonstrate properties associated with the finite size of the system which would become important over certain long-time scales and would be ignored by the use of the customary "thermodynamic limit" of statistical mechanics. The relationship between this work and that of Montroll and Mazur and of Rubin is also discussed.

I. INTRODUCTION

This paper deals with the problem of the description by nonequilibrium statistical mechanics of a finite system. The model chosen for discussion is that of the Wigner–Weisskopf atom: a two-level quantum system in interaction with a massless boson field, here 1-dimensional. The problem was brought up in an article of Zwanzig,[1] in which he suggested that the solution to an equation with the form of what is now usually called the Prigogine–Résibois master equation might demonstrate properties associated with the finite size of a system, which would become important over certain long-time scales and would be ignored by the use of the customary "thermodynamic limit" of statistical mechanics by which systems are assumed to have an infinite number of degrees of freedom and to be unlimited in extent. A further problem raised by Zwanzig in the same article, that of the existence of nonexponential behavior in the thermodynamic limit, has been discussed in a previous paper by the authors,[2] hereafter referred to as I. The same model—that of the Wigner-Weisskopf atom—was used for the discussion of that paper, and the results established there are used extensively in what follows here.

The Hamiltonian for the model [Eq. (I.6)] is

$$H = \epsilon_1 \alpha\alpha^* + \epsilon_2 \alpha^*\alpha + \sum_\lambda [\tfrac{1}{2}\hbar\omega_\lambda (a_\lambda^* a_\lambda + 1)] + \sum_\lambda (h_\lambda^* \alpha^* a_\lambda + h_\lambda \alpha a_\lambda^*),$$

where ϵ_1 is the energy of the ground state of the two-level system $|1\rangle$ and ϵ_2 that of the excited state $|2\rangle$. The fermion operators α and α^* are

$$\alpha = |1\rangle\langle 2|,$$
$$\alpha^* = |2\rangle\langle 1|.$$

The operators for the boson field are labeled by the normal modes λ, and are defined by their matrix elements in the occupation number representation:

$$\langle n_\lambda | a_\lambda | m_\lambda \rangle = [2(n_\lambda + 1)]^{\frac{1}{2}} \delta^{\mathrm{kr}}(m_\lambda - n_\lambda - 1),$$
$$\langle m_\lambda | a_\lambda^* | n_\lambda \rangle = [2(n_\lambda + 1)]^{\frac{1}{2}} \delta^{\mathrm{kr}}(m_\lambda - n_\lambda - 1).$$

The state $|n_\lambda\rangle$ is a state with n_λ, $n = 0, 1, 2, \cdots$, photons in the λth mode. The symbol $\delta^{\mathrm{kr}}(\cdots)$ is a Kronecker delta. The h_λ measure the strength of the coupling between the atom and the radiation. The states of the whole system are taken to be

$$|i; \{n_\lambda\}\rangle = |i\rangle \prod_\lambda |n_\lambda\rangle,$$

with $i = 1, 2$ and $n_\lambda = 0, 1, 2, \cdots$, for each λ. The model will be described by the density matrix ρ, which satisfies the Liouville–von Neumann equation:

$$\frac{\partial \rho}{\partial t} = \frac{1}{i\hbar}[H, \rho] = \frac{1}{i\hbar}[H\rho - \rho H].$$

From this may be derived the Prigogine–Résibois master equation [Eq. (I.24)] for the diagonal elements

of the density matrix

$$\frac{\partial}{\partial t}\rho_0(N,t)=\int_0^t d\tau \mathcal{C}(\tau)\rho_0(N,t-\tau)+\mathfrak{D}(t,\{\rho_\nu(N,0)\}), \tag{1}$$

in which the matrix elements $\rho_\nu(N,t)$ of the density matrix are given by

$$\rho_\nu(N,t)=\langle N+\tfrac{1}{2}\nu|\,\rho\,|N-\tfrac{1}{2}\nu\rangle,$$

where N and ν are shorthand for the set of quantum numbers corresponding to each degree of freedom of the system. The operators $\mathcal{C}$ and $\mathfrak{D}$ are defined in terms of a perturbation theory. The particular problem to be treated in this paper, as in I, is that of spontaneous emission, for which the initial state of the system is specified such that the atom is in its excited state. Thus,

$$\rho_\nu(N,0)=\delta^{\mathrm{kr}}(\nu)\delta^{\mathrm{kr}}(N_p-2)\prod_\lambda \delta^{\mathrm{kr}}(N_\lambda),$$

where N_p is the N number for the atom and the N_λ are those for the modes of the boson field. In this case, the term $\mathfrak{D}$ disappears in Eq. (1), which may then be solved by Laplace transform [see Eq. (I.31)] to yield

$$\rho_0(N,t)=-\frac{1}{2\pi}\int_C dz e^{-izt}[\psi(z)+iz]^{-1}\rho_0(N,0), \tag{2}$$

where C is a contour in the z plane parallel to the real axis and above all singularities of the integrand. The operator $\psi(z)$ is the Laplace transform of the operator $\mathcal{C}(\tau)$ of Eq. (1) and is given by Eq. (I.35). For the case of spontaneous emission, Eq. (2) takes the following form in the lowest order of the perturbation theory yielding $\psi(z)$, this form coming from Eqs. (I.38) and (I.39):

$$\rho_0(\mathcal{N},t)=-\frac{1}{2\pi}\int_C dz e^{-izt}\sigma(\mathcal{N}), \tag{3}$$

where

$$\sigma(\mathcal{N})=\frac{1}{iz}[1+S(z)]^{-1}$$

with

$$S(z)=\sum_\lambda\left(1-\frac{\hbar^2}{4|h_\lambda|^2}[z^2-(c|k_\lambda|-E)^2]\right)^{-1}. \tag{4}$$

In these expressions, the argument $\mathcal{N}$ denotes this choice of N variables:

$$N_p=2,\quad N_\lambda=0,\quad \text{for all } \lambda.$$

Further,

$$\hbar E=\epsilon_2-\epsilon_1\quad\text{and}\quad c|k_\lambda|=\omega_\lambda,$$

so that c is the velocity of light.

We shall consider our system to be enclosed in a well of length L. In this way, if L is allowed to tend to infinity, the results of I should be recovered. The normal modes λ are now characterized by their wavenumbers k_λ as $k_\lambda=2\pi n/L$, where the positive integer n replaces λ as the label. The specification of the problem is completed by a choice of h_λ. This will be taken, following the arguments of I, such that

$$|h_\lambda|^2=\hbar^2\alpha cE/L,$$

where α is a dimensionless coupling constant replacing the fine-structure constant of quantum electrodynamics in a 1-dimensional system.

In the next section, the expression (4) for $S(z)$ will be obtained by performing the summation over λ and, in Sec. III, it will be verified that the result yields agreement with the expressions derived in I, for the limit in which $L\to\infty$. In Sec. IV, the Laplace transform solution, Eq. (3), of the master equation is examined and some of its properties studied. An exact expression for this solution is obtained in Sec. V for the limiting case of weak coupling, which is in any event the only case legitimately treated by the lowest order of the perturbation theory leading to the Prigogine–Résibois equation (1). This weak-coupling solution, though complicated, is seen in Sec. VI to yield in the thermodynamic limit exactly the exponential decay usually predicted by calculations on large systems. However, since the principal object of the present investigation is a search for discrepancies from this exponential solution arising from the finite size of the system, in Secs. VII–IX a numerical investigation of the solution obtained in Sec. V is described. Here, it is indeed found that a wide variety of phenomena appears in the régime not only where L is small, but where it may assume a value up to 10^4 times the wavelength of the radiation in resonance with the energy gap in the two-level atom. In particular, the solution predicts the possibility of Poincaré recurrences in the system over very long-time scales. Section X is devoted to a discussion of the results obtained in the paper and the presentation of conclusions.

II. SUMMATION OVER NORMAL MODES

In Eq. (3) it is seen that, to obtain the time dependence of $\rho_0(\mathcal{N},t)$, one must evaluate the sum

$$S(z)\equiv\sum_\lambda\left(1-\frac{\hbar^2}{4|h_\lambda|^2}[z^2-(c|k_\lambda|-E)^2]\right)^{-1}, \tag{5}$$

where $|h_\lambda|^2$ has been chosen as $\hbar^2\alpha cE/L$. The sum extends over the normal modes of the boson field, for which k_λ takes the values $2\pi n/L$ with n integral. So,

explicitly, Eq. (5) becomes

$$S(z) = \sum_{n=-\infty}^{\infty}\left\{1 - \frac{L}{4\alpha cE}\left[z^2 - \left(\frac{2\pi c\,|n|}{L} - E\right)^2\right]\right\}^{-1}$$
$$= -\left(1 - \frac{L}{4\alpha cE}(z^2 - E^2)\right)^{-1}$$
$$+ 2\sum_{n=0}^{\infty}\left\{1 - \frac{L}{4\alpha cE}\left[z^2 - \left(\frac{2\pi cn}{L} - E\right)^2\right]\right\}^{-1}. \tag{6}$$

One may write the denominator of the summand in Eq. (6) in the form

$$\frac{\pi^2 c}{\alpha EL}(n + r_1)(n + r_2),$$

where

$$r_{1,2} = \frac{L}{2\pi c}\left[-E \pm \left(z^2 - \frac{4\alpha cE}{L}\right)^{\frac{1}{2}}\right].$$

The sum can thus be performed as

$$\sum_{n=0}^{\infty}\left\{1 - \frac{L}{4\alpha cE}\left[z^2 - \left(\frac{2\pi cn}{L} - E\right)^2\right]\right\}^{-1}$$
$$= \frac{\alpha EL}{\pi^2 c}\cdot\frac{1}{r_2 - r}\left(\sum_{n=0}^{\infty}\frac{1}{n + r_1} - \sum_{n=0}^{\infty}\frac{1}{n + r_2}\right)$$
$$= \frac{\alpha EL}{\pi^2 c}\cdot\frac{1}{r_2 - r_1}\cdot[\zeta(1, r_1) - \zeta(1, r_2)], \tag{7}$$

where

$$\zeta(s, a) = \sum_{n=0}^{\infty}\frac{1}{(a + n)^s}$$

is the generalized Riemann ζ function (see, for example, Ref. 3). This function has a pole, with residue 1, at $s = 1$, but the difference of the two functions in Eq. (7) is regular. This may be seen by using Hermites formula[3] or by the relation between the ζ function and the γ function:

$$\lim_{s\to 1}\left(\zeta(s, a) - \frac{1}{s - 1}\right)$$
$$= -\log a + \frac{1}{2a} + 2\int_0^{\infty}\frac{y\,dy}{(a^2 + y^2)(e^{2\pi y} - 1)}$$
$$= -\frac{\Gamma'(a)}{\Gamma(a)}. \tag{8}$$

This last expression, the logarithmic derivative of the γ function, is also called the zeroth polygamma function,[3] ϕ:

$$\phi(w) = \frac{d}{dw}\log\Gamma(w) = \frac{\Gamma'(w)}{\Gamma(w)}.$$

In terms of ϕ, Eq. (7) becomes

$$\frac{\alpha EL}{\pi^2 c}\cdot\frac{1}{r_2 - r_1}[\phi(r_2) - \phi(r_1)]. \tag{9}$$

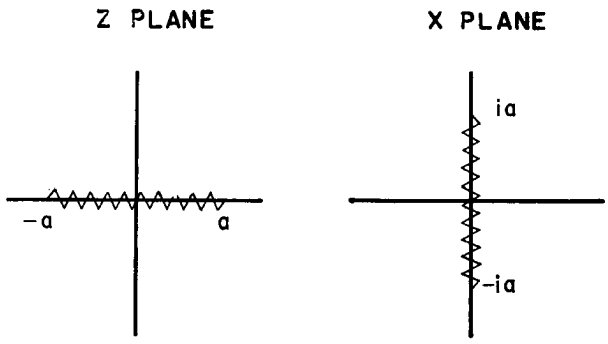


FIG. 1. The complex (x, z) planes [see the discussion following Eq. (11)].

As a function of a complex variable w, $\phi(w)$ is regular in the right-hand half-plane and has simple poles at $w = 0, -1, -2, \cdots$. It may be analytically continued into the left-hand half-plane by the functional relation

$$\phi(-w) = \phi(w + 1) + \pi\cot\pi w. \tag{10}$$

It is convenient at this stage to introduce a new complex variable x, defined in terms of the Laplace transform z as

$$x = \left(z^2 - \frac{4\alpha Ec}{L}\right)^{\frac{1}{2}}. \tag{11}$$

Clearly, as $L\to\infty$, $x\to z$. In order that Eq. (11) represent a regular single-valued mapping, the (x, z) planes will be cut as shown in Fig. 1. The branch points then map into one another as

$$a = \left(\frac{4\alpha cE}{L}\right)^{\frac{1}{2}},$$

where

$$z = a \leftrightarrow x = 0+, \tag{12a}$$
$$z = i0+ \leftrightarrow x = ia, \tag{12b}$$
$$z = i0- \leftrightarrow x = -ia, \tag{12c}$$
$$z = -a \leftrightarrow x = 0-. \tag{12d}$$

In terms of x, the expression (9) is

$$\frac{\alpha E}{\pi x}\left[\phi\left(-\frac{L}{2\pi c}(E - x)\right) - \phi\left(-\frac{L}{2\pi c}(E + x)\right)\right]. \tag{13}$$

The first term of the expression (13) is regular, for $\mathrm{Re}\,x < -E$, and the second term, for $\mathrm{Re}\,x > +E$. Using the relation (10), we may write this expression for the three ranges of x as follows:

for $\mathrm{Re}\,x < -E$:

$$-\frac{\alpha E}{\pi x}\left\{\phi\left(-\frac{L}{2\pi c}(E + x)\right) - \phi\left[\frac{L}{2\pi c}\left(E - x + \frac{2\pi c}{L}\right)\right]\right.$$
$$\left. - \pi\cot\left(\frac{L}{2c}(E - x)\right)\right\}; \tag{14}$$

for $-E < \mathrm{Re}\, x < E$:

$$-\frac{\alpha E}{\pi x}\left\{\phi\left[\frac{L}{2\pi c}\left(E+x+\frac{2\pi c}{L}\right)\right]\right.$$
$$-\phi\left[\frac{L}{2\pi c}\left(E-x+\frac{2\pi c}{L}\right)\right]$$
$$\left.+\pi\cot\left(\frac{L}{2c}(E+x)\right)-\pi\cot\left(\frac{L}{2c}(E-x)\right)\right\}; \quad (15)$$

for $\mathrm{Re}\, x > E$:

$$-\frac{\alpha E}{\pi x}\left\{\phi\left[\frac{L}{2\pi c}\left(E+x+\frac{2\pi c}{L}\right)\right]\right.$$
$$\left.-\phi\left(-\frac{L}{2\pi c}(E-x)\right)+\pi\cot\left(\frac{L}{2c}(E+x)\right)\right\}. \quad (16)$$

In the second of these ranges, we may also express this quantity in the integral representation given in Eq. (8). We obtain for Eq. (6)

$$-\frac{4\alpha Ec}{L}\left(\frac{1}{E^2-x^2}+\frac{1}{E'^2-x^2}\right)$$
$$+\frac{2\alpha E}{\pi x}\log\left(\frac{E'-x}{E'+x}\right)-\frac{16\alpha EE'}{\pi}$$
$$\times\int_0^\infty\frac{w\,dw}{(e^{Lw/c}-1)(w^2+(E'+x)^2)(w^2+(E'-x)^2)}$$
$$-\frac{2\alpha E}{x}\cot\left(\frac{L}{2c}(E'+x)\right)+\frac{2\alpha E}{x}\cot\left(\frac{L}{2c}(E'-x)\right), \quad (17)$$

where

$$E' = E + 2\pi c/L$$

and where the change of variable $w = 2\pi yc/L$ is used in the integral. The last two terms of the above expression (17) may then be grouped together to yield

$$\frac{2\alpha E}{x}\sin\left(\frac{Lx}{c}\right)\csc\left(\frac{L}{2c}(E'+x)\right)\csc\left(\frac{L}{2c}(E'-x)\right).$$

Similar integral representations may readily be written for the other two ranges of x.

III. THE INFINITE SYSTEM LIMIT

It is important to verify that in the limit, as $L\to\infty$, the expressions (14)–(16) go over into that which is calculated in I for the same problem in the limit of a large system. From Eq. (I.44), one sees that the expression which corresponds to the function $S(z)$ in Eq. (6) is

$$\frac{2\alpha E}{\pi z}\left[\log\left(\frac{E-z}{E+z}\right)+2\pi i\right], \quad (18)$$

where the branch of the logarithm is such that, when $z = 0$, we take

$$\log\left(\frac{E-z}{E+z}\right) = 0.$$

In Eq. (6) itself, the first term is zero, when $L\to\infty$.

To evaluate the limits of (14)–(16), the following results are needed:

$$\text{for Im } w > 0, \quad \lim_{L\to\infty}\cot(Lw) = -i,$$
$$\text{for Im } w < 0, \quad \lim_{L\to\infty}\cot(Lw) = +i,$$
$$\text{for Re } w > 0, \quad \phi(Lw)\sim\log(Lw), \quad \text{as} \quad L\to\infty,$$

where the branch of the logarithm is the usual one with $\log(Lw = 1) = 0$. With these relations, the expression (14) has this limit, as $L\to\infty$:

$$\lim_{L\to\infty}\left(-\frac{\alpha E}{\pi z}\left\{\log\left(-\frac{L}{2\pi c}(E+z)\right)\right.\right.$$
$$\left.\left.-\log\left[\frac{L}{2\pi c}\left(E-z+\frac{2\pi c}{L}\right)\right]-\pi i\right\}\right),$$

since $x\to z$, $E'\to E$, as $L\to\infty$. Then, for (14) we have

$$\frac{\alpha E}{\pi z}\left[\log\left(\frac{z-E}{z+E}\right)+i\pi\right], \quad (19)$$

for (15) we have

$$\frac{\alpha E}{\pi z}\left[\log\left(\frac{E-z}{E+z}\right)+2\pi i\right], \quad (20)$$

and for (16) we have

$$-\frac{\alpha E}{\pi z}\left[\log\left(\frac{z+E}{z-E}\right)-i\pi\right]. \quad (21)$$

The three expressions (19)–(21) all clearly represent the same analytic function except perhaps for the choice of the branch of the logarithm. When $z\to\infty$, (19) and (21) both behave asymptotically as $i\alpha E/z$. The expression (20) has a different asymptotic behavior for $\mathrm{Im}\, z > 0$ and $\mathrm{Im}\, z < 0$. In the former case, the expression is proportional to $i\alpha E/z$, and so in the upper half-plane, all three expressions yield for Eq. (6) in the limit, as $L\to\infty$, the result

$$\frac{2\alpha E}{\pi z}\left[\log\left(\frac{E-z}{E+z}\right)+2\pi i\right], \quad (22)$$

with

$$\log\left(\frac{E-0}{E+0}\right) = 0.$$

This is exactly (18). For $\mathrm{Im}\, z < 0$, the asymptotic expression for (20) is different if one stays on the same branch of the logarithm, and so (22) will be a

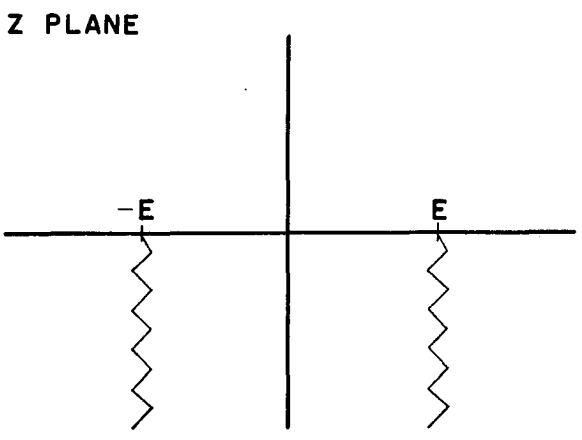


FIG. 2. The complex z plane [see the discussion following Eq. (22)].

single-valued function on the cut plane shown in Fig. 2. The cuts are equivalent to those used in I for the evaluation of $\rho_0(\mathcal{N}, t)$.

Although we have now verified completely that Eq. (6) does indeed go over to (18) when $L \to \infty$, it is instructive to see how, when $z \to i0+$, the expression (15) yields a behavior like z^{-1} when L is infinite. This property is of importance because the quantity

$$\lim_{z \to i0+} (-izS(z))^{-1} \tag{23}$$

yields the first-order relaxation time (exponential decay) of the excited state of the 2-level system (see Sec. IV of I). Since exponential decay may be expected to be a property only of infinite systems, it is of note that when $L \to \infty$, one obtains the usual answer from this calculation. First, we see from (12) that $z \to i0+$ means that $x \to +ia$. From expression (15), we have, for $L \to \infty$,

$$\begin{aligned} S(ia) \sim \frac{2i\alpha E}{\pi a}\Big[&\log\left(\frac{L}{2\pi c}(E' + ia)\right) \\ &- \log\left(\frac{L}{2\pi c}(E' - ia)\right)\Big]\frac{2\alpha E}{z} \\ &\times \left[\sin\left(\frac{iLa}{c}\right)\csc\left(\frac{L}{2c}(E' + ia)\right)\right. \\ &\left.\times \csc\left(\frac{L}{2c}(E' - ia)\right)\right]. \end{aligned} \tag{24}$$

Now when $L \to \infty$, $a \to 0$, but $aL \to \infty$. Thus, in expression (24), the logarithmic terms cancel in the limit, and the remaining term tends to $4i\alpha E/z$. The quantity (23) is thus $(4\alpha E)^{-1}$, in agreement with Eq. (I.48).

IV. THE CONTOUR INTEGRAL SOLUTION

The time dependence of $\rho_0(\mathcal{N}, t)$ is given by the contour integral as

$$\begin{aligned} \rho_0(\mathcal{N}, t) &= -\frac{1}{2\pi}\int_C dz e^{-izt}\sigma(\mathcal{N}) \\ &= -\frac{1}{2\pi}\int_C dz e^{-izt}\left(\frac{1}{iz}\,\frac{1}{1 + S(z)}\right), \end{aligned} \tag{25}$$

for the case of spontaneous emission. Let us write Eq. (25) in terms of the variable x defined in Eq. (11):

$$\begin{aligned} \rho_0(\mathcal{N}, t) = -\frac{1}{2\pi i}\int_C dx \exp\left[-i(x^2 + a^2)^{\frac{1}{2}}t\right] \\ \times \frac{x}{x^2 + a^2}\cdot\frac{1}{1 + S(x)}, \end{aligned} \tag{26}$$

where S has been expressed as a function of x [Eq. (17) and corresponding expressions]. To evaluate the integral in Eq. (26), the contour C may be closed by a large semicircle in the lower half-plane of x. Then for the application of Cauchy's theorem, it is necessary to examine all the singularities of the integrand, for by definition, C is taken above all these singularities. One sees at once that there are two simple poles at $x = \pm ia$; furthermore, one sees that, because of the square root in the exponent, a cut must be made between these poles so as to have a single-valued function in the integrand. The other singularities arise from the zeros of $1 + S(x)$, and these will be considered in detail shortly. We must consider therefore an integral round the contour of Fig. 3. Because the integral vanishes on the large semicircle, $\rho_0(\mathcal{N}, t)$ is then obtained by calculating the residues at $x = \pm ia$ and the zeros of $[1 + S(x)]$, and contributions from the branch cut mentioned or any others that arise from the properties of $S(x)$.

It is readily seen that there is, in fact, no contribution from the cut between ia and $-ia$. The integral along the cut is

$$\begin{aligned} &\frac{2i}{-2\pi i}\int_{-ia}^{ia} dx \sin\left[t(x^2 + a^2)^{\frac{1}{2}}\right]\cdot\frac{x}{x^2 + a^2}\cdot\frac{1}{1 + S(x)} \\ &= \frac{1}{\pi}\left(\int_0^a dy + \int_{-a}^0 dy\right)\sin\left[t(a^2 - y^2)^{\frac{1}{2}}\right] \\ &\quad\times \frac{y}{a^2 - y^2}\cdot\frac{1}{1 + S(iy)} = 0, \end{aligned}$$

since it is clear from expression (17), which is the relevant expression for $S(x)$ in the region of the cut, that $S(+iy) = S(-iy)$. As for the poles at $x = \pm ia$, the residue at each is easily seen to be $1/2[1 + S(ia)]$. Since we have to integrate round the poles in a

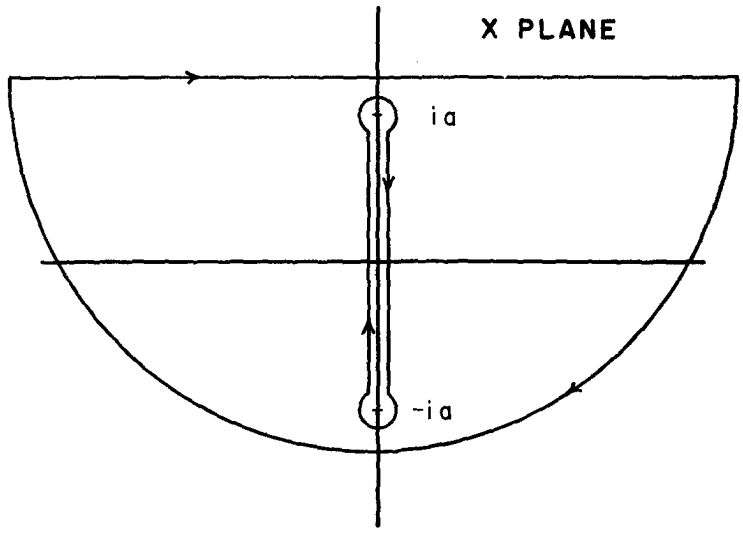


FIG. 3. The contours used in evaluating Eq. (26).

negative sense (Fig. 3), the contribution to $\rho_0(\mathcal{N}, t)$ from these is

$$1/[1 + S(ia)]. \tag{27}$$

The expression (27) gives, then, a constant (that is, time-independent) contribution to $\rho_0(\mathcal{N}, t)$. The behavior of $S(+ia)$, as $L \to \infty$, has already been calculated [Eq. (24)], and it has been seen that it becomes infinite with L. This constant contribution thus disappears in the limit of a large system, as one would expect. In this case, the two poles coalesce into a pole of $S(z)$ at $z = 0$ which, as has been remarked [see the discussion following expression (23)], is responsible for the appearance of a finite relaxation time. However, because $S(ia)$ is *not* singular when L is finite, we have that

$$\lim_{z \to i0+} [-izS(z)]^{-1}$$

is formally infinite except for $L \to \infty$. We may thus see that the notion of a relaxation time is meaningful only in the limit of a large system.

It is important to check that $S(ia)$ is real and indeed positive so that the expression (27) may yield a positive contribution to $\rho_0(\mathcal{N}, t)$. It can be shown that $S(iy)$ is positive along the entire imaginary axis and, in particular, at ia. The proof is not difficult, but tedious, and is relegated to the Appendix.

We must now consider the zeros of $1 + S(x)$. We shall consider these only in the range $-E < \mathrm{Re}\, x < E$, but similar results will pertain in the other ranges of x. It is useful to rewrite expression (17) in terms of dimensionless variables:

$$\xi = \frac{Lx}{c}, \quad \eta = \frac{LE}{c}.$$

There results

$$S(\xi) = -4\alpha\eta\left(\frac{1}{\eta^2 - \xi^2} + \frac{1}{(\eta + 2\pi)^2 - \xi^2}\right) + \frac{2\alpha\eta}{\pi\xi} \log\left(\frac{\eta + 2\pi - \xi}{\eta + 2\pi + \xi}\right) - \frac{16\alpha\eta(\eta + 2\pi)}{\pi} \times \int_0^\infty \frac{\xi\, d\xi}{(e^\xi - 1)[\xi^2 + (\eta + \xi + 2\pi)^2][\xi^2 + (\eta - \xi + 2\pi)^2]} + \frac{2\alpha\eta}{\xi} \sin\xi \csc\tfrac{1}{2}(\eta + \xi) \cdot \csc\tfrac{1}{2}(\eta - \xi). \tag{28}$$

In this expression, η is of necessity a *real* parameter and it is clear that, when ξ (and so x) is real, $S(\xi)$ itself is real. It is proved in the Appendix that $S(\xi)$ is real for purely imaginary ξ as well, but in general $S(\xi)$ is complex. Thus it is only on the real and imaginary axes of ξ that $1 + S(\xi)$ may vanish. The latter case is also impossible, for S is *positive on* the imaginary axis. On the other hand, since the last term in Eq. (28) may be written as

$$\frac{2\alpha\eta}{\xi} \cot\tfrac{1}{2}(\eta - \xi) - \frac{2\alpha\eta}{\xi} \cot\tfrac{1}{2}(\eta + \xi),$$

it can be seen that $S(\xi)$ may equal -1 at many points on the real axis because there the cotangents are rapidly varying functions of ξ which in a short range assume all values between $-\infty$ and $+\infty$. The zeros of $1 + S(x)$ have thus been shown to lie only on the real axis.

It should be remarked that, in considering all the singularities of the integrand in Eq. (26), one should consider whether any branch points arise from the logarithms appearing in $S(x)$, especially since it is

known that these do appear and are important in the limit of a large system (see I). It can be seen, however, that this is not the case, since, if $S(x)$ is expressed as in (14)–(16) in terms of the poly-γ function, the only singularities are known to be poles. It may also be checked in Eq. (28) and the corresponding expressions for the other ranges of x that the branch points of the logarithms which appear always lie outside the range of validity of the expressions. Taking into account only the zeros of $1 + S(x)$, one is thus treating all the singularities which are present. In the following sections, the contribution to $\rho_0(\mathcal{N}, t)$ associated with these zeros will be calculated for the limiting case of weak coupling, that is, for $\alpha \to 0$. This restriction affords a great simplification of the analysis and enables us to obtain a description of the time-dependent part of $\rho_0(\mathcal{N}, t)$.

V. THE WEAK-COUPLING CASE

It has been pointed out in I that, to derive a consistent weak-coupling approximation, it is necessary to introduce a dimensionless measure of the time *scaled* by the coupling constant α. As there, we shall choose this to be $\tau = \alpha E t$. In the present investigation, we have also to deal with the size of the system, measured up to now by L or η. It turns out, when one writes the constant contribution to $\rho_0(\mathcal{N}, t)$ in terms of α and η, that η occurs only in the combination $\alpha\eta$. This is, in fact, the proper variable to use in the present problem, inasmuch as $\alpha\eta = \alpha E L/c$ is the time, measured by the scaled parameter τ, required for a wave to cross the system of length L. Were we to employ α and η as independent variables then and to let α become very small, some of the full dependence of the results on η would be lost. It is evident that η, like the time t, must be scaled by α in order that the proper measure of the time required to cross the system remain finite. For reasons of convenience, the actual dimensionless measure of the length of the system which we use is

$$\sigma = (\alpha\eta)^{\frac{1}{2}}.$$

In terms of α, σ, and τ, Eq. (26) becomes

$$\rho_0(\mathcal{N}, t) = -\frac{1}{2\pi i}\int_C d\xi \cdot \exp\left(-\frac{iz}{\sigma^2}(\xi^2 + 4\sigma^2)^{\frac{1}{2}}\right) \times \frac{\xi}{\xi^2 + 4\sigma^2} \cdot \frac{1}{1 + S(\xi)}. \quad (29)$$

It is useful, as in I, to replace the Laplace transform variable ξ by

$$x/\alpha E = \xi/\sigma^2 = \theta.$$

Then, in the limit of a large system, θ corresponds to the variable ξ of Eq. (I.52). Using θ, we may write Eq. (29) as

$$\rho_0(\mathcal{N}, t) = -\frac{1}{2\pi i}\int_C d\theta \exp\left[-i\tau(\theta^2 + 4/\sigma^2)^{\frac{1}{2}}\right] \times \frac{\theta}{\theta^2 + (4/\sigma^2)} \cdot \frac{1}{1 + S(\theta)}. \quad (30)$$

We shall now write out $S(\theta)$ explicitly in the range which corresponds to

$$-1/\alpha < \mathrm{Re}\,\theta < 1/\alpha, \quad -E < \mathrm{Re}\,x < +E,$$

and see that our choice of variables enables us easily to derive the weak-coupling limit. We have

$$S(\theta) = -\frac{4\alpha^2}{\sigma^2}\left(\frac{1}{1 - \alpha^2\theta^2} + \frac{1}{(1 + 2\pi\alpha/\sigma^2)^2 - \alpha^2\theta^2}\right) + \frac{2}{\pi\theta}\log\left(\frac{1 - \alpha\theta + 2\pi\alpha/\sigma^2}{1 + \alpha\theta + 2\pi\alpha/\sigma^2}\right) - \frac{16\alpha^3(1 + 2\pi\alpha/\sigma^2)}{\pi}\int_0^\infty \frac{w\,dw}{(e^{\sigma^2 w} - 1)[\alpha^2 w^2 + (1 + \alpha\theta + 2\pi\alpha/\sigma^2)^2][\alpha^2 w^2 + (1 - \alpha\theta + 2\pi\alpha/\sigma^2)^2]} + \frac{2}{\theta}\sin(\sigma^2\theta)\cdot\csc\frac{\sigma^2}{2\alpha}(1 - \alpha\theta)\cdot\csc\frac{\sigma^2}{2\alpha}(1 + \alpha\theta). \quad (31)$$

A check that our choice of variables is reasonable may be made here by letting σ tend to infinity while keeping α and θ finite. Only the second and fourth terms contribute, and they yield

$$\frac{2}{\pi\theta}\log\left(\frac{1 - \alpha\theta}{1 + \alpha\theta}\right) + 4i/\theta,$$

which is just the expression in dimensionless variables of the expression (22). Hence, σ^2 is a suitable measure of the size of the system.

The weak-coupling limit is obtained by setting $\alpha = 0$ in Eq. (31). This limit has been discussed extensively in I, where it was pointed out that contributions to $\rho_0(\mathcal{N}, t)$ appear which are not analytic in α, at $\alpha = 0$. However, these effects arise only if one retains in Eq. (31) first-order terms in α. When this is not done, Eq. (I.59a) gives as the weak-coupling solution in the infinite limit

$$\rho_0(\mathcal{N}, t) = e^{-4\tau}. \quad (32)$$

The subsequent calculations in this paper are thus accurate to the same order as is Eq. (32) for the large system. That is, all complications associated with nonanalytic contributions to $\rho_0(\mathcal{N}, t)$ are ignored. It should nonetheless be stressed that, although in an infinite system nonexponential behavior is associated only with the nonanalytic contributions, for the finite system being studied here, exponential decay can certainly not occur (the relaxation time does not even exist), but still no account is being taken of those terms in $S(\theta)$ which yield the deviation from the purely exponential solution, Eq. (32), in a large system. Nonanalyticity at $\alpha = 0$ will indeed appear in the calculations which follow, but it is due to a quite different cause, namely, the finite extent of the system. The oscillatory behavior which we shall find is thus in no way related to that discussed in I.

When $\alpha = 0$, only the last term in Eq. (31) remains. It is

$$\frac{2}{\theta}\sin(\sigma^2\theta)\cdot\csc\left(-\tfrac{1}{2}\sigma^2\theta+\frac{\sigma^2}{2\alpha}\right)\cdot\csc\left(+\tfrac{1}{2}\sigma^2\theta+\frac{\sigma^2}{2\alpha}\right)=\frac{-4\sin(\sigma^2\theta)}{\theta[\cos(\sigma^2/\alpha)-\cos(\sigma^2\theta)]}. \quad (33)$$

It is the appearance of $\cos(\sigma^2/\alpha)$ in Eq. (33) which brings in the nonanalyticity at $\alpha = 0$, referred to above. However, $\cos(\sigma^2/\alpha)$ is always some number ϵ say, lying between -1 and $+1$. A consistent approximation scheme can be developed therefore by treating α as sufficiently small, so that Eq. (33) is the only significant contribution to $S(\theta)$, and then simply treating ϵ as a further parameter of the problem. That this is a reasonable approach will be demonstrated by the numerical calculations discussed in Secs. VIII and IX. The use of the expression (33) for $S(\theta)$ leads to a great simplification of the constant term (27) in $\rho_0(\mathcal{N}, t)$:

$$\frac{1}{1+S(x=ia)}=\frac{1}{1+S(\theta=2i/\sigma)}=\left(1+\frac{2\sigma\sinh 2\sigma}{\cosh 2\sigma-\epsilon}\right)^{-1}. \quad (34)$$

When σ is large, ϵ may be neglected with respect to $\cosh(2\sigma)$, and Eq. (34) can be approximated by

$$\frac{1}{1+2\sigma\tanh 2\sigma}. \quad (35)$$

This expression yields the asymptotic behavior of the constant term as the system becomes large. It can be seen that it falls off fairly slowly as σ $[=(\alpha\eta)^{\frac{1}{2}}]$ tends to infinity. The zeros of $1 + S(\theta)$ may also be located by using Eq. (33) for $S(\theta)$. They are at the solutions of the transcendental equation

$$4\sin(\sigma^2\theta) = \theta[\cos(\sigma^2/\alpha) - \cos(\sigma^2\theta)], \quad (36)$$

excluding the trivial solution $\theta = 0$, at which (33) does not vanish. It can be seen at once that if θ_p is a solution to Eq. (36), so is $-\theta_p$. The solutions may be found by various approximate methods, of which the simplest is a graphical one where the two sides of Eq. (36) are plotted so that the points of intersection yield those values of θ which satisfy the equation. This is illustrated in Fig. 4 and is the basis of the computation of the solutions of Eq. (36) discussed in Sec. IX.

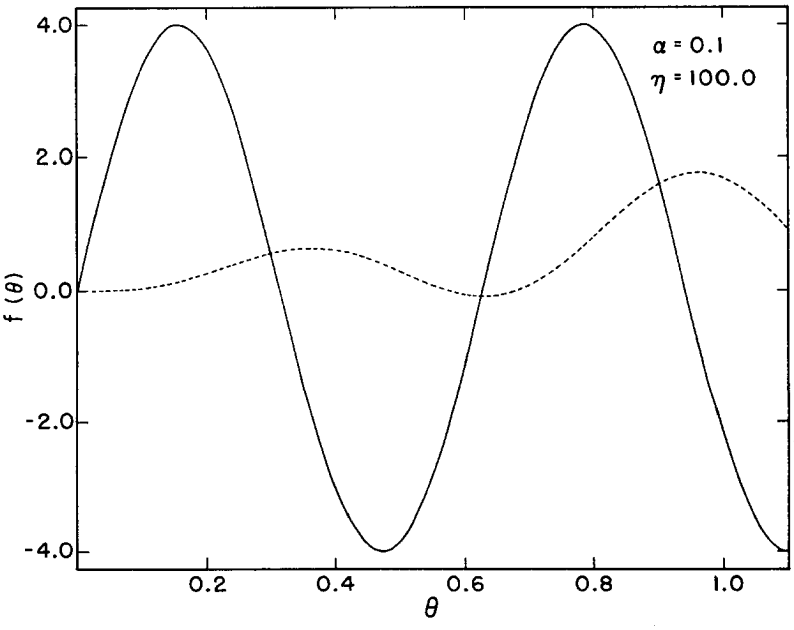


FIG. 4. A graphical illustration of the numerical method used to solve Eq. (36) for the pole locations. The dotted line represents $\theta[\cos(\sigma^2/\alpha) - \cos(\sigma^2\theta)]$ and the full line represents $4\sin(\sigma^2\theta)$.

Within the present approximation scheme, the solution (30) for $\rho_0(\mathcal{N}, t)$ is

$$\rho_0(\mathcal{N}, t) = -\frac{1}{2\pi i}\int_C d\theta \cdot \frac{\theta^2}{\theta^2 + (4/\sigma^2)} \times \exp\left[-i\tau(\theta^2 + 4/\sigma^2)^{\frac{1}{2}}\right] \cdot \left(\theta - \frac{4\sin(\sigma^2\theta)}{\epsilon - \cos(\sigma^2\theta)}\right)^{-1}. \tag{37}$$

If $\theta = \theta_p$ is a solution of Eq. (36), then the residue of the integrand in Eq. (37) is

$$\frac{\theta_p^2}{\theta_p^2 + (4/\sigma^2)} \cdot \exp\left[-i\tau(\theta_p^2 + 4/\sigma^2)^{\frac{1}{2}}\right] \times \left\{\frac{d}{d\theta}\left(\theta - \frac{4\sin(\sigma^2\theta)}{\epsilon - \cos(\sigma^2\theta)}\right)\right]_{\theta=\theta_p}\Bigg\}^{-1} = \frac{\theta_p^2 \cdot \exp\left[-i\tau(\theta_p^2 + 4/\sigma^2)^{\frac{1}{2}}\right]}{(\theta_p^2 + 4/\sigma^2) \cdot \left\{1 - \dfrac{4\sigma^2[\epsilon\cos(\sigma^2\theta_p) - 1]}{[\epsilon - \cos(\sigma^2\theta_p)]^2}\right\}}. \tag{38}$$

This expression can be expressed in terms of θ_p without trigonometric functions by using the fact that

$$\cos(\sigma^2\theta_p) = q,$$

say, satisfies the quadratic equation [from Eq. (36)]

$$(\epsilon - q)^2 = (16/\theta_p^2)(1 - q^2),$$

with the solution as

$$q = \frac{\epsilon \pm 4/\theta_p(1 + 16/\theta_p^2 - \epsilon^2)^{\frac{1}{2}}}{1 + (16/\theta_p^2)}, \tag{39}$$

where the minus sign is to be taken with the smallest value of θ_p and alternate values thereafter and the plus sign with the intermediate values. Putting Eq. (39) in Eq. (38) yields, for the residue at θ_p,

$$\frac{\theta_p^2}{\theta_p^2 + 4/\sigma^2} \cdot \exp\left[-i\tau(\theta_p^2 + 4/\sigma^2)^{\frac{1}{2}}\right] \cdot \left[1 \pm \frac{\sigma^2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}}\left(\theta_p^2 + \frac{16}{(1-\epsilon^2)}\right)^{\frac{1}{2}} - \frac{\sigma^2}{4}\left(\theta_p^2 + \frac{16}{(1-\epsilon^2)}\right)\right]^{-1}$$

with the above prescription for $\pm$. We may now write the solution (37) in the form

$$\rho_0(\mathcal{N}, t) = \left(1 + \frac{2\sigma\sinh 2\sigma}{\cosh 2\sigma - \epsilon}\right)^{-1} + \sum_{\theta_p > 0} \frac{2\theta_p^2}{\theta_p^2 + 4/\sigma^2} \cdot \cos\left[\tau(\theta_p^2 + 4/\sigma^2)^{\frac{1}{2}}\right] \times \left[1 \pm \frac{\sigma^2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}}\left(\theta_p^2 + \frac{16}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \frac{\sigma^2}{4}\left(\theta_p^2 + \frac{16}{(1-\epsilon^2)}\right)\right]^{-1}, \tag{40}$$

where the summation extends over the *positive* solutions of Eq. (36). Equation (40) is then the complete solution to the problem of spontaneous emission in the weak-coupling limit. It is the *complete* solution because, in this limit where $\alpha \to 0$, the range of validity of the expression (31) for $S(\theta)$, and hence of (33), extends over the whole range of θ. In the next section we shall verify that this solution, though complicated, goes over in the limit, as $\sigma \to \infty$, to the purely exponential solution, Eq. (32).

VI. RECOVERY OF THE EXPONENTIAL SOLUTION

An examination of Fig. 4 reveals that the locations of the zeros of $1 + S(\theta)$ are such that there is one zero in each range of length π/σ^2. In fact, the larger values of θ_p are well approximated by $n\pi/\sigma^2$ (n is a positive integer). This can be seen by noticing that as the fluctuation of $\theta[\epsilon - \cos(\sigma^2\theta)]$ becomes very large with θ, the range of θ within which the value of this expression lies between $+4$ and -4 becomes small and close to

$$\sigma^{-2}[(-1)^n\cos^{-1}\epsilon + n\pi].$$

The use of these values of θ_p should be a permissible approximation for large values of n, large values of σ^2, or in the special case that $\epsilon = \frac{1}{2}\pi$. Certainly, for the purpose of investigating the limit of the solution (40) when $\sigma \to \infty$, it is sufficient to use the values $n\pi/\sigma^2$ for the θ_p. If this is done, then the time-dependent term in Eq. (40) becomes

$$\sum_{n=1}^{\infty} \frac{2n^2\pi^2}{n^2\pi^2 + \sigma^2} \cdot \cos\left(\frac{2\tau}{\sigma^2}(n^2\pi^2 + \sigma^2)^{\frac{1}{2}}\right)\left[1 + \frac{2\sigma^2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}}\left(\frac{n^2\pi^2}{\sigma^4} + \frac{4}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \sigma^2\left(\frac{n^2\pi^2}{\sigma^4} + \frac{4}{(1-\epsilon^2)}\right)\right]^{-1} + \sum_{n=1}^{\infty} \frac{2(2n-1)^2\pi^2}{(2n-1)^2\pi^2 + 4\sigma^2} \cdot \cos\left(\frac{\tau}{\sigma^2}[(2n-1)^2\pi^2 + 4\sigma^2]^{\frac{1}{2}}\right) \times \left[1 - \frac{\sigma^2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}}\left(\frac{(2n-1)^2\pi^2}{\sigma^4} + \frac{16}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \frac{\sigma^2}{4}\left(\frac{(2n-1)^2\pi^2}{\sigma^4} + \frac{16}{(1-\epsilon^2)}\right)\right]^{-1}. \tag{41}$$

This form is of use for the evaluation of the limit of the time-dependent part of $\rho_0(\mathcal{N}, t)$ when $\sigma \to \infty$. But if one wishes to consider the case where σ is not too large, it is better to use the more exact expression

$$\sum_{n=0}^{\infty} \frac{2[(-1)^n \cos^{-1}\epsilon + n\pi]^2}{[(-1)^n \cos^{-1}\epsilon + n\pi]^2 + 4\sigma^2} \cos\left(\frac{\tau}{\sigma^2}\{[(-1)^n \cos^{-1}\epsilon + n\pi]^2 + 4\sigma^2\}^{\frac{1}{2}}\right)$$
$$\times \left[1 + (-1)^{n+1} \frac{\sigma^2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}}\left(\frac{[(-1)^n \cos^{-1}\epsilon + n\pi]^2}{\sigma^4} + \frac{16}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \frac{\sigma^2}{4}\left(\frac{[(-1)^n \cos^{-1}\epsilon + n\pi]^2}{\sigma^4} + \frac{16}{(1-\epsilon^2)}\right)\right]^{-1},$$

where

$$0 < \cos^{-1}\epsilon \leq \pi, \quad \cos^{-1} 0 = 2\pi.$$

When σ is small, clearly only the exact result, Eq. (40), can be used.

When σ becomes large, the two summations in the expression (41) appear as approximating sums to integrals. One can in the limit treat $n/\sigma^2 = k$, say, as a continuous variable, and one obtains that the limit of a summation is the integral over k of the limit, as $\sigma \to \infty$, of σ^2 times the summand. The first of the two sums becomes, in the limit,

$$2\int_0^{\infty} dk \cdot \cos(2\pi k\tau) \cdot \left[\frac{2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}} \cdot \left(\pi^2k^2 + \frac{4}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \left(\pi^2k^2 + \frac{4}{(1-\epsilon^2)}\right)\right]^{-1}$$

and the second,

$$2\int_0^{\infty} dk \cdot \cos(2\pi k\tau) \cdot \left[\frac{-2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}} \cdot \left(\pi^2k^2 + \frac{4}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \left(\pi^2k^2 + \frac{4}{(1-\epsilon^2)}\right)\right]^{-1}.$$

Since we know already that the constant contribution to $\rho_0(\mathcal{N}, t)$ vanishes as $\sigma \to \infty$, the final result for $\rho_0(\mathcal{N}, t)$ in the limit is

$$\rho_0(\mathcal{N}, t) = 4\int_0^{\infty} dk \frac{\cos(2\pi k\tau)[\pi^2k^2 + 4/(1-\epsilon^2)]}{-4\epsilon^2/(1-\epsilon^2)[\pi^2k^2 + 4/(1-\epsilon^2)] + [\pi^2k^2 + 4/(1-\epsilon^2)]^2}$$
$$= 4\int_0^{\infty} dk \frac{\cos(2\pi k\tau)}{\pi^2k^2 + 4} = 2\int_{-\infty}^{+\infty} dk \cdot \frac{e^{2\pi ik\tau}}{\pi^2k^2 + 4}. \tag{42}$$

We may notice that the extra parameter ϵ, which was introduced because of the nonanalyticity of $S(\theta)$ at $\alpha = 0$, has disappeared from Eq. (42), in accord with the statement that the weak-coupling limit contained no nonanalytic contributions in the infinite limit. Equation (42) is easily evaluated by closing the integral in the upper half-plane of k. This is permitted by Jordan's lemma, since $2\pi\tau > 0$. The value of the integral is calculated from the residue at $k = 2i/\pi$, and we get the result

$$\rho_0(\mathcal{N}, t) = e^{-4\tau}.$$

This result is precisely Eq. (32). We have thus proved that the solution, Eq. (40), does indeed yield the simple exponential solution when $\sigma \to \infty$.

VII. DISCUSSION OF NUMERICAL CALCULATIONS

The chief interest of the solution (40) lies in the information it yields about the behavior of systems where σ is not large. It is of importance to know the relative orders of magnitude of the first constant term in Eq. (40) and the second time-dependent term at the various stages of evolution of the system. In particular, the way in which the time dependence approaches or differs from the exponential form should be studied over a considerable range of the variables σ and τ, on which Eq. (40) depends. It will be useful also to see to what extent the extra parameter ϵ can influence the result. It is only through ϵ that the actual physical value of the coupling parameter α enters the weak-coupling treatment.

To this end, a numerical investigation of Eq. (40) has been undertaken. Throughout, α has been taken as small, in the sense mentioned in Sec. V. That is, $S(\theta)$ can be taken to be given by the expression (33), and its assigned value is then used to calculate ϵ. This is reasonable inasmuch as the contributions to $\rho_0(\mathcal{N}, t)$ which are thereby neglected [they come from the terms other than the last in Eq. (31)] are proportional to at least one power of α. On the other hand, the range of α for which this scheme is meaningful may be discovered by verifying that the neglected terms in Eq. (31) are, in fact, small compared with (33) within this range.

In the next section, the first term of Eq. (40) will be investigated numerically in detail as a function of

α and η. The quantity η is chosen as the variable rather than σ because it is possible to investigate this term, $\rho_c(\alpha, \eta)$, say, independently of considerations of time evolution, and hence it is irrelevant with regard to the dynamics whether or not one scales the measure of the size of the system. The suitability of η appears since it measures length in units of the wavelength of the radiation in resonance with the energy gap in the atom. On the other hand, σ^2 measures length in terms of the time, as measured by the scaled variable τ, which it takes for the radiation to cross the system. Thus, where there is no time dependence, σ^2 is not especially appropriate. A further advantage of using η is that it is possible to compute ρ_c, not only from its expression in Eq. (40), but also from the complete expression in the solution of the master equation without approximation on α. This quantity is [see (27) and (6)]

$$[1 + S(ia)]^{-1}, \tag{43}$$

where

$$S(ia) = \frac{4\alpha}{4\alpha + \eta} + 2\alpha\eta \sum_{n=1}^{\infty} [\pi^2 n^2 - \pi\eta n + \eta(\alpha + \tfrac{1}{4}\eta)]^{-1}. \tag{44}$$

The quantity (43) is calculated for a range of values of α and η, and then is compared with the two expressions (34) and (35). In this way, not only can we find how small α should be for the weak-coupling limit to be meaningful, as discussed above, but also see the size of the system beyond which the parameter ϵ becomes inimportant. Further, one may hope that these limits have meaning for the time-dependent part of Eq. (40), since the dynamics of the problem is contained entirely in the function S when it is considered over the whole complex plane.

In Sec. IX, an analysis is made of the second term in Eq. (40), ρ_T, say. For this calculation, it is necessary to use σ rather than η to measure the size of the system, since the arguments of the previous paragraph are no longer pertinent. Numerically, of course, both parameters will still be well defined. The calculation is performed both with the expression (41), which uses the assumption of even spacing of the roots of Eq. (36) at $\theta_p = n\pi/\sigma^2$, and also with the proper expression, given in Eq. (40), using numerically computed solutions of Eq. (36). We may expect that the expression (41) will afford a good estimate of ρ_T for sufficiently large σ and small τ, and in this way we shall see to what extent this is so.

VIII. ANALYSIS OF THE CONSTANT TERM

The first expression for ρ_c to be computed numerically in this study was (43) by using Eq. (44). It is readily seen that the summand in the second term of Eq. (44), when considered as a function of n, has a minimum at $n = \eta/2\pi$ and that, for values of n greater than this, the terms in the sum rapidly decrease in magnitude. For this reason, the number of terms included in the computation varied from a minimum of 100, for small η, up to a maximum of 100 000, for $\eta = 10\,000$. In all cases, it could be seen that the terms at the point of truncation were considerably smaller than the limits of accuracy which have been attained in the graphical displays of the results. These are the full curves in Figs. 5 and 6 for values of $\alpha = 0.1$ and 0.01, respectively, and for a range of η from 0 to 25.

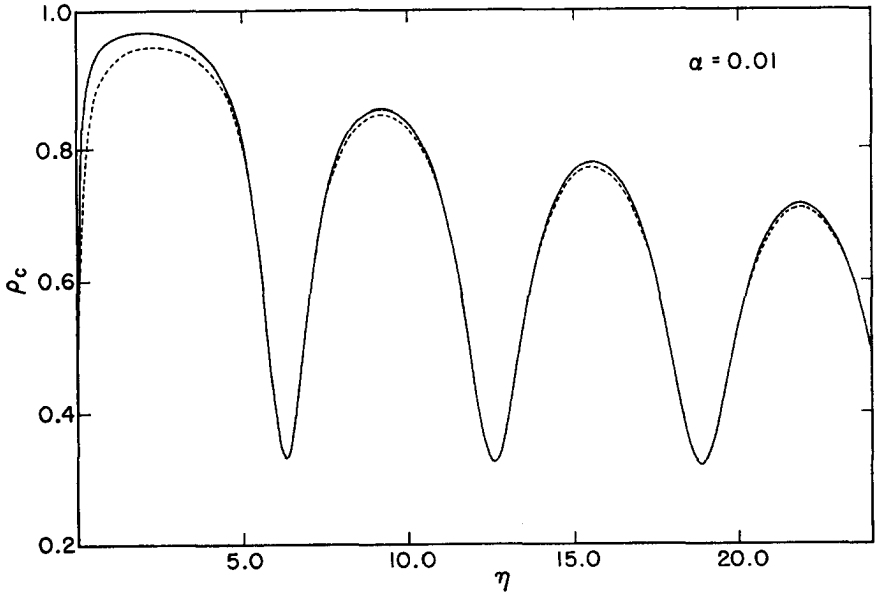


FIG. 5. A comparison of the dependence of ρ_c on η obtained by using Eq. (34) (dotted line) vs (43) with (44) (full line) for the case $\alpha = 0.01$.

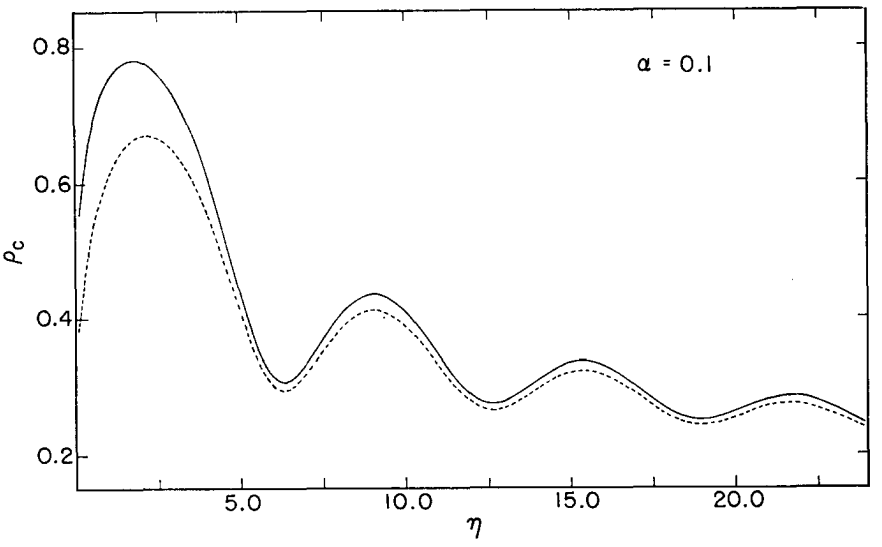


FIG. 6. A comparison of the dependence of ρ_c on η obtained by using Eq. (34) (dotted line) vs (43) with (44) (full line) for the case $\alpha = 0.1$.

The next step was to compute ρ_c using the two closed expressions (34) and (35), which are, respectively, the weak-coupling limit and its asymptotic form for large η of (43). Values of Eq. (34) are displayed. These are the dotted curves in Figs. 5 and 6 for the same values of α and range of η as before. It can be seen that the weak-coupling scheme is adequate for η greater than around 5. For values of η less than this, it is clear that terms other than that of lowest order in α become important in (43).

The weak-coupling approach fails in the régime η less than 5 not only in the sense that Eq. (44) should be used in the expression (43), but also in that for consistency higher terms in the perturbation solution of the master equation must be included. This point will be further discussed in Sec. X. In Table I, a comparison is made between the expressions (34), (35), and (43) with (44) for $\alpha = 0.1$. We see that the three expressions yield agreement for $\eta \geq 100$ roughly, and hence in this region the effect of the parameter ϵ has become negligible.

Certain general features emerge from these results. First, for any α, ρ_c tends to the value $\frac{1}{2}$, as $\eta \to 0$, when expression (44) is used in the calculation, but to the value $\frac{1}{3}$ when the expression (34) is used. These limits are apparent from the analytical form of (44) and (34). This phenomenon is discussed in Sec. X. Then it is seen that ρ_c, taken as a function of η for fixed α has a series of maxima, of which the first is found where η is approximately 2. The variation of ρ_c between these maxima is an indication of the effect of the nonanalyticity in α of the expressions (34) and (44) for $S(ia)$. This is the case since the nonanalyticity is brought in by the parameter ϵ, neglect of which yields the monotonic function (35). The third feature is that ρ_c tends fairly slowly to zero as $\eta \to \infty$, as predicted by (35), and that this expression is in fact a good estimate of ρ_c for values of η greater than around 100.

TABLE I. A comparison of the dependence of ρ_c on $\log_{10} \eta$, using Eqs. (34), (35), and (43) with (44), for $\alpha = 0.1$.

$\log_{10} \eta$	Eq. (34)	Eq. (35)	Eqs. (43), (44)
−4	0.3333	1.0000	0.5000
−3	0.3339	0.9996	0.5006
−2	0.3388	0.9960	0.5061
−1	0.3836	0.9620	0.5545
0	0.6094	0.7385	0.7528
+1	0.3881	0.3415	0.4093
+2	0.1361	0.1365	0.1386
+3	0.0476	0.0476	0.0479
+4	0.0156	0.0156	0.0156

IX. ANALYSIS OF THE TIME-DEPENDENT TERM

The computation of $\rho_T(\tau)$ was performed first of all by using the expression (41). The calculation was based on a value of 0.1, for α, and two values of σ^2 were examined, namely 1 and 10, corresponding to $\eta = 10$ and $\eta = 100$, respectively, for which values, as we have seen in the last section, the weak-coupling scheme is likely to be valid. The values of ϵ corresponding to these choices of η are, respectively, -0.8391 and $+0.8623$. The locations of the roots θ_p of Eq. (36) are estimated in the expression (41) to be the points $n\pi/\sigma^2$, for positive integral n. Thus, for $\sigma^2 = 1$, these roots are at the points $n\pi$, and for $\sigma^2 = 10$ at the points $\frac{1}{10}n\pi$. The series in expression (41) were truncated for $n = 1000$. An upper bound for the errors thus committed can be obtained by using the integral which approximates (41) for large

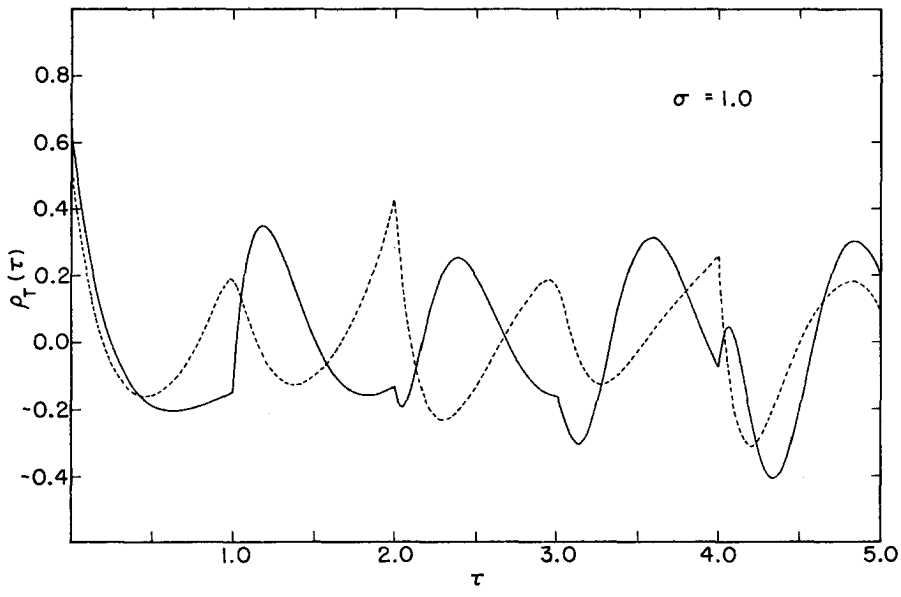


FIG. 7. A comparison of the dependence of $\rho_T(\tau)$ on τ using the approximately determined pole locations (dotted line) vs the exactly determined pole locations (full line), for $\sigma^2 = 1.0$.

σ. This integral [see Eq. (42)] is

$$4\int_0^\infty \frac{\cos(2\pi k\tau)\,dk}{\pi^2k^2+4}.$$

This integral is bounded above by its value at $\tau = 0$:

$$4\int_0^\infty \frac{dk}{\pi^2k^2+4}. \tag{45}$$

If the series in (41) are truncated at $n = N$, this corresponds to setting $k = N/\sigma^2$ in expression (45). The upper bound on the error involved in truncation is

$$4\int_{N/\sigma^2}^\infty \frac{dk}{\pi^2k^2+4} \sim \frac{4\sigma^2}{\pi^2 N}, \quad \text{for large } N. \tag{46}$$

In Figs. 7 and 8, the results of the calculation of (41) are displayed (the dotted curves) for $\sigma^2 = 1$ and $\sigma^2 = 10$, respectively. As a function of τ, ρ_T presents an extremely complicated behavior. At $\tau = 0$, it has a large value which rapidly decays to the neighborhood of zero as the arguments of the cosines move out of phase. Then, when τ reaches the vicinity of a multiple of σ^2, the phases line up somewhat and rapid fluctuations take place, the exact structure of which is very complicated and may well not be reproduced with much accuracy by the numerical calculation, even though the calculations were done in "double precision." The lining up of phases can never be more than approximate, since the periods of the cosines are not linearly related in any way.

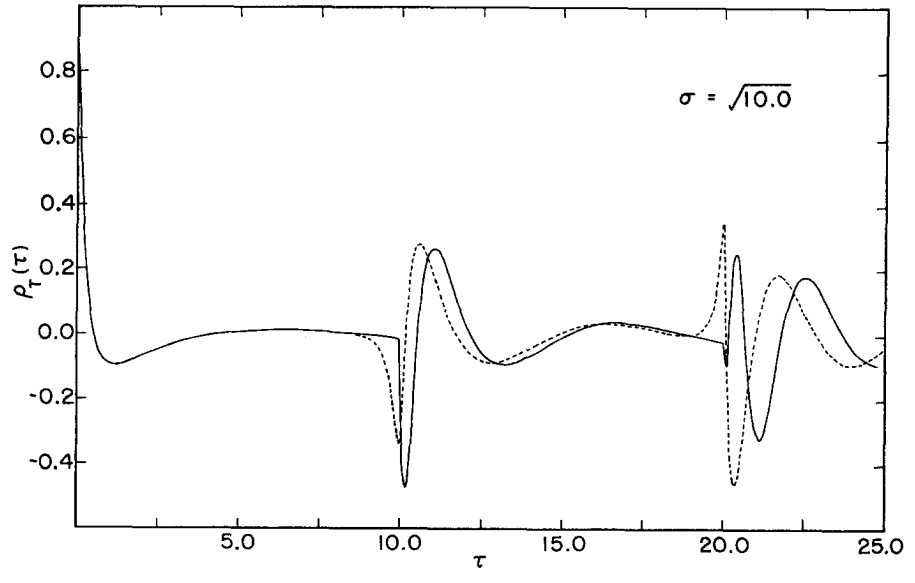


FIG. 8. A comparison of the dependence of $\rho_T(\tau)$ on τ using the approximately determined pole locations (dotted line) vs the exactly determined pole locations (full line), for $\sigma^2 = 10.0$.

It is possible to obtain numerical values for ρ_T based on the exact expression for it given in Eq. (40):

$$\rho_T(\tau) = \sum_{n=1}^{\infty} \frac{2\theta_n^2}{\theta_n^2 + 4/\sigma^2} \times \cos\left[\tau(\theta_n^2 + 4/\sigma^2)^{\frac{1}{2}}\right]\left[1 + (-1)^n \frac{\sigma^2\epsilon}{(1-\epsilon^2)^{\frac{1}{2}}} \times \left(\theta_n^2 + \frac{16}{(1-\epsilon^2)}\right)^{\frac{1}{2}} + \frac{\sigma^2}{4}\left(\theta_n^2 + \frac{16}{(1-\epsilon^2)}\right)\right]^{-1}, \quad (47)$$

where, again, the θ_n are the successive positive roots of Eq. (36). In Fig. 4, the graphical method of obtaining these roots was shown, and the numerical procedure used was analogous: The expressions $4 \sin(\sigma^2\theta)$ and $\theta[\epsilon - \cos(\sigma^2\theta)]$ were evaluated at each point of a very fine mesh, and the points at which these two expressions were most nearly equal were taken as the roots. In this way, the θ_n up to $n = 190$ and $n = 600$ were calculated for $\sigma^2 = 1$ and 10, respectively. These values were then used in Eq. (47) and the results are displayed in Figs. 7 and 8 by the full curves. It can be seen that, as we expected, agreement between these exact curves and those plotted using the expression (41) is better for $\sigma^2 = 10$ than for $\sigma^2 = 1$, and that,in both cases, for small τ (except very close to $\tau = 0$, for $\sigma^2 = 1$) agreement is satisfactory. Again, in the neighborhood of $\tau =$ an integral multiple of σ^2, rapid fluctuations occur, but their detailed structure bears little resemblance to that given by the expression (41). We may again emphasize that the numerical calculation is probably too delicate for the results to be certain in these sensitive regions.

It should be verified that at $\tau = 0$

$$\rho_c + \rho_T(0) = 1.$$

This is indeed the case for the two calculations described here, for which we have

$$\text{for } \sigma^2 = 1, \quad \rho_c = 0.3881, \quad \rho_T = 0.6076,$$

and

$$\text{for } \sigma^2 = 10, \quad \rho_c = 0.1362, \quad \rho_T = 0.8504.$$

The apparent slight discrepancy in the latter case is accounted for by the error in truncation of the series, Eq. (47), after 600 terms, since the error estimate (46), which is, of course, equally valid if N is sufficiently large for both (41) and (47), is 0.0133. It has been remarked that (46) is just the value of the contribution of the neglected terms in Eq. (47) at $\tau = 0$, and indeed the addition of 0.0133 to $\rho_T(0)$ yields the correct result. We may incidentally expect that the truncation error in ρ_T, for $\tau > 0$, should be considerably smaller than 0.0133, since the cosines in the neglected terms never are in phase after $\tau = 0$.

X. DISCUSSION AND CONCLUSIONS

In this paper a model has been studied in which it was possible to obtain an expression for the exact solution of the master equation for the problem of spontaneous emission in a finite system, within the framework of the weak-coupling approximation. It was possible to define dimensionless variables (α, σ, and τ) and to perform the calculation to all orders in both σ and τ but to zeroth order in α. The reason for this choice of variables should perhaps be stressed here: Because we have been interested in a system in which true irreversibility may manifest itself in an infinite system (even in the weak-coupling limit), it was necessary to *scale* the measures both of the time and of the size of the system so that the evolution of the diagonal element $\rho_0(\mathcal{N}, t)$ of the density matrix be given at *all* times and so that the time for information to cross the system be finite no matter what approximation be chosen for the coupling. This scaling made it possible to define the limit of weak coupling in a quite unambiguous fashion. It was further seen that the limit of a large system—usually called the thermodynamic limit—was obtained by letting *one* parameter only, namely σ, tend to infinity.

The result takes the form of a constant term plus a time-dependent one, expressed as the sum of residues at a series of poles along the real axis of a Laplace-transform variable. This series could be approximated in various ways within certain regimes, and in the last section, it was computed numerically in two ways, one based on the exact expression, the other on an approximate scheme. Although these calculations were too delicate to be quite certain around sensitive values of the time, the general picture is clear: After an initial decay to a value near zero, the series gives rise at fairly regular intervals to rapid and large fluctuations, the size of which never quite attains the initial value, but may nonetheless be large even after very long times. The constant term is also of great interest. It has already been remarked that the weak-coupling form (34) is all that may legitimately be taken from the calculations of this paper, since the terms of $S(\theta = 2i/\sigma^2)$ proportional to $\alpha, \alpha^2, \cdots$ must for consistency be taken only in conjunction with terms coming from higher orders in the perturbation series leading to the master equation. The constant term appears in place of the finite relaxation time which characterizes an infinite system. It arises

because, as has been seen, the relaxation time for a finite system is formally infinite, and it is its disappearance as the system becomes large which leads to the singularity of $S(z)$ at $z = 0$, the singularity that permits the definition of the relaxation time.

Along with the absence of time irreversibility in a finite system goes, of course, the possibility of Poincaré recurrences. The numerical estimation of the frequency of these is lengthy, and this study will be the subject of a forthcoming paper by the authors. In theory, however, their existence is apparent from the solution (40) of the master equation. It is nonetheless certain, as is seen from the possibility of retrieving purely exponential behavior in the large system limit, that the frequency of recurrence goes to zero as the system increases in size. It is of note that the calculations leading to Eq. (40) do indeed establish the existence of Poincaré recurrences in a model which is not exactly soluble and in which infinite system irreversibility cannot be shown to be the result of working in a representation whose quantum numbers are not those of the exact normal modes. This is in contrast to the work of Mazur and Montroll[4] in which the complete dynamics of an assembly of coupled harmonic oscillators can be given in simple mathematical form. The same is true of the detailed work of Rubin[5] in which the deviation from exponential behavior of the momentum autocorrelation function of a heavy particle in a finite crystal is discussed. The spirit of the calculations of this paper is nonetheless close to that of both of the studies mentioned, and the formal links are obvious.

In Sec. VIII, the constant term has been examined as a function of α and η. This particular study demonstrated various phenomena clearly. First, by maintaining α constant and varying η, one produces rapid changes in the nonanalyticity parameter ϵ, which manifest themselves by the large fluctuations of the constant term for small η. Then, for η less than 5, the growing importance of contributions proportional to α (relative to the weak-coupling expression) indicates the breakdown of the weak-coupling approximation. Another reason for doubting results for small values of η can be found in the time-dependent part of the solution. The range of validity of Eq. (31), from which Eq. (34) was derived, was

$$-1/\alpha < \mathrm{Re}\,\theta < 1/\alpha\,,$$

whereas the location θ_1 of the smallest root of Eq. (36) is about

$$\pi/\sigma^2 = \pi/\alpha\eta.$$

Thus, unless $\eta > \pi$, θ_1 lies outside the range of validity of the expression which permits its calculation. Thus, there are good reasons to suppose that when η is too small, the whole concept of weak coupling seems to be meaningless. Physically, the cause of the breakdown is likely to be that, when the emitted photon is constrained to remain in a length less than a few wavelengths, it is in effectively constant interaction with the 2-level system. Thus, the condition frequently stated[6] as necessary for any weak-coupling approach, namely, the existence of widely separated time scales for the interaction and for the relaxation (here for large change in the density matrix), is no longer satisfied. Not only does the interaction time scale become effectively infinite, but the time over which large variations of the density matrix take place becomes very short as η becomes small.

The constant term in the solution, Eq. (40), has a further significance. Since the time average of ρ_T over times considerably longer than that required for its initial decay is zero, the average of $\rho_0(\mathcal{N}, t)$ itself over such periods is just ρ_c. This quantity is thus a measure of the proportion of time which the 2-level system spends in the excited state relative to that spent in the ground state. It corresponds in this way to the observation that, whereas the spontaneous emission problem in an infinite system is essentially a problem at zero temperature, as soon as the system becomes finite in extent, there is a nonzero temperature which causes the 2-level atom to be in its excited state for a finite fraction of the time. It is even possible (this is the case when $\rho_c > \frac{1}{2}$) that the temperature be negative.

The next remark has to do with the behavior of the time-dependent part of the solution for small η. In the limit of a very small system, only one term of the series, Eq. (47), contributes, namely, the first, for which

$$\lim_{\sigma^2 \to 0} \sigma^2\theta_1^2$$

is finite. The weak-coupling result in this limit is that $\rho_c = \frac{1}{3}$, that $\rho_T(0) = \frac{2}{3}$, and that ρ_T oscillates between $\frac{2}{3}$ and $-\frac{2}{3}$, with a frequency which becomes infinite as $\sigma^2 \to 0$. This unphysical result is just another manifestation of the failure of Eq. (34) for too small systems. It is interesting to note, however, that if Eq. (44) is used to calculate ρ_c, its limiting value is $\frac{1}{2}$, and so in this case $\rho_0(\mathcal{N}, t)$ oscillates with infinite frequency between 1 and 0. This is no more than an expression of the uncertainty principle.

Another feature of the present investigation is that it shows how the well-known analytic behavior of the Laplace-transform integral for $\rho_0(\mathcal{N}, t)$ in the thermodynamic limit appears out of the much more complicated structure of the integral in the finite case.

Apart from the disappearance of the constant term, the frequently stated[7,8] remark that a series of poles on the real axis in the integral goes over to a Cauchy integral with singularities off the real axis is verified. The integral in expression (42) is indeed just the Cauchy integral limit of the sum (40), and it was crucial to the discovery of exponential behavior that indeed the integrand in (42) has a pole away from the real axis. The problem of the extent to which the thermodynamic limit, which approximates (physically) a finite system by an infinite one, can be taken as valid may be examined in the light of this analytic structure. Indeed, it can be seen from the numerical results in Secs. VIII and IX that a system of length up to 100 times the wavelength of the resonant radiation is not at all well described by the infinite-system solution. If one takes as at least one necessary condition for the success of this solution that the constant term ρ_c should be very small, then the expression (35) shows that, for $\alpha = 0.1$, η must be of order 10^4 before ρ_c falls to a value of 1% of $\rho_0(\mathcal{N}, 0)$. For weaker coupling, the system must be even larger.

It seems reasonable to state that the question raised by Zwanzig[1] as to the effect of small-scale "noise," of order $1/L$, has been answered in the context of the Wigner–Weisskopf atom. Because of the simplicity of the model, it has been possible to see how the *periodic* functions which are characteristic of the solution of a master equation for a finite system yield the exponential decay associated with an irreversible process in the thermodynamic limit.

ACKNOWLEDGMENT

The authors are indebted to Professor I. Prigogine for a number of interesting discussions.

APPENDIX

We have to prove that the quantity $S(\xi)$ of Eq. (28) is positive for all imaginary ξ and any positive η. Since $S(iy) = S(-iy)$ [see expression (17)], it is sufficient to prove that $S(i\gamma)$ is positive for positive γ. Putting $\xi = i\gamma$ into Eq. (28) yields

$$S(i\gamma,\eta) = -4\alpha\eta\left(\frac{1}{\eta^2+\gamma^2}+\frac{1}{(\eta+2\pi)^2+\gamma^2}\right)-\frac{2\alpha\eta}{\pi\gamma}\cdot\tan^{-1}\left(\frac{2\gamma(\eta+2\pi)}{(\eta+2\pi)^2-\gamma^2}\right)$$
$$-\frac{16\alpha\eta(\eta+2\pi)}{\pi}\int_0^\infty\frac{\xi\,d\xi}{(e^\xi-1)\{\xi^4+2\xi^2[(\eta+2\pi)^2-\gamma^2]+[(\eta+2\pi)^2+\gamma^2]^2\}}$$
$$+\frac{2\alpha\eta}{\gamma}\sinh\gamma\,(\cosh^2\tfrac{1}{2}\gamma-\cos^2\tfrac{1}{2}\eta)^{-1}, \tag{A1}$$

where the logarithm in Eq. (28) has been expressed by means of an inverse tangent function whose value is to lie between 0 and π, for $\gamma > 0$. All the terms in Eq. (A1) are clearly real for $\eta, \gamma > 0$, and it is evident that the first three terms, T_1, T_2, and T_3, say, are negative, and the last, T_4, is positive. We have to show, then, that

$$T_4 > -(T_1 + T_2 + T_3).$$

We may readily find an upper bound for $-T_3$. Since

$$\frac{\xi}{e^\xi-1} \le 1, \quad \text{for} \quad \xi \ge 0,$$

then

$$-T_3 \le \frac{16\alpha\eta(\eta+2\pi)}{\pi}\times\int_0^\infty\frac{d\xi}{\xi^4+2\xi^2[(\eta+2\pi)^2-\gamma^2]+[(\eta+2\pi)^2+\gamma^2]^2}$$
$$=\frac{4\alpha\eta}{(\eta+2\pi)^2+\gamma^2}.$$

So then,

$$-\frac{(T_1+T_2+T_3)}{2\alpha\eta} \le \frac{2}{\eta^2+\gamma^2}+\frac{4}{(\eta+2\pi)^2+\gamma^2}+\frac{1}{\pi\gamma}\tan^{-1}\left(\frac{2\gamma(\eta+2\pi)}{(\eta+2\pi)^2-\gamma^2}\right)$$
$$\le \frac{6}{\gamma^2}+\frac{1}{\gamma}, \quad \text{for all } \eta, \tag{A2}$$

whereas

$$\frac{T_4}{2\alpha\eta} \ge \frac{\sinh\gamma}{\gamma\cosh^2\frac{1}{2}\gamma} = \frac{\tanh\frac{1}{2}\gamma}{\frac{1}{2}\gamma}, \quad \text{for all } \eta. \tag{A3}$$

Hence, the result holds if

$$\tanh\tfrac{1}{2}\gamma > 3/\gamma + \tfrac{1}{2},$$

that is, if $\gamma > 6.1$, roughly. Similarly, for all $\gamma < 6.1$, it can be seen that the result holds if $\eta > \frac{1}{2}\pi$, roughly,

TABLE II. A comparison of Eq. (A2) and the term $T_4/2\alpha\eta$, for $\eta = 0.0$, 0.7, and $\frac{1}{2}\pi$, and for a range of values of γ.

γ	0.0	0.5	1.0	2.0	4.0	6.1
$\eta = 0.0$						
Value of Eq. (A2)		8.202	2.200	0.690	0.287	0.186
$\frac{T_4}{2\alpha\eta}$		16.332	4.328	1.313	0.519	0.329
$\eta = 0.7$						
Value of Eq. (A2)	4.254	2.875	1.889	0.616	0.266	0.175
$\frac{T_4}{2\alpha\eta}$	8.475	5.725	3.013	1.210	0.514	0.329
$\eta = \frac{1}{2}\pi$						
Value of Eq. (A2)	0.957	0.882	0.722	0.449	0.235	0.160
$\frac{T_4}{2\alpha\eta}$	2.000	1.848	1.522	0.964	0.500	0.328

for Eq. (A3), being a lower bound for $T_4/2\alpha\eta$, is never less than 0.322 in this range of γ, while the expression (A2) is equal to 0.160 for $\gamma = 6.1$, $\eta = \frac{1}{2}\pi$. Since $-T_1$, $-T_2$, and $-T_3$ are all decreasing functions of η, values of η greater than $\frac{1}{2}\pi$ will give a positive value of Eq. (A1), for $\gamma < 6.1$.

It is simplest in the range $0 < \gamma < 6.1$, $0 < \eta < \frac{1}{2}\pi$ to check numerically that the expression (A2) is smaller than $T_4/2\alpha\eta$, rather than to attempt a more sophisticated analysis. Sample results are given in Table II, from which the result is evident. Further, it is clear that Eq. (A1) is always positive if either γ or η is zero.

[1] R. W. Zwanzig, *Lectures in Theoretical Physics* (*Boulder*) (Interscience Publishers, Inc., New York, 1961), Vol. III.
[2] R. Davidson and J. J. Kozak, J. Math. Phys. **11**, 189 (1970).
[3] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, New York, 1962); *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Company, New York, 1953), Vol. 1.
[4] P. Mazur and E. Montroll, J. Math. Phys. **1**, 70 (1960).
[5] R. J. Rubin, J. Am. Chem. Soc. **90**, 3061 (1968).
[6] I. Prigogine, *Nobel Symposium 5: Fast Reactions and Primary Processes in Chemical Kinetics*, Stig Claesson, Ed. (Interscience Publishers, Inc., New York, 1967).
[7] I. Prigogine, *Nonequilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).
[8] P. Résibois, "A Perturbative Approach to Irreversible Statistical Mechanics," *Physics of Many-Particle Systems* (Gordon and Breach, Sci. Publishers, Inc., New York, 1966), Vol. I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Oblique Reflection of an Electromagnetic Plane Wave from a Striated Impedance Surface

JAMES R. WAIT
Cooperative Institute for Research in Environmental Sciences
University of Colorado, Boulder, Colorado 80302

(Received 14 October 1969)

A method is described for calculating the reflection of a plane wave of arbitrary incidence from a flat surface which has a spatially modulated surface impedance in one particular direction. Choosing a sinusoidal form for the modulation, the reflected wave spectrum is calculated. It is shown that, in general, for any arbitrary incidence, an infinite number of reflected modes are excited. Also, we find that the mode of order n is coupled to modes of order $n \pm 1$.

1. INTRODUCTION

Many interesting physical phenomena are associated with reflection and guiding of waves over surfaces which have a periodic structure. Many years ago, Lord Rayleigh[1] propounded a theory for the scattering of sound from a periodic surface with a sinusoidal profile. Since then, many related investigations have been pursued and an excellent and comprehensive review is found in the book by Beckmann and Spizzichino.[2] Many of the essential features of scattering from periodic nonuniform surfaces can be found when dealing with a flat surface that has a spatially varying impedance. In the present paper, we use such a model and focus attention on the case of a plane wave at arbitrary incidence with respect to the orientation of the periodic striations of the surface impedance. The emphasis in the present paper is on the technique used to solve the problem rather than the physical applications. A number of extensions of the theory are also described briefly.

2. FORMULATION

We consider a flat imperfectly conducting surface of infinite extent. The surface impedance characterizing the boundary is assumed to be periodic in the x-coordinate direction, while it is constant in the y-coordinate direction. The region $z > 0$ above the surface is free space. The problem posed is to calculate the reflection of a plane wave incident at an arbitrary angle. However, to simplify the discussion initially, we take the y component of the incident electric vector to be zero.

The boundary conditions on the tangential components of the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ are

$$E_x = -\eta_0 \Delta(x) H_y,$$
$$E_y = \eta_0 \Delta_1 H_x, \tag{1}$$

at $z = 0$, where $\eta_0 = (\mu_0/\epsilon_0)^{\frac{1}{2}}$ is the characteristic impedance of free space ($\simeq 120\pi\Omega$ in MKS units). Thus, $\Delta(x)$ is an x-dependent dimensionless surface impedance, while Δ_1 is a constant dimensionless surface impedance. For present purposes, we choose

$$\Delta(x) = \Delta_0(1 + M \cos ax), \tag{2}$$

where Δ_0 is constant, M is a modulation index, and the period of the modulation is $2\pi/a$.

The y component of the incident magnetic field $\mathbf{H}^p$ is written as

$$H_y^p = H_0 e^{-i\beta x} e^{-i\gamma y} e^{+ikC_0 z}, \tag{3}$$

for an implied time factor $e^{+i\omega t}$. Here, $\beta = kS_0 \cos\phi$, $\gamma = kS_0 \sin\phi$, $S_0 = \sin\theta$, $C_0 = \cos\theta$, and $k = (\epsilon_0\mu_0)^{\frac{1}{2}}\omega$ is the free-space wavenumber. We readily confirm that $(\nabla^2 + k^2)H_y^p = 0$. Also, we see that θ is the angle which the incident wave vector subtends with the z axis, while ϕ is the angle subtended by the plane of incidence and the x axis.

3. THE HERTZ POTENTIALS

In terms of electric and magnetic Hertz vectors Π and Π^*, the electromagnetic fields in the region $z > 0$ are quite generally obtained from

$$\mathbf{E} = (k^2 + \text{grad div})\,\mathbf{\Pi} - i(\mu_0\omega/\eta_0)\,\text{curl}\,\mathbf{\Pi}^*, \tag{4}$$
$$\eta_0\mathbf{H} = (k^2 + \text{grad div})\,\mathbf{\Pi}^* + i\epsilon_0\omega\eta_0\,\text{curl}\,\mathbf{\Pi}. \tag{5}$$

In order to match boundary conditions in the present problem, it is only necessary that $\mathbf{\Pi}$ and $\mathbf{\Pi}^*$ have y components which we designate Π and Π^* (omitting the subscript y). Thus, we can write

$$E_y = \left(k^2 + \frac{\partial^2}{\partial y^2}\right)\Pi, \tag{6}$$

$$\eta_0 H_y = \left(k^2 + \frac{\partial^2}{\partial y^2}\right)\Pi^*, \tag{7}$$

$$E_x = \frac{\partial^2\Pi}{\partial x \partial y} + ik\frac{\partial \Pi^*}{\partial z}, \tag{8}$$

$$\eta_0 H_x = \frac{\partial^2\Pi^*}{\partial x \partial y} - ik\frac{\partial \Pi}{\partial z}, \tag{9}$$

in terms of the "Hertz potentials" Π and Π^*.

First of all, we observe that (3) can be obtained from a Hertz potential Π_p^* given by

$$\Pi_p^* = Ke^{-i\beta x}e^{-i\gamma y}e^{ikC_0 z}, \tag{10}$$

where $K = (k^2 - \gamma^2)^{-1}H_0$. We now write suitable expressions for the total "Hertz potentials," keeping in mind that the reflecting surface is periodic in the x direction and constant in the y direction. Thus, on using a certain amount of hindsight,

$$\Pi = Ke^{-i\gamma y}e^{-i\beta x}\sum_{m=-\infty}^{+\infty} B_m e^{-imax}e^{-ikC_m z} \tag{11}$$

and

$$\Pi^* = Ke^{-i\gamma y}e^{-i\beta x}\left(e^{ikC_0 z} + \sum_{m=-\infty}^{+\infty} A_m e^{-imax}e^{-ikC_m z}\right), \tag{12}$$

where the summations are to include all integers, B_m and A_m are coefficients, and

$$\begin{aligned} kC_m &= [k^2 - (\beta + ma)^2 - \gamma^2]^{\frac{1}{2}} \\ &= -i[(\beta + ma)^2 + \gamma^2 - k^2]^{\frac{1}{2}}. \end{aligned} \tag{13}$$

We readily confirm that the right-hand sides of (11) and (12) satisfy the wave equation. Also, each expression varies with y in accordance with $e^{-i\gamma y}$ as specified by the incident wave. In addition, we see that appropriate radiation conditions are satisfied as $z \to \infty$, whether C_m is real or imaginary. The latter occurs when $(\beta + ma)^2 > \gamma^2 - k^2$, in which case the reflected modes are evanescent in the positive z direction.

4. APPLICATION OF BOUNDARY CONDITIONS

Using Eqs. (6)–(9) for the tangential fields along with (11) and (12) for the potentials, we find that the boundary conditions specified by (1) now lead to the system

$$\begin{aligned} &\sum_{m=-\infty}^{+\infty}[-\gamma(\beta + ma)B_m + k^2C_mA_m]e^{-imax} - k^2C_0 \\ &= -\Delta_0(k^2 - \gamma^2) \\ &\quad \times \sum_{m=-\infty}^{+\infty} A_m[e^{-imax} + \tfrac{1}{2}Me^{-i(m-1)ax} + \tfrac{1}{2}Me^{-i(m+1)ax}] \\ &\quad - \Delta_0(k^2 - \gamma^2)(1 + \tfrac{1}{2}Me^{iax} + \tfrac{1}{2}Me^{-iax}) \end{aligned} \tag{14}$$

and

$$\begin{aligned} &(k^2 - \gamma^2)\sum_{m=-\infty}^{+\infty} B_m e^{-imax} \\ &= -\Delta_1 \sum_{m=-\infty}^{+\infty}[\gamma(\beta + ma)A_m + k^2C_mB_m]e^{-imax} - \Delta_1\gamma\beta. \end{aligned} \tag{15}$$

We now multiply both sides of (14) and (15) by e^{inax} and integrate, with respect to x, from 0 to $2\pi/a$. At the same time, we note that

$$\begin{aligned} \frac{a}{2\pi}\int_0^{2\pi/a} e^{i(n-m)ax}\,dx &= 0, \quad \text{if} \quad n \neq m, \\ &= 1, \quad \text{if} \quad n = m, \end{aligned} \tag{16}$$

where n can be any integer including zero. Then, from (15) alone, we discover that

$$B_n = -\Delta_1\gamma[(\beta + na)A_n + \beta\delta_n]/(k^2 - \gamma^2 + \Delta_1 k^2 C_n), \tag{17}$$

where $\delta_n = 1$, if $n = 0$, and $\delta_n = 0$, if $n \neq 0$. This tells us immediately that the electric Hertz potentials vanish if either or both $\Delta_1 = 0$ and $\gamma = 0$. Also, we see, in general, that only electric modes of order n are coupled to magnetic modes of order n.

On applying the orthogonality condition (16) to (14), we obtain

$$\begin{aligned} &-\gamma(\beta + na)B_n + k^2C_nA_n - k^2C_0\delta_n \\ &\quad = -\Delta_0(k^2 - \gamma^2)(A_n + \tfrac{1}{2}A_{n+1}M + \tfrac{1}{2}A_{n-1}M) \\ &\qquad - \Delta_0(k^2 - \gamma^2)(\delta_n + \tfrac{1}{2}\delta_{n+1}M + \tfrac{1}{2}\delta_{n-1}M). \end{aligned} \tag{18}$$

We eliminate B_n from (18) by using (17) to yield

$$A_n\Gamma_n + A_{n+1} + A_{n-1} = \Gamma_-\delta_n - \delta_{n+1} - \delta_{n-1}, \tag{19}$$

where

$$\begin{aligned} \Gamma_n &= \frac{2}{M\Delta_0(k^2 - \gamma^2)} \\ &\quad \times \left(k^2C_n + \Delta_0(k^2 - \gamma^2) + \frac{\gamma^2(\beta + na)^2\Delta_1}{k^2 - \gamma^2 + \Delta_1 k^2 C_n}\right) \end{aligned} \tag{20}$$

and

$$\begin{aligned} \Gamma_- &= \frac{2}{M\Delta_0(k^2 - \gamma^2)} \\ &\quad \times \left(k^2C_0 - \Delta_0(k^2 - \gamma^2) - \frac{\gamma^2(\beta + na)\beta\Delta_1}{k^2 - \gamma^2 + \Delta_1 k^2 C_n}\right). \end{aligned} \tag{21}$$

This is an infinite set of coupled equations to solve for A_n. Two special cases of (19) are worth noting. In the relatively simple case of normal incidence (i.e., $\gamma = 0$ and $S_0 = 0$), we recover the results given by Lysanov.[3] Then, if we choose the plane of incidence to be in the (x, y) plane (i.e., $\gamma = 0$), we obtain results consistent with those given by Oliner and Hessel,[4] provided that the source plane wave is removed.

5. SOLVING THE INFINITE SYSTEM

To solve the infinite system (19), we first note that, for $n < -1$, we can write

$$\frac{A_n}{A_{n+1}} = -\left(\Gamma_n + \frac{A_{n-1}}{A_n}\right)^{-1}, \tag{22}$$

$$\frac{A_{n-1}}{A_n} = -\left(\Gamma_{n-1} + \frac{A_{n-2}}{A_{n-1}}\right)^{-1}, \tag{23}$$

$$\frac{A_{n-2}}{A_{n-1}} = -\left(\Gamma_{n-2} + \frac{A_{n-3}}{A_{n-2}}\right)^{-1}, \tag{24}$$

and so on. On inserting (23) into (22), (24) into (23) and continuing the process indefinitely, we obtain the continued-fraction expansion

$$\frac{A_n}{A_{n+1}} = -\cfrac{1}{\Gamma_n - \cfrac{1}{\Gamma_{n-1} - \cfrac{1}{\Gamma_{n-2} - \cfrac{1}{\Gamma_{n-3} - \cdots}}}}. \tag{25}$$

For finite values of the modulation index, the infinite fraction may be truncated at a negative integer $-N$ which is sufficiently large to achieve the desired accuracy. In particular, we can use (25) for $n = 2$ to calculate the ratio A_{-2}/A_{-1}.

Working with (19) for $n > +1$, we find, in a similar fashion, that

$$\frac{A_{n+1}}{A_n} = -\cfrac{1}{\Gamma_{n+1} - \cfrac{1}{\Gamma_{n+2} - \cfrac{1}{\Gamma_{n+3} - \cdots}}}, \tag{26}$$

which, in particular, can be applied to calculate A_2/A_1 to any desired precision.

We now return to (19) and explicitly display the three equations for $n = 0, +1$, and -1 in the form

$$A_0\Gamma_0 + A_1 + A_{-1} = \Gamma_{-1}, \tag{27}$$

$$A_1(\Gamma_1 + \chi_+) + A_0 = -1, \tag{28}$$

$$A_{-1}(\Gamma_{-1} + \chi_-) + A_0 = -1, \tag{29}$$

where $\chi_- = A_{-2}/A_{-1}$ and $\chi_+ = A_2/A_1$. Simple algebra now tells us that

$$A_0 = \left[\Gamma_- + \frac{1}{\Gamma_1 + \chi_+}\left(1 + \frac{\Gamma_1 + \chi_+}{\Gamma_{-1} + \chi_-}\right)\right] \Big/ \left[\Gamma_0 - \frac{1}{\Gamma_1 + \chi_+}\left(1 + \frac{\Gamma_1 + \chi_+}{\Gamma_{-1} + \chi_-}\right)\right], \tag{30}$$

$$A_1 = -(1 + A_0)/(\Gamma_1 + \chi_+), \tag{31}$$

and

$$A_{-1} = -(1 + A_0)/(\Gamma_{-1} + \chi_-). \tag{32}$$

This actually constitutes the final solution of the problem since the coefficients B_n and A_n are now determined in terms of the specified parameters.

6. SOME SPECIAL CASES

As a partial check on the results, we can let M approach zero,whence (30) reduces to

$$A_0 = \frac{k^2C_0 - \Delta_0(k^2 - \gamma^2) - \gamma^2\beta^2\Delta_1(k^2 - \gamma^2 + \Delta_1k^2C_0)^{-1}}{k^2C_0 + \Delta_0(k^2 - \gamma^2) + \gamma^2\beta^2\Delta_1(k^2 - \gamma^2 + \Delta_1k^2C_0)^{-1}} \tag{33}$$

and $A_n = 0$, if $n \neq 0$. Also, for this special case, we see that (17) gives

$$B_0 = -\Delta_1\gamma\beta(1 + A_0)/(k^2 - \gamma^2 + \Delta_1k^2C_0) \tag{34}$$

and $B_n = 0$, if $n \neq 0$.

If in addition $\gamma = 0$, these simplify even further to

$$A_0 = (C_0 - \Delta_0)/(C_0 + \Delta_0) \quad \text{and} \quad B_0 = 0, \tag{35}$$

which is the expected Fresnel reflection coefficient.

Another important special case is when the magnetic field vector of the incident plane wave is parallel to the y direction. Then, $\gamma = 0$ and the plane of incidence is perpendicular to the y axis. Then, from (17), $B_n = 0$ as noted previously. The coefficients A_n are still to be found from the coupled set (19), but now

$$\Gamma_n = 2(M\Delta_0)^{-1}(C_n + \Delta_0) \tag{36}$$

and

$$\Gamma_- = 2(M\Delta_0)^{-1}[C_0 - \Delta_0]. \tag{37}$$

Also, as indicated before, this solution is equivalent to the one derived by Oliner and Hessel[4] for the sourceless case (i.e., $A_0 = 0$).

7. SOME EXTENSIONS AND FINAL REMARKS

The oblique-incidence solution presented here is not quite perfectly general. For simplicity, we set E_y in the *incident* plane wave equal to zero. For an incident plane wave of arbitrary polarization, we should allow for the existence of both magnetic and electric Hertz potentials in the incident wave. Thus, in place of (11), we should use

$$\Pi = Ke^{-i\gamma v}e^{-i\beta x} \times \left(qe^{ikC_0z} + \sum_{m=-\infty}^{+\infty} B_m e^{-imax}e^{-ikC_mz}\right), \tag{38}$$

where q is a constant. The form for Π^* given by (12) is left unchanged. The reader would now find that (17)

is to be replaced by

$$B_n = -\{\Delta_1\gamma\beta(\beta + na)A_n - \delta_n[\Delta_1 k^2 C_0 q - (k^2 - \gamma^2)q - \gamma\beta\Delta_1]\} \times [k^2 - \gamma^2 + \Delta_1 k^2 C_n]^{-1}. \quad (39)$$

Then, on completing the derivation, it is found that the coefficients A_n are still given by the coupled infinite set (19), but now (21) is replaced by

$$\Gamma_- = 2[M(k^2 - \gamma^2)\Delta_0]^{-1}\{k^2 C_0 + \gamma\beta q - \Delta_0(k^2 - \gamma^2) - \gamma(\beta + na)[(k^2 - \gamma^2)q + \Delta_1(\gamma\beta - k^2 q C_0)] \times (k^2 - \gamma^2 + \Delta_1 k^2 C_n)^{-1}\}, \quad (40)$$

while the equation for Γ_m remains the same as (20).

The specified polarization of the incident plane wave determines uniquely the (complex) constant q. Thus, the solution is perfectly general.

Another extension which is straightforward is to allow the surface impedance to have arbitrary periodic variation in the x direction. For example, (2) could be replaced by

$$\Delta(x) = \Delta_0\left(1 + \sum_p M_p e^{ipx}\right), \quad (41)$$

where M_p is the Fourier coefficient for the spatial modulation of the impedance variation. For example, the summation in (41) could extend from $-P$, $-P + 1$, $-P + 2, \cdots, P - 2$, $P - 1$, P, where P is finite. Then the derivation proceeds as before, but the coupled set given by (19) is now of the form

$$A_{n-P} + A_{n-P+1}\Gamma_{n-P+1} + \cdots + A_n\Gamma_n + \cdots + A_{n+P-1}\Gamma_{n+P-1} + A_{n+P} = -\delta_{N-P} \cdots + \Gamma_-\delta_n + \cdots - \delta_{N+P}, \quad (42)$$

where the Γ's are complicated algebraic expressions which reduce to the forms given by (20) and (21) when $P = 1$ and $M_{+1} = M_{-1} = \frac{1}{2}M$.

Finally, we should like to mention that the present method can be applied to nonflat periodic surfaces provided that the boundary conditions at the actual (rippled) surface are transferred to a reference plane by an expansion in terms of powers of (height of surface)/(wavelength). Clearly, this approach is only useful if the expansion parameter is reasonably small compared with unity.

[1] Lord Rayleigh, *The Theory of Sound* (Dover Publications, Inc., New York, 1945), 2nd ed., Vol. 2, Sec. 272a.

[2] P. Beckmann and A. Spizzichino, *Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon Press, Ltd., London, 1963), Chap. 2.

[3] Iu. P. Lysanov, Akust. Zh. **4**, 47 (1958) [Sov. Phys.—Acoust. **4**, 45 (1958)].

[4] A. A. Oliner and A. Hessel, IRE Trans. Antennas Propagation, Suppl. 7, 5201 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Physical-Region Discontinuity Equations for Many-Particle Scattering Amplitudes. II*

JOSEPH COSTER AND HENRY P. STAPP
Lawrence Radiation Laboratory, University of California, Berkeley, California 94550

(Received 4 April 1968)

Discontinuity equations are derived for physical-region normal thresholds in all direct and crossed channels. The discontinuity is given as a unitarity-type integral with an integrand that contains, as factors, the two physical scattering functions corresponding to the two vertices of the Landau diagram associated with the normal threshold. There is also a third factor, which for the case of a leading normal threshold with any number of particles is the Hermitian conjugate of the elastic scattering matrix associated with the set of internal lines of the Landau diagram. For nonleading normal thresholds below the lowest 4-particle threshold, the extra factor is defined by an integral equation that resembles unitarity, but has a restricted set of intermediate particles.

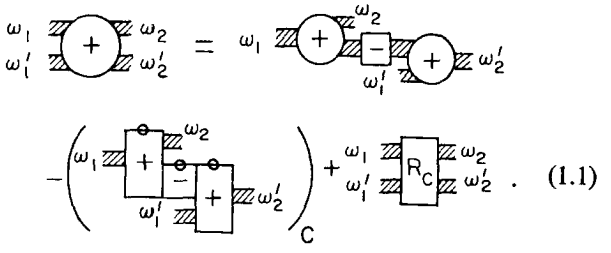
1. INTRODUCTION

This paper is the second in a series[1] devoted to calculating discontinuities around physical-region singularities of multiparticle scattering functions. The aim here is to obtain discontinuity formulas for normal thresholds. These normal-threshold formulas are important both in their own right and as the basic ingredients of the discontinuity formulas for more complicated singularities.

The main content in this paper consists in the derivation of some physical-region identities. These identities express, typically, any physical-region scattering function as a sum of terms, each consisting of a unitarity-type integral over a product of physical scattering functions or their conjugates. Each term is conveniently represented by a *bubble diagram*, in which plus and minus bubbles represent the connected parts of scattering amplitudes and their conjugates, respectively, and the lines connecting these bubbles represent physical particles. The identities are derived from unitarity and cluster properties alone, no analyticity property being invoked. Like unitarity, these identities hold at all real values of the external momentum vectors. Their importance lies in the fact that they explicitly display the discontinuity around normal thresholds.

The result that the discontinuity is explicitly displayed by certain terms in these identities follows from certain topological properties of the diagrams that represent the other terms, together with some structure theorems derived earlier[2] that specify the analytic structure of bubble-diagram functions. These structure theorems say that the bubble-diagram function M^B corresponding to the bubble diagram B has the following properties: M^B is singular only at points lying on Landau surfaces and only on those Landau surfaces $L(D)$ that correspond to Landau diagrams $D \supset\subset B$. A Landau diagram $D \supset\subset B$ is a diagram that is a contraction of a diagram D' constructed by replacing the bubbles of B by connected diagrams. Moreover, the signs of the Landau α's of the lines of this D' are restricted by the condition that the (internal) lines of diagrams replacing plus or minus bubbles must be positive or negative, respectively. If only one $D \supset\subset B$ gives an $L(D)$ passing through a point P lying inside the physical region, then the function M^B can be continued around $L(D)$ near P by passing into a well-defined "upper half-plane," which can be defined geometrically in terms of the diagram that generates P. If several surfaces pass through P, then the continuation can be made through the intersection of the various upper half-planes, provided that this intersection is nonempty. The relation between the diagram $D \supset\subset B$ and the corresponding half-plane is such that, if two diagrams are identical except for a single over-all reversal of the signs of all the α's, then the two corresponding half-planes are opposite half-planes. Thus, if a point P lies on the $L(D)$ of two such $D \supset\subset B$, then no continuation is possible, in general. The hypothesis of the structure theorems is the analyticity property of the physical-region scattering functions obtained from S-matrix macroscopic causality conditions, as is discussed in Sec. 8.

Our key identity reads, in box notation,[1]

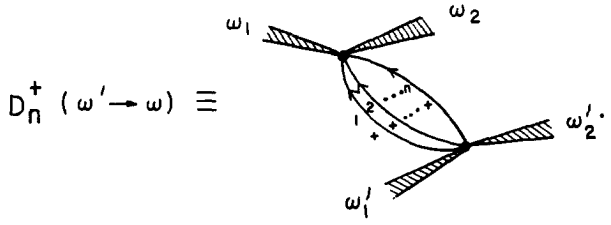


(1.1)

The shaded strips represent sets of any number of lines. A plus box represents the scattering matrix S;

a plus circle, its connected part. A minus box represents $S^\dagger$; a minus circle its connected part. A plus or minus box with a little circle on it represents S or $S^\dagger$ minus its connected part. The subscript c denotes connected part. Finally, the box R_c denotes a well-defined set of bubble diagrams, each of which represents a well-defined integral over a product of physical-region scattering amplitudes or their conjugates.

The importance of (1.1) arises from the following property of R_c: No bubble diagram B in the sum represented by the box R_c supports any Landau diagram D' that contracts to any diagram of the form $D_n^+(\omega' \to \omega)$, where

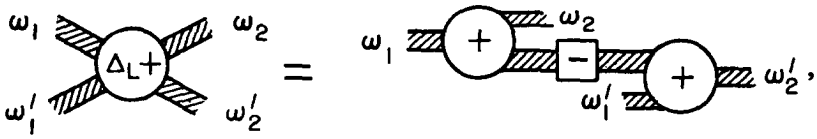


(1.2)

The plus signs on the lines indicate that the corresponding Landau α's are positive. The similar diagram with minus signs on all lines is denoted by $D_n^-(\omega' \to \omega)$. A diagram of the form $D_n^+(\omega' \to \omega)$ or $D_n^-(\omega' \to \omega)$, for any positive integer n, is called a $\omega' \to \omega$ normal threshold diagram and will be denoted by $D(\omega' \to \omega)$.

Because the box R_c has the property just described, we know from the quoted structure theorems the following fact: The function R_c represented by the box R_c continues into itself via a "minus-$i\epsilon$ continuation" past any singularity corresponding to any pure-(positive or negative)-α Landau diagram D that contracts to any $D(\omega' \to \omega)$. (The minus-$i\epsilon$ continuation is the one opposite to the physical continuation. That R_c must continue in this way follows from the fact that these pure-α diagrams must be pure-minus-α diagrams.) Using this fact, one immediately deduces from (1.1) the discontinuity formula for the leading $\omega' \to \omega$ normal threshold, near points lying "inside" the physical region. [The "leading" $\omega' \to \omega$ normal threshold is the multiparticle ($n \geq 2$) $\omega' \to \omega$ normal threshold with the smallest value M_L of the $\omega' \to \omega$ exchange c.m. energy $E(\omega' \to \omega)$. A point on this threshold lies "inside" the physical region if any neighborhood of that point contains physical points lying below the thresholds [i.e., at $E(\omega' \to \omega) < M_L$] and physical points lying above the threshold [i.e., at $E(\omega' \to \omega) > M_L$.] The argument goes as follows: Let P' be a point in the physical region lying just below the leading $\omega' \to \omega$ normal threshold. The first two terms on the right of (1.1) vanish at P'. Let P'' be a point such that there is a physical-region path $P'P''$ from P' to P'' with the following two properties: (a) Every singularity of R_c on $P'P''$ corresponds to a Landau diagram that can be contracted to a $\omega' \to \omega$ normal threshold diagram; (b) no singularity of R_c on $P'P''$ corresponding to a D having both positive and negative α's occurs. For any such P'' the function $R_c(P'')$ is an explicit expression in terms of strictly physical-region scattering functions of the continuation of the physical scattering function from P' to P'', along a path that has a minus-$i\epsilon$ continuation around every occurring singularity. This is because any pure-α singularity (i.e., one corresponding to a diagram in which all α's have the same sign) must be a pure-minus-α singularity, around which the function continues via a minus-$i\epsilon$ rule, and no mixed-α singularity lies on this path. Since the last term on the right of (1.1) is the continuation of the scattering function from P' to P'', the sum of the first two terms is precisely the discontinuity of the scattering function at P'' corresponding to the path $P'P''$. This argument is given in more detail, and is generalized, in Sec. 8.

At points P'' sufficiently near almost any point P on the threshold, the second term on the right of (1.1) is zero. Thus, the discontinuity around the threshold singularity alone is given by the first term on the right of (1.1). This result is represented by the equation

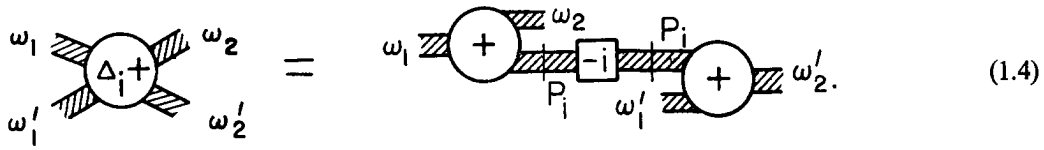


(1.3)

which is valid just above almost all points on the leading threshold T_L. The symbol on the left represents the discontinuity of the scattering amplitude around T_L. Equation (1.3) is a special case of a general rule first conjectured and discussed in Ref. 3.

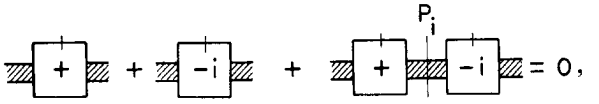
The above arguments apply equally well to the 1-particle $\omega' \to \omega$ normal threshold. In that case, the minus box in (1.3) becomes unity and the discontinuity formula becomes the well-known pole-factorization property.[4]

Results similar to those for the leading normal threshold described above are obtained for the non-leading normal thresholds lying inside the physical region below the 4-particle threshold in the $\omega' \to \omega$ channel. [*Note added in proof:* This restriction to regions lying below the 4-particle threshold is removed in the third paper of this series, (UCRL-18512, Oct. 1968), in which the general physical-region discontinuity formula is derived.] The analog of (1.3) reads



(1.4)

The P_i bar imposes the restriction that the sum of the rest masses of the particles associated with the lines cut by the bar be not less than the mass M_i associated with the threshold T_i in question. The $-i$ box is defined by the equation



(1.5)

where the bar on a box signifies that an I box has been subtracted off. This equation becomes the unitarity equation if the P_i bar is omitted, and the $-i$ boxes are replaced by minus boxes. Equation (1.4) is closely connected to a formula obtained by Olive[5] for 2-particle thresholds.

Equation (1.1) and the similar equation leading to (1.4) are physical-region identities. Thus, they may be substituted into the plus bubble appearing in the first term on the right. In this way, discontinuity equations for more complicated singularities can be obtained.

The main task in this paper is to derive the identity (1.1), and ones similar to it, and to substantiate the claims made regarding the properties of R_c. This involves repeated use of only the unitarity and cluster properties; analyticity properties are not involved. The first step is to develop a diagram calculus to deal with the cluster properties and unitarity equations for indeterminate numbers of particles. This is done in Sec. 2. Sections 3 and 4 establish some terminology and place on a firm basis some simple preliminary propositions. These two sections can be skimmed on first reading. The main proofs are in Secs. 5 and 6. Section 7 contains some incidental remarks concerning other forms of the results. The discontinuity formulas that follow from the identities established in Secs. 5 and 6 are described in detail in Sec. 8.

2. REPRESENTATION OF THE CLUSTER PROPERTY

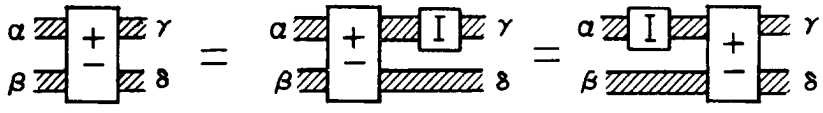
A box labeled by a symbol and connected to a set of lines represents a certain sum of bubble diagrams. The lines on the right and left will be called the incoming and outgoing lines, respectively. A plus (minus) box represents a sum over columns of plus (minus) bubbles, the sum being over all different ways that the given incoming and outgoing lines can be connected to each other by bubbles, subject only to the conditions that each line touch precisely one bubble and each bubble touch at least one incoming line and at least one outgoing line. (Each incoming and outgoing line terminates at or emerges from, respectively, the bubble it touches.)

An "I" box is constructed by the same rules, with the added condition that each bubble touch precisely one incoming line and precisely one outgoing line. No distinction is drawn between a plus, minus, or I bubble that satisfies this condition.

The unitary equation

(2.1)

is regarded as an equivalence relation connecting different box diagrams. As explained in Paper I, the rule for multiplication of diagrams is that topologically equivalent diagrams of the natural product are counted precisely once. This leads to the second fundamental equivalence relation



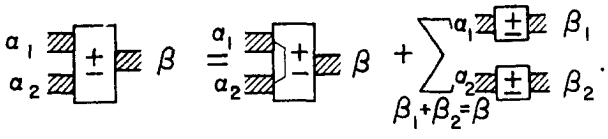
(2.2)

where the same choice of $\pm$ is to be used throughout. These two fundamental equivalences, when combined with cluster properties, will yield our results.

Often a set of intermediate lines connecting two boxes will not be explicitly shown: the two boxes will simply be moved into contact.

In the above equations and in what follows, sets of lines are labeled by Greek letters. *These sets are allowed to be empty, unless otherwise stated.* The equation $\alpha = 0$ means the set α is empty, $\alpha \neq 0$ means it is not empty, and $\alpha > 1$ means the set has more than one line. The lines are considered to run from right to left and symbols $X^+(\alpha)$ and $X^-(\alpha)$ denote the sets of leading and trailing end points, respectively, of the lines of the set α.

The cluster properties reside in the definitions of the plus and minus boxes in terms of their respective bubbles. To exploit these properties, we do not fully decompose the boxes into their constituent bubbles, but make, rather, partial decompositions into sets of terms with different connectedness properties. The first of these partial decompositions is expressed by the equation

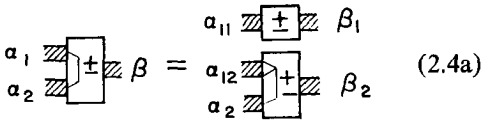


(2.3)

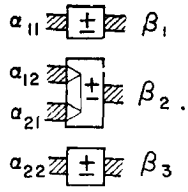
The first term on the right represents the sum of those terms of the left that contain at least one bubble touching lines in both α_1 and α_2. The second term consists of a sum over all decompositions of the set β into two sets β_1 and β_2. Each term of this sum consists of the indicated boxes combined according to a product rule of composition. This rule gives, for each pair consisting of one term from each box, a column consisting of the sum of the bubbles of the two members of this pair. The validity of (2.3) is proved in Appendix A.

Henceforth, we adopt a summation convention that serves to eliminate the summation sign in (2.3) and in similar equations that follow. If a set ϕ appearing on the left is partitioned in all possible ways into certain sets appearing on the right, then these latter will be denoted by ϕ_j: A summation over all partitions of the ϕ into the various ϕ_j is always implied. Occasionally the set on the left will already have an index, but the rule still applies. For example, α_1 is partitioned into sets α_{1j}.

A second important decomposition rule is

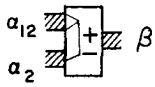


(2.4a)



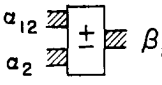
(2.4b)

Here



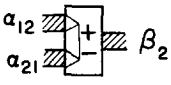
(2.5)

is defined to be empty if $\alpha_2 = 0$ or $\alpha_{12} = 0$. Otherwise, it is the sum of all contributions to



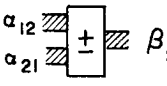
(2.6)

with the property that each line of α_{12} touches a bubble that touches at least one line of α_2. The box diagram



(2.7)

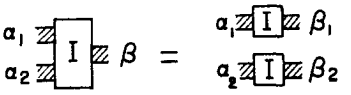
is defined to be empty if $\alpha_{12} = 0$ or $\alpha_{21} = 0$. Otherwise, it is the sum of all contributions to



(2.8)

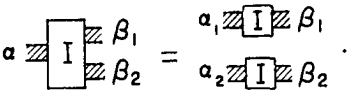
that have the property that each line of α_{12} touches a bubble that touches at least one line of α_{21}, and vice versa. Equations (2.4a) and (2.4b) are proved in Appendix A.

The decomposition (2.3) also applies to the I box, but the first term on the right of (2.3) is then empty. Thus,



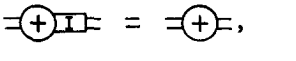
(2.9)

and



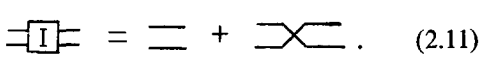
(2.9′)

The I box (2.9) or (2.9′) is equivalent to the identity when postmultiplying a bubble diagram symmetric in the set of lines α or when premultiplying a bubble diagram symmetric in the set of lines β. This follows from the definition of the I box and from the fact that in products of boxes topologically equivalent contributions are counted only once. Thus, for instance,



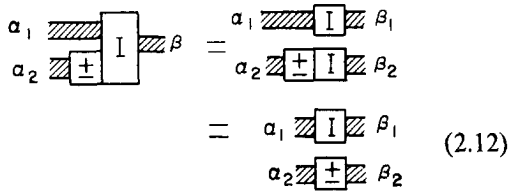
(2.10)

even though



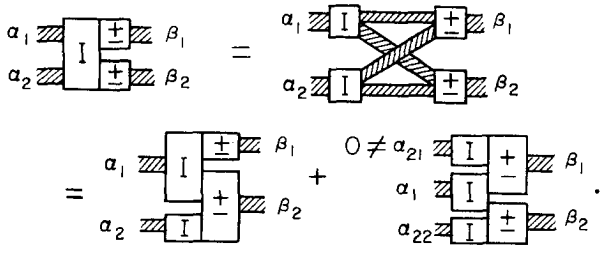
(2.11)

From (2.9), (2.9′), and (2.2) we obtain



(2.12)

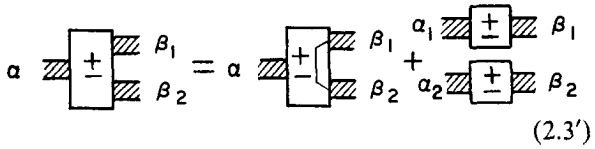
and



(2.13)

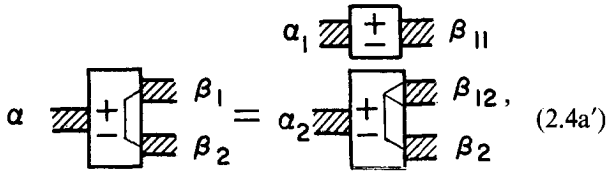
(The number of lines crossing an interface between boxes or represented by a shaded strip is always allowed to be zero, unless otherwise stated.)

Remark 2.1: Equations completely analogous to those discussed in this section, but with the role of the incoming and outgoing particles switched, are denoted by a primed equation number. We shall use, in particular, the equations



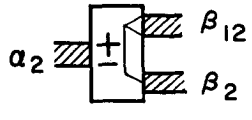
(2.3′)

and



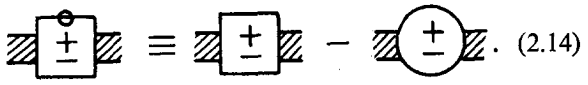
(2.4a′)

where the box diagram



(2.5′)

is defined in analogy with the definition of (2.5). We also record for future use the definition



(2.14)

Manipulations with the unitarity equations and cluster properties give equations that are represented by bubble diagram equations. The bubble diagrams B that occur all satisfy the partial ordering condition that the diagrams can be drawn with all lines of B directed from right to left. The singularities of the corresponding functions M^B (defined in I) are confined to the Landau surfaces $L(D)$ for $D \supset\subset B$, as discussed above.

These diagrams D are simply topological structures. They should not be confused with the geometric structures $\bar{D}$ obtained from them by mapping each line L_j of D into a 4-vector $\Delta_j = \alpha_j p_j$. The diagrams $\bar{D}$ are a geometric representation of all the Landau equations, including the loop equations, whereas the diagrams $D \supset\subset B$ represent only the conservation laws. In particular, the significance of the arrows on the lines of D arises from the conservation laws

$$\sum_j p_j \epsilon_{jn} = 0;$$

since each p_j^0 is by definition positive and the arrow on each line L_j points to the end at which ϵ_{jn} is $+1$, these arrows simply indicate the direction of the flow of positive energy. This interpretation of the arrows in terms of energy flow is independent of the sign of α_j.

The lines of a bubble diagram are defined to be directed from right to left. The lines in the interior of each plus bubble will also be directed from right to left, provided that the vector diagram $\bar{D}_c^b$ is mapped into the topological space by a monotonic mapping that places points of greater energy further left. The interior lines of each minus bubble will be leftward-directed if the opposite rule is used. Thus, if these rules are adopted, all the vectors of any $D' \subset B$ will point from right to left, and one has a diagrammatic representation of the conservation laws in which positive energy flows from right to left.

The convention just established plays no essential role in our arguments. But it makes geometrically obvious the fact that each point of any $D \supset\subset B$ lies on a continuous, directed path P that runs from the trailing end point of some incoming line of B to the leading end point of some outgoing line of B and is such that each line segment L_j lying on P has the same direction as P itself. That such a P exists follows analytically from energy conservation and the partial-ordering requirement imposed on bubble diagrams (see Sec. 2 of Paper I).

3. CUT SETS AND DIAGRAMS THAT CANNOT BE CONTRACTED TO POSITIVE-α NORMAL-THRESHOLD DIAGRAMS

Consider a transition from the system of m initial particles to the system of n final particles. Any separation of the $n + m$ particles into two disjoint subsets defines a channel. If the two subsets are precisely the initial and final sets, then the channel is called the direct channel. If each subset contains both initial and final particles, then the channel is called a cross channel. If one subset contains only initial or only final particles and the other subset contains both initial and final particles, then the channel is called a subchannel. The three cases are indicated in Fig. 1, where we have labeled one of the two subsets of lines ω and the other ω'. By convention the set ω does not consist only of initial lines, and ω' does not consist only of final lines. The sets ω and ω' are further subdivided into the terms $\omega_1 + \omega_2$ and $\omega_1' + \omega_2'$, respectively, where $\omega_1 + \omega_1'$ is the set of final lines and $\omega_2 + \omega_2'$ is the set of initial lines. The sets of external end points of the lines of ω and ω' are denoted by $X(\omega) \equiv X^+(\omega_1) + X^-(\omega_2)$ and $X(\omega') \equiv X^+(\omega_1') + X^-(\omega_2')$, respectively. The "$\omega' \to \omega$ channel" will mean the channel labeled in this way.

The $\omega' \to \omega$ channel energy $E(\omega' \to \omega)$ is

$$E(\omega, \omega') = \sum_{i \in I(\omega_1)} p_i^0 - \sum_{j \in I(\omega_2)} p_j^0 = \sum_{i \in I(\omega_2')} p_i^0 - \sum_{j \in I(\omega_1')} p_j^0, \tag{3.1}$$

evaluated in a frame where the corresponding 3-momentum is zero. The set $I(\alpha)$ is the set of indices labeling the lines of α. Thus, the channel energy $E(\omega' \to \omega)$ is the net center-of-mass energy flowing from $X(\omega')$ to $X(\omega)$.

Definition 3.1: An (ω, ω') *cut set* of a Landau diagram D supported[1] by the bubble diagram B (i.e., $D \supset\subset B$) is a (possibly empty) collection C of lines of D with the following properties: (1) Every path in D that starts in $X(\omega')$ and ends in $X(\omega)$ passes along the interior of some line of C, and (2) Property (1) is not satisfied by any proper subset of C.

Proposition 3.1: Let C be an (ω, ω') cut set of a $D \supset\subset B$ and let C' be the set of all interior points of all lines of C. Let $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ be the parts of D connected in $D - C'$ to $X(\omega)$ and $X(\omega')$, respectively. Then,

(i) $\bar{X}(C, \omega) \cap \bar{X}(C, \omega') = 0$ (the empty set),

(ii) each line of C has one of its end points in $\bar{X}(C, \omega)$ and its other end point in $\bar{X}(C, \omega')$, and

(iii) $D - C' = \bar{X}(C, \omega) \cup \bar{X}(C, \omega')$.

Proof: (i) If $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ had a common point, then $X(\omega)$ could be connected to $X(\omega')$ in $D - C'$. But then C could not be an (ω, ω') cut set, contrary to assumption. (ii) The lines of D can intersect (by definition) only at their end points. Let C_ω be the set of lines of C that have an end point in $\bar{X}(C, \omega)$. Since $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ are disjoint, every path in D from $X(\omega)$ to $X(\omega')$ must leave $\bar{X}(C, \omega)$ via a line of C_ω. Thus, $X(\omega)$ and $X(\omega')$ cannot be connected by any path in $D - C_\omega'$, where C_ω' is the set of all interior points of all lines of C_ω. But then C_ω must be just C, for if C_ω were a proper subset of C, then the second requirement on the (ω, ω') cut set C would be violated. Similarly, every line of C must have an end point in $\bar{X}(C, \omega')$. Since $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ are disjoint, one end point of each line of C must lie in $\bar{X}(C, \omega)$ and the other must lie in $\bar{X}(C, \omega')$. (iii) Suppose there is a part of $D - C'$ not connected in $D - C'$ to $X(\omega)$ or to $X(\omega')$. Then, by virtue of (ii), this part cannot be connected in D to $X(\omega)$ or to $X(\omega')$. But all parts of any $D \supset\subset B$ are connected in D to external lines of B and, hence, to $X(\omega)$ or to $X(\omega')$, by virtue of the conditions on $D \supset\subset B$. In particular, each bubble of B has both incoming and outgoing lines. The partial-ordering condition on the bubbles of B ensures that each bubble lies on a path that starts at some incoming line of B and ends at some outgoing line of B. In forming $D \supset\subset B$, the bubbles are replaced by connected diagrams. Thus, every point of $D \supset\subset B$ is connected to both the incoming and the outgoing lines of B and, hence, to $X(\omega) \cup X(\omega')$.

A schematic representation of Proposition 3.1 is shown in Fig. 2.

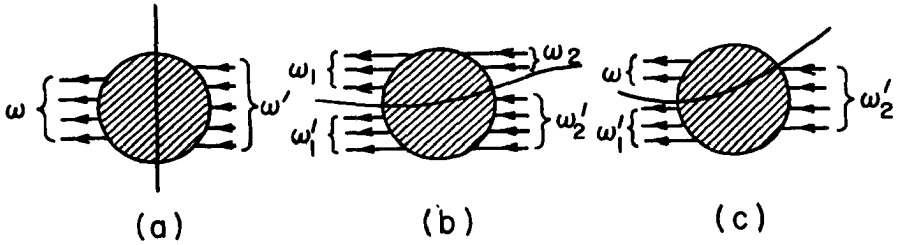


FIG. 1. Three types of channels (a) the direct channel; (b) a cross channel; (c) a subchannel.

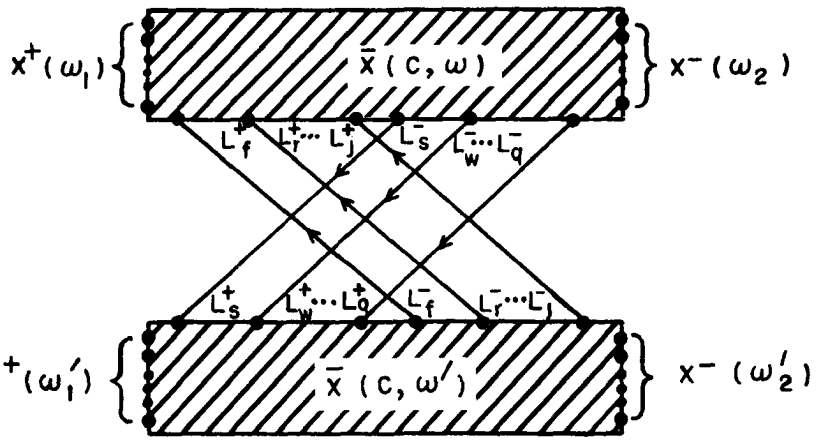


FIG. 2. The topological structure induced by an (ω, ω') cut set C. The directed line segments $L_i \in C$ are shown. The lines are all shown directed to the left in accordance with the discussion at the end of Sec. 2. The $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ parts of $D \supset\subset B$ (indicated by the shaded areas of the figure) may be disconnected diagrams. Each point of $\bar{X}(C, \omega)[\bar{X}(C, \omega')]$ is connected in $\bar{X}(C, \omega)$ $[\bar{X}(C, \omega')]$ to one or more of the external end points $X(\omega)[X(\omega')]$ exhibited in the figure. Some of the leading end points $L_f^+, L_r^+, \cdots, L_j^+$ of the upward-directed line segments of C may lie in $X^+(\omega_1)$, some of the trailing end points $L_s^-, L_w^-, \cdots, L_q^-$ of the downward-directed line segments of C may lie in $X^-(\omega_2)$, and similarly for the lower end points of these lines. The ordering of the end points (i.e., $L_i^+ > L_j^+$) is not significant. The set C is called a *simple* (ω, ω') cut set (see Definition 3.2) if and only if the set of downward-directed line segments [i.e., those leading from $\bar{X}(C, \omega)$ to $\bar{X}(C, \omega')$] is empty.

Definition 3.2: A *simple* (ω, ω') *cut set* C is an (ω, ω') cut set C such that every line of C has its leading end point in $\bar{X}(C, \omega)$ and its trailing end point in $\bar{X}(C, \omega')$.

Definition 3.3: A *simple positive* (ω, ω') *cut set* is a simple (ω, ω') cut set having no "minus lines." (A *minus line* is a line of $D \supset\subset B$ that is an internal line of a Landau diagram D_c^b corresponding to a minus bubble b of B. Such lines must carry negative Landau α's, according to the rules set down in Ref. 1.)

Definition 3.4: $\mathcal{R}(\omega, \omega')$ is the set of all bubble diagrams B with the property that no $D \supset\subset B$ contains a simple positive (ω, ω') cut set.

Definition 3.5: $\mathcal{R}_c(\omega, \omega')$ is the set of all bubble diagrams B for which at least one of the following two conditions holds:

(1) B is not a connected diagram;

(2) no $D \supset\subset B$ contains a simple positive (ω, ω') cut set such that $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ are both connected diagrams.

It is evident that

$$\mathcal{R}(\omega, \omega') \subset \mathcal{R}_c(\omega, \omega'). \tag{3.2}$$

These definitions isolate the class of bubble diagrams that cannot support any D_n^+ $(\omega' \to \omega)$: It is clear from the definition of $D \supset\subset B$ (see Sec. 1 above) that no bubble diagram belonging to $\mathcal{R}_c(\omega, \omega')$ can contain any D_n^+ $(\omega' \to \omega)$ or any D that contracts to any D_n^+ $(\omega' \to \omega)$.

The importance to us of this classification arises from the third structure theorem, which ensures that the bubble-diagram function M^B represented by any $B \in \mathcal{R}_c(\omega, \omega')$ has a minus-iϵ continuation into itself around the singularity associated with any pure-α D that can be contracted to any $\omega' \to \omega$ normal-threshold diagram. Our principal task will be to prove that certain bubble diagrams are equivalent to diagrams belonging to $\mathcal{R}_c(\omega, \omega')$.

4. A CHARACTERIZATION OF BUBBLE DIAGRAMS THAT SUPPORT NO POSITIVE-α DIAGRAMS THAT CONTRACT TO A $\omega' \to \omega$ NORMAL-THRESHOLD DIAGRAM

The characterization we seek is expressed in terms of "paths." A path in D is a continuous directed curve composed of an ordered sequence of line segments L_j of D. Neighboring segments meet at coincident end points. Each segment L_j of a path P can be independently directed either along the path or against it. The direction of the path is specified by specifying its origin and its destination. The direction of the line segment L_j of D is from L_j^- to L_j^+, as already mentioned.

Definition 4.1: A $P \in \mathfrak{P}_{om}(D \supset\subset B)$ is a path in $D \supset\subset B$ having the property that each L_j lying on P is either directed against P or is a minus line (as defined in Definition 3.3).

Proposition 4.1: If every $D \supset\subset B$ contains a $P \in \mathfrak{P}_{om}(D \supset\subset B)$ that runs from $X^+(\omega_1')$ to $X^-(\omega_2)$, then B belongs to $\mathcal{R}(\omega, \omega')$, and conversely.

Proof: If B does not belong to $\mathcal{R}(\omega, \omega')$, then some $D \supset\subset B$ must have a simple positive (ω, ω') cut set C. Any path P from $X^+(\omega_1')$ to $X^-(\omega_2)$ must contain lines of C. The first of these lines encountered as P is traced out is a nonminus line directed along P. Thus, this $D \supset\subset B$ can contain no $P \in \mathfrak{P}_{om}(D \supset\subset B)$.

To prove the converse, suppose some $D \supset\subset B$ contains no $P \in \mathfrak{P}_{om}(D \supset\subset B)$ that runs from $X^+(\omega_1')$ to $X^-(\omega_2)$. Then, any path p in this $D \supset\subset B$ from

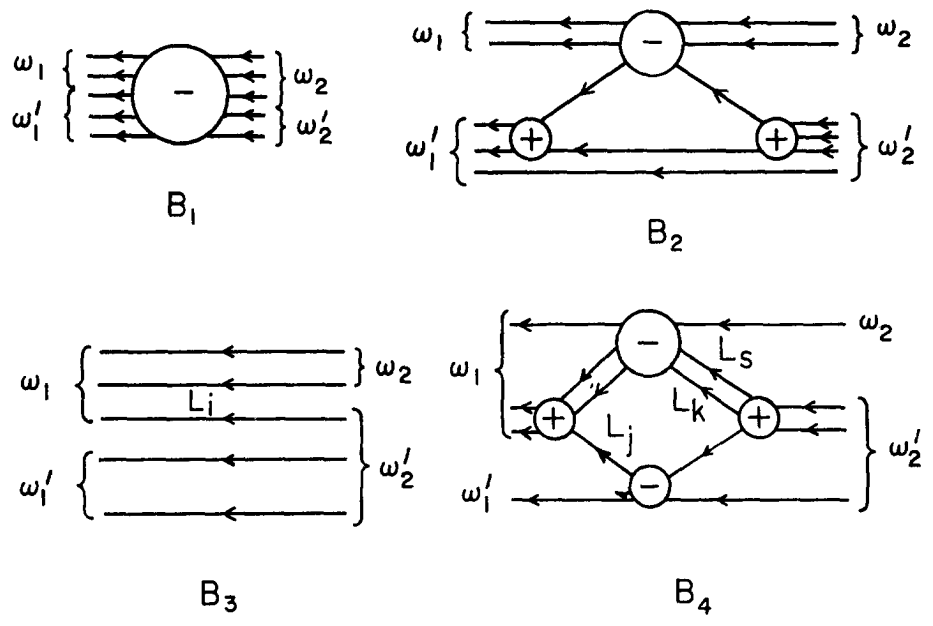


FIG. 3. The bubble diagrams B_1 and B_2 belong to $\mathcal{R}(\omega, \omega')$, whereas B_3 and B_4 do not. This follows from Proposition 4.1: For every $D \supset\subset B_1$ or $D \supset\subset B_2$, there is a path P running from $X^+(\omega_1')$ to $X^-(\omega_2)$ such that P is in $\mathfrak{I}_{om}(D \supset\subset B_1)$ or $\mathfrak{I}_{om}(D \supset\subset B_2)$, respectively. But not every $D \supset\subset B_3$ or $D \supset\subset B_4$ has a path P running from $X^+(\omega_1')$ to $X^-(\omega_2)$ such that P is in $\mathfrak{I}_{om}(D \supset\subset B_3)$ or $\mathfrak{I}_{om}(D \supset\subset B_4)$, respectively. That B_3 and B_4 do not belong to $\mathcal{R}(\omega, \omega')$ may also be seen directly from Definition 3.4, since the lines L_i of B_3 and the lines L_j, L_k, and L_s of B_4 are a simple positive (ω, ω') cut set for some $D \supset\subset B_3$ or $D \supset\subset B_4$, respectively.

$(X^+\omega_1')$ to $X^-(\omega_2)$ must have a first segment L_p that is both nonminus and directed along the path. Let the union of the L_p over all paths p from $X^+(\omega_1')$ to $X^-(\omega_2)$ be denoted by $\bigcup L_p$, and define $\bar{C} = (\bigcup L_p) \cup \omega_1 \cup \omega_2'$. The sets $X(\omega)$ and $X(\omega')$ certainly cannot be connected in $D - \bar{C}'$, where $\bar{C}'$ is the set of interior points of $\bar{C}$. Some subset C of $\bar{C}$ is, therefore, an (ω, ω') cut set. Each line of C is a nonminus line with either its trailing end point in $\bar{X}(C, \omega')$ or its leading end point $\bar{X}(C, \omega)$. Thus, by virtue of Proposition 3.1, each leading end point of C is in $\bar{X}(C, \omega)$ and each trailing end point is in $\bar{X}(C, \omega')$. Thus, C is a simple positive (ω, ω') cut set for this $D \supset\subset B$. But then B cannot belong to $\mathcal{R}(\omega, \omega')$.

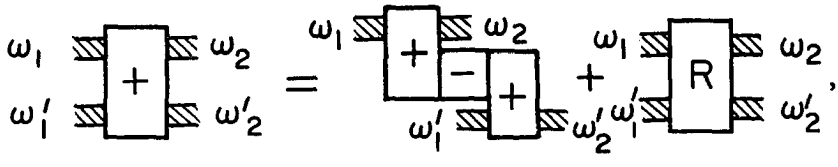
A few examples illustrating Proposition 4.1 and the various definitions of Sec. 3 are given in Fig. 3. In these examples one should insert various connected Landau diagrams for the bubbles. However, the results are independent of the form of these diagrams D_c^b. We need only the general result that there is a path from any incoming line of any D_c^b to some outgoing line such that the direction of the path agrees with the directions of all its segments. This follows from energy conservation and the fact that energy flows in along incoming lines and out along outgoing lines. (See the discussion at the end of Sec. 2.)

5. NORMAL-THRESHOLD EXPANSIONS OF SCATTERING FUNCTIONS

A. Two Basic Identities

Unitarity can be regarded as an equivalence relationship between different box diagrams. In this section certain box diagrams are converted by repeated use of unitarity and the cluster properties to certain equivalent box diagrams.

Proposition 5.1: The cluster properties and unitarity imply that

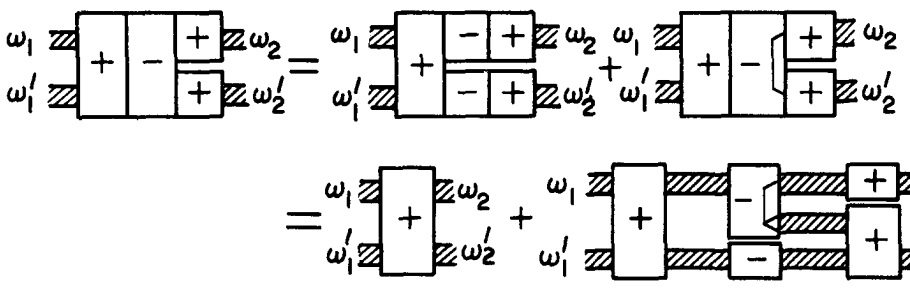


(5.1)

where the R box represents a sum of bubble diagrams all of which belong to $\mathcal{R}(\omega, \omega')$.

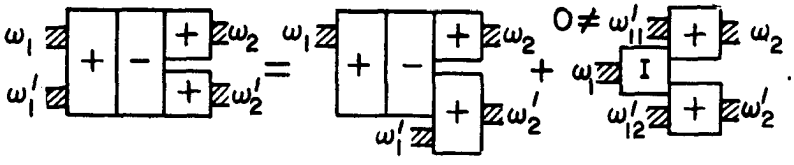
Proof: The proposition follows immediately from (2.1) and (2.2) if $\omega_2 = 0$ or if $\omega_1' = 0$, in which case $R = 0$. Thus, we assume that ω_2 and ω_1' are nonempty. The use, first, of (2.3') and, then, of (2.1), (2.2), and

(2.4a′) gives



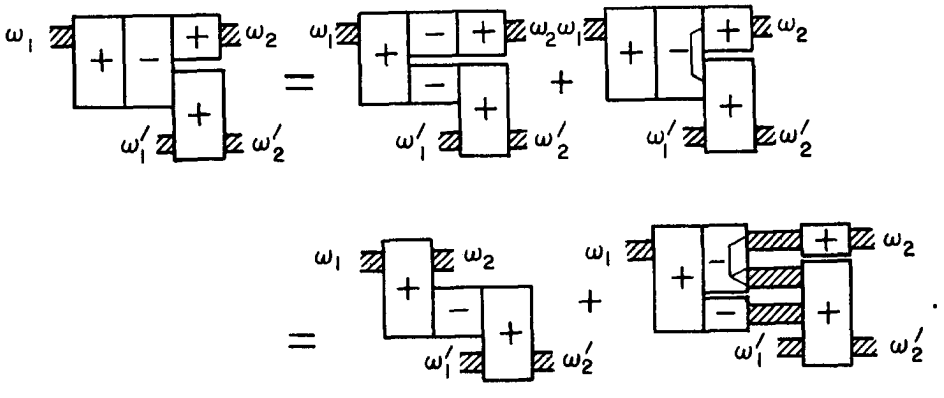
(5.2)

(The various summations required by the formulas of Sec. 2 are implied by the summation convention for internal lines introduced in Sec. II of Paper I.) Alternatively, the use, first, of unitarity, then, of (2.13), (2.2), and again unitarity gives



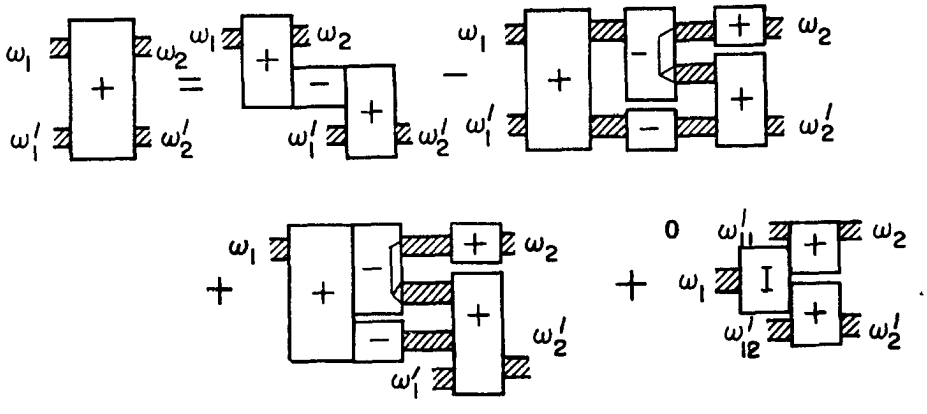
(5.3)

The application, first, of (2.3′) and, then, of (2.1), (2.2), and (2.4a′) to the first term on the right side of (5.3) gives



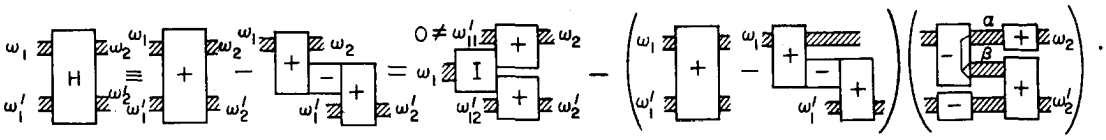
(5.4)

Combining (5.2), (5.3), and (5.4), one obtains



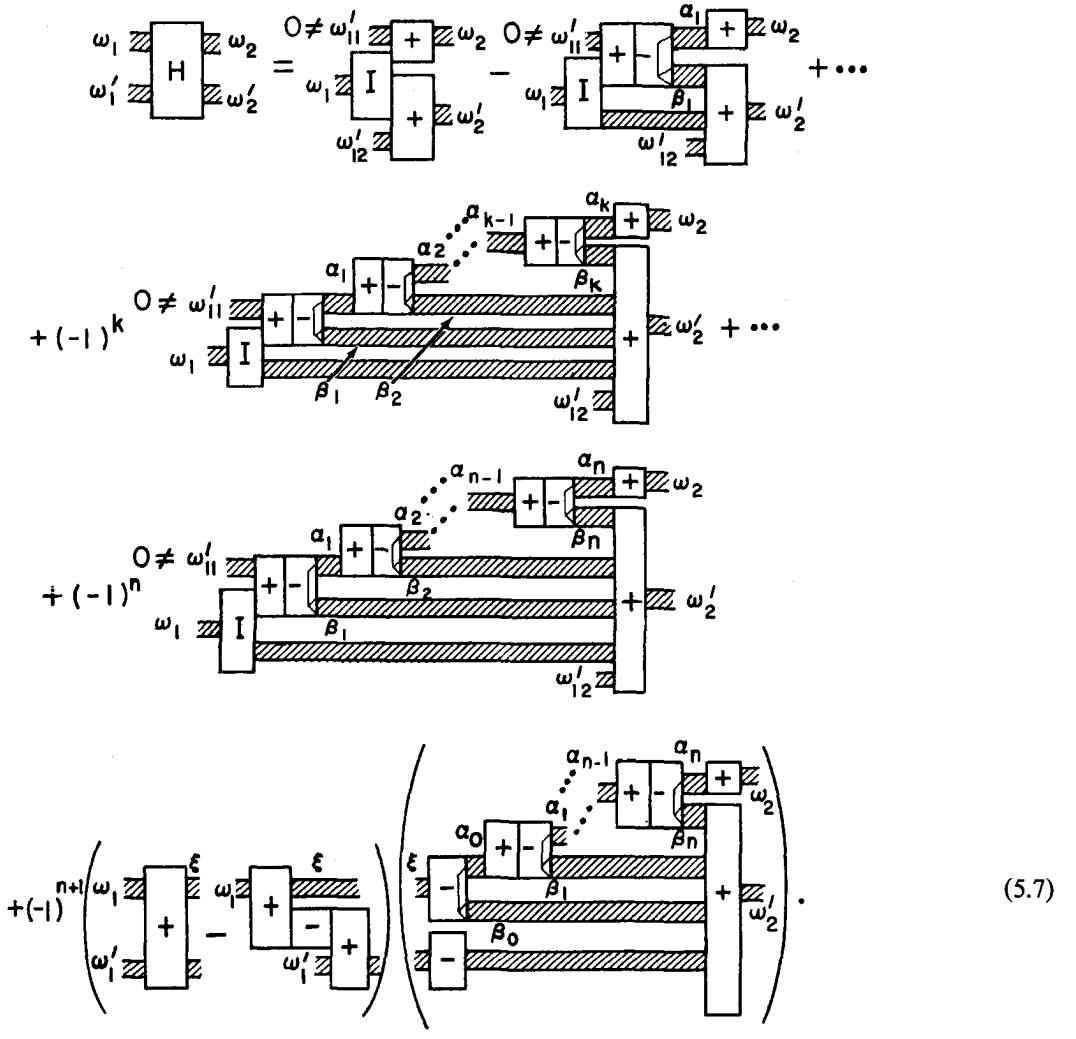
(5.5)

A rearrangement of terms and an application of (2.2) and unitarity converts this to



(5.6)

Equation (5.6) can be iterated. Iterating n times, each time simplifying by means of unitarity and (2.2), one obtains



(5.7)

The number of iterations n is fixed so that

$$E(\omega, \omega') < (n+1)E_0,$$

where E_0 is the rest energy of the lightest particle. The definition of the box diagram (2.5′) requires that the sets α_i and β_i be nonempty; otherwise, the set of diagrams containing these sets is empty.

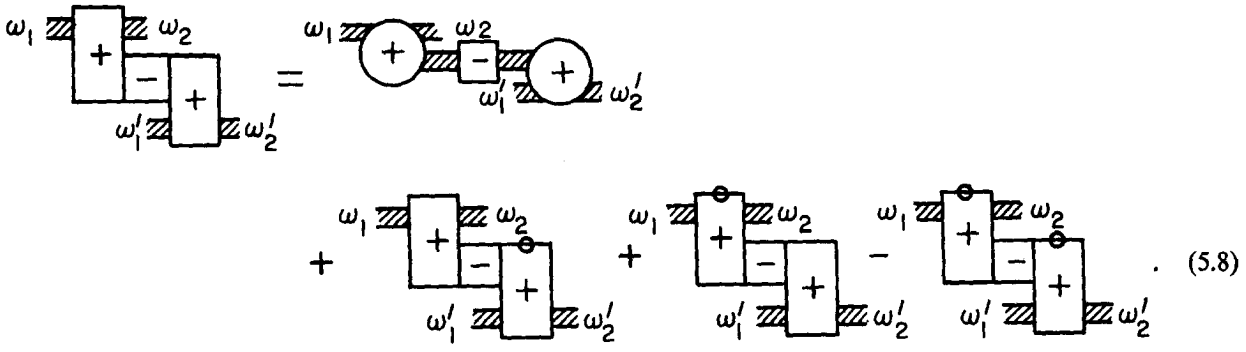
Let the box diagram on the right of (5.7) be denoted by $H_0, H_1, \cdots, H_k, \cdots, H_n, H_{n+1}$, where H_0 is the first term on the right, H_1 the second, etc. Any Landau diagram supported by any H_i contains a path $P \in \mathfrak{F}_{om}(D \supset\subset H_i)$ from $X^+(\omega_1')$ to $X^-(\omega_2)$, as will now be shown. For any H_i but H_{n+1}, each end point of $X^+(\omega_{11}')$ is connected to the trailing end point of some incoming line of the leftmost upper plus box by a path whose sense is opposed to that of each line of that plus box. This follows from energy conservation. (See the discussion at the end of Sec. 2.) This concludes the argument for H_0. For the remaining H_i, $0 < i \leq n$, the

path just constructed can, by virtue of the definition of (2.5′), be continued to some point of $X^+(\alpha_1)$ by a path composed only of minus lines and, hence, lying in $\mathfrak{I}_{om}$ $(D \supset\subset H_i)$. The same argument allows the path to be continued to α_2, then to α_3, and so on to $X^-(\omega_2)$. Thus, according to Proposition 4.1, all H_i except H_{n+1} belong to $\mathfrak{R}(\omega, \omega')$. Finally, we consider H_{n+1}. Since the sets β_i are nonempty, the energy $E(\xi)$ of the set ξ satisfies $E(\xi) \geq E(\omega_2) + (n+1)E_0$. On the other hand, the condition on n is that

$$E(\omega, \omega') \equiv E(\omega_1) - E(\xi) + E(\xi) - E(\omega_2) < (n+1)E_0.$$

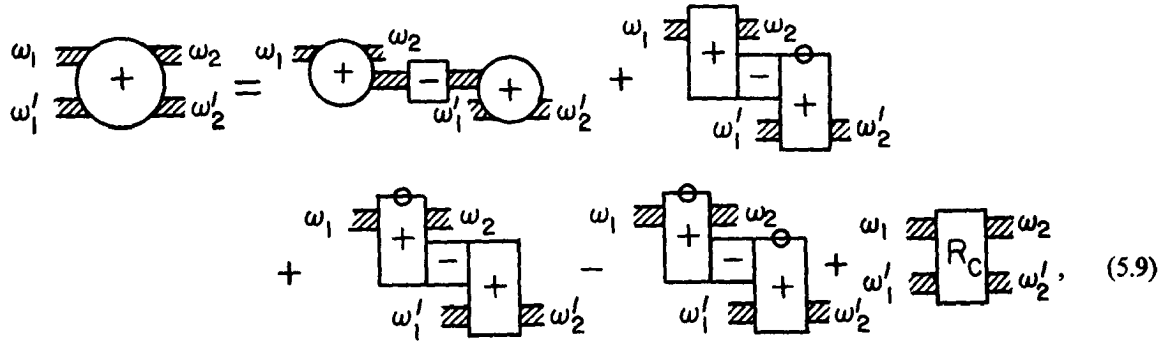
These inequalities combine to give $E(\xi) > E(\omega_1)$. But then energy conservation implies that the second term in the first parenthesis vanishes.[6] Furthermore, in the first term in the first parenthesis some energy must flow from $X^-(\xi)$ to $X^+(\omega_1')$. This ensures that there is a path $P \in \mathfrak{I}_{om}$ $(D \supset\subset H_{n+1})$ from $X^+(\omega_1')$ to $X^-(\xi)$. This path can then be extended to $X^-(\omega_2)$ by means of the same arguments as before. Thus, all terms on the right of (5.7) belong to $\mathfrak{R}(\omega, \omega')$, and Proposition 5.1 is proved.

By (2.14) the first term on the right of (5.1) can be written in the form



(5.8)

Then, Definition 3.5 [which implies that B belongs to $\mathfrak{R}_c(\omega, \omega')$ if it belongs to $\mathfrak{R}(\omega, \omega')$], together with (2.14) and (5.8), allows (5.1) to be written in the form

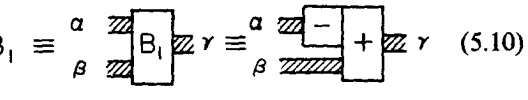


(5.9)

where the R_c box consists of a sum of terms of $\mathfrak{R}_c(\omega, \omega')$.

Our next objective is to show that the second and third terms on the right of (5.9) can be placed in the last term. To this end we first prove

Proposition 5.2: The box diagram



(5.10)

is equivalent to a diagram B_1' such that the only simple positive $(\alpha, \beta + \gamma)$ cut set C of any $D' \supset\subset B_1'$ is the set of lines α.

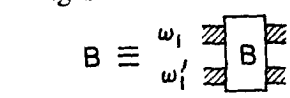
Proof: Let B_1' be the right side of (B4). In each term let α_I be the subset of α connected to the I box. The set C evidently contains α_I. For the first term on the right of (B4), the set α_I is α and the required result clearly holds. For any D' supported by any other term, the definition of (2.5) guarantees that any point of $X^-(\alpha - \alpha_I)$ lies at the end of some path

$$P \in \mathfrak{I}_{om}(D' \supset\subset B_1')$$

that starts in $X^+(\beta)$. [One uses the properties of (2.5) to trace a path P' consisting only of minus lines from each point of $X^-(\alpha_i)$ to some point of $X^-(\delta_i)$. This path can then be extended to a path P'' ending at some point of $X(\beta)$; one uses the properties of (2.5) to get through each minus box of the form (2.5) encountered on the path from $X^-(\delta_i)$ to $X(\beta)$. The desired P is the negative of P''.] No point of such a path P can belong to any simple positive cut set. Thus, all points of $X^-(\alpha)$ must belong to $\bar{X}(C, \beta + \gamma)$, by the definition of $\bar{X}(C, \beta + \gamma)$. Thus, the only simple positive $(\alpha, \beta + \gamma)$ cut set is α.

This proof also shows that for any $D' \supset\subset B_1'$, all points of $D' - \alpha$ lie in $\bar{X}(C, \beta + \gamma)$. This gives the following:

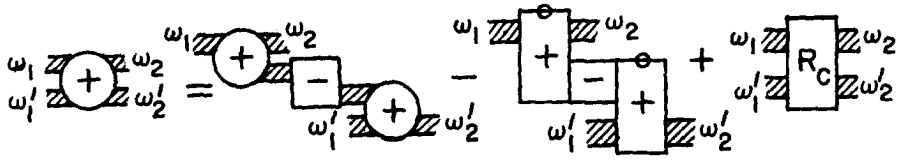
Corollary to 5.2: Suppose the B_1 of (5.10) is part of some box diagram



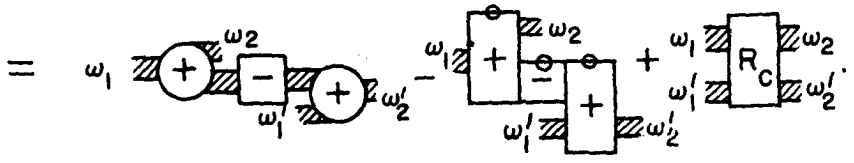
(5.11)

Then the replacement of B_1 in B by the equivalent B_1' of Proposition 5.2 gives a B' with the property that, for any simple positive (ω, ω') cut set C of any $D' \supset\subset B'$, all the points of $X^-(\alpha)$ will belong to $\bar{X}(C, \omega')$ if all points of $X^+(\beta)$ and $X^-(\gamma)$ do. In fact, all points of $D_1' - \alpha$, where D_1' is that part of D' which is supported by B_1', lie in $\bar{X}(C, \omega')$ if all points of $X^+(\beta)$ and $X^-(\gamma)$ do.

Proposition 5.3: The cluster properties and unitarity imply that



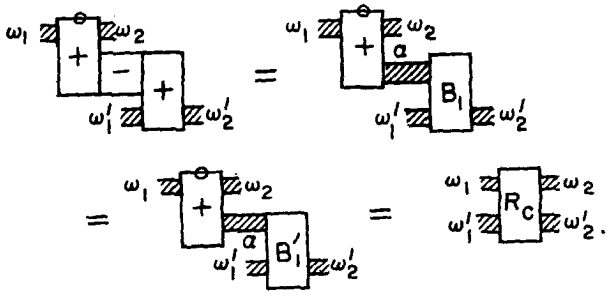
(5.12)



(5.13)

The R_c box is now, and hereafter, a generic symbol used to denote any sum of bubble diagrams each term of which belongs to $R_c(\omega, \omega')$.

Proof: The Corollary to 5.2 applied to the third term on the right of (5.9) gives

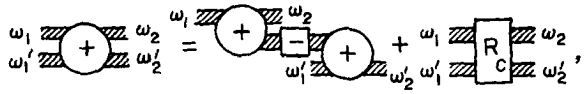


(5.14)

The last step follows from the fact that any simple positive (ω, ω') cut set C that leaves $X^-(\alpha)$ in $\bar{X}(C, \omega')$ must make $\bar{X}(C, \omega)$ a disconnected diagram. A similar argument applies to the second term on the right of (5.9). This proves (5.12). The contribution to the fourth term on the right of (5.9) coming from the connected part of the minus box clearly has no positive cut set C that leaves $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ both connected diagrams. Thus, it belongs to $\mathcal{R}_c(\omega, \omega')$ and the proof is complete.

The form of the first term on the right of (5.13) is invariant under the crossing (incoming$\leftrightarrow$ outgoing) of lines of ω or of ω'. However, no analyticity or crossing properties have been used to derive (5.13); the result is obtained strictly from unitarity and cluster properties in the direct channel.

For $E(\omega, \omega')$ less than the lowest 4-particle $\omega' \to \omega$ threshold E_4, the second term on the right side of (5.13) belongs to $\mathcal{R}_c(\omega, \omega)$. This is proved in Sec. 6. Thus, for this energy range Eq. (5.13) reduces to

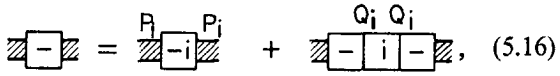


$$E(\omega, \omega') < E_4. \tag{5.15}$$

B. Expansion Exhibiting the Discontinuity of the Scattering Function for Nonleading Normal Thresholds

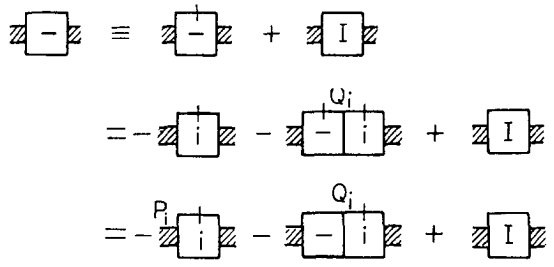
The expression for the scattering function given by Proposition 5.3 exhibits the discontinuity function for the leading normal threshold in the (ω, ω') channel. To exhibit the discontinuity for nonleading normal threshold at channel energy $E(\omega, \omega') = M_i$, we use the i-box formalism of Paper I. This formalism applies only if $E(\omega, \omega')$ is below the 4-particle threshold of the $\omega' \to \omega$ channel. Thus, the following results are similarly restricted.

The basic identity we need is



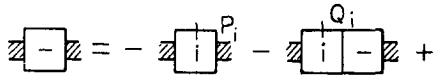
(5.16)

where P_i and Q_i are the projection operators associated with the mass M_i [see (5.5) of I]. To prove (5.16), we first use (2.9), (5.8), (5.5), and (5.19) of I to obtain



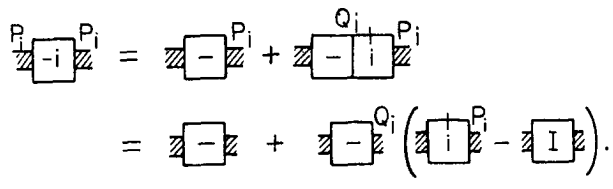
(5.17)

and



(5.18)

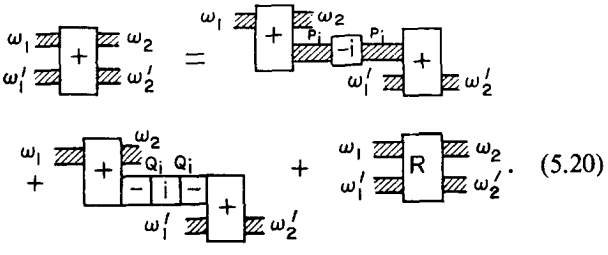
Definition (5.44) of I converts (5.17) to



(5.19)

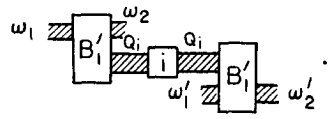
Application of (5.18) to the term in parenthesis yields (5.16).

Substitution of (5.16) into (5.1) yields



(5.20)

By Proposition 5.2, the second term on the right of (5.20) is equivalent to

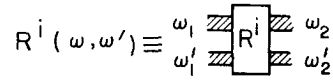


(5.21)

Some new definitions are now needed. Let

$$C^i(\omega, \omega') \equiv C^i$$

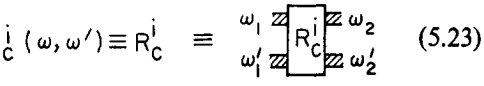
denote any simple positive (ω, ω') cut set composed of lines such that the sum of their masses is greater than or equal to M_i. Let $\mathcal{R}^i(\omega, \omega')$ represent the set of all bubble diagrams B with the property that no $D \supset\subset B$ has a cut set C^i. Let



(5.22)

represent any sum of terms of $\mathcal{R}^i(\omega, \omega')$.

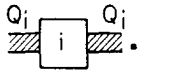
Similarly, let



(5.23)

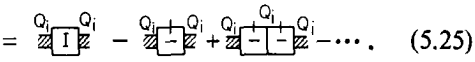
represent any sum of bubble diagrams, each term of which either is disconnected or does not support any Landau diagram containing a cut set C^i such that $\bar{X}(C^i, \omega)$ and $\bar{X}(C^i, \omega')$ are both connected diagrams. Evidently, the sets $\mathcal{R}(\omega, \omega')$, $\mathcal{R}_c(\omega, \omega')$, and $\mathcal{R}^i(\omega, \omega')$ are subsets of $\mathcal{R}_c^i(\omega, \omega')$. The function corresponding to any $B \in \mathcal{R}_c^i(\omega, \omega')$ has, by virtue of the third structure theorem, a minus-$i\epsilon$ continuation into itself past any normal-threshold singularities at $E(\omega, \omega') = M_i$, which is associated with a $\omega' \to \omega$ normal-threshold diagram.

By the Corollary to (5.2), any simple positive (ω, ω') cut set and, in particular, any $C^i(\omega, \omega')$, contained in a Landau diagram D supported by (5.21), must be composed only of lines of that part of D that is supported by



(5.24)

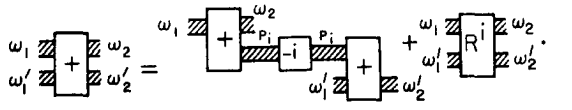
The function J_i is defined by Fredholm theory (see Appendix B of I). Its analytic structure is, however, exhibited by the terms of the formal expansion



(5.25)

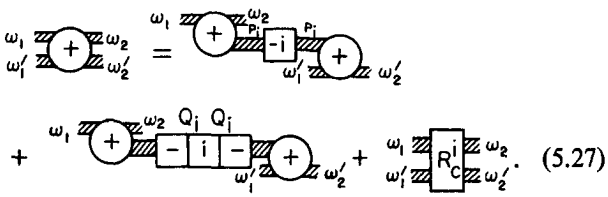
Each term in this expansion consists of minus bubbles and sets of lines restricted by the Q_i bar conditions.

Consequently, no Landau diagram supported by any term on the right of (5.25) can contain a cut set $C^i(\omega, \omega')$. Therefore, (5.21) belongs to $\mathfrak{R}^i(\omega, \omega')$, and (5.20) can be written as



(5.26)

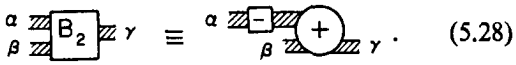
The formula just obtained is the i-box version of Proposition 5.1. It can be simplified by further manipulations. Alternatively, one can apply (5.16) directly to (5.15). Then, one obtains



(5.27)

The second term on the right can be incorporated into the last by virtue of the following proposition:

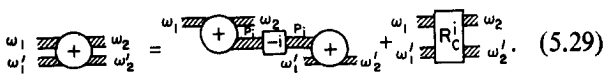
Proposition 5.2C: Consider the diagram



(5.28)

If the center-of-mass energy of the set α is below the 4-particle threshold in the $\beta + \gamma \to \alpha$ channel, then B_2 is equivalent to a bubble diagram B_2', having the property that for any $D' \supset\subset B_2'$ the only simple positive $(\alpha, \beta + \gamma)$ cut set C with connected $\bar{X}(C, \beta + \gamma)$ is α itself.

This proposition, which is similar to Proposition 5.2 upon which it is based, is proved in Appendix C. Combined with the properties of J_i of (5.25), it shows that the second term on the right of (5.27) belongs to $\mathfrak{R}_c^i(\omega, \omega')$. Thus, we obtain



(5.29)

The second term on the right has a minus-iϵ continuation into itself, past the normal threshold at $E(\omega, \omega') = M_i$. The first term is, therefore, the discontinuity of M^+ around this threshold.

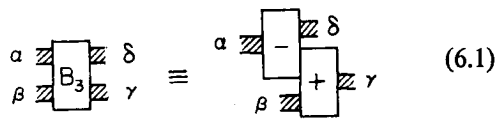
6. ANALYSIS OF THE SECOND TERM

A. General Result

In this section the second term on the right of (5.12) is further analyzed. That is, various bubble diagrams included in this term, but belonging to $\mathfrak{R}_c(\omega, \omega')$, are identified and separated out.

A generalization of Proposition 5.2 is needed.

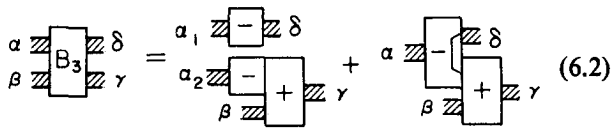
Proposition 6.1: The box diagram



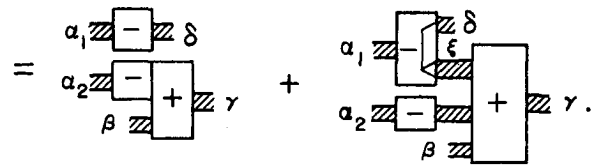
(6.1)

is equivalent to a box diagram B_3' such that, for any $D' \supset\subset B_3'$, the only simple positive $(\alpha, \beta + \gamma + \delta)$ cut set C containing no line of δ is the set α.

Proof: Equations (2.3′) and (2.4a′) give



(6.2)

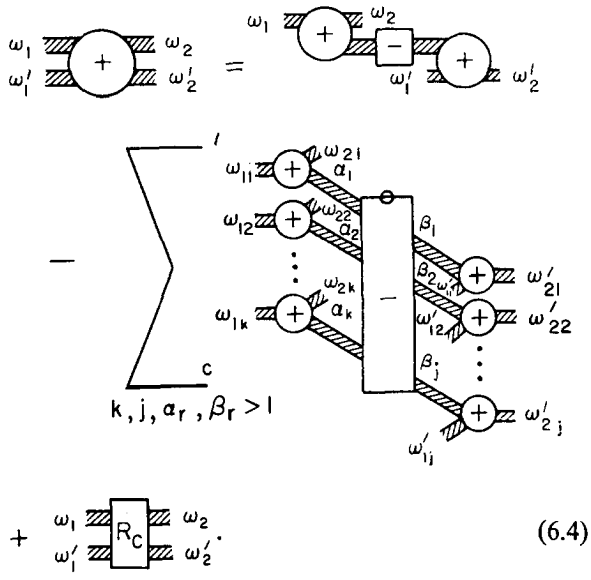


(6.3)

Since no line of δ can be in the cut set C, every line of α_1 in the first term of (6.3) evidently must be. (By definition, no line from any minus bubble can be included in this cut set.) Then the Corollary to 5.2 completes the proof for this first term. For the second term the definition of the box diagram (2.5′), together with the requirement that each bubble has both incoming and outgoing lines, ensures that each line of α_1 is connected to a minus bubble (including trivial bubbles) that is connected to a line of δ. Thus each line of α_1 must belong to C. Each point of $X^+(\xi)$ lies on a minus bubble that is connected to a line of δ, and it must, therefore, belong to $\bar{X}(C, \beta + \gamma + \delta)$. Then the Corollary to 5.2 completes the proof.

Corollary 6.1: Suppose that the B_3 of (6.1) is part of the box diagram B of (5.11). The replacement of the two B_1 parts of (6.3) by the B_1' of Proposition 5.2 converts B_3 to B_3' and converts B to an equivalent B' having the property that, for any simple positive (ω, ω') cut C of any $D' \supset\subset B'$, all the points of $X^-(\alpha)$ belong to $\bar{X}(C, \omega')$ if all points of $X^+(\beta)$, $X^-(\gamma)$, and $X(\delta)$ do. Let the part of D' supported by B_3' be called D_3'. Then all points of $D_3' - \alpha$ belong to $\bar{X}(C, \omega')$ if all points of $X^+(\beta)$, $X^-(\gamma)$, and $X(\delta)$ do.

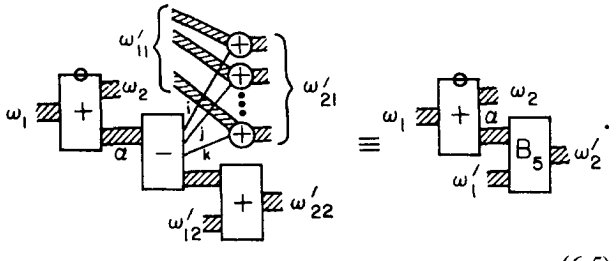
Proposition 6.2: Unitarity and the cluster properties imply that



(6.4)

The subscript c on $\sum_c$ indicates a sum over connected terms only. The condition $\alpha > 1$ means the set α must have more than one line. The prime on the summation symbol indicates that it includes only one of the $k!\,j!$ topologically equivalent terms coming from the relabeling of the sets of initial and final lines. (And, as throughout this paper, once the sets of external lines are fixed, the sums are only over topologically distinct diagrams.)

Proof: Let the right-hand plus box in the second term on the right of (5.12) be expanded according to (D11) of Appendix D, with the lines $i, j, \cdots, n$ of σ identified with the lines that go to the minus box. The terms $G^{+}_{ij\ldots k}$ of (D11) give structures of the form



(6.5)

One observes that no D supported by any term of this structure can have a simple positive (ω, ω') cut set C such that $\bar{X}(C, \omega')$ and $\bar{X}(C, \omega)$ are both connected diagrams. For if $\bar{X}(C, \omega')$ is connected, then the lines $i, j, \cdots, k$ must evidently belong to $\bar{X}(C, \omega')$. But then by Corollary 6.1 all the points of $X^{-}(\alpha)$ must belong to $\bar{X}(C, \omega')$. This means that $\bar{X}(C, \omega)$ cannot be a connected diagram. A similar argument applies to the left-hand circled plus box of (6.5). Thus all the contributions associated with the $G^{+}_{ij\ldots k}$ terms of (D11) give terms that belong to $\mathfrak{R}_c(\omega, \omega')$. The remaining term becomes the second term on the right of (6.4), after the contribution coming from the connected part of the minus box is shifted to R_c. This completes the proof of Proposition 6.2.

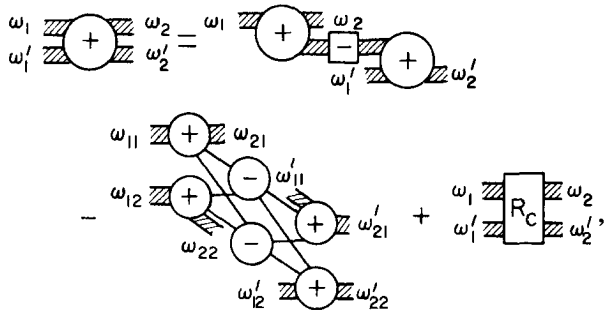
B. Special Cases

Case 1: $E(\omega, \omega')$ below the 4-particle threshold.

In this case the conditions on the second term on the right of (6.4) cannot be met. This proves (5.15).

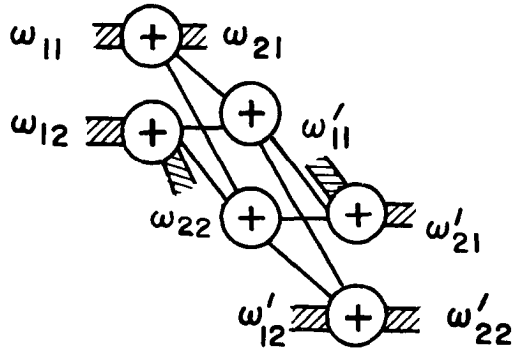
Case 2: $E(\omega, \omega')$ below the 5-particle threshold E_5.

It is shown in Appendix E that for this case (6.4) gives



$$E(\omega, \omega') < E_5. \tag{6.6}$$

The first two terms on the right contain all the singularities associated with positive-α Landau diagrams that can be contracted to any D^{+}_{n} $(\omega' \to \omega)$. The positive-α double-cross diagram, obtained by shrinking the bubbles of the second term on the right to points and assigning plus signs to the lines, is such a diagram. It is readily confirmed that the positive-α contributions to the double-cross diagram from the sum of the first two terms in (6.6) are equal to the positive-α contributions from



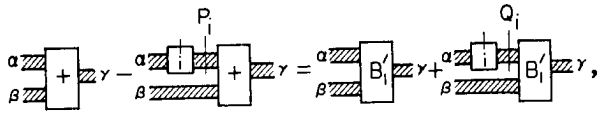
(6.7)

The significance of the results obtained in this section is this: The function represented by R_c in (6.4)

or (6.6) must, by virtue of the third structure theorem, continue into itself in a well-defined way around any Landau surface $L(D)$ corresponding to a pure-α D that contracts to a $\omega' \to \omega$ normal-threshold diagram. Moreover, for such cases this well-defined continuation is the minus-$i\epsilon$ (i.e., nonphysical) continuation. Thus, the discontinuity corresponding to any path $P'P''$ that encounters only singularities of these types in R_c is precisely the first two terms on the right of (6.4) or (6.6).

7. SHEET CONVERTERS

Below the lowest 4-particle threshold E_4^α in the channel $\beta + \gamma \to \alpha$, one obtains from (5.18), (5.10), and Proposition 5.2 the result



(7.1)

where B_1' is the right side of (B.4) and satisfies the following proposition.

Proposition 7.1: Let $\hat{\alpha}$ and $\hat{\beta}$ be sets that satisfy $\hat{\alpha} + \hat{\beta} = \alpha + \beta$, $\hat{\alpha} \subset \alpha$ and $\hat{\alpha} \cap \hat{\beta} = 0$. The only simple positive $(\hat{\alpha}, \hat{\beta} + \gamma)$ cut set C of any $D' \supset\subset B_1'$ is $\hat{\alpha}$. This proposition is a trivial extension of Proposition 5.2.

By virtue of Proposition 7.1 and the third structure theorem, we know that the function represented by the connected part of B_1' has a minus-$i\epsilon$ continuation into itself around any singularity associated with any diagram $D' \supset\subset B_1'$ that can be contracted to any $\hat{\beta} + \gamma \to \hat{\alpha}$ normal threshold diagram. Using the properties of (5.25) and B_1', one finds a similar result for the second term on the right of (7.1): The connected part of the second term on the right of (7.1) has a minus-$i\epsilon$ continuation into itself around any singularity in the interval

$$M_i \leq E(\alpha, \beta + \gamma) < E_4^\alpha$$

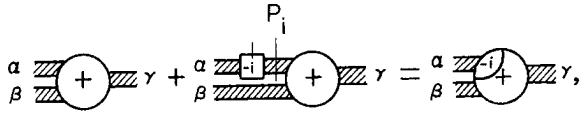
associated with any diagram $D' \supset\subset B_1'$ that can be contracted to any $\hat{\beta} + \gamma \to \hat{\alpha}$ normal-threshold diagram. Let the set of singularities just described be called $S_i(\hat{\alpha}, \hat{\beta} + \gamma)$.

Since the connected parts of both terms on the right of Eq. (7.1) continue into themselves via a minus-$i\epsilon$ rule around all singularities of $S_i(\hat{\alpha}, \hat{\beta} + \gamma)$, and since (7.1) holds identically throughout the physical region, the connected part of the left side must continue into itself via the same rule. If M_i is chosen greater than the physical threshold in the $\beta + \gamma \to \alpha$ channel, so that the physical region includes points where

$$E(\alpha, \beta + \gamma) < M_i,$$

then the connected part of either side of (7.1) can be identified as the continuation of the scattering function from physical points $E(\alpha, \beta + \gamma) < M_i$ to points in $E(\alpha, \beta + \gamma) > M_i$ lying underneath the cuts associated with the singularities of $S_i(\hat{\alpha}, \hat{\beta} + \gamma)$. Continuability past other singularities is not guaranteed, however.

The result just obtained is represented by the equation



(7.2)

where the right side represents the continuation of the scattering function to the underside of the cuts associated with the singularities of $S_i(\hat{\alpha}, \hat{\beta} + \gamma)$. {Some terms on the left of (7.1) are transferred to the right by the methods of Appendix C [see (C3)] in order to get (7.2). The symbol $-i$ in the right-hand term is simply i in Refs. 1, 2, and 5.}

The restriction that M_i be above the lowest physical threshold in the channel $\beta + \gamma \to \alpha$ means that the lines α of (7.2) can be considered cut by a Q_i bar. One would also like to have this equation with a P_i bar on these lines, since the formula could then be inserted into the discontinuity formula (5.29).

To obtain this result, one must use the pole-factorization theorem[4] in (5.26). We do not pursue the matter here, but only remark that the formulas in terms of the i boxes are the more useful ones anyway. For the expression in terms of functions on other sheets introduces, in effect, new unknown functions. And there is a different new unknown function for each choice of the "other" external lines [for example, those in the sets β and γ of (7.2)]. The i-box formula gives the discontinuity directly in terms of physical functions alone, and the i box is independent of the "other" lines.

8. ANALYTICITY PROPERTIES

The phase-space factors in unitarity equations have singularities. Corresponding singularities must appear in some scattering function. These latter singularities can combine with phase-space singularities to yield still other singularities of scattering functions, and so on. It has been shown[7] that all singularities generated in this way by unitarity must lie on Landau surfaces. The first part of our (maximal) analyticity assumption

is, accordingly, that all physical-region singularities of scattering functions lie on Landau surfaces.

Physical-region singularities lying on Landau surfaces are physically interpretable in terms of the notion that momentum–energy is transferred over macroscopic distances only by physical particles. In particular, if the singularities of scattering functions are confined to Landau surfaces, then transition amplitudes can be shown to fall off faster than any power of a scaling parameter τ, unless the trajectory regions defined by the wavepackets of the initial particles can be connected to those of the final particles by a network of physical particles.[8] In the limit of large τ, the distances involved become infinite. If one invokes the macroscopic causality requirement that the transition amplitudes fall off fast unless all the particles of the network move *forward* in time, then the singularities are confined to the positive-α branches of the Landau surfaces.[8] Moreover, the scattering functions on the two sides of these surfaces are analytically connected by a path that moves into a certain well-defined upper half-plane.[8] This rule of continuation is called continuation via a plus-$i\epsilon$ rule. Continuation through the opposite half-plane is called continuation via a minus-$i\epsilon$ rule.

The work of Ref. 8 establishes also that, if several Landau surfaces intersect, then the rules of continuation are compatible in the following sense: If the various intersecting Landau surfaces are all associated with diagrams that are contractions of any one *single* parent diagram, then the intersection of the various upper half-planes is nonempty and, hence, a region of continuation exists; if the diagrams corresponding to one subset of the surfaces intersecting at point $\bar{K}$ have a common parent, but no two that include one of the remaining diagrams have a common parent, then in a neighborhood of $\bar{K}$ the scattering function can be decomposed into a sum of functions, each having only certain of the singularities and having a well-defined rule of continuation past these singularities.

The above results provide a rule for continuation of scattering functions around all known combinations of intersecting singularity surfaces. To include any possible others, we assume that singularities that are "unrelated at $\bar{K}$" are "independent at $\bar{K}$." Singularities "unrelated at $\bar{K}$" are singularities corresponding to diagrams that have no common parent whose surface contains $\bar{K}$. Singularities "independent at $\bar{K}$" are singularities at $\bar{K}$ that can be separated into different terms of an expansion of the scattering function. The assumption that unrelated singularities are independent ensures that there is a well-defined rule of continuation past all combinations of intersecting surfaces of singularities of the scattering function. This rule is called the "general $i\epsilon$ rule." It ensures that an integral over a scattering function can be defined locally as a sum of contour integrals that detour around singularities in a manner determined by the plus-$i\epsilon$ rules for the individual singularity surfaces.

The general $i\epsilon$ rule is the hypothesis of three "structure theorems" proved in Ref. 2. These theorems provide the analytic basis of the present work. The first structure theorem says that the function corresponding to a connected bubble diagram B can have singularities only on the Landau surfaces corresponding to Landau diagrams $D \supset\subset B$. A diagram $D \supset\subset B$ is a diagram that can be obtained by inserting connected Landau diagrams for the bubbles of B and then contracting some (or no) lines. The second structure theorem says that these diagrams $D \supset\subset B$ can be further restricted by demanding that the α's on the lines of $D \supset\subset B$ that are interior lines of plus or minus bubbles of B have plus or minus signs, respectively. The third structure theorem gives the rule for continuation around the singularity associated with a given $D \supset\subset B$. It says that if $\mathfrak{L}[D]$ is the only Landau surface passing through a point $\bar{K}$, and if the momenta p_j and Feynman α_j of the internal lines of the corresponding Landau diagram are given uniquely by continuous functions $\bar{\alpha}_j(K)$ and $\bar{p}_j(K)$ for points on $\mathfrak{L}[D]$ near $\bar{K}$, then the function represented by B continues into itself when continued around $\mathfrak{L}\,[D \supset\subset B]$ near $\bar{K}$ by a path in the upper half-plane of the variable

$$\sigma_{\bar{K}}(K) \equiv \sigma(K; \bar{K}) \equiv \sum_j \bar{\alpha}_j(\bar{K})\bar{p}_j(\bar{K}) \cdot p_j(K). \quad (8.1)$$

The sum is over all internal lines, and $p_j(K)$ is any set of p_j consistent with the conservation laws at K. [A number of useful equivalent expressions for $\sigma(K; \bar{K})$ are given in Ref. 2.] The signs of the $\alpha_i(\bar{K})$ in (8.1) are fixed according to the rule given above: α's from lines lying inside plus (minus) bubbles are positive (negative). By virtue of the third structure theorem, the function represented by any bubble diagram B always has a well-defined continuation into itself past the singularity surfaces near a point P, unless there are several $D \supset\subset B$ with surfaces $\mathfrak{L}[D]$ that intersect at P and have incompatible rules for continuation.

The position of a Landau surface corresponding to a given D is not changed if the signs of all its α's are reversed. But the rule for continuation past the corresponding singularity is reversed. Thus, if both these D are supported by some B, then the third structure theorem fails to provide any (single) path of continuation for this function. This is precisely what happens at a threshold: The singularity surface corresponds to

two $D \supset\subset B$ that specify opposite rules for continuation. Accordingly, the function represented by B on one side of the threshold is not analytically connected to the function represented by B on the other side.

It should be emphasized that the contours of integration in the definitions of bubble-diagram functions are always fixed to be real, except for the infinitesimal distortions away from the singularities of the scattering functions of the integrand: One never distorts (by finite amounts) any contours, but always uses the originally defined (almost) real contours. Thus, when we say that a function corresponding to a bubble diagram B continues into itself, we mean the function defined by B with real undistorted contours of integration continues into the function defined by B with real undistorted contours, always excepting, of course, the infinitesimal distortions required by the definitions of scattering functions.

Our procedure has been to derive universal physical-region identities of the form $M^+ = T + R$, where M^+ is the physical scattering function, T is a threshold term that vanishes below the threshold t corresponding to a certain positive-α normal-threshold diagram D^+, and R is a function that has no singularity corresponding to any Landau diagram that can be contracted to D^+. Let D^- be the diagram obtained from D^+ by reversing the signs of all the α's. The singularity surface corresponding to D^- also lies at the threshold t, but the continuation past it is via the minus-$i\epsilon$ rule. Thus, if the only singularities of R near some point P on t are ones corresponding to D^+ or D^-, then R must have a minus-$i\epsilon$ continuation into itself past t near P, since the construction of R rules out singularities corresponding to D^+.

Because T vanishes below threshold, R is equal to M^+ there. Thus, R is an explicit expression in terms of physical amplitudes (i.e., scattering functions at physical points) of a function that equals M^+ below t near P, but has a minus-$i\epsilon$ continuation into itself past t near P. Since M^+ has a plus-$i\epsilon$ continuation into itself past t, the discontinuity around t near P is just T.

In the argument just given, it was assumed that each singularity of R near P corresponds to one of the threshold diagrams D^+ or D^-. The argument can be extended easily, however, to the case in which each singularity of R near P corresponds, merely, to some pure-(positive or negative)-α diagram that contracts to D^+ and D^-. Those contracting to D^+ are ruled out by the construction of R, while those that contract to D^- must be pure-negative-α diagrams. Therefore, R again continues into itself around t near P via the minus-$i\epsilon$ rule. Since only a finite number of pure-α singularity surfaces enter any bounded region,[9] the point P can be taken to lie on no pure-α surface other than t. Then, the plus- and minus-$i\epsilon$ continuations are simply into the upper and lower half E $(\omega' \to \omega)$ plane, respectively.

The above arguments cover only singularities corresponding to pure-α diagrams that contract to D^+ or D^-. However, for almost every point P on t, the contraction condition can be ignored, because, for almost every P on t, there is a neighborhood $N(P)$ of P such that each pure-α surface that intersects $N(P)$ corresponds to a diagram that contracts to D^+ or D^-. This follows from the general theory of pure-α surfaces developed in Ref. 8. That theory tells us that, apart from a set of points $\mathcal{M}_0$ (of zero measure on t) where certain external lines are parallel, the set of points lying on positive-α surfaces are the union of a set of (codimension 1) analytic manifolds, only a finite number of which pass through any bounded region. Thus, for almost every point P on t, there is a neighborhood $N(P)$ of P that intersects no Landau surface except those that coincide with t. The general theory also tells us that the normal to the surface at any point P, lying on just one surface, uniquely determines the positions of the external lines of the geometric Landau diagram $\bar{D}(P)$ that generates this point P. This means that any pure-α diagram, whose singularity surface contains a P lying on no Landau surface except t, must be such that all the lines ω intersect at one point and all the lines ω' intersect at another point. But any pure-α diagram of this kind can be contracted to D^+ or D^-. Thus, for almost all points P on t, R has minus-$i\epsilon$ continuation into itself around t near P, provided R has no mixed-α singularities passing near P.

A theory of mixed-α Landau surfaces analogous to the theory of pure-α Landau surfaces developed in Ref. 8 is not available. It seems likely, however, that almost every point P on t will have a neighborhood that contains no singularities except ones corresponding to diagrams that contract to a normal or pseudonormal $(\omega' \to \omega)$ threshold diagram.[10] Let us assume this is true.

It also seems likely that, in equations derived from unitarity and cluster properties alone, there will be no systematic cancellations between pure-α and mixed-α singularities. That is, if one side of such an equation has only pure-α singularities in some neighborhood $N(P)$, then so should the other side. In particular, we do not expect singularities associated with normal-threshold diagrams to be canceled by singularities associated with pseudonormal-threshold diagrams, since these will be moved relative to each other by small variations of the masses. The absence of such

cancellations will also be assumed. (These assumptions are either implicit or explicit in all derivations of discontinuities from unitarity.)

These two assumptions allow us to conclude that R has no mixed-α singularities in a neighborhood $N(P)$ of almost any point P on t. Energy conservation precludes the possibility that T has any pseudonormal-threshold singularities near t. Then, since $M^+ - T$ has only pure-α singularities in a neighborhood $N(P)$ of almost every P of t, so must R.

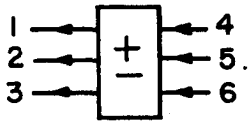
A more extensive study of the mixed-α singularities is needed, but that is a subject in itself.

APPENDIX A: PROOF OF DECOMPOSITION RULES

Every contribution to the left side of (2.3) that is not contained in the first term on the right is a column of bubbles, none of which is connected to both α_1 and α_2. Any such term occurs as a contribution to the second term on the right. No two different contributions to the left can occur as the same contribution to the second term on the right. Thus, the left side is contained in the right. Every contribution to the second term on the right occurs as a contribution to the left. No two different contributions to the second term on the right can occur as the same contribution to the left. No contribution to the second term on the right is contained in the subset of those contributions to the left that constitute the first term on the right. Thus, the right side is contained in the left side. Therefore, the two sides of (2.3) are identical sets of diagrams.

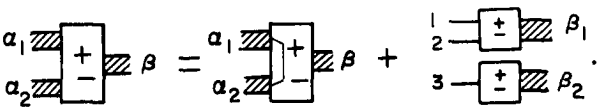
Each contribution to the left side of (2.4a) is contained in one and only one term on the right. This term is one in which α_{12} is the subset of α_1 consisting of lines connected to bubbles that are connected to α_2 and in which β_2 is the subset of β consisting of lines connected to bubbles connected to $\alpha_{12} \cup \alpha_2$. A given contribution to the left occurs as precisely one contribution to this unique term in which it appears. No two different contributions to the left occur as the same contribution on the right, and every contribution on the right occurs at least once on the left. Thus, the two sides of (2.4a) are identical. A completely similar argument proves (2.4b).

An illustration of the decomposition formulas is provided by considering the $3 \to 3$ box



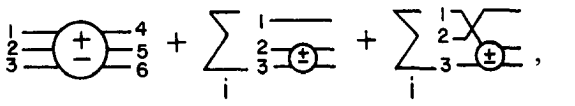
(A1)

For the case $\alpha_1 = \{1, 2\}$, $\alpha_2 = \{3\}$, and $\beta = \{4, 5, 6\}$, Eq. (2.3) reads



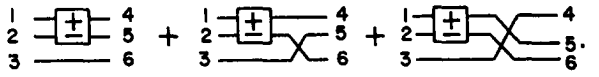
(A2)

The first term on the right of (A2) is then by definition the sum of bubble diagrams



(A3)

where the sum sign labeled "i" has the same significance as in I (see the first paragraph of Sec. 4 of I). The second term on the right of (A2) is



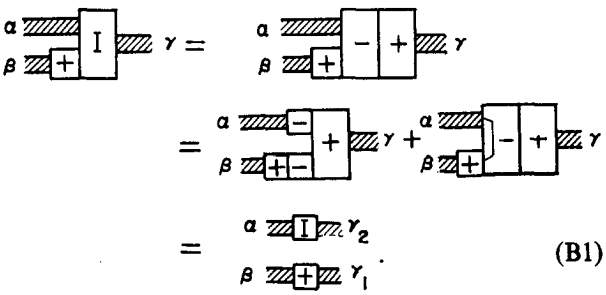
(A4)

The expressions (A3) and (A4) add up to the usual cluster expansion of the $3 \to 3$ box [Eq. (4.4) of I], as we see by substituting the expansion of the $2 \to 2$ box [Eq. (A8) of I] into (A4).

APPENDIX B: ITERATION OF CERTAIN BOX DIAGRAMS

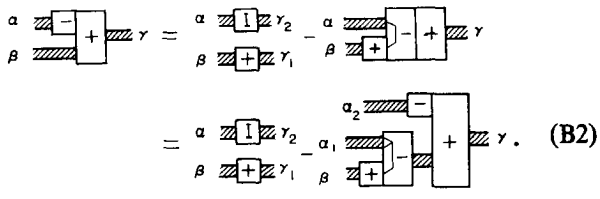
In this appendix, the box diagram B_1 defined by (5.10) is converted to an equivalent box diagram B_1' used in the proof of Proposition 5.2.

Equations (2.3) and (2.12) give



(B1)

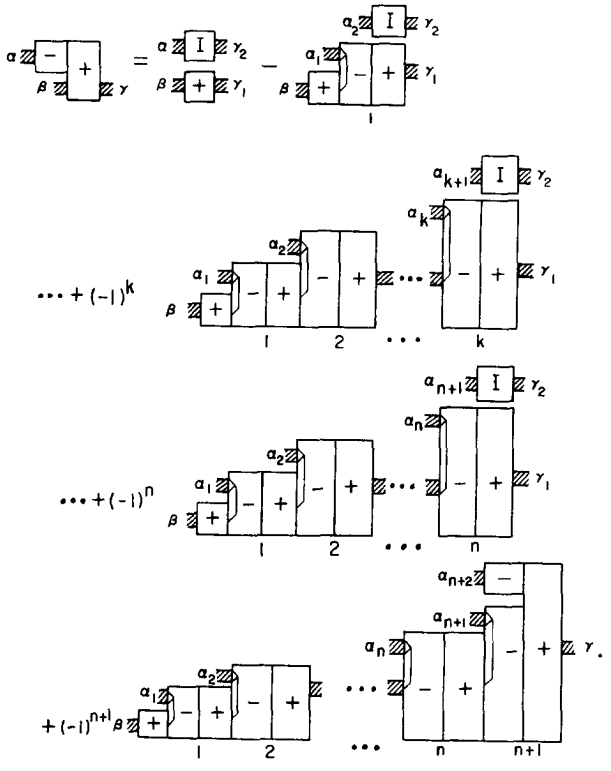
Use of unitarity and then (2.4a) converts this to



(B2)

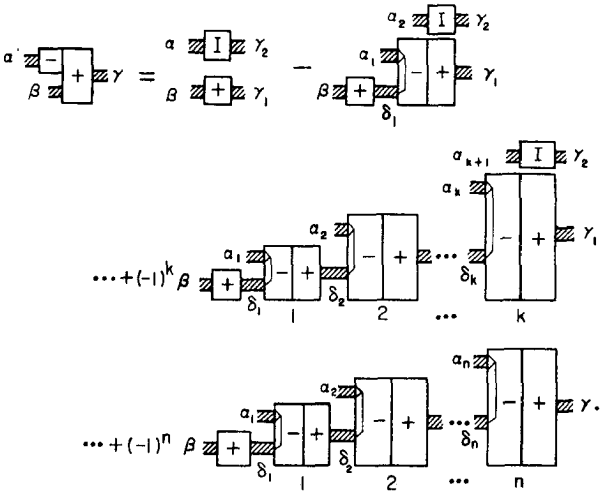
This equation can be iterated by replacing the last

factor of the last term by the entire right side. An n-fold iteration gives



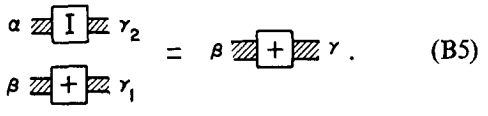
(B3)

Let the number of iterations n be at least equal to the number of lines $n(\alpha)$ of the set α. Then, the last term on the right side of (B3) is an empty set, since the definition of (2.5) implies that, if any of the sets $\alpha_1, \alpha_2, \cdots, \alpha_{n+1}$ is empty, then no diagram satisfying the required conditions exists. Thus, for any $n \geq n(\alpha)$, we have



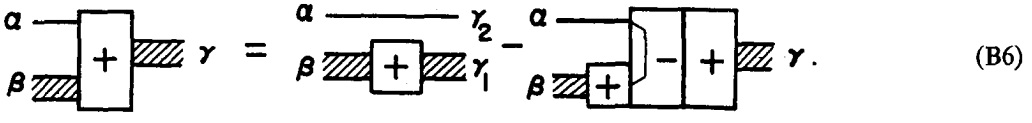
(B4)

If the set α is empty, then all terms of the right side of (B4) except the first are empty. Thus, the right side of (B4) is, in this case,



(B5)

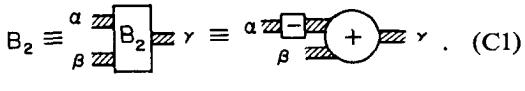
Equation (B4) is therefore a trivial identity for $n(\alpha) = 0$. If $n(\alpha) = 1$, Eq. (B4) is



(B6)

APPENDIX C: PROPOSITION 5.2C

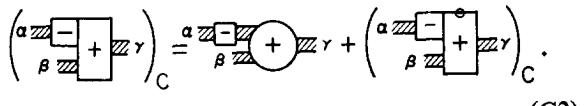
Consider the diagram



(C1)

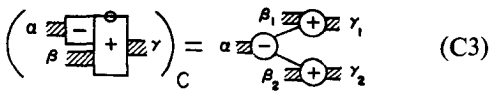
If the center-of-mass energy of the set α is below the 4-particle threshold in the channel $\beta + \gamma \to \alpha$, then B_2 is equivalent to a bubble diagram B_2' such that for any $D' \supset\subset B_2'$ the only simple positive $(\alpha, \beta + \gamma)$ cut set C with connected $\bar{X}(C, \beta + \gamma)$ is α itself. Moreover $D' - \alpha$ lies in $\bar{X}(C, \beta + \gamma)$.

Proof: The arguments establishing Proposition 5.2 apply equally well to the connected part of B_1, which is



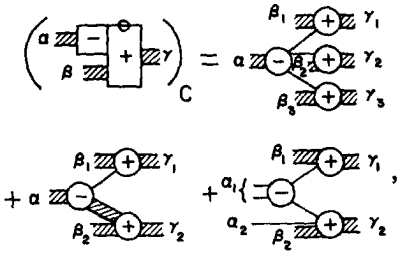
(C2)

If α is taken to be a single line, the second term on the right of (C2) drops out, and Proposition 5.2 proves the proposition. If the center-of-mass energy of the set α is less than the 3- or 4-particle threshold, respectively, then the second term on the right of (C2) is



(C3)

or



(C4)

respectively. (Topologically equivalent contributions to the right side are to be counted only once.) The only simple positive $(\alpha, \beta + \gamma)$ cut set C of any of the bubble diagrams of (C3) or (C4) such that $\bar{X}(C, \beta + \gamma)$ is connected is the set α. This follows at once from Definition 3.1, Proposition 3.1, and from the fact that no line of C can be a minus line. Proposition 5.2, applied to the left side of (C2), completes the proof.

Corollary: Suppose B_2 of (C1) is part of B of (5.11). Replacement of B_2 by the B_2' of the proposition converts B to B'. Let C be any simple positive (ω, ω') cut set of any $D' \supset\subset B'$. If $\bar{X}(C, \omega')$ is connected, then all points of $X^-(\alpha)$ belong to $\bar{X}(C, \omega')$ if all points of $X^+(\beta) \cup X^-(\gamma)$ belong to $X(\omega')$.

Proof: The above proof still applies, if the $X^+(\beta) \cup X^-(\gamma)$ belong to $X(\omega')$.

APPENDIX D: THE PRINCIPLE OF INCLUSION AND EXCLUSION AND AN EXPANSION ON UNCONNECTED LINES

Let A be a set such that each member of the set either does or does not have the property "i." The subset of A consisting of members that have the property "i" are designated by A_i. The subset of A consisting of members that do not have the property "i" are designated by A^i, $i = 1, 2, \cdots, n$. Then,

$$A = A_i + A^i. \tag{D1}$$

By a repeated application of (D1), we obtain

$$\begin{aligned} A &= A_1 + A^1 \\ &= A_{12} + A_1^2 + A_2^1 + A^{12}. \end{aligned} \tag{D2}$$

It follows from (D1) also that

$$A_1^2 = A_1 - A_{12}. \tag{D3}$$

By the definition of A_i and by (D1),

$$A_2^1 = (A - A_1)_2 = A_2 - A_{12}. \tag{D4}$$

Substituting (D3) and (D4) into (D2), we obtain

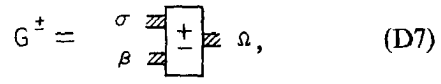
$$A = A_1 + A_2 - A_{12} + A^{12}. \tag{D5}$$

Equation (D5) can be generalized to the formula

$$\begin{aligned} A = \sum_i A_i - \sum_{i<j} A_{ij} + \cdots - (-1)^s \sum_{i<j<\cdots<k} A_{ij\cdots k} \\ + \cdots - (-1)^n A_{12\cdots n} + A^{12\cdots n}, \end{aligned} \tag{D6}$$

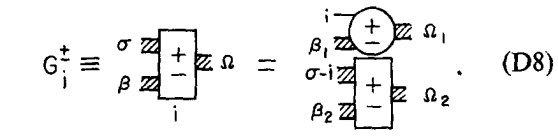
where s is the number of subscripts in $A_{ij}\ldots$ and where each summation runs from 1 to n. [For a proof of (D6), see Ref. 11.]

We consider now a plus or minus box of the form



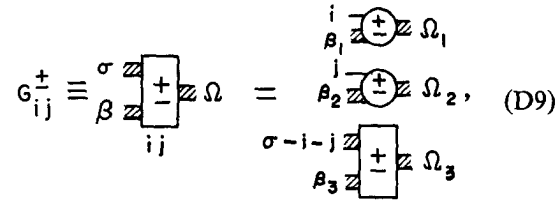
(D7)

where σ consists of precisely n lines. Let the property "i" be the property that line i of σ is connected in $G^\pm$ to no line of $\sigma - i \equiv \sigma - \{i\}$. Let $G_i^\pm$ be the subset of diagrams of $G^\pm$ with the property i. Then, $G_i^\pm$ has the form



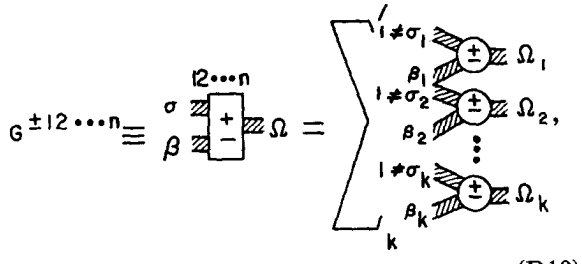
(D8)

This is because each term of $G_i^\pm$ occurs as one and only one term on the right, no two different terms of $G_i^\pm$ occur as a single term on the right, and each term on the right occurs at least once in $G_i^\pm$. Similarly, for $i \neq j$,



(D9)

etc. The term $G^{\pm 12\cdots n}$ is the term of $G^\pm$ such that each line i of σ is connected in G to some line of $\sigma - i$. Thus, it has the form



(D10)

where the prime on the summation symbol indicates that only one of the $k!$ topologically equivalent diagrams, obtained by reordering the bubbles, is to be counted and the condition $\sigma_r \neq 1$ means that the set σ_r has two or more lines. That is, the sum is over all ways of partitioning the set σ into sets $\sigma_r \neq 1$, the set β into the sets β_k, and the set Ω into the sets Ω_k, as specified by the summation convention of Sec. 2.

Let ${}^0G^\pm$ be $G^\pm$ minus its connected part. The expression given above for $G^\pm$ also applies to ${}^0G^\pm$, except

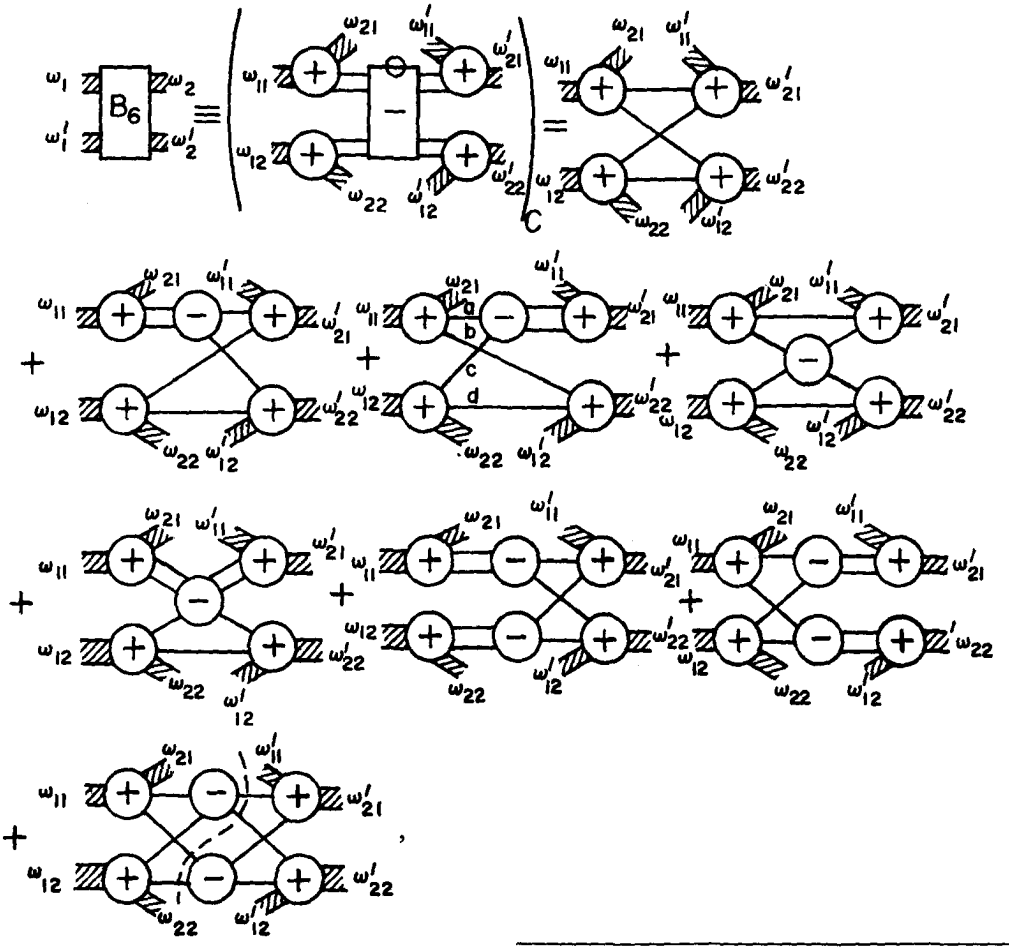
that the sum of columns of bubbles on the right of (D10) does not include a column consisting of just one bubble. Since ${}^0G^{\pm}_{ij\cdots k} = G^{\pm}_{ij\cdots k}$, the application of (D6) to the function ${}^0G^{\pm}$ yields

$$
\begin{aligned}
{}^0G^{\pm} = \sum_i G_i^{\pm} - \sum_{j<i} G_{ij}^{\pm} \\
+ \cdots + (-1)^s \sum_{i<j<\cdots<k} G_{ij\cdots k}^{\pm} \\
+ (-1)^n G_{12\cdots n}^{\pm} + {}^0G^{\pm 12\cdots n}, \quad \text{(D11)}
\end{aligned}
$$

where the $G^{\pm}_{ij\cdots k}$ are of the form indicated in (D9) and ${}^0G^{\pm\,12\cdots n}$ is given by (D10) with the restriction $k > 1$.

APPENDIX E: THE DOUBLE-CROSS TERM

In this appendix, we prove Eq. (6.6). If $E(\omega, \omega')$ is less than the 5-particle threshold, then the columns of plus bubbles in the second term on the right of (6.4) each consist of precisely two bubbles, and this term, denoted by B_6, is given by



(E1)

where topologically equivalent diagrams are to be counted only once.

Only the last bubble diagram on the right of (E1) can contain a simple positive (ω, ω') cut set C such that $\bar{X}(C, \omega)$ and $\bar{X}(C, \omega')$ are both connected. This follows directly from the properties of cut sets established by Proposition 3.1. In particular, any $D \supset\subset B$ is divided by C into three disjoint sets $\bar{X}(C, \omega)$, $\bar{X}(C, \omega')$, and C', the first two of which must be connected. Since C' contains no minus lines, the part of D contained in any minus bubble must belong to either $\bar{X}(C, \omega)$ or $\bar{X}(C, \omega')$. This precludes the possibility that the other one is connected, in the case of the fourth and fifth terms on the right of (E1). For the third term, the requirement that $\bar{X}(C, \omega)$ be connected implies that either (i) lines a and c belong to $\bar{X}(C, \omega)$ or (ii) lines b and d belong to $\bar{X}(C, \omega)$. The requirement that $\bar{X}(C, \omega')$ be connected implies that either (i') lines a and b belong to $\bar{X}(C, \omega')$, or (ii') lines c and d belong to $\bar{X}(C, \omega')$. These conditions are incompatible with the requirement $\bar{X}(C, \omega) \cap \bar{X}(C, \omega') = 0$. Except for the last term, the remaining bubble diagrams of the right side of (E1) are ruled out in the same way. The last term does not belong to $\mathcal{R}_c(\omega, \omega')$; the lines intersected by the dotted curve are a simple positive (ω, ω') cut set.

* This work was supported by the U.S. Atomic Energy Commission.

[1] The first is J. Coster and H. P. Stapp, J. Math. Phys. **10**, 371 (1969). This paper is referred to as I.

[2] H. P. Stapp, J. Math. Phys. **9**, 1548 (1968). Essentially the same results were obtained by M. J. Bloxham, D. I. Olive, and J. C. Polkinghorne [J. Math. Phys. **10**, 494, 545, 553 (1969)], though there without strict observance of the mass shell conditions. A further discussion of the structure theorems and their role in the present work is given in Sec. 8.

[3] H. P. Stapp, "Lectures on *S*-Matrix Theory," Matscience, Madras, India, 1964. See also "Lectures on the Analytic Structure of Multiparticle Scattering Amplitudes," ICTP 65/17, Trieste, 1965, and D. I. Olive, Nuovo Cimento **37**, 1422 (1965).

[4] D. I. Olive, Phys. Rev. **135**, B745 (1964); J. B. Boyling, Nuovo Cimento **33**, 1356 (1964); H. P. Stapp, in *High Energy Physics and Elementary Particles* (International Atomic Energy Assoc., Vienna, 1965), p. 41, and Ref. 2.

[5] D. I. Olive, Nuovo Cimento **29**, 326 (1963). See also R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, 1966), and J. Gunson, J. Math. Phys. **6**, 827, 845, 852 (1965).

[6] One can either consider the set of bubble diagrams to be restricted by the conservation laws or, alternatively, allow the diagrams to be restricted only by strictly topological conditions, and note that the corresponding functions will vanish if the conservation laws are not satisfied. It is in the former sense that equations such as (5.1) are strictly true as equivalence relations among diagrams.

[7] H. P. Stapp, Phys. Rev. **125**, 2139 (1962); J. C. Polkinghorne, Nuovo Cimento **25**, 901 (1962).

[8] C. Chandler and H. P. Stapp, J. Math. Phys. **10**, 826 (1969).

[9] H. P. Stapp, J. Math. Phys. **8**, 1606 (1967).

[10] D. Branson, Nuovo Cimento **44**, 1081 (1966).

[11] H. P. Stapp, ICTP 65/17, Ref. 3. The result is well known to combinatorial analysts. See John Riordan, *An Introduction to Combinatorial Analysis* (Wiley, New York, 1958), p. 50.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Limitable Dynamical Groups in Quantum Mechanics. II. A Model Including Arbitrary Spin

H. D. DOEBNER AND O. MELSHEIMER
Institut für Theoretische Physik, Universität Marburg, Germany

(Received 6 September 1967)

In the framework of limitable dynamical groups, a group

$$D_s^t = D^t \otimes SO(N)^I, \quad D^t = G_E^0(N) \textcircled{\times} Sp(2N, R),$$

is discussed. It is shown that the physical representation $U(D_s^t)$ describes interacting nonrelativistic particles of arbitrary spin in N dimensions, with linear combinations of second-order polynomials in momentum and position operators P_i, Q_i with $i = 1, \cdots, N$, and spin–flip operators as Hamiltonians. $G_E^0(N)$ is the central extension of the inhomogeneous, pure Galilei group, and $SO(N)^I$ a rotation group in N dimensions. $Sp(2N, R)$ is the noncompact form of the symplectic group. It is proven that there exists no irreducible unitary representation of D^t describing particles with spin $s > 0$. Hence it was necessary to enlarge D^t to D_s^t. The physical free-particle group $D^0 \subset D_s^t$ possesses projection parts in $Sp(2N, R)$ as well as in $SO(N)^I$. To calculate them, we decompose the physical representation $U(D^0)$ by group-theoretical methods such that the angular momentum splits into a spin and an orbital part. D^0 is isomorphic to the central extension of the inhomogeneous Galilei group. The physical representation $U(D_s^t)$ is calculated by the postulate that $U(D_s^t)$ and $U(D^0)$ can be limited into each other, and that mass and spin conservation hold. The limitation is realized as a group contraction leading to a nonfaithful representation of the contracted group. We prove that $U(D_s^t)$ is uniquely determined by mass and spin of the free particle which is obtained if the interaction in D_s^t is turned off. Furthermore, the generators are identified as functions of P_i, Q_i, and spin–flip operators. Hence the limitation postulate leads to the physical representation of the abstract group D_s^t and to the identification of the described physical system.

INTRODUCTION

1. The complete and pure group-theoretical description of interacting systems by limitable dynamical groups proposed in the previous paper[1] (hereafter referred to as I) is applied to nonrelativistic particles with arbitrary spin. Our approach is based on the following observation. Suppose that $U_{Ph}(D_s^t)$ is the unitary, irreducible, physical representation of a group D_s^t in the sense that interacting spin particles are completely described by $U_{Ph}(D_s^t)$. Denote the known dynamical group of free spin particles and its physical representations by D_s^0 and $U_{Ph}(D_s^0)$, respectively. Then it is plausible to require that the interaction contained in D_s^t can be turned off, which implies that $U_{Ph}(D_s^t)$ can be limited into $U_{Ph}(D_s^0)$,

$$U_{Ph}(D_s^t) \rightarrow U_{Ph}(D_s^0),$$

such that mass (m) and spin (s) are constant during the limitation. The reverse process, i.e., the extension of $U_{Ph}(D_s^0)$ to $U_{Ph}(D_s^t)$,

$$U_{Ph}(D_s^t) \leftarrow U_{Ph}(D_s^0),$$

* This work was supported by the U.S. Atomic Energy Commission.

[1] The first is J. Coster and H. P. Stapp, J. Math. Phys. **10**, 371 (1969). This paper is referred to as I.

[2] H. P. Stapp, J. Math. Phys. **9**, 1548 (1968). Essentially the same results were obtained by M. J. Bloxham, D. I. Olive, and J. C. Polkinghorne [J. Math. Phys. **10**, 494, 545, 553 (1969)], though there without strict observance of the mass shell conditions. A further discussion of the structure theorems and their role in the present work is given in Sec. 8.

[3] H. P. Stapp, "Lectures on *S*-Matrix Theory," Matscience, Madras, India, 1964. See also "Lectures on the Analytic Structure of Multiparticle Scattering Amplitudes," ICTP 65/17, Trieste, 1965, and D. I. Olive, Nuovo Cimento **37**, 1422 (1965).

[4] D. I. Olive, Phys. Rev. **135**, B745 (1964); J. B. Boyling, Nuovo Cimento **33**, 1356 (1964); H. P. Stapp, in *High Energy Physics and Elementary Particles* (International Atomic Energy Assoc., Vienna, 1965), p. 41, and Ref. 2.

[5] D. I. Olive, Nuovo Cimento **29**, 326 (1963). See also R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, 1966), and J. Gunson, J. Math. Phys. **6**, 827, 845, 852 (1965).

[6] One can either consider the set of bubble diagrams to be restricted by the conservation laws or, alternatively, allow the diagrams to be restricted only by strictly topological conditions, and note that the corresponding functions will vanish if the conservation laws are not satisfied. It is in the former sense that equations such as (5.1) are strictly true as equivalence relations among diagrams.

[7] H. P. Stapp, Phys. Rev. **125**, 2139 (1962); J. C. Polkinghorne, Nuovo Cimento **25**, 901 (1962).

[8] C. Chandler and H. P. Stapp, J. Math. Phys. **10**, 826 (1969).

[9] H. P. Stapp, J. Math. Phys. **8**, 1606 (1967).

[10] D. Branson, Nuovo Cimento **44**, 1081 (1966).

[11] H. P. Stapp, ICTP 65/17, Ref. 3. The result is well known to combinatorial analysts. See John Riordan, *An Introduction to Combinatorial Analysis* (Wiley, New York, 1958), p. 50.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Limitable Dynamical Groups in Quantum Mechanics. II. A Model Including Arbitrary Spin

H. D. DOEBNER AND O. MELSHEIMER
Institut für Theoretische Physik, Universität Marburg, Germany

(Received 6 September 1967)

In the framework of limitable dynamical groups, a group

$$D_s^t = D^t \otimes SO(N)^I, \quad D^t = G_E^0(N) \textcircled{\times} Sp(2N, R),$$

is discussed. It is shown that the physical representation $U(D_s^t)$ describes interacting nonrelativistic particles of arbitrary spin in N dimensions, with linear combinations of second-order polynomials in momentum and position operators P_i, Q_i with $i = 1, \cdots, N$, and spin–flip operators as Hamiltonians. $G_E^0(N)$ is the central extension of the inhomogeneous, pure Galilei group, and $SO(N)^I$ a rotation group in N dimensions. $Sp(2N, R)$ is the noncompact form of the symplectic group. It is proven that there exists no irreducible unitary representation of D^t describing particles with spin $s > 0$. Hence it was necessary to enlarge D^t to D_s^t. The physical free-particle group $D^0 \subset D_s^t$ possesses projection parts in $Sp(2N, R)$ as well as in $SO(N)^I$. To calculate them, we decompose the physical representation $U(D^0)$ by group-theoretical methods such that the angular momentum splits into a spin and an orbital part. D^0 is isomorphic to the central extension of the inhomogeneous Galilei group. The physical representation $U(D_s^t)$ is calculated by the postulate that $U(D_s^t)$ and $U(D^0)$ can be limited into each other, and that mass and spin conservation hold. The limitation is realized as a group contraction leading to a nonfaithful representation of the contracted group. We prove that $U(D_s^t)$ is uniquely determined by mass and spin of the free particle which is obtained if the interaction in D_s^t is turned off. Furthermore, the generators are identified as functions of P_i, Q_i, and spin–flip operators. Hence the limitation postulate leads to the physical representation of the abstract group D_s^t and to the identification of the described physical system.

INTRODUCTION

1. The complete and pure group-theoretical description of interacting systems by limitable dynamical groups proposed in the previous paper[1] (hereafter referred to as I) is applied to nonrelativistic particles with arbitrary spin. Our approach is based on the following observation. Suppose that $U_{Ph}(D_s^t)$ is the unitary, irreducible, physical representation of a group D_s^t in the sense that interacting spin particles are completely described by $U_{Ph}(D_s^t)$. Denote the known dynamical group of free spin particles and its physical representations by D_s^0 and $U_{Ph}(D_s^0)$, respectively. Then it is plausible to require that the interaction contained in D_s^t can be turned off, which implies that $U_{Ph}(D_s^t)$ can be limited into $U_{Ph}(D_s^0)$,

$$U_{Ph}(D_s^t) \rightarrow U_{Ph}(D_s^0),$$

such that mass (m) and spin (s) are constant during the limitation. The reverse process, i.e., the extension of $U_{Ph}(D_s^0)$ to $U_{Ph}(D_s^t)$,

$$U_{Ph}(D_s^t) \leftarrow U_{Ph}(D_s^0),$$

must also be possible and compatible with the selection rules for m and s, and the result must be unique. Hence, starting with $U_{Ph}(D_s^0)$ and turning on some known interactions t (the so-called interaction type t), we obtain the physical representation of D_s^t. The corresponding Hamiltonians H^t are assumed to be elements of the Lie algebra of $U_{Ph}(D_s^t)$. It was shown that these limitations severely restrict the group as well as its physical representations.

2. The dynamical group D_s^0 of the free particle moving in $N \leq 3$ dimensions was given by the central extension $G_E(N)$ of the inhomogeneous Galilei group $G(N)$ in N dimensions. D_s^0 is independent of s; hence the spin index was dropped. Any irreducible representation $U(G_E(N))^{[m,V,s]}$ is labeled by three numbers (m, V, s), by the eigenvalue $i \cdot m$ of the center in the Lie algebra of $G_E(N)$, by the eigenvalue V of the Casimir operator

$$C_{H_0}^{(2)} = H_0 C + \tfrac{1}{2} \sum_{i=1}^{N} P_i^2,$$

and (for $N = 3$) by the eigenvalue of the second-order Casimir operator $\mathbf{L}^2$ of the little group $SO(3)$ of $G_E(3)$. The physical identification of m and s as mass and spin was straightforward. V corresponds to a physically inessential constant in the energy operator H_0; therefore, $V = 0$ is chosen and one has $U_{Ph}(G_E(3)) = U(G_E(3))^{[m,0,s]}$. For $N > 3$, a formal particle is defined via $U(G_E(N))^{[m,0,s]}$, with s being now the spin content.

3. A limitable dynamical group for spinless particles was constructed by embedding $G_E(N)$ in

$$D^t = G_E^0(N) \textcircled{\times} Sp(2N, R),$$

with $G_E^0(N)$ being the central extension of the pure, inhomogeneous Galilei group in N dimensions, and $Sp(2N, R)$ being the noncompact form of the sympletic group. By our limitation postulates, $U_{Ph}(D^t)$ is a Nelson extension of $U(G_E^0(N))^{[m]}$ and uniquely given by m and $s = 0$. The interaction type t of D^t, i.e., the Hamiltonians in the corresponding quantum-mechanical description, is identified as the set of second-order polynomials in P_i, Q_i, $i = 1, \cdots, N$.

4. We present now a similar model for particles with arbitrary spin. In Sec. 1, the free-particle group $G_E(N)$ and its physical representations $U(G_E(N))^{[m,0,s]}$ are discussed again. It is shown that a tensor product between $U(G_E(N))^{[m,0,0]}$ and an irreducible representation $U(SO(N))^{[s]}$ of the spin group [little group for $G_E(N)$] with spin content s is unitarily equivalent to $U(G_E(N))^{[m,0,s]}$. This decomposition is the group-theoretical analog of the quantum-mechanical procedure of solving in the first step, the free-particle problem without spin and attaching, in the second step, the spin to the particle. The same method of introducing the spin is applied to the dynamical group D^t. It is shown (Sec. 2A) that D^t possesses no representation limitable into an $s = 0$ representation of $G_E(N)$. Therefore, D^t is enlarged (Sec. 2B) to

$$D_s^t = (G_E^0(N) \textcircled{\times} Sp(2N, R)) \otimes SO(N)^I,$$

such that the rotation subgroup $SO(N)$ of the physical free-particle group D^0 has projection parts in $Sp(2N, R)$ as well as in $SO(N)^I$. D^0 is isomorphic to $G_E(N)$.

D_s^t is a limitable group. Its physical representation is calculated via a limitation between $U_{Ph}(D_s^t)$ and $U(G_E)^{[m,0,s]}$, and is shown to be unique. The limitation is realized by a group contraction. The identification of the interaction Hamiltonians H^t contained in $U_{Ph}(D_s^t)$ leads to a similar result for D^t. We find H^t as a linear combination of second-order polynomials in P_i, Q_i, $i = 1, \cdots, N$, and of a set of spin–flip operators. The application of the approach to relativistic systems is discussed (Sec. 2C).

1. SPIN DECOMPOSITION FOR THE DYNAMICAL GROUP OF THE FREE PARTICLE

The spin part in $U(G_E(N))^{[m,0,s]}$, $s > 0$, is separated such that the physical angular momentum splits into a spin and orbital part. The same result is obtained if $G_E(N)$ is replaced by

$$D^0(N) = (G_E^T(N) \textcircled{\times} SO(N)) \odot \rho[SO(N)],$$

($\odot$ is defined in Sec. 1A3) with $\rho[SO(N)]$ being the image of $SO(N) \subset G_E(N)$ under a fixed isomorphism ρ. D^0 is isomorphic to $G_E(N)$. The irreducible representations of $D^0(N)$ labeled by $[m, 0, s]$ are physical and $D^0(N)$ can be used as dynamical group for free particles.

A. Spin Separation in Quantum Mechanics

1. In $U(G_E(N))^{[m,0,s]}$, spin and orbital angular momentum of the free particle are described by the same group simultaneously. The angular-momentum operators are given as generators of the physical rotation group $SO(N)^{ph}$, i.e., of the group describing rotations of the coordinate frame as it was indicated in Sec. 1C2 of I. This unified description is satisfactory from the principal point of view. However, a generalization of our oscillator model to spin particles needs a formal separation of spin and orbital angular momentum, at least for free particles.

2. In the usual quantum-mechanical treatment, this splitting is a natural consequence. We recall briefly the method. Let $\mathfrak{H}_1$ be the Hilbert space of the spinless system. $\mathfrak{H}_1$ carries a reducible representation of $SO(3)$ with generators $d_{ij} = i(P_i Q_j - P_j Q_i)$, $i, j = 1, 2, 3$.

The spin is introduced by enlarging $\mathfrak{H}_1$ to $\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2$, with $\mathfrak{H}_2 = \mathfrak{H}^{[s]}$ being a $[d(s) = 2s + 1]$-dimensional irreducible representation space of $SO(3)$ with generators l_{ij}^I. (A product space $\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2$ of two Hilbert spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$ is defined as the closure of the finite linear span of all formal products $f_i \cdot \mu_\alpha$ of basis elements $f_i \in \mathfrak{H}_1$, $\mu_\alpha \in \mathfrak{H}_2$. The elements of $\mathfrak{H}$ are denoted by $|f, \mu\rangle$. The inner product is defined as $\langle f, \mu \,|\, f', \mu'\rangle = \langle f \,|\, f'\rangle\langle \mu \,|\, \mu'\rangle$, with $\langle f \,|\, f'\rangle$ and $\langle \mu, \mu'\rangle$ being the inner products in $\mathfrak{H}_1$ and $\mathfrak{H}_2$. Note that $|f_1, \mu\rangle + |f_2, \mu\rangle = |f_1 + f_2, \mu\rangle$ and $|f, \mu_1\rangle + |f, \mu_2\rangle = |f, \mu_1 + \mu_2\rangle$ holds. Operators A in $\mathfrak{H}$ acting as $A\,|f, \mu\rangle = |Af, A\mu\rangle$ are denoted by $A = A_1 \times A_2$, e.g., the unit operator by $I = I_1 \times I_2$.) $\mathfrak{H}$ is a Hilbert space over vector-valued functions with d components. Hence two different $SO(3)$ representations appear which are not yet related to each other, and a direct product of two $SO(3)$ groups $SO(3) \otimes SO(3)^I$ is represented in $\mathfrak{H}$:

$$U(SO(3) \otimes SO(3)^I) = U_1(SO(3)) \times U_2(SO(3)^I),$$

acting as $(f \in \mathfrak{H}_1,\ \mu \in \mathfrak{H}_2)$,

$$U(SO(3) \otimes SO(3)^I)|f, \mu\rangle = |(U_1(SO(3))f), (U_2(SO(3)^I\mu)\rangle.$$

The essential point now is to identify the physical rotations group $SO(3)^{ph}$ in $SO(3) \otimes SO(3)^I$ which coincides obviously neither with $SO(3)$ nor with $SO(3)^I$. Hence it lies crosswise as a subgroup in the direct product. It is suitable to define $SO(3)^{ph}$ via its Lie algebra

$$d_{ij} = l_{ij} + l_{ij}^I,$$

acting in $\mathfrak{H}$ as

$$d_{ij}\,|f, \mu\rangle = (l_{ij} \times I_2 + I_1 \times l_{ij}^I)\,|f, \mu\rangle.$$

3. To characterize $SO(3)^{ph}$ without using its Lie algebra, let $R \in SO(3)$ and $R^I \in SO(3)^I$ be parameterized by a set (R) and (R^I) of rotation angles. Then $SO(3) \otimes SO(3)^I$ contains elements (R, R^I). Consider now an isomorphism ρ mapping $SO(3)$ in $SO(3)^I$,

$$R \underset{\rho}{\to} \rho[R] \in SO(3)^I,$$

such that

$$(\rho[R]) = (R),$$

i.e., R and $\rho[R]$ are parameterized by the same set (R). Then $SO(3)^{ph}$ is obtained as

$$SO(3)^{ph} = SO(3) \circ \rho[SO(3)],$$

where $\circ$ means not the direct product, and with Lie algebra

$$d_{ij} = l_{ij} + \rho[l_{ij}], \quad \rho[l_{ij}] = l_{ij}^I.$$

The elements of $SO(3)^{ph}$ are given by $(R, R^I = R)$. (The above discussion contains a redefinition. A straightforward calculation leads to

$$SO(3) \otimes SO(3)^I \approx SO(3) \textcircled{\times} SO(3)^{ph} \approx SO(3)^I \textcircled{\times} SO(3)^{ph}.$$

Starting with $SO(3) \textcircled{\times} SO(3)^{ph}$, we see that that implies that a redefinition exists such that a direct product decomposition with factors isomorphic to $SO(N)$ is possible. This is a special case of more general redefinition theorems. Note that also representations of $SO(3)$ and $SO(3)^{ph}$ exist in $\mathfrak{H}$; however, they are not related to physical rotations).

B. Spin Separation in $U_{Ph}(G_E(N))$

1. An angular-momentum decomposition in

$$U(G_E)^{[m,0,s]}, \quad s > 0,$$

is also possible and is related to a specific group-theoretical property of $G_E(N)$. Let $G_E(N)$ be parameterized by $r = (\theta, R, \mathbf{u}, \mathbf{v}, \tau)$ $[r \in G_E(N)]$ with real numbers θ and τ, N-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$, and (R) being a set of rotation angles describing elements of $SO(N)$. $G_E(N)$ can be written as a semidirect product,

$$G_E(N) = G_E^T(N) \textcircled{\times} SO(N),$$

and also as

$$G_E(N) = G_1 \textcircled{\times} G_2,$$

with

$$G_E^T(N) = \{(\theta, 1, \mathbf{u}, \mathbf{v}, \tau)\}, \quad G_1 = \{(\theta, 1, \mathbf{u}, 0, \tau)\}, \quad G_2 = \{(0, R, 0, \mathbf{v}, 0)\}.$$

The Lie algebra $\mathsf{G}_E(N)$ is spanned by skew-symmetric generators $\{\mathsf{C}, d_{ij} = -d_{ji}, P_i, Q_i, H; i, j = 1, \cdots, N\}$ of the corresponding subgroups $(\theta, 1, 0, 0, 0)$, etc. Then the following decomposition holds:

Lemma 1: Let $U(G_E(N))^{[m,0,s]}$ and $U(SO(N))^{[s]}$ be irreducible unitary representations of $G_E(N)$ and $SO(N)$ in representation spaces $\mathfrak{H}^{[m,0,s]}$ and $\mathfrak{H}^{[s]}$.

(i) $\mathfrak{H} = \mathfrak{H}^{[m,0,s]}$ can be decomposed into

$$\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2, \quad \mathfrak{H}_1 = \mathfrak{H}^{[m,0,0]}, \quad \mathfrak{H}_2 = \mathfrak{H}^{[s]},$$

such that $U(G_E(N))^{[m,0,s]}$ is unitarily equivalent to

$$U(\theta, R, \mathbf{u}, \mathbf{v}, \tau)^{[m,0,s]} \tilde{U} U_1(\theta, R, \mathbf{u}, \mathbf{v}, \tau)^{[m,0,0]} \times U_2(R)^{[s]}$$

or

$$U(G_E(N))^{[m,0,s]} \tilde{U} U_1(G_E^T \textcircled{\times} SO(N))^{[m,0,0]} \times U_2(\rho[SO(N)])^{[s]},$$

i.e., to a tensor product between an irreducible unitary representation of $G_E(N)$ with $s = 0$ and an irreducible

unitary representation of its little group $SO(N)$ which is mapped by a fixed isomorphism ρ^{-1} into the $SO(N)$ subgroup of $G_E(N)$.

(ii) The Lie algebra $U(\mathsf{G}_E(N))^{[m,0,s]}$ in $\mathfrak{H}$ is of the form

$$\mathsf{C} = \mathsf{C}_1 \times I_2,$$

$$d_{ij} = l_{ij} \times I_2 + I_1 \times \rho[l_{ij}], \quad l_{ij} = (i(P_iQ_j - P_jQ_i))_1,$$

$$P_i = (P_i)_1 \times I_2, \quad Q_i = (Q_i)_1 \times I_2, \quad H = (H)_1 \times I_2,$$

whereas $\rho[l_{ij}]$, $i,j = 1, \cdots, N$ is the Lie algebra of $U(\rho[SO(N)])^{[s]}$. The statement is equivalent to the desired spin separation. $l_{ij} \times I_2$ is the orbital part and $I_1 \times \rho[l_{ij}]$ the spin part of the angular momentum.

$$I_1 \times U_2(\rho[SO(N)])^{[s]} \approx U(SO(N))^{[s]}$$

can be interpreted as the spin group which is extracted from $U(G_E)^{[m,0,s]}$ such that the remaining part contains an $s = 0$ representation of $G_E(N)$. It is worthwhile to note that a similar decomposition for physical representations of the Poincaré group does not exist.

2. A proof of Lemma 1 uses some details from the construction of $U(G_E(N))$. We refer to the discussion in Appendix B of I (see also Ref. 2). Consider $G_E(N) = G_1 (\times) G_2$. Denote by $\{\chi(g_1)\} = \hat{G}_1$ the character group of G_1, and by $C(m, V)$ equivalence classes or orbits of $\hat{G}_1$ in G_E with elements $\chi_\alpha(g_1)$, $\alpha = (m, \frac{1}{2}m^{-1}p^2 + V, \mathbf{p})$, which are uniquely labeled by an N-dimensional vector $\mathbf{p}$. The numbers $i \cdot m$ and V are, as above, eigenvalues of the center C in $\mathsf{G}_E(N)$ and of $\mathsf{C}_H^{(2)}$. The set of all little groups $G_l^{m,V} \subset G_2$; i.e., the set of all groups leaving invariant at least one character $\chi_\alpha(g_1)$ is given by

$$\{(0, R, 0, m^{-1}(\mathbf{p} - R\mathbf{p}), 0)\}.$$

The little groups are independent of the m, V, $\mathbf{p}$ in the sense that they are all isomorphic to $SO(N)$. Take $\mathbf{p}$ fixed. Then $\{(0, R, 0, m^{-1}(\mathbf{p} - R\mathbf{p}), 0)\}$ is the invariance group of $\chi_\alpha(g_1)$, $\alpha = (m, \frac{1}{2}m^{-1}p^2 + V, \mathbf{p})$. An irreducible representation $U(G_l^{m,V})^{[s]}$ in $\mathfrak{H}_2 = \mathfrak{H}^{[s]}$ with basis $\{|\mu\rangle, \mu = 1, \cdots, d(s)\}$ is given by matrices

$$L^{[s]}_{\mu\mu'}(R; \mathbf{p})$$

with fixed $\mathbf{p}$. $\chi_\alpha(g_1)$ is a representation of G_1. It is convenient for our application ($V = 0$) to use $\chi_{\alpha_0}(g_1)$ with $\alpha_0 = (m, 0, 0)$. We consider now the tensor product

$$\chi_{\alpha_0}(g_1) \times L^{[s]}_{\mu\mu'}(R; \mathbf{p})$$

and choose as the little group just the invariance group of $\chi_{\alpha_0}(g_1)$, i.e., $L^{[s]}_{\mu\mu'}(R; 0)$ abbreviated as $L^{s}_{\mu\mu'}(R)$. Then

$$\chi_{\alpha_0}(g_1) \times L^{s}_{\mu\mu'}(R)$$

is a representation of $G_1 (\times) G_l^{m,0}$ in $\mathfrak{H}_2$. We decompose G_E in cosets $G_E/(G_1 (\times) G_l^{m,0}) \subset G_2$ modulo $G_1 (\times) G_l^{m,0}$. Each of them is labeled by $\mathbf{p}$ and contains elements of the form $(\hat{R})$

$$g_{\mathbf{p}}(\hat{R}) = (0, \hat{R}, 0, -m^{-1}\hat{R}\mathbf{p}, 0) \in G_E/(G_1 (\times) G_l^{m,V}).$$

We take in each coset a representative element $g_{\mathbf{p}}$, i.e., an element with a suitable choice of $\hat{R}$.

The representative of the transformed coset $r \cdot g_{\mathbf{p}}(\hat{R})$ is denoted by $g_{r\mathbf{p}}$ with $r = (\theta, R, \mathbf{u}, \mathbf{v}, \tau) \in G_E$ and $r\mathbf{p} = R^{-1}(\mathbf{p} - m\mathbf{v})$. (In Appendix B of I, $r\mathbf{p}$ was denoted by $\mathbf{p}_r$.) Now, dropping the index g in Appendix B of I, the representation space $\mathfrak{H} = \mathfrak{H}^{[m,0,s]}$ of $U(G_E)$ is spanned by vector-valued-functions f with components f_μ, $\mu = 1, \cdots, d(s)$, defined over cosets, i.e., $f_\mu = f_\mu(g_{\mathbf{p}}) = f_\mu(\mathbf{p})$. Hence $f \in \mathfrak{H}$ can be written as $|f(p), \mu\rangle$ or

$$\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2, \quad \mathfrak{H}_2 = \mathfrak{H}^{[s]}.$$

$\mathfrak{H}_1$ is the space of square-integrable functions $f(\mathbf{p})$ and $\mathfrak{H}_2$ carries $U(G_l^{m,0})^{[s]}$. In an improper basis $|\mathbf{p}, \mu\rangle$ in $\mathfrak{H}$ with $\langle \mathbf{p}, \mu \,|\, \mathbf{p}', \mu'\rangle = \delta\mu\mu'\delta^N(\mathbf{p} - \mathbf{p}')$, matrix elements of $U(r)^{[m,0,s]}$ are given by

$$\langle \mathbf{p}, \mu |\, U(r)^{[m,0,s]} \,|\mathbf{p}', \mu'\rangle = \chi_{\alpha_0}(g_{\mathbf{p}}g_1g_{\mathbf{p}}^{-1}) L^{[s]}_{\mu\mu'}(g_{\mathbf{p}}g_2g_{r\mathbf{p}}^{-1}) \langle r\mathbf{p}, \mu \,|\, \mathbf{p}', \mu'\rangle,$$

where $r = g_1g_2$.

3. Now we start with the proof.

(i) The above representation depends explicitly on the choice of the representative $g_{\mathbf{p}}$ in each coset. Different representative systems lead to different basis systems in $\mathfrak{H}$. Hence it is plausible to look for a choice of $g_{\mathbf{p}}$ such that the matrix

$$L^{[s]}_{\mu\mu'}(g_{\mathbf{p}}g_2g_{r\mathbf{p}}^{-1})$$

becomes independent of (R) in $r = (\theta, R, \mathbf{u}, \mathbf{v}, \tau)$.

The possibility is shown by putting $\hat{R} = 1$, i.e., $g_{\mathbf{p}} = (0, 1, 0, -m^{-1}\mathbf{p}, 0)$. Hence $g_{\mathbf{p}}g_2g_{r\mathbf{p}}^{-1} = (R)$ and $U(r)^{[m,0,s]}$ can be written in $\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2$ as

$$U(r)^{[m,0,s]} = U_1(r) \times U_2(R),$$

with

$$\langle \mathbf{p} |\, U_1(r) \,|\mathbf{p}'\rangle = \chi_{\alpha_0}(g_{\mathbf{p}}g_1g_{\mathbf{p}}^{-1}) \langle r\mathbf{p} \,|\, \mathbf{p}'\rangle$$

and

$$\langle \mu |\, U_2(R) \,|\mu'\rangle = L^{[s]}_{\mu\mu'}(R).$$

(ii) We now take the set $\{U_1(r)\}$ which is independent of the representation of $G_l^{m,0}$, and consider the representation of G_E which is constructed using the trivial representation of $G_l^{m,0}$, i.e., from $U(SO(N))^{[0]}$. Then

$$\mathfrak{H} = \mathfrak{H}_1 \quad \text{and} \quad U(r)^{[m,0,0]} = U_1(r)$$

holds. There is (up to unitary equivalence) only one irreducible representation labeled by $[m, 0, 0]$. Hence

a unitary matrix $\tilde{U}_1$ exists in $\mathfrak{H}_1$ with

$$\{U_1(r)\}\overset{\tilde{U}_1}{\sim}U_1(G_E(N))^{[m,0,0]}.$$

The set $\{U_2(R)\}$ is, by construction, an irreducible unitary representation of an $SO(N)^I$ group. But the transformations in this group are not independent of the $SO(N)$ transformations given by $\{(0, R, 0, 0, 0)\}$ in $U_1(r)$. If r contains (R) as $SO(N)$ part, the same transformation occurs in $U_2(SO(N)^I)$. Let (R^I) be a parametrization of $SO(N)^I$ and let ρ be a fixed isomorphism between $SO(N)^I$ and $\{(0, R, 0, 0, 0)\}$, i.e.,

$$(0, R, 0, 0, 0) \to \rho[R] \in SO(N)^I$$

such that $(R) = (\rho[R])$. Then

$$\{U_2(R)\}\overset{\tilde{U}_2}{\sim}U_2(\rho[SO(N)])^{[s]}$$

holds, with $\tilde{U}_2$ being a unitary matrix in $\mathfrak{H}_2$. The statement of Lemma 1, part (i), is obtained by putting

$$\tilde{U} = \tilde{U}_1 \times \tilde{U}_2.$$

(iii) The Lie algebra is derived via Stone's theorem. Take a 1-dimensional subgroup of $G_E(N)$ parameterized by (α). Then the corresponding generator ia is defined on a dense set of analytic vectors $\mathfrak{D}_U^{an}$:

$$\lim_{\alpha\to 0}\alpha^{-1}(U(\alpha)^{[m,0,s]} - I)\,|f,\mu\rangle = ia\,|f,\mu\rangle,$$
$$|f,\mu\rangle \in \mathfrak{D}_U^{an}.$$

Because the subgroups with generators C, P_i, Q_i, H, $i = 1, \cdots, N$, do not contain R, we have, e.g., for H that

$$(U(\tau)^{[m,0,s]} - I) = U_1(\tau)^{[m,0,s]} \times I_2 - I_1 \times I_2$$

and hence, also,

$$H = (H)_1 \times I_2.$$

The same decomposition holds for C, P_i, Q_i.

(iv) The derivation of the remaining generators needs a property of unitary representations $U(L)$ of a Lie group L on a product space $\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2$. If any $U(l)$, $l \in L$ splits in $\mathfrak{H}$ into

$$U(l) = U_1(l) \times U_2(l),$$

with $\{U_1(l)\}$ and $\{U_2(l)\}$ being unitary representations of L in $\mathfrak{H}_1$ and $\mathfrak{H}_2$, respectively, then one has, for the generator $i \cdot a$ of a 1-dimensional subgroup $\{U(\alpha)\}$ on the domain $\mathfrak{D}_U^{an}$,

$$\begin{aligned} i\cdot a\,|f,\mu\rangle &= \lim_{\alpha\to 0}(\alpha^{-1}(U_1(\alpha) \times U_2(\alpha) - I_1 \times I_2)\,|f,\mu\rangle) \\ &= \lim_{\alpha\to 0}(|\alpha^{-1}(U_1(\alpha) - I_1)f, \mu\rangle \\ &\quad + |U_1(\alpha)f, \alpha^{-1}(U_2(\alpha) - I_2)\mu\rangle). \end{aligned}$$

This leads to a special form of $i \cdot a$ with projection parts $(i \cdot a)_1$ and $(i \cdot a)_2$ in $\mathfrak{H}_1$ and $\mathfrak{H}_2$, respectively:

$$i \cdot a = (i \cdot a)_1 \times I_2 + I_1 \times (i \cdot a)_2.$$

(v) Returning to our example, we have immediately

$$d_{ij} = (d_{ij})_1 \times I_2 + I_1 \times (d_{ij})_2.$$

The form of $(d_{ij})_1 = l_{ij}$ was already derived in Paper I. $\{(d_{ij})_2\}$ is the Lie algebra of $U(\rho[SO(N)])$, i.e.,

$$(d_{ij})_2 = \rho[l_{ij}].$$

This concludes the proof.

C. An Alternative Formulation for $G_E(N)$

1. In the preceding section, a spin separation was obtained in representations $U(G_E(N))$ as a consequence of Lemma 1. For later applications it is useful to start not with $G_E(N)$, but with an abstract group $D^0(N)$ which already exhibits this separation in an obvious way. To construct $D^0(N)$, we use the fact that, in general, a dynamical group is unique only up to isomorphisms, and we replace $G_E(N)$ by an isomorphic copy such that the spin group behaves in irreducible representations similarly as a direct factor. From Lemma 1 it is plausible to choose

$$D^0(N) = (G_E^T(N) \textcircled{\times} SO(N)) \circ \rho[SO(N)],$$

where ρ denotes the isomorphism in Lemma 1. Obviously $D^0(N) \approx G_E(N)$. $D^0(N)$ is subgroup of

$$D^0(N) \subset (G_E^T(N) \textcircled{\times} SO(N)) \otimes SO(N)^I,$$

containing two rotation groups from which the physical one is selected as discussed in 1A2. Because $D^0(N) \approx G_E(N)$, Lemma 1 implies that the irreducible representations of $D^0(N)$ are of the following form:

$$U(D^0(N))^{[m,0,s]} = U(G_E^T(N)) \times SO(N))^{[m,0,0]} \times U(\rho[SO(N)])^{[s]}.$$

2. $D^0(N)$ can be chosen as a dynamical group for the free particle. It works for all $m > 0$ and s. Irreducible representations labeled by $[m, 0, s]$ are physical. The interpretation of $\rho[SO(N)]$ as a spin group and of $U(SO(N)) \subset U(G_E^T(N) \textcircled{\times} SO(N))^{[m,0,0]}$ as an orbital angular-momentum group is a result of Lemma 1. Summing up, we have

Lemma 2: Let $D^0(N)$ be the group

$$D^0(N) = (G_E^T(N) \textcircled{\times} SO(N)) \circ \rho[SO(N)],$$

with $\rho[SO(N)]$ being the image of a fixed isomorphism ρ of $\{(0, R, 0, 0, 0)\} \subset G_E(N)$.

(i) $D^0(N)$ is isomorphic to $G_E(N)$.

(ii) $D^0(N)$ is a dynamical group for the free particle. Its physical representations are given by

$$U(D^0(N))^{[m,0,s]} = U(G_E(N)^{[m,0,\sigma]} \times U(\rho[SO(N)])^{[s]}.$$

2. A DYNAMICAL GROUP FOR INTERACTING PARTICLES WITH SPIN

It is shown that the p-limitable dynamical group

$$D^t = G_E^0(N) \textcircled{\times} Sp(2N, R)$$

cannot describe particles with spin. Therefore, D^t is enlarged to

$$D_s^t = D^t \otimes SO(N).$$

The physical representation $U_{Ph}(D_s^t)$ is uniquely p-limitable into the physical representation of the free-particle group and is determined by the mass and spin of the system. The relation between limitable groups and some recent work on dynamical groups is discussed.

A. Spin Content of $D^t = G_E^0 \textcircled{\times} Sp(2N, R)$

1. We want to generalize our model with dynamical group D^t and a Nelson extension $\tilde{U}(D^t)$ of $U(G_E^0(N))^{[m]}$ as a physical representation to particles with spin. At first sight, it seems plausible to use D^t, but use another irreducible representation $\hat{U}_{ph}(D^t)$ as the physical one. $\hat{U}_{ph}(D^t)$ is exposed to our limitation postulates. Then the free and the interacting system are described in the same Hilbert space. $\hat{U}_{ph}(D^t)$ is an irreducible representation in the space $\mathfrak{H}^{[m,0,s]}$ of the free particle. The limitation between $\hat{U}_{ph}(D^t)$ and $U(G_E(N))^{[m,0,s]}$ is formulated as a restriction. Then

$$\hat{U}_{ph}(D^t) \to U(G_E(N))^{[m,0,s]}$$

leads to a condition for the restriction $\hat{U}(G_E(N))$ of $\hat{U}_{ph}(D^t)$ to $G_E(N)$,

$$\hat{U}(G_E(N)) = U(G_E(N))^{[m,0,s]}. \tag{1}$$

2. A representation of this type is possible only for $s = 0$ because of the following:

Lemma 3: A unitary representation $\hat{U}(D^t)$ of D^t in $\mathfrak{H} = \mathfrak{H}^{[m,0,s]}$ with a restriction $\hat{U}(G_E(N))$ to $G_E(N) \subset D^t$, given by

$$\hat{U}(G_E(N)) = U(G_E(N))^{[m,0,s]},$$

exists only for $s = 0$ and is unitarily equivalent to the Nelson extension of $U(G_E^0(N))^{[m]}$.

Proof:

(i) $\mathfrak{H} = \mathfrak{H}^{[m,0,s]}$ is decomposed in

$$\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2, \quad \mathfrak{H}_1 = \mathfrak{H}^{[m,0,0]}, \quad \mathfrak{H}_2 = \mathfrak{H}^{[s]}.$$

The dimension $d(s)$ of $\mathfrak{H}_2$ is finite. $\hat{U}(D^t)$ splits into

$$\hat{U}(D^t) = \hat{U}_1(D^t) \times \hat{U}_2(D^t)$$

and [see Lemma 2, part (iv)] any generator d in $\hat{U}(D^t)$ is of the form

$$d = d_1 \times I_2 + I_1 \times d_2.$$

(ii) $\hat{U}(G_E) = U(G_E)^{[m,0,s]}$ implies that the project tion part of the subalgebra G_E^T in $\mathfrak{H}_2$ vanishes. The projection part $(d_{ij})_2$ of the generators of the physical $SO(N)$ subgroup in $G_E(N)$, being a subgroup of $Sp(2N, R) \subset D^t$, form a $d(s)$-dimensional representation in $\mathfrak{H}_2$. Hence at least a group S with

$$SO(N) \subset S \subset Sp(2N, R)$$

is represented in $\mathfrak{H}_2$. Because $d(s)$ is finite, we infer from the unitarity of $U(D^t)$ that S is compact and that the projection parts of at least all noncompact generators of $Sp(2N, R)$ vanish in $\mathfrak{H}_2$. For $Sp(2N, R)$ being simple implies that also the compact generators vanish. Hence the representation of $\rho[SO(N)]$ is trivial, and $s = 0$ holds.

(iii) A proof of the last part of the lemma is contained in Paper I, Sec. 3C3.

3. A representation with the above restriction properties exists at most for $d(s) = \infty$. But the simple physical interpretation of $U(D^t)$ via a limitation procedure fails in this case. A more involved method is beyond nonrelativistic quantum mechanics and may be useful for relativistic examples, e.g., if one abandons mass and spin conservation for interacting systems or if a tower of 1-particle systems is used instead of a single particle.

B. $D_s^t = D^t(N) \otimes SO(N)$ as p-Limitable Dynamical Group

1. Now, the problem is to construct a group D_s^t or its Lie algebra D_s^t possessing a unitary irreducible representation limitable into

$$U(G_E(N))^{[m,0,s]}, \quad s > 0.$$

As in Sec. 3C of Paper I, we choose for D_s^t an embedding of the Lie algebra D^0 of the free-particle group $D^0 \approx G_E(N)$, i.e.,

$$\mathsf{D}^0 \subset \mathsf{D}_s^t.$$

The limitation

$$\hat{U}(\mathsf{D}_s^t) \to U(\mathsf{D}^0)^{[m,0,s]} \tag{2}$$

is given by a restriction of $\hat{U}(\mathsf{D}_s^t)$ to the subalgebra D^0 and can be realized as a group contraction between representations if

$$\hat{U}(\mathsf{D}^0) = U(\mathsf{D}^0)^{[m,0,s]}$$

holds. We denote by $\{d_\alpha^0\}$ a basis in D^0, and decompose D_s^t into the subalgebra D^0 and a subspace $\tilde{\mathsf{D}}$ spanned by $\{\tilde{d}_i\}$, $\mathsf{D}_s^t = \mathsf{D}^0 + \tilde{\mathsf{D}}$. Consider the algebra $\mathsf{D}_s^t[\epsilon]$ spanned by

$$\{d_\alpha^0, \epsilon\tilde{d}_i\}, \quad \text{real } \epsilon > 0.$$

$\mathsf{D}_s^t[\epsilon]$ is isomorphic to $\mathsf{D}_s^t = \mathsf{D}_s^t[1]$, but

$$\hat{U}(\mathsf{D}_s^t[\epsilon]) \xrightarrow[\epsilon\to 0]{} \hat{U}(\mathsf{D}^0) = U(\mathsf{D}^0)^{[m,0,s]}, \tag{3}$$

as is easily checked and well known.[3] The process is a p-limitation. The same contraction yields in the abstract algebra

$$\mathsf{D}_s^t[\epsilon] \xrightarrow[\epsilon\to 0]{} \mathsf{D'}_s^t = \mathsf{A}_{n-n_0} \dot{+} \mathsf{D}^0,$$

with A_{n-n_0} being an $(n-n_0)$-dimensional Abelian ideal, and n, n_0 denoting the dimension of D^t and D^0. The above contraction connects a representation of D_s^t with a nonfaithful representation of $\mathsf{D'}_s^t$ and leads to a convenient class of limitation procedures for which, in general, no corresponding mechanism exists in the abstract algebra.

The result is collected in the following:

Lemma 4: Let L_1 and L_2 be two Lie algebras with $\mathsf{L}_1 \subset \mathsf{L}_2$. Denote by $V(\mathsf{L}_1)$ and $W(\mathsf{L}_2)$ representations of L_1 and L_2. Then $V(\mathsf{L}_1)$ and $W(\mathsf{L}_2)$ are connected via an Inönü–Wigner contraction

$$W(\mathsf{L}_2) \to V(\mathsf{L}_1)$$

if and only if the restriction $W(\mathsf{L}_1)$ of $W(\mathsf{L}_2)$ to its subalgebra L_1 is given by

$$W(\mathsf{L}_1) = V(\mathsf{L}_1).$$

The contraction is a p-limitation.

2. Now let $\mathsf{D}_s^t = \mathsf{R} \dot{+} \mathsf{L}_0$ be a Levi decomposition of D_s^t. Using our result in I for the embedding of D^0, we want R to be minimal in the sense that it contains generators from $\mathsf{G}_E^T(N)$ only. There are three possibilities: R^i, $i = 1, 2, 3$. One of them, R^2, leads to a trivial result. We choose $\mathsf{R}^1 = \mathsf{G}_E^0(N)$. This choice is justified also if one demands that the spinless group D^t is subgroup in D_s^t. Then L_0 is given by

$$\mathsf{L}_0 = sp(2N, R) \oplus \mathsf{N}^1,$$

with N^1 being the algebra of an arbitrary semisimple group N^1 and

$$\mathsf{D}_s^t = \mathsf{D}^t \oplus \mathsf{N}^1 \quad \text{and} \quad D_s^t = D^t \otimes N^1. \tag{4}$$

The trivial N^1 leads back to D^t.

3. Using the limitation (1), we determine the unknown part N^1 such that

$$\hat{U}(\mathsf{D}^0) = U(\mathsf{D}^0)^{[m,0,s]} \tag{5}$$

is the restriction to the physical D^0 subalgebra. [The restriction of a representation $U(L)$ to a subgroup $L' \subset L$ is denoted by $U(L')$. The same notation is used for algebras.] There are a lot of subalgebras in D_s^t which are isomorphic to D^0 but not physical. For a location of the physical, suppose that

$$\mathsf{D}^0 \subset \mathsf{D}^t \subset \mathsf{D}_s^t.$$

Then Lemma 3 applies and leads to $s = 0$. Hence D^0 has a projection part in N^1.

To arrange the model as simple as possible, we choose for the restriction $\hat{U}(\mathsf{D}^t)$ the spinless representation already known from Paper I, Sec. 3C:

$$\hat{U}(\mathsf{D}^t) = \hat{U}(\mathsf{D}^t)^{[m]}. \tag{6}$$

Then the spin part is contained in N^1, which means that $so(N)^I \subset \mathsf{N}^1$ or $\mathsf{N}^1 = so(N)^I \oplus \mathsf{N}^2$, with N^2 being arbitrary and semisimple. We drop the direct factor which can be used to describe internal symmetries and arrive at

$$D_s^t = D^t \otimes SO(N)^I.$$

The Lie algebra $\mathsf{D}_s^t = (\mathsf{G}_E^0 \dot{+} sp(2N, R)) \oplus so(N)^I$ is spanned by

$$\{\mathsf{C}, P_i, Q_i\} = \mathsf{G}_E^0; \quad \{s_{ij}\} = sp(2N, R)$$

and

$$\{d_{ij}^I = -d_{ji}^I\} = so(N)^I,$$

and the generators of the physical $SO(N)$ subgroup of $Sp(2N, R)$ are denoted by $\{d_{ij} = -d_{ji}\} = so(N)$, $i, j = 1, \cdots, N$. $s \in sp(2N, R)$ is a generator with $[s, d_{ij}] = 0$, and $\{r_\alpha\}$ are basis elements of $sp(2N, R)$ not contained in $\{d_{ij}, s\}$, $\alpha = 1, \cdots, n$;

$$n = \tfrac{3}{2}N(N+1) - 1.$$

If $U(sp(2N, R))$ is mapped by an isomorphism σ into the set of skew-symmetric second-order polynomials in P_i, Q_i (see Paper I, Lemma 3) the generator s is identified by $\sigma(s) = \sum_{i=1}^N P_i^2$.

The essential point is now the identification of the physical subalgebra D^0. Because of (6) and of Lemma 1, $\hat{U}(\mathsf{D}^t)$ contains an irreducible representation of a subalgebra G_E, isomorphic to D^0, given by

$$U(\tilde{\mathsf{G}}_E)^{[m,0,0]}, \quad \tilde{\mathsf{G}}_E \subset \mathsf{D}^t \subset \mathsf{D}_s^t.$$

$\tilde{G}_E$ is not the free-particle group. Its generators

$$\tilde{\mathsf{G}}_E = \{\mathsf{C}, P_i, Q_i, s, d_{ij}\}$$

can be identified with those of the physical $\mathsf{G}_E^T \subset \mathsf{D}^0$ and with the orbital part of the angular momentum, respectively. Hence $\{d_{ij}^I\}$ is the corresponding spin part and the representation of $SO(N)^I$ is irreducible

with spin content s. Therefore, we choose

$$\mathsf{D}^0 = \{\mathsf{C}, P_i, Q_i, s, d_{ij} + d^I_{ij}; i, j = 1, \cdots, N\} \subset \mathsf{D}^t_s.$$

As a by-product, the physical representation of D^t_s is obtained with a representation space [$\tilde{U}(D^t)^{[m]}$ is irreducible]

$$\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2,$$
$$\hat{U}_{ph}(\mathsf{D}^t_s) = \tilde{U}_1(\mathsf{D}^t)^{[m]} \times U_2(so(N)^I)^{[s]}. \tag{7}$$

4. The physical representation is unique in the sense that there is up to unitary equivalence only one irreducible representation $\hat{U}(\mathsf{D}^t_s)$ with restriction property (5) or the equivalent (see Lemma 4) limitation (3). For the proof we use arguments similar to those for the identification of D^0. Consider the restriction $\hat{U}(\tilde{\mathsf{G}}_E)$: With (5) and with Lemma 1,

$$\hat{U}(\tilde{\mathsf{G}}_E) = U(\tilde{\mathsf{G}}_E)^{[m,0,0]}$$

holds, and Lemma 3 implies the unique result $\hat{U}(\mathsf{D}^t) = \tilde{U}(\mathsf{D}^t)^{[m]}$. The representation space $\mathfrak{H}$ splits into $\mathfrak{H} = \mathfrak{H}_1 \times \mathfrak{H}_2$ and

$$\hat{U}(\mathsf{D}^t_s) = \tilde{U}_1(\mathsf{D}^t)^{[m]} \times \hat{U}_2(so(N)^I).$$

By the application of Lemma 1, the above $so(N)^I$ representation is obtained. We remark that the spin is attached to the spinless interacting particle similarly to how it is attached to the spinless free particle.

5. $\hat{U}_{ph}(\mathsf{D}^t_s)$ is labeled uniquely by m and s,

$$\hat{U}_{ph}(\mathsf{D}^t_s) = U(\mathsf{D}^t_s)^{[m,s]}.$$

The mass and the spin of the free particle is obtained if the interaction is turned off. Because D^0 and $U(\mathsf{D}^0)^{[m,0,0]}$ remain constant during the limitation, the center $\mathsf{C} \in \mathsf{D}^0 \subset \mathsf{D}^t_s$ is identified as the mass operator, and the eigenvalue of the remaining Casimir operators of D^0, except $\mathsf{C}^{(2)}_{H_0}$, are interpreted as spin content. Hence mass and spin conservation hold during the limitation, which means that the interaction can be turned off without changing the mass and the spin of the particle.

6. For the contraction of D^t_s with respect to D^0, we introduce a new basis in

$$\mathsf{D}^t_s = \{\mathsf{C}, P_i, Q_i, s, d_{ij} + d^I_{ij}, r_\alpha, d^I_{ij}\},$$

and consider

$$\mathsf{D}^t_s[\epsilon] = \{\mathsf{C}, P_i, Q_i, s, d_{ij} + d^I_{ij}, \epsilon r_\alpha, \epsilon d^I_{ij}\}.$$

Then Lemma 4 implies

$$U(\mathsf{D}^t_s[\epsilon])^{[m,s]} \xrightarrow[\epsilon \to 0]{} U(\mathsf{D}^0)^{[m,0,s]}. \tag{8}$$

The p-limitation is unique (see the definition in I, Sec. 2C3) because of the uniqueness of the physical representation and because mass and spin conservation holds.

7. The last part of our discussion concerns the interaction type t, i.e., the set of Hamiltonians contained in $U(\mathsf{D}^t_s)^{[m,s]}$. Because of our assumption that any $H \in t$ is also a generator in $U(\mathsf{D}^t_s)^{[m,s]}$, the interaction type is spanned by a basis of D^t_s. It is convenient to select in t the set t' of those generators which are not contained in D^0, i.e., all generators related to the free particle. Then t' is given by

$$t' = \{r_\alpha, d^I_{ij}\}.$$

The interpretation of ϵ in (8) as interaction strength is obvious. The contraction can also be formulated such that different interactions $H_i \in t$, $i = 1, \cdots, p$, are porportional to different interaction strengths ϵ_i. However, then the condition that $\epsilon_i \to 0$, $i = 1, \cdots, p$, is a p-contraction[4] needs some care. The identification of r_α is known from Paper I because $r_\alpha \in \tilde{U}(\mathsf{D}^t)^{[m]}$. The d^I_{ij} are spin matrices acting only on the spin variable. They are independent of space–time coordinates and lead to spin–flip processes. Hence we have

$$t' = \{(P_iQ_j + Q_jP_i), P_kP_l, Q_iQ_j, d^I_{ij};$$
$$i, j, k, l = 1, \cdots, N, k \neq l; P^2_m; m = 2, \cdots, N\}. \tag{9}$$

8. We end this section with a collection of our results:

(i)

$$D^t_s = (G^0_E(N) \textcircled{\times} Sp(2N, R)) \otimes SO(N)^I$$

is a p-limitable dynamical group. There is only one irreducible representation $\hat{U}_{ph}(D^t_s)$ which is p-limitable via a group contraction into an irreducible representation of the free-particle group D^0 identified as subgroup of D^t_s in (6),

$$\hat{U}_{ph}(\mathsf{D}^t_s[\epsilon]) \xrightarrow[\epsilon \to 0]{} U(\mathsf{D}^0)^{[m,0,s]}.$$

Hence the p-limitation is unique. The physical representation is given by

$$U(D^t_s)^{[m,s]} = \tilde{U}(D^t)^{[m]} \times U(SO(N)^I)^{[s]},$$

with $\tilde{U}(D^t)^{[m]}$ being the Nelson extension of $U(G^0_E)^{[m]}$.

(ii) Mass and spin of the free particle are conserved during the limitation and determine the physical representation uniquely which describes particles with mass m and spin s in N dimensions with Hamiltonians given in (9).

C. Discussion

1. The groups D^t and D^t_s are examples of a pure and complete group-theoretical description of interacting nonrelativistic systems. The approach starts with the

abstract group. With physically motivated principles as restrictive conditions, the physical representation and the described interaction were calculated.

2. An application of the method to relativistic systems may be possible if mass and spin conservation are replaced by other assumptions, e.g., conservation of the total electric charge, and if the definition of the free particle is extended such that a tower of particles with given mass and spin is included as it was indicated already in 2A3. In general, such principles lead to a selection for the physical dynamical groups and their representations. We refer to a recent fruitful proposal for noncompact dynamical groups and the physical identification[5–8] of certain matrix elements as form factors and transition probabilities.

It will be shown elsewhere that the identification of the form factor in terms of generators from $SO(3, 1)$ and $SO(4, 2)$, and the physical representation $U_{P\hbar}(SO(3, 1))$ and $U_{P\hbar}(SO(4, 2))$ used in Refs. 6 and 7 can be justified via a limiting procedure which may correspond to a turning off that part of the interaction which causes the electromagnetic structure.

The choice of the physical representation of $SO(3, 1)$ is related to Lemma 2 in Paper I because $so(3, 1)$ is constructed from second-order polynomials of a finite number N' of creation and annihilation operators, or of $P_i, Q_i, i = 1, \cdots, N'$. Hence $so(3, 1)$ is a subalgebra of $sp(2N', R)$ and only those irreducible representations can occur which are available by the Nelson extension $\tilde{U}(D^t)^{[m]}$ with $m = 1$.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. L. Castell, Dr. G. C. Hegerfeldt, Dr. A. Hartkämper, Dr. J. Hennig, Dr. H. Kleinert, and Dr. H. Neumann for stimulating discussions and useful conversations.

[1] H. D. Doebner and O. Melsheimer, J. Math. Phys. **9**, 1638 (1968).
[2] J. M. Levy-Leblond, J. Math. Phys. **4**, 6 (1963); J. Voisin, J. Math. Phys. **6**, 1519 and 1822 (1965).
[3] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. US **39**, 510 (1953); E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[4] H. D. Doebner and O. Melsheimer, Nuovo Cimento **49**, 306 (1967).
[5] A. O. Barut, Phys. Rev. **156**, 1538 (1967).
[6] A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1546 (1967).
[7] H. Kleinert, thesis, University of Colorado, 1967.
[8] A. O. Barut, D. Corrigan, and H. Kleinert, University of Colorado, Preprint No. 123, 1967.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Branching Rules for $U(n) \supset SO(n)$

M. BRUNET AND M. RESNIKOFF
Department of Physics, State University of New York at Buffalo, Buffalo, New York 14214

(Received 14 April 1969)

Using Littlewood's theorem, we derive branching rules for the restriction of $U(n)$ to the special orthogonal group $SO(n)$, for the case of a 3-row Young diagram. The special cases $n = 4, 5$ are treated by use of Murnaghan's modification rules.

The general method of deriving branching rules for $U(n) \supset SO(n)$ [that is, the rules that determine which representations of the n-dimensional rotation group $SO(n)$ are contained in the representations of $U(n)$] is provided by Littlewood's theorem,[1] supplemented by Murnaghan's modification rules.[2] However, no explicit expressions exist in the literature for a case as general as a 3-row diagram of $U(n)$.[3–5] In this paper we derive the $U(n) \supset SO(n)$ results for a 3-row diagram and comment how the results may be extended to m rows, $m < n$.

The branching rules for $U(n) \supset SO(n)$ have application in nuclear physics. In the Kramer and Moshinsky[6] generalization of the Elliott[7] shell model, one must construct state vectors in the chain $U(n) \supset S(n)$, where $S(n)$ is the symmetric group in n dimensions. The state vector[8] for $U(n) \supset O(n)$ is an intermediate step in the $SU(n) \supset S(n)$ chain, and the branching rules for $n = 4$ are important for the derivation of the general $U(4) \supset U(2) \otimes U(2)$ state vector,[9] useful in the Wigner supermultiplet[10] and quasispin[11] models.

We shall present the derivation in two steps: $U(n) \downarrow O(n)$[$U(n)$ restricted to $O(n)$], using Littlewood's rules, and then the essentially trivial reduction $[U(n) \downarrow O(n)] \downarrow SO(n)$. Before starting, we shall briefly discuss the irreducible representations of $U(n)$, $O(n)$, and $SO(n)$.

To each Young diagram $Y_{h_1 h_2 \cdots h_n}$, where h_i is the

abstract group. With physically motivated principles as restrictive conditions, the physical representation and the described interaction were calculated.

2. An application of the method to relativistic systems may be possible if mass and spin conservation are replaced by other assumptions, e.g., conservation of the total electric charge, and if the definition of the free particle is extended such that a tower of particles with given mass and spin is included as it was indicated already in 2A3. In general, such principles lead to a selection for the physical dynamical groups and their representations. We refer to a recent fruitful proposal for noncompact dynamical groups and the physical identification[5–8] of certain matrix elements as form factors and transition probabilities.

It will be shown elsewhere that the identification of the form factor in terms of generators from $SO(3, 1)$ and $SO(4, 2)$, and the physical representation $U_{P\hbar}(SO(3, 1))$ and $U_{P\hbar}(SO(4, 2))$ used in Refs. 6 and 7 can be justified via a limiting procedure which may correspond to a turning off that part of the interaction which causes the electromagnetic structure.

The choice of the physical representation of $SO(3, 1)$ is related to Lemma 2 in Paper I because $so(3, 1)$ is constructed from second-order polynomials of a finite number N' of creation and annihilation operators, or of $P_i, Q_i, i = 1, \cdots, N'$. Hence $so(3, 1)$ is a subalgebra of $sp(2N', R)$ and only those irreducible representations can occur which are available by the Nelson extension $\tilde{U}(D^t)^{[m]}$ with $m = 1$.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. L. Castell, Dr. G. C. Hegerfeldt, Dr. A. Hartkämper, Dr. J. Hennig, Dr. H. Kleinert, and Dr. H. Neumann for stimulating discussions and useful conversations.

[1] H. D. Doebner and O. Melsheimer, J. Math. Phys. **9**, 1638 (1968).
[2] J. M. Levy-Leblond, J. Math. Phys. **4**, 6 (1963); J. Voisin, J. Math. Phys. **6**, 1519 and 1822 (1965).
[3] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. US **39**, 510 (1953); E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[4] H. D. Doebner and O. Melsheimer, Nuovo Cimento **49**, 306 (1967).
[5] A. O. Barut, Phys. Rev. **156**, 1538 (1967).
[6] A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1546 (1967).
[7] H. Kleinert, thesis, University of Colorado, 1967.
[8] A. O. Barut, D. Corrigan, and H. Kleinert, University of Colorado, Preprint No. 123, 1967.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Branching Rules for $U(n) \supset SO(n)$

M. BRUNET AND M. RESNIKOFF
Department of Physics, State University of New York at Buffalo, Buffalo, New York 14214

(Received 14 April 1969)

Using Littlewood's theorem, we derive branching rules for the restriction of $U(n)$ to the special orthogonal group $SO(n)$, for the case of a 3-row Young diagram. The special cases $n = 4, 5$ are treated by use of Murnaghan's modification rules.

The general method of deriving branching rules for $U(n) \supset SO(n)$ [that is, the rules that determine which representations of the n-dimensional rotation group $SO(n)$ are contained in the representations of $U(n)$] is provided by Littlewood's theorem,[1] supplemented by Murnaghan's modification rules.[2] However, no explicit expressions exist in the literature for a case as general as a 3-row diagram of $U(n)$.[3–5] In this paper we derive the $U(n) \supset SO(n)$ results for a 3-row diagram and comment how the results may be extended to m rows, $m < n$.

The branching rules for $U(n) \supset SO(n)$ have application in nuclear physics. In the Kramer and Moshinsky[6] generalization of the Elliott[7] shell model, one must construct state vectors in the chain $U(n) \supset S(n)$, where $S(n)$ is the symmetric group in n dimensions. The state vector[8] for $U(n) \supset O(n)$ is an intermediate step in the $SU(n) \supset S(n)$ chain, and the branching rules for $n = 4$ are important for the derivation of the general $U(4) \supset U(2) \otimes U(2)$ state vector,[9] useful in the Wigner supermultiplet[10] and quasispin[11] models.

We shall present the derivation in two steps: $U(n) \downarrow O(n)$[$U(n)$ restricted to $O(n)$], using Littlewood's rules, and then the essentially trivial reduction $[U(n) \downarrow O(n)] \downarrow SO(n)$. Before starting, we shall briefly discuss the irreducible representations of $U(n)$, $O(n)$, and $SO(n)$.

To each Young diagram $Y_{h_1 h_2 \cdots h_n}$, where h_i is the

ith-row length and $h_1 \geq h_2 \geq \cdots \geq h_n \geq 0$, there corresponds an irreducible representation ${}^n\Gamma_{h_1 h_2 \cdots h_n}$ of $U(n)$.

An irreducible representation[12] of $O(n)$, ${}^nD'_{m_1 m_2 \cdots m_p}$, is specified by the highest weight $(m_1, m_2, \cdots, m_p)$, where $n = 2p$ or $2p+1$ as n is even or odd, respectively. Since $m_1 \geq m_2 \geq \cdots \geq m_p$, the weights of $O(n)$ specify a partition and, hence, a Young diagram $Y_{m_1 m_2 \cdots m_p}$.

The weights of $SO(n)$ for n even may have the last component negative; i.e., for n even, the condition $m_1 \geq m_2 \geq \cdots \geq m_p$ is replaced by $m_1 \geq m_2 \geq \cdots \geq |m_p|$. Here ${}^{2p}D_{m_1 m_2 \cdots m_p}$ and ${}^{2p}D_{m_1 m_2 \cdots -m_p}$ are nonequivalent and have the same dimension, this being half of the dimension of ${}^{2p}D'_{m_1 m_2 \cdots m_p}$ if $m_p \neq 0$ and the same dimension if $m_p = 0$.

The branching problem for $U(n) \downarrow O(n)$ for irreducible representations of $U(n)$, characterized by diagrams having at most three rows, consists of finding the numbers $\nu_{m_1 m_2 \cdots m_r}$ such that

$$ {}^n\Gamma_{h_1 h_2 h_3} \downarrow O(n) = \sum_{m_1 m_2 \cdots m_p} \nu^{h_1 h_2 h_3}_{m_1 m_2 \cdots m_r} \, {}^nD'_{m_1 m_2 \cdots m_r}. $$

For $n \geq 6$, the solution of the branching problem is provided by Littlewood's theorem. We shall consider separately the modifications necessary when $n = 4, 5$. For 3-row diagrams, the theorem states

$$ {}^nD_{m_1 m_2 \cdots m_p} \in {}^n\Gamma_{h_1 h_2 h_3} \downarrow O(n) \quad \text{iff } \exists \text{ integers} $$

$p_1, p_2, \cdots, p_s$ such that

$$ Y_{h_1 h_2 h_3} \in Y_{m_1 m_2 \cdots m_r} \otimes Y_{2p_1 2p_2 \cdots 2p_s}. $$

From the rules for forming outer products, we immediately see that $r, s \leq 3$. The multiplicity $\nu^{h_1 h_2 h_3}_{m_1 m_2 m_3}$ is just the number of ways of forming $Y_{h_1 h_2 h_3}$ from the product diagrams. That is, if $g^{m_1 m_2 m_3}_{p_1 p_2 p_3}$ is the number of times $Y_{h_1 h_2 h_3}$ is contained in $Y_{m_1 m_2 m_3} \otimes Y_{2p_1 2p_2 2p_3}$, then

$$ \nu^{h_1 h_2 h_3}_{m_1 m_2 m_3} = \sum_{p_1 p_2 p_3} g^{m_1 m_2 m_3}_{p_1 p_2 p_3}. \tag{1} $$

We take the outer product $Y_{m_1 m_2 m_3} \otimes Y_{2p_1 2p_2 2p_3}$, adding to the diagram $Y_{m_1 m_2 m_3}$, $2p_1$ symbols a, $2p_2$ symbols b, and $2p_3$ symbols c. This is illustrated in Fig. 1, where

$$ x_1 + x_2 + x_3 = 2p_1, \tag{2a} $$

$$ y_1 + y_2 = 2p_2. \tag{2b} $$

From the rules for outer products and the requirement that $Y_{h_1 h_2 h_3} \in Y_{m_1 m_2 m_3} \otimes Y_{2p_1 2p_2 2p_3}$, we obtain the following independent relations:

$$ m_1 \geq m_2 + x_2, \tag{3a} $$

$$ m_2 \geq m_3 + x_3, \tag{3b} $$

$$ m_2 + x_2 \geq m_3 + x_3 + y_2 \tag{3c} $$

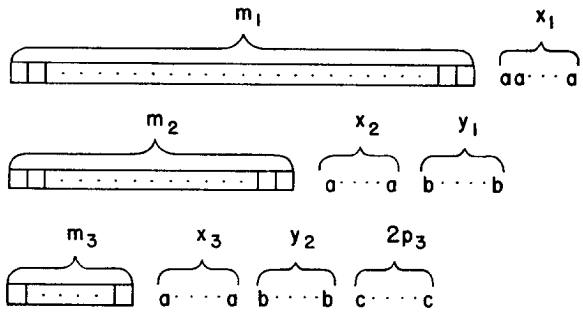


FIG. 1. The outer product $Y_{m_1 m_2 m_3} \otimes Y_{2p_1 2p_2 2p_3}$.

(from condition that no two of the same added symbols appear in the same column);

$$ x_1 \geq y_1, \tag{4a} $$

$$ y_1 \geq 2p_3, \tag{4b} $$

$$ x_1 + x_2 \geq y_1 + y_2 \tag{4c} $$

(from the condition that we have a lattice permutation); and

$$ m_1 + m_2 + m_3 + 2p_1 + 2p_2 + 2p_3 = h_1 + h_2 + h_3, \tag{5a} $$

$$ m_2 + x_2 + y_1 = h_2, \tag{5b} $$

$$ m_3 + x_3 + y_2 + 2p_3 = h_3 \tag{5c} $$

(from the condition that the product yield $Y_{h_1 h_2 h_3}$).

Using Eqs. (2) and (5), we may express x_1, x_2, y_1, and y_2 in terms of the other variables and eliminate them from the inequalities (3) and (4). The conditions $x_1, x_2, y_1, y_2 \geq 0$ then provide additional inequalities on the remaining variables. After some rearrangement, we finally obtain

$$ \min(h_2 - 2p_2, h_3 - 2p_3) \geq m_3 + x_3 $$

$$ \geq \max(h_2 + h_3 - h_1 - 2p_2 - 2p_3, h_3 - 2p_2), \tag{6a} $$

$$ h_1 - 2p_1 \geq m_3, \tag{6b} $$

$$ 2p_1 - 2p_2 \geq x_3, \tag{6c} $$

$$ \min(h_1 + x_3 - 2p_1, h_2 + h_3 - 2p_2 - 2p_3 - m_3 - x_3) $$

$$ \geq m_2 \geq \max(h_2 + h_3 - 2p_1 - 2p_2 - 2p_3 - m_3, $$

$$ m_3 + x_3, h_2 + x_3 - 2p_1), \tag{6d} $$

$$ m_1 = h_1 + h_2 + h_3 - 2p_1 - 2p_2 - 2p_3 - m_2 - m_3. \tag{6e} $$

To use expressions (6), one considers a diagram $Y_{2p_1 2p_2 2p_3}$ such that $2p_i \leq h_i$, and then computes m_3 and x_3 from (6a)–(6c), m_2 from (6d), and finally m_1 from the equality (6e). For given h_1, h_2, and h_3, one obtains numbers $p_1 p_2 p_3$, $m_1 m_2 m_3$, and x_3, and the

multiplicity of ${}^nD_{m_1m_2m_3}$ is just the number of different sets of such numbers with the same $m_1m_2m_3$ values.

For $n \geq 6$, the irreducible representations of $O(n)$ corresponding to the diagrams $Y_{m_1m_2m_3}$ occurring in Littlewood's theorem were just the irreducible representations ${}^nD_{m_1m_2m_3}$. For $n < 6$, this is no longer true, since p is now, at most, 2. The following correspondences, based on Murnaghan's modification rules,[2,13,14] furnish the irreducible representations of $O(n)$ occurring in Littlewood's theorem for $p = 2$:

(i) $U(4) \downarrow O(4)$:

for $m_3 = 0$, $\quad Y_{m_1m_20} \rightarrow D_{m_1m_2}$,

for $m_3 \neq 0$, the following conversions:

$$Y_{m_1m_21} \text{ null (to be omitted) if } m_2 > 1,$$
$$Y_{m_111} \rightarrow D_{m_10}, \quad Y_{m_1m_22} \rightarrow -D_{m_1m_2},$$
$$Y_{m_1m_2m_3} \text{ null when } m_3 > 2;$$

(ii) $U(5) \downarrow O(5)$:

for $m_3 = 0$, $\quad Y_{m_1m_20} \rightarrow D_{m_1m_2}$,

$$Y_{m_1m_21} \rightarrow D_{m_1m_2}, \quad Y_{m_1m_2m_3} \text{ null } m_3 > 1.$$

In terms of the multiplicities, we then have for case (i) that $\nu'^{h_1h_2h_3}_{m_10}$ is the number of solutions of (6) with $m_2 = m_3 = 0$ plus the number of solutions of (6) with $m_2 = m_3 = 1$, while $\nu'^{h_1h_2h_3}_{m_1m_2>1}$ is the number of solutions of (6) with $m_3 = 0$ minus the number of solutions of (6) with $m_3 = 2$. For case (ii) $\nu'^{h_1h_2h_3}_{m_1m_2>0}$ is the number of solutions of (6) with $m_3 = 0$ plus the number of solutions of (6) with $m_3 = 1$.

The final step is furnished by the branching rule for $O(n) \downarrow SO(n)$:

$${}^{2p}D'_{m_1m_2\cdots m_p} \downarrow SO(2p) = {}^{2p}D_{m_1\cdots m_p}, \quad m_p = 0,$$
$$= {}^{2p}D_{m_1\cdots m_p} + {}^{2p}D_{m_1\cdots -m_p}, \quad m_p > 0;$$

${}^{2p+1}D'_{m_1m_2\cdots m_p}$ remains irreducible upon restriction to $SO(2p+1)$.

We then see that the multiplets $\nu^{h_1h_2h_3}_{m_1m_2m_3}$ in

$${}^n\Gamma_{h_1h_2h_3} \downarrow SO(n) = \sum_{m_1m_2m_3} \nu^{h_1h_2h_3}_{m_1m_2m_3} {}^nD_{m_1m_2m_3}$$

are given by

$$\nu^{h_1h_2h_3}_{m_1m_2\pm m_3} = \nu'^{h_1h_2h_3}_{m_1m_2m_3}, \quad \text{for} \quad n = 2p \geq 6,$$

and

$$\nu^{h_1h_2h_3}_{m_1m_2m_3} = \nu'^{h_1h_2h_3}_{m_1m_2m_3}, \quad \text{for} \quad n = 2p+1;$$

For the special case $n = 4$,

$$\nu^{h_1h_2h_3}_{m_1,\pm m_2} = \nu'^{h_1h_2h_3}_{m_1m_2}.$$

ACKNOWLEDGMENT

We would like to thank Professor M. Moshinsky for pointing out the usefulness of Littlewood's theorem.

[1] D. E. Littlewood, *The Theory of Group Characters* (Oxford University Press, London, 1958), p. 240.

[2] F. D. Murnaghan, *The Theory of Group Representations* (The Johns Hopkins Press, Baltimore, 1938), pp. 282ff.

[3] M. Moshinsky and V. S. Devi, J. Math. Phys. **10**, 455 (1969). These authors have employed Littlewood's theorem to determine the branching rules and the state vector for $U(3) \supset O(3)$.

[4] M. L. Whippman [J. Math. Phys. **6**, 1534 (1965)] has surveyed branching rules for groups of low rank useful in elementary particle physics.

[5] Branching rules for $SU(3) \downarrow SO(3)$ and $SU(4) \downarrow SO(4)$ were first derived by G. Racah, Rev. Mod. Phys. **21**, 494 (1949).

[6] P. Kramer and M. Moshinsky, Nucl. Phys. **82**, 241 (1966).

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

[8] M. L. Brunet, Bull. Am. Phys. Soc., Ser. II, **14**, 503 (1969).

[9] M. Brunet and M. Resnikoff, J. Math. Phys. **11**, 1474 (1970).

[10] E. Wigner, Phys. Rev. **51**, 106 (1937).

[11] K T. Hecht, Phys. Rev. **139**, B794 (1965) and Nucl. Phys. **102**, 11 (1967); Sing Chin Pang, Dept. of Physics, University of Michigan, Ann Arbor, Mich., Technical Report, November, 1967. This latter reference has a complete bibliography.

[12] H. Boerner, *Representations of Groups* (North-Holland Publishing Co., Amsterdam, 1963).

[13] Murnaghan, Ref. 2, has the modification rules for $r = v + 2$.

[14] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan Books, Washington, D.C., 1962), pp. 129ff. This book contains implicit modification rules for $p \leq r \leq 2p$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

The Representations of $U(4) \supset U(2) \otimes U(2)$

M. BRUNET AND M. RESNIKOFF*
Department of Physics, State University of New York at Buffalo, Buffalo, New York 14214

(Received 14 April 1969)

The state vectors associated with an irreducible representation of $U(4)$ restricted to the subgroup $U(2) \otimes U(2)$ are explicitly constructed for the general 4-rowed Young diagram, employing Littlewood's rules and a theorem of Cartan. Expressions are given for the degeneracy of a $U(2) \otimes U(2)$ state in $U(4)$, and the set of $U(2) \otimes U(2)$ states obtained are shown to be independent and complete. Operators which break this degeneracy are constructed, and the behavior of the state vector under the conjugation operations is also discussed. The results are useful for the nuclear quasispin and Wigner supermultiplet models.

1. INTRODUCTION

The decomposition of unitary groups with respect to the canonical chain of subgroups $U(n) \supset U(n-1) \supset \cdots \supset U(1)$ is well known in the literature.[1] Its importance stems, of course, from the physical applications of $SU(3)$, where the isospin subgroup $SU(2)$ is fully contained. There are other physical applications, however, which employ noncanonical decompositions. In the nuclear-shell model proposed by Elliott,[2] $U(3)$ is restricted to the angular-momentum subgroup $O(3)$. The mathematical implications of the $U(3) \supset O(3)$ problem have been investigated by several authors.[3] Kramer and Moshinsky,[4] in an extension of the Elliott model, have studied the restriction of $U(n)$ to $S(n)$, the symmetric group in n dimensions.

In nuclear physics, particularly in Wigner supermultiplet theory[5] and quasispin models,[6] frequent use is made of the noncanonical decomposition $U(4) \supset U(2) \otimes U(2)$. The reduction of $U(n)$ or $SU(n)$ with respect to the product of subgroups $U(mn) \supset U(m) \otimes U(n)$, has not been extensively studied, though this decomposition presently has much greater physical application. In high-energy physics, the restriction of $SU(6)$ to the subgroup $SU(3) \otimes SU(2)$ is of major importance.[7]

In addition to the physical implications, there is considerable mathematical interest in the problem of restricting $U(mn)$ to $U(m) \otimes U(n)$. The subgroup $U(m) \otimes U(n)$ does not provide sufficient row labels to uniquely specify a state of $U(mn)$. Thus, a given state of $U(m) \otimes U(n)$ occurs in a representation of $U(mn)$ with a certain degeneracy ν. The problem then is to determine which representations of $U(m) \otimes U(n)$ occur in a given representation $U(mn)$, the degeneracy ν, and the operators from the full group $U(mn)$ which break this degeneracy.

It is the purpose of this paper to answer the above questions for the particular problem $U(4)$ restricted to the subgroup $U(2) \otimes U(2)$. We shall, in addition, explicitly construct the state vector. This state vector may then be used to calculate the general $U(4) \supset U(2) \otimes U(2)$ coupling and fractional parentage coefficients,[8] though this calculation is not carried out in this paper.

The calculation of the highest-weight state-vector for $U(4)$ restricted to $U(2) \otimes U(2)$ proceeds as follows. The branching rules for $U(4) \supset O(4)$ are set down in Sec. 2, providing a reduction of the lowest-dimensional representations of $U(4)$ with respect to $U(2) \otimes U(2)$. The highest-weight states associated with certain of these low-dimensional reduced $U(4)$ representations we shall call the fundamental states. Then, using a theorem of Cartan,[9] we construct the most general state of highest weight as a product of fundamental states. The fundamental states are independent, though certain products of fundamental states are dependent. We show that the general state vector of highest weight, subject to these dependency conditions, yields a complete set of states. In Sec. 3 we associate each fundamental state uniquely with a polynomial of the C_4 vectors $\mathbf{z}_i = (z_i^1, \cdots, z_i^4)$, $i = 1, 2, 3$, and thus construct the explicit polynomial state vector and the operators which break the degeneracy of the $U(2) \otimes U(2)$ states.

2. THE FUNDAMENTAL STATES; STATE VECTOR

An irreducible representation of $U(4)$ may be labeled by the partition (h_1, h_2, h_3, h_4), $h_1 \geq h_2 \geq h_3 \geq h_4 \geq 0$, or by a 4-rowed Young diagram where h_i is the number of boxes in the ith row. Alternatively, we label a representation with the numbers $\lambda_1 = h_1 - h_2$, $\lambda_2 = h_2 - h_3$, $\lambda_3 = h_3 - h_4$, and $\lambda_4 = h_4$, the overhang of successive rows of the Young diagram. With no loss of generality, we take $h_4 = 0$ in the following discussion and discuss the trivial modifications of the state vector for $h_4 \neq 0$ at the end of Sec. 3.

An irreducible representation of $U(2) \otimes U(2)$ is labeled by the partitions (j_1, μ_1), (j_2, μ_2), where $2j$ is the overhang of the first row, and μ is the number of boxes in the second row of the two-rowed Young diagram. A representation of $U(4)$ labeled by $(\lambda_1\ \lambda_2\ \lambda_3)$ shall be written $\mathfrak{D}_{\lambda_1\lambda_2\lambda_3}$, and the polynomial of highest weight associated with this representation $P(\lambda_1, \lambda_2, \lambda_3)$.

The general $U(4)$ representation may be constructed using Cartan's theorem[9]: An irreducible representation is determined by the highest weight,[10] which is always simple. The highest weight of a Kronecker product of representations[11] $\mathfrak{D}_{\lambda_1\lambda_2\lambda_3} \otimes \mathfrak{D}_{\lambda_1'\lambda_2'\lambda_3'}$ of highest weight is the representation $\mathfrak{D}_{\lambda_1''\lambda_2''\lambda_3''}$, where $\lambda_i'' = \lambda_i + \lambda_i'$, $i = 1, 2, 3$. An analogous theorem holds for the $U(2)$ representations. Cartan's theorem, applied to the polynomials of highest weight in the restriction $U(4)$ to the subgroup $U(2) \otimes U(2)$, reads

$$P(\lambda_1\lambda_2\lambda_3; j_1j_2)P(\lambda_1'\lambda_2'\lambda_3'; j_1'j_2') = P(\lambda_1 + \lambda_1', \lambda_2 + \lambda_2', \lambda_3 + \lambda_3'; j_1 + j_1', j_2 + j_2'). \tag{2.1}$$

Since the weights are linear forms with real coefficients, the result for the $U(2)$ subgroup indices follows.

Using Eq. (2.1), we may obtain the most general highest-weight by taking products of certain of the low-dimensional $U(4) \supset U(2) \otimes U(2)$ highest-weight states.[12,13] We shall call these states the fundamental states,[14] since they cannot be obtained as a product of other low-dimensional highest-weight states.

The fundamental states may be constructed directly from the branching rules for $U(4) \supset O(4) \simeq O(3) \otimes O(3)$ from Brunet and Resnikoff.[15] For convenience we set down the results here. A representation of $O(4)$ may be written $D_{m_1m_2}$, but for purposes of stating the branching rules we write $D_{m_1m_2m_3}$ and make the appropriate modifications to the two-weight representation below. The multiplicity $\nu(h_i, m_i)$ of an $O(4)$ representation $D_{m_1m_2m_3}$ in a $U(4)$ representation $\mathfrak{D}_{\lambda_1\lambda_2\lambda_3}$ is the number of solutions of the inequalities:

$$\begin{aligned}
&\min(h_2 - 2p_2, h_3 - 2p_3) \geq m_3 + x_3 \\
&\qquad \geq \max(h_2 + h_3 - h, -2p_2 - 2p_3, h_3 - 2p_2), \\
&\qquad h_1 - 2p_1 \geq m_3 \geq 0, \quad 2(p_1 - p_2) \geq x_3 \geq 0, \\
&\min(h_1 + x_3 - 2p_1, h_2 + h_3 - 2p_2 - 2p_3 - m_3 - x_3) \\
&\qquad \geq m_2 \geq \max(h_2 + h_3 - 2p_1 - 2p_2 - 2p_3 - m_3, \\
&\qquad\qquad m_3 + x_3, h_2 + x_3 - 2p_1), \\
&m_1 = h_1 + h_2 + h_3 - 2p_1 - 2p_2 - 2p_3 - m_2 - m_3.
\end{aligned} \tag{2.2}$$

The multiplicity of $D_{m_1m_2}$, $m_2 \neq 0$, in $\mathfrak{D}_{\lambda_1\lambda_2\lambda_3}$ is the number of solutions of (2.2) with $m_3 = 0$, minus the number of solutions with $m_3 = 2$. The multiplicity of D_{m_10} in $\mathfrak{D}_{\lambda_1\lambda_2\lambda_3}$ is the number of solutions with m_2, $m_3 = 0$, plus the number of solutions with m_2, $m_3 = 1$. All other solutions are null. Using Eq. (2.2), we have listed in Table I the possible lower-dimensional states which will be useful in the following discussion.

Using Eq. (2.1), we may obtain the highest-weight state associated with the representation $(\lambda_1\lambda_2\lambda_3; j_1j_2)$ by taking products of the fundamental states $S_1, \cdots, S_{13}$ listed in Table I. This polynomial $P(\lambda_1\lambda_2\lambda_3; j_1j_2)$ may be written

$$P(\lambda_1\lambda_2\lambda_3; j_1j_2) = T_{n_i} = \prod_{i=1}^{13} (S_i)^{n_i}, \tag{2.3a}$$

where

$$\begin{aligned}
\lambda_1 &= n_1 + n_3 + 2n_5 + n_{10} + n_{11} + n_{12} + n_{13}, \\
\lambda_2 &= n_3 + n_4 + 2n_6 + n_8 + n_9 + n_{12} + n_{13}, \\
\lambda_3 &= n_2 + n_4 + 2n_7 + n_{10} + n_{11} + n_{12} + n_{13},
\end{aligned} \tag{2.3b}$$

TABLE I. Low-dimensional $U(4) \supset U(2) \otimes U(2)$ representations. This table displays the possible $U(2) \otimes U(2)$ representations $(\lambda_1\lambda_2\lambda_3; j_1j_2)$ contained in $U(4)$ representations, and the C_4 polynomials S_i corresponding to a fundamental state.

$S_1 = (1, 0, 0; \frac{1}{2}, \frac{1}{2})$	$S_8 = (0, 1, 0; 1, 0)$
$S_2 = (0, 0, 1; \frac{1}{2}, \frac{1}{2})$	$S_9 = (0, 1, 0; 0, 1)$
$S_3 = (1, 1, 0; \frac{1}{2}, \frac{1}{2})$	$S_{10} = (1, 0, 1; 10)$
$S_4 = (0, 1, 1; \frac{1}{2}, \frac{1}{2})$	$S_{11} = (1, 0, 1; 0, 1)$
$(0, 1, 1; \frac{3}{2}, \frac{1}{2}) = (0, 0, 1; \frac{1}{2}, \frac{1}{2})(0, 1, 0; 1, 0) = S_2S_8$	$(1, 0, 1; 1, 1) = (1, 0, 0; \frac{1}{2}, \frac{1}{2})(0, 0, 1; \frac{1}{2}, \frac{1}{2}) = S_1S_2$
$(0, 1, 1; \frac{1}{2}, \frac{3}{2}) = (0, 0, 1, \frac{1}{2}, \frac{1}{2})(0, 1, 0; 0, 1) = S_2S_9$	$S_{12} = (1, 1, 1; 1, 0)$
$S_5 = (2, 0, 0; 0, 0)$	$S_{13} = (1, 1, 1; 0, 1)$
$(2, 0, 0; 1, 1) = (1, 0, 0; \frac{1}{2}, \frac{1}{2})^2 = (S_1)^2$	$(1, 1, 1; 21) = (1, 0, 0; \frac{1}{2}, \frac{1}{2})(0, 1, 0; 1, 0)(0, 0, 1; \frac{1}{2}, \frac{1}{2}) = S_1S_8S_2$
$S_6 = (0, 2, 0; 0, 0)$	$(1, 1, 1; 20) = (1, 0, 1; 1, 0)(0, 1, 0; 1, 0) = S_{10}S_8$
$(0, 2, 0; 2, 0) = (0, 1, 0; 1, 0)^2 = (S_8)^2$	$(1, 1, 1; 1, 1) = (1, 0, 0; \frac{1}{2}, \frac{1}{2})(0, 1, 1; \frac{1}{2}, \frac{1}{2}) = S_1S_4$
$(0, 2, 0; 1, 1) = (0, 1, 0; 1, 0)(0, 1, 0; 0, 1) = S_8S_9$	$= (0, 0, 1; \frac{1}{2}, \frac{1}{2})(1, 1, 0; \frac{1}{2}, \frac{1}{2}) = S_2S_3$
$(0, 2, 0; 0, 2) = (0, 1, 0; 0, 1)^2 = (S_9)^2$	$= (1, 1, 0; 1, 0)(1, 0, 1; 0, 1) = S_9S_{11}$
$S_7 = (0, 0, 2; 0, 0)$	$= (0, 1, 0; 0, 1)(1, 0, 1; 1, 0) = S_9S_{10}$
$(0, 0, 2; 1, 1) = (0, 0, 1; \frac{1}{2}, \frac{1}{2})^2 = (S_2)^2$	$(1, 1, 1; 0, 2) = (1, 0, 1; 0, 1)(0, 1, 0; 0, 1) = S_{11}S_9$
	$(1, 1, 1; 1, 2) = (1, 0, 0; \frac{1}{2}, \frac{1}{2})(0, 1, 0; 0, 1)(0, 0, 1; \frac{1}{2}, \frac{1}{2}) = S_1S_9S_2$

and

$$2j_1 = n_1 + n_2 + n_3 + n_4 + 2(n_8 + n_{10} + n_{12}),$$
$$2j_2 = n_1 + n_2 + n_3 + n_4 + 2(n_9 + n_{11} + n_{13}). \quad (2.3c)$$

Equations (2.3b) and (2.3c) are obtained by simply adding the weights according to the prescription (2.1). The partitions μ_1, μ_2 do not provide further conditions on the integers n_i, as shown in Sec. 3.

The highest weights j_i, $i = 1, 2$ are integer or half-integer depending on the parity of $\lambda_1 + \lambda_3$. The sum of λ_1, λ_3 is

$$\lambda_1 + \lambda_3 = (n_1 + n_2 + n_3 + n_4) + 2(n_5 + n_7 + n_{10} + n_{11} + n_{12} + n_{13}),$$

using Eq. (2.3b). We see that if $\lambda_1 + \lambda_3$ is even (odd), $n_1 + n_2 + n_3 + n_4$ is even (odd), and $2j_i$, using (2.3a), is even (odd).

The highest-weight polynomials (2.3a) are not yet independent, without further conditions upon the integers n_i. The reason is that certain products of fundamental states are dependent on other products of fundamental states. By considering products of fundamental states and comparing the results to the branching rules (2.2), the following products can be shown to be dependent[16]:

$$(S_3)^2,\ (S_4)^2,\ S_3S_4,\ (S_{12})^2,\ (S_{13})^2,\ S_{12}S_{13},\ S_1S_4,\ S_2S_3,\ S_{10}S_{11},\ S_1S_2S_6,\ S_1S_{12},\ S_1S_{13},\ S_3S_{12},\ S_3S_{13},\ S_{11}S_{12},\ S_2S_{13},\ S_2S_{12},\ S_4S_{13},\ S_4S_{12},\ S_{10}S_{13}. \quad (2.4)$$

The explicit functional dependence will be established in the next section where the polynomial form of the fundamental states is given.

All of the products in (2.4) may be re-expressed in terms of products of other fundamental states, and are to be eliminated in Eq. (2.3a). Equation (2.4) therefore implies further conditions on the integers n_i. Since $(S_i)^2$, $i = 3, 4, 12, 13$ is dependent, this implies that the term in (2.3a) with $n_i = 2$ may be re-expressed. Thus, we get from (2.4) the following conditions:

$$n_i = 0, 1, \quad i = 3, 4, 12, 13. \quad (2.5a)$$

The remainder of the products (2.4) have the form S_iS_j or $S_1S_2S_6$. This implies that whenever the exponents n_i and n_j are both greater than zero, the term may be re-expressed until either n_i or n_j is zero. Hence the terms $n_in_j \neq 0$, meaning both n_i and $n_j \neq 0$, is eliminated from Eq. (2.3a). This eliminates the following terms:

$$n_3n_4 \neq 0, \quad n_{12}n_{13} \neq 0, \quad n_{10}n_{11} \neq 0, \quad n_1n_2n_6 \neq 0, \quad (2.5b)$$

$$n_1n_4 \neq 0, \quad n_2n_3 \neq 0, \quad (2.5c)$$

$$n_1n_{12} \neq 0, \quad n_3n_{12} \neq 0, \quad n_{11}n_{12} \neq 0,$$
$$n_1n_{13} \neq 0, \quad n_3n_{13} \neq 0, \quad (2.5d)$$
$$n_2n_{13} \neq 0, \quad n_4n_{13} \neq 0, \quad n_{10}n_{13} \neq 0,$$
$$n_2n_{12} \neq 0, \quad n_4n_{12} \neq 0. \quad (2.5e)$$

Equations (2.3b), (2.3c), and (2.5) are constraints on the integers n_i. For a given $(\lambda_1\lambda_2\lambda_3; j_1j_2)$, the number of sets of integers solving these equations is the degeneracy $\nu(\lambda; j_1j_2)$, the number of times a representation of $U(2) \otimes U(2)$ occurs in a representation of $U(4)$.

It remains to determine whether the construction of independent polynomials T_n, as discussed above, provides a complete set of $U(2) \otimes U(2)$ states. We must verify that the sum over $U(2) \otimes U(2)$ states, times the degeneracy $\nu(\lambda; j_1j_2)$ with which each $U(2) \otimes U(2)$ state occurs, yields the correct number of $U(4)$ states, for a given partition λ,

$$\sum (2j_1 + 1)(2j_2 + 1)\nu(\lambda; j_1j_2) = N(\lambda), \quad (2.6a)$$

where

$$N(\lambda) = \tfrac{1}{12}(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)(\lambda_1 + \lambda_2 + 2) \times (\lambda_2 + \lambda_3 + 2)(\lambda_1 + \lambda_2 + \lambda_3 + 3). \quad (2.6b)$$

In terms of the integers n_i, we must show that

$$\sum_{n_i} (n_1 + n_2 + n_3 + n_4 + 2(n_8 + n_{10} + n_{12}) + 1) \times (n_1 + n_2 + n_3 + n_4 + 2(n_9 + n_{11} + n_{13}) + 1) = N, \quad (2.7)$$

subject to the constraints (2.3b) and (2.5). This is a counting operation, but it is a nontrivial calculation. To sum Eq. (2.7), it is desirable to apply condition (2.5) and separate the sum into several cases:

(i) $n_{12} = 1$ implies $n_i = 0$, $i = 1, 2, 3, 4, 11, 13$,
(ii) $n_{13} = 1$ implies $n_j = 0$, $j = 1, 2, 3, 4, 10, 12$,
(iii) $n_{12}, n_{13} = 0$:
  (a) $n_4 = 1$ implies $n_1, n_3 = 0$,
  (b) $n_3 = 1$ implies $n_2, n_4 = 0$, (2.8)
  (c) $n_3, n_4 = 0$:
    (1) $n_1 = 0$,
    (2) $n_2 = 0, n_1 > 0$,
    (3) $n_6 = 0, n_1, n_2 > 0$.

All integers n_i not equal to zero in Eq. (2.8) are summed in Eq. (2.7), according to the constraints Eq. (2.3b). The classification Eq. (2.8) incorporates all the conditions Eq. (2.5), except the condition $n_{11}n_{10} \neq 0$, which further divides the above cases into $n_{11} = 0$, $n_{10} > 0$ and $n_{11} \geq 0$, $n_{10} = 0$. We note that if $\lambda_1 + \lambda_3$ is not even, terms (i) and (ii) are not present. The summation is carried out in the Appendix for λ_i even, and does indeed reduce to (2.6b).

We note that case (i) of Eq. (2.8) has six independent integers n_i, $i = 5, 6, 7, 8, 9, 10$. Applying the constraints (2.3b) and (2.3c), we reduce the number of independent integers n_i to one. This result also holds for case (ii). In case (iii), there are eight integers $n_i \neq 0$. The condition, $n_{10}n_{11} \neq 0$ eliminated, implies that there are seven integers. Applying the constraints (2.3b) and (2.3c), we reduce the number of independent integers n_i to two. It is well known that, at most, two operators are required to break the degeneracy and characterize the polynomials completely.[16]

Given T_{n_i} as the complete set of independent highest-weight $U(4) \supset U(2) \otimes U(2)$ states, we may write the general $U(4) \supset U(2) \otimes U(2)$ state-vector in the form

$$|\lambda_1\lambda_2\lambda_3; j_1 j_2\rangle_\tau = \sum_{n_i} A_{n_i}(\tau) T_{n_i}, \tag{2.9}$$

where the index τ distinguishes between degenerate $U(4) \supset U(2) \otimes U(2)$ states and the number of sets of integers n_i for a given $(\lambda_1\lambda_2\lambda_3; j_1 j_2)$ is $\nu(\lambda; j_1 j_2)$ (see Sec. 4).

3. $U(4) \supset U(2) \otimes U(2)$ POLYNOMIAL STATE-VECTOR

The state vector associated with an irreducible representation of $U(4)$ may be constructed as a polynomial function of the C_4 vectors $\mathbf{z}_i = (z_i^{(1)}, \cdots, z_i^{(4)})$, where $i = 1, \cdots, 4$. In terms of the vectors z_i, the $U(4)$ generators may be written

$$C^{\alpha\beta} = \sum_{i=1}^{m} z_i^\alpha \frac{\partial}{\partial z_i^\beta} \quad \alpha, \beta = 1, \cdots, 4, \tag{3.1}$$

where $m \leq 4$ is the number of nonzero rows of the $U(4)$ Young diagram. If we instead contract the upper indices α and β, we may form the operators C_{ij}:

$$C_{ij} = \sum_{\alpha=1}^{n} z_i^\alpha \frac{\partial}{\partial z_j^\alpha} \equiv \mathbf{z}_i \cdot \frac{\partial}{\partial \mathbf{z}_j}, \quad i, j = 1, \cdots, m. \tag{3.2}$$

Since the operators C_{ij} and generators $C^{\alpha\beta}$ commute, the operators C_{ij} may be used as Casimir operators to construct the state vector associated with an irreducible representation of $U(4)$, labeled by a Young diagram. Moshinsky[17] has proved that, if the state vector $|h, \alpha\rangle$ satisfies the conditions

$$C_{ii}|h, \alpha\rangle = h_i |h, \alpha\rangle, \quad i = 1, \cdots, m \leq n, \tag{3.3a}$$
$$C_{ij}|h, \alpha\rangle = 0, \qquad i < j, \tag{3.3b}$$

the state $|h, \alpha\rangle$ corresponds to an irreducible representation of $U(n)$. The parameters α are the row labels. The general form of the solution of Eqs. (3.3) for $U(4)$ is

$$|h, \alpha\rangle = (\Delta_{1234}^{1234})^{h_4} \sum_{h_{ij}} C_{h_{ij}} (\mathbf{z}_1)^{\lambda_1} (\Delta_{12})^{\lambda_2} (\Delta_{123})^{\lambda_3}, \tag{3.4}$$

where the $\Delta_{12}^{\mu\nu}$ are the 2×2 antisymmetric forms, $\Delta_{12}^{\mu\nu} = z_1^\mu z_2^\nu - z_1^\nu z_2^\mu$ and $\Delta_{123}^{\mu\nu\gamma}$ are the 3×3 antisymmetric forms

$$\Delta_{123}^{\mu\nu\gamma} = \det \begin{pmatrix} z_1^\mu & z_2^\mu & z_3^\mu \\ z_1^\nu & z_2^\nu & z_3^\nu \\ z_1^\gamma & z_2^\gamma & z_3^\gamma \end{pmatrix},$$

and Δ_{1234}^{1234} is the determinant of the 16 coordinates z_i^μ. Each factor in Eq. (3.4) is a product of factors

$$(\mathbf{z}_1)^{\lambda_i} \equiv \prod_{\mu=1}^{4} (z_1^\mu)^{h_{i\mu}},$$

and similarly for the other products suitably ordered, with

$$\sum_{\mu=1}^{4} h_{i\mu} = \lambda_i, \quad i = 1, 3, \quad \text{and} \quad \sum_{\mu=1}^{6} h_{2\mu} = \lambda_2.$$

The row labels α are determined by the subgroup $U(2) \otimes U(2)$, up to the degeneracy ν mentioned in Sec. 2. We denote the $U(2) \otimes U(2)$ subgroup with the parameters $(j_1\mu_1 m_1)$, $(j_2\mu_2 m_2)$, where the m_i are the weights $-j_i \leq m_i \leq j_i$. The explicit $U(2) \otimes U(2)$ generators in terms of the vectors $\mathbf{z}_i$ are[8,18]

$$(J_1)_0 = \tfrac{1}{2}(C^{11} + C^{33} - C^{22} - C^{44}),$$
$$(J_2)_0 = \tfrac{1}{2}(C^{11} + C^{22} - C^{33} - C^{44}), \tag{3.5a}$$
$$(J_1)_+ = C^{12} + C^{34}, \quad (J_2)_+ = C^{13} + C^{24}, \tag{3.5b}$$
$$(J_1)_- = C^{21} + C^{43}, \quad (J_2)_- = C^{31} + C^{42}, \tag{3.5c}$$

and the trace operator

$$\begin{aligned} \mathrm{Tr} &= C^{11} + C^{22} + C^{33} + C^{44} \\ &= C_{11} + C_{22} + C_{33} + C_{44}. \end{aligned} \tag{3.5d}$$

It is sufficient to consider the highest-weight $U(2) \otimes U(2)$ state $m_1 = j_1$, $m_2 = j_2$, in the following, an arbitrary-weight state m_1, m_2 being obtained with the lowering operators (3.5c).

The fundamental-weight states S_k are obtained from the general $U(4)$ state Eq. (3.4), by requiring

$$(J_i)_+ S_k = 0, \quad i = 1, 2, \quad k = 1, \cdots, 13, \tag{3.6a}$$
$$(J_i)_0 S_k = j_i S_k, \tag{3.6b}$$

where j_1, j_2 for each S_k are listed in Table I. We then get the explicit fundamental-state polynomials S_i as

$$\begin{aligned}
&S_1 = z^1, \quad S_2 = \Delta^{123}, \quad S_3 = z^1\Delta^{14} - z^1\Delta^{23} - 2z^3\Delta^{12}, \\
&S_4 = \Delta^{14}\Delta^{123} + \Delta^{23}\Delta^{123} - 2\Delta^{13}\Delta^{124}, \\
&S_5 = z^1 z^4 - z^2 z^3, \\
&S_6 = (\Delta^{23})^2 + (\Delta^{14})^2 - 2\Delta^{12}\Delta^{34} - 2\Delta^{13}\Delta^{24}, \\
&S_7 = \Delta^{123}\Delta^{234} - \Delta^{124}\Delta^{134}, \quad S_8 = \Delta^{12}, \quad S_9 = \Delta^{13}, \\
&S_{10} = z^1\Delta^{124} - z^2\Delta^{123}, \quad S_{11} = z^1\Delta^{134} + z^3\Delta^{123}, \\
&S_{12} = z^1\Delta^{14}\Delta^{124} - z^1\Delta^{24}\Delta^{123} + z^2\Delta^{23}\Delta^{123} \\
&\qquad - z^2\Delta^{12}\Delta^{134} + z^1\Delta^{12}\Delta^{234} - z^2\Delta^{13}\Delta^{124}, \\
&S_{13} = z^1\Delta^{14}\Delta^{134} + 2z^1\Delta^{34}\Delta^{123} - z^2\Delta^{13}\Delta^{134} \\
&\qquad - z^3\Delta^{13}\Delta^{124} + z^3\Delta^{23}\Delta^{123}.
\end{aligned} \tag{3.7}$$

The lower indices have been suppressed.

Given the explicit polynomials S_i, we may determine the relations between products of polynomials. We use the following identities:

$$\begin{gathered} z^i\Delta^{jk} + z^k\Delta^{ij} = z^j\Delta^{ik}, \quad i<j<k, \quad ijk = 1, \cdots, 4, \\ \Delta^{12}\Delta^{34} + \Delta^{14}\Delta^{23} = \Delta^{13}\Delta^{24}, \\ \Delta^{14}\Delta^{123} + \Delta^{12}\Delta^{134} = \Delta^{13}\Delta^{124}, \\ \Delta^{24}\Delta^{123} + \Delta^{12}\Delta^{234} = \Delta^{23}\Delta^{124}, \\ z^1\Delta^{234} - z^2\Delta^{134} + z^3\Delta^{124} - z^4\Delta^{123} = 0. \end{gathered} \tag{3.8}$$

We then obtain the following relations between products of polynomials:

$$\begin{aligned} (S_3)^2 &= (S_1)^2S_6 - 4S_5S_8S_9, \\ (S_4)^2 &= (S_2)^2S_6 - 4S_7S_8S_9, \\ S_3S_4 &= -(S_9S_{12} + S_8S_{13}), \end{aligned} \tag{3.9a}$$

$$\begin{aligned} (S_{12})^2 &= S_6(S_{10})^2 + 4S_5S_7(S_8)^2, \\ (S_{13})^2 &= S_6(S_{11})^2 + 4S_5S_7(S_9)^2, \\ S_{12}S_{13} &= 4S_5S_7S_8S_9 - ((S_1)^2S_7 + (S_2)^2S_5)S_6, \end{aligned} \tag{3.9b}$$

$$\begin{aligned} S_1S_4 &= S_9S_{10} - S_8S_{11}, \\ S_2S_3 &= -(S_9S_{10} + S_8S_{11}), \end{aligned} \tag{3.9c}$$

$$\begin{aligned} S_{10}S_{11} &= (S_2)^2S_5 - (S_1)^2S_7, \\ S_1S_2S_6 &= S_9S_{12} - S_8S_{13}, \end{aligned} \tag{3.9d}$$

$$\begin{aligned} S_1S_{12} &= 2S_2S_5S_8 + S_3S_{10}, \\ S_3S_{12} &= 2S_4S_5S_8 + S_1S_6S_{10}, \\ S_1S_{13} &= -2S_2S_5S_9 + S_3S_{11}, \\ S_3S_{13} &= 2S_4S_5S_9 + S_1S_6S_{11}, \\ S_{11}S_{12} &= -(S_1S_3S_7 + S_2S_4S_5), \end{aligned} \tag{3.9e}$$

$$\begin{aligned} S_2S_{13} &= S_4S_{11} - 2S_1S_7S_9, \\ S_4S_{13} &= 2S_3S_7S_9 + S_2S_6S_{11}, \\ S_2S_{12} &= -S_4S_{10} + 2S_1S_7S_8, \\ S_4S_{12} &= 2S_3S_7S_8 - S_2S_6S_{10}, \\ S_{10}S_{13} &= -S_1S_3S_7 + S_2S_4S_5. \end{aligned} \tag{3.9f}$$

The explicit state-vector is then given by Eq. (2.9), where S_i are the polynomials (3.7). The state vector for $h_4 \neq 0$ is Eq. (2.9) multiplied by the 4×4 determinant $(\Delta^{1234}_{1234})^{h_4}$. The degeneracy ν and the highest weights j_1, j_2 remain the same, but partition numbers μ_1, μ_2 are altered.

We note, finally, a relation between the $U(4)$ trace and the highest weights j_1 and j_2. Under a unitary transformation of $U(2)_1$, the fundamental states with $j_1 = 0$, $S_5, S_6, S_7, S_9, S_{11}, S_{13}$ remain invariant. Thus,

$$\mu_1 = (n_5 + 2n_6 + 3n_7 + n_9 + 2n_{11} + 2\lambda_4 + 3n_{13}).$$

Under a unitary transformation of $U(2)_2$, the fundamental states with $j_2 = 0$ remain invariant; hence,

$$\mu_2 = (n_5 + 2n_6 + 3n_7 + n_8 + 2n_{10} + 2\lambda_4 + 3n_{12}).$$

Given μ_1, μ_2 in terms of the integers n_i, we obtain the relations, using the trace operator (3.5d),

$$\begin{aligned} \mathrm{Tr} &= \lambda_1 + 2\lambda_2 + 3\lambda_3 + 4\lambda_4 \\ &= 2[j_1 + \mu_1 + n_2 + n_3 + n_{11} + 2(n_4 + n_{13})] \\ &= 2[j_2 + \mu_2 + n_2 + n_3 + n_{10} + 2(n_4 + n_{12})]. \end{aligned} \tag{3.10}$$

4. CONJUGATION, ADDITIONAL OPERATORS

The states $|\lambda_1\lambda_2\lambda_3; j_1j_2\rangle_\tau$ [Eq. (2.9)] are degenerate, since there are ν terms in the summation with arbitrary coefficients $A_{n_i}(\tau)$. The states may be classified according to their symmetry under conjugation and then orthonormalized by the Gram–Schmidt orthogonalization procedure. This will uniquely determine the coefficients $A_{n_i}(\tau)$. On the other hand, additional operators may be constructed from the full group, and the states diagonalized with respect to them.

A. Conjugation Operations

1. C Conjugation

There are two types of conjugation operations, C and R conjugation. The C-conjugation operations are an exchange of the vector components $z_i^m \leftrightarrow z_i^k$, $m, k = 1, \cdots, 4$. Acting on T_{n_i}, they yield a polynomial with the highest weights interchanged $j_n \leftrightarrow \pm j_p$, $n, p = 1, 2$, and the space $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ invariant.

C_1: Denote the exchange of vectors $z_i^2 \leftrightarrow z_i^3$ with z_i^1, z_i^4 unchanged, as the C_1 operation. Under the C_1 operation, we see from Eqs. (3.5a)–(3.5c) that the generators of $U(2)_1$ and $U(2)_2$ are interchanged; that is, $j_1 \leftrightarrow j_2$ are interchanged. We also see from Eq. (3.7) that

$$\begin{gathered} C_1S_i = +S_i, \quad i = 1, 3, 4, 5, 6, 7, \quad C_1S_2 = -S_2, \\ C_1S_8 = S_9, \quad C_1S_{10} = S_{11}, \quad C_1S_{12} = S_{13}. \end{gathered} \tag{4.1}$$

Thus, the polynomial $C_1(T_{n_i})$ has $j_1 \leftrightarrow j_2$ interchanged. If we also denote by $\mathcal{C}_1$ the operation which interchanges $n_8 \leftrightarrow n_9$, $n_{10} \leftrightarrow n_{11}$, and $n_{12} \leftrightarrow n_{13}$, then the polynomial $\mathcal{C}_1T_{n_i}$ also has $j_1 \leftrightarrow j_2$ interchanged. The product $(C_1\mathcal{C}_1)$ on T_{n_i} reproduces T_{n_i}, except for the phase $(-1)^{n_2}$. The state vectors $|\lambda; j_1j_2\rangle_\tau$ may then be separated into two cases, those even and odd under C_1 conjugation.

C_2: Denote the exchange of vectors $z_i^1 \leftrightarrow z_i^4$, with z_i^2, z_i^3 unchanged as the C_2 operation. Under the C_2

operation, we see from Eqs. (2.5a)–(2.5c) that $(J_1)_0 \leftrightarrow (-J_2)_0$ and $(J_1)_\pm \leftrightarrow (J_2)_\mp$ are interchanged; that is, $j_1 \leftrightarrow -j_2$. Thus the polynomial $C_2 T_{n_i}$ has the highest weights $j_1 \leftrightarrow -j_2$ exchanged.

C_3: Denote the exchange of vectors $z_i^1 \leftrightarrow z_i^2$, $z_i^3 \leftrightarrow z_i^4$ as the C_3 operation. We see from Eqs. (3.5a)–(3.5c) that $(J_1)_0 \to -(J_1)_0$, $(J_1)_+ \leftrightarrow (J_1)_-$, with $(J_2)_0$, $(J_2)_\pm$ unchanged. The effect of C_3 on the polynomial T_{n_i} is the exchanged $j_1 \to -j_2$ with j_2 unchanged or, on a general state, $m_1 \to -m_1$ with m_2 unchanged.

C_4: Similar to C_3, we have the operation which exchanges $m_2 \to -m_2$ with m_1 unchanged. We denote by C_4 the operation which exchanges the vectors $z_i^1 \leftrightarrow z_i^3$, $z_i^2 \leftrightarrow z_i^4$.

C_5: Finally, we denote by C_5 the operation which exchanges the vectors $z_i^1 \leftrightarrow z_i^4$ and $z_i^2 \leftrightarrow z_i^3$. Under the C_5 operation, the generators are interchanged

$$(J_1)_0 \leftrightarrow -(J_2)_0, \quad (J_i)_+ \leftrightarrow (J_i)_-, \quad i = 1, 2. \tag{4.2}$$

The C_5 operation on T_{n_i} gives a state on lowest weight $m_1 = -j_1$, $m_2 = -j_2$, with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ invariant.

2. R Conjugation

The explicit R-conjugation operation[19–21] for $U(4)$ [generalizable to $U(n)$] may be obtained from the Laplace expansion of the 4×4 determinant Γ:

$$\Gamma = \begin{vmatrix} Z_1^1 & Z_1^2 & Z_1^3 & Z_1^4 \\ Z_2^1 & Z_2^2 & Z_2^3 & Z_2^4 \\ Z_3^1 & Z_3^2 & Z_3^3 & Z_3^4 \\ Z_4^1 & Z_4^2 & Z_4^3 & Z_4^4 \end{vmatrix}$$

$$= Z_1^1 \Delta_{234}^{234} - Z_1^2 \cdot \Delta_{234}^{134} + Z_1^3 \cdot \Delta_{234}^{124} - Z_1^4 \cdot \Delta_{234}^{123} \tag{4.3a}$$

$$= \Delta_{12}^{12}\Delta_{34}^{34} - \Delta_{12}^{13}\Delta_{34}^{24} + \Delta_{12}^{14}\Delta_{34}^{23} + \Delta_{12}^{23}\Delta_{34}^{14} - \Delta_{12}^{24}\Delta_{34}^{13} + \Delta_{12}^{34}\Delta_{34}^{12}. \tag{4.3b}$$

Under a $U(4)$ unitary transformation U,

$$z_i'^{\alpha} = \mathfrak{U}_{\alpha'\alpha} z_i^{\alpha'}, \quad i = 1, \cdots, 4, \tag{4.4}$$

Γ is invariant

$$\Gamma' = (\det \mathfrak{U})\Gamma = e^{\phi}\Gamma. \tag{4.5}$$

If we examine the expansion of Γ [Eq. (4.3a)], we see that the variables z_1^α transform according to Eq. (4.4). In order for Eq. (4.5) to hold, we see that the 3×3 antisymmetric forms

$$\Delta_{234}^{(\alpha)} = (\Delta^{234}, -\Delta^{134}, \Delta^{124}, -\Delta^{123}) \tag{4.6a}$$

must transform as

$$\Delta_{234}'^{(\alpha)} = \overline{\mathfrak{U}}_{\alpha'\alpha} \cdot \Delta_{234}^{(\alpha')} \cdot (\det \mathfrak{U}) \tag{4.6b}$$

so that Γ transforms as Eq. (4.5). In other words, Δ_{234} must transform as the minor of z_1^α. We note that the z_i^α transforms as the representation $\mathfrak{D}^\lambda$, $\lambda = (1, 0, 0, 0)$, and Δ_{234}^α as the representation $\overline{\mathfrak{D}}^\lambda$ or its equivalent $\mathfrak{D}^{\lambda'}$, $\lambda' = (0, 0, 1, 0)$. Similarly, from Eq. (4.3b), we see that, if

$$\Delta_{12}^{(\alpha)} = (\Delta^{12}, \Delta^{13}, \Delta^{14}, \Delta^{23}, \Delta^{24}, \Delta^{34}) \tag{4.7a}$$

transforms as $\mathfrak{D}^\lambda$, $\lambda = (0, 1, 0, 0)$, then the 2×2 antisymmetric form $\Delta_{34_c}^{(\alpha)}$ transforms as the conjugate

$$\Delta_{34_c}^{(\alpha)} = (\Delta^{34}, -\Delta^{24}, \Delta^{23}, \Delta^{14}, -\Delta^{13}, \Delta^{12}). \tag{4.7b}$$

Thus, $\mathfrak{D}^\lambda$ and $\overline{\mathfrak{D}}^\lambda$, $\lambda = (0, 1, 0, 0)$, are related by a similarity transformation. We define, then, the R-conjugation operation as

$$R: \quad Z_1^{(\alpha)} \leftrightarrow \Delta_{123}^{(\alpha)}, \quad \Delta_{12}^{(\alpha)} \leftrightarrow \Delta_{12_c}^{(\alpha)}, \tag{4.8}$$

where Δ_{123} and Δ_{12_c} are defined by Eqs. (4.6a) and (4.7b), respectively.

Under the product of the operations RC_5, we see that $\lambda_1 \leftrightarrow \lambda_3$ are interchanged. The effect of RC_5 on the polynomials S_i is the interchange

$$RC_5: \ S_1 \leftrightarrow -S_2, \quad S_3 \leftrightarrow S_4, \quad S_5 \leftrightarrow -S_7, \quad S_{13} \to -S_{13}, \tag{4.9}$$

with the other polynomials S_i invariant. An inspection of Eq. (3.9) shows that the dependency relations are interchanged under RC_5 (and C_1), and no new relations appear. The operations C_1 and RC_5 are used in the Appendix to relate various sums.

B. Additional Operators

It is possible to construct operators from the full group which commute with the generators of the subgroups.[22] The method of construction for $U(4) \supset U(2) \otimes U(2)$ is the same as for $U(4) \supset R(4)$ [$R(4) = R(3) \otimes R(3)$], and both cases can be considered together.

The generators of $U(4)$ are C^{ij}, $i, j = 1, \cdots, 4$. We may take the symmetric and antisymmetric combinations

$$\Lambda^{ij} = \tfrac{1}{2}(C^{ij} - C^{ji}), \tag{4.10a}$$

$$Q^{ij} = \tfrac{1}{2}(C^{ij} + C^{ji}), \quad C^{ij} = \Lambda^{ij} + Q^{ij}. \tag{4.10b}$$

The six antisymmetric operators Λ^{ij} are the generators of $O(4)$. The Casimir operators C_{ij} commute with the generators of $U(4)$, $C^{\alpha\beta}$ and, in particular, commute with the generators of $O(4)$, Λ^{ij}, and the set of Q^{ij}. However, it is only necessary that an operator commute with the generators of the subgroup $O(4)$. Hence we may break up the Casimir operators C^P into parts which commute with the generators of the

subgroup $O(4)$. It may then be shown that the operators

$$Q^3 = Q^{ij}Q^{jn}Q^{ni}, \quad Q^4 = Q^{ij}Q^{jk}Q^{km}Q^{mi} \quad (4.11)$$

commute with the generators of $R(4)$. Whereas it is true that the chain $U(4) \supset O(4)$ does not provide sufficient row labels to uniquely classify a state, at the same time, because the group $O(4)$ is smaller, it is possible to find additional operators which commute with the smaller set of generators. Similarly, in the chain $U(4) \supset U(2) \otimes U(2)$, the subgroups $U(2)$ are formed of the six generators, Eqs. (3.5).

There are ten remaining generators of $U(4)$, similar to the set Q^{ij},

$$\begin{aligned} &C^{12} - C^{34}, \quad C^{21} - C^{34}, \quad \tfrac{1}{2}(C^{11} + C^{44} - C^{22} - C^{33}), \\ &C^{13} - C^{24}, \quad C^{31} - C^{42}, \quad \tfrac{1}{2}(C^{11} + C^{22} + C^{33} + C^{44}), \\ &C^{14}, \quad C^{41}, \quad C^{23}, \quad C^{32}. \end{aligned} \quad (4.12)$$

It is possible to construct operators like Q_3, Q_4 by contracting these ten generators, as in Eq. (4.11). These operators are similar to those constructed by Moshinsky and Nagel.[16] To proceed further, the state (2.9) should be diagonalized with respect to Q_3, Q_4. However, this cannot be done with raising or lowering operators in the conventional manner since, as shown by Racah,[23] for $U(3) \supset R(3)$, a state of $U(3)$ diagonalized with respect to an operator like Q_3 always yields a nonrational number.

ACKNOWLEDGMENTS

We would like to thank Dr. M. Moshinsky for suggesting the usefulness of Littlewood's theorem and the "permissible diagram" approach. We would also like to thank Dr. E. Chacon for pointing out an error in the original manuscript.

APPENDIX

In this Appendix we perform the sums (2.7) as divided into the cases (i)–(iii) of Eq. (2.8). We need the following summation formulas:

$$\begin{aligned} &\sum_{n=0}^{N} 1 = (N+1), \quad \sum_{n=0}^{N} n = \tfrac{1}{2}N(N+1), \\ &\sum_{n=0}^{N} (n)^2 = \tfrac{1}{6}N(N+1)(2N+1), \\ &\sum_{n=0}^{N} (n)^3 = [\tfrac{1}{2}N(N+1)]^2. \end{aligned} \quad (A1)$$

We shall assume in the Appendix that λ_i is even for $i = 1, 2, 3$. Consider first the cases (ii) and (iii). We first note that (iii) is the conjugate of (ii) and gives the same result. The case (ii) is

$$N_{(ii)} = \sum_{n_i} [2(n_8 + n_{10}) + 3](2n_9 + 1), \quad (A2)$$

with the constraints

$$\begin{aligned} \lambda_1 - 1 &= 2n_5 + n_{10}, \quad \lambda_3 - 1 = 2n_7 + n_{10}, \\ \lambda_2 - 1 &= 2n_6 + (n_8 + n_9). \end{aligned} \quad (A3)$$

We see that n_{10} is odd, summed from 1 to $\lambda_3 - 1$ ($\lambda_3 \leq \lambda_1$), and $(n_8 + n_9)$ is odd, summed from 1 to $\lambda_2 - 1$. Using (A1), we have the result

$$\begin{aligned} N_{(ii)} + N_{(iii)} = {} & \tfrac{1}{12}\lambda_2(\lambda_2 + 1)(\lambda_2 + 2)\lambda_3 \\ & \times (\lambda_2 + 2\lambda_3 + 5), \quad \lambda_3 \leq \lambda_1. \end{aligned} \quad (A4)$$

If $\lambda_3 > \lambda_1$, then λ_1 and λ_3 must be exchanged in (A4).

Consider next the case (ia):

$$\begin{aligned} N_{(ia)} = {} & \sum{}' [n_2 + 2(n_8 + n_{10}) + 2] \\ & \times [n_2 + 2(n_9 + n_{11}) + 2], \end{aligned} \quad (A5)$$

where

$$\begin{aligned} \lambda_1 &= 2n_5 + (n_{10} + n_{11}), \\ \lambda_2 - 1 &= 2n_6 + (n_8 + n_9), \\ \lambda_3 - 1 &= n_2 + 2n_7 + (n_{10} + n_{11}). \end{aligned} \quad (A6)$$

We note that $(n_{10} + n_{11})$ is even; hence n_2 is odd. The integers $(n_8 + n_9)$ are also odd. n_2 is summed 1 to $\lambda_2 - 1$; $n_{10} + n_{11}$ is summed 0 to min $(\lambda_1, \lambda_3 - 2)$. We must require, according to the dependency conditions equation (2.5), that $n_{10} \geq 0$, $n_{11} = 0$, and $n_{10} = 0$, $n_{11} > 0$. Using (A1), we obtain

$$\begin{aligned} N_{(iA)} = {} & \tfrac{1}{12}\lambda_2(\lambda_2 + 2)\lambda_3(\tfrac{1}{4}(\lambda_3)^3 + \tfrac{1}{3}(\lambda_3)^2(2\lambda_2 + 5) \\ & + \tfrac{1}{4}\lambda_3(\lambda_2)^2 + 6\lambda_2 + 7) - \tfrac{1}{3}(2\lambda_2 + 5) \end{aligned} \quad (A7)$$

for $\lambda_3 - 2 \leq \lambda_1$. Under RC_5 conjugation, this also provides us with $N_{(ib)}$ for the case $\lambda_1 - 2 \leq \lambda_3$.

Consider next the case (ib) for $\lambda_3 \leq \lambda_1 - 2$. The result will provide (ia) for $\lambda_1 \leq \lambda_3 - 2$. We have

$$\begin{aligned} N_{(ib)} = {} & \sum [n_1 + 2(n_8 + n_{10}) + 2] \\ & \times [n_1 + 2(n_9 + n_{11}) + 2], \end{aligned} \quad (A8)$$

where

$$\begin{aligned} \lambda_1 - 1 &= n_1 + 2n_5 + (n_{10} + n_{11}), \\ \lambda_2 - 1 &= 2n_6 + (n_8 + n_9), \\ \lambda_3 &= 2n_7 + (n_{10} + n_{11}). \end{aligned} \quad (A9)$$

Now $(n_{10} + n_{11})$ is even, summed from 0 to min $(\lambda_1 - 2, \lambda_3)$, and n_1 is odd, summed from 1 to $\lambda_1 - 1$; $(n_8 + n_9)$ is odd, summed from 1 to $\lambda_2 - 1$.

Using (A1), we obtain

$$N_{(ia)} = (i) + (ii), \tag{A10a}$$

$$\begin{aligned}(i) &= \tfrac{1}{8}\lambda_2(\lambda_2+2)\lambda_3(\lambda_3+2)\\ &\times \{[\tfrac{1}{6}(\lambda_3+1)^2 - \tfrac{1}{3}(\lambda_3+1)(\lambda_1+2) - 2]\\ &- \tfrac{1}{9}\lambda_3(2\lambda_2-1) - \tfrac{1}{18}[3(\lambda_2)^2 + 22\lambda_2 - 14]\};\\ (ii) &= \tfrac{1}{24}\lambda_1\lambda_2(\lambda_2+2)(\lambda_3+1)\\ &\times [(\lambda_1+\lambda_2+3)(\lambda_1+\lambda_2+2) + (\lambda_2+1)].\end{aligned} \tag{A10b}$$

Finally, we have the three cases (ic). We simply state the results here for $\lambda_3 \leq \lambda_1$:

$$\begin{aligned}N_{(ic)1} &= \tfrac{1}{4}(\lambda_3+2)(\lambda_2+2)\\ &\times \{\tfrac{1}{12}(\lambda_2)^2(\lambda_2+2)(\lambda_3+2) - \tfrac{1}{6}\lambda_2(\lambda_3+2)\\ &+ \tfrac{1}{12}(\lambda_2+2)(\lambda_3+1)(\lambda_3+2)(\lambda_3+3)\\ &+ \tfrac{1}{9}\lambda_2(2\lambda_2+5)(\lambda_3+1)(\lambda_3+3)\},\end{aligned} \tag{A11}$$

$$\begin{aligned}N_{(ic)2} &= \tfrac{1}{12}\lambda_1(\lambda_2+2)[\tfrac{1}{4}(\lambda_1+2)(\lambda_2+2)(\lambda_1^2+6\lambda_1-4)\\ &+ \tfrac{1}{4}\lambda_1(\lambda_2+1)(\lambda_2+2)(\lambda_2+3)\\ &+ \tfrac{1}{3}(\lambda_1+2)(\lambda_1-\tfrac{1}{2})\lambda_2(2\lambda_2+5)],\\ &\lambda_3 \leq \lambda_1 - 2,\end{aligned} \tag{A12}$$

$$\begin{aligned}N_{(ic)3} &= \tfrac{1}{4}\lambda_3(\lambda_2+1)\{\gamma_1 + 2(\lambda_2+1)\gamma_2,\\ &+ \gamma_3[1 + \tfrac{2}{3}\lambda_2(\lambda_2+2)]\},\\ \gamma_3 &= \lambda_1(\lambda_3+1) - \tfrac{1}{3}(\lambda_3-2)(\lambda_3+2),\\ \gamma_2 &= \tfrac{1}{2}\lambda_1\lambda_3 + \tfrac{1}{2}\lambda_1(\lambda_3+1)(\lambda_1+\lambda_3+2)\\ &- \tfrac{1}{6}(\lambda_3+2)(\lambda_3-2)(\lambda_1+\lambda_3+3),\\ \gamma_1 &= \tfrac{1}{3}(\lambda_1+1)(\lambda_3+1)(2\lambda_3+1)\\ &- \tfrac{3}{2} + \tfrac{1}{3}\lambda_1(\lambda_1+1)(\lambda_1+2)(\lambda_3+1)\\ &+ \tfrac{1}{2}(\lambda_1+1)^2(\lambda_3+1)^2\\ &- \tfrac{1}{12}\lambda_1\lambda_3(\lambda_3-2)(\lambda_1+\lambda_3)\\ &+ \tfrac{1}{6}\lambda_1\lambda_3(\lambda_3+1)(\lambda_3+2)\\ &+ \tfrac{1}{6}\lambda_3(\lambda_3-2)^2(\lambda_1+1)\\ &+ \tfrac{1}{6}(\lambda_3-1)(\lambda_3-2)(\lambda_1+2)^2\\ &+ \tfrac{8}{3} - \tfrac{1}{6}\lambda_3(\lambda_1+2)^2(2\lambda_3+1)\\ &- \tfrac{1}{6}(\lambda_1+\lambda_3+2)^2 + \tfrac{1}{4}\lambda_3(\lambda_3+2)(\lambda_1+2)^2\\ &- \tfrac{1}{2}(\lambda_3)^2(\lambda_1+2)[\tfrac{1}{3}(\lambda_1+1) + \tfrac{1}{2}(\lambda_3+2)],\\ &\lambda_3 \leq \lambda_1.\end{aligned} \tag{A13}$$

Though the algebra is messy, it is possible to add Eqs. (A4), (A7), (A10)–(A13), and obtain $N(\lambda)$, Eq. (2.6b). We point out the special case $\lambda_i = 10$, $N(\lambda) = (\lambda_i+1)^6 = (11)^6$. Then, $N_{(ii)} + N_{(iii)} = 38\,500$, $N_{(ia)} + N_{(ib)} = 298\,500$, $N_{(ic)} = 247\,256$, $N_{(ic)2} = 194\,060$, $N_{(ic)3} = 993\,245$, and the sum is $1\,771\,561 = (11)^6 = N(\lambda)$. We have not summed the other cases λ_i even or odd.

* Supported in part by the Research Foundation of the State University of New York and the National Science Foundation under Grant GP-9536.

[1] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963). This paper contains an extensive bibliography of the published literature.
[2] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128,562 (1958).
[3] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961); G. Racah, Rev. Mod. Phys. **21**, 494 (1949).
[4] P. Kramer and M. Moshinsky, Nucl. Phys. **82**, 241 (1966).
[5] E. P. Wigner, Phys. Rev. **51**, 106 (1937).
[6] A. K. Kerman, Ann. Phys. (N.Y.) **12**, (1961); K. Helmers, Nucl. Phys. **23**, 594 (1961); B. H. Flowers and S. Szpikowski, Proc. Phys. Soc. **84**, 193, 673 (1964); K. T. Hecht, Phys. Rev. **139**, B794 (1965); Nucl. Phys. **102**, 11 (1967); S.-C. Pang, Dept. of Physics, University of Michigan technical report, 1967.
[7] M. A. B. Beg and V. Singh, Phys. Rev. Letters **13**, 418 (1964).
[8] K. T. Hecht, Dept. of Physics, University of Michigan technical report, 1968.
[9] This theorem is discussed for the group $O(n)$ in H. Boerner, *Representations of Groups*, (North-Holland Publ. Co., Amsterdam, 1963), pp. 249, 256.
[10] A weight $m = (m_1, \cdots, m_4)$ is called higher than a weight m' if the first nonvanishing difference $m_i - m_i$ is greater than zero.
[11] Reference 9, p. 256.
[12] M. Moshinsky and V. Syamala Devi [J. Math. Phys. **10**, 455 (1969)] construct the states of $U(3) \supset R(3)$ by means of the "elementary permissible diagrams" and Littlewood's rules.
[13] R. T. Sharp and C. S. Lam [J. Math. Phys. **10**, 2033 (1969)] construct the "elementary multiplets" for a large number of subgroup restrictions.
[14] The notation "fundamental states" is suggestive of the fundamental representations of a group; an arbitrary representation can be constructed by the reduction of direct products of the fundamental representations.
[15] M. Brunet and M. Resnikoff, J. Math. Phys. **11**, 1471 (1970).
[16] M. Moshinsky and J. G. Nagel, Phys. Letters **5**, 173 (1963).
[17] M. Moshinsky, J. Math. Phys. **4**, 1125 (1963).
[18] J. D. Louck, J. Math. Phys. **6**, 1786 (1965) or M. Resnikoff, J. Math. Phys. **11**, (1970) (to appear)
[19] M. Gell-Mann, California Institute of Technology report, CTSL-20, 1961.
[20] M. Resnikoff, J. Math. Phys. **8**, 63 (1967). This reference contains the explicit R-conjugation operation on an $SU(3)$ state vector.
[21] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1723 (1964). This reference has an implicit definition of conjugation for $SU(n)$ states.
[22] M. Resnikoff, Nucl. Phys. **83**, 632 (1966). In this reference, the additional operations Q_3, Q_4, and Q_5 are constructed for the $U(n) \supset R(n)$ problem for a 3-rowed tableau. Bargmann and Moshinsky (Ref. 3) construct the operator Ω for $U(3) \supset R(3)$.
[23] G. Racah, *Istanbul Lecture Notes*, F. Gürsey, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Simple Waves in N-Dimensional Propagation

GUY BOILLAT
Université de Clermont, Département de Mathématiques, Clermont-Ferrand, France

(Received 24 July 1969)

A field is assumed to satisfy a quasilinear system of partial differential equations. Solutions are sought which only depend on a single function of the time and space coordinates.

SIMPLE WAVES

The field equations read

$$A^\alpha(\mathbf{u})\mathbf{u}_\alpha = 0, \quad \alpha = 0, 1, 2, \cdots, n,$$
$$A^0 = I, \quad \mathbf{u}_\alpha = \partial_\alpha \mathbf{u}; \tag{1}$$

the A^α are square matrices, ∂_α means partial differentiation with respect to the variable x^α ($x^0 = t$ is a time coordinate, the x^i, $i = 1, 2, \cdots, n$, are space coordinates), and summation is always understood over the whole range of repeated indices.

We look for a solution of the form

$$\mathbf{u} = \mathbf{u}(\varphi), \quad \varphi = \varphi(x^\alpha). \tag{2}$$

Inserting into (1), we get

$$A^\alpha \varphi_\alpha \mathbf{u}' = 0, \quad \varphi_\alpha = \partial_\alpha \varphi. \tag{3}$$

We introduce the normal speed through

$$\lambda = -\varphi_t/|\nabla\varphi|, \quad n = \nabla\varphi/|\nabla\varphi|.$$

Let $\lambda^{(k)}(\mathbf{u}, n)$ be an eigenvalue of the matrix $A_n = A^i n_i$ and let $\mathbf{d}_K^{(k)}(\mathbf{u}, n)$, $K = 1, 2, \cdots, m^{(k)}$, be the corresponding right eigenvectors. Equation (3) will be satisfied by

$$\psi^{(k)} = \varphi_t + |\nabla\varphi|\, \lambda^{(k)}(\mathbf{u}, n) = 0, \tag{4}$$

$$\frac{d\mathbf{u}}{d\varphi} = u^K(\varphi)\, \mathbf{d}_K^{(k)}(\mathbf{u}, n), \tag{5}$$

where the u^K are $m^{(k)}$ arbitrary functions of φ.

In the case of 1-dimensional propagation, $n =$ const, a solution of this system of equations is called by Lax a kth simple wave.[1,2] Here we assume that n is not constant, but, since the left member of (5) only depends on φ, the same must be true of the right member. Thus, we look for solutions of (4) satisfying $n = n(\varphi)$.

This means that we shall have

$$F(x^\beta)\varphi_\alpha = f_\alpha(\varphi), \quad F > 0,$$

with $n + 1$ differentiable functions f_α. [Indeed, by (4), such an equation holds for $\alpha = 0$ as well.] Therefore,

$$F\varphi_\alpha\, dx^\alpha = f_\alpha\, dx^\alpha = d(x^\alpha f_\alpha) - x^\alpha f'_\alpha\, d\varphi,$$
$$(F + x^\alpha f'_\alpha)\, d\varphi = d(x^\alpha f_\alpha).$$

Thus,

$$g(\varphi) + x^\alpha f_\alpha(\varphi) = 0. \tag{6}$$

Equation (4) yields

$$f_0 + \lambda^{(k)}\left(\sum_i f_i^2\right)^{\frac{1}{2}} = 0,$$

so that we can rewrite (6) as follows:

$$f(\varphi) + x \cdot n(\varphi) - \lambda(\varphi)t = 0, \tag{7}$$

with

$$n^2 = 1, \quad \lambda(\varphi) = \lambda(\mathbf{u}(\varphi), n(\varphi)). \tag{8}$$

A kth simple wave is a solution of Eqs. (5), (7), and (8). The wave is plane if we choose $n =$ const; it is centered if, moreover, $f(\varphi) = 0$ and $\lambda' \not\equiv 0$ (i.e., the wave is not exceptional; see below). If $f \not\equiv 0$, we can without restriction choose $f = \varphi$. [*Note added in proof:* In a recent discussion with E. Varley we became aware of a paper where he studies simple waves in general elastic materials; cf. E. Varley, Arch. Ratl. Mech. Anal. **20**, 309 (1965).]

A solution $\mathbf{u}$ takes on a constant value on each plane (7), $\varphi =$ const. At the intersection of two planes different values of the field conflict, thus leading to shocks. This phenomenon is well known in 1-dimensional propagation and, for instance, it has been investigated in detail in hydrodynamics[3] and magnetohydrodynamics[2]; it is due to the fact that, when the field is nonlinear, λ is a function of $\mathbf{u}$ and therefore, generally, of φ. The situation is somewhat different in n-dimensional space, since shocks might arise even in linear fields: the normal n depends on φ. Similar results are derived from the study of propagation of weak disturbances.[4]

THE EXCEPTIONAL CASE

Following Lax,[4,5,6] we say that a wave is exceptional if the gradient of its velocity with respect to the field components is orthogonal to the corresponding right eigenvectors, i.e., if

$$\nabla\lambda^{(k)} \cdot \mathbf{d}_K^{(k)} = 0, \quad K = 1, 2, \cdots, m^{(k)}. \tag{9}$$

In this case, we have

$$\lambda^{(k)\prime} = \nabla\lambda^{(k)} \cdot \mathbf{u}' + \frac{\partial\lambda^{(k)}}{\partial n} \cdot n' = \Lambda^{(k)} \cdot n',$$

by virtue of (5), and by the definition of the ray velocity,[4]

$$\Lambda = \lambda n + \frac{\partial\lambda}{\partial n} - \left(n \cdot \frac{\partial\lambda}{\partial n}\right)n.$$

On the other hand,

$$\lambda = \Lambda \cdot n, \quad \lambda' = \Lambda' \cdot n + \Lambda \cdot n'.$$

Hence,

$$n \cdot \Lambda^{(k)\prime} = 0,$$

and

$$\operatorname{div} \Lambda^{(k)} = 0. \tag{10}$$

This means that the elementary volume of space is conserved by transport along rays. For exceptional disturbance waves, the crossing of the wave front has no effect on the divergence of the ray velocity; instead of (10), we have

$$[\operatorname{div} \Lambda^{(k)}] = 0.$$

THE $\lambda^{(k)}$-STATIONARY FIELD

A field which satisfies the condition

$$\frac{d\mathbf{u}}{d\sigma^{(k)}} = 0, \quad \frac{d}{d\sigma^{(k)}} = \frac{\partial\psi^{(k)}}{\partial\varphi_\alpha}\partial_\alpha,$$

may be called a $\lambda^{(k)}$-stationary field, since it seems stationary to an observer moving with the ray velocity.

It is easily seen that a kth simple wave satisfies the above condition for

$$\frac{d\varphi}{d\sigma^{(k)}} = \varphi_\alpha \frac{\partial\psi^{(k)}}{\partial\varphi_\alpha} = \psi^{(k)} = 0.$$

ACKNOWLEDGMENT

We express our thanks to Professor P. D. Lax for his kind interest in our work.

[1] P. D. Lax, Commun. Pure Appl. Math. **10**, 537 (1957).
[2] A. Jeffrey and T. Taniuti, *Non-linear Wave Propagation* (Academic Press, New York, 1964).
[3] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, New York, 1948).
[4] G. Boillat, *La propagation des ondes* (Gauthier-Villars, Paris, 1965).
[5] P. D. Lax, Ann. Math. Studies (Princeton) **33**, 211 (1954).
[6] G. Boillat, J. Math. Phys. **10**, 452 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Expansion of the Inhomogeneous Symplectic Lie Algebras $T(2n) \overleftarrow{+} Sp(n)$ to $Sp(n + 2)$

JOHN G. NAGEL AND K. TAHIR SHAH*
Institute of Theoretical Physics, Chalmers University of Technology, Göteborg, Sweden

(Received 28 April 1969)

We show that the semidirect sum $G(n + N) \cong T(Nn) \overleftarrow{+} Sp(n)$, with $T(Nn)$ the ideal, can be expanded to $Sp(n + N)$ iff $N = 2$. We derive a general formula for invariants of $T(2n) \overleftarrow{+} Sp(n)$ and, besides, two general formulas for invariants of $ISO(n_1, n_2)$.

As a conclusion we find that an expansion $E(G)$ of an arbitrary Lie algebra G, in general, is *not* isomorphic to a deformation $D(G)$ of G, but that there exists a Lie algebra $G' \supseteq G$ and a deformation $D(G')$ of G' such that $E(G) \cong D(G')$.

1. INTRODUCTION

In the sequel, we discuss the problem of expansion and contraction of the symplectic Lie algebras. As is well known, the operation of contraction of Lie groups and Lie algebras was first defined by Segal[1] and later developed by Inönü and Wigner[2] and by Saletan.[3] The importance of such an operation was emphasized by Inönü and Wigner by showing that one could pass from relativistic to nonrelativistic theories by contracting the symmetry group of Minkowski space, namely, the Poincaré group to the inhomogeneous Galilei group.

Later interest arose in the opposite problem of whether one can define an algebraic operation such that, starting from the Lie algebra of the contracted group, one can arrive at the Lie algebra of a group closely related to the original group. This is the process of expansion, which roughly can be described as the one of replacing some of the generators of the original group by new operators, which are certain

In this case, we have

$$\lambda^{(k)\prime} = \nabla\lambda^{(k)} \cdot \mathbf{u}' + \frac{\partial\lambda^{(k)}}{\partial n} \cdot n' = \Lambda^{(k)} \cdot n',$$

by virtue of (5), and by the definition of the ray velocity,[4]

$$\Lambda = \lambda n + \frac{\partial\lambda}{\partial n} - \left(n \cdot \frac{\partial\lambda}{\partial n}\right)n.$$

On the other hand,

$$\lambda = \Lambda \cdot n, \quad \lambda' = \Lambda' \cdot n + \Lambda \cdot n'.$$

Hence,

$$n \cdot \Lambda^{(k)\prime} = 0,$$

and

$$\operatorname{div} \Lambda^{(k)} = 0. \tag{10}$$

This means that the elementary volume of space is conserved by transport along rays. For exceptional disturbance waves, the crossing of the wave front has no effect on the divergence of the ray velocity; instead of (10), we have

$$[\operatorname{div} \Lambda^{(k)}] = 0.$$

THE $\lambda^{(k)}$-STATIONARY FIELD

A field which satisfies the condition

$$\frac{d\mathbf{u}}{d\sigma^{(k)}} = 0, \quad \frac{d}{d\sigma^{(k)}} = \frac{\partial\psi^{(k)}}{\partial\varphi_\alpha}\partial_\alpha,$$

may be called a $\lambda^{(k)}$-stationary field, since it seems stationary to an observer moving with the ray velocity.

It is easily seen that a kth simple wave satisfies the above condition for

$$\frac{d\varphi}{d\sigma^{(k)}} = \varphi_\alpha \frac{\partial\psi^{(k)}}{\partial\varphi_\alpha} = \psi^{(k)} = 0.$$

ACKNOWLEDGMENT

We express our thanks to Professor P. D. Lax for his kind interest in our work.

[1] P. D. Lax, Commun. Pure Appl. Math. **10**, 537 (1957).
[2] A. Jeffrey and T. Taniuti, *Non-linear Wave Propagation* (Academic Press, New York, 1964).
[3] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, New York, 1948).
[4] G. Boillat, *La propagation des ondes* (Gauthier-Villars, Paris, 1965).
[5] P. D. Lax, Ann. Math. Studies (Princeton) **33**, 211 (1954).
[6] G. Boillat, J. Math. Phys. **10**, 452 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Expansion of the Inhomogeneous Symplectic Lie Algebras $T(2n) \overleftarrow{+} Sp(n)$ to $Sp(n + 2)$

JOHN G. NAGEL AND K. TAHIR SHAH*
Institute of Theoretical Physics, Chalmers University of Technology, Göteborg, Sweden

(Received 28 April 1969)

We show that the semidirect sum $G(n + N) \cong T(Nn) \overleftarrow{+} Sp(n)$, with $T(Nn)$ the ideal, can be expanded to $Sp(n + N)$ iff $N = 2$. We derive a general formula for invariants of $T(2n) \overleftarrow{+} Sp(n)$ and, besides, two general formulas for invariants of $ISO(n_1, n_2)$.

As a conclusion we find that an expansion $E(G)$ of an arbitrary Lie algebra G, in general, is *not* isomorphic to a deformation $D(G)$ of G, but that there exists a Lie algebra $G' \supseteq G$ and a deformation $D(G')$ of G' such that $E(G) \cong D(G')$.

1. INTRODUCTION

In the sequel, we discuss the problem of expansion and contraction of the symplectic Lie algebras. As is well known, the operation of contraction of Lie groups and Lie algebras was first defined by Segal[1] and later developed by Inönü and Wigner[2] and by Saletan.[3] The importance of such an operation was emphasized by Inönü and Wigner by showing that one could pass from relativistic to nonrelativistic theories by contracting the symmetry group of Minkowski space, namely, the Poincaré group to the inhomogeneous Galilei group.

Later interest arose in the opposite problem of whether one can define an algebraic operation such that, starting from the Lie algebra of the contracted group, one can arrive at the Lie algebra of a group closely related to the original group. This is the process of expansion, which roughly can be described as the one of replacing some of the generators of the original group by new operators, which are certain

functions of the generators, and also adding further operators of this form, such that the new set of operators close under commutation. The new group which is generated by these operators is called the expansion of the original group.

This problem was treated by several authors[4–11] for various special cases of the inhomogeneous pseudo-orthogonal Lie algebras $ISO(n_1, n_2)$ and for the inhomogeneous pseudo-unitary Lie algebras $IU(n_1, n_2)$. Sankaranarayanan[5] obtained expansions of all $ISO(n)$ and showed that in these cases expansion and contraction are inverse processes. On similar lines, Rosen and Roman[8] and Rosen,[9] in a systematic way, found expansions of all $ISO(n_1, n_2)$ and $IU(n_1, n_2)$.

Expansion of the inhomogeneous symplectic Lie algebras has so far not been treated in the literature. It is well known that the maximal subgroup $\mathcal{Sp}(n-2) \times \mathcal{Sp}(2)$ ($\times$ denoting direct product) of the symplectic group $\mathcal{Sp}(n)$ is not canonical, i.e., that the subduced representation $D(\mathcal{Sp}(n)) \downarrow \mathcal{Sp}(n-2) \times \mathcal{Sp}(2)$ is not, in general, multiplicity free for $D(\mathcal{Sp}(n))$ irreducible. This structural feature of $\mathcal{Sp}(n)$ often causes trouble and it was, therefore, from the outset not clear whether expansions similar to those of $ISO(n_1, n_2)$ and $IU(n_1, n_2)$ could be found in this case.

In Sec. 2, we define a certain inhomogeneous symplectic group

$$\mathcal{G}(n+N) \cong \mathcal{T}(2n) \overleftarrow{\times} \mathcal{Sp}(n),$$

where $\overleftarrow{\times}$ denotes semidirect product, with $\mathcal{T}(Nn)$ the invariant subgroup. In Sec. 3, we show that the Lie algebra $G(n+N)$ of $\mathcal{G}(n+N)$ indeed can be expanded to $Sp(n+N)$ iff $N=2$.

In Sec. 4, we follow a method given by Rosen[10] and prove Lemma 1. Applying this lemma, we are then able to obtain general Casimir operators of $T(2n) \overleftarrow{+} Sp(n)$, where $\overleftarrow{+}$ denotes semidirect sum, from the well-known Casimir operators of $Sp(n+2)$.

Throughout the paper we have used a formalism and notation with the help of which one could reproduce immediately all the results in Refs. 9 and 10 for $ISO(n_1, n_2)$. We use this possibility, finally, to obtain two general formulas for Casimir operators of $ISO(n_1, n_2)$, since only a few of lowest degree were found in Ref. 10.

It is our impression that the process of expansion is a powerful tool for computing Casimir operators of inhomogeneous Lie algebras.

In Sec. 5, we perform the contraction of $Sp(n+2)$ with respect to the parameter of the expansion *and find in this case that expansion and contraction are not inverse processes, however, though being closely related to one another.*

Finally, we establish for an arbitrary Lie algebra G the precise relationship between expansion and the true inverse process of contraction, the abstract process of deformation.[12–15]

2. THE GROUP $\mathcal{T}(Nn) \overleftarrow{\times} \mathcal{Sp}(n)$ AND NOTATION

Let us consider the group $\mathcal{G}(n+N)$ of all real or all complex linear transformations

$$x'_\mu = \mathcal{S}^\nu_\mu x_\nu + a^i_\mu x_{n+i}, \quad \mu, \nu = 1, \cdots, n \geq 2, \quad n \text{ even},$$
$$x'_{n+i} = x_{n+i}, \quad i = 1, \cdots, N \geq 1,$$

of an $(n+N)$-dimensional vector space

$$U_{n+N} = V_n + W_N,$$

where

$$\mathcal{S}^\nu_\mu \in \mathcal{Sp}(n),$$

the n-dimensional real or complex symplectic group, and where the a^i_μ take on arbitrary real or complex values. The transformations $\mathcal{S}^\nu_\mu$, hence, leave invariant the nondegenerate skew-symmetric bilinear form

$$(x, y) \equiv g^{\mu\nu} x_\mu y_\nu.$$

Here the metric $g^{\mu\nu}$ has the following properties:

$$g^{\nu\mu} = \eta g^{\mu\nu}, \quad \eta = -1,$$
$$g_{\mu\lambda} g^{\lambda\nu} = \delta^\nu_\mu = g^{\nu\lambda} g_{\lambda\mu}.$$

Hence,

$$g_{\nu\mu} = \eta g_{\mu\nu},$$
$$g^{\cdot\nu}_{\mu} = \delta^\nu_\mu, \quad g^{\mu\cdot}_{\nu} = \eta \delta^\mu_\nu, \quad g^{\cdot\cdot}_{\mu\nu} = g_{\nu\mu},$$

and

$$g^{\mu}_{\cdot\nu} = \delta^\mu_\nu, \quad g_{\mu\cdot}^{\nu} = \eta \delta^\nu_\mu, \quad g^{\mu\nu}_{\cdot\cdot} = g^{\nu\mu},$$

where the dot indicates that we have raised or lowered the index of that place, using the metric.

It is seen that

$$\mathcal{G}(n+N) \cong \mathcal{T}(n) \overleftarrow{\times} \mathcal{Sp}(n),$$

the semidirect product of $\mathcal{Sp}(n)$ by the Nn-dimensional real or complex translation group $\mathcal{T}(Nn)$.

The generators $M_{\mu\nu}$ of $\mathcal{Sp}(n)$ and the generators $P_{i\mu}$ of $\mathcal{T}(Nn)$ form a basis for the Lie algebra

$$G(n+N) \cong T(Nn) \overleftarrow{+} Sp(n),$$

of $\mathcal{G}(n+N)$, over the real or the complex field. They satisfy the following commutation relations:

$$[M_{\mu\nu}, M_{\mu'\nu'}] = g_{\nu\mu'} M_{\mu\nu'} - g_{\mu'\mu} M_{\nu\nu'} + g_{\nu\nu'} M_{\mu'\mu} - g_{\nu'\mu} M_{\mu'\nu}, \quad (1)$$

$$[M_{\mu\nu}, P_{i\nu'}] = g_{\nu\nu'} P_{i\mu} - g_{\nu'\mu} P_{i\nu}, \quad (2)$$

$$[P_{i\nu}, P_{i'\nu'}] = 0, \quad (3)$$

where

$$M_{\nu\mu} = -\eta M_{\mu\nu}.$$

The Casimir operators of $Sp(n)$ are given by

$$C_{2m}(n) = M^{\mu_2}_{\mu_1} M^{\mu_3}_{\mu_2} \cdots M^{\mu_{2m}}_{\mu_{2m-1}} M^{\mu_1}_{\mu_{2m}}, \qquad m = 1, 2, \cdots, \tfrac{1}{2}n. \quad (4)$$

3. CONSTRUCTION OF THE GENERATORS OF $\mathfrak{S}\!\!\!/(n+2)$

In this section we shall show that out of the generators of $\mathfrak{S}(n+2)$ we can construct certain new operators which obey the commutation relations of $Sp(n+2)$.

Let us to this end define the following operators:

$$\bar{P}_{i\mu} \equiv \{P^\nu_i M_{\nu\mu}\} \equiv \tfrac{1}{2}(P^\nu_i M_{\nu\mu} + M_{\nu\mu} P^\nu_i) = P^\nu_i M_{\nu\mu} - \tfrac{1}{2}(n-\eta) P_{i\mu},$$

which also can be expressed as

$$\bar{P}_{i\mu} = \tfrac{1}{4}[P_{i\mu}, C_2(n)]$$

and

$$K_{ij} \equiv \{P^\mu_j \bar{P}_{i\mu}\} = \{P^\mu_i P^\nu_j M_{\mu\nu}\} \equiv \tfrac{1}{4}(P^\mu_i P^\nu_j M_{\mu\nu} + P^\nu_j M_{\mu\nu} P^\mu_i + P^\mu_i M_{\mu\nu} P^\nu_j + M_{\mu\nu} P^\mu_i P^\nu_j) = -\eta K_{ji}.$$

Here $\{\ \}$ denotes that the expression inside has been symmetrized with respect to the position of the P^ν_i relative to $M_{\mu\nu}$ and to $\bar{P}_{i\nu}$ (but not relative to one another as they commute) and then divided by the number of terms.

We now point out that so far we have not been considering any nontrivial transformations which leave the subspace W_N invariant. At this step, it will be necessary also to consider such transformations and, for this purpose, we introduce a metric g^{ij} in W_N with the following properties:

$$g^{ji} = \eta g^{ij},$$

similarly for g_{ij}, and

$$g_{ik} g^{kj} = \delta^j_i = g^{jk} g_{ki}.$$

Now, noticing that $P^\mu_i P_{j\mu}$ has one independent component iff $N = 2$, one verifies easily the following relation which will be needed shortly:

$$P^\mu_i P_{j\mu} = g_{ij} P^2, \quad \text{iff} \quad N = 2. \quad (5)$$

Here

$$P^2 \equiv N^{-1} P^{i\mu} P_{i\mu},$$

where use has been made of the metric mentioned above to raise the Latin index.

As the next step, we define the operators

$$\bar{M}'_{\mu\nu} \equiv M_{\mu\nu},$$
$$\bar{M}'_{n+i\,\nu} \equiv P_{i\nu} + \lambda \bar{P}_{i\nu}/\sqrt{-p^2},$$
$$\bar{M}'_{n+i\,n+j} \equiv -K_{ij}/P^2,$$

where λ is a free parameter. Also defining

$$g_{n+i\,n+j} \equiv g_{ij},$$

we obtain the following commutators, having made particular use of relation (5):

$$[\bar{M}'_{\mu\nu}, \bar{M}'_{\mu'\nu'}] = g_{\nu\mu'} \bar{M}'_{\mu\nu'} - g_{\mu'\mu} \bar{M}'_{\nu\nu'} + g_{\nu\nu'} \bar{M}'_{\mu'\mu} - g_{\nu'\mu} \bar{M}'_{\mu'\nu}, \quad (6)$$

$$[\bar{M}'_{\mu\nu}, \bar{M}'_{n+i\nu'}] = g_{\nu\nu'} \bar{M}'_{n+i\mu} - g_{\nu'\mu} \bar{M}'_{n+i\nu}, \quad (7)$$

$$[\bar{M}'_{\mu\nu}, \bar{M}'_{n+i\,n+j}] = 0, \quad (8)$$

$$[\bar{M}'_{n+i\nu}, \bar{M}'_{n+i'\nu'}] = \lambda^2(-g_{n+i'\,n+i} \bar{M}'_{\nu\nu'} + g_{\nu\nu'} \bar{M}'_{n+i'\,n+i}), \quad (9)$$

$$[\bar{M}'_{n+i\,n+j}, \bar{M}'_{n+i'\nu}] = g_{n+j\,n+i'} \bar{M}'_{n+i\nu} - g_{n+i'\,n+i} \bar{M}'_{n+j\,\nu}, \quad (10)$$

$$[\bar{M}'_{n+i\,n+j}, \bar{M}'_{n+i'\,n+j'}] = g_{n+j\,n+i'} \bar{M}'_{n+i\,n+j'} - g_{n+i'\,n+i} \bar{M}'_{n+j\,n+j'} + g_{n+j\,n+j'} \bar{M}'_{n+i'\,n+i} - g_{n+j'\,n+i} \bar{M}'_{n+i'\,n+j}. \quad (11)$$

In order to remove the parameter λ from the above commutators, we define the operators

$$\bar{M}_{\mu\nu} \equiv \bar{M}'_{\mu\nu} = M_{\mu\nu}, \quad (12)$$

$$\bar{M}_{n+i\nu} \equiv \lambda^{-1} \bar{M}'_{n+i\nu} = \lambda^{-1} P_{i\nu} + \bar{P}_{i\nu}\sqrt{-P^2} \equiv -\eta \bar{M}_{\nu\,n+i}, \quad (13)$$

$$\bar{M}_{n+i\,n+j} \equiv \bar{M}'_{n+i\,n+j} = -K_{ij}/P^2. \quad (14)$$

Also defining

$$g_{n+i\nu} = 0 = g_{\nu\,n+i},$$

we then obtain the following commutation relations;

$$[\bar{M}_{AB}, \bar{M}_{A'B'}] = g_{BA'} \bar{M}_{AB'} - g_{A'A} \bar{M}_{BB'} + g_{BB'} \bar{M}_{A'A} - g_{B'A} \bar{M}_{A'B}, \qquad A, B = 1, \cdots, n+N,$$

where

$$g_{BA} = \eta g_{AB} \quad \text{and} \quad \bar{M}_{BA} = -\eta \bar{M}_{AB}.$$

These commutation relations are evidently those of $Sp(n+2)$. *We have thus shown that* $G(n+N) \cong T(Nn) \overleftarrow{+} Sp(n)$ *by this procedure can be expanded to* $Sp(n+N)$ *iff* $N = 2$, *owing to relation* (5).

4. CASIMIR OPERATORS OF $T(2n) \overleftarrow{+} Sp(n)$ AND OF $ISO(n_1, n_2)$

In this section we shall obtain general formulas for invariants of $T(2n) \overleftarrow{+} Sp(n)$ and of $ISO(n_1, n_2)$. We start by stating the following lemma, which is proven in Appendix A and corresponds to those obtained by Rosen[10] for $ISO(n_1, n_2)$ and $IU(n_1, n_2)$.

Lemma 1: If $X = X(M_{\mu\nu}, P_{i\mu})$ is a homogeneous polynomial in $M_{\mu\nu}$ and $P_{i\mu}$ which satisfies

$$[\bar{M}_{\mu\nu}, X] = 0 = [\bar{M}_{n+i\,\mu}, X],$$

then

$$[M_{n+i\,n+j}, X] = 0 = [P_{i\mu}, X].$$

This means that, if X is an invariant of $Sp(n+2)$, it is also an invariant of $T(2n) \overleftarrow{+} Sp(n)$.

We shall also need this simple lemma.

Lemma 2: Let X be an operator which is a polynomial in a parameter α, i.e.,

$$X = X(\alpha) = \sum_i \alpha^i X_i, \quad i \geq 0, \quad -\infty < \alpha < \infty,$$

and let O be an operator which is independent of α. Then,

$$[O, X_i] = 0 \quad \text{iff} \quad [O, X(\alpha)] = 0.$$

The necessary condition is obvious and the sufficient condition follows by differentiation:

$$0 = \frac{d^i[O, X(\alpha)]}{d\alpha^i}\bigg|_{\alpha=0} = [O, X_i].$$

Defining the operators $K^j_{k\alpha;i}$, $\alpha = 1, 2$, $k \geq \alpha$, by

$$K^j_{11;i} \equiv K^j_i, \tag{15}$$

$$K^j_{21;i} \equiv -\eta \bar{P}^\mu_i \bar{P}^j_\mu, \tag{16}$$

$$K^j_{k1;i} \equiv -\eta \bar{P}^{\mu_1}_i M^{\mu_2}_{\mu_1} M^{\mu_3}_{\mu_2} \cdots M^{\mu_{k-2}}_{\mu_{k-3}} M^{\mu_{k-1}}_{\mu_{k-2}} \bar{P}^j_{\mu_{k-1}}, \quad \forall k \geq 3, \tag{17}$$

and

$$K^j_{22;i} \equiv -\eta P^\mu_i P^j_\mu, \tag{18}$$

$$K^j_{k2;i} \equiv -\eta P^{\mu_1}_i M^{\mu_2}_{\mu_1} M^{\mu_3}_{\mu_2} \cdots M^{\mu_{k-2}}_{\mu_{k-3}} M^{\mu_{k-1}}_{\mu_{k-2}} P^j_{\mu_{k-1}}, \quad \forall k \geq 3, \tag{19}$$

we can now state the following theorem.

Theorem 1: Let

$$C_{2m}(n+N) = C_{2m}(\bar{M}_{AB}), \quad N = 2,$$

be the Casimir operators of $Sp(n+2)$ given by expressions similar to those of Eq. (4). If the $\bar{M}_{AB}$ are expressed in terms of the $M_{\mu\nu}$ and $P_{i\mu}$ by Eqs. (12)–(14), then

$$C_{2m}(n+N) = \sum_{l=0}^{m} \lambda^{-2l}(-P^2)^{-N(m-l)} X_{ml},$$
$$m = 1, 2, \cdots, \nu + N - 1, \quad \nu = \tfrac{1}{2}n, \tag{20}$$

where the X_{ml} are homogeneous polynomials in $M_{\mu\nu}$ and $P_{i\mu}$ of degree $2(N+1)m - 2Nl$ and are invariants of $T(2n) \overleftarrow{+} Sp(n)$. They are given by the following expressions:

$$X_{11} = 2P^2, \tag{21}$$

$$X_{10} = K^j_i K^i_j + 2(-P^2)^{N-1} K^i_{21;i} + (-P^2)^N C_2(n), \tag{22}$$

and

$$X_{ml} = \sum_{\Pi} \sum_{p=1}^{Nm-l} (-P^2)^{Nm-l-p} K^{i_2}_{k_1\alpha_1;i_1} K^{i_3}_{k_2\alpha_2;i_2} \cdots K^{i_1}_{k_p\alpha_p;i_p} + \delta_{l0}(-P^2)^{Nm} C_{2m}(n),$$
$$\sum_{q=1}^{p} k_q = 2m, \quad m = 2, 3, \cdots, \nu + N - 1,$$
$$\sum_{q=1}^{p} \delta_{\alpha_p 2} = l, \quad l = 0, 1, \cdots, m, \tag{23}$$

where the $K^j_{k\alpha;i}$ are given by Eqs. (15)–(19) and the $C_{2m}(n)$ by Eq. (4). The summation over Π goes over all *distinct* terms which are obtained by cyclic permutations of the K^j_i, $\bar{P}_{i\mu}$, $P_{i\mu}$, and $M_{\mu\nu}$ (i.e., not only of the $K^j_{k\alpha;i}$!).

With respect to the proof of this theorem, it follows immediately from Lemmas 1 and 2 that the X_{ml} are invariants of $T(2n) \overleftarrow{+} Sp(n)$ since $(P^2)^{Nm} C_{2m}(n+N)$ is a polynomial in $M_{\mu\nu}$ and $P_{i\mu}$ and the parameter $\alpha = \lambda^{-1}$. The rest of the proof is deferred to Appendix B.

The X_{m0}, $m \geq 1$, are not only invariants of $T(2n) \overleftarrow{+} Sp(n)$ but are, in general, together with X_{11}, also Casimir operators; i.e., they can, in general, not be expressed in terms of invariants of lower degree. The remaining invariants X_{ml}, $m \geq 2$, $l \geq 1$, are not Casimir operators; i.e., they can be expressed in terms of Casimir operators of lower degree.

In order to illustrate the formula for X_{ml}, we write down explicitly

$$X_{20} = (-P^2)^{2N-4} K^j_i K^k_j K^i_k + (-P^2)^{2N-3} \sum_{\Pi=1}^{4} K^j_i K^k_j \bar{P}^\mu_k \bar{P}^i_\mu + (-P^2)^{2N-2} \left(\sum_{\Pi=1}^{4} K^j_i \bar{P}^\mu_j M^\nu_\mu \bar{P}^i_\nu + \sum_{\Pi=1}^{2} \bar{P}^\mu_i \bar{P}^j_\mu \bar{P}^\nu_j \bar{P}^i_\nu \right) + (-P^2)^{2N-1} \sum_{\Pi=1}^{4} \bar{P}^\mu_i M^\nu_\mu M^\lambda_\nu \bar{P}^i_\lambda + (-P^2)^{2N} M^\nu_\mu M^\kappa_\nu M^\lambda_\kappa M^\mu_\lambda,$$

which, since $N = 2$, is of 12th degree.

As an important point, we want to mention that all the results obtained in Refs. 9 and 10, for the case of $ISO(n_1, n_2)$, can be reproduced by over-all setting $\eta = +1$ and $N = 1$. One also finds that this value of N is the only one for which $T(Nn) \overleftarrow{+} SO(n_1, n_2)$ can be expanded to $SO(n_1 + N_1, n_2 + N_2)$, $n = n_1 + n_2$, and $N = N_1 + N_2$, owing to relation (5). In particular, this correspondence means *that our Theorem 1 also gives us a general formula for invariants of* $ISO(n_1, n_2)$, which was not obtained in Ref. 10. In addition to the changes mentioned above, one only has to set $\nu = \frac{1}{2}n$ or $\frac{1}{2}(n-1)$, for n even or odd, and note that $K^{i}_{11;i}$ of Eq. (15) vanishes in this case.

Apart from these invariants, the exceptional Casimir operator

$$C_{\nu+1}(n+N) = \epsilon_{A_1A_2\cdots A_{2\nu+1}A_{2\nu+2}}\bar{M}^{A_1A_2}\cdots\bar{M}^{A_{2\nu+1}A_{2\nu+2}}$$

of $SO(n_1 + N_1, n_2 + N_2)$, for $n + N = 2\nu + 2$, also gives rise to a general formula of invariants as follows:

$$C_{\nu+1}(n+N) = (-P^2)^{-\frac{1}{2}}X_{\nu+1\,0} + \lambda^{-1}X_{\nu+1\,1}, \qquad \forall n = 2\nu+1, \quad N = 1,$$

where

$$X_{\nu+1\,0} = 0, \quad \forall n = 2\nu + 1,$$

and where

$$X_{\nu+1\,1} = -(2\nu+2)\epsilon_{\mu_1\cdots\mu_{2\nu+1}}M^{\mu_1\mu_2}\cdots M^{\mu_{2\nu-1}\mu_{2\nu}}P^{\mu_{2\nu+1}}, \qquad \forall n = 2\nu+1,$$

is a Casimir operator of $ISO(n_1, n_2)$ of degree $2\nu + 1$.

Of the invariants X_{ml} of $ISO(n_1 + N_1, n_2 + N_2)$ only X_{10} and X_{20} were obtained in an equivalent form in Ref. 10. However, $X_{\nu+1\,0}$ was, for $\nu = 1$, erroneously stated there as nonvanishing, being equal to $-\frac{1}{4}X_{\nu+1\,1}$.

5. INTERRELATIONS OF EXPANSION, CONTRACTION, AND DEFORMATION

We shall now discuss the precise relation of expansion to the processes of contraction and deformation.

By analyzing the commutation relations (6)–(11), we find that, when taking the limit $\lambda \to 0$, $Sp(n+2)$ contracts to

$$[T(2n) \overleftarrow{+} Sp(n)] \overleftarrow{+} Sp(2) \cong T(2n) \overleftarrow{+} [Sp(n) + Sp(2)],$$

where $+$ denotes direct sum. We note, in particular, that we do not get back to the original Lie algebra $T(2n) \overleftarrow{+} Sp(n)$ (see Diagram 1).

$$T(2n) \overleftarrow{+} Sp(n) \Rightarrow Sp(n+2)$$
$$\overset{\lambda\to 0}{\rightleftarrows} T(2n) \overleftarrow{+} [Sp(n) + Sp(2)]$$

DIAGRAM 1. Expansion, contraction, and deformation diagram for the symplectic Lie algebras. Here $\Rightarrow$, $\to$, and $\leftarrow$ denote expansion, contraction, and deformation, respectively.

As a conclusion of our analysis of expansion of $T(2n) \overleftarrow{+} Sp(n)$ and contraction of the expanded Lie algebra $Sp(n+2)$, we shall give the following general remarks:

(1) An expansion $E(G)$ of a Lie algebra G is a Lie algebra whose elements are contained in an algebraic extension of the quotient division algebra[16–19] of the enveloping algebra of G;

(2) Expansion is not, in general, the inverse process of contraction in contrary to what implicitly has been assumed in some of the literature; if one first expands a Lie algebra G to a Lie algebra $E(G)$ and afterwards contracts $E(G)$, then the contracted Lie algebra $G' = C(E(G))$ is, in general, of higher dimension than G, but such that $G \subseteq G'$;

(3) The relation of expansion to deformation[12–15] is the following one: For an expansion $E(G)$ of a Lie algebra G, there exists a Lie algebra $G' \supseteq G$ and a deformation $D(G')$ of G' such that $E(G) \cong D(G')$.

ACKNOWLEDGMENTS

It is a pleasure for the authors to thank Dr. M. Levy-Nahas for a stimulating discussion and Professor M. Flato for interesting comments on this work. Furthermore, the authors would like to express their gratitude to Professor N. Svartholm and one of us (J. G. N.) also to Professor J. Nilsson for the hospitality extended to us at Chalmers University of Technology. Finally, one of us (K. T. S.) wishes to thank Professor Svartholm for many helpful discussions in general during his stay there.

APPENDIX A: PROOF OF LEMMA 1

In order to prove Lemma 1, let us consider a homogeneous polynomial in $M_{\mu\nu}$ and $P_{i\mu}$,

$$X_m = X_m(M_{\mu\nu}, P_{i\mu}),$$

of degree m which satisfies

$$[\bar{M}_{\mu\nu}, X_m] = 0 = [\bar{M}_{n+i\mu}, X_m],$$

and define

$$Y_{m;i\mu} \equiv [P_{i\mu}, X_m].$$

We then have that

$$0 = [\bar{M}_{n+i\mu}, X_m] = [\lambda^{-1}P_{i\mu} + \bar{P}_{i\mu}, X_m]$$
$$= \lambda^{-1}[P_{i\mu}, X_m] + [P^{\nu}_i, X_m]\frac{M_{\nu\mu}}{\sqrt{-P^2}} - \frac{n-\eta}{2\sqrt{-P^2}}[P_{i\mu}, X_m]$$
$$= \lambda^{-1}Y_{m;i\mu} + Y^{\nu}_{m;i}\frac{M_{\nu\mu}}{\sqrt{-P^2}} - \frac{n-\eta}{2\sqrt{-P^2}}Y_{m;i\mu}$$
$$= \lambda^{-1}Y_{m;i\mu} + \frac{1}{\sqrt{-P^2}}\left(Y^{\nu}_{m;i}M_{\nu\mu} - \frac{n-\eta}{2}Y_{m;i\mu}\right).$$

It now follows that the two terms must separately vanish for $-\infty < \lambda^{-1} < \infty$, since the first is a polynomial in $M_{\mu\nu}$ and $P_{i\mu}$ and the second is a polynomial divided by $\sqrt{-P^2}$. Hence,

$$\lambda^{-1}Y_{m;i\mu} = 0 \tag{A1}$$

and

$$Y^{\nu}_{m;i}M_{\nu\mu} - \tfrac{1}{2}(n-\eta)Y_{m;i\mu} = 0. \tag{A2}$$

Equation (A1) does not help us for $\lambda^{-1} = 0$, so we consider instead Eq. (A2). Here $Y^{\nu}_{m;i}M_{\mu\nu}$ is a homogeneous polynomial of degree $m+1$ and $Y_{m;i\mu}$ of degree m; hence, Eq. (A2) can only be fulfilled if

$$[P_{i\mu}, X_m] = Y_{m;i\mu} = 0, \quad -\infty < \lambda^{-1} < \infty.$$

We then obtain

$$[\bar{M}_{n+i\ n+j}, X_m] = -P^{-2}[K_{ij}, X_m]$$
$$= -P^{-2}[\{P\bar{P}\}, X_m] = 0,$$

which finishes the proof of the lemma.

APPENDIX B: PROOF OF THEOREM 1

In order to prove Eqs. (20)–(23), we define for an arbitrary operator Q_μ

$$\bar{Q}_\mu \equiv \tfrac{1}{2}(Q^{\nu}M_{\nu\mu} + M_{\nu\mu}Q^{\nu}) = Q^{\nu}M_{\nu\mu} - \tfrac{1}{2}(n-\eta)Q_\mu$$
$$= M_{\nu\mu}Q^{\nu} + \tfrac{1}{2}(n-\eta)Q_\mu;$$

then

$$Q_\mu\bar{R}^\mu + \bar{Q}_\mu R^\mu = Q_\mu(M^\mu_\nu R^\nu + \tfrac{1}{2}(n-\eta)R^\mu) + (Q^{\nu}M_{\nu\mu} - \tfrac{1}{2}(n-\eta)Q_\mu)R^\mu = 0.$$

Furthermore, we define

$$\vec{Q}^\mu_k \equiv Q^{\nu_1}M^{\nu_2}_{\nu_1}M^{\nu_3}_{\nu_2}\cdots M^{\mu}_{\nu_k}$$

and

$$\overleftarrow{Q}_{k\mu} \equiv M^{\nu_k}_{\mu}M^{\nu_{k-1}}_{\nu_{k-1}}\cdots M^{\nu_1}_{\nu_2}Q_{\nu_1};$$

then

$$\vec{Q}_{k\mu} \equiv \overrightarrow{(\bar{Q}_\mu)_k} = \bar{\vec{Q}}_{k\mu}, \quad \overleftarrow{Q}_{k\mu} \equiv \overleftarrow{(\bar{Q}_\mu)_k} = \bar{\overleftarrow{Q}}_{k\mu}.$$

The right-hand side of Eq. (20) must also from the beginning, apart from the $K^j_{k\alpha;i}$, $\alpha = 1, 2$, contain operators of the form

$$K^j_{k3;i} \equiv \vec{P}_{k-1;i\mu}\bar{P}^{j\mu} + \bar{P}_{i\mu}\overleftarrow{P}^{j\mu}_{k-1}, \quad k \geq 2,$$

which are to participitate in the cyclic permutations of the $\bar{P}_{i\mu}$, $P_{i\mu}$, and $M_{\mu\nu}$. We shall now show that these terms all vanish, by considering explicitly what happens under permutation to a term with at least one such factor. Using an obvious symbolic notation, we find, indeed, that

$$\sum_{\Pi} K\cdots K(\vec{P}_k\bar{P} + \bar{P}\overleftarrow{P}_k)$$
$$= K\cdots K(\vec{P}_k\bar{P} + \bar{P}\overleftarrow{P}_k) + \sum_{k_1=0}^{k}(\overleftarrow{\bar{P}}_{k_1}K\cdots K\vec{P}_{k-k_1} + \overleftarrow{P}_{k_1}K\cdots K\vec{\bar{P}}_{k_2})$$
$$= K\cdots K(\vec{P}_k\bar{P} + \bar{\vec{P}}_kP) + \sum_{k_1=0}^{k}\{K\cdots K(\bar{\overleftarrow{P}}_{k_1}\vec{P}_{k-k_1} + \overleftarrow{P}_{k_1}\bar{\vec{P}}_{k-k_1}) + [\bar{\overleftarrow{P}}_{k_1}, K\cdots K]\vec{P}_{k-k_1} + [\overleftarrow{P}_{k_1}, K\cdots K]\bar{\vec{P}}_{k-k_1}\}$$
$$= \sum_{k_1=0}^{k}\{\overline{[\overleftarrow{P}_{k_1}, K\cdots K]}\vec{P}_{k-k_1} + [P_{k_1}, K\cdots K]\bar{\vec{P}}_{k-k_1}\}$$
$$= 0.$$

Now, since the coefficients of the odd powers of λ^{-1} must contain an odd number of $K^j_{k3;i}$, they all vanish and what is left of the right-hand side of Eq. (20) is seen by inspection to be of the explicit form as given in Eqs. (21)–(23).

* Present address: Dept. of Physics, Stevens Institute of Technology, Hoboken, N.J.

[1] I. E. Segal, Duke Math. J. **18**, 221 (1951).
[2] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953).
[3] E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[4] M. A. Melvin, Bull. Am. Phys. Soc. **7**, 493 (1962).
[5] A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965).
[6] C. Fronsdal, Rev. Mod. Phys. **37**, 221 (1965).
[7] M. Y. Han, Nuovo Cimento **42B**, 367 (1966).
[8] J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966).
[9] J. Rosen, Nuovo Cimento **46B**, 1 (1966).
[10] J. Rosen, J. Math. Phys. **9**, 1305 (1968).
[11] A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).
[12] M. Gerstenhaber, Ann. Math. **79**, 59 (1964).
[13] R. Hermann, Commun. Math. Phys. **2**, 251 (1966); **3**, 53, 75 (1966); **5**, 131, 157 (1967); **6**, 205 (1967).
[14] M. Levy-Nahas, J. Math. Phys. **8**, 1211 (1967); and *Compte-rendu du Colloque de Gif-sur-Yvette* (French) (to be published).
[15] M. Levy-Nahas and R. Seneor, Commun. Math. Phys. **9**, 242 (1968).
[16] N. Bourbaki, *Algèbre* (Hermann, Paris, 1958), Chap. I, p. 145.
[17] I. M. Gel'fand and A. A. Kirillov, Dokl. Akad. Nauk SSSR **167**, 503 (1966) (not p. 407 as quoted in Ref. 19); Publ. Math. France **31**, 509 (1966).
[18] Y. Chow, J. Math. Phys. **10**, 975 (1969).
[19] A. Simoni and F. Zaccaria, Nuovo Cimento **59A**, 280 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Branching Laws, Inner Multiplicities, and Decomposition of Classical Groups

M. K. F. WONG*
Department of Physics, St. Louis University, St. Louis, Missouri

(Received 4 August 1969)

A method has been found whereby the inner multiplicity of all classical groups [all irreducible representations of $SU(n)$ and $SO(2k+1)$ and some simple irreducible representations of $SO(2k)$ and $Sp(2k)$] can be obtained easily from the branching laws of Weyl (unitary), Boerner (orthogonal), and Hegerfeldt (symplectic). Once the inner multiplicity is known, the same formula can be used again to obtain the decomposition of a classical group into its subgroups without any restriction. Finally, since the inner multiplicity is connected to the outer multiplicity through the Racah–Speiser lemma, this method enables us to obtain the Clebsch–Gordan series for the direct product of all classical groups.

INTRODUCTION

Although the inner multiplicity of classical groups can, in principle, be calculated from Kostant's formula[1] or by recurrence relations obtained by Freudenthal[2] and Racah,[3] in practice it is not well known how easily the inner multiplicity can be calculated by means of such techniques. Actually, Racah's recurrence relation is the easiest to use and, as Racah has remarked already, this relation is sufficient by itself to obtain all the inner multiplicities of all classical groups. We hope to write on this subject in a future paper. As far as Kostant's formula is concerned, the writer has devised a diagrammatic method (unpublished) to obtain the inner multiplicity by means of the formula. However, it is still quite lengthy for practical calculations. Delaney and Gruber[4] have obtained a formula relating branching (or restriction) multiplicity to inner multiplicity, a formula first obtained by Straumann,[5] and rederived independently by Klymyk.[6] In addition, Delaney and Gruber have obtained the inner multiplicity of all $SU(n)$ groups through a one-to-one mapping between the weights of $SU(n)$ and the Gel'fand pattern. Basically, they use the chain $SU(n) \supset SU(n-1) \supset SU(n-2) \supset \cdots \supset SU(2)$. So far, this method seems to be the easiest to use. Unfortunately, their method cannot be easily extended to the orthogonal and symplectic groups.

We wish to point out in this paper that the formula obtained by Delaney and Gruber [Eq. (59) in their paper, Eq. (1) in this paper] as well as Straumann and Klymyk can be used to obtain the inner multiplicities of all classical groups by means of the respective branching laws for these groups. Besides, it is not necessary to trace down the group chain to the smallest subgroup, where the multiplicity of weights is either one or zero. It is sufficient to consider only the case[7] $SU(n) \supset SU(n-1)$ or $O(n) \supset O(n-1)$ or $Sp(2n) \supset Sp(2(n-1))$. The branching multiplicity in these cases is known through the respective branching laws. In the case of $SU(n)$, it was first obtained by Weyl.[8] The branching laws for the orthogonal groups have been proved by Boerner.[9] The branching laws for the symplectic group have been obtained by Hegerfeldt.[10] It is interesting (though quite mysterious!) to note that all these branching laws look very similar to each other. They all obey the so-called "triangular inequality."

In the case of $SO(2k)$ and $Sp(2k)$, this method alone does not always give complete solutions because sometimes the number of unknowns is greater than the number of equations. (The writer wishes to thank Dr. B. Gruber for pointing this out to him.) However, it is always possible to supplement this method with Racah's recurrence relation to obtain complete solutions. We hope to write on this subject in a future paper in greater detail. In the examples considered in this paper, we use the present method alone to obtain complete solutions of some simple irreducible representations of $SO(2k)$ and $Sp(2k)$.

This paper is divided into three sections. In Sec. I, we show how to obtain the inner multiplicity of all classical groups by means of branching laws and Eq. (1). In other words, we consider the left-hand side of Eq. (1) as known and calculate the right-hand side of Eq. (1) as unknown. These are linear algebraic equations and, in all cases, we find the solutions emerging successively in the simplest fashion. It is not necessary to solve them simultaneously, though that would not have presented a problem.

In Sec. II, we use Eq. (1) again to obtain the decomposition of all classical groups without any restriction, e.g., $U(n) \supset O(n)$, $U(2n) \supset Sp(2n)$, $O(2k) \supset SU(k)$, etc. This time we regard the right-hand side of Eq. (1) as known and the left-hand side as unknown.

In Sec. III, we calculate the outer multiplicity from the inner multiplicity by means of the Racah–Speiser lemma.

I. CALCULATION OF INNER MULTIPLICITIES BY MEANS OF BRANCHING LAWS

Both Sec. I and II are essentially based on one simple equation, derived by Straumann,[5] Klymyk,[6] and Delaney and Gruber[4]:

$$\tilde{\gamma}(M') = \sum_{S' \in W'} \sum_{m \in M} \gamma(m) \delta_{S'} \delta_{L(m)+S'R_0', M'+R_0'}, \quad (1)$$

where $\tilde{\gamma}(M')$ is the branching multiplicity of the irreducible representation M' belonging to the subgroup G', $\gamma(m)$ is the inner multiplicity of weight m belonging to the irreducible representation M of the original group G, and S' is the operation of the Weyl group belonging to the subgroup G'. Also, $\delta_{S'} = +1$ or -1 according to whether S' is even or odd, repectively, and R_0 is half the sum of all the positive roots in the subgroup G'. The second δ is a Kronecker delta.

A summary of R_0 and S (the Weyl group) and the dimension formula for all classical groups is given in Table I. The only symbol in Eq. (1) that requires explanation is $L(m)$. This will be treated in detail as we discuss each case separately in the following subsections.

A. Inner Multiplicity of Weights Belonging to $SU(n)$

We begin our treatment with the group $SU(n)$ because it is the best-known group among the classical groups. Again, we wish to refer to the fact that Delaney and Gruber[4] have already found an elegant method of finding the inner multiplicity of weights belonging to $SU(n)$. As far as $SU(n)$ is concerned, we do not claim that our method is better than theirs. However, we do believe that the real advantage of our method is in its applicability to the orthogonal and symplectic groups. This we hope will become clear in Secs. IB, IC, and ID.

An irreducible representation of $SU(n)$ can be expressed by the Gel'fand pattern (M_{ni}), where $i = 1, 2, \cdots, n$, with the M_{ni} all integers. Moreover, it is understood that two irreducible representations (M_{ni}) and (M'_{ni}) are the same if each M_{ni} differs from the corresponding M'_{ni} by an integer, i.e., if $M_{ni} - M'_{ni} = k$, for all i, where k is an integer. For example, in $SU(3)$, $(1, 0, -1)$ and $(2, 1, 0)$ are the same. It is, therefore, possible to make $M_{nn} = 0$. We shall call this convention of using the Gel'fand pattern (M_{ni}) to represent the irreducible representations of $SU(n)$ convention (a).

In the case of $SU(n)$, there is a second convention which we call convention (b). In this convention the irreducible representation of $SU(n)$ is represented by n numbers not necessarily integers $(m_1, m_2, \cdots, m_n)$ subject to the condition $\sum_{i=1}^{n} m_i = 0$.

Now, in the decomposition of $SU(n) \supset SU(n')$, $n > n'$, there are two ways of expressing $L(m)$ in Eq. (1). The first way is due to Lorente,[11] who uses convention (b). In this convention, $L(m) = (M'_j)$, where $j = 1, 2, \cdots, n'$ and $\sum_{j=1}^{n'} M'_j = 0$. Then,

$$M'_j = m_j + (1/n')(m_{n'+1} + m_{n'+2} + \cdots + m_n),$$

subject to the permutation of subscripts.

The second method makes use of convention (a) and, we believe, is simpler to use. This is because in convention (a) all numbers are integers, and one can apply the branching laws to them immediately. In this method, $L(m)$ is obtained as follows: Given the weights (m_{ni}), $i = 1, 2, \cdots, n$, in $SU(n)$, we select n' of them equal to (M'_j), $j = 1, 2, \cdots, n'$, and add to these any remaining $(n - n')$ ones so that the new $L(m)$ is equal to one of the weights (m_{ni}) arranged in any order whatever.

TABLE I. Pertinent data of classical groups.

	$SU(n) = A_{n-1}$	$O(2k+1) = B_k$	$O(2k) = D_k$	$Sp(2k) = C_k$
R_0	$\frac{1}{2}(n-1, n-3, \cdots, -n+1)$	$\frac{1}{2}(2k-1, 2k-3, \cdots, 1)$	$(k-1, k-2, \cdots, 0)$	$(k, k-1, \cdots, 1)$
S (Weyl group)	permutation	permutation and any change of sign	permutation and even number of change of sign	permutation and any change of sign
Dimension	$\frac{\xi(\lambda_1, \lambda_2, \cdots, \lambda_n)}{\xi(0, 0, \cdots, 0)}$	$\frac{\xi(\lambda_1, \lambda_2, \cdots, \lambda_k)}{\xi(0, 0, \cdots, 0)}$	$\frac{\xi(\lambda_1, \lambda_2, \cdots, \lambda_k)}{\xi(0, 0, \cdots, 0)}$	$\frac{\xi(\lambda_1, \lambda_2, \cdots, \lambda_k)}{\xi(0, 0, \cdots, 0)}$
$\xi(\lambda)$	$\prod_{i<j} (l_i - l_j)$ $l_j = \lambda_j + r_j$ $r_j = n - j$	$\prod_{i<j} (l_i^2 - l_j^2) \prod_i l_i$ $l_j = \lambda_j + r_j$ $r_j = k - j + \frac{1}{2}$	$\prod_{i<j} (l_i^2 - l_j^2)$ $l_j = \lambda_j + r_j$ $r_j = k - j$	$\prod_{i<j} (l_i^2 - l_j^2) \prod_i l_i$ $l_j = \lambda_j + r_j$ $r_j = k - j + 1$

We now illustrate these two methods by means of an example. Let us consider the 15-dimensional irreducible representation in $SU(4)$. According to convention (b), this is represented by $(1, 0, 0, -1)$. According to convention (a), this can be represented by $(2, 1, 1, 0)$.

Now we use Weyl's branching law, which states that in the decomposition $SU(n) \supset SU(n-1)$, represented by their Gel'fand pattern respectively, the following "triangular inequality" should be satisfied: $M_{ni} \geq M_{n-1,i} \geq M_{n,i+1}$. In the example above, using convention(a),the only M_{3i} that satisfy the triangular inequality are $(2, 1, 0)$, $(1, 1, 1)$, $(1, 1, 0)$, and $(2, 1, 1)$. In terms of dimensions, we have $15 = 8 + 1 + 3 + \bar{3}$.

We now wish to find the inner multiplicity of the weights $(2, 1, 1, 0)$ and $(0, 0, 0, 0)$ from Eq. (1), given the branching multiplicity in $SU(3)$, i.e., $\tilde{\gamma}(2, 1, 0) = \tilde{\gamma}(0, 0, 0) = \tilde{\gamma}(1, 1, 0) = \tilde{\gamma}(1, 0, 0) = 1$. According to the second method we immediately have from Eq. (1)

$$\tilde{\gamma}(2, 1, 0) = \gamma(2, 1, 0, 1) = 1, \tag{2}$$

$$\tilde{\gamma}(0, 0, 0) = \gamma(0, 0, 0, 0) - \gamma(2, 0, 1, 1) - \gamma(1, 2, 0, 1) = 1;$$

$$\therefore \quad \gamma(0, 0, 0, 0) = 3. \tag{3}$$

Equations (2) and (3) are sufficient, in this case, to obtain all the inner multiplicities of the weights of $(2, 1, 1, 0)$ in $SU(4)$. Of course, $\gamma(2, 1, 1, 0) = 1$ can be obtained from Cartan's theorem, because the highest weight is unique (i.e., it has multiplicity one). We can also use Eq. (1) to obtain

$$\tilde{\gamma}(1, 1, 0) = \gamma(1, 1, 0, 2) = 1,$$

$$\tilde{\gamma}(1, 0, 0) = \tilde{\gamma}(2, 1, 1) = \gamma(2, 1, 1, 0) = 1,$$

which just confirms the branching law.

According to convention (b), we have to rewrite $(2, 1, 0)$, $(1, 1, 1)$, $(1, 1, 0)$, and $(2, 1, 1)$ as $(1, 0, -1)$, $(0, 0, 0)$, $(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3})$, and $(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$, respectively. Then, using Eq. (1), we have

$$\tilde{\gamma}(1, 0, -1) = \gamma(1, 0, -1, 0) = 1, \tag{2'}$$

$$\tilde{\gamma}(0, 0, 0) = \gamma(0, 0, 0, 0) - \gamma(1, -1, 0, 0) - \gamma(0, 1, -1, 0) = 1, \tag{3'}$$

or

$$\gamma(0, 0, 0, 0) = 3.$$

Equations (2') and (3') give the same results as (2) and (3). We can also use Eq. (1) to obtain

$$\tilde{\gamma}(\tfrac{2}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}) = \gamma(1, 0, 0, -1) = 1,$$

$$\tilde{\gamma}(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}) = \gamma(0, 0, -1, 1) = 1,$$

which just confirms Weyl's branching law.

Thus, we see that convention (b) is slightly more complicated to use than convention (a). However, chronologically, convention (b) was the first one used successfully by Lorente in the decomposition of $SU(n) \supset SU(n')$.

B. Inner Multiplicity of $O(2k + 1) = B_k$

To find the inner multiplicity of weights in the irreducible representation of $O(2k + 1)$, we use the restriction $O(2k + 1) \supset O(2k)$ and the branching law

$$M_{2k+1,i} \geq M_{2k,i} \geq M_{2k+1,i+1}.$$

Note that $M_{2k,k}$ can be positive or negative. If it is negative, the branching law refers to its absolute value. However, the case where $M_{2k,k}$ is negative is similar to the case where it is positive and, in trying to obtain the inner multiplicities of $O(2k + 1)$, it is sufficient to consider the case where $M_{2k,k}$ is positive only.

The procedure is the same as in Section IA. $L(m)$ is now obtained in the following way:

$$L(m_{2k+1,i}) = (M_{2k,i}), \quad \text{where} \quad M_{2k,i} = m_{2k+1,i},$$

subject to the permutation of the i's.

We give the example of $O(7) \supset O(6)$. The irreducible representation of $O(7)$ we consider is $(2, 1, 1)$. Its dimension is 189. The only M_{6i} that satisfy the triangular inequalities are $(2, 1, 1)$, $(2, 1, -1)$, $(2, 1, 0)$, $(1, 1, 1)$, $(1, 1, -1)$, and $(1, 1, 0)$. Dimensional check gives $189 = 45 + 45 + 64 + 10 + 10 + 15$.

Now we wish to find the inner multiplicities of the weights $(2, 1, 1)$, $(2, 1, 0)$, $(2, 0, 0)$, $(1, 1, 1)$, $(1, 1, 0)$, $(1, 0, 0)$, and $(0, 0, 0)$ in the irreducible representation $(2, 1, 1)$ of $O(7)$. Using Eq. (1), we have the following:

$$\tilde{\gamma}(2, 1, 1) = \gamma(2, 1, 1) = 1, \tag{4}$$

$$\tilde{\gamma}(2, 1, 0) = \gamma(2, 1, 0) = 1, \tag{5}$$

$$\tilde{\gamma}(2, 0, 0) = \gamma(2, 0, 0) - \gamma(2, 1, -1) - \gamma(2, 1, 1) = 0,$$

$$\therefore \quad \gamma(2, 0, 0) = 2; \tag{6}$$

$$\tilde{\gamma}(1, 1, 1) = \gamma(1, 1, 1) - \gamma(2, 0, 1) - \gamma(1, 2, 0) = 1,$$

$$\therefore \quad \gamma(1, 1, 1) = 3; \tag{7}$$

$$\tilde{\gamma}(1, 1, 0) = \gamma(1, 1, 0) - \gamma(2, 0, 0) - \gamma(1, 2, -1) - \gamma(1, 2, 1) = 1,$$

$$\therefore \quad \gamma(1, 1, 0) = 5; \tag{8}$$

$$\tilde{\gamma}(1, 0, 0) = \gamma(1, 0, 0) - \gamma(2, -1, 0) - \gamma(1, 1, -1) - \gamma(1, 1, 1) + \gamma(1, 2, 0) = 0,$$

$$\therefore \quad \gamma(1, 0, 0) = 6; \tag{9}$$

$$\tilde{\gamma}(0, 0, 0) = \gamma(0, 0, 0) - \gamma(1, -1, 0) - \gamma(0, 1, -1) - \gamma(0, 1, 1) + \gamma(1, 1, -2) + \gamma(0, 2, 0) + \gamma(2, -1, -1) + \gamma(1, 1, 2) + \gamma(2, 1, 1) = 0,$$

$$\therefore \quad \gamma(0, 0, 0) = 9. \tag{10}$$

We note that each solution is either immediately obtained from itself or from the preceding ones. Another example is given in Sec. IIE.

C. Inner Multiplicity of $O(2k) = D_k$

Using the restriction $O(2k) \supset O(2k-1)$ and the branching law $M_{2k,1} \geq M_{2k-1,i} \geq M_{2k,i+1}$, we can easily obtain the inner multiplicities of weights in $O(2k)$. We now obtain $L(m)$ in the following way:

$$L(m) = (m_{2k-1,i'}, m_{2k-1,k}),$$

where

$$i' = 1, 2, \cdots, k-1.$$

$(m_{2k-1,i'})$ is the irreducible representation $(M_{2k-1,i'})$ in $O(2k-1)$, and $m_{2k-1,k}$ is any number such that $(m_{2k-1,i'}, m_{2k-1,k})$ is one of the weights in $O(2k)$.

For example, in the irreducible representation $(1, 1, 0, 0)$ of $O(8)$, let us find the inner multiplicities of the weights $(1, 1, 0, 0)$ and $(0, 0, 0, 0)$. First, we find from the restriction $O(8) \supset O(7)$ that the only $O(7)$ irreducible representations contained in $(1, 1, 0, 0)$ are $(1, 1, 0)$ and $(1, 0, 0)$. Dimensional check gives $28 = 21 + 7$. Then, from Eq. (1), we have

$$\tilde{\gamma}(1, 1, 0) = \gamma(1, 1, 0, 0) = 1, \tag{11}$$

$$\begin{aligned}\tilde{\gamma}(0, 0, 0) = \gamma(0, 0, 0, 0) &- \gamma(1, -1, 0, 0) \\ &- \gamma(0, 0, 1, -1) - \gamma(0, 0, 1, 1) \\ &- \gamma(0, 1, -1, 0) = 0;\end{aligned}$$

$$\therefore \quad \gamma(0, 0, 0, 0) = 4. \tag{12}$$

Equations (11) and (12) are sufficient to determine the inner multiplicities of all the weights of $(1, 1, 0, 0)$ in $O(8)$. Of course, we can also obtain from Eq. (1)

$$\tilde{\gamma}(1, 0, 0) = \gamma(1, 0, 0, 1) + \gamma(1, 0, 0, -1) - \gamma(1, 0, 1, 0) = 1,$$

which just confirms the branching law of the orthogonal group. Another example of $O(2k)$ is given in Sec. IIF.

D. Inner Multiplicity in $Sp(2n) = C_n$

In this case, we use the restriction $Sp(2n) \supset Sp(2(n-1))$ or $C_n \supset C_{n-1}$. The branching laws have been obtained by Hegerfeldt[10]:

$$M_{ni} \geq M'_{ni} \geq M_{n,i+1},$$
$$M'_{ni} \geq M_{n-1,i} \geq M'_{n,i+1},$$

where M'_{ni} runs from $i = 1$ to $i = n$. Here $L(m)$ is obtained in the same way as in IC, i.e., as in the case of $O(2k) \supset O(2k-1)$, except, of course, that in IC fractions are allowed, whereas in this case only integers are allowed.

Thus, for example, in the irreducible representation $(1, 1, 0)$ of $Sp(6)$ we obtain for M'_{3i} $(1, 1, 0)$ and $(1, 0, 0)$. Then, from $(1, 1, 0)$, we obtain $(1, 0)$ and $(1, 1)$. From $(1, 0, 0)$ we obtain $(1, 0)$ and $(0, 0)$; thus,

$$(1, 1, 0) = (1, 1) + 2(1, 0) + (0, 0).$$

Dimensional check gives $14 = 5 + 2 \times 4 + 1$.

To obtain the inner multiplicity of the weights $(1, 1, 0)$ and $(0, 0, 0)$, we use Eq. (1) again:

$$\tilde{\gamma}(1, 1) = \gamma(1, 1, 0) = 1 \tag{13}$$

$$\tilde{\gamma}(0, 0) = \gamma(0, 0, 0) - \gamma(1, -1, 0) = 1;$$

$$\gamma(0, 0, 0) = 2. \tag{14}$$

Equations (13) and (14) are sufficient to obtain the inner multiplicities of all the weights in the irreducible representation $(1, 1, 0)$ of $Sp(6)$. Of course, we can also use Eq. (1) to obtain

$$\tilde{\gamma}(1, 0) = \gamma(1, 0, 1) + \gamma(1, 0, -1) = 2,$$

which just confirms the branching law of Hegerfeldt.

II. DECOMPOSITION OF CLASSICAL GROUPS

In the previous section, we utilized the branching laws of $SU(n) \supset SU(n-1)$, $O(n) \supset O(n-1)$, and $Sp(2n) \supset Sp(2(n-1))$ to obtain the inner multiplicity of weights of the classical groups. Once the inner multiplicity is known, we can use it to obtain the branching multiplicities of irreducible representations of any subgroups by means of Eq. (1) again. These branching rules have been considered by Klymyk,[6] Whippman,[12] and Resnikoff *et al.*[13] We shall treat, in this section, the decomposition of an arbitrary compact Lie group into its subgroups, without any restriction.

A. $U(n) \supset U(m)$, $U(n) \supset U(m) \oplus U(n-m)$, $U(mn) \supset U(m) \otimes U(n)$

These three cases have been treated by Lorente.[11] In this paper, we refer the readers to Secs. IID and IIG, where similar cases are treated in the orthogonal group and symplectic group.

B. $U(n) \supset O(n)$

We shall treat this case without using Littlewood's theorem.[14] From now on, in an attempt to achieve some notational clarity, we shall use ( ) to denote unitary groups, [ ] to denote orthogonal groups, and { } to denote symplectic groups.

1. $U(2k) \supset O(2k)$ or $A_{2k-1} \supset D_k$

This case has been treated by Klymyk.[6] We shall just give an example: $U(6) \supset O(6)$. The irreducible representation of $U(6)$ we consider is $(3, 0, 0, 0, 0, 0)$. Its dimension is 56. From Sec. I, we have $\gamma(3, 0, 0, 0, 0, 0) = 1$, $\gamma(2, 1, 0, 0, 0, 0) = 1$, and $\gamma(1, 1, 1, 0, 0, 0) = 1$. We have $L(m) = (m_{6,i} - m_{6,i+3})$, where $i = 1, 2, 3$, subject to permutation of subscripts. Then, from Eq. (1), we obtain

$$\tilde{\gamma}[3, 0, 0] = \gamma(3, 0, 0, 0, 0, 0) = 1, \tag{15}$$

$$\begin{aligned}\tilde{\gamma}[2, 1, 0] &= \gamma(2, 0, 1, 0, 0, 0) - \gamma(3, 0, 0, 0, 0, 0) \\ &= 0,\end{aligned} \tag{16}$$

$$\begin{aligned}\tilde{\gamma}[1, 0, 0] = {} & \gamma(2, 1, 0, 0, 0, 0) + \gamma(1, 0, 1, 1, 0, 0) \\ & + \gamma(1, 0, 0, 0, 1, 1) + \gamma(1, 0, 2, 0, 0, 0) \\ & - \gamma(1, 0, 1, 0, 0, 1) - \gamma(1, 0, 1, 0, 1, 0) \\ & - \gamma(2, 0, 0, 1, 0, 0) = 1,\end{aligned} \tag{17}$$

$$\begin{aligned}\tilde{\gamma}[1, 1, 1] = {} & \gamma(1, 0, 1, 0, 1, 0) - \gamma(1, 0, 2, 0, 0, 0) \\ & - \gamma(2, 0, 0, 0, 1, 0) + \gamma(3, 0, 0, 0, 0, 0) \\ = {} & 0,\end{aligned} \tag{18}$$

$$\begin{aligned}\tilde{\gamma}[1, 1, -1] = {} & \gamma(1, 0, 1, 0, 0, 1) + \gamma(3, 0, 0, 0, 0, 0) \\ & - \gamma(2, 0, 0, 0, 0, 1) \\ & - \gamma(1, 0, 2, 0, 0, 0) = 0.\end{aligned} \tag{19}$$

The result is

$$(3, 0, 0, 0, 0, 0) = [3, 0, 0] + [1, 0, 0],$$
$$56 = 50 + 6.$$

2. $U(2k + 1) \supset O(2k + 1)$ or $A_{2k} \supset B_k$

The prescription for $L(m)$ is, in fact, the same as in 11B*1* above; i.e.,

$$L(m) = (m_{2k+1,i} - m_{2k+1,i+k}, m_{2k+1,2k+1}),$$

where $(m_{2k+1,i} - m_{2k+1,i+k}) = (M_{2k+1,i})$, $i = 1, 2, \cdots, k$. $(M_{2k+1,i})$ is the irreducible representation of $O(2k + 1)$. $(m_{2k+1,i} \cdots m_{2k+1,i+k} m_{2k+1,2k+1})$ is a weight of $U(2k + 1)$, arranged in any order whatever.

For example, in $SU(3) \supset O(3)$, let us decompose the octet $(1, 0, -1)$. From Sec. I, we have

$$\gamma(1, 0, -1) = 1 \quad \text{and} \quad \gamma(0, 0, 0) = 2.$$

Then Eq. (1) gives

$$\begin{aligned}\tilde{\gamma}[2] &= \gamma(1, -1, 0) = 1, \\ \tilde{\gamma}[1] &= \gamma(1, 0, -1) + \gamma(0, -1, 1) - \gamma(1, -1, 0) \\ &= 1, \\ \tilde{\gamma}[0] &= \gamma(0, 0, 0) - \gamma(1, 0, -1) - \gamma(0, -1, 1) = 0.\end{aligned}$$

The result is

$$(1, 0, -1) = [2] + [1],$$
$$8 = 5 + 3.$$

We note that this decomposition is different from the decomposition $SU(3) \supset SU(2)$, where

$$(1, 0, -1) = (1, -1) + 2(\tfrac{1}{2}, -\tfrac{1}{2}) + (0, 0),$$
$$8 = 3 + 2 \times 2 + 1,$$

even though $O(3)$ and $SU(2)$ are homeomorphic to each other.

C. $U(2n) \supset Sp(2n)$ or $A_{2n-1} \supset C_n$

The prescription for $L(m)$ is as follows:

$$L(m) = (m_i - m_{i+k}), \quad i = 1, 2, \cdots, k,$$

where $(m_i - m_{i+k}) = (M_i)$, the irreducible representation of $Sp(2k)$.

For example, in $SU(6) \supset Sp(6)$, let the irreducible representation of $SU(6)$ be $(1, 0, 0, 0, 0, -1)$. Its dimension is 35. From Sec. I we have $\gamma(1, 0, 0, 0, 0, -1) = 1$ and $\gamma(0, 0, 0, 0, 0, 0) = 5$. Then, from Eq. (1), we have

$$\tilde{\gamma}\{2, 0, 0\} = \gamma(1, -1, 0, 0, 0, 0) = 1,$$

$$\begin{aligned}\tilde{\gamma}\{1, 1, 0\} = {} & \gamma(1, 0, 0, -1, 0, 0) \\ & + \gamma(0, -1, 1, 0, 0, 0) \\ & - \gamma(1, -1, 0, 0, 0, 0) = 1,\end{aligned}$$

$$\begin{aligned}\tilde{\gamma}\{0, 0, 0\} = {} & \gamma(0, 0, 0, 0, 0, 0) - \gamma(0, 0, 1, 0, -1, 0) \\ & - \gamma(0, 0, 0, -1, 0, 1) \\ & - \gamma(1, 0, -1, 0, 0, 0) \\ & - \gamma(0, -1, 0, 1, 0, 0) \\ & - \gamma(0, 0, 0, 0, 1, -1) = 0;\end{aligned}$$

$$\therefore \quad (1, 0, 0, 0, 0, -1) = \{2, 0, 0\} + \{1, 1, 0\},$$
$$35 = 21 + 14.$$

D. $O(n) \supset O(m)$, $O(n) \supset O(m) \oplus O(n - m)$, $O(mn) \supset O(n) \oplus O(m)$

1. $O(n) \supset O(m)$, $n > m$

There are at least two ways of decomposing $O(n)$ into $O(m)$. The first one is to apply the branching law of $O(n) \supset O(n - 1)$ successively until one reaches $O(m)$. The second one is to compute the inner multiplicity of weights belonging to $O(n)$: first, using the method of Sec. I and, then, using Eq. (1) again to obtain the branching multiplicities in $O(m)$. As an example, let us consider $O(7) \supset O(4)$. The irreducible representation of $O(7)$ we consider is $(2, 1, 1)$. Then, using either of the two methods above, one obtains

$$\begin{aligned}[2, 1, 1] = {} & 3[2, 1] + 3[2, -1] + 4[2, 0] + 9[1, 1] \\ & + 9[1, -1] + 12[1, 0] + 3[0, 0].\end{aligned}$$

Dimensional check gives

$$\begin{aligned}189 = {} & 3 \times 8 + 3 \times 8 + 4 \times 9 + 9 \times 3 \\ & + 9 \times 3 + 12 \times 4 + 3.\end{aligned}$$

2. $O(n) \supset O(m) \oplus O(n-m)$

Both in this case and in the case of Sec. IID*3* below, because there are two subgroups involved instead of one, we have to use a formula derived by Lorente[11] instead of Eq. (1) above. This formula is

$$\tilde{\gamma}(M'M'') = \sum_m \gamma(m) \sum_{S'} \sum_{S''} \delta_{S'}\delta_{S''}\delta_{L'(m)+S'R_0',M'+R_0'} \times \delta_{L''(m)+S''R_0'',M''+R_0''}, \quad (20)$$

where the first two deltas equal $+1$ or -1 according to whether $S'(S'')$ is even or odd, respectively. The third and fourth deltas are Kronecker deltas. The rest of the symbols have the same meaning as in Eq. (1). Everything primed refers to the first subgroup G' and everything double primed refers to the second subgroup G''. The inner multiplicity $\gamma(m)$ of $O(n)$ is calculated according to Sec. I. As an example, let us consider $O(7) \supset O(4) \oplus O(3)$. The irreducible representation of $O(7)$ will again be (2, 1, 1). Then, using Eq. (20), we obtain

$$\begin{aligned}[2,1,1] = {} & [2,1][1] + [2,-1][1] + [2,0][1] \\ & + [2,0][0] + [1,1][2] + [1,-1][2] \\ & + [1,1][1] + [1,-1][1] + [1,1][0] \\ & + [1,-1][0] + [1,0][2] + 2[1,0][1] \\ & + [1,0][0] + [0,0][1].\end{aligned}$$

Dimensional check gives

$$\begin{aligned}189 = {} & 8\times 3 + 8\times 3 + 9\times 3 + 9 + 3\times 5 + 3\times 5 \\ & + 3\times 3 + 3\times 3 + 3 + 3 + 4\times 5 + 2\times 4\times 3 \\ & + 4 + 3.\end{aligned}$$

3. $O(mn) \supset O(m) \otimes O(n)$

Again we use Eq. (20) and the method in Sec. I for the inner multiplicity of $O(mn)$. As an example, let us consider $O(9) \supset O(3) \otimes O(3)$. The irreducible representation of $O(9)$ we consider is [1, 1, 1, 0]. Its dimension is 84. We obtain

$$[1,1,1,0] = 5[1][1] + 12[1][0] + 3[0][0],$$

and dimensional check gives $84 = 45 + 36 + 3$.

E. $O(2k+1) \supset SU(k)$ or $B_k \supset A_{k-1}$

Here using convention a for $SU(k)$, we have

$$L(m) = (M_i') = (m_i), \quad i = 1, 2, \cdots, k.$$

As an example, let us consider $O(5) \supset SU(2)$. The irreducible representation of $O(5)$ we consider is [3, 1]. Its dimension is 81. From the branching law of $O(5) \supset O(4)$ and Eq. (1), we obtain

$$\begin{aligned}&\gamma[3,1] = 1, \quad \gamma[2,2] = 1, \quad \gamma[2,1] = 2, \\ &\gamma[2,0] = 3, \quad \gamma[1,1] = 4, \quad \gamma[1,0] = 4, \\ &\gamma[0,0] = 5.\end{aligned}$$

Now applying Eq. (1) again for $O(5) \supset SU(2)$, we have

$$\begin{aligned}\tilde{\gamma}(4,0) &= \gamma[3,-1] + \gamma[1,-3] + \gamma[2,-2] = 3, \\ \tilde{\gamma}(3,0) &= \gamma[3,0] + \gamma[2,-1] + \gamma[1,-2] + \gamma[0,-3] \\ &= 6, \\ \tilde{\gamma}(2,0) &= \gamma[3,1] + \gamma[2,0] + \gamma[1,-1] \\ &\quad + \gamma[-1,-3] + \gamma[0,-2] - \gamma[3,-1] \\ &\quad - \gamma[2,-2] - \gamma[1,-3] = 9, \\ \tilde{\gamma}(1,0) &= \gamma[2,1] + \gamma[1,0] + \gamma[0,-1] \\ &\quad + \gamma[-1,-2] - \gamma[3,0] - \gamma[2,-1] \\ &\quad - \gamma[1,-2] - \gamma[0,-3] = 6, \\ \tilde{\gamma}(0,0) &= \gamma[2,2] + \gamma[1,1] + \gamma[0,0] + \gamma[-1,-1] \\ &\quad + \gamma[-2,-2] - \gamma[1,-1] - \gamma[3,1] \\ &\quad - \gamma[2,0] - \gamma[0,-2] - \gamma[-1,-3] = 3;\end{aligned}$$

$$\therefore \quad [3,1] = 3(4,0) + 6(3,0) + 9(2,0) + 6(1,0) + 3(0,0),$$

$$81 = 3\times 5 + 6\times 4 + 9\times 3 + 6\times 2 + 3.$$

F. $O(2k) \supset SU(k)$ or $D_k \supset A_{k-1}$

$L(m)$ is the same as above. As an example, let us consider $O(6) \supset SU(3)$. The irreducible representation of $O(6)$ we consider is [2, 1, 1]. Its dimension is 45. First, we use the method of Sec. I to obtain

$$\begin{aligned}&\gamma[2,1,1] = 1, \quad \gamma[2,0,0] = 1, \\ &\gamma[1,1,0] = 2, \quad \gamma[0,0,0] = 3.\end{aligned}$$

Then, in terms of $O(6) \supset SU(3)$, we have

$$\begin{aligned}[2,1,1] = {} & (3,0,0) + (3,1,0) + (2,0,0) \\ & + (2,1,0) + (1,1,0) + (1,0,0),\end{aligned}$$

$$45 = 10 + 15 + 6 + 8 + 3 + 3.$$

G. $Sp(2n) \supset Sp(2m)$, $Sp(2n) \supset Sp(2m) \oplus Sp(2n-2m)$, $Sp(2mn) \supset Sp(2n) \otimes Sp(2m)$

The procedure is the same as in Secs. IIA and IID. We shall give only one example here: $Sp(8) \supset Sp(4) \otimes Sp(4)$. The irreducible representation in $Sp(8)$ we consider is $\{1,1,1,0\}$. From ID, we obtain $\gamma\{1,1,1,0\} = 1$ and $\gamma\{1,0,0,0\} = 2$. Then, from Eq. (20), we have

$$\{1,1,1,0\} = 2\{1,1\}\{1,0\} + 2\{1,0\}\{0,0\}.$$

Dimensional check gives $48 = 2\times 5\times 4 + 2\times 4$.

H. $Sp(2n) \supset SU(n)$ or $C_n \supset A_{n-1}$

$L(m)$ is the same as in IIE and IIF. For example, let us consider $Sp(6) \supset SU(3)$. The irreducible representation of $Sp(6)$ we consider is $\{1,1,0\}$. Its dimension is 14. From ID, we obtain $\gamma\{1,1,0\} = 1$ and

$\gamma\{0, 0, 0\} = 2$. Then, from Eq. (1), we obtain

$$\{1, 1, 0\} = (2, 1, 0) + (1, 1, 0) + (1, 0, 0),$$
$$14 = 8 + 3 + 3.$$

III. DIRECT PRODUCTS IN CLASSICAL GROUPS

For completeness, we also will give some examples of the Clebsch–Gordan series in the direct product of two irreducible representations in a group. This can be easily achieved through the Racah–Speiser lemma,[15] once the inner multiplicity is known. The Racah–Speiser lemma connects the outer multiplicity linearly with the inner multiplicity. It can be written as

$$\bar{\gamma}(M) = \sum_S \sum_m \gamma(m)\delta_S \delta_{S(\Lambda'+m+R_0),M+R_0}. \qquad (21)$$

Here $\bar{\gamma}(M)$ is the outer multiplicity for the irreducible representation (M); $\gamma(m)$ is the inner multiplicity of (m) belonging to the irreducible representation (Λ), one of the two irreducible representations whose direct product we are computing; S = operation of the Weyl group; $\delta_S = +1$ or -1, according to whether S is even or odd, respectively; R_0 = half the sum of all the positive roots; (Λ') is the other irreducible representation of the group under consideration; and the second δ is a Kronecker delta.

The Clebsch–Gordan series for the unitary group can be easily obtained through the Young tableaux, though it can also be obtained through the Racah–Speiser lemma. It is well known and will not be treated here.

We give some examples of the direct products of two irreducible representations belonging to $O(2k)$, $O(2k + 1)$, and $Sp(2n)$.

Example 1: $O(8)$: $[2, 1, 1, 1] \times [1, 1, 0, 0]$

From Sec. I, we have

$$\gamma[1, 1, 0, 0] = 1, \quad \gamma[0, 0, 0, 0] = 4.$$

Then, from Eq. (21), we have

$$\bar{\gamma}[3, 2, 1, 1] = \gamma[1, 1, 0, 0] = 1,$$
$$\bar{\gamma}[3, 1, 1, 0] = \gamma[1, 0, 0, -1] = 1,$$
$$\bar{\gamma}[1, 1, 1, 0] = \gamma[-1, 0, 0, -1] = 1,$$
$$\bar{\gamma}[2, 2, 2, 1] = \gamma[0, 1, 1, 0] = 1,$$
$$\bar{\gamma}[2, 2, 1, 0] = \gamma[0, 1, 0, -1] = 1,$$
$$\bar{\gamma}[2, 1, 1, 1] = \gamma[0, 0, 0, 0] - \gamma[0, -1, 1, 0] - \gamma[0, 0, -1, 1] = 2,$$
$$\bar{\gamma}[2, 1, 0, 0] = \gamma[0, 0, -1, -1] = 1;$$

$$\therefore \quad [2, 1, 1, 1] \times [1, 1, 0, 0] = [3, 2, 1, 1] + [3, 1, 1, 0] + [1, 1, 1, 0] + [2, 2, 2, 1] + [2, 2, 1, 0] + [2, 1, 0, 0] + 2[2, 1, 1, 1],$$

$$224 \times 28 = 2800 + 1296 + 56 + 672 + 840 + 160 + 2 \times 224 = 6272.$$

Example 2: $O(5)$: $[2, 1] \times [1, 1]$

The result is

$$[2, 1] \times [1, 1] = [3, 2] + [3, 1] + [2, 2] + 2[2, 1] + [3, 0] + [2, 0] + [1, 1] + [1, 0],$$
$$35 \times 10 = 105 + 81 + 35 + 2 \times 35 + 30 + 14 + 10 + 5.$$

Example 3: $Sp(6)$: $\{2, 1, 1\} \times \{1, 1, 0\}$

Here the result is

$$\{2, 1, 1\} \times \{1, 1, 0\} = \{3, 2, 1\} + \{3, 1, 0\} + \{2, 2, 0\} + \{2, 0, 0\} + \{2, 1, 1\} + \{1, 1, 0\} + \{2, 2, 2\},$$
$$70 \times 14 = 512 + 189 + 90 + 21 + 70 + 14 + 84.$$

ACKNOWLEDGMENTS

I wish to thank B. Gruber, R. M. Delaney, and M. Lorente for many valuable discussions.

* Present address: Fairfield University, Fairfield, Conn. 06430.

[1] N. Jacobson, *Lie Algebras* (Interscience Publishers, New York, 1965), p. 261.
[2] Reference 1, p. 247.
[3] G. Racah, *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon and Breach, Science Publishers, New York, 1962).
[4] R. M. Delaney and B. Gruber, J. Math. Phys. **10**, 252 (1969).
[5] N. Straumann, Helv. Phys. Acta **38**, 481 (1966).
[6] A. U. Klymyk, Academy of Sciences, Ukrainian SSR, Institute for Theoretical Physics, Kiev, USSR, No. 67-17, 1967.
[7] In our terminology, we make no effort to distinguish between $U(n)$ and $SU(n)$, $O(n)$ and $SO(n)$. For the sake of convenience, one can regard everything treated here as referring to $SU(n)$ and $SO(n)$. However, one can easily extend the results to $U(n)$ and $O(n)$ if one so wishes.
[8] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, New York, 1931), 2nd ed.
[9] H. Boerner, *Representations of Groups* (North-Holland Publishing Co., Amsterdam, 1963), Chap. VIII.
[10] G. C. Hegerfeldt, J. Math. Phys. **8**, 1195 (1967).
[11] M. Lorente, thesis, St. Louis University, St. Louis, Missouri, 1968.
[12] M. L. Whippman, J. Math. Phys. **6**, 1534 (1965).
[13] M. Resnikoff *et al.*, State University of New York at Buffalo, Preprints.
[14] D. E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford University Press, London, 1958).
[15] A. J. MacFarlane, L. O'Raifeartaigh, and P. S. Rao, J. Math. Phys. **8**, 536 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 4 APRIL 1970

Doppler Shift in Frequency in the Transport of Electromagnetic Waves through an Underdense Plasma*

JERRY L. PEACHER†
Space Sciences Laboratory, University of California, Berkeley, California
AND
KENNETH M. WATSON
Department of Physics and Lawrence Radiation Laboratory, University of California, Berkeley, California

(Received 3 July 1969)

In an earlier publication, the validity of the radiation transport theory was studied for the calculation of multiple scattering of electromagnetic waves by a turbulent plasma. In the present paper, we extend the transport theory to include a description of the Doppler shift in frequency caused by electron motion.

1. INTRODUCTION

In Part I of this series[1] the classical radiation transport equation was derived from Maxwell's equations for the study of scattering of electromagnetic waves by a turbulent plasma. In Part II[2] some techniques for using the transport equation were discussed. In both of these papers the Doppler shift in frequency caused by the motion of the scattering electrons was neglected. In the present paper we extend the transport theory to include any frequency shift of the scattered waves.

An exhaustive analysis of the relation between a wave equation and the corresponding classical transport approximation has yet to be made. The first such analysis seems to have been given by Foldy,[3] who discussed the scattering of scalar waves by a set of uncorrelated point scatterers, obtaining a transport equation. The quantum theory of scattering by a "weakly bound medium" was related to a classical transport theory by Watson.[4] It was an adaptation of the methods used in this work to Maxwell's equations which was given in I. A different approach was used by Barabanenkov and Finkel'berg,[5] who derived a transport equation from the scalar wave equation using a "Bethe–Salpeter" type of equation.

In Sec. II, we summarize the results derived in this paper. These lead to a radiation transport equation of conventional form,[6] the scattering kernel being explicitly expressed in terms of plasma density fluctuations. The reader who is not interested in the details of the derivation will probably find the account in Sec. II adequate for using the transport equation.

2. SUMMARY OF RESULTS

The phenomena which we wish to describe are illustrated in Fig. 1. A plasma of finite extent is illuminated by an electromagnetic wave emitted by a distant source S and propagating in the direction $\hat{\mathbf{k}}$. The intensity of the waves scattered by the plasma is measured with a receiver R, also a great distance away. (The restriction to a distant source and receiver is, of course, not required for a derivation of the transport equation.)

Several assumptions concerning the plasma were introduced in I. We shall accept these here and, in addition, explicitly suppose the plasma electrons to have nonrelativistic energies. The assumed turbulence properties of the plasma will be reviewed later in this section. The nonrelativistic assumption will be expressed by the inequality

$$(kR_c)(v_e/c) \ll 1,$$

which we will call NR. Here $k/2\pi$ is the wave number of the radiation, R_c a measure of the distance over which plasma motions are correlated, v_e the mean speed of the plasma electrons, and c the speed of light.

As in I, we suppose the plasma to be underdense (Assumption B3) and that $kR_S \gg 1$ (Assumption B4), where R_S is the "size" of the plasma. Assumption B4 allows us to ignore diffraction scattering from the entire plasma (in all but a small cone with axis parallel to $\hat{\mathbf{k}}$).

In the classical theory of radiation transport, the flow of radiant energy at a point $\mathbf{x}$ per unit area, per unit time, and traveling in the direction $\hat{\mathbf{p}}$ is

$$I(\mathbf{x}, \hat{\mathbf{p}}, \omega)\, d\Omega_{\hat{\mathbf{p}}}\, d\omega. \tag{2.1}$$

The notation here implies that the radiation has an angular frequency ω, within the interval $d\omega$, and is confined to propagation directions lying within the solid angle $d\Omega_{\hat{\mathbf{p}}}$.

For waves which have some degree of polarization, it is necessary to generalize (2.1). This was done by Chandrasekhar[6] and, in a similar manner, in I. To do this, we shall follow the notation of I and introduce the two unit vectors $\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(i)$, $i = 1, 2$, for a plane

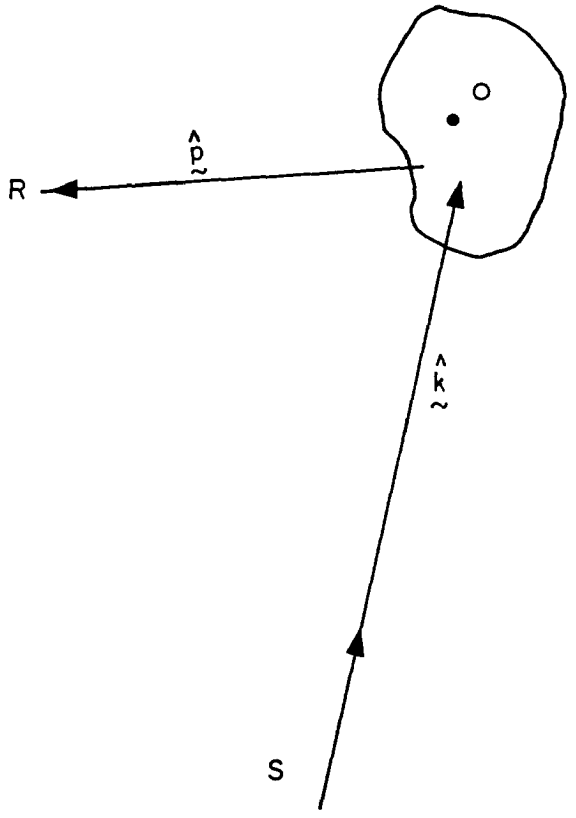


FIG. 1. Illustration of scattering from a plasma.

electromagnetic wave traveling in the direction $\hat{\mathbf{p}}$. The electric field vector for such a wave is of the form

$$\mathbf{E}_{\hat{\mathbf{p}}} = [E_{\hat{\mathbf{p}}}(1)\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(1) + E_{\hat{\mathbf{p}}}(2)\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(2)]e^{-i\omega t}. \quad (2.2)$$

The unit vectors $\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(i)$ are defined in terms of $\hat{\mathbf{k}}$, the direction of propagation of the incident wave before entering the plasma.[7] These are

$$\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(2) = C(\hat{\mathbf{p}})\hat{\mathbf{p}} \times \hat{\mathbf{k}},$$
$$\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(1) = \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(2) \times \hat{\mathbf{k}}, \quad (2.3)$$

where

$$C(\hat{\mathbf{p}}) = (|\hat{\mathbf{p}} \times \hat{\mathbf{k}}|)^{-1}. \quad (2.3')$$

To define the polarization vectors for the incident wave, we orient the z axis of a rectangular coordinate system to be parallel to $\hat{\mathbf{k}}$ and choose $\hat{\mathbf{e}}_{\hat{\mathbf{k}}}(1)$ and $\hat{\mathbf{e}}_{\hat{\mathbf{k}}}(2)$ to be parallel to the x and y axes, respectively. For backscatter we define[8]

$$\hat{\mathbf{e}}_{-\hat{\mathbf{k}}}(1) = \hat{\mathbf{e}}_{\hat{\mathbf{k}}}(1),$$
$$\hat{\mathbf{e}}_{-\hat{\mathbf{k}}}(2) = -\hat{\mathbf{e}}_{\hat{\mathbf{k}}}(2). \quad (2.4)$$

The electric field at any point can be represented as a sum of waves of the form (2.2). If we fix our attention on a single "bundle" of wavelets propagating within $d\Omega_{\hat{\mathbf{p}}}$ and $d\omega$, we may define the intensity as

$$I_{ij}(\mathbf{x}, \hat{\mathbf{p}}, \omega) = \text{const} \times \langle E^*_{\hat{\mathbf{p}}}(i)E_{\hat{\mathbf{p}}}(j)\rangle, \quad (2.5)$$

$i, j = 1, 2$. Here "$\langle \cdots \rangle$" represents an ensemble (or statistical) average over the plasma (and any source) fluctuations. The "constant" in Eq. (2.5) is defined by the following condition. We suppose that a filter at $\mathbf{x}$ passes only the component of $\mathbf{E}$ parallel to some direction $\hat{\mathbf{e}}$. Then, the power per unit area passed by the filter, corresponding to propagation within $d\Omega_{\hat{\mathbf{p}}}$ and frequency within $d\omega$, is

$$\hat{\mathbf{e}} \cdot \left[\sum_{i,j=1}^{2} \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(i)I_{ij}(\mathbf{x}, \hat{\mathbf{p}}, \omega)\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j)\right] \cdot \hat{\mathbf{e}}\, d\Omega_{\hat{\mathbf{p}}}\, d\omega. \quad (2.6)$$

We suppose the statistical properties of the plasma to be represented as a stationary random process. If the plasma contains N free electrons with coordinates $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N$, we take the probability that electron 1 is at $\mathbf{z}_1$ within d^3z_1 at time t_1, etc., to be

$$P_N(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2; \cdots; \mathbf{z}_N, t_N)\, d^3z_1 \cdots d^3z_N. \quad (2.7)$$

The statement that this is a stationary distribution function is equivalent to[9]

$$P_N(\mathbf{z}_1, t_1 + \tau; \mathbf{z}_2, t_2 + \tau; \cdots; \mathbf{z}_N, t_N + \tau) = P_N(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2; \cdots; \mathbf{z}_N, t_N). \quad (2.8)$$

We further suppose that from P_N we can define a hierarchy of distribution functions as follows:

$$P_1(\mathbf{z}_1) = \int P_N\, d^3z_2 \cdots d^3z_N, \quad (2.9a)$$

$$P_2(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2) = \int P_N\, d^3z_3 \cdots d^3z_N, \quad (2.9b)$$

etc. Here

$$P_2(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2) = P_2(\mathbf{z}_1, 0; \mathbf{z}_2, t_2 - t_1),$$

etc.

Following the notation of I, we assume that P_2, $P_3, \cdots$ may be developed in terms of correlation functions. Thus, for example,

$$P_2(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2) = P_1(\mathbf{z}_1)P_1(\mathbf{z}_2)[1 + g(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2)]. \quad (2.10)$$

Here the "pair correlation function" g is considered to vanish for $|\mathbf{z}_1 - \mathbf{z}_2| \gg R_c$, the "correlation range," or for $|t_1 - t_2| \gg t_c$, the "correlation time." Again, we write P_3 in the form

$$P_3(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2; \mathbf{z}_3, t_3) = P_1(\mathbf{z}_1)P_1(\mathbf{z}_2)P_1(\mathbf{z}_3) \times [1 + g(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2) + g(\mathbf{z}_2, t_2; \mathbf{z}_3, t_3) + g(\mathbf{z}_3, t_3; \mathbf{z}_1, t_1) + g_3(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2; \mathbf{z}_3, t_3)]. \quad (2.11)$$

The "triplet correlation function" g_3 is assumed to vanish when any pair of the three coordinates is

separated by a distance large compared to R_c or any pair of times, by an interval large compared to t_c.

Continuing as above, we can express the probability functions $P_4, P_5, \cdots$ in terms of correlation functions.[10] The n-particle correlation function

$$g_n(\mathbf{z}_1, t_1; \cdots; \mathbf{z}_n, t_n)$$

vanishes unless all n coordinates lie within a volume characterized by the linear dimension R_c and all n times within an interval of order t_c.

In the absence of significant effects from external magnetic fields and/or Coriolis forces, time-reversal invariance implies several symmetry relations for the P's and g's.[11-13] For the pair correlation we have, for example,

$$g(\mathbf{z}_1, t; \mathbf{z}_2, 0) = g(\mathbf{z}_1, -t; \mathbf{z}_2, 0). \quad (2.12a)$$

Because we have assumed a stationary random process, we may conclude that

$$\begin{aligned} g(\mathbf{z}_1, t; \mathbf{z}_2, 0) &= g(\mathbf{z}_1, 0; \mathbf{z}_2, -t) \\ &= g(\mathbf{z}_1, 0; \mathbf{z}_2, t), \end{aligned} \quad (2.12b)$$

using (2.12a). On setting $t_2 - t_1 \equiv \tau$, we obtain

$$\begin{aligned} g(\mathbf{z}_1, t_1; \mathbf{z}_2, t_2) &\equiv g(\mathbf{z}_1, \mathbf{z}_2; \tau) \\ &= g(\mathbf{z}_1, \mathbf{z}_2; -\tau). \end{aligned} \quad (2.13)$$

We finally assume, following I, that

$$\begin{aligned} g(\mathbf{z}_1, \mathbf{z}_2; \tau) &= g(\mathbf{z}_1; |\mathbf{z}_1 - \mathbf{z}_2|; \tau) \\ &\cong g(\mathbf{z}_2; |\mathbf{z}_1 - \mathbf{z}_2|; \tau). \end{aligned} \quad (2.14)$$

[The assumption (2.14) is not required for the derivation of the transport equation. It does permit us to write the scattering kernel (2.19) in "prettier" form, however.]

The mean plasma electron density at a point $\mathbf{z}$ is

$$\rho(\mathbf{z}) = NP_1(\mathbf{z}). \quad (2.15)$$

The electron collision frequency at $\mathbf{z}$ will be written as $\nu_c(\mathbf{z})$ and the plasma frequency as

$$\omega_p(\mathbf{z}) = [4\pi e^2 \rho(\mathbf{z})/m]^{\frac{1}{2}}. \quad (2.16)$$

The refractive index $n(\mathbf{z})$ of the plasma was discussed in I. The first approximation to this was written as n_1 and is given by the familiar expression

$$n_1^2(\mathbf{z}) = 1 - \omega_p^2[\omega(\omega + i\nu_c)]^{-1}. \quad (2.17a)$$

We shall, as in I, suppose the imaginary part of $n(\mathbf{z})$ to be negligible for propagation over distances comparable to R_c. This permits us to take

$$n_1^2(\mathbf{z}) \cong 1 - \omega_p^2(\omega^2 + \nu_c^2)^{-1} \quad (2.17b)$$

in Eqs. (2.20) and (2.22) below.

The absorption length $l_c(\mathbf{z})$ caused by electron collisions is expressed as

$$\frac{1}{l_c(\mathbf{z})} \cong \frac{\omega_p^2}{(\omega^2 + \nu_c^2)} \frac{\nu_c}{c}, \quad (2.18)$$

where c is the speed of light.

We now define the *scattering kernel* $\mathbf{M}$ for scattering a wave from the direction $\hat{\mathbf{p}}'$ to direction $\hat{\mathbf{p}}$ as

$$(ij|\, M(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \Omega)\, |sr) \equiv \sigma_g(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \Omega)(ij|\, m\, |sr), \quad (2.19a)$$

where

$$(ij|\, m\, |sr) = [\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(i) \cdot \hat{\mathbf{e}}_{\hat{\mathbf{p}}'}(s)][\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j) \cdot \hat{\mathbf{e}}_{\hat{\mathbf{p}}'}(r)] \quad (2.19b)$$

and

$$\begin{aligned} \sigma_g(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \Omega) = \left(\frac{r_0^2}{1 + (\nu_c/\omega)^2}\right)[\rho^2(\mathbf{z})] \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau e^{i\Omega\tau} \\ \times \int d^3R g(\mathbf{z}; R, \tau) \\ \times \exp\,[in_1(\mathbf{z})k(\hat{\mathbf{p}}' - \hat{\mathbf{p}}) \cdot \mathbf{R}]. \end{aligned} \quad (2.20)$$

For later reference we observe that because of the time reversal invariance property (2.12a) σ_g is *even* in Ω.

The absorption length $l_t(\mathbf{z})$ for scattering is defined by the equation

$$\frac{1}{l_t(\mathbf{z})} = \tfrac{1}{2} \int d\Omega_{\hat{\mathbf{p}}'} \sigma_g(\hat{\mathbf{p}} \cdot \hat{\mathbf{p}}')[1 + (\hat{\mathbf{p}} \cdot \hat{\mathbf{p}}')^2], \quad (2.21)$$

where

$$\begin{aligned} \sigma_g(\hat{\mathbf{p}} \cdot \hat{\mathbf{p}}') &\equiv \int_{-\infty}^{\infty} d\Omega \sigma_g(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \Omega) \\ &= \left(\frac{r_0^2}{1 + \nu_c^2/\omega^2}\right)\rho^2(\mathbf{z}) \int d^3R g(\mathbf{z}; R) \\ &\quad \times \exp\,[in_1 k(\hat{\mathbf{p}}' - \hat{\mathbf{p}}) \cdot \mathbf{R}] \end{aligned} \quad (2.22)$$

and

$$g(\mathbf{z}; R) \equiv g(\mathbf{z}; R, 0). \quad (2.23)$$

An elementary calculation yields

$$\sum_{s=1}^{2} \int_{-\infty}^{\infty} d\Omega \int d\Omega_{\hat{\mathbf{p}}'} (ij|\, M(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \Omega)\, |ss) = \frac{\delta_{ij}}{l_t}. \quad (2.24)$$

The net absorption length $l(\mathbf{z})$ is defined, finally, as

$$l^{-1} = l_t^{-1} + l_c^{-1}. \quad (2.25a)$$

We note that this is equivalent to the equation

$$l^{-1} = 2k \operatorname{Im} n, \quad (2.25b)$$

where n is the refractive index given to the order of accuracy obtained in I.

The transport equation for I_{ij} (to be derived in Sec. V) is

$$\hat{\mathbf{p}} \cdot \nabla \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}, \omega) + \frac{1}{l(\mathbf{x})} \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}, \omega)$$
$$= \int_0^\infty d\omega' \int d\Omega_{\hat{\mathbf{p}}'} \mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \omega - \omega') \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}', \omega'). \quad (2.26)$$

Here we have written I_{ij} as a column matrix with four elements and $(ij|\, M \,|sr)$ as a 4×4 square matrix (evaluated at the point $\mathbf{x}$, of course). The product $\mathbf{MI}$ is then the column matrix with elements

$$\sum_{s,r=1}^{2} (ij|\, M \,|sr) I_{sr}, \quad i, j = 1, 2.$$

We see from Eq. (2.20) that $\mathbf{M}$ will vanish for $|\omega - \omega'| \gg t_c^{-1}$. If $\mathbf{I}$ is nearly constant over a frequency range of order t_c^{-1}, we can rewrite Eq. (2.26) in the form

$$\hat{\mathbf{p}} \cdot \nabla \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}, \omega) + \frac{1}{l(\mathbf{x})} \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}, \omega)$$
$$= \int d\Omega_{\hat{\mathbf{p}}'} \mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}') \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}', \omega). \quad (2.27)$$

Here

$$\mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}') \equiv \int_{-\infty}^{\infty} d\Omega \mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \Omega). \quad (2.28)$$

Alternatively, if the radiation is confined to a sufficiently narrow frequency interval $\Delta\omega$, we can integrate Eq. (2.26) over frequency to obtain Eq. (2.27), as satisfied by the integrated intensity

$$\mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}) \equiv \int_0^\infty d\omega \mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}, \omega). \quad (2.29)$$

It was this equation which was obtained in I.

The fundamental assumption required to derive the classical transport equation (2.26) is that

$$R_c \ll l_t, \quad (2.30a)$$

where (we recall) R_c is the correlation length. When R_c may be taken as k^{-1}, we may rewrite (2.30a) as the condition that

$$\zeta(\omega_p^4/\omega^4) \ll 1, \quad (2.30b)$$

where

$$\zeta = \delta\rho^2/\rho^2, \quad (2.31)$$

with $\delta\rho^2$ the mean-square electron density fluctuation.

In the derivation of Eq. (2.26), it was also assumed that the paths of geometrical optics for rays propagating with the refractive index $n(\mathbf{z})$ could be approximated by straight lines. More generally, Eq. (2.26) must be integrated along curved ray paths.

3. THE POWER SPECTRUM

We consider an electric field variable $E(t)$ defined over the "long" time interval $-\frac{1}{2}T < t < \frac{1}{2}T$ and vanishing outside this interval. In representing a scattered wave, E will depend parameterically on the electron coordinates $\mathbf{z}_1, \cdots, \mathbf{z}_N$ and on any random variables characterizing the source. It will be convenient to use a complex representation for E, so the "power density" is

$$\mathcal{I}_0 = (8\pi)^{-1} \langle E^*(t) E(t) \rangle, \quad (3.1)$$

in a suitable system of units.[14] Here the average $\langle \cdots \rangle$ represents an average over both plasma electron coordinates and over source fluctuations. That is,

$$\langle E^*(t) E(t) \rangle = \int P_N \, d^3 z_1 \cdots d^3 z_N \langle E^*(t) E(t) \rangle_S, \quad (3.2)$$

where $\langle \cdots \rangle_S$ represents an average over source fluctuations only. We extend the assumption (2.8) that we are dealing with a stationary random process to include the source. Thus, for example,

$$\langle E^*(t) E(t) \rangle = \frac{1}{T} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt \langle E^*(t) E(t) \rangle. \quad (3.3)$$

The field $E(t)$ is expressed in terms of its Fourier transform $\hat{E}(\omega)$ as

$$E(t) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} \hat{E}(\omega) e^{-i\omega t} \, d\omega. \quad (3.4)$$

The *power spectrum* of $E(t)$ is then

$$\mathcal{I}(\omega) = \frac{1}{T} \left\langle \frac{|\hat{E}(\omega)|^2}{8\pi} \right\rangle$$
$$= \frac{1}{8\pi} \int \langle E^*(t) E(t + \tau) \rangle \frac{e^{i\omega\tau}}{2\pi} \, d\tau, \quad (3.5)$$

normalized to

$$\int_{-\infty}^{\infty} \mathcal{I}(\omega) \, d\omega = \mathcal{I}_0. \quad (3.6)$$

It should be noted that we are here defining the power spectrum over the interval $-\infty < \omega < +\infty$. We shall see that our transport equation is even in ω, so $\mathbf{I}$ may be defined on the interval $0 < \omega < \infty$.

The incident plane wave emitted by the distant source (see Fig. 1) is assumed, for the present, to be plane polarized[15] and of the form

$$\mathbf{E}_I(\mathbf{r}, t) = \hat{\mathbf{e}}_{\mathbf{k}}(1) E_I(\mathbf{r}, t),$$
$$E_I(\mathbf{r}, t) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int \hat{E}_0(\omega) e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \, d\omega. \quad (3.7)$$

The power spectrum of the incident wave is

$$\mathcal{I}_I(\omega) = T^{-1} \langle |\hat{E}_0(\omega)|^2 / 8\pi \rangle = \mathcal{I}_I(-\omega), \quad (3.8)$$

which must be even in ω if it is to correspond to a physical wave. The incident intensity, representing flow of power per unit area, is then (here c is again the speed of light)

$$I^0(\omega) = 2c\mathcal{J}_I(\omega), \quad 0 < \omega < \infty. \tag{3.9}$$

The total intensity is then

$$I^0 \equiv \int_0^\infty I^0(\omega)\, d\omega. \tag{3.10}$$

4. THE BORN APPROXIMATION

It is instructive to first calculate the scattered power in the Born approximation.[16] The scattered waves at a point $\mathbf{r}$ far from the plasma can then be written in the form [see Eqs. (2.3)]

$$\mathbf{E}_{sc}(\mathbf{r}, t) = \sum_{j=1}^{2} \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j) E_{sc}(j, t),$$

$$E_{sc}(j, t) = \sum_{\alpha=1}^{N} G^0_{r\alpha} f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}) E_I(\mathbf{z}_\alpha, t). \tag{4.1}$$

Here E_I is the incident field (3.7) and $\mathbf{z}_\alpha \equiv \mathbf{z}_\alpha(t_\alpha)$, $\alpha = 1, 2, \cdots, N$, is an electron coordinate evaluated at the retarded time

$$t_\alpha = t - R_\alpha/c, \tag{4.2}$$

where

$$\mathbf{R}_\alpha = \mathbf{r} - \mathbf{z}_\alpha(t_\alpha) \tag{4.3}$$

and $\hat{\mathbf{p}} = \hat{\mathbf{r}}$.

For a plane wave having wave number $k/2\pi$ and angular frequency $\omega = kc$,

$$G^0_{r\alpha} = \frac{e^{ikR_\alpha}}{R_\alpha}, \tag{4.4}$$

and

$$f_{ji}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) = -r_0\left(1 + i\frac{\nu_c}{\omega}\right)^{-1} \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j) \cdot \hat{\mathbf{e}}_{\hat{\mathbf{k}}}(i) \tag{4.5}$$

is the Thomson amplitude (here r_0 is the classical electron radius). Since E_I contains a spectrum of plane waves, we interpret (4.1) as follows:

$$G^0_{r\alpha} f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}) E_I(\mathbf{z}_\alpha, t) = \int \frac{d\omega}{(2\pi)^{\frac{1}{2}}} \frac{e^{ikR_\alpha}}{R_\alpha} f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) \hat{E}_0(\omega) e^{i(\mathbf{k}\cdot\mathbf{z}_\alpha - \omega t)}. \tag{4.6}$$

Here $\bar{k} = \bar{k}(k)$ is the wave number after scattering for an incident wave number k. Now, $|\bar{k} - k| = O(kv_e/c) \ll k$ by assumption NR made at the beginning of Sec. II. We shall interpret this to mean, for example, that

$$f_{ji}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) \cong f_{ji}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \bar{\omega}), \tag{4.7}$$

where $\bar{\omega} = \bar{k}c$.

On setting $t' \equiv t + \tau$, using Eqs. (4.1) and (4.6), and writing $R_\alpha \cong r - \hat{\mathbf{p}} \cdot \mathbf{z}_\alpha$, etc., we obtain

$$\begin{aligned}\langle E^*_{sc}(j, t) E_{sc}(l, t')\rangle &= \frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt \langle E^*_{sc}(j, t) E_{sc}(l, t')\rangle \\ &= \sum_{\alpha,\beta=1}^{N} \frac{1}{r^2} \int_{-\infty}^{\infty} d\omega f^*_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) f_{l1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) \\ &\quad \times \langle T^{-1} |\hat{E}_0(\omega)|^2 \exp(-i\omega\tau) \\ &\quad \times \exp\{i[(\mathbf{z}_\beta(t_\beta) - \mathbf{z}_\alpha(t'_\alpha)) \cdot (\bar{k}\hat{\mathbf{p}} - \mathbf{k})]\}\rangle. \end{aligned} \tag{4.8}$$

According to Eq. (3.2) the average $\langle \cdots \rangle$ here implies the integration

$$\int d^3 z_\alpha\, d^3 z_\beta P_2(\mathbf{z}_\alpha, t'_\alpha; \mathbf{z}_\beta, t_\beta) \cdots. \tag{4.9}$$

From Eq. (2.10) we see that the term *not* involving the pair correlation function does not involve the times t'_α and t_β, which could be taken to be *any* two times in the interval $-\frac{1}{2}T < t < \frac{1}{2}T$. Also, for this term we have $\bar{k} = k$, corresponding to coherent scattering.

For the other term, involving the pair correlation function, we have $|\mathbf{z}_\alpha - \mathbf{z}_\beta| \simeq O(R_c)$. Since the distribution function is stationary, we may set

$$\begin{aligned}\mathbf{z}_\alpha(t'_\alpha) - \mathbf{z}_\beta(t_\beta) &= \mathbf{z}_\alpha(t'_\alpha + r/c) - \mathbf{z}_\beta(t_\beta + r/c) \\ &= \mathbf{z}_\alpha(t') - \mathbf{z}_\beta(t) + O((v_e/c)R_c)\end{aligned} \tag{4.10}$$

within the average in Eq. (4.8). The term of $O((v_e/c)R_c k)$ may be dropped in the exponential, using assumption NR. Also, since $\bar{k} - k \cong (kv_e)/c$, we may use assumption NR to set $\bar{k} = k$ in (4.8). Finally, then, we may write this equation in the form

$$\begin{aligned}\langle E^*_{sc}(j, t) E_{sc}(l, t')\rangle &= \sum_{\alpha,\beta=1}^{N} \frac{1}{r^2} \int_{-\infty}^{\infty} d\omega' f^*_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega') f_{l1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega')[8\pi\mathcal{J}_I(\omega')] \\ &\quad \times e^{-i\omega'\tau} \int P_1(\mathbf{x}) P_1(\mathbf{x}')[1 + g(\mathbf{x}, \mathbf{x}'; \tau)] \\ &\quad \times \exp[i(\mathbf{k}' - \mathbf{p}') \cdot (\mathbf{x} - \mathbf{x}')]\, d^3x\, d^3x', \end{aligned} \tag{4.11}$$

where $\hat{\mathbf{p}}' = k'\hat{\mathbf{r}}$.

The power spectrum of the scattered waves is then

$$\begin{aligned}\mathcal{J}_{jl}(\omega) &= \frac{1}{8\pi}\int \langle E^*_{sc}(j, t) E_{sc}(l, t')\rangle \frac{e^{i\omega\tau}}{2\pi} d\tau \\ &= \frac{1}{r^2} f^*_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) f_{l1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) \int d\omega' \mathcal{J}_I(\omega') \int \frac{d\tau}{2\pi} \\ &\quad \times \int d^3x\, d^3x' \rho(\mathbf{x})\rho(\mathbf{x}')[1 + g(\mathbf{x}, \mathbf{x}'; \tau)] \\ &\quad \times \exp[i(\omega - \omega')\tau] \exp[i(\mathbf{k} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{x}')]. \end{aligned} \tag{4.12}$$

We have here used the relation (4.7) to remove the scattering amplitudes from the ω' integrand.

The *coherent* scattering in Eq. (4.12) is given by the term that does not involve g. This is immediately seen to reduce to

$$\mathfrak{I}_{jl}(\omega)\big|_{\text{coh}} = \frac{1}{r^2} f^*_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) f_{l1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}, \omega) \mathfrak{I}_I(\omega) \times \left| \int d^3x \exp\left[i(\mathbf{k} - \mathbf{p}) \cdot \mathbf{x}\right] \rho(\mathbf{x}) \right|^2, \quad (4.13)$$

where $\mathbf{p} = k\hat{\mathbf{r}}$.

The remaining part of (4.12) represents the *incoherent* scattered power. This may be written in the form

$$\mathfrak{I}_{jl}(\omega)\big|_{\text{inc}} = \frac{1}{r^2} \int d\omega' \, (jl|\, M(\hat{\mathbf{p}}, \hat{\mathbf{k}}; \omega - \omega')\, |11)\, \mathfrak{I}_I(\omega'), \quad (4.14)$$

where $\mathbf{M}$ is defined by Eqs. (2.19) and (2.20) and the refractive index n_1 is replaced by unity.

To derive the transport equation, we must consider a sequence of scatterings, just like the single one just described. In I it was shown that all coherent scatterings result in propagation with the refractive index n. The incoherent scatterings lead to the transport equation.

A sequence of coherent scatterings will not lead to a frequency shift. On the other hand, a long sequence of incoherent scatterings may lead to a large frequency shift in the wave. For each single scattering in such a sequence, we can continue to assume that the frequency shift $(\bar{\omega} - \omega) = O(\omega v_e/c)$ is small, because of the assumption NR. In particular, we can continue to use the relation (4.7), where ω and $\bar{\omega}$ are the respective frequencies *before* and *after* a given single scattering.

5. DERIVATION OF THE TRANSPORT EQUATION

The derivation of the transport equation, as given in I, needs only minor modifications to take account of the frequency shift. In this section we shall, therefore, rely heavily on the development given in I.

Following the discussion given in I, we write a particular component of the scattered electric field vector in the form

$$E(\mathbf{r}) = \sum_{n=1}^{\infty} \sum_{\alpha_1, \alpha_2, \cdots, \alpha_n} Q_n(\mathbf{r}; \mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_n}). \quad (5.1)$$

Here Q_n represents the contribution from a wave multiply scattered by electrons at $\mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_n}$ and the sum is over all electrons and numbers of scatterings.

To find the scattered power, we must evaluate such quantities as

$$\mathfrak{I}_{nm} \equiv \langle Q^*_m(\mathbf{r}; \mathbf{z}_{\beta_1}, \cdots, \mathbf{z}_{\beta_m}) Q_n(\mathbf{r}; \mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_n}) \rangle = \left\langle \int P_{m+n} Q^*_m Q_n \, d^3z_{\beta_1}, \cdots, d^3z_{\alpha_n} \right\rangle_S, \quad (5.2)$$

using the notation of Eq. (3.2). We suppose the probability function P_{m+n} to be decomposed into a cluster expansion of correlated coordinates, as in Eqs. (2.10) and (2.11). For each term of this expansion, each coordinate is a member of a correlated cluster of coordinates. First, a given coordinate $\mathbf{z}_\alpha$ may be *uncorrelated* with another coordinate. If not *uncorrelated*, $\mathbf{z}_\alpha$ is correlated with other members of the set $\mathbf{z}_{\beta_1}, \cdots, \mathbf{z}_{\alpha_n}$ in (5.2).

Let us suppose that $\mathbf{z}_\alpha$ belongs to the correlated cluster set $\mathbf{z}_{\alpha_c}, \cdots, \mathbf{z}_{\alpha_d}$, which consist of only $\mathbf{z}_\alpha$. In this case the integral over $\mathbf{z}_{\alpha_c}, \cdots, \mathbf{z}_{\alpha_d}$ involves only Q_n. This was called a "coherent part" of the average in I. Such "coherent part" averages may clearly be performed on each factor of $E(\mathbf{r})$ before squaring. It was shown in I that the effect of the "coherent part" averages is to give the plasma a refractive index. This result may be taken unchanged for our present analysis.

To see this, we note that the introduction of the time-dependent correlation functions does not modify the expressions obtained in I for the refractive index. This is obvious [because of the stationary property (2.8)] for scatterings which are uncorrelated. Scatterings which are correlated are separated by distances of the order of R_c. During the time R_c/c required for propagation across a correlated cluster, a typical electron will have moved a distance $(R_c v_e)/c$. The resulting change of phase in the exponentials is, therefore, of order

$$(kR_c)v_e/c \ll 1, \quad (5.3)$$

by assumption NR, and can be neglected.

The resulting equations for the multiply scattered waves are [see Eqs. (I.3.31), (I.3.32), (I.3.33)]

$$\mathbf{E}(\mathbf{z}_\alpha, t) = \mathbf{E}_c(\mathbf{z}_\alpha, t) + \sum_{\beta(\neq\alpha)=1}^{N} \sum_{j=1}^{2} \hat{\mathbf{e}}_{\alpha\beta}(j) E_{\alpha\beta}(j, t), \quad (5.4)$$

$$E_{\alpha\beta}(i, t) = G_{\alpha\beta} f_{i1}(\alpha\beta, \beta 0) E_c(\mathbf{z}_\beta) + \sum_{\sigma(\neq\beta)=1}^{N} \sum_{j=1}^{2} G_{\alpha\beta} f_{ij}(\alpha\beta, \beta\sigma) E_{\beta\sigma}(j, t). \quad (5.5)$$

Here

$$\hat{\mathbf{e}}_{\alpha\beta}(j) \equiv \hat{\mathbf{e}}_{\hat{\mathbf{q}}}(j), \quad (5.6)$$

where $\hat{\mathbf{q}}$ is the unit vector parallel to $(\mathbf{z}_\alpha - \mathbf{z}_\beta)$. The quantity $\mathbf{E}_c(\mathbf{z}_\alpha, t)$ represents the *coherent wave* [see

Eqs. (3.7)]

$$\mathbf{E}_c(\mathbf{z}_\alpha, t) = \hat{\mathbf{e}}_k(1)E_c(\mathbf{z}_\alpha, t),$$
$$E_c(\mathbf{z}_\alpha, t) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int d\omega \hat{E}_0(\omega)e^{i(kS_\alpha - \omega t)}, \quad (5.7)$$

where S_α is the eikonal for the coherent wave [Eq. (I.3.34)]

$$S_\alpha \equiv S(\mathbf{z}_\alpha) = \int_{-\infty}^{\mathbf{z}_\alpha}[n(\mathbf{x}) - 1]\,ds + \hat{\mathbf{k}}\cdot\mathbf{z}_\alpha. \quad (5.8)$$

Here $n(\mathbf{x})$ is the refractive index and the constant of integration has been chosen to be consistent with Eq. (3.7). The Thomson amplitude (4.5) for scattering has been rewritten in Eqs. (5.5) to indicate scattering from the direction of $(\mathbf{z}_\beta - \mathbf{z}_\sigma)$ to that of $(\mathbf{z}_\alpha - \mathbf{z}_\beta)$, etc. For a monochromatic wave, the Green's function [see Eq. (I.3.29)] is

$$G_{\alpha\beta} = \frac{e^{ikS_{\alpha\beta}}}{R_{\alpha\beta}}, \quad (5.9)$$

$$S_{\alpha\beta} = \int_{\mathbf{z}_\beta}^{\mathbf{z}_\alpha} n(\mathbf{x})\,ds, \quad (5.10)$$

the integral being taken along the straight line path from $\mathbf{z}_\beta$ to $\mathbf{z}_\alpha$ and

$$\mathbf{R}_{\alpha\beta} \equiv \mathbf{z}_\alpha - \mathbf{z}_\beta. \quad (5.11)$$

We have used the notation of Eq. (4.6) on the right-hand side of Eq. (5.5), writing

$$G_{\alpha\beta}f_{ij}(\alpha\beta, \beta\sigma)E_{\beta\sigma}(j, t) = \int\frac{d\omega}{(2\pi)^{\frac{1}{2}}}\frac{e^{ikS_{\alpha\beta}}}{R_{\alpha\beta}}f_{ij}(\alpha\beta, \beta\sigma; \omega)\hat{E}_{\beta\sigma}(j, \omega)e^{-i\omega t} \quad (5.12)$$

in terms of the Fourier transform $\hat{E}_{\beta\sigma}$ of $E_{\beta\sigma}$.

The argument given, following Eq. (4.6), lets us set $\bar{k} = k$ in the exponential in (5.12).

The "coherent part" averages in the expressions (5.2) permitted us to derive the multiple scattering equations (5.4) and (5.5). On performing the remaining averages, after using Eqs. (5.4) and (5.5), we must omit "coherent part" averages. This means that every coordinate $\mathbf{z}_{\alpha_i}$ must be now correlated with at least one of the $\mathbf{z}_\beta$ in (5.2).

Continuing to follow I, we define [a generalization of Eq. (I.5.6)]

$$U_{ij}(\alpha, \beta; \omega) \equiv \frac{1}{2\pi}\int d\tau e^{+i\omega\tau}\int d^3z_\gamma\, d^3z_\sigma\,\delta[\tfrac{1}{2}(\mathbf{z}_\gamma + \mathbf{z}_\sigma) - \mathbf{z}_\beta] \times \rho(\mathbf{z}_\gamma)\rho(\mathbf{z}_\sigma)g(\mathbf{z}_\gamma, \mathbf{z}_\sigma; \tau)\frac{1}{8\pi}\langle E^*_{\alpha\gamma}(i, t)E_{\alpha\sigma}(j, t + \tau)\rangle_{\alpha\gamma\sigma}. \quad (5.13)$$

Here the notation $\langle\cdots\rangle_{\alpha\gamma\sigma}$ means an average over all coordinates except for $\mathbf{z}_\alpha$, $\mathbf{z}_\gamma$, and $\mathbf{z}_\sigma$, which are held fixed.[17]

We now follow the derivation of Eq. (I.5.11). Equations (5.5) are substituted into the right-hand side of Eq. (5.13). Equations (I.5.9) and (I.5.10) are used to write $E_{\beta'\sigma}$ in terms of $E_{\beta\sigma}$, etc., for $|\mathbf{z}_\beta - \mathbf{z}_{\beta'}| = O(R_c)$. There finally results

$$\begin{aligned}U_{ij}(\alpha, \beta; \omega) &= (2\pi)^{-1}\int d\tau e^{i\omega\tau}\int d^3z_\gamma\, d^3z_{\gamma'}\delta(\tfrac{1}{2}(\mathbf{z}_\gamma + \mathbf{z}_{\gamma'}) - \mathbf{z}_\beta) \\ &\times \rho(\mathbf{z}_\gamma)\rho(\mathbf{z}_{\gamma'})g(\mathbf{z}_\gamma, \mathbf{z}_{\gamma'}; \tau)\,|G_{\alpha\beta}|^2 \\ &\times \Big(f^*_{i1}(\alpha\beta, \beta 0)f_{j1}(\alpha\beta, \beta 0)(8\pi)^{-1} \\ &\times \langle E^*_c(\mathbf{z}_\beta, t)E_c(\mathbf{z}_\beta, t + \tau)\rangle_S \\ &\times \exp[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\alpha\beta} - \mathbf{k})\cdot(\mathbf{R}_{\gamma\beta} - \mathbf{R}_{\gamma'\beta})] \\ &+ \sum_{s,s'=1}^{2}\int d^3z_\sigma\, d^3z_{\sigma'}\rho(\mathbf{z}_\sigma)\rho(\mathbf{z}_{\sigma'})g(\mathbf{z}_\sigma, \mathbf{z}_{\sigma'}; \tau) \\ &\times [f^*_{is}(\alpha\beta, \beta\sigma)f_{js'}(\alpha\beta, \beta\sigma)] \\ &\times \exp[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\alpha\beta} - \mathbf{k}_{\beta\sigma})\cdot\mathbf{R}_{\gamma\gamma'}](8\pi)^{-1} \\ &\times \langle E^*_{\beta\sigma}(s, t)E_{\beta\sigma'}(s', t + \tau)\rangle_{\beta\sigma\sigma'} + \text{cross terms}\Big). \end{aligned} \quad (5.14)$$

To simplify the first term above, we write

$$\begin{aligned}(2\pi)^{-1}\int d\tau e^{+i\omega\tau}(8\pi)^{-1}\langle E^*_c(\mathbf{z}_\beta, t)E_c(\mathbf{z}_\beta, t + \tau)\rangle_S &= [(2c)^{-1}I^0(\omega)]e^{ik(S_\beta - S_{\beta^*})} \\ &= [(2c)^{-1}I_c(\mathbf{z}_\beta, \omega)], \end{aligned} \quad (5.15)$$

where [see Eq. (3.9)],

$$I_c(\mathbf{z}_\beta, \omega) = I^0(\omega)\exp\left(-\int^{\mathbf{z}_\beta}\frac{ds}{l(\mathbf{x})}\right) \quad (5.16)$$

and the integral is taken along the straight line parallel to $\hat{\mathbf{k}}$.

On making use of the assumption that $R_c/l \ll 1$, we may neglect the cross terms in Eq. (5.14) and express the second term in terms of U. In so doing, we make use of our conclusion of Sec. IV that the change in frequency on a single scattering may be neglected in the Thomson amplitude and in the exponentials.

To simplify the second term in Eq. (5.14), we write it in the form

$$\begin{aligned}&\int d\tau\frac{e^{i\omega\tau}}{2\pi}\int d^3z_\gamma\, d^3z_{\gamma'}\delta(\tfrac{1}{2}(\mathbf{z}_\gamma + \mathbf{z}_{\gamma'}) - \mathbf{z}_\beta) \\ &\times \rho(\mathbf{z}_\gamma)\rho(\mathbf{z}_{\gamma'})g(\mathbf{z}_\gamma, \mathbf{z}_{\gamma'}; \tau)\,|G_{\alpha\beta}|^2 \\ &\times \sum_{s,s'=1}^{2}\int d\omega'\frac{e^{-i\omega'\tau}}{2\pi}\int d\tau' e^{i\omega'\tau'} \\ &\times \int d^3z_\sigma\, d^3z_{\sigma'}\int d^3z_\lambda\,\delta(\tfrac{1}{2}(\mathbf{z}_\sigma + \mathbf{z}_{\sigma'}) - \mathbf{z}_\lambda)\rho(\mathbf{z}_\sigma)\rho(\mathbf{z}_{\sigma'}) \\ &\times g(\mathbf{z}_\sigma, \mathbf{z}_{\sigma'}; \tau')[f^*_{is}(\alpha\beta, \beta\sigma)f_{js'}(\alpha\beta, \beta\sigma)] \\ &\times \exp[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\alpha\beta} - \mathbf{k}_{\beta\lambda})\cdot\mathbf{R}_{\gamma\gamma'}] \\ &\times \langle(8\pi)^{-1}E^*_{\beta\sigma}(s, t)E_{\beta\sigma'}(s', t + \tau')\rangle_{\beta\sigma\sigma'} \\ &= \sum_{s,s'=1}^{2}\int d^3z_\lambda\int d\omega'\,|G_{\alpha\beta}|^2\,(ij|\,M(\alpha\beta, \beta\lambda; \omega - \omega')\,|ss') \\ &\times U_{ss'}(\beta, \lambda; \omega'). \end{aligned} \quad (5.17)$$

In writing the exponential involving $\mathbf{R}_{\gamma\gamma'}$, we have replaced $\mathbf{k}_{\beta\sigma}$ in Eq. (5.14) by $\mathbf{k}_{\beta\lambda}$. This is permissible since we assume $l \gg R_c$. The quantity $\mathbf{M}$ in Eq. (5.17) is defined by Eqs. (2.19), with the obvious notational change of indicating directions of propagation as $\beta\lambda$ and $\alpha\beta$.

The results (5.15) and (5.17) permit us to write (5.14) in the form

$$U_{ij}(\alpha, \beta; \omega) = \int d\omega' |G_{\alpha\beta}|^2 \Big((ij|\, M(\alpha\beta, \beta 0; \omega - \omega')\, |11) \frac{1}{2c} I_c(\mathbf{z}_\beta, \omega') + \sum_{s,s'=1}^{2} \int d^3 z_\sigma\, (ij|\, M(\alpha\beta, \beta\sigma; \omega - \omega')\, |ss') \times U_{ss'}(\beta, \sigma; \omega') \Big). \tag{5.18}$$

Since $\mathbf{M}$ is even in $(\omega - \omega')$ [see remark following Eq. (2.20)] and I_c is even in ω', it follows that $\mathbf{U}$ is even in ω. This lets us define the intensity I_{ij} for $\omega > 0$ with the equation

$$I_{ij}(\mathbf{z}_\alpha, \hat{\mathbf{p}}, \omega) = I_c(\mathbf{z}_\alpha, \omega)\delta_{i1}\delta_{j1}\delta_{\mathbf{k},\hat{\mathbf{p}}} + 2c \int_{-\hat{\mathbf{p}}} R_{\alpha\beta}^2\, dR_{\alpha\beta} U_{ij}(\alpha, \beta; \omega). \tag{5.19}$$

The δ function here is defined by the condition that

$$\int d\Omega_{\hat{\mathbf{p}}} f(\hat{\mathbf{p}}) \delta_{\mathbf{k},\hat{\mathbf{p}}} = f(\hat{\mathbf{k}})$$

for a function $f(\hat{\mathbf{p}})$ which is regular at $\hat{\mathbf{k}} = \hat{\mathbf{p}}$. The integration in Eq. (5.19) is performed over $\mathbf{z}_\beta$ along the semi-infinite straight line beginning at $\mathbf{z}_\alpha$ and directed parallel to $-\hat{\mathbf{p}}$.

Using Eq. (5.19), we can express (5.18) in terms of I_{ij}. If we write

$$|G_{\alpha\beta}|^2 = \frac{1}{R_{\alpha\beta}^2} \exp\left(-\int_{\mathbf{z}_\beta}^{\mathbf{z}_\alpha} \frac{ds}{l(\mathbf{x})} \right),$$

we obtain [in the matrix notation of Eq. (2.26)]

$$\mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}, \omega) = \mathbf{I}_c(\mathbf{x}, \omega)\delta_{\hat{\mathbf{p}},\mathbf{k}} + \int_{-\hat{\mathbf{p}}} ds(\mathbf{x}) \exp\left(-\int_{\mathbf{z}}^{\mathbf{x}} \frac{ds'}{l(\mathbf{x}')} \right) \times \int_0^\infty d\omega' \int d\Omega_{\hat{\mathbf{p}}'} \mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}'; \omega - \omega') \mathbf{I}(\mathbf{z}, \hat{\mathbf{p}}', \omega'), \tag{5.20}$$

where now $\omega > 0$ and

$$\mathbf{I}_c(\mathbf{x}, \omega) = \mathbf{I}^0(\omega) \exp\left(-\int^{\mathbf{x}} \frac{ds}{l} \right). \tag{5.21}$$

The path integral in Eq. (5.20) extends along the straight line from $\mathbf{z}$ to ∞ in the direction $-\hat{\mathbf{p}}$.

Differentiation of Eq. (5.20) along a ray path leads to Eq. (2.26). Equation (5.20) is evidently valid for arbitrary incident polarization, to be specified by the choice of $\mathbf{I}^0(\omega)$.[8]

6. RADAR BACKSCATTER

It was pointed out in I that the transport equation is not valid for backscatter. The reason for this is illustrated in Fig. 2. To each ray path defined by a particular sequence of multiple scatterings there corresponds a path obtained by reversing all propagation vectors. These pairs of paths can interfere coherently, and this is not included in the transport equation. As was shown in I, this effect can be accounted for, however, by choosing a certain linear combination of solutions of the transport equation. The specific expression for backscatter was given in Eqs. (I.7.7) and (I.7.10).

When there is a frequency shift, Eqs. (I.7.5a) and (I.7.5b) are modified. These now read, respectively,

$$Q_n(i, s) = \sum_{j_1, \cdots, j_{n-1}} \frac{e^{ik_{n+1}S_{rn}}}{r} \times f_{ij_{n-1}}(-\hat{\mathbf{k}}, \hat{\mathbf{l}}_{n-1}) \cdots f_{j_1 s}(\hat{\mathbf{l}}_1, \hat{\mathbf{k}}) e^{ik_1 S_{1r}}, \tag{6.1a}$$

$$\tilde{Q}_n(i, s) = \sum_{j_1, \cdots, j_{n-1}} \frac{e^{ik_{n+1}S_{r1}}}{r} \times f_{ij_1}(-\hat{\mathbf{k}}, -\hat{\mathbf{l}}_1) \cdots f_{j_{n-1}s}(-\hat{\mathbf{l}}_{n-1}, \hat{\mathbf{k}}) e^{ik_1 S_{nr}}. \tag{6.1b}$$

Here k_1 is the incident wave number, k_2 that after the first scattering, $\cdots$, and k_{n+1} that after the nth scattering.

For Q_n and $\tilde{Q}_n$ to interfere coherently, the frequency (wave-number) spread must be small enough that the phase differences $[S_{ll-1}(k_l - k_{n-l+2})]$ are small compared to unity. The criterion for this is that

$$(\Delta\omega l)/c \ll 1, \tag{6.2}$$

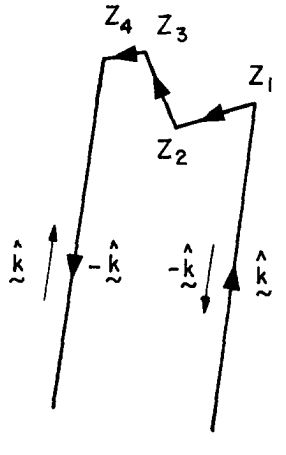


FIG. 2. Illustration of backscattering.

where $\Delta\omega$ is the total spread in frequency because of scattering. If condition (6.2) is satisfied, the *expressions* (I.7.7) and (I.7.10) *may be used.*

On the other hand, when

$$(\Delta\omega l)/c \gg 1, \tag{6.3}$$

Q_n and $\tilde{Q}_n$ will not interfere. Then, the transport equation (2.26) [or (5.20)] does tend to be valid for backscatter, without the special correction of Eqs. (I.7.7) and (I.7.10).

In intermediate cases it is not anticipated that the transport equation will be applicable to the calculation of backscatter.

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

† Supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by Army Office-Durham under Contract No. DAHCO4-67C-0070.

[1] K. M. Watson, J. Math. Phys. **10**, 688 (1969). This paper will hereafter be referred to as I. Equations in I will be referred to as "Eq. (I.3.2)," etc.

[2] K. M. Watson, "Electromagnetic Wave Scattering Within a Plasma in the Transport Approximation," Phys. Fluids (to be published). We shall refer to this paper as II.

[3] L. L. Foldy, Phys. Rev. **67**, 107 (1945).

[4] K. M. Watson, Phys. Rev. **118**, 886 (1960).

[5] Yu. N. Barabanenkov and V. N. Finkel'berg, Zh. Eksp. Teor. Fiz. **53**, 978 (1967) [Soviet Phys.—JETP **26**, 587 (1968)].

[6] See, for example, S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).

[7] Any fixed unit vector $\hat{\mathbf{k}}$ would do.

[8] The representation of polarization is discussed in II in some detail.

[9] As a practical condition, we require that (2.8) be valid for $\tau \lesssim O(\Delta\omega^{-1})$, where $\Delta\omega$ is the smallest frequency interval of interest in the intensity spectrum.

[10] This was described in more detail in I.

[11] See, for example, M. Lax, Rev. Mod. Phys. **32**, 25 (1960), or H. B. Callen, *Thermodynamics* (John Wiley & Sons, New York, 1960).

[12] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).

[13] G. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge University Press, Cambridge, 1953).

[14] Since Eq. (2.26) is homogeneous in the intensity **I**, the actual system of units used is not significant. We are here using unrationalized Gaussian units, as in I.

[15] We shall later extend the result to include arbitrary polarization of the incident wave, as was done in I.

[16] See, for example, H. G. Booker, J. Geophys. Res. **64**, 2164 (1959), and A. D. Wheelon, J. Res. Natl. Bur. Std., D, **63**, 205 (1959).

[17] This notation was described in more detail in I.