A generalization of the Gel'fand–Levitan equation for the one-dimensional Schrödinger equation[a)]

H. E. Moses

Center for Atmospheric Research, University of Lowell, Lowell, Massachusetts 01854
(Received 4 March 1977)

The Gel'fand–Levitan equation for the one-dimensional Schrödinger equation is generalized to the case that the unperturbed Hamiltonian contains part of the scattering potential, this part being denoted by $V_0(x)$, and that the direct scattering problem has been solved for this Hamiltonian. Hence one knows the reflection coefficient $b_0(k)$, the point eigenvalues E_{0i}, and the normalizations of the corresponding eigenfunctions C_{0i}. We are given $b_1(k)$, E_{1i}, C_{1i}, which are the corresponding quantities for full potential $V_1(x) = V_0(x) + \Delta V(x)$. A Gel'fand–Levitan equation is set up in terms of $b_1(k) - b_0(k)$ and the difference in measures for the discrete spectra for V_0 and V_1, respectively, from which ΔV can be found. One may regard the new algorithm as providing a means to modify a known potential to accommodate prescribed changes in the reflection coefficient and changes in the nature of the discrete spectrum. The generalization has applications to the Korteweg–de Vries equation. It is shown that a kind of "superposition" principle exists for solutions in that one can add a function of x and t to one solution and obtain a second solution. This principle can be used to separate the soliton part of a solution from the continuous spectrum part.

1. INTRODUCTION. THE ALGORITHM

In Refs. 1 and 2 the Gel'fand–Levitan equation for the one-dimensional Schrödinger equation was given from which the scattering potential could be obtained from the reflection coefficient, point eigenvalues, and normalizations of the proper eigenfunctions. The unperturbed Hamiltonian was taken as

$$H_0 = -\frac{d^2}{dx^2} \quad (-\infty < x < \infty). \tag{1}$$

For the sake of brevity we shall assume that the reader has familiarity with Refs. 1 and 2.

In the present paper we generalize the algorithm to the case where H_0 includes part of the scattering potential:

$$H_0 = -\frac{d^2}{dx^2} + V_0(x) \tag{2}$$

and we assume that the direct problem for this Hamiltonian has been solved. To be explicit we assume that the (improper) eigenfunctions belonging to the continuous spectrum which satisfy

$$H_0\psi_0(x\,|\,k) \equiv \left[-\frac{d^2}{dx^2} + V_0(x)\right]\psi_0(x\,|\,k) = k^2\psi_0(x\,|\,k), \quad \text{(all real } k\text{)} \tag{3}$$

which satisfy the boundary condition

$$\lim_{x\to-\infty} \psi_0(x\,|\,k) = (2\pi)^{-1/2}\exp(ikx) \tag{4}$$

are known.

One can show

$$\psi_0(x\,|\,-k) = \psi_0^*(x\,|\,k). \tag{5}$$

Another solution of Eq. (3), which is denoted here by $\chi_0(x\,|\,k)$, is required to satisfy the boundary conditions

$$\begin{aligned} &\lim_{x\to-\infty} \chi_0(x\,|\,k) = (2\pi)^{-1/2}[\exp(ikx) + b_0(k)\exp(-ikx)], \\ &\lim_{x\to+\infty} \chi_0(x\,|\,k) = (2\pi)^{-1/2} t_0(k)\exp(ikx). \end{aligned} \tag{6}$$

The quantities $b_0(k)$ and $t_0(k)$ are the reflection (on the left) coefficient and transmission coefficient, respectively. One has

$$b_0(-k) = b_0^*(k), \quad t_0(-k) = t_0^*(k). \tag{6a}$$

Clearly

$$\chi_0(x\,|\,k) = \psi_0(x\,|\,k) + b_0(k)\psi_0(x\,|\,-k). \tag{7}$$

Other quantities which we expect to know for the direct problem for H_0 are the point eigenvalues E_{0i}, the corresponding proper eigenfunctions $\psi_{0i}(x)$, which we require to satisfy the boundary condition

$$\lim_{x\to-\infty} \psi_{0i}(x) = \exp[(-E_{0i})^{1/2}x], \tag{8}$$

and the normalizations C_{0i} which are given by

$$C_{0i} = \int_{-\infty}^{\infty} [\psi_{0i}(x)]^2\,dx. \tag{9}$$

The eigenfunctions ψ_{0i} are real. We also take $E_{0i} < 0$.

On can show, using the methods of Ref. 2, with suitable changes in notation, that the following completeness relation holds:

$$\int_{-\infty}^{\infty} \psi_0(x\,|\,k)\psi_0^*(x'\,|\,k)\,dk + \int_{-\infty}^{\infty} \psi_0(x\,|\,k)b_0^*(k)\psi_0(x'\,|\,k)\,dk + \sum_i (C_{0i})^{-1}\psi_{0i}(x)\psi_{0i}(x') = \delta(x - x'). \tag{10}$$

It is our objective to find the potential $V_1(x)$ in the Hamiltonian

$$H_1 = -\frac{d^2}{dx^2} + V_1(x) \tag{11}$$

[a)] Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-77-3169.

from the appropriate scattering data and discrete spectrum associated with the potential $V_1(x)$.

Let $b_1(k)$ be the reflection coefficient and C_{1i} the normalizations for the proper eigenfunctions associated with the point eigenvalues E_{1i}. The subscript i will in general be used to designate a different number of point eigenvalues than that subscript used for E_{0i}. Furthermore let $\ddot{\psi}_{0i}(x)$ designate the solution of the differential equation

$$H_0\ddot{\psi}_{0i}(x) = E_{1i}\ddot{\psi}_{0i}(x) \tag{12}$$

subject to the boundary condition

$$\lim_{x\to-\infty}\ddot{\psi}_{0i}(x) = \exp[(-E_{1i})^{1/2}x]. \tag{12a}$$

(We assume that the point eigenvalues E_{1i} are negative.) It should be noted that, in general, the functions $\ddot{\psi}_{0i}$ are *not* eigenfunctions of H_0, proper or improper. However, when $E_{0i} = E_{1j}$ (with some ordering of the eigenvalues), then $\ddot{\psi}_{0j} = \psi_{0i}$.

We can now state the generalized form of the Gel'fand—Levitan algorithm: Let us define $\Omega(x\,|\,y)$ by

$$\Omega(x\,|\,y) = \int_{-\infty}^{\infty}\psi_0^*(x\,|\,k)\psi_0^*(y\,|\,k)[b_1(k) - b_0(k)]\,dk + \sum_i \frac{\ddot{\psi}_{0i}(x)\ddot{\psi}_{0i}(y)}{C_{1i}} - \sum_i \frac{\psi_{0i}(x)\psi_{0i}(y)}{C_{0i}}. \tag{13}$$

We then require the kernel $K(x\,|\,y)$ defined for $x \geq y$ to satisfy the Gel'fand—Levitan equation

$$K(x\,|\,y) = -\Omega(x\,|\,y) - \int_{-\infty}^{x}K(x\,|\,z)\Omega(z\,|\,y)\,dz. \tag{14}$$

Then the potential $V_1(x)$ of Eq. (11) is given by

$$V_1(x) = V_0(x) + \Delta V(x), \quad \Delta V(x) = 2\frac{d}{dx}K(x\,|\,x). \tag{15}$$

Furthermore the functions $\psi_1(x\,|\,k)$ defined by

$$\psi_1(x\,|\,k) = \psi_0(x\,|\,k) + \int_{-\infty}^{x}K(x\,|\,y)\psi_0(y\,|\,k)\,dy \tag{16}$$

are eigenfunctions of the Hamiltonian H_1

$$H_1\psi_1(x\,|\,k) = k^2\psi_1(x\,|\,k) \tag{17}$$

and the functions $\psi_{1i}(x)$ defined by

$$\psi_{1i}(x) = \ddot{\psi}_{0i}(x) + \int_{-\infty}^{x}K(x\,|\,y)\ddot{\psi}_{0i}(y)\,dy \tag{18}$$

are also eigenfunctions of H_1:

$$H_1\psi_{1i}(x) = E_{1i}\psi_{1i}(x). \tag{19}$$

Furthermore these eigenfunctions of H_1 satisfy the completeness relation (10) with the subscript 0 replaced by the subscript 1, i.e.,

$$\int_{-\infty}^{\infty}\psi_1(x\,|\,k)\psi_1^*(x'\,|\,k)\,dk + \int_{-\infty}^{\infty}\psi_1(x\,|\,k)b_1^*(k)\psi_1(x'\,|\,k)\,dk + \sum_i (C_{1i})^{-1}\psi_{1i}(x)\psi_{1i}(x') = \delta(x - x'). \tag{20}$$

It follows that the eigenfunctions $\psi_1(x\,|\,k)$ are eigenfunctions of the continuous spectrum of H_1 which satisfy the asymptotic condition analogous to Eq. (8)

$$\lim_{x\to-\infty}\psi_1(x\,|\,k) = (2\pi)^{-1/2}\exp(ikx), \tag{21}$$

and that the functions $\psi_{1i}(x)$ are eigenfunctions of the Hamiltonian H_1 with the eigenvalues E_{1i} and normalizations and asymptotic forms

$$C_{1i} = \int_{-\infty}^{\infty}[\psi_{1i}(x)]^2\,dx, \tag{22}$$

$$\lim_{x\to-\infty}\psi_{1i}(x) = \exp[(-E_i)^{1/2}x], \tag{23}$$

respectively.

The Gel'fand—Levitan equation (13) and (14) for the one-dimensional Schrödinger equation is analogous to the Gel'fand—Levitan equation for the radial Schrödinger equation of Ref. 3 where part of the scattering potential is also included in H_0. In Ref. 4 Newton goes into this matter in detail. The one-dimensional problem is somewhat more complicated in that the continuous spectrum of H and H_0 is degenerate (the degeneracy corresponding to the two possible directions of momentum of a given k^2). The greater complication is reflected in our having to choose $\psi_0(x\,|\,k)$ and $\psi_1(x\,|\,k)$ to satisfy the boundary condition (4) and (21), respectively. Other choices of boundary condition would have led to somewhat more complicated forms of the weight operator.

2. DISCUSSION OF THE ALGORITHM

Before deriving the algorithm, we shall discuss it briefly. One of the principal values of the algorithm is the possibility of choosing H_0 so that its reflection coefficient $b_0(k)$ and discrete spectrum approximate closely the final reflection coefficient $b_1(k)$ and the discrete spectrum of H_1. One may then hope that an expansion of the Gel'fand—Levitan kernel $K(x\,|\,y)$ in terms of $\Omega(x\,|\,y)$ will converge sufficiently rapidly to enable one to approximate the kernel by only a few terms. If the approximation is sufficiently good, perhaps even the first term would be adequate:

$$-K(x\,|\,y) \approx \Omega(x\,|\,y), \tag{24}$$

$$\Delta V(x) \approx -2\frac{d}{dx}\Omega(x\,|\,x). \tag{25}$$

In reconstructing $V_1(x)$ from measurements, for example, one will, in general, not know $b_1(k)$ for all values of k, particularly for very large values of $|k|$. If one knows a potential $V_0(x)$ which from physical reasoning can be expected to have a scattering coefficient $b_0(k)$ which is essentially identical to $b_1(k)$ for large values of $|k|$, the difference $b_1(k) - b_0(k)$, which appears in the Gel'fand—Levitan algorithm, will contribute essentially only in the finite domain of real k.

The new form of the algorithm also enables one to study the effect of a small error in one's knowledge of $b_1(k)$ on the potential $V_1(x)$. The quantity $b_1(k) - b_0(k)$ may be regarded as the error in the reflection coefficient and $\Delta V(x)$ the variation in the potential due to the error.

Let $b_1(k) = b_0(k)$. Then the difference in potentials

$$\Delta V(x) = V_1(x) - V_0(x) \tag{26}$$

is due entirely to the difference in the characters of

the discrete spectra for H_1 and H_0. This fact, of course, can be obtained from the original form of the Gel'fand–Levitan algorithm for the one-dimensional case. However, the present algorithm enables one to obtain $\Delta V(x)$ essentially in closed form, the form being analogous to that for the reflectionless potentials of Ref. 5. The closed form will involve ψ_{0i}, C_{0i}, $\ddot{\psi}_{0i}$, C_{1i}. As a special case, for example, one can determine the effect of changing the normalization of only one bound state.

Last, but not least, the general form of the algorithm reduces to the original form when $V_0(x) \equiv 0$.

The form of the generalized Gel'fand–Levitan equation has an interesting "group" or "path" property which may, in principle at least, make it easier to find the potential V_1 from $b_1(k)$ and the data on the discrete spectrum. To illustrate this property, let us assume that we are given the reflection coefficient $b_2(k)$ and data on the discrete spectrum C_{2i}, E_{2i}, $\ddot{\psi}_{0i}$ corresponding to E_{2i}. We can find the potential $V_2(x)$ in two ways. We can use Eqs. (13) and (14) directly and substitute quantities with the subscript 2 for those with the subscript 1. We shall then calculate $\Delta V(x) = V_2(x) - V_0(x)$. However, we may also calculate $V_2(x)$ in another way. We first calculate $V_1(x)$ from any reflection coefficient $b_1(k)$ and discrete spectrum data C_{1i}, E_{1i} and $\ddot{\psi}_{0i}$. We can then also calculate $\psi_1(\psi|k)$, $\psi_{1i}(x)$ from the Gel'fand–Levitan kernel $K(x|y)$. One now regards the zero-order Hamiltonian as being $-d^2/dx^2 + V_1(x)$. One can then calculate $\Delta V(x) = V_2(x) - V_1(x)$ from the generalized Gel'fand–Levitan equation by treating quantities with the subscript 1 as we previously treated quantities with the subscript 0. We then have two "paths" to reach $V_2(x)$ from $V_0(x)$ knowing $b_2(k)$ and corresponding discrete spectrum data. One path is the direct path : $0 \to 2$. The second is the indirect path : $0 \to 1 \to 2$, where the spectral quantities describing the "state" with subscript 1 are arbitrary. One might, in principle, prefer the indirect path because the quantities associated with the subscript 1 might be midway between quantities with the subscript zero and subscript 2. Each step in the path would represent a smaller change which might be significant in a numerical or analytical calculation.

One can generalize the above process and reach the final potential V_2 through a sequence of many steps in which each step differs from the previous one by a small amount as measured in terms of the reflection coefficient and discrete spectrum data. Each choice of steps between an initial "state" and a final "state" is called a "path." The final potential depends only upon the reflection coefficient and discrete spectrum data for the final potential and is thus independent of the "path." One can thus choose a "path" which is most convenient for calculations. One possible example for a path is to consider an initial state with reflection coefficient and discrete spectrum data given by $b_0(k)$, E_{0i}, C_{0i} where $i = 1, 2$. The final state data are $b_f(k)$, E_{f1}, C_{f1} in which only one bound state is assumed. As the first step in the path we take $b_1(k) = b_0(k)$, $E_{11} = E_{01}$, $C_{11} = C_{01}$. The Gel'fand–Levitan equation is now a very simple one because $\Omega(x|x')$ is simple; in fact, it can be solved in closed form. As a second step in the path we take $b_2(k) = b_1(k)$, $E_{21} = E_{f1}$, $C_{21} = C_{f1}$. Again a simple Gel'fand–Levitan equation results. Finally we use $b_f(k)$, E_{f1}, C_{f1} for the final Gel'fand–Levitan equation to find $\Delta V(x) = V_f(x) - V_2(x)$. Clearly, the interval between any two steps could also be broken up into smaller steps. Hence a subpath could be formed between any two steps of a larger path. For example, it might be useful to consider b_f to consist of a sum of functions of k. Each term in the sum could then be taken care of in a step.

The notion of "path" is treated from a more abstract point of view in the Appendix in terms of the operators U, U_0 of Ref. 2.

3. DERIVATION OF THE ALGORITHM

The derivation of the algorithm follows directly from the general form of the Gel'fand–Levitan algorithm of Parts I and II of Ref. 2 and a modification of Part III of Ref. 2 as applied to the one-dimensional problem. We shall sketch the derivation by modifying the derivation of the algorithm as given in Part III.

First of all instead of defining H_0^x and $\langle x|H_0, A_0; E, a\rangle$ by Eq. (2.1) and (2.2) of Part III of Ref. 2 we use

$$H_0^x = -\frac{d^2}{dx^2} + V_0(x), \tag{27}$$

and define $|H_0, A_0; E, a\rangle$ to be a continuous spectrum eigenstate of H_0 defined by the asymptotic condition

$$\lim_{x\to-\infty}\langle x|H_0, A_0; E, a\rangle = (E)^{-1/4}\,2^{-1}(\pi)^{-1/2}\exp[ia(E)^{1/2}x]. \tag{28}$$

If H_0 has point eigenvalues E_{0i}, we require the corresponding eigenfunctions $|H_0; E_{0i}\rangle$ to satisfy the asymptotic condition

$$\lim_{x\to-\infty}\langle x|H_0; E_{0i}\rangle = \exp[(-E_{0i})^{1/2}x]. \tag{29}$$

The normalizations of these eigenfunctions are then determined:

$$C_{0i} = \int_{-\infty}^{\infty}[\langle x|H_0; E_{0i}\rangle]^2\,dx. \tag{30}$$

[We assume $E_{0i} < 0$. The asymptotic condition (29) assures one that the eigenfunctions $\langle x|H_0; E_{0i}\rangle$ are real.]

The completeness relation (2.4) of Part III of Ref. 2 is replaced by

$$\sum_{a,a'}\int_0^{\infty} dE\,|H_0, A_0; E, a\rangle\langle a|\omega_{0c}(E)|a'\rangle\langle H_0, A_0; E, a'| + \sum_i (C_{0i})^{-1}|H_0; E_{0i}\rangle\langle H_0; E_{0i}| = \eta(H_0), \tag{31}$$

where the "matrix" $\langle a|\omega_{0c}(E)|a'\rangle$ makes its appearance because of the boundary condition (28) in the same manner as it appears in Sec. 3 of Part III of Ref. 2. In fact, in analogy to Eq. (3.16) of Part III of Ref. 2

$$\langle a|\omega_{0c}(E)|a'\rangle = \begin{pmatrix} 1 & b_0^*(k) \\ b_0(k) & 1 \end{pmatrix}, \tag{32}$$

where $k = (E)^{1/2}$. In Eq. (31) $\eta(H_0)$ is the identity operator in the Hilbert space.

In Sec. 3 of Part III of Ref. 2 we replace H by H_1 with

$$H_1^x = H_0^x + \Delta V(x) = -\frac{d^2}{dx^2} + V_1(x), \quad V_1(x) = V_0(x) + \Delta V(x). \tag{33}$$

The representation of the weight operator in the H_0 representation is the same as in Eq. (32) but with the subscript zero replaced by the subscript one [see Eq. (3.16)]. Thus

$$W_c = \sum_{a,a'} \int_0^\infty dE |H_0, A_0; E, a\rangle \times \langle a|\omega_{1c}(E)|a'\rangle \langle H_0, A_0; E, a'|. \tag{32a}$$

This form of the weight operator for the continuous spectrum follows from

$$\langle x|H_1, A_1; E, a\rangle = \langle x|H_0, A_0; E, a\rangle + \int_{-\infty}^x \langle x|K|x^\wedge\rangle dx' \langle x'|H_0, A_0; E, a\rangle \tag{34}$$

for the continuous spectrum eigenfunctions of H_1 so that we have the asymptotic condition

$$\lim_{x\to-\infty} \langle x|H_1, A_1; E, a\rangle = \lim_{x\to-\infty} \langle x|H_0, A_0; E, a\rangle = (E)^{-1/4} 2^{-1} (\pi)^{-1/2} \exp[ia(E)^{1/2}x]. \tag{35}$$

This asymptotic condition is that which leads to Eq. (3.16) of Part III of Ref. 2. Thus the continuous contribution to the weight operator is essentially the same as before.

To get the discrete portion of the weight operator we select the point eigenvalues of H_1, which we denote by E_{1i} and the normalizations of the corresponding eigenfunctions C_{1i}. The analogues of $\langle x|H_0; E\rangle$ of Eq. (5.5) of Part III of Ref. 2 are solutions of $H_0^x\langle x|H_0; E\rangle = E\langle x|H_0; E\rangle$ subject to the boundary condition

$$\lim_{x\to-\infty} \langle x|H_0; E\rangle = \exp[(-E)^{1/2}x], \quad E < 0. \tag{36}$$

Thus the discrete portion of the weight operator is

$$\langle x|W_d|x^\wedge\rangle = \sum_i (C_{1i})^{-1} \langle x|H_0; E_{1i}\rangle \langle H_0; E_{1i}|x^\wedge\rangle. \tag{37}$$

The eigenfunctions of H_1 corresponding to the eigenvalues E_{1i} are given by

$$\langle x|H_1; E_{1i}\rangle = \langle x|H_0; E_{1i}\rangle + \int_{-\infty}^x \langle x|K|x^\wedge\rangle dx' \langle x'|H_0; E_{1i}\rangle. \tag{38}$$

The operator Ω is then given by Eq. (6.2) of Part III of Ref. 2 where $\eta(H_0)$ is given by Eq. (31) and where E_i is replaced by E_{1i}.

We now set $\epsilon = 1$ and make the notational changes

$$\langle x|K|x^\wedge\rangle = K(x|x'), \quad \langle x|\Omega|x^\wedge\rangle = \Omega(x|x'). \tag{39}$$

Also

$$\psi_0(x|k) = (2|k|)^{1/2}\langle x|H_0, A_0; E, a\rangle,$$
$$\psi_1(x|k) = (2|k|)^{1/2}\langle x|H_1, A_1; E, a\rangle. \tag{40}$$

In Eq. (40) $k = a(E)^{1/2}$.

Furthermore, for the discrete spectrum

$$\psi_{0i}(x) = \langle x|H_0; E_{0i}\rangle, \quad \ddot{\psi}_{0i}(x) = \langle x|H_0; E_{1i}\rangle,$$
$$\psi_{1i}(x) = \langle x|H_1; E_{1i}\rangle. \tag{41}$$

In particular, it might be noted that Eq. (31) is the same as the completeness relation Eq. (10) which is, perhaps, the principal difference between the generalization of the present paper and Part III of Ref. 2.

It should also be noted that the proof in Part III of Ref. 2 in which it is shown that the potential V is diagonal in the x representation goes into the proof that ΔV is diagonal in the x representation. Furthermore, it is given by Eq. (15).

4. A SIMPLE VERIFICATION OF THE ALGORITHM

The algorithm has been derived in the previous section. The verification of the algorithm is automatic because it falls into the abstract formalism of Parts I and II of Ref. 2.

However, we shall now give a verification of the algorithm which is independent of the abstract formalism of Ref. 2 and is closer in spirit to the verification of Ref. 3 for the radial Schrödinger equation, though there are differences.

We shall assume the Gel'fand—Levitan equation (14) has a unique solution and show that $\psi_1(x|k)$ and $\psi_{1i}(x)$ of Eqs. (16) and (18), respectively, are eigenfunctions of H_1 with eigenvalues k^2 and E_{1i}. We shall also show that they satisfy the completeness relation (20).

It will be convenient to write

$$H_0^x = -\frac{d^2}{dx^2} + V_0(x), \quad H_0^y = -\frac{d^2}{dy^2} + V_0(y) \text{ and so on.} \tag{42}$$

We note

$$H_0^x \Omega(x|y) = H_0^y \Omega(x|y). \tag{43}$$

Now from Eq. (16)

$$H_0^x\psi_1(x|k) = H_0^x\psi_0(x|k) + H_0^x\int_{-\infty}^x K(x|y)\psi_0(y|k)\,dy$$
$$= k^2\psi_0(x|k) - \frac{d}{dx}K(x|x)\cdot\psi_0(x|k) - K(x|x)\frac{d}{dx}\psi_0(x|k) - \frac{\partial}{\partial x}K(x|x)\cdot\psi_0(x|k) - \int_{-\infty}^x \frac{\partial^2}{\partial x^2}K(x|y)\psi_0(y|k)\,dy + V_0(x)\int_{-\infty}^x K(x|y)\psi_0(y|k)\,dy. \tag{44}$$

But also

$$k^2\psi_1(x|k) = k^2\psi_0(x|k) + \int_{-\infty}^x K(x|y)H_0^y\psi_0(y|k)\,dy. \tag{45}$$

On integrating by parts twice the derivatives that occur in H_0^y.

$$k^2\psi_1(x\,|\,k) = k^2\psi_0(x\,|\,k) - K(x\,|\,x)\frac{d}{dx}\psi_0(x\,|\,k) + \frac{\partial}{\partial y}K(x\,|\,x)\cdot\psi_0(x\,|\,k) + \int_{-\infty}^{x} H_0^y K(x\,|\,y)\cdot\psi_0(y\,|\,k)\,dy. \tag{46}$$

Thus from Eq. (44) and (46)

$$(H_0^x - k^2)\psi_1(x\,|\,k) = -2\frac{d}{dx}K(x\,|\,x)\cdot\psi_0(x\,|\,k) + \int_{-\infty}^{x}(H_0^x - H_0^y)K(x\,|\,y)\cdot\psi_0(y\,|\,k)\,dy, \tag{47}$$

where we have used

$$\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\;K(x\,|\,x) = \frac{d}{dx}K(x\,|\,x).$$

We shall now show that

$$(H_0^x - H_0^y)K(x\,|\,y) = -2\frac{d}{dx}K(x\,|\,x)\cdot K(x\,|\,y). \tag{48}$$

On Substituting Eq. (48) into Eq. (47) it will follow that

$$H_1^x\psi_1(x\,|\,k) = k^2\psi_1(x\,|\,k)$$

if

$$\Delta V(x) = 2\frac{d}{dx}K(x\,|\,x).$$

That $\psi_{1i}(x)$ satisfies $H_1^x\psi_{1i}(x) = E_{1i}\psi_{1i}(x)$ is proved similarly.

On using the Gel'fand–Levitan Equation (14) and Eq. (43)

$$(H_0^x - H_0^y)K(x\,|\,y) = -H_0^x\int_{-\infty}^{x}K(x\,|\,z)\Omega(z\,|\,y)\,dz + \int_{-\infty}^{x}K(x\,|\,z)H_0^z\Omega(z\,|\,y)\,dz. \tag{49}$$

Integrating by parts the second derivative in the second integral on the right and carrying out the operation H_0^x on the first integral,

$$(H_0^x - H_0^y)K(x\,|\,y) = \Delta V(x)\Omega(x\,|\,y) - \int_{-\infty}^{x}(H_0^x - H_0^z)K(x\,|\,z)\cdot\Omega(z\,|\,y)\,dz. \tag{50}$$

Let us write

$$(H_0^x - H_0^y)K(x\,|\,y) = -\Delta V(x)L(x\,|\,y) \tag{51}$$

as we may always do and substitute into Eq. (50). It is seen that $L(x\,|\,y)$ satisfies the same Gel'fand–Levitan equation as $K(x\,|\,y)$. Since we are assuming that the solution is unique, it follows that $L(x\,|\,y) = K(x\,|\,y)$ which completes the proof of Eq. (48).

We shall now prove the completeness relation (20). In the Gel'fand–Levitan equation (14) we use Eq. (13) for $\Omega(z\,|\,y)$ and obtain

$$K(x\,|\,y) = -\Omega(x\,|\,y) - \int_{-\infty}^{\infty}dk[b_1(k) - b_0(k)]\psi_0^*(y\,|\,k) \times\int_{-\infty}^{x}K(x\,|\,z)\psi_0^*(z\,|\,k)\,dz - \sum_i (C_{1i})^{-1}\ddot{\psi}_{0i}(y) \times\int_{-\infty}^{x}K(x\,|\,z)\,\ddot{\psi}_{0i}(z)\,dz + \sum_i (C_{0i})^{-1}\psi_{0i}(y) \times\int_{-\infty}^{x}K(x\,|\,z)\psi_{0i}(z)\,dz. \tag{52}$$

On using Eq. (16) and (18) in Eq. (52) and on defining $\ddot{\psi}_{1i}(x)$ by

$$\ddot{\psi}_{1i}(x) = \psi_{0i}(x) + \int_{-\infty}^{x}K(x\,|\,y)\psi_{0i}(y)\,dy \tag{53}$$

we have

$$K(x\,|\,y) = -\int_{-\infty}^{\infty}\psi_1^*(x\,|\,k)[b_1(k) - b_0(k)]\psi_0^*(y\,|\,k)\,dk - \sum_i (C_{1i})^{-1}\psi_{1i}(x)\ddot{\psi}_{0i}(y) + \sum_i (C_{0i})^{-1}\ddot{\psi}_{1i}(x)\psi_{0i}(y). \tag{54}$$

In Eq. (54) we now use Eq. (16), (18) and (53) to replace $\psi_0^*(y\,|\,k)$, $\ddot{\psi}_{0i}(y)$ and $\psi_{0i}(y)$ respectively. On isolating the terms which contain the Gel'fand–Levitan kernel we have

$$K(x\,|\,y) = -\int_{-\infty}^{\infty}\psi_1^*(x\,|\,k)[b_1(k) - b_0(k)]\psi_1^*(y\,|\,k)\,dk - \sum_i (C_{1i})^{-1}\psi_{1i}(x)\psi_{1i}(y) + \sum_i (C_{0i})^{-1}\ddot{\psi}_{1i}(x)\ddot{\psi}_{1i}(y) + \int_{-\infty}^{y}K(y\,|\,z)\,dz\Big\{\int_{-\infty}^{\infty}\psi_1^*(x\,|\,k)[b_1(k) - b_0(k)]\psi_0^*(z\,|\,k)\,dk + \sum_i (C_{1i})^{-1}\psi_{1i}(x)\ddot{\psi}_{0i}(z) - \sum_i (C_{0i})^{-1}\ddot{\psi}_{1i}(x)\psi_{0i}(z)\Big\}. \tag{55}$$

But from Eq. (54) the quantity within the curly brackets is simply $-K(x\,|\,z)$. Hence for $x > y$

$$\int_{-\infty}^{\infty}\psi_1^*(x\,|\,k)[b_1(k) - b_0(k)]\psi_1^*(y\,|\,k)\,dk + \sum_i (C_{1i})^{-1}\psi_{1i}(x)\psi_{1i}(y) - \sum_i (C_{0i})^{-1}\ddot{\psi}_{1i}(x)\ddot{\psi}_{1i}(y) = -K(x\,|\,y) - \int_{-\infty}^{y}K(x\,|\,z)K(y\,|\,z)\,dz. \tag{56}$$

Let us define $K(x\,|\,y) = 0$ for $y > x$. We can interchange x and y in Eq. (56). The left hand side is unaltered. Hence we may write for all x and y

$$\int_{-\infty}^{\infty}\psi_1^*(x\,|\,k)[b_1(k) - b_0(k)]\psi_1^*(y\,|\,k)\,dk + \sum_i (C_{1i})^{-1}\psi_{1i}(x)\psi_{1i}(y) - \sum_i (C_{0i})^{-1}\ddot{\psi}_{1i}(x)\ddot{\psi}_{1i}(y) = -K(x\,|\,y) - K(y\,|\,x) - \eta(x-y)\int_{-\infty}^{y}K(x\,|\,z)K(y\,|\,z)\,dz - \eta(y-x)\int_{-\infty}^{x}K(x\,|\,z)K(y\,|\,z)\,dz, \tag{57}$$

where $\eta(x)$ is the Heaviside function: $\eta(x) = 1$, for $x > 0$, $\eta(x) = 0$ for $x < 0$. We note continuity for $x = y$.

However, on using Eq. (16), (18), and (53) and then (10)

$$\int_{-\infty}^{\infty}\psi_1(x|k)\psi_1^*(y|k)\,dk+\int_{-\infty}^{\infty}\psi_1^*(x|k)b_0(k)\psi_1^*(y|k)\,dk+\sum_i(C_{0i})^{-1}\ddot{\psi}_{1i}(x)\ddot{\psi}_{1i}(y)$$

$$=\int_{-\infty}^{\infty}\psi_0(x|k)\psi_0^*(y|k)\,dk+\int_{-\infty}^{\infty}\psi_0^*(x|k)b_0(k)\psi_0^*(y|k)\,dk+\sum_i(C_{0i})^{-1}\psi_{0i}(x)\psi_{0i}(y)+\int_{-\infty}^{x}K(x|z)\,dz\Big\{\int_{-\infty}^{\infty}\psi_0(z|k)\psi_0^*(y|k)\,dk$$

$$+\int_{-\infty}^{\infty}\psi_0^*(z|k)b_0(k)\psi_0^*(y|k)\,dk+\sum_i(C_{0i})^{-1}\psi_{0i}(z)\psi_{0i}(y)\Big\}+\int_{-\infty}^{y}K(y|z)\,dz\Big\{\int_{-\infty}^{\infty}\psi_0(x|k)\psi_0^*(z|k)\,dk$$

$$+\int_{-\infty}^{\infty}\psi_0^*(x|k)b_0(k)\psi_0^*(z|k)\,dk+\sum_i(C_{0i})^{-1}\psi_{0i}(x)\psi_{0i}(z)\Big\}+\int_{-\infty}^{x}K(x|z)\,dz\int_{-\infty}^{y}K(y|u)\,du\Big\{\int_{-\infty}^{\infty}\psi_0(z|k)\psi_0^*(u|k)\,dk$$

$$+\int_{-\infty}^{\infty}\psi_0^*(z|k)b_0(k)\psi_0^*(u|k)\,dk+\sum_i(C_{0i})^{-1}\psi_{0i}(z)\psi_{0i}(u)\Big\}$$

$$=\delta(x-y)+\int_{-\infty}^{x}K(x|z)\delta(z-y)\,dz+\int_{-\infty}^{y}K(y|z)\delta(z-x)\,dz+\int_{-\infty}^{x}K(x|z)\,dz\int_{-\infty}^{y}K(y|u)\,du\delta(u-z)$$

$$=\delta(x-y)+K(x|y)+K(y|x)+\eta(x-y)\int_{-\infty}^{y}K(x|z)K(y|z)\,dz+\eta(y-x)\int_{-\infty}^{x}K(x|z)K(y|z)\,dz. \tag{58}$$

On adding Eq. (57) and (58) we obtain Eq. (20).

5. APPLICATION TO THE KORTEWEG-de VRIES EQUATION

Let us consider the potential V_0 of Eq. (2). In Ref. 6 it is shown that if we include a parameter t in V_0, the reflection coefficient b_0, and the normalization C_{0i} such that

$$b_0(k;\,t)=b_0(k)\exp(-i8k^3t),$$
$$C_{0i}(t)=C_{0i}\exp[8(-E_{0i})^{3/2}t] \tag{59}$$

then $V_0(x;\,t)$ satisfies the Korteweg–deVries equation

$$\frac{\partial}{\partial t}V_0(x;t)-6V_0(x;\,t)\frac{\partial}{\partial x}V_0(x;\,t)+\frac{\partial^3}{\partial x^3}V_0(x;\,t)=0. \tag{60}$$

In Ref. 6 the initial value problem for Eq. (60) is solved from $V_0(x;\,0)$ by finding $b_0(k)$, C_{0i}, and E_{0i} from the direct problem of scattering. Then $b_0(k;\,t)$, $C_{0i}(t)$, E_{0i} are used in the original version of the inverse problem for the one-dimensional Schrödinger equation to find $V_0(x;t)$.

Let us now assume that we have a second set of data consisting of a reflection coefficient $b_1(x;\,t)$, point eigenvalues E_{1i}, and normalizations $C_{1i}(t)$ such that b_1 and C_{1i} have the same time dependence as in Eq. (59) with the subscript 0 replaced by the subscript 1. Then on defining $\Omega(x|x';\,t)$ by Eq. (13) in which $b_0(k)$, $b_1(k)$, C_{0i}, C_{1i}, $\psi_0(x|k)$, $\psi_{0i}(x)$, $\ddot{\psi}_{0i}(x)$ are replaced by $b_0(k;\,t)$, $b_1(k;\,t)$, $C_{0i}(t)$, $C_{1i}(t)$, $\psi_0(x|k;\,t)$, $\psi_{0i}(x;\,t)$, $\ddot{\psi}_{0i}(x;\,t)$ respectively, where $\psi_0(x|k;\,t)$, $\psi_{0i}(x;\,t)$, $\ddot{\psi}_{0i}(x;\,t)$ satisfy Eq. (3) with boundary condition (4), the corresponding equation for the point eigenvalue E_{0i} with the boundary condition (8), and Eq. (12) with the boundary condition (12a), resepctively, but with $V_0(x)$ replaced by $V_0(x;\,t)$. It should be mentioned that if $V_0(x;\,t)$ is obtained from the inverse method using the original form of the Gel'fand–Levitan equation for the one-dimensional Schrödinger equation, $\psi_0(x|k;\,t)$, $\psi_{0i}(x;\,t)$, $\ddot{\psi}_{0i}(x;\,t)$ can be obtained from the use of the Gel'fand–Levitan kernel acting on $(2\pi)^{1/2}\exp(ikx)$, $\exp[(-E_{0i})^{1/2}x]$, and $\exp[(-E_{1i})^{1/2}x]$ respectively in the usual way. The kernel will, of course, be time-dependent.

On solving the Gel'fand–Levitan equation with $\Omega(x|x';\,t)$, we obtain $\Delta V(x;\,t)$ from the second of Eq. (15). Then

$$V_1(x;\,t)=V_0(x;\,t)+\Delta V(x;\,t) \tag{61}$$

is also a solution of the Korteweg–deVries equation. Eq. (61) is a kind of superposition principle, though, of course, not a real one, since $\Delta V(x;\,t)$ does *not* satisfy the Korteweg–deVries equation. However, the procedure used to get to $V_1(x;\,t)$ from $b_1(k)$, E_{1i}, C_{1i} [or, equivalently from $V_1(x;\,0)$] has some advantages as described earlier in our discussion on the use of paths. We shall not go into detail, except to discuss the separation of the soliton part of the solution in $V_1(x;\,t)$ from the continuous spectrum part.

Let us choose $V_0(x;\,0)$ to be such that $b_0(k)=0$, $E_{0i}=E_{1i}$, $C_{0i}=C_{1i}$. Then $V_0(x;\,t)$ can be solved exactly as a sum of solitons, using the inverse scattering method. Furthermore, $\psi_0(x|k;\,t)$ can also be found exactly. Then $\Omega(x|x';t)$ contains only the continuous part of the data for $V_1(x;0)$: i.e., it depends only on $b_1(k)$. $V_1(x;t)$ is thus represented as the sum of the pure soliton solution V_0 and a continuous contribution ΔV.

Another application of the generalized inverse method to the Korteweg–deVries equation is to find the form of the solution for small times t. Instead of requiring V_0 to be a solution of the Korteweg–deVries equation, we take it to be time-dependent. In fact,

$$V_0(x)=V_1(x;0). \tag{62}$$

Then

$$b_0(k)=b_1(k),\quad E_{0i}=E_{1i},\quad C_{0i}=C_{1i}. \tag{63}$$

Also

$$\Omega(x|x';t)=\int_{-\infty}^{\infty}[\exp(-i8k^3t)-1]b_0(k)\psi_0^*(x|k)\psi_0^*(x'|k)\,dk$$
$$+\sum_i\{\exp[-8(-E_i)^{3/2}t]-1\}\frac{\psi_{0i}(x)\psi_{0i}(x')}{C_{0i}}. \tag{64}$$

For very small t, the time-dependent factors become proportional to t, as does $\Omega(x|x';t)$. Thus for sufficient-

ly small t, $-K(x|y) \approx \Omega(x|y;\, t)$ and

$$V_1(x;\, t) = V_0(x) + W(x)t, \tag{65}$$

where

$$W(x) = 16\frac{d}{dx}F(x),$$
$$F(x) = i\int_{-\infty}^{\infty} k^3 b_0(k)[\psi_0^*(x|k)]^2\, dk + \sum_i (-E_i)^{3/2}\frac{[\psi_{0i}(x)]^2}{C_{0i}}. \tag{66}$$

APPENDIX. THE CONCEPT OF PATHS IN TERMS OF THE ABSTRACT GEL'FAND-LEVITAN EQUATION

In Ref. 2 the Gel'fand—Levitan equation was expressed in terms of operators U, U_0 and the weight operator W.

$$UW = U_0^\dagger, \quad U_0 = U^{-1}, \quad U = I + K, \quad U_0 = I + K_0, \tag{A1}$$

where K and K_0 are required to satisfy the triangularity conditions

$$\langle x|K|x'\rangle = \langle x|K_0|x'\rangle = 0 \quad \text{for } x' > x. \tag{A1a}$$

We shall now adapt this more abstract formalism to the path problem. Let us define H_j^x by

$$H_j^x = -\frac{d^2}{dx^2} + V_j(x), \quad (j = 0, 1, 2) \tag{A2}$$

where $V_j(x)$ are scattering potentials. For simplicity we shall assume that the potentials V_j support no bound states. The eigenfunctions of the continuous spectrum will be denoted by $|H_j, A_j;\, E, a\rangle$. They are required to satisfy the boundary condition Eq. (28) where the subscript 0 is replaced by the subscript j.

For the purposes of discussing paths going from the potential V_0 to V_2, directly and by means of an intermediate potential V_1, it is convenient to write three Gel'fand—Levitan equations in the form Eq. (A1). Toward this end we introduce the operators $U_{(jl)}$, $U_{(jl)0}$ and $W_{(jl)}$ corresponding to U, U_0, and W of Eq. (A1).

The operators $U_{(jl)}$ map eigenfunctions of H_l to those of H_j:

$$|H_j, A_j;\, E, a\rangle = U_{(jl)}|H_l, A_l;\, E, a\rangle,$$
$$|H_l, A_l;\, E, a\rangle = U_{(jl)0}|H_j, A_j;\, E, a\rangle. \tag{A3}$$

The operators $W_{(jl)}$ are defined by

$$W_{(jl)} = \sum_{a,a'}\int_0^\infty dE\,|H_l, A_l;\, E, a\rangle\langle a|\,\omega_{jc}(E)\,|a'\rangle\langle H_l, A_l;\, E,\, a', \tag{A4}$$

where $\langle a|\omega_{jc}(E)|a'\rangle$ is given by Eq. (32) with the subscript 0 replaced by the subscript j.

Now the Gel'fand—Levitan equation which takes one from H_0 to H_1 and H_1 to H_2 are, respectively,

$$U_{(10)}W_{(10)} = U_{(10)0}^\dagger, \quad U_{(21)}W_{(21)} = U_{(21)0}^\dagger. \tag{A5}$$

The Gel'fand—Levitan equation for the direct path, i.e., from H_0 to H_2, is

$$U_{(20)}W_{(20)} = U_{(20)0}^\dagger. \tag{A6}$$

On using the Eq. (A4), and (A3), the second of Eq. (A5) becomes

$$U_{(21)}\sum_{a,a'}\int_0^\infty dE\,|H_1, A_1;\, E, a\rangle\langle a|\,\omega_{2c}(E)\,|a'\rangle\langle H_1, A_1;\, E, a'|$$
$$= U_{(21)}U_{(10)}\sum_{a,a'}\int_0^\infty dE\,|H_0, A_0;\, E, a\rangle$$
$$\times\langle a|\,\omega_{2c}(E)\,|a'\rangle\langle H_0, A_0;\, E, a'|\,U_{(10)}^\dagger = U_{(21)0}^\dagger, \tag{A7}$$

or

$$U_{(21)}U_{(10)}W_{(20)} = U_{(21)0}^\dagger U_{(10)0}^\dagger. \tag{A8}$$

On comparing Eq. (A8) with (A6) we see that

$$U_{(20)} = U_{(21)}U_{(10)}, \tag{A9}$$

since for a given $W_{(20)}$ Eqs. (A8) and (A6) have a unique solution because of the triangularity properties of the U operators (See Ref. 2, Part I). Equation (A9) can be generalized to paths having more steps. Perhaps the notion of paths is analogous to the concept of stochastic processes.

From Eq. (A9) we can get a relation between the kernels of the Gel'fand—Levitan equations involved in the direct and indirect paths. On using the third of Eq. (A1)

$$K_{(20)}(x|y) = K_{(21)}(x|x) + K_{(10)}(x|y) + \int_y^x K_{(21)}(x|z)K_{(10)}(z|y)\, dz, \quad \text{for } x \geq y, \tag{A10}$$

using an obvious notation.

It would be useful to know the relationship of the Gel'fand—Levitan kernels for the case that the roles of H_1 and H_0 were reversed as compared with the original formulation of the problem. One readily sees that $U_{(01)} = U_{(10)0}$ or

$$K_{(01)}(x|y) = K_{(10)0}(x|y). \tag{A11}$$

Another relation is given by Eq. (A10) when we identify the subscript 2 with the subscript 0 and use $K_{(00)}(x|y) \equiv 0$.

$$K_{(01)}(x|y) + \int_y^x K_{(01)}(x|z)K_{(10)}(z|y)$$

$$= -K_{(10)}(x|y), \quad x \geq y, \tag{A12}$$

which is of the form of an integral equation for $K_{(01)}$ in terms of $K_{(10)}$.

[1]I. Kay, Research Report No. EM-74, New York University Institute for Mathematical Sciences, Division of Electromagnetic Research (1955).

[2]I. Kay and H.E. Moses, Part I, Nuovo Cimento **2**, 917 (1955); Part II, **3**, 66 (1956); Part III, **3**, 276 (1956); Part IV, **5**, 230 (1957); Part V, **22**, 689 (1961).

[3]R. Jost and W. Kohn, Det Kgl. Danske Vidensk. Selskab. Mat.-fys. Medd. **27**, No. 9 (1953).

[4]R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), p. 610ff.

[5]I. Kay and H.E. Moses, J. Appl. Phys. **27**, 1503 (1956).

[6]C.S. Gardner, J.M. Greene, M.D. Kruskal, and R.M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).

Multi-soliton-like solutions to the Benjamin–Ono equation

R. I. Joseph

Department of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218
(Received 24 January 1977)

We outline a systematic method of obtaining particular solutions to the nonlinear, integrodifferential equation obtained by Benjamin and Ono in their study of the propagation of finite amplitude waves in fluids of great depth. These solutions have the property of asymptotically breaking up into a series of N spatially localized waves of permanent form; we loosely refer to these as "N-soliton" solutions. Detailed results for the two-soliton solution are explicitly given.

I. INTRODUCTION

In his study of internal waves of permanent form in fluids of great depth, Benjamin[1] discovered a new class of stationary wave of finite amplitude. Such waves have also been studied by Davis and Acrivos.[2] The key to understanding the difference in form between Benjamin's solution and the results of conventional shallow water theory, characterized by the solution of the Korteweg–de Vries (KdV) equation,[3] is the dispersion relation between the frequency ω and the wavenumber of infinitesimal periodic waves $u(x,t)$,

$$u(x,t)=\hat{u}\exp[i(\omega t-kx)]; \tag{1.1}$$

x is the spatial coordinate in the direction of propagation, and t the time. In the shallow water case, the KdV theory, the phase speed $c(k)=\omega/k$ has a smooth maximum c_0 at $k=0$, that is for waves of extreme length. Hence for small enough k values one has

$$c(k)\approx c_0(1-\beta k^2), \tag{1.2}$$

where β is a positive constant. However, in the case of fluids of great depth, the Benjamin theory, the dispersion relation for long waves has leading terms of the form

$$c(k)\approx c_0(1-\gamma|k|), \tag{1.3}$$

where γ is a positive constant. This nonanalyticity at $k=0$ has a profound effect on the form of the governing equation.

Assuming waves which vanish at $x=\pm\infty$, Benjamin[1] has shown by using the Fourier integral theorem in conjunction with a simple heuristic argument, that a finite amplitude wave $u(x,t)$ with dispersion given by Eq. (1.3) satisfies the equation

$$\frac{\partial u(x,t)}{\partial t}+Cu(x,t)\frac{\partial u(x,t)}{\partial x}=-c_0\frac{\partial u(x,t)}{\partial x}+\frac{c_0\gamma}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk\,dx'\,|k|\frac{\partial u(x',t)}{\partial x'}\times\exp[ik(x'-x)]. \tag{1.4}$$

A more rigorous derivation of this equation is given by Ono.[4] C is a parameter characterizing the strength of the nonlinearity and is discussed in detail by Stoker[5] and Benjamin.[6] By contrast, the corresponding equation for $u(x,t)$ when $c(k)$ is taken to be in the form given by Eq. (1.2) is[1]

$$\frac{\partial u}{\partial t}+Cu\frac{\partial u}{\partial x}=-c_0\frac{\partial u}{\partial x}-c_0\beta\frac{\partial^3 u}{\partial x^3}, \tag{1.5}$$

the well-known KdV equation. Both Eqs. (1.4) and (1.5) are special cases of a more general equation first proposed by Whitham[7]:

$$\frac{\partial u(x,t)}{\partial t}+Cu(x,t)\frac{\partial u(x,t)}{\partial x}=-\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk\,dx'\,c(k)\frac{\partial u(x',t)}{\partial x'}\exp[ik(x'-x)]; \tag{1.6}$$

Eqs. (1.5) and (1.4) result simply by substituting for $c(k)$ the expressions given in Eqs. (1.2) and (1.3), respectively, if one uses the integral representation of the delta function, $\delta(\eta)=(2\pi)^{-1}\int_{-\infty}^{\infty}\exp(ik\eta)$.

The solutions of Eqs. (1.5) and (1.4) in the form $u(x,t)=f(x-ct)$ are well known.[1,3] There are periodic solutions, and then there is a solitary wave solution. It is with the solitary wave type solutions (solitons) that the present paper is concerned. Solitons are in essence small wavenumber phenomena and their spectral energy content at large wavenumbers is extremely small. One therefore has the intuitive feeling that the form of the dispersion relation $c(k)$ as k becomes arbitrarily large should be irrelevant to their behavior.[8] Such issues are discussed in detail by a number of authors, see for example Benjamin.[9] For solutions of the form $u(x,t)=f(x-ct)$, it is straightforward to verify that Eq. (1.5) is solved by

$$u(x,t)=a\,\mathrm{sech}^2\left[\left(\frac{aC}{12c_0\beta}\right)^{1/2}(x-ct)\right] \tag{1.7}$$

with $a=3(c-c_0)/C$ while Eq. (1.4) is solved by

$$u(x,t)=a\lambda^2/[(x-ct)^2+\lambda^2] \tag{1.8}$$

with $a=4(c-c_0)/C=4c_0\gamma/C\lambda$. We shall refer to such solutions as single soliton solutions. If N denotes the number of "solitons" characterizing a solution, these are denoted $N=1$ solutions.

Considerable progress has been made in obtaining the general solution to the KdV equation, Eq. (1.5). This is based upon use of the inverse scattering technique[10] and is elegantly developed in Gardner *et al.*[11] N-soliton solutions have been obtained for a wide variety of initial functions $u(x,0)\equiv u_0(x)$. More recently Hirota[12] has developed a direct method of finding exact solutions to a wide class of nonlinear differential evolution equa-

tions. The method consists of transforming the evolution equation by $u=P/Q$ and seeking the special bilinear differential equations for P and Q, one of which is related to a linear differential operator. One then expands P and Q as power series in a parameter ϵ and determines the coefficients by perturbation theory.

However, in regard to "multi-soliton" solutions to Benjamin's equation, Eq. (1.4), the situation appears to be quite different. To the present author's knowledge, there appears to be no known analytic solutions which exhibit a multi-soliton character. In regard to the latter, we are using the word "soliton" quite loosely. What we shall here denote as an N-soliton solution for this equation is simply any function which satisfies it and which asymptotically, e.g., for large x and t, breaks up into a series of N spatially localized pieces each of which is of the functional form valid for the case $N=1$, that is, of the general form given by Eq. (1.8). Some information concerning the asymptotic characteristics of a general N-soliton solution to Eq. (1.4) can be obtained by using certain conservation laws derived by Ono.[4]

The purpose of the present paper then is to outline a possible method for finding *particular* solutions to Eq. (1.4) which have an N-soliton character, as defined in the previous paragraph. Due to the integral operator which appears in this equation, it does not appear feasible to fully utilize the techniques developed by Hirota.[12] However, the idea of writing the solution in the form $u=P/Q$ can be exploited. We carry this out as far as we can go for the N-soliton solution in Sec. 2 A; in Sec. 2 B we work out the complete details of the solution for the case $N=2$.

Using a suitably normalized spatial coordinate X, the $N=1$ case requires for u_0 a function of the form $u_0 \propto (1+X^2)^{-1}$, e.g., Eq. (1.8). For $N=2$ we shall show that in fact there is an entire family of initial functions for which we can find solutions, in particular, $u_0 \propto (1+X^2)/(1+pX^2+X^4)$, where p is an arbitrary constant (greater than -2). The case $N=1$ is included as an appropriate limit of the case $N=2$. This function $u_0(X)$ can be of two distinct natures. For $p \geq 1$ it is of the same general shape as the u_0 for $N=1$, that is, it has a single peak at $X=0$ and monotonically falls off away from it. However for $-2<p<1$, this initial function will itself have two peaks. Of particular interest is the case $p=2$ since here u_0 has the identical functional form as that for $N=1$. This situation is discussed in Sec. 3 and compared to the results which follow from Ono's conservation laws.[4]

2. N-SOLITON SOLUTIONS

A. General remarks

What we are looking for are solutions to Eq. (1.4) which asymptotically break up into a series of N spatially localized functions, each of these being of the form given by Eq. (1.8). Note that if $u(x,t)$ is a solution, then then so must be $u(-x,-t)$. With this in mind and motivated by Hirota's technique[12] we investigate the class of functions

$$u(x,t)=A\,\frac{P(x,t)}{Q(x,t)}, \tag{2.1}$$

where P and Q are bivariate polynomials in x and t of degree $2N-2$ and $2N$, respectively. We write them in the special form

$$P(x,t)=\sum_{i=0}^{2N-2} a_i(t)\,x^i, \tag{2.2}$$

$$a_i(t)=\sum_{j=\langle (i+1)/2\rangle}^{N-1} a_{i,2j-i}\,t^{2j-i}, \quad a_{2N-2}=1, \tag{2.3}$$

and

$$Q(x,t)=\sum_{i=0}^{2N} b_i(t)\,x^i, \tag{2.4}$$

$$b_i(t)=\sum_{j=\langle (i+1)/2\rangle}^{N} b_{i,2j-i}\,t^{2j-i}, \quad b_{2N}=1. \tag{2.5}$$

The symbol $\langle y\rangle$ here stands for taking the integral part of y. Our task is to now see if we can choose A and the $\{a\}$ and $\{b\}$ so as to make this u a solution of Eq. (1.4). It is considerably easier to work not with Eq. (1.4) directly, but rather with its Fourier transform. Multiplying Eq. (1.4) by $\exp(ikx)$ and then integrating over all x, gives, on using the boundary conditions $u\to 0$ as $|x|\to\infty$, the result

$$\begin{aligned} &ikc_0(1-\gamma|k|) \\ &=\left(\frac{\partial}{\partial t}\int_{-\infty}^{\infty} dx\,\exp(ikx)\,u(x,t)-\tfrac{1}{2}ikC\right. \\ &\quad\left.\times\int_{-\infty}^{\infty} dx\,\exp(ikx)\,u^2(x,t)\right)\Big/\int_{-\infty}^{\infty} dx\,\exp(ikx)\,u(x,t). \end{aligned} \tag{2.6}$$

The conditions imposed on A and the $\{a\}$, $\{b\}$ to make u a solution of Eq. (1.4) are identical to those imposed on them by this equation.

We seek solutions $u(x,t)$ which are real and finite for all real (x,t). Hence we assume that $Q\neq 0$ for all real (x,t). Consequently we can write Q in the equivalent form

$$Q(x,t)=\prod_{i=1}^{2N}[x-x_i(t)], \tag{2.7}$$

where the zeros $x_i(t)$ of Q must be complex for real t. Hence the roots $x_i(t)$ must occur in complex conjugate pairs and we adopt the notation

$$x_{N+j}(t)=x_j^*(t), \quad j=1,2,\ldots,N, \tag{2.8}$$

and assume that

$$\mathrm{Im}\,x_j(t)>0, \quad j=1,2,\ldots,N. \tag{2.9}$$

With the assumed form of $u(x,t)$ given by Eqs. (2.1)–(2.5) and Eqs. (2.7)–(2.9), the evaluation of the integrals appearing in Eq. (2.6) are now straightforward. Let

$$I_n\equiv\frac{1}{2\pi i}\int_{-\infty}^{\infty} dx\,\exp(ikx)[P(x,t)/Q(x,t)]^n. \tag{2.10}$$

We consider the cases $k\gtrless 0$ separately. For $k>0$, consider a contour which consists of the real axis from $x=-R$ to $x=+R$ closed by the semicircle C_+ in the upper half-plane of radius R. This contour encloses N

complex poles of Q, the $x_j(t)$, $j=1,2,\ldots,N$. In the limit $R\to\infty$, the integral of $\exp(ikx)[P/Q]^n$ over C_+ goes to zero whence we have that I_n is equal to $2\pi i$ times the sum of the residues of the integrand at the N poles $x_j(t)$. For $n=1$ the poles are simple while for $n=2$ we have poles of order 2. Direct evaluation of these residues for $n=1$ gives

$$I_1^{(k>0)}=\sum_{j=1}^{N} S_j(t)\exp[ikx_j(t)] \tag{2.11}$$

with

$$S_j(t)\equiv P(x_j(t),t)/\partial_x Q(x_j(t),t) \tag{2.12}$$

and

$$\partial_x Q(x,t)=Q(x,t)\sum_{j=1}^{2N}\frac{1}{x-x_j(t)}\ . \tag{2.13}$$

For $n=2$ we similarly obtain

$$I_2^{(k>0)}=\sum_{j=1}^{N} S_j^2(t)\exp[ikx_j(t)][ik+2T_j(t)], \tag{2.14}$$

where

$$T_j(t)\equiv\frac{\partial_x P(x_j(t),t)}{P(x_j(t),t)}-\sum_{i=1}^{2N}{}'\frac{1}{x_j(t)-x_i(t)}. \tag{2.15}$$

The prime on the summation excludes the term $i=j$.

When $k<0$, we use a contour which consists of the real axis from $x=+R$ to $x=-R$ closed by the semicircle C_- of radius R in the lower half-plane. This contour now enclosed the N poles $x_j^*(t)$, $j=1,2,\ldots,N$. One then similarly finds that

$$I_1^{(k<0)}=-\sum_{j=1}^{N} S_j^*(t)\exp[ikx_j^*(t)] \tag{2.16}$$

and

$$I_2^{(k<0)}=-\sum_{j=1}^{N}[S_j^*(t)]^2\exp[ikx_j^*(t)][ik+2T_j^*(t)]. \tag{2.17}$$

Substitution of these results into Eq. (2.6) then gives for $k>0$

$$\begin{aligned} ikc_0(1-\gamma k) &= \left[\sum_{j=1}^{N}\exp(ikx_j)\frac{\partial S_j}{\partial t}+ik\sum_{j=1}^{N}\exp(ikx_j)S_j\left(\frac{\partial x_j}{\partial t}-CAS_jT_j\right)\right.\\ &\quad\left.-\tfrac{1}{2}CA(ik)^2\sum_{j=1}^{N}\exp(ikx_j)S_j^2\right]\Bigg/\sum_{j=1}^{N}\exp(ikx_j)S_j \end{aligned} \tag{2.18}$$

while for $k<0$,

$$\begin{aligned} ikc_0(1+\gamma k) &= \left[\sum_{j=1}^{N}\exp(ikx_j^*)\frac{\partial S_j^*}{\partial t}+ik\sum_{j=1}^{N}\exp(ikx_j^*)S_j^*\left(\frac{\partial x_j^*}{\partial t}-CAS_j^*T_j^*\right)\right.\\ &\quad\left.-\tfrac{1}{2}CA(ik)^2\sum_{j=1}^{N}\exp(ikx_j^*)[S_j^*]^2\right]\Bigg/\sum_{j=1}^{N}\exp(ikx_j^*)S_j^*\ . \end{aligned} \tag{2.19}$$

In order that the left- and right-hand sides of each of these equations be equivalent for arbitrary k and t, that is, that our assumed u in fact be a solution, the following sets of conditions should be satisfied:

$$\frac{\partial S_j(t)}{\partial t}=0, \tag{2.20a}$$

$$\frac{\partial x_j(t)}{\partial t}-CAS_j(t)\,T_j(t)=c_0, \tag{2.20b}$$

$$\tfrac{1}{2}CAiS_j(t)=c_0\gamma, \tag{2.20c}$$

$$j=1,2,\ldots,N;$$

$$\frac{\partial S_j^*(t)}{\partial t}=0, \tag{2.21a}$$

$$\frac{\partial x_j^*(t)}{\partial t}-CAS_j^*(t)\,T_j^*(t)=c_0, \tag{2.21b}$$

$$\tfrac{1}{2}CAiS_j^*(t)=-c_0\gamma, \tag{2.21c}$$

$$j=1,2,\ldots,N.$$

The conditions imposed by Eqs. (2.20a) and (2.20c) require that all of the $S_j(t)$ be equal and time independent. Denote this common constant value by S. Then Eq. (2.20c) dictates the value of A in terms of S,

$$A=2c_0\gamma/iCS. \tag{2.22}$$

Equation (2.20b) then requires that the $x_j(t)$ obey the equation

$$\frac{\partial x_j(t)}{\partial t}=c_0[1-2i\gamma T_j(t)],\quad j=1,2,\ldots,N. \tag{2.23}$$

If Eqs. (2.20a) and (2.20b) are obeyed, then Eqs. (2.21a) and (2.21b) will automatically be satisfied. Finally the condition given by Eq. (2.21c) will be consistent with that from Eq. (2.20c) if

$$S_j(t)=S=-S^*,\quad j=1,2,\ldots,N; \tag{2.24}$$

that is, if S is purely imaginary. We now show that this is identically valid by using the following theorem.[13] Let $f(x)$ denote a polynomial in x of degree n. Denote by α_i any one of the n roots of $f(x)=0$ so that

$$f(x)=\prod_{i=1}^{n}(x-\alpha_i). \tag{2.25}$$

Let $\phi(x)$ denote an *arbitrary* polynomial in x of degree $m\leq n-2$. Then

$$\sum_{i=1}^{n}\frac{\phi(\alpha_i)}{f'(\alpha_i)}=0. \tag{2.26}$$

To use this theorem we need only identify $f(x)$ with Q [a polynomial of degree $2N$ in x] and $\phi(x)$ with P [a polynomial of degree $2N-2$ in x]. Then

$$\sum_{i=1}^{2N}\frac{P(x_i,t)}{\partial_x Q(x_i,t)}=0,\quad \text{for all } t. \tag{2.27}$$

But by the definition of the $S_j(t)$, Eq. (2.12), we then have

$$\sum_{i=1}^{2N}S_i(t)=0. \tag{2.28}$$

Now by the previous conditions, $S_i(t)=S$ for $i=1,2,\ldots,N$ and $S_i(t)=S^*$ for $i=N+1,\ N+2,\ldots,2N$, whence Eq. (2.28) becomes

$$N(S+S^*)=0 \tag{2.29}$$

or

$$S^*=-S. \tag{2.30}$$

QED

Since $S_j(t)=S$, a constant, we can in principal evaluate it for any value of t, in particular, the value $t=0$. Further details are considered in the next section.

B. Detailed results for the two-soliton case ($N=2$)

The actual labor involved in carrying out the procedure just outlined is quite considerable. Hence in the present section we shall follow it through for the specific case of $N=2$. The results for the case $N=1$ are already given by Eq. (1.8).

To start with, we want the roots of $Q=0$ for $t=0$. Denote these values of the $x_j(t)$ by y_j and in particular, set $x_1(0)=y_1$, $x_2(0)=y_2$, $x_3(0)=y_1^*$, $x_4(0)=y_2^*$. These are then obtained by solution of the quartic equation

$$0=y^4+b_{20}y^2+b_{00}. \tag{2.31}$$

We shall see that all of the rest of the members of the sets $\{a\}$ and $\{b\}$ can be expressed in terms of the two parameters b_{20} and b_{00}, which are taken as our initializing parameters. Two cases can occur which yield complex roots:

(i) $b_{00}>0$, $b_{20}>0$, $b_{20}^2-4b_{00}>0$,

(ii) $b_{00}>0$, $b_{20}^2-4b_{00}<0$.

For convenience we introduce the notation

$$\begin{aligned} &b_{00}=q^4,\\ &b_{20}=pq^2, \quad q>0. \end{aligned} \tag{2.32}$$

The requirement that u be finite for all real (x,t), e.g., $Q\neq 0$, restricts p to values such that $p+2>0$. Then for case (i) we have

$$y_1=iq\left(\frac{p+(p^2-4)^{1/2}}{2}\right)^{1/2}, \quad y_2=iq\left(\frac{p-(p^2-4)^{1/2}}{2}\right)^{1/2} \tag{2.33}$$

while for case (ii),

$$y_1=\tfrac{1}{2}q\{\lambda+i\mu\}, \quad y_2=\tfrac{1}{2}q\{-\lambda+i\mu\} \tag{2.34}$$

with

$$\lambda=(2-p)^{1/2}, \quad \mu=(2+p)^{1/2}. \tag{2.35}$$

Now, the requirement that $S_1(t)=S_2(t)=S$ becomes, on using Eq. (2.12),

$$S=\frac{x_1^2+a_1x_1+a_0}{(x_1-x_2)(x_1-x_1^*)(x_1-x_2^*)}=\frac{x_2^2+a_1x_2+a_0}{(x_2-x_1)(x_2-x_2^*)(x_2-x_1^*)}. \tag{2.36}$$

When $t=0$, for both case (i) and case (ii), it reduces to

$$\frac{y_1^2+a_{00}}{y_1}=-\left(\frac{y_2^2+a_{00}}{y_2}\right) \tag{2.37}$$

which is satisfied if $a_{00}=-y_1y_2$ and hence from Eq. (2.36), $S=1/2(y_1+y_2)$. Substituting in the values of y_1 and y_2 previously given, one finds that for either case,

$$S=1/2iq(p+2)^{1/2}, \quad a_{00}=q^2. \tag{2.38}$$

Next, we solve Eqs. (2.36) to express the unknowns $a_1(t)$ and $a_0(t)$ in terms of the $x_i(t)$ and S,

$$\begin{aligned} a_1(t)=&-(x_1+x_2)+S[(x_1-x_1^*)(x_1-x_2^*)\\ &+(x_2-x_2^*)(x_2-x_1^*)]\\ a_0(t)=&x_1x_2-S[x_2(x_1-x_1^*)(x_1-x_2^*)\\ &+x_1(x_2-x_2^*)(x_2-x_1^*)]. \end{aligned} \tag{2.39}$$

The coefficients a_{11} and a_{02} can be calculated from these expressions by using the fact that

$$a_{11}=\left.\frac{da_1(t)}{dt}\right|_{t=0}, \quad 2a_{02}=\left.\frac{d^2a_0(t)}{dt^2}\right|_{t=0} \tag{2.40}$$

and by making repeated and systematic use of the conditions imposed on the $dx_i(t)/dt$ by Eq. (2.23). After some algebra, one finds for both cases (i) and (ii) that

$$\begin{aligned} &a_{11}/c_0=-2[1+\hat{\gamma}(p+4)],\\ &a_{02}/c_0^2=1+2\hat{\gamma}(p+4)+\hat{\gamma}^2(p+3)(p+6), \end{aligned} \tag{2.41}$$

where for simplicity we have here introduced the notation

$$\hat{\gamma}\equiv\gamma/q(p+2)^{1/2}. \tag{2.42}$$

We next turn to the evaluation of the coefficients $\{b\}$. Since the $x_j(t)$ are solutions of the equation $Q=0$, or

$$0=x^4+b_3(t)x^3+b_2(t)x^2+b_1(t)x+b_0(t), \tag{2.43}$$

we may simply write

$$\begin{aligned} -b_3(t)&=\sum_{j=1}^{4}x_j(t),\\ b_2(t)&=\sum_{i<j=1}^{4}x_i(t)x_j(t),\\ -b_1(t)&=\sum_{i<j<k=1}^{4}x_i(t)x_j(t)x_k(t),\\ b_0(t)&=\prod_{i=1}^{4}x_i(t). \end{aligned} \tag{2.44}$$

Whence we can use the facts that

$$\begin{aligned} b_{31}&=\left.\frac{db_3(t)}{dt}\right|_{t=0},\\ 2b_{22}&=\left.\frac{d^2b_2(t)}{dt^2}\right|_{t=0}, \end{aligned} \tag{2.45}$$

etc.,

together with Eq. (2.23) to similarly evaluate the b_{ij}. After considerable algebra, one then finds

$$\begin{aligned} b_{31}/c_0&=-2[2+\hat{\gamma}(p+6)],\\ b_{22}/c_0^2&=6+6\hat{\gamma}(p+6)+\hat{\gamma}^2(p+6)(p+8),\\ b_{11}/c_0&=-2q^2[p+\hat{\gamma}(p^2+5p+2)],\\ b_{13}/c_0^3&=-2[2+3\hat{\gamma}(p+6)+\hat{\gamma}^2(p+6)(p+8)\\ &\quad+\hat{\gamma}^3(p+6)^2],\\ b_{02}/c_0^2&=q^2[p+2\hat{\gamma}(p^2+5p+2)\\ &\quad+\hat{\gamma}^2(p^3+10p^2+30p+20)],\\ b_{04}/c_0^4&=1+2\hat{\gamma}(p+6)+\hat{\gamma}^2(p+6)(p+8)\\ &\quad+2\hat{\gamma}^3(p+6)^2+\hat{\gamma}^4(p+6)^2. \end{aligned} \tag{2.46}$$

Consequently, for $N=2$, we have a complete specification of the function $u(x,t)$ of Eq. (2.1). The various coefficients $\{a\}$, $\{b\}$ which appear in it have all been ex-

pressed in terms of the two initializing parameters p and q. The coefficient A is given by

$$A = 4c_0\gamma q(p+2)^{1/2}/C. \tag{2.47}$$

This function of course represents a particular solution to Eq. (1.4) and is valid *only* if when $t=0$ we assume an initializing function

$$u(x,0) \equiv u_0(x) = A(x^2+q^2)/(x^4+pq^2x^2+q^4), \tag{2.48}$$

thus specifying the particular values of p and q to be used. Let us now just briefly look at the shape of this class of initializing functions $u_0(x)$ for which we have found solutions to Eq. (1.4). For convenience, we shall introduce a scaled spatial variable X by $x = qX$ so that Eq. (2.48) becomes

$$u_0(X) = (A/q^2)(1+X^2)/(1+pX^2+X^4). \tag{2.49}$$

Note that p is to be restricted to values $p > -2$ to guarantee that u is finite for all real x and t. It is now straightforward to show for $p \geq 1$ that u_0 has an absolute maximum at $X=0$ with value $u_0|_{\max} = A/q^2$. On the other hand, for $-2 < p < 1$, $X=0$ corresponds to a relative minima with absolute maxima of u_0 occurring at

$$\bar{X} = \pm[(2-p)^{1/2}-1]^{1/2} \tag{2.50}$$

at which points

$$u_0|_{\max} = (A/q^2)/(2-p)^{1/2}[2-(2-p)^{1/2}]. \tag{2.51}$$

Hence for $p \geq 1$, u_0 is a function of X which monotonically falls off from its peak value at $X=0$ while for $-2 < p < 1$, u_0 first increases as we move off in either direction from $X=0$, peaks, and then monotonically falls off for large $|X|$. Hence the shape of the functions u_0 for the class of solutions obtained is quite different as $p \geq 1$ or $-2 < p < 1$.

Before we turn to a detailed description of the complete space–time evolution of $u(x,t)$ for the class of solutions obtained, let us first consider their asymptotic behavior. Let x, t be very large but $x - ct \equiv \xi$ be finite, c being a constant as yet undetermined. We now replace x by $ct+\xi$ in the expressions for P and Q, Eqs. (2.1)–(2.5) with $N=2$, and retain only the most significant terms (in terms of powers of t). Then

$$\begin{aligned} P &= t^2(c^2 + a_{11}c + a_{02}) + O(t), \\ Q &= t^4(c^4 + b_{31}c^3 + b_{22}c^2 + b_{13}c + b_{04}) + t^3\xi(4c^3 + 3b_{31}c^2 \\ &\quad + 2b_{22}c + b_{13}) + t^2[(b_{20}c^2 + b_{11}c + b_{02}) \\ &\quad + \xi^2(6c^2 + 3b_{31}c + b_{22})] + O(t). \end{aligned} \tag{2.52}$$

If for $t \gg 0$ we want u to remain finite, we must impose the restriction that the coefficient of the terms in Q of order t^4 and t^3 identically vanish, otherwise, u is of $O(1/t^2)$. Denote the coefficient of the term in t^4 by $f(c)$. We then require that

$$\begin{aligned} f(c) &= c^4 + b_{31}c^3 + b_{22}c^2 + b_{13}c + b_{04} = 0, \\ f'(c) &= 0. \end{aligned} \tag{2.53}$$

The first of these two equations yields four possible values for c. The second equation is just the requirement that these roots be double roots, that is, that we have two double roots. Whether this can in general occur will of course depend on what the particular values of the coefficients appearing in $f(c)$ are. Denote the two double roots of $f=0$ by c_+ and c_-. It now follows directly from $f=0$ that

$$\begin{aligned} &b_{31} = -2(c_+ + c_-), \quad b_{22} = c_+^2 + 4c_+c_- + c_-^2, \\ &b_{13} = -2c_+c_-(c_+ + c_-), \quad b_{04} = c_+^2c_-^2. \end{aligned} \tag{2.54}$$

Whence we find that there will be two double roots, c_+, c_-, given by

$$c_\pm = \tfrac{1}{4}\left[-b_{31} \pm [b_{31}^2 - 16(b_{13}/b_{31})]^{1/2}\right] \tag{2.55}$$

if the coefficients satisfy the constraints

$$b_{04} = (b_{13}/b_{31})^2, \quad b_{22} = \tfrac{1}{4}b_{31}^2 + 2(b_{13}/b_{31}). \tag{2.56}$$

Substituting the actual values of these coefficients as previously determined, Eq. (2.46), into Eqs. (2.56) we directly find that they are identically satisfied for all p and q! Hence for large x and t, our $N=2$ solution breaks up into two localized waves moving at speeds given by

$$\frac{c_\pm - c_0}{c_0} = \tfrac{1}{2}\hat{\gamma}\{(p+6) \pm [(p+2)(p+6)]^{1/2}\} \tag{2.57}$$

and are of the form

$$u_\pm = AR_\pm/[(x-c_\pm t)^2 + \lambda_\pm^2], \tag{2.58}$$

where

$$\begin{aligned} R_\pm &= \frac{1}{2}\left\{1 \mp \left[\frac{p+2}{p+6}\right]^{1/2}\right\} \\ \lambda_\pm^2 &= \tfrac{1}{2}q^2\left(\frac{p+2}{p+6}\right)\left\{(p+4) \mp [(p+2)(p+6)]^{1/2}\right\}. \end{aligned} \tag{2.59}$$

The maximum value of $u_\pm$ (when $x - c_\pm t = 0$) is given by

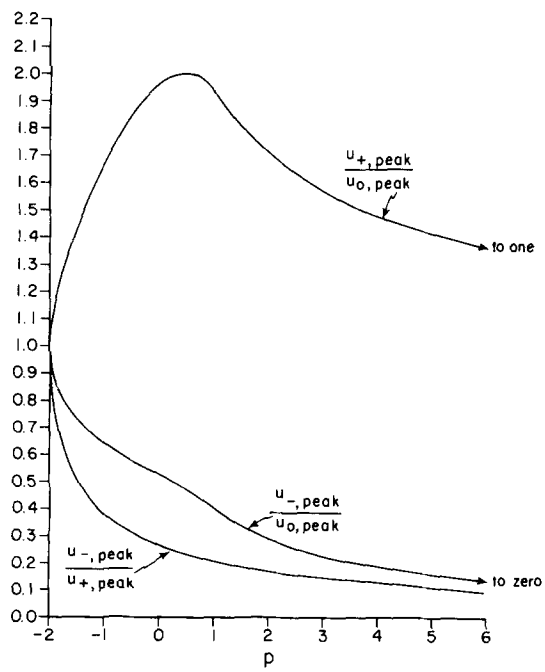


FIG. 1. Plot of asymptotic peak amplitudes (u_{peak}) relative to the peak initial amplitude ($u_{0,\text{peak}}$) for the $N=2$ solution to the Benjamin equation as a function of the initializing parameter p (>-2). Also shown is $(c_- - c_0)/(c_+ - c_0)$ $[=u_{-,\text{peak}}/u_{+,\text{peak}}]$.

$$u_\pm|_{\text{peak}} = AR_\pm/\lambda_\pm^2$$
$$= (4c_0\hat{\gamma}/C)\left(\frac{p+6\mp[(p+2)(p+6)]^{1/2}}{p+4\mp[(p+2)(p+6)]^{1/2}}\right). \quad (2.60)$$

To get an idea of how large these asymptotic waves can get, in Fig. 1 we show as a function of p a plot of the values of $u_\pm|_{\text{peak}}$ normalized by the peak value of the corresponding initializing function, $u_0|_{\text{peak}}$. Note that this normalizing quantity is different for $p \geq 1$ and $-2 < p < 1$. For all $p > -2$, this relative amplitude is always less than unity for the slow wave u_- and greater than unity for the fast wave u_+. When $p = \frac{1}{2}$ the fast wave amplitude is twice the corresponding initial amplitude and this is as large as this ratio can ever get. Also shown in this figure is a plot of the ratio of the slow to fast wave peak amplitudes. This curve is of interest since it follows directly from Eqs. (2.60) and (2.57) that

$$\frac{u_-|_{\text{peak}}}{u_+|_{\text{peak}}} \equiv \frac{c_- - c_0}{c_+ - c_0}, \quad (2.61)$$

that is, the ratio of peak wave amplitudes is equal to the corresponding ratio of the deviation of their wave speed from c_0.

We now turn to a description of the detailed space–time evolution of $u(x,t)$ and in particular how it approaches its asymptotic character. In describing these results numerically it is convenient to introduce a set of scaled variables by

$$X = x/q, \quad T = c_0 t/q,$$
$$U = (q^2/A)\,u = [qC/4c_0\gamma(p+2)^{1/2}]u. \quad (2.62)$$

We then write

$$U = \hat{P}/\hat{Q}, \quad (2.63)$$

where

$$\hat{P} = 1 + (X - VT)^2 - (p+2)\hat{\gamma}T(X - VT)$$
$$+ \tfrac{1}{4}(p+2)(p+6)\hat{\gamma}^2T^2 \quad (2.64)$$

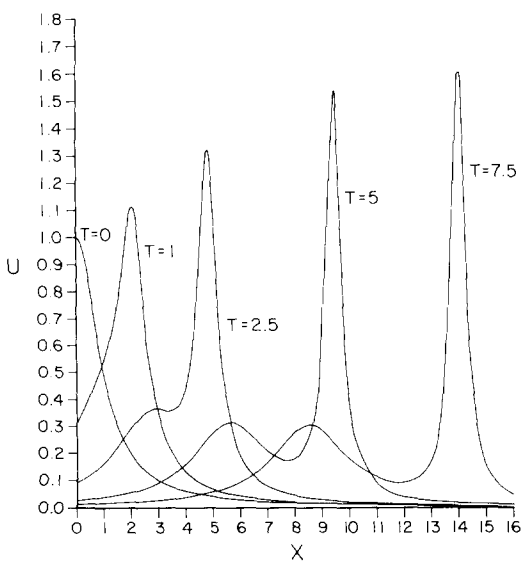


FIG. 2. Plot of normalized solution U to the Benjamin equation for $N=2$ as a function of the normalized spatial position X for various values of normalized time T for the initializing parameters $p=2$, $q=4\gamma$.

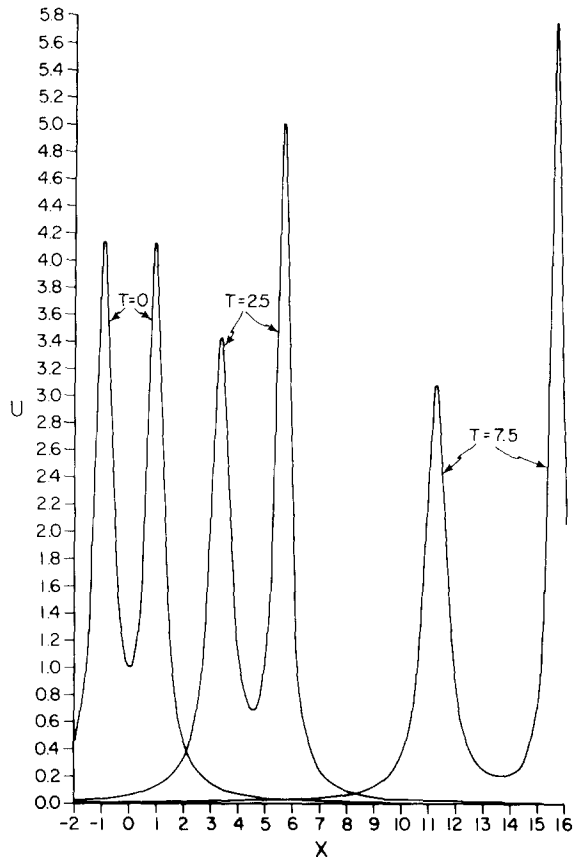


FIG. 3. Same as Fig. 2 except $p = -\frac{3}{2}$.

and

$$\hat{Q} = 1 + p(X - VT)^2 - (p+2)^2\hat{\gamma}T(X - VT)$$
$$+ \tfrac{1}{4}(p+2)(p^2+6p+16)\hat{\gamma}^2T^2$$
$$+ (X - VT)^4 - \tfrac{1}{2}(p+2)(p+6)\hat{\gamma}^2T^2(X - VT)^2$$
$$+ \tfrac{1}{16}(p+2)^2(p+6)^2\hat{\gamma}^4T^4 \quad (2.65)$$

with

$$V \equiv 1 + \tfrac{1}{2}(p+6)\hat{\gamma}. \quad (2.66)$$

The parameter V is just the average of the two asymptotic wave speeds relative to c_0,

$$V = (c_+ + c_-)/2c_0. \quad (2.67)$$

Different ways of regrouping the various powers of X and T are of course possible. Figures 2 and 3 show plots of U as given by Eqs. (2.63)–(2.66) as a function of X for various values of T for $\gamma/q = \frac{1}{4}$ and $p = 2$, $-\frac{3}{2}$, respectively. These two values of p characterize the two kinds of initial u possible, e.g., a single or double humped function. For $p=2$ we should have that the peak value of U for the asymptotic fast wave is $\frac{1}{2}(2+\sqrt{2})$ with an asymptotic speed of $\frac{1}{4}(6+\sqrt{2})$. To get some idea of how there asymptotic values are reached as a function of time, in Fig. 4 we show the actual peak value of U at each T for what we know becomes the fast wave. By a dimensionless time $T \approx 5$ we are within 10% of the correct asymptotic value. In order to get some idea of the "speed" of the wave, we can define an average speed $\langle c \rangle$ at any instant of time as the ratio of the actual

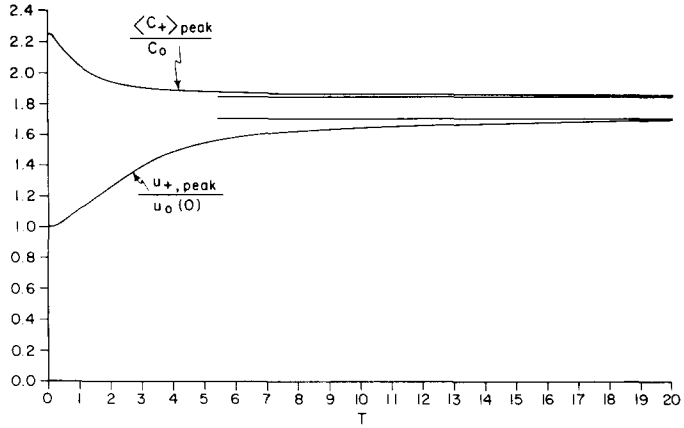


FIG. 4. Dependence on normalized time T of certain properties of that part of the $N=2$ solution to the Benjamin equation which can be identified as asymptotically becoming the fast wave. Plotted are the peak amplitude normalized by the peak initial amplitude $(u_{+,\text{peak}}/u_{0,\text{peak}})$ and the average speed of that peak $\langle c_+\rangle$ normalized by $c_0[\langle c_+\rangle/c_0 \equiv X_{+,\text{peak}}(T)/T]$. The figure corresponds to the case $q=4\gamma$, $p=2$.

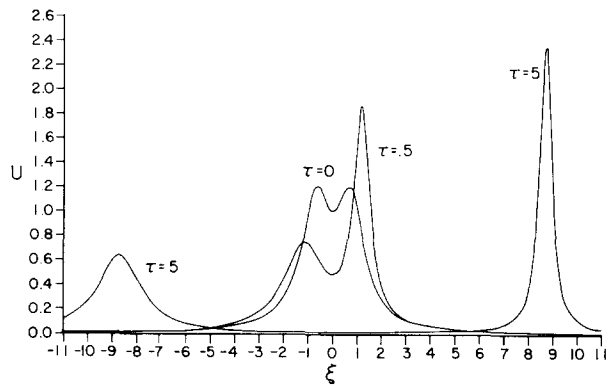


FIG. 5. Plot of the normalized solution U to the Benjamin equation for $N=2$ in a spatial frame of reference moving at the average asymptotic speed V relative to the frame (X,T) as a function of $\xi(\equiv X-VT)$ for various values of τ $(\equiv\hat{\gamma}T)$ for the case of the initialing parameter $p=0$.

location of the peak which becomes the fast wave to that value of the time. This is also shown in Fig. 4. It is clear that the peak is *decelerating* from an "initial" speed which is about 22% higher than the asymptotic speed and that by $T\approx 5$ we are within 2% of the asymptotic speed.

In order to make the plots shown in Figs. 2 and 3, we had to specify a value for γ/q in addition to one for p. One could eliminate the need to pick a particular value of γ/q in such plots by transforming to a new coordinate system which moves with the speed V relative to the original one, that is, let $\xi\equiv X-VT$ and by introducing a new scaled time by $\tau=\hat{\gamma}T$. In Fig. 5 we show a plot of U vs. ξ for various values of τ for the case $p=0$. In this plot the two waves instead of both moving off to the right at their respective speeds, appear to move off in opposite directions with similar speeds.

Finally, we now show that the single soliton solution, $N=1$, is included as a limiting case of the $N=2$ solution just presented. This should be apparent from the $p\to\infty$ limit shown in Fig. 1. Suppose that $p\to\infty$ and $q\to\infty$ but that $q/\sqrt{p}$ remains finite. Then from Eqs. (2.1), (2.47), (2.32), (2.41), and (2.46) we get

$$\begin{aligned} &P\to q^2,\\ &Q\to pq^2\{q^2/p+(x-c_0t[1+\bar{\gamma}])^2\}, \qquad (2.68)\\ &A\to(4c_0\gamma q/C\sqrt{p})\,p, \end{aligned}$$

where $\bar{\gamma}\equiv\gamma\sqrt{p}/q$. Whence we get

$$u\to(4c_0\gamma q/C\sqrt{p})/[q^2/p+(x-c_0t[1+\bar{\gamma}])^2]. \qquad (2.69)$$

Now making the identifications

$$\lambda^2=q^2/p,\quad c=c_0(1+\bar{\gamma}),$$

we can write this as

$$u=(4c_0\gamma\lambda/C)/[(x-ct)^2+\lambda^2] \qquad (2.70)$$

which is exactly the same as Eq. (1.8) since $a\lambda=4c_0\gamma/C$ and $(c-c_0)/c_0=\bar{\gamma}=\gamma/\lambda$.

3. DISCUSSION

For the two-soliton solutions just discussed, the case $p=2$ is of particular interest. In this case the $u_0(x)$ we require is of the form

$$u_0(x)=(8c_0\gamma q/C)/[x^2+q^2]. \qquad (3.1)$$

Except for the scale factor of 2, this $u_0(x)$ is the same as the one required to obtain a single soliton solution, Eq. (1.8),

$$u_0(x)=(4c_0\gamma q/C)/[x^2+q^2]. \qquad (3.2)$$

Hence by doubling the amplitude of this initial form, instead of finding a single soliton of speed c, we produce a two-soliton solution. For the single soliton case, the speed c is related to the peak value of $u_0(x)$ by

$$u_0|_{\text{peak}}=(4c_0\gamma/qC)=(4/C)(c-c_0). \qquad (3.3)$$

For the corresponding two-soliton case, the speeds are related to the corresponding peak initial value by

$$\begin{aligned} u_0|_{\text{peak}}&=(8c_0\gamma/qC)=(4/C)(c_+-c_0)(2-\sqrt{2})\\ &=(4/C)(c_--c_0)(2+\sqrt{2}). \end{aligned} \qquad (3.4)$$

For the single soliton solution the peak amplitude of the soliton is of course identical to that of $u_0(x)$. However for the two-soliton solution being considered here, the peak value of the faster moving soliton is larger by the factor $1+\frac{1}{2}\sqrt{2}\cong 1.707$ than the peak initial value while for the slower soliton, it is smaller by the factor $1-\frac{1}{2}\sqrt{2}\cong 0.293$.

Ono[4] has developed a number of conservation laws satisfied by solutions of Eq. (1.4). The first two of these are $\partial_t\int_{-\infty}^{\infty}dx\,u=0$, $\partial_t\int_{-\infty}^{\infty}dx\,\frac{1}{2}u^2=0$, and these are identical to the corresponding laws satisifed by solutions to the KdV equation. These two laws are satisfied in general by solution of the more general equation, Eq. (1.6). For this equation, it is straightforward to show that a third conservation law is

$$\partial_t\int_{-\infty}^{\infty}dx[\tfrac{1}{3}u^3+(1/C)\,u\,\mathcal{G}(u)]=0, \qquad (3.5)$$

where

$$\mathcal{G}(u(x)) \equiv \int_{-\infty}^{\infty} dx'\, u(x')\, G(x'-x). \tag{3.6}$$

$$G(z) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} c(k) \exp(ikz).$$

This result reduces to that given by Ono for Eq. (1.4) and to that valid for the KdV equation, Eq. (1.5), when G is suitably picked. A fourth conservation law satisfied by these equations is $\partial_t(\partial_t \int_{-\infty}^{\infty} dx\, xu) = 0$. See, for example, Miura *et al.*[14] *Assuming* that a two-soliton solution existed, Ono has applied these conservation laws to the case $p=2$, obtaining results in exact agreement with those just given.

Finally, it is instructive to compare the case $p=2$ with what obtains for solution of the KdV equation under similar conditions. A single, *pure*-soliton solution to the KdV equations results from the $u_0(x)$ given by Eq. (1.7),

$$u_0(x) = (12c_0\beta/C\ell^2)\,\mathrm{sech}^2(x/\ell) \tag{3.7}$$

when the soliton speed c is related the peak value of $u_0(x)$ by

$$u_0|_{\text{peak}} = (3/C)(c - c_0). \tag{3.8}$$

In order to obtain a *pure* two-soliton solution with the exact same initial function form for $u_0(x)$, the initial function must have a peak value *three* times larger than in the single soliton case,[11]

$$u_0(x) = (36c_0\beta/C\ell^2)\,\mathrm{sech}^2(x/\ell). \tag{3.9}$$

In this situation the asymptotic speeds of the two solitons are related to the peak value of the initial u by

$$u_0|_{\text{peak}} = (9/4C)(c_{\text{fast}} - c_0) = (9/C)(c_{\text{slow}} - c_0). \tag{3.10}$$

For the single soliton solution the peak amplitude of the soliton is identical to the corresponding value of $u_0(x)$ while for the two-soliton solution, the peak amplitude of the faster soliton is larger by the factor $\frac{4}{3}$ than the corresponding initial peak value, while for the slower soliton, it is smaller by the factor $\frac{1}{3}$. Note also, that just as for the two-soliton solution to the Benjamin equation, here also

$$\frac{c_{\text{slow}} - c_0}{c_{\text{fast}} - c_0} = \frac{u_{\text{slow, peak}}}{u_{\text{fast, peak}}}, \tag{3.11}$$

e.g., Eq. (2.61). Hence an interesting difference between the Benjamin and KdV equations, is that in order to produce two pure solitons one has to *double* the amplitude of the initial u for a single soliton solution for the former equation while for the latter equation, one has to *triple* it.

Still to be discovered, is a method of finding solutions to Eq. (1.4) for an arbitrary $u_0(x)$ and of course to prove that the solutions obtained in fact give *solitons*.

In regard to this latter point, it is at least clear that for the case $N=2$ discussed in Sec. 2 B that we really do have *true solitons*. That is, since $u(x,t) = u(-x,-\hat{t})$, imagine that at some large negative value of t, $t=-\hat{t}$, we have as an input to the system two waves of the form u_+ and u_-, Eq. (2.58), centered at the large negative values of position $x=-x_+, -x_-$, and moving to the right with speeds c_+ and c_-, respectively. Pick $\hat{t}$ such that $(x_+ + x_-) = \hat{t}(c_+ + c_-)$. We then write

$$u_{\text{input}} = \frac{AR_+}{[(x - c_+t + \phi_+)^2 + \lambda_+^2]} + \frac{AR_-}{[(x - c_-t + \phi_-)^2 + \lambda_-^2]}. \tag{3.12}$$

The phase factors $\phi\pm$ obey the condition $\phi_+ + \phi_- = 0$. As these wave propagate to the right they "interact." For large positive t and x these two waves re-emerge in unaltered shape, that is, $u_{\text{output}} = u_+ + u_-$, which is the same as Eq. (3.12) except for the absence of the phase factors ϕ_+ and ϕ_-. The locations of u_+ and u_- with respect to each other are reversed. That there really has been an "interaction" is simply seen by noting that

$$(u_+ + u_-)_{t=0} = A\left(\frac{x^2 + q^2\mu}{x^4 + x^2q^2\mu(p+4) + q^4\mu^2}\right), \tag{3.13}$$

where $\mu = (p+2)/(p+6)$. Except for large $|x|$, this expression is in general not equal to $u(x,t)$ for $t=0$, Eq. (2.48). If $p \to \infty$, they are the same for all x, but this just corresponds to the single soliton limit. This same idea can be seen in Fig. 4 which directly shows, or implies, that for the particular case $p=2$, the fast wave entering at large negative x and t is accelerating from the speed c_+ up to a maximum speed at $t=0$ after which time it decelerates down to its original speed c_+ at large positive t.

[1] T. B. Benjamin, J. Fluid Mech. **29**, 559 (1967).
[2] R. E. Davis and A. Acrivos, J. Fluid Mech. **29**, 593 (1967).
[3] D. J. Korteweg and G. deVries, Phil. Mag. **39**, 422 (1895).
[4] H. Ono, J. Phys. Soc. Jpn. **39**, 1082 (1975).
[5] J. J. Stoker, *Water Waves* (Interscience, New York, 1957), Chap. 10.
[6] T. B. Benjamin, J. Fluid Mech. **25**, 241 (1966).
[7] G. B. Whitham, Proc. R. Soc. A **299**, 6 (1967).
[8] O. M. Phillips, private communication.
[9] T. B. Benjamin, J. L. Bona, and J. J. Mahoney, Philos. Trans. R. Soc. A **272**, 47 (1972).
[10] I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl. Sec. 2, **1**, 253 (1955).
[11] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Commun. Pure Appl. Math. **27**, 97 (1974).
[12] R. Hirota, in *Bäcklund Transformations, the Inverse Scattering Method, Solitons and their Applications*, Lecture Notes in Mathematics, No. 515, edited by R. Miura (Springer, New York, 1976), p. 40.
[13] W. S. Burnside and A. W. Panton, *The Theory of Equations* (Longmans, Green and Co., London, 1928), Vol. I, Chap. VIII.
[14] R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 93 (1968); **11**, 952 (1970).

Continuous subgroups of the fundamental groups of physics. III. The de Sitter groups[a)]

J. Patera

Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada

R. T. Sharp

Department of Physics, McGill University, Montreal, Quebec, Canada

P. Winternitz

Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada

H. Zassenhaus

Department of Mathematics, Ohio State University, Columbus, Ohio
(Received 23 February 1977)

An algorithm for classifying the closed connected subgroups S of a given Lie group G into conjugacy classes, presented in earlier papers, is further refined so as to provide us with "normalized" lists of representatives of subalgebra classes. The normalized lists contain the subgroup normalizer $\mathrm{Nor}_G S$ ($\mathrm{Nor}_G S$ is the largest subgroup of G for which S is an invariant subgroup) for each subgroup representative. The advantage of having normalized lists is that the problem of merging several different sublists (e.g., the lists of all subgroups of each maximal subgroup of G) into a single overall list becomes greatly simplified. The method is then applied to find all closed connected subgroups of the two de Sitter groups O(3,2) and O(4,1). The classification group in each case is the group of inner automorphisms.

I. INTRODUCTION

A series of earlier papers has been devoted to the problem of classifying all continuous subgroups of Lie groups of interest in physics and finding their invariants.[1–8] All maximal solvable subgroups of the $SU(p,q)$ and $SO(p,q)$ groups have been found,[1,2] as have those of all other classical real Lie groups.[9] All continuous subgroups of the Poincaré group have been listed,[3] separated into isomorphism classes, and their invariants have been found.[6] The subgroups of the similitude groups of the four-dimensional[4] and three-dimensional[8] Minkowski space (the similitude group is the Poincaré group extended by dilations) were obtained. All subgroups of the O(4, 1) de Sitter group were constructed.[5] A general method for calculating the invariants of an arbitrary Lie algebra was developed and applied to find the invariants of all real Lie algebras of dimension $d \leq 5$ and all nilpotent Lie algebras of dimension $d = 6$. The invariants thus obtained can be polynomials in the generators of the group (these are Casimir operators), rational functions of the generators (lying in the quotient field of the enveloping algebra), or general invariants (arbitrary continuously differentiable functions of the generators).[7] Our immediate aim is to provide a list of all continuous closed subgroups of the conformal group of space–time [locally isomorphic to O(4, 2) and SU(2, 2)]. A crucial step in this direction is to find all the maximal subgroups of the conformal group and to classify their subgroups.

The present article is part of the above program, in that we obtain here all the closed continuous subgroups of O(3, 2) and also present some new results on the subgroups of O(4, 1).

[a)]Work partially supported by National Research Council of Canada and by NATO.

Our interest in the subgroups of Lie groups has been motivated earlier,[1–8] both from the physical and mathematical point of view. Let us just mention several different directions of research for which a classification of the subgroups of a given group is important:

1. A systematic study of symmetry breaking in physics. Two recent papers have been devoted to symmetry breaking in nonrelativistic quantum mechanics, making use of the knowledge of the subgroups of the Euclidean group E(3) for the time independent Schrödinger equation[10] or those of the Schrödinger group Sch_1 for the time dependent one.[11]

2. Separation of variables and symmetry properties of differential equations.[10–18]

3. Construction of specific bases for the representation theory of Lie groups.[19,20]

4. The theory of canonical transformations.[18,20–25]

5. Group theoretical expansions of physical quantities, specially scattering amplitudes.[12,26–29]

The de Sitter groups are interesting as two of the maximal subgroups of the conformal group O(4, 2) but they are also of considerable interest in their own right. They occur in relativistic cosmology as groups of motion of four-dimensional spaces of constant nonzero curvature.[30–32] They play the role of conformal groups of three-dimensional Euclidean or pseudo-Euclidean spaces. They have made their appearance in many roles in elementary particle physics, the general theory of relativity, atomic physics, special function theory, and other fields. For a brief review with numerous references we refer the reader to a previous article.[5]

In Sec. II we discuss some relevant properties of the groups $O(p,q)$ in general and O(3, 2) and O(4, 1) in

particular. The principles and methods of classification of subgroups are discussed in Sec. III, paying attention to simultaneous classifications under different locally isomorphic groups, to the procedure of finding subgroups of semisimple groups, in particular irreducibly imbedded subgroups, and to the construction of "normalized lists of subalgebras" (containing the normalizer of each subalgebra together with the subalgebra). In Sec. IV we obtain a list of representatives of the maximal subgroups of O(3, 2). All subgroups of O(3, 2) are classified in Sec. V, where we first classify the subalgebras of each maximal sublagebra of LO(3, 2) under the corresponding maximal subgroup of O(3, 2) and then merge all the lists of subgroups into an overall list of subgroups of O(3, 2). Section VI is devoted to a summary and extension of previously obtained results on subgroups of O(4,1). In Sec. VII we discuss the completeness of the sets of commuting operators provided by the invariants of subgroups in each of the found subgroup chains and the possible existence and form of missing label operators. The conclusions and future outlook are presented in the final Sec. VIII. Section III is mathematically the most demanding one. Readers interested only specifically in the results on the subgroups of O(3, 2) and O(4, 1) can skip this section and obtain all relevant information from Secs. IV—VII and Tables IV—XI.

II. PSEUDO-ORTHOGONAL GROUPS O(p, q)

A. The general case

The group $O(p,q)$ where p and q are integers satisfying $p \geq q \geq 0$, is the closed linear group of all matrices M of degree $p+q$ over the field of real numbers $\mathbb{R}$ satisfying the matrix equation

$$MD_{p,q}M^T = D_{p,q}, \tag{2.1}$$

where M^T is the matrix transposed to M and

$$D_{p,q} = \begin{pmatrix} I_p & \\ & -I_q \end{pmatrix} \tag{2.2}$$

(I_p is the identity matrix of degree p).

The group $O(p,q)$ is thus the group of linear transformations of a $(p+q)$-dimensional real vector space V leaving the real quadratic form

$$x^2 = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2 \tag{2.3}$$

invariant for $x \in V$.

In topological discussions of the $O(p,q)$ groups a decisive role is played by two multiplicative invariants that can be formed for each element $g \in O(p,q)$. The first one is the *determinant* $\det g$, satisfying

$$\det g = \pm 1, \quad \det g_1 g_2 = \det g_1 \det g_2 \tag{2.4}$$

[for any two elements g_1 and g_2 of $O(p,q)$].

The second invariant is the *spinor norm*. The spinor norm of g is defined to be $\mathrm{spn}\, g = +1$ if g is in the identity component of $O(p,q)$ which we denote by $SO_0(p,q)$ (then also $\det g = 1$), or if $\det g = -1$ and the product of g with the particular member

$$M_1 = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix} \tag{2.5}$$

of $O(p,q)$ is not in $SO_0(p,q)$. Otherwise the spinor norm is $\mathrm{spn}\, g = -1$. This definition is not an arbitrary one, since $SO_0(p,q)$ is a normal subgroup of $SO(p,q)$ and gM_1 can hence be replaced without ambiguity by $M_1 g$ or even by any product gM or Mg with M in the same coset modulo $SO_0(p,q)$ as M_1. This definition already makes it evident that the elements g of spinor norm $\mathrm{spn}\, g = 1$ form a normal subgroup $O_1(p,q)$ of $O(p,q)$.

If $q = 0$, then $SO(p,q) = SO(p)$ is compact and connected and in this case the spinor norm is always 1. If both p and q are natural numbers, then $SO(p,q)$ is not connected and elements of spinor norm -1 exist. It was already known in the nineteenth century[33] that every element of $O(p,q)$ is a product of reflections, i.e., of involutory mappings leaving a $(p+q-1)$-dimensional space invariant. Each reflection $p \in O(p,q)$ transforms a vector $x \in V = \mathbb{R}^{(p+q)\times 1}$ according to

$$x' = x - \frac{2}{(u^T D_{p,q} u)} (u^T D_{p,q} x)\, u, \tag{2.6}$$

i.e., it is a linear transformation ρ_u depending on an anisotropic element $u \in \mathbb{R}^{(p+q)\times 1}$ (with $u^T D_{p,q} u \neq 0$). The element u is determined uniquely by the reflection up to a factor of proportionality. The sign of $u^T D_{p,q} u$ is thus uniquely determined and the reflection is said to be positive or negative, depending on the sign of $u^T D_{p,q} u$. Note that M_1 is a negative reflection if $q > 0$.

The multiplicative character of the spinor norm and also the role of M_1 is elucidated by the theorem that the spinor norm of a matrix g is equal to (-1) to the power given by the number of negative reflections occurring in some presentation of g as a product of reflections.

We have thus obtained an algebraic criterion for a topological property, but we still do not have a simple method for calculating the spinor norm. Such a method has been obtained some time ago[34,35] when Lipschitz's theorem was reproven and it was shown that the spinor norm of g is equal to the sign of the determinant $\det(g + I_{p+q})$ if this determinant is not zero,

$$\mathrm{spn}\, g = \mathrm{sign}\, \det(g + I_{p+q}). \tag{2.7}$$

If, however, the matrix $g + I_{p+q}$ is singular, then a minimal exponent μ exists for which the matrices $(I_{p+q} + g)^\mu$ and $(I_{p+q} + g)^{\mu+1}$ have the same rank. We can then decompose the $(p+q)$-dimensional column space into the direct sum

$$\mathbb{R}^{(p+q)\times 1} = R_1 + R_2 \tag{2.8}$$

of the eigenspace of g corresponding to the eigenvalue -1,

$$R_1 = \ker(g + I_{p+q}) \tag{2.9}$$

and the orthogonal space

$$R_2 = 2^{-\mu} (I_{p+q} + g)^\mu \mathbb{R}^{(p+q)\times 1}. \tag{2.10}$$

An $\mathbb{R}$-basis $v_{j1}, \ldots, v_{jn_j}$ $(j = 1, 2)$ can be computed for

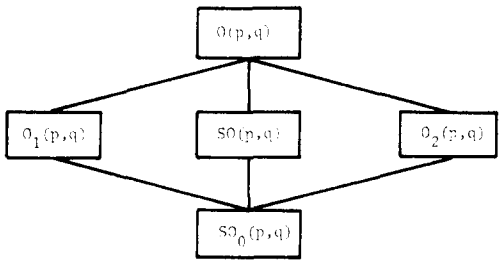


FIG. 1. Groups locally isomorphic to $O(p,q)$ (descending lines indicate subgroup inclusions).

each of these two spaces using well known methods of linear algebra.

The *discriminant* of R_1 is the number

$$d(R_1) = \det(v_{1i}^T D_{p,q} v_{1k}), \tag{2.11}$$

which is uniquely determined up to a square factor and is nonzero because of the orthogonality of the decomposition given above.

Both linear spaces R_1 and R_2 are invariant under g so that the determinant of g restricted to R_2 can be evaluated by means of the action of g on the $\mathbb{R}$-basis $v_{21}, \ldots, v_{2n_2}$ of R_2. The following formula holds for the spinor norm of g:

$$\operatorname{spn} g = \operatorname{sign}\{d(R_1) \det[I_{n_2} + g \,|\, R_2]\} \tag{2.12}$$

or in more detail,

$$\operatorname{spn} g = \operatorname{sign}\{d[\ker(I+g)^\mu] \det[I+g \,|\, (I+g)^\mu \, \mathbb{R}^{(p+q)\times 1}]\}. \tag{2.13}$$

As an example, consider the $O(3,2)$ element

$$g = \begin{bmatrix} -1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & -1 \end{bmatrix}.$$

In this case we have $\mu = 1$, the discriminant of $\ker(I+g)$ is $d(R_1) = -1$, and $I+g|R_2$ restricts to a linear transformation of $(I+g)\mathbb{R}^{5\times 1}$ with identity matrix. Hence we have $\operatorname{spn} g = \operatorname{sign}\{d(R_1) \det I\} = -1$.

For all natural numbers p and q the corresponding orthogonal groups $O(p,q)$ are closed $\mathbb{R}$-linear groups with four components forming a Klein 4-group as a factor group over the identity component of $SO_0(p,q)$. There are precisely five closed subgroups of $O(p,q)$ that are locally isomorphic to $O(p,q)$. They form the Hasse diagram[36] of a Klein 4-group (see Fig. 1). The three subgroups of index 2 can be characterized as follows:

(a) $SO(p,q)$ consists of all elements g of $O(p,q)$ with $\det g = 1$.

(b) $O_1(p,q)$ consists of all elements g of $O(p,q)$ with $\operatorname{spn} g = 1$.

(c) $O_2(p,q)$ consists of all elements g of $O(p,q)$ with $(\det g)(\operatorname{spn} g) = 1$.

The elements of $SO_0(p,q)$ are characterized as the elements g of $O(p,q)$ with $\det g = 1$ and $\operatorname{spn} g = 1$.

Many and perhaps all of these groups are of physical significance. Indeed, consider, e.g., the proper orthochronous Lorentz group $SO_0(3,1)$. If we extend this group by parity P we obtain $O_1(3,1)$, if we extend it by time reversal T we obtain $O_2(3,1)$, and if we extend it by the PT operation we obtain $SO(3,1)$. If all these operations are included we obtain the group $O(3,1)$. Similar comments could be made for the de Sitter groups $O(4,1)$ and $O(3,2)$, as well as the conformal group $O(4,2)$.

For $p > q$ the group $O(p,q)$ is the largest linear group of degree $p+q$ normalizing $SO_0(p,q)$. In the special case when $p = q$ the normalizer of $SO_0(p,p)$ is larger, namely it is the semidirect product of the group of order 2 generated by the involutory matrix

$$K_{2p} = \begin{bmatrix} & & & 1 & 1 \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix} \tag{2.14}$$

of degree $2p$ with the group $O(p,p)$. This group can be characterized by the fact that it leaves the square of the quadratic form $x^2 = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{2p}^2$ invariant. Indeed the matrix K_{2p} itself changes the sign of this form,

$$K_{2p} D_{p,p} K_{2p} = -D_{p,p}, \tag{2.15}$$

so that timelike vectors are transformed into spacelike ones and vice versa, while lightlike vectors are carried into lightlike ones.

The factor group of the extended group $\langle K_{2p}\rangle \square O(p,q)$ over its identity component $SO_0(p,q)$ is a dihedral group of order 8.

B. The de Sitter group O(3, 2)

Everything said above for $O(p,q)$ groups naturally also holds for $O(3,2)$. For further convenience we shall make some general statements more explicit. The Lie algebra of $O(p,q)$ will be denoted $LO(p,q)$ and consists of real matrices X satisfying

$$X D_{p,q} + D_{p,q} X^T = 0. \tag{2.16}$$

In particular an element of $LO(3,2)$ satisfying (2.16) with $p=3$, $q=2$ can be written in the form

$$X = \begin{bmatrix} 0 & a & b & e & f \\ -a & 0 & c & g & h \\ -b & -c & 0 & j & k \\ e & g & j & 0 & d \\ f & h & k & -d & 0 \end{bmatrix}, \tag{2.17}$$

where $a, b, \ldots, k$ are real parameters. A convenient basis for this Lie algebra consists of the ten matrices $A, B, C, D, E, F, G, H, J$, and K obtained from (2.17) by specializing respectively $a, b, c, d, e, f, g, h, j$, or k to be equal to 1, and all other parameters equal to zero.

A different but equivalent realization of $LO(3,2)$ is obtained by introducing the matrix

$$K_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.18}$$

TABLE I. Commutation relations between O(3,2) generators in the $D_{3,2}$ realization.

	A	B	C	D	E	F	G	H	J	K
A	0	$-C$	B	0	$-G$	$-H$	E	F	0	0
B	C	0	$-A$	0	$-J$	$-K$	0	0	E	F
C	$-B$	A	0	0	0	0	$-J$	$-K$	G	H
D	0	0	0	0	$-F$	E	$-H$	G	$-K$	J
E	G	J	0	F	0	D	A	0	B	0
F	H	K	0	$-E$	$-D$	0	0	A	0	B
G	$-E$	0	J	H	$-A$	0	0	D	C	0
H	$-F$	0	K	$-G$	0	$-A$	$-D$	0	0	C
J	0	$-E$	$-G$	K	$-B$	0	$-C$	0	0	D
K	0	$-F$	$-H$	$-J$	0	$-B$	0	$-C$	$-D$	0

TABLE III. Commutation relations between O(4,1) generators in the $D_{4,1}$ realization.

	A	B	C	D	E	F	G	H	J	K
A	0	$-C$	B	$-E$	D	0	$-H$	G	0	0
B	C	0	$-A$	$-F$	0	D	$-J$	0	G	0
C	$-B$	A	0	0	$-F$	E	0	$-J$	H	0
D	E	F	0	0	$-A$	$-B$	$-K$	0	0	G
E	$-D$	0	F	A	0	$-C$	0	$-K$	0	H
F	0	$-D$	$-E$	B	C	0	0	0	$-K$	J
G	H	J	0	K	0	0	0	A	B	D
H	$-G$	0	J	0	K	0	$-A$	0	C	E
J	0	$-G$	$-H$	0	0	K	$-B$	$-C$	0	F
K	0	0	0	$-G$	$-H$	$-J$	$-D$	$-E$	$-F$	0

and requiring

$$X_K K_5 + K_5 X_K^T = 0. \tag{2.19}$$

We have

$$K_5 = Z D_{3,2} Z^{-1}, \quad X_K = Z X Z^{-1} \tag{2.20}$$

with

$$Z = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.21}$$

A general LO(3,2) matrix in this realization is

$$X_K = \begin{pmatrix} -f & l & m & n & 0 \\ p & -g & q & 0 & -n \\ r & s & 0 & -q & -m \\ t & 0 & -s & g & -l \\ 0 & -t & -r & -p & f \end{pmatrix}, \tag{2.22}$$

where again all entries are real numbers and we introduce a basis $F, G, L, M, N, P, Q, R, S$, and T as in (2.17).

The commutation relations for each of the bases are given in Table I and II.

The involutory matrix K_5 of degree 5 (2.18) does not belong to O(3,2) and does not normalize $SO_0(3,2)$. It does however transform an LO(2,2) subalgebra of O(3,2) into itself and can thus play a useful role.

C. The de Sitter group O(4, 1)

An element X of LO(4,1) satisfying

$$X D_{4,1} + D_{4,1} X^T = 0 \tag{2.23}$$

can be written as

$$X = \begin{pmatrix} 0 & a & b & d & g \\ -a & 0 & c & e & h \\ -b & -c & 0 & f & j \\ -d & -e & -f & 0 & k \\ g & h & j & k & 0 \end{pmatrix}. \tag{2.24}$$

A convenient basis of the considered five-dimensional representation again consists of the matrices $A, B, \ldots, K$, this time obtained from (2.24) by setting the corresponding small letter equal to 1 and all others to 0. The corresponding commutation relations are given in Table III.

III. PRINCIPLES AND METHODS OF CLASSIFICATION

A. General comments

In order to classify the closed subgroups of a Lie group G it is necessary to construct a representative set $R_1(G)$ of the G conjugacy classes of closed subgroups of G. This task can be reduced to the following tasks:

A. To establish a representative set $R(G)$ of the G conjugacy classes of the closed connected subgroups of G.

B. To establish a representative set $RD(H)$ of the H conjugacy classes of discrete subgroups of H for certain Lie groups H attached to G.

The tasks A and B are resolved as follows.

For each member S of the set $R(G)$ we form its normalizer in G:

$$\mathrm{Nor}_G(S) = \{g \mid g \in G \text{ and } gSg^{-1} = S\}, \tag{3.1}$$

TABLE II. Commutation relations between O(3,2) generators in the K_5 realization.

	F	G	L	M	N	P	Q	R	S	T
F	0	0	$-L$	$-M$	$-N$	P	0	R	0	T
G	0	0	L	0	$-N$	$-P$	$-Q$	0	S	T
L	L	$-L$	0	0	0	$-F+G$	M	$-S$	0	0
M	M	0	0	0	0	$-Q$	$-N$	$-F$	L	S
N	N	N	0	0	0	0	0	Q	$-M$	$-F-G$
P	$-P$	P	$F-G$	Q	0	0	0	0	$-R$	0
Q	0	Q	$-M$	N	0	0	0	P	$-G$	$-R$
R	$-R$	0	S	F	$-Q$	0	$-P$	0	T	0
S	0	$-S$	0	$-L$	M	R	G	$-T$	0	0
T	$-T$	$-T$	0	$-S$	$F+G$	0	R	0	0	0

a closed subgroup of G containing S as a normal subgroup and such that the factor group $H = \mathrm{Nor}_G(S)/S$ is a Lie group.

Each member T of $RD(H)$ consists of certain cosets of $\mathrm{Nor}_G(S)$ modulo S, the union of which is a closed subgroup S_1 of G with S as its identity component. All S_1's derived from all subgroups S in the list $R(G)$ form the derived list $R_1(G)$. Indeed, if two entries S_1 and S_1' of $R_1(G)$ are G conjugate, then so are their identity components S and S'. But S and S' are by construction members of $R(G)$. Since they are G conjugate it follows that they are equal. Hence the element g of G for which $gS_1g^{-1} = S_1$ belongs to $\mathrm{Nor}_G(S)$. By construction S_1/S and S_1'/S are H conjugate members of $RD(H)$. Hence they are equal.

In other words, any two G conjugate members of $R_1(G)$ are equal. If, on the other hand, for any closed subgroup S_1' of G its identity component S' is G conjugate to some member S of $R(G)$, say $gS'g^{-1} = S$, for some $g \in G$, then $T' = gS_1'g^{-1}/S$ is a discrete subgroup of $H = \mathrm{Nor}_G(S)/S$. Hence T' is an H conjugate of some member T of $RD(H)$, say

$$\bar{h}T'\bar{h}^{-1} = T$$

with $\bar{h} = h/S$, $h \in \mathrm{Nor}_G(S)$. Hence the G conjugate $hgS_1'(hg)^{-1}$ of S_1' belongs to $R_1(G)$.

The task of establishing a list $R(G)$ of the G conjugacy classes of closed connected subgroups of G is equivalent to the task of establishing a representative set $R(L, G)$ of the G conjugacy classes of subalgebras X of the Lie algebra LG of G. However, only those subalgebras X should be included for which the connected subgroup $\langle \exp X \rangle$ of G, generated by the exponentials of the elements of X, is a closed subgroup of G. Indeed, if $R(G)$ is given, then any member S of $R(G)$ can be written in the form $\langle \exp LS \rangle$, where LS, is the Lie algebra corresponding to the closed connected subgroup S of G. Note that any closed subgroup S of G is a Lie group with a Lie algebra LS interpreted to be a subalgebra of LG. For any element g of G we have $L(gSg^{-1}) = gLSg^{-1}$.

The situation can be summarized as follows. The representative set of Lie algebras

$$R(L, G) = \{LS \,|\, S \in R(G)\} \tag{3.2}$$

is a subset of all $\mathbb{R}$ Lie subalgebras of LG and consists of real Lie algebras X with the following properties:

(a) The Lie group $S = \langle \exp X \rangle$ is a closed connected subgroup of G.

(b) Two members of $R(L, G)$ are G conjugate only if they are equal.

(c) Any $\mathbb{R}$ subalgebra Y of LG for which $\exp Y$ is closed is G conjugate to some member of $R(L, G)$.

The converse is also true and follows directly from the rule

$$g(\exp X)g^{-1} = \exp(gXg^{-1})$$

for $g \in G$ and X an $\mathbb{R}$ subalgebra of LG.

The symbol $R(L, G)$ in this article denotes a representative set of subalgebras of the Lie algebra LG, classified up to conjugacy under the group G. In cases when more than one Lie group G corresponds to the algebra LG, the classifying group G must be further specified (see below).

Let us stress here that $\mathbb{R}$ subalgebras of LG may exist for which the corresponding connected Lie groups are not closed. Indeed this happens for the de Sitter group $O(3, 2)$ and actually for all $O(p, q)$ groups with $p \geq q \geq 0$ and $p \geq 4$ or $q \geq 2$. Indeed, consider, e.g., the one-dimensional subalgebra generated by

$$X = A + dD \tag{3.3}$$

in the notation of (2.17). If d is a rational real number then $\exp X$ is a closed connected one-dimensional Lie group. On the other hand, if d is irrational, then $\exp X$ is not closed and its G closure is the two-dimensional compact Lie group corresponding to the two-dimensional Lie algebra $\{A, D\}$. Notice that both generators A and D correspond to compact subgroups of $O(3, 2)$ and that they commute.

More generally we state without proof the following theorem.

Theorem: The connected subgroup $\langle \exp L \rangle$ of $GL(n, \mathbb{C})$ generated by the exponentials of the matrices of an $\mathbb{R}$ Lie subalgebra $L\mathbb{C}^{n\times n}$ is closed relative to $GL(n, \mathbb{C})$ in the natural topology precisely if the $\mathbb{R}$ dimension of a maximal Abelian compact subalgebra S of L coincides with the maximal number of rationally independent weights of its natural representation of degree n over $\mathbb{C}$. Note that all maximal Abelian compact subalgebras of L are conjugate under $\langle \exp L \rangle$.

For a proof we refer the reader to lecture notes.[36]

B. Classification of subalgebras when conjugacy is considered under different but locally isomorphic groups

In Sec. I we discussed the five different groups locally isomorphic to $O(p, q)$. It would be impractical to provide subalgebra lists $R(L, G)$ corresponding to conjugacy under each of these groups separately. This gives rise to the problem of providing a subalgebra list $R(L, G)$ for one of these groups in such a manner that the classification for any of the other groups can be easily read off from the given list.

A very general and widely applicable principle of classification can be stated as a lemma.

Lemma 1: Given a permutation representation Π of the group G on the set M, a representative set $R(G, \Pi)$ of the G orbits of M and a subgroup G_1 of G, a representative set of the G_1 orbits of M is derived from $R(G, \Pi)$ in the form

$$R(G_1, \Pi \,|\, G_1) = \{\Pi(x)\, u \,|\, u \in R(G, \Pi) \text{ and } x \in R[G \bmod(G_1, G_u)]\}.$$

Here $G_u = \{g \,|\, g \in G \text{ and } \Pi(g)\, u = u\}$ denotes the stabilizer of u in G; (G_1, G_u) denotes the double module of G formed by the two subgroups G_1 and G_u, and $R(G \bmod(G_1, G_u))$ denotes a representative set of the double cosets $G_1 x G_u$ such that

$$G = \bigcup_{x \in R[G \bmod(G_1, G_u)]} G_1 x G_u$$

and such that the nonemptiness of the intersection

$$G_1 x G_u \cap G_1 y G_u$$

($x, y \in R[G \bmod(G_1, G_u)]$) always implies $x = y$.

Proof: If u_1 and v_1 are two G_1 conjugate elements of $R(G_1, \Pi \mid G_1)$, then the equations

$$u_1 = \Pi(x)\, u, \quad v_1 = \Pi(y)\, v = \Pi(g_1) u_1$$

with u, v in $R(G, \Pi)$, x, y in G, g_1 in G_1 hold. Hence

$$v = \Pi(y^{-1} g_1 x)\, u, \quad v, u \in R(G, \Pi)$$

and hence by construction $v = u$, $y^{-1} g_1 x \in G_u$, $y \in G_1 x G_u$ and thus $y = x$. Conversely, any element v of M is of the form $\Pi(g)u$ for some g of G and u of $R(G, \Pi)$. The element g is of the form $g_1 x h$ with some $x \in R(G \bmod(G_1, G_u))$, $g_1 \in G_1$, and $h \in G_u$. Hence

$$v = \Pi(g_1)\, \Pi(x)\, \Pi(h)\, u = \Pi(g_1)\, \Pi(x)\, u$$

with $\Pi(x)\, u$ in $R(G_1, \Pi \mid G_1)$. Thus $R(G_1, \Pi \mid G_1)$ is indeed a representative set of the G_1 orbits of M and Lemma 1 is demonstrated.

As an application of Lemma 1 set $G = O(3,2)$, $G_1 = \mathrm{SO}_0(3,2)$, let M be the set of all $\mathbb{R}$ subalgebras X of LG for which $\exp X$ is closed, and let Π be the permutation representation of G on M that maps the element X of M on gXg^{-1} for $g \in G$. Then the representative set $R(G, \Pi)$ coincides with $R(L, G)$, and $R(G_1, \Pi \mid G_1)$ coincides with $R(L, G_1)$. We observe that in this case the stabilizer G_X is the normalizer

$$\mathrm{Nor}_G X = \{ g \mid g \in G \text{ and } gXg^{-1} = X \}$$

of X under G, a closed subgroup of G. We also observe that G_1 is a normal subgroup of G so that any double coset of G modulo (G_1, G_X) is a right coset of G modulo the subgroup $G_1 G_X$ of G containing G_1.

There are only five such subgroups in the case under consideration; consequently there will be at most five distinct double coset representative lists $R[G \bmod(G_1, G_X)]$. In fact, since $-I_5$ is contained in all normalizers only two distinct cases must be considered:

Case I. $\mathrm{SO}_0(3,2)\mathrm{Nor}_G X = \mathrm{O}(3,2)$

In this case $R[G \bmod(G_1, G_X)]$ consists of the identity matrix I_5 only for all classification groups.

Case II. $\mathrm{SO}_0(3,2)\,\mathrm{Nor}_G X = \mathrm{O}_2(3,2)$

In this case $R[G \bmod(G_1, G_X)]$ consists of I_5 for $G_1 = \mathrm{O}(3,2)$, $\mathrm{SO}(3,2)$ and $\mathrm{O}_1(3,2)$. It consists of I_5 and M_1 for $G_1 = \mathrm{SO}_0(3,2)$ and $\mathrm{O}_2(3,2)$, where M_1 is the diagonal matrix

$$M_1 = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix}$$

(with $\det M_1 = -1$, $\mathrm{spn} M_1 = -1$).

In order to complete the classification task for $R[L, \mathrm{O}(3,2)]$ in such a way that it can be used with a minimum of additional effort to obtain the four derived lists for $\mathrm{SO}(3,2)$, $\mathrm{O}_1(3,2)$, $\mathrm{O}_2(3,2)$, and $\mathrm{SO}_0(3,2)$ it suffices to indicate after each member X of $R[L, \mathrm{O}(3,2)]$ to which of the above cases it belongs. If it is case I then X will figure in the same manner in all five lists. In case II, X will be replaced by two entries, namely X and $M_1 X M_1^{-1}$ in $R[L, \mathrm{O}_1(3,2)]$ and $R[L\mathrm{SO}_0(3,2)]$, but figures only as X in $R[L, \mathrm{SO}(3,2)]$ and $R[L, \mathrm{O}_2(3,2)]$.

On the other hand, we can start the classification by providing the complete list $R[L, \mathrm{SO}_0(3,2)]$, using conjugation under the connected group only. The list can be organized into subsets of *singlets* (single members) and *doublets* (ordered pairs). The singlets indicate case I whereas the doublets are of the form X, $M_1 X M_1^{-1}$ and correspond to case II. The singlets figure in the same manner in all five conjugacy lists. The doublets figure in the $\mathrm{SO}_0(3,2)$ and $\mathrm{O}_1(3,2)$ lists, but the second term should be dropped in the $\mathrm{O}_2(3,2)$, $\mathrm{SO}(3,2)$, and $\mathrm{O}(3,2)$ lists. If a subalgebra X depends on a parameter or parameters, then the range of these parameters may depend on the conjugating group. All the above comments apply in this case too. Thus, in the singlet case the range will be the same for all five groups, in the doublet case the range will be cut in half for the groups $\mathrm{O}(p,q)$, $\mathrm{O}_2(p,q)$, and $\mathrm{SO}(p,q)$ but will remain the same for $\mathrm{O}_1(p,q)$ [as for $\mathrm{SO}_0(p,q)$].

C. Classification procedure for semisimple Lie algebras

When classifying the subalgebras of a semisimple Lie algebra LG we are usually interested in determining the conjugacy classes of subalgebras of LG under some group S of automorphisms of LG. Our first step is to establish a list of representatives of the S conjugacy classes of all maximal subalgebras of LG. This makes it possible to make use of already existing classifications of subalgebras of lower dimensional Lie algebras.

To find representatives of all maximal $\mathbb{R}$ subalgebras of the given semisimple Lie algebra LG we make use of a convenient finite dimensional faithful matrix representation of LG. For the classical real semisimple Lie algebras this would usually be the defining representation, e.g., for $\mathrm{LO}(p,q)$ the $(p+q)$-dimensional real matrices X, satisfying (2.16), or for $\mathrm{LSU}(p,q)$ the complex $(p+q)$-dimensional matrices X satisfying

$$X D_{p,q} + D_{p,q} X^* = 0 \tag{3.4}$$

(X^* is the Hermitian conjugate of X), etc.

In the chosen representation the subalgebras of LG will be of two types, namely imbedded *reducibly* or imbedded *irreducibly*. The two types will be treated separately and differently. It should, however, be stressed that there is no fundamental difference between the two: The reducibility or irreducibility characterizes the chosen representation rather than the subalgebras of LG.

1. Reducible subalgebras

These leave invariant a certain nontrivial linear subspace of the representation space. To find all reducible subalgebras of LG it is sufficient to classify all vector subspaces into orbits under the group G, to choose a convenient representative for each orbit and then for each representative to find the algebra of matrices leaving invariant the representative.

We are aided in this endeavor by the Witt mapping theorem[37] for quadratic spaces which enlarges conven-

iently on the earlier Sylvester theory. It is summarized as follows.

A *quadratic space* (L,f) over a field F not of characteristic 2 is defined as a linear space L over F equipped with a symmetric bilinear form f.

Any F-linear subspace S of L is a quadratic space relative to the restriction $f|S\times S$ of f to a symmetric bilinear form on S. This is a subspace of (L,f).

Examples: (1) $f=0$: L is called an *isotropic quadratic space*.

(2) $\dim_F L=1$, $L=Fu$, $u\neq 0$, $f(\xi u,\eta u)=\gamma\xi\eta$ ($\xi,\eta\in F$; γ is a constant of F).

(3) $\dim_F L=2$, $L=Fu_1+Fu_2$, $f(\xi_1u_1+\xi_2u_2,\eta_1u_1+\eta_2u_2)=\xi_1\eta_1-\xi_2\eta_2$, L is called a *hyperbolic quadratic space*.

(4) $\dim_F L=n<\infty$, $L=Fu_1+Fu_2+\cdots+Fu_n$, $f(\sum_{i=1}^n\xi_iu_i,\sum_{k=1}^n\eta_ku_k)=\sum_{i,k}\alpha_{ik}\xi_i\eta_k$ ($\xi_i,\eta_i\in F$; $i,k=1,2,\ldots,n$), where $A=(\alpha_{ik})=A^T\in F^{n\times n}$ is a symmetric matrix of degree n over F associated with f (relative to the basis $u_1,\ldots,u_n$). Its determinant is said to be a *discriminant* of f. If another F basis $v_k=\sum_{i=1}^n\sigma_{ik}u_i$ [$k=1,2,\ldots,n$; $S=(\sigma_{ik})\in \mathrm{GL}(n,F)$] of L is chosen then the matrix $B=(\beta_{ik})=S^TAS$ is associated with f relative to it. Its determinant is equal to the determinant of A multiplied by a nonzero square element of F that can be chosen freely. There is always a basis for which B is diagonal.

(5) L is called a *core space* if it contains no isotropic subspace $\neq 0$. If $F=\mathbb{R}$ and L is finite dimensional with diagonal matrix $B=(\beta_{ii}\delta_{ik})$ associated to f, then L is a core space precisely if all β_{ii} are nonzero and of the same sign. L is said to be *positive definite* or *negative definite* depending on the sign.

Two quadratic spaces (X_1,f_1), (X_2,f_2) are said to be *isomorphic* if there is an F-linear isomorphism θ of X_1 on X_2 for which $f_2(\theta x,\theta y)=f_1(x,y)$ $(x,y\in X_1)$. The theory is concerned with the classification of the quadratic spaces up to isomorphy over F.

For each subset X of a quadratic space L the subset

$$X^\perp=\{u\,|\,u\in L \text{ and } \forall v(v\in X\Rightarrow f(u,v)=f(v,u)=0)\}$$

is a linear subspace ("X perp") often called the linear subspace of L orthogonal to X. The quadratic space L is said to be *nondegenerate* if $L^\perp=0$.

The factor space of a quadratic space L over $L^\perp$ is a nondegenerate quadratic space with respect to the f-induced symmetric bilinear form f defined by setting $f(x/L^\perp,y/L^\perp)=f(x,y)$. L is the direct sum of $L^\perp$ and a representative subspace M of L over $L^\perp$ which is isomorphic to $L/L^\perp$ no matter how M is chosen. For any nondegenerate finite dimensional subspace S of L there holds the direct decomposition $L=S\overset{\circ}{+}S^\perp$. Here $S^\perp$ is said to be the *orthogonal complement* of S in L. Conversely, if for any subspace S of L we have $L=S\overset{\circ}{+}S^\perp$, then S is nondegenerate. If L is nondegenerate finite dimensional, then the mapping of S on $S^\perp$ is an involutory correlation of the subspaces of L.

Isotropic subspaces of dimension greater than 0 are degenerate. The zero-dimensional isotropic space is said to be nondegenerate by definition. A finite dimensional quadratic space is nondegenerate precisely if its discriminants are nonzero. Hyperbolic spaces are nondegenerate.

Witt's mapping theorem[36,37] states that any isomorphism between two subspaces of a finite dimensional nondegenerate quadratic space (L,f) can be extended to an automorphism of (L,f)

As a consequence of this basic theorem any nondegenerate finite dimensional quadratic space L is the direct sum of a finite number σ of hyperbolic spaces and a core space which is unique up to an automorphism of (L,f). Hence also σ is unique.

If $F=\mathbb{R}$ and $B=(\beta_{ii}\delta_{ik})$ is a diagonal matrix associated with f, then $\Pi_{i=1}^n\beta_{ii}\neq 0$, and the number p of positive signs, q of negative signs among the signs of the n diagonal coefficients β_{ii} adds up to n so that $\sigma=q=\min(p,q)$ and the core form is positive definite (for $p>q$).

For all these statements on quadratic spaces there are corresponding statements for symplectic spaces and Hermitian symmetric spaces, but without restriction on the characteristic of F.

Let us briefly summarize the implications of Witt's theorem relevant for the purposes of this article. Consider the algebra $\mathrm{LO}(p,q)$ acting on $(p+q)$-dimensional real space M (the defining or natural representation). The following statements hold:

1. A subspace $M_r\subset M$ of dimension $r<p+q$ can be spanned by r vectors mutually orthogonal with respect to the form

$$(x,y)=x_1y_1+\cdots+x_py_p-x_{p+1}y_{p+1}-\cdots-x_{p+q}y_{p+q}. \quad (3.5)$$

Put

$$r=r_++r_-+r_0, \quad (3.6)$$

where r_+, r_-, and r_0 are the numbers of positive, negative, and zero length vectors [with respect to (3.5)], in this basis, respectively. These numbers are independent of the choice of basis. Any two subspaces M_r and $\tilde{M}_r$ for which r_+, r_-, and r_0 coincide are mutually conjugate under the group $\mathrm{O}(p,q)$ [and also under $\mathrm{SO}_0(p,q)$].

2. Let M_r be a subspace invariant under a subalgebra A of $\mathrm{LO}(p,q)$. If $r_0\geq 1$, then there always exists an r_0-dimensional isotropic (lightlike) subspace of M_r that is itself invariant under A.

3. If $r_+=r_-=1$, then there always exists a one-dimensional isotropic subspace of M, invariant under A.

4. If M_r is invariant under A, the orthogonal space $M_r^\perp$ is also invariant. If M_r is nondegenerate, then $M_r^\perp$ is its orthogonal complement.

2. Irreducible linear Lie algebras of finite degree and zero characteristic

In studying the subalgebras of one of the linear Lie algebras attached to the fundamental groups, e.g., $\mathrm{SU}(2,2)$, often the question comes up which are the ir-

reducible ones. Here LSU(2, 2) is defined as a Lie algebra of dimension 15 over the real number field, but the representation is of degree 4 *over the complex number field.* Most generally we ask what are the Lie algebras L over a field F of zero characteristic that are irreducibly embedded into the Lie algebra $LE^{n\times n}$ of finite degree n over some field extension E of F.

Here irreducibility means that the n-column space $E^{n\times 1}$ over E serving as representation space for L acting on $E^{n\times 1}$ by multiplication on the left contains no E-linear subspace invariant under L other than the trivial subspaces $E^{n\times 1}$ and 0.

According to Ref. 36, any nilideal of the linear Lie algebra EL over E is null. For example $[EL, \mathrm{Rad}(EL)]$ is a nilideal. Hence $[EL, \mathrm{Rad}(EL)]=0$. Since $\mathrm{Rad}(L) \subseteq \mathrm{Rad}(EL)$ it follows that $[L, \mathrm{Rad}(L)]\subseteq[EL, \mathrm{Rad}(EL)]$ and hence

$$[L, \mathrm{Rad}(L)]=0, \tag{3.7}$$

[$\mathrm{Rad}(L)$ denotes the radical of L, i.e., the maximal solvable ideal of L.] We see that $\mathrm{Rad}(L)$ belongs in this case to the center CL of L (for L irreducible). By Levi's theorem it follows that

$$L=CL\dotplus DL, \tag{3.8}$$

where $CL=L_0$ is the center of L and DL is a semisimple Lie algebra. By the first structure theorem of Cartan–Killing

$$DL=[L,L]=\dot{\sum_{i=1}^{s}} L_i, \tag{3.9}$$

where $L_1,\ldots,L_s$ are the minimal ideals of DL. They are finite dimensional simple Lie algebras over F.

According to Schur's lemma the centralizer algebra of L in $E^{n\times n}$,

$$C_{E^{n\times n}}(L)=\{X \mid X\in E^{n\times n} \text{ and } \forall\ Y(Y\in L \Rightarrow XY=YX)\}$$
$$=C_{E^{n\times n}}(EL),$$

is a finite dimensional division algebra over EI_n. It contains EL_0+EI_n. If it is larger than EI_n, then it contains a maximal finite extension E_1 of EI_n of degree $m>1$. The E-linear space $E^{n\times 1}$ also is an E_1-linear space, but of dimension n/m less than n. Since E_1 commutes elementwise with L acting on $E^{n\times 1}$ it follows that L also is irreducibly embedded into the full linear Lie algebra of degree $n'=n/m$ over E_1 so that a degree reduction is obtained when making the transition from E to E_1. Because of the maximality of E_1 we have

$$C_{E_1^{n\times n}}(L)=E_1I_{n'}=C_{E_1^{n'\times n'}}(E_1L). \tag{3.10}$$

This condition is tantamount to the absolute irreducibility of the E_1-linear Lie algebra E_1L. In particular we have

$$L_0\subseteq E_1I_n. \tag{3.11}$$

The obvious relations

$$[L_i, L_k]=\delta_{ik}L_i \quad (1\le i\le k\le s) \tag{3.12a}$$

imply the relation

$$[E_1L_i, E_1L_k]=\delta_{ik}E_1L_i \quad (1\le i\le k\le s)$$

so that

$$E_1L=E_1L_0\dotplus\dot{\sum_{i=1}^{s}} E_1L_i, \tag{3.12b}$$

$$E_1L_0=0 \text{ or } E_1L_0=E_1I_{n'}, \tag{3.12c}$$

$$D(E_1L)=\dot{\sum_{i=1}^{s}} E_1L_i, \tag{3.12d}$$

and the E-linear algebra $D(EL)$ is semisimple and absolutely irreducible.

If $s>1$, then the natural representation Δ of EL over E is equivalent to the Lie–Kronecker product of s absolutely irreducible representations $\Delta_1,\Delta_2,\ldots,\Delta_s$ of EL of degree n_i over E such that

$$n_i>1 \quad (1\le i\le s), \tag{3.13a}$$

$$n'=\prod_{i=1}^{s} n_i, \tag{3.13b}$$

$$\Delta_i(E_1L_k)=0 \text{ if } k\ne i \quad (i,k=1,2,\ldots,s). \tag{3.13c}$$

Thus the general discussion comes to an end.

In practice we have to consider mainly the three cases

I. $F=E=\mathbb{C}$

II. $F=\mathbb{R}$, $E=\mathbb{C}$

III. $F=E=\mathbb{R}$.

Moreover, the problem arises in conjunction with the classification of subalgebras of certain semisimple Lie algebras. Since a semisimple Lie algebra always coincides with its derived algebra, it follows that all matrices of a semisimple linear Lie algebra are of zero trace. Thus we make the additional request

$$\mathrm{Tr}X=0 \tag{3.14}$$

for all X of L. Under these circumstances we find

$L_0=0$ in cases I and II,

$L_0=0$ or L_0 is compact of dimension 1 over $\mathbb{R}$ in case III.

In case L_0 is one-dimensional over $\mathbb{R}$ the matrices of L_0 commute elementwise with L so that the square of each nonzero element of L_0 is of the form negative real number times identity matrix and L is not absolutely irreducible.

Let us consider several special cases.

a. *Application to de Sitter algebras* $LE=\mathrm{LO}(3,2)$ or $\mathrm{LO}(4,1)$: Here we have $F=E=\mathbb{R}$. We observe that $n=5$ is an odd prime number so that the natural representation of L is absolutely irreducible and $s=1$. Also $L_0=0$. In this case L is of dimension less than 10 over $\mathbb{R}$ and must be a simple absolutely irreducible linear Lie algebra of degree 5 over $\mathbb{R}$. Hence $\overline{L}=C\otimes_{\mathbb{R}} L$ is a simple Lie algebra of dimension less than 10 and of degree 5 over C. It can only be of type A_1 so that either $L\underset{\mathbb{R}}{\simeq}\mathrm{LO}(2,1)$ or $L\underset{\mathbb{R}}{\simeq}\mathrm{LO}(3)$. But in the latter case L leaves invariant only one quadratic form up to a factor of proportionality and that is bound to be positive definite. However L also is contained in LG so that L leaves invariant an indefinite quadratic form in five variables. Thus we need to discuss only the case $L\underset{\mathbb{R}}{\simeq}\mathrm{LO}(2,1)$.

Denoting by Δ_3 the natural representation of LO(2, 1) we see that $\Delta_3 \otimes \Delta_3 \cong \Delta_1 \oplus \Delta_3 \oplus \Delta_5$, where Δ_1 is the null representation of LO(2, 1) of degree 1, Δ_5 is an irreducible representation of LO(2, 1) of degree 5 over $\mathbb{R}$. Hence $\Delta_3 \otimes \Delta_3$ leaves invariant a quadratic form of signature $(2 \cdot 2 + 1, 2 \cdot 1 + 2 \cdot 1) = (5, 4)$, $\Delta_1 \oplus \Delta_3$ leaves invariant a quadratic form of signature (2, 2), Δ_5 leaves invariant a quadratic form of signature (3, 2).

Result: There is precisely one proper $\mathbb{R}$ subalgebra of LO(3, 2), up to conjugacy. It is $\mathbb{R}$ isomorphic to LO(2, 1). There is no proper irreducible $\mathbb{R}$ subalgebra of LO(4, 1).

b. *Application to* LO(2, 2): In this case we proceed "structurally." We have

$$\mathrm{LO}(2,2) = \{X \mid X \in \mathbb{R}^{4\times 4} \text{ and } X^T D + DX = 0\}$$

for

$$D = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}.$$

The matrices X of LO(2, 2) are of the form

$$X = \begin{pmatrix} 0 & b & g & h \\ -b & 0 & j & k \\ g & j & 0 & c \\ h & k & -c & 0 \end{pmatrix} \quad (b, c, g, h, j, k \in \mathbb{R}),$$

spanned over $\mathbb{R}$ by the six independent matrices B, C, G, H, J, K which are obtained by specializing b, c, g, h, j, k to be 1 in turn, all other parameters 0. Choosing the new basis

$$A_1 = (J + H)/2, \quad A_2 = (G - K)/2, \quad A_3 = (B + C)/2,$$

$$B_1 = (J - H)/2, \quad B_2 = (G + K)/2, \quad B_3 = (B - C)/2,$$

we verify by direct computation that

$$[A_i, B_k] = 0 \quad (i, k = 1, 2, 3)$$

and that

$$\mathrm{LO}(2,2) = A \oplus B,$$

$$A = \mathbb{R}A_1 + \mathbb{R}A_2 + \mathbb{R}A_3 \cong B = \mathbb{R}B_1 + \mathbb{R}B_2 + \mathbb{R}B_3$$

$$\cong \mathrm{LO}(2,1) \cong \mathrm{LSL}(2, \mathbb{R})$$

We use the Goursat twist method[38] and represent the O(2, 2) conjugacy classes of semisimple subalgebras of O(2, 2) by

(1) LO(2, 2),

(2) A,

(3) $\mathbb{R}(A_1 + B_1) + \mathbb{R}(A_2 + B_2) + \mathbb{R}(A_3 + B_3) \cong \mathrm{LO}(2, 1)$,

(4) $\mathbb{R}(A_1 + B_1) + \mathbb{R}(A_2 - B_2) + \mathbb{R}(A_3 - B_3) \cong \mathrm{LO}(2, 1)$.

Of these (1) is absolutely irreducible semisimple, not simple; (2) is reducible simple with two equivalent constituents; (3) and (4) are reducible simple with two constituents of degree 1 and 3, respectively. Applying the Goursat method we obtain another subalgebra $A \oplus \mathbb{R}B_3$ which is irreducible and not absolutely irreducible.

Result: There is precisely one proper irreducible $\mathbb{R}$ subalgebra of LO(2, 2), up to O(2, 2) conjugacy. It is $\mathbb{R}$ isomorphic to $\mathrm{LO}(2, 1) \oplus \mathrm{LO}(2)$. Under SO(2, 2) conjugacy there are two such subalgebras, the second one being $\mathbb{R}A_3 \oplus B$.

Thus, methods exist for finding all maximal subalgebras, both the irreducible and the reducible ones. Once we have a list of representatives of all maximal subalgebras of LG we proceed to find their subalgebras. For those maximal subalgebras that are semisimple we proceed in the same manner as for the original algebra. For the nonsemisimple maximal subalgebras we use a classification algorithm, making use of cohomology theory, presented in an earlier publication.[3] We thus obtain lists of representatives of all subalgebras of each maximal subalgebra LH_j, each classified under the Lie group H_j generated by the corresponding maximal subalgebra.

Two tasks still remain. The first is to further reduce the representative lists of subalgebras of each maximal subalgebra LH_j, this time making use of conjugation under elements of G, not contained in H_j. This is not hard to do. It is very helpful to construct the eigenvalues and eigenspaces of the matrices corresponding to each subalgebra. If the eigenvalues for different subalgebras of the same dimension coincide and the eigenspaces can be mapped into each other by a g transformation, then such subalgebras must be investigated for conjugacy. The second task is to merge all the different lists of subalgebras for each LH_j into a single list for LG. This involves finding all conjugacies between the different lists and thus making sure that in the final list each subalgebra is represented just once. This task is greatly facilitated by producing "normalized lists of subalgebras."

D. Normalized lists of subalgebras and the merging of several lists into a single one

Let us again consider an arbitrary Lie group G and its Lie algebra LG. We first seek a representative set $R_{00}(L, G)$ of the G conjugacy classes of maximal $\mathbb{R}$ subalgebras of LG up to G conjugacy. Each of the maximal subalgebras LH_j will either be self-normalizing, i.e., we have $L\mathrm{Nor}_G(LH_j) = LH_j$ for LH_j in $R_{00}(L, G)$, or its normalizer will be the entire group G (this is not possible if G is simple). By induction over the dimension we base the construction of the list $R(L, G)$ on the assumption that the lists $R(L, H_j)$ are already known or at any rate easier to deal with than the list $R(L, G)$ would be by a direct approach. We must, however, now face the problem of the mutual overlapping of different lists. It can and does happen that one and the same $\mathbb{R}$ subalgebra X is G conjugate both to a member X_1 of $R(L, H_1)$ and to a member X_2 of $R(L, H_2)$, where LH_j with $j = 1, 2$ are two different G nonconjugate maximal subalgebras of LG. The problem now is to find all such conjugacies and choose the one single most appropriate representative.

In order to alleviate the task of partially merging several smaller lists into one comprehensive list we make use of a normalized list of representatives.

Definition: The representative set $R(L, G)$ of the conjugacy classes of $\mathbb{R}$ subalgebras X of LG under a Lie group G such that $\exp X$ is closed in G is said to be *normalized* if the normalizer

$$\mathrm{Nor}_{LG} X = \{a \mid a \in LG \text{ and } [a, X] \subseteq X\}$$

of X in LG occurs in $R(L, G)$ jointly with X.

We make use of the following lemma.

Lemma 2: For any Lie group G a normalized list $R(L, G)$ exists.

Proof and method of construction: Our immediate goal is to produce a well-ordered list of representatives of the self-normalizers in LG under G conjugacy, say the list $R_0(L, G)$. Its first member will be LG. Before we proceed with the construction of $R_0(L, G)$ we produce a well-ordered list of representatives of the G conjugacy classes of maximal subalgebras of LG that are *not* ideals of LG, say the list $R_M(L, G)$ [if LG is simple, then $R_M(L, G) = R_{00}(L, G)$].

Any maximal subalgebra M of LG that is not an ideal of LG is a self-normalizer in LG. Moreover, it is the Lie algebra associated with $\mathrm{Nor}_G(M)$. By induction over the IR dimension, we may take for granted that for each member M of $R_M(L, G)$ we have available already a well-ordered list of representatives $R_0(M, \mathrm{Nor}_G(M))$ of the $\mathrm{Nor}_G(M)$ conjugacy classes of the self-normalizers in $L\,\mathrm{Nor}_G(M)$.

The desired list $R_0(L, G)$ is obtained as a partial merging of LG with the lists $R_0(M, \mathrm{Nor}_G(M))$, as follows. For the first member of $R_M(L, G)$, say M_1, drop from the list $R_0(M_1, \mathrm{Nor}_G(M_1))$ all those members which have a normalizer larger than themselves in LG. Among the remaining members of the list drop all those that are G conjugate to earlier ones. The remainder is joined to LG to form the new initial segment of $R_0(L, G)$. If $R_M(L, G)$ had just one member, then we have completed our task.

If there is a second member of the list $R_M(L, G)$, say M_2, then we drop again all members of the list $R_0(M_2, \mathrm{Nor}_G(M_2))$ which do not self-normalize in LG. But among the remaining members of the list we drop all those that are G conjugate to earlier members of $R_0(M_2, \mathrm{Nor}_G(M_2))$ or to any member of the initial segment of $R_0(L, G)$. What remains of $R_0(M_2, \mathrm{Nor}_G(M_2))$ is joined to the initial segment of $R_0(L, G)$ forming a new larger initial segment of $R_0(L, G)$, since at least M_2 itself will have survived. Going through the whole well-ordered list $R_M(L, G)$ our screening and partial merging method produces the desired self-normalized representative list $R_0(L, G)$.

We observe that each member X of $R_0(L, G)$ is the Lie algebra associated with $\mathrm{Nor}_G(X)$. By an induction argument over the dimension we assume that there is already available a normalized representative list $R(X, S)$ of the S conjugacy classes for each closed subgroup S of G such that $X = LS \subset LG$. Let X either be a member of $R_0(L, G)$ that is distinct from LG or let X be a proper IR ideal of LG such that $\exp X$ generates a closed subgroup of G. Let $S = \langle \exp X \rangle$.

Our next goal is to produce a normalized sublist $R_1(X, G)$ of $R(X, S)$ representing the G conjugacy classes of the subalgebras Y of X which are connected with X via the LG normalizer chain

$$Y = Y_1 \subset \mathrm{Nor}_{LG}(Y_1) = Y_2 \subset \cdots \subset \mathrm{Nor}_{LG}(Y_{r-1}) = X \tag{3.15}$$

formed by the higher normalizers of X in LG.

Note that an IR ideal of X is S conjugate only to itself. Thus the additional work involved in computing $R_1(X, G)$ is substantially less than we had to do to find $R(X, S)$.

We achieve our goal upon removal from $R(X, S)$, at step (1), all members $Y \neq X$ for which $\mathrm{Nor}_{LG}(Y) \supset X$, at step (2) all members $Y \neq X$ of the reduced list for which $\mathrm{Nor}_{LG}(Y)$ does not belong to the reduced list, and subsequently we continue as at step (2). After a finite number of steps the desired sublist $R_1(X, G)$ emerges. It is normalized because $R(X, S)$ was normalized to begin with.

Now we form the desired list $R(L, G)$ as the direct union of LG, of all lists $R_1(X, G)$ for the proper IR ideals X of LG, and of all lists $R_1(X, G)$ for the members X of $R_0(L, G)$ distinct from LG. Since the application of the conjugation by an element g of G to (3.15) produces the conjugate LG normalizer chain

$$\begin{aligned} gYg^{-1} &= gY_1g^{-1} \subset \mathrm{Nor}_{LG}(gY_1g^{-1}) \\ &= gY_2g^{-1} \subset \cdots \subset \mathrm{Nor}_{LG}(gY_{r-1}g^{-1}) = gXg^{-1}, \end{aligned} \tag{3.16}$$

it is clear that no two distinct members of $R(L, G)$ are G conjugate. On the other hand, any subalgebra Y of LG either is LG in which case it does itself occur in $R(L, G)$, or we have $Y \subset LG$ and in that case there is an LG normalizer chain (3.15) such that either X is a proper ideal of LG or else X is a self-normalizer $\neq LG$. In the first instance Y is a g conjugate (for some element $g \in G$) to a member of $R_1(X, G)$. Hence gYg^{-1} belongs to $R(L, G)$. In the second instance X is g conjugate (for some element $g \in G$) to a member gXg^{-1} of $R_0(L, G)$. Because of (3.16) there is an element s of $S = \langle \exp X \rangle$ such that Y is s conjugate to some member of $R(L, S)$ which survives the screening so as to occur also in $R(L, G)$.

IV. MAXIMAL SUBALGEBRAS OF LO(3, 2)

Following the algorithm discussed in Sec. III we proceed to construct a representative list of O(3, 2) conjugacy classes of maximal subalgebras of LO(3, 2). Since LO(3, 2) is a simple Lie algebra, each maximal subalgebra is self-normalizing. We consider the defining five-dimensional real representation of LO(3, 2) and make use of both realizations discussed in Sec. II B. Thus, we have a real five-dimensional vector space with an invariant quadratic form

$$(x, y) = (xD_{3,2}y) = x_1y_1 + x_2y_2 + x_3y_3 - x_4y_4 - x_5y_5 \tag{4.1}$$

or

$$(x, y)_K = (xK_5y) = x_1y_5 + x_2y_4 + x_3y_3 + x_4y_2 + x_5y_1. \tag{4.2}$$

Realization (4.2) is convenient when we are dealing with lightlike (isotropic) subspaces, otherwise we use realization (4.1).

A. Reducibly imbedded subalgebras

Let us first find all reducibly imbedded subalgebras, leaving (by definition) a certain linear vector space

invariant. We make use of Witt's theorem and characterize the subspaces simply by their dimension r and signature (r_+, r_-, r_0) (with $r=r_++r_-+r_0$). Here r_+, r_-, and r_0 are the numbers of positive length (spacelike), negative length (timelike), and zero length (lightlike or isotropic) vectors in any orthogonal basis for the subspace and the length (x,x) is defined by the form (4.1). Let us choose a convenient representative for each type of subspace and run through all the possibilities.

1. One-dimensional subspaces

A_1. *Timelike* [*signature*(-)]: We choose the basis vector T in the form

$$T=\begin{pmatrix}0\\0\\0\\1\\0\end{pmatrix}. \tag{4.3}$$

The condition $X\{T\}\subseteq\{T\}$, where $\{T\}$ is the vector space spanned by T and X is given by (2.17), implies $e=g=j=d=0$. We thus obtain the LO(3,1) Lie algebra of the homogeneous Lorentz group. The usual basis for this algebra consists of the three rotations

$$L_1=-C,\quad L_2=B,\quad L_3=-A, \tag{4.4a}$$

and three boosts (generators of proper Lorentz transformations)

$$K_1=F,\quad K_2=H,\quad K_3=K. \tag{4.4b}$$

They satisfy the commutation relations

$$[L_i,L_k]=\epsilon_{ikl}L_l,\quad [L_i,K_k]=\epsilon_{ikl}K_l,\quad [K_i,K_k]=-\epsilon_{ikl}L_l. \tag{4.5}$$

A_2. *Spacelike* [*signature* (+)]: We choose the basis vector S in the form

$$S=\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix}. \tag{4.6}$$

The condition $X\{S\}\subseteq\{S\}$ implies $b=c=j=k=0$ in (2.17) and we obtain an LO(2,2) algebra. Making use of the isomorphism $\mathrm{LO}(2,2)\approx\mathrm{LO}(2,1)\oplus\mathrm{LO}(2,1)$ we can choose a basis in the form

$$A_1=\frac{H-E}{2},\quad A_2=\frac{F+G}{2},\quad A_3=\frac{A+D}{2}, \tag{4.7a}$$

$$B_1=\frac{H+E}{2},\quad B_2=\frac{F-G}{2},\quad B_3=\frac{A-D}{2}. \tag{4.7b}$$

The commutation relations are

$$\begin{aligned}&[A_1,A_2]=-A_3,\quad [A_3,A_1]=A_2,\quad [A_2,A_3]=A_1,\\&[B_1,B_2]=-B_3,\quad [B_3,B_1]=B_2,\quad [B_2,B_3]=B_1,\\&[A_i,B_k]=0\quad (i,k=1,2,3).\end{aligned} \tag{4.8}$$

A_3. *Lightlike* [*signature* (0)]: In this case we use the realization (2.22), corresponding to the quadratic form (4.2). A lightlike vector can then be chosen in the form

$$L_K=\begin{pmatrix}1\\0\\0\\0\\0\end{pmatrix}. \tag{4.9}$$

The condition $X_K\{L_K\}\subseteq\{L_K\}$ implies $p=r=t=0$ in (2.22). We obtain a seven-dimensional subalgebra, namely the Lie algebra LSim(2,1) of the similitude group Sim(2,1) of three-dimensional Minkowski space (i.e., the six-dimensional Poincaré group extended by dilations). A convenient physical basis consists of the rotation L_3, two boosts K_1 and K_2, time translations P_0, space translations P_1 and P_2, and the dilation F. In terms of the two matrices (2.17) and (2.22) we can identify the generators as

$$\begin{aligned}&F,\quad L_3=C=\frac{Q-S}{\sqrt{2}},\quad P_0=D+E=\frac{-L+N}{2},\\&K_1=J=\frac{-Q-S}{\sqrt{2}},\quad P_1=-B+K=-\frac{M}{\sqrt{2}}\\&K_2=G,\quad P_2=-A+H=-\frac{L+N}{2}\end{aligned} \tag{4.10}$$

(the matrices F and G figure in both realizations). The commutation relations are

$$\begin{aligned}&[K_1,K_2]=-L_3,\quad [L_3,K_1]=K_2,\quad [L_3,K_2]=-K_1,\\&[L_3,P_0]=0,\quad [L_3,P_1]=P_2,\quad [L_3,P_2]=-P_1,\\&[K_1,P_0]=P_1,\quad [K_1,P_1]=P_0,\quad [K_1,P_2]=0,\\&[K_2,P_0]=P_2,\quad [K_2,P_1]=0,\quad [K_2,P_2]=P_0,\\&[F,L_3]=[F,K_1]=[F,K_2]=[P_\mu,P_\nu]=0,\\&[F,P_\mu]=-P_\mu\quad (\mu,\nu=0,1,2).\end{aligned} \tag{4.11}$$

2. Two-dimensional subspaces

B_1. *Timelike* [*signature* (- -)]: We choose the space in the form

$$\{TT\}=\begin{pmatrix}0\\0\\0\\t\\u\end{pmatrix}, \tag{4.12}$$

where t and u are real, $-\infty<t,u<\infty$. The condition $X\{TT\}\subseteq\{TT\}$ implies $e=f=g=h=j=k=0$ in (2.17) and we obtain the Lie algebra $\mathrm{LO}(3)\oplus\mathrm{LO}(2)$ generating the maximal compact subgroup of O(3,2). A convenient basis is

$$L_1=C,\quad L_2=B,\quad L_3=A,\quad \text{and } D. \tag{4.13}$$

B_2. *Spacelike* [*signature* (+ +)]: We choose the space in the form

$$\{SS\}=\begin{pmatrix}0\\y\\z\\0\\0\end{pmatrix}, \tag{4.14}$$

where y and z are real, $-\infty<y,z<\infty$. The condition $X\{SS\}\subseteq\{SS\}$ implies $a=b=g=j=h=k=0$ in (2.17) and we obtain the algebra $\mathrm{LO}(2)\oplus\mathrm{LO}(2,1)$. A convenient basis is

$$L_3=D,\ K_1=E,\ \ K_2=-F,\ \text{ and } C. \tag{4.15}$$

B_3. *Lightlike* [*signature* (00)]: We use the (2.22) realization of O(3,2) and choose the space in the form

$$\{LL\}_K=\begin{pmatrix} x\\ y\\ 0\\ 0\\ 0\end{pmatrix},\quad -\infty<x,y<\infty. \tag{4.16}$$

The condition $X_K\{LL\}_K\subseteq\{LL\}_K$ implies $r=s=t=0$ in (2.22). We obtain a seven-dimensional Lie algebra generating a group that we shall call the "optical group," Opt(2,1). The reason for this terminology is that when O(3,2) is considered as the conformal group of a (2+1)-Minkowski space, then Opt(2,1) is the subgroup that leaves a lightlike vector space in Minkowski space invariant.[39]

The structure of the Lie group Opt(2,1) is

$$\text{Opt}(2,1)\sim\{D\otimes \text{SL}(2,R)\}\supset W_1, \tag{4.17}$$

where D is a dilation transformation and W_1 is the Weyl group in one dimension (its Lie algebra is isomorphic to the algebra generated by a linear momentum $p=-i\partial/\partial x$, a coordinate x and a constant). The symbol $\supset$ in (4.17) indicates a semidirect product with the invariant subgroup on the right-hand side. A convenient basis for the Lie algebra LOpt(2,1) expressed in terms of the generators X_K and X of (2.22) and (2.17) is

$$\begin{aligned} &W=-\frac{F+G}{2},\quad K_1\equiv\frac{F-G}{2},\quad M=\frac{1}{\sqrt{2}}(B-K),\\ &K_2\equiv\frac{L+P}{2}=-\frac{E+H}{2},\quad Q=\frac{1}{\sqrt{2}}(C-J),\\ &L_3\equiv\frac{L-P}{2}=\frac{A-D}{2},\quad N=\frac{A+D+E-H}{2}. \end{aligned} \tag{4.18}$$

The commutation relations in this basis are

$$\begin{aligned} &[K_1,K_2]=-L_3,\ \ [L_3,K_1]=K_2,\ \ [L_3,K_2]=-K_1,\\ &[M,Q]=-N,\ \ [M,N]=[Q,N]=0,\\ &[K_1,M]=-\tfrac{1}{2}M,\ \ [K_1,Q]=\tfrac{1}{2}Q,\ \ [K_1,N]=0,\\ &[K_2,M]=\tfrac{1}{2}Q,\ \ [K_2,Q]=\tfrac{1}{2}M,\ \ [K_2,N]=0,\\ &[L_3,M]=-\tfrac{1}{2}Q,\ \ [L_3,Q]=\tfrac{1}{2}M,\ \ [L_3,N]=0,\\ &[W,K_1]=[W,K_2]=[W,L_3]=0,\ \ [W,M]=\tfrac{1}{2}M,\\ &[W,Q]=\tfrac{1}{2}Q,\ \ [W,N]=N. \end{aligned} \tag{4.19}$$

The subalgebra $\{K_1,K_2,L_3,M,Q,N\}$ generates the Schrödinger group Sch_1, i.e., the invariance group of the one-dimensional time dependent Schrödinger equation.[11]

Thus, we have obtained six maximal subalgebras of LO(3,2), all imbedded reducibly in the considered representation. It is easy to see that these are the only ones. Indeed, as stated in Sec. III C, if a two-dimensional subspace of signature (+ −), (+0), or (−0), or a three-dimensional subspace of signature (+ + 0), (+ − 0), or (− − 0) is invariant under some subalgebra, then the same subalgebra also leaves a one-dimensional subspace with signature (0) invariant. The corresponding subalgebra is hence not maximal but a subalgebra of the similitude algebra S. If a (+ + +), (+ + −), or (+ − −) subspace is invariant, then its orthogonal complement is invariant. The corresponding subalgebras coincide with those considered earlier. A (+00) invariant subspace leads to a subalgebra of the algebra leaving a (00) space invariant. Similar comments hold for all possible four-dimensional invariant subspaces.

B. Irreducibly imbedded subalgebras

It was shown in Sec. III C that the LO(3,2) algebra in the considered representation has one irreducibly imbedded subalgebra, namely LO(2,1). In the (2.17) realization the generators of this subalgebra can be written as

$$K_1=-(E+H+\sqrt{3}K),\quad K_2=F+G-\sqrt{3}J,\quad L_3=2A+D. \tag{4.20}$$

To summarize: The algebra LO(3,2) has seven O(3,2) classes of maximal subalgebras represented by formulas (4.4), (4.7), (4.10), (4.13), (4.15), (4.18), and (4.20).

Our next step is to classify the subalgebras of each maximal subalgebra, this time with respect to conjugacy under the Lie group, generated by the corresponding maximal subalgebra. To simplify the final merging into one list of representatives of O(3,2) classes of subalgebras we shall produce normalized lists of subalgebras. A further simplification is achieved by an appropriate ordering of the maximal subgroups.

The ordering we shall use is:

(a) the similitude group Sim(2,1),

(b) the optical group Opt(2,1),

(c) the maximal compact subgroup $\text{O}(3)\otimes\text{O}(2)$,

(d) the group $\text{O}(2)\otimes\text{O}(2,1)$,

(e) the group O(2,2),

(f) the Lorentz group O(3,1),

(g) the irreducible subgroup O(2,1).

Thus, we first list representatives of all subalgebras of LSim(2,1). By construction, each of them leaves a lightlike vector invariant. Some subalgebras may have higher normalizers lying in another maximal subalgebra. In the final listing these will be removed from the LSim(2,1) list to the list containing the higher normalizers. Next, we consider subalgebras of the algebra LOpt(2,1). In the final list we must omit those subalgebras, that are also contained in LSim(2,1) and also those that have a higher normalizer in a maximal subalgebra that has not yet been considered. Thus we proceed through the entire ordered list.

A general element of the optical algebra LOpt(2,1) can in the realization (2.22) be written as

$$X=\begin{pmatrix} -f & l & m & n & 0\\ p & -g & q & 0 & -n\\ 0 & 0 & 0 & -q & -m\\ 0 & 0 & 0 & g & -l\\ 0 & 0 & 0 & -p & f\end{pmatrix}. \tag{4.21}$$

The element X will also be contained in LSim(2,1) if X leaves a lightlike vector invariant. The condition for this not to happen is that the matrix

$$Y=\begin{pmatrix} -f & l \\ p & -g \end{pmatrix}$$

should have no real eigenvalues

$$\lambda=\frac{-f-g\pm[(f-g)^2+4lp]^{1/2}}{2}.$$

A necessary condition for this is that $lp<0$. By inspecting the subalgebras of GL(2,R) (the matrix Y) we see that there are exactly four subalgebras containing elements for which $lp<0$. In the notations of (4.18) these are $\{W,K_1,K_2,L_3\}$ generating GL(2,R), $\{K_1,K_2,L_3\}$ generating SL(2,R), $\{L_3\}$ generating O(2), and $\{L_3+aW\}$ with $a\neq 0$ generating a covering group of O(2). All subalgebras of LOpt(2,1), the intersections of which with GL(2,R) do not coincide with one of the above four algebras, should be eliminated from our final ordered list [since they are already contained in LSim(2,1)]. The only exceptions are those subalgebras that have their normalizer in LOpt(2,1) and not in LSim(2,1). Subalgebras of LO(3)$\oplus$LO(2) and of LO(2)$\oplus$LO(2,1) should be considered next and those already contained in LSim(2,1) or LOpt(2,1) eliminated by inspection, as well as those with normalizers elsewhere. The subalgebras of LO(3,1) and LO(2,2) already leave a timelike or a spacelike vector space invariant, respectively. Those imbedded reducibly in LO(3,1) or LO(2,2) will leave some additional space invariant and thus already be contained in the previously considered maximal subalgebras. Hence the only subalgebras that will figure in the final list are those imbedded irreducibly in the corresponding four-dimensional representation. Of these LO(3,1) has none. LO(2,2) just one—the previously established LO(2)$\oplus$LO(2,1) subalgebra of LO(2,2). The maximal subalgebras LO(3,1) and LO(2,2) themselves of course survive in the list. Finally, all nontrivial subalgebras of the irreducible LO(2,1) are imbedded reducibly in LO(3,2). Hence they have already been classified.

It follows from the above discussion that we need a complete classification of the subalgebras of the first maximal subalgebra in our ordered list and only partial lists of all the other ones. However, since each of the maximal subgroups of O(3,2) is in itself of physical interest, we shall classify all subalgebras of each maximal subalgebra and only in the end eliminate the overlap.

TABLE IV. Subalgebras of the similitude algebra.

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$a_{7,1}$	$F;K_1,K_2,L_3,P_0,P_1,P_2$	LSim(2,1)	self	$(K_1P_2-K_2P_1-L_3P_0)^2/(P_0^2-P_1^2-P_2^2)$	$a_{6,1}, a_{6,2}, a_{5,7}, a_{4,4}$
$a_{6,1}$	$;K_1,K_2,L_3, P_0,P_1,P_2$	LE(2,1)	$a_{7,1}$	$P_0^2-P_1^2-P_2^2$, $K_1P_2-K_2P_1-L_3P_0$	$a_{5,1}, a_{4,3}, a_{3,24}$
$a_{6,2}$	$F,K_2;$ K_1-L_3,P_0,P_1,P_2	$F\square A_{5,30}^0$	self	none	$a_{5,1}, a_{5,2}, a_{5,3}, a_{5,4}^a, a_{5,5}, a_{5,6}\ a_{5,8}$
$a_{5,1}$	$K_2;L_3-K_1, P_0,P_1,P_2$	$A_{5,30}^0$	$a_{6,2}$	$P_0^2-P_1^2-P_2^2$	$a_{4,1}, a_{4,7}, a_{4,14}$
$a_{5,2}$	$F-K_2;$ $-K_1+L_3,P_0,\ P_1,P_2$	$A_{5,30}^{-1}$	$a_{6,2}$	P_0-P_2	$a_{4,2}, a_{4,7}, a_{4,13}$
$a_{5,3}$	$F+K_2;$ $-K_1+L_3,P_0,P_1,P_2$	$A_{5,30}^1$	$a_{6,2}$	$(P_0^2-P_1^2-P_2^2)/(P_0-P_2)$	$a_{4,2}, a_{4,7}, a_{4,12}^\varepsilon, \tilde{a}_{4,18}$
$a_{5,4}^a$ $a\neq 0,\pm 1$	$F+aK_2;$ $-K_1+L_3,P_0,P_1,P_2$	$A_{5,30}^{1/a}$	$a_{6,2}$	$(P_0-P_2)^{2/(1+a)}/(P_0^2-P_1^2-P_2^2)$	$a_{4,7}, a_{4,10}^{-a}, a_{4,15}^a, a_{4,16}^\varepsilon (a=\frac{1}{2})$
$a_{5,5}$	$F,L_3-K_1;$ P_0,P_1,P_2	$A_{5,32}^0$	$a_{6,2}$	$(P_0^2-P_1^2-P_2^2)/(P_0-P_2)^2$	$a_{4,7}, a_{4,8}^\varepsilon, a_{4,9}, \tilde{a}_{4,14}$
$a_{5,6}$	$F,K_2;P_0,$ P_1,P_2	$A_{5,33}^{1/2,1/2}$	self	$(P_0^2-P_1^2-P_2^2)/P_1^2$	$a_{4,1}, a_{4,2}, a_{4,5}, a_{4,6}, a_{4,9}, a_{4,10}^b$
$a_{5,7}$	$F,L_3;P_0,P_1,P_2$	$A_{5,35}^{0,1}$	self	$(P_0^2-P_1^2-P_2^2)/P_0^2$	$a_{4,3}, a_{4,9}, a_{4,11}, a_{4,17}, a_{3,5}$
$a_{5,8}$	$F,K_2;L_3-K_1,$ P_0-P_2,P_1	$A_{5,36}$	self	$[(L_3-K_1)P_1/(P_0-P_2)]+K_2-F$	$a_{4,5}, \tilde{a}_{4,5}, a_{4,13}, a_{4,14}, \tilde{a}_{4,14}, a_{4,15}^b, \bar{a}_{4,18}$

TABLE IV. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$a_{4,1}$	$P_1\oplus$ $\{K_2;P_0,P_2\}$	$4_1\oplus A_{3,4}$	$a_{5,6}$	$P_1, P_0^2-P_2^2$	$a_{3,1}, a_{3,2}, a_{3,13}, a^{\epsilon}_{3,14}$
$a_{4,2}$	$P_0-P_2\oplus$ $\{F-K_2;P_0+P_2,P_1\}$	$A_1\oplus A^{1/2}_{3,5}$	$a_{5,6}$	P_0-P_2, $(P_0+P_2)/P_1^2$	$a_{3,1}, a_{3,6}, a_{3,7}$, $a_{3,16}, a^{\epsilon}_{3,17}$
$a_{4,3}$	$P_0\oplus\{L_3;P_1,P_2\}$	$A_1\oplus A_{3,6}$	$a_{5,7}$	$P_0, P_1^2+P_2^2$	$a_{3,1}, a_{3,21}, a^{\epsilon}_{3,22}, a_{2,3}$
$a_{4,4}$	$F\oplus\{;K_1,K_2,L_3\}$	$A_1\oplus A_{3,8}$	self	$F, K_1^2+K_2^2-L_3^2$	$a_{3,3}, a_{3,24}, a_{2,7}$
$a_{4,5}$	$\{K_2;P_0-P_2\}\oplus$ $\{F-K_2;P_1\}$	$A_2\oplus A_2$	self	none	$a_{3,2}, \tilde{a}_{3,3}, a_{3,4}, a_{3,7}$, $\tilde{a}_{3,10}, a_{3,15}, a_{3,16}, a^{c}_{3,19}$
$\tilde{a}_{4,5}$	$\{F;P_0-P_2\}\oplus\{F$ $-K_2, L_3-K_1\}$	$A_2\oplus A_2$	self	none	$\tilde{a}_{3,2}, a_{3,3}, a_{3,4}, \tilde{a}_{3,7}, a_{3,10}$, $\tilde{a}_{3,15}, \tilde{a}_{3,16}, \tilde{a}^{c}_{3,19}$
$a_{4,6}$	$\{F+K_2;P_0-P_2\}\oplus$ $\{F-K_2;P_0+P_2\}$	$A_2\oplus A_2$	self	none	$a_{3,4}, a_{3,6}, a_{3,12}$, $a_{3,13}, a^{c}_{3,20}$
$a_{4,7}$	$L_3-K_1, P_0+P_2;$ P_0-P_2, P_1	$A_{4,1}$	$a_{6,2}$	P_0-P_2, P_0^2 $-P_1^2-P_2^2$	$a_{3,1}, a^{\epsilon}_{3,8}, \bar{a}_{3,25}$
$a^{\epsilon}_{4,8}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F+\epsilon(L_3-K_1);$ P_0, P_1, P_2	$A_{4,4}$	$a_{5,5}$	$(P_0-P_2)^2/(P_0^2-P_1^2-P_2^2)$, $(P_0-P_2)^{\epsilon}\exp[P_1/(P_2-P_0)]$	$a_{3,1}, \tilde{a}^{\epsilon}_{3,9}$
$a_{4,9}$	$F; P_0, P_1, P_2$	$A^{1,1}_{4,5}$	$a_{7,1}$	P_1/P_0, P_2/P_0	$a_{3,1}, \tilde{a}_{3,10}$, $a_{3,11}, a_{3,12}$
$a^{b}_{4,10}$ $b>0, \neq 1$	$F-bK_2;$ P_0, P_1, P_2	$A^{1/(1-b),(1+b)/(1-b)}_{4,5}$	$a_{5,6}$	$P_1^2/(P_0^2-P_1^2-P_2^2)$, $P_1^{2b}(P_0-P_2)/(P_0+P_2)$	$a_{3,1}, a_{3,15}(b=2)$ $a^{b}_{3,19}, a^{\frac{1}{b}}_{3,20}$
$a^{b}_{4,11}$ $b>0[b\neq 0]$	$F+bL_3;$ P_0, P_1, P_2	$A^{1/b,1/b}_{4,6}$	$a_{5,7}$	$P_0^2/(P_1^2+P_2^2)$, $(P_1^2+P_2^2)^b$ $(P_1+iP_2)^i(P_1-iP_2)^{-i}$	$a_{3,1}, a^{b}_{4,23}$, $a^{b}_{3,17}$
$a^{\epsilon}_{4,12}$ $\epsilon=1^*$ $[\epsilon=\pm 1]$	$F+K_2+\epsilon(P_0+P_2);$ $-K_1+L_3, P_0-P_2, P_1$	$A_{4,7}$	$a_{5,3}$	none	$a^{\epsilon}_{3,17}, \bar{a}_{3,25}$
$a_{4,13}$	$F-K_2;$ $P_0-P_2, -K_1+L_3, P_1$	$A_{4,8}$	$a_{5,8}$	P_0-P_2, $(P_0-P_2)(F-K_2)$ $-P_1(-K_1+L_3)$	$a_{3,7}, \tilde{a}_{3,7}, \bar{a}_{3,25}$
$a_{4,14}$	$K_2, P_1;$ $-K_1+L_3, P_0-P_2$	$A^{0}_{4,9}$	$a_{5,8}$	none	$a_{3,2}, a^{\epsilon}_{3,9}, a_{3,10}$, $\bar{a}_{3,25}$
$\tilde{a}_{4,14}$	$F, -K_1+L_3;$ P_1, P_0-P_2	$A^{0}_{4,9}$	$a_{5,8}$	none	$\tilde{a}_{3,2}, \tilde{a}^{\epsilon}_{3,9}, \tilde{a}_{3,10}, \bar{a}_{3,25}$
$a^{b}_{4,15}$ $0<\lvert b\rvert<1$ $[b\neq 0,\pm 1]$	$F+bK_2; -K_1+L_3,$ P_0-P_2, P_1	$A^{h}_{4,9}$ $h=b$, $\lvert b\rvert<1$ $h=b^{-1}$, $\lvert b\rvert>1$	$a_{5,8}$	none	$a_{3,15}(b=-2), \tilde{a}_{3,15}(b=-\frac{1}{2})$, $a^{b}_{3,19}, \tilde{a}^{b}_{3,19}, \bar{a}_{3,25}$

TABLE IV. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$a^{\epsilon}_{4,16}$ $\epsilon = 1 * [\epsilon = \pm 1]$	$F + \frac{1}{2}K_2; -K_1 + L_3 + \epsilon(P_0 + P_2)$, $P_0 - P_2, P_1$	$A^{1/2}_{4,9}$	$a^{1/2}_{5,4}$	none	$a^{\epsilon}_{3,8}, a^{\epsilon}_{3,18}, a^{1/2}_{3,19}$
$a_{4,17}$	$F, L_3; P_1, P_2$	$A_{4,12}$	self	none	$a_{3,11}, a_{3,21}$, $a^{\epsilon}_{3,23}, a_{2,7}$
$\bar{a}_{4,18} = b_{4,5}$	$F + K_2; -K_1 + L_3$, $P_0 - P_2, P_1$	$A^{1}_{4,9}$	$b_{7,1}$	none	$a_{3,16}, \tilde{a}_{3,16}$, $\bar{a}_{3,25}$
$a_{3,1}$	P_0, P_1, P_2	$3A_1$	$a_{7,1}$	P_0, P_1, P_2	$a_{2,2}, a_{2,4}, a_{2,5}$
$a_{3,2}$	$P_1 \oplus \{K_2; P_0 - P_2\}$	$A_1 \oplus A_2$	$a_{4,5}$	P_1	$\tilde{a}_{2,1}, a_{2,2}, a_{2,11}$, $a^{\epsilon}_{2,12}$
$\tilde{a}_{3,2}$	$-K_1 + L_3 \oplus$ $\{F; P_0 - P_2\}$	$A_1 \oplus A_2$	$\tilde{a}_{4,5}$	$-K_1 + L_3$	$a_{2,1}, \tilde{a}_{2,2}, \tilde{a}_{2,11}$, $\tilde{a}^{\epsilon}_{2,12}$
$a_{3,3}$	$F \oplus \{K_2; -K_1 + L_3\}$	$A_1 \oplus A_2$	self	F	$a_{2,1}, a_{2,6}, a_{2,10}$, $a^{d}_{2,14}, \tilde{a}^{\epsilon}_{2,15}$
$\tilde{a}_{3,3}$	$K_2 \oplus \{F; P_1\}$	$A_1 \oplus A_2$	self	K_2	$\tilde{a}_{2,1}, a_{2,6}, \tilde{a}_{2,10}$, $\tilde{a}^{d}_{2,14}, a_{2,15}$
$a_{3,4}$	$F - K_2 \oplus$ $\{F + K_2; P_0 - P_2\}$	$A_1 \oplus A_2$	self	$F - K_2$	$a_{2,6}, a_{2,8}, a_{2,11}$, $\tilde{a}_{2,11}, a^{d}_{2,18}, \bar{a}_{2,21}$
$a_{3,5}$	$L_3 \oplus \{F; P_0\}$	$A_1 \oplus A_2$	self	L_3	$a_{2,3}, a_{2,7}, a_{2,13}$, $a^{d}_{2,17}$
$a_{3,6}$	$P_0 - P_2 \oplus$ $\{F - K_2; P_0 + P_2\}$	$A_1 \oplus A_2$	$a_{4,6}$	$P_0 - P_2$	$a_{2,5}, a_{2,8}, a^{\epsilon}_{2,19}$, $\bar{a}_{2,21}$
$a_{3,7}$	$P_0 - P_2 \oplus$ $\{F - K_2; P_1\}$	$A_1 \oplus A_2$	$a_{4,5}$	$P_0 - P_2$	$a_{2,2}, a_{2,8}, a_{2,15}$, $a^{\epsilon}_{2,16}$
$\tilde{a}_{3,7}$	$P_0 - P_2 \oplus$ $\{F - K_2; -K_1 + L_3\}$	$A_1 \oplus A_2$	$\tilde{a}_{4,5}$	$P_0 - P_2$	$\tilde{a}_{2,2}, a_{2,8}, \tilde{a}^{1}_{2,15}$, $\tilde{a}^{\epsilon}_{2,16}$
$a^{\epsilon}_{3,8}$ $\epsilon = 1 * [\epsilon = \pm 1]$	$-K_1 + L_3 + \epsilon(P_0 + P_2)$, $P_1; P_0 - P_2$	$A_{3,1}$	$a^{1/2}_{5,4}$	$P_0 - P_2$	$a_{2,2}, a^{\epsilon}_{2,9}$
$a^{\epsilon}_{3,9}$ $\epsilon = 1$ $[\epsilon = \pm 1]$	$K_2 - \epsilon P_1$; $-K_1 + L_3, P_0 - P_2$	$A_{3,2}$	$a_{4,14}$	$(P_0 - P_2)^{-\epsilon} \exp\{(-K_1 + L_3)/(P_0 - P_2)\}$	$\tilde{a}_{2,2}, a^{\epsilon}_{2,12}$
$\tilde{a}^{\epsilon}_{3,9}$ $\epsilon = 1$ $[\epsilon = \pm 1]$	$F + \epsilon(-K_1 + L_3)$; $P_1, P_0 - P_2$	$A_{3,2}$	$\tilde{a}_{4,14}$	$(P_0 - P_2)^{-\epsilon}$ $\exp\{P_1/(P_0 - P_2)\}$	$a_{2,2}, \tilde{a}^{\epsilon}_{2,12}$
$a_{3,10}$	$K_2; -K_1 + L_3$, $P_0 - P_2$	$A_{3,3}$	$a_{5,8}$	$(-K_1 + L_3)/(P_0 - P_2)$	$\tilde{a}_{2,2}, a_{2,10}, a_{2,11}$
$\tilde{a}_{3,10}$	$F; P_1, P_0 - P_2$	$A_{3,3}$	$a_{5,8}$	$P_1/(P_0 - P_2)$	$a_{2,2}, \tilde{a}_{2,10}, \tilde{a}_{2,11}$
$a_{3,11}$	$F; P_1, P_2$	$A_{3,3}$	$a_{4,17}$	P_1/P_2	$a_{2,4}, \tilde{a}_{2,10}$
$a_{3,12}$	$F; P_0, P_2$	$A_{3,3}$	$a_{4,6}$	P_0/P_2	$a_{2,5}, \tilde{a}_{2,10}, \tilde{a}_{2,11}$, $a_{2,13}$
$a_{3,13}$	$K_2; P_0, P_2$	$A_{3,4}$	$a_{5,6}$	$P_0^2 - P_2^2$	$a_{2,5}, a_{2,11}$
$a^{\epsilon}_{3,14}$ $\epsilon = 1$ $[\epsilon = \pm 1]$	$K_2 - \epsilon P_1; P_0, P_2$	$A_{3,4}$	$a_{4,1}$	$P_0^2 - P_2^2$	$a_{2,5}, a^{\epsilon}_{2,12}$
$a_{3,15}$	$F - 2K_2; P_1$, $P_0 - P_2$	$A_{3,4}$	$a_{4,5}$	$P_1(P_0 - P_2)$	$a_{2,2}, \tilde{a}^{2}_{2,14}, a^{-2}_{2,18}$

TABLE IV. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\tilde{a}_{3,15}$	$F-\frac{1}{2}K_2$; $-K_1+L_3, P_0-P_2$	$A_{3,4}$	$\tilde{a}_{4,5}$	$(-K_1+L_3)(P_0-P_2)$	$\tilde{a}_{2,2}, a_{2,14}^{1/2}, a_{2,18}^{-1/2}$
$a_{3,16}$	$F+K_2; P_1$, P_0-P_2	$A_{3,5}^{2}$	$a_{5,6}$	$P_1^2/(P_0-P_2)$	$a_{2,2}, a_{2,15}, \bar{a}_{2,21}$
$\tilde{a}_{3,16}$	$F+K_2; -K_1+L_3, P_0-P_2$	$A_{3,5}^{2}$	$\sim a_{5,6}$	$(-K_1+L_3)^2/(P_0-P_2)$	$\tilde{a}_{2,2}, \tilde{a}_{2,15}^{-1}, \bar{a}_{2,21}$
$a_{3,17}^{\epsilon}$ $\epsilon=1*[\epsilon=\pm 1]$	$F+K_2+\epsilon(P_0+P_2)$; P_0-P_2, P_1	$A_{3,5}^{2}$	$a_{4,2}$	$P_1^2/(P_0-P_2)$	$a_{2,2}a_{2,16}^{\epsilon}, a_{2,19}^{\epsilon}$
$a_{3,18}^{\epsilon}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F+\frac{1}{2}K_2; -K_1+L_3$ $+\epsilon(P_0+P_2), P_0-P_2$	$A_{3,5}^{3}$	self	$(P_0-P_2)/[-K_1+L_3$ $+\epsilon(P_0+P_2)]^3$	$a_{2,9}^{\epsilon}, a_{3,18}^{1/2}, a_{2,20}^{\epsilon}$
$a_{3,19}^{c}$ $c\neq 0,\pm 1,-2$	$F+cK_2; P_1$, P_0-P_2	$A_{3,5}^{h}$ $h=1+c, -2<c<0$ $h=(1+c)^{-1}, c<-2$ or $c>0$	$a_{4,5}$	$(P_0-P_2)P_1^{-1-c}$	$a_{2,2}, \tilde{a}_{2,14}^{c}, a_{2,18}^{c}$
$\tilde{a}_{3,19}^{c}$ $c\neq 0,-\frac{1}{2},\pm 1$	$F+cK_2$; $-K_1+L_3, P_0-P_2$	$A_{3,5}^{h}$ $h=\frac{c}{c+1}, c>-\frac{1}{2}$ $h=\frac{c+1}{c}, c<-\frac{1}{2}$	$\tilde{a}_{4,5}$	(P_0-P_2) $(L_3-K_1)^{-(1+c)/c}$	$\tilde{a}_{2,2}, a_{2,14}^{c}, a_{2,18}^{c}$
$a_{3,20}^{c}$ $c>0, c\neq 1$	$F+cK_2; P_0, P_2$	$A_{3,5}^{(1-c)/(1+c)}$	$a_{4,6}$	(P_0+P_2) $(P_0-P_2)^{(-1+c)/(1+c)}$	$a_{2,5}, a_{2,18}^{c}$
$a_{3,21}$	$L_3; P_1, P_2$	$A_{3,6}$	$a_{5,7}$	$P_1^2+P_2^2$	$a_{2,4}, \bar{a}_{1,10}$
$a_{3,22}^{\epsilon}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$L_3+\epsilon P_0; P_1, P_2$	$A_{3,6}$	$a_{4,3}$	$P_1^2+P_2^2$	$a_{2,4}, a_{1,6}^{\epsilon}$
$a_{3,23}^{c}$ $c>0$ $[c\neq 0]$	$F+cL_3; P_1, P_2$	$A_{3,7}^{1/c}$	$a_{4,17}$	$(P_1^2+P_2^2)^c$ $(P_1+iP_2)^i(P_1-iP_2)^{-i}$	$a_{2,4}, a_{1,7}^{c}$
$a_{3,24}$	$; K_1, K_2, L_3$	$A_{3,8}$	$a_{4,4}$	$K_1^2+K_2^2-L_3^2$	$a_{2,10}, \bar{a}_{1,10}$
$\tilde{a}_{3,25}=b_{3,2}$	$-K_1+L_3, P_1$: P_0-P_2	$A_{3,1}$	$b_{7,1}$	P_0-P_2	$a_{2,2}, \tilde{a}_{2,2}$
$a_{2,1}$	$F, -K_1+L_3$	$2A_1$	$a_{3,3}$	$F, -K_1+L_3$	$a_{1,1}, \tilde{a}_{1,2}, a_{1,4}^{\epsilon}$
$\tilde{a}_{2,1}$	K_2, P_1	$2A_1$	$\tilde{a}_{3,3}$	K_2, P_1	$\tilde{a}_{1,1}, a_{1,2}, \tilde{a}_{1,4}^{\epsilon}$
$a_{2,2}$	P_1, P_0-P_2	$2A_1$	$a_{6,2}$	P_1, P_0-P_2	$a_{1,2}, \bar{a}_{1,11}$
$\tilde{a}_{2,2}$	$-K_1+L_3$, P_0-P_2	$2A_1$	$b_{6,2}$	$-K_1+L_3$, P_0-P_2	$\tilde{a}_{1,2}, \bar{a}_{1,11}$
$a_{2,3}$	L_3, P_0	$2A_1$	$a_{3,5}$	L_3, P_0	$a_{1,3}, a_{1,6}^{\epsilon}, \bar{a}_{1,10}$
$a_{2,4}$	P_1, P_2	$2A_1$	$a_{5,7}$	P_1, P_2	$a_{1,2}$
$a_{2,5}$	P_0, P_2	$2A_1$	$a_{5,6}$	P_0, P_2	$a_{1,2}, a_{1,3}, \bar{a}_{1,11}$
$a_{2,6}$	F, K_2	$2A_1$	self	F, K_2	$a_{1,1}, \tilde{a}_{1,1}, a_{1,8}^{\epsilon}$, $\bar{a}_{1,12}$
$a_{2,7}$	F, L_3	$2A_1$	self	F, L_3	$a_{1,1}, a_{1,7}^{\epsilon}, \bar{a}_{1,10}$
$a_{2,8}$	$F-K_2$, P_0-P_2	$2A_1$	$a_{3,4}$	$F-K_2, P_0-P_2$	$a_{1,3}^{\epsilon}, \bar{a}_{1,11}, \bar{a}_{1,12}$
$a_{2,9}^{\epsilon}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$L_3-K_1+\epsilon(P_0+P_2)$, P_0-P_2	$2A_1$	$a_{3,18}^{\epsilon}$	$L_3-K_1+\epsilon(P_0+P_2)$, P_0-P_2	$a_{1,5}^{\epsilon}, \bar{a}_{1,11}$

TABLE IV. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$a_{2,10}$	$K_2; -K_1+L_3$	A_2	$a_{3,3}$	none	$\tilde{a}_{1,1}, \tilde{a}_{1,2}$
$\tilde{a}_{2,10}$	$F; P_1$	A_2	$\tilde{a}_{3,3}$	none	$a_{1,1}, a_{1,2}$
$a_{2,11}$	$K_2; P_0-P_2$	A_2	$a_{4,5}$	none	$\tilde{a}_{1,1}, \bar{a}_{1,11}$
$\tilde{a}_{2,11}$	$F; P_0-P_2$	A_2	$\tilde{a}_{4,5}$	none	$a_{1,1}, \bar{a}_{1,11}$
$a^{\epsilon}_{2,12}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$K_2-\epsilon P_1; P_0-P_2$	A_2	$a_{3,2}$	none	$\tilde{a}^{\epsilon}_{1,4}, \bar{a}_{1,11}$
$\tilde{a}^{\epsilon}_{2,12}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F+\epsilon(-K_1+L_3)$; P_0-P_2	A_2	$\tilde{a}_{3,2}$	none	$a^{\epsilon}_{1,4}, \bar{a}_{1,11}$
$a_{2,13}$	$F; P_0$	A_2	$a_{3,5}$	none	$a_{1,1}, a_{1,3}$
$a^{d}_{2,14}$ $d>0, d\neq 1$ $[d\neq 0, \pm 1]$	$F-dK_2$; $-K_1+L_3$	A_2	$a_{3,3}$	none	$\tilde{a}_{1,2}, a^{d}_{1,8}$
$\tilde{a}^{d}_{2,14}$ $d>0, d\neq 1$ $[d\neq 0, \pm 1]$	$F-dK_2; P_1$	A_2	$\tilde{a}_{3,3}$	none	$a_{1,2}, a^{d}_{1,8}$
$a_{2,15}$	$F-K_2; P_1$	A_2	$a_{4,5}$	none	$a_{1,2}, a_{1,12}$
$\tilde{a}^{\epsilon}_{2,15}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F-\epsilon K_2$; $-K_1+L_3$	A_2	$\tilde{a}_{4,5}$	none	$\tilde{a}_{1,2}, \bar{a}_{1,12}$
$a^{\epsilon}_{2,16}$ $\epsilon=1*$ $[\epsilon=\pm 1]$	$F-K_2+\epsilon(P_0-P_2); P_1$	A_2	$a_{3,7}$	none	$a_{1,2}, a^{\epsilon}_{1,9}$
$\tilde{a}^{\epsilon}_{2,16}$ $\epsilon=1*$ $[\epsilon=\pm 1]$	$F-K_2+\epsilon(P_0-P_2); -K_1+L_3$	A_2	$\tilde{a}_{3,7}$	none	$\tilde{a}_{1,2}, a^{\epsilon}_{1,9}$
$a^{d}_{2,17}$ $d>0$ $[d\neq 0]$	$F+dL_3; P_0$	A_2	$a_{3,5}$	none	$a_{1,3}, a^{d}_{1,7}$
$a^{d}_{2,18}$ $0<\lvert d\rvert<1$ $[d\neq 0, \pm 1]$	$F+dK_2; P_0$ $-P_2$	A_2	$a_{3,4}$	none	$a^{d}_{1,8}, \bar{a}_{1,11}$
$a^{\epsilon}_{2,19}$ $\epsilon=1*$ $[\epsilon=\pm 1]$	$F-K_2+\epsilon(P_0-P_2)$; $; P_0+P_2$	A_2	$a_{3,6}$	none	$a^{\epsilon}_{1,9}, \bar{a}_{1,11}$
$a^{\epsilon}_{2,20}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F+\frac{1}{2}K_2; -K_1$ $+L_3+\epsilon(P_0+P_2)$	A_2	self	none	$a^{\epsilon}_{1,5}, a^{1/2}_{1,8}$
$\bar{a}_{2,21}=b_{2,4}$	$F+K_2; P_0-P_2$	A_2	$b_{5,1}$	none	$\bar{a}_{1,11}, \bar{a}_{1,12}$
$a_{1,1}$	F	A_1	$a_{4,4}$	F	
$\tilde{a}_{1,1}$	K_2	A_1	$\sim a_{4,4}$	K_2	
$a_{1,2}$	P_1	A_1	$a_{5,6}$	P_1	
$\tilde{a}_{1,2}$	$-K_1+L_3$	A_1	$\sim a_{5,6}$	$-K_1+L_3$	
$a_{1,3}$	P_0	A_1	$a_{5,7}$	P_0	
$a^{\epsilon}_{1,4}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F+\epsilon(-K_1+L_3)$	A_1	$a_{2,1}$	$F+\epsilon(-K_1+L_3)$	
$\tilde{a}^{\epsilon}_{1,4}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$K_2-\epsilon P_1$	A_1	$\tilde{a}_{2,1}$	$K_2-\epsilon P_1$	
$a^{\epsilon}_{1,5}$ $\epsilon=1*$ $[\epsilon=\pm 1]$	$-K_1+L_3$ $+\epsilon(P_0+P_2)$	A_1	$a^{\epsilon}_{3,18}$	$-K_1+L_3$ $+\epsilon(P_0+P_2)$	
$a^{\epsilon}_{1,6}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$L_3+\epsilon P_0$	A_1	$a_{2,3}$	$L_3+\epsilon P_0$	
$a^{e}_{1,7}$ $e>0$ $[e\neq 0]$	$F+eL_3$	A_1	$a_{2,7}$	$F+eL_3$	
$a^{e}_{1,8}$ $0<e<1$ $[e>0]$	$F+eK_2$	A_1	$a_{2,6}$	$F+eK_2$	
$a^{\epsilon}_{1,9}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$F-K_2$ $+\epsilon(P_0-P_2)$	A_1	$a_{2,8}$	$F-K_2+\epsilon(P_0-P_2)$	
$\bar{a}_{1,10}=d_{1,1}$	L_3	A_1	$d_{4,1}$	L_3	
$\bar{a}_{1,11}=b_{1,4}$	P_0-P_2	A_1	$b_{7,1}$	P_0-P_2	
$\bar{a}_{1,12}=b_{1,5}$	$F+K_2$	A_1	$b_{4,2}$	$F+K_2$	

TABLE V. Subalgebras of the optical algebra LOpt(2,1).

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$b_{7,1}$	$W;K_1,K_2,L_3,$ M,Q,N	LOpt(2,1)	self	$\frac{1}{N}[MQK_1+\frac{1}{2}(K_2-L_3)M^2$ $-\frac{1}{2}(K_2+L_3)Q^2]_{\text{sym}}+K_1^2+K_2^2-L_3^2$	$b_{6,1}, b_{6,2}, b_{5,1}$ $b_{5,2}$
$b_{6,1}$	$;K_1,K_2,L_3,$ M,Q,N	LSch_1	$b_{7,1}$	$[MQK_1+\frac{1}{2}(K_2-L_3)M^2$ $-\frac{1}{2}(K_2+L_3)Q^2]_{\text{sym}}+N(K_1^2+K_2^2-L_3^2),N$	$b_{5,4}, b_{4,1}, b_{4,3}$
$b_{6,2}=a_{6,2}$	$W,K_1;K_2+L_3,$ M,Q,N		self	none	$b_{5,3}, b_{5,4}, b_{5,5},$ $b_{5,6}^a, b_{5,7}, b_{5,8}, b_{5,9}$
$b_{5,1}$	$\{;K_1,K_2,L_3\}$ $\oplus\{W;N\}$	$A_{3,8}\oplus A_2$	self	$K_1^2+K_2^2-L_3^2$	$b_{4,1}, b_{4,2}, b_{4,9}, b_{3,1}$
$b_{5,2}$	$W,L_3;M,Q,N$	$A_{5,37}$	self	$(Q^2+M^2+4NL_3)/N$	$b_{4,3}, b_{4,4}^b, b_{4,5}, b_{3,1}$
$b_{5,3}=a_{5,1}$	$W+K_1;K_2+L_3,$ M,Q,N	$A_{5,30}^0$	$b_{6,2}$	$M^2+2N(K_2+L_3)$	$b_{4,6}, b_{4,10}, b_{4,16}$
$b_{5,4}=a_{5,2}$	$K_1;K_2+L_3,$ M,Q,N	$A_{5,30}^{-1}$	$b_{6,2}$	N	$b_{4,7}, b_{4,10}, b_{4,15}$
$b_{5,5}=a_{5,3}$	$W,K_2+L_3;$ M,Q,N	$A_{5,30}^1$	$b_{6,2}$	$2(K_2+L_3)+M^2/N$	$b_{4,5}, \tilde{b}_{4,7}, b_{4,10}, b_{4,14}^{\epsilon}$
$b_{5,6}^a=a_{5,4}^{(1+a)/(1-a)}$ $a\neq 0,+1$	$W+aK_1;K_2+L_3,$ M,Q,N	$A_{5,30}^{(1-a)/(1+a)}$	$b_{6,2}$	$[M^2+2N(K_2+L_3)]N^{a-1}$	$b_{4,10}, b_{4,13}^a, b_{4,17}^a,$ $b_{4,18}(a=-\frac{1}{3})$
$b_{5,7}=a_{5,5}$	$W-K_1;K_2+L_3,$ M,Q,N	$A_{5,32}^0$	$b_{6,2}$	$[M^2+2N(K_2+L_3)]^2/N^2$	$b_{4,10}, b_{4,11}, b_{4,12}, b_{4,16}$
$b_{5,8}=a_{5,6}$	$W,K_1;K_2+L_3,M,N$	$A_{5,33}^{1/2}$	self	$(K_2-L_3)N/M^2$	$b_{4,6}, b_{4,7}, \tilde{b}_{4,7}, b_{4,8},$ $\tilde{b}_{4,8}, b_{4,9}, b_{4,12}, b_{4,13}$
$b_{5,9}=a_{5,8}$	$W,K_1;M,Q,N$	$A_{5,36}$	self	$[QM+MQ-2K_1(W+2K_1)]/N$	$b_{4,5}, b_{4,8}, b_{4,15},$ $b_{4,16}, b_{4,17}^b$
$b_{4,1}$	$N\oplus\{;K_1,K_2,L_3\}$	$A_1\oplus A_{3,8}$	$b_{5,1}$	$N, K_1^2+K_2^2-L_3^2$	$b_{3,3}, b_{3,8}, b_{2,1}$
$b_{4,2}$	$W\oplus\{;K_1,K_2,L_3\}$	$A_1\oplus A_{3,8}$	self	$W, K_1^2+K_2^2-L_3^2$	$b_{3,3}, b_{3,7}, b_{2,2}$
$b_{4,3}$	$L_3;Q,M,N$	$A_{4,10}$	$b_{5,2}$	$N, M^2+Q^2+4L_3N$	$b_{3,2}, b_{2,1}$
$b_{4,4}^b$ $b>0*[b\neq 0]$	$W+bL_3;Q,M,N$	$A_{4,11}^b$	$b_{5,2}$	none	$b_{3,2}, b_{2,3}^{\epsilon}$
$b_{4,5}=a_{4,18}$	$W;Q,M,N$	$A_{4,9}^1$	$b_{7,1}$	none	$b_{3,2}, b_{3,19}$
$b_{4,6}=a_{4,1}$	$M\oplus\{W+K_1;K_2$ $+L_3,N\}$	$A_1\oplus A_{3,4}$	$b_{5,8}$	$M, N(K_2+L_3)$	$b_{3,4}, b_{3,5}, \tilde{b}_{3,5},$ $b_{3,16}, b_{3,17}$
$b_{4,7}=a_{4,2}$	$N\oplus\{K_1;K_2+L_3,M\}$	$A_1\oplus A_{3,5}^{1/2}$	$b_{5,8}$	$N, (K_2+L_3)/M^2$	$b_{3,4}, b_{3,8}, b_{3,9},$ $\tilde{b}_{3,19}, \tilde{b}_{3,20}^{\epsilon}$
$\tilde{b}_{4,7}\sim a_{4,2}$	$K_2+L_3\oplus\{W;M,N\}$	$A_1\oplus A_{3,5}^{1/2}$	$b_{5,8}$	$K_2+L_3, M^2/N$	$b_{3,4}, \tilde{b}_{3,8}, \tilde{b}_{3,9},$ $b_{3,19}, b_{3,20}^{\epsilon}$

TABLE V. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\bar{b}_{4,8}=a_{4,5}$	$\{W+K_1;N\}$ $\oplus\{K_1;M\}$	$A_2\oplus A_2$	self	none	$\bar{b}_{3,5},\bar{b}_{3,6},\tilde{\bar{b}}_{3,7}$, $\bar{b}_{3,9},\bar{b}_{3,12}$, $\bar{b}_{3,18},\bar{b}_{3,19},\tilde{\bar{b}}^{\xi}_{3,22}$
$\tilde{\bar{b}}_{4,8}\sim a_{4,5}$	$\{W+K_1;K_2+L_3\}$ $\oplus\{W;M\}$	$A_2\oplus A_2$	self	none	$\tilde{\bar{b}}_{3,5},\bar{b}_{3,6},\bar{b}_{3,7},\tilde{\bar{b}}_{3,9},\tilde{\bar{b}}_{3,12}$, $\tilde{\bar{b}}_{3,18},\tilde{\bar{b}}_{3,19},\bar{b}^{\xi}_{3,22}$
$\bar{b}_{4,9}=a_{4,6}$	$\{K_1;K_2+L_3\}$ $\oplus\{W;N\}$	$A_2\oplus A_2$	self	none	$\bar{b}_{3,7},\tilde{\bar{b}}_{3,7},\bar{b}_{3,8}$, $\tilde{\bar{b}}_{3,8},\bar{b}_{3,15},\bar{b}_{3,16},\bar{b}^{\xi}_{3,22}$
$\bar{b}_{4,10}=a_{4,7}$	$K_2+L_3,Q;M,N$	$A_{4,1}$	$\bar{b}_{6,2}$	$N,M^2+2N(K_2+L_3)$	$b_{3,2},\bar{b}_{3,4},\bar{b}_{3,10}$
$\bar{b}_{4,11}\sim a^{-1}_{4,8}$	$W-K_1+Q$; K_2+L_3,M,N	$A_{4,4}$	$\bar{b}_{5,7}$	$N\exp(-M/N)$, $(2N(K_2-L_3)+M^2)/N^2$	$\bar{b}_{3,4},\bar{b}_{3,11}$
$\bar{b}_{4,12}=a_{4,9}$	$W-K_1$: K_2+L_3,M,N	$A^{1}_{4,5}$	$a_{7,1}$	$M/N,(K_2+L_3)/N$	$\bar{b}_{3,4},\bar{b}_{3,12},\tilde{\bar{b}}_{3,12}$, $\bar{b}_{3,13},\bar{b}_{3,14},\bar{b}_{3,15}$
$\bar{b}^{b}_{4,13}=a^{(b-1)/(b+1)}_{4,10}$ $0<\lvert b\rvert<1$ $[b\neq 0,\pm 1]$	$W-bK_1$: K_2+L_3,M,N	$A^{b}_{4,5}{}^{(b+1)/2}$	$b_{5,8}$	$N^{(1+b)/2}/M,N^{b}/(K_2+L_3)$	$\bar{b}_{3,4},\bar{b}_{3,18}(b=-3),\bar{b}^{b}_{3,23}$ $\tilde{\bar{b}}_{3,18}(b=-\frac{1}{3}),\bar{b}^{-b}_{3,22},\tilde{\bar{b}}^{b}_{3,22}$,
$\bar{b}^{\epsilon}_{4,14}=a^{\epsilon}_{4,12}$ $\epsilon=1*[\epsilon=\pm 1]$	$W+\epsilon(K_2+L_3)$; M,Q,N	$A_{4,7}$	$\bar{b}_{5,3}$	none	$b_{3,2},\bar{b}^{\epsilon}_{3,20}$
$\bar{b}_{4,15}\sim a_{4,13}$	$K_1;M,Q,N$	$A_{4,8}$	$\bar{b}_{5,9}$	$N,QM+MQ+4NK_1$	$b_{3,2},\bar{b}_{3,9}$
$\bar{b}_{4,16}\sim a_{4,14}$	$W+K_1,M;Q,N$	$A^{0}_{4,9}$	$\bar{b}_{5,9}$	none	$b_{3,2},\bar{b}_{3,5},\bar{b}_{3,11}$, $\bar{b}_{3,12}$
$\bar{b}^{b}_{4,17}=a^{(1-b)/(1+b)}_{4,15}$ $b>0,b\neq 1$	$W-bK_1;M,Q,N$	$A^{(1-b)/(1+b)}_{4,9}$	$\bar{b}_{5,9}$	none	$b_{3,2},\bar{b}_{3,18}$ $(b=3)$, $\tilde{\bar{b}}^{b}_{3,22}$
$\bar{b}_{4,18}\sim a^{-1}_{4,16}$	$W-\frac{1}{3}K_1$; K_2+L_3+Q,M,N	$A^{1/2}_{4,9}$	$\bar{b}^{1/3}_{5,6}$	none	$\bar{b}_{3,10},\bar{b}_{3,21},\tilde{\bar{b}}^{1/3}_{3,22}$
$b_{3,1}$	$L_3\oplus\{W;N\}$	$A_1\oplus A_2$	self	L_3	$b_{2,1},b_{2,2},b^{d}_{2,3},b_{2,4}$
$b_{3,2}=\bar{a}_{3,25}$	$Q,M;N$	$A_{3,1}$	$b_{7,1}$	N	$\tilde{\bar{b}}_{2,6}$
$\bar{b}_{3,3}=c_{3,1}$	$;K_1,K_2,L_3$	$A_{3,8}$	$e_{6,1}$	$K_1^2+K_2^2-L_3^2$	$\tilde{b}_{2,4},\bar{b}_{1,6}$
$\bar{b}_{3,4}=a_{3,1}$	K_2+L_3,M,N	$3A_1$	$a_{7,1}$	K_2+L_3,M,N	$\bar{b}_{2,6},\tilde{\bar{b}}_{2,6},\bar{b}_{2,7}$, $\bar{b}_{2,8},\tilde{\bar{b}}_{2,8}$
$\bar{b}_{3,5}=a_{3,2}$	$M\oplus\{W+K_1;N\}$	$A_1\oplus A_2$	$\bar{b}_{4,8}$	M	$\bar{b}_{2,5},\tilde{\bar{b}}_{2,6},\bar{b}_{2,13},\tilde{\bar{b}}_{2,14}$
$\tilde{\bar{b}}_{3,5}\sim a_{3,2}$	$M\oplus\{W+K_1;$ $K_2+L_3\}$	$A_1\oplus A_2$	$\tilde{\bar{b}}_{4,8}$	M	$\bar{b}_{2,5},\bar{b}_{2,6},\tilde{\bar{b}}_{2,13},\bar{b}_{2,14}$

TABLE V. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\bar{b}_{3,6}\sim a_{3,3}$	$W+K_1\oplus\{W-K_1;M\}$	$A_1\oplus A_2$	self	$W+K_1$	$\bar{b}_{2,5},\bar{b}_{2,9},\bar{b}_{2,12},\bar{b}^{d}_{2,16},\bar{b}_{2,17},\tilde{\bar{b}}_{2,17}$
$\bar{b}_{3,7}\sim a_{3,4}$	$W\oplus\{K_1;K_2+L_3\}$	$A_1\oplus A_2$	self	W	$\tilde{b}_{2,4},\bar{b}_{2,9},\tilde{\bar{b}}_{2,10},\tilde{\bar{b}}_{2,13},\tilde{\bar{b}}_{2,13},\tilde{\bar{b}}^{d}_{2,19}$
$\tilde{\bar{b}}_{3,7}=a_{3,4}$	$K_1\oplus\{W;N\}$	$A_1\oplus A_2$	self	K_1	$b_{2,4},\bar{b}_{2,9},\bar{b}_{2,10},\bar{b}_{2,13},\bar{b}^{d}_{2,19}$
$\bar{b}_{3,8}=a_{3,6}$	$N\oplus\{K_1;K_2+L_3\}$	$A_1\oplus A_2$	$\bar{b}_{4,9}$	N	$\tilde{\bar{b}}_{2,4},\tilde{\bar{b}}_{2,8},\bar{b}_{2,10},\bar{b}^{\epsilon}_{2,20}$
$\tilde{\bar{b}}_{3,8}\sim a_{3,6}$	$K_2+L_3\oplus\{W;N\}$	$A_1\oplus A_2$	$\bar{b}_{4,9}$	K_2+L_3	$b_{2,4},\tilde{\bar{b}}_{2,8},\tilde{\bar{b}}_{2,10},\tilde{\bar{b}}^{\epsilon}_{2,20}$
$\bar{b}_{3,9}=a_{3,7}$	$N\oplus\{K_1;M\}$	$A_1\oplus A_2$	$\bar{b}_{4,8}$	N	$\tilde{\bar{b}}_{2,6},\bar{b}_{2,10},\bar{b}_{2,17},\bar{b}^{\epsilon}_{2,18}$
$\tilde{\bar{b}}_{3,9}\sim a_{3,7}$	$K_2+L_3\oplus\{W;M\}$	$A_1\oplus A_2$	$\tilde{\bar{b}}_{4,8}$	K_2+L_3	$\bar{b}_{2,6},\tilde{\bar{b}}_{2,10},\tilde{\bar{b}}_{2,17},\tilde{\bar{b}}^{\epsilon}_{2,18}$
$\bar{b}_{3,10}\sim a^{-1}_{3,8}$	$K_2+L_3+Q,M;N$	$A_{3,1}$	$\bar{b}^{-1/3}_{5,6}$	N	$\tilde{\bar{b}}_{2,6},\bar{b}_{2,11}$
$\bar{b}_{3,11}\sim a^{-1}_{3,9}$	$W+K_1+M;Q,N$	$A_{3,2}$	$\bar{b}_{4,17}$	$N\exp(Q/N)$	$\tilde{\bar{b}}_{2,6},\tilde{\bar{b}}_{2,14}$
$\bar{b}_{3,12}=\tilde{a}_{3,10}$	$W-K_1;M,N$	$A_{3,3}$	$\bar{b}_{5,9}$	M/N	$\tilde{\bar{b}}_{2,6},\bar{b}_{2,12},\bar{b}_{2,13}$
$\tilde{\bar{b}}_{3,12}\sim\tilde{a}_{3,10}$	$W-K_1;K_2+L_3,M$	$A_{3,3}$	$\sim\bar{b}_{5,9}$	$(K_2+L_3)/M$	$\bar{b}_{2,6},\bar{b}_{2,12},\tilde{\bar{b}}_{2,12},\tilde{\bar{b}}_{2,13}$
$\bar{b}_{3,13}=a_{3,11}$	$W-K_1;K_2+L_3+N,M$	$A_{3,3}$	$a_{4,17}$	$(K_2+L_3+N)/M$	$\bar{b}_{2,7},\bar{b}_{2,12},\tilde{\bar{b}}_{2,12}$
$\bar{b}_{3,14}\sim a_{3,12}$	$W-K_1,K_2+L_3-N,M$	$A_{3,3}$	$\sim a_{4,6}$	$(K_2+L_3-N)/M$	$\bar{b}_{2,8},\bar{b}_{2,12},\tilde{\bar{b}}_{2,12},\tilde{\bar{b}}_{2,13},\bar{b}_{2,15}$
$\bar{b}_{3,15}=a_{3,12}$	$W-K_1;K_2+L_3,N$	$A_{3,3}$	$\bar{b}_{4,9}$	$(K_2+L_3)/N$	$\tilde{\bar{b}}_{2,8},\tilde{\bar{b}}_{2,12},\bar{b}_{2,13},\tilde{\bar{b}}_{2,13},\bar{b}_{2,15}$
$\bar{b}_{3,16}=a_{3,13}$	$W+K_1;K_2+L_3,N$	$A_{3,4}$	$\bar{b}_{5,8}$	$N(K_2+L_3)$	$\tilde{\bar{b}}_{2,8},\bar{b}_{2,13},\tilde{\bar{b}}_{2,13}$
$\bar{b}_{3,17}\sim a^{-1}_{3,14}$	$W+K_1+M;K_2+L_3,N$	$A_{3,4}$	$\bar{b}_{4,6}$	$N(K_2+L_3)$	$\tilde{\bar{b}}_{2,8},\bar{b}_{2,14},\tilde{\bar{b}}_{2,14}$
$\bar{b}_{3,18}=a_{3,15}$	$W+3K_1;M,N$	$A_{3,4}$	$\bar{b}_{4,8}$	NM	$\tilde{\bar{b}}_{2,6},\bar{b}^{-3}_{2,16},\bar{b}^{3}_{2,19}$
$\tilde{\bar{b}}_{3,18}\sim a_{3,15}$	$W+\frac{1}{3}K_1;K_2+L_3,M$	$A_{3,4}$	$\tilde{\bar{b}}_{4,8}$	$M(K_2+L_3)$	$\bar{b}_{2,6},\bar{b}^{-1/3}_{2,16},\tilde{\bar{b}}^{-1/3}_{2,19}$
$\bar{b}_{3,19}=a_{3,16}$	$W;M,N$	$A^{1/2}_{3,5}$	$\bar{b}_{5,8}$	M^2/N	$b_{2,4},\tilde{\bar{b}}_{2,6},\tilde{\bar{b}}_{2,17}$
$\tilde{\bar{b}}_{3,19}\sim a_{3,16}$	$K_1;K_2+L_3,M$	$A^{1/2}_{3,5}$	$\bar{b}_{5,8}$	$M^2/(K_2+L_3)$	$\bar{b}_{2,4},\bar{b}_{2,6},\bar{b}_{2,17}$
$\bar{b}^{\epsilon}_{3,20}=a^{\epsilon}_{3,17}$ $\epsilon=1*[\epsilon=\pm1]$	$W+\epsilon(K_2+L_3);M,N$	$A^{1/2}_{3,5}$	$\tilde{\bar{b}}_{4,7}$	M^2/N	$\tilde{\bar{b}}_{2,6},\bar{b}^{-\epsilon}_{2,18},\bar{b}^{\epsilon}_{2,20}$
$\tilde{\bar{b}}^{\epsilon}_{3,20}\sim a^{\epsilon}_{3,17}$ $\epsilon=1*[\epsilon=\pm1]$	$K_1+\epsilon N;K_2+L_3,M$	$A^{1/2}_{3,5}$	$\bar{b}_{4,7}$	$M^2/(K_2+L_3)$	$\bar{b}_{2,6},\bar{b}^{\epsilon}_{2,18},\bar{b}^{\epsilon}_{2,20}$
$\bar{b}_{3,21}\sim a^{-1}_{3,18}$	$W-\frac{1}{3}K_1;K_2+L_3+Q,N$	$A^{1/3}_{3,5}$	self	$(K_2+L_3+Q)^3/N$	$\bar{b}_{2,11},\bar{b}^{1/3}_{2,19}\bar{b}_{2,21}$

TABLE V. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\overline{b}^{c}_{3,22} \sim a^{(c+1)/(c-1)}_{3,19}$ $c \neq 0, \frac{1}{3}, \pm 1$	$W + cK_1; K_2 + L_3, M$	$A^{h}_{3,5}$ $h = \frac{2c}{c-1}$, $-1 < c < \frac{1}{3}$ $h = \frac{c-1}{2c}$, $c > \frac{1}{3}$ or $c < -1$	$\tilde{\overline{b}}_{4,8}$	$(K_2 + L_3)M^{2c/(1-c)}$	$\overline{b}_{2,6}, \overline{b}^{c}_{2,16}, \tilde{\overline{b}}^{c}_{2,19}$
$\tilde{\overline{b}}^{c}_{3,22} = a^{(1-c)/(1+c)}_{3,19}$ $c \neq 0, \pm 1, -3$	$W - cK_1; M, N$	$A^{h}_{3,5}$ $h = \frac{c+1}{2}$, $-3 < c < 1$ $h = \frac{2}{c+1}$, $c > 1$ or $c < -3$	$\overline{b}_{4,8}$	$MN^{-(c+1)/2}$	$\tilde{\overline{b}}_{2,6}, \overline{b}^{c}_{2,16}, \overline{b}^{c}_{2,19}$
$\overline{b}^{c}_{3,23} = a^{(1-c)/(1+c)}_{3,20}$ $0 < \lvert c \rvert < 1$ $[c \neq 0, \pm 1]$	$W - cK_1; K_2 + L_3, N$	$A^{c}_{3,5}$	$\overline{b}_{4,9}$	$(K_2 + L_3)N^{-c}$	$\tilde{\overline{b}}_{2,8}, \overline{b}^{c}_{2,19}, \tilde{\overline{b}}^{c}_{2,19}$
$b_{2,1}$	L_3, N	$2A_1$	$b_{3,1}$	L_3, N	$b_{1,1}, b_{1,2}, b_{1,4}, \overline{b}_{1,6}$
$b_{2,2}$	W, L_3	$2A_1$	self	W, L_3	$b^{c}_{1,3}, b_{1,5}, \overline{b}_{1,6}$
$b^{d}_{2,3}$ $d > 0$ $[d \neq 0]$	$W + dL_3; N$	A_2	$b_{3,1}$	none	$b^{d}_{1,3}, b_{1,4}$
$b_{2,4} = \overline{a}_{2,21}$	$W; N$	A_2	$b_{5,1}$	none	$b_{1,4}, b_{1,5}$
$\tilde{b}_{2,4} \sim \overline{a}_{2,21}$	$K_1; K_2 + L_3$	A_2	$\sim b_{5,1}$	none	$\tilde{b}_{1,4}, \tilde{b}_{1,5}$
$\overline{b}_{2,5} = \tilde{a}_{2,1}$	$W + K_1, M$	$2A_1$	$\overline{b}_{3,6}$	$W + K_1, M$	$\overline{b}_{1,7}, \overline{b}_{1,8}, \overline{b}_{1,10}$
$\overline{b}_{2,6} \sim a_{2,2}$	$K_2 + L_3, M$	$2A_1$	$\sim \overline{b}_{6,2}$	$K_2 + L_3, M$	$\tilde{b}_{1,4}, \overline{b}_{1,8}, \tilde{\overline{b}}_{1,8}$
$\tilde{\overline{b}}_{2,6} = a_{2,2}$	M, N	$2A_1$	$\overline{b}_{6,2}$	M, N	$b_{1,4}, \overline{b}_{1,8}$
$\overline{b}_{2,7} = a_{2,4}$	$K_2 + L_3 + N, M$	$2A_1$	$a_{5,7}$	$K_2 + L_3 + N, M$	$\overline{b}_{1,8}, \tilde{\overline{b}}_{1,8}$
$\overline{b}_{2,8} \sim a_{2,5}$	$K_2 + L_3 - N, M$	$2A_1$	$\sim \overline{b}_{5,8}$	$K_2 + L_3 - N, M$	$\tilde{b}_{1,4}, \overline{b}_{1,8}, \tilde{\overline{b}}_{1,8}, \overline{b}_{1,9}$
$\tilde{\overline{b}}_{2,8} = a_{2,5}$	$K_2 + L_3, N$	$2A_1$	$\overline{b}_{5,8}$	$K_2 + L_3, N$	$b_{1,4}, \tilde{b}_{1,4}, \tilde{\overline{b}}_{1,8}, \overline{b}_{1,9}$
$\overline{b}_{2,9} = a_{2,6}$	W, K_1	$2A_1$	self	W, K_1	$b_{1,5}, \tilde{b}_{1,5}, \overline{b}_{1,7}, \overline{b}^{c}_{1,12}$
$\overline{b}_{2,10} = a_{2,8}$	K_1, N	$2A_1$	$\tilde{\overline{b}}_{3,7}$	K_1, N	$b_{1,4}, \tilde{b}_{1,5}, \overline{b}_{1,13}$
$\tilde{\overline{b}}_{2,10} \sim a_{2,8}$	$W, K_2 + L_3$	$2A_1$	$\overline{b}_{3,7}$	$W, K_2 + L_3$	$\tilde{b}_{1,4}, b_{1,5}, \tilde{\overline{b}}_{1,13}$
$\overline{b}_{2,11} \sim a^{-1}_{2,9}$	$K_2 + L_3 + Q, N$	$2A_1$	$\overline{b}_{3,20}$	$K_2 + L_3 + Q, N$	$b_{1,4}, \overline{b}_{1,11}$
$\overline{b}_{2,12} = \tilde{a}_{2,10}$	$W - K_1; M$	A_2	$\overline{b}_{3,6}$	none	$\overline{b}_{1,7}, \overline{b}_{1,8}$
$\tilde{\overline{b}}_{2,12} \sim \tilde{a}_{2,10}$	$W - K_1; K_2 + L_3 + N$	A_2	$\sim \overline{b}_{3,6}$	none	$\overline{b}_{1,7}, \tilde{\overline{b}}_{1,8}$
$\overline{b}_{2,13} = a_{2,11}$	$W + K_1; N$	A_2	$\overline{b}_{4,8}$	none	$b_{1,4}, \overline{b}_{1,7}$
$\tilde{\overline{b}}_{2,13} \sim \tilde{a}_{2,11}$	$W - K_1; K_2 + L_3$	A_2	$\tilde{\overline{b}}_{4,8}$	none	$\tilde{b}_{1,4}, \overline{b}_{1,7}$

TABLE V. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\tilde{\tilde{b}}_{2,13} \sim a_{2,11}$	$W+K_1;K_2+L_3$	A_2	$\tilde{\tilde{b}}_{4,8}$	none	$\tilde{b}_{1,4}, \bar{b}_{1,7}$
$\bar{b}_{2,14} \sim a^1_{2,12}$	$W+K_1+M;$ K_2+L_3	A_2	$\tilde{\tilde{b}}_{3,5}$	none	$\tilde{b}_{1,4}, \bar{b}_{1,10}$
$\tilde{\tilde{b}}_{2,14} \sim a^{-1}_{2,12}$	$W+K_1+M;N$	A_2	$\bar{b}_{3,5}$	none	$b_{1,4}, \bar{b}_{1,10}$
$\bar{b}_{2,15} = a_{2,13}$	$W-K_1,$ K_2+L_3-N	A_2	$\sim a_{3,5}$	none	$\bar{b}_{1,7}, \bar{b}_{1,9}$
$\bar{b}^d_{2,16} = \tilde{a}^{(-1+d)/(1+d)}_{2,14}$ $0<\|d\|<1$ $[d \neq 0, \pm 1]$	$W-dK_1;M$	A_2	$\bar{b}_{3,6}$	none	$\bar{b}_{1,8}, \bar{b}^d_{1,12}$
$\bar{b}_{2,17} = a_{2,15}$	$K_1;M$	A_2	$\bar{b}_{4,8}$	none	$\tilde{b}_{1,5}, \bar{b}_{1,8}$
$\tilde{\tilde{b}}_{2,17} \sim a_{2,15}$	$W;M$	A_2	$\bar{b}_{4,8}$	none	$b_{1,5}, \bar{b}_{1,8}$
$\bar{b}^\epsilon_{2,18} = a^\epsilon_{2,16}$ $\epsilon = 1 * [\epsilon = \pm 1]$	$K_1+\epsilon N;M$	A_2	$\bar{b}_{3,9}$	none	$\bar{b}_{1,8}, \bar{b}_{1,13}$
$\tilde{\tilde{b}}^\epsilon_{2,18} \sim \bar{a}^\epsilon_{2,16}$ $\epsilon = 1 * [\epsilon = \pm 1]$	$W-\epsilon(K_2+L_3);M$	A_2	$\tilde{\tilde{b}}_{3,9}$	none	$\bar{b}_{1,8}, \tilde{\tilde{b}}^\epsilon_{1,13}$
$\bar{b}_{2,19} \sim a^{(1-d)/(1+d)}_{2,18}$ $d>0, \neq 1$	$W-dK_1;N$	A_2	$\tilde{\tilde{b}}_{3,7}$	none	$b_{1,4}, \bar{b}^d_{1,12}$
$\tilde{\tilde{b}}_{2,19} \sim a^{(1-d)/(1+d)}_{2,18}$ $d>0, \neq 1$ $[d \neq 0, \pm 1]$	$W-dK_1;K_2+L_3$	A_2	$\bar{b}_{3,7}$	none	$\tilde{b}_{1,4}, \bar{b}^d_{1,12}$
$\bar{b}^\epsilon_{2,20} = a^\epsilon_{2,19}$ $\epsilon = 1 * [\epsilon = \pm 1]$	$K_1+\epsilon N;K_2+L_3$	A_2	$\bar{b}_{3,8}$	none	$\tilde{b}_{1,4}, \bar{b}_{1,13}$
$\tilde{\tilde{b}}^\epsilon_{2,20} \sim a^\epsilon_{2,19}$ $\epsilon = 1 * [\epsilon = \pm 1]$	$W+\epsilon(K_2+L_3);N$	A_2	$\tilde{\tilde{b}}_{3,8}$	none	$b_{1,4}, \tilde{\tilde{b}}^\epsilon_{1,13}$
$\bar{b}_{2,21} = a^{-1}_{2,20}$	$W-\frac{1}{3}K_1;$ K_2+L_3+Q	A_2	self	none	$\bar{b}_{1,11}, \bar{b}^{1/3}_{1,12}$
$b_{1,1}$	L_3+N	A_1	$b_{2,1}$	L_3+N	
$b_{1,2}$	L_3-N	A_1	$b_{2,1}$	L_3-N	
$b^e_{1,3}$, $e>0[e \neq 0]$	$W+eL_3$	A_1	$b_{2,2}$	$W+eL_3$	
$b_{1,4} = \bar{a}_{1,11}$	N	A_1	$b_{7,1}$	N	
$\tilde{b}_{1,4} \sim \bar{a}_{1,11}$	K_2+L_3	A_1	$\sim b_{7,1}$	K_2+L_3	
$b_{1,5} = \bar{a}_{1,12}$	W	A_1	$b_{4,2}$	W	
$\tilde{b}_{1,5} \sim \bar{a}_{1,12}$	K_1	A_1	$\sim b_{4,2}$	K_1	
$\bar{b}_{1,6} = e_{1,1}$	L_3	A_1	$e_{4,1}$	L_3	
$\bar{b}_{1,7} = a_{1,1}$	$W-K_1$	A_1	$a_{4,4}$	$W-K_1$	
$\bar{b}_{1,8} = a_{1,2}$	M	A_1	$\bar{b}_{5,8}$	M	
$\tilde{\tilde{b}}_{1,8} \sim a_{1,2}$	K_2+L_3+N	A_1	$\sim \bar{b}_{5,8}$	K_2+L_3+N	
$\bar{b}_{1,9} = a_{1,3}$	K_2+L_3-N	A_1	$a_{5,7}$	K_2+L_3-N	
$\bar{b}_{1,10} \sim \tilde{a}^{-1}_{1,4}$	$W+K_1+M$	A_1	$\bar{b}_{2,5}$	$W+K_1+M$	
$\bar{b}_{1,11} \sim a^{-1}_{1,5}$	K_2+L_3+Q	A_1	$\bar{b}_{3,20}$	K_2+L_3+Q	
$\bar{b}^e_{1,12} = a^{(1-e)/(1+e)}_{1,8}$ $0<e<1$ $[e>0, \neq 1]$	$W-eK_1$	A_1	$\bar{b}_{2,9}$	$W-eK_1$	
$\bar{b}_{1,13} = a^1_{1,9}$	K_1+N	A_1	$\bar{b}_{2,10}$	K_1+N	
$\tilde{\tilde{b}}^\epsilon_{1,13} = a^\epsilon_{1,9}$ $\epsilon = 1$ $[\epsilon = \pm 1]$	$W+\epsilon(K_2+L_3)$	A_1	$\tilde{\tilde{b}}_{2,10}$	$W+\epsilon(K_2+L_3)$	

TABLE VI. Subalgebras of $LO(3) \oplus LO(2)$.

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$c_{4,1}$	$D \oplus \{;A,B,C\}$	$A_1 \oplus A_{3,9}$	self	D, $A^2+B^2+C^2$	$c_{3,1}, c_{2,1}$
$c_{3,1}$	$;A,B,C$	$A_{3,9}$	$c_{4,1}$	$A^2+B^2+C^2$	$\bar{c}_{1,3}$
$c_{2,1}$	D,C	$2A_1$	self	D,C	$c_{1,1}, c^e_{1,2}, \bar{c}_{1,3}, \bar{c}_{1,4}$
$c_{1,1}$	D	A_1	$c_{4,1}$	D	
$c^e_{1,2}$ $e>0,\neq 1$	$D+eC$	A_1	$c_{2,1}$	$D+eC$	
$\bar{c}_{1,3}=d_{1,1}$	C	A_1	$d_{4,1}$	C	
$\bar{c}_{1,4}=e_{1,1}$	$A+D$	A_1	$e_{4,1}$	$A+D$	

TABLE VII. Subalgebras of $LO(2) \oplus LO(2,1)$.

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$d_{4,1}$	$C \oplus \{;D,E,F\}$	$A_1 \oplus A_{3,8}$	self	$C, D^2-E^2-F^2$	$d_{3,1}, \bar{d}_{3,2}, \bar{d}_{2,3}$
$d_{3,1}$	$;D,E,F$	$A_{3,8}$	$d_{4,1}$	$D^2-E^2-F^2$	$\bar{d}_{2,4}, \bar{d}_{1,2}$
$\bar{d}_{3,2}=a_{3,5}$	$C \oplus \{F;D+E\}$	$A_1 \oplus A_2$	self	C	$\bar{d}_{2,1}, \bar{d}_{2,2}, \bar{d}_{2,4}, \bar{d}^q_{2,5}$
$\bar{d}_{2,1}=a_{2,7}$	C,F	$2A_1$	self	C,F	$d_{1,1}, \bar{d}_{1,4}, \bar{d}^e_{1,6}$
$\bar{d}_{2,2}=a_{2,3}$	$C,D+E$	$2A_1$	$\bar{d}_{3,2}$	$C,D+E$	$d_{1,1}, \bar{d}_{1,5}, \bar{d}^e_{1,7}$
$\bar{d}_{2,3}=c_{2,1}$	C,D	$2A_1$	self	C,D	$d_{1,1}, \bar{d}_{1,2}, \bar{d}^e_{1,3}, \bar{d}^f_{1,8}$
$\bar{d}_{2,4}=a_{2,13}$	$F;D+E$	A_2	$\bar{d}_{3,2}$	none	$\bar{d}_{1,4}, \bar{d}_{1,5}$
$\bar{d}^d_{2,5}=a^d_{2,17}$ $d>0$ $[d\neq 0]$	$F+dC;D+E$	A_2	$\bar{d}_{3,2}$	none	$\bar{d}_{1,5}, \bar{d}^q_{1,6}$
$d_{1,1}$	C	A_1	$d_{4,1}$	C	
$\bar{d}_{1,2}=c_{1,1}$	D	A_1	$c_{4,1}$	D	
$\bar{d}^e_{1,3}=c^e_{1,2}$ $e>0,\neq 1$ $[e\neq 0,\pm 1]$	$D+eC$	A_1	$\bar{d}_{2,3}$	$D+eC$	
$\bar{d}_{1,4}=a_{1,1}$	F	A_1	$a_{4,4}$	F	
$\bar{d}_{1,5}=a_{1,3}$	$D+E$	A_1	$a_{5,7}$	$D+E$	
$\bar{d}^e_{1,6}=a^e_{1,7}$, $e>0$	$F+eC$	A_1	$a_{2,7}$	$F+eC$	
$\bar{d}^\epsilon_{1,7}=a^\epsilon_{1,6}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$C+\epsilon(D+E)$	A_1	$a_{2,3}$	$C+\epsilon(D+E)$	
$\bar{d}^\epsilon_{1,8}\sim e_{1,1}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$D+\epsilon C$	A_1	$\sim e_{4,1}$	$D+\epsilon C$	

TABLE VIII. Subalgebras of $LO(2,2)$.

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$e_{6,1}$	$\{;A_1,A_2,A_3\} \oplus \{;B_1,B_2,B_3\}$	$A_{3,8} \oplus A_{3,8}$	self	$A_1^2+A_2^2-A_3^2$, $B_1^2+B_2^2-B_3^2$	$\bar{e}_{5,1}, \tilde{\bar{e}}_{5,1}, e_{4,1}, \tilde{e}_{4,1}, \bar{e}_{3,8}$ $\bar{e}_{3,9}$
$\bar{e}_{5,1}=b_{5,1}$	$\{A_2;A_1-A_3\} \oplus \{;B_1,B_2,B_3\}$	$A_2 \oplus A_{3,8}$	self	$B_1^2+B_2^2-B_3^2$	$\bar{e}_{4,2}, \bar{e}_{4,3}, \bar{e}_{4,4}, \bar{e}_{3,4}$
$\tilde{\bar{e}}_{5,1}$	$\{;A_1,A_2,A_3\} \oplus \{B_2;B_1-B_3\}$	$A_{3,8} \oplus A_2$	self	$A_1^2+A_2^2-A_3^2$	$\bar{e}_{4,2}, \tilde{\bar{e}}_{4,3}, \tilde{\bar{e}}_{4,4}, \tilde{\bar{e}}_{3,4}$
$e_{4,1}$	$A_3 \oplus \{;B_1,B_2,B_3\}$	$A_1 \oplus A_{3,8}$	self	$A_3, B_1^2+B_2^2-B_3^2$	$e_{3,1}, \tilde{\bar{e}}_{3,4}, \bar{e}_{2,6}$
$\tilde{e}_{4,1}$	$\{;A_1,A_2,A_3\} \oplus B_3$	$A_{3,8} \oplus A_1$	self	$A_1^2+A_2^2-A_3^2, B_3$	$\tilde{e}_{3,1}, \bar{e}_{3,4}, \bar{e}_{2,6}$

TABLE VIII. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\bar{e}_{4,2}=a_{4,6}$	$\{A_2;A_1-A_3\}$ $\oplus\{B_2;B_1-B_3\}$	$A_2\oplus A_2$	self	none	$\bar{e}_{3,2},\tilde{e}_{3,2},\bar{e}_{3,3},\tilde{e}_{3,3},$ $\bar{e}_{3,5},\bar{e}_{3,6},\bar{e}^{c}_{3,7}$
$\bar{e}_{4,3}=b_{4,1}$	$(A_1-A_3)\oplus$ $(;B_1,B_2,B_3\}$	$A_1\oplus A_{3,8}$	$\bar{e}_{5,1}$	$A_1-A_3,B_1^2+B_2^2-B_3^2$	$e_{3,1},\bar{e}_{3,3},\bar{e}_{2,4}$
$\tilde{e}_{4,3}$	$\{;A_1,A_2,A_3\}$ $\oplus(B_1-B_3)$	$A_{3,8}\oplus A_1$	$\sim\bar{e}_{5,1}$	$A_1^2+A_2^2-A_3^2,B_1-B_3$	$\tilde{e}_{3,1},\tilde{e}_{3,3},\tilde{e}_{2,4}$
$\bar{e}_{4,4}=b_{4,2}$	$A_2\oplus\{;B_1,B_2,B_3\}$	$A_1\oplus A_{3,8}$	self	$A_2,B_1^2+B_2^2-B_3^2$	$e_{3,1},\tilde{e}_{3,2},\bar{e}_{2,5}$
$\tilde{e}_{4,4}$	$\{;A_1,A_2,A_3\}\oplus B_2$	$A_{3,8}\oplus A_1$	self	$A_1^2+A_2^2-A_3^2,B_2$	$\tilde{e}_{3,1},\bar{e}_{3,2},\tilde{e}_{2,5}$
$e_{3,1}=\bar{b}_{3,3}$	$;B_1,B_2,B_3$	$A_{3,8}$	$e_{6,1}$	$B_1^2+B_2^2-B_3^2$	$\bar{e}_{2,7},\tilde{e}_{1,1}$
$\tilde{e}_{3,1}$	$;A_1,A_2,A_3$	$A_{3,8}$	$e_{6,1}$	$A_1^2+A_2^2-A_3^2$	$\bar{e}_{2,7},e_{1,1}$
$\bar{e}_{3,2}=a_{3,4}$	$B_2\oplus\{A_2;A_1-A_3\}$	$A_1\oplus A_2$	self	B_2	$\bar{e}_{2,2},\bar{e}_{2,3},\bar{e}_{2,7},\bar{e}_{2,10},$ $\bar{e}^{d}_{2,11}$
$\tilde{e}_{3,2}$	$A_2\oplus\{B_2;B_1-B_3\}$	$A_1\oplus A_2$	self	A_2	$\bar{e}_{2,2},\tilde{e}_{2,3},\tilde{e}_{2,7},\tilde{e}_{2,10},$ $\tilde{e}^{d}_{2,11}$
$\bar{e}_{3,3}=a_{3,6}$	$(A_1-A_3)\oplus$ $\{B_2;B_1-B_3\}$	$A_1\oplus A_2$	$\bar{e}_{4,2}$	A_1-A_3	$\bar{e}_{2,1},\bar{e}_{2,3},\tilde{e}_{2,7},\bar{e}^{c}_{2,13}$
$\tilde{e}_{3,3}$	$(B_1-B_3)\oplus$ $\{A_2;A_1-A_3\}$	$A_1\oplus A_2$	$\tilde{e}_{4,2}$	B_1-B_3	$\bar{e}_{2,1},\tilde{e}_{2,3},\tilde{e}_{2,7},\tilde{e}^{c}_{2,13}$
$\bar{e}_{3,4}=b_{3,1}$	$B_3\oplus\{A_2;A_1-A_3\}$	$A_1\oplus A_2$	self	B_3	$\bar{e}_{2,4},\bar{e}_{2,5},\bar{e}_{2,7},\bar{e}^{d}_{2,12}$
$\tilde{e}_{3,4}$	$A_3\oplus\{B_2;B_1-B_3\}$	$A_1\oplus A_2$	self	A_3	$\tilde{e}_{2,4},\tilde{e}_{2,5},\tilde{e}_{2,7},\tilde{e}^{d}_{2,12}$
$\bar{e}_{3,5}=a_{3,12}$	$A_2+B_2;A_1-A_3,$ B_1-B_3	$A_{3,3}$	$\bar{e}_{4,2}$	$(A_1-A_3)/(B_1-B_3)$	$\bar{e}_{2,1},\bar{e}_{2,8},\bar{e}_{2,9},\bar{e}_{2,10},$ $\tilde{e}_{2,10}$
$\bar{e}_{3,6}=a_{3,13}$	$A_2-B_2;A_1-A_3,$ B_1-B_3	$A_{3,4}$	$a_{5,6}$	$(A_1-A_3)(B_1-B_3)$	$\bar{e}_{2,1},\bar{e}_{2,10},\tilde{e}_{2,10}$
$\bar{e}^{c}_{3,7}=a^{(1-c)/(1+c)}_{3,20}$ $0<\lvert c\rvert<1$ $[c\neq 0,\pm 1]$	$A_2+cB_2;$ A_1-A_3,B_1-B_3	$A^{h}_{3,5}$ $h=c,\ \lvert c\rvert<1$ $h=1/c,\ \lvert c\rvert>1$	$\bar{e}_{4,2}$	$(A_1-A_3)^{-c}(B_1-B_3)$	$\bar{e}_{2,1},\tilde{e}^{c}_{2,11},\tilde{e}^{c}_{2,11}$
$\bar{e}_{3,8}=d_{3,1}$	$;A_1-B_1,$ A_2+B_2,A_3-B_3	$A_{3,8}$	$d_{4,1}$	$(A_1-B_1)^2+(A_2+B_2)^2$ $-(A_3-B_3)^2$	$\bar{e}_{2,9},\bar{e}_{1,13}$
$\bar{e}_{3,9}\sim a_{3,24}$	$;A_1+B_1,A_2+B_2,$ A_3+B_3	$A_{3,8}$	$\sim a_{4,4}$	$(A_1+B_1)^2+(A_2+B_2)^2$ $-(A_3+B_3)^2$	$\bar{e}_{2,8},\bar{e}_{1,14}$
$\bar{e}_{2,1}=a_{2,5}$	A_1-A_3,B_1-B_3	$2A_1$	$a_{5,6}$	A_1-A_3,B_1-B_3	$\bar{e}_{1,3},\bar{e}_{1,4},\bar{e}_{1,10},\tilde{e}_{1,10}$
$\bar{e}_{2,2}=a_{2,6}$	A_2,B_2	$2A_1$	self	A_2,B_2	$\bar{e}_{1,2},\bar{e}^{e}_{1,5},\bar{e}_{1,11},\tilde{e}_{1,11}$
$\tilde{e}_{2,2}=a_{2,8}$	B_2,A_1-A_3	$2A_1$	$\bar{e}_{3,2}$	B_2,A_1-A_3	$\bar{e}_{1,6},\bar{e}_{1,10},\tilde{e}_{1,11}$
$\tilde{e}_{2,3}$	A_2,B_1-B_3	$2A_1$	$\tilde{e}_{3,2}$	A_2,B_1-B_3	$\tilde{e}_{1,6},\tilde{e}_{1,10},\bar{e}_{1,11}$
$\bar{e}_{2,4}=b_{2,1}$	A_1-A_3,B_3	$2A_1$	$\bar{e}_{3,4}$	A_1-A_3,B_3	$\tilde{e}_{1,1},\bar{e}_{1,7},\bar{e}_{1,8},\bar{e}_{1,10}$
$\tilde{e}_{2,4}$	A_3,B_1-B_3	$2A_1$	$\tilde{e}_{3,4}$	A_3,B_1-B_3	$e_{1,1},\tilde{e}_{1,7},\tilde{e}_{1,8},\tilde{e}_{1,10}$
$\bar{e}_{2,5}=b_{2,2}$	A_2,B_3	$2A_1$	self	A_2,B_3	$\tilde{e}_{1,1},\bar{e}^{e}_{1,9},\bar{e}_{1,11}$
$\tilde{e}_{2,5}$	A_3,B_2	$2A_1$	self	A_3,B_2	$e_{1,1},\tilde{e}^{e}_{1,9},\tilde{e}_{1,11}$
$\bar{e}_{2,6}\sim c_{2,1}$	A_3,B_3	$2A_1$	self	A_3,B_3	$e_{1,1},\bar{e}_{1,1},\bar{e}^{e}_{1,12},\bar{e}_{1,13},$ $\bar{e}_{1,14}$
$\bar{e}_{2,7}=b_{2,4}$	$A_2;A_1-A_3$	A_2	$\bar{e}_{5,1}$	none	$\bar{e}_{1,10},\bar{e}_{1,11}$
$\tilde{e}_{2,7}=b_{2,4}$	$B_2;B_1-B_3$	A_2	$\tilde{e}_{5,1}$	none	$\tilde{e}_{1,10},\tilde{e}_{1,11}$
$\bar{e}_{2,8}\sim\tilde{a}_{2,10}$	$A_2+B_2;$ $-A_1+A_3-B_1+B_3$	A_2	$\sim\tilde{a}_{3,3}$	none	$\bar{e}_{1,2},\bar{e}_{1,3}$

TABLE VIII. (continued)

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$\bar{e}_{2,9}=a_{2,13}$	$A_2+B_2;-A_1+A_3+B_1-B_3$	A_2	$a_{3,5}$	none	$\bar{e}_{1,2},\bar{e}_{1,4}$
$\bar{e}_{2,10}=\tilde{a}_{2,11}$	$A_2+B_2;-A_1+A_3$	A_2	$\tilde{a}_{4,5}$	none	$\bar{e}_{1,2},\bar{e}_{1,10}$
$\tilde{e}_{2,10}$	$A_2+B_2;-B_1+B_3$	A_2	$\sim\tilde{a}_{4,5}$	none	$\bar{e}_{1,2},\tilde{e}_{1,10}$
$\bar{e}^d_{2,11}=a^{(1-d)/(1+d)}_{2,13}$ $d>0,\neq 1$	$A_2+dB_2;-A_1+A_3$	A_2	$\bar{e}_{3,2}$	none	$\bar{e}^d_{1,5},\bar{e}_{1,10}$
$\tilde{e}^d_{2,11}$ $d>0,\neq 1$	$A_2+dB_2;-B_1+B_3$	A_2	$\tilde{e}_{3,2}$	none	$\bar{e}^d_{1,5},\tilde{e}_{1,10}$
$\bar{e}^d_{2,12}=b^d_{2,3}$ $d>0$ $[d\neq 0]$	$A_2-dB_3;-A_1+A_3$	A_2	$e_{3,4}$	none	$\bar{e}^d_{1,9},\bar{e}_{1,10}$
$\tilde{e}^d_{2,12}$ $d>0$ $[d\neq 0]$	$B_2-dA_3;-B_1+B_3$	A_2	$\tilde{e}_{3,4}$	none	$\tilde{e}^d_{1,9},\tilde{e}_{1,10}$
$\bar{e}^\epsilon_{2,13}=a^\epsilon_{2,19}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$B_2+\epsilon(A_3-A_1);B_1-B_3$	A_2	$\bar{e}_{3,3}$	none	$\bar{e}_{1,6},\tilde{e}_{1,10}$
$\tilde{e}^\epsilon_{2,13}$ $\epsilon=1$ $[\epsilon=\pm 1]$	$A_2+\epsilon(B_3-B_1);A_1-A_3$	A_2	$\tilde{e}_{3,3}$	none	$\tilde{e}_{1,6},\bar{e}_{1,10}$
$e_{1,1}$	A_3	A_1	$e_{4,1}$	A_3	
$\tilde{e}_{1,1}$	B_3	A_1	$\tilde{e}_{4,1}$	B_3	
$\bar{e}_{1,2}=a_{1,1}$	A_2+B_2	A_1	$a_{4,4}$	A_2+B_2	
$\bar{e}_{1,3}\sim a_{1,2}$	$-A_1+A_3-B_1+B_3$	A_1	$\sim a_{5,6}$	$-A_1+A_3-B_1+B_3$	
$\bar{e}_{1,4}=a_{1,3}$	$-A_1+A_3+B_1-B_3$	A_1	$a_{5,7}$	$-A_1+A_3+B_1-B_3$	
$\bar{e}^e_{1,5}=a^{(1-e)/(1+e)}_{1,8}$, $0<e<1$ $[e>0,\neq 1]$	A_2+eB_2	A_1	$\bar{e}_{2,2}$	A_2+eB_2	
$\bar{e}_{1,6}=a^1_{1,9}$	$B_2-A_1+A_3$	A_1	$\bar{e}_{2,3}$	$B_2-A_1+A_3$	
$\tilde{e}_{1,6}$	$A_2-B_1+B_3$	A_1	$\tilde{e}_{2,3}$	$A_2-B_1+B_3$	
$\bar{e}_{1,7}=b_{1,1}$	$-A_1+A_3+B_3$	A_1	$\bar{e}_{2,4}$	$-A_1+A_3+B_3$	
$\tilde{e}_{1,7}$	$-B_1+B_3+A_3$	A_1	$\tilde{e}_{2,4}$	$-B_1+B_3+A_3$	
$\bar{e}_{1,8}=b_{1,2}$	$-A_1+A_3-B_3$	A_1	$\bar{e}_{2,4}$	$-A_1+A_3-B_3$	
$\tilde{e}_{1,8}$	$-B_1+B_3-A_3$	A_1	$\tilde{e}_{2,4}$	$-B_1+B_3-A_3$	
$\bar{e}^e_{1,9}=b_{1,3}$, $e>0$	A_2-eB_3	A_1	$\bar{e}_{2,5}$	A_2-eB_3	
$\tilde{e}^e_{1,9}$, $e>0$	B_2-eA_3	A_1	$\tilde{e}_{2,5}$	B_2-eA_3	
$\bar{e}_{1,10}=b_{1,4}$	$-A_1+A_3$	A_1	$b_{7,1}$	$-A_1+A_3$	
$\tilde{e}_{1,10}=\tilde{b}_{1,4}$	$-B_1+B_3$	A_1	$\sim b_{7,1}$	$-B_1+B_3$	
$\bar{e}_{1,11}=b_{1,5}$	A_2	A_1	$\bar{e}_{4,4}$	A_2	
$\tilde{e}_{1,11}=\tilde{b}_{1,5}$	B_2	A_1	$\tilde{e}_{4,4}$	B_2	
$\bar{e}^e_{1,12}\sim c^{(1+e)/(1-e)}_{1,2}$, $0<\lvert e\rvert<1$ $[e\neq 0,\pm 1]$	A_3+eB_3	A_1	$\bar{e}_{2,6}$	A_3+eB_3	
$\bar{e}_{1,13}=c_{1,1}$	A_3-B_3	A_1	$c_{4,1}$	A_3-B_3	
$\bar{e}_{1,14}\sim d_{1,1}$	A_3+B_3	A_1	$d_{4,1}$	A_3+B_3	

TABLE IX. Subalgebras of LO(3,1).

Name and range of parameters	Generators	Isomorphism class	Normalizer	Invariants	Maximal subalgebras
$f_{6,1}$	$;L_1,L_2,L_3,K_1,K_2,K_3$	LO(3,1)	self	$L^2-K^2,(L,K)$	$\bar{f}_{4,1},\bar{f}_{3,4},\bar{f}_{3,5}$
$\bar{f}_{4,1}=a_{4,17}$	$K_1,L_1;L_2-K_3,L_3+K_2$	$A_{4,12}$	self	none	$\bar{f}_{3,1},\bar{f}_{3,2},\bar{f}^c_{3,3},\bar{f}_{2,2}$
$\bar{f}_{3,1}=a_{3,11}$	$K_1;L_2-K_3,L_3+K_2$	$A_{3,3}$	$f_{4,1}$	$(L_2-K_3)/(L_3+K_2)$	$\bar{f}_{2,1},\bar{f}_{2,3}$
$\bar{f}_{3,2}=a_{3,21}$	$L_1;L_2-K_3,L_3+K_2$	$A_{3,6}$	$a_{5,7}$	$(L_2-K_3)^2+(L_3+K_2)^2$	$\bar{f}_{2,1},\bar{f}_{1,1}$
$\bar{f}^c_{3,3}=a^c_{3,23}$ $c>0$ $[c\neq 0]$	$K_1-cL_1;L_2-K_3,L_3+K_2$	$A^{1/c}_{3,7}$	$\bar{f}_{4,1}$	$[(L_2-K_3)^2+(L_3+K_2)^2]^c\times\left(\frac{L_2-K_3-i(L_3+K_2)}{L_2-K_3+i(L_3+K_2)}\right)^i$	$\bar{f}_{2,1},\bar{f}^c_{1,4}$
$\bar{f}_{3,4}\sim a_{3,24}$	$;K_2,K_3,L_1$	$A_{3,8}$	$a_{4,4}$	$L_1^2-K_2^2-K_3^2$	$\bar{f}_{2,3},\bar{f}_{1,1}$
$\bar{f}_{3,5}=c_{3,1}$	$;L_1,L_2,L_3$	$A_{3,9}$	$c_{4,1}$	$L_1^2+L_2^2+L_3^2$	$\bar{f}_{1,1}$
$\bar{f}_{2,1}=a_{2,4}$	L_2-K_3,L_3+K_2	$2A_1$	$a_{5,7}$	L_2-K_3,L_3+K_2	$\bar{f}_{1,3}$
$\bar{f}_{2,2}=a_{2,7}$	L_1,K_1	$2A_1$	self	L_1,K_1	$\bar{f}_{1,1},\bar{f}_{1,2},\bar{f}^e_{1,4}$
$\bar{f}_{2,3}=\tilde{a}_{2,10}$	$K_1;L_2-K_3$	A_2	$\tilde{a}_{3,3}$	none	$\bar{f}_{1,2},\bar{f}_{1,3}$
$\bar{f}_{1,1}=d_{1,1}$	L_1	A_1	$d_{4,1}$	L_1	
$\bar{f}_{1,2}=a_{1,1}$	K_1	A_1	$a_{4,4}$	K_1	
$\bar{f}_{1,3}=a_{1,2}$	L_2-K_3	A_1	$a_{5,6}$	L_2-K_3	
$\bar{f}^e_{1,4}=a^e_{1,7}$, $e>0$ $[e\neq 0]$	K_1-eL_1	A_1	$\bar{f}_{2,2}$	K_1-eL_1	

TABLE X. Subalgebras of the irreducibly embedded LO(2,1).

Name	Generators	Isomorphism class	Normalizer	Invariants
$g_{3,1}$	$;K_1,K_2,L_3$	$A_{3,8}$	self	$K_1^2+K_2^2-L_3^2$
$\bar{g}_{2,1}\sim a^1_{2,20}$	$K_2;K_1-L_3$	A_2	self	none
$\bar{g}_{1,1}\sim c^2_{1,2}$	L_3	A_1	$\sim c_{2,1}$	L_3
$\bar{g}_{1,2}\sim a^1_{1,5}$	K_1-L_3	A_1	$\sim a^1_{3,18}$	K_1-L_3
$\bar{g}_{1,3}\sim a^{1/2}_{1,8}$	K_2	A_1	$\sim a_{2,6}$	K_2

TABLE XI. All proper subalgebras of LO(4,1).

Name and range of parameters	Generators	Ismorphism class	Normalizer	Invariants
$a_{7,1}$	$G;L_1,L_2,L_3,P_1,P_2,P_3$	LSim(3)	self	$(\mathbf{L},\mathbf{P})^2/\mathbf{P}^2$
$a_{6,1}$	$;L_1,L_2,L_3,P_1,P_2,P_3$	LE(3)	$a_{7,1}$	$(\mathbf{L},\mathbf{P}),\mathbf{P}^2$
$a_{5,1}$	$G,L_3;P_1,P_2,P_3$	$A^{0,1}_{5,35}$	self	$P_3^2/(P_1^2+P_2^2)$
$a_{4,1}$	$\{L_3;P_1,P_2\}\oplus P_3$	$A_{3,6}\oplus A_1$	$a_{5,1}$	$P_3,P_1^2+P_2^2$
$a_{4,2}$	$\{;L_1,L_2,L_3\}\oplus G$	$A_{3,9}\oplus A_1$	self	$G,\mathbf{L}^2$
$a_{4,3}$	$\{G;P_1,P_2,P_3\}$	$A^{1,1}_{4,5}$	$a_{7,1}$	$P_1/P_3,P_2/P_3$
$a^b_{4,4}$, $b>0$	$L_3+bG;P_1,P_2,P_3$	$A^{b,b}_{4,6}$	$a_{5,1}$	$\frac{P_3^2}{P_1^2+P_2^2},(P_1^2+P_2^2)\left(\frac{P_1+iP_2}{P_1-iP_2}\right)^{ib}$
$a_{4,5}$	$G,L_3;P_1,P_2$	$A_{4,12}$	self	none

Table XI. (continued)

Name and range of parameters	Generators	Ismorphism class	Normalizer	Invariants
$a_{3,1}$	P_1, P_2, P_3	$3A_1$	$a_{7,1}$	P_1, P_2, P_3
$a_{3,2}$	$\{G;P_3\} \oplus L_3$	$A_2 \oplus A_1$	self	L_3
$a_{3,3}$	$G;P_1, P_2$	$A_{3,3}$	$a_{4,5}$	P_1/P_2
$a_{3,4}$	$L_3;P_1, P_2$	$A_{3,6}$	$a_{5,1}$	$P_1^2+P_2^2$
$a_{3,5}^{\epsilon}, \epsilon = 1*$	$L_3+\epsilon P_3;P_1, P_2$	$A_{3,6}$	$a_{4,1}$	$P_1^2+P_2^2$
$a_{3,6}^{c}, c>0$	$L_3+cG;P_1, P_2$	$A_{3,7}^{c}$	$a_{4,5}$	$(P_1^2+P_2^2)\left(\frac{P_1+iP_2}{P_1-iP_2}\right)^{ic}$
$a_{3,7}$	$;L_1, L_2, L_3$	$A_{3,9}$	$a_{4,2}$	$\mathbf{L}^2$
$a_{2,1}$	P_1, P_2	$2A_1$	$a_{5,1}$	P_1, P_2
$a_{2,2}$	L_3, P_3	$2A_1$	$a_{3,2}$	L_3, P_3
$a_{2,3}$	G, L_3	$2A_1$	self	G, L_3
$a_{2,4}$	$G;P_3$	A_2	$a_{3,2}$	none
$a_{2,5}^{d}, d>0$	$L_3+dG;P_3$	A_2	$a_{3,2}$	none
$a_{1,1}$	P_3	A_1	$a_{5,1}$	P_3
$a_{1,2}$	G	A_1	$a_{4,2}$	G
$a_{1,3}^{\epsilon}, \epsilon = 1*$	$L_3+\epsilon P_3$	A_1	$a_{2,2}$	$L_3+\epsilon P_3$
$a_{1,4}^{e}, e>0$	L_3+eG	A_1	$a_{2,3}$	L_3+eG
$b_{4,1}$	$\{;G, P_3, C_3\} \oplus L_3$	$A_{3,8} \oplus A_1$	self	$P_3C_3+C_3P_3-2G^2, L_3$
$b_{3,1}$	$;G, P_3, C_3$	$A_{3,8}$	$b_{4,1}$	$P_3C_3+C_3P_3-2G^2$
$b_{2,1}$	P_3+C_3, L_3	$2A_1$	self	P_3+C_3, L_3
$b_{1,1}$	L_3	A_1	$b_{4,1}$	L_3
$b_{1,2}^{e}, e>0*$	$P_3+C_3+eL_3$	A_1	$b_{2,1}$	$P_3+C_3+eL_3$
$c_{6,1}$	$;L_1, L_2, L_3, P_1+C_1, P_2+C_2, P_3+C_3$	LO(4)	self	$\mathbf{L}^2+\frac{1}{4}(\mathbf{P}+\mathbf{C})^2, (\mathbf{L}, \mathbf{P}+\mathbf{C})$
$c_{4,1}^{\epsilon}, \epsilon = 1*$	$\{;L_1+\frac{\epsilon}{2}(P_1+C_1), L_2+\frac{\epsilon}{2}(P_2+C_2), L_3+\frac{\epsilon}{2}(P_3+C_3)\} \oplus \left(L_3-\frac{\epsilon}{2}(P_3+C_3)\right)$	$A_{3,9} \oplus A_1$	self	$L_3-\frac{\epsilon}{2}(P_3+C_3), \left(\mathbf{L}+\frac{\epsilon}{2}(\mathbf{P}+\mathbf{C})\right)^2$
$c_{3,1}^{\epsilon}, \epsilon = 1*$	$L_i+\frac{\epsilon}{2}(P_i+C_i), i=1,2,3$	$A_{3,9}$	$c_{4,1}^{\epsilon}$	$\left(\mathbf{L}+\frac{\epsilon}{2}(\mathbf{P}+\mathbf{C})\right)^2$
$d_{6\,1}$	$;L_1, L_2, L_3, P_1-C_1, P_2-C_2, P_3-C_3$	LO(3,1)	self	$\mathbf{L}^2-\frac{1}{4}(\mathbf{P}-\mathbf{C})^2, (\mathbf{L}, \mathbf{P}-\mathbf{C})$

V. ALL SUBALGEBRAS OF LO(3, 2)

A. Subalgebras of the maximal subalgebras

Let us now consider the subalgebras of each of the seven maximal subalgebras of LO(3,2). Denote a maximal subalgebra M_i $(i=1,\ldots,7)$ and the identity component of the corresponding Lie group GM_i. In Tables IV–X we present a respresentative of all GM_i conjugacy classes of subalgebras of M_i for $i=1,\ldots,7$. The maximal subalgebras are ordered as in Sec. IV.

In the tables the columns ordered from left to right give the following. The first column introduces a name for each subalgebra and presents the range of values of the parameters (if any) on which the algebra depends. Thus $a_{i,k}$, $b_{i,k}, \ldots, g_{i,k}$ are subalgebras of the similitude algebra, the optical algebra, LO(3)⊕LO(2), LO(2)⊕LO(2,1), LO(2,2), LO(3,1), and LO(2,1), respectively. The first subscript gives the dimension of the subalgebra; the second enumerates the subalgebras of the same dimension. For each dimension we first list decomposable subalgebras, then indecomposable ones. If a subalgebra has a superscript, e.g., $a_{4,8}^{\epsilon}$ or $a_{4,11}^{b}$, then the algebra depends on a parameter. The range of the parameter is given in the same column in the following manner. If only one range is given (e.g., $b>0, \neq 1$ for $a_{4,10}^{b}$), then the range is the same under O(3,2) and under the identity component of the corresponding maximal subgroup [in this case Sim(2,1)]. If the range under the appropriate maximal subgroup is larger than under O(3,2), then this larger range is given in square brackets (e.g., for $a_{4,8}^{\epsilon}$ we indicate $\epsilon=1$ $[\epsilon=\pm 1]$, meaning that $a_{4,8}^{1}$ is conjugate to $a_{4,8}^{1}$ under O(3,2) [and even $SO_0(3,2)$] but not under the identity component of Sim(2,1). An asterisk after the range under O(3,2) (as in $a_{4,12}^{\epsilon}$ or $b_{4,4}^{b}$) indicates that the range should be doubled

if conjugacy is considered under $SO_0(3,2)$ [or $O_1(3,2)$], rather than under $O(3,2)$ [or $O_2(3,2)$ or $SO(3,2)$]. Thus $\epsilon = 1*$ indicates that $\epsilon = \pm 1$ under $SO_0(3,2)$, $b > 0*$ indicates that $b \neq 0$, $-\infty < b < \infty$ under $SO_0(3,2)$. If the symbol of an algebra has a "tilde" on it, e.g., $\tilde{a}_{4,14}$, then it is conjugate under $O(3,2)$ to the algebra with the same symbol but without a tilde [e.g., $\tilde{a}_{4,14}$ is $O(3,2)$ conjugate to $a_{4,14}$]. Note, however, that two such subalgebras are not mutually conjugate under the corresponding maximal subalgebra (if they were, then one of them would not be listed at all).

If the symbol of an algebra has a bar on it, e.g., $\bar{a}_{4,18}$, then in the final list of $O(3,2)$ conjugacy classes of subalgebras of $LO(3,2)$ this subalgebra will be associated with a different maximal subalgebra. This can occur for one of two reasons:

1. Its normalizer or one of its higher normalizers (i.e., a normalizer of a normalizer, or its normalizer, etc.) in $LO(3,2)$, lies in a different maximal subalgebra. This is always indicated in column 1. e.g., we have $\bar{a}_{4,18} = b_{4,5}$, meaning that these two algebras are equal [$\bar{a}_{4,18} \sim b_{4,5}$ would indicate that they are $O(3,2)$ conjugate] and that the normalizer is in the "b" list (list of subalgebras of the optical algebra).

2. The algebra is self-normalizing in $O(3,2)$ and has already occurred in an earlier list of our ordered set. Thus $\bar{b}_{6,2} = a_{6,2}$ is self-normalizing and could thus be listed in the final a or b list. By choice we define it to belong to the list in which it first occurs.

In column 2 we list the generators of each subalgebra. Those to the right of a semicolon also lie in the derived algebra. Column 3 presents the isomorphism class of the subalgebras. We make use of the classification of real Lie algebras of dimension $d \leq 5$ due to Mubarakzyanov[40] and reproduced in a modified form in a previous article.[7] The commutation relations characterizing each isomorphy class of algebras are given in Ref. 7. For $d \geq 6$ no such classification is available, so we either give the name of such an algebra [e.g., $LE(2,1)$] or indicate that the algebra is a semidirect product (e.g., $F \supset A^0_{5,30}$ indicates a semidirect product of the dilation subalgebra F with a five-dimensional ideal of isomorphy type $A^0_{5,30}$).

The normalizer in $LO(3,2)$ of each subalgebra is given in column 4. We recall that the normalizer $\mathrm{Nor}_L X$ of a subalgebra $X \subset L$ is an algebra satisfying $[\mathrm{Nor}_L X, X] \subseteq X$ and $X \subseteq \mathrm{Nor}_L X \subseteq L$. Since the list is normalized, for each algebra that stays in the final list (i.e., for all "untilded" and "unbarred" subalgebras) the normalizer is itself in the list. For tilded subalgebras this is not necessarily so, e.g., the normalizer of $a_{3,16}$ is not equal to $a_{5,6}$ but only $O(3,2)$ conjugate to $a_{5,6}$. The invariants of the algebras are given in the fifth column. Notice that we include polynomial invariants (Casimir operators), rational invariants, and general invariants, as defined and calculated earlier.[7] The maximal subalgebras of each subalgebra are given in column 6. Their classification is again under the relevant maximal subgroup of $O(3,2)$.

The derivation of the results in Tables IV—X is a straightforward application of the classification methods presented earlier[3–8] and developed further in Sec. III.

The subalgebras of $LSim(2,1)$ (Table IV) were classified earlier;[8] here we present for the first time a normalized list and present additional information on the subalgebras. Table V is new, however the subalgebras of the Schrödinger algebra itself have already been listed.[11] The classification of the subalgebras of $LO(3) \oplus LO(2)$, $LO(2) \oplus LO(2,1)$, and $LO(2,2) \sim LO(2,1) \oplus LO(2,1)$ was performed via the Goursat method[38,3,4] for direct sums of Lie algebras. The subalgebras of $LO(3,1)$ and $LO(2,1)$ have been classified a long time ago[12] (see also Ref. 41).

When constructing the maximal subalgebras of subalgebras in one of the lists (columns 6) it proved to be very helpful to make use of a recent classification[42] of subalgebras of all real Lie algebras of dimension $d \leq 4$.

B. Final list of O(3, 2) conjugacy classes of subalgebras of LO(3, 2)

Tables IV—X have been so arranged that it is a simple matter to merge them into one final list of representatives of all $O(3,2)$ classes of subalgebras of $LO(3,2)$. Indeed, all we have to do is take these lists and omit all subalgebras that have a bar or a tilde on the symbol in columns 1. Notice that Table IV of subalgebras of $LSim(2,1)$ remains essentially as it stands. The only "barred" subalgebras to be included elsewhere are $\bar{a}_{4,18}$, $\bar{a}_{3,25}$, $\bar{a}_{2,21}$, $\bar{a}_{1,11}$, and $\bar{a}_{1,12}$ (all go into the "b" list of Table V), and $\bar{a}_{1,10}$ (which goes into the "d" list of Table VII). The tilded entries must just be dropped.

Relatively few algebras from the other lists will figure in the final list. From Table V these are $b_{7,1}$, $b_{6,1}$, $b_{5,1}$, $b_{5,2}$, $b_{4,1}, \ldots, b_{4,5}$, $b_{3,1}$, $b_{3,2}$, $b_{2,1}, \ldots, b_{2,4}$, $b_{1,1}, \ldots$, and $b_{1,5}$. Table VI contributes $c_{4,1}$, $c_{3,1}$, $c_{2,1}$, $c_{1,1}$ and $c^e_{1,2}$, Table VII only $d_{4,1}$, $d_{3,1}$, and $d_{1,1}$. Table VIII contributes $e_{6,1}$, the irreducible subalgebra $e_{4,1}$, and its subalgebras $e_{3,1}$ and $e_{1,1}$. Tables IX and X contribute only $f_{6,1}$ and $g_{3,1}$, respectively. In all tables the "unbarred" algebras, that survive in the final list, precede the "barred" ones.

Notice that a closed Lie group corresponds to each Lie algebra in our lists. The only exception is in Table VI where the subalgebra $c^e_{1,2}$ is an algebra for any e, but generates a closed Lie group only if e is rational.

VI. SUBALGEBRAS OF LO(4, 1)

The subalgebras of the de Sitter algebra $LO(4,1)$ have already been classified.[5] Without going into any details we shall just present a normalized list of representatives of $O(4,1)$ conjugacy classes of $LO(4,1)$ subalgebras.

A basis for $LO(4,1)$ is introduced in (2.24); the commutation relations are in Table III. For our purposes a more convenient basis is one adapted to the similitude subalgebra $LSim(3)$,

$$G, \quad L_1 = C, \quad L_2 = -E, \quad L_3 = F,$$
$$P_1 = D - K, \quad P_2 = B - J, \quad P_3 = A - H, \tag{6.1}$$
$$C_1 = D + K, \quad C_2 = B + J, \quad C_3 = A + H.$$

In (6.1) G generates dilations, L_i rotations, P_i translations, and C_i special conformal transformations. The commutation relations are:

$$[L_i,L_k]=\epsilon_{ikl}L_l,\quad [L_i,P_k]=\epsilon_{ikl}P_l,\quad [P_i,P_k]=0,$$

$$[G,L_i]=0,\quad [G,P_i]=-P_i,$$

$$[L_i,C_k]=\epsilon_{ikl}C_l,\quad [C_i,C_k]=0,\quad [G,C_i]=C_i \qquad (6.2)$$

$$[C_i,P_k]=2(\epsilon_{ikl}L_l-\delta_{ik}G),\quad i,k=1,2,3.$$

The maximal subalgebras of LO(4,1) are:

(a) The similitude algebra LSim(3): $\{G,L_i,P_i\}$ $i=1,2,3$;

(b) LO(2,1)$\oplus$LO(2): $\{P_3,C_3,G\}\oplus\{L_3\}$;

(c) LO(4): $\{L_i,P_i+C_i\}$, $i=1,2,3$;

(d) LO(3,1): $\{L_i,P_i-C_i\}$, $i=1,2,3$.

All subalgebras of LSim(3) have normalizers in LSim(3), except for $\{L_3\}$ whose normalizer is $b_{4,1}$ $=$LO(2,1)$\oplus$LO(2). The "b" list for LO(2,1)$\oplus$LO(2) provides five new subalgebras, the "c" list for LO(4) provides three, the "d" list for LO(3,1) just one. For the individual lists and mutual inclusions we refer to our earlier paper.[5]

An overall list of LO(4,1) subalgebras is given in Table XI. The ordering is the same as in Tables IV–X. The range of parameters is given considering conjugacy under $SO_0(4,1)$ and under O(4,1). As discussed in Sec. III this automatically also provides us with conjugacy under the other locally isomorphic groups $O_1(4,1)$ $O_2(4,1)$, and SO(4,1).

VII. MISSING LABEL OPERATORS

One of the motivations for a subalgebra classification is to define bases for the representations of the Lie algebra in question. The bases are to correspond to the different chains of subalgebras found; they will be the simultaneous eigenstates of the invariant operators of all the algebras belonging to the chain concerned (these are listed for each subalgebra in Tables IV–XI).

For most chains of maximal subalgebras the bases are determined uniquely in this way, but in a few cases there are remaining degeneracies which may be resolved by the introduction of additional commuting operators, the so-called missing label operators.

With plausible assumptions about the number of labels required to determine a basis, Racah's method of counting, valid for semisimple algebras, can be extended to arbitrary Lie algebras.[43,44] Then the number of missing labels at a stage $G\supset H$ of a subalgebra chain is

$$n=\tfrac{1}{2}(r_G-l_G-r_H-l_H)+l',$$

where r_G, r_H are the dimensions and l_G, l_H are the number of invariants of G,H respectively; l' is the number of G invariants which depend on H elements only. Then it can be shown[44] that the number of available missing label operators (H invariant functions of G elements which are independent of each other and of G and H invariants) is $2n$, i.e., twice the number actually missing. The missing label operators are found by solving certain partial differential equations;[7,44] for semisimple algebra subalgebra pairs other methods may also be used.[19]

In maximal subalgebra chains of LO(4,1) there is just one missing label, at the step LO(4,1)$\supset$LO(2,1) $\oplus$LO(2). In maximal subalgebra chains of LO(3,2) there are two missing labels at the step LO(3,2)$\supset$LO(2,1) and one at each of the steps LO(3,2)$\supset$LO(3)$\oplus$LO(2), LO(3,2)$\supset$LO(2,1)$\oplus$LO(2) and $b^b_{4,4}\supset b^b_{2,3}$.

The algebra $b^b_{4,4}$ and its subalgebra $b^b_{2,3}$ are subalgebras of LOpt(2,1) and appear in Table V. Two functionally independent missing label operators are $(Q^2+M^2)/N^2$ and $N^{2b}(Q-iM)^i(Q+iM)^{-i}$.

The other missing labels all involve semisimple subalgebras of LO(3,2) or LO(4,1) and the missing label operators may be copied, *mutatis mutandi*, from the known operators for the corresponding subalgebras of LO(5). Thus for LO(3,2)$\supset$LO(3)$\oplus$LO(2) the elements of LO(3,2) decompose into three LO(3) vectors and one scalar. The vectors may be denoted by [see (4.13)] $L_1=C$, $L_2=B$, $L_3=A$ [the elements of LO(3)], $U_1=-F$, $U_2=H$, $U_3=-K$, and $V_1=-E$, $V_2=G$, $V_3=-J$; the scalar is the LO(2) element D. Corresponding components of U and V transform like the 1′- and 2′-components of a vector under the rotations generated by D about a 3′ axis. The missing label operators may be chosen as[19]

$$\sum_{ijk}\epsilon_{ijk}\{U_iV_jL_k\}_{\text{symmetrized}}\quad\text{and}\quad\{(\sum_i U_iL_i)^2+(\sum_i V_iL_i)^2\}_{\text{symmetrized}}.$$

The missing label operators for LO(3,2)$\supset$LO(2,1) $\oplus$LO(2) and for LO(4,1)$\supset$LO(2,1)$\oplus$LO(2) have a similar form.

For LO(3,2)$\supset$LO(2,1) there are two missing labels, and four available missing label operators. An integrity basis for the parallel problem LO(5)$\supset$LO(3), of interest in connection with quadrupole nuclear vibrations, has been worked out, and is to be published.[45] The details are rather complicated and are not reproduced here.

Sometimes nonmaximal subalgebra chains are of interest; perhaps because of broken symmetry, some of the algebras in a chain are not physically relevant to a particular problem. If an algebra with l independent invariants is discarded, there are at most l new missing labels, and twice as many corresponding missing lable operators. Of the new operators, half are just the invariants of the discarded algebra; the others are really new and may be the physically relevant ones. They can be determined by the standard methods.

VIII. CONCLUSIONS

The general result of this article is a significant refinement of a previously presented subgroup classification algorithm.[3,5] The refinement is the provision of normalized ordered lists of subalgebras and greatly simplifies the task of merging several sublists into one.

The specific result contained here is a complete classification of the subalgebras of both de Sitter algebras LO(3,2) and LO(4,1) and hence a classification of all closed connected Lie subgroups of these two de Sitter groups.

The ground is now well prepared for the final article of this series, namely a classification of all closed connected subgroups of the conformal group O(4,2). Work is in progress on applications of the classification of subgroups, in particular on invariant equations for each subgroup.

ACKNOWLEDGMENTS

In conclusion we thank G. Burdet, M. Moshinsky, M. Perrin, and M. Perroud for discussions.

[1]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 15, 1378 (1974).
[2]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 15, 1932 (1974).
[3]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 16, 1597 (1975).
[4]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 16, 1613 (1975).
[5]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 717 (1976).
[6]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 977 (1976).
[7]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 986 (1976).
[8]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, Can. J. Phys. 54, 950 (1976).
[9]M. Perroud, J. Math. Phys. 17, 1028 (1976).
[10]J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math. Phys. 18, 72 (1977).
[11]C.P. Boyer, R.T. Sharp, and P. Winternitz, J. Math. Phys. 17, 1439 (1976).
[12]P. Winternitz and I. Fris̆, Yad. Fiz. 1, 889 (1965) [Sov. J. Nucl. Phys. 1, 636 (1965)]; P. Winternitz, I. Lukač̆, and Ya. A. Smorodinskiĭ, Yad. Fiz. 7, 192 (1968) [Sov. J. Nucl. Phys. 7, 139 (1968)].
[13]J. Patera and P. Winternitz, J. Math. Phys. 14, 1130 (1973).
[14]J. Patera and P. Winternitz, J. Chem. Phys. 65, 2725 (1976).
[15]W. Miller, Jr., SIAM J. Math. An. 5, 626 (1974); 5, 822 (1974); in *Theory and Applications of Special Functions* (Academic, New York, 1975).
[16]E.G. Kalnins and W. Miller, Jr., J. Math. Phys. 15, 1025, 1263, 1728 (1974); 17, 331, 356, 369 (1976).
[17]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., J. Math. Phys. 16, 499, 512 (1975); Nagoya Math. J. 60, 35 (1976).
[18]E.G. Kalnins, W. Miller, Jr., and P. Winternitz, SIAM J. Appl. Math. 30, 630 (1976).
[19]B.R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. 15, 1787 (1974); R.T. Sharp, J. Math. Phys. 16, 2050 (1975).
[20]M. Moshinsky, J. Patera, R.T. Sharp, and P. Winternitz, Ann. Phys. 95, 139 (1975).
[21]P. Kramer, T. Seligman, and M. Moshinsky, in *Group Theory and Its Applications*, Vol. 3, edited by E. Loebl (Academic, New York, 1975).
[22]J.D. Louck, M. Moshinsky, and K.B. Wolf, J. Math. Phys. 14, 692, 696 (1973).
[23]M. Moshinsky, J. Patera, and P. Winternitz, J. Math. Phys. 16, 82 (1975); M. Moshinsky and J. Patera, J. Math. Phys. 16, 1866 (1975).
[24]P. Winternitz, Ya.A. Smorodinskiĭ, M. Uhlir̆, and I. Fris̆, Yad. Fiz. 4, 625 (1966) [Sov. J. Nucl. Phys. 4, 444 (1967)].
[25]A. Makarov, Ya.A. Smorodinskiĭ, Kh.V. Valiev, and P. Winternitz, Nuovo Cimento 52 A, 1061 (1967).
[26]N. Ya. Vilenkin and Ya.A. Smorodinskiĭ, Zh. Eksp. Teor. Fiz. 46, 1793 (1964) [Sov. Phys. JETP 19, 1209 (1964)].
[27]P. Winternitz, Ya.A. Smorodinskiĭ, and M.B. Shefel' Yad. Fiz. 7, 1325 (1968) [Sov. J. Nucl. Phys. 7, 785 (1968)].
[28]E.G. Kalnins, J. Patera, R.T. Sharp, and P. Winternitz, in *Group Theory and Its Applications*, Vol. 3, edited by E.M. Loebl (Academic, New York, 1975).
[29]J. Bystricky̆, F. Lehar, J. Patera, and P. Winternitz, Phys. Rev. D 13, 1276 (1976).
[30]W. de Sitter, Mon. Not. 78, 3 (1917).
[31]G.E. Lemaître, J. Math. and Phys. (MIT) 4, 188 (1925).
[32]H.P. Robertson, Phil. Mag. 5, 835 (1928).
[33]R. Lipschitz, *Untersuchungen über die Summen von Quadraten* (Cohen, Bonn, 1886).
[34]H. Zassenhaus, Arch. Math. 13, 434 (1962).
[35]J.A. Dieudonné, *La Géométrie des Groupes Classiques* (Springer, Berlin, 1975), Vol. 5.
[36]H. Zassenhaus, Lie groups and Lie Algebras (Lectures at 15th Session of the Séminaires de Mathématiques Supérieurs, Montréal, 1976).
[37]E. Witt, J. Reine Angew. Math. 176, 31 (1937).
[38]E. Goursat, Ann. Sci. Ecole Normale Supérieure, (3), 6, 9 (1889).
[39]G. Burdet, J. Patera, M. Perrin, and P. Winternitz, Ann. Sci. Math. Québec, No. 2 (1977).
[40]G.M. Mubarakzyanov, Izv. Vyshikh Uchebn. Zavedeniĭ, Mat. 1 (32) 114; 3 (34), 99; 4 (35), 104 (1963).
[41]D. Finkelstein, Phys. Rev. 100, 924 (1955); J. Segercrantz, Ann. Acad. Sci. Fenn. 6, 275 (1968); R. Shaw, Q.J. Math. Oxford 20, 333 (1969).
[42]J. Patera and P. Winternitz, J. Math. Phys. 18, 1449 (1977).
[43]G. Racah, Lecture notes "Group Theory and Spectroscopy" Institute for Advanced Study, Princeton, New Jersey, 1951. Reprinted in Ergebs. Exakt. Naturw. 37, 28 (1965).
[44]A. Peccia and R.T. Sharp, J. Math. Phys. 17, 1313 (1976).
[45]R. Gaskell, A. Peccia, and R.T. Sharp, J. Math. Phys. (submitted).

Finite and infinite measurement sequences in quantum mechanics and randomness: The Everett interpretation

Paul A. Benioff

Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439
(Received 3 November 1976)

The quantum mechanical description of both a finite and infinite number of measurement repetitions, as interactions between copies of an object system and a record system, are considered here. States describing the asymptotic situation of an infinite number of repetitions are seen to have some interesting properties. The main construction of the paper is the association of states to sequential tests for randomness. To each such test T and each positive integer m one can associate states Θ_n^{Tm} and $\Theta^{\infty Tm}$ corresponding respectively to those length-n and finite outcome sequences which pass test T at the significance level 2^{-m}. Following the methods of Martin Löf, a universal sequential test V, which includes an infinity of sequential statistical tests for randomness, is given and the corresponding states Θ_n^{Vm} and $\Theta^{\infty Vm}$ are discussed. Finally, a possible use of these states in the Everett interpretation of quantum mechanics is discussed.

I. INTRODUCTION

In many discussions of the measurement process in quantum mechanics a measurement is (partly) described as an interaction between an object system and an apparatus system in isolation from the surroundings. The corresponding object + apparatus system state evolves under the action of a unitary operator from an initial to a final state. n repetitions of the measurement can be described by a product state for n copies of the object + apparatus systems each undergoing one interaction. The state describing n repetitions of a measurement of a discrete observable can be given as a sum over components, each component corresponding to an outcome sequence of length n.

A natural requirement on the outcome sequences obtained from measurement repetitions is that they be (initial segments of) random sequences. Corresponding to this is the question of the existence of states describing measurement repetitions but which correspond, in some sense, to random outcome sequences only.

This question, which is the concern of this paper, is nontrivial. At first thought it might seem that such a question has an obvious answer. Construct a state corresponding to an infinite number of measurement repetitions as a sum over components each corresponding to an infinite outcome sequence and restrict the sum to the random sequences. The difficulty with this is that the state so constructed is, in general, the zero vector.

A related problem is the following. Definitions of randomness (for infinite sequences) are given in terms of sets of measure zero with respect to a product measure—a sequence is random if it does not lie in any set of measure zero with the set restricted to be of a certain type. Now it is possible to construct states Θ^∞ corresponding to an infinite number of repetitions of measurements (Sec. II) as well as the states Θ_n corresponding to n repetitions. This suggests that one exclude from Θ^∞ and Θ_n all components corresponding to infinite sequences lying in the sets of measure zero.

However, this is not possible—for no n does Θ_n contain such components and Θ^∞ cannot be represented in a form suitable for such an exclusion. In fact (Sec. II) Θ^∞ is orthogonal to the Hilbert space spanned by all renormalized infinite outcome sequence component states.

Still another problem is the following: DeWitt[1] has considered a particular test for randomness and has constructed states which correspond, for each n, to those sequences of length n which pass the test. He has stated without proof that the construction can be extended to any finite combination of the infinite number of tests for randomness. However, it is not clear how these states are to be constructed.

The main point of this paper is to show that if one uses the statistical idea of failing a test for randomness at some significance level, then some of the above problems can be taken care of. In particular, one can, for each significance level 2^{-m}, construct states Θ_n^{Vm} and $\Theta^{\infty Vm}$ which include respectively, component states for exactly those length $-n$ and all finite outcome sequences which pass, at level 2^{-m}, an infinity of tests for randomness. One can also associate to $\Theta^{\infty Vm}$ exactly those infinite outcome sequences which pass at level 2^{-m} all the tests for randomness.

The essentials of the construction use the methods of Martin-Löf[2] as follows:

Each set D of measure zero in a definition of randomness corresponds to a sequential test for randomness. Each such test T corresponds to a prescription, for each significance level 2^{-m} with $m = 1, 2, \cdots$ for rejecting sequences as being nonrandom. For each m, T gives a set T_m of finite sequences such that any sequence with an initial segment in T_m is rejected as failing T at level m (or, equivalently, at significance level 2^{-m}). By a canonical correspondence, each set T_m corresponds to an open Borel cover $W(T_m)$ of D of measure $\leq 2^{-m}$ such that $D \subseteq W(T_m)$. Following Martin-Löf,[2] one then constructs for the infinity of tests in the randomness definition, a universal test V which includes all the tests.

The properties of sequential tests are such that for each m one can construct a correspondence between

the sets T_m and quantum states corresponding to just those finite sequences in T_m. The application of this correspondence to the universal test gives a sequence of states $X_n^{V_m}$ with $n=0, 1, \cdots$ and $X^{\infty V_m}$ which correspond respectively to all sequences of length n in V_m. The desired result is obtained by defining the states $\Theta_n^{V_m}=\Theta_n-X_n^{V_m}$ and $\Theta^{\infty V_m}=\Theta^{\infty}-X^{\infty V_m}$ [Eqs. (20) and (21)]. For each m the sequence of states $\Theta_n^{V_m}$ with $n=0, 1, \cdots$ and $\Theta^{\infty V_m}$ correspond respectively to all sequences of length n and all finite outcome sequences which pass the universal test for randomness at level m, $\Theta^{\infty V_m}$ can also be associated with exactly those infinite sequences which are random at level m.

In Sec. II some preliminary results are given. The properties of sequential tests and universal sequential tests are given and discussed in Sec. III A, and in Sec. III B the correspondence between tests and states is given and discussed. In Sec. IV A some further aspects of the construction are discussed. Section IV B concludes with an application of the results obtained to the Everett[3] interpretation of quantum mechanics.

II. PRELIMINARIES

Let H be the Hilbert space for a single object system and $\mathcal{K}$ the Hilbert space for the record system. Then $\mathcal{H}=H\otimes\mathcal{K}$ is the overall system space. Consider the measurement of an observable A which is assumed to be purely discrete. If σ is the eigenvalue set of A, then

$$A=\sum_{a\in\sigma} aP_a,$$

where P_a is the projection operator corresponding to eigenvalue a.

One now assumes the existence of a unitary operator U acting on $\mathcal{H}$ such that for any state $\psi_a\in P_aH$ and for each a in σ

$$U\psi_a\otimes\phi_0=\psi_a\otimes\phi_a \tag{1}$$

Here ϕ_0 is the input state of the record system and ϕ_a is the final record system state corresponding to outcome a. One also has for each a, $a'\in\sigma$,

$$(\phi_0,\phi_a)=0,\quad (\phi_a,\phi_{a'})=\delta_{a,a'}. \tag{2}$$

For a general system state ψ, one has by linearity,

$$U\psi\otimes\phi_0=\sum_{a\in\sigma} P_a\psi\otimes\phi_a, \tag{3}$$

where P_a acts in H.

The extension of the above to a fixed, finite number of repetitions is well known and won't be repeated here. For an infinite number of repetitions one considers the preparation of an infinite number of copies of the object and recording systems. (It is not necessary to consider copies of the measurement apparatus; one such system, interacting with successive copies of the object + record system will suffice.)

The initial object + record system state is

$$\Theta_0=\bigotimes_{j=1}^{\infty}(\psi(j)\otimes\phi_0(j)), \tag{4}$$

where $\psi(j)=\psi\in H$ and $\phi_0(j)=\phi_0\in\mathcal{K}$ for each j. The state Θ_n corresponding to n completed repetitions is given by

$$\Theta_n=\bigotimes_{j=1}^{n}\left(\sum_{a_j\in\sigma} P_{a_j}\psi\otimes\phi_{a_j}\right)\otimes\Theta_0^n \tag{5}$$

with[4,5]

$$\Theta_0^n=\bigotimes_{k>n}(\psi\otimes\phi_0)_k. \tag{6}$$

One also has for each n,

$$(\Theta_n,\Theta_n)=(\Theta_0,\Theta_0)=1.$$

Note that Θ_n is the same as the state Θ_n' given by

$$\Theta_n'=\sum_{\nu_n\in\sigma^n}\bigotimes_{j=1}^{n}(P_{\nu_n(j)}\psi\otimes\phi_{\nu_n(j)})\otimes\Theta_0^n, \tag{7}$$

where σ^n is the set of all eigenvalue sequences of length n. Θ_n' is a sum over orthogonal component states, each corresponding to a possible outcome sequence ν_n in σ^n.

The state Θ^{∞} corresponding to an infinity of measurement repetitions is then given by extension as

$$\Theta^{\infty}=\bigotimes_{j=1}^{\infty}\left(\sum_{a_j\in\sigma} P_{a_j}\psi\otimes\phi_{a_j}\right). \tag{8}$$

Note that, as for Θ_n,

$$(\Theta^{\infty},\Theta^{\infty})=\prod_{j=1}^{\infty}\left(\sum_{a_j\in\sigma}(\psi,P_{a_j}\psi)\right)=1.$$

Before discussing some properties of Θ^{∞} it should be noted that the overall Hilbert space for an infinite number of copies of object + record systems is $\otimes_{j\in N}\mathcal{H}_j$, the complete tensor product[4] of countably many copies of $\mathcal{H}$. $\otimes_{j\in N}\mathcal{H}_j$ is spanned by the set of all product vectors $\otimes_j\lambda_j$ with $\lambda_j\in\mathcal{H}$ and is nonseparable even if H is two-dimensional. However the essentials of the discussion can be given, for the most part, in the subspaces $\mathcal{H}^0$ and $\mathcal{H}^{\infty}$ of $\otimes_{j\in N}\mathcal{H}_j$. $\mathcal{H}^0$ is spanned by all product vectors $\otimes_j\lambda_j$ with $\lambda_j=\psi\otimes\phi_0$ for all but a finite number of j values. $\mathcal{H}^{\infty}$ is spanned by all product vectors $\otimes_j\lambda_j$ with $\lambda_j=\sum_{a\in\sigma}(P_a\psi\otimes\phi_a)$ for all but a finite number of j values. $\mathcal{H}^{\infty}$ and $\mathcal{H}^0$ are separable if H is, and are orthogonal to one another (Eq. 2). It is clear from the definitions of $\mathcal{H}^0$ and $\mathcal{H}^{\infty}$ that $\Theta_n\in\mathcal{H}^0$ for each n and $\Theta^{\infty}\notin\mathcal{H}^0$ (Θ^{∞} is a product state in $\mathcal{H}^{\infty}$).

For each infinite outcome sequence ν, let Θ_ν be a product state of norm 1 given by

$$\Theta_\nu=\bigotimes_{j=1}^{\infty}(\psi'_{\nu(j)}\otimes\phi_{\nu(j)}), \tag{9}$$

where $\psi'_{\nu(j)}=P_{\nu(j)}\psi/\|P_{\nu(j)}\psi\|$. Let $\mathcal{H}^{N'}$ be the subspace of $\otimes_{j\in N}\mathcal{H}_j$ spanned by all vectors Θ_ν with $\nu\in\sigma^N$, the set of all infinite outcome sequences. Clearly $\mathcal{H}^{N'}$ is nonseparable and is the space of states interpretable as linear combinations of infinite-outcome-sequence states each of norm 1.

For any $\nu\in\sigma^N$ one has $(\Theta_\nu,\Theta^{\infty})=\Pi_j(\psi'_{\nu(j)}\otimes\phi_{\nu(j)}$, $\sum_{a\in\sigma}P_a\psi\otimes\psi_a)=\Pi_j(\psi,P_{\nu(j)}\psi)^{1/2}=0$ for most ψ. Thus in general Θ^{∞} lies in a subspace orthogonal to $\mathcal{H}^{N'}$. It follows from this that, in general, it is *not* possible to represent the asymptotic situation by a sum over component states each corresponding to an infinite outcome

sequence ν. Except for the trivial case with $(\psi, P_a\psi)=1$ for some a, the state $\sum_{\nu\in\sigma^N}\otimes_j(P_{\nu(j)}\psi\otimes\phi_{\nu(j)})$ is the zero vector. The asymptotic state differs in this respect from the state Θ_n, Eq. (5) which is the same as Θ_n', Eq. (7). Θ^∞ can be characterized as having, for each finite outcome sequence θ of length $l(\theta)$ a nonzero component state X_θ^∞ provided that $(\psi, P_a\psi)>0$ for each a in σ. However, it is not an orthogonal sum over all such components. To see this define X_θ^∞ by

$$X_\theta^\infty=\bigotimes_{j=1}^{l(\theta)}(P_{\theta(j)}\psi\otimes\phi_{\theta(j)})\otimes\Theta_{l(\theta)}^\infty \tag{10}$$

with $\Theta_{l(\theta)}^\infty$ given by

$$\Theta_{l(\theta)}^\infty=\bigotimes_{j=l(\theta)+1}^{\infty}\left(\sum_{a_j\in\sigma}P_{a_j}\psi\otimes\phi_{a_j}\right). \tag{11}$$

X_θ^∞ is the asymptotic state in $\mathcal{H}^\infty$ corresponding to all finite outcome sequences containing θ as an initial segment.

One has for each finite θ, $(X_\theta^\infty, \Theta^\infty)=(X_\theta^\infty, X_\theta^\infty)=\prod_{j=1}^{l(\theta)}(\psi, P_{\theta(j)}\psi)\geq 0$. The desired result follows from the fact that for each state ψ there exist sequences θ and θ' with θ an initial segment of θ' such that $(X_\theta^\infty, X_{\theta'}^\infty)=(X_{\theta'}^\infty, X_{\theta'}^\infty)>0$.

Define θ and θ' to be *orthogonal* if θ is not an initial segment of θ' and θ' is not an initial segment of θ. If each pair of finite sequences in a set S of finite sequences are orthogonal, then S is a set of *pairwise orthogonal* sequences. If θ and θ' are orthogonal, then $(X_\theta^\infty, X_{\theta'}^\infty)=0$. If S is a set of pairwise orthogonal sequences such that every finite sequence has an initial segment in S, then $\Theta^\infty=\sum_{\theta\in S}X_\theta^\infty$. Thus there are many decompositions of Θ^∞ into orthogonal sums over some X_θ^∞.

III. STATES CORRESPONDING TO THE SET OF RANDOM OUTCOME SEQUENCES

A. Sequential tests

From now on, to avoid some complications, the observable A is restricted to be a question observable. Then $\sigma=\{0, 1\}$ as 0 and 1 are the only outcomes possible. Also the object state ψ is restricted to be such that $(\psi, P_1\psi)$ [and $(\psi, P_0\psi)$] is an effectively computable[2] real number. That is, there must exist an effective procedure[6] for computing successive approximations to the value of $(\psi, P_1\psi)$. For example, any real number r between 0 and 1 has a binary expansion $\sum_{i=1}^\infty a_i 2^{-i}$ with $a_i=0$ or 1. r is computable iff there is an effective procedure[6] for enumerating the a_i.

Let $\sigma^{<N}$ and σ^N denote the respective sets of all finite and infinite 0, 1 sequences with N the set of positive integers and $\mathcal{P}(\sigma^{<N})$ and $\mathcal{B}(\sigma^N)$ the respective sets of all subsets of $\sigma^{<N}$ and all Borel subsets of σ^N. Define $W:\mathcal{P}(\sigma^{<N})\to\mathcal{B}(\sigma^N)$ by

$$W(E)=\bigcup_{\theta\in E}B_\theta \tag{12}$$

for each subset E of $\sigma^{<N}$. Here B_θ is the set of all infinite 0, 1 sequences having θ as an initial segment. (The variables θ, θ' denote finite 0, 1 sequences and ν, ν', ... denote infinite 0, 1 sequences.) For each E, $W(E)$ is a Borel set and is open in the usual product topology on σ^N.

Let P_Θ be the probability measure on $\mathcal{B}(\sigma^N)$ which gives the statistics of the measurement repetitions as given by the states Θ_n. That is for each θ of length $l(\theta)$, $P_\Theta B_\theta$ is the probability that in any state Θ_n with $n\geq l(\theta)$ the first $l(\theta)$ record systems have $\theta(1), \ldots, \theta(l(\theta))$ as outcomes. For the case at hand

$$P_\Theta B_\theta=\prod_{j=1}^{l(\theta)}(\psi, P_{\theta(j)}\psi), \tag{13a}$$

and

$$P_\Theta W(E)=\sum_{\theta\in E}P_\Theta B_\theta. \tag{13b}$$

Following Martin-Löf,[2] a P_Θ sequential test T is defined here as a subset of $N\times\sigma^{<N}$ such that the m slices of T defined by $T_m=\{\theta\,|\,(m,\theta)\in T\}$ satisfy $T_0=\sigma^{<N}$ and

(1) $T_1\supseteq T_2\supseteq\cdots\supseteq T_m\supseteq\cdots$,

(2) $\theta\in T_m\to\theta*\theta'\in T_m$ for all $\theta'\in\sigma^{<N}$

(3) $P_\Theta W(T_m)<2^{-m}$, $m=0, 1, 2, \cdots$,

(4) T is recursively enumerable.

Here $\theta*\theta'$ is the finite sequence obtained by adjoining θ' to the end of θ. The fourth condition means that there exists an effective procedure[6] for listing or enumerating the elements of T. By an obvious selection process, for each m such a procedure gives also an effective listing of T_m. $W(T_m)$ is called a critical region for the test and 2^{-m} the corresponding significance level. In what follows the significance level is often referred to as "level m." For each $\nu\in\sigma^N$ the *critical value* $m_T(\nu)$ is defined by $m_T(\nu)=\max\{m\,|\,\nu\in W(T_m)\}$.

As a statistical test for some property Q of infinite sequences such that $P_\theta\{\nu\,|\,\nu \text{ has } Q\}=1$, T gives for each m, a prescription as a set T_m, of finite 0, 1 outcome sequences, for rejecting or accepting Q. That is, the infinite outcome sequence ν is rejected as not having Q if $\nu\in W(T_m)$. The error probability of this estimation, i.e., one rejects ν as not having Q when it really has Q, is $\leq 2^{-m}$. However any $\nu\notin W(T_m)$ certainly has Q because $\nu\notin W(T_m)$ implies $\nu\notin W(T_n)$ for all $n\geq m$.

One can also speak of finite sequences[7] having Q at level m. Here θ is rejected by T as not having Q at level m if $\theta\in T_m$. If $\theta\notin T_m$, then we shall say that θ has Q or passes T at level m. Note that if θ is rejected by T at level m, then θ is rejected at all lower levels. Also $\cap_m T_m$ is empty as no θ is rejected by T at all levels. For no finite sequence can one be certain that it is rejected by T.

It is to be noted that deciding whether or not $\nu\in W(T_m)$ requires that ν be available only asymptotically. For if $\nu\in W(T_m)$ then this will be found out in a finite number of steps in which the successive finite sequences $\theta_1, \theta_2, \ldots, \theta_n\cdots$ in the enumeration of T_m are compared with the outcome sequence obtained from the first $l(\theta_1)$, $l(\theta_2), \ldots, l(\theta_n), \cdots$ measurement repetitions. If for some n, $\nu_{l(\theta_n)}=\theta_n$, then $\nu\in W(T_m)$. If $\nu\notin W(T_m)$, then this will be found out asymptotically only after exhausting

the enumeration and discovering that for no n does $\nu_{l(\theta_n)} = \theta_n$.

All of the usual probability theoretical tests for randomness can be given as simple sequential tests.[2,7] Consider for example the property $\bar{M}\nu = (\psi, P_1\psi)$, i.e., the limit relative frequency of occurrence of 1 in ν equals $(\psi, P_1\psi)$. A P_Θ sequential test T for this property is given by $T_m = \cup_{h=1}^{\infty} T^h_{m+h}$ [Eq. (15) and discussion] where for each h, T^h is a P_Θ sequential test for the property "there exists an n such that for all $l > n$, $|\bar{M}_l\nu - (\psi, P_1\psi)| \leq 2^{-h}$ ($\bar{M}_l\nu$ is the mean of the first l elements of ν). The effective enumerability of the T^h and T can be proved by actual construction.

As another example let ν^0 be a given effectively enumerable sequence in σ^N and Q the property $\nu \neq \nu^0$. The corresponding test T is obtained by defining T_m to be the set of all finite σ sequences which have ν^0_m as an initial segment where ν^0_m is the shortest initial segment of ν^0 such that $P_\Theta W(T_m) < 2^{-m}$.

For each P_Θ sequential test T one can associate to each T_m the set $\bar{T}_m$ defined by $\theta \in \bar{T}_m \leftrightarrow \theta \in T_m$ and for all $\theta' \in T_m$ θ' is not a proper initial segment of θ. That is, $\bar{T}_m$ is the set of all elements of T_m which are the shortest initial segments of elements of T_m. Note that $\bar{T}_m$ is a set of pairwise orthogonal sequences.

Now let C be the (countably infinite) set of all P_Θ sequential tests. A universal test[2] for the tests in C is a P_Θ sequential test V such that for each test T in C there is a constant d, depending on T but not on m, such that for each m

$$T_{m+d} \subseteq V_m. \tag{14}$$

The existence of universal tests is shown as follows: Let $T^{(1)}, \cdots, T^{(k)}, \cdots$ be an effective enumeration of the tests in C, which does exist.[2] Define V_m by[2]

$$V_m = \bigcup_{k=1}^{\infty} T^{(k)}_{m+k} \tag{15}$$

for $m \geq 1$ and $V_0 = \sigma^N$. It can be shown[2] that $V = \{(m, \theta) \mid \theta \in V_m, m = 0, 1, 2, \cdots\}$ is a P_Θ sequential test which is universal for C.

The universal property can be seen as follows: If a sequence θ passes V at level m, then for each test T in C, θ passes T at some higher level $m + k$. Note that $\nu \notin \cap_m W(V_m)$ implies that $\nu \notin {}_m W(T_m)$ for each T in C and conversely.

A sequence ν is defined to be C random if and only if $m_V(\nu) < \infty$ or equivalently $\nu \notin \cap_m W(V_m)$. That is, a sequence ν is C random if ν passes V at some level m. If ν passes V at level m, then ν passes each test T in C at some level higher than m. Conversely if ν passes each test T in C at some level, then ν passes V at some level (because V is in C).

A sequence ν is non-C-random if and only if $\nu \in \cap_m W(V_m)$, or ν is rejected by V at all levels. Since $P_\Theta(\cap_m W(V_m)) = 0$ one sees that the set of C-random sequences is a set of P_Θ measure 1. The dependence of the definition of randomness on the measure P_Θ is indicated by the term "C random" and the obvious dependence of the tests in C on P_Θ. (If desired one can remove this dependence by calling a sequence random if it is C random for some computable product measure P.)

B. States corresponding to sequential tests

It will now be seen that one can associate families of states in $\mathcal{H}^0$ and $\mathcal{H}^\infty$ to a sequential test.

Let T be a P_Θ sequential test and let $\bar{T}_m$ be defined from T as above. Define the state $X_n^{T_m}$ in $\mathcal{H}^0$ by

$$X_n^{T_m} = \sum_{\theta \in \sigma^{\leq n} \cap \bar{T}_m} X_n^\theta, \tag{16}$$

where X_n^θ is given by

$$X_n^\theta = \sum_{\theta' \in \tau_\theta^n} \overset{n}{\underset{j=1}{\otimes}} (P_{\theta'(j)}\psi \otimes \phi_{\theta'(j)}) \otimes \Theta_0^n \tag{17}$$

and τ_θ^n is the set of all σ sequences of length n which have θ as an initial segment. Θ_0^n is given by Eq. (6). X_n^θ is the state corresponding to all outcome sequences of length n which have θ as an initial segment [note that $n \geq l(\theta)$]. $X_n^{T_m}$ is the orthogonal sum over all states, corresponding to the shortest initial sequences of length $\leq n$ which are rejected by the test T at level m. From the definition of $\bar{T}_m$, and condition (2) in the definition of a sequential test it is clear that $X_n^{T_m}$ is also given by

$$X_n^{T_m} = \sum_{\theta \in \sigma^n \cap T_m} \overset{n}{\underset{j=1}{\otimes}} (P_{\theta(j)}\psi \otimes \phi_{\theta(j)}) \otimes \Theta_0^n. \tag{18}$$

Thus $X_n^{T_m}$ is also the state corresponding to all outcome sequences of length n which are rejected by the test T at level m. For each such sequence θ, $X_n^{T_m}$ includes a corresponding component state X_n^θ.

The corresponding asymptotic state $X^{\infty T_m}$ in $\mathcal{H}^\infty$ is defined by

$$X^{\infty T_m} = \sum_{\theta \in \bar{T}_m} X_\theta^\infty, \tag{19}$$

where X_θ^∞ is given by Eq. (10). By the discussion above and in Sec. II, $X^{\infty T_m}$ is an asymptotic state which includes component states, corresponding to exactly those finite outcome sequences which are rejected by the test T at level m.

Note that (1) $X^{\infty T_m}$, like Θ^∞, and X_θ^∞, cannot be written as a sum over infinite outcome sequences and (2) one cannot replace $\bar{T}_m$ by T_m in the sum. The reason for the latter is that unlike the case for T_m the terms in the $\bar{T}_m$ sum are pairwise orthogonal. Also from the definitions of a sequential test and $\bar{T}_m$ one has $\|X_n^{T_m}\| \leq \|X_{n+1}^{T_m}\| \leq \|X^{\infty T_m}\| \leq (2^{-m})^{1/2}$.

The states corresponding to those sequences which are accepted by T are obtained as follows: Define $\Theta_n^{T_m}$ and $\Theta^{\infty T_m}$ by

$$\Theta_n^{T_m} = \Theta_n - X_n^{T_m} \tag{20}$$

and

$$\Theta^{\infty T_m} = \Theta^\infty - X^{\infty T_m} \tag{21}$$

with Θ_n and Θ^∞ given by Eqs. (5) and (8). Consider first

the $\Theta_n^{T_m}$. One has from Eqs. (5) and (18),

$$\Theta_n^{T_m} = \sum_{\theta \in \sigma^n - T_m} \bigotimes_{j=1}^{n} (P_{\theta(j)}\psi \otimes \phi_{\theta(j)}) \otimes \Theta_0^n \tag{22}$$

which is a sum over all outcome sequences of length n not in T_m. That is, a sequence of length n has a component in $\Theta_n^{T_m}$ if and only if it passes test T at level m. $\Theta^{\infty T_m}$ includes component states corresponding to all and only those finite 0—1 sequences which pass test T at level m.

As was the case for Θ^∞, $X^{\infty T_m}$ and $\Theta^{\infty T_m}$ in $\mathcal{H}^\infty$ do not correspond to a sum over components each corresponding to an infinite outcome sequence. However, one can partition σ^N into two sets, each associated with $X^{\infty T_m}$ and $\Theta^{\infty T_m}$ as follows: One says that a set of infinite sequences is *associated* with a state in $\mathcal{H}^\infty$ if and only if for each sequence ν in the set and for each n, the initial segment of ν of length n, ν_n, has a component in the state. Then it is clear that $W(T_m)$ is associated with $X^{\infty T_m}$ and $\sigma^N - W(T_m)$ is associated with $\Theta^{\infty T_m}$.

To see this note that by the definitions of T_m and $\bar{T}_m$, $W(T_m) = W(\bar{T}_m) = \cup_{\theta \in \bar{T}_m} B_\theta$. From Eq. (10) one sees that each infinite sequence in σ^N with θ as an initial segment is associated with X_θ^∞. Thus by Eq. (19), $W(T_m)$ is exactly the set of infinite sequences associated with $X^{\infty T_m}$. Since σ^N is associated with Θ^∞ one has from Eq. (21) that $\sigma^N - W(T_m)$ is associated with $\Theta^{\infty T_m}$.

It should also be noted that for each infinite sequence ν in $\sigma^N - W(T_m)$ [i.e., not in $W(T_m)$] ν_n has a component in $\Theta_n^{T_m}$ *for each* n. To see this one notes that $\nu \notin W(T_m)$ implies that for all n, $\nu_n \notin T_m$. [For if $\nu_n \in T_m$ for some n, then by Eq. (12) $\nu \in W(T_m)$, which is a contradiction.] From Eq. (22) one clearly has the result that for each n, ν_n has a component state $X_n^{\nu_n}$ [Eq. (17)] in $\Theta_n^{T_m}$.

This has the following consequence. Suppose one considers the evolution of a sequence of states Θ_n, $n = 0, 1, \cdots$ as the time development of an infinite number of copies of an object + record system, each undergoing one measurement interaction. Then if the sequence of states $\Theta_n^{T_m}$ with $n = 0, 1, \cdots$ is to have meaning as a description of measurement sequences which pass T at level m there must clearly exist in the sequences $\Theta_n^{T_m}$, many possible trajectories or infinite outcome sequence tracks. The above proof shows that this is so—each ν in $\sigma^N - W(T_m)$ describes such a trajectory.

Finally one has from the construction of the states $X_n^{T_m}$ and $X^{\infty T_m}$, Eqs. (16) and (19), $[\|X_n^{T_m}\| \leq \|X^{\infty T_m}\| \leq 2^{-m/2}]$,

$$0 < \lim_{n \to \infty} \|\Theta_n - \Theta_n^{T_m}\| \leq 2^{-m/2} \tag{23}$$

for each m and

$$\lim_{m \to \infty} \|\Theta^\infty - \Theta^{\infty T_m}\| = 0. \tag{24}$$

These limits reflect the fact that for each m there exist finite outcome sequences which fail T at level m and that no finite outcome sequence fails T at all levels ($\cap_m T_m$ is empty). So the state which corresponds to exactly those sequences in $\cap_m T_m$ must be the 0 vector. However $\cap_m W(T_m)$ is far from empty. Thus there is no (nontrivial) state associated with $\cap_m W(T_m)$, the set of all infinite 0—1 random sequences which fail T with certainty even though for each m, $X^{\infty T_m}$ is associated with $W(T_m)$. Similarly there is no state associated with the set, $\sigma^N - \cap_m W(T_m)$ of all infinite sequences which pass T even though for each m, $\Theta^{\infty T_m}$ is associated with $\sigma^N - W(T_m)$.

Now let V be a universal test for the set C of all P_Θ sequential tests and define $X_n^{V_m}$, $X^{\infty V_m}$, $\Theta_n^{V_m}$, and $\Theta^{\infty V_m}$ by Eqs. (16), (19), (20), and (21) with V replacing T. Then $X_n^{V_m}$ and $\Theta_n^{V_m}$ include components corresponding to exactly those finite 0—1 sequences of length n which are non-C-random and C random at level m respectively. $X^{\infty V_m}$ and $\Theta^{\infty V_m}$ include components corresponding exactly to all finite 0—1 sequences which are non-C-random and C random at level m. Also $X^{\infty V_m}$ and $\Theta^{\infty V}$ are associated with the respective sets of infinite 0—1 sequences which are non-C-random and C random at level m. [By the previous discussion $W(V_m)$ is associated with $X^{\infty V_m}$ and $\sigma^N - W(V_m)$ is associated with $\Theta^{\infty V_m}$.] Similarly there are no nontrivial states associated with the sets $\cap_m W(V_m)$ and $\sigma^N - \cap_m W(V_m)$ of non-C-random and C-random sequences respectively.

IV. DISCUSSION

A. Other aspects of the test-state correspondence

Other aspects of sequential tests and their associated states should be noted. Since the universal test V depends on the effective enumeration of the tests in C, the states $X_n^{V_m}$, Θ^{V_m}, $X^{\infty V_m}$, and $\Theta^{\infty V_m}$ also depend on the enumeration. However it can be shown that the set of infinite C-random sequences is independent of the effective enumeration of tests. Thus if V and U are two universal tests for C which differ only in the numeration of the tests in C, $V_m \neq U_m$ and $X^{\infty V_m} \neq X^{\infty U_m}$ and $\Theta^{\infty V_m} \neq \Theta^{\infty U_m}$. However $\cap_m W(V_m) = \cap_m W(U_m)$ and so $\sigma^N - \cap_m W(V_m) = \sigma^N - \cap_m W(U_m)$.

One can extend the construction to some of the other definitions of randomness which have been proposed in the literature.[8-10] For example, C can be the set of all sequential tests T with condition (4) in the definition relaxed to require that T be hyperarithmetic,[8] T have a code in the minimal standard transitive model of Zermelo Frankel set theory,[9] or T be definable in the language of Zermelo Frankel set theory.[10] In each of these cases one can define the set of random sequences to be those which pass each test T in C at some significance level.[11] For each definition of C one can define a universal test[12] for all the tests in C and, by the construction of the previous section, give the corresponding states $X_n^{V_m}$, $X^{\infty V_m}$, $\Theta_n^{V_m}$, and $\Theta^{\infty V_m}$ for each m.

It should be noted that it is not known at present which of the proposed definitions of randomness is correct in the sense of being the weakest possible without leading to contradictions. Moreover a proof of which definition is correct would have important consequences[13] for the relationship between the foundations of physics and mathematics. Until such a proof is forthcoming, one is free to use any of the definitions he wishes.

B. The Everett interpretation

The Everett interpretation[3] of quantum mechanics

considers the whole physical universe, including all observers, to be described by a single universal state vector whose time evolution is deterministic. The universal state can be represented as a coherent sum over a large number of component states, each corresponding to a possible state of the universe as seen by an observer who is also included in the universal state. As time goes on, the universal state branches more and more in such a representation. The universe as the observer perceives it and its time evolution corresponds to one of the observer memory sequences in the universal state. All of the possible worlds, each corresponding to a branch in the universal state, have equal reality status.

In the repeated measurement model of this interpretation O_n with $n = 0, 1, \ldots$, Eqs. (14) and (5), represents the evolution of a measurement apparatus + automation undergoing successive measurement interactions with an ensemble of identically prepared systems. Each possible path in the sequence $O_0, \ldots, O_n, \cdots$ corresponds to a unique outcome sequence ν in $\{0, 1\}^N$ and represents a possible world in which $\nu(j)$ for $j = 1, 2, \cdots$ are the outcomes observed by the automation.

An essential part of the Everett interpretation is the metatheorem which states that quantum mechanics generates its own statistical interpretation. In a proof of the metatheorem, given by DeWitt,[1] a sequence of states Θ_n^ϵ, $n = 1, 2, \cdots$ is defined, which for each n, exclude from O_n Eq. (5), all components corresponding to sequences θ in $\{0, 1\}^N$ for which $|\bar{M}_n\theta - (\psi, P_1\psi)| > \epsilon$. The Θ_n^ϵ are then shown to satisfy $\|\Theta_n^\epsilon - O_n\| \to 0$ as $n \to \infty$. It is noted in passing that the same result is obtained for states defined to exclude components for memory sequences which fail to meet any finite combination of tests for randomness. The metatheorem is then stated to follow from this and the fact that nonrandom memory sequence component states are of a set of measure 0 in the limit $n \to \infty$.

The above proof is incomplete in the sense that the states relating to the other tests of randomness are not given an explicit definition. Furthermore, their definition depends on what properties, other than those given for the Θ_n^ϵ, they are required to have.

The point to be made here is that the $\Theta_n^{V_m}$ can be used to remove this incompleteness. By Eq. (22) the $\Theta_n^{V_m}$ include components for exactly those memory sequences of length n which pass the universal test V at level m. By an argument given in Sec. III, the infinite memory sequences ν in $\sigma^N - W(V_m)$ are those which, for each n, have a component state $X_n^{\nu_n}$, Eq. (17) in $\Theta_n^{V_m}$. So in the limit $n \to \infty$ the memory sequence ν_n passes at level m, all the tests for randomness. Since, Eq. (24),

$$\lim_{m\to\infty}\lim_{n\to\infty}\|\Theta_n^{V_m} - O_n\| = 0,$$

the set of memory sequences which pass V at some level m is a set of measure 1. (The set which fails V at all levels is a set of measure 0.) Thus the $\Theta_n^{V_m}$ have the desired properties for DeWitt's proof.

There is another difference between the Θ_n^ϵ as defined by DeWitt[1] and the $\Theta_n^{V_m}$. To see this consider a perceived world whose complete evolution path is given by the infinite sequence ν. (Recall that each ν corresponds to a chain of finite memory sequences ν_n, totally well ordered by the relation "is an initial segment of.") It is clear that in general this path will wander in and out of the state sequence Θ_n^ϵ as n increases. It will be inside for those n for which $|\bar{M}_n\nu_n - (\psi, P_1\psi)| < \epsilon$. It will be outside for those n for which $|\bar{M}\nu_n - (\psi, P_1\psi)| > \epsilon$.

Furthermore there are in general no paths at all which are inside the sequence Θ_n^ϵ for all n. In fact, for many n and ϵ, Θ_n^ϵ is the zero vector.[14] Thus the most one has, is that if a perceived world is such that its path ν satisfies $|\bar{M}\nu - (\psi, P_1\psi)| < \epsilon$, then there is some value of n, dependent on ν, ϵ, and $(\psi, P_1\psi)$, such that for all $l > n$, the world path lies in the state sequence Θ_l^ϵ for all $l = n+1, n+2, \cdots$.

The $O_n^{V_m}$ differ in this respect from the Θ_n^ϵ in that there are many perceived worlds whose paths have *for each* n, component states in $\Theta_n^{V_m}$. This follows from the argument given at the end of Sec. III. All these worlds satisfy the appropriate statistical laws in the sense of DeWitt since the appropriate tests are included in V. Only the exceptional or "Maverick" worlds wander and they make only one change. Initially they are in the $\Theta_n^{V_m}$. For some value of n they leave the $\Theta_n^{V_m}$ and move into $X_n^{V_m}$. Once in the $X_n^{V_m}$ they stay there and do not wander out.

[1] B. DeWitt, in *The Many Worlds Interpretation of Quantum Mechanics*, edited by B. DeWitt and N. Graham (Princeton U.P., Princeton, New Jersey, 1973), pp. 167—218; in *Battelle Recontres*, edited by C. DeWitt and J. Wheeler (Benjamin, New York, 1968), pp. 318—32.

[2] P. Martin-Löf, Inf. Control **9**, 603 (1966); "Algorithms and Random Sequences," Lectures at Erlangen University (1966), translated by W.A. Beyer, Los Alamos, New Mexico, 1976, LA-TR-67-105.

[3] H. Everett III, Rev. Mod. Phys. **29**, 454 (1957); J.A. Wheeler, Rev. Mod. Phys. **29**, 463 (1957).

[4] J. Von Neumann, "On Infinite Direct Products" in *Collected Works*, Vol. III, edited by A.H. Taub (Pergamon, New York, 1963), pp. 323—99; M.A. Guichardet, Ann. Ec. Norm. Super. **83**, 1 (1966).

[5] If O_n is considered as a state evolved from Θ_0 by the action of a unitary time evolution operator then the time evolution of $\mathcal{O}_0^n$ as describing not-yet-interacting systems should be included. However this is quite inessential for our purposes here and is thus suppressed.

[6] Hartley Rogers, Jr., *Theory of Recursive Functions and Effective Computability* (McGraw-Hill, New York, 1967), Chaps. 1—5.

[7] Terrence Fine, *Theories of Probability* (Academic, New York, 1973), Chap. V.

[8] P. Martin-Löf, *On the Notion of Randomness*, Proceedings of Summer Institute on Proof Theory and Intuitionism, State University of Buffalo, New York, 1968, edited by J. Myhill, A. Kino, and R. Vesley (North-Holland, Amsterdam, 1970).

[9] R. Solovay, Ann. Math. **92**, 1 (1970).

[10]P. A. Benioff, Phys. Rev. D 7, 3603 (1973); A. H. Kruse, Z. Math. Logik Grundl. Math. 13, 299 (1967).

[11]The randomness definitions of Refs. 8—10 are given in terms of sets of measure zero rather than in terms of sequential tests. However one can show that the two types of definitions are equivalent for the definitions of Refs. 8 and 9. For that of Ref. 10 the definition given in terms of sets of measure zero is at least as strong as that given in terms of sequential tests.

[12]For these stronger definitions a universal test V is not in general a test in C. Thus a definition of randomness in terms of a universal test V for C is at least as strong a definition given in terms of the tests in C.

[13]P. A. Benioff, J. Math. Phys. **17**, 618, 629 (1976).

[14]To see this let $\epsilon = 2^{-l}$ and the object state ψ be such that $(\psi, P_1\psi) = \frac{1}{2}$. (This choice is made for convenience only.) Let $s_n = \sum_{j=1}^{n} \theta(j)$ for each $0-1$ sequence θ length n. Then for each $n < 2^{l-n}$,

$$\left| \bar{M}_n\theta - \tfrac{1}{2} \right| = \left| \frac{s_n}{n} - \frac{1}{2} \right| = \left| \frac{2s_n - n}{2n} \right| > \frac{1}{2n} \geq 2^{-l} = \epsilon .$$

So for odd $n < 2^{l-1}$ no sequence θ of length n satisfies $|\bar{M}_n\theta - (\psi, P_1\psi)| < \epsilon$ and Θ_n^ϵ is the zero vector for these n.

Vector spherical harmonics of the unit hyperboloid in Minkowski space

Levere Hostler

Physics Department, Wilkes College, Wilkes-Barre, Pennsylvania 18703
(Received 23 February 1977)

Vector fields in Minkowski space which are simultaneous eigenfunctions of the operators $[J_x]^2+[J_y]^2+[J_z]^2$, $[J_z]$, and $-(1/2)L^{\mu\nu}L_{\mu\nu}$ are investigated using special tensor methods which exploit the properties of the intrinsic gradient operator ∇ of the unit hyperboloid $x^\mu x_\mu = 1$. A convenient representation of the simultaneous eigenfunctions is provided by the use of Helmholtz's theorem for the unit hyperboloid. The utility of this representation arises from the existence of intertwining relations such as $J^{\mu\nu}\nabla = \nabla L^{\mu\nu}$. An addition theorem for the solenoidal vector spherical harmonics of the unit hyperboloid is derived, and the Green's function of Poisson's equation on the unit hyperboloid is obtained.

I. INTRODUCTION

In the following we investigate vector fields in Minkowski space which are simultaneous eigenfunctions of the operators $[J_x]^2+[J_y]^2+[J_z]^2$, $[J_z]$ and $-\frac{1}{2}L^{\mu\nu}L_{\mu\nu}$. Here $[J_x]=L_x\mathbf{I}+[S_x]$, etc., denote the components of the ordinary total angular momentum operator associated with the vector field, and $L_{\mu\nu}=(1/i)(x_\mu\partial_\nu - x_\nu\partial_\mu)$ is the four dimensional orbital angular momentum tensor. Such simultaneous eigenfunctions are here called *vector spherical harmonics of the unit hyperboloid*. The vector spherical harmonics of the unit hyperboloid are particularly suited for applications to the electromagnetic field. One such application, for which the present paper provides necessary mathematical background, will be reported in a companion paper. Since vector spherical harmonics of the unit hyperboloid have a certain intrinsic appeal and may well find further quite different applications it was decided to report the results contained herein separately.

The 4-vector spherical harmonics considered here differ from those discussed recently by Daumens and Minnaert[1] in that the latter authors consider simultaneous eigenfunctions of $[J_x]^2+[J_y]^2+[J_z]^2$, $[J_z]$, and the three dimensional orbital angular momentum operator squared, $L_x^2+L_y^2+L_z^2$, rather than our Lorentz invariant four dimensional angular momentum operator squared.

We work in a pseudospherical coordinate system in Minkowski space, defined through the equations

$$\begin{aligned} x^0 &= s\cosh\rho, \\ x^1 &= s\sinh\rho\sin\theta\cos\varphi, \\ x^2 &= s\sinh\rho\sin\theta\sin\varphi, \\ x^3 &= s\sinh\rho\cos\theta, \end{aligned} \tag{1.1}$$

$$0<s<\infty, \quad 0\le\rho<\infty, \quad 0<\theta\le\pi, \quad 0<\varphi\le 2\pi.$$

Our attention is hereby restricted to vector fields within the forward light cone. For efficient calculations in pseudospherical coordinates a special calculus based on the properties of the intrinsic gradient operator ∇ of the unit hyperboloid $x^\mu x_\mu = 1$ is developed in Sec. II. The derivation in Appendix A of the curl–curl identity (2.17) and the derivation of the intertwining relations (3.49), (3.50) are indicative of the type of calculations the methods of Sec. II are designed to handle with relative ease.

The mathematical techniques developed in Sec. II are applied in Sec. IIIA to scalar spherical harmonics and in Sec. IIIB to the vector spherical harmonics that we are really interested in. Aside from the derivation of the Green's function (3.27) of Poisson's equation of the unit hyperboloid in Minkowski space, the final results concerning scalar spherical harmonics in Sec. IIIA are not especially new; but they are included for completeness, since the scalar spherical harmonics will form the building blocks for the construction in Sec. IIIB of the vector spherical harmonics. The discussion of Sec. IIIB is restricted to vector spherical harmonics which are tangent to the unit hyperboloid and are solenoidal in the sense of the intrinsic geometry of the unit hyperboloid, as only these solenoidal vector spherical harmonics are required for later application to the electromagnetic field. Our discussion of vector spherical harmonics of the unit hyperboloid is completed in Appendix B where the non-solenoidal case is considered. The very simple representations, (3.44) and Appendix B Eqs. (B3) and (B4), obtained for the vector spherical harmonics are obtained by exploiting a form of Helmholtz's theorem for the unit hyperboloid according to which every tangent vector field can be written as the sum of a gradient plus a curl. The utility of writing a tangent vector field as a gradient plus a curl lies in the existence of intertwining relations such as those mentioned above. Also in Appendix B we derive an addition theorem [Eq. (B9)] for the solenoidal vector spherical harmonics, a result for vector spherical harmonics which parallels an earlier scalar addition theorem for Gegebauer functions which was obtained by Durand, Fishbane, and Simmons.[2]

II. TENSOR METHODS FOR PHYSICS ON SPACELIKE HYPERBOLOIDS

The D'Alembertian operator in the pseudospherical coordinate system (1.1) takes the form

$$\Box = \hat{s}\frac{\partial}{\partial s} - \frac{\nabla}{s}, \tag{2.1}$$

in which

$$\nabla \equiv \hat{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\sinh\rho}\left(\hat{\theta}\frac{\partial}{\partial\theta} + \frac{\hat{\varphi}}{\sin\theta}\frac{\partial}{\partial\varphi}\right). \tag{2.2}$$

The unit tangent vectors to the various coordinate curves are defined through the equations

$$\hat{s} \equiv (\partial\mathbf{x}/\partial s)\,|\,\partial\mathbf{x}/\partial s\,|^{-1}, \quad \hat{\theta} \equiv (\partial\mathbf{x}/\partial\theta)\,|\,\partial\mathbf{x}/\partial\theta\,|^{-1}, \quad \text{etc.}$$

Their components are conveniently summarized in the matrix statement

$$\begin{pmatrix} \hat{s} \\ -\hat{\theta} \\ -\hat{\varphi} \\ -\hat{\rho} \end{pmatrix} = \left(\begin{array}{c|ccc} \cosh\rho & -\sinh\rho\sin\theta\cos\varphi & -\sinh\rho\sin\theta\sin\varphi & -\sinh\rho\cos\theta \\ \hline 0 & \cos\theta\cos\varphi & \cos\theta\sin\varphi & -\sin\theta \\ 0 & -\sin\varphi & \cos\varphi & 0 \\ -\sinh\rho & \cosh\rho\sin\theta\cos\varphi & \cosh\rho\sin\theta\sin\varphi & \cosh\rho\cos\theta \end{array}\right) \begin{pmatrix} \hat{i}^0 \\ \hat{i}^1 \\ \hat{i}^2 \\ \hat{i}^3 \end{pmatrix}, \tag{2.3}$$

in which $\hat{i}^0$, $\hat{i}^1$, $\hat{i}^2$, $\hat{i}^3$ denote contravariant basis vectors along the x^0, x^1, x^2, and x^3 axes, respectively. Their normalization is $\hat{i}^\mu \cdot \hat{i}^\nu = g^{\mu\nu}$, in which $g^{\mu\nu}$ are the contravariant components of the metric tensor in Minkowski space, with nonzero components $g^{00} = -g^{11} = -g^{22} = -g^{33} = 1$.

The transformation matrix (2.3) is a Lorentz transformation matrix; consequently the tangent vectors $\hat{s}$, $-\hat{\theta}$, $-\hat{\varphi}$, $-\hat{\rho}$ are a local set of contravariant and orthonormal basis vectors. Due to the not positive definite nature of the dot product in Minkowski space, we have the relations $\hat{s}\cdot\hat{s} = -\hat{\theta}\cdot\hat{\theta} = -\hat{\varphi}\cdot\hat{\varphi} = -\hat{\rho}\cdot\hat{\rho} = 1$.

The dimensionless operator ∇ distinguished in Eq. (2.2) is the intrinsic gradient operator of the unit hyperboloid $x^\mu x_\mu = 1$. When applied to a scalar function it produces a tangent vector field of the unit hyperboloid. For efficient calculations involving ∇ it is convenient to go over from the representation (2.2) to an equivalent representation exploiting the fact that the unit hyperboloid is a Riemannian manifold and is thus subject to the laws of general tensor analysis. Thus, we define the covariant basis vectors

$$\mathbf{e}_a \equiv (\partial\hat{s}/\partial u^a), \tag{2.4}$$

$a = 1, 2, 3$, $u^1 \equiv \theta$, $u^2 \equiv \varphi$, $u^3 \equiv \rho$ in the tangent space of the unit hyperboloid. The $\mathbf{e}_a$ are just the vectors $\hat{\theta}$, $\hat{\varphi}$, and $\hat{\rho}$ again but with modified normalization; and are orthogonal to $\hat{s}$:

$$\hat{s}\cdot\mathbf{e}_a = 0. \tag{2.5}$$

From the defining Eq. (2.4) it is immediate that an arbitrary infinitesimal displacement $d\mathbf{x}$ on the unit hyperboloid has the representation $d\mathbf{x} = du^a(\partial\mathbf{s}/\partial u^a) = du^a\mathbf{e}_a$. We can thus compute the squared line element on the unit hyperboloid as $d\mathbf{x}\cdot d\mathbf{x} = \mathbf{e}_a\cdot\mathbf{e}_b\, du^a du^b$. If we compare with the standard representation $d\mathbf{x}\cdot d\mathbf{x} = g_{ab}du^a du^b$ of the line element of a Riemannian manifold, we find that the unit hyperboloid has the metric tensor

$$g_{ab} = \mathbf{e}_a\cdot\mathbf{e}_b. \tag{2.6}$$

Raising and lowering of indices can be defined in the usual way; for brevity the usual operations of Riemannian geometry will be assumed as needed and without explicit comment.[3] We adopt the convention that Greek indices shall be tensor indices of Minkowski space, while Latin indices are reserved for tensor structures of the unit hyperboloid. The Einstein summation convention is adhered to with regard to both geometries. The completeness relation for our basis vectors reads

$$\mathbf{I} = \hat{s}\hat{s} + \mathbf{e}^a\mathbf{e}_a, \tag{2.7}$$

in which the $\mathbf{e}^a$ are the vectors of the dual basis, defined through the requirement that $\mathbf{e}^a\cdot\mathbf{e}_b = \delta^a_b$. Dyadic outer product in Eq. (2.7) is understood. On the left-hand side of Eq. (2.7) in the unit (metric) tensor of Minkowski space; with mixed components δ^μ_ν, covariant components $g_{\mu\nu}$. The gradient operator of the unit hyperboloid has the representation[4]

$$\nabla = -\mathbf{e}^a\partial_a. \tag{2.8}$$

We can now derive an operator commutation relation

$$[\nabla_\mu; \hat{s}_\nu] = -(g_{\mu\nu} - \hat{s}_\mu\hat{s}_\nu) \tag{2.9}$$

that will be needed subsequently. Thus,

$$\begin{aligned}[\nabla_\mu; \hat{s}_\nu] &= [-e^a_\mu\partial_a; \hat{s}_\nu] \\ &= -e^a_\mu[\partial_a; \hat{s}_\nu] = -e^a_\mu(\partial\hat{s}_\nu/\partial u^a) \\ &= -e^a_\mu e_{a\nu} = -(\mathbf{e}^a\mathbf{e}_a)_{\mu\nu} \\ &= -(\mathbf{I} - \hat{s}\hat{s})_{\mu\nu} = -(g_{\mu\nu} - \hat{s}_\mu\hat{s}_\nu). \quad \surd\end{aligned}$$

For the more complicated commutator

$$[\nabla_\mu; \nabla_\nu] = -(\hat{s}_\mu\nabla_\nu - \hat{s}_\nu\nabla_\mu) \tag{2.10}$$

derivatives of the $\mathbf{e}^a$ are required. These derivatives an and the corresponding formulae for the $\mathbf{e}_a$ are

$$(\partial\mathbf{e}_a/\partial u^b) = \left\{{c \atop ab}\right\}\mathbf{e}_c - g_{ab}\hat{s} \tag{2.11}$$

and

$$(\partial\mathbf{e}^a/\partial u^b) = -\left\{{a \atop bc}\right\}\mathbf{e}^c - \delta^a_b\hat{s}. \tag{2.12}$$

The Christoffel symbols $\left\{{a \atop bc}\right\}$ of the unit hyperboloid can hereby enter our calculations; but it has been the author's experience that these invariably disappear from final results of interest, except in combinations such as

$$\left\{{a \atop ab}\right\} = \frac{1}{(-g)^{1/2}}\frac{\partial(-g)^{1/2}}{\partial u^b}, \tag{2.13}$$

in which g denotes the determinant of the (negative definite) metric tensor g_{ab}. We illustrate with the derivation of Eq. (2.10):

$$\begin{aligned}[\nabla_\mu; \nabla_\nu] &= [-e^a_\mu\partial_a; -e^b_\nu\partial_b] \\ &= e^a_\mu[\partial_a; e^b_\nu\partial_b] + [e^a_\mu; e^b_\nu\partial_b]\partial_a \\ &= e^a_\mu(\partial e^b_\nu/\partial u^a)\partial_b - e^b_\nu(\partial e^a_\mu/\partial u^b)\partial_a \\ &= e^a_\mu\left(-\left\{{b \atop ac}\right\}e^c_\nu - \delta^b_a\hat{s}_\nu\right)\partial_b - e^b_\nu\left(-\left\{{a \atop bc}\right\}e^c_\mu - \delta^a_b\hat{s}_\mu\right)\partial_a \\ &= -\left\{{b \atop ac}\right\}e^a_\mu e^c_\nu\partial_b + \left\{{a \atop bc}\right\}e^b_\nu e^c_\mu\partial_a - \hat{s}_\nu e^a_\mu\partial_a + \hat{s}_\mu e^b_\nu\partial_b \\ &= \hat{s}_\nu\nabla_\mu - \hat{s}_\mu\nabla_\nu = -(\hat{s}_\mu\nabla_\nu - \hat{s}_\nu\nabla_\mu). \quad \surd\end{aligned}$$

The relative ease of this calculation lies in the ability to treat the three Gaussian coordinates θ, φ, and ρ of the unit hyperboloid on equal footing. Further commutators that will be needed are the following:

$$[\nabla^2;\hat{s}] = -2\nabla - 3\hat{s} \tag{2.14}$$

and

$$[\nabla^2;\nabla] = \nabla + 2\hat{s}\,\nabla^2 \tag{2.15}$$

in which $\nabla^2 \equiv \nabla\circ\nabla = (-g)^{-1/2}\partial_a(-g)^{1/2}g^{ab}\partial_b$ is the intrinsic Laplacian operator of the unit hyperboloid. The proofs of Eqs. (2.14) and (2.15) build on the commutators (2.9) and (2.10):

$$\begin{aligned}
[\nabla^2;\hat{s}_\nu] &= [\nabla^\mu\nabla_\mu;\hat{s}_\nu] \\
&= \nabla^\mu[\nabla_\mu;\hat{s}_\nu] + [\nabla^\mu;\hat{s}_\nu]\nabla_\mu \\
&= -\nabla^\mu(g_{\mu\nu} - \hat{s}_\mu\hat{s}_\nu) - (\delta^\mu_\nu - \hat{s}^\mu\hat{s}_\nu)\nabla_\mu \\
&= -2\nabla_\nu + \nabla^\mu\hat{s}_\mu\hat{s}_\nu = -2\nabla_\nu + [\nabla^\mu;\hat{s}_\mu\hat{s}_\nu] \\
&= -2\nabla_\nu + \hat{s}_\mu[\nabla^\mu;\hat{s}_\nu] + [\nabla^\mu;\hat{s}_\mu]\hat{s}_\nu \\
&= -2\nabla_\nu - \hat{s}_\mu(\delta^\mu_\nu - \hat{s}^\mu\hat{s}_\nu) - (\delta^\mu_\mu - \hat{s}^\mu\hat{s}_\mu)\hat{s}_\nu \\
&= -2\nabla_\nu - (\delta^\mu_\mu - \hat{s}^\mu\hat{s}_\mu)\hat{s}_\nu = -2\nabla_\nu - 3\hat{s}_\nu
\end{aligned}$$

and

$$\begin{aligned}
[\nabla^2;\nabla_\nu] &= [\nabla^\mu\nabla_\mu;\nabla_\nu] \\
&= \nabla^\mu[\nabla_\mu;\nabla_\nu] + [\nabla^\mu;\nabla_\nu]\nabla_\mu \\
&= -\nabla^\mu(\hat{s}_\mu\nabla_\nu - \hat{s}_\nu\nabla_\mu) - (\hat{s}^\mu\nabla_\nu - \hat{s}_\nu\nabla^\mu)\nabla_\mu \\
&= -\nabla^\mu\hat{s}_\mu\nabla_\nu + \nabla^\mu\hat{s}_\nu\nabla_\mu - \hat{s}^\mu\nabla_\nu\nabla_\mu + \hat{s}_\nu\nabla^2 \\
&= -[\nabla^\mu;\hat{s}_\mu]\nabla_\nu - \hat{s}_\mu\nabla^\mu\nabla_\nu + [\nabla^\mu;\hat{s}_\nu]\nabla_\mu + \hat{s}_\nu\nabla^2 \\
&\quad - [\hat{s}^\mu;\nabla_\nu]\nabla_\mu - \nabla_\nu\hat{s}^\mu\nabla_\mu + \hat{s}_\nu\nabla^2 \\
&= (\delta^\mu_\mu - \hat{s}^\mu\hat{s}_\mu)\nabla_\nu - 2(\delta^\mu_\nu - \hat{s}^\mu\hat{s}_\nu)\nabla_\mu + 2\hat{s}_\nu\nabla^2 \\
&= 3\nabla_\nu - 2\nabla_\nu + 2\hat{s}_\nu\nabla^2 \\
&= \nabla_\nu + 2\hat{s}_\nu\nabla^2. \surd
\end{aligned}$$

In these calculations the relation $\hat{s}\cdot\nabla = -\hat{s}\cdot\mathbf{e}^a\partial_a = 0$ was used to eliminate certain terms. Also, the differentiation property $[AB;C] = A[B;C] + [A;C]B$ of the commutator was needed.

The unit hyperboloid has an intrinsic cross product in terms of which a curl operator, designated by the symbols $\nabla\times$, can be defined:

$$\nabla\times \equiv \mathbf{e}_a\mathbf{e}_b(-g)^{-1/2}\epsilon^{abc}\partial_c. \tag{2.16}$$

Dyadic notation is assumed. Also, the epsilon density is defined through the property $\epsilon^{123} \equiv 1$. Two such curl operations performed in succession would technically be written using dyadic notation as $\nabla\times\cdot\nabla\times$; however, we usually delete the dot, writing simply $\nabla\times\nabla\times$. Useful operator relations involving the curl are

$$\nabla\times\nabla\times$$

$$= (\mathrm{I} - \hat{s}\hat{s})\nabla^2 - (\mathrm{I} - \hat{s}\hat{s}) - \nabla\nabla - \nabla\hat{s} \tag{2.17}$$

and

$$\nabla^2\nabla\times = \nabla\times\nabla^2. \tag{2.18}$$

The derivation of Eq. (2.17) is rather long, and is indicated in Appendix A. Equation (2.18) can be obtained quite simply once Eq. (2.17) has been established by expanding $\nabla\times\nabla\times\nabla\times$ either as $\nabla\times\circ(\nabla\times\nabla\times)$ or as $(\nabla\times\nabla\times)\circ\nabla\times$:

$$\nabla\times\cdot(\nabla\times\nabla\times)$$

$$= \nabla\times\circ[(\mathrm{I} - \hat{s}\hat{s})\nabla^2 - (\mathrm{I} - \hat{s}\hat{s}) - \nabla\nabla - \nabla\hat{s}]$$

$$= \nabla\times\circ\mathrm{I}\nabla^2 - \nabla\times\cdot\mathrm{I} = \nabla\times\nabla^2 - \nabla\times$$

since $\nabla\times\circ\hat{s} = 0$ and $\nabla\times\cdot\nabla = 0$.

Equating our result here to the expression obtained by expanding in the other order, find

$$\nabla\times\nabla^2 - \nabla\times$$

$$= (\nabla\times\nabla\times)\circ\nabla\times$$

$$= [(\mathrm{I} - \hat{s}\hat{s})\nabla^2 - (\mathrm{I} - \hat{s}\hat{s}) - \nabla\nabla - \nabla\hat{s}]\cdot\nabla\times$$

$$= (\mathrm{I} - \hat{s}\hat{s})\nabla^2\cdot\nabla\times - \mathrm{I}\cdot\nabla\times$$

$$= \nabla^2\mathrm{I}\cdot\nabla\times - \hat{s}\hat{s}\nabla^2\cdot\nabla\times - \nabla\times$$

$$= \nabla^2\nabla\times - \hat{s}[\hat{s};\nabla^2]\cdot\nabla\times - \hat{s}\nabla^2\hat{s}\cdot\nabla\times - \nabla\times$$

$$= \nabla^2\nabla\times + \hat{s}(-2\nabla - 3\hat{s})\circ\nabla\times - \nabla\times$$

$$= \nabla^2\nabla\times - \nabla\times.$$

Terms dropped vanish by virtue of either the above identity, $\hat{s}\cdot\nabla\times = 0$, or the new identity $\nabla\cdot\nabla\times = 0$. So far we have obtained the result $\nabla\times\nabla^2 - \nabla\times = \nabla^2\nabla\times - \nabla\times$. Cancelling the common term $\nabla\times$ on either side of this result now gives Eq. (2.18).

With the help of the curl–curl identity we are able to prove the following important closure property of ∇^2: If $\mathbf{F}$ lies in the tangent space of the hyperboloid $x^\mu x_\mu = s^2$ through the field point, then $\nabla^2\mathbf{F}$ lies in the tangent space if and only if $\nabla\cdot\mathbf{F} = 0$.

Proof: by the curl–curl identity (2.17) and the commutation rule (2.14), we find

$$\nabla\times\nabla\times\mathbf{F}$$

$$= \nabla^2\mathbf{F} - \hat{s}\hat{s}\cdot\nabla^2\mathbf{F} - \mathbf{F} - \nabla\nabla\cdot\mathbf{F}$$

$$= \nabla^2\mathbf{F} - [\hat{s}\hat{s};\nabla^2]\cdot\mathbf{F} - \nabla^2\hat{s}\hat{s}\cdot\mathbf{F} - \mathbf{F} - \nabla\nabla\cdot\mathbf{F}$$

$$= \nabla^2\mathbf{F} - \hat{s}[\hat{s};\nabla^2]\cdot\mathbf{F} - [\hat{s};\nabla^2]\hat{s}\cdot\mathbf{F} - \mathbf{F} - \nabla\nabla\cdot\mathbf{F}$$

$$= \nabla^2\mathbf{F} + \hat{s}(-2\nabla - 3\hat{s})\cdot\mathbf{F} - \mathbf{F} - \nabla\nabla\cdot\mathbf{F}$$

$$= \nabla^2\mathbf{F} - \mathbf{F} - \nabla\nabla\cdot\mathbf{F} - 2\hat{s}\nabla\cdot\mathbf{F}.$$

Terms dropped here vanish because $\hat{s}\cdot\mathbf{F} = 0$. Our result so far is equivalent to the representation $\nabla^2\mathbf{F} = \nabla\times\nabla\times\mathbf{F} + \mathbf{F} + \nabla\nabla\cdot\mathbf{F} + 2\hat{s}\nabla\cdot\mathbf{F}$ of $\nabla^2\mathbf{F}$ in which all terms on the right hand side represent tangential vectors except the $2\hat{s}\nabla\cdot\mathbf{F}$ term. The vector $\nabla^2\mathbf{F}$ is thus wholely tangential according as this term does or does not vanish, i.e., according as $\nabla\cdot\mathbf{F} = 0$ or not. $\surd$ As a by product of our proof we learn from the representation of $\nabla^2\mathbf{F}$ above that the curl–curl identity (2.17) assumes the very simple form

$$\nabla\times\nabla\times\mathbf{F} = \nabla^2\mathbf{F} - \mathbf{F}, \qquad \hat{s}\cdot\mathbf{F} = 0, \quad \nabla\cdot\mathbf{F} = 0, \tag{2.19}$$

for a solenoidal tangent vector field $\mathbf{F}$.

III. VECTOR SPHERICAL HARMONICS OF THE UNIT HYPERBOLOID IN MINKOWSKI SPACE

A. Scalar spherical harmonics of the unit hyperboloid

A particle moving in Minkowski space has an angular momentum tensor

$$L^{\mu\nu} \equiv (1/i)(x^\mu \partial^\nu - x^\nu \partial^\mu). \tag{3.1}$$

The space–space components of $L^{\mu\nu}$ are made up of the usual angular momentum operators: $L_x = L^{32}$, $L_y = L^{13}$, and $L_z = L^{21}$, describing a simple rotation of the reference frame in x, y, z space; while the components $L^{0k} = -L^{k0}$, $k = 1,2,3$, describe Lorentz transformations to a moving frame of reference (boosts). The commutation relations

$$[L^{\mu\nu}; L^{\alpha\beta}] = i(g^{\mu\alpha}L^{\nu\beta} + L^{\mu\alpha}g^{\nu\beta} - g^{\mu\beta}L^{\nu\alpha} - L^{\mu\beta}g^{\nu\alpha}) \tag{3.2}$$

can be established by use of the representation (3.1); and can be used to show that the Lorentz invariant Casimir operator $-\frac{1}{2}L^{\alpha\beta}L_{\alpha\beta}$ commutes with all $L^{\mu\nu}$. It follows that $-\frac{1}{2}L^{\alpha\beta}L_{\alpha\beta}$ commutes with $L_x^2 + L_y^2 + L_z^2$ as well as with L_z. Since the last two operators are well known to commute from standard angular momentum theory; we have the mutually commuting set

$$-\tfrac{1}{2}L^{\alpha\beta}L_{\alpha\beta}, \quad L_x^2 + L_y^2 + L_z^2, \text{ and } L_z. \tag{3.3}$$

Going over to the pseudospherical coordinates (1.1) we find

$$L^{\mu\nu} = (1/i)(\hat{s} \wedge \mathbf{e}^a)^{\mu\nu}\partial_a. \tag{3.4}$$

We employ the wedge product notation "$\wedge$" signifying the completely antisymmetrized tensor formed from the indicated vectors:

$$\mathbf{A} \wedge \mathbf{B} \equiv \mathbf{AB} - \mathbf{BA}$$

and

$$\mathbf{A} \wedge \mathbf{B} \wedge \mathbf{C} \equiv \mathbf{ABC} - \mathbf{ACB} + \mathbf{BCA} - \mathbf{BAC} + \mathbf{CAB} - \mathbf{CBA}.$$

Only coordinates referring to the unit hyperboloid $x^\mu x_\mu = 1$ appear on the right hand side of Eq. (3.4). This is made more explicit by the representations

$$\begin{aligned}
L^{01} &= i[\sin\theta\cos\varphi\partial_\rho + \coth\rho(\cos\theta\cos\varphi\partial_\theta - \sin\varphi\csc\theta\partial_\varphi)],\\
L^{02} &= i[\sin\theta\sin\varphi\partial_\rho + \coth\rho(\cos\theta\sin\varphi\partial_\theta + \cos\varphi\csc\theta\partial_\varphi)],\\
L^{03} &= i[\cos\theta\partial_\rho - \coth\rho\sin\theta\partial_\theta],\\
L^{32} &= i[\sin\varphi\partial_\theta + \cot\theta\cos\varphi\partial_\varphi],\\
L^{13} &= i[-\cos\varphi\partial_\theta + \cot\theta\sin\varphi\partial_\varphi],
\end{aligned} \tag{3.5}$$

and

$$L^{21} = -i\partial_\varphi,$$

obtained by evaluating the wedge products $\hat{s} \wedge \mathbf{e}^a = \hat{s}\mathbf{e}^a - \mathbf{e}^a\hat{s}$ using matrix algebra, and then substituting in Eq. (3.4). For this calculation the components (2.3) and the explicit expressions for the metric tensor[5] are required. The above representations show that the $L^{\mu\nu}$ can be viewed as linear operators acting on the Hilbert space of scalar fields defined over the unit hyperboloid $x^\mu x_\mu = 1$. Indeed, it can be shown that the $L^{\mu\nu}$ are self-adjoint operators on this Hilbert space; if the latter is assumed to be endowed with the unitary dot product

$$(g;f) \equiv \int d\Sigma\, g^* f. \tag{3.6}$$

Here $f(\theta,\varphi,\rho)$ and $g(\theta,\varphi,\rho)$ are possibly complex-valued functions defined on the unit hyperboloid, and $d\Sigma = (-g)^{1/2}d^3u = (-g)^{1/2}d\theta\,d\varphi\,d\rho = \sinh^2\rho\sin\theta d\theta d\varphi\,d\rho$ is the three dimensional area element on the unit hyperboloid. In consequence of the self-adjointness of the $L^{\mu\nu}$, our operators (3.3) above are self-adjoint. Since they mutually commute; they admit a complete orthogonal set of simultaneous eigenfunctions. These simultaneous eigenfunctions of the operators (3.3) are the "scalar spherical harmonics of the unit hyperboloid" that we are concerned with here.

To calculate these generalized spherical harmonics we require the explicit representation of the operator $-\frac{1}{2}L^{\alpha\beta}L_{\alpha\beta}$ in the variables θ, φ, ρ. This can be obtained rather efficiently from the representation (3.4) of the $L^{\mu\nu}$. Thus,

$$\begin{aligned}
&-\tfrac{1}{2}L^{\mu\nu}L_{\mu\nu}\\
&= -\tfrac{1}{2}(1/i)(\hat{s}\wedge\mathbf{e}^a)^{\mu\nu}\partial_a(1/i)(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&= \tfrac{1}{2}(-g)^{-1/2}(\hat{s}\wedge(-g)^{1/2}\mathbf{e}^a)^{\mu\nu}\partial_a(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&= \tfrac{1}{2}(-g)^{-1/2}[(\hat{s}\wedge(-g)^{1/2}\mathbf{e}^a)^{\mu\nu};\partial_a](\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&\quad + \tfrac{1}{2}(-g)^{-1/2}\partial_a(-g)^{1/2}(\hat{s}\wedge\mathbf{e}^a)^{\mu\nu}(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&= -\frac{1}{2(-g)^{1/2}}\frac{\partial(\hat{s}\wedge(-g)^{1/2}\mathbf{e}^a)^{\mu\nu}}{\partial u^a}(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&\quad + \frac{1}{(-g)^{1/2}}\partial_a(-g)^{1/2}\hat{s}^\mu\mathbf{e}^{a\nu}(\hat{s}_\mu\mathbf{e}^b_\nu - \hat{s}_\nu\mathbf{e}^b_\mu)\partial_b\\
&= -\frac{1}{2(-g)^{1/2}}(\mathbf{e}_a\wedge(-g)^{1/2}\mathbf{e}^a)^{\mu\nu}(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&\quad - \frac{1}{2}\left\{\hat{s}\wedge\frac{1}{(-g)^{1/2}}\frac{\partial((-g)^{1/2}\mathbf{e}^a)}{\partial u^a}\right\}^{\mu\nu}(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b\\
&\quad + \frac{1}{(-g)^{1/2}}\partial_a(-g)^{1/2}(\hat{s}\cdot\hat{s})(\mathbf{e}^a\cdot\mathbf{e}^b)\partial_b\\
&= -\tfrac{1}{2}\{\hat{s}\wedge(-3\hat{s})\}^{\mu\nu}(\hat{s}\wedge\mathbf{e}^b)_{\mu\nu}\partial_b + (-g)^{-1/2}\partial_a(-g)^{1/2}g^{ab}\partial_b\\
&= (-g)^{-1/2}\partial_a(-g)^{1/2}g^{ab}\partial_b,
\end{aligned}$$

showing that the Casimir operator $-\frac{1}{2}L^{\mu\nu}L_{\mu\nu}$ is the intrinsic Laplacian operator $\nabla^2 \equiv (-g)^{-1/2}\partial_a(-g)^{1/2}g^{ab}\partial_b$ of the unit hyperboloid. In the above calculation, the identity [derived from Eqs. (2.12) and (2.13)] $(-g)^{-1/2}\partial\{(-g)^{1/2}\mathbf{e}^a\}/\partial u^a = -3\hat{s}$ was required.

Also, some terms could be dropped because the wedge product vanishes if two factors are equal or parallel. One case in point are the wedge products $\mathbf{e}_1\wedge\mathbf{e}^1 = \mathbf{e}_2\wedge\mathbf{e}^2 = \mathbf{e}_3\wedge\mathbf{e}^3 = 0$; the vector $\mathbf{e}^a$ differs from $\mathbf{e}_a$ only by a scalar factor by virtue of the fact that the metric tensor is diagonal. Substituting the values $g^{11} = -(\sinh\rho)^{-2}$, $g^{22} = -(\sinh\rho\sin\theta)^{-2}$, $g^{33} = -1$, and $(-g)^{1/2} = \sinh^2\rho\sin\theta$ in the expression for the Laplacian now gives the Casimir operator in pseudospherical coordinates:

$$-\tfrac{1}{2}L^{\mu\nu}L_{\mu\nu}=\nabla^2=-\left\{\frac{\partial^2}{\partial\rho^2}+2\coth\rho\frac{\partial}{\partial\rho}-\frac{L_x^2+L_y^2+L_z^2}{\sinh^2\rho}\right\} \quad (3.7)$$

in which the identification (familiar from standard angular momentum theory)

$$L_x^2+L_y^2+L_z^2=-\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right)$$

has been made.

The details of the derivation of the simultaneous eigenfunctions of the three operators $-\frac{1}{2}L^{\mu\nu}L_{\mu\nu}$, $L_x^2+L_y^2+L_z^2$, and L_z are quite standard; consequently, their derivation will be discussed only briefly. Solutions of the simultaneous eigenvalue equations can be found in the separable form $P(\rho)Y_{lm}(\theta,\varphi)$, in which $Y_{lm}(\theta,\varphi)$ denotes the usual spherical harmonics of the Euclidean 2-sphere.[6] Substituting into the eigenvalue equation associated with the operator (3.7) gives a second order ordinary differential equation for $P(\rho)$ which can be converted into the form

$$\left\{(x^2-1)\frac{d^2}{dx^2}+(2l+3)x\frac{d}{dx}+(l+1)^2+\alpha^2\right\}C(\rho)=0, \quad (3.8)$$

$$x\equiv\cosh\rho,$$

of Gegenbauer's equation, by writing $P(\rho)=(\sinh\rho)^lC(\rho)$. In Eq. (3.8) the eigenvalue of ∇^2 is represented in the form $1+\alpha^2$. Initially one must assume α real or pure imaginary, in order that an aribitrary real number admit a representation of the form $1+\alpha^2$. Positive definiteness of ∇^2 imposes the restriction $1+\alpha^2>0$. Hence imaginary values of α, if they occur, must be less than one in absolute magnitude. This possibility of imaginary values of α whose magnitudes are less than one is ruled out by the failure of the corresponding eigenfunctions to vanish sufficiently rapidly in the neighborhood of infinity, $\rho\to\infty$. We are thus left with real α only.[7] Solutions obtained for negative real α are not independent of those for positive real α, hence the further restriction $\alpha>0$. The resulting eigenfunctions, denoted $Y_{\alpha lm}(\rho,\theta,\varphi)$, are

$$Y_{\alpha lm}(\rho,\theta,\varphi)=P_{l\alpha}(\cosh\rho)Y_{lm}(\theta,\varphi), \quad (3.9)$$

where

$$P_{l\alpha}=\left|\frac{\Gamma(l+1+i\alpha)}{\Gamma(i\alpha)\Gamma(l+\frac{3}{2})2^{l+1/2}}\right|(\sinh\rho)^l \times{}_2F_1\left(l+1+i\alpha,l+1-i\alpha,l+\tfrac{3}{2};\frac{1-\cosh\rho}{2}\right). \quad (3.10)$$

This form of $P_{l\alpha}$, involving the Gaussian hypergeometric function ${}_2F_1$, results from an investigation of the series solutions of Eq. (3.8) in ascending powers of $(1-x)/2$. Other forms of $P_{l\alpha}$ are[8]

$$P_{l\alpha}=i(-1)^{l+1}\left|\frac{\Gamma(1+i\alpha)}{\Gamma(1+l+i\alpha)}\right|\left(\frac{2}{\pi}\right)^{1/2}l!2^l(\sinh\rho)^l \times C_{-l-1+i\alpha}^{l+1}(\cosh\rho), \quad (3.11)$$

and

$$P_{l\alpha}=(-1)^l\left|\frac{\Gamma(1+i\alpha)}{\Gamma(1+l+i\alpha)}\right|\left(\frac{2}{\pi}\right)^{1/2} \times(\sinh\rho)^l\frac{d^l}{d(\cosh\rho)^l}\left(\frac{\sin\alpha\rho}{\sinh\rho}\right). \quad (3.12)$$

The form (3.11) can be obtained from Eq. (3.10) by use of the relation[9]

$$C_\lambda^\nu(z)=\frac{\Gamma(\lambda+2\nu)}{\Gamma(\lambda+1)\Gamma(2\nu)}{}_2F_1\left(-\lambda,\lambda+2\nu,\nu+\tfrac{1}{2};\frac{1-z}{2}\right) \quad (3.13)$$

between the Gegenbauer function C_λ^ν and the hypergeometric function. Also, in addition to Eq. (3.13), the usual properties of the gamma function are required.[10] In order to obtain the representation (3.12) one can use the differentiation formula[11]

$$\frac{\Gamma(a+l)\Gamma(b+l)}{\Gamma(c+l)}{}_2F_1(a+l,b+l,c+l;z) =\left(\frac{d}{dz}\right)^l\frac{\Gamma(a)\Gamma(b)}{\Gamma(c)}{}_2F_1(a,b,c;z) \quad (3.14)$$

to rewrite the hypergeometric function in Eq. (3.10) in terms of the $l=0$ function ${}_2F_1(1+i\alpha,\ 1-i\alpha,\ \frac{3}{2};\frac{1}{2}(1-\cosh\rho))$. The latter is then expressed in terms of elementary functions by means of the identity[12]

$${}_2F_1\left(1+i\alpha,\ 1-i\alpha,\ \tfrac{3}{2};\frac{1-\cosh\rho}{2}\right)=\frac{\sin\alpha\rho}{\alpha\sinh\rho}. \quad (3.15)$$

As indicated above the $P_{l\alpha}$ are eigenfunctions having a continuous eigenvalue spectrum. Their normalization as defined in the above equations (3.10) through (3.12) is adjusted to make

$$\int_0^\infty d\rho\sinh^2\rho P_{l\alpha_2}^*(\cosh\rho)P_{l\alpha_1}(\cosh\rho)=\delta(\alpha_2-\alpha_1). \quad (3.16)$$

The normalization of the full generalized spherical harmonics (3.9) is then

$$\int d\Sigma Y_{\alpha_2 l_2 m_2}^*Y_{\alpha_1 l_1 m_1}=\delta_{l_2 l_1}\delta_{m_2 m_1}\delta(\alpha_2-\alpha_1). \quad (3.17)$$

With this normalization, the completeness relation for the $Y_{\alpha lm}$ reads

$$\delta(21)=\sum_{l=0}^\infty\sum_{m=-l}^{l}\int_0^\infty d\alpha\, Y_{\alpha lm}(2)Y_{\alpha lm}^*(1). \quad (3.18)$$

The delta function of position on the unit hyperboloid $\delta(21)$, is defined in terms of Dirac delta functions by

$$\delta(21)=\delta(\rho_2-\rho_1)\delta(\theta_2-\theta_1)\delta(\varphi_2-\varphi_1)/\sinh^2\rho_2\sin\theta_2,$$

and has the characteristic property

$$f(2)=\int d\Sigma_1 f(1)\delta(21), \quad (3.19)$$

for any sufficiently smooth function f. For completeness, we record explicitly the defining equations

$$\nabla^2Y_{\alpha lm}=(1+\alpha^2)Y_{\alpha lm},$$
$$(L_x^2+L_y^2+L_z^2)Y_{\alpha lm}=l(l+1)Y_{\alpha lm}, \quad (3.20)$$
$$L_zY_{\alpha lm}=mY_{\alpha lm},$$

of the scalar spherical harmonics of the unit hyperboloid.

We close this section with an application of the theory of spherical harmonics of the unit hyperboloid to the problem of calculating the Green's function of Poisson's

equation on the unit hyperboloid. This Green's function, denoted $G(21)$, is the solution of the differential equation

$$\nabla_2^2 G(21) = \delta(21). \tag{3.21}$$

The point source $\delta(21)$ in this equation is the delta function of position on the unit hyperboloid defined above. Equation (3.21) will be solved by expanding $G(21)$, regarded as a function of its arguments ρ_2, θ_2, φ_2, as a linear superposition of the functions $Y_{\alpha lm}(2)$,

$$G(21) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^{\infty} d\alpha\, Y_{\alpha lm}(2) C_{\alpha lm}, \tag{3.22}$$

and then substituting into the differential equation (3.21). The result of this substitution

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^{\infty} d\alpha (1+\alpha^2) Y_{\alpha lm}(2) C_{\alpha lm} = \delta(21) \tag{3.23}$$

makes use of the first eigenvalue equation (3.20) of the spherical harmonics of the unit hyperboloid. The generalized angular momentum expansion of $\delta(21)$ is provided by the completeness relation (3.18). Using Eq. (3.18) we can write Eq. (3.23) in the form

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^{\infty} d\alpha (1+\alpha^2) Y_{\alpha lm}(2) C_{\alpha lm} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^{\infty} d\alpha\, Y_{\alpha lm}(2) Y^*_{\alpha lm}(1).$$

By equating coefficients of corresponding orthogonal functions $Y_{\alpha lm}(2)$ on both sides of this equation, we find the expansion coefficients of $G(21)$ in the form $C_{\alpha lm} = Y^*_{\alpha lm}(1)/(1+\alpha^2)$. Going back to Eq. (3.22), we find the following eigenfunction expansion of $G(21)$:

$$G(21) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^{\infty} d\alpha\, \frac{Y_{\alpha lm}(2) Y^*_{\alpha lm}(1)}{1+\alpha^2}\ . \tag{3.24}$$

The discrete sum in Eq. (3.24) can be performed by means of the following special case:

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_{\alpha lm}(2) Y^*_{\alpha lm}(1) = \frac{\alpha}{2\pi^2} \frac{\sin(\alpha\rho_{21})}{\sinh\rho_{21}}\ , \tag{3.25}$$

$$\cosh\rho_{21} \equiv \hat{s}_2 \cdot \hat{s}_1,$$

of the general addition theorem for Gegenbauer functions of Durand, Fishbane, and Simmons.[13] The quantity ρ_{21} in Eq. (3.25) can be viewed as the ρ coordinate of one of the two timelike vectors $\hat{s}_2$, $\hat{s}_1$ measured in a Lorentz frame of reference for which the other vector is aligned along the time axis. Using the addition theorem (3.25), our eigenfunction expansion (3.24) becomes

$$G(21) = \frac{1}{\sinh\rho_{21}} \int_0^{\infty} \frac{\alpha d\alpha}{2\pi^2} \frac{\sin(\alpha\rho_{21})}{1+\alpha^2}\ . \tag{3.26}$$

This integral can be evaluated by means of the calculus of residues: We write it in the form $G(21) = \operatorname{csch}\rho_{21}(4\pi^2)^{-1} \operatorname{Im}\int \alpha\, d\alpha (1+\alpha^2)^{-1} \exp(i\alpha\rho_{21})$ and close the contour in the upper half α plane, obtaining

$$G(21) = \operatorname{csch}\rho_{21}(4\pi^2)^{-1} \operatorname{Im} \oint_{(i+)} (\alpha - i)^{-1} d\alpha\, \alpha(\alpha+i)^{-1} \times \exp(i\alpha\rho_{21})$$

$$= \operatorname{csch}\rho_{21}(4\pi^2)^{-1} \operatorname{Im} 2\pi i (2i)^{-1} \exp(ii\rho_{21})$$

$$= \operatorname{csch}\rho_{21}(4\pi)^{-1} \exp(-\rho_{21}),$$

which can be rewritten

$$G(21) = \frac{1}{2\pi} \frac{1}{(e^{2\rho_{21}} - 1)}\ . \tag{3.27}$$

This is the Green's function of Poisson's equation on the unit hyperboloid. It can be verified by explicit calculation that the function (3.27) obeys Laplace's equation $\nabla_2^2 G(21) = 0$, for $\mathbf{x}_2 \neq \mathbf{x}_1$, and has a singularity at $\mathbf{x}_2 = \mathbf{x}_1$ of the strength and type demanded by the delta function source term in Eq. (3.21).

B. Vector spherical harmonics of the unit hyperboloid

Lorentz transformations of a vector field in Minkowski space are described by a total angular momentum tensor $[J^{\mu\nu}]$ consisting of the orbital angular momentum tensor $L^{\mu\nu}\mathbf{I}$ plus a spin angular momentum tensor $[S^{\mu\nu}]$:

$$[J^{\mu\nu}] = L^{\mu\nu}\mathbf{I} + [S^{\mu\nu}]. \tag{3.28}$$

The bracket around a quantity signifies that the quantity enclosed is a 4×4 matrix. An exception is the 4×4 unit matrix I, for which the bracket is deleted. The operator $L^{\mu\nu}$ is the orbital angular momentum operator of Sec. IIIA. The action of the spin matrix $[S^{\mu\nu}]$ on a 4-vector field **A** is described by the equation

$$([S^{\mu\nu}]\mathbf{A})^{\alpha} = [S^{\mu\nu}]^{\alpha}_{\beta} A^{\beta} \tag{3.29}$$

in which the $\alpha-\beta$ matrix element of $[S^{\mu\nu}]$ appears. The matrix element $[S^{\mu\nu}]^{\alpha}_{\beta}$ has the representation

$$[S^{\mu\nu}]^{\alpha}_{\beta} = (1/i)(g^{\mu\alpha}\delta^{\nu}_{\beta} - g^{\nu\alpha}\delta^{\mu}_{\beta}), \tag{3.30}$$

which can be used to show that the components of the spin tensor $[S^{\mu\nu}]$ obey the same Lie algebra (3.2) as the $L^{\mu\nu}$. The components of the total angular momentum tensor $[J^{\mu\nu}] = L^{\mu\nu}\mathbf{I} + [S^{\mu\nu}]$ then likewise obey this Lie algebra. (The action of $L^{\mu\nu}\mathbf{I}$ on a 4-vector field is simply that the differential operator $L^{\mu\nu}$ acts on each Cartesian component:

$$\{(L^{\mu\nu}\mathbf{I})\mathbf{A}\}^{\alpha} \equiv L^{\mu\nu}\{\mathbf{I}\mathbf{A}\}^{\alpha} = L^{\mu\nu}A^{\alpha}.)$$

Of particular interest are the components $[S_x] \equiv [S^{32}]$, $[S_y] \equiv [S^{13}]$, and $[S_z] \equiv [S^{21}]$ of the spin tensor. Explicitly

$$[S_x]^{\alpha}_{\beta} = \begin{array}{c} \beta\rightarrow \\ \alpha\downarrow \end{array} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix},$$

$$[S_y]^{\alpha}_{\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix},$$

$$[S_z]^{\alpha}_{\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The operators $[J_x] \equiv L_x \mathbf{I} + [S_x]$, $[J_y] \equiv L_y \mathbf{I} + [S_y]$, and $[J_z] \equiv L_z \mathbf{I} + [S_z]$ describe simple rotations of the 4-vector field in the xyz space. It is readily shown that the set of operators [analogous to the set (3.3)]

$$\nabla^2, \quad [J_x]^2 + [J_y]^2 + [J_z]^2, \quad [J_z] \tag{3.31}$$

is a mutually commuting set. Again, our angular momentum operators can be viewed as linear operators acting on a space of functions defined on the unit hyperboloid, the space of functions consisting this time of the set of 4-vector fields defined on the unit hyperboloid. A 4-vector field of the unit hyperboloid which is a simultaneous eigenfunction of the operators (3.31) will be called a *vector spherical harmonic of the unit hyperboloid*. The set of 4-vector fields of the unit hyperboloid endowed with the dot product

$$(\mathbf{A}_2; \mathbf{A}_1) \equiv \int d\Sigma \, \mathbf{A}_2^* \cdot \mathbf{A}_1, \tag{3.32}$$

fails to form a Hilbert space with positive definite metric: the indefinite metric in Minkowski space makes the dot product (3.32) in the Hilbert space not positive definite. Because of this lack of positive definiteness, it is not straightforward to apply the usual spectral theorem to the set of commuting operators (3.31). In the following we investigate the possibility of finding a complete orthogonal set of simultaneous eigenfunctions of the operators (3.31); i.e., of finding a complete orthogonal set of vector spherical harmonics of the unit hyperboloid.

In order to cope with the lack of positive definiteness of the dot product (3.32) we start out by restricting consideration to the set of *tangent vector fields* of the hyperboloid, for which the dot product (3.32) will be negative definite. Application of the usual spectral theorem would now become straightforward; but a new difficulty is encountered: the operator ∇^2 does not always change a tangent vector field into another tangent vector field. In short, ∇^2 cannot be viewed as a linear transformation of our new Hilbert space into itself! The closure property of ∇^2 proved at the end of Sec. II indicates a way out of this last difficulty, however. If we further restrict our Hilbert space to be the Hilbert space of solenoidal tangent vector fields, then we shall have both the needed positive definiteness and closure. The usual spectral theorem now guarantees us the existence of a complete orthogonal set of solenoidal tangent vector fields which are simultaneous eigenfunctions of the mutually commuting self-adjoint operators (3.31). The term "complete set" must here be interpreted to mean that any solenoidal tangent vector field can be expanded in a linear superposition of the eigenfunctions. The eigenfunctions referred to here (solenoidal eigenfunctions tangent to the unit hyperboloid) turn out to be the only eigenfunctions needed for our future application to the electromagnetic field. Therefore, as indicated in the introduction, the discussion of this section will be limited to the calculation of the solenoidal vector spherical harmonics. See Appendix B for the treatment of the nonsolenoidal vector spherical harmonics. Also, as mentioned in the introduction, Appendix B contains the derivation of an addition theorem for vector spherical harmonics of the unit hyperboloid which parallels the scalar spherical harmonic addition theorem (3.25) of Durand, Fishbane, and Simmons.

It is convenient to characterize all possible tangent vector fields by means of a form of Helmholtz's theorem. Any tangent vector field $\mathbf{A}$ can be written as a radial part plus a transverse part

$$\mathbf{A} = \hat{\rho} A\rho + \mathbf{A}^T, \quad \mathbf{A}^T \equiv \hat{\theta} A_\theta + \hat{\varphi} A\varphi.$$

For a fixed ρ the transverse part $\mathbf{A}^T$, regarded as a function of θ and φ, can be regarded as a tangent vector field for the unit 2-sphere described by the equations $x^0 = 0$, $x^1 = \sin\theta\cos\varphi$, $x^2 = \sin\theta\sin\varphi$, $x^3 = \cos\theta$. To see this one verifies that the unit vectors tangent to the coordinate curves $\varphi = \text{const}$, $\theta = \text{const}$, of the above 2-sphere are the same vectors $\hat{\theta}$ and φ respectively defined earlier. The form of Helmholtz's theorem in question states that any tangent vector field of the 2-sphere has a unique decomposition into a conservative and solenoidal part

$$\begin{aligned} &\mathbf{A}^T = \partial V + \mathbf{A}_1^T, \\ &\partial \cdot \mathbf{A}_1^T = 0, \end{aligned} \tag{3.33}$$

in which $\partial = (\hat{\theta}(\partial/\partial\theta) + \hat{\varphi}\csc\theta(\partial/\partial\varphi))$ is the intrinsic gradient operator of the unit 2-sphere. It is readily shown that the solenoidal part $\mathbf{A}_1^T$ admits a representation of the form

$$\mathbf{A}_1^T = -\hat{\varphi}\frac{\partial W}{\partial\theta} + \frac{\hat{\theta}}{\sin\theta}\frac{\partial W}{\partial\varphi}, \tag{3.34}$$

for some scalar function, W. For, let $\mathbf{A}_1^T = \hat{\theta}A_{1\theta} + \hat{\varphi}A_{1\varphi}$. The divergence condition $\partial \cdot A_1^T = 0$ reads $-\csc\theta(\partial/\partial\theta)\sin\theta A_{1\theta} - \csc\theta(\partial/\partial\varphi)A_{1\varphi} = 0$, which is equivalent to the vanishing of a curl $(\partial/\partial\varphi)B_\theta = (\partial/\partial\theta)B_\varphi$, $B_\theta \equiv -A_{1\varphi}$, $B_\varphi \equiv \sin\theta A_{1\theta}$. But the vanishing of the curl implies that there exists a potential function W such that $B_\theta = (\partial/\partial\theta)W$ and $B_\varphi = (\partial/\partial\varphi)W$, leading to the equations $A_{1\varphi} = -(\partial/\partial\theta)W$ and $A_{1\theta} = \csc\theta(\partial/\partial\varphi)W$, in agreement with Eq. (3.34). Combining Eq. (3.34) and Eq. (3.33), we find $\mathbf{A}^T$ in the form

$$A^T = \partial V + (-\hat{\varphi}(\partial/\partial\theta)W + \hat{\theta}\csc\theta(\partial/\partial\varphi)W).$$

To this we add the radial part $\hat{\rho}A_\rho$ of $\mathbf{A}$, and find the representation

$$\mathbf{A} = \hat{\rho}A_\rho + \partial V + \left(-\hat{\varphi}\frac{\partial W}{\partial\theta} + \frac{\hat{\theta}}{\sin\theta}\frac{\partial W}{\partial\varphi}\right) \tag{3.35}$$

of a general vector field tangent to the unit hyperboloid. An equivalent form of Eq. (3.35)

$$\mathbf{A} = \hat{\rho}A_\rho + \partial V + \partial \times (\hat{\rho}W) \tag{3.36}$$

exhibits the part of $\mathbf{A}^T$ which is solenoidal with respect to ∂ as a curl type of structure formed from ∂. The curl type operator in Eq. (3.36) is defined in conformity with Eq. (2.16) for $\nabla\times$:

$$\partial \times \equiv \mathbf{e}_a(-g)^{-1/2}\epsilon^{abc}\partial_b\mathbf{e}_c.$$

For applications to the vector spherical harmonics of the unit hyperboloid it is advantageous to go over from a representation of A involving ∂ to a representation involving the full gradient operator ∇. This is made possible by the identities $\partial \times (\hat{\rho}W) = \nabla \times (\hat{\rho}W\sinh\rho)$ and $\hat{\rho}A_\rho + \partial V = \hat{\rho}U\sinh\rho + \nabla(V\sinh\rho)$, in which

$$U \equiv (\sinh\rho)^{-1}(A_\rho - \partial(V\sinh\rho)/\partial\rho).$$

In terms of ∇ the representation (3.36) becomes

$$\mathbf{A} = \hat{\rho} U \sinh\rho + \nabla(V \sinh\rho) + \nabla\times(\hat{\rho} W \sinh\rho). \tag{3.37}$$

Taking the curl of the general tangent vector field (3.37) will provide us with a general representation of a solenoidal tangent vector field. Since the curl of the gradient is zero; the representation obtained is

$$\mathbf{Y} = \nabla\times(\hat{\rho}\sinh\rho U) + \nabla\times\nabla\times(\hat{\rho}\sinh\rho W), \tag{3.38}$$

in which $\mathbf{Y}$ denotes any solenoidal tangent vector field. In particular, our solenoidal vector spherical harmonics of the unit hyperboloid must have this form.

Proceeding, we investigate the effect of applying the operator $J^{\mu\nu}$ to a structure of the type (3.38). The commutation relation

$$J^{\mu\nu}\nabla\times = \nabla\times J^{\mu\nu} \tag{3.39}$$

enables us to bring the operator $J^{\mu\nu}$ in $J^{\mu\nu}\nabla\times(\hat{\rho}\sinh\rho U)$ and $J^{\mu\nu}\nabla\times\nabla\times(\hat{\rho}\sinh\rho W)$ to the arguments $\hat{\rho}\sinh\rho U$ and $\hat{\rho}\sinh\rho W$ of the curls. It follows that $\nabla\times(\hat{\rho}\sinh\rho U)$ and $\nabla\times\nabla\times(\hat{\rho}\sinh\rho W)$ individually would be simultaneous eigenfunctions of $[J_x]^2+[J_y]^2+[J_z]^2$ and $[J_z]$ if $\hat{\rho}\sinh\rho U$ and $\hat{\rho}\sinh\rho W$ were. In fact, the intertwining relations

$$J^{lm}\hat{\rho} = \hat{\rho} L^{lm}, \quad l,m = 1,2,3. \tag{3.40}$$

show that $\hat{\rho}\sinh\rho U$ and $\hat{\rho}\sinh\rho W$ become simultaneous eigenfunctions of $[J_x]^2+[J_y]^2+[J_z]^2$ and $[J_z]$ on simply choosing U and W to be simultaneous eigenfunctions of $L_x^2+L_y^2+L_z^2$ and L_z; i.e., if U and W are taken to have the form $f(\rho)Y_{lm}(\theta,\varphi)$. We shall complete our calculation of the solenoidal vector spherical harmonics of the unit hyperboloid and then shall discuss the meaning of the intertwining relations (3.40).

We come now to the investigation of the effect of ∇^2 on the structures $\nabla\times(\hat{\rho}\sinh\rho U)$ and $\nabla\times\nabla\times(\hat{\rho}\sinh\rho W)$. According to the commutation relation (2.18) ∇^2 may also be moved to the arguments $\hat{\rho}\sinh\rho U$ and $\hat{\rho}\sinh\rho W$ of the curls: $\nabla^2\nabla\times(\hat{\rho}\sinh\rho U) = \nabla\times(\nabla^2\hat{\rho}\sinh\rho U)$, and similarly for the other function $\nabla\times\nabla\times(\hat{\rho}\sinh\rho W)$. In the arguments of the curls we write

$$\nabla^2\hat{\rho}\sinh\rho = [\nabla^2;\hat{\rho}\sinh\rho] + \hat{\rho}\sinh\rho\nabla^2,$$

giving

$$\nabla^2\nabla\times(\hat{\rho}\sinh\rho U) = \nabla\times[\nabla^2;\hat{\rho}\sinh\rho]U + \nabla\times(\hat{\rho}\sinh\rho\nabla^2 U), \tag{3.41}$$

with a similar equation for the function $\nabla\times\nabla\times(\hat{\rho}\sinh\rho W)$. The commutator $[\nabla^2;\hat{\rho}\sinh\rho]$ would be lengthy to evaluate but it will be shown that the factor $\nabla\times$ in Eq. (3.41) annuls the commutator, giving simply

$$\nabla^2\nabla\times(\hat{\rho}\sinh\rho U) = \nabla\times(\hat{\rho}\sinh\rho\nabla^2 U). \tag{3.42}$$

Thus

$$\begin{aligned}
&\nabla\times[\nabla^2;\hat{\rho}\sinh\rho] \\
&= [\nabla^2;\nabla\times\hat{\rho}\sinh\rho] - [\nabla^2;\nabla\times]\hat{\rho}\sinh\rho \\
&= [\nabla^2;\nabla\times\hat{\rho}\sinh\rho] \\
&= [\nabla^2;\mathbf{e}_a(-g)^{-1/2}\epsilon^{abc}\partial_c\mathbf{e}_b\cdot\hat{\rho}\sinh\rho] \\
&= [\nabla^2;\mathbf{e}_a(-g)^{-1/2}\epsilon^{a3c}\partial_c\hat{\rho}\cdot\hat{\rho}\sinh\rho] \\
&= \left[\nabla^2;\frac{\mathbf{e}_1\epsilon^{132}}{\sinh^2\rho\sin\theta}\frac{\partial}{\partial\varphi}(-\sinh\rho)\right. \\
&\quad \left. + \frac{\mathbf{e}_2\epsilon^{231}}{\sinh^2\rho\sin\theta}\frac{\partial}{\partial\theta}(-\sinh\rho)\right] \\
&= \left[\nabla^2;\frac{\sinh\rho\hat{\theta}(-1)(-\sinh\rho)}{\sinh^2\rho\sin\theta}\frac{\partial}{\partial\varphi}\right. \\
&\quad \left. + \frac{\sinh\rho\sin\theta\hat{\varphi}(-\sinh\rho)}{\sinh^2\rho\sin\theta}\frac{\partial}{\partial\theta}\right] \\
&= \left[\nabla^2;\frac{\hat{\theta}}{\sin\theta}\frac{\partial}{\partial\varphi} - \hat{\varphi}\frac{\partial}{\partial\theta}\right] \\
&= \left[\nabla^2;\frac{1}{i}(\hat{i}_1 L_x + \hat{i}_2 L_y + \hat{i}_3 L_z)\right] \\
&= \left[-\frac{\partial^2}{\partial\rho^2} - 2\coth\rho\frac{\partial}{\partial\rho} + \frac{L_x^2+L_y^2+L_z^2}{\sinh^2\rho};\right. \\
&\quad \left.\frac{1}{i}(\hat{i}_1 L_x + \hat{i}_2 L_y + \hat{i}_3 L_z)\right] = 0.
\end{aligned}$$

This calculation justifies the transition from Eq. (3.41) to Eq. (3.42).[14] In the last step of the calculation the fact, familiar from standard angular momentum theory, that $L_x^2+L_y^2+L_z^2$ commutes with the Cartesian components L_x, L_y, and L_z was used. The analogous result for W can be obtained by changing U into W in Eq. (3.42) and applying $\nabla\times$ to both sides of the equation. This result is

$$\nabla^2\nabla\times\nabla\times(\hat{\rho}\sinh\rho W) = \nabla\times\nabla\times(\hat{\rho}\sinh\rho\nabla^2 W). \tag{3.43}$$

We are now in a position to write down the solenoidal vector spherical harmonics of the unit hyperboloid denoted $\mathbf{Y}^{\rho}_{\alpha lm}(\rho,\theta,\varphi)$. These will be defined as

$$\mathbf{Y}^1_{\alpha lm} \equiv \frac{\nabla\times(\hat{\rho}\sinh\rho Y_{\alpha lm})}{[l(l+1)]^{1/2}} \tag{3.44}$$

and

$$\mathbf{Y}^2_{\alpha lm} \equiv \frac{\nabla\times\nabla\times(\hat{\rho}\sinh\rho Y_{\alpha lm})}{\alpha[l(l+1)]^{1/2}}$$

$$l = 1,2,3,\cdots;$$

in which $Y_{\alpha lm}$ denotes the scalar spherical harmonic of Eqs. (3.9) through (3.12). It is perhaps worthwhile summarizing our arguments briefly. Let $\mathcal{O}$ be one of the three operators $[J_x]^2+[J_y]^2+[J_z]^2$, $[J_z]$, and ∇^2. To a particular $\mathcal{O}$ associate the operator $\mathcal{T} = L_x^2+L_y^2+L_z^2$, L_z, or ∇^2, respectively. By use of the intertwining relation (3.40) or Eq. (3.42) as appropriate, we find

$$\begin{aligned}
&\mathcal{O}\nabla\times(\hat{\rho}\sinh\rho Y_{\alpha lm}) = \nabla\times(\hat{\rho}\sinh\hat{\rho}\mathcal{T}Y_{\alpha lm}) \\
&= \nabla\times(\hat{\rho}\sinh\rho t Y_{\alpha lm}) = t\nabla\times(\hat{\rho}\sinh\rho Y_{\alpha lm}),
\end{aligned}$$

showing that the vector field $\nabla\times(\hat{\rho}\sinh\rho Y_{\alpha lm})$ is an eigenfunction of $\mathcal{O}$ to the eigenvalue t. Corresponding eigenvalues of the vector and scalar spherical harmonics are equal: the eigenvalue t of $\mathcal{O}$ is equal to $l(l+1)$, m, or $1+\alpha^2$, respectively. The simplified curl–curl identity (2.19) used in conjunction with Eqs. (3.44) enables one to demonstrate the identities

$$\nabla\times\mathbf{Y}^2_{\alpha lm}=\alpha\mathbf{Y}^1_{\alpha lm} \tag{3.45}$$

and

$$\nabla\times\mathbf{Y}^1_{\alpha lm}=\alpha\mathbf{Y}^2_{\alpha lm}.$$

In order to determine the normalization of the $\mathbf{Y}^1_{\alpha lm}$ it is simplest to go back to a representation in terms of ∂:

$$\mathbf{Y}^1_{\alpha lm}=\frac{\partial\times(\hat\rho Y_{\alpha lm})}{[l(l+1)]^{1/2}}. \tag{3.46}$$

To proceed, we note the operator identity $\partial\times\hat\rho = -i\hat\rho\times(1/i)\partial = -i\mathbf{L}$; which uses the representation $\mathbf{L}=\hat\rho\times(1/i)\partial$ of the orbital angular momentum operator $\mathbf{L}=\hat i_1 L_x+\hat i_2 L_y+\hat i_3 L_z$. Thus

$$\begin{aligned}
&\int d\Sigma\,\mathbf{Y}^{1*}_{\alpha_2 l_2 m_2}\cdot\mathbf{Y}^1_{\alpha_1 l_1 m_1}\\
&\quad=\int d\Sigma(-i\mathbf{L}Y_{\alpha_2 l_2 m_2})^*\cdot(-i\mathbf{L}Y_{\alpha_1 l_1 m_1})[l(l+1)]^{-1}\\
&\quad=\int d\Sigma\, Y^*_{\alpha_2 l_2 m_2}\mathbf{L}\cdot\mathbf{L}Y_{\alpha_1 l_1 m_1}[l(l+1)]^{-1}\\
&\quad=-\int d\Sigma\, Y^*_{\alpha_2 l_2 m_2}(L_x^2+L_y^2+L_z^2)Y_{\alpha_1 l_1 m_1}[l(l+1)]^{-1}\\
&\quad=-\int d\Sigma\, Y^*_{\alpha_2 l_2 m_2}Y_{\alpha_1 l_1 m_1}\\
&\quad=-\delta(\alpha_2-\alpha_1)\delta_{l_2 l_1}\delta_{m_2 m_1}.
\end{aligned}$$

The self-adjointness of $\mathbf{L}$ was required here in moving $\mathbf{L}$ from one side of a unitary dot product to the other. The final minus sign is inherited from the indefinite Minkowski metric, according to which $\hat i_1\cdot\hat i_1=\hat i_2\cdot\hat i_2=\hat i_3\cdot\hat i_3=-1$. Hence $\mathbf{L}\cdot\mathbf{L}=-(L_x^2+L_y^2+L_z^2)$. In order to show that $\mathbf{Y}^1_{\alpha_2 l_2 m_2}$ is orthogonal to $\mathbf{Y}^2_{\alpha_1 l_1 m_1}$ we form their unitary dot product using the representations (3.44). The curl operator of $\mathbf{Y}^1_{\alpha_2 l_2 m_2}$ is moved to the other side of the dot product, by integration by parts, producing a $\nabla\times\nabla\times\nabla\times$ structure. The latter can be evaluated as $(\nabla^2-1)\nabla\times$ by use of the simplified curl–curl identity (2.19). Thus:

$$\begin{aligned}
&\int d\Sigma\,\mathbf{Y}^{1*}_{\alpha_2 l_2 m_2}\cdot\mathbf{Y}^2_{\alpha_1 l_1 m_1}\\
&\quad=\int d\Sigma\,\hat\rho\sinh\rho Y^*_{\alpha_2 l_2 m_2}\cdot(\nabla^2-1)\nabla\times\\
&\qquad\times(\hat\rho\sinh\rho Y_{\alpha_1 l_1 m_1})[\alpha l(l+1)]^{-1}\\
&\quad=\int d\Sigma\,\sinh\rho Y^*_{\alpha_2 l_2 m_2}\hat\rho\cdot\partial\times(\hat\rho\, Y_{\alpha_1 l_1 m_1})\alpha^2[\alpha l(l+1)]^{-1}\\
&\quad=-\int d\Sigma\,\sinh\rho Y^*_{\alpha_2 l_2 m_2}\hat\rho\cdot(\hat\rho\times\partial)Y_{\alpha_1 l_1 m_1}\alpha[l(l+1)]^{-1}\\
&\quad=0,
\end{aligned}$$

by virtue of the operator identity $\hat\rho\cdot(\hat\rho\times\partial)=0$. We omit the few remaining details of the normalization calculations, and just state the final result; that the functions (3.44) have the normalization

$$\begin{aligned}
&\int d\Sigma\,\mathbf{Y}^{\rho_2 *}_{\alpha_2 l_2 m_2}\cdot\mathbf{Y}^{\rho_1}_{\alpha_1 l_1 m_1}\\
&\quad=-\delta(\alpha_2-\alpha_1)\delta_{\rho_2\rho_1}\delta_{l_2 l_1}\delta_{m_2 m_1}.
\end{aligned} \tag{3.47}$$

The completeness relation for the solenoidal vector spherical harmonics is

$$\begin{aligned}
&-\sum_{\rho lm}\int_0^\infty d\alpha\, Y^\rho_{\alpha lm}(2)_\mu Y^{\rho *}_{\alpha lm}(1)_\nu\\
&\quad=\{\delta(21)(g_{\mu\nu}-\hat s_{2\mu}\hat s_{1\nu})+\nabla_{2\mu}\nabla_{1\nu}G(21)\}.
\end{aligned} \tag{3.48}$$

In Eq. (3.48) $G(21)$ is the Green's function of Poisson's equation on the unit hyperboloid investigated in Sec. IIIA and given by Eq. (3.27). Equation (3.48) can be proven by showing that both sides of the equation provide a representation of the same kernel $K(21)_{\mu\nu}$ of a linear operator K which acts on an arbitrary 4-vector field $\mathbf{F}$ according to the rule $\{K\mathbf{F}(2)\}_\mu\equiv\int d\Sigma_1 K(21)_{\mu\nu}F(1)^\nu$ and projects $\mathbf{F}$ onto the solenoidal part of its tangential component.

In closing this section we shall discuss the meaning of the intertwining relations (3.40). The existence of these relations was responsible above for the utility of the representation (3.38) of a general solenoidal tangent vector field. Equations (3.40) are one of a family of intertwining relations. Other examples are

$$[J^{\mu\nu}]\hat s=\hat s L^{\mu\nu}, \tag{3.49}$$

$$[J^{\mu\nu}]\nabla=\nabla L^{\mu\nu}. \tag{3.50}$$

The relations (3.49) and (3.50) will be required in Appendix B to derive the remaining vector spherical harmonics of the unit hyperboloid. The intertwining relations (3.40), (3.49), and (3.50) all have a similar origin. Their meaning is indicated by the following study of Eq. (3.49). We begin by showing that $[J^{\mu\nu}]\hat s=0$.

Using Eqs. (3.4) and (3.30), we find

$$\begin{aligned}
L^{\mu\nu}\hat s^\alpha&=(1/i)(\hat s^\mu e^{a\nu}-\hat s^\nu e^{a\mu})(\partial\hat s^\alpha/\partial u^a)\\
&=(1/i)(\hat s^\mu e^{a\nu}e_a^\alpha-\hat s^\nu e^{a\mu}e_a^\alpha)\\
&=(1/i)(\hat s^\mu(g^{\nu\alpha}-\hat s^\nu\hat s^\alpha)-\hat s^\nu(g^{\mu\alpha}-\hat s^\mu\hat s^\alpha))\\
&=(1/i)(\hat s^\mu g^{\nu\alpha}-\hat s^\nu g^{\mu\alpha})\\
&=-(1/i)(g^{\mu\alpha}\delta^\nu_\beta-g^{\nu\alpha}\delta^\mu_\beta)\hat s^\beta\\
&=-[S^{\mu\nu}]^\alpha_{\ \beta}\hat s^\beta=-\{[S^{\mu\nu}]\hat s\}^\alpha.
\end{aligned}$$

Transposing the spin term gives the desired result:

$$L^{\mu\nu}\hat s+[S^{\mu\nu}]\hat s=[J^{\mu\nu}]\hat s=0. \quad\surd$$

When acting on $\hat s$ the effect of the derivatives in $L^{\mu\nu}$ is annulled by the effect of the spin matrix $[S^{\mu\nu}]$. In view of the interpretation of $[J^{\mu\nu}]$ as generating Lorentz transformations, the equation $[J^{\mu\nu}]\hat s=0$ expresses the fact that $\hat s$ is a Lorentz invariant vector field. Next we investigate the effect of $L^{\mu\nu}$ and $[S^{\mu\nu}]$ on a product sU. In accordance with the law of differentiating a product, the first order derivatives in $L^{\mu\nu}$ produce a sum of two terms $L^{\mu\nu}(sU)=(L^{\mu\nu}\hat s)U+\hat s(L^{\mu\nu}U)$. In the first of these terms $L^{\mu\nu}$ acts only on $\hat s$; in the second, only on U. When the constant matrix $[S^{\mu\nu}]$ acts, it acts effectively only on the factor $\hat s$, $[S^{\mu\nu}](\hat sU)=([S^{\mu\nu}]\hat s)U$. Now when $[J^{\mu\nu}](\hat sU)$ is formed, we again have a sum of two terms; but $\hat s$ is acted on by the full operator $[J^{\mu\nu}]$, while U is acted on only by $L^{\mu\nu}$:

$$[J^{\mu\nu}](\hat sU)=([J^{\mu\nu}]\hat s)U+\hat s(L^{\mu\nu}U).$$

By the invariance of $\hat s$, the first term here vanished. Therefore, the effect of $[J^{\mu\nu}]$ on the vector field $\hat sU$ is expressed through the action of $L^{\mu\nu}$ on the component U, $[J^{\mu\nu}]\hat sU=\hat sL^{\mu\nu}U$. Dropping the arbitrary argument of the operators here gives the intertwining relation

(3.49). In order to prove Eq. (3.50) the representation [c.f. Eq. (3.4)] $L^{\mu\nu}=i(\hat{s}^{\mu}\nabla^{\nu}-\hat{s}^{\nu}\nabla^{\mu})$ of $L^{\mu\nu}$ is used. One begins by reducing the commutator $[L^{\mu\nu};\nabla^{\alpha}]$ using the identities (2.9) and (2.10). The appearance of the spin operator in the resulting expression is recognized through the use of Eq. (3.30). Equation (3.40) is a bit unusual. The vector field $\hat{\rho}$ is invariant under ordinary rotations of xyz space, but not under the full Lorentz group. This accounts for the restriction $l, m = 1, 2, 3$ only in Eq. (3.40). The $\hat{\rho}$ direction in the tangent space of the unit hyperboloid is now a distinguished direction; and we require the special identities $\partial\hat{\rho}/\partial u^{1,2}=\coth\rho\,\mathbf{e}_{1,2}$; and $\coth\rho\hat{s}^{k}=\hat{\rho}^{k}$, $k=1,2$, and 3. Thus

$$
\begin{aligned}
[L^{lm};\hat{p}^{\alpha}] &= [-i(\hat{s}^{l}e^{am}-\hat{s}^{m}e^{al})\partial_{a};\hat{p}^{\alpha}] \\
&= -i(\hat{s}^{l}e^{am}-\hat{s}^{m}e^{al})(\partial\hat{\rho}^{\alpha}/\partial u^{a}) \\
&= \sum_{a=1}^{2}(-i)(\hat{s}^{l}e^{am}-\hat{s}^{m}e^{al})\coth\rho e_{a}^{\alpha}-i(\hat{s}^{l}e^{3m}-\hat{s}^{m}e^{3l})\hat{s}^{\alpha} \\
&= \sum_{a=1}^{2}(-i)(\hat{\rho}^{l}e^{am}-\hat{\rho}^{m}e^{al})e_{a}^{\alpha} \\
&\quad -i\tanh\rho(\hat{\rho}^{l}e^{3m}-\hat{\rho}^{m}e^{3l})\hat{s}^{\alpha} \\
&= -i(\hat{\rho}^{l}e^{am}-\hat{\rho}^{m}e^{al})e_{a}^{\alpha}=-i\hat{\rho}^{l}(g^{m\alpha}-\hat{s}^{m}\hat{s}^{\alpha}) \\
&\quad +i\hat{\rho}^{m}(g^{l\alpha}-\hat{s}^{l}\hat{s}^{\alpha}) \\
&= i(g^{l\alpha}\delta_{\mu}^{m}-g^{m\alpha}\delta_{\mu}^{l})\hat{\rho}^{\mu}+i\tanh\rho(\hat{\rho}^{l}\hat{\rho}^{m}-\hat{\rho}^{m}\rho^{l})\hat{s}^{\alpha} \\
&= -[S^{lm}]^{\alpha}{}_{\mu}\hat{\rho}^{\mu}=-\{[S^{lm}]\hat{\rho}\}^{\alpha}.
\end{aligned}
$$

So far, we have $L^{lm}\hat{\rho}=\hat{\rho}L^{lm}-[S^{lm}]\hat{\rho}$. Transposing the spin term now gives $(L^{lm}+[S^{lm}])\hat{\rho}=\hat{\rho}L^{lm}$, in agreement with Eq. (3.40).

APPENDIX A: THE CURL-CURL FORMULA

Here the derivation of the curl–curl identity (2.17) will be indicated briefly. We begin by using the two representations [Eq. (2.16) is repeated here for convenience]

$$\nabla\times=\mathbf{e}_{a}\mathbf{e}_{b}(-g)^{-1/2}\epsilon^{abc}\partial_{c} \tag{A1}$$

and

$$\nabla\times=-(-g)^{-1/2}\partial_{k}(-g)^{1/2}g^{kh}\mathbf{e}^{d}\mathbf{e}^{f}(-g)^{1/2}\epsilon_{dfh} \tag{A2}$$

of $\nabla\times$. The representation (A2) is obtained from A1) by moving the factor $(-g)^{-1/2}\partial_{c}$ to the extreme left followed by suitable raising and lowering of indices. Correction terms due to moving ∂_{c} involve the symmetric quantities $\partial\mathbf{e}_{a}/\partial u^{c}$ and $\partial\mathbf{e}_{b}/\partial u^{c}$ and vanish upon contracting with the antisymmetric ϵ^{abc}. We here adopt the convention that $\epsilon_{123}\equiv 1$. Since we had earlier defined $\epsilon^{123}\equiv 1$ and since our $g<0$ an explicit minus sign appears on raising or lowering indices. For example,

$$(-g)^{-1/2}\epsilon^{abc}g_{ad}g_{bf}g_{ck}=-(-g)^{1/2}\epsilon_{dfk}.$$

Dotting (A2) into (A1), we find

$$\nabla\times\nabla\times=(-g)^{-1/2}\partial_{k}(-g)^{1/2}g^{kh}(\mathbf{I}-\hat{s}\hat{s})\partial_{h} - (-g)^{-1/2}\partial_{k}(-g)^{1/2}\mathbf{e}^{c}\mathbf{e}^{k}\partial_{c}. \tag{A3}$$

In the next step in the reduction the factor $\mathbf{I}-\hat{s}\hat{s}$ in Eq. (A3) is moved to the extreme left. The commutator correction terms thereby encountered can be reduced by use of the equations $\partial\hat{s}/\partial u^{a}=\mathbf{e}_{a}$ and $\mathbf{e}^{a}\mathbf{e}_{a}=\mathbf{I}-\hat{s}\hat{s}$, leading to the formula

$$\nabla\times\nabla\times=(\mathbf{I}-\hat{s}\hat{s})\nabla^{2}+(\mathbf{I}-\hat{s}\hat{s})+\nabla\hat{s}+\hat{s}\nabla - (-g)^{-1/2}\partial_{k}(-g)^{1/2}\mathbf{e}^{c}\mathbf{e}^{k}\partial_{c}. \tag{A4}$$

The next step in the derivation involves the identity

$$-(-g)^{-1/2}\partial_{k}(-g)^{1/2}\mathbf{e}^{c}\mathbf{e}^{k}\partial_{c} = -\nabla\nabla-\hat{s}\nabla-2\nabla\hat{s}-2(\mathbf{I}-\hat{s}\hat{s}). \tag{A5}$$

In order to obtain Eq. (A5) the operator ∂_{k} is first moved to the right of the two factors $\mathbf{e}^{c}$ and $(-g)^{1/2}\mathbf{e}^{k}$ with commutator correction terms involving $\partial\mathbf{e}^{c}/\partial u^{k}=-\{{}^{c}_{kl}\}\mathbf{e}^{l}-\delta^{c}_{k}\hat{s}$ and $\partial((-g)^{1/2}\mathbf{e}^{k})/\partial u^{k}=-3(-g)^{1/2}\hat{s}$.

Christoffel symbols which appear at this step are cancelled by commutator correction terms from the next shift, which is a shift of the operator ∂_{c} to the left of the factor $\mathbf{e}^{k}$. When Eq. (A5) is used in conjunction with Eq. (A4), the identity (2.17) results. √

APPENDIX B: FURTHER RESULTS ON VECTOR SPHERICAL HARMONICS

An arbitrary vector field $\mathbf{F}$ in Minkowski space can be written $\mathbf{F}=\hat{s}F_{s}+\mathbf{F}^{T}$, where $\mathbf{F}^{T}$ is tangent to the unit hyperboloid. As indicated in Sec. IIIB, $\mathbf{F}^{T}$ can be expressed as a conservative plus a solenoidal part $\mathbf{F}^{T}=\nabla V+\nabla\times\mathbf{A}$, for suitable scalar and vector potentials V and $\mathbf{A}$. We thus have the representation

$$\mathbf{F}=\hat{s}F_{s}+\nabla V+\nabla\times\mathbf{A} \tag{B1}$$

of the full vector field $\mathbf{F}$. Now $\nabla\times\mathbf{A}$ can be expanded in a series of the solenoidal vector spherical harmonics $Y^{\rho}_{\alpha lm}$. Since V and F_{s} can be expanded in a series of the scalar spherical harmonics $\mathbf{Y}_{\alpha lm}$, it follows that the set of vector fields

$$\hat{s}Y_{\alpha lm},\quad \nabla Y_{\alpha lm},\quad \mathbf{Y}^{\rho}_{\alpha lm} \tag{B2}$$

is a complete set. The set (B2) is in fact a complete orthogonal set. It is readily verified by use of the intertwining relations (3.49), (3.50) that the new vector fields $\hat{s}Y_{\alpha lm}$ and $\nabla Y_{\alpha lm}$ are simultaneous eigenfunctions of the operators $[J_{x}]^{2}+[J_{y}]^{2}+[J_{z}]^{2}$ and $[J_{z}]$. On the other hand $\hat{s}Y_{\alpha lm}$ and $\nabla Y_{\alpha lm}$ are not eigenfunctions of ∇^{2}. However, linear combinations can be taken which are eigenfunctions of ∇^{2}. These linear combinations are

$$\mathbf{Y}^{0}_{\alpha lm}\equiv(1+i\alpha)\hat{s}Y_{\alpha lm}+\nabla Y_{\alpha lm} \tag{B3}$$

and

$$\mathbf{Y}^{3}_{\alpha lm}\equiv(1-i\alpha)\hat{s}Y_{\alpha lm}+\nabla Y_{\alpha lm}, \tag{B4}$$

and these are the additional vector spherical harmonics of the unit hyperboloid that we are looking for. The complex eigenvalues which appear in the eigenvalue equations

$$\nabla^{2}\mathbf{Y}^{0}_{\alpha lm}=[(\alpha-i)^{2}+1]\mathbf{Y}^{0}_{\alpha lm} \tag{B5}$$

and

$$\nabla^{2}\mathbf{Y}^{3}_{\alpha lm}=[(\alpha+i)^{2}+1]\mathbf{Y}^{3}_{\alpha lm} \tag{B6}$$

are a manifestation of the breakdown of the usual spectral theorem in a Hilbert space in which the dot product fails to obey the positive definiteness condition. Another manifestation of this breakdown is the lack of orthogonality of the vectors (B3), (B4). The lack of orthogonality is ameliorated somewhat by the existence of a convenient dual basis: If we define

$$\mathbf{Z}^0_{\alpha l m} \equiv \mathbf{Y}^3_{\alpha l m}[2\alpha(\alpha - i)]^{-1} \text{ and } \mathbf{Z}^3_{\alpha l m} \equiv \mathbf{Y}^0_{\alpha l m}[2\alpha(\alpha + i)]^{-1}, \tag{B7}$$

then we have

$$\int d\Sigma\, \mathbf{Z}^{\rho_2 *}_{\alpha_2 l_2 m_2} \cdot \mathbf{Y}^{\rho_1}_{\alpha_1 l_1 m_1} = -\delta_{\rho_2 \rho_1}\delta(\alpha_2 - \alpha_1)\delta_{l_2 l_1}\delta_{m_2 m_1}. \tag{B8}$$

We close this appendix by indicating the derivation of the new addition theorem

$$\sum_{\rho=1}^{2}\sum_{lm} \mathbf{Y}^{\rho}_{\alpha l m}(2)\mathbf{Y}^{\rho *}_{\alpha l m}(1) = \frac{1}{2\pi^2}\,\nabla_2 \times \nabla_2 \times \hat{s}_1 \hat{s}_2 \cdot (\mathbf{I} - \hat{s}_1 \hat{s}_1) \times \frac{d}{d\cosh\rho_{21}}\left(\frac{\sin\alpha\,\rho_{21}}{\alpha \sinh\rho_{21}}\right). \tag{B9}$$

This addition theorem involves the solenoidal vector spherical harmonics only. The derivation of Eq. (B9) exploits the Lorentz transformation law

$$Y^{\rho_1}_{\alpha l_1 m_1}(2')^{\mu}\hat{i}'_{\mu} = \sum_{\rho_2 l_2 m_2} \mathbf{Y}^{\rho_2}_{\alpha l_2 m_2}(2) L^{\alpha}_{\rho_2 l_2 m_2; \rho_1 l_1 m_1} \tag{B10}$$

of the solenoidal vector spherical harmonics. On the left-hand side Eq. (B10) we have the usual vector spherical harmonic, but formed entirely from quantities referring to a different Lorentz frame of reference. In order that a normalization condition like Eq. (3.47) shall hold for the new functions as well as the old, the coefficients $L^{\alpha}_{\rho_2 l_2 m_2; \rho_1 l_1 m_1}$ must form a unitary matrix. This unitarity in turn implies the Lorentz invariance of the dyadic structure on the left hand side of Eq. (B9):

$$\sum_{\rho l m} \mathbf{Y}^{\rho}_{\alpha l m}(2)\mathbf{Y}^{\rho *}_{\alpha l m}(1) \sum_{\rho l m} Y^{\rho}_{\alpha l m}(2')^{\mu}\hat{i}'_{\mu} Y^{\rho *}_{\alpha l m}(1')^{\nu}\hat{i}'_{\nu}\,. \tag{B11}$$

The addition theorem (B9) will be obtained by evaluating the right hand side of Eq. (B11) in a special Lorentz frame for which the time axis is aligned along the time-like vector $\hat{s}_1$. In this frame of reference the only non-zero values of $Y^{\rho}_{\alpha l m}(1')^{\nu}\hat{i}'_{\nu}$ are

$$Y^2_{\alpha 10}(1')^{\nu}\hat{i}'_{\nu}\big|_{\rho'=0} = \hat{i}'_3\,\frac{(1+\alpha^2)^{1/2}}{\pi(3)^{1/2}}$$

and

$$Y^2_{\alpha 1,\pm 1}(1')^{\nu}\hat{i}'_{\nu}\big|_{\rho'=0} = \frac{\mp \hat{i}'_1 - i\hat{i}'_2}{(2)^{1/2}}\,\frac{(1+\alpha^2)^{1/2}}{\pi(3)^{1/2}}\,. \tag{B12}$$

The infinite series on the right hand side of Eq. (B11) thus collapses into only three terms. To proceed, we require the explicit expression (3.44) for $\mathbf{Y}^2_{\alpha l m}$, formed using quantities referring to the primed frame of reference. Here the Lorentz invariance of the operator $\nabla\times$ comes to our aid. Because of this Lorentz invariance only the argument of $\nabla\times\nabla\times$ in the expression for $\mathbf{Y}^2_{\alpha l m}$ need be formed from quantities referring to the primed frame of reference, and we find

$$Y^2_{\alpha l m}(2')^{\mu}\hat{i}'_{\mu} = \frac{1}{\alpha(2)^{1/2}}\,\nabla_2 \times \nabla_2 \times \left(\left[\sinh\rho'_2 Y_{\alpha l m}(2')\,\frac{\partial \hat{s}_2}{\partial \rho'_2}\right). \tag{B13}$$

If we write down the now finite sum on the right-hand side of Eq. (B11), using Eqs. (B12) and (B13), and convert all expressions occurring back to a form involving only quantities referring to the original Lorentz frame, the addition theorem (B9) will be obtained. The rather lengthy calculation needed to write all expressions obtained in the finite sum in terms of quantities referring to the original Lorentz frame is outlined below. [We note that these same Lorentz invariance considerations can be used to provide an alternate derivation of the scalar spherical harmonic addition theorem (3.25) originally established by Durand, Fishbane, and Simmons by a method of analytic continuation starting with the addition theorem for the Gegenbauer polynomials.] In terms of quantities referring to the original Lorentz frame, ρ'_2 is simply $\rho'_2 = \rho_{21} = \cosh^{-1}(\hat{s}_2 \cdot \hat{s}_1)$. The derivation of the equation

$$\frac{\partial \hat{s}_2}{\partial \rho'_2} = \frac{-\hat{s}_1 + \cosh\rho_{21}\hat{s}_2}{\sinh\rho_{21}} \tag{B14}$$

requires an explicit transformation equation

$$\begin{pmatrix} x^{0\prime} \\ x^{1\prime} \\ x^{2\prime} \\ x^{3\prime} \end{pmatrix} = \begin{pmatrix} \cosh\rho_1 & -\sinh\rho_1 \sin\theta_1 \cos\varphi_1 & -\sinh\rho_1 \sin\theta_1 \sin\varphi_1 & -\sinh\rho_1 \cos\theta_1 \\ 0 & \cos\theta_1 \cos\varphi_1 & \cos\theta_1 \sin\varphi_1 & -\sin\theta_1 \\ 0 & -\sin\varphi_1 & \cos\varphi_1 & 0 \\ -\sinh\rho_1 & \cosh\rho_1 \sin\theta_1 \cos\varphi_1 & \cosh\rho_1 \sin\theta_1 \sin\varphi_1 & \cosh\rho_1 \cos\theta_1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{B15}$$

to a primed frame of reference whose time axis is in the direction $\hat{s}_1$. If in Eq. (B15) both new and old components of x_2 are written in the form (1.1), we find the relations

$$\cosh\rho'_2 = \hat{s}_1 \cdot \hat{s}_2, \quad \sinh\rho'_2 \sin\theta'_2 \cos\varphi'_2 = -\hat{\theta}_1 \cdot \hat{s}_2, \tag{B16}$$

$\sinh\rho_2' \sin\theta_2' \sin\varphi_2' = -\hat{\varphi}_1 \circ \hat{s}_2$, and $\sinh\rho_2' \cos\theta_2' = -\hat{\rho}_1 \cdot \hat{s}_2$.

We apply the derivative $\partial/\partial\rho_2'$ to both sides of Eqs. (B16), with $\partial\hat{s}_2/\partial\rho_2'$ expanded by the chain rule as

$$\partial\hat{s}_2/\partial\rho_2' = \sinh\rho_2 \hat{\theta}_2 \partial\theta_2/\partial\rho_2' + \sinh\rho_2 \sin\theta_2 \hat{\varphi}_2 \partial\varphi_2/\partial\rho_2' + \hat{\rho}_2 \partial\rho_2/\partial\rho_2'.$$

The resulting equations are arranged in the matrix form

$$\begin{pmatrix} \sinh\rho_{21} \\ \coth\rho_{21}\hat{\theta}_1 \cdot \hat{s}_2 \\ \coth\rho_{21}\hat{\varphi}_1 \cdot \hat{s}_2 \\ \coth\rho_{21}\hat{\rho}_1 \cdot \hat{s}_2 \end{pmatrix} = \begin{pmatrix} \hat{s}_1 \cdot \hat{s}_2 & \hat{s}_1 \circ \hat{\theta}_2 & \hat{s}_1 \cdot \hat{\varphi}_2 & \hat{s}_1 \cdot \hat{\rho}_2 \\ \hat{\theta}_1 \circ \hat{s}_2 & \hat{\theta}_1 \cdot \hat{\theta}_2 & \hat{\theta}_1 \circ \hat{\varphi}_2 & \hat{\theta}_1 \cdot \hat{\rho}_2 \\ \hat{\varphi}_1 \cdot \hat{s}_2 & \hat{\varphi}_1 \cdot \hat{\theta}_2 & \hat{\varphi}_1 \cdot \hat{\varphi}_2 & \hat{\varphi}_1 \circ \hat{\rho}_2 \\ \hat{\rho}_1 \circ \hat{s}_2 & \hat{\rho}_1 \cdot \hat{\theta}_2 & \hat{\rho}_1 \cdot \hat{\varphi}_2 & \hat{\rho}_1 \cdot \hat{\rho}_2 \end{pmatrix} \begin{pmatrix} 0 \\ \sinh\rho_2 \partial\theta_2/\partial\rho_2' \\ \sinh\rho_2 \sin\theta_2 \partial\varphi_2/\partial\rho_2' \\ \partial\rho_2/\partial\rho_2' \end{pmatrix} \tag{B17}$$

and solved for the important unknowns $\sinh\rho_2 \partial\theta_2/\partial\rho_2'$, $\sinh\rho_2 \sin\theta_2 \partial\varphi_2/\partial\rho_2'$, and $\partial\rho_2/\partial\rho_2'$ in terms of quantities referring to the original Lorentz frame of reference. Matrix inversion in Eq. (B17) is readily carried out, since the matrix is a Lorentz transformation matrix. Equation (B14) can now be obtained by substituting the expressions for the unknowns thus obtained in the above chain rule formula for $\partial\hat{s}_2/\partial\rho_2'$, and simplifying. When Eq. (B14) is used in Eq. (B13), we find simply

$$Y^2_{\alpha l m}(2')^{\mu} \hat{i}'_{\mu} = -\nabla_2 \times \nabla_2 \times [\hat{s}_1 Y_{\alpha l m}(2')] \alpha^{-1} 2^{-1/2}. \tag{B18}$$

To complete the derivation we need the relations

$$\sin\theta_2' e^{\pm i\varphi_2'} = -\hat{s}_2 \cdot (\hat{\theta}_1 \pm i\hat{\varphi}_1)/\sinh\rho_{21} \quad \text{and} \quad \cos\theta_2' = -\hat{s}_2 \circ \hat{\rho}_1/\sinh\rho_{21}, \tag{B19}$$

which follow from Eqs. (B16); and the relations

$$\hat{i}_1' = -\hat{i}'^1 = -L^1{}_{\alpha}\hat{i}^{\alpha} = \hat{\theta}_1, \quad \hat{i}_2' = -\hat{i}'^2 = -L^2{}_{\alpha}\hat{i}^{\alpha} = \hat{\varphi}_1, \quad \text{and} \quad \hat{i}_3' = -\hat{i}'^3 = -L^3{}_{\alpha}\hat{i}^{\alpha} = \hat{\rho}_1. \tag{B20}$$

[1]M. Dalumens and P. Minnaert, J. Math. Phys. **17**, 2085 (1976).

[2]L. Durand, P.M. Fishbane, and L.M. Simmons, Jr., J. Math. Phys. 17, 1933 (1976).

[3]See, for example, R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1975).

[4]In general we adhere to the convention that the notation ∂_a shall signify a differential operator, while $\partial/\partial u^a$ will be reserved for ordinary differentiation and will be followed by an explicit argument. In this notation for example $\partial_\theta \sin\theta = \cos\theta + \sin\theta\partial_\theta$, but $(\partial/\partial\theta)\sin\theta = \cos\theta$.

[5]The relations needed are: $\mathbf{e}_1 = \sinh\rho\hat{\theta}$, $\mathbf{e}_2 = \sinh\rho\sin\theta\hat{\varphi}$, $\mathbf{e}_3 = \hat{\rho}$; $g^{11} = -(\sinh\rho)^{-2}$, $g^{22} = -(\sinh\rho\sin\theta)^{-2}$, $g^{33} = -1$; $\mathbf{e}^1 = -\hat{\theta}(\sinh\rho)^{-1}$, $\mathbf{e}^2 = -\hat{\varphi}(\sinh\rho\sin\theta)^{-1}$, $e^3 = -\hat{\rho}$.

[6]We use the definitions of A. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University, Princeton, N.J., 1957).

[7]It follows that not only ∇^2 but also $\nabla^2 - 1$ is positive definite. More perspective on the positive definiteness of $\nabla^2 - 1$ is given in footnote Ref. 14 to Sec. IIIB.

[8]The form (3.11) is in the notation of Ref. 2, which contains a rather extensive account of the properties of Gegenbauer functions.

[9]Reference 2, Eq. (2.5).

[10]H. Bateman, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, Chap. I.

[11]Reference 10, Chap. II, Eq. (7), p. 50.

[12]Reference 10, Chap. II, Eq. (12) (second form, with $z \to i\rho/2$ and $\alpha \to 2i\alpha$), p. 101.

[13]Reference 2, Eq. (8.3).

[14]Equations (3.42) can be combined with the simplified curl–curl identity, Eq. (2.19), to give the relation $\nabla\times\nabla\times\nabla\times(\hat{\rho}\sinh\rho U) = \nabla\times(\hat{\rho}\sinh\rho(\nabla^2 - 1)U)$. By use of this relation it follows that if U is an eigenfunction of ∇^2 to eigenvalue $\lambda + 1$ for some λ, then $\nabla\times(\hat{\rho}\sinh\rho U)$ is an eigenfunction of $\nabla\times\nabla$ to eigenvalue λ. But it can be shown that the operator $\nabla\times\nabla\times$ is a positive definite operator on the Hilbert space of solenoidal tangent vector fields on the unit hyperboloid. The restriction $\lambda > 0$ and the positive definiteness of $\nabla^2 - 1$ viewed as an operator on scalar functions now follows. This result provides some added perspective on the positive definiteness of $\nabla^2 - 1$. In order to gain this perspective we have had to go beyond the scalar case to the theory of vector spherical harmonics.

Classical mechanics, the diffusion (heat) equation, and the Schrödinger equation

Aubrey Truman

Mathematics Department, Heriot-Watt University, Edinburgh, Scotland
(Received 28 February 1977)

We consider the limiting case $\lambda \to 0$ of the Cauchy problem
$\partial u_\lambda/\partial t = (\lambda/2\mu)\nabla_x^2 u_\lambda + [V(x)/\lambda]u_\lambda$,
$u_\lambda(x,0) = \exp[-S_0(x)/\lambda]T_0(x)$; S_0, T_0 independent of λ, for both real and pure imaginary λ. We prove two new theorems relating the limiting solution of the above Cauchy problem to the corresponding equations of classical mechanics
$\mu(d^2x/d\tau^2)(\tau) = -\nabla_x V[x(\tau)]$, $\tau \in (0,t)$.
These relationships include the physical result quantum mechanics → classical mechanics as $\hbar \to 0$.

1. INTRODUCTION

We consider the limiting case $\lambda \to 0$ of the Cauchy problem for the equation

$$\frac{\partial u}{\partial t} = \frac{\lambda}{2\mu}\nabla_x^2 u + \frac{V(x)}{\lambda}u, \tag{1}$$

where ∇^2 is the n-dimensional Laplacian, V is a real-valued potential, $V \in L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)$, and μ is a positive constant. Here $u = u_\lambda(x,t)$ is the solution of the above equation defined on $\mathbb{R}^n \times [0,\infty)$ with Cauchy data $u_\lambda(x,0) = \exp[-S_0(x)/\lambda]T_0(x)$, S_0 and T_0 being independent of λ.

When λ is real, the above equation is a diffusion (heat) equation with potential V. This equation arises in stochastic mechanics with $\lambda = \hbar$, Planck's constant divided by 2π, for stationary states u. When λ is pure imaginary, $\lambda = i\hbar$, the equation above is the Schrödinger equation for the wavefunction u of a particle of mass μ (or suitably transformed for several particles) moving in the force field $(-\nabla V)$ in the n-dimensional Euclidean space $\mathbb{R}^n$.

Here we derive a relationship between the limiting solution, as $\lambda \to 0$, of Eq. (1), in the case of both real and pure imaginary λ, and the solution to the corresponding equations of classical mechanics

$$\mu\frac{d^2X(x,\tau)}{d\tau^2} = -\nabla_x V[X(x,\tau)], \tag{2}$$

$X(x,\tau) \in \mathbb{R}^n$, $\tau \in (0,t)$, with boundary conditions $X(x,t) = x$, $\mu\dot{X}(x,0) = \nabla S_0[x_0(x,t)]$, where $x_0(x,t)$ is defined below.

The study undertaken here was motivated in part by our previous investigations[1] of Feynman's path integral.[2] We were concerned too with the problem of obtaining, in as simple a manner as possible, the limiting case of quantum mechanics when $\hbar$ tends to zero. This problem lies at the very foundation of quantum mechanics and has previously been investigated by many authors.[3] In this paper we present some new results in this context—Theorems 1 and 2. We believe that these two theorems give a better understanding of how classical mechanics is achieved as the limiting case of quantum mechanics when $\hbar$ tends to zero. We also feel that our results are less formal and easier to derive than those of previous authors.

Theorem 1 is a new result for the diffusion (heat) equation and probably has applications to stochastic mechanics.[4] Here, however, our interest in Theorem 1 centers on the method of proof which uses functional integration (the Feynman–Kac formula) in an intrinsic way and the fact that it gives an explicit form for the limiting solution to Eq. (1), which is valid for both real and pure imaginary λ. Similar results to Theorem 2 have previously been obtained by Maslov[5] but our proof is much shorter than Maslov's. In spite of the strong similarities between Theorems 1 and 2 we have been unable to obtain both results by a common method.

We need to make some modest assumptions about the potential V. In the diffusion (heat) equation case we assume that in addition to V being real, $V \in L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)$, V has continuous second order partial derivatives $(\partial^2 V/\partial x_i \partial x_j)(x)$, $i,j = 1,2,\ldots,n$, at all points $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and

$$\sup_{i,j,x}\left|\frac{\partial^2 V}{\partial x_i \partial x_j}(x)\right| = K < \infty.$$

When these last two conditions on V are satisfied we write $V \in C^{2\prime}$. We assume S_0 is real, and again in the diffusion (heat) equation case $S_0 \in C^{2\prime}$.

The above assumptions imply that Eq. (2) has a unique local solution $x[x_0, p_0, \tau] \in \mathbb{R}^n$,

$$\mu\frac{d^2x}{d\tau^2} = -\nabla_x V[x], \tag{3}$$

$\tau \in [0, T']$, satisfying $x[x_0, p_0, 0] = x_0 \in \mathbb{R}^n$, $\mu\dot{x}[x_0, p_0, 0] = p_0 \in \mathbb{R}^n$. We also need to assume that the equation $x[x_0, \nabla S_0(x_0), t] = x$ can be solved uniquely to yield $x_0 = x_0(x,t)$, $t \in [0, T'']$. The required classical solution $X(x,\tau)$ is then defined by

$$X(x,\tau) = x[x_0(x,t), \nabla S_0[x_0(x,t)], \tau], \tag{4}$$

$\tau \in (0,t)$, $t \in (0,T)$, where $T \le T'$, $T \le T''$.

We shall say that the trajectory $X(x,\tau)$ is well behaved if it does not pass through any focus of the classical problem so that $|\partial X^i(x,\tau)/\partial x_j^0| \neq 0$, $\tau \in (0,t)$ and if, in addition, $X(x,\tau)$ is such that $V[X(x,\tau)]$ has continuous third order partial derivative with respect to space variables for $\tau \in (0,t)$. For the diffusion (heat) equation case we must assume $X(x,\tau)$ is well behaved for $\tau \in (0,t)$, $t \in (0,T)$, where $T \le T'$, $T \le T''$. In this case it is necessary to further restrict T. We denote by $\lambda(V)$ the maximum eigenvalue of the matrix $\|[\partial^2 V(x)/\partial x_i \partial x_j]\|$. Putting $\lambda_+ = \sup_x\{\max[\lambda(V), 0]\}$, it follows that $0 \le \lambda_+ < \infty$. Simi-

larly denoting by $\lambda(-S_0)$ the maximum eigenvalue of the matrix $\|-[\partial^2 S_0(x)/\partial x_i \partial x_j]\|$ and $\lambda_- = \sup_x\{\max[\lambda(-S_0), 0]\}$ we obtain $0 \leq \lambda_- < \infty$. We define the time $T''' \leq \infty$, by $T''' = (\mu/\lambda_+)^{1/2} \tan^{-1}[(\mu\lambda_+)^{1/2}/\lambda_-]$. For the diffusion (heat) equation we put $T = \min[T', T'', T''']$.

We are now ready to state our main result for the diffusion (heat) equation.

Theorem 1: Let $S(x, t)$ be the solution of the Hamilton–Jacobi equation

$$\frac{|\nabla_x S(x,t)|^2}{2\mu} + V(x) + \frac{\partial S(x,t)}{\partial t} = 0, \quad t \in (0, T), \tag{5}$$

with $S(x, 0) = S_0(x) \in C^{2\prime}$, so that

$$S(x,t) = S_0[x_0(x,t)] + \int_0^t \left[\frac{\mu}{2}\sum_{j=1}^{n}\left(\frac{dX_j}{d\tau}\right)^2 - V[X]\right] d\tau, \tag{6}$$

where $X = X(x, \tau) = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^n$, is well behaved.

Let $u_\lambda(x, t)$ be the solution of the diffusion (heat) equation

$$\frac{\partial u_\lambda}{\partial t} = \frac{\lambda}{2\mu}\nabla_x^2 u_\lambda + \frac{V(x)}{\lambda} u_\lambda, \tag{7}$$

with Cauchy data $u_\lambda(x, 0) = \exp[-S_0(x)/\lambda] T_0(x) \in L^2(\mathbb{R}^n)$, where T_0 is bounded and continuous $V \in C^{2\prime} \cap [L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)]$.

Then, for each fixed $t \in (0, T)$,

$$\exp[S(x,t)/\lambda] u_\lambda(x,t) - J_t^{1/2}(x) T_0[x_0(x,t)]$$

pointwise, as $\lambda \to 0$, where $J_t(x) = |\partial x_0^i/\partial x_j|$, the Jacobian of the transformation $x \to x_0(x, t) = (x_0^1, x_0^2, \ldots, x_0^n) \in \mathbb{R}^n$.

The last result shows that the diffusion process determined by the above Cauchy problem, for small λ, follows the lines of the classical flow. Hence, when $u_\lambda(x, 0)$ has support contained in a small neighborhood of the point x_0, the support of $u_\lambda(x, t)$, for small λ, is concentrated around the classical flow $x[x_0, \nabla S_0(x_0), t]$.

To deduce the results corresponding to Theorem 1 for the Schrödinger equation we slightly change our assumptions regarding the classical flow. We use the above notation except that $T = \min[T', T'']$. The time T is the maximum time for which the above uniqueness and existence theorems obtain for the classical problem. These uniqueness and existence results are still required for the Schrödinger equation, but we do not require explicitly $V \in C^{2\prime}$ and $S_0 \in C^{2\prime}$.

For each fixed $t \in (0, T)$, we assume that $D_t : x_0 \to x[x_0, \nabla S_0(x_0), t]$ is a C^1 diffeomorphism $D_t : \mathbb{R}^n \to \mathbb{R}^n$ with C^1 inverse $D_t^{-1} : x \to x_0(x, t)$, $D_t^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. Denote by $J_t(x)$ or $J_t(x_0)$ the Jacobian of D_t^{-1}, $J_t(x_0) = J_t(x) = |\partial x_0^i/\partial x^j|$, where it is understood that, if $J_t(x_0)$ is required, we put $x = x[x_0, \nabla S_0(x_0), t]$. Then we assume that $S = \{x \in \mathbb{R}^n \mid J_t(x) = 0\}$ has Lebesgue measure zero. We then have the theorem.

Theorem 2: Let $\psi_\hbar(x, t)$ be the solution of the Schrödinger equation

$$\frac{\partial \psi_\hbar}{\partial t} = \frac{i\hbar}{2\mu}\nabla_x^2 \psi_\hbar + \frac{V(x)}{i\hbar}\psi_\hbar, \tag{8}$$

with Cauchy data $\psi_\hbar(x, 0) = \exp\{iS_0(x)/\hbar\}\phi_0(x) \in L^2(\mathbb{R}^n)$, where $V \in \{L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)\}$ is real-valued and ϕ_0 (independent of $\hbar$) is such that $F(\tau) = \|\nabla_x^2 J_\tau^{1/2}(x)\phi_0[x_0(x,\tau)]\|_{L_2} \in L^1(0, t)$, $\|\ \|_{L_2}$ being the L^2 norm with respect to x.

Then, for each fixed $t \in (0, T)$,

$$\exp[-iS(x,t)/\hbar]\psi_\hbar(x,t) \to J_t^{1/2}(x)\phi_0[x_0(x,t)],$$

in the L^2 norm with respect to x, as $\hbar \to 0$.

We shall show in the last section of this paper that Theorem 2 establishes that for any finite number of spinless nonrelativistic particles with potential interaction, quantum mechanics tends to classical mechanics as $\hbar$ tends to zero.

Theorem 1 is proved by establishing a quasiclassical representation for the diffusion equation similar to that in Ref. 1. The quasiclassical representation in Ref. 1(a) was derived from the Feynman–Ito formula for the Schrödinger equation using the translational properties of the Feynman path integral $\mathcal{J}$. In this case a quasiclassical representation is obtained for the diffusion equation by exploiting the Feynman–Kac formula and the translational properties of the Wiener integral E. The quasiclassical representation leads to a certain Wiener integral which is evaluated using the Jacobi fields of Ref. 2. This is carried out in Secs. 2 and 3.

In Sec. 4 the explicit form of Theorem 1 leads us to change the dependent variable in the Schrödinger equation to a new variable ϕ. For the new variable ϕ we show there is an isometric evolution operator $\tilde{U}$, whose time-dependent generator A is linearly related to $\hbar$. The linear relationship between A and $\hbar$ enables us to prove Theorem 2 in Sec. 4.

2. QUASICLASSICAL REPRESENTATION FOR THE DIFFUSION EQUATION

We derive the quasiclassical representation in the next lemma. We use the notation of the previous section.

Lemma 1: Let $\Delta^2 V[X, (2\lambda/\mu)^{1/2} Y]$ be defined by

$$\begin{aligned}\Delta^2 V\left[X, \left(\frac{2\lambda}{\mu}\right)^{1/2} Y\right] &= V\left[X(x, t-\tau) + \left(\frac{2\lambda}{\mu}\right)^{1/2} Y(\tau)\right] - V[X(x, t-\tau)] \\ &\quad - \left(\frac{2\lambda}{\mu}\right)^{1/2} Y(\tau)\cdot\nabla V[X(x, t-\tau)],\end{aligned} \tag{9}$$

where $Y(\tau) = (Y_1(\tau), Y_2(\tau), \ldots, Y_n(\tau)) \in C_0(0, t)^{\otimes n}$, so that $Y_j(\tau)$ is continuous on $(0, t)$, $\lim_{\tau \to 0^+} Y_j(\tau) = 0$, $j = 1, 2, \ldots, n$ and here $X = X(x, t-\tau)$. Denoting $X_0 = X(x, 0) = x_0(x, t)$ and $Y_t = Y(t)$, similarly let

$$\begin{aligned}\Delta^2 S_0\left[X_0, \left(\frac{2\lambda}{\mu}\right)^{1/2} Y_t\right] &= S_0\left[X(x,0) + \left(\frac{2\lambda}{\mu}\right)^{1/2} Y(t)\right] - S_0[X(x,0)] \\ &\quad - \left(\frac{2\lambda}{\mu}\right)^{1/2} Y(t)\cdot\nabla S_0[X(x,0)].\end{aligned} \tag{10}$$

Then, for $t < T$,

$$\exp[S(x,t)/\lambda]u_\lambda(x,t) = E\left[\exp\left\{\frac{1}{\lambda}\int_0^t \Delta^2 V\left[X,\left(\frac{2\lambda}{\mu}\right)^{1/2}Y\right]d\tau\right\} \times\exp\left\{\frac{-1}{\lambda}\Delta^2 S_0\left[X_0,\left(\frac{2\lambda}{\mu}\right)^{1/2}Y_t\right]\right\}T_0\left[X_0+\left(\frac{2\lambda}{\mu}\right)^{1/2}Y_t\right]\right], \tag{11}$$

the Wiener integral being taken with respect to the variables $Y(\tau)$.

Proof: Put $u_\lambda(x,t)=v_\lambda(X,t)$, where v_λ is the solution of the Cauchy problem

$$\frac{\partial v_\lambda}{\partial t} = \tfrac{1}{4}\nabla_X^2 v_\lambda + \lambda^{-1}V\left[\left(\frac{2\lambda}{\mu}\right)^{1/2}X\right]v_\lambda, \tag{12}$$

$v_\lambda(X,0)=u_\lambda[(2\lambda/\mu)^{1/2}X,0]$, X being given by $X=(\mu/2\lambda)^{1/2}x$.

Then, because $V(\cdot)\in L^2(\mathbb{R}^n)+L^\infty(\mathbb{R}^n)$ and $v_\lambda(\cdot,0)\in L^2(\mathbb{R}^n)$, the Trotter product formula implies

$$v_\lambda(X,t) = E\left[\exp\left\{\frac{1}{\lambda}\int_0^t V\left[\left(\frac{2\lambda}{\mu}\right)^{1/2}(X+X(\tau))\right]d\tau\right\} \times v_\lambda(X+X(t),0)\right]. \tag{13}$$

This is the Feynman–Kac formula.[6]

Thus, we obtain

$$u_\lambda(x,t) = E\left[\exp\left\{\frac{1}{\lambda}\int_0^t V\left[x+\left(\frac{2\lambda}{\mu}\right)^{1/2}X(\tau)\right]d\tau\right\} \times u_\lambda\left(x+\left(\frac{2\lambda}{\mu}\right)^{1/2}X(t),0\right)\right]. \tag{14}$$

We now make a parallel translation in the Wiener measure argument. According to Koval'chik,[7] if $X\to Y+a$, $a(\tau)\in C_0(0,t)^{\otimes n}$, $\dot a(\tau)\in L_2(0,t)^{\otimes n}$, for continuous bounded functionals F,

$$E[F]=\int F(X)\,dw(X) = \exp\left[-\int_0^t\sum_{j=1}^n\left(\frac{da_j}{d\tau}\right)^2 d\tau\right] \times E\left[F(Y+a)\exp\left(-2\int_0^t\sum_{j=1}^n\frac{da_j}{d\tau}dY_j(\tau)\right)\right], \tag{15}$$

where on rhs E is taken with respect to the variables $Y(\tau)$.

Putting $a=(\mu/2\lambda)^{1/2}X_{c1}$ where $X_{c1}=(X_1,X_2,\ldots,X_n)$ is as yet unspecified, we obtain

$$\exp\left[\frac{\mu}{2\lambda}\int_0^t\sum_{j=1}^n\left(\frac{dX_j}{d\tau}\right)^2 d\tau\right]u_\lambda(x,t) = E\left[\exp\left\{\frac{1}{\lambda}\int_0^t V\left[x+X_{c1}(\tau)+\left(\frac{2\lambda}{\mu}\right)^{1/2}Y(\tau)\right]d\tau - 2\left(\frac{\mu}{2\lambda}\right)^{1/2}\int_0^t\sum_{j=1}^n\frac{dX_j}{d\tau}dY_j(\tau)\right\} \times u_\lambda\left(x+X_{c1}(t)+\left(\frac{2\lambda}{\mu}\right)^{1/2}Y(t),0\right)\right]. \tag{16}$$

We now choose $x+X_{c1}(\tau)=X(x,t-\tau)=x[x_0(x,t), \nabla S_0[x_0(x,t)],t-\tau]$. Partial integration of the Stieltjes integral[8] gives

$$\int_0^t\sum_{j=1}^n\frac{dX_j}{d\tau}dY_j(\tau) = \left[\sum_{j=1}^n\frac{dX_j}{d\tau}(x,t-\tau)Y_j(\tau)\right]_0^t - \int_0^t\sum_{j=1}^n\frac{d^2X_j}{d\tau^2}(x,t-\tau)Y_j(\tau)d\tau = \frac{-Y(t)}{\mu}\cdot\nabla S_0[x_0(x,t)] + \int_0^t\frac{Y(\tau)}{\mu}\cdot\nabla V[X(x,t-\tau)]\,d\tau. \tag{17}$$

Substitution of this result and $u_\lambda(x,0)=\exp[-S_0(x)/\lambda]\times T_0(x)$ into the above equation gives the quasiclassical representation.

The above lemma has a simple corollary.

Corollary 1: With the above notation, for $t<T$,

$$\exp[S(x,t)/\lambda]u_\lambda(x,t) \to E\left[\exp\left(\frac{1}{\mu}\int_0^t\sum_{i,j=1}^n\frac{\partial^2 V}{\partial X_i\partial X_j}[X(x,t-\tau)]Y_i(\tau)Y_j(\tau)d\tau\right) \times\exp\left(\frac{-1}{\mu}\sum_{i,j=1}^n\frac{\partial^2 S_0}{\partial x_i\partial x_j}[x_0(x,t)]Y_i(t)Y_j(t)\right)\right] \times T_0[x_0(x,t)], \tag{18}$$

as $\lambda\to 0_+$, the Wiener integral being with respect to the variables $Y_i(\tau)$.

Proof: Consider the functional $F(\lambda,Y)$,

$$F(\lambda,Y) = \exp\left\{\frac{1}{\lambda}\int_0^t\Delta^2 V\left[X,\left(\frac{2\lambda}{\mu}\right)^{1/2}Y\right]d\tau\right\} \times\exp\left\{\frac{-1}{\lambda}\Delta^2 S_0\left[X_0,\left(\frac{2\lambda}{\mu}\right)^{1/2}Y_t\right]\right\}T_0\left[X_0+\left(\frac{2\lambda}{\mu}\right)^{1/2}Y_t\right]. \tag{19}$$

Then $V, S_0\in C^{2'}$ and T_0 continuous imply that, as $\lambda\to 0_+$,

$$F(\lambda,Y)\to\exp\left\{\frac{1}{\mu}\int_0^t\sum_{i,j}\frac{\partial^2 V}{\partial X_i\partial X_j}[X(x,t-\tau)]Y_i(\tau)Y_j(\tau)\,d\tau\right\} \times\exp\left\{\frac{-1}{\mu}\sum_{i,j}\frac{\partial^2 S_0}{\partial x_i\partial x_j}[x_0(x,t)] \times Y_i(t)Y_j(t)\right\}T_0[x_0(x,t)], \tag{20}$$

a.e. with respect to Wiener measure.

Also, because $V, S_0\in C^{2'}$ and T_0 is bounded, $\exists\, M$ such that

$$|F(\lambda,Y)|\le M\exp\left\{\frac{\lambda_+}{\mu}\int_0^t\sum_{j=1}^n Y_j(\tau)Y_j(\tau)\,d\tau + \frac{\lambda_-}{\mu}\sum_{j=1}^n Y_j(t)Y_j(t)\right\}=F(Y)M. \tag{21}$$

It is not difficult to show that

$$E[F]=[D(0)]^{-n/2}\quad[D(0)\neq 0], \tag{22}$$

where $D(\tau)\in C(0,t)$ is the unique solution of

$$\ddot D(\tau)+\frac{\lambda_+}{\mu}D(\tau)=0, \tag{23}$$

$D(t)=1$, $\dot D(t)=\lambda_-/\mu$.[9] Thus, for $t<T<T'''$, $E[F]<\infty$. Applying the dominated convergence theorem for Wiener measure in the quasiclassical representation proves the corollary.

We show how to evaluate the Wiener integral on the rhs of Eq. (18) in the next section.

3. EVALUATION OF THE WIENER INTEGRAL

The key to the evaluation of the Wiener integral is contained in the next lemma.

Lemma 2: Let $p(\tau)$ be a real $(n\times n)$ symmetric matrix continuously differentiable on $(0,t)$ and let q be a real $(n\times n)$ symmetric matrix. Then, if the equation

$$\frac{d^2K(\tau)}{d\tau^2}+\lambda p(\tau)K(\tau)=0 \tag{24}$$

admits a unique nonsingular $(n\times n)$ matrix solution $K(\tau)$, $\tau\in(0,t)$, with $K(\tau=t)=1$ and $(dK/d\tau)(\tau=t)=q$,

$$E\left[\exp\lambda\left\{\int_0^t\sum_{i,j=1}^n x_i(\tau)p_{ij}(\tau)x_j(\tau)\,d\tau\right\}\times\exp\left\{\sum_{i,j=1}^n x_i(t)\,q_{ij}x_j(t)\right\}\right]=|D|^{-1}, \tag{25}$$

where the Wiener integral is with respect to the variables $x_i(\tau)$ and $|D|$ is the Fredholm determinant of the Volterra transformation

$$x(\tau)\to y(\tau)=x(\tau)-\int_0^\tau\frac{dK(\sigma)}{d\sigma}K^{-1}(\sigma)x(\sigma)\,d\sigma. \tag{26}$$

Proof: Let F be some bounded continuous functional $F: C_0(0,t)\otimes^n\to\mathbb{C}$. We consider the Wiener integral of F

$$E[F]=\int F(y)\,dw(y), \tag{27}$$

when we make the transformation $y=x+Ax$, A being the linear transformation

$$(Ax)(\tau)=-\int_0^\tau\frac{dK(\sigma)}{d\sigma}K^{-1}(\sigma)x(\sigma)\,d\sigma, \tag{28}$$

for the nonsingular $K(\sigma)$ defined above. This transformation is a bijection, mapping $C_0(0,t)\otimes^n\to C_0(0,t)\otimes^n$, with inverse

$$x(\tau)=y(\tau)-K(\tau)\int_0^\tau\frac{dK^{-1}(\sigma)y(\sigma)\,d\sigma}{d\sigma}.$$

Let $L_0^{2,1}$ be the Sobolev space of continuous functions $x_j(\tau)$ on $(0,t)$ with $x_j(0)=0$ and with weak derivative $dx_j/d\tau\in L^2(0,t)$. Then, if $A\upharpoonright L_0^{2,1}\otimes^n$ is nuclear,

$$E[F]=|D|\,E\left[F(x+Ax)\times\exp\left(-\int_0^t\sum_{i=1}^n\frac{d}{d\tau}(Ax)_i(\tau)\frac{d}{d\tau}(Ax)_i(\tau)\,d\tau\right)\times\exp\left(-2\int_0^t\sum_{i=1}^n\frac{d}{d\tau}(Ax)_i(\tau)\,dx_i(\tau)\right)\right], \tag{29}$$

$|D|$ being the Fredholm determinant above.[10] However, $A\upharpoonright L_0^{2,1}\otimes^n$ is nuclear if $k(\sigma)=-(dK/d\sigma)(\sigma)K^{-1}(\sigma)$ is two times continuously differentiable, which is the case if $p(\sigma)$ is one times continuously differentiable. Assume for the moment that $k(\sigma)$ is symmetric as well. Then, observing that

$$\frac{d}{d\tau}(Ax)_i(\tau)=\sum_{j=1}^n k_{ij}(\tau)x_j(\tau)$$

and

$$2\int_0^t\sum_{i,j=1}^n k_{ij}(\tau)x_j(\tau)\,dx_i(\tau)=\int_0^t d\sum_{i,j=1}^n x_i(\tau)k_{ij}(\tau)x_j(\tau)-\int_0^t\sum_{i,j=1}^n x_i(\tau)\frac{dk_{ij}}{d\tau}(\tau)x_j(\tau)\,d\tau, \tag{30}$$

we obtain

$$E[F]=|D|\,E\left[F(x+Ax)\exp\left\{\int_0^t\left[\sum_{i,j=1}^n\frac{dk}{d\tau}(\tau)-k^2(\tau)\right]_{ij}\times x_i(\tau)x_j(\tau)\,d\tau\right\}\exp\left[-\sum_{i,j=1}^n x_i(t)k_{ij}(t)x_j(t)\right]\right]. \tag{31}$$

However, $K(\tau)K^{-1}(\tau)=1$ implies $K(\tau)[dK^{-1}(\tau)/d\tau]+[dK(\tau)/d\tau]K^{-1}(\tau)=0$, so that

$$\frac{dK^{-1}(\tau)}{d\tau}+K^{-1}(\tau)\frac{dK(\tau)}{d\tau}K^{-1}(\tau)=0 \tag{32}$$

and

$$K(\tau)\frac{d^2K^{-1}(\tau)}{d\tau^2}+2\frac{dK(\tau)}{d\tau}\frac{dK^{-1}(\tau)}{d\tau}+\frac{d^2K(\tau)}{d\tau^2}K^{-1}(\tau)=0. \tag{33}$$

Thus,

$$k^2(\tau)-\frac{dk(\tau)}{d\tau}=K(\tau)\frac{dK^{-1}(\tau)}{d\tau}K(\tau)\frac{dK^{-1}(\tau)}{d\tau}-\frac{dK(\tau)}{d\tau}\frac{dK^{-1}(\tau)}{d\tau}-K(\tau)\frac{d^2K^{-1}(\tau)}{d\tau^2}, \tag{34}$$

or from above

$$\begin{aligned}k^2(\tau)-\frac{dk(\tau)}{d\tau}&=\frac{dK(\tau)}{d\tau}K^{-1}(\tau)\frac{dK(\tau)}{d\tau}K^{-1}(\tau)+\frac{d^2K(\tau)}{d\tau^2}K^{-1}(\tau)+\frac{dK(\tau)}{d\tau}\frac{dK^{-1}(\tau)}{d\tau}\\&=\frac{d^2K(\tau)}{d\tau^2}K^{-1}(\tau)=-\lambda p(\tau),\end{aligned} \tag{35}$$

where we have combined first and third terms to obtain penultimate equality. Putting $F\equiv 1$ the result now follows from Eq. (31).

It remains to prove that $k(\tau)$ is symmetric. We observe that $k^T(\tau)$, the transpose of $k(\tau)$, satisfies

$$[k^T(\tau)]^2-\frac{dk^T(\tau)}{d\tau}=-\lambda p(\tau). \tag{36}$$

Define $K_{1q}(\tau)$, $\tau\in(0,t)$, by $(dK_{1q}(\tau)/d\tau)=-k^T(\tau)K_{1q}(\tau)$, $K_{1q}(\tau=t)=1$, $(dK_{1q}(\tau=t)/d\tau)=q$. Then

$$\begin{aligned}\frac{d^2K_{1q}(\tau)}{d\tau^2}+\lambda p(\tau)K_{1q}(\tau)&=-\frac{dk^T(\tau)}{d\tau}K_{1q}(\tau)-k^T(\tau)\frac{dK_{1q}(\tau)}{d\tau}+\lambda p(\tau)K_{1q}(\tau)\\&=[k^T(\tau)]^2K_{1q}(\tau)-k^T(\tau)\frac{dK_{1q}(\tau)}{d\tau}+\lambda p(\tau)K_{1q}(\tau).\end{aligned} \tag{37}$$

Hence,

$$\frac{d^2K_{1q}(\tau)}{d\tau^2}+\lambda p(\tau)K_{1q}(\tau)=\{-[k^T(\tau)]^2-\lambda p(\tau)\}K_{1q}(\tau)+[k^T(\tau)]^2K_{1q}(\tau)+\lambda p(\tau)K_{1q}(\tau)=0. \tag{38}$$

By the assumed uniqueness of $K(\tau)$ we have $K_{1q}(\tau)=K(\tau)\Rightarrow k^T(\tau)=k(\tau)$. This proves the lemma.

Lemma 2 has the following corollary.

Corollary 2: For $t < T$ and well behaved $X(x, \tau)$, S_0, $V \in C^{2\prime}$,

$$E\left[\exp\left(\frac{1}{\mu}\int_0^t \sum_{i,j=1}^n \frac{\partial^2 V}{\partial x_i \partial x_j}[X(x, t-\tau)]X_i(\tau)X_j(\tau)\,d\tau\right)\right.$$
$$\left.\times\exp\left(\sum_{i,j=1}^n -\frac{1}{\mu}\frac{\partial^2 S_0}{\partial x_i \partial x_j}[x_0(x, t)]X_i(t)X_j(t)\right)\right]$$
$$=\left[\frac{\partial(x_0^1, x_0^2, \ldots, x_0^n)}{\partial(x_1, x_2, \ldots, x_n)}\right]^{1/2}, \tag{39}$$

the Wiener integral being with respect to the variables $X_i(\tau)$.

Proof: As a preliminary we prove that $|\mathcal{D}|$, the Fredholm determinant of the Volterra transformation in the above lemma, is given by

$$|\mathcal{D}| = \left(\frac{\det K(0)}{\det K(t)}\right)^{1/2}. \tag{40}$$

For the Volterra transformation $x(\tau) \to x(\tau) + \int_0^\tau k(\tau, \sigma) \times x(\sigma)\,d\sigma$, a simple calculation implies that

$$|\mathcal{D}| = \exp\left(\tfrac{1}{2}\int_0^t \mathrm{tr}[k(s, s)]\,ds\right), \tag{41}$$

where $k(s, s) = \lim_{\sigma \to s-} k(s, \sigma)$. In the particular case $k(\tau, \sigma) = K(\sigma)[dK^{-1}(\sigma)/d\sigma]\,\theta(\tau - \sigma)$, θ being the Heaviside function, we obtain

$$\tfrac{1}{2}\int_0^t \mathrm{tr}[k(s, s)]\,ds = \tfrac{1}{2}\int_0^t \sum_{i,j=1}^n K_{ij}(s)\frac{dK_{ji}^{-1}(s)}{ds}\,ds$$
$$= -\tfrac{1}{2}\int_0^t \frac{d\ln[\det K(s)]}{ds}\,ds = -\tfrac{1}{2}\ln\left(\frac{\det K(t)}{\det K(0)}\right), \tag{42}$$

where we have used the well-known identity[11]

$$\sum_{i,j=1}^n K_{ij}(s)\frac{dK_{ji}^{-1}(s)}{ds} = -\frac{d}{ds}\ln[\det K]. \tag{43}$$

This proves Eq. (40).

Hence, from the previous lemma, for symmetric p and q,

$$E\left[\exp\left\{\lambda\int_0^t \sum_{i,j=1}^n x_i(\tau)p_{ij}(\tau)x_j(\tau)\,d\tau\right\}\right.$$
$$\left.\times\exp\left\{\sum_{i,j=1}^n x_i(t)q_{ij}x_j(t)\right\}\right] = \left(\frac{\det K(t)}{\det K(0)}\right)^{1/2}, \tag{44}$$

where K is the unique nonsingular matrix solution of

$$\frac{d^2K(\tau)}{d\tau^2} + \lambda p(\tau)K(\tau) = 0, \quad \tau \in (0, t),$$

$$K(t) = 1, \quad \frac{dK}{d\tau}(\tau = t) = q,$$

$p(\tau)$ being continuously differentiable on $(0, t)$.

Putting

$$p_{ij}(\tau) = \frac{\partial^2 V}{\partial x_i \partial x_j}[X(x, t-\tau)], \quad q_{ij} = \frac{-1}{\mu}\frac{\partial^2 S}{\partial x_i \partial x_j}[x_0(x, t)]$$

and $\lambda = \mu^{-1}$, we obtain

$$\mu\frac{d^2K_{ij}(\tau)}{d\tau^2} + \sum_{j'=1}^n \frac{\partial^2 V}{\partial x_i \partial x_{j'}}[X(x, t-\tau)]K_{j'j}(\tau) = 0, \tag{45}$$

$\tau \in (0, t)$, with $K_{ij}(t) = \delta_{ij}$, $(dK_{ij}/d\tau)(\tau = t) = (-1/\mu) \times(\partial^2 S/\partial x_i \partial x_j)[x_0(x, t)]$. (In this case p is one time continuously differentiable because X is well-behaved and does not pass through a singularity of $\partial^2 V/\partial x_i \partial x_j$). However, $V \in C^{2\prime}$ implies global uniqueness and existence of the above K.[12] We now find K and show that it is nonsingular for well-behaved trajectories X. Partially differentiating the classical equations of motion

$$\mu\frac{d^2x_i}{d\tau^2}[x_0, \nabla S_0(x_0), t-\tau] = \frac{-\partial V}{\partial x_i}[x, [x_0, \nabla S_0(x_0), t-\tau]] \tag{46}$$

with respect to x_0^j and putting $x_0 = x_0(x, t)$, gives a solution $K_{ij}(\tau) = (\partial x_i/\partial x_0^j)[x_0(x, t), \nabla S_0(x_0(x, t)), t-\tau]$ of the above equation. This solution $\|K_{ij}(\tau)\|$ is nonsingular $\tau \in (0, t)$, because X does not pass through a focus of classical problem. It is simple to check that $K_{ij}(\tau)$ satisfies the correct boundary conditions.

Hence, we arrive at the result

$$E\left[\exp\left(\frac{1}{\mu}\int_0^t \sum_{i,j=1}^n \frac{\partial^2 V}{\partial x_i \partial x_j}[X(x, t-\tau)]x_i(\tau)x_j(\tau)\,d\tau\right)\right]$$
$$\times\exp\left(\frac{-1}{\mu}\sum_{i,j=1}^n \frac{\partial^2 S_0}{\partial x_i \partial x_j}[x_0(x, t)]x_i(t)x_j(t)\right]$$
$$= \left|\frac{\partial x_i}{\partial x_0^j}[x_0, \nabla S_0(x_0), t]\right|^{-1/2}_{x_0 = x_0(x,t)}, \tag{47}$$

the Wiener integral being with respect to the variables $x_i(\tau)$. Recalling that $x[x_0(x, t), \nabla S_0[x_0(x, t)], t] = x$, the corollary follows from the implicit function theorem. This concludes the proof of Theorem 1.

We have seen that the solution of the Cauchy problem for Eq. (1), for real λ, with initial data $u_\lambda(x, 0) = \exp[-S_0(x)/\lambda]T_0(x)$, S_0, T_0 independent of λ, is such that $u_\lambda(x, t) \sim \exp[-S(x, t)/\lambda]J_t^{1/2}(x)T_0[x_0(x, t)]$, as $\lambda \sim 0$. Thus, if we consider the case where T_0 has its support concentrated in the neighborhood of some point, $N(x_0)$, we see that, for small λ, the support of $u_\lambda(x, t)$ is contained in a neighborhood of the points $x = x[y_0, \nabla S_0(y_0), t]$, $y_0 \in N(x_0)$. Thus, we can think of the diffusion or heat flow governed by (1) following these lines of classical flow.[13] This result probably has applications to stochastic mechanics, but we do not pursue these here. In the next section we show how the last result can be generalized to the Schrödinger equation arising when λ is pure imaginary. Here the physical import of this result is that classical mechanics can be viewed as the limiting case of quantum mechanics when $\hbar$ tends to zero.

4. THE SCHRÖDINGER EQUATION

We consider the Schrödinger equation

$$\frac{\partial \psi_\hbar}{\partial t} = \frac{i\hbar}{2\mu}\nabla_x^2\psi_\hbar + \frac{V(x)}{i\hbar}\psi_\hbar \tag{48}$$

with Cauchy data $\psi_\hbar(x, 0) = \exp[iS_0(x)/\hbar]\phi_0(x) \in L^2(\mathbb{R}^n)$, where S_0 and ϕ_0 are independent of $\hbar$ and $V \in \{L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)\}$.

If $V \in L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)$, the Hamiltonian $H_0 = [(-\hbar^2/2\mu)\nabla^2 + V]$ is essentially self-adjoint on some suitable domain in $L^2(\mathbb{R}^n)$ and H, the extension of H_0, is such that (iH) generates a continuous unitary one-parameter group $U(t)$ on $L^2(\mathbb{R}^n)$.[14] Writing $\psi_t(x) = \psi(x, t)$ and $U(t) = \exp(-itH/\hbar)$ we obtain

$$\psi_t(x) = [U(t)\psi_0](x). \tag{49}$$

The diffeomorphism $D_t : \mathbb{R}^n \to \mathbb{R}^n$ induces a unitary map $U_0(t) : L^2(\mathbb{R}^n) \to L^{2'}(\mathbb{R}^n)$. We define the map $U_0(t)$ by $\psi_t \in L^2(\mathbb{R}^n)$ and $U_0(t)\psi_t = \phi_t \in L^{2'}(\mathbb{R}^n)$, according to

$$\phi_t(x_0) = J_t^{-1/2}(x) \exp[-iS(x, t)/\hbar]\psi_t(x), \quad \text{a.e.}, \tag{50}$$

where on the rhs $x = x[x_0, \nabla S_0(x_0), t]$ and $\|\phi_t\|^2_{L^2_t} = \int |\phi_t(x_0)|^2 d^n x_0$. The assumptions on D_t ensure that $U_0(t)$ is an isometry

$$\|\phi_t\|_{L^2_t} = \|U_0(t)\psi_t\|_{L^2_t} = \|\psi_t\|_{L^2}. \tag{51}$$

It is a simple matter to check that $\mathrm{Ran}U_0(t) = L^{2'}(\mathbb{R}^n)$. Indeed, defining $U_0^{-1}(t)$ by, $\phi \in L^{2'}(\mathbb{R}^n)$,

$$[U_0^{-1}(t)\phi](x) = J_t^{1/2}(x) \exp[iS(x, t)/\hbar]\phi[x_0(x, t)], \quad \text{a.e.}, \tag{52}$$

$U_0^{-1}(t)\phi \in L^2(\mathbb{R}^n)$ if $\phi \in L^{2'}(\mathbb{R}^n)$. Hence, $U_0^{-1}(t) = U_0^*(t)$, where U_0^* denotes the adjoint of U_0, $U_0^* : L^{2'}(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$.

We define the putative evolution operator $\tilde{U}(t, s)$, for $\phi_s \in L^{2'}(\mathbb{R}^n)$, according to

$$\tilde{U}(t, s) = U_0(t)U(t - s)U_0^*(s). \tag{53}$$

Here, in $\tilde{U}(t, s)$, s denotes the initial time and t denotes the final time ($t > s$) and $\tilde{U}(t, s)$ has the evolution property ($u > t > s$)

$$\tilde{U}(u, t)\tilde{U}(t, s) = \tilde{U}(u, s). \tag{54}$$

Evidently $\tilde{U}(t, s)$ is unitary on $L^{2'}(\mathbb{R}^n)$, because $U(t - s)$ is unitary on $L^2(\mathbb{R}^n)$.

The infinitesimal generator ($iA(t)$) of the evolution operator $\tilde{U}(t, s)$ is defined by

$$iA(t)\tilde{U}(t, s) = \frac{d\tilde{U}(t, s)}{dt} = \lim_{k \to 0^+} \frac{\tilde{U}(t + k, s) - \tilde{U}(t, s)}{k}, \tag{55}$$

so that

$$A(t) = \left\{ -i\frac{dU_0(t)}{dt}\tilde{U}(t, s)U_0^*(s) - iU_0(t)\frac{dU(t - s)}{dt}U_0^*(s) \right\} \tilde{U}^*(t, s), \tag{56}$$

where $d/dt = \partial/\partial t|_{x_0}$ is the time derivative at constant x_0,

$$\left.\frac{\partial}{\partial t}\right|_{x_0} = \left.\frac{\partial}{\partial t}\right|_x + \mu^{-1}\nabla S \cdot \nabla_x.$$

Using the fact that J satisfies the continuity equation

$$\frac{\partial J_t}{\partial t} + \mu^{-1}\nabla S \cdot \nabla J_t + \mu^{-1}J_t\nabla^2 S = 0, \tag{57}$$

we obtain

$$-i\frac{dU_0(t)}{dt} = -\hbar^{-1}\left(\frac{\partial S}{\partial t} + \frac{|\nabla S|^2}{\mu} + \frac{i\hbar\nabla^2 S}{2\mu}\right)U_0(t). \tag{58}$$

Also, from the above we obtain

$$-i\frac{dU(t - s)}{dt} = \left(\frac{-H}{\hbar} - \frac{i\nabla S}{\mu}\nabla\right)U(t - s). \tag{59}$$

Combining these results, gives the operator identity

$$A(t) = -\hbar^{-1}\left\{\frac{\partial S}{\partial t} + \frac{|\nabla S|^2}{\mu} + \frac{i\hbar}{2\mu}\nabla^2 S\right\} - \hbar^{-1}U_0(t)HU_0^*(t) - i\mu^{-1}U_0(t)\nabla S \cdot \nabla U_0^*(t). \tag{60}$$

Putting $H = [-(\hbar^2/2\mu)\nabla^2 + V]$ and recalling that S satisfies the Hamilton–Jacobi equation, we arrive at

$$A(t) = -\hbar^{-1}\left(\frac{|\nabla S|^2}{2\mu} + \frac{i\hbar}{2\mu}\nabla^2 S\right) + (2\mu)^{-1}\hbar \times U_0(t)\nabla^2 U_0^*(t) - i\mu^{-1}U_0(t)\nabla S \cdot \nabla U_0^*(t). \tag{61}$$

However,

$$U_0(t)\nabla^2 U_0^*(t) = J_t^{-1/2}\nabla^2 J_t^{1/2} + U_0(t)[\nabla^2, \exp(iS/\hbar)]J_t^{1/2} = J_t^{-1/2}\nabla^2 J_t^{1/2} + J_t^{-1/2}i\hbar^{-1}\nabla S \cdot \nabla J_t^{1/2} + U_0(t)\nabla \cdot i\hbar^{-1}\nabla S U_0^*(t), \tag{62}$$

where $[\nabla^2, \exp(iS/\hbar)] = \nabla^2 \exp(iS/\hbar) - \exp(iS/\hbar)\nabla^2$, is the commutator of ∇^2 and $\exp(iS/\hbar)$.

Hence, using the operator identities $\nabla \cdot \nabla S - \nabla S \cdot \nabla = \nabla^2 S$ and $\nabla S \cdot \nabla U_0^*(t) = \exp(iS/\hbar)[\nabla S \cdot \nabla J_t^{1/2} + i\hbar^{-1}|\nabla S|^2 J_t^{1/2}]$, and splitting up the last term in Eq. (61), we finally obtain

$$A(t) = \frac{+\hbar}{2\mu}J_t^{-1/2}\nabla^2 J_t^{1/2}. \tag{63}$$

This expresses $A(t)$ as a differential operator on a sufficiently small domain $\mathcal{D}_t \subset L^{2'}(\mathbb{R}^n)$, $\mathcal{D}_t = \{\phi \in L^{2'}(\mathbb{R}^n) | A(t)\phi \in L^{2'}(\mathbb{R}^n)\}$. Here $\nabla^2 = \nabla_x^2$ must be expressed in the curvilinear coordinates $x_0(x, t)$, so too with J_t.

It is not difficult to show that $A(t)$ as defined above is symmetric. [We can extend the domain of definition of $A(t)$ by defining ∇^2 as a pseudodifferential operator by taking Fourier transforms. In this way we can make $A(t)$ self-adjoint, but this is hardly worthwhile here.]

Putting $A(t) = -\hbar H(t)/2\mu$, we see that

$$i\frac{\partial\phi_t}{\partial t} = \frac{+\hbar}{2\mu}H(t)\phi_t. \tag{64}$$

Integrating and using the symmetry of $H(\tau)$ gives, for $\phi_0 \in \cap_{\tau \in (0,t)}\mathcal{D}_\tau$,

$$(\phi_0, \phi_t)_{L^2_t} - (\phi_0, \phi_0)_{L^2_t} = \frac{-i\hbar}{2\mu}\int_0^t (\phi_0, H(\tau)\phi_\tau)_{L^2_t}\, d\tau = \frac{-i\hbar}{2\mu}\int_0^t (H(\tau)\phi_0, \tilde{U}(\tau, 0)\phi_0)_{L^2_t}\, d\tau. \tag{65}$$

Using the Cauchy–Schwarz inequality and the isometric property of $\tilde{U}(\tau, 0)$,

$$|(\phi_0, \phi_t)_{L^2_t} - (\phi_0, \phi_0)_{L^2_t}| \le \frac{\hbar\|\phi_0\|_{L^2_t}}{2\mu}\int_0^t \|H(\tau)\phi_0\|_{L^2_t}\, d\tau = \frac{\hbar\|\phi_0\|_{L^2_t}}{2\mu}\int_0^t F(\tau)\, d\tau, \tag{66}$$

where $F(\tau) = \|\nabla_x^2 J_\tau^{1/2}(x)\phi_0[x_0(x, \tau)]\|_{L^2}$. Hence, if $F(\tau) \in L^1(0, t)$,

$$|(\phi_0, \phi_t)_{L^2_{\frac{1}{2}}} - (\phi_0, \phi_0)_{L^2_{\frac{1}{2}}}| \to 0, \tag{67}$$

as $\hbar \to 0$.

Using the isometric property of $\tilde{U}(t, 0)$, we obtain

$$\|\phi_t - \phi_0\|^2_{L^2_{\frac{1}{2}}} = 2(\phi_0, \phi_0)_{L^2_{\frac{1}{2}}} - (\phi_0, \phi_t)_{L^2_{\frac{1}{2}}} - (\phi_t, \phi_0)_{L^2_{\frac{1}{2}}}$$
$$\leq 2|(\phi_0, \phi_t)_{L^2_{\frac{1}{2}}} - (\phi_0, \phi_0)_{L^2_{\frac{1}{2}}}| \to 0, \tag{68}$$

as $\hbar \to 0$. Finally, changing integration variables once more, we arrive at

$$\|\exp[-iS(x, t)/\hbar]\psi_\hbar(x, t) - J_t^{1/2}(x)\phi_0[x_0(x, t)]\|_{L_2} \to 0, \tag{69}$$

as $\hbar \to 0$. This proves Theorem 2.

To see the physical significance of the last result for a single particle in $\mathbb{R}^n$ ($n \leq 3$) we choose $S_0(x) = p_0 \cdot x$, where $p_0 = (p_0^1, \ldots, p_0^n) \in \mathbb{R}^n$, $p_0 \cdot x = p_0^1 x^1 + p_0^2 x^2 + \cdots + p_0^n x^n$. Then the Cauchy data $\psi(x, 0) = \exp(ip_0 \cdot x/\hbar)\phi_0(x)$ corresponds to $(p\psi)(x, 0) = (\hbar/i)\nabla\psi(x, 0) = p_0\psi + O(\hbar)$. Hence, as $\hbar \to 0$, the Cauchy data corresponds to giving the quantum mechanical particle a fixed initial momentum $p_0 \in \mathbb{R}^n$. The diffeomorphism corresponding to the equivalent classical flow we denote by $T_t : \mathbb{R}^n \to \mathbb{R}^n$, $T_t x_0 = x[x_0, p_0, t]$. Let Ω be any measurable subset of $\mathbb{R}^n$ and let the limiting quantum mechanical probability of finding the particle (with the given initial conditions) in Ω at time t be $P(t, \Omega)$. Then a simple consequence of Theorem 2 is that

$$P(t, \Omega) = \lim_{\hbar \to 0} \int_\Omega |\psi_\hbar(x, t)|^2 d^n x$$
$$= \lim_{\hbar \to 0} \int_{T_t^{-1}\Omega} |\psi_\hbar(x_0, 0)|^2 d^n x_0 = P(0, T_t^{-1}\Omega) \tag{70}$$

for $t < T$ and Ω any measurable subset of $\mathbb{R}^n$. This is tantamount to quantum mechanics $\to$ classical mechanics as $\hbar \to 0$.

An analogous result holds for any finite number, say, r particles in configuration space $\mathbb{R}^{3r}$. We put $n = 3r$. Then we must consider the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2m_1}\nabla_1^2\psi\frac{-\hbar^2}{2m_2}\nabla_2^2\psi - \cdots\frac{-\hbar^2}{2m_r}\nabla_r^2\psi + V(y)\psi, \tag{71}$$

for $\psi(y, t)$ with $\psi(y, 0) = \exp(i\sum_{j=1}^r \mathbf{p}_j \cdot \mathbf{y}_j/\hbar)\phi_0(y)$, where $y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r) \in \mathbb{R}^{3r}$, $p = (\mathbf{p}_1, \ldots, \mathbf{p}_r) \in \mathbb{R}^{3r}$ and $\Delta_j^2 = \nabla^2 y_j$ is the three-dimensional Laplacian with respect to $\mathbf{y}_j$. We simply put $x = (\sqrt{m_1}\mathbf{y}_1, \ldots, \sqrt{m_r}\mathbf{y}_r)$ to obtain

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2}\nabla_x^2\psi + V\left[\frac{x_1}{\sqrt{m_r}}, \ldots, \frac{x_r}{\sqrt{m_r}}\right]\psi. \tag{72}$$

This corresponds to the above theorem with $\mu = 1$. Here the classical equations now read

$$\dot{x}_j = -\nabla_{x_j}V(y), \quad \dot{x}_j(0) = \frac{p_j}{m_j^{1/2}} \Leftrightarrow m_j\ddot{y}_j = -\nabla_j V(y),$$

$$m_j\dot{y}_j(0) = p_j, \quad j = 1, 2, \ldots, r. \tag{73}$$

Solving the above equations with $y_j(0) = y_0^j$ induces the diffeomorphism $T_t : \mathbb{R}^{3r} \to \mathbb{R}^{3r}$, $T_t y_0 = y[y_0, p, t]$. Let $P(t, \Omega)$ be the limiting quantum mechanical probability of observing the r particles at time t in $\Omega \subset \mathbb{R}^{3r}$, a measurable subset of the configuration space. Then from the above theorem for the appropriate initial conditions

$$P(t, \Omega) = P(0, T_t^{-1}\Omega). \tag{74}$$

Thus, we see that as $\hbar \to 0$ quantum mechanics $\to$ classical mechanics.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. David Elworthy of the Mathematical Institute of Warwick University for technical advice concerning the transformation formulas for Wiener integrals and Dr. John Ball of the Mathematics Department of Heriot–Watt University, Edinburgh, for a helpful suggestion. I must point out that a special instance of the Wiener integral evaluated in Sec. 3 was first calculated from an infinite determinant by Professor Cécile Morette DeWitt of the Department of Physics and Astronomy, University of Texas at Austin, and communicated to me by letter. Although the evaluation given here is quite different, I must stress that it would not have been possible without Professor DeWitt's calculation for which I am most grateful. Finally I would like to thank Professor John T. Lewis of the Dublin Institute of Advanced Studies for his constant interest and encouragement and for the opportunity to lecture on some of the above results to the School of Theoretical Physics at the Dublin Institute of Advanced Studies.

[1] (a) A. Truman, J. Math. Phys. **17**, 1852 (1976); (b) A. Truman, "The classical action in nonrelativistic quantum mechanics," to appear in J. Math. Phys.

[2] See also: (a) Cécile Morette DeWitt, Commun. Math. Phys. **28**, 47 (1972), (b) Cécile Morette DeWitt, Commun. Math. Phys. **37**, 63 (1974), (c) Cécile Morette DeWitt, Ann. Phys. (N.Y.) **97**, 367–99 (1976), (d) S. Albeverio and R. Hoegh-Krohn, *The Mathematical Theory of Feyman Path Integrals* (Springer lecture notes in Math., 1976).

[3] (a) J.H. Van Vleck, Proc. Nat. Acad. Sci. **14**, 178 (1928). This paper makes elegant use of the Jordan–Dirac transformation theory to obtain classical mechanics as a limiting case of quantum mechanics when $\hbar \to 0$. I am grateful to Professor Cécile Morette DeWitt for pointing out this reference. (b) G. Wentzel, Z. Phys. **38**, 518 (1926), (c) L. Brillouin, J. Phys. **7**, 353 (1926); (d) C. Eckart, Proc. Nat. Acad. Sci. **12**, 684 (1926), (e) J. Marsden, *Applications of Global Analysis in Mathematical Physics* (Publish or Perish, Boston, 1974). See Appendix. Here the hydrodynamic formulation of quantum mechanics and some recent results in global analysis are used to good effect to obtain an analogous result. (f) For a very detailed analysis see S. Albeverio and R. Hoegh-Krohn, "Oscillatory integrals and the method of stationary phase in infinitely many dimensions with applications to the classical limit of quantum mechanics. I," University of Oslo preprint 1975, to appear in Invent. Math. (g) See especially Ref. 5 below.

[4] (a) E. Nelson, Phys. Rev. **150**, 1079 (1966), (b) E. Nelson, *Dynamical Theories of Brownian Motion, Mathematical Notes* (Princeton U.P., Princeton, New Jersey, 1967), (c) S. Albeverio and R. Hoegh-Krohn, J. Math. Phys. **15**, 1745 (1974).

[5] (a) V.P. Maslov, Zh. Vych. Mat. **1**, 638–63 (1961), (b) V.P. Maslov, Zh. Vych. Mat. **1**, 113–28 (1961), (c) V.P. Maslov, *Theorie des Perturbations et Methods Asymptotiques* (Dunod, Paris, 1972).

[6] (a) M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972), p. 295 and M. Reed and B. Simon, *Fourier Analysis, Self Adjointness* (Academic, New York, 1975).

(b) The original references are: M. Kac, Trans. Amer. Math. Soc. **65**, 1 (1949); in *Proceedings of the Second Berkeley Symposium in Mathematical Statistics and Probability*, 1950 (University of California Press, Berkeley, 1951); and *Probability and Related Topics in Physical Sciences* (Interscience, London, 1959).

[7](a) I.M. Koval'chik, Russ. Math. Surv. **18**, 97 (1963).
(b) The original reference here is R.H. Cameron and W.T. Martin, Trans. Am. Math. Soc. **58**, 184 (1945).

[8]E. Nelson, Ref. 3(b), p. 40, Theorem 7.4 proves this is valid a.e. w.r.t. Wiener measure if $dX_j/d\tau$ is absolutely continuous on $[0,t]$, for $j=1, 2,\ldots,n$.

[9]I.M. Gelfand and A.M. Yaglom, J. Math. Phys. **1**, 48—69 (1960).

[10]A. Guichardet, Springer Lecture Notes in Mathematics Vol. 261.

[11]J.L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1964), see, e.g., p. 32.

[12]J. Dieudonné, *Foundations of Modern Analysis* (Academic, New York, 1969), Chap. X, Sec. 6.

[13]These results are somewhat reminiscent of S.A. Molchanov, Russ. Math. Surv. 30, 1—63 (1975).

[14]M. Reed and B. Simon, Ref. 5, Theorem X.15.

Proof that the H^- ion has only one bound state. Details and extension to finite nuclear mass

Robert Nyden Hill

Physics Department, University of Delaware, Newark, Delaware 19711
(Received 23 June 1977)

It is rigorously demonstrated that the H^- ion, treated in nonrelativistic approximation with Coulomb interactions only, has only one bound state for the electron to nucleus mass ratio less than 0.21010636. This extends earlier work which had proven the result in the fixed (infinite mass) nucleus approximation. The method used can, if desired, also be used to calculate rigorous lower bounds to the energies of those bound states of two electron atomic systems which do exist.

I. INTRODUCTION

The H^- ion, made up of a proton and two electrons, has long been known to have one bound state.[1] The first rigorous proof that additional bound states do not exist in the fixed (infinite mass) nucleus approximation with Coulomb interactions only appeared recently in abbreviated form.[2] The present paper provides additional details of that proof, and extends it to finite nuclear mass to prove that the H^- ion has only one bound state if (1) only Coulomb interactions are included and (2) the electron—proton mass ratio is less than 0.21010636. The methods used to obtain this result can also, if desired, be used to calculate rigorous lower bounds to the energies of those bound states of two electron atomic systems which do exist.

An effort has been made to make the present paper accessible to a wide audience. Theorems which are not a part of traditional graduate training in theoretical physics and chemistry are discussed briefly before they are applied. Standard Dirac notation is employed. When operators are written out explicitly, continuous matrix notation is used where convenient. The mathematican who is uncomfortable with such notation should have no difficulty rewriting the operators as tensor products.

The mathematical foundation for the study of N-particle quantum mechanical Coulomb systems was laid by Kato,[3] who gave the first rigorous definition of the Hamiltonian, proved that it is essentially self-adjoint (Hermitian), proved that eigenfunctions belonging to whatever discrete eigenvalues the Hamilton has will satisfy the Schrödinger equation as a differential equation except at singular points of the potential, and proved that the discrete eigenvalues (below the continuum) and their eigenfunctions could be characterized by the familiar Rayleigh—Ritz variational principle. It has subsequently been shown that the essential spectrum of the Hamiltonian (i.e., the continuous spectrum plus limit points of the discrete spectrum and discrete eigenvalues of infinite multiplicity) consists of the half-line $[\mu, \infty)$ where μ is the lowest threshold for breakup into subsystems.[4] This result, which most physicists would consider intuitively obvious, holds in each subspace of particular permutation and rotation—reflection symmetry. It has also been proven, for classes of potentials which include the Coulomb potential, that eigenfunctions belonging to discrete eigenvalues fall off exponentially,[5] and that there are no positive discrete eigenvalues embedded in the continuous spectrum.[6] Rigorous results in the theory of N-particle Schrödinger Hamiltonians have been reviewed by Sigalov,[7] Kato,[8] Simon,[9] Jorgens and Wiedmann,[10] and Hunziker.[11]

The bound state spectrum of negative ions is very different from the bound state spectrum of positive ions and neutrals. Positive atomic ions and neutral atoms have an infinite number of bound states,[12] with a spectrum which is qualitatively similar to that of hydrogen. A stable negative ion, however, has only a finite number of bound states,[13] for which electron correlation is of decisive importance.[14] H^-, for example, is believed not to be bound in Hartree—Fock approximation.[15] In what follows (and in the above), the term "bound state" refers to a discrete eigenvalue of finite multiplicity *below the continuum*; discrete eigenvalues embedded in the continuum will, by definition, not be bound states. Methods for counting the number of bound states of a Schrödinger Hamiltonian are by now well developed for one particle in an external potential and, equivalently, for two particles without external forces.[16] However, the counting of the bound states for the more general N-body problem is in a somewhat more primitive state. The Rayleigh—Ritz variational procedure, which gives upper bounds to bound state eigenvalues, can be used to prove the existence of bound states and, more generally, to obtain a lower bound on the number of bound states. Rayleigh—Ritz is, however, powerless to establish the nonexistence of bound states or, more generally, to give an upper bound on the number of bound states. An adiabatic approach, which can in principle give rigorous lower bounds to ground state eigenvalues and/or prove the nonexistence of bound states, has been used by Gertler, Snodgrass, and Spruch and extended by Aronson, Kleinman, and Spruch,[17] but is powerless to show that H^- has at most one bound state. An alternative approach of Aronson, Kleinman, and Spruch[18] has not yet been made rigorous. The methods of Refs. 12 and 13, although rigorous, have not been shown capable of establishing the number of bound states of H^-.

The results of the present paper are obtained via generalizations of methods introduced by Bazley[19] and by Bazley and Fox[20] to construct lower bounds to helium eigenvalues. The methods of Bazley and Fox in turn have their roots in the Weinstein method of intermediate problems, which is discussed in monographs by Gould,[21] by Weinstein and Stenger,[22] and by Weinberger.[23] Sec-

tion II establishes notation and constructs the internal Hamiltonian in a convenient dimensionless form. Section III shows how to obtain one-particle Schrödinger equations whose eigenvalues are sufficiently good lower bounds to the eigenvalues of the original internal Hamiltonian to show that H^- has at most one bound state in the fixed (infinite mass) nucleus approximation. Section IV contains the proofs that the single particle equations derived in III have one bound state for the singlet sector and no bound states for the triplet sector. Section V extends the results to finite nuclear mass. An understanding of the basic ideas can be had from Secs. II, III, and V A; Section IV and the remainder of V contains details which can be omitted at a first reading.

II. NOTATION: THE HAMILTONIAN

The Schrödinger Hamiltonian for two-electron atomic systems such as H^-, He, or Li^+ in nonrelativistic approximation with Coulomb interactions only, is

$$H_{\text{total}} = (2M)^{-1}\pi_0^2 + (2m)^{-1}(\pi_1^2+\pi_2^2) - Ze^2(|\mathbf{x}_1-\mathbf{x}_0|^{-1}+|\mathbf{x}_2-\mathbf{x}_0|^{-1}) + e^2|\mathbf{x}_1-\mathbf{x}_2|^{-1}, \tag{2.1}$$

where $\mathbf{x}_1$, π_0 are the nuclear coordinate and momentum, $\mathbf{x}_1, \mathbf{x}_2, \pi_1, \pi_2$ are electron coordinates and momenta, Ze and M are the nuclear charge and mass, and $-e$ and m are the electron charge and mass. The center of mass motion can be separated off and the internal Hamiltonian reduced to dimensionless form by making the definitions

$$\mathbf{P} = \pi_0+\pi_1+\pi_2, \tag{2.2}$$

$$\mathbf{R} = [M\mathbf{x}_0 + m(\mathbf{x}_1+\mathbf{x}_2)]/(M+2m), \tag{2.3}$$

$$\mathbf{p}_1 = a\hbar^{-1}[(m+M)\pi_1 - m(\pi_0+\pi_2)]/(M+2m), \tag{2.4}$$

$$\mathbf{r}_1 = a^{-1}(\mathbf{x}_1-\mathbf{x}_0), \tag{2.5}$$

$$\mathbf{p}_2 = a\hbar^{-1}[(m+M)\pi_2 - m(\pi_0+\pi_1)]/(M+2m), \tag{2.6}$$

$$\mathbf{r}_2 = a^{-1}(\mathbf{x}_2-\mathbf{x}_0). \tag{2.7}$$

Here $\hbar$ is Planck's constant divided by 2π, $a=\hbar^2/(\mu e^2)$ is the reduced mass Bohr radius, and $\mu = mM/(m+M)$ is the reduced mass. It is easy to show that the transformation from $\pi_0, \mathbf{x}_0, \pi_1, \mathbf{x}_1, \pi_2, \mathbf{x}_2$ to $\mathbf{P}, \mathbf{R}, \hbar\mathbf{p}_1, \mathbf{r}_1, \hbar\mathbf{p}_2, \mathbf{r}_2$ is canonical, and that the Hamiltonian can be expressed as

$$H_{\text{total}} = [2(M+2m)]^{-1}P^2 + \tfrac{1}{2}e^2a^{-1}H_{\text{int}}, \tag{2.8}$$

where the internal Hamiltonian H_{int} is

$$H_{\text{int}} = H_0 + 2\gamma\mathbf{p}_1\cdot\mathbf{p}_2 + V, \tag{2.9}$$

with

$$H_0 = p_1^2 - 2Zr_1^{-1} + p_2^2 - 2Zr_2^{-1}, \tag{2.10}$$

$$V = 2|\mathbf{r}_1-\mathbf{r}_2|^{-1}, \tag{2.11}$$

and

$$\gamma = m/(m+M) = \mu/M. \tag{2.12}$$

The term $2\gamma\mathbf{p}_1\cdot\mathbf{p}_2$ in the internal Hamiltonian (2.9), which vanishes in the fixed (infinite mass) nucleus approximation, is known as the Hughes–Eckart term.[24]

III. ONE-PARTICLE EQUATIONS

This section will show how to obtain one-particle Schrödinger equations whose eigenvalues are lower bounds to the eigenvalues of the internal Hamiltonian H_{int}. The analysis is carried out in the fixed (infinite mass) nucleus approximation $\gamma=0$. Energy eigenvalues are in units of $h^{-1}\,\text{Ry} = e^2/(2a)$; this place the hydrogen atom's ground state energy at -1.

A. Lower bounds

The basic tool to be used is a comparison theorem[25] well known among mathematicians who work on eigenvalue problems:

Theorem 1: Let $H^{(1)}$ and $H^{(2)}$ be two essentially self-adjoint (Hermitian) Hamiltonians whose discrete eigenvalues below the bottom of the of the essential spectrum can be characterized by the familiar variational principle $E = \min\langle\psi|H|\psi\rangle/\langle\psi|\psi\rangle$, with the minimization for excited states carried out subject to the constraint that $|\psi\rangle$ be orthogonal to preceding eigenvectors. Denote the ordered eigenvalues of $H^{(i)}$ by $E_1^{(i)} \leq E_2^{(i)} \leq \cdots \leq E_n^{(i)} \leq \cdots \leq E_{\text{ess}}^{(i)}$, where $E_{\text{ess}}^{(i)}$ is the energy at which the essential spectrum (if any) begins. Assume $\langle\psi|H^{(1)}|\psi\rangle$ is defined for all vectors $|\psi\rangle$ for which $\langle\psi|H^{(2)}|\psi\rangle$ is defined. Then if $\langle\psi|H^{(1)}|\psi\rangle \leq \langle\psi|H^{(2)}|\psi\rangle$ holds for all admissible state vectors $|\psi\rangle$, $E_n^{(1)} \leq E_n^{(2)}$ holds for all n, and $E_{\text{ess}}^{(1)} \leq E_{\text{ess}}^{(2)}$.

The result $E_1^{(1)} \leq E_1^{(2)}$ for the ground state energy follows immediately from the "familiar variational principle." Proofs of the results for the excited states are usually based on one of the minimax characterizations[26] of eigenvalues. In practical applications of Theorem 1 to the computation of lower bounds, $H^{(2)}$ is the original Hamiltonian, while $H^{(1)}$ is something more tractable. The results of the present paper will be obtained by letting $H^{(2)}$ be the internal Hamiltonian H_{int} while $H^{(1)}$ is something for which the Schrödinger equation is reducible to one-particle Schrödinger equations.

The lower bounding Hamiltonian $H^{(1)}$ for the case $\gamma=0$ will be constructed by generalizing a method introduced by Bazley[19] to construct lower bounds to helium eigenvalues: Replace V in $H^{(2)} = H_{\text{int}} = H_0 + V$ by $V^{1/2}PV^{1/2}$ where P is a projection operator. The positive square root of $2|\mathbf{r}_1-\mathbf{r}_2|^{-1}$ is to be taken when constructing $V^{1/2}$. The fact that a projection operator such as P cannot increase the length of a vector such as $V^{1/2}|\psi\rangle$ implies that $\langle\psi|V^{1/2}PV^{1/2}|\psi\rangle \leq \langle\psi|V|\psi\rangle$. The eigenvalues of $H^{(1)} = H_0 + V^{1/2}PV^{1/2}$ are then lower bounds to the eigenvalues of $H_{\text{int}} = H_0 + V$. Bazley constructed his $V^{1/2}PV^{1/2}$ by starting with the first N eigenvectors $|\xi_i\rangle$ of H_0 and using Schmidt orthogonalization to construct vectors $|\xi_i'\rangle$ such that

$$\langle\xi_i'|V^{-1}|\xi_j'\rangle = \delta_{i,j}. \tag{3.1}$$

His $V^{1/2}PV^{1/2}$ then took the form

$$V^{1/2}P_{\text{Bazley}}V^{1/2} = \sum_{i=1}^{N}|\xi_i'\rangle\langle\xi_i'|, \tag{3.2}$$

for small finite N. It follows immediately from (3.1) and (3.2) that $P_{\text{Bazley}}^\dagger = P_{\text{Bazley}}$ and $P_{\text{Bazley}}^2 = P_{\text{Bazley}}$, so that P_{Bazley} is a projection operator as is required. Be-

cause $|\xi_i'\rangle$ is a linear combination of the first N $|\xi_i\rangle$, $V^{1/2}P_{\text{Bazley}}V^{1/2}$ couples only the first N states of H_0, and the eigenvalue problem for $H_0+V^{1/2}P_{\text{Bazley}}V^{1/2}$ reduces to an eigenvalue problem for an $N\times N$ matrix.

B. The construction of P

The discrete spectrum of H_0, given by

$$E^{(0)}_{n_1,n_2}=-Z^2(n_1^{-2}+n_2^{-2}), \tag{3.3}$$

where (n_1,n_2) is any pair of positive integers, is known from the theory of the hydrogen atom. Because the two electron atomic system can dissociate into a free electron plus an electron bound in a hydrogenic ground state with energy $-Z^2$, both H_0 and H_0+V have a continuous spectrum beginning at $-Z^2$ (see Refs. 4 and paragraph 3 of the Introduction). The spectrum (3.3) has an infinite number of levels belonging to pairs of quantum numbers of the form $(1,n_2)$ and $(n_1,1)$ which lie below $-Z^2$ and accumulate at $-Z^2$. The associated eigenfunctions have the form $\phi_1(r_1)\chi_{n_2}(\mathbf{r}_2)$ and $\chi_{n_1}(\mathbf{r}_1)\phi_1(r_2)$ where

$$\phi_1(r)=Z^{3/2}\pi^{-1/2}\exp(-Zr) \tag{3.4}$$

is the normalized hydrogenic ground state wavefunction and $\chi_n(\mathbf{r})$ is a hydrogenic bound state wavefunction belonging to the principle quantum number n. In order to prove that $H=H_0+V$ has only one bound state for $Z=1$, $V^{1/2}PV^{1/2}$ must contain enough of the original repulsive V to push all but one of these levels up to -1; in particular $V^{1/2}PV^{1/2}$ must couple to all of these levels of H_0 which lie below the continuum. Stated another way, $V^{1/2}PV^{1/2}$ must retain enough of the original V to preserve shielding: If one electron is in a hydrogenic ground state with the second electron far out, the far out electron must see, after the replacement of V by $V^{1/2}PV^{1/2}$, a potential which cannot support an infinite number of bound states. Clearly $V^{1/2}P_{\text{Bazley}}V^{1/2}$, which couples only a finite number of low-lying states of H_0, will not work here; a generalization is needed.

Bazley's procedure will now be generalized to obtain a $V_1=V^{1/2}P_1V^{1/2}$ which couples to all states of the form $f(\mathbf{r}_1)\phi_1(r_2)$ where f is arbitrary. Start with the wavefunction

$$\chi_{\mathbf{r}}(\mathbf{r}_1,\mathbf{r}_2)=\delta(\mathbf{r}-\mathbf{r}_1)\phi_1(r_2) \tag{3.5}$$

corresponding to particle one in a position eigenstate at $\mathbf{r}$ and particle two in the hydrogenic ground state. Replace the orthogonality condition (3.1) by

$$\int\overline{\chi_{\mathbf{r}}'(\mathbf{r}_1,\mathbf{r}_2)}^{\frac{1}{2}}\,|\,\mathbf{r}_1-\mathbf{r}_2\,|\,\chi_{\mathbf{r}'}'(\mathbf{r}_1,\mathbf{r}_2)d^3\mathbf{r}_1d^3\mathbf{r}_2=\delta(\mathbf{r}-\mathbf{r}'), \tag{3.6}$$

and the expression (3.2) for $V^{1/2}PV^{1/2}$ by

$$\begin{aligned}V_1(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_1',\mathbf{r}_2')&\equiv(V^{1/2}P_1V^{1/2})(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_1',\mathbf{r}_2')\\&=\int\chi_{\mathbf{r}}'(\mathbf{r}_1,\mathbf{r}_2)\overline{\chi_{\mathbf{r}}'(\mathbf{r}_1',\mathbf{r}_2')}\,d^3\mathbf{r}.\end{aligned} \tag{3.7}$$

One then easily finds that

$$\chi_{\mathbf{r}}'(\mathbf{r}_1,\mathbf{r}_2)=[U(r_1)]^{1/2}\delta(\mathbf{r}-\mathbf{r}_1)\phi_1(r_2), \tag{3.8}$$

and

$$V_1(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_1',\mathbf{r}_2')=U(r_1)\delta(\mathbf{r}_1-\mathbf{r}_1')\phi_1(r_2)\overline{\phi_1(r_2')}, \tag{3.9}$$

where

$$U(r)\equiv[\int\tfrac{1}{2}|\mathbf{r}-\mathbf{r}'|\,|\phi_1(r')|^2d^3\mathbf{r}']^{-1}. \tag{3.10}$$

From a rigorous point of view, the difficulty that $\chi_{\mathbf{r}}$ and $\chi_{\mathbf{r}}'$ are not in the Hilbert space of square integrable functions within which one would like to work is most easily circumvented by regarding the steps which lead to (3.9) and (3.10) as a purely formal heuristic argument. Equation (3.9), which shows that V_1 is the tensor product of the ordinary potential (multiplicative operator) U in the space of $\mathbf{r}_1$ with the dyad $|\phi_1\rangle\langle\phi_1|$ in the space of $\mathbf{r}_2$, is taken as the definition of V_1, with the projection operator P_1 defined by $P_1=V^{-1/2}V_1V^{-1/2}$. It is then easy to see that P_1 is a bounded Hermitian operator, and to verify directly from (3.9) and (3.10) that $V_1V^{-1}V_1=V_1$, which is equivalent to $P_1^2=P_1$, holds.

Unfortunately V_1 couples only to those levels of H_0 below the continuum which have the form $\chi_{n_1}(\mathbf{r}_1)\phi_1(r_2)$. Interchange of 1 and 2 yields the operator

$$\begin{aligned}V_2(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_1',\mathbf{r}_2')&\equiv(V^{1/2}P_2V^{1/2})(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_1',\mathbf{r}_2')\\&=\phi_1(r_1)\overline{\phi_1(r_1')}U(r_2)\delta(\mathbf{r}_2-\mathbf{r}_2'),\end{aligned} \tag{3.11}$$

which couples to the levels of the form $\phi_1(r_1)\chi_{n_2}(\mathbf{r}_2)$. A $V^{1/2}PV^{1/2}$ which couples to *all* levels of H_0 below the continuum (actually, to everything below $-Z^2/2$) results from choosing P to be the projection onto the span of the ranges of P_1 and P_2. This P is given by[27]

$$\begin{aligned}P=\tfrac{1}{2}(P_1+P_2)+\tfrac{1}{2}\sum_{n=0}^{\infty}[&(I-P_1)P_2K_1^nP_2(I-P_1)\\&+(I-P_2)P_1K_2^nP_1(I-P_2)],\end{aligned} \tag{3.12}$$

where K_1 and K_2 are the operators

$$K_1=P_2P_1P_2,\quad K_2=P_1P_2P_1. \tag{3.13}$$

Convergence of the infinite series of operators in P is implied by the following observations. (1) Since K_1 and K_2 are products of projection operators, their eigenvalues cannot exceed 1. (2) Eigenvectors of K_1 and K_2 with eigenvalue 1 must be simultaneous eigenvectors of P_1 and P_2 (proof for K_1: Suppose $P_2P_1P_2|\psi\rangle=|\psi\rangle$. Multiply on the left by $\langle\psi|$ and by $\langle\psi|P_2$ to obtain $\langle\psi|P_2P_1P_2|\psi\rangle=\langle\psi|\psi\rangle$ and $\langle\psi|P_2^2P_1P_2|\psi\rangle=\langle\psi|P_2|\psi\rangle$, from which $\langle\psi|\psi\rangle=\langle\psi|P_2|\psi\rangle$. This implies that the inner product of $|\psi\rangle-P_2|\psi\rangle$ with itself is zero, so that $|\psi\rangle-P_2|\psi\rangle$ is the null vector and $P_2|\psi\rangle=|\psi\rangle$. This reduces $\langle\psi|P_2P_1P_2|\psi\rangle=\langle\psi|\psi\rangle$ to $\langle\psi|P_1|\psi\rangle=\langle\psi|\psi\rangle$, from which $P_1|\psi\rangle=|\psi\rangle$ follows by the same argument. Q.E.D.) (3) Simultaneous eigenvectors of P_1 and P_2 are annihilated by $(I-P_1)P_2$, $(I-P_2)P_1$, and their adjoints. (4) K_1 and K_2 are Hilbert–Schmidt operators (as will be seen), so that eigenvalues of K_1 and K_2 cannot accumulate at 1. Hence only eigenvalues of K_1 and K_2 which are strictly less than 1 contribute to the sum in P, which therefore converges by comparison with the geometric series. Thus (3.12) and (3.13) are a proper definition of P. It is straightforward to verify that P is Hermitian, that $PP_1=P_1P=P_1$, that $PP_2=P_2P=P_2$, and that $P^2=P$.

It should be noted that the terms in the sum in P have the form $A^\dagger A$ [with $K_1^{n/2}P_2(I-P_1)$ or $K_2^{n/2}P_1(I-P_2)$ for A when n is even and $P_1P_2K_1^{(n-1)/2}P_2(I-P_1)$ or $P_2P_1K_2^{(n-1)/2}P_1(I-P_2)$ for A when n is odd]. Thus lower

bounds to the eigenvalues of $H = H_0 + V$ are obtained even if all but a finite number of the terms in the infinite sum in P are discarded.

C. Analysis of $V^{1/2}PV^{1/2}$

The structure of $V^{1/2}PV^{1/2}$ with P given by (3.12) can be made more explicit by examining K_1 and K_2. Equations (3.9), (3.11), and (3.13) imply that

$$K_1(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') = [\tfrac{1}{2}|\mathbf{r}_1 - \mathbf{r}_2|U(r_2)]^{1/2}\phi_1(r_1) \times \int d^3\mathbf{r} k(\mathbf{r}_2, \mathbf{r}) k(\mathbf{r}, \mathbf{r}_2')\phi_1(r_1')[\tfrac{1}{2}|\mathbf{r}_1' - \mathbf{r}_2'|U(r_2')]^{1/2}, \tag{3.14}$$

where

$$k(\mathbf{r}, \mathbf{r}') = [U(r)]^{1/2}\phi_1(r)\tfrac{1}{2}|\mathbf{r} - \mathbf{r}'|\overline{\phi_1(r')}[U(r')]^{1/2}. \tag{3.15}$$

The corresponding expression for K_2 can be obtained by interchanging 1 and 2 on the right-hand side of (3.14). The fact that K_1 is a Hilbert–Schmidt kernel follows easily from (3.14) and (3.15). As a consequence of (3.10) and (3.14), the solutions of the eigenvalue problem $K_1|\chi_i^{(1)}\rangle = \mu_i^{(1)}|\chi_i^{(1)}\rangle$ are related to the solutions of the eigenvalue problem $k|h_i\rangle = \nu_i|h_i\rangle$ via $\mu_i^{(1)} = \nu_i^2$ and

$$\chi_i^{(1)}(\mathbf{r}_1, \mathbf{r}_2) = [\tfrac{1}{2}|\mathbf{r}_1 - \mathbf{r}_2|U(r_2)]^{1/2}\phi_1(r_1)h_i(\mathbf{r}_2). \tag{3.16}$$

It is straightforward to verify that

$$h_1(r) = (32Z)^{1/2}[35U(r)]^{-1/2}\phi_1(r) \tag{3.17}$$

is a normalized eigenfunction of k with eigenvalue $\nu_1 = 1$. The corresponding eigenfunction of K_1, which is

$$\chi_1(\mathbf{r}_1, \mathbf{r}_2) = [16Z|\mathbf{r}_1 - \mathbf{r}_2|/35]^{1/2}\phi_1(r_1)\phi_1(r_2), \tag{3.18}$$

is a simultaneous eigenfunction of P_1 and P_2, as it must be in light of the above discussion of the convergence of the sum in (3.12).

It will now be shown that $\mu_1^{(1)} = \nu_1^2 = 1$ is the only eigenvalue of K_1 which equals 1 and therefore does not contribute to the sum in (3.12). All other eigenvalues of K_1 are strictly less than 1 and do contribute, as a consequence of the fact that the eigenvalues of the operator M, defined by

$$M(\mathbf{r}, \mathbf{r}') = h_1(r)h_1(r') - k(\mathbf{r}, \mathbf{r}'), \tag{3.19}$$

are strictly less than 1. This can be proven by using the representation[28] $k = k_1 - k_2$ where

$$k_1(\mathbf{r}, \mathbf{r}') = \tfrac{1}{4}[U(r)]^{1/2}\phi_1(r)(1 + r)(1 + r')\overline{\phi_1(r')}[U(r')]^{1/2}, \tag{3.20}$$

and

$$k_2(\mathbf{r}, \mathbf{r}') = \tfrac{1}{4}[U(r)]^{1/2}\phi_1(r)\{(1 - r)(1 - r') + \pi^{-1}\int (|\mathbf{r} - \mathbf{r}''|^{-1} - r^{-1})(|\mathbf{r}' - \mathbf{r}''|^{-1} - r'^{-1})\, d^3\mathbf{r}''\}\overline{\phi_1(r')}[U(r')]^{1/2}. \tag{3.21}$$

Expectation values of k_1 and of k_2 cannot be negative because k_1 is a dyad of the form $|\alpha\rangle\langle\alpha|$ with positive coefficient while k_2 is a sum of such dyads with all coefficients positive. Theorem 1 then implies that the eigenvalues of k are bounded above by the eigenvalues of k_1. But k_1 has only one positive eigenvalue. Hence k has only one positive eigenvalue, which must be the eigenvalue 1 associated with the eigenfunction h_1. Therefore, all eigenvalues of M are nonnegative. Armed with this fact, the eigenvalues of M can be proven strictly less than 1 by considering the trace of k. It follows from (3.15) that

$$\sum_{i=1}^{\infty} \nu_i = \mathrm{Tr}k = 0. \tag{3.22}$$

Hence the sum of the eigenvalues of M is given by

$$\sum_{i=2}^{\infty} (-\nu_i) = \mathrm{Tr}M = 1. \tag{3.23}$$

Since all eigenvalues of M are nonnegative, each eigenvalue must be strictly less than 1 if no single eigenvalue exhausts the sum rule (3.23). That this is so follows from the partial wave decomposition of k given in Appendix A, which implies that $\mathrm{Tr}M$ has the partial wave decomposition

$$\mathrm{Tr}M = \sum_{l=0}^{\infty} (2l + 1)\,\mathrm{Tr}M_l, \tag{3.24}$$

where

$$\mathrm{Tr}M_l = \delta_{l,0} - \mathrm{Tr}k_l = \delta_{l,0} - \int_0^{\infty} k_l(r, r)r^2\, dr = \delta_{l,0} + 8[(2l - 1)(2l + 1)(2l + 3)]^{-1}C \tag{3.25}$$

with

$$C \equiv Z^3 \int_0^{\infty} U(r)\exp(-2Zr)r^3\, dr. \tag{3.26}$$

Since C is clearly positive [Eq. (3.10) shows that $U(r) \geq 0$; numerical integration yields $C \approx 0.318780514$], no single partial wave (and therefore no single eigenvalue) exhausts the sum rule (3.23).

With the aid of Eqs. (3.9)–(3.15), (3.18), and (3.19), $V^{1/2}PV^{1/2}$ can be written out explicitly as

$$(V^{1/2}PV^{1/2})(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') = k_d(\mathbf{r}_1, \mathbf{r}_1')\phi_1(r_2)\overline{\phi_1(r_2')} + k_d(\mathbf{r}_2, \mathbf{r}_2')\phi_1(r_1)\overline{\phi_1(r_1')} + k_e(\mathbf{r}_1, \mathbf{r}_2')\phi_1(r_2)\overline{\phi_1(r_1')} + k_e(\mathbf{r}_2, \mathbf{r}_1')\phi_1(r_1)\overline{\phi_1(r_2')}, \tag{3.27}$$

where k_d and k_e are the one-particle operators

$$k_d = U - (16Z/35)|\phi_1\rangle\langle\phi_1| + \sum_{n=1}^{\infty} U^{1/2}M^{2n}U^{1/2}, \tag{3.28}$$

$$k_e = \sum_{n=1}^{\infty} U^{1/2}M^{2n-1}U^{1/2}. \tag{3.29}$$

The physical effect of $V^{1/2}PV^{1/2}$ is clear from (3.27). $V^{1/2}PV^{1/2}$ couples only to states with at least one particle in a hydrogenic ground state. One of the particles is scattered while the other remains in the hydrogenic ground state. The terms containing k_d are direct terms while those containing k_e are exchange.

D. One-particle equations

The eigenvalue problem for $H_0 + V^{1/2}PV^{1/2}$ is simpler than the eigenvalue problem for $H_{\rm int} = H_0 + V$. Let $\mathcal{S}_S$ be the space spanned by functions of the form

$$\psi_S(\mathbf{r}_1, \mathbf{r}_2) = f(\mathbf{r}_1)\phi_1(r_2) + \phi_1(r_1)f(\mathbf{r}_2), \tag{3.30}$$

$\mathcal{S}_A$ the space spanned by functions of the form

$$\psi_A(\mathbf{r}_1, \mathbf{r}_2) = g(\mathbf{r}_1)\phi_1(r_2) - \phi_1(r_1)g(\mathbf{r}_2), \tag{3.31}$$

and $\mathcal{S}_\perp$ the orthogonal complement of the space spanned by functions which are linear combinations of functions of the form ψ_S and ψ_A. Each of the spaces $\mathcal{S}_S$, $\mathcal{S}_A$, and $\mathcal{S}_\perp$ is mapped into itself by $H_0 + V^{1/2}PV^{1/2}$, which is to say that these spaces are reducing spaces for $H_0 + V^{1/2}PV^{1/2}$. $V^{1/2}PV^{1/2}$ is zero on $\mathcal{S}_\perp$, so that eigenfunctions in $\mathcal{S}_\perp$ have their (discrete) eigenvalues given by (3.3) with $n_1 \geq 2$, $n_2 \geq 2$; the continuous spectrum in $\mathcal{S}_\perp$ begins at $-Z^2/4$. Because the bottom of the spectrum in $\mathcal{S}_\perp$ is at $-Z^2/2$, which is above the bottom of the continuum for the full problem at $-Z^2$, $\mathcal{S}_\perp$ need not be considered further.

The eigenvalue problems for $H_0 + V^{1/2}PV^{1/2}$ on $\mathcal{S}_S$ and $\mathcal{S}_A$ are equivalent to the following one-particle eigenvalue problems for f and g:

$$\begin{aligned}(I + |\phi_1\rangle\langle\phi_1|)[p^2 - ZV_0 + \tfrac{1}{2}Z^2|\phi_1\rangle\langle\phi_1| \\ + k_d + k_e](I + |\phi_1\rangle\langle\phi_1|)|f\rangle \\ = (E + Z^2)(I + |\phi_1\rangle\langle\phi_1|)|f\rangle,\end{aligned} \tag{3.32}$$

and

$$\begin{aligned}(I - |\phi_1\rangle\langle\phi_1|)(p^2 - ZV_0 + k_d - k_e)(I - |\phi_1\rangle\langle\phi_1|)|g\rangle \\ = (E + Z^2)(I - |\phi_1\rangle\langle\phi_1|)|g\rangle,\end{aligned} \tag{3.33}$$

where

$$p^2 = -\nabla^2 \tag{3.34}$$

is the one-particle kinetic energy, and

$$V_0 = 2r^{-1}, \tag{3.35}$$

so that $-ZV_0$ is the Coulomb interaction with the nucleus. I is the identity operator. Equations (3.32) and (3.33) are most easily derived by writing out the Schrödinger equation for $H_0 + V^{1/2}PV^{1/2}$ in configuration space with the aid of (3.27), (3.30), and (3.31), multiplying by $\overline{\phi_1(r_2)}$, and integrating $\mathbf{r}_2$ over all space. Occurrence of the operators $(I + |\phi_1\rangle\langle\phi_1|)$ and $(I - |\phi_1\rangle\langle\phi_1|)$ in (3.32) and (3.33) is not surprising once the normalization integrals for ψ_S and ψ_A have been written out: It follows directly from (3.30) and (3.31) that

$$\langle\psi_S|\psi_S\rangle = 2\langle f|(I + |\phi_1\rangle\langle\phi_1|)|f\rangle, \tag{3.36}$$

and

$$\langle\psi_A|\psi_A\rangle = 2\langle g|(I - |\phi_1\rangle\langle\phi_1|)|g\rangle. \tag{3.37}$$

Equations (3.36) and (3.37) give the natural metric for use with $|f\rangle$ and with $|g\rangle$. Occurrence of the operator $I + |\phi_1\rangle\langle\phi_1|$ is a manifestation of the well-known fact that bosons prefer to be in the same state. The operator $I - |\phi_1\rangle\langle\phi_1|$ is a manifestation of the Pauli exclusion principle; it arises because only the part of $|g\rangle$ which is orthogonal to $|\phi_1\rangle$ has any significance. An alternative method of calculating the result of applying $V^{1/2}PV^{1/2}$ to functions of the form ψ_S or ψ_A, which was communicated to me by W. Hunziker after he had read a preprint of the abreviated Phys. Rev. Lett.[2] version of the present proof, is outlined in Appendix B.

Equations (3.32) and (3.33) can be thought of as approximate Schrödinger equations for the "outer" electron, where the approximation has been constructed so that the eigenvalues of (3.32) are lower bounds to the eigenvalues of the original problem in the singlet sector (symmetric spatial wavefunction and antisymmetric spin function), with the eigenvalues of (3.33) lower bounds in the triplet sector (antisymmetric spatial wavefunction and symmetric spin function). The terms $k_d \pm k_e$ are a kind of "effective interaction" with the inner electron. All of the terms in k_e and k_d except U fall off exponentially at large distances due to the presence of the hydrogenic ground state function $\phi_1(r)$. Carrying out the integration in the definition (3.10) of U shows that

$$U(r) = \{\tfrac{1}{2}r + (2Z^2r)^{-1} - [(4Z)^{-1} + (2Z^2r)^{-1}]\exp(-2Zr)\}^{-1}. \tag{3.38}$$

Combining the large $-r$ expansion of (3.38) with $-ZV_0$ and setting $Z = 1$ yields

$$[-ZV_0(r) + U(r)]_{Z=1} = -2r^{-3} + O(r^{-5}). \tag{3.39}$$

Thus the outer electron sees, in the present approximation, an attractive r^{-3} potential at large r. The r^{-3} falloff is fast enough to avoid the infinite number of bound states associated with an attractive Coulomb potential, but is not the r^{-4} falloff of the induced dipole potential which the outer electron should see at large r in the original physical problem. An r^{-3} falloff instead of the physically correct r^{-4} arises because polarization is not properly treated when the inner electron is forced to remain in the hydrogenic ground state. Nevertheless, the present approximation *is* good enough to count the number of bound states correctly, as will be shown in the next section.

IV. THE NUMBER OF BOUND STATES FOR $Z = 1$ AND $\gamma = 0$

This section will show that, for $Z = 1$, the singlet equation (3.32) has at most one bound state and the

triplet equation (3.33) has none. This will in turn imply that the singlet state of H^- has at most one bound state and the triplet has none in the fixed (infinite mass) nucleus approximation with Coulomb interactions only. The analysis of (3.32) and (3.33) will be carried out by first discarding some terms with nonnegative expectation values to simplify the equations. A coupling constant will then be introduced; the number of bound states will be studied as a function of the coupling constant by sitting at the bottom of the continuum and counting the bound states as they emerge from the continuum when the coupling constant is increased.[29]

A. The singlet equation

The singlet equation (3.32) will be considered first. It follows from (3.28) and (3.29) that

$$\begin{aligned}&\tfrac{1}{2}Z^2|\phi_1\rangle\langle\phi_1| + k_d + k_e \\ &\quad = U + \{[\tfrac{1}{2}Z^2 - (16Z/35)]|\phi_1\rangle\langle\phi_1| \\ &\qquad + \sum_{n=1}^{\infty} U^{1/2}M^nU^{1/2}\}. \end{aligned} \tag{4.1}$$

Because all eigenvalues of M are real and nonnegative, the term in curly brackets in (4.1) has a nonnegative expectation value for $Z=1$. Furthermore, for $Z=1$, it follows from (3.35) and (3.38) that

$$\begin{aligned}-V_0(r) + U(r) &\geq -2r^{-1} + 2(r + r^{-1})^{-1} \\ &= -4r^{-1}[(1+r)^2 + (1-r)^2]^{-1} \\ &\geq -4r^{-1}(1+r)^{-2} \equiv -U_1(r). \end{aligned} \tag{4.2}$$

Hence the terms $-ZV_0 + \frac{1}{2}Z^2|\phi_1\rangle\langle\phi_1| + k_d + k_e$ in (3.32) can, as a consequence of Theorem 1, be replaced by $-U_1$ without raising the eigenvalues or decreasing the number of bound states. Making this replacement and introducing a coupling constant λ yields

$$\begin{aligned}&(I + |\phi_1\rangle\langle\phi_1|)(p^2 - \lambda U_1)(I + |\phi_1\rangle\langle\phi_1|)|f\rangle. \\ &\quad = (E+1)(I + |\phi_1\rangle\langle\phi_1|)|f\rangle. \end{aligned} \tag{4.3}$$

The number of bound states of (4.3) for $\lambda = 1$ is thus an upper bound to the number of bound states of (3.32) for $Z=1$. The number of bound states of (4.3) will be counted by the standard trick of sitting at the bottom of the continuum, cranking up the coupling constant λ, and counting the bound states as they emerge from the bottom of the continuum. One lets E approach -1 from below and regards λ as the eigenvalue; the number of eigenvalues λ which do not exceed 1 is then the same as the number of bound states for $\lambda = 1$. The justification for this trick rests on Theorem 1, which, since expectation values of U_1 must be nonnegative, implies that the number of bound states of (4.3) is a nondecreasing function of λ and that the eigenvalues E of (4.3) are nonincreasing functions of λ.

The configuration space representatives of bound state eigenvectors of (4.3) must decrease exponentially at large r. The fate of this boundary condition as $E \to -1$ from below can be seen by rewriting (4.3) as the (integral) equation

$$|f\rangle = \lambda G U_1(I + |\phi_1\rangle\langle\phi_1|)|f\rangle, \tag{4.4}$$

where G is the inverse operator (Green's function)

$$G = [p^2(I + |\phi_1\rangle\langle\phi_1|) - (E+1)I]^{-1}, \tag{4.5}$$

which is well defined for $E < -1$. It is easy to show that

$$\begin{aligned}G = G_0 + [1 + \tfrac{1}{2}(E+1)\langle\phi_1|G_0|\phi_1\rangle]^{-1}[-\tfrac{1}{2}|\phi_1\rangle\langle\phi_1|G_0 \\ - \tfrac{1}{2}(E+1)G_0|\phi_1\rangle\langle\phi_1|G_0], \end{aligned} \tag{4.6}$$

where

$$G_0 = [p^2 - (E+1)I]^{-1}. \tag{4.7}$$

But for $E < -1$, the configuration space representative of G_0 is

$$G_0(\mathbf{r}, \mathbf{r}') = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_{l,m}(\theta,\phi)\overline{Y_{l,m}(\theta',\phi')}g_l(r,r'), \tag{4.8}$$

where $Y_{l,m}$ is a spherical harmonic and

$$g_l(r,r') = (rr')^{-1/2}I_{l+1/2}(\kappa r_<)K_{l+1/2}(\kappa r_>), \tag{4.9}$$

with $r_<$ the smaller of the pair (r,r'), $r_>$ the larger of the pair (r,r'), and $I_{l+1/2}$, $K_{l+1/2}$, modified Bessel functions of the first and third kinds respectively.[30] κ is defined by

$$\kappa = [-(E+1)]^{1/2}. \tag{4.10}$$

As $E \to -1$ and $\kappa \to 0$,

$$I_{l+1/2}(\kappa r_<) \to [\Gamma(l+\tfrac{3}{2})]^{-1}(\kappa r_</2)^{l+1/2},$$

and

$$K_{l+1/2}(\kappa r_>) \to \tfrac{1}{2}\Gamma(l+\tfrac{1}{2})(\kappa r_>/2)^{-l-1/2},$$

so that

$$g_l(r,r') \to (2l+1)^{-1}(r_<^l/r_>^{l+1}). \tag{4.11}$$

Because $G \to (I - \frac{1}{2}|\phi_1\rangle\langle\phi_1|)G_0$ as $E \to -1$, Eqs. (4.4) and (4.11) imply that the radial part of a bound state wavefunction in the lth partial wave must behave like r^{-l-1} for large r when E is at the bottom of the continuum at -1.

When $E=-1$, it is convenient to simultaneously decompose (4.3) into partial waves and eliminate the $|\phi_1\rangle\langle\phi_1|$ terms (which contribute only in the $l=0$ partial wave) by putting

$$f(\mathbf{r}) + \phi_1(r)\int\overline{\phi_1(r')}f(\mathbf{r}')d^3\mathbf{r}' = R(r)Y_{l,m}(\theta,\phi). \tag{4.12}$$

R then satisfies the ordinary differential equation

$$\frac{d^2}{dr^2}[rR(r)] + \left(\frac{4\lambda}{r(1+r)^2} - \frac{l(l+1)}{r^2}\right)[rR(r)] = 0. \tag{4.13}$$

Because this radial equation has regular singular points at 0, -1, and ∞ it is transformable into the hypergeometric equation [this, of course, is the reason for employing the inequalities (4.2)]. The change of variables $r = z/(1-z)$, $R(r) = r^l w(z)$ brings it to the standard form of the hypergeometric equation

$$z(1-z)\frac{d^2w}{dz^2} + 2(l+1-z)\frac{dw}{dz} + 4\lambda w = 0. \tag{4.14}$$

As $r \to 0$, $R(r)$ must behave like r^l; the solution which behaves like r^{-l-1} as $r \to 0$ must be rejected to avoid a source at $r=0$. Hence $w(z)$ must approach a constant as $z \to 0$; the solution of the hypergeometric equation which behaves like z^{-2l-1} as $z \to 0$ must be rejected. Therefore, w is given, in standard hypergeometric notation,[31] by

$$w(z) = {}_2F_1(a, b; c; z), \tag{4.15}$$

where

$$c = 2l+2, \quad a+b=1, \quad ab = -4\lambda. \tag{4.16}$$

As has been argued above, $R(r)$ must behave like r^{-l-1} as $r \to \infty$; the solution which behaves like r^l as $r \to \infty$ must be rejected. Hence $w(z)$ must behave like $(1-z)^{2l+1}$ as $z \to 1$; the solution of the hypergeometric equation which approaches a constant as $z \to 1$ must be rejected. One of the linear transformations for the hypergeometric function[32] yields

$$\begin{aligned} &{}_2F_1(a, b; a+b+2l+1; z) \\ &= \frac{\Gamma(2l+1)\Gamma(a+b+2l+1)}{\Gamma(a+2l+1)\Gamma(b+2l+1)} \sum_{n=0}^{2l} \frac{(a)_n (b)_n (n-2l)!}{n!\,(2l)!} (1-z)^n \\ &\quad - \frac{\Gamma(a+b+2l+1)}{\Gamma(a)\Gamma(b)} (z-1)^{2l+1} \sum_{n=0}^{\infty} \frac{(a+2l+1)_n (b+2l+1)_n}{n!\,(n+2l+1)!} \\ &\quad \times (1-z)^n [\ln(1-z) - \psi(n+1) + \psi(a+2l+n+1) \\ &\quad + \psi(b+2l+n+1) - \psi(n+2l+2)]. \end{aligned} \tag{4.17}$$

Here $(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1) = \Gamma(\alpha+n)/\Gamma(\alpha)$ is Pochhammer's symbol. Equation (4.17) makes it clear that the boundary condition at $z=1$ can be satisfied only if either $a+2l+1$ or $b+2l+1$ is a negative integer or zero. Without loss of generality, then, $a = -2l-1-n$, $n = 0, 1, 2, \cdots$. Equations (4.16) then imply that the eigenvalues λ are

$$\lambda = \tfrac{1}{4}(n+2l+1)(n+2l+2), \quad n = 0, 1, 2, \cdots. \tag{4.18}$$

Since only one of these eigenvalues is less than 1, (4.3), and therefore also (3.32) and the singlet state of H^-, has at most one bound state. Since the singlet sector is already known to have at least one bound state, this completes the proof that the singlet state of H^- has exactly one bound state in the fixed (infinite mass) nucleus approximation with Coulomb interactions only.

B. The triplet equation

The triplet equation (3.33) will now be considered. It follows from (3.28) and (3.29) that

$$\begin{aligned} k_d - k_e &= U - U^{1/2} M U^{1/2} - (16Z/35) |\phi_1\rangle\langle\phi_1| \\ &\quad + \sum_{n=1}^{\infty} U^{1/2} M^n (I-M) M^n U^{1/2}. \end{aligned} \tag{4.19}$$

Because all eigenvalues of M are real, nonnegative, and strictly less than 1, the sum in (4.19) has a nonnegative expectation value and can therefore be discarded without raising eigenvalues or increasing the number of bound states. The term $-(16Z/35)|\phi_1\rangle\langle\phi_1|$ in (4.19) does not contribute when (4.19) is inserted in (3.33) because of the presence of the projection operator $I - |\phi_1\rangle\langle\phi_1|$ in (3.33). Discarding these terms and introducing a coupling constant λ yields

$$\begin{aligned} &(I - |\phi_1\rangle\langle\phi_1|)[p^2 - \lambda(V_0 - U + U^{1/2} M U^{1/2})](I - |\phi_1\rangle\langle\phi_1|)|g\rangle \\ &\quad = (E+1)(I - |\phi_1\rangle\langle\phi_1|)|g\rangle. \end{aligned} \tag{4.20}$$

The number of bound states of (4.20) for $\lambda = 1$ is an upper bound to the number of bound states of (3.33) for $Z=1$.

Because p^2 is a differential operator in configuration space while $U^{1/2} M U^{1/2}$ is an integral operator, (4.20) is an integrodifferential equation. It is most easily handled by converting it to the integral equation

$$|g\rangle = \lambda G'(V_0 - U + U^{1/2} M U^{1/2})|g\rangle, \tag{4.21}$$

where the generalized inverse G', which is well defined for $E < -1$, is that solution of

$$\begin{aligned} &(I - |\phi_1\rangle\langle\phi_1|)[p^2 - (E+1)I](I - |\phi_1\rangle\langle\phi_1|)G' \\ &\quad = I - |\phi_1\rangle\langle\phi_1|, \end{aligned} \tag{4.22}$$

which satisfies $G'|\phi_1\rangle =$ the null vector. G' is given explicitly by

$$G' = G_0 - G_0|\phi_1\rangle(\langle\phi_1|G_0|\phi_1\rangle)^{-1}\langle\phi_1|G_0, \tag{4.23}$$

where G_0 is defined in (4.7).

The integral equation (4.21) will now be decomposed in partial waves. The partial wave expansion of G_0 has already been recorded in (4.8) and (4.9). It is straightforward to show that

$$\begin{aligned} \int G_0(\mathbf{r}, \mathbf{r}')\phi_1(r')\,d^3\mathbf{r}' &= 2\pi^{-1/2}\{(1-\kappa^2)^{-2} r^{-1} \exp(-\kappa r) \\ &\quad - [\tfrac{1}{2}(1-\kappa^2)^{-1} + (1-\kappa^2)^{-2} r^{-1}] \\ &\quad \times \exp(-r)\}, \end{aligned} \tag{4.24}$$

and that

$$\langle\phi_1|G_0|\phi_1\rangle = (1+\kappa)^{-4}(5+4\kappa+\kappa^2). \tag{4.25}$$

It follows from (4.8), (4.9), and (4.23)–(4.25) that G' has the partial wave expansion

$$G'(\mathbf{r}, \mathbf{r}') = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_{l,m}(\theta, \phi)\overline{Y_{l,m}(\theta', \phi')}\, g_l'(r, r'), \tag{4.26}$$

where

$$g_l'(r,r') = (rr')^{-1/2} I_{l+1/2}(\kappa r_<) K_{l+1/2}(\kappa r_>)$$
$$- 16(5+4\kappa+\kappa^2)^{-1}\{(1-\kappa)^{-2} r^{-1}\exp(-\kappa r)$$
$$- [\tfrac{1}{2}(1+\kappa)(1-\kappa)^{-1}$$
$$+ (1-\kappa)^{-2} r^{-1}]\exp(-r)\}\{(1-\kappa)^{-2}(r')^{-1}\exp(-\kappa r')$$
$$- [\tfrac{1}{2}(1+\kappa)(1-\kappa)^{-1} + (1-\kappa)^{-2}(r')^{-1}]$$
$$\times \exp(-r')\}\,\delta_{l,0}. \tag{4.27}$$

The partial wave expansion of $U^{1/2}MU^{1/2}$ follows from (3.4), (3.17), (3.19), and the partial wave expansion of k given in Appendix A. It is

$$[U(r)]^{1/2} M(\mathbf{r},\mathbf{r}')[U(r)]^{1/2} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_{l,m}(\theta,\phi) \times \overline{Y_{l,m}(\theta',\phi')} N_l(r,r'), \tag{4.28}$$

where

$$N_l(r,r') = 2(2l+1)^{-1} U(r)\exp(-r)[(2l-1)^{-1} r_<^l r_>^{-l+1},$$
$$- (2l+3)^{-1} r_<^{l+2} r_>^{-l-1}]\exp(-r')U(r')$$
$$+ \tfrac{128}{35}\exp(-r-r')\delta_{l,0}. \tag{4.29}$$

The integral equation (4.21) can be decomposed in partial waves with the aid of the partial wave expansions (4.26)–(4.29). Put $g(\mathbf{r}) = R(r)Y_{l,m}(\theta,\phi)$. R then satisfies the radial (integral) equation

$$R(r) = \lambda\int_0^{\infty}\{g_l'(r,r')[V_0(r') - U(r')] + \int_0^{\infty} g_l'(r,r'')N_l(r'',r')r''^2\,dr''\}R(r')r'^2\,dr'. \tag{4.30}$$

The integral equation (4.30) does not have a symmetric kernel, which means that the Hilbert–Schmidt theory of integral equations[33] cannot be applied directly to (4.30). Integral equations with unsymmetric kernels are also harder to handle numerically, because conversion to a matrix problem via a numerical integration rule[34] leads to an eigenvalue problem for an unsymmetric matrix; the diagonalization of an unsymmetric matrix on a computer requires considerably more computer time than the diagonalization of a symmetric matrix of the same size. For these reasons it is desirable to transform (4.30) into a symmetric Hilbert–Schmidt equation. This can be done by finding a kernel η_l with adjoint $\eta_l^\dagger$ such that

$$rg_l'(r,r')r' = \int_0^{\infty}\eta_l^\dagger(r,r'')\eta_l(r'',r')\,dr''. \tag{4.31}$$

Such kernels η_l (which are not unique) are found explicitly in Appendix C. Only the existence of η_l is needed for the following argument. Given η_l, define ρ by

$$\rho(r) \equiv \lambda\int_0^{\infty}\{\eta_l(r,r')[V_0(r') - U(r')] + \int_0^{\infty}\eta_l(r,r'')[r''N_l(r'',r')r']\,dr''\}r'R(r')\,dr'. \tag{4.32}$$

It follows that

$$r'R(r') = \int_0^{\infty}\eta_l^\dagger(r',r)\rho(r)\,dr, \tag{4.33}$$

and that ρ satisfies the integral equation

$$\rho(r) = \lambda\int_0^{\infty}[Q_1^{(l)}(r,r') + Q_2^{(l)}(r,r')]\rho(r')\,dr', \tag{4.34}$$

where

$$Q_1^{(l)}(r,r') = \int_0^{\infty}\eta_l^\dagger(r,r'')[V_0(r'') - U(r'')]\eta_l^\dagger(r'',r')\,dr', \tag{4.35}$$

and

$$Q_2^{(l)}(r,r') = \int_0^{\infty}\int_0^{\infty}\eta_l(r,r'')[r''N_l(r'',r''')r'''] \times \eta_l^\dagger(r''',r')\,dr''\,dr'''. \tag{4.36}$$

The kernels $Q_1^{(l)}$ and $Q_2^{(l)}$ are symmetric and square integrable, so that (4.34) *is* a Hilbert–Schmidt equation. It is straightforward to show, from (3.35) and (3.38), that $V_0(r) - U(r) \geq 0$. Furthermore, (4.28) and the fact that all eigenvalues of M are positive implies that N_l is an operator with nonnegative expectation value. Hence $Q_1^{(l)}$ and $Q_2^{(l)}$, defined by (4.35) and (4.36), are operators with nonnegative expectation values, which in turn implies that all eigenvalues of (4.34) are nonnegative. These eigenvalues are, of course, the same as the eigenvalues of (4.20).

The proof that the triplet state of H^- is not bound will now be completed by showing that, for $E \leq -1$, all eigenvalues λ of (4.34) exceed 1. This will be accomplished with the aid of a well-known eigenvalue bounding trick from the theory of Hilbert–Schmidt integral equations.[35] It is a standard result of Hilbert–Schmidt theory that the eigenvalues $\lambda_n^{(l)}$ of (4.34) satisfy the sum rule

$$\sum_{n=1}^{\infty}(\lambda_n^{(l)})^{-2} = \mathrm{Tr}[(Q_1^{(l)} + Q_2^{(l)})^2]. \tag{4.37}$$

Because all $\lambda_n^{(l)}$ are nonnegative, discarding all but the $n=1$ term on the left-hand side of (4.37) shows that the smallest eigenvalue $\lambda_1^{(l)}$ satisfies the inequality

$$\lambda_1^{(l)} \geq \{\mathrm{Tr}[(Q_1^{(l)} + Q_2^{(l)})^2]\}^{-1/2}. \tag{4.38}$$

But $\mathrm{Tr}[(Q_1^{(l)} + Q_2^{(l)})^2] = \mathrm{Tr}[(Q_1^{(l)2})] + 2\,\mathrm{Tr}(Q_1^{(l)}Q_2^{(l)}) + \mathrm{Tr}[(Q_2^{(l)})^2]$. The Schwarz inequality implies that $\mathrm{Tr}(Q_1^{(l)}Q_2^{(l)}) \leq \{\mathrm{Tr}[(Q_1^{(l)})^2]\,\mathrm{Tr}[(Q_2^{(l)})^2]\}^{1/2}$. Since all eigenvalues of $Q_2^{(l)}$ are nonnegative, $\mathrm{Tr}[(Q_2^{(l)})^2] \leq (\mathrm{Tr}Q_2^{(l)})^2$. Using all of these inequalities in (4.38) yields

$$\lambda_1^{(l)} \geq \lambda_{1bd}^{(l)} \equiv \{[\mathrm{Tr}((Q_1^{(l)})^2)]^{1/2} + \mathrm{Tr}Q_2^{(l)}\}^{-1}. \tag{4.39}$$

It is straightforward to show that $\mathrm{Tr}(Q_1^{(l)})^2 \leq \mathrm{Tr}(Q_1^{(1)})^2$ and $\mathrm{Tr}Q_2^{(l)} \leq \mathrm{Tr}Q_2^{(1)}$ for $l \geq 1$, so that

$$\lambda_1^{(l)} \geq \lambda_{1bd}^{(1)} \quad \text{for} \quad l \geq 1. \tag{4.40}$$

Thus it is necessary to calculate the traces explicitly only for $l = 0, 1$. It is also straightforward to show, with the aid of Theorem 1, that $\lambda_1^{(l)}$ is a nonincreasing function of E. Thus it is necessary to calculate the traces explicitly only for $E = -1$.

Explicit calculation of these traces can be simplified

by using invariance of the trace under cyclic permutation, which here amounts to a (justifiable) interchange of the orders of integration, to collect η_l and $\eta_l^\dagger$ together in the form $\eta_l^\dagger\eta_l = r g_l' r$. Thus, for example, $\mathrm{Tr}Q_2^{(1)} = \mathrm{Tr}(\eta_l r N_l r \eta_l^\dagger) = \mathrm{Tr}(\eta_l^\dagger \eta_l r N_l r) = \mathrm{Tr}(r g_l' r^2 N_l r)$. Numerical integration for $E=-1$ shows that $\mathrm{Tr}[(Q_1^{(0)})^2] \approx 0.428239$, $\mathrm{Tr}Q_2^{(0)} \approx 0.1162229$, $\mathrm{Tr}[(Q_1^{(1)})^2] \approx 0.011974739$, and $\mathrm{Tr}Q_2^{(1)} \approx 0.013762233$. It follows that

$$\lambda_1^{(0)} \geq \lambda_{1bd}^{(0)} \approx 1.297652, \quad \text{and} \tag{4.41}$$

$$\lambda_1^{(l)} \geq \lambda_{1bd}^{(1)} \approx 2.0675353, \quad l \geq 1. \tag{4.42}$$

The bounds (4.41) and (4.42) imply that (4.20), and therefore also (3.33) and the triplet state of H^- in the fixed (infinite mass) nucleus approximation with Coulomb interactions only, have no bound states.

V. FINITE NUCLEAR MASS

This section will extend the results of the preceding sections to finite nuclear mass [$\gamma \neq 0$ in Eq. (2.9)]. Before outlining the method to be used, it is in order to ask what one can reasonably expect to prove. Note that interchange of the electron mass m and the nuclear mass M puts γ into $1-\gamma$ and transforms the internal Hamiltonian H_{int} into the internal Hamiltonian of the hydrogen molecular ion H_2^+. The H_2^+ ion, which is well described by the adiabatic or Born–Oppenheimer approximations, has a number of (rotational and vibrational) bound states. This makes it clear that one cannot expect to extend the result that H^- has at most one bound state to finite nuclear mass for all values of the electron to nuclear mass ratio; as γ increases from zero, there should be some value of γ at which a second bound state appears. The estimates used in the present section are sufficient to show that H^- has one bound state for $(m/M) < (m/M)_{\max} \approx 0.210106366$ (for $\gamma < \gamma_S \approx 0.1736263618$). This does not mean that a second bound state comes in for m/M greater than this limit; it only means that the estimates of the present paper are not sufficiently good to exclude a second bound state for $m/M > (m/M)_{\max}$. Complementary variational (Rayleigh–Ritz) upper bounds to eigenvalues together with lower bounds such as those of the present paper are needed to get rigorous estimates of the value of m/M for which a second bound state appears. The present section begins with an outline of the method used to get lower bounds for $(m/M) > 0$. Detailed calculations to show that H^- has at most one bound state are then presented. The section concludes with a simple variational calculation to show that H^- has at least one bound state for $m/M > 0$.

A. The basic idea

The basic idea to be used to extend the results to finite nuclear mass is most easily described with the aid of the following projection operators:

$$\begin{aligned} P_S(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') &= \tfrac{1}{2}[\delta(\mathbf{r}_1 - \mathbf{r}_1')\phi_1(r_2)\overline{\phi_1(r_2')} \\ &\quad + \delta(\mathbf{r}_2 - \mathbf{r}_1')\phi_1(r_1)\overline{\phi_1(r_2')} + \delta(\mathbf{r}_2 - \mathbf{r}_2')\phi_1(r_1)\overline{\phi_1(r_1')} \\ &\quad + \delta(\mathbf{r}_1 - \mathbf{r}_2')\phi_1(r_2)\overline{\phi_1(r_1')}] - \phi_1(r_1)\phi_1(r_2)\overline{\phi_1(r_1')}\overline{\phi_1(r_2')}, \end{aligned} \tag{5.1}$$

$$\begin{aligned} P_A(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') &= \tfrac{1}{2}[\delta(\mathbf{r}_1 - \mathbf{r}_1')\phi_1(r_2)\overline{\phi_1(r_2')} \\ &\quad - \delta(\mathbf{r}_2 - \mathbf{r}_1')\phi_1(r_1)\overline{\phi_1(r_2')} + \delta(\mathbf{r}_2 - \mathbf{r}_2')\phi_1(r_1)\overline{\phi_1(r_1')} \\ &\quad - \delta(\mathbf{r}_1 - \mathbf{r}_2')\phi_1(r_2)\overline{\phi_1(r_1')}], \end{aligned} \tag{5.2}$$

$$P_{\|}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') = P_S(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') + P_A(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_1', \mathbf{r}_2'), \tag{5.3}$$

and

$$\begin{aligned} P_\perp(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') &= [\delta(\mathbf{r}_1 - \mathbf{r}_1') - \phi_1(r_1)\overline{\phi_1(r_1')}] \\ &\quad \times [\delta(\mathbf{r}_2 - \mathbf{r}_2') - \phi_1(r_2)\overline{\phi_1(r_2')}]. \end{aligned} \tag{5.4}$$

P_S, P_A, and $P_\perp$ project onto the spaces $\mathcal{S}_S$, $\mathcal{S}_A$, and $\mathcal{S}_\perp$ defined in Sec. III D. $P_{\|}$ projects onto the span of the ranges of P_S and P_A, which will be called $\mathcal{S}_{\|}$. The basic difficulty to be overcome when $m/M \neq 0$ is the fact that the Hughes–Eckart term $2\gamma \mathbf{p}_1 \cdot \mathbf{p}_2$ in H_{int} couples $\mathcal{S}_{\|}$ and $\mathcal{S}_\perp$. This difficulty will be handled by splitting H_{int} as follows[36]:

$$\begin{aligned} H_{\mathrm{int}} &= P_{\|}H_{\mathrm{int}}P_{\|} + P_{\|}H_{\mathrm{int}}P_\perp + P_\perp H_{\mathrm{int}}P_{\|} \\ &\quad + P_\perp H_{\mathrm{int}}P_\perp. \end{aligned} \tag{5.5}$$

$P_\perp H_{\mathrm{int}} P_\perp$ will be replaced by $BP_\perp$ where B is a lower bound to the spectrum of $P_\perp H_{\mathrm{int}} P_\perp$; Theorem 1 guarantees that this replacement cannot raise eigenvalues or decrease the number of bound states. The Coulomb repulsion will be handled as in Sec. III. The resulting Hamiltonian H_{int}' is

$$\begin{aligned} H_{\mathrm{int}}' &= P_{\|}(H_0 + 2\gamma \mathbf{p}_1 \cdot \mathbf{p}_2 + V^{1/2}PV^{1/2})P_{\|} \\ &\quad + 2\gamma(P_{\|}\mathbf{p}_1 \cdot \mathbf{p}_2 P_\perp + P_\perp \mathbf{p}_1 \cdot \mathbf{p}_2 P_{\|}) + BP_\perp. \end{aligned} \tag{5.6}$$

The Schrödinger equation $H_{\mathrm{int}}'|\psi\rangle = E|\psi\rangle$ splits into two equations:

$$\begin{aligned} &P_{\|}(H_0 + 2\gamma \mathbf{p}_1 \cdot \mathbf{p}_2 + V^{1/2}PV^{1/2})P_{\|}|\psi\rangle + 2\gamma P_{\|}\mathbf{p}_1 \cdot \mathbf{p}_2 P_\perp |\psi\rangle \\ &\quad = EP_{\|}|\psi\rangle, \end{aligned} \tag{5.7}$$

and

$$2\gamma P_\perp \mathbf{p}_1 \cdot \mathbf{p}_2 P_{\|}|\psi\rangle + BP_\perp|\psi\rangle = EP_\perp|\psi\rangle. \tag{5.8}$$

Equation (5.8) can be readily solved for $P_\perp|\psi\rangle$ if $E \neq B$. Inserting the result in (5.7) yields

$$\begin{aligned} &P_{\|}[H_0 + 2\gamma \mathbf{p}_1 \cdot \mathbf{p}_2 + V^{1/2}PV^{1/2} \\ &\quad - 4\gamma^2(B-E)^{-1}\mathbf{p}_1 \cdot \mathbf{p}_2 P_\perp \mathbf{p}_1 \cdot \mathbf{p}_2]P_{\|}|\psi\rangle \\ &\quad = EP_{\|}|\psi\rangle. \end{aligned} \tag{5.9}$$

The Schrödinger equation (5.9) reduces to one-particle equations in essentially the same way as the Schrödinger equation which arose in the fixed (infinite mass) nucleus case. If $E=B$, one obtains B as an eigenvalue of infinite multiplicity, with any vector $|\psi\rangle$ which satisfies $P_{\|}|\psi\rangle = P_{\|}\mathbf{p}_1 \cdot \mathbf{p}_2 P_\perp |\psi\rangle =$ the null vector as an eigenvector.

B. The lower bound B

This subsection will derive the lower bound B to the spectrum of $P_\perp H_{\mathrm{int}} P_\perp$. Clearly this bound B must lie above the bottom of the continuum at $-Z^2$ if the method outlined in the preceding section is to work. As will be seen, the methods used in this section will yield a bound B which satisfies $B > -Z^2$ only for $\gamma < \gamma_b$ where

$$\gamma_b \approx 0.330485327. \tag{5.10}$$

Construction of the lower bound B is facilitated by the definitions

$$h_\gamma(\mathbf{r},\mathbf{r}') \equiv [-(1-\gamma)\nabla^2 - 2Zr^{-1}]\delta(\mathbf{r}-\mathbf{r}'), \tag{5.11}$$

and

$$h'_\gamma(\mathbf{r},\mathbf{r}') \equiv -Z^2(1-\gamma)^{-1}[\tfrac{3}{4}\phi_{1-\gamma}(r)\overline{\phi_{1-\gamma}(r')} + \tfrac{1}{4}\delta(\mathbf{r}-\mathbf{r}')], \tag{5.12}$$

where

$$\phi_{1-\gamma}(r) \equiv (1-\gamma)^{-3/2}\phi_1[r/(1-\gamma)], \tag{5.13}$$

with ϕ_1 defined by (3.4). $\phi_{1-\gamma}$ is the ground state eigenfunction of the hydrogenic Hamiltonian h_γ, with ground state energy $-Z^2(1-\gamma)^{-1}$. The expectation values of h_γ and of h'_γ with respect to $\phi_{1-\gamma}$ are the same; the expectation value of h'_γ with respect to any wavefunction perpendicular to $\phi_{1-\gamma}$ is $-Z^2/[4(1-\gamma)]$, which is the energy of the first excited state of h_γ. It follows from the variational principle that $\langle\psi|h_\gamma|\psi\rangle \geq \langle\psi|h'_\gamma|\psi\rangle$ for all $|\psi\rangle$.[37]

Equations (2.9)—(2.11) imply that

$$H_{\rm int}(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}'_1,\mathbf{r}'_2) = H''_{\rm int}(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}'_1,\mathbf{r}'_2) + H'''_{\rm int}(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}'_1,\mathbf{r}'_2), \tag{5.14}$$

where

$$H''_{\rm int}(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}'_1,\mathbf{r}'_2) = h'_\gamma(\mathbf{r}_1,\mathbf{r}'_1)\delta(\mathbf{r}_2-\mathbf{r}'_2) + \delta(\mathbf{r}_1-\mathbf{r}'_1)h'_\gamma(\mathbf{r}_2,\mathbf{r}'_2), \tag{5.15}$$

and

$$\begin{aligned} H'''_{\rm int}&(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}'_1,\mathbf{r}'_2) \\ &= (h_\gamma - h'_\gamma)(\mathbf{r}_1,\mathbf{r}'_1)\delta(\mathbf{r}_2-r'_2) \\ &\quad + \delta(\mathbf{r}_1-\mathbf{r}'_1)(h_\gamma - h'_\gamma)(\mathbf{r}_2,\mathbf{r}'_2) + [-\gamma(\nabla_1+\nabla_2)^2 \\ &\quad + 2|\mathbf{r}_1-\mathbf{r}_2|^{-1}]\delta(\mathbf{r}_1-\mathbf{r}'_1)\delta(\mathbf{r}_2-\mathbf{r}'_2). \end{aligned} \tag{5.16}$$

Because the expectation value of $H'''_{\rm int}$ cannot be negative, the lowest eigenvalue of $P_\perp H''_{\rm int}P_\perp$ will furnish a lower bound B to the spectrum of $P_\perp H_{\rm int}P_\perp$. The eigenvalue problem for $P_\perp H''_{\rm int}P_\perp$ is separable; its eigenvalues can be obtained from the eigenvalues of

$$(I - |\phi_1\rangle\langle\phi_1|)h'_\gamma(I - |\phi_1\rangle\langle\phi_1|)|\psi\rangle = b|\psi\rangle. \tag{5.17}$$

The lowest eigenvalue b_1 of (5.17) is readily found with the aid of (5.12); it is

$$b_1 = -Z^2(1-\gamma)^{-1}[1 - \tfrac{3}{4}|\langle\phi_1|\phi_{1-\gamma}\rangle|^2]. \tag{5.18}$$

The desired lower bound B is equal to $2b_1$. Evaluation of the overlap integral $\langle\phi_1|\phi_{1-\gamma}\rangle$ in (5.18) then leads to

$$B = -2Z^2(1-\gamma)^{-1}[1 - 48(1-\gamma)^3(2-\gamma)^{-6}]. \tag{5.19}$$

It is easy to show that B is a monotone decreasing function of γ for $0 \leq \gamma < 1$. Numerical methods show that the root γ_b of $B = -Z^2$ is given by (5.10); the inequality $B > -Z^2$ is satisfied for $\gamma < \gamma_b$.

C. The singlet state

The Schrödinger equation (5.9) is equivalent to two one-particle equations, one for the singlet state (from $\mathcal{S}_S$) and one for the triplet (from $\mathcal{S}_A$). The reduction procedes in essentially the same way as in Sec. III D; the one-particle equation which replaces (3.32) is

$$\begin{aligned} (I + |\phi_1\rangle\langle\phi_1|)&\{p^2 - ZV_0 + \tfrac{1}{2}Z^2|\phi_1\rangle\langle\phi_1| + k_d + k_e \\ &+ 6\gamma Z^{-2}T_1 - 4\gamma^2(B-E)^{-1}[\tfrac{1}{3}Z^2p^2 + T_0 - \tfrac{1}{6}Z^4|\phi_1\rangle\langle\phi_1| \\ &- T_1 + T_2]\}(I + |\phi_1\rangle\langle\phi_1|)|f\rangle \\ &= (E+Z^2)(I + |\phi_1\rangle\langle\phi_1|)|f\rangle, \end{aligned} \tag{5.20}$$

where

$$T_0 \equiv \tfrac{1}{3}(I - |\phi_1\rangle\langle\phi_1|)p^2|\phi_1\rangle\langle\phi_1|p^2(I - |\phi_1\rangle\langle\phi_1|), \tag{5.21}$$

$$T_1 \equiv \tfrac{1}{3}Z^2\sum_{i=1}^{3} p_i|\phi_1\rangle\langle\phi_1|p_i, \tag{5.22}$$

$$T_2 \equiv -\tfrac{1}{3}p^2|\phi_1\rangle\langle\phi_1|p^2 + \sum_{i=1}^{3}\sum_{j=1}^{3} p_ip_j|\phi_1\rangle\langle\phi_1|p_jp_i. \tag{5.23}$$

Here p_i is the ith Cartesian component of the momentum operator $\mathbf{p} = -i\nabla$. It should be noted that the finite rank operators T_i have nonnegative expectation values, and that they can be written out explicitly in coordinate space in the form

$$T_l(\mathbf{r},\mathbf{r}') = t_l(r,r')\sum_{m=-l}^{l} Y_{l,m}(\theta,\phi)\overline{Y_{l,m}(\theta',\theta')}, \tag{5.24}$$

where

$$t_0(r,r') = (16\pi Z^2/3)(r^{-1}-Z)\phi_1(r)\phi_1(r')(r'^{-1}-Z), \tag{5.25}$$

$$t_1(r,r') = (4\pi Z^4/9)\phi_1(r)\phi_1(r'), \tag{5.26}$$

and

$$t_2(r,r') = (8\pi Z^2/15)(r^{-1}+Z)\phi_1(r)\phi_1(r')(r'^{-1}+Z). \tag{5.27}$$

The analysis of (5.20) is similar to the analysis of Sec. IV. Setting $Z=1$, discarding certain terms which have nonnegative expectation values [note that $(B-E)^{-1}$ is positive for $E \leq -1$ and $\gamma < \gamma_b$], and introducing a coupling constant λ changes (5.20) into

$$\begin{aligned} (I + |\phi_1\rangle\langle\phi_1|)&\{[1 - (4\gamma^2/3)(B-E)^{-1}]p^2 \\ &- \lambda[U_1 + 4\gamma^2(B-E)^{-1}(T_0+T_2)]\}(I + |\phi_1\rangle\langle\phi_1|)|f\rangle. \\ &= (E+1)(I + |\phi_1\rangle\langle\phi_1|)|f\rangle. \end{aligned} \tag{5.28}$$

Equation (5.28), which replaces (4.3) for $\gamma \neq 0$, is easily handled by converting it to an integral equation. Proceeding as in the derivation of (4.4) and letting $E \to -1$ yields

$$\begin{aligned} (I + |\phi_1\rangle\langle\phi_1|)|f\rangle = \lambda[1 - (4\gamma^2/3)(B+1)^{-1}]^{-1}\,&G_0[U_1 + 4\gamma^2 \\ &\times (B+1)^{-1}(T_0+T_2)](I + |\phi_1\rangle\langle\phi_1|)|f\rangle. \end{aligned} \tag{5.29}$$

Equation (5.29) can be decomposed in partial waves by giving f the form (4.12) and using the partial wave decompositions (4.8) and (5.24). The resulting radial integral equation do not have symmetric kernels, but can be transformed into equations with symmetric kernels with the aid of (C1) in essentially the same way as (4.30) was transformed into the symmetric equation (4.34) with the aid of (4.31). The result is

$$rR(r)=\int_0^\infty \zeta_l^\dagger(r,s)\rho(s)\,ds, \tag{5.30}$$

where ρ satisfies the symmetric Hilbert–Schmidt integral equation

$$\rho(r)=\lambda[1-(4\gamma^2/3)(B+1)^{-1}]^{-1}\int_0^\infty [S_1^{(l)}(r,r') + S_2^{(l)}(r,r')]\rho(r')\,dr', \tag{5.31}$$

with

$$S_1^{(l)}(r,r')=\int_0^\infty \zeta_l(r,r'')U_1(r'')\zeta_l^\dagger(r'',r')\,dr'', \tag{5.32}$$

and

$$S_2^{(l)}(r,r')=4\gamma^2(B+1)^{-1}\int_0^\infty\int_0^\infty \zeta_l(r,r'')r''[t_0(r'',r''')\delta_{l,0} + t_2(r'',r''')\delta_{l,2}]r'''\zeta_l^\dagger(r''',r')\,dr''\,dr'''. \tag{5.33}$$

The eigenvalues $\lambda_n^{(l,0)}$ [given by (4.18)] and eigenfunctions $u_n^{(l)}$ of the kernel $S_1^{(l)}$ are known from the analysis of the singlet case for $\gamma=0$ in Sec. IV C. The solutions to the integral equation (5.31) are therefore also known except for $l=0$ and $l=2$. Lower bounds to the eigenvalues $\lambda_n^{(l,\gamma)}$ of (5.31) for $l=0$ and $l=2$ can be readily obtained by truncation of $S_1^{(l)}$. Replace $S_1^{(l)}$ in (5.31) by $S_t^{(l)}$ where

$$\begin{aligned} S_t^{(0)}(r,r') &\equiv (1/\lambda_0^{(0,0)})u_0^{(0)}(r)u_0^{(0)}(r') \\ &\quad + (1/\lambda_1^{(0,0)})[\delta(r-r') - u_0^{(0)}(r)u_0^{(0)}(r')] \\ &= (4/3)u_0^{(0)}(r)u_0^{(0)}(r') + (2/3)\delta(r-r'), \end{aligned} \tag{5.34}$$

and

$$S_t^{(2)}(r,r')=(1/\lambda_0^{(2,0)})\delta(r-r')=(2/15)\delta(r-r'), \tag{5.35}$$

with

$$u_0^{(0)}(r)=\tfrac{1}{2}\int_0^\infty \zeta_0(r,s)U_1(s)[3^{1/2}s(1+s)^{-1}]\,ds. \tag{5.36}$$

The quantity in square brackets in Eq. (5.36) is the (properly normalized) solution $sR(s)$ of the differential eigenvalue problem (4.13) for $n=l=0$. Because $\langle\psi|S_t^{(l)}|\psi\rangle \geq \langle\psi|S_1^{(l)}|\psi\rangle$, this replacement produces new integral equations with degenerate kernels whose eigenvalues $\lambda_n^{(l,lb)}$ are lower bounds to the eigenvalues of (5.31). Integral equations with degenerate kernels can be readily solved by reducing them to eigenvalue problems for finite dimensional matrices.[38] The results for the eigenvalues $\lambda_n^{(l,\gamma)}$ of (5.31) are

$$\lambda_n^{(0,\gamma)} \geq \lambda_n^{(0,lb)} = (\mu_n^{(0)}+\tfrac{2}{3})^{-1}[1-(4\gamma^2/3)(B+1)^{-1}], \tag{5.37}$$

$$\lambda_n^{(2,\gamma)} \geq \lambda_n^{(2,lb)} = (\mu_n^{(2)}+\tfrac{2}{15})^{-1}[1-(4\gamma^2/3)(B+1)^{-1}], \tag{5.38}$$

$$\lambda_n^{(l,\gamma)}=\tfrac{1}{4}(n+l+1)(n+l+2)[1-(4\gamma^2/3)(B+1)^{-1}], \quad l\neq 0,2. \tag{5.39}$$

All but two of the $\mu_n^{(0)}$ which appear in the lower bound (5.37) are zero: the remaining two, obtained from a 2×2 matrix eigenvalue problem, are roots of the quadratic equation

$$\mu^2-\tfrac{4}{3}[1+4\gamma^2(B+1)^{-1}]\mu + (64\gamma^2/9)(B+1)^{-1}\{1-12[1-2eE_1(1)]^2\}=0, \tag{5.40}$$

where[39]

$$eE_1(1)=\int_0^\infty (1+x)^{-1}\exp(-x)\,dx = 0.596347361. \tag{5.41}$$

All but one of the $\mu_n^{(2)}$ which appear in (5.38) are zero; the remaining one is given by

$$\mu_0^{(2)}=\tfrac{8}{15}\gamma^2(B+1)^{-1}. \tag{5.42}$$

Numerical evaluation of the bounds (5.37)–(5.39) shows that: (a) all but one of the bounds $\lambda_n^{(0,lb)}$ of (5.37) exceeds 1 for $0\leq\gamma\leq\gamma_s$ where

$$\gamma_s \approx 0.173626362, \tag{5.43}$$

(b) all of the bounds $\lambda_n^{(2,lb)}$ of (5.38) exceed 1 for $0\leq\gamma\leq\gamma_{s,2}\approx 0.273244889$, and (c) all of the $\lambda_n^{(l,\gamma)}$ of (5.39) exceed 1 for $0\leq\gamma\leq\gamma_{s,1}\approx 0.240923098$. Hence (5.31), and therefore also the singlet sector of (5.9) and of the H^- ion, has at most one bound state for $0\leq\gamma\leq\gamma_s$, that is to say for $0\leq(m/M)\leq(m/M)_{\max}\cong 0.210106366$.

D. The triplet state

Reduction of the Schrödinger equation (5.9) for the triplet state proceeds in essentially the same way as the reduction of the preceding section for the singlet state; the one-particle equation which replaces (3.33) is

$$\begin{aligned} &(I-|\phi_1\rangle\langle\phi_1|)[p^2-ZV_0+k_d-k_e-6\gamma Z^{-2}T_1-4\gamma^2(B-E)^{-1} \\ &\quad\times(\tfrac{1}{3}Z^2p^2-T_0-T_1-T_2)](I-|\phi_1\rangle\langle\phi_1|)|g\rangle \\ &\quad=(E+Z^2)(I-|\phi_1\rangle\langle\phi_1|)|g\rangle, \end{aligned} \tag{5.44}$$

where T_0, T_1, and T_2 were defined in Eqs. (5.21)–(5.23). The analysis of (5.44) procedes by setting $Z=1$, discarding certain terms which have nonnegative expectation values [note that $(B-E)^{-1}$ is positive for $E\leq -1$ and $\gamma<\gamma_b$], and introducing a coupling constant λ to obtain

$$\begin{aligned} &(I-|\phi_1\rangle\langle\phi_1|)\{[1-(4\gamma^2/3)(B-E)^{-1}]p^2-\lambda(V_0-U \\ &\quad+U^{1/2}MU^{1/2}+6\gamma T_1)\}(I-|\phi_1\rangle\langle\phi_1|)|g\rangle \\ &\quad=(E+1)(I-|\phi_1\rangle\langle\phi_1|)|g\rangle, \end{aligned} \tag{5.45}$$

Equation (5.45) is analyzed in the same way as Eq. (4.20) which it replaces. The term T_1 contributes an additional kernel

$$Q_3^{(l)}(r,r')\equiv 6\gamma\int_0^\infty\int_0^\infty n_1(r,r'')[r''t_1(r'',r''')r'''] \times\eta_l^\dagger(r''',r')\,dr''\,dr''', \tag{5.46}$$

which is to be added to $Q_1^{(l)}+Q_2^{(l)}$ when constructing the analog of (4.34) for $l=1$. It is straightforward to show that

$$\mathrm{Tr}Q_3^{(l)}=2\gamma/3. \tag{5.47}$$

It follows that the lower bound (4.39) is now replaced by

$$\lambda_1^{(l)}\geq[1-(4\gamma^2/3)(B+1)^{-1}]\{[\mathrm{Tr}((Q_1^{(l)})^2)]^{1/2} + \mathrm{Tr}Q_2^{(l)}+(2\gamma/3)\delta_{l,1}\}^{-1}. \tag{5.48}$$

Numerical evaluation of the right-hand side of (5.48) with the aid of the values for $\mathrm{Tr}[(Q_1^{(0)})^2]$, $\mathrm{Tr}Q_2^{(0)}$, $\mathrm{Tr}[(Q_1^{(1)})^2]$, and $\mathrm{Tr}Q_2^{(1)}$ quoted in Sec. IV B shows that (a) $\lambda_1^{(0)}\geq 1$ for $0\leq\gamma\leq\gamma_t$ where

$$\gamma_t\approx 0.217891, \tag{5.49}$$

(b) $\lambda_1^{(1)} \geq 1$ for $0 \leq \gamma \leq \gamma_{t,1} \approx 0.24433520$, and (c) $\lambda_1^{(l)} \geq 1$ for $l \geq 2$ and $0 \leq \gamma \leq \gamma_{t,2} \approx 0.26468945$. Hence (5.45), and therefore also the triplet sector of (5.9) and of the H^- ion, have no bound states for $0 \leq \gamma < \gamma_t$. Comparing (5.10), (5.43), and (5.49) shows that $\gamma_s < \gamma_t < \gamma_b$. Thus the result that H^- has at most one bound state has been proven for $0 \leq \gamma < \gamma_s$.

E. At least one bound state

To complete the analysis and prove that H^- has exactly one bound state for $0 \leq \gamma \leq \gamma_s$, it is necessary to show that H^- has at least one bound state for finite γ. This is easily done with the aid of the so-called $(1s, 1s')$ function ψ_{SL} of Shull and Löwdin,[40] which is

$$\psi_{SL}(\mathbf{r}_1, \mathbf{r}_2) = N[\exp(-ar_1 - br_2) + \exp(-ar_2 - br_1)]. \tag{5.50}$$

Shull and Löwdin state that the optimum choice $a = 1.0392$, $b = 0.2832$ yields a value of -1.0266 for the expectation of the internal Hamiltonian H_{int} when $Z = 1$ and $\gamma = 0$. The Hughes–Eckart term $2\gamma \mathbf{p}_1 \cdot \mathbf{p}_2$ has zero expectation value with respect to ψ_{SL}. Since -1.0266 is below -1, H^- must therefore have at least one bound state for $0 \leq \gamma < 1$.

APPENDIX A: THE PARTIAL WAVE EXPANSION OF k

The generating function for the Legendre polynomials is

$$(1 + z^2 - 2z\cos\Theta)^{-1/2} = \sum_{l=0}^{\infty} z^l P_l(\cos\Theta). \tag{A1}$$

It (A1) is inserted in the identity

$$\frac{d}{dz}[z^{-1/2}(1 + z^2 - 2z\cos\Theta)^{1/2}] = \tfrac{1}{2}(z^{1/2} - z^{-3/2}) \times (1 + z^2 - 2z\cos\Theta)^{-1/2}, \tag{A2}$$

and the result is integrated term by term with respect to z, the expansion

$$(1 + z^2 - 2z\cos\Theta)^{1/2} = \sum_{l=0}^{\infty} \left(\frac{z^{l+2}}{2l+3} - \frac{z^l}{2l-1}\right) P_l(\cos\Theta), \tag{A3}$$

is obtained. The partial wave expansion of $|\mathbf{r} - \mathbf{r}'|$ follows from (A3) by letting Θ be the angle between $\mathbf{r}$ and $\mathbf{r}'$ and setting $z = r_</r_>$ where $r_<$ is the smaller of the pair r, r' and $r_>$ is the larger of r, r'. Inserting the result in the definition (3.15) of k and using the addition theorem for the spherical harmonics yields

$$k(\mathbf{r}, \mathbf{r}') = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_{l,m}(\theta, \phi)\overline{Y_{l,m}(\theta', \phi')} k_l(r, r'), \tag{A4}$$

where

$$k_l(r, r') = \frac{2Z^3}{2l+1}[U(r)]^{1/2} \exp(-Zr)\left(-\frac{r_<^l}{(2l-1)r_>^{l-1}} + \frac{r_<^{l+2}}{(2l+3)r_>^{l+1}}\right) \exp(-Zr')[U(r')]^{1/2}. \tag{A5}$$

APPENDIX B; HUNZIKER'S CONSTRUCTION OF $V^{1/2}PV^{1/2}$

This appendix will sketch an alternative derivation of $V^{1/2}PV^{1/2}$, due to Walter Hunziker, which he was kind enough to allow me to include. Consider first the subspace $\mathcal{S}_S$, spanned by functions of the form (3.30). Let ψ_{S1} and ψ_{S2} be two functions of the form (3.30):

$$\psi_{S1}(\mathbf{r}_1, \mathbf{r}_2) = f_1(\mathbf{r}_1)\phi_1(r_2) + \phi_1(r_1)f_1(\mathbf{r}_2), \tag{B1}$$

$$\psi_{S2}(\mathbf{r}_1, \mathbf{r}_2) = f_2(\mathbf{r}_1)\phi_1(r_2) + \phi_1(r_1)f_2(\mathbf{r}_2). \tag{B2}$$

The projection operator P is to be chosen so that for each f_1 there exists f_2 such that

$$V^{1/2}PV^{1/2}|\psi_{S1}\rangle = |\psi_{S2}\rangle. \tag{B3}$$

Equation (B3) is equivalent to

$$PV^{1/2}|\psi_{S1}\rangle = V^{-1/2}|\psi_{S2}\rangle. \tag{B4}$$

For any vector $|\psi\rangle$, let $\||\psi\rangle\| = \langle\psi|\psi\rangle^{1/2}$, and let I be the identity. Then (B4) implies that

$$\|V^{1/2}|\psi_{S1}\rangle - V^{-1/2}|\psi_{S2}\rangle\|^2 = \|(I-P)V^{1/2}|\psi_{S1}\rangle\|^2 + \|PV^{1/2}|\psi_{S1}\rangle - V^{-1/2}|\psi_{S2}\rangle\|^2. \tag{B5}$$

It follows that with f_1 given, f_2 can be characterized by

$$\|V^{1/2}|\psi_{S1}\rangle - V^{-1/2}|\psi_{S2}\rangle\|^2 = \text{minimum}. \tag{B6}$$

Equations (B1), (B2), (3.10), and (3.15) can be used to show that

$$\begin{aligned}\|V^{1/2}|\psi_{S1}\rangle - V^{-1/2}|\psi_{S2}\rangle\|^2 &= \langle\psi_{S1}|V|\psi_{S1}\rangle - 2\langle f_1|(I + |\phi_1\rangle\langle\phi_1|)|f_2\rangle \\ &\quad - 2\langle f_2|(I + |\phi_1\rangle\langle\phi_1|)|f_1\rangle \\ &\quad + 2\langle f_2|U^{-1}|f_2\rangle + 2\langle f_2|U^{-1/2}kU^{-1/2}|f_2\rangle.\end{aligned} \tag{B7}$$

Varying (B7) with respect to $|f_2\rangle$ yields, via a standard argument of the calculus of the variations, the following equation for $|f_2\rangle$:

$$|f_2\rangle = U(I + |\phi_1\rangle\langle\phi_1|)|f_1\rangle - U^{1/2}kU^{-1/2}|f_2\rangle. \tag{B8}$$

Equations (3.17) and (3.19) can be used to bring (B8) to the form

$$|f_2\rangle = U(I + |\phi_1\rangle\langle\phi_1|)|f_1\rangle - (32Z/35)|\phi_1\rangle \times \langle\phi_1|U^{-1}|f_2\rangle + U^{1/2}MU^{-1/2}|f_2\rangle. \tag{B9}$$

It also follows from (3.17) and (3.19) that

$$MU^{-1/2}|\phi_1\rangle = \text{the null vector}. \tag{B10}$$

Equations (B9), (B10), and the fact that $\langle\phi_1|U^{-1}|\phi_1\rangle = 35/(32Z)$ can be used to show that $\langle\phi_1|U^{-1}|f_2\rangle = \langle\phi_1|f_1\rangle$. This can in turn be used to rewrite (B9) in the form

$$|f_2\rangle = [U - (16Z/35)|\phi_1\rangle\langle\phi_1|] \times (I + |\phi_1\rangle\langle\phi_1|)|f_1\rangle + U^{1/2}MU^{-1/2}|f_2\rangle. \tag{B11}$$

The Neumann series solution of (B11) in powers of M is readily calculated with the aid of (B10). Comparing the result with the definitions (3.28) and (3.29) of k_d and k_e shows that this Neumann series solution is

$$|f_2\rangle = (k_d + k_e)(I + |\phi_1\rangle\langle\phi_1|)|f_1\rangle. \tag{B12}$$

With $|f_2\rangle$ known, the action of $V^{1/2}PV^{1/2}$ in $\int_S$ is known. The expression (B12) for $|f_2\rangle$ is the same as the one which follows from Eqs. (B1)–(B3) and (3.27). The action of $V^{1/2}PV^{1/2}$ in $\int_A$ can be obtained from a variational principle like (B6) in similar fashion. Once the action of $V^{1/2}PV^{1/2}$ in $\int_S$ and in $\int_A$ is known, $V^{1/2}PV^{1/2}$ is itself known, and can be written in the explicit form (3.27) if desired.

APPENDIX C: THE DECOMPOSITIONS $n_l\,{}^\dagger n_l$ AND $\zeta_l\,{}^\dagger\zeta_l$

This appendix shows how the η_l of the decomposition (4.31) can be explicitly constructed. The construction proceeds by first finding a decomposition

$$rg_l(r,r')r' = \int_0^\infty \zeta_l^\dagger(r,s)\zeta_l(s,r')\,ds. \tag{C1}$$

Given a decomposition of the form $\zeta_l^\dagger\zeta_l$ for the ordinary inverse $rg_l(r,r')r'$, the corresponding decomposition $\eta_0^\dagger\eta_0$ of the generalized inverse $rg_0'(r,r')r'$ can be found with the aid of the formula

$$\zeta_0^\dagger\zeta_0 - \zeta_0^\dagger\zeta_0|v\rangle(\langle v|\zeta_0^\dagger\zeta_0|v\rangle)^{-1}\langle v|\zeta_0^\dagger\zeta_0 = \eta_0^\dagger\eta_0, \tag{C2}$$

where

$$\eta_0 = \zeta_0 - \zeta_0|v\rangle(\langle v|\zeta_0^\dagger\zeta_0|v\rangle)^{-1}\langle v|\zeta_0^\dagger\zeta_0. \tag{C3}$$

The kernel ζ_l will be obtained by observing that the differential equation

$$\left(-\frac{\partial^2}{\partial s^2} + \frac{l(l+1)}{s^2} + \kappa^2\right) sg_l(s,r')r' = \delta(s-r'), \tag{C4}$$

can be used to express $rg_l(r,r')r'$ in the form

$$rg_l(r,r')r' = \int_0^\infty rg_l(r,s)s\left(-\frac{\partial^2}{\partial s^2} + \frac{l(l+1)}{s^2} + \kappa^2\right) sg_l(s,r')r'\,ds. \tag{C5}$$

If a function $U_2(s)$ can be found such that

$$-\frac{\partial^2}{\partial s^2} + \frac{l(l+1)}{s^2} + \kappa^2 = \left(-\frac{\partial}{\partial s} + U_2(s)\right)\left(\frac{\partial}{\partial s} + U_2(s)\right), \tag{C6}$$

then substituting (C6) into (C8) and integrating by parts once yields (C1) with

$$\zeta_l(s,r') = \left(\frac{\partial}{\partial s} + U_2(s)\right) sg_l(s,r')r', \tag{C7}$$

and

$$\zeta_l^\dagger(r,s) = \zeta_l(s,r). \tag{C8}$$

It follows immediately from (C6) that the U_2 needed for the factorization (C6) satisfies the Riccati equation

$$\frac{-dU_2(s)}{ds} + [U_2(s)]^2 = l(l+1)s^{-2} + \kappa^2. \tag{C9}$$

The standard substitution

$$U_2 = -\frac{d\ln u(s)}{ds}, \tag{C10}$$

converts the Riccati equation (C9) into

$$\frac{-d^2u}{ds^2} + [l(l+1)s^{-2} + \kappa^2]u = 0, \tag{C11}$$

whose general solution is an arbitrary linear combination of $s^{1/2}I_{l+1/2}(\kappa s)$ and $s^{1/2}K_{l+1/2}(\kappa s)$. Choosing a particular solution u determines ζ_l via (C7) and (C10); clearly ζ_l is not unique. The particular choise $u(s) = s^{1/2}K_{l+1/2}(\kappa s)$ leads to

$$\zeta_l(s,r') = \begin{cases} r'^{1/2}K_{l+1/2}(\kappa r')/[s^{1/2}K_{l+1/2}(\kappa s)], & s < r', \\ 0, & s > r'. \end{cases} \tag{C12}$$

It is easy to verify that (C12) works: Eq. (C12) implies that

$$\int_0^\infty \zeta_l^\dagger(r,s)\zeta_l(s,r')\,ds = r^{1/2}K_{l+1/2}(\kappa r)\,r'^{1/2} \times K_{l+1/2}(\kappa r')\int_0^{r_<}[s^{1/2}K_{l+1/2}(\kappa s)]^{-2}\,ds \tag{C13}$$

Here $r_<$ is the smaller of the pair (r,r'). The integral in (C13) can be evaluated with the aid of the Wronskian relation[41]

$$I'_{l+1/2}(x)K_{l+1/2}(x) - I_{l+1/2}(x)K'_{l+1/2}(x) = x^{-1}. \tag{C14}$$

One has

$$\begin{aligned}\int_0^{r_<}[s^{1/2}K_{l+1/2}(\kappa s)]^{-2}\,ds &= \int_0^{\kappa r_<}[K_{l+1/2}(x)]^{-2}[I'_{l+1/2}(x)K_{l+1/2}(x) \\ &\quad - I_{l+1/2}(x)K'_{l+1/2}(x)]\,dx \\ &= I_{l+1/2}(\kappa r_<)/K_{l+1/2}(\kappa r_<).\end{aligned} \tag{C15}$$

Using (C15) in (C13) and comparing with (4.9) shows that the particular ζ_l given in (C12) satisfies (C1).

With ζ_l known, the η_l needed for the decomposition (4.31) can be constructed with the aid of (C3); the particular choice (C12) yields

$$\begin{aligned}\eta_l(s,r) = \zeta_l(s,r) &- 4(5+4\kappa+\kappa^2)^{-1}[(1+\kappa)s+1]\exp(-s) \\ &\times\{2(1-\kappa)^{-2}\exp(-\kappa r) - [(1+\kappa)(1-\kappa)^{-1} \\ &\times r + 2(1-\kappa)^{-2}]\exp(-r)\}\,\delta_{l,0}.\end{aligned} \tag{C16}$$

It is straighforward to verify that the η_l given by (C16) satisfies (4.31).

[1]Variational (Rayleigh–Ritz) upper bounds of sufficient accuracy to demonstrate that H^- has at least one bound state were obtained only a few years after the discovery of the Schrödinger equation. H. Bethe, Z. Phys. **57**, 815–21 (1929) obtained an upper bound of $-1.05056\ h^{-1}$ Ry, and E. A. Hylleraas, Z. Phys. **63**, 291–2 (1930) obtained an upper bound of $-1.0528\ h^{-1}$ Ry. Both bounds lie below the hydrogen ground state energy of $-1\ h^{-1}$ Ry. A large scale variational calculation by C. L. Pekeris, Phys. Rev. **126**, 1470–6 (1962) used a 444 term trial function to obtain an upper bound of $-1.0555020125908\ h^{-1}$ Ry to the nonrelativistic energy. The H^- ground state has been discussed recently by C. D. Lin, Phys. Rev. A **12**, 493–7 (1975).

[2]R. N. Hill, Phys. Rev. Lett. **38**, 643—6 (1977).
[3]Tosio Kato, Trans. Am. Math. Soc. **70**, 195—211 (1951).
[4]The first results of this kind were obtained by G. M. Zhislin, Dokl. Akad. Nauk SSSR **128**, 231—4 (1959), Trudy Moskov Mat. Obsc. **9**, 81—120 (1960), Dokl. Akad. Nauk SSSR **175**, 521—524 (1967) [Soviet Math. Dokl. **8**, 878—82 (1967)], Izv. Akad. Nauk SSSR, Ser. Mat. **33**, 590—649 (1969) [Math. USSR, Izvestija **3**, 559—616 (1969/70)], and by G. M. Zhislin and A. G. Sigalov, Izv. Akad. Nauk SSSR, Ser. Mat. **29**, 835—60 [AMS Transl. Ser. II, **91**, 263—95 (1970)], Izv. Akad. Nauk SSSR, Ser. Mat. **29**, 1261—72 (1965) [AMS Transl. Ser. II, **91**, 297—310 (1970)]. Extensions of the results and alternate methods of proof may be found in W. Hunziker, Helv. Phys. Acta. **39**, 451—62 (1966), B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, New Jersey, 1971), Chap. VII; K. Jörgens and J. Weidmann, *Spectral Properties of Hamiltonian Operators* (Lecture Notes in Mathematics, Vol. 313, Springer-Verlag, 1973), J. Uchiyama, Publ. Res. Inst. Math. Sci. (Kyoto) A **2**, 117—32 (1966); V. Enss, Commun. Math. Phys. **52**, 233—8 (1977).
[5]Very general results of this kind have been obtained by A. O'Connor, Commun. Math. Phys. **32**, 319—40 (1973). See also B. Simon, Proc. Am. Math. Soc. **42**, 395—401 (1974).
[6]The earliest result of this kind for multiparticle atomic systems is due to J. Weidmann, Commun. Pure Appl. Math. **19**, 107—10 (1966). See also J. Weidmann, Bull. Am. Math. Soc. **73**, 452—6 (1967); S. Albeverio, Ann. Phys. (N. Y.) **71**, 167—276 (1972); B. Simon, Math. Ann. **207**, 133—8 (1974).
[7]A. G. Sigalov, Uspekhi Mat. Nauk **22**, No. 2 (134), 3—20 (1967) [Russ. Math. Surveys **22**, No. 2, 1—18 (1967)].
[8]T. Kato, Suppl. Prog. Th. Phys. **40**, 3—19 (1967).
[9]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, New Jersey, 1971); B. Simon, in *Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann* (Princeton U. P., Princeton, New Jersey, 1976), pp. 305—26.
[10]K. Jörgens and J. Weidmann, *Spectral Properties of Hamiltonian Operators* (Lecture Notes in Mathematics, Vol. 313, Springer Verlag, 1973).
[11]W. Hunziker, Acta Phys. Austr., to appear.
[12]T. Kato, Trans. Am. Math. Soc. **70**, 212—18 (1951), first showed that helium in the infinite nuclear mass approximation has an infinity of bound states. For general atoms, it is a result of G. M. Zhislin, Trudy Moskov. Mat. Obsc. **9**, 81—128 (1960). See also E. Balslev, Ann. Phys. (N. Y.) **73**, 49—107 (1972); B. Simon, Helv. Phys. Acta **43**, 607—30 (1970); J. Uchiyama, Publ. Res. Inst. Math. Sci. (Kyoto) A **2**, 117—132 (1966/67).
[13]The finiteness of the number of bound states for negative ions has been discussed in a rigorous way by J. Uchiyama, Publ. Res. Inst. Math. Sci. (Kyoto) A **5**, 51—63 (1969); G. M. Zhislin, Teor. Mat. Fiz. **7**, 332—341 (1971) [Theor. Math. Phys. **7**, 571—578 (1971)]; M. A. Antonets, G. M. Zhislin, and I. A. Shereshevskii, Teor. Mat. Fiz. **16**, 235—46 (1973) [Theor. Math. Phys. **16**, 800—9 (1974)]; D. R. Yafeev, Funkcional. Anal. Prilozen. **6**, 103—4 (1973) [Funct. Anal. Appl. **6**, 349—50 (1972)]. In particular, Yafeev has proven that H^- has only finitely many bound states even when the nuclear mass is finite.
[14]See H. S. W. Massey, *Negative Ions* (Cambridge U. P., Cambridge, 1976), 3rd ed., pp. 9—11 and pp. 25—30.
[15]This belief is based on computer solutions to the Hartree—Fock equations. To the best of the author's knowledge, a rigorous proof has not been given.
[16]This case has been reviewed by B. Simon, in *Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann* (Princeton U. P., Princeton, New Jersey, 1976), pp. 305—26 and by A. Martin (CERN preprint, Ref. TH. 2294-CERN 10 March 1977).
[17]F. H. Gertler, H. B. Snodgrass, and L. Spruch, Phys. Rev. **172**, 110—8 (1968); I. Aronson, C. J. Kleinman, and L. Spruch, Phys. Rev. A **4**, 841—6 (1971). Some of their results are not rigorous because they are based on approximate adiabatic potentials for which rigorous error bounds are not supplied.
[18]I. Aronson, C. J. Kleinman, and L. Spruch, Phys. Rev. A **12**, 349—52 (1975).
[19]N. W. Bazley, Proc. Natl. Acad. Sci. U. S. A. **45**, 850—3 (1959); Phys. Rev. **120**, 144—9 (1960).
[20]N. W. Bazley and D. W. Fox, Phys. Rev. **124**, 483—92 (1961); J. Res. Nat. Bur. Standards Sect. B **65**, 105—11 (1961). Certain difficulties of principle in extending the Bazley—Fox method to lithium are resolved in D. W. Fox, SIAM J. Math. Anal. **3**, 617—24 (1972); D. W. Fox and V. G. Sigillito, Chem. Phys. Lett. **13**, 85—7 (1972); **14**, 583—5 (1972); D. W. Fox and V. G. Sigillito, J. Appl. Math. Phys. **23**, 392—411 (1972). See also C. E. Reid, Int. J. Quantum Chem. (U. S. A.) **6**, 793—5 (1972).
[21]S. H. Gould, *Variational Methods for Eigenvalue Problems: An Introduction to the Weinstein Method of Intermediate Problems* (University of Toronto Press, Toronto, 1966), 2nd ed.
[22]A. Weinstein and W. Stenger, *Methods of Intermediate Problems for Eigenvalues* (Academic, New York, 1972).
[23]H. F. Weinberger, *Variational Methods for Eigenvalue Approximation* (SIAM Regional Conference Series in Applied Mathematics, 1974).
[24]D. S. Hughes and Carl Eckart, Phys. Rev. **36**, 694—8 (1930).
[25]Essentially this theorem appears as Theorem 1 on p. 21 of Weinstein and Stenger, Ref. 22, as Theorem 9 on p. 62 of Weinberger, Ref. 23, and is implicit in the discussion of intermediate problems of the second type in Gould, Ref. 21.
[26]There are actually two distinct minimax principles, one going back to H. Weyl and R. Courant, the other due to Poincaré. See Weinstein and Stenger, Ref. 22, Chap. 3, for a careful discussion of the distinction and of the relation between the two principles. See also Weinberger, Ref. 23, pp. 55—57.
[27]A formula such as (3.12) for the projection onto the span of the ranges of two projection operators is apparently not in the literature. A formula equivalent to (3.12) has been obtained independently by Percy Deift (private communication).
[28]The decomposition $k = k_1 - k_2$ can be verified by carrying out the integration indicated in (3.21). The existence of such a decomposition is suggested by the fact that $(4\pi|\mathbf{r}-\mathbf{r}'|)^{-1}$ is an inverse of $-\nabla^2$ while $|\mathbf{r}-\mathbf{r}'|/(8\pi)$ is an inverse of $-\nabla^4$. The decomposition is essentially the formal relation $\nabla^{-4} = (-\nabla^{-2})^2$ corrected by the subtractions needed to obtain a convergent integral in (3.21).
[29]This method of counting bound states is originally due independently to J. Schwinger, Proc. Natl. Acad. Sci. U. S. A. **47**, 122—9 (1961) and M. S. Birman, Mat. Sb. **55**, 124—74 (1961) [AMS Transl. Ser. II **53**, 23—80 (1966)]. See B. Simon, Ref. 16, for a review.
[30]The notation for the modified Bessel functions is that used in W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), 3rd ed., p. 66.
[31]Ref. 30, p. 37.
[32]Ref. 30, p. 49.
[33]See, for example, S. G. Mikhilin, *Integral Equations and Their Applications to Certain Problems in Mechanics, Mathematical Physics, and Technology* (Pergamon, New York, 1957), 2nd revised ed., Chap. II or F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957), Chap. III.
[34]A useful general reference is L. M. Delves and J. Walsh, *Numerical Solution of Integral Equations* (Oxford U. P., New York, 1974). Rigorous error bounds to the eigenvalues of integral equations which are solved by numerical methods are discussed by I. P. Mysovskih, Mat. Sb. (N. S.) **48** (90), 137—148 (1959) [AMS Transl. Ser. II, **35**, 237—50 (1964)].
[35]This trick, which is outlined below, is discussed in S. G. Mikhilin, Ref. 33, pp. 88—94 and in F. G. Tricomi, Ref. 33, pp. 119—24.
[36]The method has some features in common with the partitioning technique of P. O. Lowdin, J. Math. Phys. **3**, 969 (1962); Phys. Rev. **139**, A357 (1965); J. Chem. Phys. **43**, 5175 (1965). See also T. M. Wilson, J. Chem. Phys. **47**, 3912, 4706 (1967).
[37]This trick for constructing lower bounds is known as trunca-

tion. It was introduced in H. F. Weinberger, "A Theory of Lower Bounds for Eigenvalues," Tech. Note BN-183, IFDAM, Univ. of Maryland, College Park, Maryland (1959). It was developed for quantum mechanical problems by N. W. Bazley and D. W. Fox, Ref. 20. See also Refs. 21—23.

[38]See S. G. Mikhilin, Ref. 33, pp. 19—22 or F. G. Tricomi, Ref. 33, pp. 55—64.

[39]See Ref. 30, p. 344, or Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1965), p. 230. $E_1(x)$ is tabulated on pp. 238—51 of Abramowitz and Stegun.

[40]H. Shull and P.-O. Löwdin, J. Chem. Phys. 25, 1035—40 (1956). See also H. S. W. Massey, Ref. 14, pp. 9—10.

[41]Ref. 30, p. 68.

The generalized Langevin equation with Gaussian fluctuations

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 10 May 1977)

It is shown that all statistical properties of the generalized Langevin equation with Gaussian fluctuations are determined by a single, two-point correlation function. The resulting description corresponds with a *stationary, Gaussian, non-Markovian* process. Fokker–Planck-like equations are discussed, and it is explained how they can lead one to the erroneous conclusion that the process is *nonstationary, Gaussian, and Markovian*.

I. INTRODUCTION

The generalized Langevin equation provides a stochastic description of Brownian motion. In one dimension, it has the form

$$\frac{d}{dt}u(t) = -\int_0^t \beta(t-s)u(s)\,ds + \frac{1}{m}\tilde{f}(t) \tag{1}$$

in which $u(t)$ is the velocity of the Brownian particle at time t, m is its mass, $\beta(t-s)$ is the dissipative "memory kernel," and $\tilde{f}(t)$ is a Gaussian fluctuating driving force. It is assumed that $\tilde{f}(t)$ possesses first and second moments given by

$$\langle \tilde{f}(t)\rangle = 0 \quad \text{and} \quad \langle \tilde{f}(t)\tilde{f}(s)\rangle = k_B T m \beta(|t-s|), \tag{2}$$

in which k_B is Boltzmann's constant and T is the temperature of the fluid in which the Brownian particle is immersed. A Markovian limit of this description is obtained when $\beta(t-s) = 2\beta\delta(t-s)$ in which β is a constant and $\delta(t-s)$ is the Dirac delta function.

In recent papers, Adelman[1] and Fox[2] have derived Fokker–Planck-like equations corresponding to the process described by (1) and (2). It was even asserted,[2] on the basis of the Fokker–Planck-like equation, that the process being described must be a *nonstationary, Gaussian, Markovian* process. Here, it will be shown that the process is in fact a *stationary, Gaussian, non-Markovian* process, and that the Fokker-Planck-like equations of Adelman[1] and Fox[2] are not properly Fokker–Planck equations after all. It will be clearly indicated how the confusion arises, and the distinction between bona fide Fokker–Planck equations for bona fide *nonstationary, Gaussian, Markov* processes, and Fokker–Planck-like equations which arise in the study of *stationary, Gaussian, non-Markovian* processes will be elucidated.

II. THE SOLUTION TO THE GENERALIZED LANGEVIN EQUATION

Even though the process described by (1) and (2) will be seen to be non-Markovian, the Gaussian property of the fluctuating driving force leads to a complete stochastic description in terms of a single, two-point correlation function. Therefore, the Gaussianness provides a description which has a property usually associated with Markovianness, i.e., a single two-point function determines everything. In the case of Markovianness the two point function is the two-point conditional distribution P_2, which will be discussed below.

Using Laplace transforms and the definition

$$\hat{\beta}(z) = \int_0^\infty e^{-zt}\beta(t)\,dt \tag{3}$$

enables one to obtain the solution to (1) in the form

$$u(t) = \chi(t)u(0) + \frac{1}{m}\int_0^\infty \chi(t-s)\tilde{f}(s)\,ds, \tag{4}$$

in which $\chi(t)$ is defined through its Laplace transform

$$\hat{\chi}(z) \equiv [z + \hat{\beta}(z)]^{-1}. \tag{5}$$

While the Laplace transform method of treatment of the generalized Langevin equation is standard in the literature, some of the results to be given below appear not to have been previously published and greatly clarify the discussion.

In their pioneering work on Brownian motion, Uhlenbeck and Ornstein[3] observed that two types of averaging are necessary in a discussion of Brownian motion using the Langevin equation. The first type of averaging is with respect to the stochastic driving force, $\tilde{f}(t)$, and is denoted, as in (2), by $\langle\cdots\rangle$. The second type of averaging is with respect to the initial velocity $u(0)$, which appears in the solution (4) and will be denoted by $\{\cdots\}$. The distribution for $u(0)$ will be the Maxwellian

$$W(u(0)) = (2\pi k_B T/m)^{-1/2}\exp[-mu^2(0)/2k_B T]. \tag{6}$$

Using the solution (4), we can compute the velocity autocorrelation function for $t_2 \geq t_1$

$$\begin{aligned}\{\langle u(t_1)u(t_2)\rangle\} &= \chi(t_1)\chi(t_2)\{u^2(0)\} \\ &\quad + (k_B T/m)(\chi(t_2 - t_1) - \chi(t_1)\chi(t_2)) \\ &= (k_B T/m)\chi(t_2 - t_1).\end{aligned} \tag{7}$$

To get (7), we have used an identity which is proved in Appendix A, which states, for $t_2 \geq t_1$,

$$\begin{aligned}&\langle\left(\int_0^{t_2}\chi(t_2 - s_2)\tilde{f}(s_2)\,ds_2\right)\left(\int_0^{t_1}\chi(t_1 - s_1)\tilde{f}(s_1)\,ds_1\right)\rangle \\ &\quad = k_B T m(\chi(t_2 - t_1) - \chi(t_1)\chi(t_2))\end{aligned} \tag{8}$$

and which is not found in the usual treatments of the

problem by the Laplace transform methods. Equation (7) makes it quite plain that the process is *stationary*. For $t_2 = t_1$, (7) reduces to precisely the same result obtained from (6) for $\{u^2(0)\}$. Stationarity means that the Maxwellian persists.

Using the autocorrelation given by (7), we can construct the unconditioned two-point distribution function from the correlation matrix, by following a procedure discussed by Wang and Uhlenbeck.[4] The correlation matrix is

$$\frac{k_B T}{m}\begin{pmatrix} 1 & \chi(t_2 - t_1) \\ \chi(t_2 - t_1) & 1 \end{pmatrix} \tag{9}$$

and its inverse is easily seen to be

$$\left(\frac{k_B T}{m}\right)^{-1} \frac{1}{1-\chi^2(t_2-t_1)}\begin{pmatrix} 1 & -\chi(t_2 - t_1) \\ -\chi(t_2 - t_1) & 1 \end{pmatrix}. \tag{10}$$

This implies the two-point distribution function

$$W_2(u_1 t_1; u_2 t_2)$$

$$= \left[\left(2\pi \frac{k_B T}{m}\right)^2 (1 - \chi^2(t_2 - t_1))\right]^{-1/2}$$

$$\times \exp\left(-\frac{m(u_1^2 + u_2^2 - 2u_1 u_2 \chi(t_2 - t_1))}{2k_B T(1-\chi^2(t_2-t_1))}\right). \tag{11}$$

The validity of this result follows from the fact that $u(t)$, as given by (4), inherits the Gaussianness of $f(t)$ as a consequence of linearity, and $\{\langle u(t)\rangle\} = 0$.

If we want the conditioned two-point distribution, then we can use the definition[4]

$$P_2(u_1 t_1; u_2 t_2) \equiv W_2(u_1 t_1; u_2 t_2)/W_1(u_1 t_1). \tag{12}$$

However, from above we have $\{\langle u^2(t_1)\rangle\} = k_B T/m$, and with $\{\langle u(t_1)\rangle\} = 0$, it follows that

$$W_1(u_1 t_1) = W_1(u_1) = (2\pi k_B T/m)^{-1/2}$$

$$\times \exp(-m u_1^2/2k_B T). \tag{13}$$

As already mentioned, this persistence of the Maxwellian distribution exhibits the *stationarity* of the process. From (12) and (13) it follows that

$$P_2(u_1 t_1; u_2 t_2)$$

$$= [2\pi(k_B T/m)(1 - \chi^2(t_2 - t_1))]^{-1/2}$$

$$\times \exp[-m(u_2 - \chi(t_2 - t_1)u_1)^2/2k_B T(1 - \chi^2(t_2 - t_1))]. \tag{14}$$

Higher order distributions can also be constructed and they all depend upon $\chi(t - t')$, the two-point correlation function. In particular, the three-point, unconditioned distribution, $W_3(u_1 t_1; u_2 t_2; u_3 t_3)$ for $t_3 \geq t_2 \geq t_1$ is determined from the corelation matrix

$$\frac{k_B T}{m}\begin{pmatrix} 1 & \chi(t_2 - t_1) & \chi(t_3 - t_1) \\ \chi(t_2 - t_1) & 1 & \chi(t_3 - t_2) \\ \chi(t_3 - t_1) & \chi(t_3 - t_2) & 1 \end{pmatrix} \tag{15}$$

by computing its inverse, a somewhat laborious but straightforward procedure.[4]

III. NON-MARKOVIANNESS OF THE SOLUTION

If the process were Markovian, then the Smoluchowski,[5] or Chapman–Kolmogorov,[6] equation would have to hold:

$$P_2(u_1 t_1; u_3 t_3) = \int_{-\infty}^{\infty} dv\, P_2(vs; u_3 t_3) P_2(u_1 t_1; vs) \tag{16}$$

for $t_3 \geq s \geq t_1$. For the result in (14), however, this requirement leads to the requirement

$$\chi(t_3 - t_2)\chi(t_2 - t_1) = \chi(t_3 - t_1). \tag{17}$$

Equation (17) is only satisfied by

$$\chi(t - t') = \exp[-(t - t')D] \tag{18}$$

according to Doob's theorem.[7,8] But this implies, when (5) is used, that $\hat{\beta}(z) = D$, so that

$$\beta(t) = 2D\delta(t). \tag{19}$$

This is simply a Markovian limit of the generalized Langevin equation. Therefore, (17) is not satisfied and neither is (16). The process is *non-Markovian*. This is surely hardly a surprise given the presence of the "memory kernel" in (1).

IV. FOKKER-PLANCK-LIKE EQUATIONS FOR THE SOLUTION

Associated with (14) is the partial differential equation

$$\frac{\partial}{\partial t_2} P_2(u_1 t_1; u_2 t_2)$$

$$= -\frac{\dot{\chi}(t_2 - t_1)}{\chi(t_2 - t_1)} \frac{\partial}{\partial u_2}(u_2 P_2(u_1 t_1; u_2 t_2))$$

$$- \frac{k_B T}{m} \frac{\dot{\chi}(t_2 - t_1)}{\chi(t_2 - t_1)} \frac{\partial^2}{\partial u_2^2} P_2(u_1 t_1; u_2 t_2) \tag{20}$$

subject to the initial condition $P_2(u_1 t_1; u_2 t_1) = \delta(u_2 - u_1)$. $\dot{\chi}(t_2 - t_1)$ denotes the derivative of $\chi(\tau)$ with respect to τ. If we consider the special case $t_1 = 0, t_2 = t, u_1 = u(0)$, and $u_2 = u$, then (20) looks like

$$\frac{\partial}{\partial t} P_2(u(0); ut)$$

$$= -\frac{\dot{\chi}(t)}{\chi(t)} \frac{\partial}{\partial u}(u P_2(u(0); ut))$$

$$- \frac{k_B T}{m} \frac{\dot{\chi}(t)}{\chi(t)} \frac{\partial^2}{\partial u^2} P_2(u(0); ut), \tag{21}$$

with the initial condition $P(u(0); u0) \equiv \delta(u - u(0))$.

Equation (21) looks very much like a bona fide Fokker–Planck equation for a *nonstatinary*, *Gaussian*, *Markov* process and is precisely the equation both Adelman and Fox obtained earlier by a different procedure. Below, it will be shown that bona fide *nonstationary*, *Gaussian*, *Markov* processes do lead to Fokker–Planck equations of precisely the form of (21) *but* with *less* stringent initial conditions. It will also be shown that (21) will *not* lead to results consonant with (20) and (14) if it is treated as a bona fide Fokker–Planck equation. The reasons for these distinctions are mani-

fest in (20) wherein the coefficients $\dot{\chi}(t_2-t_1)/\chi(t_2-t_1)$ exhibit explicit dependence on *both* t_2 and t_1.

A bona fide *nonstationary, Gaussian, Markov* process is described by the equation

$$\frac{d}{dt}u(t) = -\beta(t)u(t) + \frac{1}{m}\tilde{f}(t) \tag{22}$$

with a Gaussian fluctuating force $\tilde{f}(t)$ possessing first and second moments

$$\langle\tilde{f}(t)\rangle = 0 \quad \text{and} \quad \langle\tilde{f}(t)\tilde{f}(s)\rangle = 2k_BTm\beta(t)\delta(t-s). \tag{23}$$

The solution to (22) is

$$u(t) = \exp[-\int_0^t \beta(s)ds]u(0) + \int_0^t \exp[-\int_s^t \beta(s')ds'][\tilde{f}(s)/m]ds. \tag{24}$$

The velocity autocorrelation function is, for $t_2 \geq t_1$,

$$\begin{aligned}\{\langle u(t_2)u(t_1)\rangle\} &= \exp[-\int_0^{t_2}\beta(s)\,ds]\exp[-\int_0^{t_1}\beta(s)\,ds]\{u^2(0)\} + (k_BT/m)\\ &\times\{\exp[-\int_{t_1}^{t_2}\beta(s)\,ds] - \exp[-\int_0^{t_2}\beta(s')\,ds']\\ &\times\exp[-\int_0^{t_1}\beta(s'')\,ds'']\}\\ &= (k_BT/m)\exp[-\int_{t_1}^{t_2}\beta(s)\,ds].\end{aligned} \tag{25}$$

This result is proved in Appendix B.

Proceeding as in (7)–(14) leads to the conditioned two-point distribution

$$P_2(u_1t_1;u_2t_2) = (2\pi(k_BT/m)\{1-\exp[-2\int_{t_1}^{t_2}\beta(s)\,ds]\})^{-1/2} \times\exp\left(-\frac{m(u_2-\exp[-\int_{t_1}^{t_2}\beta(s)\,ds]u_1)^2}{2k_BT(1-\exp[-2\int_{t_1}^{t_2}\beta(s)\,ds])}\right). \tag{26}$$

Associated with this P_2 is the partial differential equation

$$\frac{\partial}{\partial t_2}P_2(u_1t_1;u_2t_2) = \beta(t_2)\frac{\partial}{\partial u_2}(u_2P_2(u_1t_1;u_2t_2)) + \frac{k_BT}{m}\beta(t_2)\frac{\partial^2}{\partial u_2^2}P_2(u_1t_1;u_2t_2) \tag{27}$$

with the initial equation $P_2(u_1t_1;u_2t_1) = \delta(u_2-u_1)$. Now, again we make the substitutions $t_1=0$, $t_2=t$, $u_1=u(0)$, and $u_2=u$. Then (27) looks like

$$\frac{\partial}{\partial t}P_2(u(0);ut) = \beta(t)\frac{\partial}{\partial u}(uP_2(u(0);ut)) + \frac{k_BT}{m}\beta(t)\frac{\partial^2}{\partial u^2}P_2(u(0);ut) \tag{28}$$

with initial condition $P_2(u(0);u0) = \delta(u-u(0))$. The big difference between these results and those in (20) and (21) is

$$-\dot{\chi}(t_2-t_1)/\chi(t_2-t_1) \to \beta(t_2). \tag{29}$$

The explicit t_1 dependence of (20) disappears in (27). Moreover, we can always solve (28) starting at any time t_1 with $P_2(u_1;ut_1) = \delta(u-u_1)$, and (28) then yields precisely (26). Therefore, (28) *with appropriate initial conditions* yields (26) for *any* time interval. This is a major characteristic of a bona fide Fokker–Planck equation. However, Eq. (21) does not behave this way. If the interval from t_1 to t_2 is considered and $P_2(u_1;ut_1) = \delta(u-u_1)$ is assumed, then (21) yields

$$P_2(u_1t_1;u_2t_2) = \left[2\pi\frac{k_BT}{m}\left(1-\frac{\chi^2(t_2)}{\chi^2(t_1)}\right)\right]^{-1/2} \times\exp\left(-\frac{m(u_2-[\chi(t_2)/\chi(t_1)]u_1)^2}{2k_BT(1-\chi^2(t_2)/\chi^2(t_1))}\right). \tag{30}$$

This is *not* equal to (14) except when $\chi(t_2-t_1) = \chi(t_2)/\chi(t_1)$ which only holds when (18) is true, that is, only in the Markovian situation.

V. DISCUSSION OF RESULTS

In this paper we have developed the explicit solution for the generalized Langevin equation using the distribution functions which all depend upon a single, two-time correlation function. In the earlier work of Adelman[1] and Fox[2] the approach was based upon obtaining the Fokker–Planck equation representation. In fact both Adelman and Fox only obtained the Fokker–Planck-like equation identical with (21). They did not obtain (20) which would have made it clear that these equations require a very special initial condition tailored to the specific time interval between t_1 and t_2.

Adelman does suggest that the non-Markovian Langevin description, as in (1), is more fundamental than the Fokker–Planck-like equation description, as in (21). However, after obtaining (21) Adelman overlooks the fact that the Fokker–Planck-like equation only generates the P_2 function for the time interval from 0 to t, and for *no* other interval. He also does not obtain the complete description for arbitrary intervals which is exhibited in (14) and (20). Fox compounds this confusion by noting that the solutions to (21), which he mistakenly takes to be valid for *arbitrary* intervals, generates (30) for arbitrary intervals. While (30) clearly does not describe the actual process given by (1) and (4), as has been pointed out above, it does, *unfortunately*, satisfy identically the Chapman–Kolmogorov–Smoluchowski identity (16) *and* the Doob identity, which in that case is simply

$$\frac{\chi(t_3)}{\chi(t_2)}\frac{\chi(t_2)}{\chi(t_1)} = \frac{\chi(t_3)}{\chi(t_1)}. \tag{31}$$

This "verifies" the *Markov* property! Thus, it appears that the process is really *nonstationary, Gaussian, Markovian*.

The remarkable feature, which is valid in the Gaussian case anyway, is that the description of all the statistics for the generalized Langevin equation depends on only a single, two-point correlation, $\chi(t_2-t_1)$. Markov processes are always determined completely by a single two-point distribution, P_2. Gaussian, *non*-Markovian processes are *also* completely determined by

$\chi(t_2-t_1)$, when there is a fluctuation–dissipation relation as in (2). Therefore, nothing is really lost by using a non-Markovian description in place of a Markov description as long as it is Gaussian!

None of the preceding considerations are substantially altered in the multicomponent generalization of (1). One still gets a *stationary, Gaussian, non-Markovian* process which is determined completely by a single, two-point correlation matrix.

ACKNOWLEDGMENTS

The author is indebted to Dr. Mark Kac and Dr. Joel Keizer for numerous communications relating to this paper. He also has been supported by the Alfred P. Sloan Foundation and a grant from the National Science Foundation.

APPENDIX A: PROOF OF (8)

$$\langle[\int_0^{t_2}\chi(t_2-s_2)\tilde f(s_2)\,ds_2][\int_0^{t_1}\chi(t_1-s_1)\tilde f(s_1)\,ds_1]\rangle$$
$$=\int_0^{t_2}ds_2\int_0^{t_1}ds_1\,\chi(t_2-s_2)\chi(t_1-s_1)k_BTm\beta(|s_2-s_1|) \tag{A1}$$

follows from (2).

The double Laplace transform of the right-hand side is

$$\int_0^\infty dt_2\exp(-zt_2)\int_0^\infty dt_1\exp(-z't_1)\int_0^{t_2}ds_2\int_0^{t_1}ds_1$$
$$\times\chi(t_2-s_2)\chi(t_1-s_1)k_BTm\beta(|s_2-s_1|)$$
$$=k_BTm\int_0^\infty ds_2\int_{s_2}^\infty dt_2\int_0^\infty ds_1\int_{s_1}^\infty dt_1$$
$$\times\exp[-z(t_2-s_2)]\chi(t_2-s_2)$$
$$\times\exp[-z'(t_1-s_1)]\chi(t_1-s_1)\exp(-zs_2)$$
$$\times\exp(-z's_1)\beta(|s_2-s_1|)$$
$$=k_BTm\int_0^\infty ds_2\int_0^\infty d\tau_2\int_0^\infty ds_1\int_0^\infty d\tau_1$$
$$\times\exp(-z\tau_2)\chi(\tau_2)\exp(-z'\tau_1)$$
$$\times\chi(\tau_1)\exp(-zs_2)\exp(-z's_1)\beta(|s_2-s_1|)$$
$$=k_BTm\hat\chi(z)\hat\chi(z')\int_0^\infty ds_2\int_0^\infty ds_1\exp(-zs_2)$$
$$\times\exp(-z's_1)\beta(|s_2-s_1|). \tag{A2}$$

Now,

$$\int_0^\infty ds_2\int_0^\infty ds_1\exp(-zs_2)\exp(-z's_1)\beta(|s_2-s_1|)$$
$$=\int_0^\infty ds_2\int_0^\infty ds_1\exp[-z(s_2-s_1)]\beta(|s_2-s_1|)$$
$$\times\exp[-(z+z')s_1]$$
$$=\int_0^\infty ds_1\int_{-s_1}^\infty d\sigma\exp(-z\sigma)\beta(|\sigma|)\exp[-(z+z')s_1]$$
$$=\int_0^\infty ds_1\Big(\hat\beta(z)+\int_{-s_1}^0 d\sigma\exp(-z\sigma)\beta(|\sigma|)$$
$$\times\exp[-(z+z')s_1]$$
$$=\frac{\hat\beta(z)}{z+z'}+\int_0^\infty ds_1\left(-\frac{1}{z+z'}\frac{d}{ds_1}\right.$$
$$\left.\times\exp[-(z+z')s_1]\right)\int_{-s_1}^0\exp(-z\sigma)\,\beta(|\sigma|)\,d\sigma$$
$$=\frac{\hat\beta(z)}{z+z'}+\int_0^\infty ds_1\,\frac{\exp[-(z+z')s_1]}{z+z'}\exp(zs_1)\beta(s_1)$$
$$=\frac{\hat\beta(z)+\hat\beta(z')}{z+z'}\ . \tag{A3}$$

Therefore, we get the identity

$$\left\langle\left[\int_0^{t_2}\chi(t_2-s_2)\tilde f(s_2)\,ds_2\right]\left[\int_0^{t_1}\chi(t_1-s_1)\tilde f(s_1)\,ds_1\right]\right\rangle$$
$$=k_BTm\hat\chi(z)\hat\chi(z')\frac{\hat\beta(z)+\hat\beta(z')}{z+z'}\ . \tag{A4}$$

Using the definition of $\hat\chi(z)$ in (5) gives

$$\hat\chi(z)\hat\beta(z)=1-z\hat\chi(z)\quad\text{and}\quad\hat\chi(z')\hat\beta(z')=1-z'\hat\chi(z'). \tag{A5}$$

These two identities yield

$$k_BTm\hat\chi(z)\hat\chi(z')\frac{\hat\beta(z)+\hat\beta(z')}{z+z'}$$
$$=k_BTm\left(\frac{\hat\chi(z')+\hat\chi(z)}{z+z'}-\hat\chi(z)\hat\chi(z')\right). \tag{A6}$$

In parallel with the identity in (A3), we conclude that (A6) is the double Laplace transform of

$$k_BTm(\chi(|t_2-t_1|)-\chi(t_2)\chi(t_1))$$

which completes the proof of (8) because $t_2\geq t_1$.

APPENDIX B; PROOF OF (25)

It will suffice to verify, for $t_2\geq t_1$, that

$$\langle\{\int_0^{t_2}ds_2\exp[-\int_{s_2}^{t_2}\beta(s')\,ds']\tilde f(s_2)/m\}$$
$$\times\{\int_0^{t_1}ds_1\exp[-\int_{s_1}^{t_1}\beta(s')\,ds']\tilde f(s_1)/m\}\rangle$$
$$=k_BTm[\exp[-\int_{t_1}^{t_2}\beta(s)\,ds]-\exp[-\int_0^{t_2}\beta(s)\,ds)]$$
$$\times\exp[-\int_0^{t_1}\beta(s')\,ds']]. \tag{B1}$$

From (23) we get

$$\langle\{\int_0^{t_2}ds_2\exp[-\int_{s_2}^{t_2}\beta(s')\,ds']\tilde f(s_2)/m\}$$
$$\times\{\int_0^{t_1}ds_1\exp[-\int_{s_1}^{t_1}\beta(s')\,ds']\tilde f(s_1)/m\}\rangle$$
$$=2(k_BT/m)\int_0^{t_2}ds_2\int_0^{t_1}ds_1\exp[-\int_{s_2}^{t_2}\beta(s')\,ds']$$
$$\times\exp[-\int_{s_1}^{t_1}\beta(s)ds]\beta(s_2)\delta(s_2-s_1)$$
$$=2(k_BT/m)\exp[-\int_{t_1}^{t_2}\beta(s)ds]\int_0^{t_1}ds_2\int_0^{t_1}ds_1$$
$$\times\exp[-\int_{s_2}^{t_1}\beta(s')\,ds']$$
$$\times\exp[-\int_{s_1}^{t_1}\beta(s)\,ds]\beta(s_2)\delta(s_2-s_1)$$
$$=2(k_BT/m)\exp[-\int_{t_1}^{t_2}\beta(s)\,ds]\int_0^{t_1}ds_2\beta(s_2)$$
$$\times\exp[-2\int_{s_2}^{t_1}\beta(s)\,ds]$$

$$= 2(k_B T/m) \exp[-\int_{t_1}^{t_2} \beta(s)\, ds] \int_0^{t_1} ds_2 \tfrac{1}{2}(d/ds_2)$$
$$\times \exp[-2\int_{s_2}^{t_1} \beta(s)\, ds]$$
$$= 2(k_B T/m) \exp[-\int_{t_1}^{t_2} \beta(s)\, ds]\{\tfrac{1}{2} - \tfrac{1}{2} \exp[-2\int_0^{t_1} \beta(s)\, ds]\}$$
$$= (k_B T/m)[\exp \int_{t_1}^{t_2} \beta(s)\, ds] - \exp[\int_0^{t_2} \beta(s)\, ds]$$
$$\times \exp[-\int_0^{t_1} \beta(s')\, ds']], \qquad \text{(B2)}$$

which completes the proof.

[1] S. A. Adelman, J. Chem. Phys. **64**, 124 (1976).
[2] R. F. Fox, J. Stat. Phys. **16**, 259 (1977).
[3] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).
[4] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945). See note II of the Appendix.
[5] See Ref. 4, Sec. 3.
[6] L. Arnold, *Stochastic Differential Equations: Theory and Applications* (Wiley-Interscience, New York, 1974).
[7] See Ref. 4, Sec. 7.
[8] J. L. Doob, Ann. Math. **43**, 351 (1942).

N-body quantum scattering theory in two Hilbert spaces. I. The basic equations

Colston Chandler

Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131

A. G. Gibson

Centre de Physique Théorique, C.N.R.S., 13274 Marseille, France and Department of Mathematics and Statistics, [a] *University of New Mexico, Albuquerque, New Mexico 87131*

(Received 28 December 1976)

Derivations are given for some transition and resolvent operator equations for multichannel quantum scattering with short-range potentials. The basic difference between these and previous equations is that the unknown operators act only on the channel subspaces. This is made possible by utilizing, and extending, the two-Hilbert-space formulation previously given by the authors [in J. Math. Phys. **14**, 1328 (1973)]. The equations in abstract form are of the Lippmann-Schwinger type, differing only in the appearance of certain injection operators from one Hilbert space to the other. When applied to multichannel quantum scattering, the abstract theory yields a new system of equations for the transition and resolvent operators. Uniqueness of the solution to the equations is proved.

1. INTRODUCTION

Since the pioneering work of Faddeev on the three-body problem,[1,2] nonrelativistic scattering processes with three or more particles have been studied with considerable interest. Several alternatives to the Faddeev equations and their many-body generalizations have been proposed.[3–16] The goal is to solve for the N-body scattering and/or resolvent operator. This is normally done indirectly by writing equations for transition operators $U_{\beta\alpha}$, or some related operators. A common feature of all of these equations is that the unknown quantities, say $U_{\beta\alpha}$, are operators acting in the full N-body Hilbert space H_N for all channels α, β. At the end of the calculation it is then recognized that the operators $U_{\beta\alpha}$ really provide more information than needed (cf. Refs. 14–18). All that is really needed are the operators $T_{\beta\alpha}=P_\beta U_{\beta\alpha}P_\alpha$, where P_α and P_β are projections onto the initial and final channel subspaces $H_\alpha=P_\alpha H_N$ and $H_\beta=P_\beta H_N$, respectively. In order to compensate for this excess information, and to simplify the theory, bound state pole approximation or quasiparticle methods are then applied, one purpose of which is to obtain effective equations for the operators $T_{\beta\alpha}$ instead of $U_{\beta\alpha}$.

We ask the question: If all that are needed from the beginning are the operators $T_{\beta\alpha}$ acting on the "small" spaces H_α and H_β, then why use equations for operators which act on the big space H_N? An examination of the derivations given in Refs. 1–18 reveals that the basic reason why equations are not written directly for the $T_{\beta\alpha}$ operators, or some related channel operators, is that this would require something analogous to inserting $P_\beta^{-1}P_\beta$ into the kernels of the equations. Since P_β is a singular operator for all channels except the free channel, the operator P_β^{-1}, of course, does not exist.

It is our thesis that several of the problems of N-body scattering theory have been magnified by always working in the full Hilbert space H_N. In this paper we present the beginning of a theory of multichannel scattering in which the projections onto the asymptotic channel states are incorporated throughout the calculation.

Channel projections have previously been included in the equations of Osborn–Kowalski,[14] and Karlsson–Zeiger.[15,16] However, what we are proposing here is quite different. The equations derived in Refs. 14–16 incorporate the channel projections P_α and P_β only at the input and output stages of the calculation. We propose to work only on the channel subspaces from the beginning to the end of the calculation.

The mathematical vehicle for our work is the two-Hilbert-space formulation of multichannel scattering given in Ref. 19. The results of paper[19] form the starting point of our present work, and the two together constitute a rigorous derivation of our equations, complete from the time-dependent theory of multichannel scattering developed in Refs. 20–24.

In the two-Hilbert-space formulation, the second "auxiliary" Hilbert space may be the direct sum of the channel spaces H_α or the direct sum of certain cluster spaces H_A. Its function is to simultaneously keep track of all the subspaces in an efficient manner. Communication between the two Hilbert spaces is provided by an injection operator J and its adjoint J^*. The key to deriving equations which preserve the channel or cluster subspaces is to invert the operator JJ^*. The operator JJ^* is the complete system analog of the single channel projection operator P_β. In contrast to the operator P_β, however, the operator JJ^* is nonsingular. Once all possible cluster arrangements of the N-body system are specified, then JJ^* and $(JJ^*)^{-1}$ are given explicitly. They are fixed operators for all channels of the system. The operator JJ^* is the sum of all channel projection operators for the system, and $(JJ^*)^{-1}$ is a converging Neumann series of products of projection operators.

We begin in Sec. 2 by presenting in abstract form a

[a] Present address.

two-Hilbert-space theory of scattering which is sufficiently general to include the case of multichannel quantum scattering theory. Except for the presence of the injection operators J, J^*, and $(JJ^*)^{-1}$, the basic equations derived in Sec. 2 take on the form of the usual Lippmann–Schwinger and resolvent operator equations. The derivations of these equations, and the proofs that their solutions are unique, are, however, considerably complicated by the presence of the J operators.

In Sec. 3 we consider the situation of N-body quantum mehcanical scattering with short-range potentials. The assumptions of Sec. 2 are shown to be satisfied if the auxiliary Hilbert space H is a direct sum of either the channel spaces H_α or of certain cluster spaces H_A. The abstract equations of Sec. 2 then give some new systems of N-body equations for the transition operators and resolvent operators. The distinctive feature of these equations is that the unknown operators act only on the channel (or cluster) subspaces of H_N, thus fulfilling our goal.

The paper is concluded in Sec. 4 with an additional discussion of the results.

2. THE ABSTRACT THEORY

A. Basic assumptions

The following abstract assumptions define the time-dependent scattering theory to be studied in this paper. These assumptions are motivated by considerations of multichannel quantum scattering theory, the relevant aspects of which are discussed at greater length in Sec. 3. The abstract theory, rather than the specific quantum mechanical theory that motivates it, is pursued in order to render the basic structure of the theory as clear as possible, and also to have the results as widely applicable as possible.

(A1) A spectral family $E_N(\lambda)$ defines a self-adjoint linear operator (total Hamiltonian) $H_N \equiv \int \lambda\, dE_N(\lambda)$ with domain $\mathcal{D}(H_N)$ dense in a separable Hilbert space H_N.

(A2) A spectral family $E(\lambda)$ defines a self-adjoint linear operator $H \equiv \int \lambda\, dE(\lambda)$ with domain $\mathcal{D}(H)$ dense in a separable Hilbert space H. The operator H has an absolutely continuous spectrum consisting of a half-line.

(A3) A bounded linear operator $J: H \to H_N$ is defined. The operator J maps $\mathcal{D}(H)$ into $\mathcal{D}(H_N)$, and the adjoint operator J^* maps $\mathcal{D}(H_N)$ into $\mathcal{D}(H)$. The operator $JJ^*: H_N \to H_N$ has a bounded inverse.

(A4) Potential operators $V: \mathcal{D}(V) \to H_N$ and $V^*: \mathcal{D}(V^*) \to H$ are defined by

$$V \equiv H_N J - JH \tag{2.1}$$

and

$$V^* \equiv J^* H_N - HJ^*. \tag{2.2}$$

A consequence of Assumption (A3) is that $\mathcal{D}(V) \supset \mathcal{D}(H)$ and $\mathcal{D}(V^*) \supset \mathcal{D}(H_N)$. For all positive ϵ and ϵ^* there exist finite b and b^* such that

$$\|V\Phi\| \leq \epsilon \|H\Phi\| + b\|\Phi\|, \tag{2.3}$$

for all $\Phi \in \mathcal{D}(H)$, and

$$\|V^*\psi\| \leq \epsilon^* \|H_N \psi\| + b^* \|\psi\|, \tag{2.4}$$

for all $\psi \in \mathcal{D}(H_N)$.

(A5) Wave operators $\Omega^\pm: H \to H_N$,

$$\Omega^\pm \equiv \operatorname*{s-lim}_{t \to \pm\infty} \exp(iH_N t) J \exp(-iHt), \tag{2.5}$$

are defined on H. The adjoint wave operators $\Omega^{\pm *}: H_N \to H$ have on H_N the representation

$$\Omega^{\pm *} = \operatorname*{w-lim}_{t \to \pm\infty} \exp(iHt) J^* \exp(-iH_N t). \tag{2.6}$$

The wave operators are partial isometries,

$$\Omega^{\pm *}\Omega^\pm = I \quad \text{and} \quad \Omega^\pm \Omega^{\pm *} = E_N^\pm, \tag{2.7}$$

where I is the identity on H, and where the operators $E_N^\pm$ are the orthogonal projections of H_N onto the ranges of $\Omega^\pm$. The operators $\Omega^\pm$, in addition, map $\mathcal{D}(H)$ into $\mathcal{D}(H_N)$, and on $\mathcal{D}(H)$ they satisfy the intertwining relation

$$H_N \Omega^\pm = \Omega^\pm H. \tag{2.8}$$

(A6) The scattering operator $S: H \to H$ is defined by

$$S = \Omega^{+ *}\Omega^-. \tag{2.9}$$

It is unitary if and only if $E_N^+ = E_N^-$.

In these assumptions and throughout this paper, we have adopted the notational convention that all operators written with a subscript [for example, $H_N, E_N(\lambda)$, and $E_N^\pm$] map H_N, or a subspace of H_N, into H_N. All operators written without a subscript (for example, H, J, J^*, $V, \Omega^\pm$, and S) map from and/or to H.

The validity of Assumptions (A1)–(A6) with the exception of the existence of a bounded inverse of JJ^* and of (A4), has been established previously[19] for a class of scattering systems that includes systems of distinguishable spinless particles interacting via square integrable potentials. The validity of (A4) and the boundedness of $(JJ^*)^{-1}$ is established in Sec. 3 for a similar class of systems.

It is to be emphasized that for multichannel quantum mechanical systems the weak convergence in Eq. (2.6) cannot be replaced by strong convergence.[19] This point, which is commonly overlooked, means that most of the elegant two-Hilbert-space results of Refs. 25 and 26 are unavailable for our use.

B. Time-independent transition operators

Assumptions (A1)–(A6) define a time-dependent theory. Using techniques of spectral integration, one can rigorously derive time-independent formulas for $\Omega^\pm$ and S. The result which we need is given in the following theorem.

Theorem 1[19,27,28]: Assume (A1)–(A3), (A5), and (A6). Define

$$\delta_{\epsilon_1}(\lambda - H) \equiv \int dE(\mu)(\epsilon_1/\pi)[(\lambda - \mu)^2 + \epsilon_1^2]^{-1} \tag{2.10}$$

for $\epsilon_1 > 0$. Then $S - I$ is given on H by the formulas

$$S - I = \operatorname*{w-lim}_{\epsilon_1 \to 0^+} \operatorname*{s-lim}_{\epsilon_2 \to 0^+} (-2\pi i) \int \delta_{\epsilon_1}(\lambda - H) T(\lambda + i\epsilon_2)\, dE(\lambda) \tag{2.11}$$

$$= \text{s-}\lim_{\epsilon_1 \to 0^+} \text{w-}\lim_{\epsilon_2 \to 0^+} (-2\pi i) \int dE(\lambda)\, T(\lambda + i\epsilon_2)\delta_{\epsilon_1}(\lambda - H). \tag{2.12}$$

The operator $T: \mathcal{D}(H) \to \mathcal{H}$ is defined by

$$T(z) \equiv (z - H)J^*(z - H_N)^{-1}J(z - H) - (z - H) \tag{2.13}$$

for all $z \in \rho(H_N)$, the resolvent set of H_N.

The operator $T(z)$ corresponds to the symmetric transition operators of Alt, Grassberger, and Sandhas.[6,7] It is not the only candidate for the transition operator. Other, nonsymmetric, candidates are $T^{(\pm)}(z): \mathcal{D}(H) \to \mathcal{H}$ defined by

$$T^{(+)}(z) \equiv (z - H)J^*(z - H_N)^{-1}J(z - H) - J^*J(z - H) \tag{2.14}$$

and

$$T^{(-)}(z) \equiv (z - H)J^*(z - H_N)^{-1}J(z - H) - (z - H)J^*J. \tag{2.15}$$

These operators correspond to the nonsymmetric operators used by Lovelace.[8]

Theorem 2: Under the assumptions of Theorem 1, Eq. (2.11) is also true with T replaced by $T^{(+)}$, and Eq. (2.12) is true with T replaced by $T^{(-)}$.

Proof: Since

$$T(z) - T^{(+)}(z) = (J^*J - I)(z - H) \tag{2.16}$$

and

$$T(z) - T^{(-)}(z) = (z - H)(J^*J - I) \tag{2.17}$$

the theorem follows from Theorem 1 and the following lemma.

Lemma 1: Under the assumptions of Theorem 1,

$$\text{s-}\lim_{\epsilon_2 \to 0^+} \int \delta_{\epsilon_1}(\lambda - H)(J^*J - I)(\lambda + i\epsilon_2 - H)\, dE(\lambda) = 0 \tag{2.18}$$

and

$$\text{s-}\lim_{\epsilon_2 \to 0^+} \int dE(\lambda)(\lambda + i\epsilon_2 - H)(J^*J - I)\delta_{\epsilon_1}(\lambda - H) = 0. \tag{2.19}$$

Proof: The operator $\delta_{\epsilon_1}(\lambda - H)(J^*J - I)$ is a bounded operator for all $\epsilon_1 > 0$. By Ref. 19, Lemma 5 the factor $(\lambda + i\epsilon_2 - H)$ in Eq. (2.18) may be replaced by $i\epsilon_2$. The result is $i\epsilon_2$ times a bounded operator which, therefore, converges strongly to zero as $\epsilon_2 \to 0^+$. This proves (2.18). The limit (2.19) is proved in the same way.

Q.E.D.

Remark: It can be shown by other methods that Eq. (2.11) is also true with T replaced by $T^{(-)}$. If the strong limit in Eq. (2.12) is replaced by a weak limit, then it is also true with T replaced by $T^{(+)}$. Proofs, as well as a discussion of the possibility of replacing all weak limits in Theorems 1 and 2 by strong limits, are contained in a separate paper in preparation. Alternative forms of the operators $T(z)$, $T^{(\pm)}(z)$ can be found by use of the resolvent equations. The resolvents $R(z) \equiv (z - H)^{-1}$ and $R_N(z) \equiv (z - H_N)^{-1}$ are related by the equations

$$R_N(z)J = JR(z) + R_N(z)VR(z) \tag{2.20}$$

and

$$J^*R_N(z) = R(z)J^* + R(z)V^*R_N(z), \tag{2.21}$$

where V and V^* are defined by Eqs. (2.1) and (2.2).

Since Assumption (A5) implies that the spectrum of H_N contains the spectrum of H, Eqs. (2.20) and (2.21) are valid for $z \in \rho(H_N)$.

Lemma 2: The operators T and $T^{(\pm)}$ equal, respectively,

$$T(z) = Z(z) + V^*R_N(z)V, \tag{2.22}$$

$$T^{(+)}(z) = V^*J + V^*R_N(z)V, \tag{2.23}$$

$$T^{(-)}(z) = J^*V + V^*R_N(z)V, \tag{2.24}$$

where

$$Z(z) \equiv (J^*J - I)(z - H) + V^*J = (z - H)(J^*J - I) + J^*V. \tag{2.25}$$

Proof: By Eqs. (2.13), (2.21), and (2.20),

$$\begin{aligned} T &= R^{-1}(RJ^* + RV^*R_N)JR^{-1} - R^{-1} \\ &= (J^*J - I)R^{-1} + V^*(JR + R_NVR)R^{-1} \\ &= Z + V^*R_NV. \end{aligned} \tag{2.26}$$

The two formulas for Z in Eq. (2.25) are equal because they both expand to

$$Z(z) = J^*H_NJ + H - J^*JH - HJ^*J + z(J^*J - I). \tag{2.27}$$

The formulas for $T^{(\pm)}$ are proved similarly. Q.E.D.

Time-independent scattering theory consists of the study of the operator $T(z)$, or of $T^{(\pm)}(z)$. If one of these operators is known, then Theorem 1 or 2 tells us how to calculate the scattering operator. The operator $J^*R_N(z)J$ could also be determined from one of the Eqs. (2.13)—(2.15). Our goal then is to derive equations for $T(z)$ and $T^{(\pm)}(z)$. We first look more closely at the operators JJ^* and $(JJ^*)^{-1}$.

C. The operator JJ^* and its inverse

By Assumption (A3) the operator JJ^* is bounded, has a bounded inverse, and maps $\mathcal{D}(H_N)$ into $\mathcal{D}(H_N)$. We would also like to know that the operator $(JJ^*)^{-1}$ maps $\mathcal{D}(H_N)$ into $\mathcal{D}(H_N)$. In order to show this, it is necessary for us to prove that JJ^* maps $\mathcal{D}(H_N)$ onto $\mathcal{D}(H_N)$.

Theorem 3 (Deift): Assume (A1)—(A4). Then JJ^* and $(JJ^*)^{-1}$ map $\mathcal{D}(H_N)$ onto $\mathcal{D}(H_N)$.

Proof: Define $K_N(y): \mathcal{H}_N \to \mathcal{H}_N$ by

$$K_N(y) \equiv R_N^{-1}(iy)JJ^*R_N(iy) \tag{2.28}$$

and let $L_N(y) \equiv K_N(y) - JJ^*$. Substituting

$$H_NJ = JH + V \tag{2.29}$$

and Eq. (2.21) into Eq. (2.28) yields

$$L_N(y) = JV^*R_N(iy) - VR(iy)J^* - VR(iy)V^*R_N(iy). \tag{2.30}$$

By Assumption (A4), for every $\epsilon > 0$ there is a b such that for all $\Phi \in \mathcal{H}$

$$\|VR(iy)\Phi\| \leq \epsilon\|HR(iy)\Phi\| + b\|R(iy)\Phi\|. \tag{2.31}$$

It follows that $VR(iy)$ is bounded, with the bound satisfying

$$\|VR(iy)\| \leq \epsilon + by^{-1}. \tag{2.32}$$

Therefore, $\|VR(iy)\|$ can be made as small as desired by choosing y sufficiently large. A similar conclusion can be drawn for $\|V^*R_N(iy)\|$. Hence, for sufficiently large y, one can conclude that $\|(JJ^*)^{-1}L_N(y)\| < 1$. It follows that

$$K_N(y) = (JJ^*)[I_N + (JJ^*)^{-1}L_N(y)] \tag{2.33}$$

is invertible for y sufficiently large, and K_N maps $\mathcal{H}_N$ onto $\mathcal{H}_N$. Therefore, for a given vector $\psi \in \mathcal{D}(H_N)$, there is a vector $\varphi \in \mathcal{H}_N$ such that $R_N^{-1}(iy)\psi = K_N(y)\varphi$. But this implies that $\psi = JJ^*\zeta$ for $\zeta = R_N(iy)\varphi \in \mathcal{D}(H_N)$. Since JJ^* maps $\mathcal{D}(H_N)$ into $\mathcal{D}(H_N)$ by Assumption (A3), we can now conclude that JJ^* maps $\mathcal{D}(H_N)$ onto $\mathcal{D}(H_N)$. Since $(JJ^*)^{-1}$ is assumed bounded, similar conclusions about $(JJ^*)^{-1}$ follow immediately. Q.E.D.

Related to the operator $(JJ^*)^{-1}$ are certain projection operators P and Q. Define Q to be the orthogonal projection of $\mathcal{H}$ onto $\mathcal{N}(J)$, the null space of J. Let $P \equiv I - Q$. Then P is the orthogonal projection of $\mathcal{H}$ onto $\mathcal{N}(J)^\perp = \overline{\mathcal{R}}(J^*)$, where $\mathcal{N}(J)^\perp$ denotes the orthogonal complement of $\mathcal{N}(J)$, and $\overline{\mathcal{R}}(J^*)$ denotes the closure of the range of J^*.

Lemma 3: $P = J^*(JJ^*)^{-1}J$. (2.34)

Proof: Let $\tilde{P} \equiv J^*(JJ^*)^{-1}J$. Clearly $\tilde{P}^2 = \tilde{P}$. Also,

$$\tilde{P}^* = [(JJ^*)^{-1}J]^*J = J^*(JJ^*)^{-1}J = \tilde{P} \tag{2.35}$$

(cf. Ref. 29, Prob. III-5.26]). Therefore, $\tilde{P}$ is an orthogonal projection. If $\Phi \in \mathcal{N}(P) = \mathcal{N}(J)$, then $\tilde{P}\Phi = 0$ and $\Phi \in \mathcal{N}(\tilde{P})$. On the other hand, if $\Phi \in \mathcal{N}(\tilde{P})$, then

$$J^*(JJ^*)^{-1}J\Phi = 0. \tag{2.36}$$

Multiplication of Eq. (2.36) on the left by J yields $J\Phi = 0$. Hence, $\Phi \in \mathcal{N}(J)$. Therefore, P and $\tilde{P}$ are both orthogonal projections with null space $\mathcal{N}(J)$, and $P = \tilde{P}$.

Q.E.D.

D. Lippman–Schwinger-type equations

The first step in deriving equations of the Lippmann–Schwinger type is summarized in the following lemma.

Lemma 4: The transition operator $T(z)$ satisfies the identities

$$I + T(z)R(z) = J^*J + V^*R_N(z)J \tag{2.37}$$

and

$$I + R(z)T(z) = J^*J + J^*R_N(z)V. \tag{2.38}$$

The transition operators $T^{(\pm)}(z)$ satisfy the identities

$$T^{(+)}(z)R(z) = V^*R_N(z)J \tag{2.39}$$

and

$$R(z)T^{(-)}(z) = J^*R_N(z)V, \tag{2.40}$$

respectively.

Proof: By Eqs. (2.13) and (2.21)

$$I + TR = R^{-1}J^*R_NJ = J^*J + V^*R_NJ. \tag{2.41}$$

The other identities are proved similarly. Q.E.D.

Converting these equations into equations of the Lippmann–Schwinger type is now straightforward.

Theorem 4: The operator $T = T(z)$ is, for $z \in \rho(H_N)$, a solution of the equations

$$T = (J^*J - I)R^{-1} + V^*(JJ^*)^{-1}J + V^*(JJ^*)^{-1}JRT \tag{2.42}$$

and

$$T = R^{-1}(J^*J - I) + J^*(JJ^*)^{-1}V + TRJ^*(JJ^*)^{-1}V. \tag{2.43}$$

The operators $T^{(\pm)}(z)$ are solutions of

$$T^{(+)} = V^*J + T^{(+)}RJ^*(JJ^*)^{-1}V \tag{2.44}$$

and

$$T^{(-)} = J^*V + V^*(JJ^*)^{-1}JRT^{(-)}, \tag{2.45}$$

respectively.

Proof: Let T^τ be defined by the right side of Eq. (2.42),

$$T^\tau = (J^*J - I)R^{-1} + V^*(JJ^*)^{-1}J[I + RT], \tag{2.46}$$

with the T on the right side of Eq. (2.46) being given by the defining equation (2.13). Then by Lemma 4,

$$T^\tau = (J^*J - I)R^{-1} + V^*(JJ^*)^{-1}J[J^*J + J^*R_NV] \tag{2.47}$$

$$= (J^*J - I)R^{-1} + V^*J + V^*R_NV. \tag{2.48}$$

Comparing Eqs. (2.48) and (2.22), we see that $T^\tau = T$, and, hence, that T is a solution of Eq. (2.42). The other equations are verified in a similar manner.

Q.E.D.

Corresponding equations also hold for resolvent operators. In particular, if the "resolvent" operator $G: \mathcal{H}_N \to \mathcal{H}$ is defined for $z \in \rho(H_N)$ by

$$G(z) \equiv J^*R_N(z), \tag{2.49}$$

then the following theorem is immediate from Eq. (2.21).

Theorem 5: The operator $G = G(z)$ is a solution of

$$G = RJ^* + RV^*(JJ^*)^{-1}JG. \tag{2.50}$$

The importance of Theorem 3 is now apparent. The operator V^*, for example, is well defined on $\mathcal{D}(H_N)$. But for Eqs. (2.42), (2.45), and (2.50) to be well defined it is necessary that $V^*(JJ^*)^{-1}J$ be well defined on $\mathcal{D}(H)$. By Assumption (A3) the operator J maps $\mathcal{D}(H)$ into $\mathcal{D}(H_N)$. The fact that $V^*(JJ^*)^{-1}J$ is well defined on $\mathcal{D}(H)$ then depends on $(JJ^*)^{-1}$ mapping $\mathcal{D}(H_N)$ into $\mathcal{D}(H_N)$. This is just the link provided by Theorem 3.

E. Uniqueness of solutions

We now turn to the question of uniqueness for the equations of Theorems 4 and 5. An outline of the proof of uniqueness for Eq. (2.42) was given in Ref. 30. For the first steps of a complete proof, we establish two lemmas.

Lemma 5: If Q is the orthogonal projection of $\mathcal{H}$ onto $\mathcal{N}(J)$, then the operator QHQ is self-adjoint. The spectrum of QHQ is contained in the spectrum of H.

Proof: The operator QHQ is clearly symmetric. To prove it is self-adjoint, it is sufficient (Ref. 29, Theorem V-3.16) to prove that for some $y > 0$ the ranges of the operators

$$W^\pm \equiv QHQ \pm iyI \tag{2.51}$$

are the entire space $\mathcal{H}$.

To prove this we write $W^{\pm}$ as

$$W^{\pm} = W^{\pm}P + W^{\pm}Q, \tag{2.52}$$

where $P \equiv I - Q$. Multiplying Eq. (2.51) from the right by P yields

$$W^{\pm}P = \pm iyP. \tag{2.53}$$

It is obvious that for any $y > 0$, the operator $W^{\pm}P$ maps $\mathcal{H}$ onto $P\mathcal{H}$.

The operator $W^{\pm}Q$ can be rewritten in the form

$$W^{\pm}Q = Q[HQ \pm iyI] \tag{2.54}$$

$$= Q[H \pm iyI - HP] \tag{2.55}$$

$$= Q[I - HP(H \pm iyI)^{-1}](H \pm iyI). \tag{2.56}$$

By Lemma 3 and Eq. (2.2),

$$HP = HJ^*(JJ^*)^{-1}J \tag{2.57}$$

$$= J^*H_N(JJ^*)^{-1}J - V^*(JJ^*)^{-1}J. \tag{2.58}$$

Substitution of this into Eq. (2.56), and noting that $QJ^* = 0$, yields

$$W^{\pm}Q = Q[I + V^*(JJ^*)^{-1}J(H \pm iyI)^{-1}](H \pm iyI). \tag{2.59}$$

Since $(H \pm iyI)^{-1}$ maps $\mathcal{H}$ into $\mathcal{D}(H)$, Assumption (A4) implies that for every $\epsilon > 0$ there is a finite b such that

$$\|V^*(JJ^*)^{-1}J(H \pm iyI)^{-1}\Phi\| \leq \epsilon \|H_N(JJ^*)^{-1}J(H \pm iyI)^{-1}\Phi\| + b\|(JJ^*)^{-1}J(H \pm iyI)^{-1}\Phi\|. \tag{2.60}$$

By Assumption (A3), Theorem 3, and Ref. 19, Lemma 1, the operators $H_N(JJ^*)^{-1}J(H \pm iyI)^{-1}$ are bounded in norm for any $y > 0$. Let the bounds be $b^{\pm}(y)$. Further, since

$$\|(H \pm iy)(H \pm iy')^{-1}\Phi\| \leq \|\Phi\| \tag{2.61}$$

for all $y' \geq y$, it follows that $b^{\pm}(y') \leq b^{\pm}(y)$ for $y' \geq y$. Choose now some $y_0 > 0$ and choose ϵ so that $\epsilon b^{\pm}(y_0) < \frac{1}{2}$. Then $\epsilon b^{\pm}(y) < \frac{1}{2}$ for all $y > y_0$. Now choose y so that $y > 2b\|(JJ^*)^{-1}\|\|J\|$. Then the right side of Eq. (2.60) is strictly less than $\|\Phi\|$. It follows that the operator $I + V^*(JJ^*)^{-1}J(H \pm iyI)^{-1}$ in Eq. (2.59) has a bounded inverse and maps $\mathcal{H}$ onto $\mathcal{H}$. Since H is self-adjoint, the operator $H \pm iyI$ in Eq. (2.59) maps $\mathcal{D}(H)$ onto $\mathcal{H}$. By following the sequence of operations in Eq. (2.59), one now sees that $W^{\pm}Q$ maps $\mathcal{D}(H)$ onto $Q\mathcal{H}$.

Choose now some Ψ in $\mathcal{H}$. Then $P\Psi$ lies in $P\mathcal{H}$ and is the image under $W^{\pm}P$ of vectors $\Theta^{\pm}$ in $\mathcal{H}$. The vector $Q\Psi$ lies in $Q\mathcal{H}$ and is the image under $W^{\pm}Q$ of vectors $\Phi^{\pm}$ in $\mathcal{D}(H)$. Since $PQ = QP = 0$, it follows that $\Psi = P\Psi + Q\Psi$ is the image under $W^{\pm}$ of $P\Theta^{\pm} + Q\Phi^{\pm}$. Since Ψ was arbitrary, the operator QHQ is self-adjoint.

Let $\sigma(\cdot)$ and $\bar{\omega}(\cdot)$ denote the spectrum and closure of the numerical range of an operator, respectively. Since QHQ is self-adjoint, $\sigma(QHQ) \subset \bar{\omega}(QHQ)$ (cf. Ref. 31, p. 309). But $\bar{\omega}(QHQ) \subset \bar{\omega}(H)$, since $(\Phi, QHQ\Phi) = (\Psi, H\Psi)$ for $\Psi = Q\Phi$. Finally, it follows from Assumption (A2) that $\bar{\omega}(H) = \sigma(H)$. Q.E.D.

Let $\mathcal{N}(K)$, $\mathcal{R}(K)$, and $\bar{\mathcal{R}}(K)$, respectively, denote the null space, range, and closure of the range of an operator K.

Lemma 6: Let

$$\hat{H} \equiv H + J^*(JJ^*)^{-1}V \tag{2.62}$$

and

$$\hat{H}^* \equiv H + V^*(JJ^*)^{-1}J. \tag{2.63}$$

If $z \in \rho(H_N)$, then

$$\mathcal{N}(z - \hat{H}) = \mathcal{N}(z - \hat{H}^*) = \{0\} \tag{2.64}$$

and

$$\bar{\mathcal{R}}(z - \hat{H}) = \bar{\mathcal{R}}(z - \hat{H}^*) = \mathcal{H}. \tag{2.65}$$

Proof: Let $\Phi \in \mathcal{D}(H) = \mathcal{D}(z - \hat{H})$, and suppose that

$$(z - \hat{H})\Phi = 0. \tag{2.66}$$

Multiplication of Eq. (2.66) on the left by J yields

$$(z - H_N)J\Phi = 0, \tag{2.67}$$

and a multiplication by $R_N(z)$ gives $J\Phi = 0$. Then $\Phi \in \mathcal{N}(J) = Q\mathcal{H}$, and $\Phi = Q\Phi$. Since $JQ = 0$, and $Q^2 = Q$, replacing Φ by $Q\Phi$ in Eq. (2.66) gives

$$(z - QHQ)Q\Phi = 0. \tag{2.68}$$

By Lemma 5, the operator QHQ is self-adjoint, and $z - QHQ$ is invertible for $z \in \rho(H_N) \subset \rho(QHQ)$. Hence, Eq. (2.68) has only the trivial solution $Q\Phi = \Phi = 0$. This proves that $\mathcal{N}(z - \hat{H}) = \{0\}$.

Now suppose that $\Phi \in \mathcal{N}(z - \hat{H}^*)$. Then

$$(z - HQ)\Phi = J^*H_N(JJ^*)^{-1}J\Phi. \tag{2.69}$$

Multiplication of Eq. (2.69) on the left by Q again gives Eq. (2.68), and $Q\Phi = 0$. Substituting $Q\Phi = 0$ into Eq. (2.69) and multiplying on the left by $(JJ^*)^{-1}J$ reduces it to

$$(z - H_N)(JJ^*)^{-1}J\Phi = 0. \tag{2.70}$$

A further multiplication by $R_N(z)$ gives $(JJ^*)^{-1}J\Phi = 0$. Hence, $P\Phi = 0$. Consequently, $\Phi = P\Phi + Q\Phi = 0$. This proves Eqs. (2.64).

Since the operators $(z - \hat{H})$ and $(\bar{z} - \hat{H}^*)$ are densely defined and adjoint to each other, Eqs. (2.65) follow from Eqs. (2.64) by taking orthogonal complements (cf. Ref. 29, Sec. III-5.5). Q.E.D.

Theorem 6: If $z \in \rho(H_N)$, then each of Eqs. (2.42)–(2.45) has a unique solution defined on $\mathcal{D}(H)$, and Eq. (2.50) has a unique solution defined on $\mathcal{H}_N$.

Proof: Suppose there are two solutions of Eq. (2.42) or (2.45), and let X denote their difference. Then

$$(I - V^*(JJ^*)^{-1}JR)X\Phi = 0 \tag{2.71}$$

for all $\Phi \in \mathcal{D}(H)$. Let $\Psi = RX\Phi$. Then $\Psi \in \mathcal{D}(H)$, and $(z - H)\Psi = X\Phi$. Eq. (2.71) becomes

$$(z - \hat{H}^*)\Psi = 0, \tag{2.72}$$

with $\hat{H}^*$ defined by Eq. (2.63). By Lemma 6, $\Psi = 0$. Thus $X\Phi = 0$ for all $\Phi \in \mathcal{D}(H)$, and the solutions of Eqs. (2.42) and (2.45) are unique.

Suppose next that there are two solutions, defined on $\mathcal{D}(H)$, of Eq. (2.43) or (2.44). Let Y denote their difference. Then

$$Y(I - RJ^*(JJ^*)^{-1}V)\Phi = 0 \tag{2.73}$$

for all Φ in $\mathcal{D}(H)$. This implies that

$$(\Phi, [I - V^*(JJ^*)^{-1}JR(\bar{z})]Y^*\Psi)$$
$$= (Y[I - R(z)J^*(JJ^*)^{-1}V]\Phi, \Psi) = 0 \tag{2.74}$$

for all Φ and Ψ in $\mathcal{D}(H)$. Since $\mathcal{D}(H)$ is dense in $\mathcal{H}$, it follows that

$$[I - V^*(JJ^*)^{-1}JR(\bar{z})]Y^*\Psi = 0 \tag{2.75}$$

for all Ψ in $\mathcal{D}(H)$. But then $Y^*\Psi = 0$, as before. Hence,

$$(Y\Phi, \Psi) = (\Phi, Y^*\Psi) = 0 \tag{2.76}$$

for all Φ and Ψ in $\mathcal{D}(H)$, which implies $Y\Phi = 0$.

Similarly, if there are two solutions of Eq. (2.50), with a difference X, then

$$(I - RV^*(JJ^*)^{-1}J)X\varphi = 0 \tag{2.77}$$

for all φ in $\mathcal{H}_N$. Eq. (2.77) implies that $X\varphi$ lies in $\mathcal{D}(H)$. We can then multiply on the left by $(z - H)$ to obtain

$$(z - \hat{H}^*)X\varphi = 0. \tag{2.78}$$

By Lemma 6, $X\varphi = 0$. Q.E.D.

3. MULTICHANNEL QUANTUM SCATTERING THEORY

A. Properties of N-body systems

To lend substance to the abstract development of the last section we review in some detail the pertinent features of quantum scattering theory for systems of N distinguishable spinless particles that interact pairwise via short-range forces.

The system as a whole is described by a wavefunction $\psi(\mathbf{X}, t)$ that depends on the time t and the positions $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ of the particles. For each fixed time t the wavefunction ψ belongs to the separable Hilbert space $\mathcal{H}_N = L^2(\mathbf{R}^{3N})$ of functions square integrable in the variable $\mathbf{X}$.

The dynamics of the system is specified by the N-body Hamiltonian

$$H_N = \sum_i (2m_i)^{-1}(-\Delta_i) + \sum_{i<j} V_{ij} = H_0 + V_N, \tag{3.1}$$

where units with $\hbar = 1$ have been adopted. The symbol m_i denotes the mass of particle i, and Δ_i denotes the Laplacian with respect to $\mathbf{x}_i$. H_0 is the sum over Laplacians in Eq. (3.1), and V_N is the total N-body potential given by the sum over pair potentials V_{ij} in Eq. (3.1).

Assumption B: The pair potentials $V_{ij}(\mathbf{x}_i - \mathbf{x}_j) = V_{ji}$ are assumed to be real-valued and to satisfy the condition

$$V_{ij} \in L^2(\mathbf{R}^3) + L^p(\mathbf{R}^3) \tag{3.2}$$

for some p, $2 \leq p < 3$.

Asymptotically the particles are arranged in clusters, each of which is in a specific quantum mechanical bound state. We denote different clusterings by $A, B, C, \cdots$. A specification of the clustering together with the bound states is said to specify a channel. We denote different channels by $\alpha, \beta, \gamma, \cdots$.

A clustering A is specified by a partitioning of the index set $\{1, \ldots, N\}$ into disjoint subsets (fragments) $a_1, \ldots, a_k$. We indicate that particle i belongs to fragment a_l by writing $i \in a_l$.

To specify a particular set of bound states for the clustering A one first defines a cluster Hamiltonian

$$H_A = \sum_{l=1}^{k} H_A(a_l), \tag{3.3}$$

where

$$H_A(a_l) \equiv \sum_{i \in a_l} (2m_i)^{-1}(-\Delta_i) + \sum_{\substack{i<j \\ i,j \in a_l}} V_{ij}. \tag{3.4}$$

Second, the operators $H_A(a_l)$ are written as

$$H_A(a_l) = H_A^0(a_l) + \hat{H}_A(a_l), \tag{3.5}$$

where the Hamiltonian $H_A^0(a_l)$ governs the free motion of the center of mass of fragment a_l and $\hat{H}_A(a_l)$ governs the motion of the particles in fragment a_l relative to their center of mass. The normalized eigenvectors $\hat{\varphi}_l$ of $\hat{H}_A(a_l)$ then describe the bound states of the fragment a_l. By convention, $\hat{\varphi}_l \equiv 1$ if a_l contains only one particle, and $\hat{\varphi}_l \equiv 0$ if fragment a_l has more than one particle but $\hat{H}_A(a_l)$ has no eigenvectors.

The channel wavefunctions ψ_α for a channel α are then of the form

$$\psi_\alpha = f(\mathbf{Y}, t) \prod_{l=1}^{k} \hat{\varphi}_l, \tag{3.6}$$

where the $\hat{\varphi}_l$ are the bound state wavefunctions appropriate to the channel α. The function f describes the motion of the centers of mass of the clusters and at each time t is a square integrable function of the variables $\mathbf{Y}$. The functions $\hat{\varphi}_l$ are, of course, also square integrable in the variables on which they depend. At each time t, therefore, the channel wavefunctions ψ_α belong to the subspace $\mathcal{H}_\alpha$ of $\mathcal{H}_N$ consisting of the closed linear span of vectors of the form (3.6).

The dynamics of channel α is governed by a channel Hamiltonian H_α, which is the restriction to $\mathcal{H}_\alpha$ of the appropriate cluster Hamiltonian H_A.

The standard scattering theory is now formulated by requiring wavefunctions $\psi(t)$ of the full system to evolve as $t \to \pm\infty$ to channel wavefunctions $\psi_\alpha(t)$. This leads to the definition and study of the Møller wave operators

$$\Omega_\alpha^\pm \equiv \operatorname*{s-lim}_{t \to \pm\infty} \exp(iH_N t) \exp(-iH_\alpha t) P_\alpha, \tag{3.7}$$

where P_α is the orthogonal projection of $\mathcal{H}_N$ onto $\mathcal{H}_\alpha$.

Several general properties of the theory can now be usefully abstracted.

Properties:

(P1) The temporal evolution of the complete N-particle system is governed by a one-parameter group $\exp(-iH_N t)$. The Hamiltonian H_N is a linear self-adjoint operator with domain $\mathcal{D}(H_N)$ dense in a separable Hilbert space $\mathcal{H}_N$.

(P2) The temporal evolution of each channel α is governed by a one-parameter group $\exp(-iH_\alpha t)$. The channel Hamiltonians H_α are linear self-adjoint opera-

tors with domains $\mathcal{D}(H_\alpha)$ dense in separable Hilbert spaces $\mathcal{H}_\alpha \subset \mathcal{H}_N$. The Hamiltonians H_α are bounded from below and have absolutely continuous spectra consisting of half-lines. The Hilbert space $\mathcal{H}_0$ corresponding to a clustering with only one particle per cluster is the entire space $\mathcal{H}_N$. The Hamiltonians H_α for all channels α with fixed clustering A have a common self-adjoint extension H_A, called the cluster Hamiltonian, with domain $\mathcal{D}(H_A) = \mathcal{D}(H_N)$ dense in $\mathcal{H}_N$.

(P3) The orthogonal projections P_α of $\mathcal{H}_N$ onto $\mathcal{H}_\alpha$ satisfy the limited orthogonality relation

$$P_\alpha P_\beta = \delta_{\alpha\beta} P_\beta, \tag{3.8}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta, for channels α and β with the same clustering. This orthogonality relation is not in general true if α and β do not have the same clustering. The operator P_0 which projects $\mathcal{H}_N$ onto $\mathcal{H}_0$ is the identity I_N.

(P4) For all clusterings A the cluster potential $\bar{V}_A \equiv H_N - H_A$ is symmetric and has domain $\mathcal{D}(\bar{V}_A)$ that contains $\mathcal{D}(H_N) = \mathcal{D}(H_A)$. For every $\epsilon > 0$ there exists $b > 0$ such that the inequality

$$\|\bar{V}_A \psi\| \leq \epsilon \|H_A \psi\| + b\|\psi\| \tag{3.9}$$

is true for all ψ in $\mathcal{D}(H_A) = \mathcal{D}(H_N)$.

(P5) The channel wave operators $\Omega_\alpha^\pm : \mathcal{H}_\alpha \to \mathcal{H}_N$ defined by Eq. (3.7) exist for all channels α. The orthogonal projections $E_\alpha^\pm$ of $\mathcal{H}_N$ onto the ranges of $\Omega_\alpha^\pm$ satisfy

$$E_\alpha^\pm E_\beta^\pm = \delta_{\alpha\beta} E_\alpha^\pm \tag{3.10}$$

for all channels α and β. For all channels α the intertwining relation

$$H_N \Omega_\alpha^\pm = \Omega_\alpha^\pm H_\alpha \tag{3.11}$$

holds on $\mathcal{D}(H_\alpha)$.

(P6) The scattering operators $S_{\beta\alpha} : \mathcal{H}_\alpha \to \mathcal{H}_\beta$ are defined by

$$S_{\beta\alpha} \equiv \Omega_\beta^{+*} \Omega_\alpha^- . \tag{3.12}$$

All of these properties, except perhaps Property (P4), are well-established in the literature (cf. Refs. 20–24). Property (P4) is also well known, but to our knowledge it has not been stated in quite the same way as here. For this reason, we include a proof.

The proof follows from an easily proved general property of relatively bounded operators with relative bound zero.

Definition 1[29,32,33]: Let K and L be two linear operators with the same domain space (but not necessarily with the same range space). Suppose

(i) the domain $\mathcal{D}(K)$ of K contains the domain $\mathcal{D}(L)$ of L, and

(ii) for every $\epsilon > 0$ there exists $b(\epsilon) > 0$ such that the inequality

$$\|K\psi\| \leq \epsilon \|L\psi\| + b\|\psi\|$$

holds for every $\psi \in \mathcal{D}(L)$.

Then the operator K is said to be *relatively bounded with respect to L with relative bound zero* (*or infinitesimally small with respect to L*), written $K \ll L$.

Lemma 7: Let $K_i : \mathcal{X} \to \mathcal{Y}$, where $1 \leq i \leq n$, and $L : \mathcal{X} \to \mathcal{Y}$ be a finite number of operators with domains in a Banach space $\mathcal{X}$ and ranges in a Banach space $\mathcal{Y}$. If $K_i \ll L$ for all values of i, then

$$\sum_{i=1}^n c_i K_i \ll L + \sum_{i=1}^n d_i K_i, \tag{3.13}$$

where the c_i and d_i are any complex numbers.

Proof: The domain property (i) of Definition 1 is clearly satisfied. Suppose $\epsilon > 0$ is given, and choose ϵ_1 so that

$$\epsilon_1 < \epsilon \left\{ \epsilon \sum_i |d_i| + \sum_i |c_i| \right\}^{-1}. \tag{3.14}$$

Then there exists b_1 such that

$$\|K_i \psi\| \leq \epsilon_1 \|L\psi\| + b_1 \|\psi\| \tag{3.15}$$

for all ψ in $\mathcal{D}(L)$ and all i. By the triangle inequality and Eq. (3.15) we have

$$\|\sum c_i K_i \psi\| \leq \sum |c_i| \, \|K_i \psi\| \leq \sum |c_i| \{\epsilon_1 \|L\psi\| + b_1 \|\psi\|\} \tag{3.16}$$

for all ψ in $\mathcal{D}(L)$. On the other hand, $L = L + \sum d_i K_i - \sum d_i K_i$. Again the triangle inequality and Eq. (3.15) imply that

$$\|L\psi\| \leq \|(L + \textstyle\sum d_i K_i)\psi\| + \sum |d_i| \, \|K_i \psi\| \tag{3.17}$$

$$\leq (1 - \epsilon_1 \textstyle\sum |d_i|)^{-1} \{\|(L + \sum d_i K_i)\psi\| + b_1 \sum |d_i| \, \|\psi\|\}. \tag{3.18}$$

Substitution of Eq. (3.18) into Eq. (3.16) yields

$$\|\sum c_i K_i \psi\| \leq \epsilon_2 \|(L + \textstyle\sum d_i K_i)\psi\| + b\|\psi\|, \tag{3.19}$$

where

$$\epsilon_2 \equiv (1 - \epsilon_1 \textstyle\sum |d_i|)^{-1} \epsilon_1 \sum |c_i| \tag{3.20}$$

and

$$b \equiv b_1 (1 - \epsilon_1 \textstyle\sum |d_i|)^{-1} \sum |c_i| . \tag{3.21}$$

The restriction on ϵ_1 given by Eq. (3.14) is such that $\epsilon_2 < \epsilon$. Since ϵ was arbitrary, the proof is complete.

Q.E.D.

Proof of Property (P4): It is well known[22] that potentials V_{ij} satisfying Eq. (3.2) have the property that $V_{ij} \ll H_0$, and that the domains of H_N, H_0, and all the cluster Hamiltonians H_A are the same. The potentials $\bar{V}_A \equiv H_N - H_A$ and $V_A \equiv H_A - H_0$ are then different linear sums of the operators V_{ij}. Property (P4) thus follows from Lemma 7 upon identification of the potentials V_{ij} with the operators K_i of the lemma, and of H_0 with L.

Q.E.D.

We now turn to verifying the abstract assumptions of Sec. 2 on the basis of Properties (P1)–(P6).

B. Verification of assumptions (A1)–(A6)

In the Appendix to Ref. 19 it was shown how most of the abstract assumptions in Sec. 2 follow from the properties of N-body systems in the last subsection. In that paper a space $\mathcal{H}'$ was defined by

$$\mathcal{H}' \equiv \bigoplus_{\alpha} \mathcal{H}_\alpha . \tag{3.22}$$

An operator H' was defined by

$$H'\Phi' \equiv \bigoplus_{\alpha} H_\alpha \varphi_\alpha \tag{3.23}$$

for all $\Phi' = \oplus_\alpha \varphi_\alpha \in \mathcal{H}'$ such that $\varphi_\alpha \in \mathcal{D}(H_\alpha)$. Finally, an operator J' (J in Ref. 19): $\mathcal{H}' \to \mathcal{H}_N$ was defined by

$$J'\Phi' \equiv \sum_{\alpha} \varphi_\alpha . \tag{3.24}$$

The space $\mathcal{H}'$ and the operators H' and J' play the role of $\mathcal{H}, H$, and J of Sec. 2.

Another formulation which emphasizes the cluster structure rather than the channels is also possible. Define the operator

$$P_A \equiv \sum_{\alpha}{}^{(A)} P_\alpha , \tag{3.25}$$

where $\sum^{(A)}$ denotes a strong topology sum over all channels with clustering A. Since the P_α in Eq. (3.25) are a countable sequence of projections of $\mathcal{H}_N$ onto orthogonal subspaces $\mathcal{H}_\alpha$ [cf. Property (P3)], it follows that P_A is an orthogonal projection. Define

$$\mathcal{H}_A \equiv P_A \mathcal{H}_N \tag{3.26}$$

and

$$\mathcal{H} \equiv \bigoplus_{A} \mathcal{H}_A . \tag{3.27}$$

The space $\mathcal{H}_A$ is easily seen to be given alternatively by

$$\mathcal{H}_A = J'^A \mathcal{H}', \tag{3.28}$$

where $J'^A : \mathcal{H}' \to \mathcal{H}_A$ is defined by

$$J'^A \Phi' \equiv \sum_{\alpha}{}^{(A)} \varphi_\alpha . \tag{3.29}$$

Further define the operator $H : \mathcal{D}(H) \subset \mathcal{H} \to \mathcal{H}$ by

$$H\Phi \equiv \bigoplus_{A} H_A \varphi_A , \tag{3.30}$$

for all $\Phi = \oplus_A \varphi_A \in \mathcal{H}$ such that $\varphi_A \in \mathcal{D}(H_A)$. Finally, define $J : \mathcal{H} \to \mathcal{H}_N$ by

$$J\Phi \equiv \sum_{A} \varphi_A . \tag{3.31}$$

The space $\mathcal{H}$ and the operators H and J are then candidates for the $\mathcal{H}, H$, and J of Sec. 2.

Indeed, examination of the proofs of Ref. 19 show that the abstract assumptions of Sec. 2A that were verified on the basis of $\mathcal{H}', H'$, and J' were simultaneouly verified for the $\mathcal{H}, H$, and J defined in Eqs. (3.25)–(3.31).

Not demonstrated in Ref. 19 were Assumption (A4) and the property in Assumption (A3) that JJ^* has a bounded inverse.

Consider first the operator JJ^*. Let m_N denote the number of possible cluster arrangements of the N particles. The number m_N is thus a constant for any given scattering problem. For each clustering A define the operator Q_A by

$$Q_A \equiv I_N - P_A , \tag{3.32}$$

where I_N is the identity on $\mathcal{H}_N$. Then the following theorem establishes the boundedness of $(JJ^*)^{-1}$.

Theorem 7: The operator JJ^* is invertible, and

$$(JJ^*)^{-1} = m_N^{-1} \sum_{n=0}^{\infty} \left(m_N^{-1} \sum_{A \neq 0} Q_A \right)^n \tag{3.33}$$

$$= m_N^{-1} \sum_{n=0}^{\infty} \left(I_N - m_N^{-1} \sum_{A} P_A \right)^n . \tag{3.34}$$

Specifically, $\|(JJ^*)^{-1}\| \leq 1$.

Proof: The adjoint operator $J^* : \mathcal{H}_N \to \mathcal{H}$ is given for $\psi \in \mathcal{H}_N$ by

$$J^*\psi = \bigoplus_{A} P_A \psi . \tag{3.35}$$

Therefore,

$$JJ^* = \sum_{A} P_A = I_N + \sum_{A \neq 0} P_A . \tag{3.36}$$

Substituting Eq. (3.32) and dividing by m_N gives

$$m_N^{-1} JJ^* = I_N - m_N^{-1} \sum_{A \neq 0} Q_A . \tag{3.37}$$

Now

$$\| m_N^{-1} \sum_{A \neq 0} Q_A \| \leq m_N^{-1} \sum_{A \neq 0} \| Q_A \| \leq m_N^{-1} (m_N - 1) < 1 . \tag{3.38}$$

The inverse of $m_N^{-1} JJ^*$ is thus the absolutely convergent Neumann series

$$m_N (JJ^*)^{-1} = \sum_{n=0}^{\infty} \left(m_N^{-1} \sum_{A \neq 0} Q_A \right)^n . \tag{3.39}$$

Equation (3.33) follows immediately from Eq. (3.39), and Eq. (3.34) follows upon the further substitution of Eq. (3.32) into Eq. (3.33). Finally, from Eq. (3.33) and inequality (3.38),

$$\| (JJ^*)^{-1} \| \leq m_N^{-1} \sum_{n=0}^{\infty} \| m_N^{-1} \sum_{A \neq 0} Q_A \|^n$$

$$\leq m_N^{-1} \sum_{n=0}^{\infty} [m_N^{-1} (m_N - 1)]^n = 1 . \tag{3.40}$$

Q.E.D.

We remark that there are two simple cases of $(JJ^*)^{-1}$: (1) In the case of simple scattering systems one has $m_N = 1$ and $JJ^* = I_N$. In this case $(JJ^*)^{-1} = I_N$ trivially. (2) In the case in which the scattering subspaces $\mathcal{H}_A$ are orthogonal for $A \neq 0$ one has

$$(JJ^*)^{-1} = I_N - \tfrac{1}{2} \sum_{A \neq 0} P_A . \tag{3.41}$$

We emphasize that Eq. (3.41) is not generally true but holds only if for all $A \neq 0$ and $B \neq 0$ the orthogonality relation $P_A P_B = \delta_{AB} P_B$ holds. In particular, if there is only one possible nonfree cluster (bound state) arrangement A of the scattering system, then $m_N = 2$, and $(JJ^*)^{-1} = I_N - \frac{1}{2} P_A$.

A similar result holds for $J'J'^*$.

Corollary 1: The operator $J'J'^*$ is invertible, and $\|(J'J'^*)^{-1}\| \leq 1$.

Proof: The operator $J'J'^*$ is given by

$$J'J'^* = \sum_\alpha P_\alpha . \tag{3.42}$$

The sum in Eq. (3.42) is commutatively convergent, and hence, the order of summation can be rearranged.[34] Thus,

$$J'J'^* = \sum_A \sum_\alpha {}^{(A)}P_\alpha = \sum_A P_A = JJ^* . \tag{3.43}$$

The result now follows from Theorem 7. Q.E.D.

It is also interesting from a computational point of view that the operator $(JJ^*)^{-1}$ can be approximated by finite sums of powers of the operators Q_A and hence also of P_A.

Corollary 2: If

$$(JJ^*)_k^{-1} \equiv m_N^{-1} \sum_{n=0}^{k} \left(m_N^{-1} \sum_{A \neq 0} Q_A \right)^n , \tag{3.44}$$

then

$$\|(JJ^*)^{-1} - (JJ^*)_k^{-1}\| \leq (1 - m_N^{-1})^{k+1} . \tag{3.45}$$

Proof: Since

$$(JJ^*)^{-1} - (JJ^*)_k^{-1} = m_N^{-1} \sum_{n=k+1}^{\infty} \left[m_N^{-1} \sum_{A \neq 0} Q_A \right]^n , \tag{3.46}$$

one has

$$\|(JJ^*)^{-1} - (JJ^*)_k^{-1}\| \leq m_N^{-1} \sum_{n=k+1}^{\infty} \|m_N^{-1} \sum_{A \neq 0} Q_A\|^n \tag{3.47}$$

$$\leq m_N^{-1} \sum_{n=k+1}^{\infty} (1 - m_N^{-1})^n = (1 - m_N^{-1})^{k+1} . \tag{3.48}$$

Q.E.D.

We now consider Assumption (A4). With J defined by Eq. (3.31) and V defined by Eq. (2.1), we see that

$$V\Phi = \sum_A (H_N - H_A)\varphi_A = \sum_A \bar{V}_A \varphi_A , \tag{3.49}$$

for $\Phi \in \mathcal{D}(H)$. It now follows from Property (P4) that for any $\epsilon > 0$, there exists $b(\epsilon) > 0$ such that

$$\|V\Phi\| \leq \epsilon \sum_A \|H_A \varphi_A\| + b \sum_A \|\varphi_A\| . \tag{3.50}$$

Since $\|\Phi\|^2 = \sum_A \|\varphi_A\|^2$, and the number m_N of clustering A is finite, the Cauchy–Schwarz inequality gives

$$\|V\Phi\| \leq \epsilon m_N^{1/2} \|H\Phi\| + b m_N^{1/2} \|\Phi\| . \tag{3.51}$$

It follows that inequality (2.3) is satisfied, and $V \ll H$.

If V^* is defined by Eq. (2.2) and $\psi \in \mathcal{D}(H_N)$, then

$$V^*\psi = \bigoplus_A P_A \bar{V}_A \psi . \tag{3.52}$$

Hence,

$$\|V^*\psi\| = \left\{ \sum_A \|P_A \bar{V}_A \psi\|^2 \right\}^{1/2} \leq \sum_A \|P_A \bar{V}_A \psi\| \leq \sum_A \|\bar{V}_A \psi\| . \tag{3.53}$$

Since $H_A = H_N - \bar{V}_A$, the triangle inequality and inequality (3.9) give

$$\|H_A \psi\| \leq (1 - \epsilon)^{-1} (\|H_N \psi\| + b\|\psi\|) . \tag{3.54}$$

Substitution of inequality (3.54) into inequality (3.9) shows that $\bar{V}_A \ll H_N$. It then follows from inequality (3.53) that inequality (2.4) is satisfied, and $V^* \ll H_N$.

The same proofs apply to prove Assumption (A4) for $V' \equiv H_N J' - J'H'$. The proof for the cluster formalism carries over virtually intact because of the relation between H and H' given by Eqs. (3.27) and (3.28), and because $H_A P_\alpha = H_\alpha P_\alpha$ if channel α has clustering A.

Other possible spaces H could be defined by letting H_α correspond to the space of functions $f(\mathbf{Y}, t)$ in Eq. (3.6),[35] or to functions which are (anti)symmetrized versions of functions in H_α.[36]

C. Matrix elements

Definition 2: (a) If an operator M maps $H_N \to H$ or $H \to H_N$, its restrictions which map $H_N \to H_A$ or $H_A \to H_N$ are called the *cluster components* of M. If an operator M' maps $H_N \to H'$ or $H' \to H_N$, its restrictions which map $H_N \to H_\alpha$ or $H_\alpha \to H_N$ are called the *channel components* of M'.

(b) If an operator M (or M') maps $H \to H$ ($H' \to H'$), its restrictions which map $H_A \to H_B$ ($H_\alpha \to H_\beta$) are called the *cluster matrix elements* (*channel matrix elements*) of $M(M')$.

We note that the words "matrix elements" are used here only because of their historical origin,[20] and that there are no matrices appearing anywhere in this paper.

Since P_A commutes with H_A, the cluster components of $V = H_N J - JH$ and $V^* = J^* H_N - HJ^*$ are given by

$$(V)_A = \bar{V}_A P_A \tag{3.55}$$

and

$$(V^*)_A = P_A \bar{V}_A . \tag{3.56}$$

The cluster matrix elements $T_{BA}(z)$ of the transition operator $T(z)$ defined in Eq. (2.13) are

$$T_{BA}(z) = P_B [R_B^{-1}(z) R_N(z) R_A^{-1}(z) - \delta_{BA} R_A^{-1}(z)] P_A , \tag{3.57}$$

where $R_A(z) \equiv (z - H_A)^{-1}$, $R_N(z) \equiv (z - H_N)^{-1}$, and δ_{BA} is the Kronecker delta. The cluster matrix elements $T_{BA}^{(\pm)}(z)$ of $T^{(\pm)}(z)$ defined by Eqs. (2.14) and (2.15) can be similarly obtained.

Let $T'(z) : \mathcal{D}(H') \to H'$ be defined by

$$T'(z) \equiv (z - H')J'^* R_N(z) J'(z - H') - (z - H') . \tag{3.58}$$

If α has clustering A and β has clustering B, then $\delta_{BA} P_\beta P_\alpha = \delta_{\beta\alpha} P_\alpha$ [cf. Eq. (3.8)], and $H_A P_\alpha = H_\alpha P_\alpha$. Hence, the channel matrix elements $T'_{\beta\alpha}(z)$ of $T'(z)$ are

$$T'_{\beta\alpha}(z) = P_\beta T_{BA}(z) P_\alpha . \tag{3.59}$$

If $T'^{(\pm)}(z)$ are defined by adding primes in Eqs. (2.14) and (2.15), then the channel matrix elements $T'^{(\pm)}_{\beta\alpha}(z)$ are analogously related to $T^{(\pm)}_{BA}(z)$. Consequently, it suffices to consider only the cluster transition operators T and $T^{(\pm)}$. The channel matrix elements of the

corresponding channel transition operators T' and $T'^{(\pm)}$ can then be easily obtained by adding projections onto the channel subspaces $\mathcal{H}_\alpha$ of $\mathcal{H}_A$ as in Eq. (3.59).

Define the transition operators $U_{BA}(z)$ and $U_{BA}^{(\pm)}(z)$ mapping $\mathcal{D}(H_N) \to \mathcal{H}_N$ by

$$U_{BA}(z) \equiv R_B^{-1}(z) R_N(z) R_A^{-1}(z) - \delta_{BA} R_A^{-1}(z), \tag{3.60}$$

$$U_{BA}^{(+)}(z) \equiv \bar{V}_B + \bar{V}_B R_N(z) \bar{V}_A, \tag{3.61}$$

$$U_{BA}^{(-)}(z) \equiv \bar{V}_A + \bar{V}_B R_N(z) \bar{V}_A, \tag{3.62}$$

respectively, for all clusterings A and B.

The operators U_{BA} are the symmetrical Alt–Grassberger–Sandhas transition operators,[6,7] and the operators $U_{BA}^{(\pm)}$ are the nonsymmetrical transition operators used by Lovelace *et al.*[8–10] Their relationship to the T and $T^{(\pm)}$ operators is shown by the following theorem.

Theorem 8: The cluster matrix elements of $T(z)$ and $T^{(\pm)}(z)$ are

$$T_{BA}(z) = P_B U_{BA}(z) P_A \tag{3.63}$$

and

$$T_{BA}^{(\pm)}(z) = P_B U_{BA}^{(\pm)}(z) P_A, \tag{3.64}$$

respectively.

Proof: Equation (3.63) is immediate from Eqs. (3.57) and (3.60). Equations (3.64) are also immediate from Lemma 2 and Eqs. (3.55) and (3.56). Q.E.D.

D. Systems of N-body equations

Theorem 8 shows that the cluster matrix elements of $T(z)$ and $T^{(\pm)}(z)$ are related to previously used transition operators by the addition of the projections P_A and P_B. By taking cluster matrix elements of the Lippmann–Schwinger type Eqs. (2.42)–(2.45), we obtain systems of equations which leave in the P_A and P_B operators.

Let $\bar{\delta}_{BA} \equiv 1 - \delta_{BA}$. Then the cluster transition operators $T_{BA} = T_{BA}(z)$ are, for $z \in \rho(H_N)$, the unique solution of the systems of equations

$$T_{BA} = \bar{\delta}_{BA} P_B R_A^{-1} P_A + P_B \bar{V}_B (JJ^*)^{-1} P_A + P_B \bar{V}_B (JJ^*)^{-1} \sum_C R_C P_C T_{CA} \tag{3.65}$$

and

$$T_{BA} = \bar{\delta}_{BA} P_B R_B^{-1} P_A + P_B (JJ^*)^{-1} \bar{V}_A P_A + \sum_C T_{BC} R_C P_C (JJ^*)^{-1} \bar{V}_A P_A. \tag{3.66}$$

The operators $T_{BA}^{(\pm)} = T_{BA}^{(\pm)}(z)$ are, for $z \in \rho(H_N)$, the unique solutions of the systems of equations

$$T_{BA}^{(+)} = P_B \bar{V}_B P_A + \sum_C T_{BC}^{(+)} R_C P_C (JJ^*)^{-1} \bar{V}_A P_A \tag{3.67}$$

and

$$T_{BA}^{(-)} = P_B \bar{V}_A P_A + P_B \bar{V}_B (JJ^*)^{-1} \sum_C R_C P_C T_{CA}^{(-)} \tag{3.68}$$

respectively.

Remark: The operators P_C are written in Eqs. (3.65)–(3.68) for emphasis only. Since $T_{BA} = P_B T_{BA} P_A$ and $T_{BA}^{(\pm)} = P_B T_{BA}^{(\pm)} P_A$, these P_C operators are already present without writing them explicitly.

Let $G_B(z) = P_B R_N(z)$ denote the cluster components of the operator $G(z)$ defined in Eq. (2.49). Then the operators $G_B = G_B(z)$ are, for $z \in \rho(H_N)$, the unique solution of the system of equations

$$G_B = P_B R_B + P_B R_B \bar{V}_B (JJ^*)^{-1} \sum_C G_C. \tag{3.69}$$

The uniqueness of the solution of Eqs. (3.65)–(3.69) follows from Theorem 6, because one of these equations will have a unique solution if and only if the corresponding Eq. (2.42)–(2.45) or (2.50) has a unique solution.

4. DISCUSSION

In the preceding sections we have derived some abstract equations of the Lippmann–Schwinger type, and then applied them to N-body quantum scattering theory. The starting point for this derivation was the two-Hilbert-space formulation given in Ref. 19. In this formulation an auxiliary direct sum Hilbert space $\mathcal{H}$ is used as a "bookkeeping" tool to simultaneously keep track of all the channels. Mappings between $\mathcal{H}$ and the N-body Hilbert space $\mathcal{H}_N$ are provided by the injection operator J and its adjoint J^*.

In order to derive equations within the two-Hilbert-space framework, it was necessary to invert the operator JJ^*. The presence of the J operators, however, considerably complicated the derivation. It was necessary to prove certain properties about how these operators interact with the domains and spectra of the Hamiltonians H_N and H (cf. Theorem 3 and Lemmas 5 and 6). For this we needed to make some relative boundedness assumptions on the potentials [cf. Assumption (A4) and Property (P4)]. Assumptions of this type have been used previously[22,29,32,35] for other purposes in scattering theory, and they do not severely restrict the generality of the theory. They have been shown to be satisfied if, for example, the pair potentials satisfy Assumption B.

In regard to Lemma 5, we remark that we have only shown that the spectrum of QHQ is contained in the spectrum of H. It is an open question as to the nature of this spectrum. We do not know if the spectrum of QHQ is absolutely continuous, or even if it consists of a half-line.

The unknown operators in the systems of N-body equations given in Sec. 3D depend only on the various possible clusterings of the N particles. By replacing $\mathcal{H}$ by the Hilbert space $\mathcal{H}'$ defined in Eq. (3.22) and taking channel matrix elements of Eqs. (2.42)–(2.45) and channel components of Eq. (2.50), systems of equations of the same type can be obtained for the corresponding channel operators $T'_{\beta\alpha}(z)$, $T'^{(\pm)}_{\beta\alpha}(z)$, and $G'_\beta(z) = P_\beta R_N(z)$. However, if some clusterings have more than one possible bound state, the number of unknowns in

these systems would be larger, and possibly even infinite. For this reason, we have chosen to use the Hilbert space H and to distinguish only the different clusterings in the equations. Once the cluster operators have been obtained, the corresponding channel operators can be obtained by adding the projection operators P_α as in Eq. (3.59). [It should be noted that $\alpha, \beta, \gamma, \cdots$ are sometimes used as subscripts by other authors to distinguish between different clusterings in systems of N-body equations (e.g., Refs. 2, 6–8). The reason that we have used $\alpha, \beta, \gamma, \cdots$ to denote different channels, instead of different clusterings, is because our derivation began with the time-dependent theory of scattering, and we have adopted the notation commonly used there (e.g., Refs. 21–24, 35).]

The systems of equations given in Sec. 3D are our basic transition and resolvent operator equations. For $N=2$ they reduce to the usual Lippmann–Schwinger and resolvent equations. For $N>2$ the complexity of the equations increases as the number of possible clusterings increases. We have shown that there exists a unique solution of these equations.

The distinctive feature of Eqs. (3.65)–(3.69) is that the unknown operators act only on the cluster subspaces H_A of H_N. An advantage arises because the cluster Hamiltonians H_A and their resolvents $R_A(z)$ are relatively simple operators on H_A. In particular, suppose that H_A is given by Eqs. (3.3)–(3.5),

$$H_A^0 \equiv \sum_{l=1}^{k} H_A^0(a_l), \quad \hat{H}_A \equiv \sum_{l=1}^{k} \hat{H}_A(a_l), \tag{4.1}$$

and α is one of the channels with clustering A. Then

$$H_A P_\alpha = H_A^0 P_\alpha + \hat{H}_A P_\alpha = H_A^0 P_\alpha + \lambda_\alpha P_\alpha, \tag{4.2}$$

where

$$\lambda_\alpha \equiv \sum_{l=1}^{k} \lambda_\alpha(a_l) \tag{4.3}$$

and $\lambda_\alpha(a_l)$ are the eigenvalues corresponding to the bound states of the fragments a_l in channel α. The operators $H_A P_A$ and $R_A(z)P_A$ are then of the form

$$H_A P_A = \sum_{\alpha}{}^{(A)} (H_A^0 + \lambda_\alpha) P_\alpha \tag{4.4}$$

and

$$R_A(z) P_A = \sum_{\alpha}{}^{(A)} (z - \lambda_\alpha - H_A^0)^{-1} P_\alpha . \tag{4.5}$$

The operator H_A^0 is the Laplacian operator governing the free motion of the centers of mass of the fragments in clustering A. If $H_N = L^2(\mathbf{R}^{3N})$, and A is a k-fragment clustering, then $R_A(z)P_A$ is effectively a free resolvent in a space of only $3k$ dimensions ($3k$-3 after the center of mass is removed), $k=2, 3, \ldots, N$. In momentum space, the H_A^0 in Equations (4.4) and (4.5) can be replaced by a certain quadratic polynomial. In contrast, the operators $R_A(z)Q_A$ with $Q_A = I_N - P_A$, which never arise in our equations, are considerably more complicated.

One of the nice features of the two-Hilbert-space methods used in Sec. 2 is that the abstract equations take on the same form for all N. A correct system of N-body equations can be obtained at any stage of the derivation by taking cluster components or cluster matrix elements.

On the negative side, Eqs. (3.65)–(3.69) have two drawbacks. First, the presence of the operator $(JJ^*)^{-1}$ is an added complication. In practical computations this operator will no doubt have to be approximated (cf. Corollary 2 of Theorem 7); however, we believe this is quite feasible.

Second, the kernel of Eqs. (3.65)–(3.69) is connected, and compact, only for $N=2$. The connectedness structure is, however, improved by the presence of the P_A operators. We note, for example, that the operator $P_B \bar{V}_B (JJ^*)^{-1} R_C$ in the kernel of Eq. (3.65) is connected for all N-body clusterings B which have only two fragments. Hence, for the three-body problem the kernel operator is connected for all clusterings except the free particle clustering $B=0$. The equations do not have to be iterated to obtain this connectedness for the two fragment clusterings.

In order to derive N-body equations with a connected kernel, it is necessary to first solve all subsystem problems for the clusterings A in terms of the clusterings which are refinements of A, and then to incorporate these solutions into the equations. This will be done in a subsequent paper.

ACKNOWLEDGMENTS

This work was partially completed while one of us (AGG) was on a sabbatical leave of absence at the Centre National de la Recherche Scientifique, Marseille, France, and receiving financial support from the D.G.R.S.T., France. We would like to thank the Centre de Physique Théorique, CNRS, as well as the Centre Universitaire de Toulon for their kind hospitality. Special thanks go to Professor J.M. Combes and Professor R. Stora for making this visit possible. We have benefited greatly from the opportunity to discuss this work with others, including the participants of a seminar at CNRS. We are especially indebted to Professor P. Deift, who gave us numerous suggestions and showed us how to prove Theorem 3, and to Professor H. Ekstein, Professor A. Grossman, and Professor T. Trucano for valuable discussions.

[1] L. D. Faddeev, Zh. Eksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys. JETP **12**, 1014 (1961)].
[2] L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering Theory* (Israel Program for Scientific Translations, Jerusalem, 1965).
[3] S. Weinberg, Phys. Rev. **133**, B232 (1964).
[4] C. van Winter, Mat. Fys. Skr. Dan. Vid. Selsk. **2**, No. 8, 1 (1964).
[5] O. A. Yakubovskiĭ, Sov. J. Nucl. Phys. **5**, 937 (1967).
[6] E. O. Alt, P. Grassberger, and W. Sandhas, Nucl. Phys. B **2**, 167 (1967).
[7] P. Grassberger and W. Sandhas, Nucl. Phys. B **2**, 181 (1967).
[8] C. Lovelace, Phys. Rev. **135**, B1225 (1964).
[9] D. J. Kouri and F. S. Levin, Nucl. Phys. A **250**, 127 (1975); **253**, 395 (1975).
[10] W. Tobocman, Phys. Rev. C **9**, 2466 (1974); **11**, 43 (1975).

[11]I. M. Narodetskiĭ and O. A. Yakubovskiĭ, Sov. J. Nucl. Phys. **14**, 178 (1972).
[12]Gy. Bencze, Nucl. Phys. A **210**, 568 (1973).
[13]E. F. Redish, Nucl. Phys. A **225**, 16 (1974).
[14]T. A. Osborn and K. L. Kowalski, Ann. Phys. (N. Y.) **68**, 36 (1971).
[15]B. R. Karlsson and E. M. Zeiger, Phys. Rev. D **11**, 939 (1975).
[16]B. R. Karlsson and E. M. Zeiger, in *Few Body Dynamics*, edited by A. N. Mitra *et al.* (North-Holland, Amsterdam, 1976), pp. 91—3.
[17]W. Sandhas, Acta Phys. Austr. Suppl. **9**, 57 (1972); **13**, 679 (1974).
[18]V. Vanzani, "The Nuclear Three-Body Problem," Preprint, Miramare-Trieste, 1974.
[19]C. Chandler and A. G. Gibson, J. Math. Phys. **14**, 1328 (1973).
[20]H. Ekstein, Phys. Rev. **101**, 880 (1956).
[21]J. M. Jauch, Helv. Phys. Acta **31**, 661 (1958).
[22]W. Hunziker, in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Britten (Gordon and Breach, New York, 1968), Vol. X-A.
[23]J. M. Combes, Nuovo Cimento A **64**, 111 (1969).
[24]J. D. Dollard, Rocky Mt. J. Math. **1**, 5 (1971).
[25]T. Kato, J. Func. Anal. **1**, 342 (1967).
[26]A. L. Belopol'skii and M. S. Birman, Math. USSR Izv. **2**, 1117 (1968).
[27]W. O. Amrein, V. Georgescu, and J. M. Jauch, Helv. Phys. Acta **44**, 407 (1971).
[28]D. B. Pearson, Nuovo Cimento A **2**, 853 (1971).
[29]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[30]C. Chandler and A. G. Gibson, in *Few Body Dynamics*, edited by A. N. Mitra *et al.* (North-Holland, Amsterdam, 1976), pp. 123—25.
[31]M. Schechter, *Principles of Functional Analysis* (Academic, New York, 1971).
[32]J. M. Combes, Commun. Math. Phys. **12**, 283 (1969).
[33]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[34]G. Bachman and L. Narci, *Functional Analysis* (Academic, New York, 1966).
[35]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, in preparation), Vol. III.
[36]F. Coester, Helv. Phys. Acta **38**, 7 (1965).

Analytic evaluation of an important integral in collision theory

John Detrich and Robert W. Conn

Nuclear Engineering Department, University of Wisconsin, Madison, Wisconsin 53706
(Received 18 April 1977)

A commonly occurring integral in collision theory when the radial part of the potential has the form $r^n e^{-\alpha r}$ is

$$I_{n;ll'}(\alpha;k,k') = \int_0^\infty dr\, r^{n+1}\, e^{-\alpha r} J_{l+1/2}(kr)\, J_{l'+1/2}(k'r).$$

Analytic results for this integral have been found for the most important cases, including those with negative values of n. This permits efficient evaluation of Born matrix elements required in many scattering theory applications based on perturbation methods.

I. INTRODUCTION

We report here analytic results for integrals of the form

$$I_{n;ll'}(\alpha;k,k') = \int_0^\infty dr\, r^{n+1} \exp(-\alpha r) J_{l+1/2}(kr) J_{l'+1/2}(k'r), \tag{1}$$

where $J_{l+1/2}(kr)$ denotes a Bessel function and l and l' are restricted to nonnegative integer values. Such integrals occur in the evaluation of first Born scattering matrix elements when the potential has a radial part proportional to $r^n \exp(-\alpha r)$. This is a very frequently encountered situation since most typical potentials can be represented in terms of a linear combination of such functions.

Results for integrals of the type given in Eq. (1) have been given previously,[1] but they are both unnecessarily general and unnecessarily complicated. Much simpler results may be obtained by specializing to the case where l and l' are nonnegative integers. This is important, since the efficient evaluation of first Born matrix elements is a critical consideration for the successful application of many collision theories including, for example, the operator decomposition approach to inelastic molecular scattering proposed by Conn and Rabitz.[2]

II. MATHEMATICAL DERIVATION

We assume here that the parameters α, k, and k' in Eq. (1) are each real and positive. We also take l and l' to be nonnegative integers, while n is real, but not restricted to integer values. Then the integral in Eq. (1) converges whenever we have

$$n + l + l' + 3 > 0. \tag{2}$$

It is convenient to express the Bessel functions in terms of the integral representation[3]

$$J_{l+1/2}(kr) = \frac{(kr/2)^{l+1/2}}{l!\sqrt{\pi}} \int_{-1}^{1} dt\,(1-t^2)^l \exp(-ikrt). \tag{3}$$

Then, after reversing the order of integration and integrating over r, we find

$$I_{n;ll'}(\alpha,k,k') = (k/2)^{l+1/2}(k'/2)^{l'+1/2}\Gamma(n+l+l'+3)[\pi l!\, l'!\,]^{-1} \times \int_{-1}^{1} ds(1-s^2)^l \int_{-1}^{1} dt(1-t^2)^{l'}(\alpha + iks + ik't)^{-n-l-l'-3}, \tag{4}$$

where $\Gamma(n+l+l'+3)$ indicates the usual gamma function.

The remaining integrations can be accomplished by a repeated integration by parts, a procedure that terminates after a finite number of steps when l and l' are nonnegative integers. With the aid of the result

$$[d^p(1-s^2)^l/ds^p]_{s=1} = (-1)^p [d^p(1-s^2)^l/ds^p]_{s=-1} = \begin{cases} 0, & l > p \text{ or } p > 2l, \\ (-1)^l \dfrac{2^{2l-p} l!\, p!}{(2l-p)!\,(p-l)!}, & 2l \geq p \geq l, \end{cases}$$

we obtain

$$I_{n;ll'}(\alpha;k,k') = \pi^{-1}(kk')^{-1/2} \sum_{p=0}^{l} \sum_{q=0}^{l'} (l+p)!\,(l'+q)! \times [p!\,q!\,(l-p)!\,(l'-q)!\,]^{-1}(2k)^{-p}(2k')^{-q} \times \mathrm{Re}[i^{l+l'-p-q+2} X_{n+1-p-q}(\alpha + ik + ik') + i^{l-p-l'+q} X_{n+1-p-q}(\alpha + ik - ik')], \tag{5, 6}$$

where we use

$$X_\nu(z) = \Gamma(\nu) z^{-\nu}. \tag{7}$$

This expression for $X_\nu(z)$ loses its meaning when ν is an integer less than or equal to zero, in which case Eq. (7) should be replaced by

$$X_{-p}(z) = (-1)^{p+1} z^p (\ln z - \sigma_p)/p!, \quad p \geq 0, \tag{8}$$

where

$$\sigma_p = \begin{cases} 0, & p = 0, \\ \sum_{q=1}^{p} q^{-1}, & p \geq 1. \end{cases} \tag{9}$$

When n is an integer less than $l + l'$, Eqs. (6) and (8) indicate that two arctangents must be computed in order to evaluate $I_{n;ll'}(\alpha;k,k')$ since z is in general complex. On the other hand, only the real part of *one* logarithm is required, since only the combination

$$\ln|\alpha + ik + ik'| - \ln|\alpha + ik - ik'| = \ln\left(\frac{|\alpha + ik + ik'|}{|\alpha + ik - ik'|}\right) \tag{10}$$

actually occurs in Eq. (6). One can see that this situation should hold by considering how the integral will be affected if each of the parameters α, k, k' is multiplied by a common factor. We give a more straightforward demonstration in Appendix A.

Additional rearrangements of Eq. (6) do not appear to be useful from the standpoint of actual computation, except for some special cases.[4] When l and l' are small, Eq. (6) clearly yields a compact evaluation of $I_{n;ll'}(\alpha;k,k')$. Even for large l and/or large l', the evaluation of the integral is an efficient process, since sucessive terms $X_{n+1-p-q}(\alpha+ik\pm ik')$ can be computed from the preceding one by only a few arithmetic operations, and this also applies to the terms $(l+p)!\,[p!\,(l-p)!\,]^{-1}(2k)^{-p}$.

Recurrence relations for the integrals $I_{n;ll'}(\alpha;k,k')$ are readily obtained with the help of the recurrence relations satisfied by the Bessel functions, and these are discussed by Eason, Noble, and Sneddon.[5] We shall not present a similar discussion here, since these recurrence relations do not appear to be useful from the standpoint of practical computation.

III. THE CASE $\alpha = 0$

The case $\alpha = 0$ corresponds to potentials with a radial part proportional to r^n. In addition to Eq. (2), we must impose the condition

$$n < -1, \tag{11}$$

so we are dealing with inverse-power potentials, a case of considerable practical interest. We give the definition

$$I_{n;ll'}(0;k,k') = \int_0^\infty dr\, r^{n+1} J_{l+1/2}(kr) J_{l'+1/2}(k'r). \tag{12}$$

As our notation suggests, we may give

$$I_{n;ll'}(0;k,k') = \lim_{\alpha\to 0+} I_{n;ll'}(\alpha;k,k'), \tag{13}$$

provided that Eq. (11) holds. This permits us to apply Eq. (6) to the evaluation of the integral in Eq. (12).

The integral $I_{n;ll'}(0;k,k')$ is discussed in detail by Watson,[6] with the result

$$\begin{aligned} &I_{n;ll'}(0;k,k') \\ &\quad = 2^{n+1}k^{-n-2}x^{l'+1/2}\Gamma(m)[\Gamma(l+\tfrac{3}{2}-m)\Gamma(l'+\tfrac{3}{2})]^{-1} \\ &\qquad \times F(m, m-l-\tfrac{1}{2}; l'+\tfrac{3}{2}; x^2), \end{aligned} \tag{14}$$

where we use the abbreviations

$$m = (l+l'+n+3)/2, \quad x = k'/k, \tag{15}$$

and $F(a,b;c;z)$ denotes the Gauss hypergeometric function. This result applies when $k' < k$; to handle the case $k' > k$, one interchanges the roles of the pairs k', l' and k, l. In the general case, the hypergeometric function in Eq. (14) is defined in terms of an infinite series, while Eq. (6) always contains only a finite number of terms. Thus it may frequently be advantageous to evaluate $I_{n;ll'}(0;k,k')$ in terms of Eqs. (6) and (13) instead of Eq. (14).

Of particular interest is the case where n is integer. If $n+l+l'$ is even, $m-l-\frac{1}{2}$ is integer, and the hypergeometric series in Eq. (14) truncates for $m \leq l+\frac{1}{2}$, while in case $m > l+\frac{1}{2}$, Eq. (14) gives zero for the integral. Thus Eq. (14) always gives a simple result when $n+l+l'$ is even. Of course, Eq. (6) can be rearranged to give the same result, but improvement can hardly be expected.

Our results are more helpful in the case $n+l+l'$ odd. It is not obvious from Eq. (14) that $I_{n;ll'}(0;k,k')$ can now be evaluated using only a finite number of terms, but this becomes clear by examining Eqs. (6) and (13). We can use Eq. (6) as it stands, but a simpler expression can be achieved when $\alpha = 0$. The necessary reduction is carried out in Appendix B, and this leads to the result

$$\begin{aligned} &I_{n;ll'}(0;k,k') \\ &\quad = 2^{n+1}k^{-n-2}x^{-l'-1/2}(-1)^{m-1}\left[\frac{F(l+\frac{3}{2}-m, 1-m; \frac{1}{2}-l'; x^2)}{(m-n-2)!\,\Gamma(l+\frac{3}{2}-m)\Gamma(\frac{1}{2}-l')}\right. \\ &\qquad \times (1-x^2)^{-n-1}\ln\left(\frac{k+k'}{|k-k'|}\right) - \left.\sum_{p=0}^{m-n-3} a_{n;ll';p}x^{2p+1}\right], \end{aligned} \tag{16}$$

where the coefficients $a_{n;ll';p}$ are given by

$$\begin{aligned} a_{n;ll';p} = \sum_{q=0}^{p} &[q!\,(m-n-2-q)!\,(p-q+\tfrac{1}{2}) \\ &\times \Gamma(q+\tfrac{1}{2}-l')\Gamma(l'-\tfrac{3}{2}-m-q)]^{-1}. \end{aligned} \tag{17}$$

This expression applies only in case

$$l'-1 \geq p, \quad (l'-n-2)/2 \geq p; \tag{18}$$

the remaining values of p are handled by means of the relation

$$a_{n;ll';p} = a_{n;l'l;m-n-3-p}. \tag{19}$$

Comparison of Eqs. (14) and (16) shows that the first l' terms in an expansion of Eq. (16) in a power series in x^2 will vanish. This indicates that inaccuracy can arise when using Eq. (16), due to roundoff error in the subtraction of nearly equal terms, especially when x is small. Computations should be arranged so as to avoid this difficulty. Similar remarks apply to the more general case handled in Eq. (6).

ACKNOWLEDGMENT

This research was partially supported by the U.S. Energy research and Development Administration.

APPENDIX A

To establish that the term given in Eq. (10) is the only logarithmic term which occurs in Eq. (6) in case n is an integer less than $l+l'$, it is sufficient to establish the relation

$$A_{n;ll'}(\alpha;k,k') = -A_{n;ll'}(\alpha;k,-k'), \tag{A1}$$

where

$$\begin{aligned} &A_{n;ll'}(\alpha;k,k') \\ &\quad = (-1)^n \sum_p \sum_q C_{l,p}C_{l',q}k^{-l-q-1}\,(k')^{-l'-q-1} \\ &\qquad \times \mathrm{Re}[i^{l+l'-p-q+2}(\alpha+ik+ik')^{p+q-n-1}/(p+q-n-1)!\,]. \end{aligned} \tag{A2}$$

Here the summations over p and q range over all values for which p, q, $(l-p)$, $(l'-q)$, and $(p+q-n-1)$ are all positive or zero, and we use the abbreviation

$$C_{l,p}=(-2)^{-p}(l+p)!/[p!(l-p)!]. \tag{A3}$$

In view of Eq. (8), we see that $A_{n;ll'}(\alpha;k,k')$ and $A_{n;ll'}(\alpha;k,-k')$ are, respectively, the coefficients of $\ln|\alpha+ik+ik'|$ and $\ln|\alpha+ik-ik'|$ in Eq. (6), after multiplication by $\pi k^{-l-1/2}(k')^{-l'-1/2}$.

We establish Eq. (A1) with the aid of the differential recurrence relations

$$\frac{1}{k}\left(\frac{\partial}{\partial k}\right)A_{n;ll'}(\alpha;k,k')=-A_{n+1;l+1,l'}(\alpha;k,k'), \tag{A4a}$$

$$\frac{1}{k'}\left(\frac{\partial}{\partial k'}\right)A_{n;ll'}(\alpha;k,k')=-A_{n+1;l,l'+1}(\alpha;k,k'), \tag{A4b}$$

which can be verified in straightforward fashion using Eqs. (A2) and (A3). In view of these relations, we may give

$$\begin{aligned}A_{n;ll'}&(\alpha;k,k')\\&=\left(-\frac{1}{k}\frac{\partial}{\partial k}\right)^{l}\left(\frac{1}{k'}\frac{\partial}{\partial k'}\right)^{l'}A_{n-l-l';00}(\alpha;k,k')\\&=\left(\frac{1}{k}\frac{\partial}{\partial k}\right)^{l}\left(-\frac{1}{k'}\frac{\partial}{\partial k'}\right)^{l'}(-1)^{n+l-l-l'}(kk')^{-1}\\&\quad\times\mathrm{Re}(\alpha+ik+ik')^{-n-1+l+l'}/(l+l'-n-1)!.\end{aligned} \tag{A5}$$

We expand to obtain

$$A_{n;ll'}(\alpha;k,k')=\sum_{p=0}^{l+l'-n-1}(-\alpha)^{p}A_{n+p;ll'}(0;k,k')/p!. \tag{A6}$$

From Eq. (A5) we see

$$A_{n;ll'}(0;k,k')=0,\quad n+l+l'\text{ even}, \tag{A7}$$

and, after expanding

$$\begin{aligned}A_{n;ll'}&(0;k,k')\\&=(-1)^{m-1}\sum_{p=0}^{l+l'-n-1}[p!(l+l'-n-p-1)!]^{-1}\left(\frac{1}{k}\frac{\partial}{\partial k}\right)^{l}\\&\quad\times\left(\frac{1}{k'}\frac{\partial}{\partial k'}\right)^{l'}k^{p-1}(k')^{l+l'-n-2-p},\quad n+l+l'\text{ odd},\end{aligned} \tag{A8}$$

where m is the abbreviation introduced in Eq. (15). In Eq. (A8), even powers of k occur only in terms containing even powers of k', and such terms vanish unless both of the conditions

$$2l<p-1,\quad 2l'<l+l'-n-2-p$$

can be satisfied. But this possibility is ruled out by Eq. (2), so $A_{n;ll'}(0;k,k')$ contains only odd powers of k and k'. Therefore, we have

$$A_{n;ll'}(0;k,k')=-A_{n;ll'}(0;k,-k'), \tag{A9}$$

which we can combine with Eq. (A6) to establish Eq. (A1).

APPENDIX B

It is convenient to use the definitions for $A_{n;ll'}(0;k,k')$ and $C_{l,p}$ given, respectively, by Eqs. (A2) and (A3) in Appendix A. Then, in view of Eq. (A1) in Appendix A, Eq. (6) yields, in case $n+l+l'$ is odd,

$$\begin{aligned}I_{n;ll'}&(0;k,k')\\&=\pi^{-1}(kk')^{-1/2}\{k^{l+1}(k')^{l'+1}A_{n;ll'}(0;k,k')\\&\quad\times\ln[(k+k')/|k-k'|)-B_{n;ll'}(k,k')\},\end{aligned} \tag{B1}$$

where

$$\begin{aligned}B_{n;ll'}&(k,k')\\&=(-1)^{m-1}\sum_{p=0}^{l}\sum_{q=0}^{l'}C_{l,p}C_{l',q}k^{-p}(k')^{-q}\sigma_{p+q-n-1}\\&\quad\times[(k+k')^{p+q-n-1}+(-1)^{l'+q+1}(k-k')^{p+q-n-1}]/\\&\quad(p+q-n-1)!.\end{aligned} \tag{B2}$$

Most of the reduction of $A_{n;ll'}(0;k,k')$ is given in Appendix A, so we need merely carry out the differentiations indicated in Eq. (A8). With the help of the doubling formula

$$\Gamma(p+1)=2^{p}\Gamma(p/2+1/2)\,\Gamma(p/2+1)/\sqrt{\pi}\,, \tag{B3}$$

we may cast the result into the form

$$\begin{aligned}A_{n;ll'}&(0;k,k')=\frac{2^{n+1}(-1)^{m-1}k^{-2m}x^{-2l'-1}\pi}{(m-n-2)!\,\Gamma(l'+\frac{3}{2}-m)\Gamma(\frac{1}{2}-l')}\\&\quad\times F(n+2,m,\ m-l'-\tfrac{1}{2};\tfrac{1}{2}-l';x^{2}).\end{aligned} \tag{B4}$$

The Euler relation

$$F(a,b;c;z)=(1-z)^{c-a-b}F(c-a,c-b;c;z) \tag{B5}$$

can be used to rewrite Eq. (B4) as

$$\begin{aligned}A_{n;ll'}&(0;k,k')=\frac{2^{n+1}(-1)^{m-1}k^{-2m}x^{-2l'-1}(1-x^{2})^{-n-1}}{(m-n-2)!\,\Gamma(l'+\frac{3}{2}-m)\Gamma(\frac{1}{2}-l')}\\&\quad\times F(l+\tfrac{3}{2}-m,\ 1-m;\tfrac{1}{2}-l';x^{2}).\end{aligned} \tag{B6}$$

According to Eq. (2), m is a positive integer, since $l+l'+n$ is odd, so the hypergeometric series here again truncates, and is shorter than the one in Eq. (B4), since $n+2<1$, according to Eq. (11).

The term $B_{n;ll'}(k,k')$ can be reduced by applying a binomial expansion to obtain

$$B_{n;ll'}(k,k')=(-1)^{m-1}k^{-n-1}x^{1-l'}\sum_{p=0}^{m-n-3}b_{n;ll';p}\,x^{2p}, \tag{B7}$$

where

$$\begin{aligned}b_{n;ll';p}&=2\sum_{q}\sum_{r}C_{l,q}C_{l',r}\sigma_{q+r-n-1}\\&\quad\times[(q+l'-n-2-2p)!\,(2p+r-l'+1)!\,]^{-1},\end{aligned} \tag{B8}$$

with the summations over q and r running over all values for which q, r, $(l-q)$, $(l'-r)$, $(2p+r-l'+1)$, and $(q+l'-n-2-2p)$ are all positive or zero. Since we have

$$\sigma_{q}=\int_{0}^{1}dt\,(t-1)^{-1}(t^{q}-1),$$

we may rewrite Eq. (B8) to give

$$\begin{aligned}B_{n;ll';p}&=2\int_{0}^{1}dt\,(t-1)^{-1}[G_{l,l'-n-2-2p}(t)G_{l',2p+1-l'}(t)\\&\quad-G_{l,l'-n-2-2p}(1)G_{l',2p+1-l'}(1)],\end{aligned} \tag{B9}$$

where

$$G_{l,q}(t)=\sum_{r}C_{l,r}t^{q+r}/(q+r)!. \tag{B10}$$

It is useful to note the relations

$$d^{n}G_{l,q}(t)/dt^{n}=G_{l,q-n}(t), \tag{B11}$$

$$G_{l,q}(0)=0,\quad q>0. \tag{B12}$$

Also, we have

$$G_{l,q}(t) = (t^q/q!)F(l+1,-l;q+1;t/2), \quad q \geq 0. \tag{B13}$$

and in particular,

$$G_{l,l}(t) = t^l(1-t/2)^l/l! = (2^l l!)^{-1}[1-(t-1)^2]^l. \tag{B14}$$

For $q \geq 0$, we may obtain $G_{l,q}(t)$ from Eq. (B13), since $F(l+1,-l;q+1;\frac{1}{2})$ can be evaluated in closed form.[7] With the aid of Eq. (B3) we find

$$G_{l,q}(1) = \pi^{1/2}2^{-q}[\Gamma(q/2+l/2+1)\,\Gamma(q/2-l/2+1/2)]^{-1}. \tag{B15}$$

For $q \leq l$, this result may also be obtained using Eqs. (B11) and (B14), which also shows that it remains valid for $q < 0$. In addition, we note

$$G_{l,q}(1) = 0 \quad \text{for } q < l \text{ and } l-q \text{ odd}. \tag{B16}$$

We first consider the case where we have the the two conditions

$$p \leq l'-1, \quad l'-n-2-2p \geq 0. \tag{B17}$$

Then we may expand $G_{l',2p+1-l}(t)$ in a Taylor series about $t=1$ and apply Eq. (B16) to bring Eq. (B9) to the form

$$b_{n;ll';p} = 2\sum_{q=0}^{p} G_{l',2p-2q-l}(1) \times \int_0^1 dt\, G_{l,l'-n-2-2p}(t)\,(t-1)^{2q}/(2q+1)!. \tag{B18}$$

The integrals here are readily reduced by repeated integration by parts, using Eq. (B11). In view of the second condition in Eq. (B17), Eq. (B12) applies, and we obtain

$$b_{n;ll';p} = \sum_{q=0}^{p} G_{l';2p-2q-l}(1)G_{l,l'-n-1+2p-2p}(1)/(q+\tfrac{1}{2}) = 2^{n+1}\pi\sum_{q=0}^{p}[q!\,(m-n-2-q)!\,(p-q+1/2) \times \Gamma(q+1/2-l')\,\Gamma(l'+3/2-m-q)]^{-1}. \tag{B19}$$

We claim that if either condition in Eq. (B17) is invalid, then *both* of the conditions

$$m-n-3-p \leq l-1, \quad 2p+1-l' \geq 0, \tag{B20}$$

will be valid. If the first condition in Eq. (B17) is invalid, we have $p \geq l'$, which assures $2p+1-l \geq 0$. The other condition in Eq. (20) must also hold, because the alternative would contradict Eq. (2). Also, if $l'-n-1-2p \leq 0$, this insures the first condition in Eq. (B20) , and the second condition must also hold, in order to avoid contradicting Eq. (11). Therefore, the cases not handled by Eqs. (B17) and (B19) may be disposed of by the relation

$$b_{n;ll';p} = b_{n;l'l;\,m-n-3-p}. \tag{B21}$$

[1] A. Erdelyi, *et al.*, *Higher Transcendental Functions, Vol. II* (McGraw-Hill, New York, 1954), p. 48.
[2] R. Conn and H. Rabitz, J. Chem. Phys. **61**, 600 (1974).
[3] G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1944), p. 48.
[4] For the case $l=l'$, see Ref. 3, p. 389. For other special cases (most notably, $n=\frac{3}{2}$), see W.N. Bailey, Proc. Lond. Math. Soc. (2) **40**, 37 (1936).
[5] G. Eason, B. Noble, and I.N. Sneddon, Phil. Trans. Roy. Soc. London A **247**, 529 (1955).
[6] Reference 3, p. 401.
[7] L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966), p. 32.

Mirror planes in Newtonian stars with stratified flows[a)]

Lee Lindblom

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 7 June 1977)

This paper shows that a certain class of Newtonian stellar models must possess a plane of mirror symmetry. A corollary of this result is that static Newtonian stars must be spherical. The new features of the results given here are that: (a) The assumptions about the velocity distribution of the fluid are weaker than previous treatments and (b) the method of proof given here does not depend as strongly on the linearity of the gravitational field equations as the previously published treatments. Therefore, this proof may serve as a model for a general relativistic generalization of the mirror plane theorem.

1. INTRODUCTION

An interesting feature of equilibrium stellar models is that extra symmetries are acquired from the field equations and the boundary conditions by the stationary equilibrium configurations. Perhaps the oldest known result of this type is Lichtenstein[1,2] and Wavre's[3] proof that rotating Newtonian stellar models must have a plane of mirror symmetry which is perpendicular to the rotation axis of the star. A related theorem by Carleman[4] and Lichtenstein[2] shows that static Newtonian stellar models must be spherical. A few results are also known for general relativistic models: static black holes are spherical[5]; stationary black holes are axisymmetric[6]; and stationary viscous stars are axisymmetric.[7]

This paper will present a new type of proof of the mirror plane theorem for Newtonian stellar models. The assumptions on which the present proof is based are somewhat weaker than those used previously. It had been assumed that the fluid motion in the star was purely azimuthal; here we assume that there is a Cartesian coordinate system in which the z component of the velocity vanishes. (Thus, the velocity field of the fluid will be called stratified.) This weaker assumption allows us to consider somewhat more complex velocity distributions such as those in the Dedekind ellipsoids.[8] Furthermore, we do not make any assumption about stationarity here. Thus, we are able to prove the existence of mirror symmetry for objects which are nonaxisymmetric and rotate with respect to the inertial frame of reference (e.g., the Jacobi and Riemann S ellipsoids[8]).

The method of proof employed in the present work may also be of some interest. This proof is based on the maximum principles (see the Appendix) which must be satisfied by the solutions to certain elliptic differential equations. This proof depends in a less crucial way on the linearity of the gravitational field equations than the Green's function approach taken by Lichtenstein[2] and Wavre.[3] Therefore, the present type of proof is more likely to form the basis for a general relativistic generalization of this theorem than the previous approaches to this problem.

We now give a qualitative outline of the proof which is given in detail in the following sections. Section 2 makes explicit the physical and mathematical assumptions on which the proof of this theorem is based. The purpose of Sec. 3 is to construct the plane which is shown to be a mirror plane in Sec. 4. We begin by considering the set of chords which are parallel to the z axis, and which have both endpoints on the same level surface of the gravitational potential function. Lemma 2 is used to show that every point is the endpoint of some such chord. Next we consider the set of midpoints of those chords. For this purpose we define a function m_ϕ, which maps the endpoints of chords into their midpoints. In Lemma 3 we show that there is a chord whose midpoints z component, z_m, is larger than or equal to the z component of the midpoint of any other chord. We will decompose each of the functions into even and odd parts with respect to reflection about the plane $z = z_m$; and we will show that this plane is a mirror plane of the star. In Lemma 4 we derive the important fact that the odd part of the mass density, ϵ^-, is negative for all z exceeding z_m. In Sec. 4 we prove the main theorem. We show that the odd part of the gravitational potential, ϕ^-, must satisfy the differential equation $\nabla_i \nabla^i \phi^- = -4\pi G \epsilon^- \geq 0$ for all $z \geq z_m$; this follows from Lemma 4. In addition we argue that ϕ^- must have a maximum in the half space $z > z_m$. The maximum principles for this type of differential equation are then invoked to show that in fact $\phi^- = 0$ everywhere. It follows that the odd parts of the mass density and pressure must vanish also. Thus the star must have a plane of mirror symmetry.

2. NEWTONIAN STELLAR MODELS

We will be considering the properties of stratified Newtonian stellar models. These models are completely defined by the following functions of the Cartesian coordinates x, y, z:

ϕ—gravitational potential,

ϵ—mass density,

p—pressure,

v_i—velocity of fluid,

a_i—acceleration of fluid.

[a)]This research was supported by the National Science Foundation Grant GP-25548.

These functions are assumed to satisfy the usual differential equations which describe a Newtonian fluid stellar model:

$$\nabla_i \nabla^i \phi = -4\pi G\epsilon, \tag{1}$$

$$\epsilon a_i = -\nabla_i p + \epsilon \nabla_i \phi, \tag{2}$$

$$a_i = \partial v_i/\partial t + v^j \nabla_j v_i. \tag{3}$$

In addition we make the following assumptions:

(a) The z component of the velocity of the fluid vanishes (this defines our meaning of stratified).

(b) The density is a function of the pressure: $\epsilon(p)$ with $d\epsilon/dp \geq 0$, $\epsilon \geq 0$, and $p \geq 0$.

(c) The density has compact support at every instant of time.

(d) The gravitational potential, ϕ, vanishes as $x^2+y^2+z^2 \to \infty$.

(e) The gravitational potential is C^3 except at the boundary of the star, where it is C^1 with respect to normal derivatives and C^2 with respect to tangential derivatives.

(f) The magnitudes of the functions ϵ, p, and ϕ are bounded.

The following lemma shows that if the velocity field of the star is stratified, then the Euler's equation (2) can be written in an important simplified form.

Lemma 1: The Euler's equation for a Newtonian stellar model which satisifes assumptions (a) and (b) may be written in the form

$$\nabla_i p = \epsilon \nabla_i \psi, \tag{4}$$

where $\psi = \phi - \Upsilon$ and Υ is some function which is independent of z.

Proof: Equation (2) may be written in the form $a_i = -\epsilon^{-1}\nabla_i p + \nabla_i \phi$. When assumption (b) is satisfied, the right-hand side is a gradient, thus $a_i = \nabla_i \Upsilon \equiv -\epsilon^{-1}\nabla_i p + \nabla_i \phi$. This can be re-arranged into the form of Eq. (4). Also since $v_z = 0$ by assumption (a), it follows that $a_z = 0$ by Eq. (3). Therefore $a_z = \partial\Upsilon/\partial z = 0$. ∎

We note that for the special case of an azimuthal velocity field, $v^\varphi = \Omega$, the potential Υ takes the familiar form of the centrifugal potential, $\nabla_i \Upsilon = -\frac{1}{2}\Omega^2 \nabla_i (x^2+y^2)$.

3. PRELIMINARY LEMMAS

To construct the plane, $z = \text{const}$, which we show in Sec. 4 is a plane of mirror symmetry of the stellar model, we need to classify the points in the star, based on the nearby behavior of the gravitational potential ϕ.

Definition: A point (x, y, z) will be called normal if $\partial\phi/\partial z(x, y, z) \neq 0$; and a point will be called special if $\partial\phi/\partial z(x, y, z) = 0$.

Lemma 2: Let ϕ be the gravitational potential of a stratified Newtonian stellar model satisfying assumptions (a) through (f). For every normal point (x, y, z) there exists a unique associated point $(x, y, \bar{z})$ which has the property $\phi(x, y, z) = \phi(x, y, \bar{z})$ and $\phi(x, y, z) < \phi(x, y, z')$ for all z' between z and $\bar{z}$. (A special point is said to be associated with itself.)

Proof: Let us first show that $\phi(x, y, z) > 0$ everywhere. If there is a point with $\phi(x, y, z) \leq 0$, then we could find some point, say (x', y', z'), with $\phi(x', y', z') \leq \phi(x, y, z)$ for all points (x, y, z). By Eq. (1) and assumption (b) we have $\nabla_i \nabla^i \phi \leq 0$. Using Theorem 2A (see the Appendix) one can show that if the point (x', y', z') exists, then $\phi = 0$ everywhere. If the point (x', y', z') lies on the boundary of the star, a slightly different argument using Theorem 1A gives the same result, $\phi = 0$. Thus we can conclude that ϕ must be positive everywhere.

We next consider the normal point (x, y, z). One can start at (x, y, z) and proceed along the line $(x, y) = \text{const}$ in the direction of increasing ϕ. When one reaches points having sufficiently large values of $x^2+y^2+z^2$, the potential ϕ will become arbitrarily small. This guarantees that a point, say $(x, y, \hat{z})$, will be reached along the line at which $\phi(x, y, z) = \phi(x, y, \hat{z})$. If one takes the first such point reached along the line, say $(x, y, \bar{z})$, then $\phi(x, y, z') > \phi(x, y, z)$ for all z' between z and $\bar{z}$. Thus $(x, y, \bar{z})$ is associated with (x, y, z) and the lemma is proved. ∎

To assist in the construction of the plane which is shown to be a symmetry plane of the model in Sec. 4, we will consider the following function.

Definition: The function m_ϕ maps points (x, y, z) from the support of the mass density function into some subset of R^3. We define

$$m_\phi(x, y, z) \equiv (x, y, \tfrac{1}{2}[z+\bar{z}]), \tag{5}$$

where $(x, y, \bar{z})$ is the point associated with (x, y, z).

The next lemma will derive an important property of the function m_ϕ.

Lemma 3: There exists a point (x_0, y_0, z_0) in the domain of m_ϕ, whose image $(x_0, y_0, z_m) \equiv m_\phi(x_0, y_0, z_0)$ is a least upper bound of the z component of the range of m_ϕ; i.e., for every point (x, y, z) in the range of m_ϕ, $z_m \geq z$.

Proof: Let us first argue that the z components of the range of m_ϕ are bounded. We can consider the total potential ψ, defined in Lemma 1. The function m_ψ, constructed using ψ rather than ϕ, is identical to the function m_ϕ because $\phi - \psi = \Upsilon$ is independent of z. By Eq. (4) the level surfaces of ψ coincide with the level surfaces of the functions ϵ and p. Therefore the points which are associated with normal points within the support of the density will also lie within the support of the density. Thus, the range of m_ϕ must be bounded since the domain is bounded by assumption (c). Since the range of m_ϕ is bounded, the z component of the range must also be bounded and therefore must have a least upper bound, say z_m.

We will now show that z_m is the z component of some element in the range of m_ϕ. In any case, there must be a sequence of numbers ζ_n each of which is the z component of some element of the range of m_ϕ, and $\lim \zeta_n = z_m$. There must also be a corresponding sequence of points ξ_n in the domain of m_ϕ whose images have ζ_n as z com-

ponents: $m_\phi(\xi_n)=(x_n, y_n, \zeta_n)$. The domain of m_ϕ is compact, therefore, there is a subsequence ξ_n' of ξ_n which converges to a point in the domain, say $\lim\xi_n' = (x_0, y_0, z_0)$. It follows that $\lim m_\phi(\xi_n')=(x_0, y_0, z_m)$. The prime will henceforth be dropped from the name of the sequence of points ξ_n'. If m_ϕ were a continuous function, it would follow that $m_\phi(x_0, y_0, z_0)=(x_0, y_0, z_m)$ and the proof would be complete. m_ϕ is not necessarily continuous however.

Let us first consider the case where there is a subsequence ξ_n'' of ξ_n which are all special points. At each of these points we have $\partial\phi/\partial z(\xi_n'')=0$; and since $\partial\phi/\partial z$ is a continuous function, $(\partial\phi/\partial z)(z_0, y_0, z_0)=0$. For special points $m_\phi(x, y, z)=(x, y, z)$, therefore $\lim m_\phi(\xi_n'')=(x_0, y_0, z_0)=(x_0, y_0, z_m)$. Therefore (x_0, y_0, z_m) must be an element of the domain of m_ϕ with the property $m_\phi(x_0, y_0, z_m)=(x_0, y_0, z_m)$. Thus we have shown that the lemma follows if there exists a subsequence ξ_n'' of special points.

The other case we need to consider is when ξ_n are all normal points when n becomes sufficiently large. To each of the normal points ξ_n (with z component ω_n) there is an associated point $\bar\xi_n$ (with z component $\bar\omega_n$). We also know that $\lim\omega_n=z_0$ and $\lim\frac{1}{2}(\omega_n+\bar\omega_n)=z_m$, thus $\lim\bar\omega_m = 2z_m - z_0$. There are three possibilities: $z_0=z_m$, $z_0 > z_m$, and $z_0 < z_m$. We will consider first the case where $z_0=z_m$. The chord connecting each pair of points ξ_n to $\bar\xi_n$ in our sequence must contain a point ξ_n'', where $\partial\phi/\partial z(\xi_n'')=0$. Thus, the sequence ξ_n'' are all special points. Furthermore $\lim\xi_n''=\lim\xi_n=\lim\bar\xi_n=(x_0, y_0, z_0)$. Thus, we have a sequence of special points whose limit point is (x_0, y_0, z_0). We have shown above that the lemma follows in this case. We next consider the case where $z_0 > z_m$; then (x_0, y_0, z_0) must be a normal point with associated point $(x_0, y_0, \bar z_0)$. It follows that $\bar z_0 \le 2z_m - z_0$ because z_m is the least upper bound. Since ϕ is a continuous function $\lim\phi(\xi_n)=\phi(x_0, y_0, z_0)=\lim\phi(\bar\xi_n) = \phi(x_0, y_0, 2z_m - z_0)$. Therefore the point $(x_0, y_0, 2z_m - z_0)$ must be the point associated with (x_0, y_0, z_0) and as a result $m_\phi(x_0, y_0, z_0)=(x_0, y_0, z_m)$ and the lemma follows. The last possibility is that $z_0 < z_m$. In this case the sequence of associated points $\bar\xi_n$ must converge to $(x_0, y_0, 2z_m - z_0)$ and $2z_m - z_0 > z_m$. The same argument as the one given for the case $z_0 > z_m$ shows that (x_0, y_0, z_0) is the point associated with $(x_0, y_0, 2z_m - z_0)$. In this case $m_\phi(x_0, y_0, 2z_m - z_0)=(x_0, y_0, z_m)$ and the lemma follows. ∎

We can now derive a very important inequality for the old part of the density function, when it is taken with respect to the plane $z=z_m$.

Lemma 4: Let ϵ be the mass density of a stratified Newtonian stellar model satisfying assumptions (a) through (f). Then,

$$\epsilon^-(x, y, z)\equiv\tfrac{1}{2}\epsilon(x, y, z) - \tfrac{1}{2}\epsilon(x, y, 2z_m - z)\le 0 \quad \forall \quad z \ge z_m.$$

Proof: Consider a point (x, y, z) with $z > z_m$. If (x, y, z) is not in the support of ϵ, then $\epsilon^-(x, y, z) = -\frac{1}{2}\epsilon(x, y, 2z_m - z)\le 0$ by assumption (b). Next suppose that (x, y, z) is in the support of ϵ. Since z_m is the least upper bound of the midpoints, (x, y, z) must be a normal point and the associated point $(x, y, \bar z)$ must satisfy $\bar z \le 2z_m - z \le z$. Lemma 2 implies $\phi(x, y, 2z_m - z) \ge \phi(x, y, z)$ so that $\phi^-(x, y, z)=\frac{1}{2}\phi(x, y, z) - \frac{1}{2}\phi(x, y, 2z_m - z)\le 0$. The total potential ψ, defined in Lemma 1 satisfies $\psi^-=\phi^-$, because Υ is independent of z; consequently $\psi^-(x, y, z)\le 0$. From Eq. (4) it follows that the level surfaces of ϵ, p, and ψ all coincide. This fact and the requirement that $\epsilon\ge 0$, $p\ge 0$ and $d\epsilon/dp\ge 0$ from assumption (b) imply that $\epsilon^-(x, y, z)\le 0$ for all $z\ge z_m$. ∎

4. THE MAIN THEOREM

We can now prove that these stratified Newtonian stellar models have a plane of mirror symmetry.

Theorem: Consider a stratified Newtonian stellar model which satisfies assumptions (a) through (f). There exists a plane $z=z_m$, such that the odd parts of the functions ϕ, ϵ and p vanish when taken with respect to the plane $z=z_m$. Thus, the star has a plane of mirror symmetry for these functions

Proof: From Lemma 3 we know that there is a point (x_0, y_0, z_0) such that $m_\phi(x_0, y_0, z_0)=(x_0, y_0, z_m)$. We will consider two separate cases. In the first case (x_0, y_0, z_0) is assumed to be a normal point, in the second case it is assumed to be a special point.

Case 1: Associated with the point (x_0, y_0, z_0) is the point $(x_0, y_0, \bar z_0)$ with $\bar z_0=2z_m - z_0$. Since $\phi^-(x_0, y_0, z_0) = \frac{1}{2}\phi(x_0, y_0, z_0) - \frac{1}{2}\phi(x_0, y_0, \bar z_0)=0$, there exists a point [either (x_0, y_0, z_0) or $(x_0, y_0, \bar z_0)$] say (x_0, y_0, z_0) with $z_0 > z_m$, where ϕ^- vanishes. The function ϕ^- vanishes on the boundary of the half space $z > z_m$. In the interior of this region ϕ^- is bounded due to assumption (f); therefore there must exist a point $(\hat x, \hat y, \hat z)$ in this half space where ϕ^- is maximal. The odd part of Eq. (1) is given by $\nabla_i\nabla^i\phi^- = -4\pi G\epsilon^-$. From Lemma 4 we have $\nabla_i\nabla^i\phi^-\ge 0$ for all $z > z_m$. This inequality, the existence of a point where ϕ^- is maximal and Theorem 2A (see the Appendix) guarantee that $\phi^-=0$ everywhere. That $\epsilon^-=p^-=0$ follows trivially.

The argument given above is not strictly correct for the case where the maximum of ϕ^- lies on the boundary of the star. The density ϵ need not be continuous at the surface of the star, and consequently the potential ϕ need not be sufficiently differentiable there to apply Theorem 2A. Consider now the case where the maximum of ϕ^-, $(\hat x, \hat y, \hat z)$ lies on the boundary of the star. Find an open ball B which has $(\hat x, \hat y, \hat z)$ as a point on its boundary, which is tangent to the surface of the star at $(\hat x, \hat y, \hat z)$ and which is sufficiently small that all of the points of B lie in the exterior of the star. Within B, ϕ^- will be C^3, and ϕ^- is C^1 at (x, y, z). Furthermore $\phi^-\le\phi^-(\hat x, \hat y, \hat z)$ at all points in B and $\nabla_i\phi^-(\hat x, \hat y, \hat z)=0$, since ϕ^- is a maximum at $(\hat x, \hat y, \hat z)$. From Theorem 1A if follows that ϕ^- has the constant value $\phi^-(\hat x, \hat y, \hat z)$ everywhere in B and consequently everywhere. This constant value must be zero since ϕ^- vanishes on the boundary of the half-space $z > z_m$.

Case 2: We now consider the case where (x_0, y_0, z_0) is a special point. We have shown that $\phi^-\le 0$ and $\epsilon^-\le 0$ for all $z\ge z_m$. Similarly $\phi^-\ge 0$ and $\epsilon^-\ge 0$ for all $z\le z_m$. It follows that there is a neighborhood U of the plane $z=z_m$ in which the following inequalities must hold: $\partial\phi^-/\partial z\le 0$, $\partial\epsilon^-/\partial z\le 0$. From Eq. (1) it follows that

$\nabla_i \nabla^i(\partial\phi^-/\partial z) = -4\pi G \partial\epsilon^-/\partial z$, hence $\nabla_i \nabla^i(\partial\phi^-/\partial z) \geq 0$ in U. At a special point $\partial\phi/\partial z = 0 = \partial\phi^+/\partial z + \partial\phi^-/\partial z$, but at $z = z_m$, $\partial\phi^+/\partial z$ vanishes, therefore $\partial\phi^-/\partial z(x_0, y_0, z_m) = 0 \geq \partial\phi^-/\partial z$ for all points in U. By Theorem 2A it follows that $\partial\phi^-/\partial z = 0$ everywhere in U, and consequently $\phi^- = 0$ everywhere in U, and as a result $\phi^- = 0$ everywhere.

As in Case 1 special consideration must be given to the case that (x_0, y_0, z_m) is on the boundary of the star. From assumption (e) we know that ϕ must be at least C^1 in the normal direction, and C^2 in the tangential direction at the surface of the star. Therefore Theorem 2A cannot be applied and Theorem 1A must be used. Since (x_0, y_0, z_m) is a special point, it follows that $\partial\phi/\partial z = \partial\psi/\partial z = 0$. There $\partial/\partial z$ is a tangential derivative to the surface at this point; it follows that $\partial\phi^-/\partial z$ is C^1 at (x_0, y_0, z_m). We have argued that $\partial\phi^-/\partial z \leq 0$ in the set U. Thus, $\partial\phi^-/\partial z$ will be a maximum at (x_0, y_0, z_m) so that $\nabla_i(\partial\phi^-/\partial z) = 0$ there also. Construct an open ball B which contains (x_0, y_0, z_m) as one of its boundary points, which is tangent to the surface of the star at (x_0, y_0, z_m), and which is sufficiently small that B lies completely within U and completely within the exterior of the star. Within B, $\nabla_i \nabla^i(\partial\phi^-/\partial z) = 0$ and $\partial\phi^-/\partial z$ is C^2. Thus by Theorem 1A, $\partial\phi^-/\partial z = 0$ in B, and therefore $\phi^- = 0$ in B (the plane $z = z_m$ intersects the center of B). It follows that $\phi^- = 0$ everywhere since it vanishes at an interior point of the half space $z > z_m$. ∎

5. DISCUSSION

In the special case of static stellar models ($v_i = 0$) there is no orientation picked out by the velocity stratification. Therefore, the Theorem proved in the last section shows that a mirror plane must exist for any choice of orientation. As a result, one can show that the star must be spherical.[2,4] We also note that the mirror plane theorem in the last section is in a sense incomplete. We have shown that the functions ϵ, p, and ϕ must all have mirror symmetry. However, it appears that no simple analogous result exists for the velocity field of the fluid, v^i. For example, consider a stationary axisymmetric star with azimuthal velocity field. An infinite number of related stellar models may be constructed by keeping the functions ϵ, p, and ϕ fixed while defining a new velocity field $v'^i = hv^i$, where h is an arbitrary function which is independent of azimuthal angle and $h^2 = 1$. Note that h may be discontinuous, so that parts of the fluid may rotate one direction while other parts rotate the other way. These related stellar models need not have simple mirror symmetry in the velocity field. A final point to note is that asumption (a), that the velocity field is stratified, is only used to prove Lemma 1. This assumption could be replaced by the weaker (but physically less transparent) assumption $0 = a_z = \partial v_z/\partial t + v^j \nabla_j v_z$.

ACKNOWLEDGMENT

I would like to thank D. Brill, W. Hiscock, J. Isenberg, and P. Yasskin for helpful suggestions and comments.

APPENDIX

We reproduce here the version of the maximum principle on which the proof of the mirror plane theorem is based. Reference 9 may be consulted for discussions of these results, and also for stronger versions of the theorems than are needed here.

Theorem 1*A*: Let B be an open ball, and x_0 a point on its boundary. Assume that f is a C^2 function everywhere in B, and C^0 in the closure of B. Let $\nabla_i \nabla^i f \geq 0$ and $f \leq f(x_0)$ everywhere in B. Then the outward normal derivative $df/dn > 0$ at x_0, or $f = f(x_0)$ everywhere in B.

Theorem 2*A*: Assume that f is a C^2 function everywhere in a bounded open neighborhood U, and that $\nabla_i \nabla^i f \geq 0$ everywhere in U. If there is a point x_0 in U such that $f(x_0) \geq f(x)$ for all x in U, then $f(x_0) = f(x)$ for all x in U.

[1] L. Lichtenstein, Sitzungsber. Preuss. Akad. Wiss., Phys.-Math. Kl. 1120 (1918).
[2] L. Lichtenstein, Gleichgewichtsfiguren Rotierender Flüssigkeiten (Springer Verlag, Berlin, 1933).
[3] R. Wavre, *Figures Planetaries et Geodesie* (Gauthiar-Villars, Paris, 1932).
[4] T. Carleman, Math. Z. B 3, 1 (1919).
[5] W. Israel, Phys. Rev. **164**, 1776 (1967).
[6] S. Hawking, Commun. Math. Phys. **25**, 152 (1972).
[7] L. Lindblom, Ap. J. **208**, 873 (1976).
[8] S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale University, New Haven, 1969).
[9] L. Bers, F. John, and M. Schecter, *Lectres in Applied Mathematics: Partial Differential Equations* (Interscience, New York, 1964), Vol. 3, p. 150.

Some solutions of stationary, axially-symmetric gravitational field equations[a)]

Metin Gürses

Physics Department, Middle East Technical University, Ankara, Turkey
(Received 9 March 1977)

Stationary, axially-symmetric solutions of the gravitational field equations for vacuum, perfect fluid, and massless scalar field are considered. For the vacuum case, a similiar formulation to the one introduced by Ernst is presented by use of quarternions. Null dust solutions are found, and it is shown that they match with the van Stockum exterior solutions. An extension of the theorem by Eriş and Gürses is given which enables one to construct solutions to the gravitational field equations coupled with a charged dust and a massless scalar field from the solutions of the field equations coupled only with a charged dust.

1. INTRODUCTION

Stationary, axially-symmetric solutions of the vacuum and electrovacuum gravitational field equations have been extensively studied, and certain classes of solutions have been found. Some of these classes are the Lewis,[1] van Stockum,[2] Papapetrou,[3] and Tomimatsu–Sato[4] solutions. The first three classes assume a functional relationship between the metric coefficients while the last one has been obtained by using computer logic. It is a three-parameter solution class with parameters describing deformation, mass, and angular momentum where the sum of the squares of the last two parameters is unity. Among all formulations of the vacuum and electrovacuum field equations, the Ernst[5] complex-potential formalism has certain advantages, one of these being generation of new vacuum and electrovacuum solutions from the old ones. Such a generation of solutions follows from the invariance of the Ernst equation under a bilinear transformation. Geroch[6] has shown that the same invariance exists in the vacuum field equations of any space–time having only a nonnull Killing vector.

In the second section of the this work we present a similar formulation of the vacuum field equations to the one introduced by Ernst, by use of quaternionic potentials. In this formalism the Lewis and van Stockum classes of solutions follow immediately. We obtain a Tomimatsu–Sato type class where in this case the difference of the squares of the parameters corresponding to mass and rotation is unity. The field equations are invariant under a quaternionic bilinear transformation.

Construction of stationary, axially-symmetric interior solutions to the gravitational field equations is one of the most difficult problems in general relativity. The difficulty arises from the complexity of the field equations. Some approximate[7] and nonfluid[8] solutions have been found but unfortunately no exact rotating fluid solution exists which matches with the Kerr metric. In order to approach such a solution, one should start with simple systems, such as null and nonnull dust distributions. In the third section we study the interior gravitational field equations for the null-dust case and give a complete solution. We obtain the general relativistic form of the Euler equation for a null rotating perfect fluid and also prove that the energy conservation equation (the Euler equation) is nothing but the integrability condition for one of the metric functions. This equation turns out to be very useful in the integration of the field equations for the case of dust distributions. It is interesting that all null dust solutions can be matched with an appropriate exterior solution of the van Stockum class.

Works on the solutions of the gravitational field equations coupled with a massless scalar field are quite recent and most of them have considered static[9] or conformally flat[10,11] space–times. Recently,[12] it was shown that it is possible to generate the axially-symmetric solutions of the field equations coupled with the electromagnetic and scalar fields from the Einstein–Maxwell solutions. In the fourth section, a generalization of this theorem to gravitational field equations coupled with charged dust and massless scalar field is given.

The components of the Ricci tensor in an orthonormal tetrad are given in the Appendix.

2. QUATERNIONIC POTENTIALS FOR VACUUM FIELD EQUATIONS

Gravitational field equations for vacuum are (see Appendix)

$$\nabla^2\psi + \frac{\exp(4\psi)}{2\rho^2}(\nabla\omega)^2 = 0, \tag{1}$$

$$\nabla\cdot(\rho^{-2}\exp(4\psi)\nabla\omega) = 0. \tag{2}$$

Once ω and ψ are found the remaining metric coefficient γ can be found by use of quadratures (A10) and (A11). Defining a new function f

$$\psi = \tfrac{1}{2}\ln(\rho/f). \tag{3}$$

Eqs. (1) and (2) become

$$f\nabla^2 f - (\nabla f)^2 - (\nabla\omega)^2 = 0, \tag{4}$$

$$\nabla\cdot(f^{-2}\nabla\omega) = 0. \tag{5}$$

We now introduce a quaternionic potential ε such that

$$\varepsilon = f + e\omega, \tag{6}$$

where e is one of the three quaternionic units with $e^2 = 1$. Similar to complex conjugation we define the quaternionic conjugation as

$$\varepsilon^* = f - e\omega. \tag{7}$$

Functions f and ω are, respectively, the scalar and

[a)]Research partially supported by the Scientific and Technical Research Council of Turkey.

vector parts of the quaternionic function ε. Hence it is clear that Eqs. (4) and (5) are the scalar and vector parts of the following differential equation:

$$(\varepsilon + \varepsilon^*)\nabla^2\varepsilon = 2\nabla\varepsilon \cdot \nabla\varepsilon. \tag{8}$$

In terms of a new quaternionic function ξ defined by

$$\varepsilon = (\xi - 1)/(\xi + 1), \tag{9}$$

Eq. (8) is written as

$$(\xi\xi^* - 1)\nabla^2\xi = 2\xi^*\nabla\xi \cdot \nabla\xi. \tag{10}$$

Equation (10) can be obtained from the Lagrangian density

$$\mathcal{L} = \nabla\xi \cdot \nabla\xi^*/(\xi\xi^* - 1)^2, \tag{11}$$

which is nothing but the Einstein Lagrantian density $\sqrt{-g}R$. This Lagrangian density is invariant under the bilinear transformation

$$\xi' = (a\xi + b)/(c\xi + d) \quad (ad - bc \neq 0) \tag{12}$$

where a, b, c, and d represent eight real constants of which only three are independent. The group of this three-parameter transformation is isomorphic to one of the well-known two-dimensional noncompact groups. Invariance of the Lagrangian density (11) under the transformations (12) leads to the generation of the new solutions of the vacuum field equations from the known solutions. Using the Ernst trick, one may also obtain solutions to the electrovacuum field equations.

The norm of a quaternion $Q = Q_s + eQ_v$ is defined as

$$QQ^* = Q_s^2 - Q_v^2. \tag{13}$$

Hence, vanishing of the norm does not necessarily lead to the vanishing of the scalar and vector parts of Q. With the properties of quaternions, we give three simple solutions to Eq. (10). They are

$$\xi = \exp(e\alpha)\coth\Theta, \tag{14}$$

$$\xi = e\exp(e\alpha)\tan\Theta, \tag{15}$$

$$\xi = (1 \pm e)\Theta, \tag{16}$$

where Θ is a real function satisfying

$$\nabla^2\Theta = 0, \tag{17}$$

and α is an arbitrary real constant. The solutions (14), (15), and (16) are called Lewis, Lewis, and van Stockum classes,[13] respectively. These classes of solutions are not easily seen in the Ernst complex potential formulation. On the other hand, the Papapetrou class can not be directly obtained in our formulation. The other well-known class is the Tomimatsu–Sato solutions. These solutions are the twisting generalization of the Weyl static vacuum metrics. They have three parameters, δ (distortion), p (mass), and q (twist) with $p^2 + q^2 = 1$. We have a similar class of solutions with $p^2 - q^2 = 1$. For example, in the oblate spheroidal coordinates

$$\xi = px + eqy, \quad (p^2 - q^2 = 1)$$

is one of the solutions of (10), corresponding to $\delta = 1$. One can also generate the NUT parameter using the invariance of (10) under the transformation $\xi' = \exp(e\alpha)\xi$.

3. NULL FLUID

The Einstein field equations with a null fluid in a tetrad basis are

$$R_b^a = -\kappa[(p + \epsilon)u^a u_b - p\delta_b^a], \tag{18}$$

where p and ϵ are the pressure and the energy density of the fluid and u^a are the components of the fluid 4-velocity, with $u^a u_a = 0$. In a stationary, axially symmetric space–time the only nonvanishing components of the fluid velocity vector u^a are u^0 and u^3. For the case of null fluid

$$u^0 = \eta u^3 = u, \quad \eta = \pm 1,$$

and the field equations are

$$\rho\lambda^{-1}\nabla \cdot (\rho^{-1}\lambda\nabla\psi) + (\exp(4\psi)/2\lambda^2)(\nabla\omega)^2 = \kappa\exp(2\gamma - 2\psi)(p_* u^2 + p), \tag{19}$$

$$\lambda^{-1}\rho\nabla \cdot (\rho^{-1}\nabla\lambda) = 2\kappa\exp(2\gamma - 2\psi)p, \tag{20}$$

$$\rho\nabla \cdot (\rho^{-1}\lambda^{-1}\exp(4\psi)\nabla\omega) = -2\kappa\eta\exp(2\gamma)p_* u^2. \tag{21}$$

Equations (13) and (14) will be considered as the equations to determine the metric function γ. Instead of (A11) we take the following equation as one of the field equations which is nothing but the energy conservation equation (contracted Bianchi identity):

$$\nabla p/p_* = u^2\lambda^{-1}\exp(2\psi)\nabla(\lambda\exp(-2\psi) - \eta\omega), \tag{22}$$

where $p_* = p + \epsilon$. The ρ component of this equation is obtained through the addition of the derivatives of (A13) and (A14) with respect to z and ρ, respectively, while the z component is obtained by subtracting the derivatives of (A13) and (A14) with respect to ρ and z, respectively. It is obvious that when $p = 0$ (null-dust or noninteracting null gaseous), the field equations become much simpler. First we can use the coordinate condition (A12) and obtain the following relation between ω and ψ by use of Eq. (22)

$$\omega = \eta\rho\exp(-2\psi). \tag{23}$$

This relation enables us to integrate completely the quadratures for γ given in (A13) and (A14). The result of the integration is

$$\exp(2\gamma - 2\psi) = \rho^{-1/2}. \tag{24}$$

Then the line element becomes

$$ds^2 = -\exp(2\psi)\,dt^2 + 2\eta\rho\,d\phi\,dt + \rho^{-1/2}(d\rho^2 + dz^2), \tag{25}$$

where ψ satisfies

$$\nabla^2 U + VU = 0, \tag{26}$$

with

$$U = \rho^{-1}\exp(2\psi) \tag{27}$$

$$V = 2\kappa\rho^{-1/2}\epsilon u^2. \tag{28}$$

Hence for a given source ϵu^2, the metric in (25) with (26) defines the gravitational field of a noninteracting null dust (or simply the null electromagnetic field). When ϵu^2 is a constant, the solution of (26) may be reduced to one type of a Bessel function. These solutions may be matched exactly to the van Stockum exterior metrics which are of the form given in Eq. (25) with

$$\nabla^2 U = 0. \tag{29}$$

In the coordinate basis the only nonvanishing covariant component of the Ricci tensor is $R_{\hat{0}\hat{0}}$; hence

$$R_{\mu\nu} = -\epsilon_0 n_\mu n_\nu, \tag{30}$$

where

$$n_\mu = e^{\psi}\delta^{\hat{0}}_{\mu}, \quad n^\mu = \rho^{-1}\delta^{\mu}_{\hat{3}}, \tag{31}$$

and

$$\epsilon_0 = \kappa\epsilon u^2. \tag{32}$$

4. MASSLESS SCALAR FIELD

Recently[13] it was shown that one can generate the solutions to the coupled Einstein–Maxwell massless scalar field equations from the known solutions of the Einstein–Maxwell equations. Here, we give an extension of the above theorem to the case when the source is a charged dust and a massless scalar field.

Theorem: Let $\psi, \omega, \gamma, A_0, A_3$, and μ be a solutions of the Einstein field equations coupled with an electromagnetic field, and a dust distribution, where A_0 and A_3 are the nonzero components of the electromagnetic vector potential and μ is the energy density of the dust distribution. Then $\psi, \omega, \gamma+\gamma^{\Phi}, A_0, A_3, \mu\exp(-2\gamma^{\Phi})$, and Φ form a solution to the Einstein field equations coupled with an electromagnetic field, a dust distribution, and a massless scalar field Φ, where

$$2\gamma^{\Phi}_{,\rho} = \rho[(\Phi_{,\rho})^2 - (\Phi_{,z})^2],$$

$$2\gamma^{\Phi}_{,z} = 2\rho\Phi_{,\rho}\Phi_{,z},$$

and

$$\nabla^2\Phi = 0.$$

Using this theorem, one may obtain solutions to the gravitational field equations coupled to a null dust and a massless scalar field once a solution of Eq. (26) is given.

5. CONCLUDING REMARKS

We presented a quaternionic potential formulation of the stationary, axially-symmetric vacuum gravitational field equations and obtained a class of "Tomimatsu–Sato"-like solutions which does not contain any algebraically special metrics.

We gave the complete solution of the stationary, axially-symmetric gravitational field equations coupled with a null dust or a null-electromagnetic field. For the case of the null-electromagnetic field, $F_{ab} = 2\epsilon^{1/2}un_{[a}l_{b]}$, where l_b is a unit spacelike vector which is orthogonal to n_a.[14] These solutions match with the van Stockum exterior solutions on a cylindrical boundary. We showed that the energy conservation or the contracted Bianchi identity is the integrability condition for the metric coefficient γ. This is, in fact, true for any energy–momentum distribution.

We presented a theorem to produce the solutions of the gravitational field equations coupled to a charged dust and a massless scalar field from the solutions of the field equations coupled with a charged dust. In fact, it is possible to extend this theorem further for any covariantly conserved energy–momentum tensor T^a_b plus a massless scalar field when the condition (A12) is satisfied for stationary, axially-symmetric space–times.

ACKNOWLEDGMENTS

We thank Dr. A. Eriş and Dr. F. Öktem for helpful discussions.

APPENDIX

The stationary, axially symmetric metric is given by

$$ds^2 = -\exp(2\psi)(dt - \omega\,d\phi)^2 + \exp(-2\psi)[\exp(2\gamma)(d\rho^2 + dz^2) + \lambda^2\,d\phi^2]. \tag{A1}$$

Choosing the basis 1-forms as

$$\omega^a = h^a{}_\mu\,dx^\mu, \tag{A2}$$

the *vier bein* components $h^a{}_\mu$ are

$$h^{\hat{0}}{}_{\hat{0}} = \exp(\psi), \quad h^{\hat{0}}{}_{\hat{3}} = -\omega\exp(\psi),$$
$$h^{\hat{3}}{}_{\hat{3}} = \lambda\exp(-\psi), \quad h^{\hat{1}}{}_{\hat{1}} = h^{\hat{2}}{}_{\hat{2}} = \exp(\gamma-\psi), \tag{A3}$$

and the inverse components $h_a{}^\mu$ can be found using relations

$$h^a{}_\nu h_a{}^\mu = \delta^\mu{}_\nu \tag{A4}$$

Here the Latin indices denote the orthonormal tetrad components and run from 0 and 3 and the Greek indices denote the coordinate components running from $\hat{0}$ to $\hat{3}$. Nonvanishing components of the Ricci tensor with the convention

$$R^a{}_b = R^{ac}{}_{cb}, \tag{A5}$$

are found as

$$R^0{}_0 = \exp(2\psi-2\gamma)[\rho\lambda^{-1}\nabla\cdot(\rho^{-1}\lambda\nabla\psi) + (\exp(4\psi)/2\lambda^2)(\nabla\omega)^2], \tag{A6}$$

$$R^0{}_3 = -\tfrac{1}{2}\rho\exp(-2\gamma)\nabla\cdot(\rho^{-1}\lambda^{-1}\exp(4\psi)\nabla\omega), \tag{A7}$$

$$R^0{}_0 + R^3{}_3 = \rho\lambda^{-1}\exp(2\psi-2\gamma)\nabla\cdot(\rho^{-1}\nabla\lambda), \tag{A8}$$

$$R^1{}_2 = \exp(2\psi-2\gamma)[2\psi_{,\rho}\psi_{,z} - (\exp(4\psi)/2\lambda^2)\omega_{,\rho}\omega_{,z} + \lambda^{-1}(\lambda_{,\rho z} - \lambda_{,\rho}\gamma_{,z} - \lambda_{,z}\gamma_{,\rho})], \tag{A9}$$

$$R^1{}_1 - R^2{}_2 = \exp(2\psi-2\gamma)[2(\psi^2_{,\rho} - \psi^2_{,z}) - (\exp(4\psi)/2\lambda^2)(\omega^2_{,\rho} - \omega^2_{,z}) + \lambda^{-1}(\lambda_{,\rho\rho} - \lambda_{,zz} - 2\lambda_{,\rho}\gamma_{,\rho} + 2\lambda_{,z}\gamma_{,z})], \tag{A10}$$

$$R^1{}_1 + R^2{}_2 = \exp(2\psi-2\gamma)[-2\rho\lambda^{-1}\nabla\cdot(\rho^{-1}\lambda\nabla\psi) + 2(\nabla\psi)^2 - (\exp(4\psi)/2\lambda^2)(\nabla\omega)^2 + 2(\gamma_{,\rho\rho} + \gamma_{,zz}) + \lambda^{-1}(\lambda_{,\rho\rho} + \lambda_{,zz})], \tag{A11}$$

where ∇, $\nabla\cdot$, and ∇^2 are the grad, the divergence, and the Laplace operators defined in flat space cylindrical coordinates, respectively. Throughout this work we consider the case

$$R^0{}_0 + R^3{}_3 = 0, \tag{A12}$$

which enables us to use the coordinate condition $\lambda = \rho$. We will discuss the other possible case $\lambda = \text{const}$ in a later communication.

For the perfect fluid case $R^1{}_2$ and $R^1{}_1 - R^2{}_2$ vanish; hence we have

$$\lambda_{,\rho}\gamma_{,z} + \lambda_{,z}\gamma_{,\rho} = \lambda(2\psi_{,\rho}\psi_{,z} - (\exp(4\psi)/2\lambda^2)\omega_{,\rho}\omega_{,z}) + \lambda_{,\rho z}, \tag{A13}$$

$$\lambda_{,\rho}\gamma_{,\rho} - \lambda_{,z}\gamma_{,z} = \lambda(\psi^2_{,\rho} - \psi^2_{,z}) - (\exp(4\psi)/4\lambda)(\omega^2_{,\rho} - \omega^2_{,z}) + \tfrac{1}{2}(\lambda_{,\rho\rho} - \lambda_{,zz}). \tag{A14}$$

[1]T. Lewis, Proc. R. Soc. (London) Ser. A **136**, 176 (1932).
[2]W. J. van Stockum, Proc. R. Soc. Edinburg **57**, 135 (1937).
[3]A. Papapetrou, Ann. Phys. (Leipzig) **12**, 309 (1953).
[4]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972); Prog. Theor. Phys. **50**, 95 (1973).
[5]F. J. Ernst, Phys. Rev. **168**, 1415 (1968); Phys. Rev. D **7**, 2520 (1973).
[6]R. Geroch, J. Math. Phys. **10**, 953 (1972).
[7]D. R. Brill and J. M. Cohen, Phys. Rev. **143**, 1011 (1966); P. S. Florides and J. L. Synge, Proc. R. Soc. (London) Ser. A **280**, 499 (1964); J. B. Hartle and K. S. Thorne, Ap. J. **153**, 807 (1968).
[8]M. Gürses and F. Gürsey, J. Math. Phys. **16**, 2385 (1975).
[9]A. I. Janis, E. T. Newman, and S. Winicour, Phys. Rev. Lett. **20**, 878 (1968); R. Penney, Phys. Rev. **174**, 1578 (1968); **182**, 1383 (1969); A. I. Janis, D. C. Robinson, and J. Winicour, *ibid.* **186**, 1729 (1969).
[10]R. V. Penney, Phys. Rev. D **14**, 910 (1976).
[11]M. Gürses, Phys. Rev. D **15**, 2731 (1977).
[12]A. Eriş and M. Gürses, J. Math. Phys. **18**, 1303 (1977).
[13]W. J. Sarill, Masters thesis, submitted in partial fullfillment of the requirements for the degree of Master of Arts, Boston University (September 1973) (unpublished).
[14]L. Witten, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1963), Chap. 9.

Universal singular functions in local field theory

Alfred C. T. Wu

Department of Physics, University of Michigan, Ann Arbor, Michigan 48109
(Received 11 April 1977)

The singularity structure of the universal singular functions in local field theory is simply seen by the stationary-phase method applied to the Lorentz group manifold.

I. INTRODUCTION

The universal singular functions in local field theory were defined about twenty years ago.[1,2] One assumes that it is important to know the set of vacuum expectation values of products of local field operators, in the sense that knowing all the moments of a field essentially characterizes the field itself.[3] In the Fourier analysis of these Wightman functions, a covariant decomposition of the momentum-space-support permits a group theoretic extraction of the "angular" part from the volume integral while leaving behind the "radial" (or scalar product) part where the dynamics enters.

The extraction of the universal singular functions is done as follows.[2] Consider the $(n+1)$-fold product as a function of n 4-vectors,

$$F^{(n)}(\zeta_k) \equiv \langle 0 | A_1(x_1) \cdots A_{n+1}(x_{n+1}) | 0 \rangle, \quad \zeta_k = x_{k+1} - x_k, \quad k = 1, \ldots, n, \tag{1}$$

$$= \text{const} \cdot \int \Pi dp_k \exp(-i \sum p_k \cdot \zeta_k) G^{(n)}(p_1, \ldots, p_n) \tag{2}$$

$$= \text{const} \cdot \int \Pi d(\Lambda p_k) \exp[-i \sum (\Lambda p_k) \circ \zeta_k] \times G^{(n)}(\Lambda p_1, \ldots, \Lambda p_n) \tag{3}$$

$$= \text{const} \cdot \int d\nu(\omega) \tilde{G}^{(n)}(\omega) \Delta(M). \tag{4}$$

In going from (2) to (3), we emphasize the invariance under $p \to \Lambda p$; Λ is an element of the homogeneous proper orthochronous Lorentz group. The universal singular function $\Delta(M)$ is simply the properly written covariant expression of the "angular" integral of the Fourier exponential factors, namely

$$\Delta(M) = \int d\mu(\Lambda) \exp[-i \text{tr}(\Lambda M)], \tag{5}$$

where $d\mu(\Lambda)$ denotes the Haar measure on the Lorentz group, and M is a matrix whose elements are formed by the mixed products of p and ζ,

$$M^{\mu}{}_{\nu} = \sum_{k=1}^{n} p_k{}^{\mu} \zeta_{k\nu}. \tag{6}$$

The formalism can be set up for Lorentz spaces of arbitrary $(m+1)$ space–time dimensions where M and Λ are $(m+1) \times (m+1)$ matrices. For the cases $n \leq m$, there is a certain economy in studying $F^{(n)}$ in a $(n+1)$-dimensional space–time, the nontrivial cases are therefore those for which the number of independent vectors equals the rank of the matrices, i.e., $n = m+1$.

Analytic structure and explicit evaluation of $\Delta(M)$ would be items of obvious interest. Let us first summarize what is known on these.

(a) Explicit forms of $\Delta(M)$ are known only for some special cases, namely $n = 1, 2$ for $m = 3$ [1,2] and $n \leq 3$ for $m = 1$.[2] There have been attempts to evaluate the integrals for higher n and m, but the results are by no means complete.[4]

(b) The analytic structure of $\Delta(M)$ of course can be read off in cases where the singularities of the integrand can be suitably displayed. The Lorentz group manifold technique (4) is an alternate (and hopefully more transparent) approach to the analyticity of the vacuum expectation values. The other approach is the so-called generalized singular functions Δ^{+}_{n+1}.[5–9] Of course, the $\Delta(M)$ function and the $\Delta^{+}_{n+1}(z;a)$ function are closely related, but the spirit and the techniques involved are sufficiently different to warrant a separate analysis of $\Delta(M)$. The main result on the $\Delta^{+}_{n+1}(z;a)$ functions is that their singularity domains are given by the following trace manifold[5–9]

$$\sum_{k=1}^{n} (\pm \sqrt{\tau_k}) = \text{real}, \tag{7}$$

where τ_k's are the eigenvalues of the product matrix Za, Z and a are the Gram matrices in the x space and p space respectively,

$$Z_{ij} = \zeta_i \circ \zeta_j, \quad a_{ij} = p_i \circ p_j. \tag{8}$$

The present work is prompted by the desire of finding a simpler way to see the sources of singularities of $\Delta(M)$, namely the manifold

$$\text{Im}\{\text{tr}[\pm (\tilde{M}M)^{1/2}]\} = 0. \tag{9}$$

$\tilde{M}$ is the transpose with the built in metric,

$$(\tilde{M})^{\mu}{}_{\nu} \equiv M_{\nu}{}^{\mu}. \tag{10a}$$

In matrix notation, we have

$$\tilde{M} = GM^{T}G, \tag{10b}$$

where M^{T} denotes the ordinary matrix transpose and G is the matrix of $g_{\mu\nu}$ ($g_{00} = 1$, $g_{ii} = -1$, and zero otherwise). We note in passing that the result (9) stated above for $\Delta(M)$ is consistent with the result (7) on Δ^{+}_{n+1} by virtue of the following identity,

$$\text{tr}(\tilde{M}M) = \text{tr}(Za). \tag{11}$$

The method of stationary phase is used here to establish the following relation,

$$\text{tr}(\Lambda M)_{\substack{\text{stationary} \\ \text{phase}}} = \text{tr}[\pm (\tilde{M}M)^{1/2}]. \tag{12}$$

In this way, the singularity source of Eq. (9) is viewed as that coming from the ceasing of the exponential damping subject to the stationary phase prescription. The present analysis shows that result (9) can be established directly from (5) in a reasonably transparent manner. A simple way of visualizing result (12) is that

the right-hand side of (12) is actually a minimum as a consequence of the extremum principle. This minimum can perhaps be seen by invoking the Schwarz inequality.[10]

The motivation for the stationary phase method is briefly discussed in Sec. II. For the sake of readability, the parametrization of the Lorentz transformation Λ is stated in Sec. III. The two-dimensional case is reviewed in Sec. IV where the result of the stationary phase method is compared with that obtained by the explicit evaluation. In Sec. V, higher rank cases are discussed with the aid of the diagonalization procedure on the matrix M.

II. MOTIVATION FOR THE STATIONARY PHASE METHOD

The method of stationary phase[11] is used here as a device of handling oscillatory exponential terms. Technically, the applicability of the method requires a limiting procedure such as a large parameter which enhances the oscillation. This may be understood by an appeal to the classical path integral by recovering a $\hbar^{-1}$ factor in the exponential and taking the limit $\hbar \to 0$. A similar approach to the Feynman integral was discussed by Nakanishi.[12] Application to the Δ^{+}_{n+1} functions was studied by Fäldt.[9]

III. LORENTZ GROUP MANIFOLD: PARAMETRIZATION OF Λ AND $d\mu(\Lambda)$

For a given space–time dimension $m+1$, the parametrization of the Lorentz transformation Λ and the invariant volume element $d\mu(\Lambda)$ can be worked out by standard procedure. One convenient scheme of parametrizing a general Lorentz transformation Λ is to decompose it into product form of a pure boost sandwiched between appropriate space rotations.[13] As is well known, this is a judicious generalization of the Euler angle decomposition for the rotation in 3-space. Explicitly, we have

(a) $m=1$ (1-space, 1-time)

$$\Lambda = \begin{pmatrix} \cosh\chi & \sinh\chi \\ \sinh\chi & \cosh\chi \end{pmatrix} \equiv \Lambda_{01}, \tag{13a}$$

$$d^{(2)}\mu(\Lambda) = d\chi, \quad -\infty < \chi < \infty; \tag{13b}$$

(b) $m=2$ (2-space, 1-time)

$$\Lambda = R_{12}(\theta)\,\Lambda_{01}(\chi)\,R_{12}(\phi), \tag{14a}$$

$$R_{12} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}, \tag{14b}$$

$$\Lambda_{01}(\chi) = \begin{pmatrix} \cosh\chi & \sinh\chi & 0 \\ \sinh\chi & \cosh\chi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{14c}$$

$$d^{(3)}\mu(\Lambda) = \sinh\chi\, d\chi\, d\theta\, d\phi, \quad 0 \le \theta,\ \phi \le 2\pi, \quad -\infty < \chi < \infty; \tag{14d}$$

(c) $m=3$ (3-space, 1-time)

$$\Lambda = R_{23}(\theta)\,R_{13}(\psi)\,R_{23}(\phi)\,\Lambda_{01}(\chi)\,R_{13}(\alpha)\,R_{23}(\beta), \tag{15a}$$

where $R_{ij}(\theta)$ denotes a rotation in the (x^i, x^j) plane, e.g.,

$$R_{23}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta \\ 0 & 0 & -\sin\theta & \cos\theta \end{pmatrix}, \tag{15b}$$

$\Lambda_{01}(\chi)$ denotes a boost [like (14c)] except here it is 4×4, and

$$d^{(4)}\mu(\Lambda) = \sinh^2\chi\, d\chi\, \sin\alpha\, d\alpha\, \sin\psi\, d\psi\, d\theta\, d\phi\, d\beta, \tag{15c}$$

$$0 < \chi < \infty, \quad 0 \le \alpha, \psi \le \pi, \quad 0 \le \theta, \beta, \phi \le 2\pi.$$

(d) For general $(m+1)$-space–time, we may write

$$\Lambda = R_1\,\Lambda_{01}\,R_2, \tag{16a}$$

where R_1 denotes a rotation in m-space, R_2 would be a rotation in m-space also except for two reasons. One reason is that this would yield $\frac{1}{2}(m-1)(m-2)$ excessive parameters. The other reason is that the Λ_{01} matrix has a $(m-1)\times(m-1)$ identity submatrix in it which permits a tunnelling of a rotation in $(m-1)$-space. These two situations can be reconciled if R_2 is the quotient of R_m/R_{m-1}, which yields $(m-1)$ parameters. [Cf. Eq. (15a).] The volume element may be written as

$$d^{(m+1)}\mu(\Lambda) = \sinh^{m-1}\chi\, d\chi\, d\mu(R_m)\, d\mu(R_m/R_{m-1}). \tag{16b}$$

In the next section, we study the simplest case $m=1$, $n=2$.

IV. A TWO-DIMENSIONAL EXAMPLE REVISITED: EXPLICIT EVALUATION VERSUS STATIONARY PHASE PRESCRIPTION

For the sake of illustration, consider the following example $m=1$, $n=2$, namely, the case of a three-point function in two-dimensional space–time.

A. Explicit evaluation

Write

$$M = \begin{pmatrix} M^0{}_0 & M^0{}_1 \\ M^1{}_0 & M^1{}_1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{17}$$

with Λ given by (13a), we have

$$\begin{aligned} \mathrm{tr}(\Lambda M) &= (\mathrm{tr}M)\cosh\chi + \mathrm{tr}(M\sigma_1)\sinh\chi \\ &= [(\mathrm{tr}M)^2 - (\mathrm{tr}(M\sigma_1))^2]^{1/2}\cosh(\chi+\chi_0) \\ &= [\mathrm{tr}(\tilde{M}M) + 2\det M]^{1/2}\cosh(\chi+\chi_0) \\ &= \mathrm{tr}[\pm(\tilde{M}M)^{1/2}]\cosh(\chi+\chi_0), \end{aligned} \tag{18, 19}$$

where

$$\tanh\chi_0 = \mathrm{tr}(M\sigma_1)/\mathrm{tr}M. \tag{20}$$

The Lorentz transpose $\tilde{M}$ was defined in (10). In the last step leading to (19), use has been made of the trace identity for a rank two matrix, namely

$$\mathrm{tr}A^2 - (\mathrm{tr}A)^2 + 2\det A = 0, \tag{21}$$

with $A = (\tilde{M}M)^{1/2}$.

In this case, evaluation of $\Delta(M)$ readily gives

$$\begin{aligned} \Delta^{(3)}_2(M) &= \int_{-\infty}^{\infty} d\chi\, \{\exp i(\mathrm{tr}[\pm(\tilde{M}M)^{1/2}])\}\cosh(\chi+\chi_0) \\ &= i\pi H^{(1)}_0(\mathrm{tr}[\pm(\tilde{M}M)^{1/2}]). \end{aligned} \tag{22}$$

From (22) it is seen that the source of singularities is at those points where there is no damping for the Hankel function, and that is where the argument for the Hankel function becomes real and positive, which lies on the manifold (9).

B. Stationary phase prescription

We indicate here how the stationary phase method can be made to give the desired result. The zeros of $(\partial/\partial\chi)\,\mathrm{tr}(\Lambda M)$ may be found from (18) or (19). The stationary phase prescription gives

$$\chi = -\chi_0 \tag{23}$$

and

$$\mathrm{tr}(\Lambda M)\big|_{\mathrm{SP}} = \mathrm{tr}[\pm(\tilde{M}M)^{1/2}], \tag{24}$$

verifying (12) for this simple example.

It is obvious that as (n, m) goes up, the control of the integration scheme becomes increasingly difficult. On the other hand, the stationary phase prescription which is easier to handle, hopefully will give the desired result on the singularity structure.

We note in passing that on account of the extreme simplicity of the above example, there is no need to invoke the diagonalization of the M matrix in (18). However, the diagonalization technique will become highly desirable in the treatment of higher rank cases.

IV. HIGHER RANK CASES

A. Two lemmas on diagonalization

To facilitate algebraic manipulations, it will be convenient to utilize two lemmas on matrix diagonalization. The first is due to Hall,[14] and the second one is an obviously parallel statement.

Lemma 1 (*Hall*[14]): A 2×2 matrix M, $M^{\mu}{}_{\nu} = \sum_{k=1}^{2} p_k{}^{\mu}\eta_{k\nu}$, with positive timelike p_k, η_k can be diagonalized by two boosts

$$M' = \Lambda(\chi_1)\, M \Lambda(\chi_2), \tag{25}$$

where Λ is of the form (13a).

Proof: We simply exhibit the angles that will render M' diagonal. They are

$$\tanh 2\chi_1 = -\mathrm{tr}(\tilde{M} i\sigma_2 M)/\mathrm{tr}(\tilde{M}\sigma_3 M), \tag{26a}$$

$$\tanh 2\chi_2 = \mathrm{tr}(M i\sigma_2 \tilde{M})/\mathrm{tr}(M\sigma_3\tilde{M}), \tag{26b}$$

where

$$i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{26c}$$

Remark: The timelike condition on p_k and η_k guarantees that such real $\Lambda(\chi_1)$ and $\Lambda(\chi_2)$ of the form (13a) exist. The usefulness of this lemma lies in the invariance of the volume element and the integrand in (5) under (25). In applying this lemma to the configuration space difference vectors lying in the tube domain,[15] it suffices to take $\xi_k = -i\eta_k$ with η_k positive timelike and to continue analytically from there.

Lemma 2: A 2×2 matrix N can be diagonalized by two rotations [N is understood to consist of spatial indices only]

$$N' = R(\theta_1)\, N R(\theta_2), \tag{27}$$

where $R(\theta)$ is of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Proof: The diagonalization is achieved with the angles θ_1 and θ_2 given by

$$\tan 2\theta_1 = -\mathrm{tr}(N^T\sigma_1 N)/\mathrm{tr}(N^T\sigma_3 N), \tag{28a}$$

$$\tan 2\theta_2 = \mathrm{tr}(N\sigma_1 N^T)/\mathrm{tr}(N\sigma_3 N^T). \tag{28b}$$

Remark: Actually, the spatial part N of the matrix M being a symmetric submatrix can be diagonalized by orthogonal similarity transformations (spatial rotations). Then Lemma 1 can be repeatedly applied in the $(0k)$ planes to remove the nondiagonal elements in the first row and the first column. It should be realized that prior spatial diagonalization is essential because of the noncommutativity of boosts Λ_{0i} and Λ_{0j} in the different directions.

B. Successive rotations and boosts

As stated in Sec. III, a general Lorentz transformation in the $(m+1)$-dimensional space-time may be parametrized as a product of a pure boost [say in the (01) plane] sandwiched between two sets of spatial rotations. The latter may be decomposed into suitable products of plane rotations. We may utilize the rotational freedom to effect first the block diagonalization of M in the spatial indices in accordance with the remark following Lemma 2. Thus

$$\begin{aligned}\mathrm{tr}(\Lambda M) &= \mathrm{tr}(R_1\,\Lambda_{01}\,R_2 M) = \mathrm{tr}(\Lambda_{01}\,R_2 M R_1) \\ &= \mathrm{tr}(\Lambda_{01}\,\hat{R}_2\,\hat{M}\hat{R}_1) = \mathrm{tr}(\hat{R}_1\,\Lambda_{01}\,\hat{R}_2\,\hat{M}) \\ &= \mathrm{tr}(\hat{\Lambda}\hat{M}),\end{aligned} \tag{29}$$

where

$$\hat{M} = R_2'\, M R_1'$$

is diagonal in the spatial indices

$$\hat{R}_2 = R_2 R_2'^{-1}, \quad \hat{R}_1 = R_1'^{-1} R_1,$$
$$\hat{\Lambda} = \hat{R}_1\,\Lambda_{01}\,\hat{R}_2.$$

What remains to be done is to use Hall's lemma repeatedly to render $\hat{M}$ diagoanl in the time components also.

We conclude that without loss of generality the M matrix in the integrand of (5) can be suitably diagonalized.

C. Stationariness of the trace manifold

Suitable parametrization of the Lorentz transformation and the proper diagonalization of the matrix M are thus two technical devices which considerably simplify the evaluation of the trace quantity. We have (the matrix M below is understood to be diagonal)

$$\begin{aligned}\mathrm{tr}(\Lambda M) &= \mathrm{tr}(\Lambda_{01}\,RMS), \quad R, S \text{ are rotations} \\ &= (M^0{}_0 + R^1{}_k\, M^k{}_j\, S^j{}_1)\cosh\chi + \mathrm{tr}(\bar{\bar{R}}\,\bar{\bar{M}}\,\bar{\bar{S}}),\end{aligned}$$

where the double bar quantities denote the corresponding $(m-2)\times(m-2)$ submatrices with indices running from 2 to m. The $\sinh\chi$ term is absent on account of the accomplished diagonalization of M. The demand of the stationary phase in each of the angle variable (Lorentz as well as Euclidean) then results in the following. First the stationariness in the boost angle χ which implies $\chi=0$ gives

$$\mathrm{tr}(\Lambda M)\big|_{\text{stat. boost}} = M^0{}_0 + \mathrm{tr}(\bar{\bar{R}}'\bar{\bar{M}}), \tag{30}$$

where the single bar quantities denote the corresponding $(m-1\ )\times(m-1)$ submatrices with the indices running from 1 to m, and $\bar{R}'=\bar{S}\bar{R}$. For the remaining spatial indices, it is obvious that the orthogonal manifold for the rotational matrices would simply reduce the trace quantity in (30) to assume the value $\mathrm{tr}[\pm(\bar{M}^T\bar{M})^{1/2}]$. Hence

$$\mathrm{tr}(\Lambda M)\big|_{\text{SP}} = \mathrm{tr}[\pm(\tilde{M}M)^{1/2}]$$

which is the result stated in (12).

ACKNOWLEDGMENT

The author is indebted to Professor A.S. Wightman for introducing him to the subject matter.

[1]A.S. Wightman and D. Hall, Phys. Rev. **99**, 674 (1955).
[2]D. Hall, thesis, Princeton University, 1956, unpublished; esp. Chaps. 4—6.
[3]A.S. Wightman, Phys. Rev. **101**, 860 (1956); R. Jost, *The General Theory of Quantum Fields* (Am. Math. Soc., Providence, R.I., 1965).
[4]E.g., A.S. Wightman (private communication, 1962).
[5]G. Källén and A.S. Wightman, K. Dan. Vidensk. Selsk. Mat. Fys. Skr. **1**, No. 6 (1958).
[6]G. Källén and H. Wilhelmsson, K. Dan. Videns. Selsk. Mat. Fys. Skr. **1**, No. 9 (1958).
[7]G. Källén, Lecture Notes (Les Houches, 1960), in *Relation de dispersion et particules élémentaires* (Hermann, Paris, 1960), pp. 229—313.
[8]For completeness, other work by Källén's students on the Δ^+ functions may be cited here: I. Nieminen, Nucl. Phys. **37**, 250 (1962) studied the Δ_4^+ in lower dimensional spaces; G. Fäldt, *ibid.* **59**, 449 (1964) discussed the Δ_{n+1}^+ in n-dimensional space—time.
[9]G. Fäldt, Nucl. Phys. B **59**, 449 (1964).
[10]The author is indebted to Prof. L. Michel for a remark on this.
[11]See, e.g., H. Jeffreys and B.S. Jeffreys, *Methods of Mathematical Physics* (Cambridge U.P., Cambridge, 1956), 3rd ed., esp. p. 506.
[12]N. Nakanishi, Prog. Theor. Phys. **39**, 768 (1968).
[13]See, e.g., A.S. Wightman, Lecture Notes (Les Houches, 1960), volume cited in Ref. 7, esp., p. 167.
[14]D. Hall, Ref. 2, pp. VI. 3
[15]A.S. Wightman, Proc. Indian Math. Soc. **24**, 625 (1960).

KMS condition for stable states of infinite classical systems*

M. Pulvirenti

Istituto Matematico dell'Università, Camerino (MC), Italy

G. Riela

Istituto di Fisica dell'Università di Palermo, Palermo, Italy
(Received 22 June 1976; revised manuscript received 9 February 1977)

Using an abstract algebraic approach, we obtain a new derivation of the KMS condition from a stability property of an infinite system via a classical version of the Tomita–Takesaki theorem.

1. INTRODUCTION

In a recent work[1] the problem of characterizing the equilibrium state of an infinite classical particle system has been studied. Invariant states under the infinite Hamiltonian dynamics have been analyzed and it has been shown that an invariant state "regular enough" must satisfy the equilibrium condition. An analogous result had been previously obtained for quantum systems[2] and later improvements had been given in Refs. 3 and 4, where it was shown that it is possible to relax the order of a cluster assumption; the explicit use of the modular automorphism made clearer in some sense the proof. For classical systems an equivalent of the Tomita–Takesaki theorem was given in Ref. 5, and the main theorem there can be summarized as follows:

Theorem 1.1: Let (K, M, ω) be a probability space, iL be a self-adjoint operator, and $\mathcal{A}$ be a self-adjoint algebra of essentially bounded functions with identity contained in the domain of iL, such that $\mathcal{A}$ is a core for iL and

$$L fg = gLf + fLg \quad \forall f, g \in \mathcal{A} .$$

Then there exists a one-parameter (mod0) family of measure preserving transformations T_t, $t \in \mathrm{R}$ of K such that

$$f(T_t x) = (e^{Lt} f)(x) , \quad x \in K , \quad f \in L_2(K, \omega) , \quad t \in \mathrm{R} .$$

In our paper we mimic the proofs in Refs. 3 and 4 in order to obtain an analogous result for classical systems using Theorem 1.1 instead of the Tomita–Takesaki theorem; unfortunately we are not able to relax the order of clustering assumption. We need the same condition of threefold mixing as in Ref. 1; this is not too bad however, since there is not any known mixing dynamical system which is not also n-mixing.

Further remarks on the difference between our paper and Ref. 1 will be found in Sec. 2.

2. NOTATION, DEFINITIONS, AND RESULTS

Let (K, Σ, ω) be a probability Lebesgue space and let $\alpha_t : K \to K$, $t \in \mathrm{R}$ be a representation of R as one-to-one ω-preserving transformation of K into itself. We require that

$$t \to \omega(A \cap \alpha_t B)$$

be measurable for all A and B in Σ.

We denote by U_t the action of the group transformation on the functions

$$(U_t f)(x) = f(\alpha_t x) \quad x \in K , \quad f \in L_2(\omega) .$$

The measurability requirement implies the weak measurability of U_t. Therefore, there exists a self-adjoint operator H on the separable Hilbert space $L_2(\omega)$ such that $U_t = e^{iHt}$ (see Ref. 6).

We suppose that a $L_2(\omega)$-dense, self-adjoint algebra with identity, $\mathcal{A} \subset L_\infty(\omega)$ exists with the following properties:

(1) $\mathcal{A} \subset D(H)$,

(2) $U_t \mathcal{A} \subset \mathcal{A} \quad \forall t \in \mathrm{R}$.

Remark: It follows that $\mathcal{A}$ is a core for H.

A bilinear map (Poisson brackets)

$$\{\cdot, \cdot\} : \mathcal{A} \times \mathcal{A} \to L_2(\omega)$$

with the following properties is supposed to exist:

(i) $\{f, g\} = -\{g, f\}$,

(ii) $\{\overline{f}, \overline{g}\} = \{\overline{f}, \overline{g}\}$,

(iii) $\{f, gh\} = \{f, g\}h + \{f, h\}g$,

(iv) $\{U_t f, U_t g\} = U_t\{f, g\}$.

Definition 2.1: A system $Y \equiv (K, \alpha, \omega, \mathcal{A}, \{\cdot\})$ with the above properties will be called a Poisson system.

Definition 2.2: For all $f, g \in \mathcal{A}$ two functions are defined as follows:

$$F_{f,g}(t) = \omega(f U_t g) - \omega(f)\omega(g) ,$$

$$G_{f,g}(t) = \omega(\{f, U_t g\}) .$$

Definition 2.3: A Poisson system is said to satisfy the KMS condition at some temperature β^{-1} if the following holds:

$$G_{f,g}(t) = -\beta \frac{d}{dt} F_{f,g}(t) .$$

Definition 2.4: A dynamical flow (K, α, ω) is said to be k-fold mixing if

*Partially supported from CNR under contract no. 75/0042202.

$$\omega(U_{t_1}f_1\cdots U_{t_K}f_K)\xrightarrow[\min_{ij}|t_i-t_j|\to\infty]{}\omega(f_1)\cdots\omega(f_K)$$

$$\forall f_1\cdots f_K\in L_\infty(\omega)\,.$$

A twofold mixing system is said to be simply mixing.

Definition 2.5: A Poisson system Y is said to have correlations L_1 decreasing in time if

$$F_{f,g}(t)\in L_1(\mathbb{R})\quad\forall f,g\in\mathcal{A}\,.$$

Definition 2.6: A Poisson system Y is said to be p-dispersive if

$$\|\{f,U_tg\}\|_p\in L_1(\mathbb{R})\quad\forall f,g\in\mathcal{A}\,.$$

Definition 2.7: A Poisson system Y is said to be stable if:

(i) $\forall f\in\mathcal{A}$, $f=\bar f$, $\exists\,\delta\in\mathbb{R}:\forall\lambda\in(0,\delta)$ there exists a probability measure $\omega^{\lambda f}$ formally invariant for the perturbed dynamics generated by $\mathcal{L}+\lambda\{f,\cdot\}$, where $\mathcal{L}=iH$ is the Liouville operator, i.e.,

$$\omega^{\lambda f}(\mathcal{L}g+\lambda\{f,g\})=0\quad\forall g\in Q\,.$$

(ii) $\omega^{\lambda f}\ll\omega$; we denote by $\rho_{\lambda f}$ the $L_1(\omega)$ density giving $\omega^{\lambda f}$.

(iii) $\rho_{\lambda f}\in L_\infty(\omega)$ and $\lim_{\lambda\to 0}\omega^{\lambda f}(g)=\omega(g)\ \ \forall g\in L_1(\omega)$.

Our main result is the following:

Theorem 2.1: Let Y be a Poisson system which is stable, 1-dispersive, threefold mixing and with correlations L_1 decreasing in time. If the functions $\omega(\{U_tf,U_sgU_uh\})$ are polynomially bounded in t,s,u for all $f,g,h\in\mathcal{A}$, then there exists a temperature β^{-1} for which Y is KMS.

We shall prove the above theorem in the next section. It may be useful now to say a few words about the definitions given above and to compare our result with the similar one in Ref. 1. In our approach we deal with the abstract structure of the Poisson system without any reference to concrete physical systems. Nevertheless we have in mind, as in Ref. 1, an infinite classical system, where K is the phase space, α is the almost everywhere defined time evolution, and ω is an invariant measure. $\mathcal{A}$ is an algebra of nice observables on which the formal Poisson brackets make sense. The choice of $\mathcal{A}$ is delicate and not unique. If $\mathcal{A}$ is large one can show that the KMS condition is equivalent to the equilibrium condition.[7,8] However in our present context $\mathcal{A}$ may not be chosen too large since it would then be difficult to prove the dispersivity and clustering properties. Actually such difficulty can be overcome only in the free gas case. For a deeper discussion of this point see Ref. 1. Finally we observe that we use the weaker 1-dispersivity rather than the 2-dispersivity as in Ref. 1; the price we pay is the strengthening of the stability property, requiring $\rho_{\lambda f}\in L_\infty(\omega)$ rather than $\rho_{\lambda f}\in L_2(\omega)$. The polynomial boundedness in Theorem 2.1 and the L_1 decrease in time of the correlation functions lead to a KMS condition valid for all t rather than for almost all t as in Ref. 1 where such a property is not required. On the other hand, it would be sufficient there to require the continuity of the G functions to recover the same stronger result, which is needed in order to show the equivalence between the KMS and the equilibrium conditions.

We conclude this section deriving the only consequence of the stability we will use in the sequel. For all $f,g\in\mathcal{A}$, $f=\bar f$ the invariance property gives

$$\frac{d}{dt}\omega^{\lambda f}(U_tg)=-\lambda\omega^{\lambda f}(\{f,U_tg\})\,.$$

Then

$$\omega(\rho_{\lambda f}U_tg)-\omega(\rho_{\lambda f}U_\tau g)=-\lambda\int_{-\infty}^{+\infty}\chi_\tau\omega(\rho_{\lambda f}\{f,U_tg\})dt\,,$$

where $\chi_\tau\equiv\chi_{[-\tau,\tau]}$ is the characteristic function of the set $[-\tau,\tau]$. Taking the limit $\tau\to\infty$ and using the dominated convergence theorem and 1-dispersivity we obtain

$$\int_{-\infty}^{\infty}\omega(\rho_{\lambda f}\{f,U_tg\})dt=0\,.$$

When $\lambda\to 0$ the sequence of L_1 functions $\omega(\rho_{\lambda f}\{f,U_tg\})$ converges pointwise to the measurable function $\omega(\{f,U_tg\})$; the uniform boundedness theorem gives the following estimate:

$$|\omega(\rho_{\lambda f}\{f,U_tg\})|\le c\|\{f,U_tg\}\|_1\in L_1(\mathbb{R})\,,$$

and so

$$\int_{-\infty}^{\infty}G_{f,g}(t)dt=0\quad\forall f,g\in\mathcal{A}\,,\ f=\bar f\,.\tag{2.1}$$

Finally one gets the same equality for complex f using the self-adjointness of $\mathcal{A}$ and the linearity of the Poisson brackets.

3. PROOFS

In the proof of Theorem 2.1 we need two lemmas which we state separately and will prove later. Lemma 3.1 below is essentially contained in Ref. 1 and Lemma 3.2 is based on an Abelian version of arguments given in Ref. 4. We give here the proofs for sake of completeness.

Lemma 3.1: Let Y be a Poisson system verifying the hypotheses of Theorem 2.1. Then the following identity holds:

$$\hat G_{g_1g_2}(\lambda)\hat F_{f_1f_2}(\lambda)=\hat G_{f_1f_2}(\lambda)\hat F_{g_1g_2}(\lambda)\,.$$

$\hat G$ and $\hat F$ are the Fourier transforms of G and F, respectively.

Lemma 3.2: Let (K,ω) be a measure space and H be a self-adjoint operator on $L_2(\omega)$ such that e^{iHt} gives a mixing automorphism of $L_\infty(\omega)$ for all $t\in\mathbb{R}$. Let ϕ be a real measurable function such that $e^{it\phi(H)t}$ gives an automorphism of $L_\infty(\omega)$ for all $t\in\mathbb{R}$. Then there exists $\beta\in\mathbb{R}$ such that $\phi(x)=\beta x$.

Proof of Theorem 2.1: Let $D\subset R$ be the set of all those λ for which there exists some $f,g\in\mathcal{A}$ with $\hat F_{f,g}(\lambda)\neq 0$. D is open because all the $\hat F$'s are continuous. As a consequence of Lemma 3.1, for all λ in D there exist $f_1,f_2\in\mathcal{A}$ such that

$$\hat G_{g_1g_2}(\lambda)=[\hat G_{f_1f_2}(\lambda)\backslash\hat F_{f_1f_2}(\lambda)]\hat F_{g_1g_2}(\lambda)\quad\forall g_1g_2\in\mathcal{A}\,.$$

The function $\lambda\to\hat G_{f_1f_2}\backslash F_{f_1f_2}\equiv\psi$ does not depend clearly on the choice of f_1,f_2; it is continuous because of the continuity of the $\hat F_{f,g}$'s and $\hat G_{f,g}$'s, $f,g\in\mathcal{A}$, and purely imaginary because $G_{fg}(t)=-G_{gf}(-t)$ due to the symmetry pro-

perties of the Poisson brackets. If $0 \in D$ it follows from Eq. (2.1) that $\psi(0)=0$. We extend ψ (denoting it also by ψ) on the whole real line by putting $\psi(\lambda)=0$ if $\lambda \notin D$. Let us put $\phi=-i\psi$. In order to investigate the self-adjointness property of $\phi(H)$ we introduce two sets of functions, $\mathcal{D}_D, \hat{\mathcal{D}}_D$. $\mathcal{D}_D$ is the set of functions in $\mathcal{L}(\mathrm{R})$ with compact support contained in D; $\hat{\mathcal{D}}_D$ is the set of functions whose Fourier transforms are in $\mathcal{D}_D$. We define

$$\mathcal{D}_\infty \equiv \{ f_\varphi \,|\, f_\varphi \equiv \int \varphi(t) U_t f dt\ ,\ \varphi \in \hat{\mathcal{D}}_D\ ,\ f \in \mathcal{A} \}.$$

$\mathcal{D}_\infty$ is dense in $L_2(\omega)$. In fact, take a sequence $\{\varphi_n\}_{n=1}^\infty$ such that $\hat{\varphi}_n \nearrow \chi_D$, $\hat{\varphi}_n \in \hat{\mathcal{D}}_D$, where χ_D is the characteristic function of D. Then for all $f \in \mathcal{A}$,

$$\lim_{n\to\infty} f\varphi_n = \lim_{n\to\infty} \hat{\varphi}_n(H) f = \chi_D(H) f$$

and, if $\omega(f)=0$

$$\|\chi_D(H)f - f\|^2 \leq \int |\chi_D(\lambda) - 1|\, \|dE_\lambda\|^2 \leq \int_{R\setminus D} F_{f,f}(\lambda) d\lambda = 0\,.$$

The density of $\mathcal{D}_\infty$ follows from the density of Q. $\mathcal{D}_\infty$ is a set of analytic vectors for $\phi(H)$. This follows from the estimate

$$\int |\phi(\lambda)|^{2n} (f\varphi, dE_\lambda f\varphi) = \int |\phi|^{2n} |\hat{\varphi}|^2 (f, dE_\lambda f) \leq \|f\|_2^2 \|\varphi\|_1^2 \sup_{\lambda \in \operatorname{supp}\hat{\varphi}} |\phi(\lambda)|^{2n}.$$

$\phi(H)$ is therefore essentially self-adjoint on $\mathcal{D}_\infty$.

We show next that, if $\mathcal{A}_\infty$ is the algebra generated by $\mathcal{D}_\infty$, then $\phi(H)$ is defined on $\mathcal{A}_\infty$ and acts on it as a derivation. If $f, g \in \mathcal{A}$

$$\omega(\{\bar{f}, g\}) = i \int \phi(\lambda)(f, dE_\lambda g)\,. \tag{3.1}$$

Then, if $\theta \in \hat{\mathcal{D}}_D$

$$(g, \phi(H) f_\theta) = \int \hat{\theta}(\lambda)\phi(\lambda)(g, dE_\lambda f) = i \int \hat{\theta}(\lambda) \hat{G}_{\bar{g},f}(\lambda) d\lambda = i \int \theta(t) G_{\bar{g},f}(t) dt = \int dt\, \theta(t) \int (g, dE_\lambda U_t f)\phi(\lambda)\,. \tag{3.2}$$

Let us now put $\tilde{f} = f_\theta$, $\tilde{g} = g_\varphi$, $\tilde{h} = h_\xi$, $\theta, \varphi, \xi \in \hat{\mathcal{D}}_D$. Then using (3.2),

$$\begin{aligned} i(\tilde{g}\tilde{h}, \phi(H)\tilde{f}) &= i \int ds\, \varphi(s) \int du\, \xi(u) (U_s g U_u h, \phi(H) f\theta) \\ &= i \int ds\, \varphi(s) \int du\, \xi(u) \int dt\, \theta(t) \\ &\quad \times \int (U_s g U_u h, dE_\lambda U_t f)\phi(\lambda) \\ &= \int ds\, \varphi(s) \int du\, \xi(u) \int dt\, \theta(t) \omega(\{U_s \bar{g} U_u \bar{h}, U_t f\}) \\ &= \int ds\, \varphi(s) \int du\, \xi(u) \int dt\, \theta(t) [\omega(\{u_s \bar{g}, u_u \bar{h} U_t f\}) \\ &\quad + \omega(\{U_u \bar{h}, U_s \bar{g} U_t f\})]\,. \end{aligned}$$

Because of the polynomial boundedness of the ω's we may change the order of integration and obtain, using (3.1),

$$\begin{aligned} i(\tilde{g}\tilde{h}, \phi(H)\tilde{f}) &= i \int dt\, \theta(t) \int ds\, \varphi(s) \int du\, \xi(u) \\ &\quad \times \int \phi(\lambda)(dE(\lambda) U_u h, U_s \bar{g} U_t f) \\ &\quad + i \int dt\, \theta(t) \int du\, \xi(u) \int ds\, \varphi(s) \\ &\quad \times \int \phi(\lambda)(dE(\lambda) U_s g, U_u \bar{h} U_t f)\,. \end{aligned}$$

Finally, using (3.2),

$$i(\tilde{g}\tilde{h}, \phi(H)\tilde{f}) = i(\tilde{f}, \tilde{g}\phi(H)\tilde{h}) + i(\tilde{f}, \tilde{h}\phi(H)\tilde{g})\,.$$

We have shown that $\phi(H)$ is an essentially self-adjoint derivation on the algebra of bounded functions $\mathcal{A}_\infty$ and therefore, by Theorem 1.1 $e^{i\phi(H)t}$ implements a group of automorphisms of $L_\infty(\omega)$. From Lemma 3.2 and Eq. (3.2) the thesis of the theorem follows, since there exists a real β such that

$$\omega(\{\bar{f}, g\}) = -\beta(f, Lg) \quad \forall f, g \in \mathcal{A}\,.$$

Proof of Lemma 3.2: We follow closely the proof of Lemma 1 in Ref. 1. Let us define

$$\tilde{G}(t,u,s) = \omega(\{g_1 U_u f_2, U_t(g, U_{u-s} f_2)\})\,.$$

From (2.1)

$$\begin{aligned} \int \tilde{G}(t,u,s)dt &= \int dt\, \omega(U_u\{f_1, U_{t-s} f_2\} g_1 U_t g_2) \\ &\quad + \int dt\, \omega(\{g_1, U_t g_2\} U_u(f_1 U_{t-s} f_2)) \\ &\quad + \int dt\, \omega(\{U_u f_1, U_t g_2\} g_1 U_{t+u-s} f_2) \\ &\quad + \int dt\, \omega(\{g_1, U_{t+u-s} f_2\}\, U_u f_1\, U_t g_2) = 0\,. \end{aligned}$$

Let us take the limit $u \to \infty$; the integrands of the first two terms are bounded respectively by

$$|g_1|_\infty |g_2|_\infty \omega(|\{f_1, U_{t-s} f_2\}|) \in L_1(\mathrm{R})\,,$$
$$|f_1|_\infty |f_2|_\infty \omega(|\{g_1, U_t g_2\}|) \in L_1(\mathrm{R})\,.$$

The mixing property, the invariance of the state, the bounded convergence theorem and (2.1), when $u \to \infty$, make the first two terms equal to

$$-\int G_{f_1 f_2}(s-t) F_{g_1 g_2}(t) dt + \int F_{f_1 f_2}(s-t) G_{s_1 s_2}(t) dt\,.$$

The last two terms can be written as

$$\int dt\, \omega(\{f_1, U_t g_2\} u_{-u} g_1 U_{t+u-s} f_2)\,,$$
$$\int dt\, \omega(\{g_1, U_{t-s} f_2\} u_u f, U_{t-u} g_2)$$

The integrands are bounded respectively by

$$|g_1|_\infty |f_2|_\infty \omega(|\{f_1, U_t g_2\}|) \in L_1(\mathrm{R})\,,$$
$$|f_1|_\infty |g_2|_\infty \omega(|\{g_1, U_{t-s} f_2\}|) \in L_1(\mathrm{R})\,,$$

so the last two terms converge to 0 because their integrands converge pointwise (threefold mixing), respectively to

$$\omega(\{f_1, U_t g_2\})\omega(g_1)\omega(f_2) \quad \text{and} \quad \omega(\{g_1, U_{t-s} f_2\})\omega(f_1)\omega(g_2)\,.$$

Finally we have

$$(F_{g_1 g_2} * G_{f_1 f_2})(t) = (F_{f_1 f_2} * G_{g_1 g_2})(t) \quad \forall f, g \in \mathcal{A}$$

and this completes the proof.

Proof of Lemma 3.2: Let us define the two-parameter group $V(t,s) = e^{-iHt} e^{-i\phi(H)s}$. The closed set

$$S_p V \equiv \{\lambda \in R^2 \,|\, \hat{\varphi}(\lambda) = 0 \ \ \forall \varphi \in L_1 \text{ such that } V\varphi = 0\}\,,$$

where $V_\varphi \equiv \int \varphi(\tau) V_\tau d\tau$ is additive, as can be seen adapting to the Abelian case Theorem 4.1 in Ref. 4. Then,

$$V_\varphi = \int \hat{\varphi}(\lambda, \phi(\lambda)) dE_\lambda ,$$

where E_λ is the spectral family of H. Hence,

$$S_p V = \{(\lambda, \phi(\lambda)) \,|\, \lambda \in S_p H\} .$$

The additivity of $S_p V$ implies $\phi(\lambda_1 + \lambda_2) = \phi(\lambda_1) + \phi(\lambda_2)$. The lemma is then proved.

ACKNOWLEDGMENTS

We thank G. Gallavotti who suggested the problem to us and whose continuous criticism has been valuable for the completion of this work. One of us (M.P.) is also indebted to D. Kastler for useful explanations.

[1]M. Aizenman, G. Gallavotti, S. Goldstein, and J. Lebowitz, Commun. Math. Phys. **48**, 1–14 (1976).
[2]R. Haag, D. Kastler, E. Trych-Pohlmeyer, Commun. Math. Phys. **38**, 173 (1974).
[3]O. Bratteli and D. Kastler, Commun. Math. Phys. **46**, 37 (1976).
[4]D. Kastler, "Equilibrium states of matter and operator algebras" Proc. of the Rome Conf. on C Algebras, 1st. Naz. Alta Matematica, 1975.
[5]G. Gallavoti and M. Pulvirenti, Commun. Math. Phys. **46**, 1 (1976).
[6]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1973), Vol. 1.
[7]G. Gallavotti and E. Verboven, Nuovo Cimento B **28**, 274 (1975).
[8]B. M. Gurevich and J. M. Sukhov, Commun. Math. Phys. **49**, 63 (1976).

Classical particles with spin. I. The WKBJ approximation

John Stachel[a)]

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540

Jerzy Plebański[b)]

Centro de Investigación y de Estudios Avanzados, México, Mexico
(Received 18 May 1977)

This is the first of a series of papers developing the classical theory of a spinning particle. The equations of motion will be derived from a Lagrangian, and solutions for the classical trajectory and spin precession in external fields will be given. In this paper an abstract spin vector is introduced to characterize the spin of a classical particle. Lagrangians for the classical trajectories and for the motion of the abstract spin vector are derived from corresponding quantum-mechanical Lagrangians by the WKBJ approximation method for nonrelativistic and relativistic particles. The equations of motion for the trajectory and the abstract spin vector following from the extremalization of these Lagrangians are given. The equations of motion for the precession in an external electromagnetic field of the spin vector (or tensor) in space–time is derived from the equations of motion for the abstract spin vector. In the relativistic case, they are equivalent to the Bargmann–Michel–Telegdi equations [Phys. Rev. Lett. **2**, 435 (1959)]. The relationship between the ensemble and single-particle points of view is also elucidated.

I. INTRODUCTION

In this series of papers we discuss the theory of classical particles with spin. That subject has a long history, which we will not review here, although we will give some references to the most relevant literature in the course of our work.[1] What distinguishes our approach from most of the existing work in the field is that we do not consider the spin tensor (or vector) as a primary quantity in defining the theory, but rather as derived from some more fundamental representation of the rotation or Lorentz group, depending on whether it is a nonrelativistic or special-relativistic particle that is being treated. This, of course, is the way that spin enters into quantum mechanics, where the wavefunction corresponding to a particle with spin is taken to be a multicomponent entity, with the approriate transformation properties under the relevant groups. The point is that there is nothing fundamentally quantum mechanical about such a concept of a particle, and the same ideas may be applied at the classical level.

So our basic concepts are a trajectory in space—time, to be picked out by some equation of motion, and a spinor, vector, tensor—what have you—attached to each point of that trajectory with appropriate transformation properties under the rotation group (for nonrelativistic theories) or the homogeneous Lorentz group (for special-relativistic theories) which also obeys some equation of motion along the trajectory. We shall refer to this entity as the *abstract spin—vector* since it is a vector in some abstract space on which a representation of the appropriate group acts. Then, the usual spin tensor (or vector) is derived from this basic spin representation by some operation on it which produces an antisymmetric tensor (or vector) in the Galilei—Newtonian or Minkowski space of the trajectory.

In the earlier literature, we find this point of view in Schiller,[2,3] who starts from such a classical theory of the electron; and implicitly in Pauli,[4] Rubinow and Keller,[5] who treat the classical motion by means of a WKB expansion of the Dirac equation. Rafanelli and Schiller[6] show that the classical equations of motion for the electron may be derived from the WKB approximation to either the Dirac equation or the squared Dirac equation.

Our approach is also characterized by the assumption that the trajectory is not influenced by the spin characteristics of the particle. Thus, we eschew all those theories of the spinning particle in which momentum need not be parallel to velocity, with their accompanying classical Zitterbewegungen. Such theories have their interest, and indeed may also be motivated by certain types of approximation to quantum mechanical equations of motion, just as we shall motivate our approach in this paper, by a discussion of the WKBJ or eikonal type of approximation. However, they are not the type of theory that we wish to develop here, in which the trajectory of the particle is not affected by its spin.[7]

We could, at this point, just begin to consider such classical systems, for example, by writing down a Lagrangian giving rise to the desired equations of motion. However, we shall motivate our approach by showing that the equations that we shall consider can be looked upon as the WKBJ or quasiclassical limit of well-known quantum mechanical equations.

The WKBJ approximation[4] consists in making an asymptotic expansion of the wavefunction in powers of $\hbar$,

$$\psi = \exp(iS/\hbar) = \exp(i/\hbar)[S_0 + (\hbar/i)S_1 + (\hbar/i)^2 S_2 + \cdots], \tag{1.1}$$

with the assumption that S_0 is a real scalar function of the coordinates and time, while S_i $(i = 1, 2, 3, \cdots)$ are abstract spin vectors like ψ itself. This is equivalent

[a)]On leave from the Department of Physics, Boston University. This material is based upon research supported by the National Science Foundation under Grant No. INT76-05769.
[b)]On leave from the University of Warsaw, Warsaw, Poland. Partially supported by C.O.N.A.C.Y.T. de México.

to assuming the expansion to be of the form

$$\psi = [R_0 + (\hbar/i)R_1 + \cdots] \exp(iS_0/\hbar) \tag{1.2}$$

as Pauli[4] points out, where the R_i $(i = 0, 1, 2, \cdots)$ are abstract spin vectors. This expansion is then usually inserted into the wave equation for ψ. However, we shall insert it into the variational principle for the wave equation, thus getting an expansion of the variational principle to various orders in $\hbar$. We shall refer to the terms of this expansion as zeroth order, first order, etc., meaning order in powers of $\hbar$, and not the order of the highest derivatives in the variational integrand. Variation of the nth order term will then yield the corresponding order in the expansion of the wave equation. Of course, since the zeroth order in the WKBJ expansion of a wavefunction corresponds to a classical ensemble,[8] we must expect to get the ensemble form of our classical equations, involving the action function S, Hamilton–Jacobi equations, etc., rather than getting the trajectories directly. But, of course, since any solution to the Hamilton–Jacobi equation corresponds to an ensemble of mechanical trajectories which can be derived from it, this constitutes no problem.

In this paper, we shall first discuss the spinless particle, nonrelativistic and relativistic, in an external electromagnetic field in order to demonstrate some features of our approach, which works from the action principle directly, in the simplest possible context. Then we shall discuss the Pauli and Dirac equations for nonrelativistic and relativistic particles of spin $\frac{1}{2}$, both interacting with external electromagnetic fields. We could easily extend our results formally to particles of any spin interacting with the electromagnetic field. However, in view of the well known difficulties with the external field problem for higher spin particles,[9] it is doubtful if these results would have more than formal significance. It is interesting, of course, that these difficulties do not manifest themselves at the level of the quasiclassical approximation. Of course, there is no difficulty with extending the results of this paper to free particles of arbitrary spin, but the results then are rather trivial: The abstract spin–vector is just parallel transported along the free particle trajectory. Finally, we shall consider the transition from the ensemble to the single-particle Lagrangian.

In the next paper we shall generalize the particle Lagrangian for the relativistic particle with spin $\frac{1}{2}$ interacting with an external electromagnetic field, developed here, to the most general possible relativistically invariant interaction, and discuss the solution of the resulting equations of motion.

II. NONRELATIVISTIC SPINLESS PARTICLE

We start from the well-known variational principle for the Schrödinger equation,

$$\delta \int \psi^* \left(i\hbar \frac{\partial}{\partial t} - \hat{H} \right) \psi \, d^3x \, dt = 0, \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian for the particle

$$\hat{H} = \frac{1}{2m} \left[\nabla \frac{\hbar}{i} - \frac{e}{c} \mathrm{A} \right]^2 + V, \tag{2.1'}$$

where V includes $e\phi$, the electric potential energy, as well as any other scalar potentials, and A is the magnetic potential. We now insert the WKBJ ansatz[10] $\psi = R_0 \exp(iS/\hbar)$ directly into the variational principle, giving

$$\delta \int R_0^* \left[-\frac{\partial S}{\partial t} - \frac{1}{2m} \left(\nabla S - \frac{e}{c} \mathrm{A} \right)^2 - V \right] R_0 \, d^3x \, dt = 0, \tag{2.2}$$

where we have omitted all terms of first or higher order in $\hbar$.

Variation of (2.2) with respect to R_0^*, yields

$$\left[\frac{\partial S}{\partial t} + \frac{1}{2m} \left(\nabla S - \frac{e}{c} \mathrm{A} \right)^2 + V \right] R_0 = 0, \tag{2.3}$$

and we see that if R_0 does not vanish, the nonrelativistic Hamilton–Jacobi equation must hold for S. Variation of (2.2) with respect to S gives

$$\frac{\partial |R_0|^2}{\partial t} + \nabla \cdot \left[\frac{|R_0|^2}{m} \left(\nabla S - \frac{e}{c} \mathrm{A} \right) \right] = 0. \tag{2.4}$$

The Hamilton equations for the trajectories corresponding to solutions of the Hamilton–Jacobi equation (2.3) show that

$$m\mathbf{v} = \nabla S - \frac{e}{c} \mathrm{A}, \tag{2.5}$$

so that (2.4) is just the equation of continuity for $|R_0|^2$, the density of trajectories in configuration space.

Thus, we have derived the equations of motion for an ensemble of trajectories, from the zeroth-order WKBJ approximation to the Lagrangian for the Schrödinger equation. The density of trajectories $|R_0|^2$ is also determined by the equation of continuity, which is easily converted into an equation for the ordinary derivative of $|R_0|^2$ along a mechanical trajectory determined by S,

$$\frac{d|R_0|^2}{dt} + \frac{|R_0|^2}{m} \nabla \cdot \left[\nabla S - \frac{e}{c} \mathrm{A} \right] = 0. \tag{2.6}$$

But this does not enable us to determine R_0 itself, which contains a phase factor, undetermined so far. As we shall see, this phase factor can be determined from the first-order approximation to the Lagrangian. This is the reflection, at the spinless particle level, of the same feature we shall find for particles with spin: To determine the trajectories, we only need S, which is fixed by the zeroth-order approximation to the Lagrangian as a solution to the Hamilton–Jacobi equation. However, to fix the motion of the abstract spin vector (in this case just the phase of R_0), the next approximation must be calculated, even though the resulting equation of motion for the spin is independent of $\hbar$, and indeed of any other quantities characterizing the next approximation.

We shall not bother to give the derivation of the equation of motion for R_0 from the first-order approximation, since it can be deduced immediately from our discussion for the Pauli equation in Sec. III, by setting the terms with $\sigma = 0$ in Eq. (4.7). We merely note that the result is

$$\frac{dR_0}{dt} = -\frac{1}{2m} R_0 \nabla \cdot \left(\nabla S - \frac{e}{c} \mathrm{A} \right). \tag{2.7}$$

Comparing (2.6) and (2.7), it follows that the phase of R_0 is constant along a mechanical trajectory,

$$\frac{d}{dt}\frac{R_0}{|R_0|}=0. \tag{2.8}$$

Summarizing our results, we see that the zeroth-order WKBJ approximation yields the nonrelativistic Hamilton–Jacobi equation; a class of mechanical trajectories can be derived from a solution to this in the well-known way. It also gives the continuity equation, which enables us to determine the evolution of the magnitude of R_0, along a mechanical trajectory, given a solution to the Hamilton–Jacobi equation. Thus, the magnitude of R_0 is determined for an ensemble of trajectories; it corresponds to the density in configuration space of the particles in the ensemble, but it is not a quantity which has any meaning for an individual trajectory independently of an ensemble. So it is not surprising that its evolution cannot be determined independently of S. On the other hand, the phase of R_0 along a mechanical trajectory is obtained from the first-order WKBJ approximation. Its evolution is meaningful for an individual trajectory, quite apart from any ensemble to which the latter may belong. In our case, this equation is trivial—the phase stays constant. But this feature of the results will generalize to other cases with spin: The magnitude of the abstract spin–vector will be meaningful only for the ensemble point of view, while the evolution of the "unit" abstract spin vector will be determined by an equation of motion along a single trajectory.

III. RELATIVISTIC SPINLESS PARTICLE

We start from the Lagrangian for the Klein–Gordon equation with external electromagnetic field[11]:

$$\int\left[\left(\frac{-h}{i}\nabla-\frac{e}{c}\mathrm{A}\right)\psi^*\cdot\left(\frac{\hbar}{i}\nabla-\frac{e}{c}\mathrm{A}\right)\psi-m^2c^2\psi^*\psi\right]d^4x. \tag{3.1}$$

Inserting the WKBJ ansatz $\psi=\phi\exp(iS/\hbar)$ into (3.1) and again keeping only terms independent of $\hbar$, we get

$$\int\phi^*\left[\left(\nabla S-\frac{e}{c}\mathrm{A}\right)\cdot\left(\nabla S-\frac{e}{c}\mathrm{A}\right)-m^2c^2\right]\phi d^4x. \tag{3.2}$$

Variation with respect to ϕ^* gives

$$\left[\left(\nabla S-\frac{e}{c}\mathrm{A}\right)^2-m^2c^2\right]\phi=0; \tag{3.3}$$

and again, if ϕ does not vanish, the relativistic Hamilton–Jacobi equation must hold (variation of ϕ again leads to the conjugate equation). Variation with respect to S leads to

$$\nabla\cdot\left[(\phi^*\phi)\left(\nabla S-\frac{e}{c}\mathrm{A}\right)\right]=0. \tag{3.4}$$

Since the Hamilton–Jacobi equation (3.3) implies Hamilton's equations of motion for the trajectories, we again see that (2.5) holds, now as a 4-vector equation; and thus (3.4) is a continuity equation for $\phi^*\phi$, the density of trajectories. Thus, the zeroth-order WKBJ approximation again determines the relativistic Hamilton–Jacobi equation, a solution to which yields an ensemble of mechanical trajectories; as well as the equation of continuity, which determines the evolution of the magnitude of ϕ along each trajectory, given a solution S. We omit the details of the proof that the first-order WKBJ approximation determines the evolution of ϕ along a trajectory given S; from which it follows that the phase along each trajectory is constant.

IV. NONRELATIVISTIC PARTICLE OF SPIN $\frac{1}{2}$ (PAULI EQUATION)

We start from a variational principle for the Pauli equation,

$$\delta\quad\psi^*\left(\pm i\hbar\frac{\partial}{\partial t}-\hat{H}\right)\psi\,d^3x\,dt=0, \tag{4.1}$$

where

$$\hat{H}=\frac{1}{2m}\left[\sigma\cdot\left(\frac{\hbar}{i}\nabla-\frac{e}{c}\mathrm{A}\right)\right]\left[\sigma\cdot\left(\frac{\hbar}{i}\nabla-\frac{e}{c}\mathrm{A}\right)\right]+V. \tag{4.2}$$

Here, ψ is a two-component spinor, ψ^* its Hermitian adjoint, σ the Pauli matrices, A the electromagnetic vector potential, and V a scalar potential energy which includes $e\phi$, where ϕ is the scalar electrostatic potential. We again insert the WKBJ ansatz, but this time the coefficient of $\exp(iS/\hbar)$ is a two-component spinor. Since we shall have to consider both the zero- and first-order approximations, we include two terms in our ansatz:

$$\psi=(D_0+{}^*y_iD_1)\exp(iS/\hbar),\qquad \psi^*=\left(D_0^*-\frac{\hbar}{i}D_1^*\right)\exp(-iS/\hbar), \tag{4.3}$$

where D_0 and D_1 are two-component spinor fields. Inserting this into (4.1) we expand the Lagrangian up to first order in $\hbar$,

$$\begin{aligned}L=&\int D_0^*\left[-\frac{\partial S}{\partial t}-\frac{1}{2m}\left(\nabla S-\frac{e}{c}\mathrm{A}\right)^2-V\right]D_0\,d^3x\\&+\int\left\{\frac{\hbar}{i}(D_0^*D_1-D_1^*D_0)\left[-\frac{\partial S}{\partial t}-\frac{1}{2m}\left(\nabla S-\frac{e}{c}\mathrm{A}\right)^2-V\right]\right.\\&-\frac{\hbar}{i}D_0^*\left[\frac{\partial D_0}{\partial t}+\frac{1}{2m}(\nabla^2S)D_0+\frac{1}{m}\left(\nabla S-\frac{e}{c}\mathrm{A}\right)\cdot\nabla D_0\right.\\&\left.\left.-\frac{ie}{2mc}\sigma\cdot\mathrm{B}\,D_0\right]\right\}d^3x.\end{aligned} \tag{4.4}$$

Variation of the first-, or zeroth-order term gives

$$\delta D_0^*\Rightarrow\left[\frac{\partial S}{\partial t}+\frac{1}{2m}\left(\nabla S-\frac{e}{c}\mathrm{A}\right)^2+V\right]D_0=0, \tag{4.5a}$$

$$\delta S\Rightarrow\frac{\partial}{\partial t}(D_0^*D_0)+\nabla\cdot\left[D_0^*D_0\left(\frac{\nabla S-(e/c)\mathrm{A}}{m}\right)\right]=0. \tag{4.5b}$$

Variation with respect to D_0 yields the Hermitian conjugate of (4.5a) and thus nothing new. If $D_0\neq0$, we see that (4.5a) implies that S obeys the nonrelativistic Hamilton–Jacobi equation; while (4.5b) is again a conservation law for the magnitude of D_0 from which it follows that

$$\frac{d}{dt}(D_0^*D_0)^{-1/2}=\frac{1}{2m}\nabla^2S(D_0^*D_0)^{-1/2}. \tag{4.6}$$

Thus, the evolution of the magnitude of D_0 along each mechanical trajectory is again fixed by S which determines an ensemble of trajectories. To determine the evolution of $d_0=D_0/(D_0^*D_0)^{1/2}$, the "unit" abstract spin

vector ($d_0^\dagger d_0 = 1$), we need to look at the first-order variation of (4.4).

Notice that variation of the term in $(D_0^\dagger D_1 - D_1^\dagger D_0)$ with respect to the D's will yield nothing new, since its coefficient vanishes by virtue of the zeroth-order equation (4.5a). Thus, it may be omitted from the Lagrangian if only results involving D_0 are desired.[12] Variation of the remaining term with respect to $D_0^\dagger$ yields

$$\frac{\partial D_0}{\partial t} + \frac{1}{m}\left(\nabla S - \frac{e}{c}\mathbf{A}\right)\cdot\nabla D_0 + \frac{1}{2m}(\nabla^2 S)D_0 - \frac{ie}{2mc}(\sigma\cdot\mathbf{B})D_0 = 0. \tag{4.7}$$

Variation with respect to D_0 yields the Hermitian conjugate of (4.7) [it is easily checked that (4.5b) actually follows as a consequence of these two equations]. Now, the first two terms in (4.7) are seen to equal dD_0/dt, since $m\mathbf{v} = \nabla S - (e/c)\mathbf{A}$ along a mechanical trajectory, as a consequence of the Hamilton–Jacobi equation. Thus, (4.7) is indeed the equation required for determining the evolution of D_0, given a solution of the Hamilton–Jacobi equation. Our previous work suggests that the equation of evolution of d_0 will be independent of S. Indeed it is easily shown that

$$\frac{d}{dt}d_0 = \frac{ie}{2mc}(\sigma\cdot\mathbf{B})\,d_0 \tag{4.8}$$

Thus, the motion of the particle and of its abstract spin–vector in an external electromagnetic field are given. It only remains to see how the spin–vector in Galilei space–time is determined as a consequence of this equation of motion. Since $d_0^\dagger \sigma d_0 \equiv \mathbf{S}$ will transform as a 3-vector in space as a result of the transformation properties of the two-component spinors and the σ matrices, it is natural to take this as the definition of the spin–vector (actually, any multiple of this could be used, since the resulting equation is linear homogeneous in S). Using (4.8) and its Hermitian conjugate for $d_0^\dagger$, we find immediately that

$$\frac{d\mathbf{S}}{dt} = \frac{-e}{mc}(\mathbf{S}\times\mathbf{B}), \tag{4.9}$$

the equation of motion for spin precession in a magnetic field, for a particle with gyromagnetic ratio two, as might be expected from the Pauli equation.

Note that if we break up D_0 into an amplitude R times d_0,

$$D_0 = Rd_0, \quad R = (D_0^\dagger D_0)^{1/2}, \tag{4.10}$$

we get a representation of D_0 similar to the amplitude-phase representation of a complex number. We shall use this breakup in our discussion of particle Lagrangians in Sec. VI.

V. RELATIVISTIC PARTICLE OF SPIN $\frac{1}{2}$ (DIRAC EQUATION)

Now that our approach is (hopefully) clear, we take up the most complicated example we shall consider in this paper, the relativistic spin-$\frac{1}{2}$ particle in an external electromagnetic field, described quantum mechanically by the Dirac equation. We start from the variational principle,

$$\delta\left\{\frac{1}{2}\left[\frac{\hbar}{i}\left(\partial_\kappa + \frac{e}{c}A_\kappa\right)\psi^*\gamma^\kappa\psi - \psi^*\gamma^\kappa\left(\frac{\hbar}{i}\partial_\kappa - \frac{e}{c}A_\kappa\right)\psi\right] + mc\psi^*\psi\right\}d^4x = 0, \tag{5.1}$$

where ψ is a four-component spinor field, ψ^* its "adjoint" field defined by

$$\psi^* = \psi^*\gamma^4, \tag{5.2}$$

and ψ^* is the Hermitian adjoint of ψ.[13] Inserting the WKBJ ansatz

$$\psi = \left(D_0 + \frac{\hbar}{i}D_1\right)\exp(-iS/\hbar), \quad \psi^* = \left(D_0^* - \frac{\hbar}{i}D_1\right)\exp(-iS/\hbar) \tag{5.3}$$

(note that we have used the same notation, D_0 and D_1, for the 4-spinors here that we used for the 2-spinors of the last section), into the variational principle, we get the expansion of the Lagrangian up to first order in $\hbar$,

$$\int D_0^*\left[-\gamma^\kappa\left(\partial_\kappa S - \frac{e}{c}A_\kappa\right) + mc\right]D_0\,d^4x + \int\frac{\hbar}{i}\Big[\Big(D_1^*\gamma^\kappa D_0 - D_0^*\gamma^\kappa D_1\Big)\left(\partial_\kappa S - \frac{e}{c}A_\kappa\right) + mc(D_0^*D_1 - D_1^*D_0) + \tfrac{1}{2}(\partial_\kappa D_0^*\gamma^\kappa D_0 - D_0^*\gamma^\kappa\partial_\kappa D_0)\Big]d^4x. \tag{5.4}$$

Variation of the zeroth-order terms in (5.4) gives

$$\delta D_0^* \Rightarrow \left[\gamma^\kappa\left(\partial_\kappa S - \frac{e}{c}A_\kappa\right) - mc\right]D_0 = 0, \tag{5.5a}$$

$$\delta S \Rightarrow \partial_\kappa(D_0^*\gamma^\kappa D_0) = 0. \tag{5.5b}$$

[Again, variation with respect to D_0 yields the adjoint equation to (5.5a).][14] Equation (5.5a) will not have any solutions, for nonvanishing D_0, unless the determinant of the matrix in brackets vanishes. This condition is easily seen to be equivalent to the relativistic Hamilton–Jacobi equation

$$\eta^{\mu\nu}\left(\partial_\mu S - \frac{e}{c}A_\mu\right)\left(\partial_\nu S - \frac{e}{c}A_\nu\right) - m^2c^2 = 0. \tag{5.6}$$

The matrix is of rank two, as Rubinow and Keller noted,[5] so that there are only two linearly independent solutions to (5.5a), once S satisfies (5.6). We will not have to use the form of these solutions, given by Rubinow and Keller, but will continue to work with an arbitrary solution.

Now we look at the variations of the first order terms in (5.4). Again, variation with respect to D_1 and D_1^* merely reproduce equations (5.5a) and its adjoint. Thus, the first equations we require result from the variation of the first-order part of (5.4) with respect to D_0 and D_0^*,

$$\gamma^\kappa\partial_\kappa D_0 + \left[\gamma^\kappa\left(\partial_\kappa S - \frac{e}{c}A_\kappa\right) - mc\right]D_1 = 0, \tag{5.7a}$$

$$\partial_\kappa D_0^*\gamma^\kappa + D_1^*\left[\gamma^\kappa\left(\partial_\kappa S - \frac{e}{c}A_\kappa\right) - mc\right] = 0. \tag{5.7b}$$

Note that, because of the first-order derivative form of the Dirac equation, we cannot avoid the appearance

of D_1 and $D_1^\dagger$ in our first-order equations, as we could in the previous second-order wave equations. Our task is to derive equations of motion for D_0 which will not include D_1. We can do this by straightforward computation of $d/d\tau(D_0)$, using our previous equations and a little manipulation of γ matrices. By definition,

$$\frac{dD_0}{d\tau} = \partial_\kappa D_0 \frac{dx^\kappa}{d\tau} \tag{5.8}$$

along any mechanical trajectory, where τ is the proper time along the trajectory. But the equations for the trajectory following from the relativistic Hamilton–Jacobi equation show that

$$m\frac{dx^\kappa}{d\tau} = \eta^{\kappa\lambda}\left(\partial_\lambda S - \frac{e}{c}A_\lambda\right); \tag{5.9}$$

substituting this into (5.8), and remembering that $\eta^{\kappa\lambda} = \frac{1}{2}(\gamma^\kappa\gamma^\lambda + \gamma^\lambda\gamma^\kappa)$, we get

$$\frac{dD_0}{d\tau} = \frac{(\gamma^\kappa\gamma^\lambda + \gamma^\lambda\gamma^\kappa)}{2m}\partial_\kappa D_0\left(\partial_\lambda S - \frac{e}{c}A_\lambda\right). \tag{5.10}$$

When we expand the parenthesis in (5.10), we get two terms, the second of which is

$$\frac{\gamma^\lambda\gamma^\kappa\partial_\kappa D_0}{2m}\left(\partial_\lambda S - \frac{e}{c}A_\lambda\right). \tag{5.11}$$

By using (5.7a), and (5.6) this can be reduced to

$$-\frac{c}{2}\gamma^\kappa\partial_\kappa D_0. \tag{5.12}$$

The first term in (5.10) is

$$\frac{\gamma^\kappa\gamma^\lambda}{2m}\partial_\kappa D_0\left(\partial_\lambda S - \frac{e}{c}A_\lambda\right). \tag{5.13}$$

By writing this as

$$\frac{1}{2m}\left\{\partial_\kappa\left[\gamma^\kappa\gamma^\lambda\left(\partial_\lambda S - \frac{e}{c}A_\lambda\right)D_0\right] - \frac{\gamma^\kappa\gamma^\lambda}{2m}\partial_\kappa\left(\partial_\lambda S - \frac{e}{c}A_\lambda\right)D_0\right\}, \tag{5.14}$$

and by using (5.5a) (and adopting the Lorentz gauge condition $\partial_\kappa A^\kappa = 0$ to avoid some additional steps) the first term reduces to

$$\frac{c}{2}\gamma^\kappa\partial_\kappa D_0 - \frac{1}{2m}(\Box S)D_0 + \frac{e}{2mc}F_{\kappa\lambda}\sigma^{\kappa\lambda}D_0, \tag{5.15}$$

where $\Box S$ means the D'Alembertian of S, and $\sigma^{\kappa\lambda} \equiv \frac{1}{2}(\gamma^\kappa\gamma^\lambda - \gamma^\lambda\gamma^\kappa)$ means the commutator of the γ's. So finally, the required equation of motion for D_0 is

$$\frac{dD_0}{d\tau} = -\frac{\Box S}{2m}D_0 + \frac{e}{2mc}F_{\kappa\lambda}\sigma^{\kappa\lambda}D_0. \tag{5.16}$$

Again, S is required to determine the evolution of D_0 along a mechanical trajectory. However, it is easily shown that

$$\frac{d}{d\tau}(D_0^\dagger D_0) = -\frac{\Box S}{m}(D_0^\dagger D_0); \tag{5.17}$$

so, again defining $d_0 = D_0/(D_0^\dagger D_0)^{1/2}$, we find the equation of motion for d_0,

$$\frac{d}{d\tau}d_0 = \frac{e}{2mc}F_{\kappa\lambda}\sigma^{\kappa\lambda}d_0. \tag{5.18}$$

Note the close analogy with (4.8), which can be made closer by using the operator $s^{\kappa\lambda} = i\sigma^{\kappa\lambda}$, in Eq. (5.18), which is more directly related to the spin–tensor in Minkowski space,

$$\frac{d}{d\tau}d_0 = -\frac{ie}{2mc}F_{\kappa\lambda}s^{\kappa\lambda}d_0. \tag{5.18a}$$

Indeed, we must now relate this equation to the equation of motion of the spin–tensor in Minkowski space. As is well known, $\psi^\dagger s^{\mu\nu}\psi$ transforms like an antisymmetric tensor of second rank, so that it seems natural to define the spin–tensor by

$$S^{\mu\nu} = d_0^\dagger s^{\mu\nu} d_0. \tag{5.19}$$

Thus,

$$\frac{d}{d\tau}S^{\mu\nu} = \left(\frac{d}{d\tau}d_0^\dagger\right)s^{\mu\nu}d_0 + d_0^\dagger s^{\mu\nu}\left(\frac{d}{d\tau}d_0\right), \tag{5.20}$$

and substituting (5.18a) and its adjoint equation into (5.20), we get

$$\frac{d}{d\tau}S^{\mu\nu} = \frac{ie}{2mc}d_0^\dagger[s^{\kappa\lambda}s^{\mu\nu} - s^{\mu\nu}s^{\kappa\lambda}]d_0F_{\kappa\lambda}. \tag{5.21}$$

Using the commutation relations between $s^{\mu\nu}$, which are essentially those for the generators of the homogeneous Lorentz transformations

$$s^{\kappa\lambda}s^{\mu\nu} - s^{\mu\nu}s^{\kappa\lambda} = i(s^{\lambda\nu}\eta^{\kappa\mu} - s^{\kappa\nu}\eta^{\lambda\mu} + s^{\mu\lambda}\eta^{\kappa\nu} - s^{\mu\kappa}\eta^{\lambda\nu}), \tag{5.22}$$

we finally get

$$\frac{d}{d}S^{\mu\nu} = -\frac{e}{mc}(S^{\lambda\nu}\eta^{\kappa\mu} - S^{\lambda\mu}\eta^{\kappa\nu})F_{\kappa\lambda}. \tag{5.23}$$

This is the required equation of motion for the spin tensor, which is seen to be the relativistic generalization of (4.9) for the nonrelativistic spin vector. Indeed, we may introduce a relativistic spin vector S_μ by

$$S_\mu = e_{\mu\nu\kappa\lambda}S^{\nu\kappa}v^\lambda, \tag{5.24}$$

where $e_{\mu\nu\kappa\lambda}$ is the Levi-Civita tensor, equal to $(-\eta)^{1/2}\epsilon_{\mu\nu\kappa\lambda}$. Clearly, $S_\mu v^\mu = 0$, and the **S** of Sec. IV represents the nonrelativistic version of S^μ. On the other hand, it is easily shown that $S^{\mu\nu}p_\mu^{\text{mech}} = 0$, where $p_\mu^{\text{mech}} = \partial_\mu S - (e/c)A_\mu$:

$$\begin{aligned} 2i(D_0^\dagger D_0)S^{\mu\nu}p_\mu^{\text{mech}} &= D_0^\dagger(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)p_\mu^{\text{mech}}D_0 \\ &= D_0^\dagger(\gamma^\mu\gamma^\nu p_\mu^{\text{mech}} - mc\gamma^\nu)D_0 = D_0^\dagger(\gamma^\mu p_\mu^{\text{mech}} - mc)\gamma^\nu D_0 = 0, \end{aligned} \tag{5.25}$$

where we have used (5.5a) and its Hermitian conjugate. This is the well-known Frenkel condition on the spin tensor[1]; it also guarantees, as we shall see in the next section, that p^{mech} and v^μ are parallel. Note that it is not an additional postulate here, but a consequence of the equations of motion. When the Frenkel condition holds, $S^{\mu\nu}$ can be derived from S_μ, so that the two are entirely equivalent,

$$S^{\mu\nu} = e^{\mu\nu\kappa\lambda}S_\kappa v_\lambda, \tag{5.26}$$

where $e^{\mu\nu\kappa\lambda}$ is again the tensor formed from the Levi-Civita tensor density: $e^{\mu\nu\kappa\lambda} = (-\eta)^{-1/2}\epsilon^{\mu\nu\kappa\lambda}$. Equations (5.23), or the corresponding equations for S_μ are equivalent to the Bargmann–Michel–Telegdi equations.[14]

VI. SINGLE-PARTICLE LAGRANGIANS

In the last five sections, we have seen how to develop ensemble Lagrangians, the variation of which lead to partial differential equations of motion for functions describing ensembles of classical particles without and with spin, by WKBJ expansions of the quantum mechanical Lagrangians for relativistic and nonrelativistic particles of spin zero and spin $\frac{1}{2}$. We are now ready to discuss the transition to single-particle Lagrangians, whose variation leads to ordinary differential equations for the mechanical trajectories and abstract spin vector. As we have seen, we cannot hope to find equations of motion for the magnitude of the abstract spin vector along a single trajectory, as this is basically a characteristic of an ensemble density. Thus, we must expect the magnitude of the abstract spin vector to be left undetermined by the equations of motion; however, this indeterminacy can be absorbed by a reparametrization of the equations as we shall see.

If we remember that the integrand of the zeroth order part of our variational principle is essentially the Hamiltonian written in terms of S plus $\partial S/\partial t$ for the nonrelativistic Lagrangians—and a similar expression in the relativistic case—it will not be surprising that we can form a homogeneous particle Hamiltonian by taking this expression, and replacing all derivatives of S by the corresponding particle variables. That is, by letting $\partial S/\partial t \Rightarrow -E$, $\nabla S \Rightarrow \mathbf{p}$ in the nonrelativistic action principles, and letting $\partial_\mu S \Rightarrow p_\mu$ in the relativistic cases, we get a particle Hamiltonian. Subtracting this from $\mathbf{p}\cdot(d\mathbf{r}/d\lambda) - E(dt/d\lambda)$ in the nonrelativistic cases; and from $p_\mu(dx^\mu/d\lambda)$ in the relativistic cases we get a particle Lagrangian. We have here introduced a parameter λ along the path in space–time, to enable us to vary with respect to the time t in the nonrelativistic case, and with respect to all four x^μ in the relativistic case, without worrying about constraints. The variational principle is now homogeneous in λ, and precisely this enables us to get rid of the unwanted freedom in the length of the abstract spin–vector.

We proceed to write down the variation of the Lagrangian for each of our four cases, and briefly discuss the resulting equations of motion.

(a) Nonrelativistic spinless particle:

$$\delta\int\left[\mathbf{p}\cdot\frac{d\mathbf{r}}{d\lambda} - E\frac{dt}{d\lambda} - R_0^*R_0\left(\frac{[\mathbf{p}-(e/c)\mathbf{A}]^2}{2m} + V - E\right)\right]d\lambda = 0. \tag{6.1}$$

Variation with respect to:

$$\delta E \Rightarrow \frac{dt}{d\lambda} = R_0^*R_0 = |R_0|^2, \tag{6.2}$$

which relates the parameter λ to the norm of R;

$$\delta\mathbf{p} \Rightarrow \frac{d\mathbf{r}}{d\lambda} = |R_0|^2\frac{[\mathbf{p}-(e/c)\mathbf{A}]}{m}. \tag{6.3}$$

Using (6.2), this reduces to

$$\frac{d\mathbf{r}}{dt} = \frac{[\mathbf{p}-(e/c)\mathbf{A}]}{m} \tag{6.3'}$$

(from now on, we shall omit this intermediate step, and write time derivatives directly);

$$\delta\mathbf{r} \Rightarrow \frac{d\mathbf{p}}{dt} = -\nabla V + \frac{e}{mc}\,\mathbf{p}\cdot\nabla\mathbf{A} - \frac{e^2}{2mc^2}\mathbf{A}\cdot\nabla\mathbf{A}, \tag{6.4}$$

which, using (6.3′) is easily proved equivalent to the Lorentz force law.

$$\delta R_0^* \Rightarrow \left(\frac{[\mathbf{p}-(e/c)\mathbf{A}]^2}{2m} + V - E\right)R_0 = 0, \tag{6.5}$$

the expression for the total energy as sum of kinetic plus potential energy (with a similar expression from δR_0),

$$\delta t \Rightarrow \frac{dE}{dt} = \frac{\partial V}{\partial t} - \frac{1}{m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)\cdot\frac{\partial\mathbf{A}}{\partial t}, \tag{6.6}$$

which expresses the rate at which the particle's energy changes in a time-dependent external electric field.

Similarly, the first-order term in the expansion of the Lagrangian for the spinless particle can be converted into a Lagrangian for the phase of R_0. But since the equation of motion for the phase is so trivial (phase = const), and the result can be obtained from the Pauli equation results to be given later, we omit the details.

(b) Relativistic spinless particle:

$$\delta\int\left\{p_\mu\frac{dx^\mu}{d\lambda} - \frac{1}{2}\phi^*\phi\left[\eta^{\mu\nu}\left(p_\mu - \frac{e}{c}A_\mu\right)\times\left(p_\nu - \frac{e}{c}A_\nu\right) - m^2c^2\right]\right\}d\lambda = 0, \tag{6.7}$$

$$\delta p_\mu \Rightarrow \frac{dx^\mu}{d\lambda} = \phi^*\phi\left(p^\mu - \frac{e}{c}A^\mu\right), \tag{6.8}$$

$$\delta x^\mu \Rightarrow \frac{dp_\mu}{d\lambda} = -\frac{e}{c}\left(p_\nu - \frac{e}{c}A_\nu\right)\frac{\partial A^\nu}{\partial x^\mu}, \tag{6.9}$$

which is again easily proved equivalent to the Lorentz force law.

$$\delta\phi^* \Rightarrow \left[\eta^{\mu\nu}\left(p_\mu - \frac{e}{c}A_\mu\right)\left(p_\nu - \frac{e}{c}A_\nu\right) - m^2c^2\right]\phi = 0, \tag{6.10}$$

the relativistic energy–momentum relation for a particle of mass m. It follows from (6.8) and (6.10) that

$$\eta_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = (mc\phi^*\phi)^2, \tag{6.11}$$

so that $d\tau/d\lambda = m\phi^*\phi$, where τ is the proper time along the world line. Thus, all λ derivatives can be converted to τ derivatives, giving the correct relativistic relationships.

Again, we omit details of the derivation of the trivial equation of motion for the phase of ϕ from the first-order Lagrangian.

(c) Nonrelativistic particles of spin $\frac{1}{2}$:

$$\delta\int\left[\mathbf{p}\cdot\frac{d\mathbf{r}}{d\lambda}-E\frac{dt}{d\lambda}-D_0^\dagger\left(\frac{[\boldsymbol{\sigma}\cdot(\mathbf{p}-(e/c)\mathbf{A})]^2}{2m}+V-E\right)D_0\right]d\lambda=0. \tag{6.12}$$

The analysis for the mechanical trajectories goes much as in the previous cases, except that now variation with respect to E gives

$$\delta E \Rightarrow \frac{dt}{d\lambda}=D_0^\dagger D_0, \tag{6.13}$$

so that it is the norm of the abstract spin vector which is related to λ.

Now we shall derive the Lagrangian for the evolution of the unit abstract spin vector along the trajectory. To do this, we need to consider the first-order terms in the Lagrangian (4.4). As noted in Sec. IV, the term in $(D_0^\dagger D_1 - D_1^\dagger D_0)$ may be omitted, since its coefficient vanishes by virtue of the zero-order equations of motion. The second term may be rewritten

$$\int D_0^\dagger\left(\frac{d}{dt}D_0-\frac{ie}{2mc}(\boldsymbol{\sigma}\cdot\mathbf{B})D_0+\frac{1}{2m}(\nabla^2 S)D_0\right)d^3x. \tag{6.14}$$

Breaking up D_0 into an amplitude times d_0 ($D_0 = Rd_0$), and inserting this into (6.14), we get (remembering that $d_0^\dagger d_0 = 1$)

$$\int\left[R\left(\frac{d}{dt}R+\frac{1}{2m}(\nabla^2 S)R\right)+R^2 d_0^\dagger\left(\frac{d}{dt}d_0-\frac{ie}{2mc}(\boldsymbol{\sigma}\cdot\mathbf{B})\,d_0\right)\right]d^3x. \tag{6.15}$$

But the first term vanishes, using (4.6), and since we are interested in a single-particle Lagrangian we may take R^2 as a delta function centered on the position of the particle. So we finally arrive at the variational principle for d_0,[15]

$$\delta\int d_0^\dagger\left(\frac{d}{dt}d_0-\frac{ie}{2mc}(\boldsymbol{\sigma}\cdot\mathbf{B})\,d_0\right)dt=0. \tag{6.16}$$

(d) Relativistic particle of spin $\frac{1}{2}$:

$$\delta\int\left\{p_\mu\,dx^\mu-D_0^\dagger\left[\gamma^\kappa\left(p_\kappa-\frac{e}{c}A_\kappa\right)-mc\right]D_0\right\}d\lambda=0, \tag{6.17}$$

$$\delta p^\mu \Rightarrow \frac{dx^\mu}{d\lambda}=D_0^\dagger\gamma^\mu D_0, \tag{6.18}$$

$$\delta x^\mu \Rightarrow \frac{dp_\mu}{d\lambda}=\frac{e}{c}D_0^\dagger\gamma^\kappa D_0\partial_\mu A_\kappa, \tag{6.19}$$

$$\delta D_0^\dagger \Rightarrow \left[\gamma^\kappa\left(p_\kappa-\frac{e}{c}A_\kappa\right)-mc\right]D_0=0. \tag{6.20}$$

(6.20) can only hold for nonvanishing D_0 if the determinant of the matrix in brackets vanishes, which gives the relativistic energy–momentum relation

$$\eta^{\mu\nu}\left(p_\mu-\frac{e}{c}A_\mu\right)\left(p_\nu-\frac{e}{c}A_\nu\right)-m^2c^2=0. \tag{6.21}$$

Multiplication of (6.20) from the left by $D_0^\dagger\gamma^\mu$ gives [remembering (5.25)]

$$\eta^{\mu\kappa}\left(p_\kappa-\frac{e}{c}A_\kappa\right)D_0^\dagger D_0=mc\frac{dx^\mu}{d\lambda}, \tag{6.22}$$

and we see that by choosing $d\tau/d\lambda = D_0^\dagger D_0$ we can go over to the proper time parameterization of the equations of motion. Letting $v^\mu \equiv dx^\mu/d\tau$, it is easily seen that (6.19) is the Lorentz force law of motion. The equations of motion for the 4-spinor d_0 may be obtained by adjoining the Lagrangian

$$\int d_0^\dagger\left(\frac{d}{d\tau}d_0+\frac{ie}{2mc}F_{\kappa\lambda}S^{\kappa\lambda}\,d_0\right)d\tau, \tag{6.23}$$

which may again be derived from the first-order terms in (5.4).[16] We omit the details.

ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with Drs. Havas, Hacyan, and Letelier.

Note: After submission of this paper, Professor Peter Havas kindly brought to our attention the paper of Choquard,[17] which contains some of the results of Schiller,[2,3,6] and Rubinow and Keller,[5] as well as some of our own.

[1]Much of the literature may be traced through the references in H. C. Corben, *Classical and Quantum Theory of Spinning Particles* (Holden-Day, San Francisco, 1968), which surveys this work.

[2]R. Schiller, Phys. Rev. **125**, 1116 (1962).

[3]R. Schiller, Phys. Rev. **128**, 1402 (1962).

[4]W. Pauli, Helv. Phys. Acta **5**, 179 (1932).

[5]S. I. Rubinow and J. B. Keller, Phys. Rev. **131**, 2789 (1963).

[6]K. Rafanelli and R. Schiller, Phys. Rev. **135**, B279 (1964).

[7]For a discussion of the conditions of validity of these different types of classical theories, considered as approximations to the quantum mechanical equations of motion of the electron, see Ref. 5, Sec. II.

[8]See, e.g., J. Frenkel, *Wave Mechanics: Advanced General Theory* (Clarendon, Oxford, 1974), Chap. I, especially Secs. 3 and 4; W. Pauli, "Die alligemeinen Prinzipien der Wellenmechanik" in S. Flügge, Ed., *Handbuch der Physik*, *Band V*, *Teil* 1. *Prinzipien der Quantentheorie I* (Springer-Verlag, Berlin, 1958), Sec. 12.

[9]See, for example, the article: Arthur S. Wightman, "Instability Phenomena in the External Field Problem for Two Classes of Relativistic Wave Equations," in E. H. Lieb, B. Simon, and A. S. Wightman, *Studies in Mathematical Physics* (Princeton U. P., Princeton, New Jersey, 1976), pp. 423–60, and the references to his earlier work therein, for an introduction to this problem.

[10]From now on we drop the subscript from S_0, in the WKBJ approximation, since no confusion can arise.

[11]Note that throughout this section vector notation denotes a 4-vector, $\nabla = \partial/\partial x^\mu \equiv \partial_\mu$, the four-dimensional operator, and a dot product is to be taken with the Minkowski metric, $\mathbf{A}\cdot\mathbf{B} = \eta_{\mu\nu}A^\mu B^\nu$.

[12]Its variation with respect to S will yield a term in a conservation equation involving D_1 that would be needed if we were to go to the second order WKBJ approximation.

[13]We have not included an anomalous magnetic moment in the Dirac equation, although it could be included with a small amount of extra work. In the next paper, we shall consider the case of a classical particle of spin $\frac{1}{2}$ in the most general possible interaction with external fields.

[14]These equations, and/or equations equivalent to them and to the first order equations that will follow were derived by Keller and Rubinow in their paper on the WKBJ approximation to the Dirac equation,[5] following the early work by Pauli[4] on this subject. They also derived the Bargmann–Michel–Telegdi equations for the spin vector by a noncovariant approach: V. Bargmann, L. Michel, V. L. Telegdi, Phys. Rev. Lett. 2, 435 (1959).

[15]Schiller, in Ref. 2, basing himself on the work of H. A. Kramers in his book *Quantum Mechanics* (North-Holland, Amsterdam, 1957), has written down this Lagrangian for the spin equations of motion without deriving it from the WKBJ approximation to the Lagrangian for the Pauli equation.

[16]Schiller, in Ref. 3, has again written down this Lagrangian without deriving it from the WKBJ approximation to the Dirac Lagrangian.

[17]Ph. Choquard, Helv. Phys. Acta, 28, 89 (1955).

Properties of causally continuous closed universes

Hirohisa Ishikawa

Department of Physics, Nagoya University, Nagoya, Japan
(Received 15 July 1977)

We consider the properties of causally continuous space-time with a closed spacelike hypersurface S, i.e., a closed universe. We show that a closed universe does not collide with other universes.

1. INTRODUCTION

There are several stages of causality condition and they are classified by the following causality hierarchy: Globally hyperbolic → causally simple → causally continuous → stably causal → strongly causal → distinguishing → causal → chronological. From the physical point of view the causality condition should be loose, and, to give a definite property (for example, to prove the exis-existence of singularities in an open universe), it should be to some extent strong. Up to now, the stable causality[1] is the acceptable strongest causality condition from the physical point of view (the stable causality is equivalent to the existence of a cosmic time). But the stable causality is too wide to discuss the general properties of space–time structure; for example, the manifold structure of space–time is completely arbitrary. The global hyperbolicity has very nice properties, but this condition is too restrictive from the physical point of view; for example, the Kerr–Newmann rotating charged black hole solution[2] or plane wave space–time[3] is not globally hyperbolic.

Several years ago Hawking and Sachs[4] introduced the concept "causal continuity." This condition is weaker than the global hyperbolicity but stronger than the stable causality. The causal continuity is acceptable from the physical point of view in some sense, as was discussed by Hawking and Sachs, because, if the space–time is not causally continuous in the past direction, some observer finds that all his predictions are upset by the new information which come from the indenfinitely large regions of the space–time. Such behavior destroys the possibility of doing physics. Thus we may assume that $I^-(z)$ is continuous and, if we accept time symmetry, $I^+(z)$ is also continuous. Since it is reasonable to assume that the space–time is to be distinguishing, causal continuity of the space–time is a reaonable condition. Furthermore, as was discussed by Budic and Sachs,[5] the causally continuous space–time has very nice properties in constructing the causal relations on causal boundaries. In this paper, we want to concern our attention to a causally continuous space–time with closed spacelike hypersurface.

In Sec. 2 we review briefly the fundamental concepts and properties which are used in Sec. 3. In Sec. 3 we give the theorems.

2. BASIC CONCEPTS

The manifold we consider here is always connected, Hausdorff, paracompact and differentiable. The word "closed" means compact and without boundaries. The notations, sign conventions, etc., used in this paper are the same as those used in Hawking and Ellis.[6]

A space–time is said to be *distinguishing* if either $I^+(x)=I^+(y)$ or $I^-(x)=I^-(y)$ implies $x=y$.

Suppose F is a function which assigns to each point p in $\mathcal{M}$ an open set $F(p)$ in $\mathcal{M}$. F is called *inner continuous* if, for any p and any compact set K in $F(p)$, there is an open neighborhood U of p such that K is contained in $F(q)$ for all q in U. Apparently I^+ and I^- are inner continuous. F is called *outer continuous* if, for any p and any compact set K in $\mathcal{M}$-Clouser $[F(p)]$, there is a neighborhood U of p such that K is contained in $\mathcal{M}$-Clouser $[F(q)]$ for all q in U.

A causally continuous space-time is defined as follows[4]:

Definition 2.1: A space-time $\mathcal{M}$ is said to be *causally continuous* if it is distinguishing and I^+ and I^- are outer continuous. This condition is equivalent to the condition that the space–time is reflecting, i.e., for all events x and y in M, $I^+(y)\supset I^+(x)$ if and only if $I^-(x)\supset I^-(y)$.

3. THEOREMS

Our result is the following one:

Proposition 3.1: Let the space–time $(\mathcal{M},g)$ be causally continuous; then for any closed spacelike hypersurface S in $\mathcal{M}$, $\mathrm{Bd}[I^+(S)\cup I^-(S)]=S$.

Before proving the proposition we prove the following lemma.

Lemma 3.2: For any closed spacelike hypersurface S, there exists hypersurfaces S^+ and S^- such that S^+ and S^- are strictly located to the future and past of S respectively and such that they are both diffeomorphic to S and every future and past directed timelike curve from S intersects S^+ and S^- respectively. Furthermore, the boundary of $I^+(S)\cup I^-(S)$ is disjoint union of $\mathrm{Bd}[I^+(S)]-S$, $\mathrm{Bd}[I^-(S)]-S$ and S.

Proof: By using the compactness of S, we can cover S with finite number of normal neighborhoods. We can obtain S^+ and S^- by using the orthogonal trajectories of the surface S and the partition of unity. Since S is boundaryless and spacelike, a sufficiently small neighorhood of S is divided into two parts by S, and the surfaces S^+ and S^- can be chosen such that they are the boundaries of the neighborhood divided into two parts by S (i.e., S is two-sided in the sense of Milner[7]). Thus every future and past directed timelike curve from S intersects S^+ and S^- respectively. The last statement follows from the fact that $S\cup I^+(S)\cup I^-(S)$ is an open neighborhood of S and the boundary of $I^+(S)$ is disconnected union of $\mathrm{Bd}[I^+(S)]-S$ and S. □

By using the fact that S^+ is compact and every future

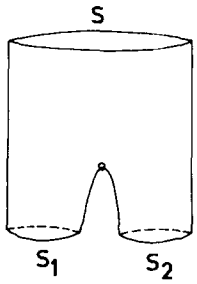


FIG. 1.

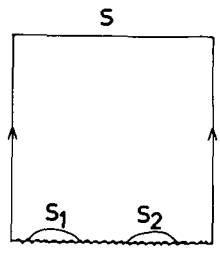


FIG. 2.

directed timelike curve from S intersects S^+ we can prove the proposition.

Proof of the proposition: Suppose there exists a point x in $\mathrm{Bd}[I^+(S)]-S$. From Lemma 3.2 we may assume that x is not in $J^-(S^+)$. Choose a timelike curve γ through x; then for any point y in γ with $x \ll y$, Clouser $[I^-(y)] \cap S^+ \neq \emptyset$. Let $\mathcal{A}$ be a collection of the subsets of the form $\{\text{Clouser } [I^-(y)] \cap S^+ : x \ll y,\ y \in \gamma\}$, then $\mathcal{A}$ has a finite intersection property.[8] Thus there exists a point p common to all members of $\mathcal{A}$ in S^+. If we assume the causal continuity, then p is contained in Clouser $[I^-(x)]$. Since p lies in $I^+(S)$, $I^-(p) \cap I^+(S)$ cannot be contained in any $I^-(z)$ for $z \ll x$. Thus at the point x causal continuity does not hold, which leads to a contradiction. Thus $\mathrm{Bd}[I^+(S)]-S$ is empty. Similarly $\mathrm{Bd}[I^-(S)]-S$ is empty. By Lemma 3.2 this proposition is trivial. □

Since we assumed that M is to be connected following corollary may be easily derived from Proposition 3.1.

Corollary 3.3: If there exists a closed spacelike hypersurface S such that $S \cup I^+(S) \cup I^-(S) \neq M$, then M is not causally continuous. □

In other words, in causally continuous space–time with closed spacelike hypersurface S, someone on S can see every spacelike singularities[9] in the past, if they exist, and for every event in the future of S, one can send a message to the event from S.

As a simple consequence of the Corollary 3.3 we discuss a trouser world.[10] According to Kundt,[11] one may define a trouser world as the space–time M with closed spacelike hypersurfaces S_i $(i=1,2,\ldots,n,\ 2 \leq n)$ such that the space–time is divided into $(n+1)$ parts by suppressing $S_1 \cdots S_n$, the pasts of S_i, the future of $\cup S_i$. This space–time apparently satisfies the condition of Corollary 3.3; thus it is not causally continuous. That is to say, a closed universe does not collide with other universes if we accept the causal continuity. If we drop the condition of the compactness of S_i, there can exist a trouser world with the property of causal continuity. This can be seen in the usual Friedmann universe.

In conclusion, we mention the following point. If we allow noncausally continuous space–time a pathological space–time may be possible. Suppose C^- is the set of points where I^- is not outer continuous. The set C^- is not an open subset of M,[13] but it can be almost open, i.e., dense in some open subset of M. In this kind of space–time the observers who travel in such a region cannot perform physics during some time interval because of the indefinitely large, unpredictable amount of information. This kind of behavior should be excluded. This fact gives some reason to assume that the space–time should be causally continuous.

ACKNOWLEDGMENT

The author expresses his thanks to Dr. M. Ninomiya for his careful reading of the manuscript.

[1] S. Hawking, Proc. Roy. Soc. Lond. A **308**, 433 (1968).
[2] B. Carter, Phys. Rev. **174**, 1559 (1968).
[3] R. Penrose, Rev. Mod. Phys. **37**, 215 (1965).
[4] S. Hawking and R. K. Sachs, Comm. Math. Phys. **35**, 287 (1974).
[5] R. Budic and R. K. Sachs, J. Math. Phys. **15**, 1302 (1974).
[6] S. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[7] T. Milner, *Lecture on the h-Cobordism Theorem* (Princeton U. P., Princeton, N. J., 1965), p. 23.
[8] J. Nagata, *Modern General Topology* (North-Holland, Amsterdam, 1974), p. 84.
[9] Spacelikeness is well defined by using the methods described in Budic and Sachs.[5]
[10] W. Kundt, Comm. Math. Phys. **4**, 143 (1967).
[11] This definition is the stronger version of Kundt's definition. From our definition, according to the discussion of Tipler,[12] the trouser world corresponds to several big bangs if the cosmological constant is negative and the strong energy condition is satisfied.
[12] F. J. Tipler, Astrophys. J. **209**, 12 (1976).
[13] To prove this fact it is sufficient to prove $\gamma \cap C^-$ is countable for any timelike curve γ. If $\gamma \cap C^-$ is not countable, we can construct an uncountable open covering of M. However, this contradicts the Lindelöf-ness of the space–time.

Convergence acceleration technique for lattice sums arising in electronic-structure studies of crystalline solids[a)]

Frank E. Harris[b)]

Department of Chemistry, University of Hawaii at Manoa, 2545 The Mall, Honolulu, Hawaii 96822
(Received 24 June 1977)

Slowly convergent lattice summations arise when *ab initio* quantum-mechanical studies of electronic structure in crystalline solids are carried out by Fourier representation methods. Summations of this type are identified and discussed, and it is shown how a technique related to, but not identical with, that of Ewald can be used to accelerate their convergence. The presentation is illustrated with numerical examples.

I. INTRODUCTION

It is well known that lattice sums of electrostatic energy contributions (Madelung sums) converge so slowly that convergence acceleration schemes are of great importance. This observation is not only relevant for the conditionally convergent summations describing the potentials of charge arrays, but also applies to systems of higher-order multipoles for which there are no formal convergence difficulties. Illustrative of the problem is the summation of r^{-8} over the nonzero points of a simple cubic lattice. To obtain this sum to five significant figures, it is necessary to include points out to approximately $r=10$, a total of over 4,000 points. Allowance for the crystal symmetry would reduce the sum to that of approximately 150 inequivalent points, but in actual applications the distances may be measured from a low-symmetry point or occur with an offset (e.g., $|r^2+\delta^2|^{-4}$). We see that from a practical viewpoint the convergence difficulties of inverse power summations persist to surprisingly high powers.

The two best-known methods for accelerating convergence of Madelung sums are those of Evjen[1] and Ewald.[2] The Ejven method consists of grouping together the contributions of shells of points in such a way that the low-order multipole moments of each shell vanish. The result is that the contributions of shells fall off with increasing distance more rapidly than do the contributions of individual points. We shall not discuss the Evjen method further in this paper, because it is not directly applicable to summations where all points make contributions of the same sign. The Ewald method involves the introduction of an integral transform for the potential, followed by a division of the transform integration into two ranges, each of which is then treated separately. From one integration range there emerges a summation which converges more rapidly than the original sum. For the other integration range, the Poisson summation formula[3] is used to replace the summands by their Fourier transforms, after which that sum also becomes rapidly convergent.

When *ab initio* quantum-mechanical studies of the electronic structures of crystalline solids are developed in a Fourier representation formulation,[4] there arise Madelung-type summations of the sorts and with the problems identified above. It is therefore highly desirable to apply convergence acceleration techniques to such summations, but they differ from those previously studied to an extent which renders impractical the usual acceleration techniques. The purpose of this paper is to present and illustrate a method by which the most time-consuming summations in *ab initio* quantum-mechanical calculations on solids can be evaluated more accurately and conveniently.

In succeeding sections of this paper we describe briefly the physical problems giving rise to the summations whose evaluations we seek, we give alternative (unaccelerated) forms of these sums, and we present an acceleration technique in the spirit of, but not identical with that of Ewald. Illustrative results indicate the effectiveness of the method.

II. PHYSICAL BACKGROUND

Fourier representation techniques have been used for the evaluation of the multicenter integrals arising in electronic structure calculations.[5] In such approaches a key quantity is the Fourier transform of a product of atomic orbitals. For example, the electron repulsion integral $\langle \varphi_a\varphi_c | r_{12}^{-1} | \varphi_b\varphi_d \rangle$ is given by the well-known formula

$$\langle \varphi_a\varphi_c | r_{12}^{-1} | \varphi_b\varphi_d \rangle = \frac{1}{2\pi^2}\int \frac{d\mathbf{q}}{q^2} \exp(-i\mathbf{q}\cdot\mathbf{R}_{ac}) \times \phi_{ab}^T(\mathbf{q})\phi_{cd}^T(-\mathbf{q}), \quad (1)$$

with $\mathbf{R}_{ac}=\mathbf{R}_c-\mathbf{R}_a$, $\mathbf{R}_a$ the center for orbital φ_a, and

$$\phi_{ab}^T(\mathbf{q}) = \langle \varphi_a(\mathbf{r}) | \exp(i\mathbf{q}\cdot\mathbf{r}) | \varphi_b(\mathbf{r}-\mathbf{R}_{ab}) \rangle. \quad (2)$$

Equation (2) shows ϕ_{ab}^T to be the Fourier transform of $\varphi_a^*\varphi_b$ in a coordinate system with origin at $\mathbf{R}_a$. When φ_a and φ_b are Slater-type orbitals (STO's), the right-hand side of Eq. (2) is cumbersome to evaluate but can be reduced either to a single quadrature[5] or to an infinite series of Bessel functions.[6] However, for $q=0$, ϕ_{ab}^T assumes a simple form; it is then the overlap integral $\langle \varphi_a | \varphi_b \rangle$.

Expressions parallel to Eqs. (1) and (2) arise when Fourier representation methods are used for electronic-structure calculations on crystalline solids.[4] As an example, consider the use of Bloch-wave crystal

[a)]Supported in part by the National Science Foundation, under Grant No. CHE-7501284.
[b)]Permanent address: Department of Physics, University of Utah, Salt Lake City, UT 84112.

orbitals $|\mathbf{k}_a\rangle$, defined as

$$|\mathbf{k}_a\rangle = \exp(i\mathbf{k}\cdot\mathbf{r})\sum_\mu \varphi_a(\mathbf{r}-\mathbf{R}_\mu), \tag{3}$$

where $\mathbf{k}$ is the Bloch wave vector and the $\mathbf{R}_\mu$ are the lattice vectors. The index "a" denotes the functional form of φ_a. This definition of $|\mathbf{k}_a\rangle$ is of what we call "mudulated plane wave" type, and shares with the usual linear-combination-of-atomic-orbital (LCAO) or "tight-binding" orbitals the calculational features described below. In terms of these crystal orbitals, the electron repulsion integrals of interest are $\langle \mathbf{k}_a\mathbf{k}'_c|r_{12}^{-1}|\mathbf{k}_b\mathbf{k}'_d\rangle$ and $\langle \mathbf{k}_a\mathbf{k}'_c|r_{12}^{-1}|\mathbf{k}'_b\mathbf{k}_d\rangle$; the first of these integrals diverges faster than the sample size increases, but the divergence is offset by those of the electron–nuclear attraction integrals $\langle \mathbf{k}_a|\sum_\mu|\mathbf{r}-\mathbf{R}_\mu|^{-1}|\mathbf{k}_b\rangle$ and the nuclear–nuclear repulsions.

The nondivergent (i.e., properly extensive) part of $\langle \mathbf{k}_a\mathbf{k}'_c|r_{12}^{-1}|\mathbf{k}_b\mathbf{k}'_d\rangle$ may be reduced to the form

$$\langle \mathbf{k}_a\mathbf{k}'_c|r_{12}^{-1}|\mathbf{k}_b\mathbf{k}'_d\rangle = \frac{4\pi N}{v_0}\sum_\mu{}' \frac{1}{q_\mu^2}\Phi_{ab}^T(\mathbf{q}_\mu)\Phi_{cd}^T(-\mathbf{q}_\mu), \tag{4}$$

where N is the number of unit cells in the sample, the $\mathbf{q}_\mu$ are the vectors of the lattice reciprocal to the $\mathbf{R}_\mu$, and v_0 is the unit-cell volume of the $\mathbf{R}_\mu$ lattice. The prime on the summation sign indicates that the point $\mathbf{q}_\mu = 0$ is to be omitted from the sum. The Φ^T are transforms of lattice sums of atomic–orbital products, and are therefore linear combinations of the ϕ^T appearing in Eq. (2):

$$\Phi_{ab}^T(\mathbf{q}) = \sum_\mu{}' \langle \varphi_a(\mathbf{r})|\exp(i\mathbf{q}\cdot\mathbf{r})|\varphi_b(\mathbf{r}-\mathbf{R}_\mu)\rangle. \tag{5}$$

The integrals $\langle \mathbf{k}_a\mathbf{k}'_c|r_{12}^{-1}|\mathbf{k}'_b\mathbf{k}_d\rangle$ and the nondivergent part of $\langle \mathbf{k}_a|\sum_\mu|\mathbf{r}-\mathbf{R}_\mu|^{-1}|\mathbf{k}_b\rangle$ also reduce to summations involving Φ^T. The overlap integral $\langle \mathbf{k}_a|\mathbf{k}_b\rangle$ assumes the simple form

$$\langle \mathbf{k}_a|\mathbf{k}_b\rangle = N\Phi_{ab}^T(0). \tag{6}$$

Crystalline solids appear to be far more economically described when the atomic orbitals appearing in Eq. (3) are chosen to be STO's rather than the historically more popular Gaussian-type orbitals. Fortunately, an STO-based formulation is practical, as the quantities Φ_{ab}^T of Eq. (5) can be reduced to readily manipulable forms. Although the individual summands of Eq. (5) are difficult to calculate [cf. the discussion immediately following Eq. (2)], the equation can be made tractable by taking advantage of the presence of the lattice sum. A number of calculations based on Eqs. (4) and (5) have now been reported.

While the work done to date suffices to demonstrate the practicality of Fourier representation methods for solid-state electronic structure studies, it has also shown that the evaluation of Eq. (5) consumes the bulk of the required computational effort, and that the root of the problem is the slow rate of convergence of the lattice sum involved. We therefore turn our attention to methods for the evaluation of Eq. (5), and more specifically to the introduction of convergence acceleration techniques of the kinds already found to be useful in Madelung summations.

III. EVALUATION OF $\Phi_{ab}^T(\mathbf{q})$

We now consider more specifically the evaluation of Eq. (5) when φ_a and φ_b are normalized ls STO's:

$$\varphi_a = (\zeta_a^3/\pi)^{1/2}\exp(-\zeta_a r). \tag{7}$$

Expressions involving STO's of higher quantum numbers can be derived by analogy or by differentiating the results given here with respect to the screening parameters ζ_a and ζ_b. Inserting Eq. (7) into Eq. (5),

$$\Phi_{ab}^T(\mathbf{q}) = [(\zeta_a\zeta_b)^{3/2}/\pi]\sum_\mu \int d\mathbf{r} \times\exp(-\zeta_a r - \zeta_b|\mathbf{r}-\mathbf{R}_\mu| + i\mathbf{q}\cdot\mathbf{r}). \tag{8}$$

The integral on the right-hand side of Eq. (8) is that which was earlier identified as cumbersome to evaluate. For simplicity we assume a simple cubic lattice.

The most straightforward evaluation of Φ_{ab}^T is obtained by using the Fourier convolution theorem to write

$$\Phi_{ab}^T(\mathbf{q}) = [1/(2\pi)^3]\sum_\mu \int d\mathbf{p}\,\varphi_a^T(\mathbf{q}-\mathbf{p})[\varphi_b(\mathbf{r}-\mathbf{R}_\mu)]^T(\mathbf{p}). \tag{9}$$

Introducing the expression for the transform of the ls STO,

$$\varphi_a^T(\mathbf{q}) = 8\pi^{1/2}\zeta_a^{5/2}/(q^2+\zeta_a^2)^2, \tag{10}$$

and noting that

$$[\varphi(\mathbf{r}-\mathbf{R}_\mu)]^T(\mathbf{p}) = \exp(i\mathbf{R}_\mu\cdot\mathbf{p})\varphi^T(\mathbf{p}), \tag{11}$$

we have

$$\Phi_{ab}^T(\mathbf{q}) = \frac{8(\zeta_a\zeta_b)^{5/2}}{\pi^2}\sum_\mu \int d\mathbf{p} \times\frac{1}{(|\mathbf{q}-\mathbf{p}|^2+\zeta_a^2)^2}\,\frac{\exp(i\mathbf{R}_\mu\cdot\mathbf{p})}{(p^2+\zeta_b^2)^2}. \tag{12}$$

Next, we interchange the order of summation and integration in Eq. (12), reaching thereby a lattice sum satisfying[7]

$$\sum_\mu \exp(i\mathbf{R}_\mu\cdot\mathbf{p}) = (8\pi^3/v_0)\sum_\mu \delta(\mathbf{p}-\mathbf{p}_\mu), \tag{13}$$

where the $\mathbf{p}_\mu$ are reciprocal-lattice vectors. Equation (13) is sometimes referred to as a "lattice orthogonality relation", and is a special case of a Poisson summation formula, the summands on its right-hand side being Fourier transforms of those on its left. When the right side of Eq. (13) is substituted into Eq. (12), the $\mathbf{p}$ integration reduces to a lattice sum, and we have the final result

$$\Phi_{ab}^T(\mathbf{q}) = \frac{64\pi(\zeta_a\zeta_b)^{5/2}}{v_0} \times\sum_\mu \frac{1}{(|\mathbf{q}-\mathbf{p}_\mu|^2+\zeta_a^2)^2(p_\mu^2+\zeta_b^2)^2}. \tag{14}$$

Equations (8) and (14) may be regarded as the two "standard" ways to express Φ_{ab}^T as a lattice sum. To simplify further discussion, we recapitulate these equations in a dimensionless notation in which the summations are over a unit lattice of vectors μ, and write $v_0 = a^3$, $\delta_i = a\zeta_i/2\pi$, $\mathbf{q} = 2\pi\nu/a$:

$$\Phi_{ab}^T(\nu) = \frac{(\delta_a\delta_b)^{3/2}}{\pi}\sum_\mu \int d\mathbf{r}$$

$$\times \exp(-\delta_a r - \delta_b |\mathbf{r} - 2\pi\mu| + i\nu\cdot\mathbf{r}) \tag{15}$$

$$= \frac{8(\delta_a\delta_b)^{5/2}}{\pi^2} \sum_\mu \frac{1}{(|\nu-\mu|^2+\delta_a^2)^2(\mu^2+\delta_b^2)^2}. \tag{16}$$

Leaving aside for the moment the very real differences in ease of evaluation of the summands of Eqs. (15) and (16), we note that the summation of Eq. (15) will converge rapidly when δ_a or δ_b is large, but only slowly when both δ_a and δ_b are small. We note also that the convergence will be exponential (as that of a three-dimensional geometric series). On the other hand, Eq. (16) will converge rapidly when δ_a or δ_b is small, but will become inappropriate when δ_a and δ_b are both large. The summation of Eq. (16) approaches for large μ that of μ^{-8}, with the disadvantages identified for such sums in the introduction.

Remembering now that the summands of Eq. (15) are relatively difficult to evaluate, we may appreciate the real problem in the evaluation of Φ_{ab}^T. We should like to have at least part of the convenience of Eq. (16), but in a context improving its convergence, particularly for large δ_a and δ_b. We cannot really afford to use Eq. (15) unless δ_a and δ_b are so large that extremely few terms are required.

IV. ACCELERATION OF CONVERGENCE

Following the general idea of the Ewald method, we introduce integral transforms for φ_a and φ_b in Eq. (5), planning to divide the integrations into regions each of which will receive optimal further processing. The transform we have found most suitable is that suggested by Kikuchi[8] and subsequently used by Shavitt and Karplus[9]:

$$\exp(-\delta r) = \frac{\delta}{\pi^{1/2}} \int_0^\infty dx\, x^{-1/2} \exp(-x\delta^2 - r^2/4x). \tag{17}$$

Insertion of Eq. (17) for $\exp(-\delta_a r)$ and $\exp(-\delta_b|\mathbf{r}-2\pi\mu|)$ in Eq. (15) [the dimensionless equivalent of Eq. (5)] yields

$$\Phi_{ab}^T(\nu) = \frac{(\delta_a\delta_b)^{5/2}}{\pi^2} \sum_\mu \int_0^\infty dx \int_0^\infty dy (xy)^{-1/2} \times \exp(-\delta_a^2 x - \delta_b^2 y) \int d\mathbf{r} \exp\left(-\frac{r^2}{4x} - \frac{|\mathbf{r}-2\pi\mu|^2}{4y} + i\nu\cdot\mathbf{r}\right). \tag{18}$$

The r integration may now be carried out, leading to the result

$$\Phi_{ab}^T(\nu) = \frac{8(\delta_a\delta_b)^{5/2}}{\pi^{1/2}} \sum_\mu \int_0^\infty dx \int_0^\infty dy \frac{xy}{(x+y)^{3/2}} \times \exp\left(-\delta_a^2 x - \delta_b^2 y - \frac{\nu^2 xy}{x+y} - \frac{\pi^2\mu^2}{x+y} + \frac{2\pi i x \mu\cdot\nu}{x+y}\right). \tag{19}$$

We next notice that the μ summation will converge at a rate mainly determined by the magnitude of $x+y$. We therefore change variables from x, y to s, t, where $s = x+y$, $t = (x-y)/(x+y)$, with limits $0 \le s < \infty$, $-1 \le t \le 1$, and with $dx\,dy = \frac{1}{2} s\,ds\,dt$:

$$\Phi_{ab}^T(\nu) = \frac{(\delta_a\delta_b)^{5/2}}{\pi^{1/2}} \sum_\mu \int_0^\infty ds \int_{-1}^1 dt\, s^{3/2}(1-t^2) \times \exp(-\gamma s - \pi^2\mu^2/s + \pi i(1+t)\mu\cdot\nu), \tag{20}$$

with

$$\gamma = \tfrac{1}{2}\delta_a^2(1+t) + \tfrac{1}{2}\delta_b^2(1-t) + \tfrac{1}{4}\nu^2(1-t^2). \tag{21}$$

We are now ready to divide the range of s into the two intervals $(0, Z)$ and (Z, ∞), where Z is arbitrary and will be specified later. For the first of these intervals the μ sum will be strongly convergent, in fact converging as $\exp(-\pi^2\mu^2/Z)$. For the remaining interval, the convergence is poor, but can be improved by use of the Poisson summation formula

$$\sum_\mu \exp(-\pi^2\mu^2/s + \pi i(1+t)\mu\cdot\nu) = \left(\frac{s}{\pi}\right)^{3/2} \sum_\mu \exp[-s|\mu - \tfrac{1}{2}(1+t)\nu|^2]. \tag{22}$$

After substitution of Eq. (22) for (Z, ∞), the integrations in both s and t may be carried out, leading to the final result

$$\Phi_{ab}^T(\nu) = \frac{(\delta_a\delta_b)^{5/2}}{\pi^{1/2}} \sum_\mu \int_0^Z ds \int_{-1}^1 dt\, s^{3/2}(1-t^2) \times \exp[-\gamma s - \pi^2\mu^2/s + \pi i(1+t)\mu\cdot\nu] + \frac{8(\delta_a\delta_b)^{5/2}}{\pi^2} \sum_\mu \left[\frac{\exp(-\alpha Z)}{\alpha(\alpha-\beta)^2}\left(\frac{2}{\alpha-\beta} + \frac{1}{\alpha} + Z\right) + \frac{\exp(-\beta Z)}{\beta(\beta-\alpha)^2}\left(\frac{2}{\beta-\alpha} + \frac{1}{\beta} + Z\right)\right], \tag{23}$$

with

$$\alpha = |\mu - \nu|^2 + \delta_a^2, \tag{24}$$

$$\beta = \mu^2 + \delta_b^2. \tag{25}$$

If there is a μ value such that $\alpha = \beta$, the square bracket on the right-hand side of Eq. (23) must be replaced for that μ value by

$$[\ ] = \frac{\exp(-\alpha Z)}{\alpha}\left(\frac{Z^3}{6} + \frac{Z^2}{2\alpha} + \frac{Z}{\alpha^2} + \frac{1}{\alpha^3}\right) \quad (\alpha = \beta). \tag{26}$$

We note that the portion of Eq. (23) corresponding to the interval (Z, ∞) now exhibits convergence as $\exp(-\mu^2 Z)$.

Equation (23) possesses the convergence acceleration properties we seek. The value of Z remains to be chosen; it determines the relative importance of the two μ summations. In the limit $Z \to 0$, the first sum vanishes while the second approaches that of Eq. (16); in the limit $Z \to \infty$ only the first sum survives, and it can be shown equivalent to Eq. (15). Intermediate values of Z give pairs of series corresponding roughly to Eqs. (15) and (16), but each converging more rapidly than the limiting series to which it corresponds. Remembering that the first summation contains summands which are harder to evaluate than those of the second sum, we may choose Z small enough that very few terms of the first summation are needed, but with Z sufficiently far from zero that we have effective exponential convergence in the second sum.

V. NUMERICAL EXAMPLES

For illustrative purposes we elect to aim for seven significant figures and to choose Z small enough that in the first sum of Eq. (23) we will need to keep only $\mu = 0$ and the six members of the (1, 0, 0) star

$[\mu=(1,0,0),(-1,0,0),(0,1,0),(0,-1,0),(0,0,1),(0,0,-1)]$. The leading terms omitted from the first sum will have $\mu^2=2$, and can be expected to be smaller than the $\mu=0$ term by approximately the factor $\exp(-2\pi^2/Z)$, ignoring any additional convergence which may be produced by large values of δ_a or δ_b. We therefore tentatively take $Z=1$, causing the six omitted terms with $\mu^2=2$ to aggregate to about 10^{-8} or less of the leading term. Details of the calculation of the above-described terms are given in an appendix.

In evaluating the second sum, we note that the $\exp(-\alpha Z)$ portion of the summand is maximal when $\mu\approx\nu$, while the $\exp(-\beta Z)$ portion peaks at $\mu=0$. We therefore reorganize the sum so as to keep for each portion a set of terms radiating spherically out from its respective maximum. With $Z=1$ our accuracy requirements will be met if we keep μ values within spheres of radius $\sqrt{18+}$ about the maxima. There are 20 stars of μ values within such a sphere.

For the purpose of presenting numerical results, we write Eq. (23) in the form

$$\Phi^T_{ab}(\nu)=SZ(0)+SZ(1,0,0)+LZ, \tag{27}$$

where $SZ(0)$ and $SZ(1,0,0)$ are the respective contributions of the $\mu=0$ and $(1,0,0)$ stars to the first (small-Z) summation, and LZ is the contribution of the second (large-Z) summation. Table I gives the results of a number of calculations for various δ_a, δ_b, and ν values, both for $Z=1$ and for other Z values. The accuracy of the results is attested to by their substantial independence of Z; the effectiveness of the formulation is indicated by the fact that some of the quoted results would have required thousands of terms if evaluated without a convergence acceleration technique.

ACKNOWLEDGMENTS

The author gratefully acknowledges the hospitality of the Chemistry Department of the University of Hawaii, and wishes particularly to thank Professor C. S. Fadley for his part in making arrangements conducive to a productive and enjoyable visit.

APPENDIX: EVALUATION OF "SMALL-Z" SUMMANDS

The $\mu=0$ term, denoted $SZ(0)$ in Eq. (27), can be completely reduced to a closed expression involving the error function. It is convenient to introduce the definitions

$$F(u)=u^{-1/2}\operatorname{erf}(u^{1/2})=\frac{2}{\sqrt{\pi}}\int_0^1\exp(-uv^2)\,dv, \tag{A1}$$

$$G(u)=u^{-1/2}e^u\operatorname{erf}(u^{1/2})=\frac{2e^u}{\sqrt{\pi}}\int_0^1\exp(-uv^2)\,dv. \tag{A2}$$

In the material to follow, $G(u)$ occurs for negative u, and therefore involves error functions of imaginary argument. We have prepared a computer program for rapid and accurate evaluation of $G(u)$ with negative u.

The actual reduction of $SZ(0)$ is tedious, but can be accomplished with a judicious choice of standard procedures. The results can be expressed in terms of the auxiliary quantities $\alpha=\frac{1}{2}\nu(1+d)$, $\beta=\frac{1}{2}\nu(1-d)$, $a=\frac{1}{2}(\delta_a^2+\delta_b^2)+\frac{1}{4}\nu^2$, $b=\frac{1}{4}(\delta_a^2-\delta_b^2)$, $c=\frac{1}{2}(\delta_a^2+\delta_b^2+\alpha^2+\beta^2)$, and $d=(\delta_b^2-\delta_a^2)/\nu^2$. The final formulas depend upon whether ν or b vanish; we distinguish three cases as set forth below:

$$\begin{aligned}SZ(0)=&[4(\delta_a\delta_b)^{5/2}Z^{1/2}/\pi^{1/2}c\nu^3]\\&\times((\pi^{1/2}\nu/2c)\{[\alpha^2+d\delta_a^2]F(\delta_a^2Z)+[\beta^2-d\delta_b^2]F(\delta_b^2Z)\}\\&+\exp(-\delta_a^2Z)\{\beta-\alpha[c-(4\alpha\beta/\nu^2)(1+Zc)]G(-\alpha^2Z)\}\\&+\exp(-\delta_b^2Z)\{\alpha-\beta[c-(4\alpha\beta/\nu^2)\\&\times(1+Z_c)]G(-\beta^2Z)\}\quad(\nu\neq0)\\=&[(\delta_a\delta_b)^{5/2}Z^{1/2}/2b^3]\end{aligned} \tag{A3}$$

TABLE I. Lattice sums $\Phi^T_{ab}(\nu)$ of Eqs. (23) and (27) and the "small-Z" and "large-Z" contributions thereto for various values of the convergence acceleration parameter Z.

δ_a	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
δ_b	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
ν_x	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	8.0
ν_y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0
ν_z	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0
Z	0.50	0.75	1.00	1.10	1.20	0.80	1.00	1.20	1.00
$SZ(0)$	0.13143383	0.25054432	0.36571232	0.40785898	0.44726064	0.24930055	0.32635939	0.39299372	3.8222328(−3)
$SZ(1,0,0)$	0.00000000	0.00000027	0.00001192	0.00003330	0.00007819	0.00000036	0.00000615	0.00004025	0.0000146(−3)
LZ	0.75903396	0.63993161	0.52475204	0.48258395	0.44313733	0.45821558	0.38115115	0.31448285	0.5628282(−3)
$\Phi^T_{ab}(\nu)$	0.89046779	0.89047620	0.89047628	0.89047623	0.89047615	0.70751649	0.70751669	0.70751681	4.3850756(−3)

δ_a	2.0	6.0	6.0	6.0	0.2	0.2	0.2	0.2
δ_b	1.0	1.0	1.0	1.0	0.1	0.1	0.1	0.1
ν_x	8.0	0.0	0.0	0.0	0.0	0.0	8.0	8.0
ν_y	6.0	0.0	0.0	0.0	0.0	0.0	6.0	6.0
ν_z	4.0	0.0	0.0	0.0	0.0	0.0	4.0	4.0
Z	1.20	0.80	1.00	1.20	0.50	1.00	0.50	1.00
$SZ(0)$	3.9458438(−3)	0.22931254	0.25550774	0.27513503	0.00000	0.00002	0.0049480(−5)	0.0072523(−5)
$SZ(1,0,0)$	0.0000892(−3)	0.00000024	0.00000333	0.00001895	0.00000	0.00000	0.0000000(−5)	0.0000000(−5)
LZ	0.4391423(−3)	0.12086336	0.09466506	0.07502213	286.57986	286.57985	3.6241185(−5)	3.6218148(−5)
$\Phi^T_{ab}(\nu)$	4.3850753(−3)	0.35017614	0.35017613	0.35017610	286.57986	286.57986	3.6290665(−5)	3.6290671(−5)

$$\times\{(1/\pi^{1/2}Z)[\exp(-\delta_b^2 Z)-\exp(-\delta_a^2 Z)]$$
$$+(a-b)F(\delta_b^2 Z)-(a+b)F(\delta_a^2 Z)\}\quad(\nu=0,\ b\neq 0)\qquad(A4)$$
$$=(\delta_a^5 Z^{1/2}/a^2)[F(\delta_a^2 Z)-(2/\pi^{1/2})(1+2aZ/3)$$
$$\times\exp(-\delta_a^2 Z)]\quad(\nu=b=0).\qquad(A5)$$

The other contribution we require is that when μ is a member of the (1, 0, 0) star; this contribution was denoted $SZ(1, 0, 0)$. Starting from Eq. (23), we proceed by expanding $\exp(-\gamma s)$ in a Taylor series about $s=Z$, after which the s and t integrations separate in each term:

$$SZ(1,0,0)=\frac{(\delta_a\delta_b)^{5/2}}{\pi^{1/2}}\sum_{n=0}^{\infty}A_n(Z)\int_{-1}^{1}dt\,\gamma^n(1-t^2)\exp(-\gamma Z)$$
$$\times\{\cos[\pi(1+t)\nu_x]+\cos[\pi(1+t)\nu_y]$$
$$+\cos[\pi(1+t)\nu_z]\},\qquad(A6)$$

with

$$A_n(Z)=\frac{2}{n!}\int_0^Z ds\,s^{3/2}(Z-s)^n\exp(-\pi^2/s).\qquad(A7)$$

The quantities $A_n(Z)$ depend only upon n and Z, so that once Z is fixed they may be determined once and for all. For $Z=1$ they decrease rapidly with increasing n; for seven-significant-figure results they are not needed beyond $n=4$.

The t integration could be reduced analytically to error functions of complex argument, but it is never needed with an accuracy of more than about four significant figures and is easily handled by numerical integration. We have obtained satisfactory results both from the use of Simpson's rule and from Lobatto quadratures.[10]

[1]H. M. Evjen, Phys. Rev. **39**, 675 (1932).
[2]P. P. Ewald, Ann. Physik **64**, 253 (1921).
[3]See, for example, F. Oberhettinger, *Fourier Expansions, A Collection of Formulas* (Academic, New York, 1973), p. 5.
[4]See, for example, F. E. Harris, in *Advances in Theoretical Chemistry*, edited by D. Henderson and H. Eyring (Academic, New York, 1975), Vol. 1, pp. 147–218.
[5]R. A. Bonham, J. L. Peacher, and H. L. Cox, Jr., J. Chem. Phys. **40**, 3083 (1964).
[6]H. J. Monkhorst and F. E. Harris, Int. J. Quantum Chem. **6**, 601 (1972).
[7]See, for example, J. Callaway, *Quantum Theory of the Solid State* (Academic, New York, 1974), pp. 352ff.
[8]R. Kikuchi, J. Chem. Phys. **22**, 148 (1954).
[9]I. Shavitt and M. Karplus, J. Chem. Phys. **43**, 398 (1965).
[10]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Nat. Bur. Stds., Washington, D. C., 1964), p. 888.

Casimir invariants and vector operators in simple and classical Lie algebras [a)]

Susumu Okubo

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 17 January 1977)

A method of computing eigenvalues of certain types of Casimir invariants has been developed for simple and classical Lie algebras. Especially these eigenvalues for algebras A_n, B_n, C_n, D_n, and G_2 have been computed in closed terms. We also enumerate numbers and functional forms of all linearly independent vector operators in terms of generators in any irreducible representation of these algebras. Some polynomial identities among infinitesimal generators of these algebras are derived by means of the same technique.

INTRODUCTION

The Lie group and Lie algebra are very useful mathematical tools in dealing with various branches[1] of theoretical physics. It is known[2] that any simple Lie algebra L of rank n has precisely n algebraically independent Casimir invariants whose structures are determined by Betti numbers and Poincaré polynomial of the group. We also know from pioneering works of Racah[3] and Gel'fand[4] that any irreducible representation of L can be labelled by eigenvalues of these n invariants. Since then, the problem has been exhaustively investigated by many authors,[5-12] especially for algebras A_n, B_n, C_n, D_n, and G_2 because of its physical relevance to many-body problems. However, some related problems are not yet completely solved. First, eigenvalues of these Casimir invariants for the algebra A_n have been computed in a closed form by Popov and Perelomov[13] and by Louck and Biedenharn,[10] while those of B_n, C_n, and D_n have been calculated by Wong and Yeh[14] and by Nwachuku and Rashid.[15] The methods employed by these authors utilize heavily special properties inherent in these particular algebras, so that its extension to other algebras does not appear to be straightforward. So far, no general unified formula for this problem seems to exist. Here in this paper, we shall present a partial solution for the problem in the sense that we can compute eigenvalues of certain types of Casimir invariants for any simple and classical Lie algebras. As examples, we computed them in closed forms for algebras A_n, B_n, C_n, D_n, and G_2, reproducing some of the previously known results.

Secondly, we often have a problem of finding numbers and explicit forms (as functions of generators of the algebra) of all linearly independent vector operators of L. The problem has been solved for the algebra A_n by the present author,[16] and subsequently the same technique has been extended with some partial success by Nwachuku and Rashid[17] for the algebras B_n, C_n, and D_n. Again, the method depends heavily upon special properties enjoyed by these algebras. In this paper, we shall develop another method applicable to more general class of algebras and will enumerate numbers and functional forms (in terms of generators) of all linearly independent vector operators of algebras A_n, B_n, C_n, D_n, and G_2 in any given irreducible representation.

Third, the infinitesimal generators of some Lie algebras are known[16,18-20] to satisfy various polynomial identities in given irreducible representations. We shall present a more systematic way of finding such identities, which is applicable to any simple Lie algebra.

The main advantage of the technique to be developed in this paper is that it is straightforward and can be applicable to a wider class of algebras. Also, we can simultaneously treat three problems described above.

2. CASIMIR INVARIANTS AND VECTOR OPERATORS

Let L be a finite-dimensional Lie algebra over the complex number field, and let X_μ $(\mu=1,2,\ldots,b)$ be a basis of L. Then we have the Lie equation

$$[X_\mu, X_\nu] = C^\lambda_{\mu\nu} X_\lambda, \tag{2.1}$$

where the repeated index over λ automatically implies a summation of λ over values $1,2,\ldots,b$. Then, a set consisting of b elements T_μ $(\mu=1,2,\ldots,b)$ which are members of the universal enveloping algebra $\hat{L}$ of L is called a vector operator in $\hat{L}$, if they satisfy the commutation relation

$$[X_\mu, T_\nu] = C^\lambda_{\mu\nu} T_\lambda. \tag{2.2}$$

Comparing (2.1) and (2.2), we see that the generator X_μ is a vector operator. However, it is in general not the only one. Similarly, a Casimir invariant K in $\hat{L}$, is defined as any element K of $\hat{L}$, satisfying

$$[X_\mu, K] = 0 \tag{2.3}$$

for all $\mu=1,2,\ldots,b$. We may also define vector operators in a given representation of L. If d is the dimension of the representation, and if X_μ are now $d\times d$ matrix representations of L, then any $d\times d$ matrices T_μ will be called a vector operator in the representation, provided that the commutation relation (2.2) is satisfied.

Analogous to T_μ, we may define a covariant vector operator T^μ by the commutation relation

$$[X_\mu, T^\nu] = -C^\nu_{\mu\lambda} T^\lambda \tag{2.4}$$

either in $\hat{L}$ or in a given representation of L. If T^μ and S_μ are covariant vector and vector operators respectively, then the product

[a)]Work supported in part by the U.S. Energy Research and Development Administration.

$$K = T^{\lambda} S_{\lambda} \text{ or } S_{\lambda} T^{\lambda} \tag{2.5}$$

is a Casimir invariant, as we may check easily. Conversely, if K is a Casimir invariant in $\hat{L}$, then we can find a covariant vector operator T^{μ} such that

$$K = T^{\lambda} X_{\lambda} = X_{\lambda} T^{\lambda}. \tag{2.6}$$

However, since we do not directly utilize this fact in the present paper, we will not prove it.

Let us, as usual, define the adjoint representation $\mathrm{ad}X_{\lambda}$ by

$$(\mathrm{ad}X_{\lambda})_{\mu\nu} = C^{\mu}_{\lambda\nu}. \tag{2.7}$$

Moreover, let us assume that its conjugate representation specified by $-(\mathrm{ad}X_{\lambda})^{T}$ (where the superscript T stands for the transpose matrix) is equivalent to $\mathrm{ad}X_{\lambda}$, i.e., that we have a nonsingular $b \times b$ matrix $\overline{g}_{\mu\nu}$ with its inverse $(\overline{g}^{-1})_{\mu\nu} = \overline{g}^{\mu\nu}$ satisfying

$$\overline{g}(\mathrm{ad}X_{\lambda})\overline{g}^{-1} = -(\mathrm{ad}X_{\lambda})^{T}. \tag{2.8}$$

Then we shall call any such L to be a classical Lie algebra, following Freund and Kaplansky.[21] In terms of component, (2.8) is rewritten as

$$\overline{g}_{\mu\lambda} C^{\lambda}_{\nu\tau} = -C^{\lambda}_{\nu\mu} \overline{g}_{\lambda\tau}. \tag{2.9}$$

A sufficient condition for L being classical is as follows. If we can find at least one nontrivial representation $\{\lambda_0\}$ of L such that a symmetric bilinear form

$$\overline{g}_{\mu\nu} = \overline{g}_{\nu\mu} = \mathrm{tr}(x_{\mu} x_{\nu}) \tag{2.10}$$

is nondegenerate, then L is classical. Here x_{μ} $(\mu = 1, 2, \ldots, b)$ are the representation matrices of X_{μ} in $\{\lambda_0\}$. The fact that $\overline{g}_{\mu\nu}$ defined by (2.10) satisfies (2.9) is a simple consequence of an identity

$$\mathrm{tr}(x_{\mu}[x_{\nu}, x_{\tau}]) = \mathrm{tr}(x_{\tau}[x_{\mu}, x_{\nu}]).$$

Especially, any semisimple Lie algebra is automatically classical, since we can choose $\{\lambda_0\}$ to be the adjoint representation with $x_{\mu} = \mathrm{ad}X_{\mu}$ and we use Cartan's criteria of semisimplicity. Actually, for semisimple Lie algebras, any nontrivial representation $\{\lambda_0\}$ can be used[22] for this purpose to define $\overline{g}_{\mu\nu}$ as in (2.10), leading to the same conclusion. Moreover, if L is simple, then we can prove that any $\overline{g}_{\mu\nu}$ defined by (2.10) [or more generally by (2.9)] can be expressed as

$$\overline{g}_{\mu\nu} = C\,\mathrm{tr}(\mathrm{ad}X_{\mu}\,\mathrm{ad}X_{\nu}) \tag{2.10'}$$

where C is a nonzero constant which depends upon $\{\lambda_0\}$. The Lie algebra of l-dimensional unitary group $\mathrm{U}(l)$ is not semisimple but classical. This fact will be used in Sec. 4.

If L is classical, then there is a one-to-one correspondence between a vector operator T_{μ} and a covariant vector operator T^{μ} by

$$T^{\mu} = \overline{g}^{\mu\nu} T_{\nu}, \quad T_{\mu} = \overline{g}_{\mu\nu} T^{\nu}. \tag{2.11}$$

Indeed, Eqs. (2.2) and (2.4) are mutually consistent with (2.11), if we note (2.9). Hereafter, we assume L to be classical, and designate X_{μ} to be either the basis of L or its representation matrices in a generic irreducible representation $\{\lambda\}$ of L. Let x_{μ} be an arbitrary but fixed nontrivial representation $\{\lambda_0\}$ of L, which we shall call the reference representation. In order to avoid possible confusion, we identify this reference representation with the one in which $\overline{g}_{\mu\nu}$ given by (2.10) is non-degenerate, whenever such a $\{\lambda_0\}$ exists. We now define

$$g_{\mu_1\mu_2\cdots\mu_p} = \mathrm{tr}(x_{\mu_1} x_{\mu_2} \cdots x_{\mu_p}), \tag{2.11a}$$

$$g^{\mu_1\mu_2\cdots\mu_p} = \mathrm{tr}(x^{\mu_1} x^{\mu_2} \cdots x^{\mu_p}), \tag{2.11b}$$

$$x^{\mu} = \overline{g}^{\mu\nu} x_{\nu}, \tag{2.11c}$$

for $p \geq 1$. If L is semisimple, then we have

$$g_{\mu} = g^{\mu} = 0 \tag{2.12}$$

for the case $p = 1$, since $\mathrm{tr}x_{\mu} = 0$, but this is not necessarily valid for cases of classical Lie algebras. Moreover, if we define

$$K^{(p)} = g^{\mu_1\mu_2\cdots\mu_p} X_{\mu_1} X_{\mu_2} \cdots X_{\mu_p}, \tag{2.13}$$

$$S^{(p)}_{\lambda} = g_{\lambda\mu_1\mu_2\cdots\mu_p} X^{\mu_1} X^{\mu_2} \cdots X^{\mu_p}, \tag{2.14}$$

then $K^{(p)}$ and $S^{(p)}_{\lambda}$ are, respectively, Casimir invariants and vector operators. Also, note

$$K^{(p+1)} = X^{\lambda} S^{(p)}_{\lambda} = S^{(p)}_{\lambda} X^{\lambda}, \tag{2.15}$$

which is the analog of (2.6). The proof is simple. Let X_{μ} be representation matrices of a generic representation $\{\lambda\}$. Consider a tensor product space $\{\lambda\} \otimes \{\lambda_0\}$, and set

$$Y_{\mu} = X_{\mu} \otimes E_0 + E \otimes x_{\mu}, \tag{2.16}$$

where E and E_0 are identity matrices in the spaces $\{\lambda\}$ $\{\lambda_0\}$, respectively. Then Y_{μ} are easily seen to satisfy

$$[Y_{\mu}, Y_{\nu}] = C^{\lambda}_{\mu\nu} Y_{\lambda} \tag{2.17}$$

so that Y_{μ} are the representation matrices of L in the product space $\{\lambda\} \otimes \{\lambda_0\}$. Moreover, let us set

$$Q = X_{\mu} \otimes x^{\mu} = \overline{g}^{\mu\nu} X_{\mu} \otimes x_{\nu}. \tag{2.18}$$

Then we find

$$[Y_{\mu}, Q] = 0. \tag{2.19}$$

When we define the pth power Q^{p} $(p \geq 0)$ by

$$Q^{p+1} = QQ^{p},$$

$$Q^{0} = E \otimes E_0 \equiv \tilde{E} \quad (p = 0), \tag{2.20}$$

this leads to

$$[Y_{\mu}, Q^{p}] = 0, \tag{2.21a}$$

$$[Y_{\mu}, Q^{p}(E \otimes x_{\nu})] = C^{\lambda}_{\mu\nu} Q^{p}(E \otimes x_{\lambda}). \tag{2.21b}$$

If we now take the partial trace $\mathrm{tr}_{\{\lambda_0\}}$ of both sides of (2.21) with respect to the reference representation $\{\lambda_0\}$ (but *not* with respect to the generic space $\{\lambda\}$), we obtain the desired relations

$$[X_{\mu}, K^{(p)}] = 0, \tag{2.22a}$$

$$[X_{\mu}, S^{(p)}_{\nu}] = C^{\lambda}_{\mu\nu} S^{(p)}_{\lambda}. \tag{2.22b}$$

We notice that we can define $K^{(0)}$ and $S^{(0)}_{\nu}$ for the case $p = 0$ by this procedure to be

$$K^{(0)} = d_0 E, \tag{2.23a}$$

$$S^{(0)}_{\mu} = E\,\mathrm{tr}x_{\mu}, \tag{2.23b}$$

where d_0 is the dimension of the reference space $\{\lambda_0\}$. If L is semisimple, then $S^{(0)}_{\mu} = 0$ identically. We can re-

peat essentially the same argument when X_μ are basis of L.

Let us now prescribe a procedure to enable us to compute eigenvalues of $K^{(p)}$. To this end, let us assume that L is either semisimple or a classical Lie algebra with symmetric $\overline{g}_{\mu\nu}$ which can be derived from a compact Lie group. For the latter case, we assume also that all representations such as $\{\lambda_0\}$ and $\{\lambda\}$ refer to infinitesimal representations of the corresponding classical compact Lie group. Then, any representation of L is fully reducible by Weyl's theorem, so that we can decompose the product $\{\lambda\}\otimes\{\lambda_0\}$ as a direct sum of irreducible representations $\{\hat{\lambda}\}$:

$$\{\lambda\}\otimes\{\lambda_0\}=\sum_{\hat{\lambda}}\oplus C(\hat{\lambda};\lambda,\lambda_0)\{\hat{\lambda}\}, \tag{2.24}$$

where the summation is over all inequivalent irreducible representations $\{\hat{\lambda}\}$ of L, and the constant $C(\hat{\lambda};\lambda,\lambda_0)$ is the multiplicity of the representation $\{\hat{\lambda}\}$ contained in the reduction of the product $\{\lambda\}\otimes\{\lambda_0\}$. It is convenient for our purpose to proceed as follows. For a given $\{\lambda\}$, let us enumerate all irreducible representations $\{\hat{\lambda}\}$ appearing nontrivially [i.e., those with $C(\hat{\lambda};\lambda,\lambda_0)\neq 0$] in the right side of (2.24) as $\{\lambda_1\}$, $\{\lambda_2\}$, ..., $\{\lambda_N\}$. If the multiplicity of $\{\hat{\lambda}\}$ is more than one, then we count the same irreducible representation more than once as many times as the multiplicity, labelling them by different indices. In this way, we rewrite (2.24) as

$$\{\lambda\}\otimes\{\lambda_0\}=\sum_{j=1}^{N}\oplus\{\lambda_j\}, \tag{2.25}$$

$$N=\sum_{\hat{\lambda}}C(\hat{\lambda};\lambda,\lambda_0). \tag{2.26}$$

Note that N depends in general upon $\{\lambda\}$ and $\{\lambda_0\}$, and that two representations $\{\lambda_j\}$ and $\{\lambda_k\}$ for $j\neq k$ in (2.25) could be equivalent to each other.

Let P_j and $X_\mu^{(j)}$ $(j=1,2,\ldots,N)$ be respectively the projection operators and the representation matrices of X_μ in the jth irreducible representation space $\{\lambda_j\}$, appearing in the right side of (2.25). We have then, of course,

$$Y_\mu=\sum_{j=1}^{N}X_\mu^{(j)}P_j, \tag{2.27a}$$

$$P_jP_k=\delta_{jk}P_j, \tag{2.27b}$$

$$\sum_{j=1}^{N}P_j=E\otimes E_0\equiv\tilde{E}. \tag{2.27c}$$

In a suitably chosen matrix basis, they have forms of

$$Y_\mu=\begin{pmatrix}X_\mu^{(1)} & & & 0\\ & X_\mu^{(2)} & & \\ & & \ddots & \\ 0 & & & X_\mu^{(N)}\end{pmatrix}, \tag{2.28a}$$

$$P_j=\begin{pmatrix}0 & & & & 0\\ & 0 & & & \\ & & \ddots & & \\ & & & E_j & \\ & & & & \ddots \\ 0 & & & & & 0\end{pmatrix}, \tag{2.28b}$$

where E_j is the unit matrix in the jth subspace.

Let us denote eigenvalues of the second order Casimir invariants in an irreducible representation $\{\lambda'\}$ by $\tilde{I}_2(\lambda')$. Then, of course,

$$\overline{g}^{\mu\nu}X_\mu^{(j)}X_\nu^{(j)}=\tilde{I}_2(\lambda_j)E_j,\quad j=1,2,\cdots,N, \tag{2.29a}$$

$$\overline{g}^{\mu\nu}X_\mu X_\nu=\tilde{I}_2(\lambda)E, \tag{2.29b}$$

$$\overline{g}^{\mu\nu}x_\mu x_\nu=\tilde{I}_2(\lambda_0)E_0. \tag{2.29c}$$

When we rewrite Q as

$$Q=\overline{g}^{\mu\nu}X_\mu\otimes x_\nu=\tfrac{1}{2}\{\overline{g}^{\mu\nu}Y_\mu Y_\nu - (\overline{g}^{\mu\nu}X_\mu X_\nu)\otimes E_0 - E\otimes(\overline{g}^{\mu\nu}x_\mu x_\nu)\}$$

and note Eqs. (2.27), (2.28), and (2.29), this leads to

$$Q^p=\sum_{j=1}^{N}(\xi_j)^pP_j,\quad p=0,1,2,\cdots, \tag{2.30}$$

where we have set for simplicity

$$\xi_j=\tfrac{1}{2}\{\tilde{I}_2(\lambda_j)-\tilde{I}_2(\lambda)-\tilde{I}_2(\lambda_0)\}. \tag{2.31}$$

Let $d(\lambda)$, $d(\lambda_0)$, and $d(\lambda_j)$ be dimensions of the representation spaces $\{\lambda\}$, $\{\lambda_0\}$, and $\{\lambda_j\}$ $(j=1,2,\ldots,N)$ and take the full trace of both sides of (2.30) with respect to $\{\lambda\}\otimes\{\lambda_0\}$. Then, we find

$$I_p(\lambda)=\frac{1}{d(\lambda)}\sum_{j=1}^{N}(\xi_j)^pd(\lambda_j),\quad p\geq 0, \tag{2.32}$$

where $I_p(\lambda)$ is the eigenvalue of the pth Casimir invariant $K^{(p)}$ defined by (2.13) in the generic representation $\{\lambda\}$, i.e.,

$$K^{(p)}=I_p(\lambda)E. \tag{2.33}$$

Since the eigenvalue $I_2(\lambda_j)$ of the second-order Casimir invariant $K^{(2)}$ is easy to compute, the formula (2.32) tells us that we could compute eigenvalues of all higher order Casimir invariants, once the Clebsch–Gordon series (2.25) is known. If two irreducible representations $\{\lambda_j\}$ and $\{\lambda_k\}$ for $j\neq k$ are equivalent, then contributions from two such terms in (2.32) give the same result. Therefore, returning to the original notation, we can rewrite (2.32) as

$$I_p(\lambda)=\sum_{\hat{\lambda}}C(\hat{\lambda};\lambda,\lambda_0)[\phi(\hat{\lambda};\lambda,\lambda_0)]^p\frac{d(\hat{\lambda})}{d(\lambda)}, \tag{2.34}$$

$$\phi(\hat{\lambda};\lambda,\lambda_0)=\tfrac{1}{2}[\tilde{I}_2(\hat{\lambda})-\tilde{I}_2(\lambda)-\tilde{I}_2(\lambda_0)], \tag{2.35}$$

where the summation in (2.34) runs over all inequivalent irreducible representations $\{\hat{\lambda}\}$.

Hereafter, let us restrict ourselves to the case that L is a simple Lie algebra, and we choose the reference representation $\{\lambda_0\}$ to be an arbitrary but fixed nontrivial irreducible representation of L. Then the multiplicity factor $C(\hat{\lambda};\lambda,\lambda_0)$ can be computed in principle by means of the Kostant–Steinberg formula.[23] However, for the present purpose, another formula, essentially due to Weyl[24] is more convenient. We follow Racah[25] in describing the formula. Also, because of the remark made just before (2.11), we can set hereafter $\overline{g}_{\mu\nu}=g_{\mu\nu}$, $\overline{g}^{\mu\nu}=g^{\mu\nu}$, and $\tilde{I}_2(\lambda)=I_2(\lambda)$.

Let Λ_0, Λ, and $\hat{\Lambda}$ be the highest weights of the irreducible representations $\{\lambda_0\}$, $\{\lambda\}$, and $\{\hat{\lambda}\}$, respectively, and set

$$K_0=\Lambda_0+\delta,\quad K=\Lambda+\delta,\quad \hat{K}=\hat{\Lambda}+\delta \tag{2.36}$$

where δ is given by

$$\delta=\tfrac{1}{2}\sum_{\alpha>0}\alpha \tag{2.37}$$

in terms of positive roots α. Let $\gamma_0(M)$ be the multiplicity of the weight M of the reference representation $\{\lambda_0\}$. Then, Weyl's formula[25] is expressed as

$$C(\hat{\lambda};\lambda,\lambda_0)=\sum_S(\det S)\gamma_0(S\hat{K}-K), \tag{2.38}$$

where the summation is over all Weyl-reflection operations S, and $\det S$ is the parity of S. Moreover, the dimensional formula of Weyl reads as

$$d(\hat{\lambda})=\prod_{\alpha>0}(\hat{K},\alpha)/(\delta,\alpha). \tag{2.39}$$

Finally, the eigenvalue of the second order Casimir invariant is readily computed to be

$$I_2(\hat{\lambda})=\theta[(\hat{K},\hat{K})-(\delta,\delta)], \tag{2.40}$$

$$\theta=\frac{b}{d(\lambda_0)}[(K_0,K_0)-(\delta,\delta)]^{-1}. \tag{2.41}$$

Inserting all these expressions into (2.34), we find

$$I_p(\lambda)=\sum_{\hat{K}}\sum_S(\det S)\gamma_0(S\hat{K}-K)[\phi(\hat{K};K,K_0)]^p \times\prod_{\alpha>0}(\hat{K},\alpha)/(K,\alpha), \tag{2.42}$$

$$\phi(\hat{K};K,K_0)=\tfrac{1}{2}\theta[(\hat{K},\hat{K})+(\delta,\delta)-(K,K)-(K_0,K_0)]. \tag{2.43}$$

We remark that conditions for $p=0$ and $p=1$, i.e.,

$$I_0(\lambda)=d(\lambda_0),\quad I_1(\lambda)=0 \tag{2.44}$$

as well as the consistency requirement for $p=2$, i.e.,

$$I_2(\lambda)=\theta[(K,K)-(\delta,\delta)]$$

impose some constraints for the weight $\gamma_0(M)$, if we set $p=0,1$, and 2 in (2.42). Some of such constraints for $\gamma_0(M)$ are already found and discussed by Racah.[25]

Since the multiplicity factor $\gamma_0(M)$ can be computed from Freudental's formula,[26] we can compute all $I_p(\lambda)$ for any p and any $\{\lambda\}$. Unfortunately, we cannot determine from our formula which choice of the reference space $\{\lambda_0\}$ and which values of p could give the desired n algebraically independent Casimir invariants. For this, we have to utilize the previously known results of Refs. 2–10. However, for algebras A_n, B_n, C_n, D_n, and G_2, we can rather easily compute eigenvalues $I_p(\lambda)$ in closed terms from (2.42), if we choose $\{\lambda_0\}$ to be one of the fundamental representations of these algebras. In this ad hoc way, we find all n algebraically independent Casimir invariants for these cases in Secs. 4–8. In this connection, we need not consider $I_p(\lambda)$ with large values of p, as we shall see below.

Let us classify N irreducible representations $\{\lambda_j\}$ appearing in (2.25) by introducing the following equivalence relation. We define that two irreducible representations $\{\lambda_j\}$ and $\{\lambda_k\}$ are related to each other if and only if, we have $I_2(\lambda_j)=I_2(\lambda_k)$ (or $\xi_j=\xi_k$ equivalently), i.e., if and only if they have the same values for the second-order Casimir invariants. Let us suppose that we have $N-\Delta$ inequivalent classes by this equivalence relation. We shall call Δ ($\Delta\geq 0$) be the deficient index, and set

$$\tilde{P}_\alpha=\sum_j P_j,\quad \xi_j=\tilde{\xi}_\alpha \tag{2.45}$$

for $\alpha=1,2,\ldots,N-\Delta$, where the summation is over all j such that $\{\lambda_j\}$ belongs to the same αth equivalent class and where $\tilde{\xi}_\alpha$ is the class function defined by its representative ξ_j of the class. Then, $\tilde{P}_\alpha$ is evidently the projection operator for a space spanned by members of the αth equivalent class and we have

$$\tilde{P}_\alpha\tilde{P}_\beta=\delta_{\alpha\beta}\tilde{P}_\alpha, \tag{2.46}$$

$$\sum_{\alpha=1}^{N-\Delta}\tilde{P}_\alpha=\tilde{E}=E\otimes E_0, \tag{2.47}$$

$$Q=\sum_{\alpha=1}^{N-\Delta}\tilde{\xi}_\alpha\tilde{P}_\alpha. \tag{2.48}$$

As a result, Q satisfies a polynomial equation of order $N-\Delta$,

$$\prod_{\alpha=1}^{N-\Delta}\{Q-\tilde{\xi}_\alpha\tilde{E}\}=0, \tag{2.49}$$

which may be rewritten as

$$Q^{N-\Delta}=\sum_{k=0}^{N-\Delta-1}a_kQ^k. \tag{2.50}$$

Note that a_k ($k=0,1,\ldots,N-\Delta-1$) are polynomials of $\xi_1,\ldots,\xi_N$. Repeated uses of (2.50) lead to

$$Q^p=\sum_{k=0}^{N-\Delta-1}a_k^{(p)}Q^k \tag{2.51}$$

for all $p\geq N-\Delta$, where $a_k^{(p)}$ are polynomials again of $\xi_1,\ldots,\xi_N$. Taking the trace of both sides of (2.51) with respect to the full space $\{\lambda\}\otimes\{\lambda_0\}$, this gives the desired formula

$$I_p(\lambda)=\sum_{k=0}^{N-\Delta-1}a_k^{(p)}I_k(\lambda),\quad p\geq N-\Delta, \tag{2.52}$$

which expresses $I_p(\lambda)$ for large values of p in terms of $I_k(\lambda)$ with smaller k and of $\xi_1,\ldots,\xi_N$.

Next, we shall show that Eq. (2.50) leads to polynomial identities among infinitesimal generators X_μ in the given representation $\{\lambda\}$. To this end, we shall take (α,β) matrix element of both sides of (2.50) with respect to the reference space $\{\lambda_0\}$ (but *not* on $\{\lambda\}$), then it gives

$$(Q^{N-\Delta})_{\alpha\beta}=\sum_{k=0}^{N-\Delta-1}a_k(Q^k)_{\alpha\beta} \tag{2.53}$$

for $\alpha,\beta=1,2,\ldots,\ d_0=d(\lambda_0)$. Since we have

$$Q_{\alpha\beta}=(X_\lambda\otimes x^\lambda)_{\alpha\beta}=(x^\lambda)_{\alpha\beta}X_\lambda, \tag{2.54}$$

$$(Q^{k+1})_{\alpha\beta}=\sum_{\gamma=1}^{d_0}Q_{\alpha\gamma}(Q^k)_{\gamma\beta},$$

Eq. (2.53) implies the desired polynomial identity among X_λ of order $N-\Delta$. As we shall observe in Sec. 4, this reproduces the minimum polynomial identities[16] discovered for the algebra A_n.

In ending this section, we can express $\tilde{P}_\alpha$ in terms of

Q by

$$\tilde{P}_\alpha = \prod_{\beta=1}^{N-\Delta}{}' \frac{Q-\tilde{\xi}_\beta \tilde{E}}{\tilde{\xi}_\alpha - \tilde{\xi}_\beta} \tag{2.55}$$

for $\alpha = 1,2,\ldots,N-\Delta$, where the product on β omits the singular point $\beta=\alpha$. Especially, if we have $\Delta=0$, then P_j itself for all $j=1,2,\ldots,N$ can be expressed as a polynominal of Q of the order $N-1$. This fact will be utilized in the next section.

3. VECTOR OPERATORS AND VECTOR IDENTITIES

Here in this section, we assume L to be a simple Lie algebra of rank n over the complex number field. Let $\Lambda_1,\Lambda_2,\ldots,\Lambda_n$ be the n fundamental weight system of L, so that the highest weight Λ of any $\{\lambda\}$ is given by

$$\Lambda = \sum_{j=1}^{n} m_j \Lambda_j \tag{3.1}$$

where m_j are nonnegative integers specifying the irreducible representation $\{\lambda\}$. We shall prove in the Appendix A that the number $n_V(\lambda)$ of all linearly independent vector operators in $\{\lambda\}$ is given by

$$n_V(\lambda) = n - n_0(\lambda) \tag{3.2}$$

where $n_0(\lambda)$ is the number of m_j's which are equal to zero. In other words, (3.2) states that $n_V(\lambda)$ is equal to the number of m_j's which are not zero. Especially, we find $n_V(\lambda) \leq n$, which reproduces the result found by Michel.[27]

We have found in the previous section that $S_\mu^{(p)}$ defined by (2.14) are vector operators for all $p \geq 1$ and for all $\{\lambda_0\}$. Therefore, they must satisfy some linear identity equations among themselves because of the theorem mentioned above. It is easy to find some examples of such relations. Multiply $E \otimes x_\mu$ to both sides of (2.51) and take the partial trace with respect to the reference space $\{\lambda_0\}$. This immediately leads to the desired linear relations for all $p \geq N-\Delta$,

$$S_\mu^{(p)} = \sum_{k=0}^{N-\Delta-1} a_k^{(p)} S_\mu^{(k)}. \tag{3.3}$$

Moreover, noting

$$S_\mu^{(0)} = 0, \quad S_\mu^{(1)} = X_\mu, \tag{3.4}$$

this shows that all vector operators $S_\mu^{(p)}$ for a given $\{\lambda_0\}$ are expressible in terms of at most $N-\Delta-1$ vector operators $S_\mu^{(k)}$ with $N-\Delta-1 \geq k \geq 1$.

Next, we would like to find explicit functional forms of all linearly independent vector operators in a given irreducible representation $\{\lambda\}$. To this end, we assume that for a given $\{\lambda\}$, we can find a reference representation $\{\lambda_0\}$ such that all irreducible representations $\{\lambda_j\}$ appearing in the right side of (2.25) for the decomposition of $\{\lambda\} \otimes \{\lambda_0\}$ are inequivalent to each other. In other words, we choose $\{\lambda_0\}$ in such a way that the multiplicity factor $C(\hat{\lambda};\lambda,\lambda_0)$ appearing in (2.24) is at most one whenever this is possible. We call this condition the nondegeneracy condition. As we will see in the next sections, this condition is satisfied for generic irreducible representation $\{\lambda\}$ of the algebras A_n, B_n, C_n, D_n, and G_2 if we choose the reference representation $\{\lambda_0\}$ to be one of the fundamental representations of these algebras. Suppose that $\{\lambda_0\}$ satisfy the nondegeneracy condition, and let T_μ be an arbitrary vector operator in the representation $\{\lambda\}$. Setting

$$T = T_\mu \otimes x^\mu = g^{\mu\nu} T_\mu \otimes x_\nu, \tag{3.5}$$

it is easy to verify

$$[Y_\mu, T] = 0. \tag{3.6}$$

In block-matrix form analogous to (2.28), we can express T as

$$T = \begin{pmatrix} T_{11}, T_{12}, \cdots, T_{1N} \\ T_{21}, T_{22}, \cdots, T_{2N} \\ \cdots\cdots\cdots\cdots \\ T_{N1}, T_{N2}, \cdots, T_{NN} \end{pmatrix}, \tag{3.7}$$

where T_{jk} is $d(\lambda_j) \times d(\lambda_k)$ matrix.

Equation (3.6) is now rewritten in terms of irreducible components as

$$X_\mu^{(j)} T_{jk} = T_{jk} X_\mu^{(k)} \quad (j,k=1,2,\cdots,N). \tag{3.8}$$

But the nondegeneracy ansatz implies that $X_\mu^{(j)}$ is inequivalent to $X_\mu^{(k)}$ whenever $j \neq k$. Therefore, the Schur's lemma applied to (3.8) enables us to conclude

$$T_{jk} = C_j \delta_{jk} E_j \tag{3.9}$$

where C_j $(j=1,2,\ldots,N)$ are some constants. Inserting this into (3.7), we find

$$T = T_\mu \otimes x^\mu = \sum_{j=1}^{N} C_j P_j. \tag{3.10}$$

By multiplying $E \otimes x_\mu$ by both sides of (3.10) and taking the partial trace with respect to the space $\{\lambda_0\}$, this gives

$$T_\mu = \sum_{j=1}^{N} C_j R_\mu^{(j)}, \tag{3.11}$$

$$R_\mu^{(j)} = \mathrm{tr}_{\{\lambda_0\}}\{P_j (E \otimes x_\mu)\}. \tag{3.12}$$

Note that $R_\mu^{(j)}$ $(j=1,2,\ldots,N)$ do not depend upon T_μ at all. Also, it is easy to prove that they are vector operators. Therefore, (3.11) implies that any vector operator in the given representation $\{\lambda\}$ is a linear combination of at most N vector operators $R_\mu^{(j)}$ $(j=1,2,\ldots,N)$ defined by (3.12).

The next task is to determine more explicit forms of $R_\mu^{(j)}$. First, let us assume that the deficiency index Δ is zero, i.e., that we have $I_2(\lambda_j) \neq I_2(\lambda_k)$ for $j \neq k$, $j,k = 1,2,\ldots,N$. In that case, (2.55) is rewritten as

$$P_j = \prod_{k=1}^{N}{}' \frac{Q-\xi_k \tilde{E}}{\xi_j - \xi_k}, \quad j=1,2,\cdots,N. \tag{3.13}$$

Expanding the product in powers of Q, inserting this into (3.12), we see that any T_μ must be expressed as

$$T_\mu = \sum_{p=1}^{N-1} C'_p S_\mu^{(p)}. \tag{3.14}$$

This implies that any vector operator in $\{\lambda\}$ is a linear combination of $N-1$ special vector operators $S_\mu^{(p)}$ $(1 \leq p \leq N-1)$ defined in (2.14) when the reference representation $\{\lambda_0\}$ is now chosen to satisfy both nondegeneracy condition and $\Delta=0$. Further among these $N-1$ vector operators, only $n_V(\lambda)$ of them are in reality linearly independent by the theorem quoted in the beginning.

When we have $\Delta \neq 0$, then (3.13) is not correct. However, Eq. (2.55) still determines P_k $(1 \leq k \leq N-\Delta)$ as polynomials of Q, so that we can still evaluate $R_\mu^{(j)}$ for $1 \leq j \leq N-\Delta$ in terms of $S_\mu^{(p)}$ with $1 \leq p \leq N-\Delta-1$ as before. But the remaining $R_\mu^{(j)}$ $(N-\Delta+1 \leq j \leq N)$ cannot be computed in this way. In such cases, we proceed as follows. Instead of Q, let us consider higher-order Casimir invariants $\tilde{Q}_p$ in $\{\lambda\}\otimes\{\lambda_0\}$;

$$\tilde{Q}_p = g^{\mu_1\mu_2\cdots\mu_p} Y_{\mu_1} Y_{\mu_2} \cdots Y_{\mu_p}. \tag{3.15}$$

In view of (2.27) and (2.13), we can decompose $\tilde{Q}_p$ in terms of irreducible components

$$\tilde{Q}_p = \sum_{j=1}^{N} I_p(\lambda_j) P_j, \tag{3.16}$$

where $I_p(\lambda_j)$ is the eigenvalue of the pth Casimir invariant $K^{(p)}$ in the jth irreducible representation $\{\lambda_j\}$. If, for some $p \geq 3$, $I_p(\lambda_j)$ are all distinct, i.e., if we have

$$I_p(\lambda_j) \neq I_p(\lambda_k), \quad \text{if } j \neq k,$$

then we can express P_j in a form

$$P_j = \prod_{k=1}^{N}{}' \frac{\tilde{Q}_p - I_p(\lambda_k)\tilde{E}}{I_p(\lambda_j) - I_p(\lambda_k)}. \tag{3.17}$$

Therefore, we can express $R_\mu^{(j)}$ as a linear combination of vector operators

$$\tilde{S}_\mu^{(j)} = \mathrm{tr}_{\{\lambda_0\}}[(\tilde{Q}_p)^j (E \otimes x_\mu)]. \tag{3.18}$$

As an illustration, we note that $\tilde{S}_\mu^{(1)}$, for example, is furthermore expressible as a linear combination of the following vector operators:

$$g_{\mu\nu_1\cdots\nu_k} g^{\nu_1\cdots\nu_k\mu_1\cdots\mu_l} X_{\mu_1} X_{\mu_2} \cdots X_{\mu_l}. \tag{3.19}$$

In the next section, we choose the reference representation $\{\lambda_0\}$ to be one of the fundamental representations. Then the nondegeneracy condition is satisfied for any generic irreducible representation $\{\lambda\}$ of the algebra A_n, B_n, C_n, D_n and G_2. Moreover, we have $\Delta = 0$ for all $\{\lambda\}$ in A_n, B_n, C_n, and G_2. Also, for the algebra D_n, we have $\Delta = 0$ in general for the majority of irreducible representations $\{\lambda\}$. However, for some special representations, we could have $\Delta = 1$, and this causes some slight complication in finding explicit forms of all vector operators for this case, as we may see in Sec. 7.

In ending this section, we note the following. Although the algebra L has been assumed in this section to be simple, the same reasoning is equally applicable to the case when L is classical, provided that the product $\{\lambda\}\otimes\{\lambda_0\}$ is fully reducible as in Eq. (2.24) with the nondegeneracy condition. However, in this instance, $S_\mu^{(0)} = E\mathrm{tr}x_\mu$ is not necessarily zero and we have to take into account this extra term corresponding to $p=0$ in all equations such as (3.14). A similar remark applies to the Casimir invariant $I_1(\lambda)$, i.e., $I_1(\lambda) \neq 0$ in general for such cases.

4. ALGEBRAS A_n AND LIE ALGEBRA OF THE $U(n)$

The Lie algebra of the $(n+1)$-dimensional unitary group $U(n+1)$ is characterized by generators $A_{\mu\nu}$ satisfying

$$[A_{\mu\nu}, A_{\alpha\beta}] = \delta_{\nu\alpha} A_{\mu\beta} - \delta_{\mu\beta} A_{\alpha\nu} \tag{4.1}$$

for all $\mu, \nu, \alpha, \beta = 1, 2, \ldots, n+1$. The Lie algebra A_n corresponding to the $SU(n+1)$ group can be obtained from $A_{\mu\nu}$ by

$$B_{\mu\nu} = A_{\mu\nu} - \frac{1}{n+1}\delta_{\mu\nu}\sum_{\lambda=1}^{n+1} A_{\lambda\lambda}. \tag{4.2}$$

Note that $B_{\mu\nu}$ satisfies the same commutation relation (4.1) together with traceless condition

$$\sum_{\lambda=1}^{n+1} B_{\lambda\lambda} = 0. \tag{4.3}$$

Any irreducible representation $\{\lambda\}$ of the $U(n+1)$ group is characterized by $n+1$ integers satisfying[28]

$$f_1 \geq f_2 \geq \cdots \geq f_n \geq f_{n+1}. \tag{4.4}$$

Corresponding to (4.2) and (4.3), any irreducible representation of the A_n is specified by $n+1$ rational numbers $\tilde{f}_j$ given by

$$\tilde{f}_j = f_j - \frac{1}{n+1}\sum_{\lambda=1}^{n+1} f_\lambda. \tag{4.5}$$

Therefore, we notice

$$\tilde{f}_1 \geq \tilde{f}_2 \geq \cdots \geq \tilde{f}_n \geq \tilde{f}_{n+1}, \tag{4.6a}$$

$$\sum_{j=1}^{n+1} \tilde{f}_j = 0. \tag{4.6b}$$

Also, m_j's defined in (3.1) are given by

$$m_j = \tilde{f}_j - \tilde{f}_{j+1} = f_j - f_{j+1} \tag{4.6c}$$

for $j = 1, 2, \ldots, n$. Since all relations to be given in this section for the generators $A_{\mu\nu}$ are equally valid for $B_{\mu\nu}$ by substituting f_j by $\tilde{f}_j$, we will consider only the former case here.

It is convenient to define l_j by

$$l_j = f_j + n + 1 - j, \quad n+1 \geq j \geq 1. \tag{4.7}$$

Then the dimension of the representation $\{\lambda\}$ specified by the signature $(f_1, f_2, \ldots, f_{n+1})$ is given by

$$d(\lambda) = \frac{\prod_{j<k}^{n+1}(l_j - l_k)}{1!\,2!\cdots n!}. \tag{4.8}$$

Although the Lie algebra of the $U(n+1)$ group is not semisimple, it is nevertheless classical. Indeed, we can choose the reference representation $\{\lambda_0\}$ to be either

$$\{\lambda_0\} = (1, 0, 0, \cdots, 0) \tag{4.9a}$$

or

$$\{\lambda_0\} = (0, 0, \cdots, 0, -1). \tag{4.9b}$$

Both choices not only assure the existence of $g^{\mu\nu}$ but also satisfy the nondegeneracy condition. Here in this note, we choose the case (4.9b), where (α, β) matrix element of $x_{\mu\nu}$ is given by

$$(x_{\mu\nu})_{\alpha\beta} = -\delta_{\mu\beta}\delta_{\nu\alpha} \tag{4.10}$$

for $\mu, \nu, \alpha, \beta = 1, 2, \ldots, n+1$. Note that we have

$$d(\lambda_0) = n+1, \tag{4.11a}$$

$$g^{(\mu\nu),(\alpha\beta)} = g_{(\mu\nu),(\alpha\beta)} = \delta_{\nu\alpha}\delta_{\mu\beta}. \tag{4.11b}$$

Then, the product is reduced to

$$\{\lambda\}\otimes\{\lambda_0\}=(f_1,f_2,\cdots,f_{n+1})\otimes(0,0,\cdots,0,-1)$$
$$=(f_1-1,f_2,\cdots,f_{n+1})\oplus(f_1,f_2-1,f_3,\cdots,f_{n+1})\oplus\cdots\oplus$$
$$\oplus(f_1,f_2,\cdots,f_n,f_{n+1}-1) \tag{4.12}$$

from the Weyl's formula.[24] In the right side of (4.12), we simply drop any term,[29] if it corresponds to unnatural ordering, i.e., if its signature factors do not satisfy the natural ordering (4.4). Then we may easily find $N=n+1-n_0(\lambda)$, where $n_0(\lambda)$ is the number of m_j's which are zero. We notice that N is also identical to the number of nonredundant f_j's.[29]

The eigenvalue of the second-order Casimir invariant can be easily computed to be

$$I_2(\lambda)=\sum_{j=1}^{n+1} l_j(l_j-n)+\tfrac{1}{6}n(n-1)(n+1). \tag{4.13}$$

From this, we easily find that the deficiency index Δ is always zero for any generic representation $\{\lambda\}$. The higher-order Casimir invariants $I_p(\lambda)$ can be found now easily from (2.32) to be

$$I_p(\lambda)=(-1)^p\sum_{j=1}^{n+1}(l_j)^p\prod_{k=1}^{n+1}{}'\frac{l_k+1-l_j}{l_k-l_j}, \tag{4.14}$$

where the product omits the singular point $k=j$. The formula (4.14) reproduces the result of Ref. 10 and 13. In spite of its apparent fractional form, (4.14) can be proved[10] to be a symmetric polynomial of $l_1,l_2,\ldots,l_{n+1}$ of degree p. Therefore, we can specify any generic representation $\{\lambda\}$ by $n+1$ Casimir invariants $K^{(p)}$ $(1\leq p\leq n+1)$.

When we take (α,β) matrix element of Q with respect to the space $\{\lambda_0\}$ as in Eq. (2.54), it leads to

$$Q_{\alpha\beta}=(X_B\otimes x^B)_{\alpha\beta}=-A_{\alpha\beta} \tag{4.15}$$

for $\alpha,\beta=1,2,\ldots,n+1$ with $B=(\mu,\nu)$. Then, we can easily verify that the corresponding identity (2.53) or equivalently (2.49) with $\Delta=0$, exactly reproduces the minimum polynomial vector identity found previously.[16] Since we have $\Delta=0$, all vector operators are linear combinations of N special vector operators $S_B^{(p)}$ $(0\leq p\leq N)$ with $B=(\mu,\nu)$. This conclusion again agrees with that reached in Ref. 16. For the algebra A_n, we only need $N-1=n-n_0(\lambda)$ of $S_B^{(p)}$ in agreement with the theorem of Sec. 3.

In order to give an example of more general tensor identity equations, let us consider the case of the completely symmetric representation $\{\lambda\}=(f,0,0,\ldots,0)$, i.e., $f_1=f$ but $f_j=0$ $(j\neq 1)$. In this case, it is convenient to choose $\{\lambda_0\}$ to be

$$\{\lambda_0\}=(1,1,0,0,\cdots,0),$$

i.e., $f_1=f_2=1$, but $f_j=0$ $(j\geq 3)$. Then the Clebsch–Gordon series is now[30]

$$\{\lambda\}\otimes\{\lambda_0\}=(f,0,0,\cdots,0)\otimes(1,1,0,0,\cdots,0)$$
$$=(f+1,1,0,\cdots,0)\oplus(f,1,1,0,\cdots,0) \tag{4.16}$$

so that we have $N=2$ with $\xi_1=+f$ and $\xi_2=-2$. Then (2.53) is rewritten as

$$(A_{\mu\nu}+\delta_{\mu\nu}E)A_{\alpha\beta}=(A_{\mu\beta}+\delta_{\mu\beta}E)A_{\alpha\nu}, \tag{4.17}$$

which has been discovered already by many authors.[18,19,16]

5. ALGEBRA B_n

This corresponds to the orthogonal group $O(2n+1)$, where the generators X_A with $A=(\mu,\nu)$, $\mu,\nu=1,2,\ldots,2n+1$, satisfy

$$X_{\mu\nu}=-X_{\nu\mu}, \tag{5.1a}$$

$$[X_{\mu\nu},X_{\alpha\beta}]=\delta_{\nu\alpha}X_{\mu\beta}+\delta_{\mu\beta}X_{\nu\alpha}-\delta_{\mu\alpha}X_{\nu\beta}-\delta_{\nu\beta}X_{\mu\alpha} \tag{5.1b}$$

in a non-Cartan form. Instead of m_j's given in (3.1), it is more convenient to consider $f_1,f_2,\ldots,f_n$ by

$$f_j=m_j+m_{j+1}+\cdots+m_{n-1}+\tfrac{1}{2}m_n \quad (j\neq n),$$
$$f_n=\tfrac{1}{2}m_n \quad (j=n). \tag{5.2}$$

Noting $m_j\geq 0$, this gives the natural ordering

$$f_1\geq f_2\geq\cdots\geq f_n\geq 0. \tag{5.3}$$

These f_j must be simultaneously either all integers or all half-integers. The latter corresponds to the spinor representations, while the former case are tensor representations. It is often more convenient to define

$$l_j=f_j+n-j+\tfrac{1}{2} \quad (n\geq j\geq 1), \tag{5.4}$$

then the dimension of the representation $\{\lambda\}$ is

$$d(\lambda)=\frac{2^n}{1!3!\cdots(2n-1)!}\prod_{j=1}^{n}l_j\prod_{j<k}^{n}[(l_j)^2-(l_k)^2]. \tag{5.5}$$

We choose the reference representation $\{\lambda_0\}$ to be $\{\Lambda_1\}$, i.e.,

$$\{\lambda_0\}=(1,0,0,\cdots,0) \tag{5.6}$$

with $f_1=1$, and $f_j=0$ $(j\geq 2)$. The (α,β) matrix element of the generator $x_{\mu\nu}=-x_{\nu\mu}$ in $\{\lambda_0\}$ is given by

$$(x_{\mu\nu})_{\alpha\beta}=\delta_{\mu\alpha}\delta_{\nu\beta}-\delta_{\mu\beta}\delta_{\nu\alpha}. \tag{5.7}$$

We now have

$$d(\lambda_0)=2n+1, \tag{5.8}$$

$$g_{(\mu\nu),(\alpha\beta)}=4g^{(\mu\nu),(\alpha\beta)}=2[\delta_{\mu\beta}\delta_{\nu\alpha}-\delta_{\mu\alpha}\delta_{\nu\beta}]. \tag{5.9}$$

We compute the eigenvalue of the second-order Casimir invariant to be

$$I_2(\lambda)=\tfrac{1}{2}\{\sum_{j=1}^{n}(l_j)^2-\tfrac{1}{12}n(2n+1)(2n-1)\}. \tag{5.10}$$

The product decomposition rule is[31]

$$\{\lambda\}\otimes\{\lambda_0\}=(f_1,f_2,\cdots,f_n)\otimes(1,0,\cdots,0)$$
$$=(f_1+1,f_2,\cdots,f_n)\oplus(f_1,f_2+1,f_3,\cdots,f_n)\oplus\cdots$$
$$\oplus(f_1,\cdots,f_{n-1},f_n+1)\oplus(f_1-1,f_2,\cdots,f_n)$$
$$\oplus(f_1,f_2-1,f_3,\cdots,f_n)\oplus\cdots\oplus(f_1,\cdots,f_{n-1},f_n-1)$$
$$\oplus(f_1,f_2,\cdots,f_n). \tag{5.11}$$

Again if some signatures on the right side of (5.11) do not satisfy the natural ordering (5.3), then we simply omit all such terms.[31] Moreover, when we have $f_n=0$, then we must delete the last two terms although the last

of them satisfies the natural ordering. In this way we find $N=\epsilon_0+2[n-n_0(\lambda)]$, where $n_0(\lambda)$ is the number of m_j's which are zero and where we defined ϵ_0 to be $\epsilon_0=1$ for $f_n\neq\frac{1}{2}$ and $\epsilon_0=0$ for $f_n=\frac{1}{2}$.

The Casimir invariants $I_p(\lambda)$ are easily computed now as

$$2^p I_p(\lambda)=\sum_{j=1}^{n}\Bigg\{(l_j+\tfrac{1}{2}-n)^p\,\frac{l_j+1}{l_j}\times\prod_{k=1}^{n}{}'\,\frac{(l_j+1)^2-(l_k)^2}{(l_j)^2-(l_k)^2}+(-1)^p(l_j-\tfrac{1}{2}+n)^p\times\frac{l_j-1}{l_j}\prod_{k=1}^{n}{}'\,\frac{(l_j-1)^2-(l_k)^2}{(l_j)^2-(l_k)^2}\Bigg\}+(-n)^p, \tag{5.12}$$

where the product on k in the right side omits the singular point $k=j$. In spite of its apparent fractional form, we can prove that $I_p(\lambda)$ is a symmetric polynomial of $(l_1)^2,(l_2)^2,\ldots,(l_n)^2$ of degree $[p/2]$, where

$$[p/2]=\begin{cases}p/2, & \text{if } p \text{ is an even integer,}\\ (p-1)/2, & \text{if } p \text{ is an odd integer.}\end{cases} \tag{5.13}$$

We may also verify the validity of (2.44), i.e., $I_0(\lambda)=d(\lambda_0)=2n+1$ and $I_1(\lambda)=0$. We may use n Casimir invariants $I_2(\lambda)$, $I_4(\lambda),\ldots,I_{2n}(\lambda)$ to specify[7,8] any irreducible representation of B_n.

From (5.10) and (5.11), we can check that the deficiency index Δ is zero for all irreducible representations $\{\lambda\}$. When we note

$$(Q)_{\alpha\beta}=(X_B\otimes x^B)_{\alpha\beta}=-\tfrac{1}{2}X_{\alpha\beta} \tag{5.14}$$

for (α,β) matrix element in Eq. (2.54), then Eq. (2.53) or (2.49) with $\Delta=0$ give vector identity equations of a kind discovered in Refs. 17 and 19. Also, the result of the Sec. 3 implies that all vector operators are linear combinations of special vector operators $S_B^{(p)}$ $(1\leq p\leq N-1)$;

$$S_B^{(p)}=g_{BB_1B_2\cdots B_p}X^{B_1}X^{B_2}\cdots X^{B_p}, \tag{5.15}$$

where we have set for simplicity $B=(\mu,\nu)$, and $B_j=(\mu_j,\nu_j)$ with

$$g_{BB_1\cdots B_p}=\mathrm{tr}(x_B x_{B_1}\cdots x_{B_p}). \tag{5.16}$$

However, not all of $S_B^{(p)}$ for $1\leq p\leq N-1$ are linearly independent by the following reason. From (5.7), we readily see that x_B satisfies

$$(x_B)^T=-x_B,\quad B=(\mu,\nu) \tag{5.17}$$

for its transpose matrix so that we find

$$g_{BB_1\cdots B_p}=(-1)^{p+1}g_{B_pB_{p-1}\cdots B_1B}. \tag{5.18}$$

Because of (5.18), we can always reduce $S_B^{(p)}$ with even integer p in terms of $S_B^{(q)}$ with $q\leq p-1$, and q being odd integer. Therefore, all vector operators are linear combinations of $n-n_0(\lambda)$ vector operators

$$S_B^{(2q+1)},\quad 0\leq q\leq n-n_0(\lambda)-1. \tag{5.19a}$$

Especially, for any generic representation, any vector operators are linear combinations of n special vector operators:

$$S_B^{(1)}=X_B, S_B^{(3)}, S_B^{(5)},\cdots, S_B^{(2n-1)} \tag{5.19b}$$

in conformity with the theorem stated in Sec. 3.

6. ALGEBRA C_n

This corresponds to the symplectic group $\mathrm{Sp}(2n)$ where generators X_B with $B=(\mu,\nu)$, $\mu,\nu=1,2,\ldots,2n$, satisfy

$$X_{\mu\nu}=X_{\nu\mu}, \tag{6.1}$$

$$[X_{\mu\nu},X_{\alpha\beta}]=\epsilon_{\nu\alpha}X_{\mu\beta}+\epsilon_{\mu\beta}X_{\nu\alpha}+\epsilon_{\nu\beta}X_{\mu\alpha}+\epsilon_{\mu\alpha}X_{\nu\beta}. \tag{6.2}$$

Here $\epsilon_{\mu\nu}$ is antisymmetric with values 0 and ±1,

$$\epsilon_{\mu\nu}=-\epsilon_{\nu\mu}, \tag{6.3a}$$

$$\epsilon_{2j-1,2j}=-\epsilon_{2j,2j-1}=1,\qquad \epsilon_{\mu\nu}=0,\quad \text{if } |\mu-\nu|>1. \tag{6.3b}$$

Again instead of m_j's given in (3.1), we use f_j defined by

$$f_j=m_j+m_{j+1}+\cdots+m_n \tag{6.4}$$

so that they satisfy

$$f_1\geq f_2\geq\cdots\geq f_n\geq 0. \tag{6.5}$$

Setting moreover

$$l_j=f_j+n-j+1, \tag{6.6}$$

the dimension of the representation is

$$d(\lambda)=\frac{1}{1!3!\cdots(2n-1)!}\Big(\prod_{j=1}^{n}l_j\Big)\prod_{j<k}^{n}[(l_j)^2-(l_k)^2]. \tag{6.7}$$

Note that (6.7) has exactly the same form as (5.5) apart from the factor 2^n, although the meaning of l_j differs for two cases.

We choose the reference representation $\{\lambda_0\}$ to be $\{\Lambda_1\}$, i.e.,

$$\{\lambda_0\}=(1,0,0,\cdots,0), \tag{6.8}$$

i.e., $f_1=1$, but $f_j=0$ $(j\geq 2)$. Then its explicit matrix representation of the infinitesimal generator $x_{\mu\nu}=x_{\nu\mu}$ $(\mu,\nu=1,2,\ldots,2n)$ is given by

$$(x_{\mu\nu})_{\alpha\beta}=\epsilon_{\nu\beta}\delta_{\mu\alpha}+\epsilon_{\mu\beta}\delta_{\nu\alpha} \tag{6.9}$$

and we have

$$d(\lambda_0)=2n, \tag{6.10a}$$

$$g_{(\mu\nu),(\alpha\beta)}=-2(\epsilon_{\mu\alpha}\epsilon_{\nu\beta}+\epsilon_{\mu\beta}\epsilon_{\nu\alpha}). \tag{6.10b}$$

The second order Casimir invariant is easily computed to be

$$I_2(\lambda)=\tfrac{1}{2}\Big\{\sum_{j=1}^{n}(l_j)^2-\tfrac{1}{6}n(n+1)(2n+1)\Big\}. \tag{6.11}$$

With the choice (6.8) for $\{\lambda_0\}$, the nondegeneracy condition is always satisfied with $\Delta=0$ for any irreducible representation $\{\lambda\}$. The Clebsch–Gordan decomposition is

$$\begin{aligned}\{\lambda\}\otimes\{\lambda_0\}&=(f_1,f_2,\cdots,f_n)\otimes(1,0,\cdots,0)\\&=(f_1+1,f_2,\cdots,f_n)\oplus(f_1,f_2+1,f_3,\cdots,f_n)\oplus\cdots\\&\quad\oplus(f_1,\cdots,f_{n-1},f_n+1)\oplus(f_1-1,f_2,\cdots,f_n)\\&\quad\oplus(f_1,f_2-1,f_3,\cdots,f_n)\oplus\cdots\oplus(f_1,\cdots,f_{n-1},f_n-1)\end{aligned} \tag{6.12}$$

by the Weyl' formula.[24] Again, we omit all terms in the right side of (6.12) if the signature factors of the term violate the natural ordering (6.5). We have in the present case $N=\epsilon_0+2(n-n_0(\lambda))$, where $\epsilon_0=0$ for $f_n\neq 0$ and $\epsilon_0=1$ for $f_n=0$. The pth order Casimir invariant is readily computed to be

$$2^p I_p(\lambda)=\sum_{j=1}^{n}\left\{(l_j-n)^p\,\frac{l_j+1}{l_j}\prod_{k=1}^{n}{}',\ \frac{(l_j+1)^2-(l_k)^2}{(l_j)^2-(l_k)^2}\right.$$
$$\left.+(-1)^p\,(l_j+n)^p\,\frac{l_j-1}{l_j}\prod_{k=1}^{n}{}',\ \frac{(l_j-1)^2-(l_k)^2}{(l_j)^2-(l_k)^2}\right\}, \tag{6.13}$$

where we omit the singular term $k=j$ in the product over k. Again, we can prove that $I_p(\lambda)$ is a symmetric polynomial of $l_1^2,l_2^2,\ldots,l_n^2$ of order $[p/2]$, so that n terms $I_2(\lambda)$, $I_4(\lambda),\ldots,I_{2n}(\lambda)$ can specify[2] any irreducible representations. We can also check the condition $I_0(\lambda)=d(\lambda_0)=2n$ and $I_1(\lambda)=0$. Also, Eq. (6.13) reproduces the formula (6.11) for $p=2$.

Since the deficiency index Δ is zero, all vector operators are linear combinations of $S_B^{(p)}$ as in (5.15) and (5.16). Also, x_B satisfies now

$$x_B^T=-\epsilon^{-1}x_B\epsilon \tag{6.14}$$

where ϵ is the $2n\times 2n$ matrix whose (μ,ν) component is given by $\epsilon_{\mu\nu}$. This replaces (5.17) but the same argument is applicable to prove (5.18). Therefore, again all vector operators are linear combinations of at most n vector operators of the form (5.19) again. More precisely, only first $n-n_0(\lambda)$ of them are required, in conformity with the theorem of Sec. 3.

7. ALGEBRA D_n

This is the Lie algebra corresponding to the group SO($2n$). The generators x_B, $B=(\mu,\nu)$, satisfy the same relations (5.1) although μ,ν,α, and β assume now values only $1,2,\ldots,2n$ but not $2n+1$.

We define f_j's by

$$f_j=m_j+m_{j+1}+\cdots+m_{n-2}+\tfrac{1}{2}(m_{n-1}+m_n),\quad 1\leq j\leq n-2,$$
$$f_{n-1}=\tfrac{1}{2}(m_{n-1}+m_n),\quad f_n=\tfrac{1}{2}(-m_{n-1}+m_n) \tag{7.1}$$

so that they satisfy now

$$f_1\geq f_2\geq\cdots\geq f_{n-1}\geq|f_n|. \tag{7.2}$$

Note that the last entry in (7.2) is not f_n but its absolute value $|f_n|$. Again, f_j's assume values simultaneously all integers or all half-integers, corresponding to tensor or spinor representations.

When we set moreover

$$l_j=f_j+n-j,\quad 1\leq j\leq n, \tag{7.3}$$

the dimension formula is now given by

$$d(\lambda)=\frac{2^{n-1}}{2!4!\cdots(2n-2)!}\prod_{j<k}^{n}[(l_j)^2-(l_k)^2]. \tag{7.4}$$

We choose again the reference representation $\{\lambda_0\}$ to be $\{\Lambda_1\}$, i.e., $\{\lambda_0\}=(1,0,\ldots,0)$ with $f=1$ and $f_j=0$ $(j\geq 2)$. Its explicit matrix representation is still given by Eq. (5.7) with (5.9). However, (5.8) is now replaced by

$$d(\lambda_0)=2n. \tag{7.5}$$

The Clebsch–Gordan decomposition rule[32] is formally equivalent to Eq. (6.12), although we must delete all terms which are unnatural now under the ordering (7.2). Then we find $N=2[n-n_0(\lambda)]$.

The second-order Casimir invariant is computed to be

$$I_2(\lambda)=\tfrac{1}{2}\{\sum_{j=1}^{n}(l_j)^2-\tfrac{1}{6}n(n-1)(2n-1)\}. \tag{7.6}$$

We observe that the nondegeneracy condition is satisfied by our choice of $\{\lambda_0\}$ for all generic representations $\{\lambda\}$. We have $\Delta=0$ for any representation $\{\lambda\}$ such that $f_n\neq 0$ or $f_n=f_{n-1}=0$. However, the deficiency index assumes $\Delta=1$ for any representation $\{\lambda\}$ with $f_n=0$ but $f_{n-1}\geq 1$. This latter case adds a new complication in determining all vector operators.

The pth order Casimir invariant is readily computed to be

$$2^p I_p(\lambda)=\sum_{j=1}^{n}\left\{(l_j-n+1)^p\prod_{k=1}^{n}{}',\ \frac{(l_j+1)^2-(l_k)^2}{(l_j)^2-(l_k)^2}\right.$$
$$\left.+(-1)^p(l_j+n-1)^p\prod_{k=1}^{n}{}',\ \frac{(l_j-1)^2-(l_k)^2}{(l_j)^2-(l_k)^2}\right\}. \tag{7.7}$$

Again, we can prove that $I_p(\lambda)$ is a symmetric polynomial of $(l_1)^2$, $(l_2)^2,\ldots,(l_n)^2$ of the degree $[p/2]$. We may also verify $I_0(\lambda)=d(\lambda_0)=2n$, and $I_1(\lambda)=0$. We can determine $l_1^2,l_2^2,\ldots,(l_n)^2$ from n quantities $I_2(\lambda)$, $I_4(\lambda),\ldots,I_{2n}(\lambda)$. But this does not specify the sign of f_n, since $l_n=f_n$. In this sense, the choice of the above set is insufficient. However, we have another Casimir invariant

$$\tilde{K}_n=(-1)^{n(n+2)/2}\frac{1}{2^n n!}\epsilon_{\mu_1\nu_1\mu_2\nu_2\cdots\mu_n\nu_n}X_{\mu_1\nu_1}X_{\mu_2\nu_2}\cdots X_{\mu_n\nu_n}, \tag{7.8}$$

where $\epsilon_{\mu_1\nu_1\mu_2\nu_2\cdots\mu_n\nu_n}$ is the completely antisymmetric Levi-Civita tensor, with values 0 and ± 1. We may note that (7.8) can be obtained from $K^{(n)}$ when we choose $\{\lambda_0\}$ to be a spinor representation $\{\lambda_0\}=(\frac{1}{2},\frac{1}{2},\cdots,\frac{1}{2})$, i.e., $f_j=\frac{1}{2}$ for all j. At any rate, its eigenvalue $\tilde{I}_n(\lambda)$ has been computed by Louck[9] to be

$$\tilde{I}_n(\lambda)=l_1l_2\cdots l_n. \tag{7.9}$$

We shall present a simpler proof of (7.9) in the Appendix B. Note that $(\tilde{I}_n)^2$ is a polynomial of $I_2,I_4,\ldots,I_{2n}$, but $\tilde{I}_n$ itself is linearly independent of them. Therefore,[2] if we choose $I_2,I_4,\ldots,I_{2n-2},\tilde{I}_n$ to be the n-algebraically independent Casimir invariant, then we can completely specify any irreducible representation of the D_n.

Next, let us consider the problem of finding all vector operators. As we noted earlier, we have $\Delta=0$ if $f_n\neq 0$ or if $f_{n-1}=f_n=0$. Hence as long as we have $f_n\neq 0$, the same argument presented for the case of the algebra B_n is applicable and we conclude that any vector operators are linear combinations of at most n special vectors $X_B,S_B^{(3)},\ldots,S_B^{(2n-1)}$ as in (5.19). However, the same argument does not apply to the case $f_n=0$ but $f_{n-1}\neq 0$ since we have $\Delta=1$, and we have to resort to the method discussed in Sec. 3. All Casimir invariants of

the form $\tilde{Q}$ ($p \geq 2$) discussed there do not still uplift the degeneracy of two subspaces specified by P_{2n-1} and P_{2n}. Because of this, we consider a Casimir invariant $\tilde{Q}$ in $\{\lambda\} \otimes \{\lambda_0\}$ given by

$$\tilde{Q} = (-1)^{n(n+2)/2} \frac{1}{2^n n!} \epsilon_{\mu_1\nu_1\mu_2\nu_2 \cdots \mu_n\nu_n} Y_{\mu_1\nu_1} Y_{\mu_2\nu_2} \cdots Y_{\mu_n\nu_n}, \tag{7.10}$$

$$Y_{\mu\nu} = X_{\mu\nu} \otimes E_0 + E \otimes x_{\mu\nu} \tag{7.11}$$

then an analog of (3.16) is now given by

$$\tilde{Q} = \sum_{j=1}^{2n} \tilde{I}_n(\lambda_j) P_j. \tag{7.12}$$

But for $f_n = 0$, and $f_{n-1} \neq 0$ only the projection operators P_{2n-1} and P_{2n} corresponding to representations $(f_1, f_2, \ldots, f_{n-1}, f_n - 1)$ and $(f_1, f_2, \ldots, f_{n-1}, f_n + 1)$ in the right side of the decomposition to $\{\lambda\} \otimes \{\lambda_0\}$ can contribute, since for all other cases $\tilde{I}_n(\lambda_j) = 0$. The result is

$$\tilde{Q} = (l_1 l_2 \cdots l_{n-1})[P_{2n-1} - P_{2n}]. \tag{7.13}$$

Since the sum $P_{2n-1} + P_{2n} = \tilde{P}_{2n-1}$ can be computed as in Sec. 3, we can completely specify all P_j now as functions of Q and $\tilde{Q}$. Then, as in (3.18), all vector operators must be linear combinations of n special vector operators,

$$X_{\mu\nu}, S^{(3)}_{\mu\nu}, S^{(5)}_{\mu\nu}, \cdots, S^{(2n-3)}_{\mu\nu}, \tilde{S}_{\mu\nu}, \tag{7.14}$$

where $\tilde{S}_{\mu\nu}$ is defined by

$$\tilde{S}_{\mu\nu} = \frac{1}{2^n n!} \epsilon_{\mu\nu\mu_1\nu_1\mu_2\nu_2 \cdots \mu_{n-1}\nu_{n-1}} X_{\mu_1\nu_1} X_{\mu_2\nu_2} \cdots X_{\mu_{n-1}\nu_{n-1}}. \tag{7.15}$$

Actually, we can show that, for the case $f_n \neq 0$, $S^{(2n-1)}_{\mu}$ can be expressed as a linear combination of (7.14). Therefore, for all cases, we can say that n-linearly independent vector operators of the algebra D_n is given by (7.14). To prove this fact, we note that we can easily express

$$\tilde{K}_n \tilde{S}_{\mu\nu} = \sum_{j=1}^{n} a_j S^{(2j-1)}_{\mu\nu}, \quad a_n \neq 0 \tag{7.16}$$

since a product of two Levi-Civita tensors can be reduced to a sum of products of Kroneckers deltas, $\delta_{\mu\nu}$. For the case $f_n = 0$, we know $\tilde{K}_n = 0$ so that (7.16) simply reproduces the vector identity for $\Delta = 1$. But if we have $f_n \neq 0$, then $\tilde{K}_n \neq 0$ so that we can express $\tilde{S}_{\mu\nu}$ in terms of $S^{(2j-1)}_{\mu}$ ($1 \leq j \leq n$), and conversely $S^{(2n-1)}_{\mu}$ in terms of (7.14).

8. ALGEBRA G_2

The discussions for five exceptional algebras G_2, F_4, E_6, E_7, and E_8 are quite involved. Here, we shall consider the simplest case G_2 as an illustration. We use now the standard Cartan—Weyl basis for G_2. If α_1 and α_2 are two fundamental roots of G_2, then two fundamental weights Λ_1 and Λ_2 are given by

$$\Lambda_1 = 2\alpha_1 + 3\alpha_2, \quad \Lambda_2 = \alpha_1 + 2\alpha_2. \tag{8.1}$$

Then, the highest weight Λ of a generic irreducible representation $\{\lambda\}$ is expressed as

$$\Lambda = m_1\Lambda_1 + m_2\Lambda_2 \tag{8.2}$$

where m_1 and m_2 are nonnegative integers.[33] The dimension of the representation is

$$d(\lambda) = \frac{1}{120}(m_1+1)(m_2+1)(m_1+m_2+2)(2m_1+m_2+3) \times (3m_1+m_2+4)(3m_1+2m_2+5). \tag{8.3}$$

We choose the reference representation $\{\lambda_0\}$ to be the seven-dimensional representation

$$\{\lambda_0\} = \{\Lambda_2\} = (0,1), \quad d(\lambda_0) = 7. \tag{8.4}$$

The Clebsch—Gordan series can be computed by means of the formula (2.38), since the twelve elements of the Weyl group as well as the weight diagram of $\{\lambda_0\}$ are well known.[34] The result is

$$\begin{aligned}\{\lambda\} \otimes \{\lambda_0\} &= (m_1, m_2) \otimes (0,1) \\ &= (m_1, m_2+1) \oplus (m_1, m_2) \oplus (m_1, m_2-1) \\ &\quad \oplus (m_1+1, m_2-2) \oplus (m_1+1, m_2-1) \\ &\quad \oplus (m_1-1, m_2+2) \oplus (m_1-1, m_2+1)\end{aligned} \tag{8.5}$$

as long as $m_1 \geq 1$ and $m_2 \geq 2$ with $N = 7$. When we have $m_1 = 0$ and/or $m_2 = 1$, we simply omit any terms such as $(m_1', -1)$ and $(-1, m_2')$ from the right side of (8.5). However, for $m_2 = 0$, the situation is more involved. The correct formula for this case is[35]

$$(m_1, 0) \otimes (0,1) = (m_1, 1) \oplus (m_1 - 1, 1) \oplus (m_1 - 1, 2) \tag{8.6}$$

for $m_1 \geq 1$ with $N = 3$. Then, we see that the choice of this reference representation $\{\lambda_0\}$ satisfies the nondegeneracy condition with $\Delta = 0$ for all irreducible representations $\{\lambda\}$. The second-order Casimir invariant is computed to be

$$3I_2(\lambda) = 3(m_1)^2 + (m_2)^2 + 3m_1m_2 + 9m_1 + 5m_2. \tag{8.7}$$

Therefore, we can compute all higher-order Casimir invariants from (2.32) or (2.34). However, the final expression is very complicated and we will not reproduce them here. As before, all Casimir invariants $I_p(\lambda)$ with odd p's are expressed in terms of $I_p(\lambda)$ with even p's. One peculiar aspect of G_2 is the fact that the fourth-order Casimir invariant is a function of I_2, i.e., we have identically

$$I_4(\lambda) = \tfrac{1}{4}[I_2(\lambda)]^2 + \tfrac{7}{6} I_2(\lambda). \tag{8.8}$$

However, the six-order Casimir invariants $I_6(\lambda)$ is algebraically independent[2] of $I_2(\lambda)$ as has also been observed by Scheibling and Umezawa.[8] Therefore, any irreducible representation of the G_2 is characterized[2] by $I_2(\lambda)$ and $I_6(\lambda)$.

Regarding the determination of all vector operators, we also encounter a similar trait. Since we have $\Delta = 0$ and $N \leq 7$, the discussion of Sec. 3 shows that all vector operators must be a linear combination of vectors $S^{(p)}_{\mu}$ with $1 \leq p \leq 6$. But the theorem stated in Sec. 3 demands that only two of them are in general linearly independent. Since $\{\lambda_0\} = (0,1)$ is the only seven-dimensional irreducible representation of G_2, its transpose representation must be equivalent to $\{\lambda_0\}$ itself. In other words, we must have a nonsingular 7×7 matrix g such that

$$-x^T_{\mu} = g x_{\mu} g^{-1}, \quad \mu = 1, 2, \cdots, 14. \tag{8.9}$$

As its consequence, we have the validity of (5.18) so that only $X_\mu = S_\mu^{(1)}$, $S_\mu^{(3)}$, and $S_\mu^{(5)}$ could be linearly independent. Moreover, we can prove that $S_\mu^{(3)}$ must be proportional to X_μ, i.e., we have

$$S_\mu^{(3)} = cX_\mu, c = \tfrac{1}{4}I_2(\lambda) + \tfrac{7}{6}. \tag{8.10}$$

Therefore, we conclude that only X_μ and $S_\mu^{(5)}$ are two linearly independent vector operators in generic irreducible representations of G_2. For the cases with $m_1 = 0$ or $m_2 = 0$, only X_μ is the sole[33] linearly independent vector operator of G_2, since we have $N = 5$ or 3 for these cases. The validity of (8.8) and (8.10) is a consequence of the following identity in G_2. Let ξ_μ ($\mu = 1, 2, \ldots, 14$) be real arbitrary constants. Then, we can prove an identity

$$4\mathrm{tr}[(\xi x)^4] = [\mathrm{tr}(\xi x)^2]^2 \tag{8.11}$$

where we have set

$$(\xi x) = \sum_{\mu=1}^{14} \xi_\mu x_\mu. \tag{8.12}$$

The proof of (8.11) is rather involved and will not be given here.

ACKNOWLEDGMENT

The author would like to express his thanks to the referee of this paper for informing him of Refs. 2, 5, 6, and 33.

APPENDIX A

Here we shall prove the theorem Eq. (3.2). To that end, let

$$\pi = \{\alpha_1, \alpha_2, \cdots, \alpha_n\} \tag{A1}$$

be the fundamental (or simple) root system of L, and let

$$S_j \equiv S_{\alpha_j}, \quad \alpha_j \in \pi, \tag{A2}$$

be the special Weyl operation with respect to plane perpendicular to the root $\alpha_j \in \pi$. If $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ are the fundamental weight system of L, then we find[36]

$$2(\Lambda_j, \alpha_k) = \delta_{jk}(\alpha_k, \alpha_k), \tag{A3}$$

$$2(\delta, \alpha_k) = (\alpha_k, \alpha_k), \tag{A4}$$

$$S_j \Lambda_k = \Lambda_k - \delta_{jk}\alpha_k, \tag{A5}$$

$$S_j \delta = \delta - \alpha_j, \tag{A6}$$

where $\delta = \frac{1}{2}\sum_{\alpha>0}\alpha$ is defined by (2.37) and δ_{jk} is the Kronecker symbol, i.e., $\delta_{jk} = 1$, if $j = k$ and $\delta_{jk} = 0$, if $j \neq k$. Therefore, if

$$\Lambda = m_1\Lambda_1 + m_2\Lambda_2 + \cdots + m_n\Lambda_n \tag{A7}$$

is the highest weight of the irreducible representation $\{\lambda\}$, then we compute

$$2(K, \alpha_j) = (m_j + 1)(\alpha_j, \alpha_j), \tag{A8}$$

$$S_j K = K - (m_j + 1)\alpha_j \tag{A9}$$

for $K = \Lambda + \delta$.

According to the well-known Wigner–Eckart theorem,[37] the number of linearly independent vector operators in $\{\lambda\}$ is precisely equal to the multiplicity of the representation $\{\lambda\}$ occurring in the product $\{\lambda\} \otimes \{\lambda_0\}$, where $\{\lambda_0\}$ is now the adjoint representation of L. Hence by the formula (2.38) of Weyl, we have

$$n_V(\lambda) = C(\lambda; \lambda, \lambda_0) = \sum_S (\det S)\gamma_0(SK - K). \tag{A10}$$

The multiplicity $\gamma_0(M)$ of the weight M in the adjoint representation $\{\lambda_0\}$ is evidently given by

$$\gamma_0(M) = \begin{cases} n, & \text{if } M = 0, \\ 1, & \text{if } M = \alpha \text{ is any nonzero root}, \\ 0, & \text{if } M \text{ is not a root}. \end{cases} \tag{A11}$$

Note that K is dominant[36] with

$$SK < K, \quad \text{if } S \neq I = \text{identity}. \tag{A12}$$

so that we separate the case $S = I = $ identity in (A10) to find

$$n_V(\lambda) = n + \sum_{\alpha>0} \det S(\alpha). \tag{A13}$$

Here the summation is over all positive root α such that

$$K - SK = \alpha \tag{A14}$$

has a solution $S = S(\alpha)$ for some Weyl symmetry $S(\alpha)$. Note that the solution, if it exists, is unique because of (A12). We shall prove below that (A14) has a solution if and only if α coincides with $\alpha_j \in \pi$ such that $m_j = 0$. Then, the solution is given by $S = S_j$ with $\det S = -1$. Therefore, (A13) gives the desired formula, Eq. (3.2), i.e.,

$$n_V(\lambda) = n - n_0(\lambda), \tag{A15}$$

where $n_0(\lambda)$ is the number of m_j's which is zero.

To prove this, we note first that (A14) leads to

$$(K, K) = (SK, SK) = (K - \alpha, K - \alpha)$$

so that the root α must satisfy a condition

$$(\alpha, \alpha) = 2(K, \alpha). \tag{A16}$$

Second, since α is a positive root, we can express it as

$$\alpha = \sum_{j=1}^{n} p_j\alpha_j \tag{A17}$$

where p_j are nonnegative integers. In view of (A8), we can rewrite (A16) as

$$(\alpha, \alpha) = \sum_{j=1}^{n} (m_j + 1)p_j(\alpha_j, \alpha_j). \tag{A18}$$

On the other hand, any non-zero root α can be written[36] in a form of

$$\alpha = T\alpha_l, \quad \alpha_l \in \pi, \tag{A19}$$

in terms of a Weyl operation T and $\alpha_l \in \pi$. This leads to

$$(\alpha, \alpha) = (\alpha_l, \alpha_l) \tag{A20}$$

for some $\alpha_l \in \pi$.

We now consider three different cases, separately.

Case (I) (Algebras A_n, D_n, E_6, E_7, and E_8)

In this case a glance on Dynkin diagrams of these algebras gives

$$(\alpha_j,\alpha_j)=1, \quad \alpha_j\in\pi$$

for all j in a suitable unit. Therefore, we have $(\alpha,\alpha)=1$ also from (A20) and (A18) becomes to

$$\sum_{j=1}^{n}(m_j+1)p_j=1. \tag{A21}$$

But since m_j and p_j are nonnegative integers, this is possible only if one of p_j alone is nonzero and we have

$$\alpha=\alpha_l, \quad m_l=0.$$

Then, $S=S_l$ is a solution of (A14) in view of (A9). Conversely, if we have $m_l=0$ for some m_l, then we find that $\alpha=\alpha_l$ and $S=S_l$ satisfy (A14). Therefore, (A15) follows.

Case (II) (Algebras B_n, C_n, and F_4)

In this case we know

$$(\alpha_j,\alpha_j)=1 \text{ or } 2$$

in a suitable unit so that from (A20) we find

$$(\alpha_j,\alpha_j)\geq\tfrac{1}{2}(\alpha,\alpha) \tag{A22}$$

and (A18) is satisfied only if

$$2\geq\sum_{j=1}^{n}(m_j+1)p_j. \tag{A23}$$

A solution $\sum_{j=1}^{n}(m_j+1)p_j=1$ leads to the same conclusion as in the Case I, and we only need to prove the impossibility of another alternative

$$\sum_{j=1}^{n}(m_j+1)p_j=2. \tag{A24}$$

This may be possible only if we have either

$$\alpha=\alpha_l, \quad m_l=1, \quad p_l=1, \tag{A25}$$

or

$$\alpha=2\alpha_k, \quad m_k=0, \quad p_k=2 \tag{A26}$$

or

$$\alpha=\alpha_j+\alpha_k \quad (j\neq k), \quad m_j=m_k=0, \quad p_j=p_k=1. \tag{A27}$$

But the first possibility (A25) contradicts (A22) and/or (A18). The second possibility $\alpha=2\alpha_k$ is also ruled out, since $2\alpha_k$ cannot be a root. For the last case, (A27), the equality in (A22) and (A23) can happen only if

$$(\alpha,\alpha)=2(\alpha_j,\alpha_j)=2(\alpha_k,\alpha_k). \tag{A28}$$

Comparing this with (A27), we must have

$$\alpha=\alpha_j+\alpha_k \quad (j\neq k), \quad (\alpha_j,\alpha_k)=0. \tag{A29}$$

But an inspection of Dynkin diagrams shows that (A28) and (A29) are impossible for B_n and F_4. For the algebra C_n (as well as B_n and F_4), we proceed as follows. Since α is assumed to be a root, so must be

$$\beta=S_j\alpha=\alpha-[2(\alpha,\alpha_j)/(\alpha_j,\alpha_j)]\alpha_j.$$

But (A29) implies

$$\beta=\alpha_k-\alpha_j \quad (k\neq j).$$

Since β is neither absolutely positive nor negative, β cannot be a root, and this is a contradiction. This implies that α of the form (A29) is *not* a root.

Case (III) (Algebra G_2)

In this case, we know

$$(\alpha_1,\alpha_1)=3, \quad (\alpha_2,\alpha_2)=1, \quad (\alpha_1,\alpha_2)=-\tfrac{3}{2}. \tag{A30}$$

A similar study of (A18) and (A20) enables us to conclude that (A18) is possible only for $\alpha=\alpha_1$, with $m_1=0$ or for $\alpha=\alpha_2$ with $m_2=0$. Therefore, just as in the Case (I), we obtain (A15).

We remark that the validity of the present theorem for the special case of the algebra A_n has been noted by Pais.[30]

APPENDIX B

We shall prove here (7.9). For this purpose, it is convenient to introduce K^μ_ν, $R_{\mu\nu}$, and $R^{\mu\nu}$, for $\mu,\nu=1,2,\ldots,n$ by[38]

$$\begin{aligned}
K^\mu_\nu&=\tfrac{1}{2}\{-X_{\mu\nu}-X_{\mu+n,\nu+n}-i[X_{\mu,\nu+n}-X_{\mu+n,\nu}]\},\\
R^{\mu\nu}&=-R^{\nu\mu}=\tfrac{1}{2}\{X_{\mu\nu}-X_{\mu+n,\nu+n}-i[X_{\mu,\nu+n}+X_{\mu+n,\nu}]\},\\
R_{\mu\nu}&=-R_{\nu\mu}=\tfrac{1}{2}\{-X_{\mu\nu}+X_{\mu+n,\nu+n}-i[X_{\mu,\nu+n}+X_{\mu+n,\nu}]\}.
\end{aligned} \tag{B1}$$

Then, the Cartan subalgebra may be generated by n elements,

$$H_\mu=K^\mu_\mu=-iX_{\mu,\mu+n}, \quad \mu=1,2,\cdots,n, \tag{B2}$$

and all other K^μ_ν $(\mu\neq\nu)$, $R_{\mu\nu}$, and $R^{\mu\nu}$ correspond to some E_α's with nonzero root α.

We can rewrite the Casimir invariant (7.8) in a form[8]

$$\tilde{K}_n=H_1H_2\cdots H_n+\text{lesser polynomials of } H_j \text{ and } E_\alpha. \tag{B3}$$

Let v be the highest weight state of the generic representation $\{\lambda\}$ with signature (7.2). Then, as usual,[8] we compute

$$\tilde{K}_nv=[l_1l_2\cdots l_n+P_{n-1}(l)]v, \tag{B4}$$

where $P_{n-1}(l)$ is a polynomial of $l_1,l_2,\ldots,l_n$ of at most degree $n-1$. In deriving (B4), we used the fact

$$H_jv=f_jv, \quad E_\alpha v=0 \quad (\alpha>0). \tag{B5}$$

Equation (B4) implies that the eigenvalue $\tilde{I}_n(\lambda)$ of the Casimir invariant $\tilde{K}_n$ has a form

$$\tilde{I}_n(\lambda)=l_1l_2\cdots l_n+P_{n-1}(l). \tag{B6}$$

However, $P_{n-1}(l)$ must be a symmetric polynomial of $l_1,l_2,\ldots,l_n$ since $(\tilde{K}_n)^2$ can be easily shown to be polynomial of $K^{(p)}$ $(1\leq p\leq 2n)$.

Next, consider a linear mapping of D_n into itself, given by

$$X_{\mu\nu}\to X'_{\mu\nu}=\theta(X_{\mu\nu})\equiv\begin{cases}X_{\mu\nu}, & \text{if } \mu\neq 2n \text{ and } \nu\neq 2n,\\ -X_{\mu\nu}, & \text{if } \mu=2n \text{ or } \nu=2n.\end{cases} \tag{B7}$$

We may easily verify that θ is an automorphism of D_n. Moreover, we can prove that it is outer.[39] Indeed, the effect of θ is to change

$$l_j \to l_j \quad (j \neq n),$$
$$l_n \to -l_n \quad (j = n), \tag{B8}$$

as we see from (B5). A more careful analysis shows that θ interchanges[39] the two roots α_{n-1} and α_n in the Dynkin diagram of the D_n. Also, the operation of θ by (B7) changes the sign of the Casimir invariant $\tilde{K}_n$. This implies that the polynomials $P_{n-1}(l)$ must change its sign under (B8). But since $P_{n-1}(l)$ is a symmetric polynomial of $l_1, l_2, \ldots, l_n$ of degree at most $n-1$, this is not possible unless $P_{n-1}(l)$ is identically zero. Therefore, we must have

$$\tilde{I}_n(\lambda) = l_1 l_2 \cdots l_n, \tag{B9}$$

which is (7.9).

[1]E.g., see *Group Theory and Its Applications*, *Vols. I* and *II*, edited by E.M. Loebl (Academic, New York and London, 1968).
[2]H. Samuelson, Ann. Math. **42**, 1091 (1941); A. Borel and C. Chevalley, Amer. Math. Soc. Mem. **14**, 1 (1955); B. Kostant, Amer. J. Math. **85**, 327 (1963).
[3]G. Racah, Rend. Lincei **8**, 108 (1950) and Princeton Lecture Notes (1951).
[4]I.M. Gel'fand, Mat. Sb. **26**, (68) No. 1, 103 (1950).
[5]L.C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
[6]L. Michel and L.A. Radicatti, Ann. Inst. H. Poincaré A **XVIII**, 185 (1973).
[7]B. Gruber and L. O'Raifeartaigh, J. Math. Phys. **5**, 1796 (1964).
[8]M. Umezawa, Koninkl. Nederl. Akad. Wetensch. B **69**, 579 (1966); F. Scheibling and M. Umezawa, *ibid.* B **73**, 22 (1970).
[9]J.D. Louck, "Theory of Angular Momentum in N-dimensional Space," Los Alamos Scientific Lab. (1960).
[10]J.D. Louck and L.C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[11]L.C. Biedenharn, J.D. Louck, F. Chacón, and M. Ciftan, J. Math. Phys. **13**, 1957 (1972); L.C. Biedenharn and J.D. Louck, *ibid.* **13**, 1985 (1972); J.D. Louck and L.C. Biedenharn, *ibid.* **14**, 1336 (1973).
[12]There are many other works not listed above. See Ref. 1 as well as those quoted below.
[13]V.S. Popov and A.M. Perelomov, Yad. Fiz. **5**, 693 (1967) [Sov. J. Nucl. Phys. **5**, 489 (1967)]; A.M. Perelomov and V.S. Popov, *ibid.* **3**, 924, 1127 (1966) [*ibid.* **3**, 676, 819 (1966)].
[14]M.X.F. Wong and H.Y. Yeh, J. Math. Phys. **16**, 1239 (1975). This paper contains further earlier references on the subject.
[15]C.O. Nwachuku and M.A. Rashid, J. Math. Phys. **17**, 1611 (1976).
[16]S. Okubo, J. Math. Phys. **16**, 528 (1975).
[17]C.O. Nwachuku and M.A. Rashid, ICTP Report IC/75/144 (1975). However, the conclusion reached in this paper for the algebra D_n is, I believe, incomplete. As we shall see in Sec. 7 of the present paper, the case of D_n is more complicated. Also, the same remark may apply to some representations of B_n and C_n, where these irreducible representations are degenerate, i.e., some of m_j are zero.
[18]J.D. Louck, Amer. J. Phys. **31**, 378 (1963) and J. Math. Phys. **6**, 1786 (1965); N. Mukunda, J. Math. Phys. **8**, 1069 (1967).
[19]A.J. Bracken and H.S. Green, J. Math. Phys. **12**, 2099 (1971); H.S. Green, *ibid.* **12**, 2107 (1971).
[20]See also Ref. 15. Such identities for the algebra A_n are useful to simplify mass formula in particle physics, e.g., see S. Okubo, Phys. Rev. D **11**, 3261 (1975).
[21]P.G.O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976), where the terminology is suggested for cases when the nondegenerate bilinear form (2.10) exists for a true or projective representation $\{\lambda_0\}$.
[22]N. Jacobson, *Lie Algebra* (Interscience, New York, 1962), p. 69.
[23]Ref. 22, p. 261.
[24]H. Weyl, *Classical Group* (Princeton U.P., Princeton, N.J., 1939). See the theorem (7.10A), p. 231, as well as the comment (22) on p. 295.
[25]G. Racah, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York and London, 1964), p. 1. See also D. Speiser, *ibid.*, p. 201 and R.E. Behrends, in Ref. 1, p. 541.
[26]N. Jacobson, Ref. 22, p. 247.
[27]L. Michel, see Ref. 1, Vol. II.
[28]H. Weyl, see Ref. 24, Chap. IV.
[29]Note that $f_{j+1} = f_j$ is equivalent to $m_j = 0$ for this case. This is the reason why we obtain the redundant factor rule of Ref. 16. For other algebras, the above two conditions, however, are not necessarily equivalent to each other.
[30]This formula can be more easily derived from the so-called Young tableau technique; see, e.g., A. Pais, Rev. Mod. Phys. **38**, 215 (1966); C. Itzykson and M. Nauenberg, *ibid.* **38**, 95 (1966).
[31]See F.D. Murnaghan, *Theory of Group Representations* (Johns Hopkins U.P., Baltimore, 1938), pp. 278 and 282, where the Clebsch–Gordan decomposition (5.11) is given for cases of any tensor representation $\{\lambda\}$.
[32]This formula differs from the one given by Murnaghan (see Ref. 31) for the case $f_n \neq 0$. The reason is that Murnaghan is discussing the representations of the full $O(2n)$ group, but *not* the $SO(2n)$ subgroup which is related to the algebra D_n. For the case $f_n > 0$, any irreducible representation of the full $O(2n)$ group is a direct sum of two irreducible representations $(f_1, f_2, \ldots, f_{n-1}, f_n)$ and $(f_1, f_2, \ldots, f_{n-1}, -f_n)$ of the $SO(2n)$ group. However, for the case $f_n = 0$, irreducible representations are the same for both groups. This is also the reason why the dimensional formula of Ref. 31 for $O(2n)$ gives answers differing by a factor 2, depending upon $f_n = 0$ or $f_n \neq 0$.
[33]B.R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963), specifies the irreducible representation of G_2 by two integers u_1 and u_2 $(u_1 \geq u_2)$ which are related to our m_1 and m_2 by $u_1 = m_1 + m_2$ and $u_2 = m_1$. However, over-all normalization constants for the second-order Casimir invariant differs from that given in this paper since we used the seven-dimensional representation for the definition of $g_{\mu\nu}$, rather than the adjoint representation.
[34]E.g., M. Konuma, K. Shima, and M. Wada, Prog. Theor. Phys. Suppl. **28**, 1 (1963); R.E. Behrends, J. Dreitlein, C. Fronsdal, and B.W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[35]In comparison to (8.5), the result (8.6) implies that we should delete $(m_1+1, -2)$ together with $(m_1, 0)$ representations in addition to all $(m_1', -1)$ terms. We may remark that the dimensional formula (8.3) shows that the formal dimension of $(m_1+1, -2)$ representation is negative but has the same magnitude as that of the $(m_1, 0)$ representation so that the sum of two dimensions is zero. Similarly, formal dimensions of $(m_1', -1)$ and $(-1, m_2')$ representations are identically zero. These are the reasons why the formula (8.6) results from (8.5) for $m_2 = 0$.
[36]E.g. see N. Jacobson, Ref. 22. However, the Weyl reflection S operates here from the left to the right, instead of the contrary convention there. The results quoted here are scattered over in this reference.
[37]E.g. see L. O'Raifeartaigh, *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1968), p. 469.
[38]S. Okubo, Phys. Rev. C **10**, 2048 (1974).
[39]Note that the algebra D_n $(n \geq 5)$ as well as $D_3 = A_3$ has effectively only one outer automorphism, while D_4 has five. The connection between outer automorphism and the invariance group of Dynkin diagram is well known. See N. Jacobson, Ref. 22, p. 281–85.

A new class of superalgebras and local gauge groups in superspace[a)]

Freydoon Mansouri[b)]

Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305
(Received 5 July 1977)

It is shown that there is a new class of superalgebras associated with a given Lie algebra or a superalgebra. The structure constants of the new algebras either vanish or else are directly related to those of the original algebra. The new algebraic structures provide a possible link between the local gauge groups constructed over superspace and those over ordinary space-time.

I. INTRODUCTION

The purpose of this paper is to construct a new class of superalgebras associated with a given Lie algebra or with another known superalgebra of simple or semisimple type. The latter algebras have been extensively studied in the literature.[1-3] The proof of the existence of the new algebras is given by constructing explicit representations for them. This is carried out in Sec. II. Although the new algebraic structures are independent of any specific application, the interest in them is not purely from the mathematical point of view. There have been a number of suggestions[4-10] about the construction of locally supersymmetric gauge theories both in ordinary space—time and in superspace. One would therefore want to have an understanding of the meaning of local gauge group elements in superspace. It is shown in Sec. III that a local gauge group element in superspace can be written in terms of a set of elements of a *different* local gauge group in space—time, which is based on one of the new algebras. We use the notation and conventions of Refs. 5, 7, and 10.

II. THE NEW SUPERALGEBRAS

Let G be a Lie algebra with basis elements $\{X_A\}$ satisfying the commutation relations

$$[X_A, X_B] = f_{AB}^C X_C , \tag{2.1}$$

where f_{AB}^C are the structure constants of the associated Lie group. Let $\theta^{\alpha 1}, \theta^{\alpha 2}, \ldots, \theta^{\alpha m}$, be elements of a Grassmann algebra,

$$\{\theta^{\alpha i}, \theta^{\alpha j}\} = 0, \quad i,j = 1, \ldots, N. \tag{2.2}$$

Construct the elements

$$X_A, \quad X_A^i = \theta^{\alpha i} X_A, \quad X_A^{ij} = \theta^{\alpha i}\theta^{\alpha j} X_A, \quad i \neq j,$$
$$X_A^{ij\cdots km} = \theta^{\alpha i}\theta^{\alpha j}\ldots\theta^{\alpha k}\theta^{\alpha m} X_A, \qquad i \neq j \neq \cdots \neq k \neq m. \tag{2.3}$$

By construction all elements $X_A^{ij\cdots}$ involving products of more than N $\theta^{\alpha k}$'s vanish identically. Recall[5] the definition of a superbracket

$$[X_A, X_B\} \equiv X_A X_B - (-)^{\sigma_A \sigma_B} X_B X_A , \tag{2.4}$$

where $\sigma_A = 0$ if X_A is an even element of the algebra, and $\sigma_A = 1$ if X_A is an odd element. Then, by straightforward computation one can show that

$$[X_A^{\cdots ij\cdots}, X_B^{\cdots kl\cdots}\} = \begin{cases} 0 \text{ if there are more } ijkl \text{ indices than } N, \\ f_{AB}^C X_C^{\cdots ij\cdots kl\cdots} \text{ otherwise.} \end{cases} \tag{2.5}$$

The order of the indices $\cdots ij\cdots kl\cdots$ are the same on both sides of the equality sign.

Similarly, if the Lie algebra (2.1) is replaced by a superalgebra

$$[X_A, X_B\} = f_{AB}^C X_C , \tag{2.6}$$

a representation of the elements of the new superalgebra will be of the form given by (2.3). But this time

$$[X_A^{\cdots ij\cdots}, X^{\cdots kl\cdots}\} = \begin{cases} 0 \text{ if there are more } ijkl \text{ indices than } N, \\ (-)^{\sigma_A(\cdots+\sigma_k+\sigma_l+\cdots)} f_{AB}^C X_C^{\cdots ij\cdots kl\cdots} \text{ otherwise.} \end{cases} \tag{2.7}$$

Notice that in this case even for nonvanishing brackets the sign of some structure coefficients will change when X_A is an odd element of the algebra.

As an example, suppose $N = 4$. Then the basis elements of the new algebra are the set

$$\{X_A, X_A^i, X_A^{ij}, X_A^{ijk}, X_A^{ijkl}\}. \tag{2.8}$$

The nonvanishing superbrackets are listed below:

$$\begin{aligned} &[X_A, X_B\} = f_{AB}^C X_C, \quad [X_A, X_B^j\} = (-)^{\sigma_A \sigma_j} f_{AB}^C X_C^j , \\ &[X_A, X_B^{jk}\} = f_{AB}^C X_C^{jk}, \quad [X_A, X_B^{jkl}\} = (-)^{\sigma_A \sigma_j} f_{AB}^C X_C^{jkl}, \\ &[X_A, X_B^{jklm}\} = f_{AB}^C X_C^{jklm}, \quad [X_A^i, X_B^j\} = (-)^{\sigma_A \sigma_j} f_{AB}^C X_C^{ij}, \\ &[X_A^i, X_B^{jk}\} = f_{AB}^C X_C^{ijk}, \quad [X_A^i, X_B^{jkl}\} = (-)^{\sigma_A \sigma_j} f_{AB}^C X_C^{jkl}, \\ &[X_A^{ij}, X_B^{kl}\} = f_{AB}^C X^{ijkl} . \end{aligned} \tag{2.9}$$

A typical vanishing bracket is

$$[X_A^{ij}, X_B^{klm}\} = 0.$$

Consider some of the subalgebras of the superalgebra (2.8). Clearly, the elements $\{X_A\}$ of the original superalgebra form a subalgebra. The remaining elements, i.e., when excluding $\{X_A\}$, also form a subalgebra, in fact an invariant subalgebra. The latter subalgebra contains a hierarchy of invariant subalgebras obtained by first excluding $\{X_A^i\}$ then $\{X_A^i, X_A^{ij}\}$, etc. These properties are, of course, not limited to this illustrative example but are shared by the general algebras (2.5) and (2.7).

a)Supported in part by the Energy Research and Development Administration.
b)Yale Junior Fellow, on leave of absence from the Department of Physics, Yale University, New Haven, CT 06520.

III. LOCAL GROUP ELEMENTS IN SUPERSPACE

In the neighborhood of the identity any element g of a symmetry group G can be written in the form

$$g = \exp(\epsilon^A X_A),$$

where ϵ^A are the group parameters. A local group element over ordinary space—time can be constructed by requiring that

$$\epsilon^A \to \epsilon^A(x).$$

Thus, one obtains a group with a continuous infinity of parameters, one for each $\{x^\mu\}$. At every value of $\{x^\mu\}$, the algebra of the group remains the same,

$$[X_A, X_B] = f_{AB}^C X_C, \quad \text{for every } \{x^\mu\}. \tag{3.1}$$

Now suppose one wants to construct a local group element over superspace. Then one must require

$$\epsilon^A \to \epsilon^A(x, \theta),$$

where θ is an element of Grassmann algebra (2.2). For definiteness suppose the set $\{\theta^{\alpha i}\}$ consists of four elements. Then, suppressing the x dependence, the exponent $\epsilon^A(\theta) X_A$ can be expanded in powers of θ,

$$\epsilon^A(\theta) X_A = \epsilon^A X_A + \epsilon^A_\alpha X_A^\alpha + \epsilon^A_{\alpha\beta} X_A^{\alpha\beta} + \epsilon^A_{\alpha\beta\gamma} X_A^{\alpha\beta\gamma} + \epsilon^A_{\alpha\beta\gamma\delta} X_A^{\alpha\beta\gamma\delta}, \tag{3.2}$$

where, as in (2.3)

$$X_A^\alpha = \theta^\alpha X_A, \quad X_A^{\alpha\beta} = \theta^\alpha \theta^\beta X_A, \quad \text{etc.}$$

Thus one can regard a local group element with generators $\{X_A\}$ and parameters $\{\epsilon^A(x, \theta)\}$ as a set of elements with generators $\{X_A^{\cdots\alpha}\}$ and parameters $\{\epsilon^A_{\cdots\alpha}(x)\}$. The new generators form an algebra of the type discussed in Sec. II.[11]

The above correspondence applies to any local group or supergroup in superspace. Since local gauge transformations in real space—time have definite meaning and physical implications, one is faced with the problem of providing a justification, from the physical point of view, for using local symmetry groups based on the new algebras; the full implications of this observation remains to be explored.

ACKNOWLEDGMENTS

Conversations with F. Gursey and B. Baaquie are gratefully acknowledged. I would also like to thank Professor S. Drell for the hospitality extended to me at the theory group at SLAC, where this work was done.

[1] V.G. Kac, Commun. Math. Phys. **53**, 31 (1977). References to earlier works of Kac and others are given in this work.
[2] P.G.O. Freund and I. Kaplanski, J. Math. Phys. **17**, 228 (1976).
[3] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626, 1640 (1976).
[4] L.N. Chang, K. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976).
[5] F. Mansouri, J. Math. Phys. **18**, 52 (1977).
[6] P. Nath and R. Arnowitt, Phys. Lett. B **56**, 81 (1975).
[7] S.W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[8] M. Gell'Mann *et al.*, private communication.
[9] J. Wess and B. Zumino, Phys. Lett. B **66**, 361 (1977).
[10] F. Mansouri, SLAC-PUB-1940, April, 1977, "Superunified theories based on the geometry of local (super-) gauge invariance," to be published in Phys. Rev. D **15**, (Oct., 1977).

[11] Of course, this does not mean that the groups based on the two algebras are identical. Since, in obtaining the new superalgebras, the number field is changed from real or complex to Grassman, these algebras can have representations other than those given by (2.3).

The diffraction of sound pulses by a circular cylinder[a)]

Pocheng Chen and Yih-Hsing Pao

Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853
(Received 23 November 1976)

The diffraction of pulses in acoustic medium (scalar waves) by a circular cylinder is analyzed by applying the Cagniard method. Solutions for the incident, reflected, diffracted, and creeping pulses in the illuminated and shadow zones are all obtained by a unified approach. Numerical results are shown for the forward, backward, and side scattering of an incident pulse with a step or square time function.

INTRODUCTION

The scattering of acoustic pulses by smooth obstacle has been investigated by many authors.[1-5] A simple model which reveals important features of scattering is a circular cylinder whose axis is parallel to a line source.

The commonly used method of analysis is as follows: The steady state scattered waves are first expanded in a series of eigenfunctions; the sum of the eigenfunctions is converted to a series of complex integrals by applying Poisson's summation formula; each integral is then evaluated either by the calculus of residues or by the method of saddle point. An integration of the steady-state waves over all frequencies yields the solution for the pulses scattered by the circular cylinder. This last step is often replaced by applying the technique of Laplace transform.

This paper presents an alternative approach. The complex integral and the integration over all frequencies are evaluated jointly with the aid of Cagniard's method.[6] A suitable asymptotic approximation for Bessel functions[7] is adopted such that the Cagniard's method can be applied to obtain a closed form solution for the scattered pulses. This method avoids the tedious calculation of the residues for Bessel functions with complex order. It unifies the analyses of diffracted waves in the shadow zone and in the lit zone; and delineates the arrival times of all pulses exactly as predicted by Fermat's principle.

Gilbert and Helmberger[8] used a similar method in the problem of an spherical inclusion. However, they used a modified Cagniard method,[9] and considered only the pulses reflected by the illuminated side of the sphere.

The final form of the solution is expressed in terms of incident, reflected, and diffracted rays in lit and shadow zones. The last named ray includes the pulses diffracted into the shadow zone by the curved edge, and the pulses that creep around the circumference of the cylinder. The incident ray contributes to the total waves in both lit and shadow zones. Its arrival is marked by a weak discontinuity which resembles the "contact discontinuity" of the shock waves in nonlinear acoustics.[10]

2. WAVE EQUATION AND SOLUTION BY LAPLACE TRANSFORM

The problem under consideration is the scattering of cylindrical waves by an acoustically weak or hard circular cylinder of radius $r=a$ as shown in Fig. 1. A line source which generates the cylindrical wave is located at $r=r'$, $\theta=0$. The time function for the source is $f(t)$. The initial condition is that, at $t=0$, the system is at rest, and the boundary condition at $r=a$ is either the Neumann (weak cylinder) or Dirichlet condition (hard cylinder). As $r\to\infty$, the solution should satisfy Sommerfeld's radiation condition.

Within the framwork of linear theory, the velocity potential ϕ for the acoustic wave satisfies the wave equation:

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\delta(\mathbf{r}-\mathbf{r}')f(t), \tag{1}$$

where $c=(B/\rho)^{1/2}$ is the sound speed which depends on the modulus of compressibility B and density ρ of the medium; $\delta(\mathbf{r})=\delta(r)\delta(\theta)/r$ is the delta function. The ϕ is related to pressure p and particle velocities $\mathbf{v}$ by

$$\mathbf{v}=\nabla\phi, \quad p=\rho\frac{\partial\phi}{\partial t}. \tag{2}$$

In the sequel, we set

$$\phi(r,\theta,t)=\int_0^t f(t-\tau)\psi(r,\theta,\tau)\,d\tau, \tag{3}$$

where $\psi(r,\theta,t)$ is the solution of Eq. (1) when $f(t)=\delta(t)$. Define the Laplace transform of $\psi(t)$ by $\Psi(s)$,

$$\Psi(s)=\int_0^\infty \psi(t)\exp(-st)\,dt.$$

The $\Psi(s)$ then satisfies the reduced wave equation

$$\nabla^2\Psi(s)-(s^2/c^2)\Psi(s)=-\delta(\mathbf{r}-\mathbf{r}'). \tag{4}$$

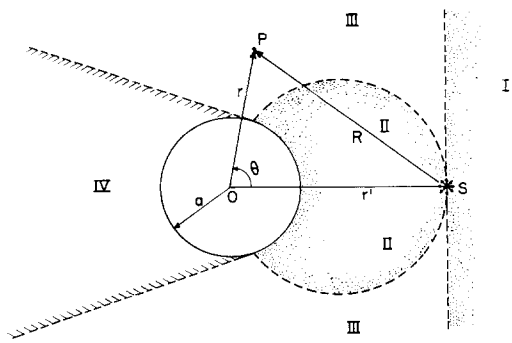


FIG. 1. Lit zone (I, II, III) and shadow zone (IV) outside a cylindrical scatterer.

[a)]This research is supported by a grant of National Science Foundation (ENG 75-13703).

The initial conditions of $\psi(\mathbf{r}, t)=0$ and $\partial\psi/\partial t=0$ at $t=0$ have been assumed in deriving Eq. (4). The boundary condition for $\psi(r, \theta, t)$ or $\Psi(r, \theta, s)$ are

$$\text{Dirichlet condition: } \Psi(r, \theta, s)=0 \text{ at } r=a, \tag{5a}$$

$$\text{Neumann condition: } \partial\Psi(r, \theta, s)/\partial r=0 \text{ at } r=a. \tag{5b}$$

The solution for Eq. (4) is composed of two parts, a particular solution Ψ^i, which corresponds to the incident wave generated by the source, and a homogeneous solution Ψ^s which represents the scattered wave. They are

$$\Psi^i(r, \theta, s)=\frac{1}{2\pi}K_0(kR), \tag{6}$$

$$\Psi^s(r, \theta, s)=\frac{1}{2\pi}\sum_{m=0}^{\infty}\epsilon_m\frac{-\Omega I_m(ka)}{\Omega K_m(ka)}K_m(kr')K_m(kr)\cos m\theta, \tag{7}$$

where ϵ_m is the Neumann factor with $\epsilon_m=1$ when $m=0$ and $\epsilon_m=2$ when $m>0$. In the above equation, I_m and K_m are the mth order modified Bessel functions of the first and second kinds respectively, and

$$k=s/c, \quad R=\sqrt{(r^2+r'^2-2rr'\cos\theta)}. \tag{8}$$

The operator Ω is defined according to boundary conditions (5),

$$\text{Dirichlet: } \Omega\equiv 1, \tag{9a}$$

$$\text{Neumann: } \Omega\equiv\partial/\partial(ka). \tag{9b}$$

In deriving the solutions, the addition theorem[11] of Bessel functions has been applied to express Eq. (6) as

$$\Psi^i(r, \theta, s)=\frac{1}{2\pi}\sum_{m=0}^{\infty}\epsilon_m K_m(kr_>)I_m(kr_<)\cos m\theta, \tag{10}$$

where $r_>=\max(r, r')$ and $r_<=\min(r, r')$. The total wave outside the cylinder is

$$\Psi(r, \theta, s)=\Psi^i(r, \theta, s)+\Psi^s(r, \theta, s). \tag{11}$$

The inverse Laplace transform of Ψ^i in Eq. (6) is[12]

$$\psi^i(R, t)=\frac{1}{2\pi}\frac{c}{\sqrt{(c^2t^2-R^2)}}H(ct-R), \tag{12}$$

where $H(t)$ is the Heaviside step function.

In the lit zone (I, II, and III in Fig. 1), R is the least-time path between the source $(r', 0)$ and receiver (r, θ), and ψ^i represents the influence due to the source alone. However, the cylindrical wavefront as defined by Eq. (12) does not reach the shadow zone (IV in Fig. 1), because the ray emitted from the source is diffracted by the cylinder. The effect of this diffracted source ray in the shadow zone may be determined by an inverse transform of the series representation of $\Psi^i(s)$ as given by Eq. (10).

3. SUMMATION OF THE SERIES BY POISSON'S FORMULA

In practice, for an accurate evaluation of the series in Eqs. (7) and (10) the number of terms that must be retained should be larger than ka. Therefore, for harmonic waves with high frequencies or for pulse with short duration, it is convenient to convert them to another series by applying the Poisson formula[13]

$$\sqrt{\alpha}\sum_{m=0}^{\infty}\epsilon_m f(m\alpha)=\sqrt{\beta}\sum_{n=0}^{\infty}\epsilon_n F_c(n\beta), \tag{13}$$

where $\alpha\beta=2\pi$, $\alpha>0$, and

$$F_c(x)=\sqrt{2/\pi}\int_0^{\infty}f(\nu)\cos(x\nu)\,d\nu. \tag{14}$$

Applying the Poisson formula to (7), (10), (11) and setting $\alpha=1$, $\beta=2\pi$, we obtain for $-\pi\le\theta\le\pi$

$$\Psi(r, \theta, s)=\frac{1}{\pi}\sum_{n=0}^{\infty}\epsilon_n\left\{\int_0^{\infty}[E_g(\nu)+E_h(\nu)]\times\cos\nu\theta\cos 2n\pi\nu\,d\nu\right\}, \tag{15}$$

where

$$\begin{aligned}E_g(r, s;\nu)&=K_\nu(sr_>/c)I_\nu(sr_</c),\\ E_h(r, s;\nu)&=-\frac{\Omega I_\nu(sa/c)}{\Omega K_\nu(sa/c)}K_\nu(sr'/c)K_\nu(sr/c).\end{aligned} \tag{16}$$

The product of cosine function in Eq. (15) may further be reduced to

$$\begin{aligned}\epsilon_n\cos\nu\theta\cos 2n\pi\nu&=\mathrm{Re}[\exp(i\nu\theta_n)+\exp(i\nu\theta_n')] \quad n>0,\\ &=\mathrm{Re}\exp(i\nu\theta), \qquad\qquad n=0,\end{aligned} \tag{17}$$

where Re means "the real part of" and

$$\theta_n=2n\pi+\theta, \quad \theta_n'=-2n\pi+\theta. \tag{18}$$

It will be shown later that each term in the n series of Eqs. (15) and (17) represent the propagation of a pulse. Because of symmetry, we consider only the region $0\le\theta\le\pi$. It then follows that $\theta_n>0$ and $\theta_n'<0$. The terms associated with θ_n in Eq. (18) represent counterclockwise propagation of a wave around the circular cylinder, n being the number of complete turns, whereas the terms associated with θ_n' represent clockwise wave motion, $n-1$ being the number of complete turns. Since the arrival time of each pulse is longer when n is larger, the series in Eq. (15) may be truncated at will if pulses with later arrival times are excluded from the solution.

4. EARLY TIME APPROXIMATIONS

The inverse Laplace transform of $\Psi(r, \theta, s)$ in Eq. (15) is very difficult to evaluate if an exact solution is sought. In the literature, the integrals in Eq. (15) are evaluated by calculus of residues in the ν plane for diffracted waves, and by the method of saddle points for reflected waves in the illuminated zone. At this stage, the analytic properties of functions K_ν and I_ν in complex ν plane must be carefully examined and the poles of the denominator $\Omega K_\nu(sa/c)$ must be calculated with great accuracy.[1-5] The integral in s for the inverse transform is then evaluated either numerically, or analytically when all modified Bessel functions in E_g and E_h are replaced by their asymptotic representation for large s. For a certain cases, the exact inverse Laplace transforms of the asymptotic functions can be found.[3,4,12]

The main purpose of this paper is to show that if the same asymptotic representations of Bessel function are substituted into Eq. (16) at an early stage, the inverse Laplace transform of the $\Psi(r, \theta, s)$ in asymptotic form can be completed by applying the Cagniard method. The final answer, for a simple time function like $\delta(t)$, is in closed form. This analysis, we believe, is numerically more accurate than the existing ones, and avoids the tedious calculation of residues. Furthermore, it delineates the arrival times of all pulses exactly as predicted by the theory of geometric acoustics (ray theory). Finally, it unifies the treatments of diffracted waves in the shadow zone and in the lit zone.

The asymptotic representations of the modified Bessel functions to be used are[7]

$$I_\nu(\nu z) \sim \frac{1}{\sqrt{2\pi\nu}} \frac{\exp(\nu\eta)}{(1+z^2)^{1/4}}[1+O(\nu^{-1})],$$
$$K_\nu(\nu z) \sim \left(\frac{\pi}{2\nu}\right)^{1/2} \frac{\exp(-\nu\eta)}{(1+z^2)^{1/4}}[1+O(\nu^{-1})], \tag{19}$$

$$I_\nu'(\nu z) \sim \frac{1}{\sqrt{2\pi\nu}} \frac{(1+z^2)^{1/4}}{z} \exp(\nu\eta)[1+O(\nu^{-1})],$$
$$K_\nu'(\nu z) \sim -\left(\frac{\pi}{2\nu}\right)^{1/2} \frac{(1+z^2)^{1/4}}{z} \exp(-\nu\eta)[1+O(\nu^{-1})], \tag{20}$$

where

$$\eta = \sqrt{(1+z^2)} - \sinh^{-1}(1/z). \tag{21}$$

Substituting (19)–(21) into Eq. (15) and setting $\xi = c\nu/s$, we obtain

$$\Psi^i(r, \theta, s) \sim \frac{1}{\pi}\mathrm{Re}\int_0^\infty E_g(\xi, r)\exp[-sg(\xi, \theta)]\,d\xi + \frac{1}{\pi}\sum_{n=1}^{\infty}\mathrm{Re}\int_0^\infty E_g(\xi, r)\{\exp[-sg(\xi, \theta_n)] + \exp[-sg(\xi, \theta_n')]\}\,d\xi, \tag{22a}$$

$$\Psi^s(r, \theta, s) \sim -\frac{1}{\pi}\mathrm{Re}\int_0^\infty E_h(\xi, r)\exp[-sh(\xi, \theta)]\,d\xi - \frac{1}{\pi}\sum_{n=1}^{\infty}\mathrm{Re}\int_0^\infty E_h(\xi, r)\{\exp[-sh(\xi, \theta_n)] + \exp[-sh(\xi, \theta_n')]\}\,d\xi. \tag{22b}$$

Each integral in the series is called a *ray*, which, as shown later, represents a pulse propogating along a *ray path*.

In Eq. (22), the two phase functions are

$$g(\xi, \theta_n) = c^{-1}[-i\xi\theta_n + \sqrt{(\xi^2+r_>^2)} - \sqrt{(\xi^2+r_<^2)} + \xi(\sinh^{-1}\xi/r_< - \sinh^{-1}\xi/r_>)], \tag{23}$$

$$h(\xi, \theta_n) = c^{-1}[-i\xi\theta_n + \sqrt{(\xi^2+r^2)} + \sqrt{(\xi^2+r'^2)} - 2\sqrt{(\xi^2+a^2)} + \xi(2\sinh^{-1}\xi/a - \sinh^{-1}\xi/r - \sinh^{-1}\xi/r')]. \tag{24}$$

The amplitude for the source ray is

$$E_g(\xi, r) \approx \tfrac{1}{2}(\xi^2+r'^2)^{-1/4}(\xi^2+r^2)^{-1/4}. \tag{25}$$

The $E_h(\xi)$ which is the amplitude of the scattered waves is different for the two types of boundary condition. However, when terms up to $O(\nu^{-1})$ are considered, they differ only by a sign,

$$\text{Dirichlet:}\quad E_h(\xi) \approx E_g(\xi), \tag{26a}$$

$$\text{Neumann:}\quad E_h(\xi) \approx -E_g(\xi). \tag{26b}$$

In deriving Eq. (22), we noticed that the amplitude functions E_g and E_h, after the substituting of the asymptotic formula for I_ν and K_ν, are homogeneous functions of $c\nu/s$. A change of variable $\xi = c\nu/s$ thus removes the s from $E(r, s; \nu)$, and the Laplace transform parameter appears only in the exponential functions as $-sg(\xi)$, and $-sh(\xi)$.

Furthermore, if one is able to transform the complex variable ξ to another variable t, such that $g(\xi, \theta) = t$ or $h(\xi, \theta) = t$, the integrals in Eq. (22) can all be expressed as

$$\int_{t_A}^\infty E[\xi(t), r]\exp(-st)(d\xi/dt)\,dt.$$

The inverse Laplace transformation of this integral is, by inspection,

$$E[\xi(t), r](d\xi/dt)H(t-t_A),$$

where $H(t)$ is the step function in time. This is, in essence, the *Cagniard method*. In the next two sections, we present separately the transformations for the scattered rays and the incident rays.

5. EVALUATION OF SCATTERED RAYS

In Eq. (22b), the scattered waves are represented by terms with E_h. A typical ray integral is

$$\Psi_n^s(r, \theta, s) = (1/\pi)\mathrm{Re}[\textstyle\int_0^\infty E_h(\xi, r)\exp(-sh(\xi, \theta_n))d\xi]$$
$$n = 0, 1, 2, \cdots. \tag{27}$$

In the phase function $h(\xi, \theta_n)$, the θ_n may be replaced by θ or θ_n'. We shall find the inverse Laplace transform of $\Psi_n(s)$ by setting in Eq. (27)

$$t = h(\xi, \theta_n). \tag{28}$$

Assume, for the moment, that an inverse transformation

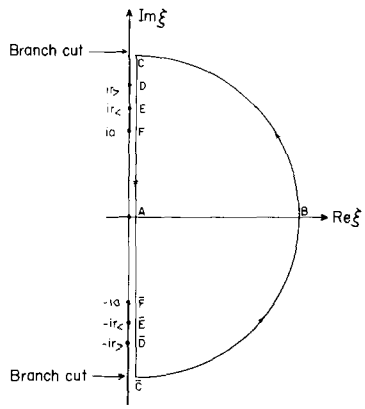


FIG. 2. Complex ξ plane and branch cuts for the phase function $h(\xi, \theta_n)$.

TABLE I. Mapping of the imaginary ξ axis ($\xi = \pm il$) onto the complex t plane for function $t = h(\xi, \theta_n)$.

ξ plane	$t = h(\eta, \theta_n)$
$\bar{C} \to B \to C$	$\xi[\ln(rr'/a^2) - i\theta_n]$
$C \to D$	$l\theta_n + i\{\sqrt{(l^2 - r^2)} + \sqrt{(l^2 - r'^2)} - 2\sqrt{(l^2 - a^2)} - l[\cosh^{-1}(l/r) + \cosh^{-1}(l/r') - 2\cosh^{-1}(l/a)]\}$
D	$r_>\theta_n + i[\sqrt{(r_>^2 - r_<^2)} - 2\sqrt{(r_>^2 - a^2)} - r_>\cosh^{-1}(r_>/r_<) + 2r_>\cosh^{-1}(r_>/a)]$
$D - E$	$\sqrt{(r_>^2 - l^2)} + l\sin^{-1}(l/r_>) + l(\theta_n - \pi/2) + i[\sqrt{(l^2 - r_<^2)} - 2\sqrt{(l^2 - a^2)} - l\cosh^{-1}(l/r_<) + 2l\cosh^{-1}(l/a)]$
E	$\sqrt{(r_>^2 - r_<^2)} + r_<\sin^{-1}(r_</r_>) + r_<(\theta_n - \pi/2) + i[2r_<\cosh^{-1}(r_</a) - 2\sqrt{(r_<^2 - a^2)}]$
$E \to F$	$\sqrt{(r^2 - l^2)} + \sqrt{(r'^2 - l^2)} + l[\sin^{-1}(l/r')] + l(\theta_n - \pi) + i[2l\cosh^{-1}(l/a) - 2\sqrt{(l^2 - a^2)}]$
F	$\sqrt{(r^2 - a^2)} + \sqrt{(r'^2 - a^2)} + a[(\sin^{-1}(a/r) + \sin^{-1}(a/r')] + a(\theta_n - \pi)$
$F \to A$	$\sqrt{(r^2 - l^2)} + \sqrt{(r'^2 - l^2)} - 2\sqrt{(a^2 - l^2)} + l[\sin^{-1}(l/r) + \sin^{-1}(l/r') - 2\sin^{-1}(l/a) + \theta_n]$
A	$r + r' - 2a$
$A \to \bar{F}$	$\sqrt{(r^2 - l^2)} + \sqrt{(r'^2 - l^2)} - 2\sqrt{(a^2 - l^2)} + l[\sin^{-1}(l/r) + \sin^{-1}(l/r') - 2\sin^{-1}(l/a) - \theta_n]$
$\bar{F}$	$\sqrt{(r^2 - a^2)} + \sqrt{(r'^2 - a^2)} + a[\sin^{-1}(a/r) + \sin^{-1}(a/r')] + a(-\theta_n - \pi)$
$\bar{F} \to \bar{E}$	$\sqrt{(r^2 - l^2)} + \sqrt{(r'^2 - l^2)} + l[\sin^{-1}(l/r) + \sin^{-1}(l/r')] - l(\theta_n + \pi) - i[2l\cosh^{-1}(l/a) - 2\sqrt{(l^2 - a^2)}]$
$\bar{E}$	$\sqrt{(r_>^2 - r_<^2)} + r_<\sin^{-1}(r_</r_>) + r_<(\theta_n - \pi/2) - i[2r_<\cosh^{-1}(r_</a) - 2\sqrt{(r_<^2 - a^2)}]$
$\bar{E} \to \bar{D}$	$\sqrt{(r_>^2 - l^2)} + l\sin^{-1}l/r_> - l(\theta_n + \pi/2) - i[\sqrt{(l^2 - r_<^2)} - 2\sqrt{(l^2 - a^2)} - l\cosh^{-1}(l/r_<) + 2l\cosh^{-1}(l/a)]$
$\bar{D}$	$-r_>\theta_n + i[-\sqrt{(r_>^2 - r_<^2)} + 2\sqrt{(r_>^2 - a^2)} + r_>\cosh^{-1}(r_>/r_<) - 2r_>\cosh^{-1}(r_>/a)]$
$\bar{D} \to \bar{C}$	$-l\theta_n - i\{\sqrt{(l^2 - r^2)} + \sqrt{(l^2 - r'^2)} - 2\sqrt{(l^2 - a^2)} - l[\cosh^{-1}(l/r) + \cosh^{-1}(l/r') - 2\cosh^{-1}(l/a)]\}$

of the variable from ξ to t can be found,

$$\xi = h^{-1}(t, \theta_n). \tag{29}$$

For a complex variable ξ, $h(\xi, \theta_n)$ is multivalued with branch points at $\pm ir'$, $\pm ir$, and $\pm ia$, as shown in Fig. 2. We render it a single-value function by introducing branch cuts along the imaginary axis, starting from $\pm i\infty$. The branches are chosen such that if ξ is real and positive, the radicals are positive and inverse hyperbolic sines assume the principal values. The right half plane of $\xi(-\pi/2 < \arg\xi < \pi/2)$ is enclosed by a large hemicircle with contour $\bar{C}BCA\bar{C}$.

As $|\xi| \to \infty$, we find in $h(\xi, \theta_n)$

$$\sqrt{(\xi^2 + r^2)} \to \xi + O(\xi^{-1}), \quad \sinh^{-1}(\xi/r) \to \ln(2\xi/r) + O(\xi^{-2}). \tag{30}$$

Thus the t in (28) approaches, as $|\xi| \to \infty$,

$$t \to \xi[\ln(rr'/a^2) - i\theta_n] \tag{31}$$

The infinite semicircle $\bar{C}BC$ in ξ plane (Fig. 2) is then mapped as $\bar{C}'B'C$ in the complex t plane as shown in Fig. 3(A).

For the phase function $h(\xi, \theta_n')$, a transformation

$$t = h(\xi, \theta_n') \tag{32}$$

is introduced. It then maps the semicircle $\bar{C}BC$ onto $\bar{C}'B'C'$ in Fig. 3(B). Since these two cases are mathematically analogous, we shall consider in the sequel the case of $h(\xi, \theta_n)$ only.

To determine the mapping of the imaginary axis of ξ, we set $\xi = il$ along CA and $\xi = -il$ along $A\bar{C}$ in Eq. (24). The results are shown in Table I. The corresponding curves on t-plane are shown for two different cases in Figs. 4 and 5.

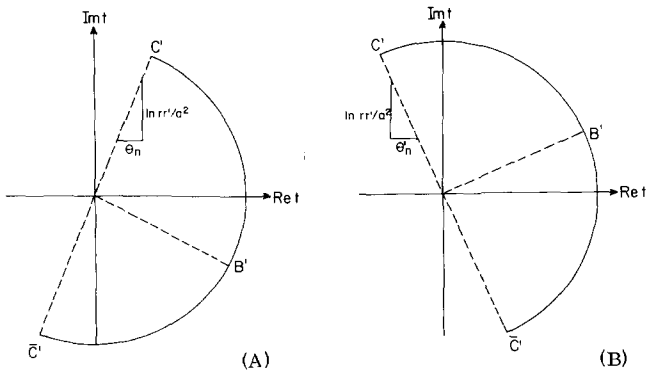


FIG. 3. (A) Complex t plane and mapping of hemicircle $\bar{C}BC$, when $\theta = \theta_n$. (B) Complex t plane and mapping of hemicircle $CB\bar{C}$, when $\theta = \theta' n$.

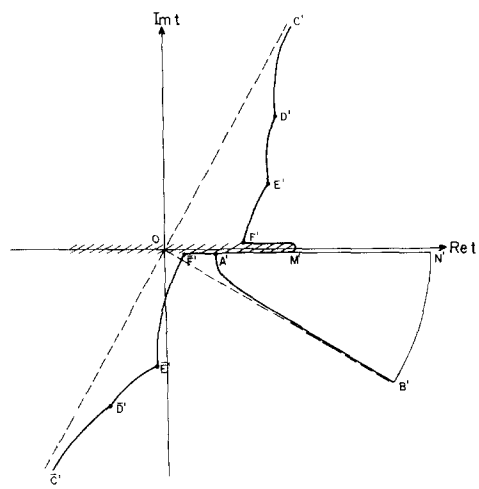


FIG. 4. Mapping of imaginary ξ axis with a stationary point M' and the integration contour AB in t plane.

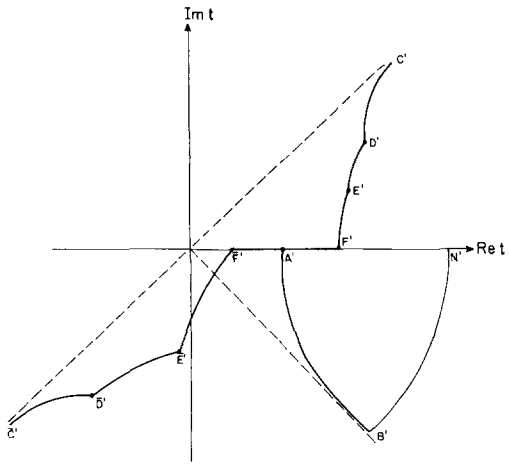


FIG. 5. Mapping of imaginary ξ axis without a stationary pooint and the integration contour AB in t plane.

From Table I, it is seen that the mapping of C to F and $\bar{F}$ to $\bar{C}$ do not cross the real axis of t plane, since their imaginary parts never vanish because $r\cosh^{-1}(l/r) > \sqrt{(l^2 - r^2)}$ and $l > r > 0$.

The mapping of $FA\bar{F}(F'A'\bar{F}')$ is real in the t plane and may be double valued along the real t aixs, if there exists a stationary point M' in between F' and $\bar{F}'$ (Fig. 4). A branch cut is thus introduced along the real t axis. At this stationary point, the value of the function $h(\xi, \theta_n)$ is at its extremum, and it gives rise to the *reflected ray*.

When there exists no stationary point, the mapping of $FA\bar{F}$ is continuous and single valued (Fig. 5). This mapping gives rise to the *diffracted rays*. Thus for either case, every point on the right half ξ plane is uniquely mapped onto a point on the t plane bounded by the closed contour $C'F'\bar{F}'\bar{C}'B'C'$, and an inverse transformation of Eq. (29) is assured.

A. The reflected ray

To determine the extreme values, we set along FA

$$\frac{dh(il, \theta_n)}{dl} = \theta_n - \left(\cos^{-1}\frac{l}{r'} + \cos^{-1}\frac{l}{r} - 2\cos^{-1}\frac{l}{a}\right) = 0 \tag{33a}$$

and along $A\bar{F}$

$$\frac{dh(-il, \theta_n)}{dl} = -\theta_n - \left(\cos^{-1}\frac{l}{r'} + \cos^{-1}\frac{l}{r} - 2\cos^{-1}\frac{l}{a}\right) = 0. \tag{33b}$$

There exists a root for either Eqs. (33a) or (33b), if and only if $|\theta_n| < \cos^{-1}a/r' + \cos^{-1}a/r$. We denote the root, if it exists, as $l = d$, at which the function h is extreme. Furthermore, since the principal values of arccosines are taken, Eq. (33) can only be satisfied when $n = 0$ $(\theta_n \to \theta)$. The geometric interpretation of the root $l = d$ is shown in Fig. 6. The d is called the ray parameter.

When $l = d$ is a solution of Eq. (33), the $h(\xi, \theta_n)$ between F and A is simplified to

$$t = t_M = c^{-1}[\sqrt{(r'^2 - d^2)} + \sqrt{(r^2 - d^2)} - 2\sqrt{(a^2 - d^2)}]. \tag{34}$$

As can be easily seen from Fig. 6, t_M is the arrival time for a pulse along the reflected ray path.

Note that only the first integral of Ψ^s in Eq. (22b) $[n = 0$ in Eq. (27)] contributes to the pulse directly reflected by the cylinder in the lit zone. This integral is then transformed into one with respect to t,

$$\Psi_0^s(r, \theta, s) = \frac{1}{\pi}\mathrm{Re}\int_{A'B'} E_h[\xi(t), r]\frac{\partial \xi(t)}{\partial t}\exp(-st)\,dt, \tag{35}$$

where the path of integration is along the $A'B'$ curve shown in Fig. 4. Since there is no singularity for the integrand in the region bounded by the curve $A'B'N'M'A'$, and the integrand vanishes along the arc $B'N'$ when it is removed to infinity, the path of integral of Eq. (35) may be changed to $A'M'N'$ along the real t aixs:

$$\Psi_0^s(r, \theta, s) = \frac{1}{\pi}\mathrm{Re}\int_{t_A}^{\infty} E_h[\xi(t), r]\frac{\partial \xi}{\partial t}\exp(-st)\,dt. \tag{36}$$

By inspection, the inverse Laplace transform is

$$\psi_0^s(r, \theta, t) = \frac{1}{\pi}\mathrm{Re}\left(E_h[\xi(t), r]\frac{\partial \xi(t)}{\partial t}\right)H(t - t_A). \tag{37}$$

Furthermore, between $t = t_A$ and t_M, the product $E_h(\partial\xi/\partial t)$ is imaginary. Hence we obtain in the lit zone (Regions I, II, III),

$$\psi_0^s(r, \theta, t) = \frac{1}{\pi}\mathrm{Re}\left(E_h[\xi(t, \theta), r]\frac{\partial \xi(t, \theta)}{\partial t}\right)H(t - t_M). \tag{38}$$

It is seen that the Cagniard method has been applied successfully to invert the Laplace transforms. The $\xi(t, \theta)$ function is given by Eq. (29), which in this case can be evaluated numerically; $\partial\xi(t, \theta)/\partial t$ is obtained by differentiating Eq. (28). The time t_M is given by Eq. (34).

B. Diffracted rays

If Eq. (33) cannot be satisfied for a given angle θ_n, the phase function $h(\xi, \theta_n)$ has no stationary point. We define an angle δ_n,

$$\delta_n = |\theta_n| - \cos^{-1}\frac{a}{r'} - \cos^{-1}\frac{a}{r}. \tag{39}$$

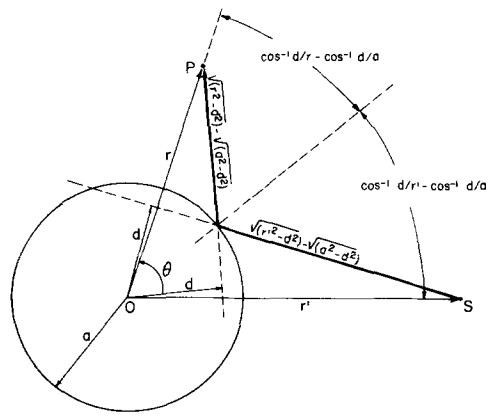


FIG. 6. Geometry of reflected ray.

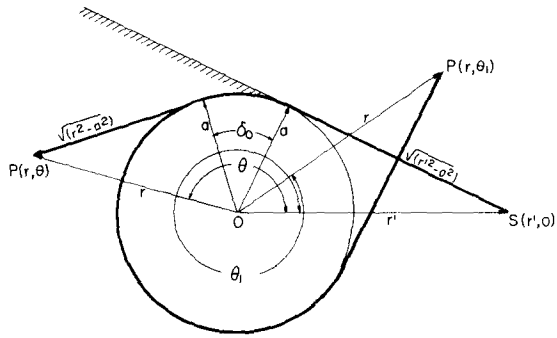


FIG. 7. Geometry of diffracted rays for observer P in the shadow zone and lit zone.

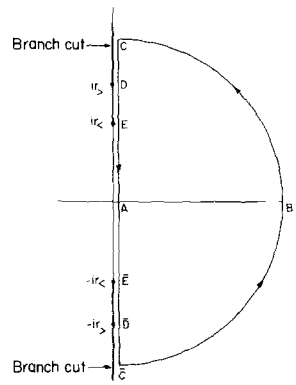


FIG. 8. Complex ξ-plane and branch cuts for the function $g(\xi, \theta_n)$.

For $n=0$, the δ_0 is the angle subtended by the ray diffracted around the cylinder into the shadow zone IV as shown in Fig. 7. This ray is diffracted continuously into the lit zone in lower half plane as θ increases. For $n \geq 1$, δ_n is the angle of a ray path creeping around the cylinder n times before it reaches the receiver (Fig. 7), which may be either in the shadow or in the lit zone.

From Table I, the t at the point F or F' is

$$t_F = c^{-1}[\sqrt{(r'^2 - a^2)} + \sqrt{(r^2 - a^2)} + \delta_n a]. \tag{40}$$

It is the time taken by a pulse to travel from S to P along the path shown in Fig. 7.

The inverse transform of Ψ_n^s in Eq. (27) is then evaluated as that of Ψ_0^s in Eq. (35). The answer is

$$\psi_n^s(r, \theta_n, t) = \frac{1}{\pi} \mathrm{Re}\left(E_n[\xi(t, \theta_n), r] \frac{\partial \xi(t, \theta_n)}{\partial t}\right) H(t - t_F),$$

$$n = 0, 1, \cdots. \tag{41}$$

The $\xi(t, \theta_n)$ function is given by Eq. (29), and $\partial \xi(t, \theta_n)/\partial t$ is obtained by differentiating Eq. (28). The time t_F is given by Eq. (40).

6. EVALUATION OF THE INCIDENT RAY

Since the inverse transform of the source ray $\Psi^i(s)$ in Eq. (6) can be obtained exactly, the transient waves in the lit zone (I, II, III) is known in closed form as given by Eq. (12). However, we shall investigate the inverse transform of the eigenfunction expansion for $\Psi^i(s)$ in Eq. (10). The asymptotic value for $\Psi^i(s)$ at large s is given in Eq. (22a), a typical integral of which is

$$\Psi_n^i(s) = \frac{1}{\pi} \mathrm{Re} \int_0^\infty E_g(\xi, r) \exp[-sg(\xi, \theta_n)]\, d\xi,$$

$$n = 0, 1, 2, \cdots. \tag{42}$$

We render the function $g(\xi, \theta_n)$ single valued by introducing the branch cuts as shown in Fig. 8. A transformation

$$t = g(\xi, \theta_n) \tag{43}$$

maps the semicircle $CB\bar{C}$ in the ξ plane to a semicircle $C'B'\bar{C}'$ in the t plane. The mapping of the imaginary ξ axis is given in Table II for various intervals. Along $E\bar{E}$, the function $g(\xi, \theta_n)$ is real and it may or may not have a stationary value, depending on the angle θ_n. The mapping of $\mathrm{Im}\,\xi$ on the t plane is similar to that shown in Fig. 4 or Fig. 5. A branch cut along the real t axis (Fig. 4) is introduced if there exists a stationary point for the function $g(\xi, \theta_n)$. Thus the mapping from ξ to t is unique, and an inverse solution $\xi = g^{-1}(t, \theta_n)$ is assured.

TABLE II. Mapping of the imaginary ξ axis ($\xi = \pm il$) onto the complex t plane for function $t = g(\xi, \theta_n)$.

ξ plane	$t = g(\xi, \theta_n)$
$\bar{C} \to B \to C$	$\xi[\ln(r_>/r_<) - i\theta_n]$
$C \to D$	$l\theta_n + i[\sqrt{(l^2 - r_>^2)} - \sqrt{(l^2 - r_<^2)} - l\cosh^{-1}(l/r_>) + l\cosh^{-1}(l/r_<)$
D	$r_>\theta_n + i[-\sqrt{(r_>^2 - r_<^2)} + r_>\cosh^{-1}(r_>/r_<)]$
$D \to E$	$[l(\theta_n - \pi/2 + \sin^{-1}(l/r_>)) + \sqrt{(r_>^2 - l^2)}] + i[-\sqrt{(l^2 - r_<^2)} + l\cosh^{-1}(l/r_<)]$
E	$\sqrt{(r_>^2 - r_<^2)} + r_<\sin^{-1}(r_</r_>) + r_<(\theta - \pi/2)$
$E \to A$	$\sqrt{(r_>^2 - l^2)} - \sqrt{(r_<^2 - l^2)} + l\sin^{-1}(l/r_>) - l\sin^{-1}(l/r_<) + l\theta_n$
A	$r_> - r_<$
$A \to \bar{E}$	$\sqrt{(r_>^2 - l^2)} - \sqrt{(r_<^2 - l^2)} + l\sin^{-1}(l/r_>) - l\sin^{-1}(l/r_<) - l\theta_n$
$\bar{E}$	$\sqrt{(r_>^2 - r_<^2)} + r_<\sin^{-1}(r_</r_>) - r_<(\theta_n - \pi/2)$
$\bar{E} \to \bar{D}$	$\{\sqrt{(r_>^2 - l^2)} + l[\sin^{-1}(l/r_>) - \theta_n - \pi/2]\} + i[\sqrt{(l^2 - r_<^2)} - l\cosh^{-1}(l/r_<)]$
$\bar{D}$	$-r_>\theta_n + i[-\sqrt{(r_>^2 - r_<^2)} - r_>\cosh^{-1}(r_>/r_<)]$
$\bar{D} \to \bar{C}$	$-l\theta_n + i[\sqrt{(l^2 - r_>^2)} - \sqrt{(l^2 - r_<^2)} + l\cosh^{-1}(l/r_>) - l\cosh^{-1}(l/r_<)]$

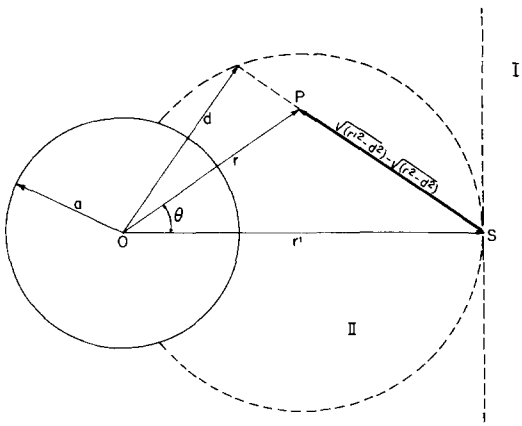


FIG. 9. Geometry of incident ray in Region II.

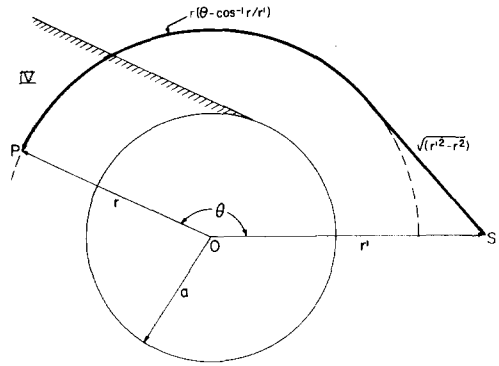


FIG. 10. A ray-path of the incident wave in the shadow zone.

To determine the stationary point, we set, along EA,

$$\frac{dg(il, \theta_n)}{dl} = \theta_n - \left(\cos^{-1}\frac{l}{r_>} - \cos^{-1}\frac{l}{r_<}\right) = 0 \tag{44}$$

and, along $A\bar{E}$,

$$\frac{dg(-il, \theta_n)}{dl} = -\theta_n - \left(\cos^{-1}\frac{l}{r_>} - \cos^{-1}\frac{l}{r_<}\right) = 0. \tag{45}$$

These equations may possess a root, $l = d$, if and only if $|\theta_n| < \cos^{-1}(r_</r_>)$. Since only the principal value of the arccosine is allowed, this condition implies that $\theta_n = \theta(n=0)$, and defines the regions I and II (Figs. 1 and 9), both in the lit zone. The stationary point M' in these regions is

$$t = t_M = c^{-1}[\sqrt{(r_>^2 - d^2)} - \sqrt{(r_<^2 - d^2)}]. \tag{46}$$

It is the arrival time along the incident ray as shown in Fig. 9.

Based on this analysis of stationary point and the Cagniard method discussed in the previous section, we find an approximate value for the first integral of Eq. (22a) in regions I and II,

$$\psi_0^i(r, \theta, t) \approx \frac{1}{\pi}\mathrm{Re}\left(E_g[\xi(t, \theta), r]\frac{\partial \xi(t, \theta)}{\partial t}\right) H(t - t_M). \tag{47}$$

The function $\xi(t, \theta)$ is the inverse transformation of Eq. (43), which can be solved numerically. The $\partial\xi/\partial t$ is given by differentiating Eqs. (43) and (23),

$$\frac{\partial \xi}{\partial t} = c\left[-i\theta + \left(\sinh^{-1}\frac{\xi}{r_<} - \sinh^{-1}\frac{\xi}{r_>}\right)\right]^{-1}. \tag{48}$$

For an observation point in region III of the lit zone and in the shadow zone IV, there exists no stationary point for the function $g(\xi, \theta_n)$. The inverse transform of $\Psi_n^i(s)$ is then evaluated by the same procedure that leads to Eq. (41). The answer is

$$\psi_n^{(i)}(r, \theta_n, t) = \frac{1}{\pi}\mathrm{Re}\left(E_g[\xi(t, \theta_n), r]\frac{\partial \xi(t, \theta_n)}{\partial t}\right) H(t - t_E),$$
$$n = 0, 1, \cdots. \tag{49}$$

where, from Table II,

$$t_E = c^{-1}\left[\sqrt{(r_>^2 - r_<^2)} - r_< \cos^{-1}\frac{r_<}{r_>} + r_<\theta_n\right]. \tag{50}$$

The function $\partial\xi/\partial t$ is the same as that in Eq. (48) with $\theta = \theta_n$. The ray path for $r < r'$ and $n = 0$ in the shadow zone is shown in Fig. 10.

The results for $\psi_n^{(i)}(t)$ in regions III and IV are very interesting. In the shadow zone where the exact solution Eq. (12) is not applicable, the $\psi_n^i(t)$ in Eq. (49) should be combined with the diffracted ray, $\psi_n^s(t)$, in Eq. (41) to yield the total wave $\psi_n(t)$, the arrival time of $\psi_n^s(t)$ is ahead of that of $\psi_n^i(t)$. Thus behind the diffracted wavefront, there follows another signal, $\psi_n^i(t)$, after a time interval $t_E - t_F$. The arrival of the second signal is marked only by a change of the slope of the total pressure, as shown later.

In region III of the lit zone, the locus marked by the arrival of the signal $\psi_n^{(i)}(t)$ falls far behind the incident wavefront as given by the exact representation $\psi^i(t)$. Thus $\sum\psi_n^{(i)}(t)$ should be considered only as an approximation to the incident wave behind the wavefront. This locus in both zone III and zone IV is traced out in Fig. 11, along with the incident [Eq. (46)], reflected [Eq. (34)], and diffracted [Eq. (40)] wavefronts. It is interesting to note that this locus resembles the "contact discontinuity" or "slip-stream front" of the diffracted shock waves in nonlinear acoustics (see pp. 299—305

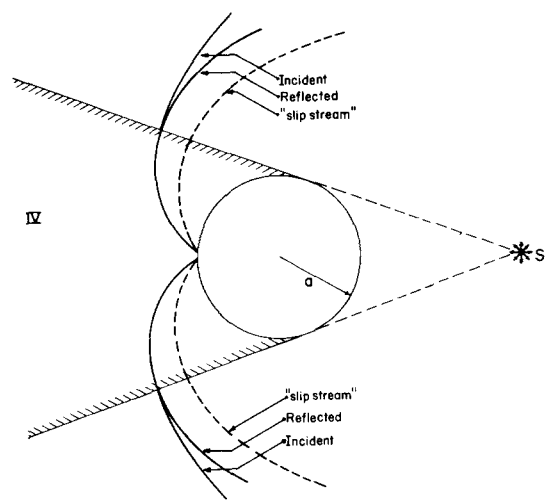


FIG. 11. Wave fronts of incident, reflected, diffracted waves and "slip stream" front.

TABLE III. Arrival times for various rays.

Receiver at $(5a, \pi/2)$—Fig. 12			Receiver at $(5a, \pi)$—Fig. 13		
Rays	Arrival time (ct/a)	Type of rays	Rays	Arrival time (ct/a)	Type of rays
1	7.07	incident	1, 2	10.20	creep (1st counterclockwise and clockwise)
2	8.70	reflected			
3	11.77	creep (1st counterclockwise)	3, 4	16.48	creep (2nd counterclockwise and clockwise)
4	14.91	creep (1st clockwise)			
5	18.05	creep (2nd counterclockwise)	5, 6	22.76	creep (3rd counterclockwise and clockwise
6	21.19	creep (2nd clockwise)			
			7, 8	29.05	creep (4th counterclockwise and clockwise)
			9, 10	35.33	creep (5th counterclockwise and clockwise)

Receiver at $(20a, \pi)$—Fig. 14.			Receiver at $(5a, 0)$—Fig. 15		
Rays	Arrival time (ct/a)	Type of rays	Rays	Arrival time (ct/a)	Type of rays
1, 2	25.12	creep (1st counterclockwise and clockwise)	1	0	source
			2	8	reflected
1s, 2s	28.48	1st slip stream	3	13.34	creep (1st round)
3, 4	31.40	creep (2nd counterclockwise and clockwise)	4	19.62	creep (2nd round)
			5	25.90	creep (3rd round)
5, 6	37.69	creep (3rd counterclockwise and clockwise)	6	32.19	creep (4th round)
7, 8	43.97	creep (4th counterclockwise and clockwise)			
9, 10	50.25	creep (5th counterclockwise and clockwise)			
3s, 4s	59.89	2nd slip stream			

of Ref. 10). At the slip-stream front the pressure of the shock is continuous but the mass density is not.

Strictly speaking, the path of the slip-stream front (Figs. 10 and 11) is not a "ray path" as it does not conform to Fermat's principle. However, we should not construe from this that the results in Eqs. (47) and (49) are in violation of Fermat's principle because $\Psi_n^i(t)$ forms only a part of the total wave which is the only physical quantity measurable at an observation point.

7. SUMMARY

To summarize the results in Secs. 5 and 6, we write, according to Eq. (11), the total wave outside the cylinder as

$$\psi(r, \theta, t) = \psi^i(r, \theta, t) + \psi^s(r, \theta, t), \quad -\pi \leq \theta \leq \pi. \tag{51}$$

Lit zone (Regions I, II, III)

$$\psi^i(r, \theta, t) = \frac{1}{2\pi} \frac{c}{\sqrt{(c^2t^2 - R^2)}} H\left(t - \frac{R}{c}\right), \tag{52}$$

$$\psi^s(r, \theta, t) = \psi_0^s(r, \theta, t) + \sum_{n=1}^{\infty} [\psi_n^s(r, \theta_n, t) + \psi_n^s(r, \theta_n', t)]. \tag{53}$$

Equation (52) is the same as Eq. (12). In Eq. (53), ψ_0^s is given in Eq. (38) which represents the reflected ray (Fig. 6); $\psi_n^s(r, \theta_n, t)$ is given by Eq. (41), which represents the diffracted ray that creeps around the cylinder (Fig. 7); the $\psi_n^s(r, \theta_n', t)$ is obtained from Eq. (41) by changing θ_n to θ_n'.

Approximate values for ψ^i in various regions are:

Regions I and II:

$$\psi^i(r, \theta, t) \approx \psi_0^i(r, \theta, t) + \sum_{n=1}^{\infty} [\psi_n^i(r, \theta_n, t) + \psi_n^i(r, \theta_n', t)]. \tag{54}$$

Region III:

$$\psi^i(r, \theta, t) \approx \sum_{n=1}^{\infty} [\psi_n^i(r, \theta_n, t) + \psi_n^i(r, \theta_n', t)]. \tag{55}$$

The ψ_0^i is given in Eq. (47), and ψ_n^i in Eq. (49). The ψ_0^i in Eq. (54) defines the incident wavefront which coincides with that given by Eq. (52). The locus of $\psi_n^i(r, \theta_n, t)$ for various (r, θ) falls behind the incident wavefront (Fig. 11) and resembles the slip stream in shock waves.

Shadow zone (Region IV)

$$\psi^i(r, \theta, t) = \sum_{n=0}^{\infty} [\psi_n^i(r, \theta_n, t) + \psi_n^i(r, \theta_n', t)], \tag{56}$$

$$\psi^s(r, \theta, t) = \sum_{n=0}^{\infty} [\psi_n^s(r, \theta_n, t) + \psi_n^s(r, \theta_n', t)]. \tag{57}$$

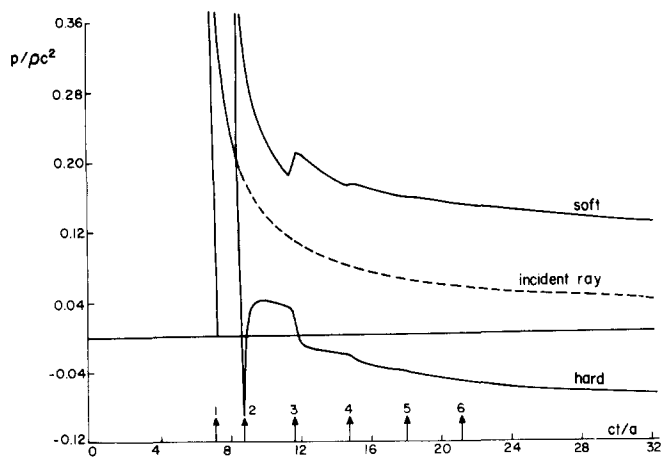


FIG. 12. Side scattering $(5a, \pi/2)$ due to a soft or hard cylinder.

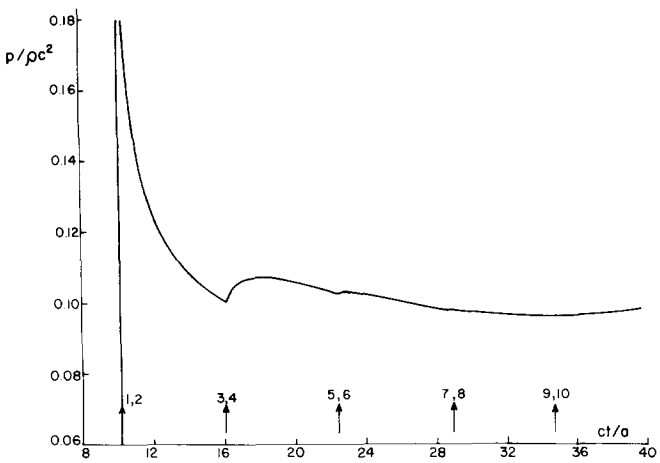


FIG. 13. Forward scattering at $(5a, \pi)$ due to a hard cylinder.

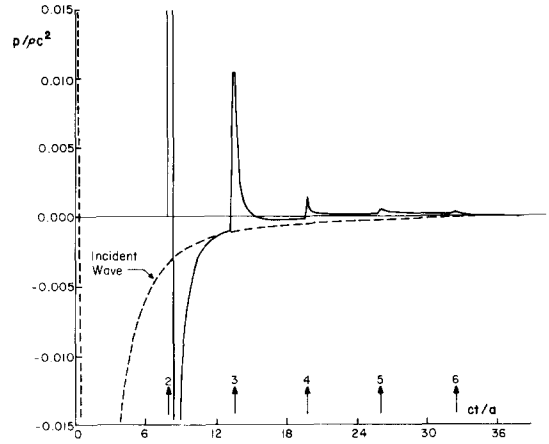


FIG. 15. Backscattering at $(5a, 0)$ by a hard cylinder for an incident square pulse.

The ψ_n^i is given in Eq. (49), and ψ_n^s in Eq. (41). The ψ_n^i defines a slip-stream locus in the shadow zone (Fig. 10), and ψ_n^s gives rise to the diffracted wavefront which wraps around the cylinder (Fig. 7).

8. NUMERICAL EXAMPLES

In this section, we show some numerical results for the pressure field p generated by a source with a step-time function or a square-time function.

If $f(t) = H(t)$, we find, from Eqs. (2) and (3),

$$p(r, \theta, t) = \rho\psi(r, \theta, t), \tag{58}$$

where $\psi(r, \theta, t)$ are given in Eqs. (52)—(57).

The calculations are done in two parts. First we calculate the arrival times t_E and t_M of ψ_n^i and t_F and t_M of ψ_n^s in order to decide the number of rays that must be included within a prescribed time interval of observation. Next we calculate the contribution from each ray from its onset to the end of time of observation. This is done at each location (r, θ), by first calculating $\xi = h^{-1}(r, \theta, t)$ and $\xi = g^{-1}(r, \theta, t)$ for each instant t, which is obtained from an numerical procedure for inverting Eqs. (28) and (43), and then calculating E_h, E_g, $\partial\xi/\partial t$ numerically.

In all examples shown, the source is placed at $r' = 5a$ and $\theta = 0$. Several observation points at (r, θ) are selected to illustrate the back scattering $(5a, 0)$, side scattering $(5a, \pi/2)$, near field forward scattering $(5a, \pi)$ and far field forward scattering $(20a, \pi)$. The number of rays and their arrival times for each example are summarized in Table III.

Figure 12 shows the normalized pressure of the total wave at $(5a, \pi/2)$ due to the scattering by a hard or soft cylinder. The pressure due to the incident wave, $\psi^i(r, \theta, t)$ of Eq. (52), is shown in dashed lines.

Two cases of forward scattering are shown in Figs. 13 and 14. At $(5a, \pi)$ which is the image point of the source, the ψ^i in Eq. (56) vanishes identically, and ψ^s is shown in Fig. 13 for a hard cylinder (Neumann condition). The pressure of the scattered wave due to a soft cylinder (Dirichlet condition) is the negative of that due to a hard cylinder.

At a far field point $(20a, \pi)$ in the shadow zone (Fig. 14), the incident wave contributes to the total pressure through the creeping rays in Eq. (56). Because the contributions from the incident ray and the scattered rays are additive for a hard cylinder, the total pressure at $(20a, \pi)$ is larger than that at $(5a, \pi)$. This examplifies the difference between the forward scattering at a "deep shadow" point $(5a, \pi)$ and a "light shadow" point $(20a, \pi)$. The "arrival" time of the slip-stream front are also marked on the figure.

Finally, we show in Fig. 15 the results for backscattering at $(5a, 0)$. To approximate the actual experimental observation reported by Barnard and McKinney[14] (see also the review articel by Neubauer[15]), we assume a square wave function for $f(t)$,

$$f(t) = H(t) - H(t - \Delta t). \tag{59}$$

In Fig. 15, $\Delta tc/a = 0.2$. Since the receiver is at the same point as the transmitter, only the scattered waves ψ^s of Eq. (55) are shown for a hard cylinder. ψ^s for a

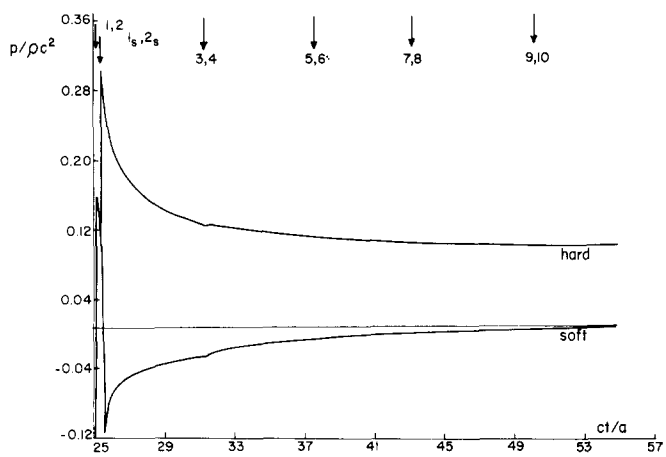


FIG. 14. Forward scattering at $(20a, \pi)$ due to a soft or hard cylinder.

soft cylinder is the negative of that shown in Fig. 15. The pressure generated by the source function alone is also shown in dashed lines. The presence of creeping waves (Ray No. 3, 4; 5, 6; etc.) is clearly evident in this figure.

[1]J.B. Keller, "A Geometrical Theory of Diffraction," *Symposium on Microwave Optics*, Eaton Electronic Research Lab., McGill Univ., Montreal, Canada, 1953.
[2]W. Franz, *Beugung elektromagnetischer Wellan* (Springer-Verlag, Berlin, 1957).
[3]F.G. Friedlander, *Sound Pulses* (Cambridge U.P., New York, 1958).
[4]F. Gilbert, "Scattering of Impulsive Elastic Waves by a Smooth Convex Cylinder," J. Acoust. Soc. Am. 32, 841—57 (1960).
[5]H. Überall, "Use of Sound Pulses for a Study of Circumferential Waves," J. Acoust. Soc. Am. 39, 564—78 (1966).
[6]L. Cagniard, *Reflection and Refraction of Progressive Seismic Waves*, translated by Flinn and Dix (McGraw-Hill, New York, 1962).
[7]Abramowitz and Stegun, *Handbook of Mathematical Function* (Dover, New York, 1968).
[8]F. Gilbert and D.V. Helmberger, "Generalized Ray Theory for a Layered Sphere", Geophys. J.Roy. Astron. Soc. 27, 57—80 (1972).
[9]A.T. de Hoop, "A Modification of Cagniard Method for Solving Seismic pulse problems," Appl. Sci.Res. B 8, 349—56 (1960).
[10]G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1973).
[11]G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., New York, 1966).
[12]A. Erdélyi *et al.*, *Tables of Integral Transforms*, Vol. 1 (McGraw-Hill, New York, 1954).
[13]E.C. Titchmarsh, *Theory of Fourier Integrals* (Oxford U.P., London, 1948).
[14]G.R. Barnard and C.M. McKinney, "Scattering of Acoustic Energy by Solid and Air-Filled Cylinders in Water," J. Acoust. Soc. Am. 33, 226—38 (1961).
[15]W.G. Neubauer, "Observation of Acoustic Radiation from Pane and Curved Surfaces," Phys. Acoust. X, 61—126 (1973).

Regularization of the Roy equations with a smooth cutoff

D. Atkinson, T. P. Pool,[a] and H. A. Slim

Instituut voor Theoretische Natuurkunde, Hoogbouw WSN, Postbus 800, Universiteitscomplex Paddepoel, Groningen, the Netherlands

(Received 26 April 1977)

The Roy equations for $\pi\pi$ scattering are combined with unitarity to give a nonlinear system of equations for the determination of the low-energy amplitudes. A Hölder continuous interpolation between the input high-energy absorptive parts and the output low-energy absorptive parts is implemented; and the resultant singular equations are regularized by means of an effective inelastic N/D method. If the scattering lengths, the CDD parameters, and the high-energy absorptive parts satisfy certain constraints, then there exists a locally unique solution of the system.

1. INTRODUCTION

In this paper we continue the study of the system of nonlinear, singular integral equations that results from a combination of the Roy equations[1] with elastic unitarity. It is assumed that the partial wave absorptive parts above a certain point, say $s=s_0$, are given, and that the S-wave scattering lengths are held fixed. The problem is to prove the existence of solutions for the partial waves in the domain $4 \leq s \leq s_0$ and to investigate the question of the nonuniqueness of such solutions.

It has already been shown[2] that if the input quantities are small enough, there exists a locally unique solution in a suitable space of Hölder-continuous functions. However, it is known that the physically interesting solution lies outside the scope of the above proof unless one chooses s_0 to be such that all the phase shifts remain small in $[4, s_0]$. A first step has been made[3] towards the removal of this limitation, in which a finite-interval version of the N/D method was used to regularize the singular equations. The new equations contained the customary CDD poles, and a further free parameter entered the solution in some cases, due to the marginally singular nature of the N integral equation.[4] The fact that this equation is not subject to Fredholm theory is an embarrassment for numerical work. Although an explicit integral representation for a resolvent kernel of the dominant part of the singular equation has been constructed and the homogeneous equation has been exhaustively studied, nevertheless, it is rather awkward to have to program this resolvent and to use it every time that the N equation is solved in the course of iterating the nonlinear system.

These problems are sidestepped in the present paper by the expedient of introducing a smooth instead of a sharp cutoff. By means of a Hölder-continuous cutoff function $h(s)$ we effect a homotopy from the elastically unitary "output" absorptive part below s_0 to the prescribed"input" absorptive part above s_1, where s_1 is greater than s_0, but still within the domain of validity of the Roy equations. The equations are again regularized by means of the N/D method; but the fact that the amplitude is not strictly unitary in $[s_0, s_1]$ leads to a Frye–Warnock[5] system with an effective elasticity that is a function of the input absorptive part and of $h(s)$. The new N integral equation is Fredholm and is eminently suited to numerical treatment.

In this paper we shall work with the reduced partial-wave amplitude,

$$f_l^I(s)=\left(\frac{s^{1/2}+2}{s^{1/2}-2}\right)^l F_l^I(s), \tag{1.1}$$

where F_l^I is the projection of the usual invariant scattering amplitude, for isospin I, onto the Legendre polynomial. This partial-wave amplitude satisfies elastic unitarity for $s \in [4, 16]$:

$$\mathrm{Im} f_l^I(s)=q_l(s)\,|f_l^I(s)|^2, \tag{1.2}$$

where

$$q_l(s)=\left(\frac{s-4}{s}\right)^{1/2}\left[\frac{s^{1/2}-2}{s^{1/2}+2}\right]^l. \tag{1.3}$$

In practice there is little inelasticity below the $K\bar{K}$ threshold, and we shall assume (1.2) to be correct in the domain of validity of the Roy equations, which includes the interval $[4, 32)$. The advantage of using the reduction factor of (1.1) is that this ensures that $F_l^I(s)$ has the correct behavior as $s \to \infty$, if $f_l(s)$ is bounded uniformly with respect to s and l.

The Roy equations for the partial waves can be written

$$f_l(s)=\frac{s^2}{\pi}\int_4^\infty \frac{ds'}{s'^2(s'-s)}\,\mathrm{Im} f_l(s')+b_l(s), \tag{1.4}$$

where isospin has been made an implicit variable and where

$$b_l(s)=C_l(s)+R_l(s)+S_l(s)+U_l(s)+T_l(s), \tag{1.5}$$

where

$$C_l(s)=\frac{1}{4}\left[\frac{s^{1/2}+2}{s^{1/2}-2}\right]^l\left\{\delta_{l0}\left[s-\frac{s-4}{2}\,(C_{st}+C_{su})\right]+\delta_{l1}\,\frac{s-4}{6}\,(C_{st}-C_{su})\right\}\alpha, \tag{1.6}$$

$$R_l(s)=\frac{s^2}{\pi}\int_4^\infty \frac{ds'}{s'^2(s'-s)}\left\{\left[\frac{s^{1/2}+2}{s'^{1/2}+2}\right]^{2l}-1\right\}\mathrm{Im} f_l(s'), \tag{1.7}$$

$$S_l(s)=\left[\frac{s^{1/2}+2}{s^{1/2}-2}\right]^l\frac{s^2}{\pi}\int_4^\infty \frac{ds'}{s'^2(s'-s)}\sum_{l'=l+2}^{\infty}(2l'+1)\times V_{ll'}(s,s')\left[\frac{s'^{1/2}-2}{s'^{1/2}+2}\right]^{l'}\mathrm{Im} f_{l'}(s'), \tag{1.8}$$

[a] Present address: Université Pierre et Marie Curie, Laboratoire de Physique Théorique et Hautes Energies, Tour 16—1er etage, 4, Place Jussieu, 75230 Paris-Cedex 05, France.

where

$$V_{ll'}(s,s') = \tfrac{1}{2}\int_{-1}^{1} dz\, P_l(z) P_{l'}(z'), \tag{1.9}$$

with

$$z = 1 + 2t/(s-4), \tag{1.10}$$

$$z' = 1 + 2t/(s'-4), \tag{1.11}$$

$$U_l(s) = \left[\frac{s^{1/2}+2}{s^{1/2}-2}\right]^l \frac{1}{\pi}\int_4^\infty \frac{ds'}{s'^2} \sum_{l'=0}^{\infty} (2l'+1) \times W_{ll'}(s,s') \left[\frac{s'^{1/2}-2}{s'^{1/2}+2}\right]^{l'} \mathrm{Im} f_{l'}(s'), \tag{1.12}$$

where

$$W_{ll'}(s,s') = \tfrac{1}{2} C_{su}\int_{-1}^{1} dz\, P_l(z) \frac{u^2}{s'-u} P_{l'}(z'), \tag{1.13}$$

$$T_l(s) = \left[\frac{s^{1/2}+2}{s^{1/2}-2}\right]^l \frac{1}{\pi}\int_4^\infty \frac{ds'}{s'} \sum_{l'=0}^{\infty} (2l'+1) \times \left[X_l(s,s') + Y_{ll'}(s,s')\right] \left[\frac{s'^{1/2}-2}{s'^{1/2}+2}\right]^{l'} \times \mathrm{Im} f_{l'}(s'), \tag{1.14}$$

where

$$X_l(s,s') = \tfrac{1}{2} C_{st}\int_{-1}^{1} dz\, P_l(z) \left[1 + \frac{s}{t-4}(1-C_{tu})\right] \times \left[\frac{t^2}{s'-t} - \frac{4(t-4)}{s'-4}(1-C_{su}) - \frac{16}{s'-4}\right], \tag{1.15}$$

$$Y_{ll'}(s,s') = \tfrac{1}{2}\int_{-1}^{1} dz\, P_l(z) \left\{ C_{st}\left[1 + \frac{s}{t-4}(1-C_{tu})\right] C_{su} - \left[C_{su} + \frac{s}{t-4}(C_{su}-1)\right] P_{l'}(z') \right\} \frac{(4-t)^2}{s'-4+t}, \tag{1.16}$$

In the above, C_{st}, C_{tu}, and C_{su} are the usual isospin crossing matrices and α is a constant vector, in which the $I=0$ and $I=2$ components are the corresponding S-wave scattering lengths, while the $I=1$ component is zero.

2. *N/D* EQUATIONS WITH A SMOOTH CUTOFF

In this section, we shall replace the abrupt cutoff of Ref. 3 by a gradual one that begins at $s=s_0$ and ends at $s=s_1 > s_0$. To be precise, we write

$$f(s) = \frac{s^2}{\pi}\int_4^\infty \frac{ds'}{s'^2(s'-s)} \mathrm{Im} f(s') + b(s), \tag{2.1}$$

where we have suppressed the suffix l and where

$$\mathrm{Im} f(s) = h(s) q(s) |f(s)|^2 + [1-h(s)] a(s). \tag{2.2}$$

Here $a(s)$ is a function that is specified for $s \geq s_0$, which is fixed in advance, and $h(s)$ is a monotonic cutoff function with the properties

$$h(s) = 1, \quad 4 \leq s \leq s_0 \tag{2.3a}$$

$$h(s) = 0, \quad s \geq s_1 \tag{2.3b}$$

$$|h(s_a) - h(s_b)| \leq |s_a - a_b|^\mu, \quad s_0 \leq s_a < s_b \leq s_1, \quad 0 < \mu < 1. \tag{2.3c}$$

Thus (2.2) effects a Hölder continuous interpolation from the elastic unitarity output expression for $s \leq s_0$, to the high-energy input model for $s \geq s_1$. For the purposes of the proof we need only the properties (2.3); but numerically it is convenient to make $h(s)$ a thrice-differentiable function for which, in addition to (2.3a) and (2.3b),

$$h'(s_0) = h'(s_1) = h''(s_0) = h''(s_1) = 0. \tag{2.3d}$$

An example of such a function is

$$h(s) = x^3(4-3x), \quad x = \cos\left(\frac{\pi}{2}\frac{s-s_0}{s_1-s_0}\right), \tag{2.4}$$

for $s_0 \leq s \leq s_1$, and this is eminently suited to numerical computations in which cubic splines are employed.

The expression (2.2) has a formal resemblance to the inelastic condition, in which the elasticity function is given. In fact we may rewrite (2.1) in the form

$$g(s) = \frac{1}{\pi}\int_4^\infty \frac{ds'}{s'-s}\left[\rho(s')|g(s')|^2 + \frac{1-\eta^2(s')}{4\rho(s')}\right] + c(s), \tag{2.5}$$

where we have written

$$g(s) = s^{-2} f(s) \tag{2.6}$$

and

$$c(s) = s^{-2} b(s), \tag{2.7}$$

in order to absorb the subtraction factor s^2, and where

$$\rho(s) = s^2 h(s) q(s) \tag{2.8}$$

and

$$\eta(s) = \{1 - 4h(s)[1-h(s)] q(s) a(s)\}^{1/2}. \tag{2.9}$$

Equation (2.5) mimics exactly the standard form of a dispersion relation for a partial wave amplitude, $g(s)$, with born term $c(s)$, for which inelastic unitarity holds, with phase space $\rho(s)$ and elasticity $\eta(s)$.

In order to find all the solutions of (2.5), for a given $c(s)$, we apply the standard Frye–Warnock method,[5] in which one writes

$$g(s) = N(s)/D(s). \tag{2.10}$$

Here

$$D(s) = 1 + \sum_n \frac{r_n}{s-t_n} - \frac{1}{\pi}\int_4^\infty \frac{ds'}{s'-s} \rho(s') n(s') \tag{2.11}$$

and

$$N(s) = \frac{1+\eta(s)}{2} n(s) + i\frac{1-\eta(s)}{2\rho(s)} \mathrm{Re} D(s), \tag{2.12}$$

for $s \geq 4$, in which the real function $n(s)$ is the solution of the nonsingular integral equation

$$\eta(s) n(s) = \bar{c}(s) + \sum_n r_n \frac{\bar{c}(s) - \bar{c}(t_n)}{s-t_n} + \frac{1}{\pi}\int_4^\infty ds' \frac{\bar{c}(s') - \bar{c}(s)}{s'-s} \rho(s') n(s') \tag{2.13}$$

and

$$\bar{c}(s)=c(s)+\frac{P}{\pi}\int_4^\infty \frac{ds'}{s'-s}\,\frac{1-\eta(s')}{2\rho(s')}\,. \tag{2.14}$$

The t_n are the positions of the CDD poles, r_n being the residues. For the sake of formal elegance, and to allow us to take over the standard formulas without change, we have written all integrals from $s=4$ to $s=\infty$, but because of the support properties of $h(s)$, actually $\rho(s)$ vanishes for $s\geq s_1$, so the integration domain in (2.11) and (2.13) is $4\leq s\leq s_1$. According to (2.9), $1-\eta(s)$ vanishes both for $s\leq s_0$ and for $s\geq s_1$, and we have to define the integrand in (2.14) by continuity at the point $s=s_1$. In fact

$$\frac{1-\eta(s)}{2\rho(s)}=\frac{2[1-h(s)]\,a(s)\,s^{-2}}{1+\{1-4h(s)[1-h(s)]\,q(s)\,a(s)\}^{1/2}}\,, \tag{2.15}$$

and one can see that this expression changes continuously from 0 for $s\leq s_0$ to $s^{-2}a(s)$ for $s\geq s_1$. Thus $\bar{c}(s)$ is also continuous at $s=s_1$, and so by means of the smooth cutoff we have removed the logarithmic singularity that complicated the earlier method.

Since $c(s)$ is assumed to be known for $4\leq s\leq s_1$ and $a(s)$ is known for $s\geq s_0$, we know $\eta(s)$ and hence $\bar{c}(s)$ for $4\leq s\leq s_1$, and $\bar{c}(s)$ is in fact Hölder continuous on this interval. It may be shown[6] that any Hölder continuous solution of the nonlinear equation (2.5) has a representation of the form (2.10)–(2.14), on condition that the phase shift of the solution tends to a limit as $s\to\infty$, and on condition that $\eta(s)$ has no zeros in $[4,\infty)$, since such zeros would introduce singularities of the third kind[7] into the integral equation (2.13). In our case, $\eta(s)=1$ for $s\leq s_0$ and for $s\geq s_1$, so any possible zero can only lie in the interval (s_0,s_1). However, since h falls monotonically from 1 to 0, it is easy to see that $4h(1-h)\leq 1$, the equality being reached only once, at the point at which $h=\frac{1}{2}$. On the other hand, $qa\leq 1$, the equality being reached at the position of an elastic resonance. Hence we need only choose s_0 and s_1 in such a way that the given function, $a(s)$, is not equal to $1/q(s)$ at precisely the point for which $h(s)=\frac{1}{2}$, in order to ensure that

$$\eta(s)>0 \tag{2.16}$$

for $s_0\leq s\leq s_1$. The simplest way to do this is to choose the interval $[s_0,s_1]$ in such a way that $a(s)$ does not have a resonance in it. Thus we can avoid third kind singularities and make (2.13) Fredholm. Conversely, one can show that any solution of (2.10)–(2.14) satisfies the original nonlinear equation (2.5), on condition that $D(s)$ has no zeros in the cut plane.

To conclude this section, let us study more closely the connection between the method of this paper and that of Ref. 3. Suppose that an amplitude $f(s)$ is given that satisfies (2.1), and which is elastically unitary up to s_1. Such an amplitude is a solution of the sharp cutoff equations of Ref. 3, if the cutoff is placed at s_1 and if the CDD parameters are chosen appropriately. It has been shown that the amplitude, considered as a solution of the sharp cutoff equations, is embedded in a continuum of solutions of dimension

$$d=[2\delta(s_1)/\pi]. \tag{2.17}$$

The reason for this is that the sharp cutoff equations contain n CDD poles, where

$$n=[\delta(s_1)/\pi], \tag{2.18}$$

and each CDD pole carries two real parameters, the position and residue. Further, if $\alpha\geq\frac{1}{2}$, where

$$\alpha=\delta(s_1)/\pi-n, \tag{2.19}$$

then an additional degree of freedom arises, because an arbitrary multiple of the homogeneous solution of the marginally singular N integral equation is allowed.

The above amplitude is also a solution of the equations of this section, again if the CDD pole parameters are chosen appropriately (because of the generality of the N/D representation). In this special case, there will be no difference between $\mathrm{Im}f(s)$ and $a(s)$ for s in (s_0,s_1), since they are both equal to $q(s)|f(s)|^2$, and one can always find a $\bar{\delta}(s)$ such that

$$q(s)f(s)=\exp(i\bar{\delta}(s))\sin\delta(s)=\frac{\eta(s)\exp(2i\bar{\delta}(s))-1}{2ih(s)}\,, \tag{2.20}$$

where $\delta(s)$ is the actual phase shift. For $4\leq s\leq s_0$, $\bar{\delta}(s)$ is the same as $\delta(s)$, but since $h(s_1)=0$ and $\eta(s_1)=1$, it follows that $\bar{\delta}(s_1)=n_c\pi$, where n_c is an integer. In fact (2.20) can be solved for η and $\bar{\delta}$, yielding

$$\eta^2(s)=1-4h(s)[1-h(s)]\sin^2\delta(s), \tag{2.21}$$

which is consistent with (2.9), and

$$\tan[2\bar{\delta}(s)]=\frac{h(s)\sin[2\delta(s)]}{1-h(s)\{1-\cos[2\delta(s)]\}}\,. \tag{2.22}$$

Now we have agreed to choose s_0 and s_1 such that $\delta(s)$ does not attain a multiple of $\pi/2$ in the interval $[s_0,s_1]$. Then we define the integer

$$n=[\delta(s_1)/\pi]=[\delta(s_0)/\pi], \tag{2.23}$$

and a fraction

$$\alpha=\delta(s_1)/\pi-n. \tag{2.24}$$

Then if $\alpha<\frac{1}{2}$, $\bar{\delta}(s_1)=n\pi$, whereas if $\alpha>\frac{1}{2}$, $\bar{\delta}(s_1)=(n+1)\pi$. Now we define in the standard manner

$$D(s)=\exp\left[-\frac{1}{\pi}\int_4^{s_1}\frac{ds'}{s'-s}\,\bar{\delta}(s')\right]$$

$$\underset{s\sim s_1}{\sim}(s-s_1)^{-n_c}, \tag{2.25}$$

where

$$n_c=n+\theta(\alpha-\tfrac{1}{2}). \tag{2.26}$$

Then the D function that satisfies (2.11) is

$$D(s)=(s-s_1)^{n_c}\left[\prod_{n=1}^{n_c}(s-t_n)\right]^{-1}D(s) \tag{2.27}$$

where the t_n, the CDD pole positions, are n_c distinct points for which $\sin\bar{\delta}(s)$ vanishes. In the case $\alpha<\frac{1}{2}$, there are at least n_c points below s_0 where this happens; however, if $\alpha>\frac{1}{2}$, there may be only n_c-1 such points, but in this case we can always take the n_cth point to be s_1. For definiteness we stipulate that no CDD poles are to be placed in the interval $[s_0,s_1]$ if $\alpha<\frac{1}{2}$, and that just one is to be placed in this interval, at s_1, if $\alpha\geq\frac{1}{2}$. It is interesting to note that, for a given $c(s)$ and a given $a(s)$ for $s\geq s_0$, the dimension of the manifold of solutions of

the N/D system is $2n_c$ if $\alpha < \frac{1}{2}$, since each CDD pole carries two real parameters (the position and residue), but the dimension is $2n_c - 1$ if $\alpha > \frac{1}{2}$, since in this case the position of the last CDD pole is frozen at $s = s_1$. In general the dimension is then

$$2n_c - \theta(\alpha - \tfrac{1}{2}) = 2[\delta(s_1)/\pi] + \theta(\alpha - \tfrac{1}{2}) = [2\delta(s_1)/\pi]. \tag{2.28}$$

This is a satisfying result, since it agrees precisely with the dimension found in the case of the sharp cutoff.[3,8]

It must be stressed that the above discussion is somewhat artificial, in the sense that most solutions of the equations of this section will not be elastically unitary for $s_0 < s < s_1$. Our purpose was simply to make contact with the earlier results. In an autonomous application of the present method, the final amplitude would be in general unitary only below s_0. In the interval (s_0, s_1), the amplitude would neither be unitary, nor would its imaginary part be equal to the input function $a(s)$.

3. SOLUTION OF THE NONLINEAR SYSTEM

In this section we shall treat the nonlinear system, incorporating the N/D equations, as a nonlinear mapping, on the assumption that the following input quantities have been specified: the scattering lengths, the CDD parameters, and $a_l(s)$ for $s \geq s_0$. We shall show that the mapping is contractive if the inputs satisfy certain conditions.

It is of some importance to choose a well-behaved function as the basic quantity to be determined. Let us write

$$b_l(s) = \tilde{b}_l(s) + \hat{b}_l(s), \tag{3.1}$$

where $\tilde{b}_l(s)$ corresponds to the expressions (1.5)—(1.16), in which however $\mathrm{Im} f_l(s)$ is replaced by $[1 - h(s)]\, a_l(s)$, the known input quantity, and where $\hat{b}_l(s)$ is the remainder, namely the corresponding formula (1.5) [omitting $C_l(s)$], in which $\mathrm{Im} f_l(s)$ is replaced by $h(s)\, q_l(s) |f_l(s)|^2$. Now $\tilde{b}_l(s)$ is wholly known, and we shall seek to make $\hat{b}_l(s)$ uniformly small.

Consider the following mapping for $\hat{b}_l(s)$, at fixed $\tilde{b}_l(s)$:

$$\hat{b}'_l(s) = \Phi[\hat{b}; l, s] \equiv \hat{R}_l(s) + \hat{S}_l(s) + \hat{U}_l(s) + \hat{T}_l(s), \tag{3.2}$$

where $\hat{R}_l$, $\hat{S}_l$, $\hat{U}_l$, and $\hat{T}_l$ are written as in (1.7)—(1.16), but with

$$h(s)\, q_l(s)\, s^4 |N_l(s)/D_l(s)|^2 \tag{3.3}$$

in place of $\mathrm{Im} f_l(s)$. Here $\rho_l(s)$ and $\eta_l(s)$ are defined as in (2.8)—(2.9), and $n_l(s)$ is the solution of (2.13), where

$$\bar{c}_l(s) = s^{-2}[\tilde{b}_l(s) + \hat{b}_l(s)] + \frac{P}{\pi} \int_4^\infty \frac{ds'}{s' - s} \frac{1 - \eta_l(s')}{2\rho_l(s')}; \tag{3.4}$$

moreover, $D_l(s)$ is defined by (2.11), and $N_l(s)$ by (2.12). We have reinstated the angular momentum suffix, but isospin remains implicit.

Suppose that $\hat{b}_l(s)$ belongs to the Banach space of sequences of functions that have continuous second derivatives, with norm

$$\|\hat{b}\| = \sup_{I,l,s} |\hat{b}^I_l(s)| + \sup_{I,l,s} \left| \frac{d}{ds} \hat{b}^I_l(s) \right| + \sup_{I,l,s} \left| \frac{d^2}{ds^2} \hat{b}^I_l(s) \right|, \tag{3.5}$$

where the suprema are taken over $I = 0, 1, 2$ and $l = 0, 1, 2, \cdots$, and $s \in [4, s_1]$. We suppose that the input quantities $a_l(s)$, $s \geq s_0$, are such that $\|\tilde{b}\|$ is finite. This is not unreasonable, since $b_l(s)$ is actually analytic in a neighborhood of $4 \leq s \leq s_0 - \epsilon$, $\epsilon > 0$, and although $\tilde{b}_l(s)$ certainly has branch points at $s = s_0$ and $s = s_1$, it will have a bounded second derivative at these points if the given function $a_l(s)$ is sufficiently smooth. We shall show in this section that if $\hat{b}_l(s)$ belongs to our space and is small enough in norm, then $\hat{b}'_l(s)$ also belongs to the space. In view of the quadratic nature of the mapping Φ, one can then show easily that the contraction mapping theorem applies if the inhomogeneities are small enough.

The first step in the proof consists in showing that (2.13) has a unique solution $n_l(s)$, given $\hat{b}_l(s)$ and so $\bar{c}_l(s)$. Let us consider this linear equation as a mapping on the subsidiary Banach space of sequences of continuous functions with norm

$$\|n\|_1 = \sup_{I,l,s} |n^I_l(s)|. \tag{3.6}$$

Since

$$\left| \frac{1}{\pi} \frac{\bar{c}(s') - \bar{c}(s)}{s' - s} \rho(s') \right| \leq \frac{1}{\pi} \left| \int_s^{s'} dx \frac{d}{dx} \bar{c}(x) \right| \cdot |s' - s|^{-1} s'^2 h(s') \leq \frac{1}{\pi} s_1^2 \|\bar{c}\|, \tag{3.7}$$

it follows that if

$$\|\bar{c}\| < \pi s_1^{-2} \inf_{s_0 \leq s \leq s_1} \eta(s), \tag{3.8}$$

then (2.13) defines a contraction mapping on the space (3.6), and so the solution $n_l(s)$ is unique in this space. This solution is not merely continuous, but is actually differentiable, as we can see by differentiating both sides of (2.13) and by using the fact that

$$\left| \frac{d}{ds} \frac{\bar{c}(s') - \bar{c}(s)}{s' - s} \right| \leq \tfrac{1}{2} \|\bar{c}\|. \tag{3.9}$$

Hence the singular integral in (2.11) is well defined, and in fact $D_l(s)$ is not merely bounded on $[4, s_1]$, but is Hölder continuous as well.

It is not sufficient that $D_l(s)$ is bounded; we must show that it has no zeros on the first Riemann sheet. In the case that there are no CDD poles, this is easy enough. When there are CDD poles however, we expect that the real part of $D_l(s)$ will have zeros on the real axis. If the CDD pole residues are small, there will be one zero close to each pole, at the mass squared of a resonance. $D_l(s)$ will itself have a complex zero nearby, and we must ensure that such a zero is on the second Riemann sheet.

If we choose $\hat{b}_l(s)$ and $\tilde{b}_l(s)$ to be small enough in

norm, then the solution $n_l(s)$ of (2.13) will be dominated by the inhomogeneous term,

$$\{\bar{c}_l(s)+\sum_n r_{l,n}[\bar{c}_l(s)-\bar{c}_l(t_{l,n})]/(s-t_{l,n})\}/\eta_l(s). \quad (3.10)$$

For $\|\tilde{b}\|$ sufficiently small, we can be sure that the norm of the fixed point, $\hat{b}_l(s)$, which is quadratic in $\|\tilde{b}\|$, is still smaller. It then follows from (3.4) that the dominant part of (3.10) will be obtained by replacing $\bar{c}_l(s)$ by

$$\tilde{c}_l(s)=s^{-2}\tilde{b}_l(s)+\frac{P}{\pi}\int_4^\infty \frac{ds'}{s'-s}\,\frac{1-\eta_l(s')}{2\rho_l(s')}, \quad (3.11)$$

and this is a known input function. We are free to choose this known function to be such that the dominant part of (2.11),

$$1+\sum_n \frac{r_{l,n}}{s-t_{l,n}}-\frac{1}{\pi}\int_4^\infty \frac{ds'}{s'-s}\,\rho_l(s')\tilde{c}_l(s'), \quad (3.12)$$

has no zeros on the first Riemann sheet. For our purposes it is even necessary to suppose that $\tilde{c}_l(s)$ has been chosen such that the modulus of (3.12) has a positive lower bound which is uniform with respect to l. Then it is clear that we can arrange that $|D_l(s)|$ has also such a uniform positive lower bound. Hence it is possible to rule out ghosts by restricting the input suitably.

The fact that $D_l(s)$ is dominated by the known function, (3.12), means that we can exclude first-sheet zeros of $D_l(s)$, but not zeros of $\mathrm{Re}D_l(s)$ on the real axis. Generally, for small values of the $r_{l,n}$, there will be a zero of $\mathrm{Re}D_l(s)$ near each CDD pole. However, since we have agreed that it is possible to choose the function (3.12) in such a way that $|D_l(s)|$ has a uniform lower bound, it follows that, at a zero of $\mathrm{Re}D_l(s)$, $|n_l(s)|$ is uniformly bounded below, since it is simply $|\mathrm{Im}D_l(s)|$. Hence we have no difficulty in obtaining an estimate of the Lipschitz coefficient in our contraction mapping proof. Detailed conditions which ensure that $n_l(s)$ does not have zeros near the zeros of $\mathrm{Re}D_l(s)$ have been given in the literature[9]; but for our purposes such fulsomeness is unnecessary, since we are not trying to calculate the maximal radius of a ball on which the mapping Φ is contractive. We are content to show that the radius is nonzero if the inhomogeneities are small enough, and if they are chosen such that the modulus of the function (3.12) has a uniform lower bound.

Consider now the expression (3.3), which has to be injected into $\hat{R}_l$, $\hat{S}_l$, $\hat{U}_l$, and $\hat{T}_l$ in order finally to yield $\hat{b}'_l$, the image of $\hat{b}_l$ under the mapping Φ. We have now to show that

$$\|\hat{b}'\|\leq\kappa\|\hat{b}+\tilde{b}\|^2, \quad (3.13)$$

where κ is a constant. Since we know that $N_l(s)$ has a uniform upper bound, that is proportional to $\|\hat{b}+\tilde{b}\|$, and we have ensured that $|D_l(s)|$ has a uniform positive lower bound, it will be enough if we can show that P_l, S_l, U_l, and T_l are bounded in norm if we replace $\mathrm{Im}f_{l'}(s')$ by

$$h(s')\left(\frac{s'-4}{s'}\right)^{1/2}\left[\frac{s'^{1/2}-2}{s'^{1/2}+2}\right]^{l'}, \quad (3.14)$$

in Eqs. (1.7), (1.8), (1.12), and (1.14). The rest of this section and the Appendix are devoted to this demonstration.

All the integrals are over the finite domain, $4\leq s\leq s_1$, and none of them are singular. The only nontrivial point concerns the infinite l' series, and we can prove convergence only if $s_1<32$, this being the limit of validity of the Roy equations. The term $R_l(s)$ is trivial in this respect, since it only contains one partial wave. At first sight it looks as though $R_l(s)$ might not be uniformly bounded as $l\to\infty$. However, in view of the bound (3.14), what one has to maximize is

$$[(s^{1/2}+2)^2(s'^{1/2}-2)/(s'^{1/2}+2)^3]^l \quad (3.15)$$

for s and s' in $[4,s_1]$. This quantity remains smaller than unity for $s\lesssim 70$, and any $s'\geq 4$. For $s<32$, it is less than $(16/27)^l$, which is certainly bounded as $l\to\infty$. In fact we have to estimate

$$\left[\left(\frac{s^{1/2}+2}{s'^{1/2}+2}\right)^{2l}-1\right]\Big/(s'-s), \quad (3.16)$$

which involves differentiating the numerator with respect to s. Since we must also consider the second derivative of $R_l(s)$, in order to be able to bound $\|R\|$, we have finally to majorize the first three derivatives of the numerator in (3.16). Aside from trivial factors, these derivatives involve (3.15) again, with, however, factors of l, up to the third power. Clearly these powers are tamed by the bound $(16/27)^l$, and we conclude that $|R_l(s)|$, $|R'_l(s)|$, and $|R''_l(s)|$ are uniformly bounded.

Let us consider next $S_l(s)$. For $l'\geq l+2$, it is clear from the definition (1.9) that $V_{ll'}(s',s')=0$, and so if we define

$$\bar{V}_{ll'}(s,s')=\frac{1}{s'-s}\left[\frac{s^{1/2}+2}{s^{1/2}-2}\right]^l V_{ll'}(s,s'), \quad (3.17)$$

then it is easy to see that

$$\left(\frac{\partial}{\partial s}\right)^n \bar{V}_{ll'}(s,s')=-\int_0^1 y^n\,dy\left(\frac{\partial}{\partial x}\right)^{n+1} \times\left[\left(\frac{x^{1/2}+2}{x^{1/2}-2}\right)^l V_{ll'}(x,s')\right]_{x=s'+y(s-s')}, \quad (3.18)$$

for $n=0,1,2$. In the Appendix, we prove that

$$\left|\left(\frac{\partial}{\partial s}\right)^n\left[\left(\frac{s^{1/2}+2}{s^{1/2}-2}\right)^l V_{ll'}(s,s')\right]\right| \leq\kappa\exp(-\epsilon l')\left[\frac{s'^{1/2}+2}{s'^{1/2}-2}\right]^{2l'}, \quad (3.19)$$

for any $s\in[4,s_1]$, with $s_1<32$, where κ is a constant and ϵ is a small positive constant, both independent of s, s', l, and l'. Hence

$$\left|\left(\frac{d}{ds}\right)^n S_l(s)\right|\leq\kappa\int_4^{s_1}\left[\frac{s'-4}{s'}\right]^{1/2}\sum_{l'=l+2}^{\infty}(2l'+1) \times\left|\left(\frac{\partial}{\partial s}\right)^n[s^2\bar{V}_{ll'}(s,s')]\right|\left[\frac{s'^{1/2}-2}{s'^{1/2}+2}\right]^{2l'} \leq\kappa\int_4^{s_1}ds'\left(\frac{s'-4}{s'}\right)^{1/2}\sum_{l'=0}^{\infty}(2l'+1)e^{-\epsilon l'}, \quad (3.20)$$

for $n=0,1,2$, and this is clearly bounded.

The terms $U_l(s)$ and $T_l(s)$, defined in (1.12) and (1.14), are somewhat easier to treat, since there is no vanishing denominator. It can be shown that $W_{ll'}(s,s')$ and $Y_{ll'}(s,s')$ satisfy inequalities of the type (3.19), while

$$\left|\left(\frac{\partial}{\partial s}\right)^n\left[\left(\frac{s^{1/2}+2}{s^{1/2}-2}\right)^l X_l(s,s')\right]\right| \leq \kappa, \tag{3.21}$$

for $n=0,1,2$. The methods are similar to those given in the appendix, and details may be found in Ref. 10. These inequalities suffice to bound $U_l(s)$ and $T_l(s)$, and their first two derivatives; and this concludes the proof that $\hat{b}_l'(s)$ belongs to the Banach space. Thus Φ is contractive if the inhomogeneities are small enough.

ACKNOWLEDGMENT

We would like to thank R. L. Warnock for some useful suggestions.

APPENDIX

We shall sketch the derivation of certain bounds[10] for the functions V, W, X, and Y, defined in (1.9), (1.13), (1.15), and (1.16). From (1.10) and (1.11), we can write

$$z' = 1 + (z-1)/\alpha, \tag{A1}$$

where

$$\alpha = (s'-4)/(s-4), \tag{A2}$$

and so $P_{l'}(z')$ is in fact a polynomial in z. Let us write a Cauchy integral for this polynomial around the following elliptical contour in the z plane:

$$\partial\epsilon(z_1) = \{z : |z + (z^2-1)^{1/2}| = z_1 + (z_1^2-1)^{1/2}, z_1 > 1\}. \tag{A3}$$

The z integration in (1.9) can be performed under the contour integral, the result being

$$V_{ll'}(s,s') = \frac{\alpha}{2\pi i}\oint_{\partial\epsilon(z_1)} d\xi\, P_{l'}(\xi)\, Q_l(\bar{\xi}), \tag{A4}$$

where Q_l is the Legendre function of the second kind and where

$$\bar{\xi} = 1 + \alpha(\xi - 1). \tag{A5}$$

If $\xi \in \partial\epsilon(z_1)$, it may be shown from the Laplace representations that $|P_{l'}(\xi)|$ is bounded above by $P_{l'}(z_1)$ and $|Q_l(\xi)|$ by $Q_l(z_1)$. However, we have to majorize $|Q_l(\bar{\xi})|$ and not $|Q_l(\xi)|$, and $\bar{\xi}$ describes an ellipse that lies wholly outside $\partial\epsilon(z_1)$ if $\alpha > 1$, and whooly inside $\partial\epsilon(z_1)$ if $\alpha < 1$. In the former case, for a given ξ on $\partial\epsilon(z_1)$, $|Q_l(\bar{\xi})|$ is bounded by $Q_l(\eta)$, where η is the rightmost extremity of the ellipse with foci at ± 1 that passes through $\bar{\xi}$. For any $\xi \in \partial\epsilon(z_1)$, it is easy to check that the corresponding η satisfies

$$\alpha z_1 - (\alpha - 1) \leq \eta \leq \alpha z_1 + (\alpha - 1), \tag{A6}$$

and since $Q_l(z)$ is a monotonically decreasing function of z, for $z > 1$, it follows that $|Q_l(\bar{\xi})|$ is majorized by $Q_l(\alpha z_1 - \alpha + 1)$ for all $\xi \in \partial\epsilon(z_1)$. In the case $\alpha < 1$, $\bar{\xi}$ describes an ellipse that lies wholly within $\partial\epsilon(z_1)$, and we may now conclude that the corresponding η satisfies

$$\alpha z_1 - (1-\alpha) \leq \eta \leq \alpha z_1 + (1-\alpha); \tag{A7}$$

and so $|Q_l(\bar{\xi})|$ is bounded by $Q_l(\alpha z_1 + \alpha - 1)$, on condition that the argument of the latter function is greater than unity. We may combine both results as follows:

$$\sup_{\xi \in \partial\epsilon(z_1)} |Q(\bar{\xi})| = Q_l(\alpha z_1 - |\alpha - 1|), \tag{A8}$$

for any $\alpha > 2/(1+z_1)$. Hence we have

$$|V_{ll'}(s,s')| \leq \alpha z_1 P_{l'}(z_1)\, Q_l(\tilde{z}), \tag{A9}$$

where we have majorized the circumference of the ellipse by that of its circumscribing circle and where

$$\tilde{z} = \alpha z_1 - |\alpha - 1| = [(s'-4)z_1 - |s'-s|]/(s-4). \tag{A10}$$

We wish now to motivate a choice for z_1, in order to make (A9) as useful as possible, and to maximize s_1, the largest value of s for which the inequalities hold. In the first place we write

$$z_1 = 2z_0^2 - 1, \tag{A11}$$

and require

$$z_0 \leq \left(\frac{s'+4}{s'-4}\right)\exp(-\epsilon/2), \tag{A12}$$

where ϵ is a small, positive constant. This ensures that

$$P_{l'}(z_1) \leq [z_0 + (z_0^2-1)^{1/2}]^{2l'} \leq \exp(-\epsilon l')\left[\frac{s'^{1/2}+2}{s'^{1/2}-2}\right]^{2l'}. \tag{A13}$$

Suppose further that z_1 is such that

$$\tilde{z} \geq \frac{s+4}{s-4}\exp(\epsilon). \tag{A14}$$

Then

$$Q_l(\tilde{z}) \leq [\tilde{z} + (\tilde{z}^2-1)^{1/2}]^{-l} Q_0(\tilde{z}) \leq \exp(-\epsilon l)\left(\frac{s^{1/2}-2}{s^{1/2}+2}\right)^l \frac{s-4}{8}. \tag{A15}$$

Now in the case $s' \leq s$, the inequality (A14) implies the following constraint upon s:

$$s \leq [1+\exp(\epsilon)]^{-1}[(s'-4)z_1 + s' - 4\exp(\epsilon)]. \tag{A16}$$

The minimum value of the right-hand side, as a function of s', leads to

$$s \leq s_1 \equiv [\exp(\epsilon)+1]^{-1}\cdot[64\exp(-\epsilon) - 4(\exp(\epsilon)-1)], \tag{A17}$$

in view of (A11) and (A12). The limit of this bound as $\epsilon \to 0$ is 32, and this is the maximum value of s for which the Roy equations are valid. Since we wish to retain the exponential factor in (A15), we need $\epsilon > 0$, and this means that s_1 will be less than 32, although we can make it as close to 32 as we like by making ϵ small enough

The inequality (A14) implies

$$z_1 \geq [s(\exp(\epsilon)-1) + s' + 4\exp(\epsilon)]/(s'-4) \tag{A18}$$

in the case $s' \geq s$; and this can be combined with (A11)

and (A12) to yield

$$\frac{(s'+4)^2}{s'-4}\exp(-\epsilon) \geq \tfrac{1}{2}(s'-4)(z_1+1) \geq s' + \frac{s+4}{2}(\exp(\epsilon)-1). \tag{A19}$$

The minimum of the left term is $32\exp(-\epsilon)$, so if we choose z_1 such that the middle term is equal to this constant, then the first inequality is satisfied also, on condition that we restrict both s' and s to the interval $[4, s_1]$ where s_1 was defined in (A17). Thus we choose

$$z_1 = -1 + \frac{64}{s'-4}\exp(-\epsilon), \tag{A20}$$

and on combining (A9), (A13), (A15), and (A20), we find

$$\left|\left[\frac{s^{1/2}+2}{s^{1/2}-2}\right]^l V_{ll'}(s,s')\right| \leq 8\exp[-\epsilon(l+l')]\left(\frac{s'^{1/2}+2}{s'^{1/2}-2}\right)^{2l'}, \tag{A21}$$

for any $s, s' \in [4, s_1]$.

By similar techniques one can show that

$$\left|\left(\frac{\partial}{\partial s}\right)^n\left[\left(\frac{s^{1/2}+2}{s^{1/2}-2}\right)^l V_{ll'}(s,s')\right]\right| \leq \kappa\, l^n \exp[-\epsilon(l+l')]\left(\frac{s'^{1/2}+2}{s'^{1/2}-2}\right)^{2l'}, \tag{A22}$$

for $n = 1, 2, 3$, where κ is a constant. To prove this, $Q_l(\xi)$ must be differentiated repeatedly with respect to s under the integral in (A4). The recurrence relations for the Legendre functions are then used to make all cancellations explicit. Details of some of the necessary calculations are to be found in Ref. 10; the remainder are very similar, and we shall not reproduce them here. The factor l^n arises because of the nth order derivative with respect to s; but it can be removed, thanks to the term $\exp(-\epsilon l)$, at the expense of an adjustment of the constant κ. The inequality (3.19) then follows immediately.

[1] D. Atkinson and T. P. Pool, Nucl. Phys. B **81**, 502 (1974). S. M. Roy, Phys. Lett. B **36**, 353 (1971).
[2] T. P. Pool, thesis, University of Groningen (1977).
[3] D. Atkinson and R. L. Warnock, "Multiple Solutions of the Roy Equations, and the N/D Method on a Finite Interval," Groningen preprint (1976).
[4] D. Atkinson and A. P. Contogouris, Nuovo Cimento **39**, 1082 (1965).
[5] G. Frye and R. L. Warnock, Phys. Rev. **130**, 478 (1963).
[6] R. L. Warnock, in *Lectures in Theoretical High Energy Physics*, edited by H. H. Aly (Interscience, New York, 1968).
[7] G. Bart and R. L. Warnock, SIAM J. Math. Anal. **4**, (1973) 609.
[8] C. Pomponiu and G. Wanders, Nucl. Phys. B **103**, 172 (1976).
[9] H. Mc. Daniel and R. L. Warnock, Phys. Rev. **180**, 1433 (1969); D. Atkinson and R. L. Warnock, Phys. Rev. **188** 2098 (1969).
[10] H. Slim, "The Existence of a Certain Class of Solutions of the Roy Equation," Groningen Internal Report no. 110 (1976).

Perturbation theory for Green's functions as an effective mass formalism

F. Krause

Physikalisches Institut der Universität Bonn, Bonn, F. R. Germany
(Received 7 February 1977)

A formal solution for Green's functions of the type $[\partial^2 + m^2 + gV(x)]G(x,y) = \delta(x-y)$ is presented which has the structure of an effective mass formalism. One first solves the free case $[gV(x) = 0]$ for given boundary conditions and then replaces the parameter m in the solution by a quantity depending on $V(x)$. The rules for this replacement are given, a connection with the Baker–Campbell–Hausdorff formula is established, and it is shown how the formalism unites different perturbation and approximation schemes.

INTRODUCTION

There exist quite a few methods in theoretical physics which employ Green's functions of the type

$$[\partial^2 + m^2 + gV(x)]G(x,y) = \delta(x-y).$$

Here ∂^2 is a second order differential operator with constant coefficients ($\partial^2 = \sum_{\nu=1}^{n} \partial_\nu \partial_\nu^2$, $\partial_\nu = \partial/\partial x^\nu$). For a complete specification of $G(x,y)$ one has to fix some additional boundary conditions.

Generally there is no problem to evaluate $G(x,y)$ for the "free case" $[gV(x) \equiv 0]$ for example by the use of Fourier-like representations. If $gV(x)$ is not a constant, however, the solution becomes difficult due to the noncommutativity of ∂^2 and $V(x)$. Apart from general existence proofs one has to take recourse to approximations and perturbation expansions which mostly take the form of a formal solution of the problem.

In this paper it is shown that one can mould all these different solution formulas (with corresponding different boundary conditions, dimensions, differential operators, etc.) into one unified formalism which can be comprised as follows:

Calculate the Green's function for the free case $[gV(x) \equiv 0]$ with the desired boundary conditions. Then replace (in the solution) the constant m^2 by some sort of effective mass (squared) which depends on $gV(x)$. The resulting expression is a formal solution of $G(x,y)$. Formal solutions like the Born series, the lightcone or Hadamard expansion, the eikonal approximation, etc. then simply arise via different expansions of the resulting expressions.

In Sec. 1 of this paper the rules for the construction of the effective mass are described, and a proof is given that one indeed gets a formal solution. The results are somewhat more general than sketched in this introduction. Questions of convergence are not touched in this paper.

In Sec. 2 it is shown that the method may be interpreted as a special generalization of the Baker–Campbell–Hausdorff formula. This interpretation leads to some further insight and results for the formalism. For instance, one can easily derive the connection with Schwinger's proper time formalism. In Sec. 4 it is sketched how one can get the results from a path integral representation.

In Sec. 3 the method is applied to the harmonic oscillator which allows explicit evaluation for arbitrary dimension.

In Secs. 5 and 6 the Hadamard or lightcone expansion is derived, and a second representation is given which is suited to study the coincidence limit $x - y \to 0$ for $G(x,y)$ necessary for field theoretic applications. For both representations the method easily fixes the boundary conditions which are otherwise difficult to include in these expansions.

In Sec. 7 the connection with the eikonal approximation for $G(x,y)$ is given.

1. RESULTS FOR GREEN'S FUNCTIONS

In this section the main result is stated and proved as a formula for certain Green's functions. The assumptions regarding these Green's functions are the following:

(1) *Differential equation*: $G(x,y)$ has to be a solution of

$$[\partial^2 + m^2 + gV(x)]G(x,y) = \delta(x,y),$$
$$x, y \in \mathbb{R}^n, \quad \delta(x,y) \equiv \delta(x-y). \tag{1.1}$$

The differential operator ∂^2 is given by $\partial^2 = \sum_{\nu=1}^{n} a_\nu \partial_\nu^2$ with $\partial_\nu = \partial/\partial x^\nu$ and $a_\nu \in \mathbb{C}$, $a_\nu \neq 0$. To simplify the notation, a metric $g^{\mu\nu} = a_\mu \delta^{\mu\nu}$ is introduced and employed to define scalar products like $\partial^2 = \sum_{\mu,\nu} g^{\mu\nu} \partial_\mu \partial_\nu$. Special cases are the Laplace operator $-\Delta = -\sum \partial_\nu^2$ and the wave operator $\Box = \partial_0^2 - \vec{\partial}^2$. For the Laplace operator the one-dimensional case ($n = 1$) is allowed, leading to an ordinary differential equation.

The real parameter m^2 is separated for convenience from $gV(x)$, and $m^2 < 0$ is admitted.

The resulting formula for $G(x,y)$ will have a formal character which means that all questions of convergence are left aside in this paper. Therefore, no restrictions are imposed upon the "potential function" $gV(x)$ except that it is assumed to be C^∞. The possibility that this property is simply enforced in more general cases by interpreting $V(x)$ as a distribution is included.

(2) *Boundary conditions*: (1.1) is not sufficient to specify $G(x,y)$. One has to supply certain boundary conditions.

For a precise formulation let the "*generalized free propagator* $G_z^{(0)}(x,y)$ corresponding to $G(x,y)$" be de-

fined as a solution of

$$(\partial^2 + m^2)^z G_z^{(0)}(x,y) = \delta(x,y) \quad (z \in \mathbb{C}). \tag{1.2a}$$

Contrary to (1.1) this equation can be transformed into a division problem by Fourier methods. Therefore, it is assumed that (1.2a) can be solved by an integral representation

$$G_z^{(0)}(x,y) = (m^2)^{-z} \int^{(R)} \frac{d^n p}{(2\pi)^n} \exp[-ip(x-y)] \left(\frac{m^2}{-p^2+m^2}\right)^z, \tag{1.2b}$$

where (R) stands for a prescription of integration which is (a) independent of z and (b) fixes the boundary conditions for $G_z^{(0)}(x,y)$. (R) generally describes a contour of integration in complexified p-space.

Now the *main assumption* is that *the boundary condition for $G(x,y)$ shall be the same as that for $G_z^{(0)}(x,y)$.*

Many boundary conditions which are frequently used in physical applications admit a representation (1.2b). In the Euclidean case with $\partial^2 = -\Delta$ and $m^2 > 0$ ordinary Fourier transformation is a simple example leading to L^2 boundary conditions for $\mathrm{Re} z > n/2$. If $m^2 < 0$, the $\pm i\epsilon$ prescriptions lead to boundary conditions of the radiation type. In the Minkowski case one can use for example Feynman or retarded boundary conditions. A compilation of corresponding prescriptions of integration can be found in Ref. 1.

Equation (1.2b) implies that $G_z^{(0)}$ is an analytic family of distributions[2] in z. For these distributions the following three identities hold which prove to be crucial for the following in the sense that they may be abstracted from (1.2b) and then suffice to establish the solution formula for $G(x,y)$:

$$\frac{\partial}{\partial m^2} G_z^{(0)}(x,y) = (-z) G_{z+1}^{(0)}(x,y), \tag{1.3a}$$

$$(\partial^2 + m^2) G_z^{(0)}(x,y) = G_{z-1}^{(0)}(x,y), \tag{1.3b}$$

$$2\frac{\partial}{\partial x^\mu} G_{z+1}^{(0)}(x,y) = \frac{1}{z}(x-y)_\mu G_z^{(0)}(x,y), \quad z \neq 0. \tag{1.3c}$$

It proves useful to make explicit the dependence of $G_z^{(0)}(x,y)$ on the parameter m^2. The notation therefore is changed to $G_z^{(0)}(x,y \mid m^2)$. Now the main result can be stated as follows:

The Green's function $G(x,y)$ obeying (1.1) and the same boundary conditions as (1.2b) is given by the formal substitution:

$$G(x,y) = G_1^{(0)}(x,y \mid E^2 + \partial_v^2 + g\textstyle\int_0^1 d\alpha V_\alpha) \tag{1.4}$$

This equation has to be interpreted as follows:

(1) V_α is a shorthand notation for $V(\alpha x + (1-\alpha)y)$. One should remark that $\int_0^1 d\alpha V(\alpha x + (1-\alpha)y)$ depends on the variables x and y in a symmetric way contrary to $V(x)$ contained in the differential operator. Furthermore, $\int_0^1 d\alpha V_{\alpha|_{x=y}} = V(x)$.

(2) The right-hand side of (1.4) has to be expanded into a formal series of powers $(\partial_v^2)^L (\int_0^1 d\alpha V_\alpha)^N$. For this expansion the quantity ∂_v^2 acts as an ordinary number commuting with all other quantities. All products have to be ordered in the standard form given above. If a product does not contain any factor $\int_0^1 d\alpha V_\alpha$, one has to choose $N=0$. With regard to powers of $\int_0^1 d\alpha V_\alpha$, however, ∂_v^2 acts as an operator whose precise action will be given below.

(3) In general there are available different representations for $G_z^{(0)}(x,y \mid m^2)$ (integrals, series, etc.). The expansion described in (2) may be done for the integrands, single terms, etc. of these representations.

Example: Starting from the Fourier representation (1.2b) the substitution rule (1.4) leads to

$$\begin{aligned} G(x,y) &= \int^{(R)} \frac{d^n p}{(2\pi)^n} \frac{\exp[-ip(x-y)]}{(-p^2 + m^2 + \partial_v^2 + g\int_0^1 d\alpha V_\alpha)} \\ &= \sum_{n=0}^{\infty} (-)^n \left(\partial_v^2 + g\int_0^1 d\alpha V_\alpha\right)^n \\ &\quad \times \int^{(R)} \frac{d^n p}{(2\pi)^n} \frac{\exp[-ip(x-y)]}{(-p^2+m^2)^{n+1}}. \end{aligned} \tag{1.5}$$

This formula elucidates the way the boundary conditions for $G(x,y)$ are implemented via (1.2b).

If one collects terms with equal power of the coupling constant g, it is possible to show the identity of the above expression and the Born series for $G(x,y)$.[3]

One can give two equivalent definitions for the action of ∂_v^2. In Sec. 2 there will be a short indication how one can verify this equivalence.

First definition:

$$\begin{aligned} &(\partial_v^2)^L (\textstyle\int_0^1 d\alpha V_\alpha)^0 = (\partial_v^2)^L 1 = 0, \\ &(\partial_v^2)^L (\textstyle\int_0^1 d\alpha V_\alpha)^N = \int_0^1 d\alpha_1 \cdots \int_0^1 d\alpha_N \\ &\qquad \times \left[\sum_{i,j=1}^N g(\alpha_i, \alpha_j) \partial_i \circ \partial_j\right]^L V_{\alpha_1} \cdots V_{\alpha_N}. \end{aligned} \tag{1.6}$$

Here ∂_i designs the derivation of the argument of V_{α_i}, that is, $\partial_i V_{\alpha_i} = (\partial V)_{\alpha_i} = (\partial V)[\alpha_i x + (1-\alpha_i)y]$. Furthermore, $\partial_i \circ \partial_j$ contains a scalar product with respect to the metric: $\partial_i \circ \partial_j = \sum_{\mu,\nu} \partial_{i\mu} g^{\mu\nu} \partial_{j\nu}$. The kernel g is given by

$$\begin{aligned} g(\alpha,\beta) &= \Theta(\alpha-\beta)\beta(1-\alpha) + \Theta(\beta-\alpha)\alpha(1-\beta) \\ &= \tfrac{1}{2}\{\alpha + \beta - 2\alpha\beta - (\alpha-\beta)\epsilon(\alpha-\beta)\}, \end{aligned} \tag{1.7}$$

which implies $g(\alpha,\alpha) = \alpha(1-\alpha)$.

To understand the structure of (1.6), one should note the following: If one replaces $g(\alpha,\beta)$ by $\alpha(1-\beta)$, then ∂_v^2 turns into $\partial_x \circ \partial_y$. One has $\alpha_i \partial_i V_{\alpha_i} = \partial_x V_{\alpha_i}$ and $(1-\alpha_i)\partial_i V_{\alpha_i} = \partial_y V_{\alpha_i}$, and Eq. (1.6) after this change simply states the result of the application of the product rule for $[\partial_x \circ \partial_y]^L$. One should also remark that $g(\alpha,\beta)$ is the unique solution of the following boundary problem [$g(\alpha,\beta)$ is also known as "der Hamelsche Musterkern"]:

$$\begin{aligned} &-\frac{d^2}{d\alpha^2} g(\alpha,\beta) = \delta(\alpha,\beta), \quad g(\alpha,\beta) = g(\beta,\alpha) \\ &(0 < \alpha,\beta < 1) \quad g(0,\alpha) = g(1,\alpha) = 0. \end{aligned} \tag{1.8}$$

The second announced representation for the action of ∂_v^2 originates from a generalization of the identity

$$(\textstyle\int_0^1 d\alpha V_\alpha)^N = N! \int_0^1 d\alpha_1 \int_0^1 d\alpha_2 \cdots \int_0^1 d\alpha_N \Theta_\alpha V_{\alpha_1} \cdots V_{\alpha_N},$$

where the abbreviation $\Theta_\alpha = \Theta(\alpha_1 - \alpha_2)\Theta(\alpha_2 - \alpha_3)\cdots \times \Theta(\alpha_{N-1} - \alpha_N)$ is introduced.

Second definition:

$$(\partial_v^2)^L \left(\int_0^1 d\alpha V_\alpha\right)^N = N! \int_0^1 d\alpha_1 \cdots \int_0^1 d\alpha_N \Theta_\alpha \times \left[\sum_{i=1}^n \alpha_i(1-\alpha_i)\partial_i^2 + 2\sum_{i>j} \alpha_i(1-\alpha_j)\partial_i\partial_j\right]^L \times V_{\alpha_1}\cdots V_{\alpha_N} \tag{1.9}$$

Replacing in this formula $2\sum_{i>j}$ by $\sum_{i\neq j}$ one again finds a representation of $(\partial_x \cdot \partial_y)^L$. Because of the factor Θ_α the integrand is not invariant with respect to the replacement $\alpha_i \leftrightarrow \alpha_j$, and therefore this modification changes the value of (1.9).

In this sense ∂_v^2 may be interpreted as a nontrivial modification of the second order differential operator $\partial_x \cdot \partial_y$. The meaning of this modification will become clearer by the proof of Proposition 1 while the reason for the appearance of the kernel $g(\alpha, \beta)$ will be explained in Sec. 4.

One should remark that (1.6) or (1.9) imply that $(\partial_v^2)^L (\int_0^1 d\alpha V_\alpha)^N$ always has a finite limit for $y \to x$ [if $V(x)$ and its derivatives are finite].

There are quite a few possibilities of generalizing the method. One can take a more general (for example, x-dependent) matrix $g_{\mu\nu}$ or one can replace (1.1) by a matrix equation.[3] Furthermore, one can regard equations of the Dirac type or equations of the structure $[\Box + m^2 + A(x)\cdot\partial + gV(x)]G(x,y) = \delta(x,y)$ instead of (1.1). Such equations arise for example from a minimal coupling of the electromagnetic field or more generally in field theories of the Yang–Mills type. In both cases one has to introduce a quantity ∂_v besides ∂_v^2 obeying similar rules. The technical details will be given in a forthcoming paper of the author. Two generalizations of a different kind are already included in the following proposition. The first is the possibility to choose another inhomogeneous term $G_0^{(0)}$ in (1.1) instead of $\delta(x,y)$. The proposition even comprises the homogeneous case ($G_0^{(0)} = 0$). The second is the generalization of the method to iterated differential operators $(\partial^2 + m^2 + gV(x))^k$.

Proposition: Let $G^{(0)}(x,y\,|\,m^2)$ with $z \in \mathbb{C}$ be an analytic family of distributions obeying relations (1.3a)–(1.3c) for $z \in \mathbb{C}$, $\mathrm{Re}\, z > -\epsilon$, $\epsilon > 0$, $-\eta < \mathrm{Im}\, z < +\eta$, $\eta > 0$. Let G_z be given by

$$G_z(x,y) = G_z^{(0)}(x,y\,|\,m^2 + \partial_v^2 + g\textstyle\int_0^1 d\alpha V_\alpha),$$

where the right-hand side is defined as described above. Then for $\mathrm{Re}\, z > 1 - \epsilon$ one has

$$[\partial^2 + m^2 + gV(x)]G_z(x,y) = G_{z-1}(x,y).$$

Proof: Expansion of $G_z^{(0)}$ around m^2 and application of (1.3a) leads to

$$G_z(x,y) = \sum_{l,k=0}^{\infty} (-)^{l+k} \frac{(z)_{l+k}}{l!k!} (\partial_v^2)^k \left(g\int_0^1 d\alpha V_\alpha\right)^l \times G^{(0)}_{l+k+z}(x,y), \tag{1.10}$$

with $(z)_N = z(z+1)(z+2)\cdots(z+N-1)$. One has to evaluate $[\partial^2 + m^2 + gV(x)]G_z(x,y)$. There is just one contribution which is problematic arising from the application of ∂^2, namely,

$$M = \sum_{\substack{k=0\\ l=1}}^{\infty} (-)^{l+k} \frac{(z)_{l+k}}{l!k!}\, 2\left[\partial_x (\partial_v^2)^k \left(g\int_0^1 d\alpha V_\alpha\right)^l\right] \times \partial_x G^{(0)}_{l+k+z}(x,y).$$

Application of (1.3c) and (1.9) and taking into regard the identity $(x-y)\partial_x V_\alpha = \alpha(d/d\alpha)V_\alpha$ implies

$$M = \sum_{\substack{k=0\\ l=1}}^{\infty} (-)^{l+k} \frac{g^l (z)_{l+k}}{l!k!(l+k+z-1)}\, l! \int_0^1 d\alpha_1 \cdots \int_0^1 d\alpha_l \Theta_\alpha \times [\cdots]^k \left[\sum_m \alpha_m \frac{d}{d\alpha_m}\right] V_{\alpha_1}\cdots V_{\alpha_l} G^{(0)}_{l+k+z-1}(x,y),$$

where $[\cdots]$ stands for $\sum_i \alpha_i(1-\alpha_i)\partial_i^2 + 2\sum_{i>j} \times \alpha_i(1-\alpha_j)\partial_i\partial_j$. Now the integrand can be rewritten as follows:

$$[\cdots]^k \left(\sum \alpha_m \frac{d}{d\alpha_m}\right) = \left(\sum \alpha_m \frac{d}{d\alpha_m}\right)[\cdots]^k - k[\cdots]^k + k[\cdots]^{k-1}\left(\sum \alpha_i\alpha_j\partial_i\partial_j\right).$$

The last two terms lead to a contribution M_2 to M which is given by

$$M_2 = \sum_{\substack{l=1\\ k=0}}^{\infty} (-)^{l+k} \frac{(z)_{l+k}k}{l!k!(l+k+z-1)} \left[-(\partial_v^2)^k + \partial_x^2(\partial_v^2)^{k-1}\right] \times \left(g\int_0^1 d\alpha V_\alpha\right)^l G^{(0)}_{l+k+z-1}(x,y).$$

The remaining contribution M_1 is evaluated by partial integration. The lower boundary ($\alpha_m = 0$) does not contribute because of the factor α_m. The upper boundary contributes once: The factor $\Theta_\alpha = \Theta(\alpha_1 - \alpha_2)\Theta(\alpha_2 - \alpha_3)\cdots$ admits a nonzero term only if $m = 1$. In this way V_{α_1} is changed into a factor $V(x)$ which may be commuted with $[\cdots]^k_{\alpha_1=1}$ because this expression does not contain any derivative ∂_1 operating on V_{α_1} and therefore on $V(x)$. This very fact is achieved by the condition $i > j$ for the sum $\alpha_i(1-\alpha_j)\partial_i\partial_j$ in (1.9) which eliminates the critical terms $\alpha_1(1-\alpha_j)\partial_1\partial_j$ ($j \neq 1$) which do not vanish for $\alpha_1 = 1$.

The partially integrated term contains the factor $\sum_{m=1}^{l}\{\Theta_\alpha + \alpha_m d\Theta_\alpha/d\alpha_m\}$. It is easy to see that the second contribution vanishes implying a value $l\Theta_\alpha$ for this sum. Altogether one finds for M_1

$$M_1 = \sum_{\substack{k=0\\ l=1}}^{\infty} (-1)^{l+k} \frac{(z)_{l+k}l}{l!k!(l+k+z-1)} \left[gV(x)(\partial_v^2)^k \times \left(g\int_0^1 d\alpha V_\alpha\right)^{l-1} - (\partial_v^2)^k \left(g\int_0^1 d\alpha V_\alpha\right)^l\right] G^{(0)}_{l+k+z-1}(x,y).$$

Using this expression for $M = M_1 + M_2$ most contributions of $[\partial^2 + m^2 + gV(x)]G(x,y)$ cancel after an application of (1.3a). There remain two sums which can be combined. The resulting sum is again (1.10) with z replaced by $z-1$.

2. CONNECTION WITH THE BAKER-CAMPBELL-HAUSDORFF FORMULA

In this section the results of the preceding section are interpreted within the context of semigroup theory. To avoid technical complications, the metric is restricted

to the Euclidean case: $g^{\mu\nu} = -\delta^{\mu\nu}$ or $\partial^2 = -\Delta$. Furthermore, $m^2 > 0$ is assumed. The results, however, again hold for more general cases.

The leading idea is to introduce an integral representation of $G_z^{(0)}$ with an integrand containing an expression $\exp[\epsilon t(\partial^2 + m^2)]$ ($\epsilon = \pm 1, \pm i$). This means that one integrates over the elements of a semigroup $t \to \exp[\epsilon t(\partial^2 + m^2)]$ with infinitesimal generator $(\partial^2 + m^2)$. The substitution rule leads to an integrand $\exp[\epsilon t(\partial_v^2 + g \times \int_0^1 d\alpha V_\alpha)] \exp[\epsilon t(\partial^2 + m^2)]$ for $G_z(x,y)$. On the other hand this quantity is obviously related to the semigroup $\exp[\epsilon t(\partial^2 + m^2 + gV(x))]$ via the corresponding integral representation for $G_z(x,y)$.

For a Euclidean metric, $m^2 > 0$ and $\mathrm{Re}\, z > 0$ one can represent $G_z^{(0)}$ as follows:

$$G_z^{(0)}(x,y) = \frac{1}{\Gamma(z)} \int_0^\infty dt\, t^{z-1} \exp[-t(-\Delta + m^2)]\delta(x,y)$$
$$= \left(\frac{1}{4\pi}\right)^{n/2} \frac{1}{\Gamma(z)} \int_0^\infty dt\, t^{z-1-n/2} \exp(-tm^2) \times \exp\left(-\frac{(x-y)^2}{4t}\right). \quad (2.1)$$

Then the proposition of the preceding section implies (with an obvious change of notation $\partial_v^2 \to -\Delta_v$):

$$G_z(x,y) = \frac{1}{\Gamma(z)} \int_0^\infty dt\, t^{z-1} \exp\left[-t\left(-\Delta_v + g\int_0^1 d\alpha V_\alpha\right)\right] \times \exp[-t(-\Delta + m^2)]\delta(x,y)$$
$$= \left(\frac{1}{4\pi}\right)^{n/2} \frac{1}{\Gamma(z)} \int_0^\infty dt\, t^{z-1-n/2} \cdot \exp\left[-t\left(-\Delta_v + g\int_0^1 d\alpha V_\alpha\right)\right] \cdot \exp\left(-tm^2 - \frac{(x-y)^2}{4t}\right). \quad (2.2)$$

On the other hand G_z is given at least formally by

$$G_z(x,y) = [-\Delta + m^2 + gV(x)]^{-z}\, \delta(x,y)$$
$$= \frac{1}{\Gamma(z)} \int_0^\infty dt\, t^{z-1} \exp\{-t[-\Delta + m^2 + gV(x)]\}\delta(x,y)$$
$$= \left(\frac{1}{\sqrt{4\pi}}\right)^n \frac{1}{\Gamma(z)} \int_0^\infty dt\, t^{z-1-n/2} \times (\exp\{-t[-\Delta + m^2 + gV(x)]\} \exp[t(\Delta + m^2)]) \exp\left(-tm^2 - \frac{(x-y)^2}{4t}\right). \quad (2.3)$$

Comparison shows

$$\exp\{-t[-\Delta + m^2 + gV(x)]\}\delta(x,y) = \exp[-t(-\Delta_v + g\int_0^1 d\alpha V_\alpha)] \exp[-t(-\Delta + m^2)]\delta(x,y) \quad (2.4)$$

Within the framework of quantum mechanics $\exp[-it(-\Delta_v + g\int_0^1 d\alpha V_\alpha)]$ therefore can be interpreted as the configuration space representation $\langle x | U(t) | y\rangle$ of the evolution operator $U(t) = \exp(-itH)\exp(itH_0)$ leading to the Möller operators.

If one expands $\exp[-t(\Delta_v + g\int_0^1 d\alpha V_\alpha)]$ into a power series, one finds a representation which is used in Schwinger's proper time formalism[1,4] modified for the integral representation (2.1).

In Schwinger's method one introduces (for $z - 1$) the quantity

$$\langle xt | y0\rangle = \exp[-t[-\Delta + gV(x)]]\delta(x,y),$$

which obeys $\langle x0 | y0\rangle = \delta(x,y)$ and $-(\partial/\partial t)\langle xt | y0\rangle = [-\Delta + gV(x)]\langle xt | y0\rangle$. For the ansatz

$$\langle xt | y0\rangle = \frac{1}{(4\pi)^{n/2}}\, t^{-n/2} \exp\left(-\frac{(x-y)^2}{4t}\right) \sum_{n=0}^\infty a_n(-t)^n$$

the coefficients $a_n = a_n(x,y)$ are computed recursively. Comparing this with (2.4) and (2.1), one finds

$$\exp[-t(-\Delta_v + g\int_0^1 d\alpha V_\alpha)] = \sum_{n=0}^\infty a_n(-t)^n,$$

which means that our formalism solves the mentioned recursion relations. Equation (2.4) suggests a comparison with the Baker–Campbell–Hausdorff (BCH) formula.[5] Qualitatively one could proceed as follows to derive the above result: First one applies the BCH formula to achieve a factorization of the form $\exp[t\Delta - gtV(x)] = \exp(t\Delta)\exp(Z)$. The quantity Z introduced in this way depends on the derivatives of V. Hereafter one rewrites all terms of Z by use of the product rule in such a way that one can apply the identity $f(x)\delta(x,y) = \int d\alpha f(\alpha x + (1-\alpha)y)\delta(x,y)$ to all arguments. For example, $[\Delta, V(x)] = 2\partial(\partial V)(x) - (\Delta V)(x)$. Finally $\exp(t\Delta)$ is commuted to the right.

The remarkable fact is that the resulting expression can be written in such a simple closed form as (2.4) using the quantity Δ_v. In this way one may regard (2.4) as a special generalization of the Baker–Campbell–Hausdorff formula. From (2.4) one immediately derives a set of identities which hold for the general metric and can be checked explicitly for low N:

Corollary: For $N = 1, 2, 3, \cdots$ one has

$$(\partial^2 + m^2 + gV(x))^N \delta(x,y) = (\partial_0^2 + m^2 + \partial_v^2 + g\int_0^1 d\alpha V_\alpha)^N \delta(x,y), \quad (2.5)$$

with

$$[\partial_0^2, \partial_v^2] = [\partial_0^2, g\int_0^1 d\alpha V_\alpha] = [\partial_v^2, g\int_0^1 d\alpha V_\alpha] = 0$$

and

$$\partial_0^2 \delta(x,y) = \delta^2\delta(x,y).$$

These identities comprise the algebraic part of the results of Sec. 1. The last commutator $[\partial_v^2, g\int_0^1 d\alpha V_\alpha] = 0$ represents the demand that all products should be ordered in the standard form $(\partial_v^2)^L (\int_0^1 d\alpha V_\alpha)^N$ and that this ordering never introduces any commutation terms contrary to ordering differential operators. For example, $(\partial_v^2 + g\int_0^1 d\alpha V_\alpha)^2 = (\partial_v^2)^2 \cdot 1 + 2\partial_v^2 g\int_0^1 d\alpha V_\alpha + (g\int_0^1 d\alpha V_\alpha)^2$. The corollary once more illustrates the following essential point: Let f be some function of the operator $\partial^2 + m^2 + gV(x)$. Then the kernel $\langle x | f(\partial^2 + m^2 + gV(x)) | y\rangle$ of the configuration space representation of f can be reduced to the corresponding free kernel $\langle x | f(\partial^2 + M^2) | y\rangle$ with an "effective mass term" $M^2 = m^2 + \partial_v^2 + g\int_0^1 d\alpha V_\alpha$. All this holds provided the algebraic identities are not spoiled by topological properties of the respective operators.

Furthermore, one should remark that the separation of $\partial^2 + m^2 + gV(x)$ into a "free" and an "interaction" part

given by (2.2) makes possible a continuation of the coefficients of the metric $g^{\mu\nu}$ which only refers to the free part while keeping fixed the interacting part.

Next it is shown that a partial evaluation of $\exp[t(\partial_v^2 + g\int_0^1 d\alpha V_\alpha)]$ is possible for general V. From (1.6) one finds

$$(\partial_v^2)^L (g\int_0^1 d\alpha V_\alpha)^N = \int_0^1 d\alpha_1 \cdots \int_0^1 d\alpha_N \times \left[\sum_{i=1}^N g(\alpha_i,\alpha_i)\partial_i^2 + 2\sum_{i>j} g(\alpha_i,\alpha_j)\partial_i\partial_j\right]^L V_{\alpha_1}\cdots V_{\alpha_N}. \tag{2.6}$$

Application of the multinomial theorem leads to a sum each term of which may be represented by a graph G with N *ordered* vertics (labeled $1,2,\ldots,N$) and L lines. Each vertex i represents a factor V_{α_i} and a line between two vertices i and j describes a factor $g(\alpha_i,\alpha_j)\partial_i\cdot\partial_j$. Self-contractions (tadpoles) are admitted. Conversely each such graph corresponds to a unique contribution to the sum.

Graphs which only differ by a permutation of their vertex ordering lead to equal contributions, as such a permutation can be achieved by a renaming of the variables of integration α_i. Now let G be a given graph (with ordered vertices) and suppose that there are $N!/\kappa_G$ permutations of the vertices which leave G invariant. ($N!/\kappa_G$ therefore is the order of the invariance group of G). Then there exist κ_G equal terms in the above sum which may be added up. The resulting contribution can be described by a graph with *unordered* vertices in a unique way.

For each graph G with unordered vertices an arbitrary enumeration $1,2,\ldots,N$ of the vertics is fixed. Then the incidence matrix M^G of G is defined by

$$M^G_{ij} = M^G_{ji} = \text{number of lines connecting vertices } i \text{ and } j. \tag{2.7}$$

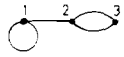
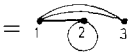
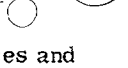
A tadpole line is counted only once.

Employing this matrix, one can immediately write down the contribution $B(G)$ corresponding to a given graph. One finds:

$$\boxed{\begin{aligned} B(G) &= K(G)\int_0^1 d\alpha_1\cdots\int_0^1 d\alpha_N \\ &\quad\times \prod_{i>j}\left[g(\alpha_i,\alpha_j)\partial_i\partial_j\right]^{M^G_{ij}} V_{\alpha_1}\cdots V_{\alpha_N} \\ K(G) &= \frac{L!\,2^{\sum_{i>j} M^G_{ij}}}{\prod_{i\geq j}(M^G_{ij})!}\kappa_G \end{aligned}} \tag{2.8}$$

and

$$(\partial_v)^L (g\int_0^1 d\alpha V_\alpha)^N = \sum_{G\in G(N,L)} B(G).$$

Example: The two graphs $G_1 =$  and $G_2 =$  with ordered vertices give rise to the same contribution. The ordering is represented as geometrical order from left to right. Both correspond to $G =$  with unordered (though labelled) vertices and

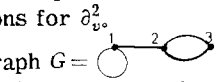
$$M^G = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}, \quad \kappa_G = 3! = 6,$$

$$B(G) = \frac{4!\cdot 2^3\cdot 3!}{2!}\int_0^1 d\alpha_1\int_0^1 d\alpha_2\int_0^1 d\alpha_3 \times g(\alpha_1\alpha_2)\partial_1^2\cdot g(\alpha_1\alpha_2)\partial_1\partial_2\cdot(g(\alpha_2\alpha_3)\partial_2\partial_3)^2 \times V_{\alpha_1}V_{\alpha_2}V_{\alpha_3}.$$

Inserting the representation

$$g(\alpha_i,\alpha_j) = \Theta(\alpha_i-\alpha_j)\alpha_j(1-\alpha_i) + \Theta(\alpha_j-\alpha_i)\alpha_i(1-\alpha_j)$$

into (8), one gets a sum of 2^{L-T} terms where each line between *different* vertices produces a factor $\Theta(\alpha_i-\alpha_j) = \Theta_{ij}$. T is equal to the number of tadpoles in G. The resulting products of Θ functions are either zero or can be broken up into a sum of Θ functions representing respectively a complete ordering of the variable α_i. After a renaming of the α_i and some combinatorial considerations one can prove that (1.6) and (1.9) give identical expressions for ∂_v^2.

Example: The graph $G =$  divides into $2^3 = 8$ parts. One part for instance contains the factor $\Theta_{12}\Theta_{23}\Theta_{32} = 0$. Another contains $\Theta_{12}(\Theta_{32})^2 = \Theta_{12}\Theta_{32} = \Theta_{13}\Theta_{32} + \Theta_{31}\Theta_{12}$. Suitable renamings of the α_i lead to contributions which contain $\Theta_\alpha \equiv \Theta_{12}\Theta_{23} = \Theta(\alpha_1-\alpha_2)\times\Theta(\alpha_2-\alpha_3)$ characteristic for the representation (1.9) of ∂_v^2.

Next one observes that the graphs G may be disconnected. Let $\mathcal{G}_c$ be the set of all connected graphs and write $\mathcal{G}_c = \{g_0, g_1, g_2, \cdots\}$. For each $g_i \in \mathcal{G}_c$ an enumeration of the vertices is fixed. Furthermore, let p_i be the number of vertices and l_i be the number of lines of g_i.

Now each graph G (possibly disconnected) can be characterized by giving the number of its respective connected components. This is denoted symbolically by

$$G(g_0^{\lambda_0}, g_1^{\lambda_1}, g_2^{\lambda_2}, \cdots).$$

Remembering that $N!/\kappa_G$ is equal to the order of the invariance group of G, one finds

$$\kappa_G = \frac{N!\,(\kappa_{g_0})^{\lambda_0}(\kappa_{g_1})^{\lambda_1}\cdots}{(p_0!)^{\lambda_0}(p_1!)^{\lambda_1}\cdots\lambda_0!\,\lambda_1!\cdots}.$$

The structure of this expression—especially the appearence of the factors $1/\lambda_i!$—together with the possibility of factorizing the α integrations into integrals corresponding to the connected parts of G—indicate that (2.8) might be a contribution to a multipole exponential sum.

Adding up all terms of

$$\exp(t\partial_v^2)\exp\left(tg\int_0^1 d\alpha V_\alpha\right) = \sum_{L,N}\frac{t^{L+N}g^N}{L!N!}(\partial_v^2)^L\left(\int_0^1 d\alpha V_\alpha\right)^N,$$

one indeed finds a complete multiple exponential sum, namely:

Proposition:

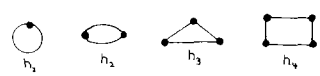
$$\exp(t\partial_v^2)\exp\left(tg\int_0^1 d\alpha V_\alpha\right) = \exp\left(\sum_{g_i\in\mathcal{G}_c}\frac{t^{l_i+p_i}g^{p_i}}{p_i!l_i!}B(g_i)\right) \tag{2.9}$$

Here $g_i\in\mathcal{G}_c$ runs over all *connected* graphs and the contribution $B(g_i)$ again is given by (2.8). Furthermore, l_i is the number of lines and p_i the number of vertices of g_i.

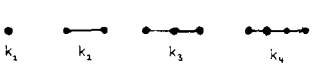
3. THE HARMONIC OSCILLATOR

In this section the formalism is applied to the case of the harmonic oscillator allowing complete evaluation. One finds a summation procedure leading from the formal Born series of the propagator to the energy values and illustrating the method.

One may start from (2.1). In the case of dimension one ($n=1$) the free propagator is simply given by $G^{(0)}(x,y)=(1/2m)\exp(-m|x-y|)$. The substitution rule leads to (2.2) where one has to replace $gV(x)$ by gx^2. Finally $\exp(t\Delta_v - tg\int_0^1 d\alpha(x^2)_\alpha)$ has to be evaluated by (2.9). Each graph containing a vertex with more then two derivatives vanishes. Therefore only two types of graphs contribute:



$$\kappa(h_i)=\tfrac{1}{2}(i-1)!,\quad i\neq 1,2,$$



$$\kappa(k_i)=\tfrac{1}{2}i!,\quad i\neq 1.$$

It is easy to evaluate the combinational factors corresponding to these graphs. They are

$$K(h_i)/p_i!l_i! = 2^{i-1}/i,\quad i=1,2,\cdots,$$
$$K(k_1)=1,\quad K(k_i)/p_i!l_i!=2^{i-2},\quad i=2,3,\cdots. \tag{3.1}$$

Corresponding to these two types of graphs one finds from (2.9) of the preceding section $\exp[t\Delta_v\int_0^1 d\alpha(x^2)_\alpha] = \exp(S_h+S_k)$ with

$$S_h=\sum_{i=1}^{\infty}\frac{(-)^i}{i}t^{2i}g^i2^{i-1}\int_0^1 d\alpha_1\cdots\int_0^1 d\alpha_i \times g(\alpha_1,\alpha_2)g(\alpha_2,\alpha_3)\cdot\cdots\cdot g(\alpha_{i-1}\alpha_i)g(\alpha_i,\alpha_1)\cdot 2^i\cdot n. \tag{3.2}$$

The factor n (equal to the dimension) arises from a trace operation due to the scalar products $\partial_i\cdot\partial_j$.

S_h is independent of the configuration space variables x and y while S_k is quadratic in these quantities. Restoring factors of $\hbar$ one finds that S_k is proportional to $\hbar^{-1}$ while S_h is independent of $\hbar$. More generally tree graphs of $\mathcal{G}_c$ always represent the leading contribution in $\hbar$.

The integrals may be evaluated by observing from (1.8) that one has the representation:

$$g(\alpha,\beta)=\frac{2}{\pi^2}\cdot\sum_{k=1}^{\infty}\frac{1}{k^2}\sin(k\pi\alpha)\sin(k\pi\beta) \tag{3.3}$$

From the orthogonality of $\sin(k\pi\alpha)$ on $0\leq\alpha\leq 1$ one concludes that

$$S_h=\tfrac{1}{2}n\left(\sum_{\nu=1}^{\infty}\frac{2i\sqrt{gt}}{\pi}\right)^{2\nu}\frac{\zeta(2\nu)}{\nu} = -\tfrac{1}{4}n\sum_{\nu=1}^{\infty}(4\sqrt{gt})^{2\nu}\frac{B_{2\nu}}{\nu(2\nu)!}.$$

ζ is the Riemannian ζ function and B_n are the Bernoulli numbers. The sum is convergent if $4\sqrt{gt}<\pi$ and can be evaluated explicitly:

$$S_h=-\tfrac{1}{2}n[\ln\sinh(2\sqrt{gt})-\ln(2\sqrt{gt})]$$

or

$$\exp(S_h)=[2\sqrt{gt}/\sinh(2\sqrt{gt})]^{n/2}, \tag{3.4}$$

which is regular for $t\to 0$ and decreases exponentially for $t\to\infty$.

The second type of graphs leads to the sum

$$S_k=\sum_{\nu=2}^{\infty}(-)^\nu t^{2\nu-1}\cdot g^\nu\cdot 2^{\nu-2}\int_0^1 d\alpha_1\cdots\int_0^1 d\alpha_\nu 2^\nu g(\alpha_1,\alpha_2)\cdots \times g(\alpha_{\nu-1},\alpha_\nu)[\alpha_1 x+(1-\alpha_1)y][\alpha_\nu x+(1-\alpha_\nu)y].$$

From this one finds analogously after including the contribution due to k_1

$$\exp(S_k)=\exp\{-\tfrac{1}{2}\sqrt{g}(x^2+y^2)[\coth(2\sqrt{gt})-1/2\sqrt{gt}] + \sqrt{g}x\cdot y[\operatorname{csch}(2\sqrt{gt})-1/2\sqrt{gt}]\}, \tag{3.5}$$

which again is regular for $t\to 0$.

Collecting all results for $z=1$, one finds for the n-imensional harmonic oscillator with $\kappa^2=\sqrt{g}$

$$G^{(n)}(x,y)=\frac{\kappa^{n-2}}{(2\pi)^{n/2}}\int_0^\infty dt\left[\frac{1}{\sinh(2t)}\right]^{n/2}\exp\left(-\frac{m^2}{\kappa^2}-t\right) \times\exp\left\{-\tfrac{1}{2}\kappa^2\cdot(x^2+y^2)\coth(2t)-\frac{2xy}{\sinh(2t)}\right\}. \tag{3.6}$$

This formula can be verified for $n=1$ by explicit calculation of the resolvent. Cf., for example, the corresponding formulae in the book of Feynman and Hibbs.[6] One should remark that the factor $\exp[-\kappa^2(x-y)^2/4t]$, which arises from the free part of the Hamiltonian is still present in form of the singular parts ($t\to 0$) of $\coth(2t)$ and $1/\sinh(2t)$. This factor achieves convergence of the integral for $t\to 0$ and $x=y$ by compensating the singularity of $t^{-n/2}$ which is still present too. If $n>1$, this factor causes a singularity of $G^{(n)}(x,y)$ for $x-y\to 0$. For $n=1$ formula (3.6) even allows a continuation from $g>0$ to $-g$ (avoiding $g=0$), which means $\kappa^2\to\pm i\kappa^2$. The resulting eigenfunctions naturally are not L^2-integrable and the corresponding eigenvalues become complex.

The energy values for $g>0$ may be evaluated as follows: Let $R(x,y|E)$ be the resolvent which one gets from $G^{(n)}(x,y)$ by the replacement $m^2\to m^2-E$. Assume $n=1$ and define $\Delta^{(1)}R(x,y|E)=R(x,y|E)-R(x,y|0)$. Now one calculates $\int_{-\infty}^{+\infty}dx\,\Delta^{(1)}R(x,x|E)$ with the following result:

$$\int_{-\infty}^{+\infty}dx\,\Delta^{(1)}R(x,x|E) = \frac{E}{\kappa^4}\sum_{n=0}^{\infty}\frac{1}{(2n+m^2/\kappa^2-E/\kappa^2+1)(2n+m^2/\kappa^2+1)}.$$

It is easy to convince oneself that the same result follows from

$$R(x,y\,|\,E)=\sum_{n=0}^{\infty}\frac{\psi_n(x)\psi_n^*(y)}{\kappa^2(1+2n)+m^2-E},$$

where the $\psi_n(x)$ are properly normalized oscillator functions.

For $n=2$ one analogously defines $\Delta^{(2)}R(x,y\,|\,E)=R(x,y\,|\,E)-R(x,y\,|\,0)-E(\partial/\partial E)R(x,y\,|\,0)$. One finds

$$\int d^2x\,\Delta^{(2)}R(x,x\,|\,E)=\frac{1}{\kappa^2}\left(\frac{E}{\kappa^2}\right)^2\sum_{n=0}^{\infty}\frac{(n+1)}{(2+2n+m^2/\kappa^2-E/\kappa^2)(2+2n+m^2/\kappa^2)^2}$$

showing the correct multiplicities and zero point energy.

Finally one should remark that there are two different mechanisms which lead to the divergence of $\int dx\,R(x,x\,|\,E)$. First there is a possible singularity arising from the coincidence limit $x-y\to 0$. This singularity corresponds to an eventual divergence of the integral in (3.6) at its lower boundary and depends of the dimension n. It is the singularity of the free propagator. The second divergence is due to the fact that the sum over the reciprocal energy values diverges for the harmonic oscillator. This divergence is already present for $n=1$, where $G(x,x)$ is regular.

4. THE CONNECTION WITH PATH INTEGRALS

The formalism developed in Sec. 1 gives certain Green's functions as a series of derivatives of the "potential" $gV(x)$ taken along the straight path connecting the two points x and y. Therefore, it is suggestive to try to derive this expansion from a path integral representation[6] of the Green's functions at least in those cases where such a representation exists. A derivation of this kind is possible and described in this section. To avoid technical complications again, a restriction to the Euclidean case is performed.

The Feynman kernel is given by the path integral[6]

$$K(x,y\,|\,T)=\frac{1}{N}\int D\hat{r}\exp(iS[\hat{r}]) \tag{4.1}$$

$$S[\hat{r}]=\frac{M}{2}\int_0^T dt[\dot{\hat{r}}(t)]^2-g\int_0^T dtV(\hat{r}(t)),\quad T>0. \tag{4.1a}$$

Here the integration is over all paths $t\to\hat{r}(t)$ with $\hat{r}(0)=y$ and $\hat{r}(1)=x$. The metric is $g^{\mu\nu}=a_\mu\delta^{\mu\nu}$ with $a_\mu>0$. This action leads to a Hamiltonian $H=-\Delta+gV(x)$ for $M=\frac{1}{2}$.

The Feynman kernel K is connected with the resolvent and therefore with $G(x,y)$ by the formula

$$R(x,y\,|\,E)=\int_0^\infty dt\exp[i(E+i0)t]K(x,y\,|\,t). \tag{4.2}$$

To proceed, the scaled variable $\tau=(1/T)t$ is introduced into (4.1) and the straight line path $q_0(\tau)=\tau x+(1-\tau)y$ is separated from $r(\tau)\equiv\hat{r}(t)$. That is $r(\tau)=q_0(\tau)+q(\tau)$. The action becomes

$$S[r]=\frac{M}{2T}\int_0^1 d\tau\dot{q}_0^2(\tau)+\frac{M}{2T}\int_0^1 d\tau\dot{q}^2(\tau)-gT\int_0^1 d\tau\times V(q_0(\tau)+q(\tau)).$$

The important point is that the mixed term $\int_0^1 d\tau\,\dot{q}_0(\tau)\dot{q}(\tau)$ vanishes because firstly $\dot{q}_0(\tau)$ is independent of τ and secondly $q(0)=q(1)=0$ due to the boundary conditions of $r(\tau)$. The usual procedure achieving the vanishing of this term is an expansion around the "classical path."[7]

After expanding V around q_0 and replacing the corresponding exponential by its series representation, one finds

$$K(x,y\,|\,T)=\exp\left(\frac{iM}{2T}(x-y)^2\right)\sum_{\lambda_0,\lambda_1,\ldots}\frac{(-igT)^{\sum\lambda_k}}{\prod_k\lambda_k!(k!)^{\lambda_k}}\frac{1}{N}\times\int Dq\exp\left(\frac{i}{2}\int_0^1 d\tau q_\mu(\tau)S^{\mu\nu}q_\nu(\tau)\right)\times\prod_k\left\{\int_0^1 d\tau[q(\tau)\cdot\partial]^k V(q_0(\tau))\right\}^{\lambda_k}$$

with $S^{\mu\nu}=-(M/T)g^{\mu\nu}(d^2/d\tau^2)$.

All the functional integrals are of the Gaussian type and may be evaluated. The normalization is fixed by the free case ($g=0$). Details of the combinatorics are given in Ref. 7. Each contraction of two factors $q(\tau_1)$ and $q(\tau_2)$ gives rise to a propagator kernel

$$i\bar{g}_{\mu\nu}(\tau_1,\tau_2)=i(T/M\alpha_\mu)\delta_{\mu\nu}g(\tau_1,\tau_2),$$

where $g(\tau_1,\tau_2)$ is defined by (1.7). This explains the appearance of the kernel $g(\alpha,\beta)$ in the definition of ∂_ν^2.

As usual each term of the resulting series can be described by a graph. Collecting all terms which correspond to the same graph, one finds after some combinatorial considerations indeed a result which is almost identical to the results of Sec. 2, formulas (2) and (8). (One has to replace the Laplace representation of Sec. 2 by a representation with an additional factor i in the exponential and take the special case $z=1$.)

5. LIGHTCONE OR HADAMARD EXPANSION OF GREEN FUNCTIONS

In this section the lightcone or Hadamard expansion of Green's functions is derived by application of the formalism of Sec. 1. The achievement compared with the usual treatments of Hadamards' expansion[8] is the explicit incorporation of the boundary conditions into this expansion, that is, the determination of the corresponding solution of the homogeneous equation which otherwise can be rather difficult.

To avoid notational and technical complications, the results are presented for the Minkowski case [$\partial^2=\Box$, $n\geq 2$, $g^{\mu\nu}=\mathrm{diag}(1,-1,\ldots,-1)$, and $m^2>0$].

The corresponding generalized free propagator introduced in Sec. 1 is given by

$$G_z^{(0)}(x,y\,|\,m^2)=(m^2)^{-z}\int\frac{d^np}{(2\pi)^n}\exp[-ip(x-y)]\times\left[\frac{m^2}{-p^2+m^2-i0}\right]^z=i\frac{(m^2)^{n/2-z}}{(2\pi)^{n/2}}\frac{2^{1-z}}{\Gamma(z)}\frac{K_{n/2-z}(m[-(x-y)^2+i0])^{1/2}}{(m[-(x-y)^2+i0]^{1/2})^{n/2-z}}. \tag{5.1}$$

K_ν is the Bessel function of the MacDonald type. The structure of $G_z^{(0)}$ for $(x-y)^2\to 0$ is exposed best by a

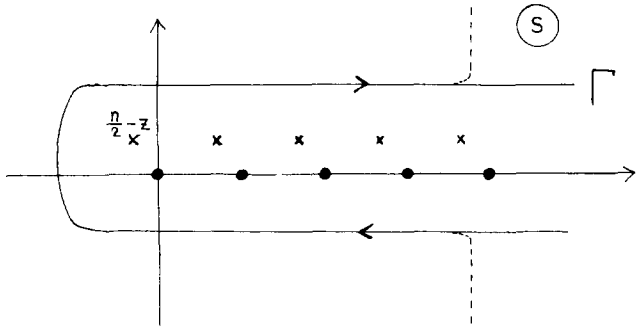


FIG. 1.

suited integral representation allowing application of the substitution rule. One finds

$$G_z^{(0)}(x,y\,|\,m^2) = i\frac{(m^2)^{n/2-z}}{(4\pi)^{n/2}}\int_\Gamma \frac{ds}{2\pi i} \times \frac{\Gamma(n/2-z-s)\Gamma(-s)}{\Gamma(z)}\left[\frac{m^2}{4}[-(x-y)^2+i0]\right]^{s-n/2+z}. \tag{5.2}$$

The integrand contains two series of poles running parallel to the positive real axis. The path of integration Γ has to enclose these two series (see Fig. 1). Γ may be opened provided one has asymptotically $2\,\mathrm{Res} > n/2 - \mathrm{Re}z$. This restriction arises only for timelike distances $(x-y)^2 > 0$. After opening of the contour one can pick up the residuum contributions and estimate the remainder integral.

These contributions have the following structure:

$$\begin{aligned} &\Gamma(-s)\text{ series: const}\cdot(m^2)^N[-(x-y)^2+i0]^{N-n/2+z} \\ &\qquad N=0,1,2,\ldots, \\ &\Gamma(n/2-z-s)\text{ series:} \\ &\qquad \text{const}\cdot(m^2)^{N+n/2-z}[-(x-y)^2+i0]^N, \\ &\qquad N=0,1,2,\cdots. \end{aligned} \tag{5.3}$$

The contributions of the first series are proportional to an integer power of m^2. Therefore, their modification due to the interaction immediately can be written down applying the rules of Sec. 1. Especially the term $N=0$ is not modified at all. Generally one gets a factor $(m^2+\Box_v+g\int_0^1 d\alpha V_\alpha)^N$. For $\mathrm{Re}(n/2-z)>0$ the $N=0$ term gives the leading lightcone contribution, which therefore remains that of the free case.

The second series contains contributions with powers of m^2 which are generally not integers and therefore require more qualified methods of evaluation. Nevertheless, if the corresponding series converge at all in some sense these expressions are regular for $x-y\to 0$ as the definition of ∂_v^2 shows. In certain cases the two series become degenerate. This happens for the physical Feynman propagator ($n=4$, $z=1$).

The resulting double poles give rise to terms

$$\text{const}(m^2)^N\ln\{(m^2/4)[-(x-y)^2+i0]\}\cdot[-(x-y)^2+i0]^N.$$

The modification due to the interaction therefore again is of the nonpolynomial type containing a factor $\ln(m^2+\Box_v+g\int_0^1 d\alpha V_\alpha)$.

The well-known differences[8] between even and odd dimensions for Hadamard's expansion (logarithmic terms—no such terms) are reflected in this behavior of the two pole series.

6. THE COINCIDENCE LIMIT FOR GREEN'S FUNCTIONS

One can derive another integral representation for the Green's functions which always leads to simple poles and for which all contributions are proportional to *integer* powers of m^2. The logarithmic terms which arise in the degenerate case are already contained in the kernel function and do not arise via double poles.

This representation is especially well suited if one wants to study the coincidence limit of the Green's functions which, for instance, is important for applications of the formalism to field theoretic perturbation theory.[1,3]

The idea is to make use of the fact that the partition

$$\partial^2+m^2+gV(x) = \partial_0^2+m^2+\partial_v^2+g\int_0^1 d\alpha V_\alpha$$

of the differential operator into a free and an interacting part discussed in Sec. 2 does not prescribe to which part the m^2 term belongs as it commutes with ∂^2 and with $V(x)$. Therefore, one has the freedom to introduce a parameter $\mu^2 \geq 0$ by

$$\partial^2+m^2+gV(x) = (\partial_0^2+\mu^2)+(m^2-\mu^2+\partial_v^2+g\int_0^1 d\alpha V_\alpha), \tag{6.1}$$

which fixes the splitting completely.

Again the discussion is restricted to the Minkowski case. The appertaining kernel function is defined as follows:

$$\begin{aligned} K_z^{(n)}(x-y\,|\,\mu^2) &= \int \frac{d^np}{(2\pi\mu^2)^{n/2}}\exp[-ip(x-y)] \\ &\quad\times\left[\frac{\mu^2}{-p^2+\mu^2-i0}\right]^z \\ &= i\frac{2^{1-z}}{\Gamma(z)}\,\frac{K_{n/2-z}(\mu[-(x-y)^2+i0]^{1/2})}{\{\mu[-(x-y)^2+i0]^{1/2}\}^{n/2-z}}. \end{aligned} \tag{6.2}$$

$K_z^{(n)}$ is proportional to the corresponding generalized free propagator $G_z^{(0)}$ of Sec. 1. Furthermore, $(\mu^2)^{n/2-z}\cdot K_z^{(n)}$ has a limit for $\mu^2\to 0$ leading to the representation of the preceding section. For $z=0,-1,-2,\cdots$ the factor $1/\Gamma(z)$ extracts the residuum from the remainder which becomes singular in this limit. This residuum is concentrated at the origin and can be found by direct evaluation as

$$K_{-N}^{(n)}(x-y\,|\,\mu^2) = (2\pi)^{n/2}(\mu^2)^{-N-n/2}(\Box+\mu^2)^N\delta(x-y). \tag{6.3}$$

For $\mathrm{Re}z > n/2$ the coincidence limit $x-y\to 0$ of $K_z^{(n)}$ exists and one finds

$$K_z^{(n)}(0\,|\,\mu^2) = i2^{-n/2}\frac{\Gamma(z-n/2)}{\Gamma(z)}. \tag{6.4}$$

Now the desired integral representation for the Green's functions is derived by formal application of the following inverse Mellin representation[3]:

$$\frac{1}{(x+c)^z} = \int_{\Gamma_0} \frac{ds}{2\pi i} \frac{\Gamma(z+s)\Gamma(-s)}{\Gamma(z)} c^s x^{-s-z} \quad (\mathrm{Re}z > 0). \tag{6.5}$$

The path of integration is $\int_{\Gamma_0} = \int_{c-i\infty}^{c+i\infty}$ with $-\mathrm{Re}z < c < 0$. This implies

$$G_z(x,y) = \left[(\Box + \mu^2 - i0) + \left(m^2 - \mu^2 + \Box_v + g\int_0^1 d\alpha V_\alpha\right)\right]^{-z} \delta(x,y)$$
$$= \int_{\Gamma_0} \frac{ds}{2\pi i} \frac{\Gamma(z+s)\Gamma(-s)}{\Gamma(z)} \left[m^2 - \mu^2 + \Box_v + g\int_0^1 d\alpha V_\alpha\right]^s$$
$$\times \frac{(\mu^2)^{n/2-s-z}}{(2\pi)^{n/2}} K_{z+s}^{(n)}(x-y\,|\,\mu^2). \tag{6.6}$$

The validity of this representation can also be verified by methods analogous to those used for the proof of Proposition 1.[3]

By shifting the path of integration to the right, one gets a series representation of $G_z(x,y)$ with the announced properties. The question of convergence of these integrals and series is somewhat delicate and will not be touched.

The representation proves very useful for a configuration space formulation of field theoretic perturbation theory,[1] where one needs coincidence limits for Green's functions.

For example, if $\mathrm{Re}z > n/2$ application of (6.4) leads to

$$G_z(x,x) = \frac{i}{(4\pi)^{n/2}} (\mu^2)^{n/2-z} \int_{\Gamma_0} \frac{ds}{2\pi i} \frac{\Gamma(z-n/2+s)\Gamma(-s)}{\Gamma(z)}$$
$$\times \frac{1}{\mu^2}\left[m^2 - \mu^2 + \Box_v + g\int_0^1 d\alpha V_\alpha\right]^s_{|x=y}$$
$$= \frac{i}{(4\pi)^{n/2}} \frac{\Gamma(z-n/2)}{\Gamma(z)} \left[m^2 - \mu^2 + \Box_v + g\int_0^1 d\alpha V_\alpha\right]^{z-n/2}_{|x=y}. \tag{6.7}$$

If $\mathrm{Re}z < n/2$, one has to shift the contour of integration sufficiently far the right. The emerging residuum contributions are singular parts which have to be subtracted by counterterms in field theoretic applications.

7. THE EIKONAL APPROXIMATION

In this section the connection of the effective mass formalism and the eikonal approximation method is discussed. As an example, a Schrödinger theory with Hamiltonian $H = -\Delta + m^2 + gV(x)$ is noted. The corresponding T matrix for outgoing radial waves obeys

$$\langle \mathbf{k}\,|\,T\,|\,\mathbf{p}\rangle = \langle \mathbf{k}\,|\,V\,|\,\mathbf{p}\rangle - \langle \mathbf{k}\,|\,VG(E)V\,|\,\mathbf{p}\rangle \tag{7.1}$$

with $E = \mathbf{p}^2 = \mathbf{k}^2$.

Furthermore,

$$\langle \mathbf{k}\,|\,VGV\,|\,\mathbf{p}\rangle = (2\pi)^{-3}\int d^3x \int d^3y \exp(-i\mathbf{p}\mathbf{y} + i\mathbf{k}\mathbf{x})$$
$$\times V(\mathbf{x})V(\mathbf{y})G(\mathbf{x},\mathbf{y}\,|\,E), \tag{7.2}$$

where $G(\mathbf{x},\mathbf{y}\,|\,E) = \langle \mathbf{x}\,|\,(H - Ei0)^{-1}\,|\,\mathbf{y}\rangle$ is one of the Green's functions the effective mass formalism deals with.

Application of the substitution rule implies

$$G(\mathbf{x},\mathbf{y}\,|\,E) = (H_0 + \tilde{M}^2 - i0)^{-1}\delta(\mathbf{x}-\mathbf{y})$$
$$= \int \frac{d^3q}{(2\pi)^3} \exp[-i\mathbf{q}(\mathbf{x}-\mathbf{y})] \frac{1}{-\mathbf{q}^2 + \tilde{M}^2 - i0}. \tag{7.3}$$

with $H_0 = -\Delta_{\mathbf{x}}$ and the effective mass $\tilde{M}^2 = m^2 + H_{0v} - E + g\int_0^1 d\alpha V_\alpha$. Eikonal methods now typically introduce a modified propagator[9]

$$G_1(\mathbf{x},\mathbf{y}\,|\,E) = \exp[-i\mathbf{l}(\mathbf{x}-\mathbf{y})]G(\mathbf{x},\mathbf{y}\,|\,E) \quad \text{with } \mathbf{l}^2 = E.$$

For this propagator one finds

$$G_1(\mathbf{x},\mathbf{y}\,|\,E) = \int \frac{d^3q}{(2\pi)^3} \frac{\exp[-iq(x-y)]}{-\mathbf{q}^2 + 2\mathbf{q}\mathbf{l} + M^2 - i0} \tag{7.4}$$

where $M^2 = \tilde{M}^2 - \mathbf{l}^2 = m^2 + H_{0v} + g\int_0^1 d\alpha V_\alpha$. This can be rewritten as follows:

$$G_1(\mathbf{x},\mathbf{y}\,|\,E) = i\int_0^\infty ds \exp(-isM^2)$$
$$\exp[-is(\Delta_{\mathbf{x}} - i0)]\delta(\mathbf{x}-\mathbf{y}+2s\mathbf{l}).$$

The interaction now is concentrated in the exponential $\exp(-isM^2) = \exp[-is(m^2 + H_{0v} + g\int_0^1 d\alpha V_\alpha)]$, which is determined by (2.9). The leading term corresponding to the graph · is simply given by $\exp(-isg\int_0^1 d\alpha V_\alpha)$. The eikonal approximation amounts to replacing first $\exp(-siM^2)$ by this leading term and second $\exp(-is\Delta_x)$ by one, corresponding to the usual linearization of denominators. One finds

$$G_1^{(\mathrm{Eik})}(\mathbf{x},\mathbf{y}\,|\,E) = i\int_0^\infty ds\, \delta(\mathbf{x}-\mathbf{y}+2s\mathbf{l})$$
$$\times \exp[-ig\int_0^s d\tau V(\mathbf{x}+2\tau\mathbf{l})]. \tag{7.5}$$

Introducing this approximation into (7.1) and (7.2) and choosing $\mathbf{l} = \mathbf{k}$ or $\mathbf{l} = \mathbf{p}$, one immediately gets the usual eikonal approximations for $\langle \mathbf{k}\,|\,T\,|\,\mathbf{p}\rangle$.[9] The significance of the above approximations are not pursued here any further. Obviously the method is not restricted to the Schrödinger case but can also be applied to propagators arising, for example, in field theories.[3]

ACKNOWLEDGMENTS

The author would like to thank M. Scheunert and H. Römer for useful discussions on the subject of this paper. The introduced formalism was discovered in cooperation with M. Scheunert by explicitly guessing a formal solution of Hadamard's expansion.

[1] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[2] I. M. Gel'fand and G. E. Shilov, *Generalized Functions I* (Academic, New York, 1964).
[3] F. Krause, Habilitationsschrift, Bonn 1975 (unpublished).
[4] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[5] G. Hochschild, *The Structure of Lie Groups* (Holden-Day, San Francisco, 1965); R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).
[6] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[7] J. Honerkamp, F. Krause, and M. Scheunert, Nucl. Phys. B **69**, 618 (1974).
[8] P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964).
[9] R. L. Sugar and R. Blancenbecler, Phys. Rev. **183**, 1387 (1969).

Massless quantum sine-Gordon equation in two space-time dimensions: Correlation inequalities and infinite volume limit

Yong Moon Park

Department of Mathematics, Yonsei University, Seoul, Korea
(Received 2 May 1977)

We prove new correlation inequalities for the massive and massless quantum sine-Gordon equations. These results are then used to construct infinite volume limit theory for the massless $(S\text{–}G)_2$ model that satisfies the Osterwalder–Schrader axioms. As consequences, infinite volume limit theories for the classical, neutral, statistical mechanical systems with two-body Coulomb potentials and for the massive Thirring model exist.

1. INTRODUCTION AND MAIL RESULTS

We study correlation inequalities and infinite volume limits of the Euclidean Green's functions for the massless (and massive) quantum sine-Gordon equations in two space–time dimensions. The $(S\text{–}G)_2$ interacting actions are given by

$$V(\Lambda)=\int_\Lambda d^2x : \cos\epsilon(\phi+\theta) :_1(x). \tag{1.1}$$

Here ϕ is a real, scalar Euclidean field on R^2 with bare mass $m_0 \geq 0$, Λ denotes a space cutoff, the double colon $: :_1$ denotes Wick ordering with respect to bare mass $m_0=1$, $\epsilon \in (-2\sqrt{\pi}, 2\sqrt{\pi})$, and $\theta \in [0, 2\pi)$.

The massive $(S\text{–}G)_2$ model has been constructed in Refs. 1–4. It has been proven that for sufficiently large m_0, Feynman perturbation expansion of the Euclidean Green's functions converges.[3] For arbitrary $m_0 > 0$, infinite volume theory has been constructed in Ref. 4 by means of GKS inequalities,[5] weak boundary conditions,[6] and results in Refs. 2 and 3. In Ref. 4 existence of infinite volume limit theories for the massless $(S\text{–}G)_2$ model has been shown by a standard compactness argument (Cantor's diagonal process) together with uniform bounds of correlation functions with respect to m_0. Hence, the theories may depend on how one chooses a sequence of bare mass m_0 which tends to zero. Our main purpose is to show that the space-cutoff Euclidean Green's functions for $m_0=0$ converges as the space cutoff is removed. To do this, we will derive correlation inequalities for the $(S\text{–}G)_2$ model and we will use a standard monotonicity argument with respect to Λ. Our results then imply existence of the infinite volume limit theories for the classical statistical mechanical Coulomb system and the massive Thirring model.

A. Definitions and notation

Let $C_{m_0}(x\text{-}y)$ denote the kernel of the operator $(-\Delta + m_0)^{-1}$ where Δ is the two-dimensional Laplacian. The Gaussian measure on $\mathcal{S}'(R^2)$ with mean zero and covariance $C_{m_0}(x\text{-}y)$ is denoted by $d\mu_0^{(m_0)}(\phi)$, and $\langle\cdot\rangle^{(m_0)}$ denotes expectations with respect to $d\mu_0^{(m_0)}$. Let $: :_1$ denote Wick ordering with respect to $d\mu_0^{(1)}$. We define:

$$: \phi^2 :_1(f)=\int d^2x : \phi(x)^2 :_1 f(x),$$

$$c(f,\theta)=\int d^2x : \cos\epsilon(\phi(x)+\theta) :_1 f(x),$$

$$s(f,\theta)=\int d^2x : \sin\epsilon(\phi(x)+\theta) :_1 f(x), \tag{1.2}$$

$$-\nabla\phi(x)=\int d^2x\phi(x)\nabla f(x),$$

$$\chi(f,\epsilon)=\int d^2x : \exp[i\epsilon\phi(x)] :_1 f(x),$$

for $f \in \mathcal{S}(R^2)$. The interacting measure is defined by

$$d\mu(\Lambda, m_0)=Z(\Lambda, m_0)^{-1}\exp[\lambda V(\Lambda)]d\mu_0^{(m_0)}(\phi), \tag{1.3}$$

where $V(\Lambda)$ is the interacting action defined in (1.1) and

$$Z(\Lambda, m_0)=\langle\exp[\lambda V(\Lambda)]\rangle^{(m_0)} \tag{1.4}$$

is the partition function. For convenience we write

$$\langle\cdot\rangle^{(\Lambda, m_0)}=\int\cdot\, d\mu^{(\Lambda, m_0)}. \tag{1.5}$$

Throughout this paper we do Wick ordering with respect to bare mass $m_0=1$ $(d\mu_0^{(1)})$. *From now on we suppress m_0 in the notation.*

B. Main results

We summarize our results. In Sec. 2 we will prove

Theorem A. [correlation inequalities for the $(S\text{–}G)_2$ model]: We assume that $\lambda \geq 0$ and $\theta=0$ in (1.1). We also assume that $f \in \mathcal{S}_{\text{real}}(R^2)$ for $m_0>0$ and $f=-\mathbf{a}\cdot\nabla g$, $g \in \mathcal{S}_{\text{real}}(R^2)$, for $m_0=0$ where $\mathbf{a}$ is a constant vector. Then the following inequalities hold for the $(S\text{–}G)_2$ model:

(a) $$\left\langle\left(\prod_{j=1}^n : \exp[i\epsilon_j\phi(x_j)] :_1\right)\exp[i\phi(f)]\right\rangle^{(\Lambda)} \geq 0,$$

(b) $$\left\langle\left(\prod_{j=1}^n : \exp[i\epsilon_j\phi(x_j)] :_1\right)\exp[i\phi(f)]c(h,0)\right\rangle^{(\Lambda)} - \left\langle\left(\prod_{j=1}^n : \exp[i\epsilon_j\phi(x_j)] :\right)\exp[i\phi(f)]\right\rangle^{(\Lambda)}\langle c(h,0)\rangle^{(\Lambda)} \geq 0,$$

(c) for $m_0>0$,

$$\left\langle\left(\prod_{j=1}^n : \exp[i\epsilon_j\phi(x_j)] :_1\right)\exp[i\phi(f)] : \phi^2 :_1(h)\right\rangle^{(\Lambda)} - \left\langle\left(\prod_{j=1}^n : \exp[i\epsilon_j\phi(x_j)] :_1\right)\exp[i\phi(f)]\right\rangle^{(\Lambda)}\langle : \phi^2 :_1(h)\rangle^{(\Lambda)} \leq 0,$$

where ϵ_j is either ϵ or else $-\epsilon$, and $0 \leq h \in \mathcal{S}_{\text{real}}(R^2)$ or $h=\chi_\Lambda$ (the characteristic function of Λ).

Remark 1: (a) Each term in the above inequalities is well defined by the assumption on f and h.[2,4,7] (b) The inequalities can be easily extended to periodic, Dirichlet and Neumann boundary conditions,[2,5,8] and also to the lattice $(S-G)_2$ model by following a method similar to that used in Sec. 2. (c) The inequalities for $m_0=0$ hold for $\lambda \in R$ and $\theta \in [0, 2\pi)$ by the invariance under field translation $\phi \to \phi + \alpha$ (See, e.g., Sec. 3 for the detailed discussions).

In Sec. 3 we will use the above inequalities [Theorem A(b)] and a standard monotonicity argument to obtain the following theorem.

Theorem B. [infinite volume limit theory for the massless $(S-G)_2$ model]: For $m_0=0$ and for f_i, g_j, $h_k \in \mathcal{S}(R^2)$, the limit

$$\left\langle \prod_{i=1}^{n} c(f_i,\theta) \prod_{j=1}^{m} s(g_j,\theta) \prod_{k=1}^{l} \mathbf{a}_k \cdot \nabla\phi(h_k) \right\rangle = \lim_{\Lambda \uparrow R^2} \left\langle \prod_{i=1}^{n} c(f_i,\theta) \prod_{j=1}^{m} s(g_j,\theta) \prod_{k=1}^{l} \mathbf{a}_k \cdot \nabla\phi(h_k) \right\rangle^{(\Lambda)}$$

exists, where the $\mathbf{a}_k$'s are constant vectors. These limits are Euclidean Green's functions that satisfy all the Osterwalder–Schrader axioms with the possible exception of the clustering properties.

Remark 2: (a) One may be able to construct an infinite volume limit theory for $m_0>0$ by using Theorem A, a method similar to that for $m_0=0$ and an introduction of a linear external field. Because of notational complications, we do not produce a construction for $m_0>0$. Notice that the construction for $\lambda>0$ and $\theta=0$ is very simple by Theorem A(b). (b) For $\lambda \geq 0$ and $\theta=0$ the infinite volume limit for $m_0=0$ with weak boundary conditions[4] is unique by Theorem A(c).

Combining Theorem B and the isomorphism between the massless sine-Gordon equation, the classical Coulomb system, and the massive Thirring model (see Refs. 1–3 for the details) we immediately obtain the following.

Corollary C: (a) Let β, $\pm e$, and z be the inverse temperature, the charge, and the fugacity for the two-component classical Coulomb system. Then, for $\beta e^2 < 4\pi$, the thermodynamic limts of the correlation functions for the two-component, neutral, classical statistical mechanical systems with two-body Coulomb potentials exist for arbitrary fugacity z. These are Euclidean invariant. (b) The infinite volume limit theory for the massive Thirring model exists and satisfies all the (O–S) axioms possible with the exception of the clustering.

2. CORRELATION INEQUALITIES: PROOF OF THEOREM A

In this section we prove Theorem A. In proving the theorem one may assume $m_0>0$, because each term for $m_0=0$ in the inequalities can be approximated by that of $m_0>0$. The above fact follows from Lemma IV.4 and its proof in Ref. 1, and Sec. 4 of Ref. 2. Here one may have used the "neutrality property[1,2]" (see also Sec. 3 for the discussions).

We first prove Theorem A(a). We note that

$$\exp[i\phi(f)] = \exp[-(f, Cf)/2] : \exp[i\phi(f)] :_C . \tag{2.1}$$

By re-Wick ordering and by a direct computation (see also Lemma 2.2 of Ref. 2), one obtains

$$\left\langle \prod_{j=1}^{n} : \exp[i\epsilon_j \phi(x_j)] :_1 \right\rangle_0 = K \exp\left[- \sum_{1 \leq i < j \leq n} \epsilon_i \epsilon_j C(x-y) \right] \geq 0,$$

where K is a constant produced by re-Wick ordering. On the other hand, we have

$$\left\langle \left(\prod_{j=1}^{n} : \exp[i\epsilon_j \phi(x_j)] :_1 \right) \exp[i\phi(f)] \right\rangle^{(\Lambda)} = Z(\Lambda)^{-1} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \left\langle \left(\prod_{j=1}^{n} : \exp[i\epsilon_j \phi(x_j)] :_1 \right) \exp[i\phi(f)] V(\Lambda)^k \right\rangle_0 . \tag{2.2}$$

The above series is summable absolutely,[2] and so one only needs to show that

$$\left\langle \prod_{j=1}^{n} : \exp[i\epsilon_j \phi(x_j)] :_1 \quad \exp[i\phi(f)] V(\Lambda)^k \right\rangle_0 \geq 0 \tag{2.3}$$

for $f \in \mathcal{S}_{\text{real}}(R^2)$. We substitute $V(\Lambda) = \frac{1}{2}[\chi(\Lambda,\epsilon) + \chi(\Lambda, -\epsilon)]$ in (2.3) and use (2.1) and (2.2) to obtain (2.3). Here we have used the fact that $\int C(x-y) f(y) d^2y$ is real. This completes the proof of Theorem A(a).

To prove Theorems A(b) and A(c) we use a method of duplicated variables introduced by Percuse.[9] (See also Refs. 5 and 10 and the references therein.) Let $\phi_1(x)$ and $\phi_2(x)$ be duplicated Euclidean free fields; this means that $\langle \phi_i(x)\phi_j(y)\rangle_0 = \delta_{ij} C(x,y)$. We define

$$\phi'(x) = [\phi_1(x) + \phi_2(x)]/\sqrt{2}, \qquad \phi''(x) = [\phi_1(x) - \phi_2(x)]/\sqrt{2}. \tag{2.4}$$

Since joint covariance of ϕ_1 and ϕ_2 is the same as that of ϕ' and ϕ'', we have that

$$\int F(\phi_1,\phi_2) d\mu_0(\phi_1) d\mu_0(\phi_2) = \int F((\phi'+\phi'')/\sqrt{2}, (\phi'-\phi'')/\sqrt{2}) d\mu_0(\phi') d\mu_0(\phi''). \tag{2.5}$$

We will use the following identities:

$$\cos(x-y) + \cos(x+y) = 2\cos(x)\cos(y),$$

$$\cos(x-y) - \cos(x+y) = 2\sin(x)\sin(y).$$

We now prove Theorem A(b). We note that

$$\left\langle \left(\prod_{j=1}^{n} : \exp[i\epsilon_j \phi(x_j)] :_1 \right) \exp[i\phi(f)] c(h,0) \right\rangle^{(\Lambda)} - \left\langle \left(\prod_{j=1}^{n} : \exp[i\epsilon_j \phi(x_j)] :_1 \right) \exp[i\phi(f)] \right\rangle^{(\Lambda)} \langle c(h,0) \rangle^{(\Lambda)}$$
$$= Z(\Lambda)^{-2} \int \left(\prod_{j=1}^{n} : \exp[i\epsilon_j \phi_2(x_j)] :_1 \right) \times \exp[i\phi_2(f)] (c(\phi_2,h,0) - c(\phi_1,h,0)) \times \exp[\lambda V(\phi_1,\Lambda) + V(\phi_2,\Lambda)] d\mu_0(\phi_1) d\mu_0(\phi_2)$$

$$= Z(\Lambda)^{-2}\int \left(\prod_{j=1}^{n} :\exp[(i/\sqrt{2})\epsilon_j \phi'(x_j)]:_1\right)$$

$$\times \exp[(i/\sqrt{2})\phi'(f)]$$

$$\times \left(\prod_{j=1}^{n} :\exp[(-i/\sqrt{2})\epsilon_j \phi''(x_j)]:_1\right)$$

$$\times \exp[(-i/\sqrt{2})\phi''(f)] s(\phi', \phi'', h)$$

$$\times \exp[\lambda c(\phi', \phi'', \Lambda)] d\mu_0(\phi')\, d\mu_0(\phi'') \qquad (2.6)$$

by (2.5), where

$$s(\phi', \phi'', h) = 2\int d^2x :\sin(\epsilon\phi'(x)/\sqrt{2}):_1 \sin(\epsilon\phi''(x)/\sqrt{2}):_1 h(x),$$

$$c(\phi', \phi'', \Lambda) = 2\int_\Lambda d^2x :\cos(\epsilon\phi'(x)/\sqrt{2}):_1 :\cos(\epsilon\phi''(x)/\sqrt{2}):_1.$$

We expand $\exp[\lambda c(\phi', \phi'', \Lambda)]$ by Tayler's series (the series converges absolutely[1]). The problem then reduces to showing that each term in the series is nonnegative. But

$$\int \left(\prod_{j=1}^{n} :\exp[(i/\sqrt{2})\epsilon_j \phi'(x_j)]:_1\right) \exp[(i/\sqrt{2})\phi'(f)]$$

$$\times \left(\prod_{j=1}^{n} :\exp[(-i/\sqrt{2})\epsilon_j \phi''(x_j)]:_1\right) \exp[(-i/\sqrt{2})\phi''(f)]$$

$$\times\ s(\phi', \phi'', h) c(\phi', \phi'', \Lambda)^l d\mu_0(\phi')\, d\mu_0(\phi'') \qquad (2.7)$$

$$= 2^{l+1} \int d^2x' h(x') \int_{\Lambda^l} \prod_{k=1}^{l} d^2x_k \left\{ \int \left(\prod_{j=1}^{n} :\right.\right.$$

$$\exp[(i/\sqrt{2})\epsilon_j \phi(x_j)]:_1\Big)$$

$$\times \exp[(i/\sqrt{2})\phi(f)] :\sin(\epsilon\phi(x')/\sqrt{2}):_1$$

$$\left.\left. \times \left(\prod_{k=1}^{l} :\cos(\epsilon\phi(x_k)/\sqrt{2}):_1\right) d\mu(\phi)\right\}^2 \geq 0.$$

This proves Theorem A(b) completely.

We next prove Theorem A(c). We note that Theorem A(b) and Theorem A(c) are related by the following fact:

$$\lambda^{-2}[1 - C(\lambda\phi, h)] \to \tfrac{1}{2} :\phi^2:_1(h) \quad \text{as } \lambda \to 0.$$

Therefore, Theorem A(b) implies Theorem A(c). This proves Theorem A completely.

3. INFINITE VOLUME LIMIT: PROOF OF THEOREM B

Rather than proving the theorem for general $\lambda \in R$ and $\theta \in [0, 2\pi)$, we first consider the case for $\lambda \geq 0$ and $\theta = 0$. The theorem for general cases will follow from a translation of fields by $\phi \to \phi + \alpha$ for a suitable constant α.

We consider the expectation

$$\left\langle \left(\prod_{i=1}^{n} \chi(f_i, \epsilon_i)\right) \exp[\zeta\, \mathbf{a} \cdot \nabla\phi(g)] \right\rangle^{(\Lambda)} \qquad (3.1)$$

where $f_i, g \in \mathcal{S}_{\text{real}}(R^2)$, a is a constant vector and $\zeta \in C$. For such f_i and g we prove the existence of the limit as $\Lambda \uparrow R^2$ by inclusion. The existence of these limits is proven by monotonicity property in Λ for pure imaginary ζ, uniform bounds of (3.1) in Λ, and a standard application of Vitali's theorem.

We give the proof more detaily. We are interested in the objects

$$c(f_1, \theta), s(f_2, \theta), \text{ and } \mathbf{a} \cdot \nabla\phi(f_3).$$

In [1.2.4] it has been shown that there exists a Schwartz space norm $\|\cdot\|$ such that for $f_i \in \mathcal{S}_{\text{real}}(R^2)$

$$\langle \exp[c(f_1, \theta) + s(f_2, \theta) + \mathbf{a} \cdot \nabla\phi(f_3)\rangle^{(\Lambda)}$$

$$\leq \exp K(\|f_1\| + \|f_2\| + a^2\|f_3\|_2^2) \qquad (3.2)$$

uniformly in Λ and $m_0 \in [0, m]$.

Remark: The uniform bounds in (3.2) has been proven for periodic boundary conditions.[1,2,4,7,8] A similar version of uniform bounds for free boundary conditions may follow from a method similar to that used in Refs. 7, 11 and the results in Refs. 1, 2.

We first consider the case for $\lambda \geq 0$ and $\theta = 0$. From Theorem A(b) it follows that for $f_i,\ g \in \mathcal{S}_{\text{real}}(R^2)$ the limit

$$\left\langle \prod_{i=1}^{n} \chi(f_i\epsilon) \exp[\not{g}\, \mathbf{a} \cdot \nabla\phi(g)] \right\rangle$$

$$= \lim_{\Lambda \uparrow R^2} \left\langle \prod_{i=1}^{n} \chi(f_i, \epsilon) \exp[\zeta\, \mathbf{a} \cdot \nabla\phi(g)] \right\rangle^{(\Lambda)} \qquad (3.3)$$

exists for pure imaginary ζ. We combine the uniform bounds in (3.2) [these imply uniform bounds of the expressions in (3.1)] and a standard application of Vitali's theorem (see, e.g., Ref. 12) to conclude that (3.3) holds for any $\zeta \in C$. Since

$$c(f, \epsilon, 0) = \tfrac{1}{2}[\chi(f, \epsilon) + \chi(f, -\epsilon)],$$

$$s(f, \epsilon, 0) = [\chi(f, \epsilon) - \chi(f, -\epsilon)]/2i, \qquad (3.4)$$

we obtain Euclidean Green's functions by differentiating (3.3). Euclidean invariance and (O—S) positivity follow from the monotonicity property (the inclusion).[5,11] This completes the proof of Theorem B for $\lambda \geq 0$ and $\theta = 0$.

We now consider the cases for $\lambda \in R$ and $\theta \in [0, 2\pi)$. In Lemma IV.4 of Ref. 1 and Sec. 4 of Ref. 2 it has been shown that

$$\left\langle \prod_{j=1}^{n} :\exp[i\epsilon_j(\phi(x_j) + \boldsymbol{\alpha})]:_1 \right\rangle^{(m_0=0)}$$

$$= \begin{cases} 0 & \text{if } \sum_{j=1}^{n} \epsilon_j \neq 0 \quad \text{(neutrality!)}, \\ \left\langle \prod_{j=1}^{n} :\exp[i\epsilon_j \phi(x_j)]:_1 \right\rangle^{(m_0=0)} & \text{if } \sum_{j=1}^{n} \epsilon_j = 0. \end{cases} \qquad (3.5)$$

We expand

$$\left\langle \left(\prod_{j=1}^{n} :\exp[i\epsilon_j(\phi(x_j) + \alpha)]:_1\right) \exp[i\mathbf{a} \cdot \nabla\phi(g)] \right.$$

$$\left. \times \exp[\lambda :\cos\epsilon(\phi + \alpha):(\Lambda)] \right\rangle_0$$

by Tayler's series in λ. We then use (3.5), a method similar to that in the proof of (3.5) in [1.2] and the fact that

$$\nabla\phi(g) = \nabla(\phi + \alpha)(g),$$

$$\langle \phi(f)\nabla\phi(g)\rangle_C = \langle \phi(f)\nabla\phi(g)\rangle_{C+b},$$

where b is a constant and $\langle\cdot\rangle_C$ is the Gaussian expectation with respect to convariance C, to conclude that

$$\left\langle\left(\prod_{j=1}^{n} :\exp[i\epsilon_j(\phi(x_j)+\alpha)]:_1\right)\exp[ia\cdot\nabla\phi(g)\right\rangle^{(\Lambda,\alpha)}$$

$$=\left\langle\left(\prod_{j=1}^{n} :\exp[i\epsilon_j\phi(x_j)]:_1\right)\exp[ia\cdot\nabla\phi(g)]\right\rangle^{(\Lambda,0)} \qquad (3.6)$$

(invariance under a field translation $\phi\to\phi+\alpha$), where $\langle\cdot\rangle^{(\Lambda,\alpha)}$ is the interacting expectation with $\theta=\alpha$.

To prove the theorem for $\lambda\in R$ and $\theta\in[0,2\pi)$, we translate fields by $\phi\to\phi+n\pi/\epsilon+\theta$, which gives a change of the interacting action from $V(\Lambda,0)$ to $(-1)^n V(\Lambda,\theta)$. The expectations in (3.3) are invariant (with possible exception of signs) under the translation by the neutrality in (3.5) and the invariance in (3.6). Hence, to prove the theorem for $\lambda\in R$ and $\theta\in[0,2\pi)$ it is sufficient to show the theorem for $\lambda\geq 0$ and $\theta=0$. This completes the proof of Theorem B.

[1]J. Fröhlich, "Quantum Sine—Gordon Equation and Quantum Solitons in Two Space—Time Dimensions," in *Renormalization Theory*, Int. School of Math. Phys. Ettrore Majorana, 1975 (Reidel, Dordrecht, 1976).

[2]J. Fröhlich, "Classical and Quantum Statistical Mechanics in One and Two Dimensions," Commun. Math. Phys. **47**, 233 (1976).

[3]J. Fröhlich and E. Seiler, "The Massive Thirring Schwinger Model (QED)$_2$: Convergence of Perturbation Theory and Particle Structure," Helv. Phys. Acta **49**, 889 (1976).

[4]J. Fröhlich and Y. M. Park, "Remarks on Exponential Interactions and the Quantum Sine—Gordon Equation in Two Space—Time Dimensions," to appear in Helv Phys. Acta.

[5]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, N. J., 1974).

[6]J. Glimm and A. Jaffe, "On the Approach to the Critical Points," Ann. Inst. H. Poincaré **22**, 109 (1975).

[7]J. Fröhlich, B. Simon, and T. Spencer, "Infrared Bounds, Phase Transitions and Continuous Symmetry Breaking," Commun. Math. Phys. **50**, 79 (1976).

[8]Y. M. Park, "Uniform bounds of the Schwinger functions in boson field models," J. Math. Phys. **17**, 1143 (1976).

[9]J. K. Percus, "Correlation Inequalities for Ising Spin Lattices," Commun. Math. Phys. **40**, 283 (1975).

[10]J. L. Lebowitz, "G. H. S. and Other Inequalities," Commun. Math. Phys. **35**, 87 (1974).

[11]J. Fröhlich and B. Simon, "Pure States for General $P(\phi)_2$ Theories: Construction Regularity and Variational Equality," submitted to Ann. Math.

[12]J. Fröhlich, "Schwinger Functions and Their Generating Functionals, II," Adv. Math. **23** (2), 119 (1977).

Klein–Gordon kinks with fourth order derivative self-coupling[a)]

Antonio F. Rañada

Departamento de Física Teórica, Universidad de Madrid, Madrid 3, Spain

Manuel F. Rañada

Departamento de Física Teórica, Universidad de Zaragoza, Zaragoza 9, Spain
(Received 7 March 1977)

A Klein–Gordon field with a derivative fourth order self-coupling is studied. It is shown that the kinks of the model form singlets, doublets, or triplets of electric charge, according to the values of the coupling constants.

I. INTRODUCTION

A great deal of attention is presently being paid[1] to the study of the kinks or solitons in nonlinear field theories. In this work we will consider the case of a scalar field with a derivative self-coupling.

Many authors have proposed the use of the localized solutions of some field theories, as a tool to study elementary particles. The history of this approach begins with Rosen[2] and goes on through Finkelstein[3] *et al.*,[4] who considered nonlinear spinor fields, and Wakano,[5] who tried unsuccessfully to construct a model of stable particle with interacting Dirac and Maxwell fields. In 1970 Soler[6] considered a nonlinear Dirac field and proposed a model of elementary fermions in which the physical frequency corresponds to a minimum of the energy $E=E(\omega)$. The same approach was followed subsequently to construct a model of charged particle[7] and of the nucleon.[8] Recently it was refined by the introduction of a scalar field to represent the meson cloud.[9] Other physicists considered nonlinear scalar fields with ϕ^4 or ϕ^6 self-coupling.[10–14] The stability properties of some of these models have also been investigated.[15–17] In this connection the paper by Anderson[17] shows that the effect of the addition of a ϕ^6 term increases the stability of the kink of the ϕ^4 model. This means that the stability properties may change with the addition of new nonlinear terms.

In this paper we call kink a solitary wave which is regular and has finite energy. We reserve the term soliton for a kink which preserves its individuality upon collision with other analogous waves.

We consider in this work a nonlinear model in which the scalar field has, in addition to the ϕ^4 nonlinearity, a derivative self-coupling. We will show that, contrary to the ϕ^4 model, it can describe charge multiplets of particles, because the energy can have several minima as a function of the frequency. In other words, the scalar and electromagnetic fields do not tend to decouple.

[a)]Work supported in part by Junta de Energía Nuclear, Madrid.

II. DESCRIPTION OF THE MODEL AND SOLUTION OF THE FIELD EQUATIONS

We will study the kinks of the model based on the Lagrangian:

$$L=\partial_\mu\phi^*\partial^\mu\phi-m^2\phi^*\phi+\tfrac{1}{2}g_1(\phi^*\phi)^2+g_2J_\mu J^\mu, \tag{1}$$

where

$$J_\mu=\phi^*\partial_\mu\phi-\phi\partial_\mu\phi^*. \tag{2}$$

J_μ is a vector which is closely related to the electromagnetic current. It could be said thus that L includes a kind of electromagnetic structure. A term as $J_\mu J^\mu$ appears in some cases[18] as a consequence of the invariance under the Poincaré gauge group. It corresponds to

$$(\bar\psi\gamma^\mu\gamma^5\psi)(\bar\psi\gamma_\mu\gamma^5\psi)=(\bar\psi\gamma^\mu\psi)(\bar\psi\gamma_\mu\psi),$$

which appears in the case of the spinor field.[19,20] When $g_2=0$, we have the ϕ^4 model, which, as we know, has two drawbacks: The kinks are unstable and it can only represent neutral particles.

The field equation is

$$(\partial_\mu\partial^\mu+m^2)\,\phi=g_1(\phi^*\phi)\,\phi+g_2(4\phi^*\partial_\mu\phi\partial^\mu\phi-4\phi\partial_\mu\phi^*\partial^\mu\phi+2\phi^*\phi\partial_\mu\partial^\mu\phi-2\phi^2\partial_\mu\partial^\mu\phi^*). \tag{3}$$

We look for stationary, spherically symetric solutions, which we write as

$$\phi(x)=(m/\sqrt{|g_1|})\,S(\rho)\exp(-i\omega t),\quad \rho=mr; \tag{4}$$

the field equation takes the following form ($D=d/d\rho$):

$$[D^2+(2/\rho)D+\Lambda^2-1]S=\eta(-1+8\Lambda^2B)S^3, \tag{5}$$

where

$$B=m^2g_2/g_1,\quad \eta=\mathrm{sgn}(g_1),\quad \Lambda=\omega/m,\quad |\Lambda|<1. \tag{6}$$

In order to simplify (5), we make the following change of variables and parameters

$$S=\alpha T,\quad \rho=k\sigma,\quad \alpha^2=(1-\Lambda^2)/|1-8\Lambda^2B|,$$
$$k^2=1/(1-\Lambda^2),\quad \delta=\mathrm{sgn}(1-8\Lambda^2B), \tag{7}$$

which leads to ($D=d/d\sigma$)

$$[D^2+(2/\sigma)D-1]T=-\eta\delta T^3. \tag{8}$$

This is the well-known radial equation of the ϕ^4 model. It admits kinks if and only if

$$\eta\delta = +1. \tag{9}$$

In that case there exists a discrete infinity of solutions in L^2. We will restrict ourselves to the ground state which is obtained for $T(0) = 4.3441$. $T(\sigma)$ being well known, $S(\rho)$ can be easily determined for all values of Λ and B.

In order to obtain the energy, we use the energy–momentum tensor, which has the form

$$T^{\mu\nu} = [\partial^\mu\phi - 2g_2(\phi^*\phi\partial^\mu\phi - \phi^2\partial^\mu\phi^*)]\partial^\nu\phi^* + \text{c. c.} - g^{\mu\nu}L.$$

We obtain (10)

$$E = (\pi m/|g_1|)\mathcal{E},$$
$$\mathcal{E} = 4k^3[2\Lambda^2\alpha^2 I(T^2) + \eta(\tfrac{1}{2} - 12B\Lambda^2)\alpha^4 I(T^4)], \tag{11}$$

where α, β, η are given in (6), (7) and

$$I(T^2) = \int_0^\infty T^2\sigma^2\, d\sigma = 1.499,$$
$$I(T^4) = \int_0^\infty T^4\sigma^2\, d\sigma = 5.998; \tag{12}$$

these integrals do not depend on Λ, B. It can be shown[16] that the quotient of their values must be 4. In our case it is 4.001, which shows that our numerical precision is enough for our purposes. The formula (11) provides thus an analytic expression for the energy

$$E = E(m, g_1, g_2, \Lambda) = E(m, B, \eta, \Lambda), \tag{13}$$

valid except when $g_1 = 0$, a case which we will treat separately.

Another important quantity is the charge of the solution. If we integrate the electromagnetic current

$$j^\mu = -ie(1 - 4g_2\phi^*\phi)(\phi\partial^\mu\phi^* - \phi^*\partial^\mu\phi),$$

we obtain the charge

$$Q = e(8\pi\Lambda/|g_1|)k^3[\alpha^2 I(T^2) - 4B\eta\alpha^4 I(T^4)]. \tag{14}$$

As we see, the properties of the solutions are given by Λ, B, η while m, $|g_1|$ are scale parameters. It is convenient to consider a plane with coordinates $x = g_1$, $y = m^2 g_2$. We have then $B = \tan\theta$, θ being the polar angle. Any direction of this plane represents a value of B and η, and therefore, a family of solutions which depend continuously on Λ. We will consider the functions $E(\Lambda)$, where

$$E(\Lambda) = E(m, B, \eta, \Lambda). \tag{15}$$

First of all, let us remark that the plane is divided into four regions according to the value of B and the fulfillment of condition (9):

(i) Strictly forbidden region: $(g_1, \theta/g_1 < 0,\ \tan\theta < 1/8)$. As $\eta\delta = -1$ there are no kinks there

(ii) First partially forbidden region: $(g_1, \theta/g_1 > 0,\ \tan\theta > 1/8)$. In it (9) only holds for $|\Lambda| < (8\tan\theta)^{-1/2}$.

(iii) Second partially forbidden region: $(g_1, \theta/g_1 < 0,\ \tan\theta > 1/8)$. In it (9) only holds for $|\Lambda| > (8\tan\theta)^{-1/2}$.

(iv) Allowed region $(g_1, \theta/g_1 > 0,\ \tan\theta < 1/8)$. In it the condition (9) is verified for $|\Lambda| < 1$.

In regions ii and iii there are kinks for some values of Λ. However, the curves $E(\Lambda)$ show no minima. In region iv there are kinks for any Λ such that $|\Lambda| < 1$. The examination of formula (11) shows that this region can be divided in three sectors:

(I) Doublet sector, which corresponds to $7.125° > \theta > 3.57°$. $E(\Lambda)$ has a maximum at $\Lambda = 0$ and two minima at $\Lambda = \pm\Lambda_m$. When $\theta \to \arctan\frac{1}{8} = 7.125°$, $\Lambda_m \to 1$.

(II) Singlet sector, which corresponds to $3.57° > \theta > -8°$. $E(\Lambda)$ has only one minimum at $\Lambda = 0$. If $\theta = 0$, we have the ϕ^4 model.

(III) Triplet sector, which corresponds to $-8° > \theta > -90°$. $E(\Lambda)$ has three minima at $\Lambda = 0$ and $\Lambda = \pm\Lambda m$. If $\theta \to -90°$, $\Lambda_m \to 1$ and $E(\Lambda_m) \to 0$.

The plane $(g_1, m^2 g_2)$ appears in Fig. 1. In Figs. 2(a) and 2(b) we can see some curves $\mathcal{E}(\Lambda)$ for several values of θ.

Let us now consider the case in which $g_1 = 0$. We look for solutions of the form

$$\phi(x) = (1/\sqrt{|g_2|})S(\rho)\exp(-i\omega t) \tag{16}$$

and make the change of variables

$$S = \alpha' T, \quad \rho = k\sigma, \quad \alpha'^2 = (1 - \Lambda^2)/8\Lambda^2, \quad k^2 = 1/(1 - \Lambda^2) \tag{17}$$

the equation takes the same form as (8) with $\text{sgn}(g_2)$ instead of $-\eta\delta$, which indicates that there are kinks if and only if $g_2 < 0$, a condition that we admit from now on. The energy and the charge are given by

$$E = (4\pi/m|g_2|)\mathcal{E},$$
$$\mathcal{E} = 4k^3[2\Lambda^2\alpha'^2 I(T^2) + 12\Lambda^2\alpha'^4 I(T^4)], \tag{18}$$

$$Q = e(8\pi\Lambda/m^2|g_2|)k^3[\alpha'^2 I(T^2) + 4\alpha'^2 I(T^4)]. \tag{19}$$

As we see, m and $|g_2|$ are scale parameters. The curve $\mathcal{E} = \mathcal{E}(\Lambda)$ appears in Fig. 3. It has two minima at $\Lambda = \pm 0.902$ and a divergence at $\Lambda = 0$.

III. INTERPRETATION OF THE RESULTS

One of the problems which poses a classical theory in which the kinks are interpreted as particles is the determination of the frequency. In quantum theory it is determined by the Planck relation which is not valid classically. We have found families of kinks which depend continuously on $\Lambda = \omega/m$. In order to overcome

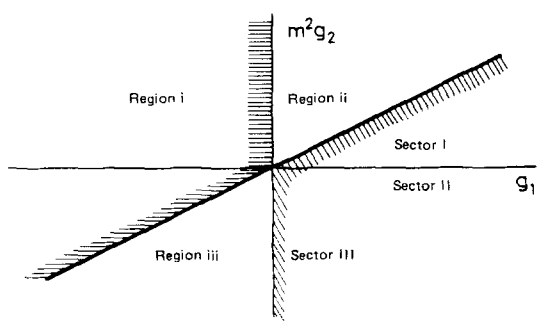


FIG. 1. $(g_1, m^2 g_2)$ plane showing the different regions and sectors.

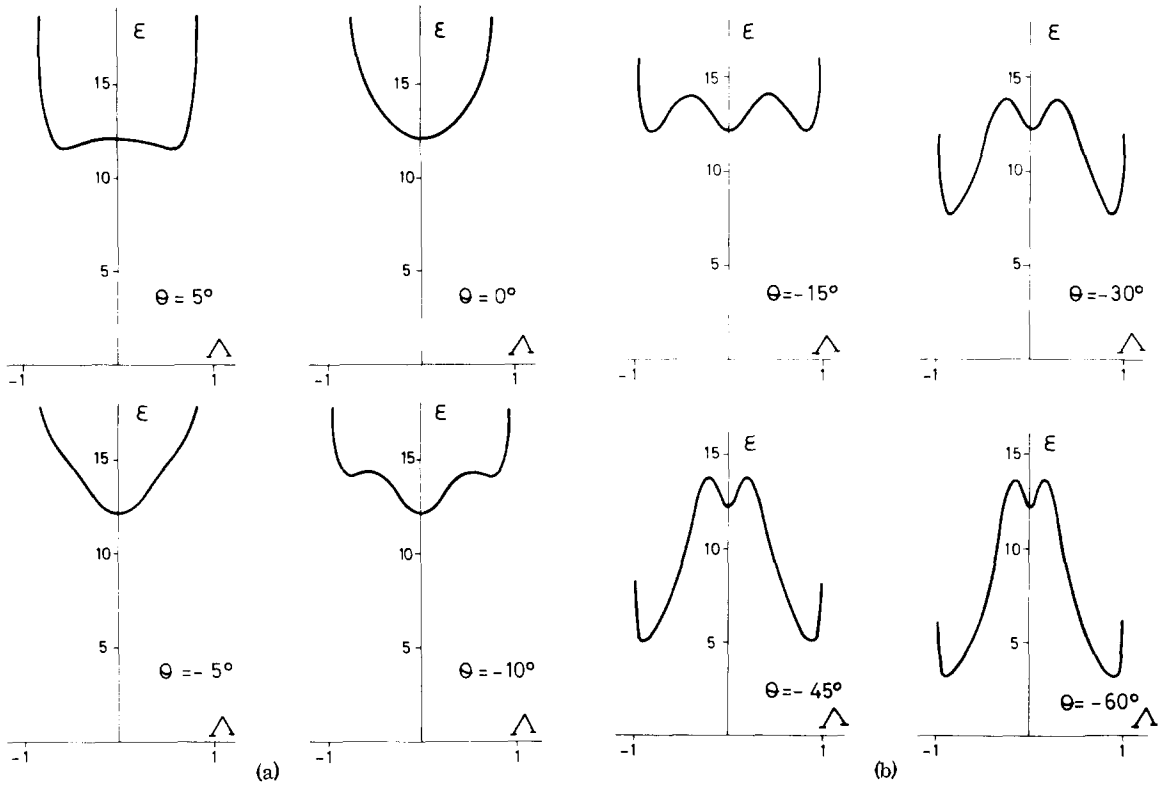


FIG. 2. (a) Shape of the energy $\mathcal{E}$ for different values of the θ angle. (b) Same as in Fig. 2(a) for other values of the θ angle.

the indeterminacy, we adopt the point of view proposed in Soler's paper[6] and define the physical frequencies as those which correspond to minima of $E(\Lambda)$. As is known, the ϕ^4 model has only one minimum for $\Lambda = 0$. This implies that it can only describe time independent states with zero charge. On the other hand the model presented in this paper has one, two, or three minima according to the value of B. We can thus say that it can describe singlets, doublets, or triplets of particles. It is convenient to stress that the existence of charge multiplets is not primarily related to the electromagnetic interaction, but appears as a consequence of the nonlinearity. As the electric charge is proportional to Λ the doublets contain always particles with charge $\pm Q$, because the curves $E(\Lambda)$ are symmetrical with respect to $\Lambda = 0$. In this they differ from the real doublets as (K^+, K^0).

The real elementary particles are too complex to be represented by such a simple model. However, as an illustration, we can compare the doublets or triplets that we have obtained with the pions or kaons. Let us begin with the pion. In order to obtain the values of m, g_1, and g_2 which give an appropriate kink, we have three data $m(\pi^\pm)$, $m(\pi^0)$, and $Q(\pi^\pm)$. If

$$m = 129.6 \text{ MeV}, \quad g_1 = 36.19, \quad g_2 = -5.28\times10^{-4} \text{ MeV}^{-2}, \tag{20}$$

the corresponding triplet has the same energy and charge as the pions. The value of the angle is $\theta \cong -14°$. The mass difference $m(\pi^\pm) - m(\pi^0)$ appears as a non-electromagnetic effect. Other observable quantities are the mean square charge radius and the form factor. With the values (20) we have $\langle r^2\rangle^{1/2} = 2.2$ fm, which is too big as compared with the experimental value ($\cong 0.7$ fm). We have calculated the form factor of the kink as

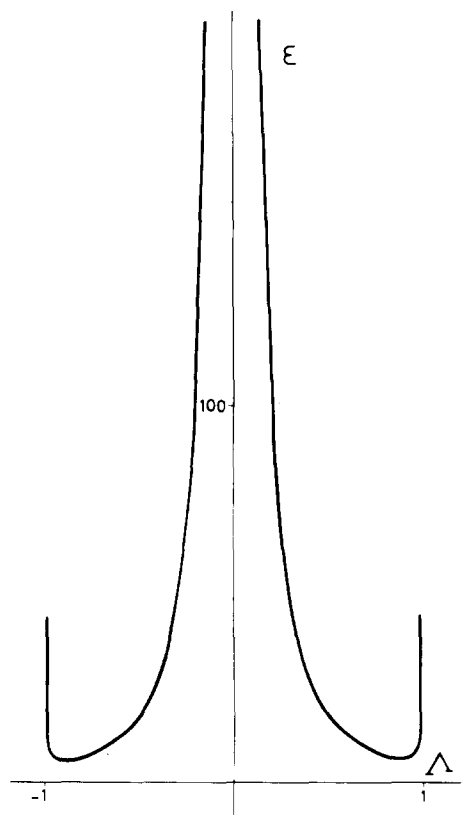


FIG. 3. Shape of the energy $\mathcal{E}$ for $g_1 = 0$.

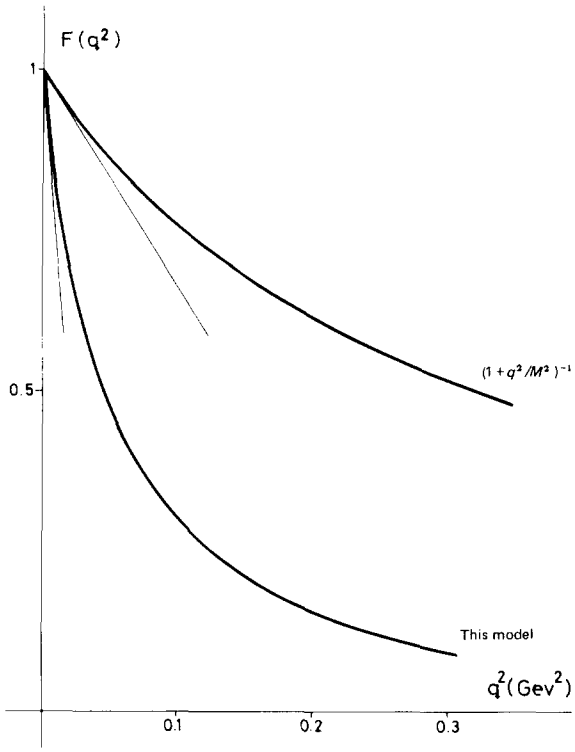


FIG. 4. Pion form factor for this model compared with the parametrized function $(1+q^2/M^2)^{-1}$, $M=0.56$ MeV.

$$F(q^2)=(1/Q)\int j^0 \exp(i\mathbf{qr})\,d^3r;$$

it turns out to decrease faster than the experimental value, in agreement with the previous result on $\langle r^2\rangle^{1/2}$. Figure 4 shows $F(q^2)$ and the pion form factor, approximated[21] as $(1+q^2/M^2)^{-1}$, $M=0.56$ MeV.

As another example, let us consider the pair $K^{\pm}$ (neglecting the K^0, $\bar{K}^0$) and take $g_1=0$. The kinks have the same mass and charge as the kaons if

$$m=473 \text{ MeV}, \quad g_2=-7.8\times10^{-2}\ \text{MeV}^{-2}. \tag{21}$$

IV. ELECTROMAGNETIC STRUCTURE

If one wants to use this model to represent elementary particles, it is convenient to know about its electromagnetic properties, which appear as a consequence of the minimal coupling principle. Instead of L we will have a new Lagrangian L':

$$L'=L+L_{EM}+L_I, \tag{22}$$

where

$$L_{EM}=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{23}$$

$$L_I=(1-4g_2\phi^*\phi)[ie(\phi\partial_\mu\phi^*-\phi^*\partial_\mu\phi)A^\mu+e^2\phi^*\phi A_\mu A^\mu]. \tag{24}$$

If the scalar field has the same form as before and if $A=(m/e)\{V(r),\mathbf{0}\}$, the field equations are $(D=d/d\rho)$.

$$[D^2+(2/\rho)D+(\Lambda-V)^2-1]S=-S^3+8(\Lambda-V)^2BS^3, \tag{25}$$

$$[D^2+(2/\rho)D]V=\epsilon S^2(1-4BS^2)(V-\Lambda), \tag{26}$$

where $\epsilon=2e^2/g_1=8\pi\alpha/g_1$.

We have solved Eqs. (25) and (26) by a perturbative method,[22] and we have found that, as a result of the electromagnetic interaction,

(I) In the doublet sector the minima become deeper and remain displaced to higher values of $|\Lambda|$.

(II) In the singlet sector the minimum becomes deeper and remains at $\Lambda=0$.

(III) In the triplet sector the charged minima become softer and are displaced to higher values of $|\Lambda|$, while the neutral one remain at $\Lambda=0$.

If $g_1=0$, the problem can be treated by a similar method. The energy $\mathcal{E}_1$ is given in Fig. 5. As we see, it has four minima which correspond to $\Lambda_0=\pm0.714$ and $\Lambda_0=\pm0.983$ and it is divergent at $\Lambda_0=0$. As a consequence for high values of ϵ the first order correction transforms the doublet into a quadruplet. We have found that this happens when $\epsilon>0.3$. For the values (21) we obtain $\epsilon=0.010$.

V. SUMMARY AND CONCLUSIONS

We have shown that the model based on the Lagrangian (1) can represent charge multiplets of elementary particles. According to the values of g_1 and g_2 one has a singlet, a doublet, or a triplet. The model can be applied to the pion which appears thus as a kink triplet with the right charge and mass but with a radius bigger than the experimental value.

The class of nonlinear theories is too large to make a systematic study of all the possibilities. Some kind of guiding principle is necessary, and perhaps the consideration of the geometry could shed some light on the problem. But in the absence of an unambiguous mechanism which generates the self-coupling the study of all the lowest order self-coupling could be very useful.

It is worthwhile to make a last remark. For B near to -0.2 the mass of the charged states of the triplet is very close to of the neutral one. This could be interpreted as a breaking of a unitary symmetry in a three

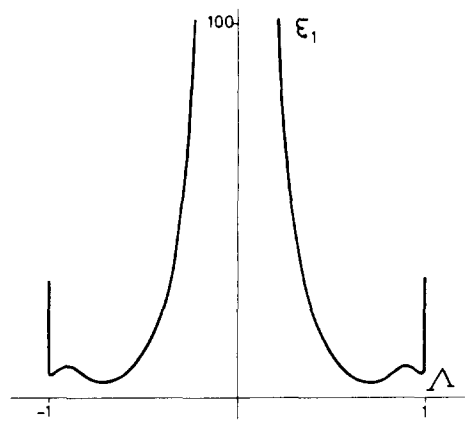


FIG. 5. Shape of the order one energy $\mathcal{E}_1$ for $g_1=0$.

component theory of the scalar field. There is, however no such breaking.

[1]A. Scott, F. Chu, and D. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[2]N. Rosen, Phys. Rev. **55**, 94 (1939).
[3]R. Finkelstein, R. Lelevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
[4]R. Finkelstein, C. Fronsdal, and P. Kaus, Phys. Rev. **103**, 571 (1956).
[5]T. Wakano, Prog. Theor. Phys. **35**, 1117 (1966).
[6]M. Soler, Phys. Rev. D **1**, 2766 (1970).
[7]M. Soler, Phys. Rev. D **8**, 3424 (1973).
[8]A. F. Rañada, M. F. Rañada, M. Soler, and L. Vazquez, Phys. Rev. D **10**, 517 (1974).
[9]A. F. Rañada and L. Vazquez, Prog. Theor. Phys. **56**, 311 (1976).
[10]N. Rosen and H. Rosenstock, Phys. Rev. **82**, 257 (1952).
[11]D. Anderson and G. Derrick, J. Math. Phys. **11**, 1336 (1970).
[12]G. Derrick and W. Kay-Kong, J. Math. Phys. **9**, 232 (1968).
[13]Pinski, J. Math. Phys. **9**, 1323 (1968).
[14]G. Rosen, J. Math. Phys. **6**, 1269 (1965); **8**, 573 (1967).
[15]G. Derrick, J. Math. Phys. **5**, 1252 (1964).
[16]G. Rosen, J. Math. Phys. **7**, 2066, 2071 (1966).
[17]D. Anderson, J. Math. Phys. **12**, 945 (1971).
[18]A. F. Rañada and M. Soler, Nuovo Cimento **8**, 421 (1973).
[19]H. Weyl, Phys. Rev. **77**, 699 (1950).
[20]T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[21]C. Mistretta, D. Imrie, J. A. Appel, R. Budnitz, L. Carroll, M. Goitein, K. Hanson, and R. Wilson, Phys. Rev. Lett. **20**, 1532 (1968).
[22]A. F. Rañada and M. Soler, Phys. Rev. D **8**, 3430 (1973).

Time dependent canonical transformations and the symmetry-equals-invariant theorem[a)]

John R. Cary

Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720
(Received 4 May 1977)

Expressions for the remainder function of a time dependent infinitesimally generated canonical transformation have recently been found by Dewar, who considered the action of the transformation operators on Liouville's equation. Here an alternate proof of the remainder function expression is given, based on the transformations of particle trajectories. Then, using this expression, a proof of the symmetry-equals-invariant theorem is given.

I. INTRODUCTION

In the canonical transformation theory presented in most texts,[1,2] the generating function $F(q, P, t)$ of mixed variables plays a major role. Knowledge of this function allows one to calculate the new Hamiltonian using the equation:

$$K = H + \frac{\partial F}{\partial t}. \tag{1}$$

$\partial F/\partial t$ is known as the remainder function of the transformation. Another topic presented in most texts is that of infinitesimal canonical transformations. By successively doing infinitesimal transformations, one can generate a family of canonical transformations. The formula corresponding to Eq. (1) was not known for a family of canonical transformations until recently, when Deprit[3] found such an expression in terms of a power series expansion. Then Dewar[4] cast Deprit's theory in operator form, and found an expression for the Hamiltonian by considering the action of these operators on Liouville's equation.

The first part of this paper is devoted to deriving Dewar's result by considering the individual particle trajectories rather than Liouville's equation. In this formulation it is seen that finding the remainder function is a calculus problem. The final result of the transformation theory is then used to prove the symmetry-equals-invariant theorem. This theorem has been discussed previously,[5,6] but its proof can be made more rigorous by using the new transformation theory.

II. THEORY OF INFINITESIMAL CANONICAL TRANSFORMATIONS

This section begins with the introduction of notation and the statement of elementary facts concerning canonical transformations. Then the fundamental theorem will be stated and proven.

Following Saletan and Cromer,[1] the set of canonical variables is denoted by the vector $\mathbf{z}$, such that $q_1, \ldots, q_n = z_1, \ldots, z_n$, and $p_1, \ldots, p_n = z_{n+1}, \ldots, z_{2n}$. The matrix γ is defined to contain the Poisson bracket relations:

$$\gamma_{ij} \equiv \{z_i, z_j\} = \begin{cases} 1, & \text{for } j = i+n, \\ -1, & \text{for } i = j+n, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The matrix γ is seen to be antisymmetric and invertible.

$$\gamma_{ij} = -\gamma_{ji} \tag{3a}$$

$$-\sum_k \gamma_{ik}\gamma_{kj} = \delta_{ij}. \tag{3b}$$

It will be necessary to consider time-dependent canonical transformations which depend differentially on a parameter θ. A transformation is canonical if it preserves the Poisson bracket relations:

$$\{Z_m(\mathbf{z}, t, \theta), Z_l(\mathbf{z}, t, \theta)\} \equiv \sum_{ij} \frac{\partial Z_m}{\partial z_i} \gamma_{ij} \frac{\partial Z_l}{\partial z_j} = \gamma_{ml}. \tag{4}$$

In addition to being canonical, the transformations $\mathbf{Z}(\mathbf{z}, t, \theta)$ are required to be invertible, twice differentiable in all arguments simultaneously, and to reduce to the identity when $\theta = 0$:

$$\mathbf{Z}[\mathbf{Z}^{-1}(\mathbf{z}, t, \theta), t, \theta] = \mathbf{Z}^{-1}[\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta] = \mathbf{z} \tag{5a}$$

$$\mathbf{Z}(\mathbf{z}, t, 0) = \mathbf{z}. \tag{5b}$$

It will also be necessary to consider functions of the phase space variables $\mathbf{z}$, the time t, and the parameter θ. By transforming the variables, new functions can be formed from old. As an example, the function $f(\mathbf{z}, t, \theta)$ can be defined by transforming the function $F(\mathbf{z}, t)$ according to

$$f(\mathbf{z}, t, \theta) \equiv F(\mathbf{Z}(\mathbf{z}, t, \theta), t). \tag{6}$$

To avoid ambiguities in taking derivatives, a very explicit notation must be introduced. The symbol

$$\left.\frac{\partial F}{\partial z_l}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta} \tag{7}$$

means: take the derivative of the function $F(\mathbf{z}, t)$ with respect to the variable z_l, then for the variables $\mathbf{z}$, substitute $\mathbf{Z}(\mathbf{z}, t, \theta)$. When the arguments are not explicitly written, they are assumed to be $\mathbf{z}$. This notation is illustrated by applying the chain rule to equation (6):

$$\frac{\partial f}{\partial t} = \left.\frac{\partial F}{\partial t}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t} + \sum_l \left.\frac{\partial F}{\partial z_l}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t} \times \frac{\partial Z_l}{\partial t}. \tag{8}$$

[a)]Work done under the auspices of the U.S. Energy Research and Development Administration.

Finally, one more fact is needed which can be stated in the form of a lemma.

Lemma: Given a differentiable family of invertible canonical mappings $\mathbf{Z}(\mathbf{z}, t, \theta)$. there exists a function $w(\mathbf{z}, t, \theta)$ such that

$$\frac{\partial Z_i}{\partial \theta} = +\{w(\mathbf{z}, t, \theta), Z_i(\mathbf{z}, t, \theta)\}. \tag{9}$$

This statement is shown to be true in Ref. 1, p. 222. For completeness, a proof is included in the Appendix. Using (9) and (5), it is possible to show that the inverse transformation satisfies the following relation with the same function w:

$$\frac{\partial Z_i^{-1}}{\partial \theta} = -\{w(\mathbf{Z}^{-1}(\mathbf{z}, t, \theta), t, \theta), Z_i^{-1}(\mathbf{z}, t, \theta)\}. \tag{10}$$

The function $w(\mathbf{z}, t, \theta)$ is here known as the generating function of the transformation $\mathbf{Z}(\mathbf{z}, t, \theta)$. This function is not to be confused with the generating functions of mixed variables used by Goldstein,[2] which are known here as the "mixed generating functions."

Now that the basic properties of canonical transformations have been discussed, it is possible to discuss the problem at hand. First it is assumed that the evolution in time of the variables $\mathbf{z}$ is given by a Hamiltonian $h(\mathbf{z}, t)$. Then it is known (Ref. 1, Chap. VI) that there exists a function K which gives the evolution of the transformed variables according to

$$\dot{Z}_I(\mathbf{z}, t, \theta) = \{Z_I, K(\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta)\}. \tag{11}$$

The objective here is to find the new Hamiltonian K.

Consider the standard expression for computing the time derivative of the function $Z_I(\mathbf{z}, t, \theta)$:

$$\dot{Z}_I(\mathbf{z}, t, \theta) = \frac{\partial Z_I}{\partial t} + \{Z_I, h\}. \tag{12}$$

Suppose a function $r(\mathbf{z}, t, \theta)$ can be found such that the partial derivative of Z_I with respect to time can be written in the form:

$$\frac{\partial Z_I}{\partial t} = \{Z_I, r\}. \tag{13}$$

Then Eq. (12) becomes

$$\dot{Z}_I = \{Z_I, k\}, \tag{14a}$$

where

$$k = h + r. \tag{14b}$$

Thus the function K which is in (11) is given by

$$K(\mathbf{z}, t, \theta) = k(\mathbf{Z}^{-1}(\mathbf{z}, t, \theta), t, \theta). \tag{15}$$

Now it is seen that to complete the transformation theory, the function $r(\mathbf{z}, t, \theta)$ which satisfies (13) must be found. The function r is found by differentiating (9) with respect to time.

$$\frac{\partial}{\partial \theta}\left(\frac{\partial Z_i}{\partial t}\right) = \left\{\frac{\partial w}{\partial t}, Z_i\right\} + \left\{w, \frac{\partial Z_i}{\partial t}\right\} \tag{16}$$

Equation (16) is a differential equation in θ for the function $\partial Z_i/\partial t$. This equation, together with a boundary condition, uniquely specifies $\partial Z_i/\partial t$. The appropriate boundary condition follows from (5b):

$$\left.\frac{\partial Z_i}{\partial t}\right|_{\mathbf{z}, t, 0} = 0. \tag{17}$$

I now assert that the following set of formulas gives a solution to (16) and (17):

$$\frac{\partial Z_i}{\partial t} = \{Z_i, r\}, \tag{18}$$

$$r(\mathbf{z}, t, \theta) = R(\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta), \tag{19a}$$

$$R(\mathbf{z}, t, \theta) = -\int_0^\theta d\theta' \left.\frac{\partial w}{\partial t}\right|_{\mathbf{Z}^{-1}(\mathbf{z}, t, \theta'), t, \theta'}. \tag{19b}$$

Proving this assertion completes the task of finding K.

To prove that (18) is a solution, it must first be noted that when θ is zero, R vanishes. Using (19a) and (18), this implies that (17) is satisfied. To prove that (16) is satisfied by (18) and (19), I will calculate both sides of equation (16) using (18) and (19), and show them to be equal.

Using (18), the left-hand side (lhs) of Eq. (16) can be put in the form:

$$\text{lhs} = \left\{\frac{\partial Z_i}{\partial \theta}, r\right\} + \left\{Z_i, \frac{\partial r}{\partial \theta}\right\}. \tag{20}$$

Rewriting the second term using (19a) and the chain rule gives

$$\text{lhs} = \left\{\frac{\partial Z_i}{\partial \theta}, r\right\} + \left\{Z_i, \left.\frac{\partial R}{\partial \theta}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta}\right\} + \sum_I \left\{Z_i, \left.\frac{\partial R}{\partial z_I}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta} \times \frac{\partial Z_I}{\partial \theta}\right\}. \tag{21}$$

Then using (9) on the first and third terms, Eq. (21) becomes

$$\text{lhs} = \{\{w, Z_i\}, r\} + \left\{Z_i, \left.\frac{\partial R}{\partial \theta}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta}\right\} + \sum_I \left\{Z_i, \left.\frac{\partial R}{\partial z_I}\right|_{\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta} \times \{w, Z_I\}\right\}. \tag{22}$$

Recognizing the chain rule in the following form:

$$\sum_I \left.\frac{\partial R}{\partial z_I}\right|_{Z_I(\mathbf{z}, t, \theta), t, \theta} \times \{w, Z_I\} = \{w, R(\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta)\}, \tag{23}$$

and inserting (19b) into the second term of (22) results in

$$\text{lhs} = \{\{w, Z_i\}, r\} + \left\{\frac{\partial w}{\partial t}, Z_i\right\} + \{Z_i, \{w, r\}\}. \tag{24}$$

Jacobi's identity allows equation (24) to be written in its final form.

$$\text{lhs} = \left\{\frac{\partial w}{\partial t}, Z_i\right\} + \{w, \{Z_i, r\}\}. \tag{25}$$

This expression is seen to equal the right-hand side of (16) upon using (18), proving the assertion.

The results of this section show the existence of a formula giving the new Hamiltonian in terms of the infinitesimal generating function w. Combining (14b), (15), and (19), the final result is.

$$K(\mathbf{z}, t, \theta) = h(\mathbf{Z}^{-1}(\mathbf{z}, t, \theta), t) - \int_0^\theta d\theta' \left.\frac{\partial w}{\partial t}\right|_{\mathbf{Z}^{-1}(\mathbf{z}, t, \theta'), t, \theta'}. \tag{26}$$

To connect these results to Dewar's,[4] operators corresponding to the transformation are defined by

$$(T(\theta)f)(\mathbf{z}, t, \theta) = f(\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta). \tag{27}$$

Then Eq. (26) becomes

$$K = T^{-1}(\theta)h - \int_0^\theta d\theta' \, T^{-1}(\theta') \frac{\partial w}{\partial t}(\theta'). \tag{28}$$

Though this equation appears to differ from Dewar's equation (27),[4] it is only because of differences in conventions.

As Dewar points out, by expanding w in a power series in θ, Deprit's perturbation theory can be derived. Since the operator T is also a power series, this way of doing perturbation theory involves multiplying series. I would like to point out that in practical calculation, (28) is more convenient than Dewar's formula since there is one less operator series to multiply.

III. THE SYMMETRY-EQUALS-INVARIANT THEOREM

Now I would like to consider the application of (26) to the symmetry-equals-invariant theorem. In its time dependent form, this theorem was partially discussed by Whittaker.[6] Recently, Anderson gave a more complete discussion.[5] Here I would like to show that the proof of this theorem need not be based on expansions; in fact, its proof for a finite composition of infinitesimal transformations becomes straightforward using (26).

First, definitions for the terms used must be given. A family of canonical transformations is said to be a symmetry, if the new Hamiltonian K is identical in form to the old Hamiltonian h up to the addition of an arbitrary function of t and θ alone.

$$K(\mathbf{z}, t, \theta) = h(\mathbf{z}, t) + f(t, \theta). \tag{29}$$

An invariant of the motion $g(\mathbf{z}, t)$ is any function whose total time derivative is zero.

$$\dot{g} \equiv \frac{\partial g}{\partial t} + \{g, h\} = 0. \tag{30}$$

With these definitions, the following theorem is proven.

The Symmetry-Equals-Invariant Theorem: Given a family of canonical transformations which is a symmetry of the Hamiltonian h, one can construct an invariant of the motion g. Conversely, the canonical transformation generated by any invariant g is a symmetry of the Hamiltonian.

Proof: To prove the first statement, we assume that we know the symmetry $\mathbf{Z}(\mathbf{z}, t, \theta)$, and we have constructed the generating function $w(\mathbf{z}, t, \theta)$ as in the Appendix. Using the symmetry property (29) in equation (26) gives

$$h(\mathbf{z}, t) + f(t, \theta) = h(\mathbf{Z}^{-1}(\mathbf{z}, t, \theta), t) - \int_0^\theta d\theta' \left. \frac{\partial w}{\partial t} \right|_{\mathbf{z}^{-1}(\mathbf{z}, t, \theta'), t, \theta'}. \tag{31}$$

Differentiating (31) with respect to θ results in

$$\frac{\partial f}{\partial \theta} = \sum_l \left. \frac{\partial h}{\partial z_l} \right|_{\mathbf{z}^{-1}(\mathbf{z}, t, \theta), t} \times \frac{\partial Z_l^{-1}}{\partial \theta} - \left. \frac{\partial w}{\partial t} \right|_{\mathbf{z}^{-1}(\mathbf{z}, t, \theta), t, \theta}. \tag{32}$$

Now using equation (10) and transforming $\mathbf{z}$, (32) becomes

$$\frac{\partial w}{\partial t} + \{w, h\} + \frac{\partial f}{\partial \theta} = 0. \tag{33}$$

The function $g(\mathbf{z}, t, \theta) = w(\mathbf{z}, t, \theta) - \int^t dt' (\partial f / \partial \theta)(t', \theta)$, is seen to be an invariant of the motion for all θ.

To prove the second statement, I assume a function $g(\mathbf{z}, t)$ is known which is an invariant. Then, the transformation $\mathbf{Z}(\mathbf{z}, t, \theta)$ is determined by integrating (9), using

$$w(\mathbf{z}, t, \theta) = g(\mathbf{z}, t). \tag{34}$$

To prove this transformation is a symmetry, differentiate (26) with respect to θ.

$$\frac{\partial K}{\partial \theta} = \sum_l \left. \frac{\partial h}{\partial z_l} \right|_{\mathbf{z}^{-1}(\mathbf{z}, t, \theta), t} \times \frac{\partial Z_l^{-1}}{\partial \theta} - \left. \frac{\partial g}{\partial t} \right|_{\mathbf{z}^{-1}(\mathbf{z}, t, \theta), t}. \tag{35}$$

Using (10), this becomes

$$\frac{\partial K}{\partial \theta} = -\left(\{g, h\} + \frac{\partial g}{\partial t} \right) \Big|_{\mathbf{z}^{-1}(\mathbf{z}, t, \theta), t}, \tag{36}$$

but since g is an invariant

$$\frac{\partial K}{\partial \theta} = 0. \tag{37}$$

This of course tells us that

$$K(\mathbf{z}, t, \theta) = K(\mathbf{z}, t, 0) = h(\mathbf{z}, t), \tag{38}$$

proving the theorem.

IV. CONCLUSIONS

It has been shown that Dewar's formula for the remainder function for a succession of infinitesimal transformations can be proven by consideration of particle trajectories rather than Liouville's equation. In the process, an equation has been derived which is simpler to use when doing perturbation theory. Finally, this equation has been used to give a rigorous proof of the symmetry-equals-invariant theorem.

ACKNOWLEDGMENTS

I would like to acknowledge long and helpful discussions with A.N. Kaufman and E.H. Wichmann. I also want to acknowledge a helpful correspondence with R.L. Dewar. I have learned that he has now proven (26) in the context of the transformation theory given in Goldstein's text. I also want to thank C. Cary, R. Hagstrom, J. Hammer, D. Judd, and G. Smith for reading various drafts of this manuscript.

APPENDIX: PROOF OF THE LEMMA

Lemma: Given a differentiable family of invertible mappings $\mathbf{Z}(\mathbf{z}, t, \theta)$, there exists a function $w(\mathbf{z}, t, \theta)$ such that

$$\frac{\partial Z_k}{\partial \theta} = \{w, Z_k\}. \tag{A1}$$

Proof: The lemma will be proven by construction. Consider first the vector $\mathbf{V}(\mathbf{z}, t, \theta)$ given by

$$V_k(\mathbf{z}, t, \theta) = \sum_l \gamma_{lk} \frac{\partial Z_l}{\partial \theta}\bigg|_{\mathbf{z}^{-1}(\mathbf{z},t,\theta),t,\theta}. \tag{A2}$$

Suppose $\mathbf{V}$ can be shown to be the gradient of a potential, i.e.,

$$V_k = -\frac{\partial W}{\partial z_k}. \tag{A3}$$

Then w is given by

$$w(\mathbf{z}, t, \theta) = W(\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta). \tag{A4}$$

This can be seen by inserting (A2) into (A3), multiplying the result by γ_{km}, summing over k and using (3b) to get

$$\frac{\partial Z_m}{\partial \theta}\bigg|_{\mathbf{z}^{-1}(\mathbf{z},t,\theta),t,\theta} = \sum_k \frac{\partial W}{\partial z_k}\gamma_{km} = \{W, z_m\}. \tag{A5}$$

Upon transforming $\mathbf{z}$, this becomes

$$\frac{\partial Z_m}{\partial \theta} = \{W(\mathbf{Z}(\mathbf{z}, t, \theta), t, \theta), Z_m(\mathbf{z}, t, \theta)\}. \tag{A6}$$

So it is seen that once the potential W has been found, the lemma has been proven.

To find the potential W, first the symmetry of the partial derivatives of $\mathbf{V}$ must be shown:

$$\frac{\partial V_k}{\partial z_l} \stackrel{?}{=} \frac{\partial V_l}{\partial z_k}. \tag{A7}$$

Once this is proven, W is found by integrating $\mathbf{V}$:

$$W = -\int^{\mathbf{z}} \sum_l V_l(\mathbf{z}', t, \theta)\, dz_l'. \tag{A8}$$

To prove the symmetry (A7), the partial derivative must be calculated. This is done by differentiation of (A2).

$$\frac{\partial V_k}{\partial z_l} = -\sum_m \gamma_{km} \sum_r \frac{\partial^2 Z_m}{\partial \theta \partial z_r}\bigg|_{\mathbf{z}^{-1}(\mathbf{z},t,\theta),t,\theta} \times \frac{\partial Z_r^{-1}}{\partial z_l}. \tag{A9}$$

Digressing for a moment, it is noted that (3b) and (4) can be combined to give

$$\delta_{jl} = \sum_r \frac{\partial Z_j}{\partial z_r}\left(-\sum_{ip} \gamma_{li}\frac{\partial Z_i}{\partial z_p}\gamma_{pr}\right). \tag{A10}$$

Also, differentiating (5a) and transforming $\mathbf{z}$ gives

$$\delta_{jl} = \sum_r \frac{\partial Z_j}{\partial z_r} \cdot \frac{\partial Z_r^{-1}}{\partial z_l}\bigg|_{\mathbf{Z}(\mathbf{z},t,\theta),t,\theta}. \tag{A11}$$

At this point the fact that the matrix $\partial Z_j/\partial z_r$ is invertible, since the transformation $\mathbf{Z}(\mathbf{z}, t, \theta)$ is invertible, is used to imply, from (A10) and (A11), the following relationship:

$$\frac{\partial Z_r^{-1}}{\partial z_l}\bigg|_{\mathbf{Z}(\mathbf{z},t,\theta),t,\theta} = -\sum_{ip} \gamma_{li}\frac{\partial Z_i}{\partial z_p}\gamma_{pr}. \tag{A12}$$

Transforming $\mathbf{z}$ in (A12) and inserting the result into (A9) gives the final form for the partial derivative:

$$\frac{\partial V_k}{\partial z_l} = \sum_{mrip} \gamma_{km}\frac{\partial^2 Z_m}{\partial \theta \partial z_r}\gamma_{rp}\frac{\partial Z_i}{\partial z_p}\cdot\gamma_{il}\bigg|_{\mathbf{z}^{-1}(\mathbf{z},t,\theta),t,\theta}. \tag{A13}$$

To prove that the right-hand side of (A13) is symmetric in k and l, Eq. (4) is differentiated with respect to θ.

$$\sum_{rp}\frac{\partial^2 Z_m}{\partial \theta \partial z_r}\gamma_{rp}\frac{\partial Z_i}{\partial z_p} = -\sum_{rp}\frac{\partial Z_m}{\partial z_r}\gamma_{rp}\frac{\partial^2 Z_i}{\partial \theta \partial z_p}. \tag{A14}$$

Upon inserting this relation into (A13) and using the antisymmetry property of the γ matrix, the symmetry (A7) is seen to be true, proving the lemma.

[1] E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971).
[2] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950).
[3] A. Deprit, Celestial Mech. 1, 12 (1969).
[4] R. L. Dewar, J. Phys. A 9, 2043 (1976).
[5] J. L. Anderson, Am. J. Phys. 40, 541 (1972).
[6] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944), Sec. 144.

Neutron slowing down and transport in a medium of constant cross section. I. Spatial moments[a)]

Dan G. Cacuci[b)] and Herbert Goldstein

Department of Mechanical and Nuclear Engineering, Columbia University, New York, New York 10027
(Received 14 March 1977)

Some aspects of the problem of neutron slowing down and transport have been investigated in an infinite medium consisting of a single nuclide scattering elastically and isotropically without absorption and with energy-independent cross sections. The method of singular eigenfunctions has been applied to the Boltzmann equation governing the Laplace transform (with respect to the lethargy variable) of the neutron flux. Formulas have been obtained for the lethargy dependent spatial moments of the scalar flux applicable in the limit of large lethargy. In deriving these formulas, use has been made of the well-known connection between the spatial moments of the Laplace-transformed scalar flux and the moments of the flux in the "eigenvalue space." The calculations have been greatly aided by the construction of a closed general expression for these "eigenvalue space" moments. Extensive use has also been made of the methods of combinatorial analysis and of computer evaluation, via FORMAC, of complicated sequences of manipulations. It has been possible to obtain for materials of any atomic weight explicit corrections to the age theory formulas for the spatial moments $M_{2n}(u)$, of the scalar flux, valid through terms of order of u^{-5}. Higher order correction terms could be obtained at the expense of additional computer time. The evaluation of the coefficients of the powers of n, as explicit functions of the nuclear mass, represent the end product of this investigation.

1. INTRODUCTION

Perhaps the simplest problem of linear transport theory involving transport combined with energy change is the slowing down and transport of neutrons by purely elastic scattering in an infinite homogeneous medium with cross section independent of energy.

Yet, it has not been possible to find analytic solutions for the neutron flux density as a function of space and energy even when the problem is further restricted to scattering isotropic in the center-of-mass system. Attention has for the most part been concentrated at finding approximate solutions for certain limiting cases. Thus, for energies very small compared to the source energy (i.e., at large lethargy) the neutron must have suffered many collisions. Its track in phase space is then described by a random walk with many steps, and the law of large numbers may be used to describe the energy and spatial distribution of the neutrons, at least not too far from the source. The resultant predictions, known as "age-theory," have been widely used in practrical applications.

Very little has been done to define precisely the limits of validity of age theory or to obtain corrections that can be used for a wider range of conditions. It is likely that the requirements for age theory as given in the older literature, mostly dating from studies during World War II, are not correct. Thus Davison[1] states explicitly "that age theory... is nowhere valid in hydrogen" (p. 407), while McInerney[2] and Amster[3] have shown that even in hydrogen, age theory is the correct limiting form for large lethargy.

Marshak in his 1947 review paper on neutron transport[4] made brief reference to some remarkable results obtained by Placzek in a then unpublished 1942 report.[5] Placzek concentrated on the set of equations, obtained from the Boltzmann equation, for the spatial moments of the various components of the angular flux density. By working with the Laplace transform of the lethargy dependent Fourier space transform of the flux, he was able to derive formulas for the spatial moments of the scalar flux which incorporate the first correction term in reciprocal lethargy (i.e., in u^{-1}) to the age theory prediction. Using the same methods Huang[6] extended Placzek's results to correction terms of the order of u^{-2} for scatterers heavier than hydrogen.

The method of singular eigenfunctions,[7,2] applied to the Laplace transformed Boltzmann equation, provides a powerful technique for seeking higher order corrections to the spatial moments. This is the approach chosen in the present investigation. Formulas have been obtained for the lethargy dependent spatial moments of the scalar flux applicable in the limit of large lethargy. In deriving these formulae, use has been made of the well-known[8] connection between the spatial moments of the Laplace-transformed scalar flux and the moments of the flux in the "eigenvalue space." The calculations have been greatly aided by the construction of a closed general expression for these "eigenvalue space" moments [i.e., Eq. (40), below]. Extensive use has also been made of the methods of combinatorial analysis and of computer evaluation of complicated sequences of manipulations. It has been possible to obtain for materials of any atomic weight, explicit corrections to the age theory formulas for the spatial moments, $M_{2n}(u)$, of the scalar flux, valid through terms of the order of u^{-5}. Each of these terms can be represented as a finite power series in n, and the explicit formulas in terms of the nuclear mass, of the coefficients of the powers of n represent the end product of this investigation.

2. THE SPATIAL MOMENTS OF THE LAPLACE TRANSFORMED BOLTZMANN EQUATION

We consider the time independent homogeneous

[a)]Research partially supported under USERDA Contract EY-76-S-02-2509.
[b)]Present address: Oak Ridge National Laboratory, Neutron Physics Division, Oak Ridge, Tennessee 37830.

Boltzmann equation for the case of neutron transport under the following conditions: infinite medium, plane geometry, azimuthally independent scattering, and cross sections that are constant in energy. The explicit form is

$$\left(\mu\frac{\partial}{\partial x}+1\right)\Psi(x,\mu,u)$$
$$=c\int_{\Omega'}\int_{u'} f(\Omega\cdot\Omega',u-u')\,\Psi(x,\mu',u')\,du'\,d\Omega', \tag{1}$$

where $\Psi(x,\mu,u)$ is the angular flux density of the neutrons per unit solid angle and lethargy. All other quantities in (1) are in the standard notation of transport theory and have their usual meaning.

The scattering kernel for isotropic scattering in the c. m. system by a single element of mass A is:

$$f(\Omega\cdot\Omega',u-u')$$
$$=\frac{\exp[-(u-u')]}{8\pi}\,\frac{(A+1)^2}{A}\cdot\delta\left[\Omega\cdot\Omega'-\left(\frac{A+1}{2}\times\exp[-(u-u')/2]-\frac{A-1}{2}\exp[(u-u')/2]\right)\right],$$
$$\text{for } (u-u')\in[0,\epsilon] \text{ with } \epsilon=2\ln\frac{A+1}{A-1},$$
$$=0 \text{ for } (u-u')>\epsilon. \tag{2}$$

By expanding the scattering kernel in spherical harmonics, applying a one-sided Laplace transform in lethargy and using the convolution theorem, we obtain from (1) and (2):

$$\left(\mu\frac{\partial}{\partial x}+1\right)\Psi(x,\mu,\eta)$$
$$=c\sum_{n=0}^{\infty}\frac{2n+1}{2}\,g_n(\eta)\int_{-1}^{1}P_n(\mu')\,\Psi(x,\mu',\eta)\,d\mu', \tag{3}$$

where

$$\Psi(x,\mu,\eta)\equiv\mathcal{L}[\Psi(x,\mu,u)] \tag{4}$$

and

$$g_n(\eta)\equiv\mathcal{L}[f_n(u)]=\int_0^{\epsilon}\exp(-\eta u)f_n(u)\,du. \tag{5}$$

Equation (3) has the same form as the one speed transport equation with anisotropic scattering to all orders in terms of Legendre polynomials, depending, moreover, on the additional parameter η.[2,7]

Substituting

$$\Psi(x,\mu,\eta)=\phi(\nu,\mu,\eta)\exp(-x/\nu) \tag{6}$$

and defining

$$\phi_n(\nu,\eta)\equiv\int_{-1}^{1}\phi(\nu,\mu,\eta)\,P_n(\mu)\,d\mu \tag{7}$$

we obtain the well-known eigenvalue equation[2,7,9] for $\phi(\nu,\mu,\eta)$

$$(\nu-\mu)\,\phi(\nu,\mu,\eta)=\tfrac{1}{2}c\nu\sum_{n=0}^{\infty}(2n+1)\,g_n(\eta)\,P_n(\mu)\,\phi_n(\nu,\eta), \tag{8}$$

where the normalization of $\phi(\nu,\mu,\eta)$ being, at this stage arbitrary, can be chosen such that

$$\phi_0(\nu,\eta)\equiv\int_{-1}^{1}\phi(\nu,\mu,\eta)\,d\mu=1. \tag{9}$$

Multiplying Eq. (8) by $P_k(\mu)$, integrating over μ from -1 to $+1$ and using the recurrence relation for the Legendre polynomials, we obtain the recurrence relation for the polynomials $\phi_n(\nu,\eta)$:

$$\xi_n\nu\phi_n(\nu,\eta)=(n+1)\,\phi_{n+1}(\nu,\eta)+n\phi_{n-1}(\nu,\eta), \tag{10}$$

where

$$\xi_n\equiv\xi_n(\eta)=(2n+1)\,\gamma_n(\eta) \tag{11}$$

and

$$\gamma_n(\eta)=1-cg_n(\eta). \tag{12}$$

Relation (10), together with Eq. (9) and $\phi_{-1}\equiv 0$ (9′) clearly define all ϕ_n's.

We now recall some essential (and well-know[2,7,9]) properties of the eigensolutions $\phi(\nu,\mu,\eta)$.

The eigenvalue spectrum contains a continuous portion extending over $(-1,+1)$, while the discrete eigenvalues are the roots of

$$\Lambda(\nu,\eta)\equiv 1-\tfrac{1}{2}c\nu\int_{-1}^{1}\frac{M(\nu,\mu,\eta)}{\nu-\mu}\,d\mu$$
$$=1-\tfrac{1}{2}c\nu\int_{-1}^{1}\frac{M(\mu,\mu,\eta)}{\nu-\mu}\,d\mu=0, \tag{13}$$

where

$$M(\mu,\nu,\eta)=\sum_{l=0}^{\infty}(2l+1)\,g_l(\eta)\,P_l(\mu)\,\delta(\mu-\nu), \tag{14}$$

The series in Eq. (14) will be uniformly convergent if[9] $\sum_{l=0}^{\infty}(2l+1)\,|g_l(\eta)|\leq\infty$ since both $P_l(\mu)$ and $\phi_l(\mu,\eta)$ are bounded for $\mu\in[-1,1]$.

The solutions belonging to the discrete spectrum are given by

$$\phi(\pm\nu_j,\mu,\eta)=\pm\tfrac{1}{2}\nu_j\,\frac{M(\mu,\pm\nu_j,\eta)}{\pm\nu_j-\mu}, \tag{15}$$

where ν_j depends on η, while the continuous spectrum is expressed as

$$\phi(\nu,\mu,\eta)=\tfrac{1}{2}c\nu P\,\frac{M(\mu,\nu,\eta)}{\nu-\mu}+\lambda(\nu,\eta)\,\delta(\mu-\nu), \tag{16}$$

with

$$\lambda(\nu,\eta)=1-\tfrac{1}{2}c\nu P\int_{-1}^{1}\frac{M(\mu,\nu,\eta)}{\nu-\mu}\,d\mu \tag{17}$$

(P denotes Cauchy principal value).

The normalization integrals for the discrete and continuous spectrum eigenfunctions are defined by

$$N_j\pm(\pm\nu_j,\eta)=\int_{-1}^{1}\mu\phi^2(\pm\nu_j,\mu,\eta)\,d\mu \tag{18}$$

and

$$N(\nu,\eta)\,\psi(\nu,\eta)=\int_{-1}^{1}\mu\phi(\nu,\mu,\eta)\times\int_{-1}^{1}\psi(\xi,\eta)\,\phi(\xi,\mu,\eta)\,d\xi\,d\mu, \tag{19}$$

where ψ is a function which may be expanded in terms of the eigenfunctions $\phi(\nu_j,\mu,\eta)$ and $\phi(\nu,\mu,\eta)$.

We now consider a plane, monoenergetic source S isotropically emitting neutrons:

$$S(x,\mu,u)=\delta(x)\,\delta(u)/2. \tag{20}$$

By Laplace transforming the inhomogeneous Boltzmann equation, by employing appropriate boundary conditions and using the above listed properties, one can show[2,7] that the Green function for this problem can be written as

$$\Psi^G(x,\mu,\eta)=\sum_j \frac{\phi(\nu_j,\mu,\eta)}{2N_j(\nu_j,\eta)}\exp(-x/\nu_j) + \int_0^1 \frac{\phi(\nu,\mu,\eta)}{2N(\nu,\eta)}\exp(-x/\nu)\,d\nu, \quad \text{for } x>0,\ \mathrm{Re}(\nu_j)>0. \tag{21}$$

The scalar flux is

$$\Psi_0^G(x,\eta)\equiv\int_{-1}^{1}\Psi^G(x,\mu,\eta)\,d\mu = \sum_j \frac{\exp(-x/\nu_j)}{2N_j(\nu_j,\eta)} + \int_0^1 \frac{\exp(-x/\nu)}{2N(\nu,\eta)}\,d\nu. \tag{22}$$

We can define the spatial moments of the scalar flux as

$$M_n(\eta)\equiv\int_{-\infty}^{\infty}x^n\Psi_0^G(x,\eta)\,dx. \tag{23}$$

Because of symmetry, i.e., $\Psi_0^G(x,\eta)=\Psi_0^G(-x,\eta)$, we have

$$M_{2k}(\eta)=2\int_0^\infty x^{2k}\Psi_0^G(x,\eta)\,dx;\quad n=2k. \tag{24}$$

Using (22), Eq. (24) becomes

$$M_{2n}(\eta)=(2n)!\times\left[\sum_j\frac{\nu_j^{2n+1}}{N_j(\nu_j,\eta)}+\int_0^1\frac{\nu^{2n+1}}{N(\nu,\eta)}\,d\nu\right];\quad n=0,1,2,\cdots. \tag{25}$$

In order to proceed further with the evaluation of the spatial moments of the scalar flux, we shall turn our attention to the polynomials $\phi_n(\nu,\eta)$, and discuss some of their most important properties in relation to the associated moment problem.

From (10) and the initial values (9), (9′) we can write $\phi_n(\nu,\eta)$ as the continuant[9–11]:

$$\phi_n(\nu,\eta)=\frac{1}{n!}\begin{vmatrix} \xi_0\nu & 1 & 0 & 0 & \cdots & \cdot & \cdot \\ 1 & \xi_1\nu & 2 & 0 & \cdots & \cdot & \cdot \\ 0 & 2 & \xi_2\nu & 3 & \cdots & \cdot & \cdot \\ 0 & 0 & 3 & \xi_3\nu & \cdots & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cdot & n-1 \\ 0 & 0 & 0 & 0 & \cdots & n-1 & \xi_{n-1}\nu \end{vmatrix},\quad \text{for } n\geq 1. \tag{26}$$

Expanding the determinant, we obtain[9]

$$\phi_n(\nu,\eta)=\frac{1}{n!}\prod_{i=0}^{n-1}\xi_n\times\left[\nu^n-\nu^{n-2}\sum_{j_1=0}^{n-2}w_{j_1}+\nu^{n-4}\sum_{j_1=0}^{n-4}\sum_{j_2=j_1+2}^{n-2}w_{j_1}w_{j_2}\cdots+(-1)^m\nu^{n-2m}\sum_{j_1=0}^{n-2m}w_{j_1}\sum_{j_2=j_1+2}^{n-2m+2}w_{j_2}\cdots\sum_{j_m=j_{m-1}+2}^{n-2}w_{j_m}+\cdots\right.$$
$$\left.+\begin{cases}(-1)^{n/2}w_0w_2\cdots w_{n-2}\big] & \text{for } n=\text{even}\\ (-1)^{(n-1)/2}\nu\sum_{j_1=0}^{1}w_{j_1}\sum_{j_2=j_1+2}^{3}w_{j_2}\cdots\sum_{\substack{j_m=j_{m-1}+2\\ m=1/2(n-1)}}^{n-2}w_{j_m}\big] & \text{for } n=\text{odd}\end{cases}\right. \tag{27}$$

where

$$w_j\equiv(j+1)^2/\xi_j\xi_{j+1} \tag{28}$$

is a function of j and η and ξ_j is as defined in (11).

As in monoenergetic transport, the polynomials $\phi_n(\nu,\eta)$ are orthogonal, in the Stieltjes sense, over the interval $(-\nu_1,+\nu_1)$ (what we shall call the "eigenvalue space"), where $|\nu_1|=\max_j|\nu_j|$, with a proper weight function $\omega(\nu,\eta)$. The actual proof of the orthogonality relations can be carried out along the same lines as has been variously used in monoenergetic transport theory.[2,7,9,10,12–14] The validity of these approaches is not affected by the presence of the additional parameter η. We write the orthogonality relations as[9]:

$$\frac{2\delta_{kl}}{\xi_l}=\int_{-1}^{1}\phi_k(\nu,\eta)\,\phi_l(\nu,\eta)\,\frac{\nu}{N(\nu,\eta)}\,d\nu + \sum_{j=1}^{M}\left[\phi_k(\nu_j,\eta)\,\phi_l(\nu_j,\eta)\,\frac{\nu_j}{N_{j+}(\eta)} + \phi_k(-\nu_j,\eta)\,\phi_l(-\nu_j,\eta)\left(\frac{-\nu_j}{N_{j-}(\eta)}\right)\right], \tag{29}$$

or, more concisely, as

$$\int_{-\nu_1}^{\nu_1}\phi_k\phi_l\,d\omega=2\delta_{kl}/\xi_l, \tag{30}$$

where the function $\omega(\nu,\eta)$ is nondecreasing in the interval $(-\nu_1,+\nu_1)$ and may be expressed[9] as the continuous distribution:

$$\omega(\nu,\eta)=\int_0^{\nu}\frac{d\nu}{\Lambda^+(\nu,\eta)\,\Lambda^-(\nu,\eta)}$$
$$=\int_0^{\nu}\frac{d\nu}{\lambda^2(\nu,\eta)+\frac{1}{4}\pi^2\nu^2c^2M^2(\nu,\nu,\eta)}\quad\text{for } |\nu|\le 1 \tag{31}$$

and the sequence of horizontal lines

$$\omega(\mp\nu,\eta)=\omega(\mp 1,\eta)\mp\sum_{i=m}^{M}\frac{\nu_i}{N_i}\quad\text{for } \nu_m<|\nu|<\nu_{m-1}, \tag{32}$$

with

$$\omega(\mp\nu,\eta)=\omega(\mp 1,\eta)\mp\sum_{i=1}^{M}\frac{\nu_i}{N_i}\quad\text{for } \nu_1<|\nu|. \tag{33}$$

A. Properties of the "Eigenvalue space" moments m_k (η)

We can now introduce moments m_k related to the orthogonal polynomials $\varphi_n(\nu,\eta)$, defined as

$$m_k\equiv m_k(\eta)=\int_{-\nu_1}^{\nu_1}\nu^k\,d\omega(\nu,\eta),\quad\text{for } k=0,1,2,\cdots. \tag{34}$$

Because of symmetry, the odd order moments vanish and Eq. (34) becomes

$$m_k=\begin{cases}0 & \text{for } k=\text{odd}\\ 2\int_0^{\nu_1}\nu^k\,d\omega(\nu,\eta) & \text{for } k=\text{even}.\end{cases} \tag{35}$$

Previous investigations (e.g., Inonu,[9] etc.) have not led to a closed and general formula for the m_k's. We propose instead a method which will be shown to lead to a closed form, explicitly displaying the dependence of $m_{2n}(\eta)$ on n and on certain w's.

For this purpose, we multiply Eq. (27) by $d\omega(\nu,\eta)$ and integrate it over the interval $(0,\nu_1)$. Taking also into account the orthogonality relations (29) and Eq. (35), we obtain

$$m_{2p}=m_{2(p-1)}A_{1,p}-m_{2(p-2)}A_{2,p}+m_{2(p-3)}A_{3,p}$$
$$+\cdots+(-1)^{k+1}m_{2(p-k)}A_{k,p}+\cdots+(-1)^{p+1}m_0A_{p,p}$$
$$\text{for } p>0, \tag{36}$$

with

$$m_0=2/\xi_0=2/\gamma_0(\eta)\quad\text{for } p=0, \tag{37}$$

where

$$A_{k,r}=\sum_{j_1=0}^{2(r-k)}w_{j_1}\sum_{j_2=j_1+2}^{2(r-k+1)}w_{j_2}\cdots\sum_{j_k=j_{k-1}+2}^{2(r-1)}w_{j_k}. \tag{38}$$

By successively setting $p=1,2,\ldots,n$ (36), we obtain a system of n linear equations, which can be solved for m_{2n} by Cramer's rule. The determinant of this set of equations is automatically unity, since all terms on the main diagonal are $+1$ and all elements on one side of the main diagonal are zero. Hence m_{2n} is given by

$$m_{2n}=\begin{vmatrix}(-1)^{n+1}A_{n,n} & -A_{1,n} & A_{2,n} & \cdots & (-1)^{n-2}A_{n-2,n} & (-1)^{n-1}A_{n-1,n}\\ (-1)^{n}A_{n-1,n-1} & 1 & -A_{1,n-1} & \cdots & (-1)^{n-3}A_{n-3,n-1} & (-1)^{n-2}A_{n-2,n-1}\\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot\\ -A_{2,2} & 0 & 0 & \cdots & 1 & -A_{1,2}\\ A_{1,1} & 0 & 0 & \cdots & 0 & 1\end{vmatrix}. \tag{39}$$

By reducing the first column to zeros, except for A_{11}, then using Laplace's theorem[11] and repeating the process on the resulting determinants, Eq. (39) can be reduced to

$$m_{2n}=m_0\sum_{i_0=0}^{0}w_{i_0}\sum_{i_1=0}^{i_0+1}w_{i_1}\sum_{i_2=0}^{i_1+1}w_{i_2}\cdots\sum_{i_{n-1}=0}^{i_{n-2}+1}w_{i_{n-1}}, \tag{40}$$

where the m_{2n}'s and w's are functions of η and the nuclear mass A. The values for m_{2n} obtained through Eq. (40) satisfy identically relationships among m_{2n} as presented by Inonu[9] and as can be found in the standard theory of orthogonal polynomials.[10,13-15] Equation (40) is of basic importance, insofar as it gives the dependence of $m_{2n}(\eta)$ on the coefficients $g_n(\eta)$ [i.e., the Laplace transformed coefficients of the Legendre polynomial expansion of the scattering kernel $f(\bar{\Omega}\cdot\bar{\Omega}',\mu)$], in a closed and structured way, facilitating the investigations of the most important characteristics of $m_{2n}(\eta)$, which in turn, will translate [via Eq. (48), below] into properties of the spatial moments $M_{2n}(\eta)$. Based on Eq. (40), we have calculated moments up to, and including m_{18}, using the symbolic computer language FORMAC,[16] on Columbia's IBM 360/91. A listing is presented in Appendix A.[17]

We now present some of the properties of $m_{2n}(\eta)$ particularly pertinent to our goal of evaluating the spatial moments. These characteristics are derived from Eq. (40) using the methods of combinatorial analysis. Examination of the nested sums in Eq. (40) indicates that the right-hand side of the equation can be written as a single sum of terms all of the form

$$w_{j_1}w_{j_2}w_{j_3}w_{j_4}\cdots w_{j_n},$$

where the indices are subject to the restrictions

$$0\le j_i\le i-1$$
$$j_i\le j_{i-1}+1. \tag{41}$$

It is obviously consistent with these restrictions for there to be terms in which an index j_i can be equal to another index j_k, so that powers of various w's can occur. Indeed, all powers through n can occur, e.g., w_0^n is a valid term in Eq. (40). Details of the procedure are omitted in the interests of brevity, but the following explicit formulas for the coefficients of some specific terms are given by way of illustrations:

$$\text{coeff}(w_0^{n-k}w_1^k) = (n-1)!/(n-k-1)!\,k!, \tag{42}$$

$$\text{coeff}(w_0^{n-k}w_1w_2\cdots w_k) = n-k, \tag{43}$$

$$\text{coeff}(w_0^{n-k}w_1^{k-j}w_2\cdots w_{j+1} = (n-j-1)!/ \times (n-1-k)!\,(k-j-1)!. \tag{44}$$

[By coefficient of a term is meant the number of times the term appears in the expansion of the right-hand side of Eq. (40).] In addition, Appendix B presents the first 32 terms in Eq. (40) in order of decreasing powers of w_0, starting with w_0^n.

Some other properties of m_{2n} may be mentioned.

Let E_{2n} be the number of terms in the fully expanded expression of m_{2n} [i.e., sum of all coefficients in Eq. (40)]. Then E_{2n} is given by

$$E_{2n} = \sum_{i=0}^{n-1} (-1)^{n-i+1} E_{2i} K_{n-i,2i+1} \tag{45}$$

where $K_{i,j}$ is the jth triangular number of i, i.e.,

$$K_{i,j} = \sum_{r=1}^{j} K_{i-1,r}, \quad K_{0,j} = 1, \quad K_{1,j} = j. \tag{46}$$

Also, the number of distinct terms in m_{2n} is given by 2^{n-1}. Moreover, for a particular m_{2n}, the number of distinct terms that contain a particular w_k^{n-r} ($k=0$ to $n-1$; $r=0$ to n) is given by 2^r. These properties of Eq. (40) enable us to recast the expansion of m_{2n} in the form:

$$m_{2n}(\eta) = m_0 w_0^n \sum_{r=0}^{n-1} \frac{1}{w_0^r} \sum_{s=0}^{r} n^s B_{r,s}(\eta), \tag{47}$$

where the quantities $B_{r,s}(\eta)$ are related in an obvious way to the coefficients of w_0^k ($k=0$ to n). A listing of $B_{r,s}$ from $r, s = 0$ to 5 is presented in Appendix C.

A simple relation exists between the "eigenvalue space" moments m_{2n} and the spatial moments M_{2n} defined by (23), as has been demonstrated by Case[8] for monoenergetic transport. The proof used there depends basically on the identity of the density functions $\omega(\nu,\eta)$ appearing in the orthogonality relations Eq. (30) and the solution representation Eq. (22). In as much as η appears here solely as a parameter, the derivation can be taken over without change, leading to

$$M_{2n}(\eta) = (2n)!\ \tfrac{1}{2}m_{2n}(\eta), \tag{48}$$

where

$$m_0(\eta) = 2/\gamma_0(\eta). \tag{49}$$

3. THE LAPLACE INVERSION: FORM OF $M_{2n}(\mu)$ FOR $C = 1$

As we have already pointed out, the lethargy-dependent spatial moments of the scalar flux $\Psi_0^G(x,u)$ are given by

$$M_{2n}(u) = \mathcal{L}^{-1}[M_{2n}(\eta)] = (2n)!\ \mathcal{L}^{-1}[\tfrac{1}{2}m_{2n}(\eta)]$$

$$= (2n)! \sum \text{Res}[\exp(\eta u)\,\tfrac{1}{2}m_{2n}(u)]_{\text{poles}}. \tag{50}$$

In order to carry out the inversion, properties of the relevant poles of $m_{2n}(\eta)$ have to be discussed.

From Eqs. (5) and (12) one can easily see that the $\gamma_n(\eta)$ are entire functions and have the following properties[18,19]:

$$\lim_{n\to\infty} g_n(\eta) = 0 \quad \text{for } \text{Re}(\eta) > 0, \tag{51}$$

$$\lim_{\text{Re}(\eta)\to\infty} g_n(\eta) = 0, \tag{52}$$

for $c \leq 1$:

$$|\gamma_n(\eta)| < 1 \quad \text{for real } \eta > 0, \tag{53}$$

and for $c = 1$:

$\gamma_0(\eta) = 0$ at $\eta = 0$ and $\eta = 0$ is the zero with the largest real part. (54)

It is now seen from Eqs. (28) and (47) that $m_{2n}(\eta)$ has poles at the zeros of $\gamma_i(\eta)$ ($i=0$ to n). While it seems impossible to obtain the exact Laplace inverse of (47), Tauberian theorems[20,21] permit the investigation of $M_{2n}(u)$ in an asymptotic sense. Specifically, the behavior of $M_{2n}(u)$ for small values of u (at or near the source) can be found by studying the behavior of $m_{2n}(\eta)$ for $\eta \to \infty$ and, correspondingly, the behavior of $M_{2n}(u)$ for large u is dictated by the pole of $m_{2n}(\eta)$ that has the largest real part.

Since $m_{2n}(\eta)$ has a pole of order $(n+1)$ at $\eta = 0$, as can be readily seen from Eqs. (47), (37), (28), and (54), the residue appearing in the right-most member of Eq. (50) evaluated at $\eta = 0$ is

$$\text{Res} = \frac{1}{n!}\frac{d^n}{d\eta^n}[\eta^{n+1}\exp(\eta u)\,\tfrac{1}{2}m_{2n}(\eta)]_{\eta=0}. \tag{55}$$

We expand Eq. (47) in a power series of η. The leading term in this expansion of m_{2n} is of order $\eta^{-(n+1)}$ which cancels the corresponding factor in Eq. (55), which can now be written as

$$\text{Res} = \frac{1}{n!}\frac{1}{\gamma_0'(0)}\left[\frac{1}{3\gamma_0'(0)\gamma_1(0)}\right]^n \frac{d^n}{d\eta^n}\left[\exp(\eta u)\,h(\eta)\right]_{\eta=0}, \tag{56}$$

where

$$h(\eta) = \left(\sum_{r=0}^{n-1}\frac{1}{w_0^r}\sum_{s=0}^{r} n^s B_{r,s}\right) \Bigg/ \left[\left(\sum_{k=0}^{\infty} a_k\eta^k\right)^{n+1}\left(\sum_{k=0}^{\infty} b_k\eta^k\right)^n\right], \tag{57}$$

with

$$a_0 = 1, \quad a_k = \frac{1}{\gamma_0'(0)}\frac{1}{(k+1)!}\left[\frac{d^{k+1}}{d\eta^{k+1}}\gamma_0(\eta)\right]_{\eta=0}$$

$$\text{for } k = 1, 2, \cdots, \tag{58}$$

and

$$b_0 = 1, \quad b_k = \frac{1}{\gamma_1(0)}\frac{1}{k!}\left[\frac{d^k}{d\eta^k}\gamma_1(\eta)\right]_{\eta=0}$$

$$\text{for } k = 1, 2, \cdots. \tag{59}$$

Applying Leibnitz' differentiation formula to (56) we obtain

$$\text{Res} = \frac{1}{\gamma_0'(0)}\frac{\tau^n}{n!}\sum_{m=0}^{n}\binom{n}{m}\frac{1}{u^m}h_{(0)}^{(m)}, \tag{60}$$

where

$$\tau = u/3\gamma_0'(0)\gamma_1(0) \tag{61}$$

and

$$h_{(0)}^{(m)} = \left[\frac{d^m}{d\eta^m} h(\eta)\right]_{\eta=0}. \tag{62}$$

In the following, the function $h_{(0)}^{(m)}$ appearing in (62) will be cast into a form suitable for practical applications. Specifically, we will show that

$$h_{(0)}^{(m)} = (m)!\, C_m(n) = m! \sum_{j=0}^{m} C_{m,j} n^j. \tag{63}$$

The coefficients $C_{m,j}$ in (63) are functions of the nuclear mass only. The values of $C_{m,j}$ can be obtained through a recursion relation which can be derived by the following sequence of steps:

(i) The numerator in (57) is expanded in an infinite power series in η, around $\eta = 0$. The coefficient of η^m in this power series is seen to be a polynomial in n with coefficients $t_{m,j}$. The explicit form of the $t_{m,j}$'s, after invoking Leibnitz' differentiation rule and after some algebraic manipulations, is found to be

$$t_{m,j} = \frac{1}{m!} \sum_{s=j}^{m} \sum_{r=j}^{s} \binom{m}{s} \left[\left(\frac{1}{w_0^r}\right)^{(s)} B_{r,j}^{(m-s)}\right]_{\eta=0},$$
$$\text{for } m, j = 0, 1, 2, \cdots \text{ and } j \le m$$
$$\text{with } t_{0,0} = 1. \tag{64}$$

It will be recalled that w_0 is defined in Eq. (28) and that the quantities $B_{r,j}$ are defined by Eq. (47) and examples of $B_{r,j}$ are listed in Appendix C.

(ii) The same procedure is used on the denominator: A single power series in η (around $\eta = 0$) is formed from the two factors of the denominator. The coefficients of this power series are again polynomials in n with coefficients

$$Z_{m,j} = \sum_{k=0}^{m} \sum_{s=\max(0,j-k)}^{s=\min(j,m-k)} X_{k,j-s}\, Y_{m-k,s},$$
$$\text{for } m, j = 0, 1, 2, \circ\circ\circ \text{ and } j \le m$$
$$\text{with } Z_{0,0} = 1. \tag{65}$$

Here the $X_{m,j}$'s in Eq. (65) come from the first factor in the denominator of (57) and obey the recursion relation

$$X_{m,j} = \frac{1}{m} \sum_{k=1}^{m} a_k [k X_{m-k,j-1} + (2k - m) X_{m-k,j}],$$
$$\text{for } m, j = 0, 1, 2, \circ\circ\circ \text{ and } j \le m$$
$$\text{with } X_{0,0} = 1 \text{ and } X_{m,j} = 0 \text{ for } j > m. \tag{66}$$

Similarly, the quantities $Y_{m,j}$ that appear in (65) come from an expansion of the second term of the denominator of (57) and obey the recursion relation

$$Y_{m,j} = \frac{1}{m} \sum_{k=1}^{m} b_k [k Y_{m-k,j-1} + (k - m) Y_{m-k,j}],$$
$$m, j = 0, 1, 2, \circ\circ\circ \text{ and } j \le m \tag{67}$$
$$\text{with } Y_{0,0} = 1 \text{ and } Y_{m,j} = 0 \text{ for } j > m.$$

(iii) In consequence of the manipulations described in the previous two steps, $h(\eta)$ in Eq. (57) can be written as the ratio of two power series in η. Evaluating the mth derivative of $h(\eta)$ with respect to η at $\eta = 0$ we are led, after some algebraic manipulation to the polynomial form of $h_{(0)}^{(m)}$ given in Eq. (63), with the coefficients $C_{m,j}$ determined by a recursion relation of the form

$$C_{m,j} = t_{m,j} - \sum_{k=1}^{m} \sum_{s=\max(0,j-k)}^{s=\min(m-k,j)} C_{m-k,s} Z_{k,j-s},$$
$$\text{for } m, j = 0, 1, 2, \cdots, \text{ where } j \le m$$
$$\text{with } C_{0,0} = 1 \text{ and } C_{m,j} = 0 \text{ if } m < j. \tag{68}$$

Equation (68), in combination with Eqs. (64)–(67), has been programmed in FORMAC for use on the IBM 360/91. We have by this means obtained algebraic expressions for the coefficients $C_{m,j}$ as functions of nuclear mass A, from $m, j = 0$ to $m, j = 5$. A partial listing of the results is presented in Appendix D.[17]

Substituting now Eq. (63) in (60), we can write

$$\text{Res} = \frac{1}{\gamma_0'(0)} \frac{\tau^n}{n!} \sum_{m=0}^{n} \frac{n!}{(n-m)!} \frac{1}{u^m} \sum_{j=0}^{m} n^j C_{m,j} \tag{69}$$

and, by Eq. (50), we find the expression for the spatial moments of the scalar flux, valid for large lethargies, as

$$M_{2n}(u) = (2n)!\, \frac{1}{\gamma_1'(0)} \frac{\tau^n}{n!} \sum_{m=0}^{n} \frac{n!}{(n-m)!} \frac{1}{u^m} \sum_{j=0}^{m} n^j C_{m,j}. \tag{70}$$

It should be emphasized again that the coefficients $C_{m,j}$ are independent of n and are functions only of the running indices m and j and of the nuclear mass. Suppose, for example, it is desired to calculate correction terms through u^{-5} consistently for *all* even moments. Then the necessary coefficients $C_{m,j}$ need be obtained through the analysis of only the even spatial moments through M_{10}. The results will be valid nonetheless for all higher moments to the given order of approximation.

4. CONCLUSION

The form of the lethargy-dependent spatial moments of the scalar flux, $M_{2n}(u)$ given in Eq. (70) is similar to that of the results obtained by Placzek,[5] and our $C_{m,j}$ correspond to his $A_{m,j}$. Placzek explicitly calculated $A_{0,0}$, $A_{1,0}$, and $A_{1,1}$ and also obtained the special relation

$$A_{m,m} = (1/m!)(A_{1,1})^m. \tag{71}$$

As expected, our values agree with his for the above three coefficients and Eq. (68) can be simplified for the particular case $j = m$ to recover Placzek's expression for $A_{m,m}$. It would be very difficult to carry Placzek's method beyond correction terms in u^{-2}; the technique described here can be used in a straightforward, albeit lengthy, procedure to go to any desired order of correction. It does not seem likely, however, that significant useful information could be obtained from corrections beyond u^{-5} and therefore only coefficients through this order were explicitly obtained. It may be noted that the computer time on the IBM 360/91 needed to calculate the coefficients presented here was about seven minutes.

The results for $M_{2n}(u)$, Eq. (70) and Appendix D may be used to explore analytically and numerically the

range of validity in n-u space of the simple age theory. Useful as these conclusions might be, it would be even more desirable to analyze the convergence to age theory in $x-u$ space.

It is therefore intended to present in a subsequent paper results for a direct evaluation of the scalar flux $\Psi_0^G(x,u)$ carried to corresponding correction terms relative to age theory.

ACKNOWLEDGMENTS

The authors wish to acknowledge gratefully the assistance of Mr. Waldo Patton in guiding them through the intricacies of FORMAC and in devising some generalizations of the procedures of combinatorial analysis. In particular, Eqs. (42)—(46) resulted from his suggestions.

APPENDIX A

Values of $m_{2n}(\eta)$ for $n=0,1,2,$ etc, as calculated with the help of FORMAC, from Eq. (40).

$m_0=2/\gamma_0(\eta)$ [see Eq. (37)]

$m_2=m_0w_0$

$m_4=m_0(w_1w_0+w_0^2)$

$m_6=m_0(w_2w_1w_0+w_1^2w_0+2w_1w_0^2+w_0^3)$

$m_8=m_0(w_3w_2w_1w_0+w_2^2w_1w_0+2w_2w_1^2w_0+w_1^3w_0+2w_2w_1w_0^2+3w_1^2w_0^2+3w_1w_0^3+w_0^4)$

$$m_{10}=m_0(w_4w_3w_2w_1w_0+w_3{}^2w_2w_1w_0+2w_3w_2^2w_1w_0+w_2^3w_1w_0+2w_3w_2w_1^2w_0+3w_2^2w_1^2w_0+3w_2w_1^3w_0+w_1^4w_0+2w_3w_2w_1w_0^2$$
$$+2w_2^2w_1w_0^2+6w_2w_1^2w_0^2+4w_1^3w_0^2+3w_2w_1w_0^3+6w_1^2w_0^3+4w_1w_0^4+w_0^5)$$

$$m_{12}=m_0(w_5w_4w_3w_2w_1w_0+2w_4w_3^2w_2w_1w_0+2w_4w_3w_2^2w_1w_0+2w_4w_3w_2w_1^2w_0+2w_4w_3w_2w_1w_0^2+w_3^3w_2w_1w_0+3w_3^2w_2^2w_1w_0$$
$$+3w_3w_2^3w_1w_0+w_2^4w_1w_0+2w_3^2w_2w_1^2w_0+6w_3w_2^2w_1^2w_0+4w_2^3w_1^2w_0+3w_3w_2w_1^3w_0+6w_2^2w_1^3w_0+4w_2w_1^4w_0+w_1^5w_0$$
$$+w_4^2w_3w_2w_1w_0+2w_3^2w_2w_1w_0+4w_3w_2w_1w_0^2+2w_2^3w_1w_0^2+6w_3w_2w_1^2w_0^2+9w_2^2w_1^2w_0^2+12w_2w_1^3w_0^3+5w_1^4w_0^2+3w_3w_2w_1w_0^3$$
$$+3w_2^2w_1w_0^3+12w_2w_1^2w_0^3+10w_1^3w_0^3+4w_2w_1w_0^4+10w_1^2w_0^4+5w_1w_0^5+w_0^6).$$

See Ref. 17 of the main text for information on the availability of computer listings for higher terms through m_{18}.

APPENDIX B

First 32 terms in Eq. (40), in order of decreasing powers of w_0, starting with w_0^n.

w_0 power	Multiplied by	Coefficient	w_0 power	Multiplied by	Coefficient
w_0^n	1	1	w_0^{n-5}	w_1^5	$\frac{(n-1)(n-2)(n-3)(n-4)(n-5)}{1\cdot2\cdot3\cdot4\cdot5}$
w_0^{n-1}	w_1	$(n-1)$	w_0^{n-5}	$w_1^4w_2$	$\frac{(n-2)(n-3)(n-4)(n-5)}{1\cdot2\cdot3}$
w_0^{n-2}	w_1^2	$(n-1)(n-2)/2$	w_0^{n-5}	$w_1^3w_2^2$	$(n-3)(n-4)(n-5)$
w_0^{n-2}	w_1w_2	$(n-2)$	w_0^{n-5}	$w_1^3w_2w_3$	$(n-3)(n-4)(n-5)/2$
w_0^{n-3}	w_1^3	$(n-1)(n-2)(n-3)/(1\cdot2\cdot3)$	w_0^{n-5}	$w_1^2w_2^3$	$2(n-4)(n-5)$
w_0^{n-3}	$w_1^2w_2$	$(n-2)(n-3)$	w_0^{n-5}	$w_1^2w_2^2w_3$	$3(n-4)(n-5)$
w_0^{n-3}	$w_1w_2^2$	$(n-3)$	w_0^{n-5}	$w_1^2w_2w_3^2$	$(n-4)(n-5)$
w_0^{n-3}	$w_1w_2w_3$	$(n-3)$	w_0^{n-5}	$w_1^2w_2w_3w_4$	$(n-4)(n-5)$
w_0^{n-4}	w_1^4	$\frac{(n-1)(n-2)(n-3)(n-4)}{1\cdot2\cdot3\cdot4}$	w_0^{n-5}	$w_1w_2^4$	$(n-5)$
w_0^{n-4}	$w_1^3w_2$	$(n-2)(n-3)(n-4)/2$	w_0^{n-5}	$w_1w_2^3w_3$	$3(n-5)$
w_0^{n-4}	$w_1^2w_2^2$	$(n-3)(n-4)3/2$	w_0^{n-5}	$w_1w_2^2w_3^2$	$3(n-5)$
w_0^{n-4}	$w_1^2w_2w_3$	$(n-3)(n-4)$	w_0^{n-5}	$w_1w_2^2w_3w_4$	$2(n-5)$
w_0^{n-4}	$w_1w_2^3$	$(n-4)$	w_0^{n-5}	$w_1w_2w_3^3$	$(n-5)$
w_0^{n-4}	$w_1w_2^2w_3$	$2(n-4)$	w_0^{n-5}	$w_1w_2w_3^2w_4$	$2(n-5)$
w_0^{n-4}	$w_1w_2w_3^2$	$(n-4)$	w_0^{n-5}	$w_1w_2w_3w_4^2$	$(n-5)$
w_0^{n-4}	$w_1w_2w_3w_4$	$(n-4)$	w_0^{n-5}	$w_1w_2w_3w_4w_5$	$(n-5)$

APPENDIX C

The coefficients $B_{r,s}(\eta)$ from $r,s=0$ to $r,s=5$

$$B_{0,0}=1,\quad B_{1,0}=-w_1,\quad B_{1,1}=w_1,$$

$$B_{2,0}=w_1^2-w_1w_2,\quad B_{2,1}=-\tfrac{3}{2}w_1^2+w_1w_2,\quad B_{2,2}=(1/2!)\,w_1^2,$$

$$B_{3,0}=-w_1^3+6w_1^2w_2-3w_1w_2^2-3w_1w_2w_3,$$

$$B_{3,1}=\tfrac{11}{6}w_1^3-5w_1^2w_2+w_1w_2^2+w_1w_2w_3,$$

$$B_{3,2}=-w_1^3+w_1^2w_2,$$

$$B_{3,3}=(1/3!)\,w_1^3,$$

$$B_{4,0}=w_1^4-12w_1^3w_2+18w_1^2w_2^2+12w_1^2w_2w_3-4(w_1w_2^3+2w_1w_2^2w_3+w_1w_2w_3^2+w_1w_2w_3w_4),$$

$$B_{4,1}=-\tfrac{25}{12}w_1^4+13w_1^3w_2-\tfrac{21}{2}w_1^2w_2^2-7w_1^2w_2w_3+w_1w_2^3+2w_1w_2^2w_3+w_1w_2w_3^2+w_1w_2w_3w_4,$$

$$B_{4,2}=(35/4!)\,w_1^4-\tfrac{9}{2}w_1^3w_2+\tfrac{3}{2}w_1^2w_2^2+w_1^2w_2w_3,$$

$$B_{4,3}=-\tfrac{5}{12}w_1^4+\tfrac{1}{2}w_1^3w_2,$$

$$B_{4,4}=(1/4!)\,w_1^4.$$

Let

$$E_1=w_1^3w_2^2+\tfrac{1}{2}w_1^3w_2w_3,$$

$$E_2=2w_1^2w_2^3+3w_1^2w_2^2w_3+w_1^2w_2w_3^2+w_1^2w_2w_3w_4,$$

$$E_3=w_1w_2^4+3w_1w_2^3w_3+3w_1w_2^2w_3^2+2w_1w_2^2w_3w_4+w_1w_2w_3^3+2w_1w_2w_3^2w_4+w_1w_2w_3w_4^2+w_1w_2w_3w_4w_5.$$

Then

$$B_{5,0}=-w_1^5+20w_1^4w_2-60E_1+20E_2-5E_3,$$

$$B_{5,1}=(274/5!)w_1^5-(154/3!)w_1^4w_2+47E_1-9E_2+E_3,$$

$$B_{5,2}=-(225/5!)w_1^5+(71/3!)w_1^4w_2-12E_1+E_2,$$

$$B_{5,3}=(85/5!)w_1^5-(14/3!)w_1^4w_2+E_1,$$

$$B_{5,4}=-(15/5!)w_1^5+(1/3!)w_1^4,$$

$$B_{5,5}=(1/5!)w_1^5.$$

APPENDIX D: LISTING OF COEFFICIENTS $C_{m,j}$ FOR $m,j=0$ TO $m,j=5$

The coefficients $C_{m,j}$ will be given in terms of the quantities γ_i^m defined as:

$$\gamma_i^{(n)}=\frac{d^n}{d\eta^n}\,[\gamma_i(\eta)]_{\eta=0};\quad \begin{matrix} n=1,2,3,\cdots \\ i=0,1,2,3,\cdots \end{matrix}.$$

Since $\gamma_i^{(n)}$, as defined above, are functions of the nuclear mass A only [by virtue of Eqs. (12), (5), and (2)], the coefficients $C_{m,j}$ will themselves be functions of the nuclear mass A only.

$$C_{0,0}=1\ \text{[see also Eq. (68)]},$$

$$C_{1,0}=-\tfrac{4}{5}\gamma_0^{(1)}/\gamma_2-\tfrac{1}{2}\gamma_0^{(2)}/\gamma_0^{(1)},$$

$$C_{1,1}=-\gamma_1^{(1)}/\gamma_1+\tfrac{4}{5}\gamma_0^{(1)}/\gamma_2-\tfrac{1}{2}\gamma_0^{(2)}/\gamma_0^{(1)},$$

$$C_{2,0}=-\frac{216}{175}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2\frac{\gamma_1}{\gamma_3}+\frac{16}{25}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2+\frac{4}{5}\,\frac{\gamma_2^{(1)}\gamma_0^{(1)}}{(\gamma_2)^2}+\frac{1}{4}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^2-\frac{1}{6}\,\frac{\gamma_0^{(3)}}{\gamma_0^{(1)}},$$

$$C_{2,1}=\frac{108}{175}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2\frac{\gamma_1}{\gamma_3}+\frac{1}{2}\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^2+\frac{4}{5}\,\frac{\gamma_1^{(1)}\gamma_0^{(1)}}{\gamma_2\gamma_1}+\frac{1}{2}\,\frac{\gamma_1^{(1)}\gamma_0^{(2)}}{\gamma_1\gamma_0^{(1)}}-\frac{1}{2}\,\frac{\gamma_1^{(1)}}{\gamma_1}-\frac{24}{25}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2-\frac{4}{5}\,\frac{\gamma_2^{(1)}\gamma_0^{(1)}}{(\gamma_2)^2}+\frac{2}{5}\,\frac{\gamma_0^{(2)}}{\gamma_2}$$
$$+\frac{3}{8}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^2-\frac{1}{6}\,\frac{\gamma_0^{(3)}}{\gamma_0^{(1)}},$$

$$C_{2,2}=\frac{1}{2}\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^2-\frac{4}{5}\,\frac{\gamma_1^{(1)}\gamma_0^{(1)}}{\gamma_2\gamma_1}+\frac{1}{2}\,\frac{\gamma_1^{(1)}\gamma_0^{(2)}}{\gamma_1\gamma_0^{(1)}}+\frac{8}{25}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2-\frac{2}{5}\,\frac{\gamma_0^{(2)}}{\gamma_2}+\frac{1}{8}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^2=\frac{1}{2!}\,(C_{1,1})^2,$$

$$C_{3,0}=\frac{432}{175}\,\frac{\gamma_2^{(1)}(\gamma_0^{(1)})^2\gamma_1}{\gamma_3(\gamma_2)^3}+\frac{2592}{875}\,\frac{\gamma_1}{\gamma_3}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3+\frac{216}{175}\,\frac{\gamma_1\gamma_3^{(1)}}{(\gamma_3)^2}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2-\frac{108}{175}\,\frac{\gamma_0^{(2)}\gamma_0^{(1)}\gamma_1}{\gamma_3(\gamma_2)^2}-\frac{8748}{6125}\left(\frac{\gamma_1}{\gamma_3}\right)^2\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3$$
$$-\frac{1728}{1225}\,\frac{(\gamma_0^{(1)})^3(\gamma_1)^2}{\gamma_4(\gamma_3)^2(\gamma_2)^2}-\frac{32}{25}\,\frac{\gamma_2^{(1)}(\gamma_0^{(1)})^2}{(\gamma_2)^3}-\frac{64}{125}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3-\frac{4}{5}\,\frac{(\gamma_2^{(1)})^2\gamma_0^{(1)}}{(\gamma_2)^3}-\frac{216}{175}\,\frac{\gamma_1^{(1)}}{\gamma_3}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2+\frac{8}{25}\,\frac{\gamma_0^{(2)}\gamma_0^{(1)}}{(\gamma_2)^2}$$

$$+\frac{2}{5}\frac{\gamma_2^{(2)}\gamma_0^{(1)}}{(\gamma_2)^2}-\frac{1}{8}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^3+\frac{1}{6}\frac{\gamma_0^{(3)}\gamma_0^{(2)}}{(\gamma_0^{(1)})^2}-\frac{1}{24}\frac{\gamma_0^{(4)}}{\gamma_0^{(1)}},$$

$$C_{3,1}=-\frac{216}{175}\frac{\gamma_2^{(1)}(\gamma_0^{(1)})^2\gamma_1}{\gamma_3(\gamma_2)^3}-\frac{432}{175}\frac{\gamma_1}{\gamma_3}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3-\frac{108}{175}\frac{\gamma_3^{(1)}\gamma_1}{(\gamma_3)^2}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2+\frac{162}{175}\frac{\gamma_0^{(2)}\gamma_0^{(1)}\gamma_1}{\gamma_3(\gamma_2)^2}-\frac{1}{3}\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^3-\frac{2}{5}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^2$$
$$-\frac{1}{4}\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^2+\frac{1}{2}\frac{\gamma_1^{(2)}\gamma_1^{(1)}}{(\gamma_1)^2}-\frac{16}{25}\frac{\gamma_1^{(1)}}{\gamma_1}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2-\frac{4}{5}\frac{\gamma_2^{(1)}\gamma_1^{(1)}\gamma_0^{(1)}}{(\gamma_2)^2\gamma_1}+\frac{2}{5}\frac{\gamma_1^{(2)}\gamma_0^{(1)}}{\gamma_2\gamma_1}-\frac{1}{4}\frac{\gamma_1^{(1)}}{\gamma_1}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^2$$
$$+\frac{1}{4}\frac{\gamma_1^{(2)}\gamma_0^{(2)}}{\gamma_1\gamma_0^{(1)}}+\frac{1}{6}\frac{\gamma_1^{(1)}\gamma_0^{(3)}}{\gamma_1\gamma_0^{(1)}}-\frac{1}{6}\frac{\gamma_1^{(3)}}{\gamma_1}+\frac{2916}{6125}\left(\frac{\gamma_1}{\gamma_3}\right)^2\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3+\frac{576}{1225}\frac{(\gamma_0^{(1)})^3(\gamma_1)^2}{\gamma_4(\gamma_3)^2(\gamma_2)^2}+\frac{48}{25}\frac{\gamma_2^{(1)}(\gamma_0^{(1)})^2}{(\gamma_2)^3}$$
$$+\frac{352}{375}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3+\frac{4}{5}\frac{(\gamma_2^{(1)})^2\gamma_0^{(1)}}{(\gamma_2)^3}+\frac{324}{175}\frac{\gamma_1^{(1)}}{\gamma_3}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2-\frac{4}{5}\frac{\gamma_0^{(2)}\gamma_0^{(1)}}{(\gamma_2)^2}-\frac{2}{5}\frac{\gamma_2^{(2)}\gamma_0^{(1)}}{(\gamma_2)^2}-\frac{2}{5}\frac{\gamma_2^{(1)}\gamma_0^{(2)}}{(\gamma_2)^2}-\frac{1}{10}\frac{(\gamma_0^{(2)})^2}{\gamma_0^{(1)}\gamma_2}$$
$$+\frac{2}{15}\frac{\gamma_0^{(3)}}{\gamma_2}-\frac{11}{48}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^3+\frac{1}{4}\frac{\gamma_0^{(3)}\gamma_0^{(2)}}{(\gamma_0^{(2)})^2}-\frac{1}{24}\frac{\gamma_0^{(4)}}{\gamma_0^{(1)}},$$

$$C_{3,2}=\frac{432}{875}\frac{\gamma_1}{\gamma_3}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3-\frac{54}{175}\frac{\gamma_0^{(2)}\gamma_0^{(1)}\gamma_1}{\gamma_3(\gamma_2)^2}-\frac{1}{2}\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^3-\frac{1}{2}\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\left(\frac{\gamma_1^{(1)}}{\gamma_1}\right)^2+\frac{1}{2}\frac{\gamma_1^{(2)}\gamma_1^{(1)}}{(\gamma_1)^2}+\frac{24}{25}\frac{\gamma_1^{(1)}}{\gamma_1}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2$$
$$+\frac{4}{5}\frac{\gamma_2^{(1)}\gamma_1^{(1)}\gamma_0^{(1)}}{\gamma_1(\gamma_2)^2}-\frac{2}{5}\frac{\gamma_1^{(2)}\gamma_0^{(1)}}{\gamma_2\gamma_1}-\frac{2}{5}\frac{\gamma_1^{(1)}\gamma_0^{(2)}}{\gamma_2\gamma_1}-\frac{3}{8}\frac{\gamma_1^{(1)}}{\gamma_1}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^2+\frac{1}{4}\frac{\gamma_1^{(2)}\gamma_0^{(2)}}{\gamma_1\gamma_0^{(1)}}+\frac{1}{6}\frac{\gamma_1^{(1)}\gamma_0^{(3)}}{\gamma_1\gamma_0^{(1)}}$$
$$-\frac{16}{25}\frac{\gamma_2^{(1)}(\gamma_0^{(1)})^2}{(\gamma_2)^3}-\frac{64}{125}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^3-\frac{108}{175}\frac{\gamma_1^{(1)}}{\gamma_3}\left(\frac{\gamma_0^{(1)}}{\gamma_2}\right)^2+\frac{16}{25}\frac{\gamma_0^{(2)}\gamma_0^{(1)}}{(\gamma_2)^2}+\frac{2}{5}\frac{\gamma_2^{(1)}\gamma_0^{(2)}}{(\gamma_2)^2}-\frac{2}{15}\frac{\gamma_0^{(3)}}{\gamma_2}$$
$$-\frac{1}{8}\left(\frac{\gamma_0^{(2)}}{\gamma_0^{(1)}}\right)^3+\frac{1}{12}\frac{\gamma_0^{(3)}\gamma_0^{(2)}}{(\gamma_0^{(1)})^2},$$

$$C_{3,3}=(1/3!)(C_{1,1})^3.$$

See Ref. 17 of the main text for information on the availability of computer listings for higher terms through $C_{5,5}$.

[1]B. Davison, *Neutron Transport Theory* (Oxford University, London, 1958).
[2]J. J. McInerney, Nucl. Sci. Eng. **22**, 215 (1965).
[3]H. J. Amster and K. C. Chan, Nucl. Sci. Eng. **61**, 388 (1976).
[4]R. E. Marshak, Rev. Mod. Phys. **19**, 185 (1947).
[5]G. Placzek, Manhattan Project Report A-25. Apparently it never was published, but copies can be obtained from Technical Information Center, ERDA, Oak Ridge Tennessee.
[6]L. Huang (private communication).
[7]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Welsey, Reading, MA, 1967).
[8]K. M. Case, Phys. Fluids **16**, 1607 (1973).
[9]E. Inonu, J. Math. Phys. **11**, 568 (1970).
[10]N. I. Akhiezer, *The Classical Moment Problem* (Olyver and Boyd, Edinburgh, 1965).
[11]T. Muir, *A Treatise on the Theory of Determinants* (Longman, Green, London, 1933).
[12]K. M. Case, J. Math. Phys. **15**, 974 (1974).
[13]G. Freud, *Orthogonale Polynome* (Birkhauser Verlag, Basel, 1975).
[14]G. Szego, *Orthogonal Polynomials* (American Mathematical Society, Providence, RI, 1975), 4th ed., Vol. XXIII.
[15]H. S. Wall, *Analytical Theory of Continued Fractions* (Van Nostrand, Princeton, N.J.,1948).
[16]PL/1-FORMAC Symbolic Mathematic Interpreter, IBM Corp. 360-03.3.004 (1971).
[17]A full listing of the computer produced expressions for m_{2n} and C_{mj} (beyond those given in Appendices A and D) will be found in AIP document No. PAPS JMAPA-18-2436-97. Order by PAPS number and journal reference from American Institute of Physics, Physics Auxiliary Publication Service, 335 East 45th Street, New York, NY 10017. The price is $1.50 for each microfiche (98 pages), or $5 for photocopies of up to 30 pages with $0.15 for each additional page over 30 pages. Airmail additional. Make checks payable to the American Institute of Physics. This material also appears in *Current Physics Microfilm*, the monthly microfilm edition of the complete set of journals published by AIP, on the frames immediately following this journal article.
[18]G. C. Wick, Phys. Rev. **75**, 738 (1949).
[19]G. Holte, Ark. Fys. Band **2**, No. 48 (1950).
[20]G. Doetsch, *Guide to the Application of the Laplace and Z-Transforms* (Van Nostrand, Princeton, N. J., 1971).
[21]D. W. Widder, *The Laplace Transform* (Princeton U. P., Princeton, N. J., 1972).

A variational principle for the Gel'fand–Levitan equation and the Korteweg–de Vries equation

M. Kanal and H. E. Moses[a)]

Center for Atmospheric Research, College of Pure and Applied Science, University of Lowell, Lowell, Massachusetts 01845
(Received 15 March 1977)

A functional is constructed from the spectral density used in the general Gel'fand–Levitan equation and an arbitrary function N of two sets of variables. This functional is shown to be an absolute maximum when N satisfies the Gel'fand–Levitan equation. In the case of the Gel'fand–Levitan equation for the one-dimensional and radial Schrödinger equations and certain generalizations, this result can be translated into a theorem about the area under a curve to a given point (x or r), considered as a functional of N. This curve is given by the scattering potential to the given point when the functional takes on its maximum value. The functional may thus be considered a method of obtaining the scattering potential from the spectral data through a variational technique. In the case that the Gel'fand–Levitan equation is that for the one-dimensional Schrödinger equation the results can be interpreted as a theorem about the area to a given point x under the curve given by the solution of the Korteweg–de Vries equation. That is, at a given time t the area under a curve to a given point x, considered as a known functional of N, takes on its maximum value for all x and t when the curve represents the solution of the Korteweg–de Vries equation with appropriate initial conditions.

1. THE VARIATIONAL PRINCIPLE FOR THE FREDHOLM EQUATION

We shall give a general variational principle for the Fredholm equation in which the kernel satisfies a positive definite condition. Our proof follows that of Ref. 1 (p. 318) and is included in the present paper mainly to make the paper more self contained.

Let us consider the Fredholm equation

$$f(y)=g(y)-\int_c^d \sigma(y,\ z)f(z)\,dz, \tag{1}$$

where f is the unknown function, g is the given function, and σ is the kernel. The functions f, g, and σ are complex in general. Though we shall proceed in our proof as though the variables y and z and the constants c and d are one-dimensional, they may actually be generalized to be collections of variables and constants in an obvious way.

We require that σ be Hermitian and satisfy a positive definite condition,

$$\sigma(y,z)=\sigma^*(z,y),$$

$$\int_c^d\int_c^d m^*(y)\sigma(y,z)m(z)dy\,dz+\int_c^d |m(y)|^2\,dy\geq 0, \tag{2}$$

where m is one of a set of complex functions of suitable integrability and the equality holds only for $m(y)\equiv 0$.

It will now be convenient to introduce a more abstract notation. The inner product of two functions m and n will be defined by

$$(m,n)=\int_c^d m^*(y)n(y)\,dy. \tag{3}$$

Likewise, the operator A acting on functions $m(y)$ is given by

$$Am(y)=m(y)+\int_c^d \sigma(y,z)m(z)\,dz. \tag{4}$$

From Eq. (2) the operator A is positive definite,

$$(m,Am)\geq 0, \tag{5}$$

for all functions m, where the equality holds only for $m(y)\equiv 0$.

The Fredholm equation (1) now takes the form

$$Af(y)=g(y). \tag{6}$$

The fundamental theorem for the variational principle is the following:

From a class of suitable functions $\{m(y)\}$, construct the functional

$$F(m)=(m,g)+(g,m)-(m,Am). \tag{7}$$

Then $F(m)$ is an absolute maximum if and only if $m(y)=f(y)$, where f is a solution of the Fredholm equation (1) or (6).

Let us first prove the sufficient condition. We set

$$m(y)=f(y)+n(y). \tag{8}$$

Then

$$F(m)=F(f)-(n,An)$$

$$=(f,Af)-(n,An), \tag{9}$$

where we have used (6) and the Hermitian character of A

$$(m,Ap)=(Am,p), \tag{10}$$

for any two functions $m(y)$ and $p(y)$. The sufficient condition then follows from the positive definite character of A given by Eq. (5).

To show the necessity replace the function $n(y)$ by $\epsilon n(y)$ where ϵ is complex. We do not assume f satisfies Eq. (6) but wish to prove that it must if $F(f)$ is a maximum. Thus we assume

[a)]Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-77-3169.

$$F(f) \geq F(f+\epsilon n) \tag{11}$$

or

$$2\,\mathrm{Re}\epsilon(n, Af-g) + |\epsilon|^2(n, An) \geq 0, \tag{12}$$

for all ϵ and all functions $n(y)$. Let ϵ be real.

Then

$$\mathrm{Re}(n, Af-g) = 0, \tag{13}$$

for, if it were not, then one could pick ϵ so that the left-hand side of Eq. (12) is less than zero, which is a contradiction. Similarly by making ϵ purely imaginary,

$$\mathrm{Im}(n, Af-g) = 0 \tag{14}$$

so that

$$(n, Af-g) = 0 \tag{15}$$

for all functions $n(y)$.

Hence we may take $n = Af - g$ and Eq. (6) follows.

2. APPLICATION TO THE GEL'FAND-LEVITAN EQUATION

We refer to Ref. 2, Part I for a general discussion of the Gel'fand—Levitan equation. We shall use essentially the notation for the one-dimensional Gel'fand—Levitan equation (Parts III and IV of Ref. 2). However, the results hold with obvious modifications for the radial equation (Refs. 3 and 4) and for the three-dimensional equation (Part V of Ref. 2). Recently one of us (H. E. M.) has generalized the one-dimensional Gel'fand—Levitan equation so that H_0 includes part of the scattering potential and the radial equation so that H_0 operates in a more general space than heretofore (Refs. 5 and 6). The variational principle holds for these cases also. Indeed, the variational principle holds for the most general Gel'fand—Levitan equation (defined in terms of "triangularity" conditions on the Gel'fand—Levitan kernel) as discussed in Part I of Ref. 2.

We now proceed to discuss the one-dimensional case as a prototype of the other cases. The Gel'fand—Levitan equation is

$$K(x|y) = -\Omega(x|y) - \int_{-\infty}^{x} K(x|z)\Omega(z|y)\,dy. \tag{16}$$

If x is taken to be a fixed parameter, Eq. (16) is a Fredholm equation of the form of Eq. (1) and (6) with

$$\sigma(x|y) = \Omega(y|x), \quad f(y) = K(x|y), \quad g(y) = -\Omega(x|y), \tag{17}$$
$$c = -\infty, \quad d = x.$$

[In the one-dimensional case all quantities are real. In the general case Ω and K are complex.]

The function $\Omega(x|y)$ satisfies the positive definite condition (2). Hence the variational principle is valid for the Gel'fand—Levitan equation.

We shall now express the variational principle explicitly in terms of a trial kernel $N(x|y)$ which is to approximate the Gel'fand—Levitan kernel $K(x|y)$. Since the Gel'fand—Levitan equation for the one-dimensional and radial equation cases and their generalizations are real, all quantities will be taken as real. Let the functional to be maximized be called $F(N, x)$ to express its dependence on the trial kernel N and on x which is regarded as a parameter. From Eq. (7)

$$F(N,x) = -\int_{-\infty}^{x} N(x|y)\,dy\,[2\Omega(x|y) + N(x|y) + \int_{-\infty}^{x} \Omega(y|z)N(x|z)\,dz]. \tag{18}$$

Let us now evaluate the maximum value of $F(N, x)$ which, by the theorem of the preceding section, is given by $F(K, x)$. From the Gel'fand—Levitan equation (16) and the symmetry of Ω, we obtain the surprising result,

$$F(K,x) = -\int_{-\infty}^{x} K(x|y)\Omega(y|x)\,dy = K(x|x) + \Omega(x|x). \tag{19}$$

Thus we have as our principal result

$$K(x|x) = F(K,x) - \Omega(x|x) \tag{20}$$

or

$$K(x|x) \approx F(N,x) - \Omega(x|x), \tag{21}$$

where the kernel N approximates the Gel'fand—Levitan kernel K. To use the variational principle, one would insert a trial kernel N on the left-hand side of Eq. (21). This kernel would depend on various parameters. One would maximize $F(N, x)$ with respect to these parameters to obtain the best approximation to $K(x|x)$. It is clear that the optimum parameters will, in general, depend on x, since x is a fixed parameter with respect to the variational principle.

It should be noted from Eq. (21) and the Gel'fand—Levitan Eq. (16) that $F(N, x)$ gives the correction to the lowest order approximation to $K(x|x)$ which is the inhomogeneous term for $y = x$ in Eq. (16), namely $-\Omega(x|x)$. From Eq. (16) $K(x|x) \to -\Omega(x|x)$ as $x \to -\infty$. Thus $F(N, x)$ adds a correction for finite x and for $x \to +\infty$.

Since the scattering potential $V(x)$ is given by

$$V(x) = 2\frac{d}{dx}K(x|x), \tag{22}$$

we have

$$\int_{-\infty}^{x} V(x')\,dx' \approx 2[F(N,x) - \Omega(x|x)]. \tag{23}$$

The variational principle can now be stated in terms of the scattering potential. Consider functions $V_N(x)$ defined by the requirement that the area under the curve $y = V_N(x)$ from $-\infty$ to x is given by

$$\int_{-\infty}^{x} V_N(x')\,dx' = 2[F(N,x) - \Omega(x|x)]. \tag{24}$$

That function V_N, for which the area is a maximum for all x, is the scattering potential associated with the function $\Omega(x|y)$.

We shall give a simple example of the use of the variational principle. Let us consider the case where

$$\Omega(x|y) = \eta(x+y)\sinh(x+y), \tag{25}$$

where $\eta(x)$ is the Heaviside function

$$\eta(x) = \begin{cases} 1 & \text{for } x > 0, \\ 0 & \text{for } x < 0. \end{cases} \tag{26}$$

This example has been treated in Part III of Ref. 2 where it is shown that

$$K(x\,|\,x) = -(2)^{1/2}\eta(x)\tanh[(2)^{1/2}x]. \tag{27}$$

As a trial function let us use

$$N(x\,|\,y) = C\eta(x+y). \tag{28}$$

In Eq. (28) C is a constant (actually a function of the parameter x) which is to be found by maximizing $F(N,x)$. The form of the trial function is suggested by the form of $\Omega(x\,|\,y)$ in Eq. (25) but is otherwise very crude. The approximate kernel $K_N(x\,|\,x)$ given by the variational principle Eq. (21), after choosing C appropriately, is

$$K_N(x\,|\,x) = -2\eta(x)\tanh x. \tag{29}$$

Clearly $K(x\,|\,x) \geq K_N(x\,|\,x)$. K_N reproduces the function shape very well and is always of the correct order of magnitude.

The case in which $\Omega(x\,|\,y)$ has the form

$$\Omega(x\,|\,y) = \eta(x+y)g(x+y), \tag{30}$$

where $g(x+y)$ is a real function, and the example of Eq. (25) is a particular case is particularly interesting for the one-dimensional case, for the form (30) is a necessary and sufficient condition for $K(x\,|\,x)$ to vanish for $x<0$, i.e., for the case that $V(x)=0$ for $x<0$. In this case the trial function Eq. (28) is indicated where generally, however, C is a function of y to get a better trial function than for the case C is a constant.

As another example, let us consider the case

$$\Omega(x\,|\,y) = -\frac{A}{2}\eta(x+y)\exp\left(-\frac{A}{2}(x+y)\right) \tag{31}$$

which is also treated in Part III of Ref. 2.

Using as a trial function N as given by Eq. (28), it is found that K_N is exact when the functional is maximized with respect to C,

$$K_N(x\,|\,x) = K(x\,|\,x) = \frac{A}{2}\eta(x). \tag{32}$$

In this case $N(x\,|\,y) = K(x\,|\,y) = (A/2)\eta(x+y)$ because of the theorem.

3. APPLICATION TO THE KORTEWEG-de VRIES EQUATION

The variational principle for the Gel'fand–Levitan equation leads to a variational principle for the Korteweg–de Vries equation. Let us consider the Korteweg–de Vries equation in the form

$$\frac{\partial}{\partial t}V(x,t) - 6V(x,t)\frac{\partial}{\partial x}V(x,t) + \frac{\partial^3}{\partial x^3}V(x,t) = 0. \tag{33}$$

In Ref. 7 a method is given for solving this equation under the condition that $V(x,t)$ dies down sufficiently rapidly as $x \to \pm\infty$.

One gives $V(x,0) = V(x)$ and solves the direct problem of scattering for the one-dimensional Schrödinger equation. Thus one solves for the continuous spectrum eigenfunctions $\psi(x\,|\,k)$

$$\left(-\frac{d^2}{dx^2} + V(x)\right)\psi(x\,|\,k) = k^2\psi(x\,|\,k), \tag{34}$$

under the boundary conditions

$$\begin{aligned} &\lim_{x\to-\infty}\psi(x\,|\,k) = \exp(ikx) + b(k)\exp(-ikx), \\ &\lim_{x\to+\infty}\psi(x\,|\,k) = t(k)\exp(ikx). \end{aligned} \tag{35}$$

The quantities $b(k)$ and $t(k)$ are the reflection and transmission coefficients respectively.

One also solves for the eigenvalues and eigenfunctions of the discrete spectrum,

$$\left(-\frac{d^2}{dx^2} + V(x)\right)\psi_i(x) = E_i\psi_i(x), \tag{36}$$

subject to the boundary conditions

$$\lim_{x\to-\infty}\psi_i(x) = \exp[\kappa_i x] \quad (E_i = -\kappa_i^2). \tag{37}$$

The normalization C_i are then obtained from

$$C_i = \int_{-\infty}^{+\infty}[\psi_i(x)]^2\,dx. \tag{38}$$

We now treat time t as a parameter and construct $\Omega(x\,|\,y)$,

$$\Omega(x\,|\,y) = (2\pi)^{-1}\int_{-\infty}^{+\infty} b(k)\exp[i(-8k^3t - kx - ky)]\,dk + \sum_j \frac{\exp[(\kappa_j x + \kappa_j y - 8\kappa_j^3 t)]}{C_j}. \tag{39}$$

One now solves the Gel'fand–Levitan equation (16) using $\Omega(x\,|\,y)$. Then the solution of Eq. (33) is

$$V(x,t) = 2\frac{d}{dx}K(x\,|\,x). \tag{40}$$

Of course the Gel'fand–Levitan kernel $K(x\,|\,y)$ is a function of the time parameter t.

We define the functional $F(N,x)$ as before (N may now depend on t and F does depend on this parameter). Let $V_N(x,t)$ be defined by

$$\int_{-\infty}^{x} V_N(x',t)\,dx' = 2[F(N,x) - \Omega(x\,|\,x)]. \tag{41}$$

We then have the theorem: The function $V_M(x,t)$ is a solution for all x and t of Eq. (33) if and only if M taken from the kernels N is such that the area given by the left-hand side of Eq. (41) is a maximum for all x and t.

This theorem dispenses with the explicit use of the Gel'fand–Levitan equation, just as Eq. (24) derives the scattering potential without its use. These maximum principles are analogous to Hamilton's principle in mechanics which replaces the differential equations of motion but is equivalent to them.

[1] S. G. Miklin, *Variational Methods in Mathematical Physics*, translated by T. Boddington, edited by L. I. G. Chambers (MacMillan, New York, 1964).
[2] I. Kay and H. E. Moses, Part I, Nuovo Cimento 2, 917 (1955); Part II, Nuovo Cimento 3, 66 (1956); Part III, Nuovo Cimento 3, 276 (1956); Part IV, Nuovo Cimento 5, 230 (1957); Part V, Nuovo Cimento 22, 689 (1961).
[3] I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Trans. 1, 253 (1951).
[4] L. D. Faddeyev, J. Math. Phys. 4, 72 (1963).
[5] H. E. Moses, "A Generalization of the Inverse Scattering Problem for the One-Dimensional Schrödinger Equation" (preprint, 1977).
[6] H. E. Moses, "A Generalization of the Direct and Inverse Problems for the Radial Schrödinger Equation" (in preparation).
[7] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. 19, 1095 (1967).

Integral theory of radiative heat transfer with anisotropic scattering and general boundary conditions

V. C. Boffi and G. Spiga

Laboratorio di Ingegneria Nucleare dell'Università di Bologna, Bologna, Italy
(Received 6 April 1977)

An integral formulation of the theory of planar radiative heat transfer with anisotropic scattering and general boundary conditions is proposed in the paper. Explicit expressions for the relevant kernels are obtained, and the correlations between the structure of these kernels and the properties of the bounding surfaces as well as the interaction between the bounding surfaces themselves are investigated for a variety of physical situations. The lines along which exact and approximate analytical solutions can be constructed are indicated and commented upon.

INTRODUCTION

Radiative transfer is one of the classical and most expanded transport theories. The literature on the subject is so vast that it is not easy to review it consistently without a long and special preparation. In view of this complexity, we restrict ourselves to recall that, for the field of astrophysics, the theory and applications of radiative transfer in participating media are illustrated in the renowned monographs by Chandrasekhar,[1] Kourganoff,[2] Sobolev,[3] and Busbridge.[4] For the engineering applications, ranging from radiative heat transfer problems in scattering, absorbing, emitting media up to the study of the interaction of radiation with other modes of heat transfer, the reader can refer to the equally well-known monographs by Sparrow and Cess,[5] and Özisik.[6] Additional material is, naturally, represented by the numerous papers that—since the appearance of the basic bibliography quoted above—have been published in the various journals and magazines devoted to the subject. Leaving apart the contributions of numerical type, we shall now give a short account of the analytical methods proposed for the solution to the functional equations of radiative transfer. (In this account we shall emphasize the case of radiative heat transfer, that will be the final task of the present work. Heat problems can also be regarded, in fact, as an excellent basis for treating extensions and applications to other problems of radiative transfer.) After the classical methods illustrated in Refs. 1—4 (the H-equation method, the variational method, etc.), then, by referring to the last few years, we recall that analytical solutions to planar radiative heat transfer problems have been systematically obtained by Case's method[7] (see, for instance, the literature quoted in Refs. 6 and 8). Case's method, however, can supply reliable analytical solutions only for ideal problems. For more realistic problems, Case's method discloses, indeed, a rather limited field of applicability. As a typical example of an application of Case's method to radiative heat transfer, we quote the paper by Beach, Özisik, and Siewert,[9] where the case of a linearly anisotropic scattering slab with reflecting boundary is treated through Case's normal mode technique: A semi-analytical solution plus simple approximate analytical solutions are discussed there, and completed by numerical results. A more recent, and even more sophisticated, evolution of the approaches based on Case's method is the one in which one employs a contour integral method rather than the singular eigenfunction Case's formalism.[10,11]

In this paper—on the basis of standard methods of classical mathematical analysis—we succeed in developing a compact, exact, integral theory of the general radiative heat transfer problem that takes place in an anisotropically scattering, absorbing, emitting, nonisothermal, gray medium confined between two plane-parallel bounding surfaces. These surfaces, that are separated by the optical distance $2a$, are diffusely emitting, and diffusely and specularly reflecting surfaces of the general type in the sense that in the boundary conditions we incorporate also the effects of the angular distribution of the radiation diffusely emitted and reflected by the bounding surfaces themselves. We assume the azimuthal symmetry scattering as well as the azimuthal symmetry of the internal source and of the boundary conditions.

The paper consists of two parts.

In Part I, the way according to which the starting linear integro-differential Boltzmann equation governing the problem under consideration can be handled is discussed. The first steps of the theory—leading to the formulation of a general iterative integral scheme for the evaluation of the angular radiation intensity—are made for the case of a general anisotropy. Then, the usual representation of the angular transfer function of the scattered radiation in terms of Legendre polynomials is invoked. As the kernels relevant to the previous iterative integral scheme are now separable, it can thus be shown that the core of all the theory consists of a system of infinite linear integral Fredholm equations for the infinite Legendre moments of the unknown angular radiation intensity. In terms of these Legendre moments not only the angular radiation intensity can be easily built up upon the prescribed general boundary conditions, but also the whole problem under consideration can be fully described. How the theory can be exploited in practice is at last considered upon the assumption of an anisotropy of arbitrary finite order, say $0 \leq L < \infty$ ($L = 0, 1, 2 \cdots$). In this case, in fact, the infinite system of linear integral equations for the relevant Legendre moments reduces to a system of finite order $L + 1$.

In Part II, we carry out some applications of the theory. Rather than going on studying analytical solu-

tions to the general case discussed in Part I, we prefer to analyze—for a variety of physical situations—the correlations between the structure of the relevant kernels and the properties of the bounding surfaces as well as the interaction between the bounding surfaces themselves. Hints are then given for constructing analytical solutions for the cases of both isotropic and linearly anisotropic scattering, and for some interesting specialization of the interaction between the bounding surfaces.

PART I: THEORY

1. THE CASE OF A GENERAL ANISOTROPY

The linear integro-differential Boltzmann equation for radiative transfer which we start with is

$$\mu\frac{\partial I(\tau,\mu)}{\partial\tau}+I(\tau,\mu)=Q(\tau,\mu)+\int_0^{2\pi}d\varphi'\int_{-1}^{1}c(\tau)\Pi(\tau,\mu_0)\times I(\tau,\mu')d\mu' \qquad (1)$$
$$(-a\leq\tau\leq a,\ -1\leq\mu\leq 1),$$

where $I(\tau,\mu)$ is the unknown angular radiation intensity, $Q(\tau,\mu)$ is the internal source, and

$$\mu_0=\mu\mu'+(1-\mu^2)^{1/2}(1-\mu'^2)^{1/2}\cos(\varphi-\varphi') \qquad (1a)$$

is the scattering angle in the laboratory system. The other symbols have the usual meaning. In Eq. (1) we account, however, for two further generalizations, namely, both the albedo $0\leq c\leq 1$ and the angular transfer function $\Pi(\tau,\mu_0)$ of the scattered radiation are taken to be functions of the optical variable τ, the extinction coefficient being not necessarily constant.

The bounding surfaces 1 and 2, positioned at $\tau=-a$ and $\tau=a$, respectively, are kept at uniform temperatures T_1 and T_2. The surfaces are diffusely emitting with emissivities ϵ_1 and ϵ_2, and diffusely and specularly reflecting with reflectivities $\rho_i=\rho_i^d+\rho_i^s$ $(i=1,2)$. In formulating the boundary conditions at these surfaces, we account also for the angular distributions of the diffusely emitted and reflected radiation so that, with

$$\alpha_i=\epsilon_i\sigma T_i^4,\ \beta_i=\rho_i^s,\ \gamma_i=\rho_i^d \quad (i=1,2), \qquad (1b)$$

Eq. (1) will be integrated upon the most general boundary conditions, $\mu\in(0,1)$,

$$I(-a,\mu)=\frac{\alpha_1}{2\pi}h_1(\mu)+\beta_1 I(-a,-\mu)+\gamma_1 g_1(\mu)\times\int_0^1\mu' I(-a,-\mu')\,d\mu', \qquad (2)$$

$$I(a,-\mu)=\frac{\alpha_2}{2\pi}h_2(\mu)+\beta_2 I(a,\mu)+\gamma_2 g_2(\mu)\int_0^1\mu' I(a,\mu')\,d\mu',$$

where the positive functions $h_i(\mu)$ and $g_i(\mu)$ $(i=1,2)$ are normalized to

$$\int_0^1\mu h_i(\mu)\,d\mu=\int_0^1\mu g_i(\mu)\,d\mu=1. \qquad (2')$$

Some special cases treated in the literature can be obtained from the problem described by Eqs. (1) and (2). Thus, for instance, the case $\beta_i=\gamma_i=0$ $(i=1,2)$ has been considered in the allied field of neutron transport theory.[12] When $h_i(\mu)=g_i(\mu)=2$, then the boundary conditions of Eqs. (2) coincide with those usually adopted in the literature for radiative heat transfer problems.[6,9]

Resorting to the Green's function method, it can now be verified that the original integro-differential problem, Eqs. (1) and (2), is equivalent to the integral problem, $\mu\in(0,1)$,

$$I(\tau,\mu)=e^{-(a+\tau)/\mu}I(-a,\mu)+\frac{1}{\mu}\int_{-a}^{\tau}e^{-(\tau-\tau')/\mu}Q^*(\tau',\mu)\,d\tau',$$
$$I(\tau,-\mu)=e^{-(a-\tau)/\mu}I(a,-\mu)+\frac{1}{\mu}\int_{\tau}^{a}e^{(\tau-\tau')/\mu}Q^*(\tau',-\mu)\,d\tau', \qquad (3)$$

where by $Q^*(\tau,\mu)$ we denote the rhs of Eq. (1). For $\mu\to 0$ we get

$$I(\tau,0)=Q^*(\tau,0) \qquad (4)$$

as follows from Eq. (1) itself or by letting $\mu\to 0^+$ from above in either of Eqs. (3). Correspondingly, the functions $I(-a,\mu)$ and $I(a,-\mu)$ are assumed to be appropriately shaped in the neighborhood of $\mu=0^+$. Setting now

$$I_1(\tau,\mu)=I(\tau,\mu),\quad I_2(\tau,\mu)=I(\tau,-\mu), \qquad (5)$$

the pair of Eqs. (3) can be converted into the following system of two linear integral equations for I_1 and I_2:

$$I_1(\tau,\mu)=e^{-(a+\tau)/\mu}I_1(-a,\mu)+\frac{1}{\mu}\int_{-a}^{\tau}e^{-(\tau-\tau')/\mu}Q(\tau',\mu)\,d\tau'+\int_{-a}^{\tau}\int_0^1 k_{11}^*(\tau,\tau';\mu,\mu')I_1(\tau',\mu')\,d\tau'\,d\mu'+\int_{-a}^{\tau}\int_0^1 k_{12}^*(\tau,\tau';\mu,\mu')I_2(\tau',\mu')\,d\tau'\,d\mu',$$
$$I_2(\tau,\mu)=e^{-(a-\tau)/\mu}I_2(a,\mu)+\frac{1}{\mu}\int_{\tau}^{a}e^{-(\tau-\tau')/\mu}Q(\tau',-\mu)\,d\tau'+\int_{\tau}^{a}\int_0^1 k_{21}^*(\tau,\tau';\mu,\mu')I_1(\tau',\mu')\,d\tau'\,d\mu'+\int_{\tau}^{a}\int_0^1 k_{22}^*(\tau,\tau';\mu,\mu')I_2(\tau',\mu')\,d\tau'\,d\mu'. \qquad (6)$$

In Eqs. (6) both the unknowns I_1 and I_2 are defined in the same bi-dimensional domain $D_2=(-a,a)\otimes(0,1)$, and the kernels are given by

$$k_{ij}^*(\tau,\tau';\mu,\mu')=\frac{c(\tau')}{\mu}\exp\left((-1)^i\frac{\tau-\tau'}{\mu}\right)\times k(\tau';(-1)^{i+j}\mu,\mu')\quad(i,j=1,2), \qquad (7)$$

where

$$k(\tau;\mu,\mu')=\int_0^{2\pi}\Pi(\tau,\mu_0)\,d\varphi', \qquad (7a)$$

with

$$k(\tau;-\mu,-\mu')=k(\tau;\mu,\mu'),\quad k(\tau;-\mu,\mu')=k(\tau;\mu,-\mu'). \qquad (7b)$$

We note the following relationships:

$$e^{(\tau-\tau')/\mu}k_{11}^*(\tau,\tau';\mu,\mu')=e^{-(\tau-\tau')/\mu}k_{22}^*(\tau,\tau';\mu,\mu'),$$
$$e^{(\tau-\tau')/\mu}k_{12}^*(\tau,\tau';\mu,\mu')=e^{-(\tau-\tau')/\mu}k_{21}^*(\tau,\tau';\mu,\mu'),$$
$$k_{12}^*(\tau,\tau';-\mu,\mu')=-k_{22}^*(\tau,\tau';\mu,\mu'),$$
$$k_{12}^*(\tau,\tau';\mu,-\mu')=k_{11}^*(\tau,\tau';\mu,\mu'),$$

$$k_{21}^*(\tau,\tau';-\mu,\mu')=-k_{11}^*(\tau,\tau';\mu,\mu'),$$
$$k_{21}^*(\tau,\tau';\mu,-\mu')=k_{22}^*(\tau,\tau';\mu,\mu'),$$

(8)

$$k_{12}^*(\tau,\tau';\mu,-\mu')=-k_{21}^*(\tau,\tau';-\mu,\mu'),$$
$$k_{21}^*(\tau,\tau';\mu,-\mu')=-k_{12}^*(\tau,\tau';-\mu,\mu'),$$
$$k_{12}^*(\tau,\tau';-\mu,-\mu')=-k_{21}^*(\tau,\tau';\mu,\mu'),$$
$$k_{21}^*(\tau,\tau';-\mu,-\mu')=-k_{12}^*(\tau,\tau';\mu,\mu'),$$

which essentially follow from the rotational invariance of the scattering, as expressed by the first of Eqs. (7b). These relationships can be regarded as a generalization of those considered, in a similar context, by Chandrasekhar.[1]

The system of Eqs. (6) can be interpreted as a system of mixed type: Its equations are, in fact, of Volterra's type with respect to the space variable τ, whereas they are of Fredholm's type with respect to the angular variable μ. For an assigned $k(\tau;\mu,\mu')$, the system of Eqs. (6) could be evidently solved by iteration in connection with Eqs. (2). If the source terms in Eqs. (6) were of class $L_2(D_2)$ and the general k_{ij}^* $(i,j=1,2)$ were of class $L_2(D_2\otimes D_2)$, then the iterated solutions should be sought in the same Hilbert space $L_2(D_2)$. We shall not discuss here these requirements that could also be of interest in predicting some appropriate expressions for $k(\tau;\mu,\mu')$. We prefer instead to use for $k(\tau;\mu,\mu')$ the alternative separable expression that holds when the angular transfer function $\Pi(\tau,\mu_0)$ of the scattered radiation is represented in terms of a Legendre polynomial expansion.

2. GENERAL ANISOTROPY IN TERMS OF LEGENDRE POLYNOMIALS

Let us set

$$\Pi(\tau,\mu_0)=\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\,\Pi_l(\tau)P_l(\mu_0), \tag{9}$$

where

$$\Pi_l(\tau)=2\pi\int_{-1}^{1}P_l(\mu_0)\Pi(\tau,\mu_0)\,d\mu_0 \tag{9a}$$

with

$$\Pi_0(\tau)=1,\quad \Pi_1(\tau)=\bar{\mu}_0(\tau), \tag{9b}$$

$P_l(u)$ standing for the lth Legendre polynomial. On the basis of the addition theorem for Legendre polynomials, we then obtain for $k(\tau;\mu,\mu')$ the separable form

$$k(\tau;\mu,\mu')=\sum_{l=0}^{\infty}\frac{2l+1}{2}\,\Pi_l(\tau)P_l(\mu)P_l(\mu'), \tag{10}$$

as follows from Eqs. (1a) and (7a). If we now denote by

$$I_l(\tau)=2\pi\int_{-1}^{1}P_l(\mu)I(\tau,\mu)\,d\mu\quad(l=0,1,\cdots) \tag{11a}$$

the lth Legendre moment of the unknown angular radiation intensity, and in Eqs. (2) we set

$$J_1(\mu)=I(-a,-\mu),\quad J_2(\mu)=I(a,\mu), \tag{11b}$$

then the system of Eqs. (3) becomes

$$I(\tau,\mu)=\sum_{l=0}^{\infty}\frac{2l+1}{2}\,\frac{P_l(\mu)}{\mu}\int_{-a}^{\tau}c(\tau')\Pi_l(\tau')e^{-(\tau-\tau')/\mu}I_l(\tau')\,d\tau'$$
$$+\frac{1}{\mu}\int_{-a}^{\tau}e^{-(\tau-\tau')/\mu}Q(\tau',\mu)d\tau'+\frac{\alpha_1}{2\pi}h_1(\mu)e^{-(a+\tau)/\mu}$$
$$+\beta_1e^{-(a+\tau)/\mu}J_1(\mu)+\gamma_1g_1(\mu)e^{-(a+\tau)/\mu}\int_0^1\mu'J_1(\mu')\,d\mu',$$

(12)

$$I(\tau,-\mu)=\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\,\frac{(-1)^lP_l(\mu)}{\mu}\int_{\tau}^{a}c(\tau)\Pi_l(\tau')e^{(\tau-\tau')/\mu}I_l(\tau')\,d\tau'$$
$$+\frac{1}{\mu}\int_{\tau}^{a}e^{-(\tau-\tau')/\mu}Q(\tau',-\mu)\,d\tau'+\frac{\alpha_2}{2\pi}h_2(\mu)e^{-(a-\tau)/\mu}$$
$$+\beta_2e^{-(a-\tau)/\mu}J_2(\mu)+\gamma_2g_2(\mu)e^{-(a-\tau)/\mu}\int_0^1\mu'J_2(\mu')\,d\mu'.$$

Also this system is of mixed type, but with both Volterra and Fredholm kernels separable. From this system we might calculate the angular radiation intensity $I(\tau,\mu)$ for $-a\le\tau\le+a$, $-1\le\mu\le1$, once the moments $I_l(\tau)$ $(l=0,1,2,\cdots)$ and the $J_i(\mu)$'s $(i=1,2)$ were known. If we now observe that

$$I_l(\tau)=2\pi\int_0^1[I(\tau,\mu)+(-1)^lI(\tau,-\mu)]P_l(\mu)\,d\mu, \tag{12'}$$

multiply the first of Eqs. (12) by $2\pi P_k(\mu)\,d\mu$ and the second by $2\pi(-1)^kP_k(\mu)d\mu$, integrate over $\mu\in(0,1)$, and sum up the resulting equations, we then recognize that the Legendre moments $I_0(\tau)$, $I_1(\tau)$, $\cdots$ are in turn solutions to the system of linear integral equations

$$I_k(\tau)=\sum_{l=0}^{\infty}\frac{2l+1}{2}\int_{-a}^{a}c(\tau')\Pi_l(\tau')\,\mathrm{sgn}(\tau-\tau')^{k+l}K_{kl}(|\tau-\tau'|)$$
$$\times I_l(\tau')\,d\tau'+\sum_{i=1}^{2}\int_0^1H_{ki}(\tau,\mu)J_i(\mu)\,d\mu+F_k(\tau)$$
$$(k=0,1,\cdots). \tag{13}$$

For $J_1(\mu)$ and $J_2(\mu)$ we obtain instead the system

$$J_i(\mu)=\sum_{l=0}\frac{2l+1}{4\pi}\int_{-a}^{a}\tilde{H}_{il}(\mu,\tau)\,I_l(\tau)\,d\tau+\tilde{F}_i(\mu)$$
$$+\beta_{i'}e^{-2a/\mu}J_{i'}(\mu)+\gamma_{i'}g_{i'}(\mu)e^{-2a/\mu}\int_0^1\mu'J_{i'}(\mu')\,d\mu'$$
$$(i'\neq i,\ i,i'=1,2) \tag{14}$$

as follows by setting $\tau=-a$ and $\tau=a$ in the second and in the first of Eqs. (12), respectively. The functions $K_{kl}(\tau)$, $H_{ki}(\tau,\mu)$, $F_k(\tau)$ appearing in Eqs. (13), and $\tilde{H}_{il}(\mu,\tau)$, $\tilde{F}_i(\mu)$ appearing in Eqs. (14) are known functions, and are listed in Appendix A. Equations (13) and (14) all together thus constitute a system of infinite linear integral equations for the unknowns $J_1(\mu)$, $J_2(\mu)$, $I_0(\tau)$, $I_1(\tau)$, $I_2(\tau)$, $\cdots$. However, $J_1(\mu)$ and $J_2(\mu)$ can be eliminated, and therefore a full description of the problem is obtained in terms of only the moments $I_0(\tau)$, $I_1(\tau)$, $I_2(\tau)$, $\cdots$ of the unknown angular radiation intensity. For this purpose we take advantage of the separability of the kernels in Eqs. (14). We set

$$\eta_i=\int_0^1\mu J_i(\mu)\,d\mu\quad(i=1,2) \tag{15}$$

so that from Eqs. (14) we extract $(i=1,2)$

$$J_i(\mu)=(1-\beta_1\beta_2e^{-4a/\mu})^{-1}$$
$$\times\sum_{j=1}^{2}\left\{\Delta_{i,3-j}^{(j)}(\mu)\tilde{F}_j(\mu)+\Delta_{i,j}^{(3-j)}\eta_j\gamma_jg_j(\mu)e^{-2a/\mu}\right.$$
$$\left.+\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\Delta_{i,3-j}^{(j)}(\mu)\int_{-a}^{a}\tilde{H}_{jl}(\mu,\tau)I_l(\tau)\,d\tau\right\} \tag{16}$$

in which we use

$$\Delta_{i,l}^{(j)}(\mu)=1-\delta_{il}(1-\beta_j e^{-2a/\mu}) \qquad (i,j,l=1,2). \tag{16'}$$

If now we multiply Eqs. (16) by $\mu d\mu$ and integrate over $\mu\in(0,1)$, we find for η_i $(i=1,2)$ the following algebraic system

$$\left\{1-\int_0^1 \frac{\beta_{i'}\gamma_i g_i(\mu)e^{-4a/\mu}}{1-\beta_1\beta_2 e^{-4a/\mu}}\,d\mu\right\}\eta_i$$
$$=\left\{\int_0^1 \mu\frac{\gamma_{i'}g_{i'}(\mu)e^{-2a/\mu}}{1-\beta_1\beta_2 e^{-4a/\mu}}\,d\mu\right\}\eta_{i'}$$
$$+\sum_{j=1}^{2}\left\{\int_0^1 \mu\frac{\Delta_{i,3-j}^{(j)}(\mu)\tilde{F}_j(\mu)}{1-\beta_1\beta_2 e^{-4a/\mu}}\,d\mu\right.$$
$$\left.+\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\int_0^1 \mu\frac{\Delta_{i,3-j}^{(j)}(\mu)}{1-\beta_1\beta_2 e^{-4a/\mu}}\,d\mu\int_{-a}^{a}\tilde{H}_{jl}(\mu,\tau)I_l(\tau)d\tau\right\}$$
$$(i'\neq i,\ i,\quad i'=1,2). \tag{17}$$

If the determinant D of the coefficients is different from zero, the solution to the system of Eqs. (17) exists and is unique, and can be cast in the form as

$$\eta_i=\delta_i+\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\int_{-a}^{a}\tilde{G}_{il}(\tau)I_l(\tau)\,d\tau \qquad (i=1,2), \tag{18}$$

the constant δ_i and the function $\tilde{G}_{il}(\tau)$ being explicitly given in Appendix B. Using now Eq. (18) in Eq. (16) we obtain

$$J_i(\mu)=j_i(\mu)+\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\int_{-a}^{a}\tilde{K}_{il}(\mu,\tau)I_l(\tau)\,d\tau, \tag{19}$$

where the functions $j_i(\mu)$ and $\tilde{K}_{il}(\mu,\tau)$ are also listed in Appendix B. By means of Eqs. (19), $J_1(\mu)$ and $J_2(\mu)$ can be thus eliminated from the system of Eqs. (13), and we get for the moments $I_0(\tau)$, $I_1(\tau)$, $I_2(\tau),\cdots$ the sought autonomous system

$$I_k(\tau)=\sum_{l=0}^{\infty}\frac{2l+1}{2}\int_{-a}^{a}H_{kl}^{*}(\tau,\tau')I_l(\tau')\,d\tau'+F_k^{*}(\tau)$$
$$(-a\leq\tau,\ \tau'\leq a,\qquad k=0,1,\cdots) \tag{20}$$

with

$$H_{kl}^{*}(\tau,\tau')=H_{kl}^{(d)}(\tau,\tau')+\sum_{i=1}^{2}H_{kl}^{(i)}(\tau,\tau')$$
$$=c(\tau')\Pi_l(\tau')\,\mathrm{sgn}(\tau-\tau')^{k+l}K_{kl}(|\tau-\tau'|)$$
$$+\frac{1}{2\pi}\sum_{i=1}^{2}\int_0^1 H_{ki}(\tau,\mu)\tilde{K}_{il}(\mu,\tau')\,d\mu \tag{21a}$$

and

$$F_k^{*}(\tau)=F_k^{(a,Q)}(\tau)+\sum_{i=1}^{2}F_k^{(i)}(\tau)$$
$$=F_k(\tau)+\sum_{i=1}^{2}\int_0^1 H_{ki}(\tau,\mu)j_i(\mu)\,d\mu. \tag{21b}$$

It is interesting to comment on the structure of the system of Eqs. (20), with particular regard to the role played by the emitting and reflecting properties of the bounding surfaces. As for the general kernel $H_{kl}^{*}(\tau,\tau')$, we observe that if the bounding surface i $(i=1,2)$ were not reflecting, then the corresponding $H_{kl}^{(i)}(\tau,\tau')$ would vanish. Consequently, we may refer to $H_{kl}^{(d)}(\tau,\tau')$ as the direct component of the kernel $H_{kl}^{*}(\tau,\tau')$ inasmuch as it represents the contribution—from $I_l(\tau)$ to $I_k(\tau)$—due to the scattering collision, whereas $H_{kl}^{(i)}(\tau,\tau')$ is the contribution due to the reflections by the bounding surface i. An analogous interpretation holds for the general source term $F_k^{*}(\tau)$ of Eq. (20). It can be, in fact, split into the sum of $F_k^{(a,Q)}(\tau)\equiv F_k(\tau)$, which is the contribution coming directly from both the internal source $Q(\tau,\mu)$ and the emitting properties of the bounding surfaces, plus $\sum_{i=1}^{2}F_k^{(i)}(\tau)$, the general addend of which is just the contribution due to the reflections by the bounding surface i.

3. THE CASE OF A FINITE ORDER ANISOTROPY

The application of the system of Eqs. (20) becomes more practicable if, as usually done, we restrict the infinite sum in the rhs of Eq. (9) [and consequently in the rhs of Eq. (20) itself] at the term with $l=L$, $0\leq L<\infty$, that is, if we refer to the case of an anisotropy of arbitrary finite order L. In this case, the system reduces to a system of $L+1$ linear integral equations in the $L+1$ unknowns $I_0^{(L)}(\tau),\ldots,I_L^{(L)}(\tau)$,

$$I_k^{(L)}(\tau)=\sum_{l=0}^{L}\frac{2l+1}{2}\int_{-a}^{a}H_{kl}^{*}(\tau,\tau')I_l^{(L)}(\tau')\,d\tau'+F_k^{*}(\tau)$$
$$(k=0,1,\ldots,L), \tag{22}$$

where by $I_k^{(L)}(\tau)$ we just denote the general solution associated with the assigned order of anisotropy. (We remark, however, that, for a fixed L, the first $L+1$ moments are still sufficient for the exact determination of the unknown angular radiation intensity.) It could be easily proved that the kernel $H_{kl}^{*}(\tau,\tau')$ and the source $F_k^{*}(\tau)$ of Eq. (22) are elements of $L_2(D_1\otimes D_1)$ and of $L_2(D_1)$, with $D_1=(-a,a)$, respectively. They are, in fact, bounded continuous functions for $-a\leq\tau,\ \tau'\leq a$, except for some singular terms which, however, exhibit only a weak (logarithmic) singularity in their domains of definition. The hypotheses of regularity requested for the functions $h_i(\mu)$ and $g_i(\mu)$ $(i=1,2)$ are, in turn, very weak (summability is sufficient "*ad abundantiam*"), and are broadly satisfied owing to the physical meaning of the functions $h_i(\mu)$ and $g_i(\mu)$ themselves. The system of Eqs. (22)—as well as the preceding one of Eqs. (20)—is then of Fredholm's type, and its general solution $I_k^{(L)}(\tau)$ $(k=0,1,\ldots,L)$ will belong to $L_2(D_1)$, if 1 is not an eigenvalue of the matrix integral operator defined by the $(L+1)^2$ kernels $H_{kl}^{*}(\tau,\tau')$. The problem of showing that for $L\to\infty$ $I_k^{(L)}(\tau)$ tends to the solution $I_k(\tau)\equiv I_k^{(\infty)}(\tau)$ of Eqs. (20) in the norm of the Hilbert space $L_2(D_1)$ should not present particular difficulties once one resorted to the classical projection methods.[13] This problem will be, be, however, the object of a future separate paper. We conclude by underlining an important feature of the theory. Let us take $L=1$. Then, the system of Eqs. (22) reduces to a system of two linear integral equations, whereas for the same linearly anisotropic case, and independently of the quite different meaning of the unknowns, the application of Case's method as discussed in Ref. 9 leads to a system of four linear integral equations. (The situation gets even more complicated in Chandrasekhar's approach, in which, however, a different problem with zero boundary conditions is considered.[1]) Thus the present theory halves the order of the system to be considered, with all the consequent advantages on a practical ground.

PART II: APPLICATIONS

1. THE CASE OF ISOTROPIC SCATTERING

Henceforth we assume that both c and Π do not depend on τ, the internal source $Q(\tau,\mu)$ is zero, and the boundign surfaces are diffusely emitting and reflecting in an isotropic way. Then, in the general theory of Part I,

$$c(\tau),\Pi(\tau,\mu_0),\ \Pi_i(\tau)\Rightarrow c,\ \Pi(\mu_0),\ \Pi_i,\quad Q(\tau,\mu)=0,$$
$$h_i(\mu)=g_i(\mu)=2\qquad(i=1,2). \tag{23}$$

Let us first examine the case of isotropic scattering. With $L=0$, the system of Eqs. (22) reduces to the single linear integral equation

$$I_0(\tau)=\tfrac{1}{2}\int_{-a}^{a}H^*_{00}(\tau,\tau')I_0(\tau')\,d\tau'+F^*_0(\tau) \tag{24}$$

for the zeroth moment $I_0(\tau)\equiv I_0^{(0)}(\tau)$ of the angular radiation intensity. [We omit the superscript (L).] In Eq. (24) we have

$$H^*_{00}(\tau,\tau')=cK_{00}(|\tau-\tau'|)+\frac{1}{2\pi}\sum_{i=1}^{2}\int_0^1 H_{0i}(\tau,\mu)\tilde{K}_{i0}(\mu,\tau')d\mu \tag{25a}$$

for the kernel, and

$$F^*_0(\tau)=F_0(\tau)+\sum_{i=1}^{2}\int_0^1 H_{0i}(\tau,\mu)j_i(\mu)\,d\mu \tag{25b}$$

for the known term. We recognize that $(i,i'=1,2)$:

$$\begin{aligned}
&K_{00}(|\tau|)=E_1(|\tau|),\\
&H_{0i}(\tau,\mu)=2\pi\beta_i e^{-[a-(-1)^i\tau]/\mu}+4\pi\gamma_i E_2[a-(-1)^i\tau]\mu,\\
&\tilde{H}_{i0}(\mu,\tau)=\frac{c}{\mu}\,e^{-[a-(-1)^i\tau]/\mu},\\
&\tilde{F}_i(\mu)=\frac{\alpha_{i'}}{\pi}e^{-2a/\mu}\qquad(i'\neq i),\\
&F_0(\tau)=2\alpha_1E_2(a+\tau)+2\alpha_2E_2(\alpha-\tau).
\end{aligned} \tag{25c}$$

For the kernel of Eq. (24) we get, after some manipulations,

$$\begin{aligned}
H^*_{00}(\tau,\tau')&=H^{(d)}_{00}(\tau,\tau')+\{H^{(\beta)}_{00}(\tau,\tau')+H^{(\beta,\gamma)}_{00}(\tau,\tau')\\
&\quad+H^{(1,2)}_{00}(\tau,\tau')+H^{(2,1)}_{00}(\tau,\tau')\}\\
&=cE_1(|\tau-\tau'|)\\
&\quad+c\Big\{\beta_1\int_0^1\frac{e^{-(2a+\tau+\tau')/\mu}+\beta_2e^{-(4a+\tau-\tau')/\mu}}{\mu(1-\beta_1\beta_2e^{-4a/\mu})}\,d\mu\\
&\quad+\beta_2\int_0^1\frac{e^{-(2a-\tau-\tau')/\mu}+\beta_1e^{-(4a-\tau+\tau')/\mu}}{\mu(1-\beta_1\beta_2e^{-4a/\mu})}\,d\mu\Big\}\\
&\quad+2c\Big\{\gamma_1E_2(a+\tau)\int_0^1\frac{e^{-(a+\tau')/\mu}+\beta_2e^{-(3a-\tau')/\mu}}{1-\beta_1\beta_2e^{-4a/\mu}}\,d\mu\\
&\quad+\gamma_2E_2(a-\tau)\int_0^1\frac{e^{-(a-\tau')/\mu}+\beta_1e^{-(3a+\tau')/\mu}}{1-\beta_1\beta_2e^{-4a/\mu}}\,d\mu\Big\}\\
&\quad+2\gamma_1\tilde{G}_{10}(\tau')\Big\{\beta_2\\
&\quad\times\int_0^1\frac{e^{-(3a-\tau)/\mu}+\beta_1e^{-(5a+\tau)/\mu}+2\gamma_1E_2(a+\tau)\mu e^{-4a/\mu}}{1-\beta_1\beta_2e^{-4a/\mu}}\,d\mu\\
&\quad+2\gamma_2E_2(a-\tau)\int_0^1\frac{\mu e^{-2a/\mu}}{1-\beta_1\beta_2e^{-4a/\mu}}\,d\mu\Big\}\\
&\quad+2\gamma_2\tilde{G}_{20}(\tau')\Big\{\beta_1\\
&\quad\times\int_0^1\frac{e^{-(3a+\tau)/\mu}+\beta_2e^{-(5a-\tau)/\mu}+2\gamma_2E_2(a-\tau)\mu e^{-4a/\mu}}{1-\beta_1\beta_2e^{-4a/\mu}}\,d\mu\\
&\quad+2\gamma_1E_2(a+\tau)\int_0^1\frac{\mu e^{-2a/\mu}}{1-\beta_1\beta_2e^{-4a/\mu}}\,d\mu\Big\}.
\end{aligned} \tag{26}$$

As already commented, if the bounding surfaces are not reflecting $(\beta_i=\gamma_i=0,\ i=1,2)$, the kernel $H^*_{00}(\tau,\tau')$ reduces simply to the direct kernel

$$H^*_{00}(\tau,\tau')\Rightarrow H^{(d)}_{00}(\tau,\tau')=cE_1(|\tau-\tau'|), \tag{26'}$$

which diverges logarithmically as $\tau-\tau'\to 0$. For the other components of the kernel $H^*_{00}(\tau,\tau')$ we observe that $H^{(\beta)}_{00}(\tau,\tau')$ takes into account the quantity of radiation contributed by the specular reflections from the surfaces, whereas $H^{(\beta,\gamma)}_{00}(\tau,\tau')$ refers to the contribution in which only the last reflection is of the diffusive type. $H^{(1,2)}_{00}(\tau,\tau')$ and $H^{(2,1)}_{00}(\tau,\tau')$ represent, at last, the mixed contributions due to the diffusive and specular reflections at the bounding surfaces. For a more comprehensive illustration of the mechanism of interaction between the bounding surfaces, we refer to the next section, where we shall investigate the expressions that the kernel $H^*_{00}(\tau,\tau')$ of Eq. (26) takes in connection with some interesting specialization of the properties of the bounding surfaces under consideration.

2. EXPRESSIONS FOR THE KERNEL $H^*_{00}(\tau,\tau')$, Eq. (26)

(i) A first meaningful example of interaction between the bounding surfaces is the one when

$$\alpha_1\neq 0,\ \beta_1=\gamma_1=0,\ \alpha_2=0,\ \beta_2,\gamma_2\neq 0, \tag{27a}$$

that is, the bounding surface 1 is black and the bounding surface 2 is diffusely and specularly reflecting (the assumption $\alpha_2=0$ is unessential). In this case the kernel $H^*_{00}(\tau,\tau')$ of Eq. (26) becomes

$$\begin{aligned}
H^*_{00}(\tau,\tau')&=cE_1(|\tau-\tau'|)+c\beta_2E_1(2a-\tau-\tau')\\
&\quad+2c\gamma_2E_2(a-\tau)E_2(a-\tau').
\end{aligned} \tag{27b}$$

A first feature of this kernel is that it is a symmetric kernel, that is, $H^*_0(\tau,\tau')=H^*_0(\tau',\tau)$. Other features are the following ones. Beside the direct kernel $cE_1(|\tau-\tau'|)$ we have the kernels $c\beta_2E_1(2a-\tau-\tau')$ and $2c\gamma_2E_2(a-\tau)E_2(a-\tau')$ which are associated with the specular and the diffuse reflectivity, respectively, of the bounding surface 2. The kernel $c\beta_2E_1(2a-\tau-\tau')$—in the argument of which we recognize the sum $(a-\tau')+(a-\tau)$ of the forward and backward optical paths—means that the radiation passes from τ' to τ after a single specular reflection on the surface 2, with attenuation β_2. When both τ and τ' tend to a, this kernel diverges logarithmically. In turn, the kernel $2c\gamma_2E_2(a-\tau)E_2(a-\tau')$—which is separable—means that the radiation passes from τ' to τ after a single diffuse reflection on the surface 2. In this kernel we observe not only the separation between the forward and the backward optical path, but also the regularization due to the circumstance that now the function $E_2(u)$ is a bounded continuous function, tending to 1 as the argument vanishes.

The known term $F^*_0(\tau)$ of Eq. (24) in correspondence

to Eq. (27a) is

$$F_0^*(\tau)=2\alpha_1E_2(a+\tau)+2\beta_2\alpha_1E_2(3a-\tau)+4\gamma_2\alpha_1E_3(2a)E_2(a-\tau), \tag{27c}$$

as follows from Eq. (25b).

In passing to examine other physical situations, we first realize that the explicit knowledge of the different components of the kernel $H_{00}^*(\tau,\tau')$, Eq. (26), requires the evaluation of integrals of the type as

$$\int_0^1 \mu^r \frac{e^{-s/\mu}}{1-\beta_1\beta_2 e^{-4a/\mu}}\,d\mu, \tag{28a}$$

where $r=-1$, 0 and s is, in general, a function of τ, τ', and a. (In the case of anisotropic scattering r would also take the values 1, 2, $\cdots$.) Since $\beta_1\beta_2 e^{-4a/\mu}<1$, the integral of Eq. (28a) can be always performed by series yielding

$$\int_0^1 \mu^r e^{-s/\mu}\sum_{j=0}^{\infty}(\beta_1\beta_2 e^{-4a/\mu})^j\,d\mu=\sum_{j=0}^{\infty}(\beta_1\beta_2)^j E_{r+2}(s+4ja). \tag{28b}$$

The general term of this infinite sum is thus related to the radiation which comes back to one of the two bounding surfaces after having crossed j times the slab ($4a$ being, in fact, the length of each crossing) and having undergone reflection at the opposite surface. The product $\beta_1\beta_2$ is the attenuation connnected with each crossing. On the basis of these results we are now able to process the investigation of the following two other physical situations,

(ii) $\gamma_1=\gamma_2=0$: Equation (26) for $H_{00}^*(\tau,\tau')$, then, reduces to

$$\begin{aligned}H_{00}^*(\tau,\tau')&=H_{00}^{(d)}(\tau,\tau')+H_{00}^{(\beta)}(\tau,\tau')\\&=cE_1(|\tau-\tau'|)\\&\quad+c\Big\{\sum_{j=0}^{\infty}\beta_1^{j+1}\beta_2^{j}E_1(2a+\tau+\tau'+2j2a)\\&\quad+\sum_{j=0}^{\infty}\beta_1^{j}\beta_2^{j+1}E_1(2a-\tau-\tau'+2j2a)\\&\quad+\sum_{j=0}^{\infty}\beta_1^{j+1}\beta_2^{j+1}E_1(2a+\tau-\tau'+(2j+1)2a)\\&\quad+\sum_{j=0}^{\infty}\beta_1^{j+1}\beta_2^{j+1}E_1(2a-\tau+\tau'+(2j+1)2a)\Big\}.\end{aligned} \tag{29a}$$

The first (second) series refers to the radiation which has undergone the first and last reflection from the bounding surface 1 (2), and has traversed the separation distance $2a$ an even number of times, plus the first path $a+\tau'$ ($a-\tau'$), and the last one $a+\tau(a-\tau)$. The third (fourth) series refers, instead, to the case when the first reflection takes place at the surface 2(1) and the last one at the surface 1(2), with the same number of impacts against each of the two surfaces, and an odd number of crossings of the separation distance $2a$, plus the first path $a-\tau'(a+\tau')$ and the last one $a+\tau(a-\tau)$. According to the circumstance that β_1, $\beta_2<1$ and that the function E_1 is rapidly decreasing, it is plausible to expect that good results can be obtained by retaining a finite and small number of terms in the infinite series in the rhs of Eq. (29a). When $\beta_1=\beta_2$, then the kernel of Eq. (29a) is symmetric. Other functions of interest for the case $\gamma_1=\gamma_2=0$ are ($i=1,2$)

$$\begin{aligned}\tilde K_{i0}(\mu,\tau)&=\frac{\sum_{j=1}^{2}\Delta_{i,3-j}^{(j)}(\mu)\tilde H_{j0}(\mu,\tau)}{1-\beta_1\beta_2 e^{-4a/\mu}},\\ j_i(\mu)&=\frac{\sum_{j=1}^{2}\Delta_{i,3-j}^{(j)}(\mu)\tilde F_j(\mu)}{1-\beta_1\beta_2 e^{-4a/\mu}},\end{aligned} \tag{29b}$$

whereas the determinant D of the system of Eqs. (17) is simply 1.

(iii) $\beta_1=\beta_2=0$. In this case the determinant D is $D=1-4\gamma_1\gamma_2E_3^2(2a)>0$ since $E_3(2a)<\frac{1}{2}$, and Eq. (26) for $H_{00}^*(\tau,\tau')$ becomes

$$\begin{aligned}H_{00}^*(\tau,\tau')&=H_{00}^{(d)}(\tau,\tau')+[H_{00}^{(\beta,r)}(\tau,\tau')]_{\beta_i=0}+[H_{00}^{(1,2)}(\tau,\tau')\\&\quad+H_{00}^{(2,1)}(\tau,\tau')]_{\beta_i=0}\\&=cE_1(|\tau-\tau'|)+2c[1-4\gamma_1\gamma_2E_3^2(2a)]^{-1}\\&\quad\times\{\gamma_1E_2(a+\tau')[E_2(a+\tau)+2\gamma_2E_3(2a)E_2(a-\tau)]\\&\quad+\gamma_2E_2(a-\tau')[E_2(a-\tau)+2\gamma_1E_3(2a)E_2(a+\tau)]\}\\&=cE_1(|\tau-\tau'|)+c\Big\{\sum_{j=0}^{\infty}2^{2j+1}\gamma_1^{j+1}\gamma_2^{j}E_3^{2j}(2a)E_2(a+\tau)\\&\quad\times E_2(a+\tau')+\sum_{j=0}^{\infty}2^{2j+1}\gamma_1^{j}\gamma_2^{j+1}E_3^{2j}(2a)E_2(a-\tau)\\&\quad E_2(a-\tau')\\&\quad+\sum_{j=0}^{\infty}2^{2j+2}\gamma_1^{j+1}\gamma_2^{j+1}E_3^{2j+1}(2a)E_2(a-\tau)E_2(a+\tau')\\&\quad+\sum_{j=0}^{\infty}2^{2j+2}\gamma_1^{j+1}\gamma_2^{j+1}E_3^{2j+1}(2a)E_2(a+\tau)E_2(a-\tau')\Big\}.\end{aligned} \tag{30a}$$

The first (second) series in the rhs of Eq. (30a) refers to the case of first and last collision against the bounding surface 1(2). The general jth term, in particular, refers to the case of $2j+1$ diffuse reflections, $j+1$ against the surface 1(2), and j against the surface 2(1), when the radiation has traversed $2j$ times the separation distance $2a$, with attenuation $E_3(2a)$ at each crossing. The third (fourth) series represents the case of first reflection against 1(2) and last one against 2(1), with $2j+2$ impacts, $j+1$ against each surface, and with $2j+1$ crossing from $\tau=-a$ to $\tau=a$. Also here as γ_1, $\gamma_2<1$ and $E_3(2a)<\frac{1}{2}$, all the series considered should be rapidly convergent. This problem is, however, now less urgent than before since one is dealing with numerical geometrical series. We further observe that if $\gamma_1=\gamma_2$, then the kernel $H_{00}^*(\tau,\tau')$, Eq. (30a), is symmetric. We also observe that all the components of the kernels, but the direct one, are separable.

For the known term of Eq. (24) in the case $\beta_1=\beta_2=0$ we get

$$\begin{aligned}F_0^*(\tau)=\{&[2\alpha_1+4\gamma_1\alpha_2E_3(2a)]E_2(a+\tau)\\&+[2\alpha_2+4\gamma_2\alpha_1E_3(2a)]E_2(a-\tau)\}/1-4\gamma_1\gamma_2E_3^2(2a).\end{aligned} \tag{30b}$$

3. THE CASE OF LINEARLY ANISOTROPIC SCATTERING

This case can be discussed as done in the preceding section for the isotropic scattering. We restrict ourselves to reporting the results connected with the case represented by Eq. (27a). With $L=1$, we need therefore the four kernels $H_{00}^*(\tau,\tau')$, $H_{01}^*(\tau,\tau')$, $H_{10}^*(\tau,\tau')$, $H_{11}^*(\tau,\tau')$, as prescribed by Eqs. (22). Now, $H_{00}^*(\tau,\tau')$ is

the same as in Eq. (27b) whereas for the other kernels we find

$$H^*_{01}(\tau,\tau')=c\bar{\mu}_0\{\mathrm{sgn}(\tau-\tau')E_2(|\tau-\tau'|)+\beta_2E_2(2a-\tau-\tau')+2\gamma_2E_2(a-\tau)E_3(a-\tau')\},$$
$$H^*_{10}(\tau,\tau')=c\{\mathrm{sgn}(\tau-\tau')E_2(|\tau-\tau'|)-\beta_2E_2(2a-\tau-\tau')-2\gamma_2E_3(a-\tau)E_2(a-\tau')\},$$
$$H^*_{11}(\tau,\tau')=c\bar{\mu}_0\{E_3(|\tau-\tau'|)-\beta_2E_3(2a-\tau-\tau')-2\gamma_2E_3(a-\tau)E_3(a-\tau')\}. \tag{31a}$$

It is interesting to observe that $H^*_{11}(\tau,\tau')$ is a symmetric kernel like $H^*_{00}(\tau,\tau')$, Eq. (27b). We recall that $\bar{\mu}_0$, the average cosine of scattering in the laboratory system, follows from the second of Eqs. (9b) in the limit of constant $\Pi(\tau,\mu_0)$.

For the source terms $F^*_0(\tau)$ and $F^*_1(\tau)$ required in Eqs. (22) with $L=1$, we note that $F^*_0(\tau)$ is the same as in Eq. (27c), whereas $F^*_1(\tau)$ is given by

$$F^*_1(\tau)=2\alpha_1[E_3(a+\tau)-\beta_2E_3(3a-\tau)-2\gamma_2E_3(2a)E_3(a-\tau)]. \tag{31b}$$

4. HINTS FOR THE SOLUTION TO THE SYSTEM OF EQS. (22)

It would be possible to show that, for a large class of kernels $H^*_{kl}(\tau,\tau')$, the system of Eqs. (22) can be solved by means of a general constructive procedure. In order to focus the essence of this procedure let us refer to the case of isotropic scattering with $\beta_1=\beta_2=0$. In this case the kernel and the known terms of Eq. (24) are given by Eqs. (30a) and (30b), respectively. If, for the sake of simplicity, we also take $\gamma_1=\gamma_2=\gamma$ and $\alpha_1=\alpha_2=\alpha$, Equation (24) explicitly becomes

$$I_0(\tau)=\frac{c}{2}\int_{-a}^{a}E_1(|\tau-\tau'|)I_0(\tau')\,d\tau'+c\gamma\frac{E_2(a+\tau)+2\gamma E_3(2a)E_2(a-\tau)}{1-4\gamma^2E_3^2(2a)}\int_{-a}^{a}E_2(a+\tau')I_0(\tau')d\tau'+c\gamma\frac{E_2(a-\tau)+2\gamma E_3(2a)E_2(a+\tau)}{1-4\gamma^2E_3(2a)}\int_{-a}^{a}E_2(a-\tau')I_0(\tau')d\tau'+2\alpha\frac{E_2(a+\tau)+E_2(a-\tau)}{1-2\gamma E_3(2a)}. \tag{32}$$

We now set

$$E_1(|\tau-\tau'|)=\sum_{n=0}^{\infty}\frac{2n+1}{2a}Q_n(\tau)P_n\left(\frac{\tau'}{a}\right), \tag{33}$$

where

$$Q_n(\tau)=\int_{-a}^{a}E_1(|\tau-\tau'|)P_n\left(\frac{\tau'}{a}\right)d\tau'. \tag{33'}$$

With the positions

$$\xi_n=\int_{-a}^{a}P_n\left(\frac{\tau'}{a}\right)I_0(\tau')\,d\tau',\quad \zeta_+=\int_{-a}^{a}E_2(a+\tau')I_0(\tau')\,d\tau',\quad \zeta_-=\int_{-a}^{a}E_2(a-\tau')I_0(\tau')\,d\tau', \tag{34}$$

we then deduce that the solution to Eq. (32) is of the form

$$I_0(\tau)=\frac{c}{2}\sum_{n=0}^{\infty}\frac{2n+1}{2a}\xi_nQ_n(\tau)+\left[c\gamma\frac{\zeta_++2\gamma E_3(2a)\zeta_-}{1-4\gamma^2E_3^2(2a)}+\frac{2\alpha}{1-2\gamma E_3(2a)}\right]E_2(a+\tau)+\left[c\gamma\frac{\zeta_-+2\gamma E_3(2a)\zeta_+}{1-4\gamma^2E_3^2(2a)}+\frac{2\alpha}{1-2\gamma E_3(2a)}\right]E_2(a-\tau). \tag{35}$$

This form becomes explicit once the infinite constants ζ_+, ζ_-, ξ_0, ξ_1, $\cdots$ are known. They can be determined by adopting the theory of the linear integral equations with separable kernels, as permissible here in force of the functional properties of the kernel $E_1(|\tau-\tau'|)$. Projecting in fact the solution, Eq. (35), over the sequence of the linearly independent functions $E_2(a+\tau)$, $E_2(a-\tau)$, $P_0(\tau/a)$, $P_1(\tau/a)$, $\cdots$, one gets the coupled infinite algebraic system from which the unknown coefficients ζ_+, ζ_-, ξ_0, ξ_1, ξ_2, $\cdots$ of Eq. (35) can be extracted. Practical solutions—the sequence of which converges in the mean of $L_2(D_1)$ to the exact and unique solution of Eq. (32)—can be obtained by restricting the infinite series in the rhs of Eq. (33) to a finite number of terms.

APPENDIX A

In Eq. (13):

$$K_{kl}(|\tau|)=\int_0^1\frac{e^{-|\tau|/\mu}}{\mu}P_k(\mu)P_l(\mu)\,d\mu \tag{A1}$$

with

$$K_{00}(|\tau|)=E_1(|\tau|),\quad K_{01}(|\tau|)=K_{10}(|\tau|)=E_2(|\tau|),\quad K_{11}(|\tau|)=E_3(|\tau|), \tag{A1'}$$

and so on. [$E_n(z)$ is the exponential integral of order n.]

$$H_{ki}(\tau,\mu)=2\pi(-1)^{(1+i)k}\{\beta_iP_k(\mu)e^{-[a-(-1)^i\tau]/\mu}+\gamma_iG_k(\tau)\mu\}\quad(i=1,2) \tag{A2}$$

with

$$G_{ki}(\tau)=\int_0^1P_k(\mu)g_i(\mu)e^{-[a-(-1)^i\tau]/\mu}\,d\mu, \tag{A2'}$$

$$F_k(\tau)=2\pi\int_0^1\frac{P_k(\mu)}{\mu}d\mu\left\{\int_{-a}^{\tau}e^{-(\tau-\tau')/\mu}Q(\tau',\mu)\,d\tau'+(-1)^k\int_{\tau}^{a}e^{(\tau-\tau')/\mu}Q(\tau',-\mu)\,d\tau'+\alpha_1\int_0^1P_k(\mu)e^{-(a+\tau)/\mu}h_1(\mu)\,d\mu+(-1)^k\alpha_2\int_0^1P_k(\mu)\times e^{-(a-\tau)/\mu}h_2(\mu)\,d\mu\right\}\quad(k=0,1,\cdots). \tag{A3}$$

In Eq. (14):

$$\tilde{H}_{il}(\mu,\tau)=(-1)^{il}\frac{P_l(\mu)}{\mu}c(\tau)\Pi_l(\tau)e^{-[a-(-1)^i\tau]/\mu}, \tag{A4}$$

$$\tilde{F}_i(\mu)=\frac{1}{\mu}\int_{-a}^{a}e^{-[a-(-1)^i\tau]/\mu}Q(\tau,(-1)^i\mu)\,d\tau+\frac{\alpha_{i'}}{2\pi}h_{i'}(\mu)e^{-2a/\mu}\quad(i'\neq i,\quad i,i'=1,2). \tag{A5}$$

APPENDIX B

In Eq. (18):

$$\delta_i = \frac{1}{D}\int_0^1 \mu \left\{ \left[1 - \int_0^1 \mu' \frac{\beta_i \gamma_{i'} g_{i'}(\mu') e^{-4a/\mu'}}{1 - \beta_1\beta_2 e^{-4a/\mu'}} d\mu' \right] \times \frac{\sum_{j=1}^{2} \Delta_{i,3-j}^{(j)}(\mu)\tilde{F}_j(\mu)}{1 - \beta_1\beta_2 e^{-4a/\mu}} + \left(\int_0^1 \mu' \frac{\gamma_{i'} g_{i'}(\mu') e^{-2a/\mu}}{1 - \beta_1\beta_2 e^{-4a/\mu}} d\mu' \right) \times \frac{\sum_{j=1}^{2} \Delta_{i,j}^{(j)}(\mu)\tilde{F}_j(\mu)}{1 - \beta_1\beta_2 e^{-4a/\mu}} \right\} d\mu \qquad (i' \neq i, \quad i, i' = 1, 2), \tag{B1}$$

$$\tilde{G}_{il}(\tau) = \frac{1}{D}\left\{ \left[1 - \int_0^1 \mu' \frac{\beta_i \gamma_{i'} g_{i'}(\mu') e^{-4a/\mu'}}{1 - \beta_1\beta_2 e^{-4a/\mu'}} d\mu' \right] \times \int_0^1 \mu \frac{\sum_{j=1}^{2} \Delta_{i,3-j}^{(j)}(\mu)\tilde{H}_{jl}(\mu,\tau)}{1 - \beta_1\beta_2 e^{-4a/\mu}} d\mu + \left(\int_0^1 \mu' \frac{\gamma_{i'} g_{i'}(\mu') e^{-2a/\mu'}}{1 - \beta_1\beta_2 e^{-4a/\mu'}} d\mu' \right) \times \int_0^1 \mu \frac{\sum_{j=1}^{2} \Delta_{i,j}^{(j)}(\mu)\tilde{H}_{jl}(\mu,\tau)}{1 - \beta_1\beta_2 e^{-4a/\mu}} d\mu \right\}$$

$$(i' \neq i, \quad i, i' = 1, 2). \tag{B2}$$

In Eq. (19):

$$j_i(\mu) = (1 - \beta_1\beta_2 e^{-4a/\mu})^{-1} \sum_{j=1}^{2} [\Delta_{i,3-j}^{(j)}(\mu)\tilde{F}_j(\mu) + \Delta_{i,j}^{(3-j)}(\mu)\delta_j \gamma_j g_j(\mu) e^{-2a/\mu}], \tag{B3}$$

$$\tilde{K}_{jl}(\mu,\tau) = (1 - \beta_1\beta_2 e^{-4a/\mu})^{-1} \sum_{j=1}^{2} [\Delta_{i,3-j}^{(j)}(\mu)\tilde{H}_{jl}(\mu,\tau) + \Delta_{i,j}^{(3-j)}(\mu)\tilde{G}_{jl}(\tau)\gamma_j g_j(\mu) e^{-2a/\mu}]. \tag{B4}$$

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford U. P., London, 1950, also Dover, New York, 1960).
[2] V. Kourganoff, *Basic Methods in Transfer Problems* (Oxford U. P., London, 1952, also Dover, New York, 1963).
[3] V. V. Sobolev, *A Treatise on Radiative Transfer* (Van Nostrand, Princeton, New Jersey, 1963).
[4] L. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge U. P., London, 1960).
[5] E. M. Sparrow and R. D. Cess, *Radiation Heat Transfer*, revised edition (Brooks/Cole, Belmont, California, 1970).
[6] N. Necati Özisik, *Radiative Transfer, and Interactions with Conduction and Convention* (Wiley, New York, 1973).
[7] K. M. Case, Ann. Phys. (N. Y.) **9**, 1 1960.
[8] N. J. McCormick and I. Kuscer, Adv. Nucl. Sci. Technol. **7**, 181 (1973).
[9] H. L. Beach, M. M. Özisik, and C. E. Siewert, Int. J. Heat Mass Transfer **14**, 1551 (1971).
[10] R. L. Bowden, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. **17**, 81 (1976).
[11] P. F. Zweifel and E. W. Larsen, J. Math. Phys. **17**, 1812 (1976).
[12] V. C. Boffi, V. G. Molinari, and G. Spiga, "Anisotropy of Scattering and Fission in Neutron Transport Theory," Nucl. Sci. Eng., in press.
[13] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, Oxford, 1964).

On the properties of collision probability integrals in annular geometry. I. Analysis

Michael S. Milgram

Applied Mathematics Branch, Atomic Energy of Canada, Ltd., Chalk River, Ontario, Canada K0J 1J0
(Received 23 May 1977)

Two integrals fundamental to the analysis of the integral transport equation in infinitely long, annular geometry are defined and represented as sums of Meijer's G-function. Transformations useful for thick or thin annuli are derived, limiting forms for voided annuli are recorded, and asymptotic series are investigated. Special results include the identification of probability integrals for infinite cylinders and generalized, associated Bickley-Nayler functions as G-functions with appropriate parameters and variables. The incomplete, generalized, associated Bickley-Nayler functions are defined for annuli of finite length, and represented as sums of G-functions for long and short annuli. Asymptotic forms are given for both integrals, one of which reduces to Sievert's integral. Some illustrative examples important for numerical analysis, evaluation of multiple integrals, summation of series, and the study of the transport equation are reported.

1. INTRODUCTION

In the solution of the integral transport equation, escape and collision type probability integrals are frequently encountered. In the simplest cases, for homogeneous convex bodies of volume V, these integrals are of the form[1]

$$P_0 = \frac{1}{4\pi V}\iint_V \exp(-\Sigma R)\, d\Omega\, dV$$
$$= \frac{1}{4\pi V}\iiint_0^{R_s} \exp(-\Sigma R)\hat{\Omega}\cdot\hat{n}\, d\Omega\, dS\, dR,$$

$\hat{\Omega}\cdot\hat{n}$ being the direction cosine at the body surface and Σ the total macroscopic cross section of the medium. In more general cases, the integrals may involve various trigonometric and other factors in the integrand[2] and extend to complicated geometries, e.g., annular regions.

When considering homogeneous slabs, the integrals reduce to the well-known exponential integral, and so the analytic properties of the probability integral for the slab may be derived from the known properties of this function. In cylindrical geometry the integrals are more complicated and the associated general theory less well developed. For infinitely long homogeneous cylindrical regions the Bickley–Nayler functions[3] $\mathrm{Ki}_n(x)$ appear in the integrand. These functions have been partially studied analytically and adequate numerical representations are known.[4] However, other than expressing them as repeated Bessel function integrals,[4,5] the identification of $\mathrm{Ki}_n(x)$ as special functions in mathematical physics has not been previously established. For the case of the infinite cylinder, certain of the probability integrals have been evaluated explicitly and shown to be related to the sums and products of the modified Bessel functions $I_0(x)$, $K_0(x)$, and their derivatives.[1,2,6,7] However, no general formulation has been offered previously.

In the case of an infinitely long annular region, the probability integrals become incomplete integrals of the homogeneous cylindrical case. The integrals are functions of two variables, the macroscopic cross section Σ and some geometrical variable κ. This is the simplest idealization of the real-life situation in nuclear reactor analysis. Because these integrals have not been well studied, it is difficult to obtain efficient numerical procedures for their calculation. Additionally, the effort required to obtain analytic properties for numerical studies, theoretical modeling and limiting values of the variables, can be quite taxing,[2,7] because no general theory has been established.

It is the purpose of this paper to develop a theory of the generalized probability integrals in annular geometry within the confines of the theory of special functions. Two fundamental forms are studied, and it is shown that these may be identified with Meijer's G function and sums thereof, and hence as generalized hypergeometric functions.[8] From these results properties may be obtained in special cases from the theory of G functions. Manipulations are simplified, limiting and

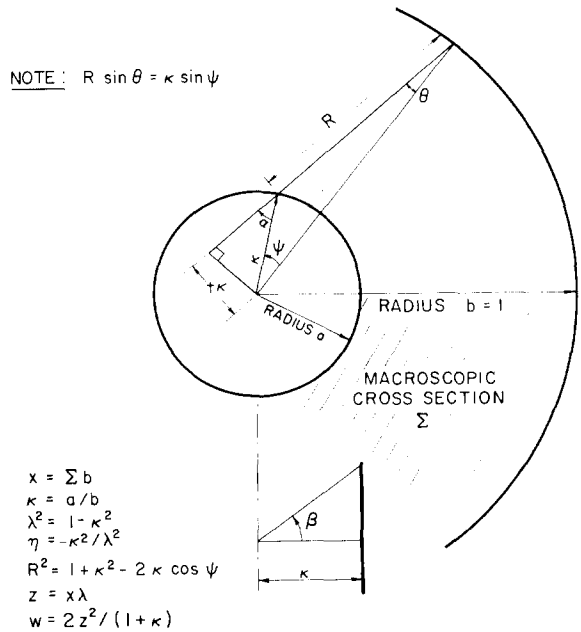


FIG. 1. Illustration of the geometry and notation used throughout the text.

asymptotic forms may be found, and, from the analytic structure, hints may be gleaned as to the best way to obtain efficient numerical approximations.

Sections 3 and 4 of this paper will treat the manipulations necessary to identify the integrals under study as G functions. The variables and parameters are more general than need be for applications, but the price of this is little added complexity. A companion paper[9] considers the problem of developing numerical methods for the evaluation of the integrals for a specific set of parameter values often encountered. In Sec. 5, asymptotic forms are investigated. Associated, generalized Bickley–Nayler functions are defined, identified, and manipulated in Sec. 6, where an analysis is also given for the incomplete integrals of these functions, one of which reduces to Sievert's integral.[10] Some simple applications are presented in Sec. 7.

2. THE FUNDAMENTAL FORMS

The geometry shown in Fig. 1 will generate integrals of the following form when the transport equation is studied; define

$$L^{\tau;\xi}_{\nu;\mu}(x,\kappa) = \int_{\sin^{-1}\kappa}^{\pi/2} d\theta \int_0^{\pi/2} d\beta \sin^{\xi}\theta \cos^{\mu}\theta \times \cos^{\nu-1}\beta \sin^{\tau-1}\beta \exp(-2x\cos\theta/\cos\beta), \quad (2.1)$$

$$^{\epsilon}M^{\tau;\xi}_{\nu;\mu,\delta}(x,\kappa) = \int_0^{\sin^{-1}\kappa} d\theta \int_0^{\pi/2} d\beta \sin^{\mu}\psi \cos^{\delta}\alpha \cos^{\epsilon}\theta \times \sin^{\xi}\theta \cos^{\nu-1}\beta \sin^{\tau-1}\beta \exp[-xR_{-}(\psi)/\cos\beta] \quad (2.2)$$

with

$$R_{-}(\psi) = (1 - 2\kappa\cos\psi + \kappa^2)^{1/2}$$

and $\epsilon = 0$ or 1. The variables x, κ, and angles θ, ψ, α, and β are depicted in Fig. 1 in terms of the geometry and cross section.

In applications, we usually find that $\tau = 1$, $\xi = 0$ and demand that $\mu = m$, $\nu = n$, $\delta = j$ (m,n,j integers); in this sense the following developments are quite general. The restriction $\epsilon = 0$ or 1 is necessary only to eliminate complicated triple sums.

The generalization to arbitrary values of the parameters is the key to the analysis, however, because the case of integral parameters is usually the "exceptional" one, requiring limiting processes and delicate cancellations. The results obtained are usually not amenable to further analytic manipulation so a tacit, but vital tactic used throughout this work is that whenever a relationship is not valid for particular (integral) values of some parameter, *even a summation index*, the offending parameter is assigned an arbitrary (complex) value within a range where the relationship is valid; the parameter is then allowed to approach its desired (integral) value. The justification for this methodology is a well-known part of the process of analytic continuation. For example, the form of Eq. (2.2) requires the conditions $\mathrm{Re}(\xi) > -1$, $\mathrm{Re}(\tau) > 0$, which may be relaxed when a representation valid for other values of these parameters is obtained.

The integrals may also be written in a more familiar form

$$L^{\tau;\xi}_{\nu;\mu}(x,\kappa) = \int_{\sin^{-1}\kappa}^{\pi/2} \cos^{\mu}\theta \sin^{\xi}\theta \, \mathrm{Ki}^{\tau}_{\nu}(2x\cos\theta)\, d\theta \quad (2.3)$$

$$^{\epsilon}M^{\tau;\xi}_{\nu;\mu,\delta}(x,\kappa) = \int_0^{\sin^{-1}\kappa} \sin^{\mu}\psi \cos^{\epsilon}\theta \cos^{\delta}\alpha \sin^{\xi}\theta \times \mathrm{Ki}^{\tau}_{\nu}(x(1+\kappa^2-2\kappa\cos\theta)^{1/2})\, d\theta, \quad (2.4)$$

where we define the associated generalized Bickley–Nayler function

$$\mathrm{Ki}^{\tau}_{\nu}(x) = \int_0^{\pi/2} \cos^{\nu-1}\beta \sin^{\tau-1}\beta \exp(-x/\cos\beta)\, d\beta \quad (2.5)$$

for $\mathrm{Re}(\nu) > 0$, $\mathrm{Re}(\tau) > 0$, degenerating into the usual[3] Bickley–Nayler functions when $\tau = 1$, $\nu = n$, an integer. These functions may be obtained from Eq. (2.2) by the limiting process (for either value of ϵ)

$$\mathrm{Ki}^{\tau}_{\nu}(x) = \lim_{\kappa\to 0} {}^{1}M^{\tau;0}_{\nu;0,0}(x,\kappa)/\kappa. \quad (2.6)$$

In a more applied context, identify[11] the probabilities

$$P^{oo} = (4/\pi) L^{1;0}_{3;1}(x,\kappa) \quad (2.7)$$

and

$$P^{io} = (4/\pi)\, {}^{1}M^{1;0}_{3;0,0}(x,\kappa)/\kappa, \quad (2.8)$$

where P^{oo} is the radial component of the (outer–outer) transmission probability; P^{io} is the radial component of the (inner–outer) transmission probability. In each case, an isotropic bath of particles at the respective entrant surface is assumed. Six other fundamental probabilities for annuli may be obtained[1,12] from conservation of probability arguments and reciprocity relations. A detailed examination is given elsewhere.[9] Note that the (escape probability) P_0 for the infinite homogeneous cylinder is

$$P_0 = (1/2x)[1 - (4/\pi) L^{1;0}_{3;1}(x,0)]$$

in this notation.

3. IDENTIFICATION

The obvious difficulty in analyzing the functions is the existence of the singularities in the integrands of Eq. (2.1) and (2.2). The strategy employed here is to invert the order of the integrals, and express the exponential as a power series, treating $\sec\beta$ as a parameter lying in a range where the θ integration is valid. Afterwards, proceed by analytic continuation, allowing β to lie in the range $[0,\pi/2]$, and evaluate the β integral explicitly. To do this, we must frequently consult various formulae from the theory of functions. The most pertinent of these are reproduced in Appendix A, and cited in the text; others are simply referenced from the literature.

The notation is that of Luke[8] except that (extraneous) semicolons are used in the parameter list of specialized G functions to simplify reading and manipulation. The placement of the semicolons is self-evident. Also, $\sum_l$ means $\sum_{l=0}^{\infty}$ throughout.

In this section the two basic forms will be identified as a sum of particular cases of Meijer's G function, valid for $0 \le \mathrm{Re}(x)$, $|\lambda^2| < 1$. Transformations to other ranges of x and λ^2 may be derived from the definition of the G function. Useful alternative representations are given in Sec. 4.

To facilitate the analysis, let $\cos\theta = \lambda t$ in Eq. (2.1); in Eq. (2.2) let $\sin\theta = t\kappa$, and set $\cos\beta = v$, $\sigma = x/v$ throughout, to obtain the integrals in the form of Euler transforms

$$L^{\tau,\xi}_{\nu,\mu}(x,\kappa)=\lambda^{\mu+1}\int_0^1 dt\int_0^1 dv\, v^{\nu-1}(1-v^2)^{\tau/2-1}\times t^{\mu}(1-\lambda^2t^2)^{(\xi-1)/2}\exp(-2x\lambda t/v) \tag{3.1}$$

$$^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa)=\kappa^{\xi+1}\int_0^1 dt\int_0^1 dv\, v^{\nu-1}(1-v^2)^{\tau/2-1}t^{\xi+\mu}\times(1-t^2)^{\delta/2}\,(1-\kappa^2t^2)^{(\epsilon-1)/2}R_-^{\mu}(t)\times\exp[-xR_-(t)/v] \tag{3.2}$$

with

$$R_-(t)=(1-\kappa^2t^2)^{1/2}-\kappa(1-t^2)^{1/2}$$

and $\epsilon=0$ or 1.

To evaluate $L^{\tau,\xi}_{\nu,\mu}(x,\kappa)$ when $0\le\mathrm{Re}(x)$, $|\lambda^2|<1$ consider the t integral, writing

$$(1-\lambda^2t^2)^{(\xi-1)/2}=\sum_l\frac{(\lambda t)^{2l}\Gamma((1-\xi)/2+l)}{\Gamma((1-\xi)/2)\Gamma(l+1)}={}_1F_0((1-\xi)/2;\lambda^2t^2) \tag{3.3}$$

when $|\lambda t|<1$. After interchanging the integral and the sum, and using Eq. (A1) to evaluate the simplified integral, obtain

$$\int_0^1 dt\,t^{\mu}(1-\lambda^2t^2)^{(\xi-1)/2}\exp(-2\sigma\lambda t)=\frac{1}{2\sqrt{\pi}\,\Gamma((1-\xi)/2)}\sum_l\frac{\lambda^{2l}\Gamma((1-\xi)/2+l)}{\Gamma(l+1)}\times G^{2,1}_{1,3}(\sigma^2\lambda^2\,|\,{}^{(1-\mu)/2-l;}_{0,1/2;-(\mu+1)/2-l}), \tag{3.4}$$

which is a well-defined function for all the variables and parameters when $|\lambda|<1$, particularly for $\sigma=x/v$ in the range of the v integral. Insert Eq. (3.4) into the v integral of Eq. (3.1), and again interchange the order of summation and integration; let $v^2=1/t$ to obtain an integral of the form of Eq. (A3). The final result, setting $z=x\lambda$, is

$$L^{\tau,\xi}_{\nu,\mu}(x,\kappa)=\frac{\lambda^{\mu+1}\Gamma(\tau/2)}{4\sqrt{\pi}\,\Gamma((1-\xi)/2)}\sum_l\frac{\lambda^{2l}\Gamma((1-\xi)/2+l)}{\Gamma(l+1)}\times G^{3,1}_{2,4}\left(z^2\,\middle|\,{}^{(1-\mu)/2-l;\ (\nu+\tau)/2}_{\nu/2,0,\frac12;-(\mu+1)/2-l}\right). \tag{3.5}$$

To evaluate $^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa)$, again for $0\le\mathrm{Re}(x)$, $|\lambda^2|<1$, consider first the integral

$$J^-=\kappa^{\xi+1}\int_0^1 t^{\xi+\mu}(1-\kappa^2t^2)^{(\epsilon-1)/2}(1-t^2)^{\delta/2}[(1-\kappa^2t^2)^{1/2}-\kappa(1-t^2)^{1/2}]^{\mu}\exp\{-\sigma[(1-\kappa^2t^2)^{1/2}-\kappa(1-t^2)^{1/2}]\}dt=\kappa^{\xi+1}\sum_n\frac{(-\sigma)^n}{\Gamma(n+1)}I^{\epsilon}_n(\kappa,\xi,\mu,\delta) \tag{3.6}$$

by expanding $\exp[-\sigma R_-(t)]$ as a power series in $\sigma R_-(t)$. Thence equate

$$I^{\epsilon}_n(\kappa,\xi,\mu,\delta)=\int_0^1 dt\,t^{\xi+\mu}(1-\kappa^2t^2)^{(\epsilon-1)/2}(1-t^2)^{\delta/2}\times[(1-\kappa^2t^2)^{1/2}-\kappa(1-t^2)^{1/2}]^{\mu+n}=\frac{\lambda^{\epsilon+\mu+n-1}\Gamma(\mu+n+1)\Gamma((1+\xi+\mu)/2))}{2^{\mu+n+1}\Gamma((\mu+n+1)/2)\Gamma(1-\epsilon+(\mu+n)/2)}$$

$$\times G^{2,2}_{3,3}\left(-\eta\,\middle|\,{}^{-\delta/2,\,(1+\epsilon-\mu-n)/2;\,(1+\epsilon+\mu+n)/2}_{0,\frac12;-(1+\xi+\mu+\delta)/2}\right) \tag{3.7}$$

from Appendix B where $\eta=-\kappa^2/\lambda^2$. Express the G function as a sum of two ${}_3F_2$'s which can each be expanded as a power series for $|\eta|<1$ and interchanged with the n sum in Eq. (3.6); this sum is then further split into its even and odd components.

After some manipulation (Appendix B) we find

$$J^-=\kappa^{\xi+1}\lambda^{\epsilon+\mu-1}\Gamma((1+\xi+\mu)/2)/2\times\left[\sum_l\frac{(-\eta)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac12+l)\Gamma(l+1)\Gamma((3+\xi+\mu+\delta)/2+l)}\times G^{2,2}_{2,4}\left(\frac{z^2}{4v^2}\,\middle|\,{}^{-\mu/2,\ (1-\mu+\epsilon)/2+l;}_{0,\ \frac12;\ \epsilon-\mu/2,\ (1-\mu-\epsilon)/2+l}\right)-\sum_l\{l\to l+\tfrac12\}\right]. \tag{3.8}$$

The term $\{l\to l+1/2\}$ means the first l sum with l replaced by $l+\frac12$. The G function is defined for all values of v, and so the v integral in Eq. (3.2) can be evaluated, after the l sum and integral are interchanged. The final result is

$$^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa)=\kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\Gamma((1+\xi+\mu)/2)\Gamma(\tau/2)/4\times\left[\sum_l\frac{(-\eta)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac12+l)\Gamma(l+1)\Gamma((3+\xi+\mu+\delta)/2+l)}\times G^{3,2}_{3,5}\left(\frac{z^2}{4}\,\middle|\,{}^{-\mu/2,\ (1-\mu+\epsilon)/2-l;\ (\nu+\tau)/2}_{\nu/2,0,\ \frac12;\ \epsilon-\mu/2,(1-\mu-\epsilon)/2+l}\right)-\sum_l\{l\to l+\tfrac12\}\right]. \tag{3.9}$$

Note added in proof: Note that J^+ of Eq. (38) and Sec. 7 D generates similar expressions for transport integrals in spheres and slabs, cf. G. Bittelli and A. Turrin, Nucl. Sci. Eng. 60, 3, 324 (1976) and this author's comments submitted to Nucl. Sci. Eng. (1977).

Considerable simplification is possible here for particular values of the parameters.

A few points are worth noting about these forms. First, the G functions can be identified as the sum of generalized hypergeometric functions [Eq. (A2)] of the form ${}_1F_2(\cdots)$ or ${}_3F_4(\cdots)$ and hence there will be rapid convergence for a large range of z, because of the form of Eq. (A10). Secondly, the λ^2 sum and the sum in the definition of the ${}_pF_q$'s can be interchanged, to produce a result of the form $\sum_i E_i(\lambda^2)z^i$, where $E_i(\lambda^2)$ is some function only of the geometry of the problem. Since in most practical cases (e.g., nuclear transmutation or slowing-down calculations), the geometry is invariant with only z changing because of changes in Σ, it is sufficient to compute the functions $E_i(\lambda^2)$ only once per problem, and thus obtain efficient and rapid numerical evaluations of the integrals. Thirdly, although Eqs. (3.5) and (3.9) give valid representations for complex values of all variables with the exception that $|\lambda^2|<1$, in the case that $\nu=n$ an integer, we face an exceptional case of the G function. To simplify this form, utilize a well-defined limiting process and obtain both a power series and logarithmic term. This is discussed in more detail in Sec. 7 and elsewhere.[9]

4. VARIATIONS

The results summarized by Eqs. (3.5) and (3.9) are most useful for thin rings ($\lambda^2 \sim 0$) or thick rings ($\eta^2 \sim 0$) respectively. Each representation may be transformed to an expansion valid for thick rings ($\lambda^2 \sim 1$) or thin rings ($\eta^2 \sim \infty$) respectively, by rewriting the intermediate Eqs. (3.3) or (3.7) using known transformation formulas for hypergeometric series or for G functions as illustrated by Eq. (A6). Otherwise, work with the final representation themselves. The technique of transforming sums of G functions has been described elsewhere.[13] In this section we record expansions valid for alternate ranges of λ^2 as well as explicit forms for annuli of small macroscopic cross section (voids).

For the function defined by Eq. (3.5), we use[13,14] the transformation law $\lambda^2 \to 1-\lambda^2$ for an ${}_2F_1$ and eventually obtain

$$L^{\tau,\xi}_{\nu,\mu}(x,\kappa) = \frac{\Gamma(\tau/2)\Gamma((1+\xi)/2)}{4\sqrt{\pi}}\, G^{3,1}_{2,4}\left(x^2 \middle| \begin{matrix}(1-\mu)/2;(\nu+\tau)/2 \\ \nu/2, 0, \tfrac{1}{2}; -(\mu+\xi)/2\end{matrix}\right) - \frac{\lambda^{\mu+1}(1-\lambda^2)^{(1+\xi)/2}\Gamma(\tau/2)\Gamma((1+\xi)/2)}{4\sqrt{\pi}} \times \sum_l \frac{(1-\lambda^2)^l}{\Gamma((3+\xi)/2+l)}\, G^{3,1}_{2,4}\left(z^2 \middle| \begin{matrix}-(\mu+\xi)/2-l;(\nu+\tau)/2 \\ \nu/2, 0, \tfrac{1}{2}; -(\mu+\xi)/2\end{matrix}\right), \tag{4.1}$$

a form most useful for thick annuli ($\lambda^2 \sim 1$).

If $\lambda^2 = 1$, the second term in Eq. (4.1) vanishes and reveals a simple representation of the probability integral for an infinite cylinder. For particular choices of the parameters, this reproduces specialized results given by Murray[7] when the G function is expanded according to Eq. (A2).

From Davison's analysis[6] identify

$$L^{1,0}_{n,0}(x,0) = \tfrac{1}{4}\sqrt{\pi}\lim_{\nu\to n} G^{3,1}_{2,4}\left(x^2 \middle| \begin{matrix}\tfrac{1}{2};(\nu+1)/2 \\ \nu/2, 0, \tfrac{1}{2}; 0\end{matrix}\right) = \tfrac{1}{2}\pi\{A_n(x)I_0(x)K_0(x) + B_n(x)[I_0(x)K_0'(x) + I_0'(x)K_0(x)] + C_n(x)I_0'(x)K_0'(x) + D_n(x)\}, \tag{4.2}$$

where $A_n(x), B_n(x), C_n(x)$, and $D_n(x)$ are polynomials in x, $I_0(x)$ and $K_0(x)$ are modified Bessel functions of the first and second kinds, and the prime denotes differentiation with respect to x. Use Eq. (A2) to unveil a simple power series representation for the right-hand side of this equality, illustrated in Eq. (7.1).

In addition, a comparison of Eqs. (4.1) and (2.1) immediately identifies the second sum in Eq. (4.1) as a representation of the complement of the integral in Eq. (2.1) with limits $(0, \sin^{-1}\kappa)$.

In the case that $\xi = 0$, the intermediate ${}_2F_1$ in the contour integral representation of the G function in Eq. (3.5) admits a quadratic transformation[13,15] of the form $\lambda^2 \to [1-(1-\lambda^2)^{1/2}]/2$. This gives

$$L^{\tau,0}_{\nu,\mu}(x,\kappa) = \frac{\lambda^{\mu+1}\Gamma(\tau/2)}{4\sqrt{\pi}}\left(\frac{1+\kappa}{2}\right)^{-(1+\mu)/2}\sum_l \left(\frac{\kappa-1}{2}\right)^l \frac{1}{\Gamma(l+1)} \times G^{3,2}_{3,5}\left(w \middle| \begin{matrix}(1-\mu)/2, (1-\mu)/2-l;(\nu+\tau)/2 \\ \nu/2, 0, \tfrac{1}{2}; (1-\mu)/2+l, -(1+\mu)/2-l\end{matrix}\right), \tag{4.3}$$

where $w = 2z^2/[1+(1-\lambda^2)^{1/2}]$. This representation is particularly useful for annuli of all sizes, since there are no numerical singularities or cancellations present for $|\lambda^2| \le 1$. For special values of the parameters, further simplification is possible.

Similarily, the equality (3.9), useful for $|\eta| \le 1$, may be transformed to the range $|\eta| \ge 1$ by using Eq. (A2). Intermediate representations worth recording are

$${}^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa) = \kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\Gamma\left(\frac{1+\xi+\mu}{2}\right)\frac{\Gamma(\tau/2)}{(4\pi)} \times \frac{1}{2\pi i}\int_L ds\, \frac{\Gamma(\nu/2-s)\Gamma(-s)\Gamma(\tfrac{1}{2}-s)\Gamma(1+\mu/2+s)}{\Gamma(1-\epsilon+\mu/2+s)\Gamma((\nu+\tau)/2-s)}\left(\frac{z^2}{4}\right)^s \times G^{2,2}_{3,3}\left(-\frac{1}{\eta} \middle| \begin{matrix}1, \tfrac{1}{2};(3+\xi+\mu+\delta)/2 \\ 1+\delta/2, (1+\mu-\epsilon)/2+s;(1-\mu-\epsilon)/2-s\end{matrix}\right), \tag{4.4}$$

where the contour L encloses the real s axis in a negative sense, crossing at $\mathrm{Re}(s) < \min(0, \mathrm{Re}(\nu/2))$. This form is equivalent to

$$\begin{aligned}{}^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa) &= \kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\Gamma\left(\frac{1+\xi+\mu}{2}\right)\frac{\Gamma(\tau/2)}{4}\left[\frac{\sin(\pi(\nu+\tau)/2)}{\sin(\pi\nu/2)}\right. \\ &\times \sum_j \frac{\Gamma(1+\mu/2+j)\Gamma(1-(\nu+\tau)/2+j)(z^2/4)^j H^0_j(\eta)}{\Gamma(1-\nu/2+j)\Gamma(\tfrac{1}{2}+j)\Gamma(1-\epsilon+\mu/2+j)\Gamma(j+1)} \\ &- \frac{\sin\pi[(\nu+\tau-1)/2]}{\sin\pi[(\nu-1)/2]}\left(\frac{z^2}{4}\right)^{1/2} \\ &\times \sum_j \frac{\Gamma((3+\mu)/2+j)\Gamma((3-\nu-\tau)/2+j)(z^2/4)^j H^1_j(\eta)}{\Gamma((3-\nu)/2+j)\Gamma(\tfrac{3}{2}+j)\Gamma((3+\mu)/2-\epsilon+j)\Gamma(j+1)} \\ &- \frac{\sin(\pi\tau/2)}{\sin(\pi\nu/2)\sin\pi[(1-\nu)/2]}\left(\frac{z^2}{2}\right)^{\nu/2} \\ &\times \sum_j \frac{\Gamma(1+(\nu+\mu)/2+j)}{\Gamma(1+\nu/2+j)\Gamma((1+\nu)/2+j)} \\ &\times \left.\frac{\Gamma(1-\tau/2+j)(z^2/4)^j H^\nu_j(\eta)}{\Gamma(1+(\nu+\mu)/2-\epsilon+j)\Gamma(j+1)}\right],\end{aligned} \tag{4.5}$$

where

$$H^\nu_j(\eta) = G^{2,2}_{3,3}\left(-\eta \middle| \begin{matrix}-\delta/2, (1-\mu-\nu+\epsilon)/2-j;(1+\nu+\mu+\epsilon)/2+j \\ 0, \tfrac{1}{2}; -(1+\xi+\mu+\delta)/2\end{matrix}\right). \tag{4.6}$$

Employ Eq. (A6) to arrive at an expansion valid for $|\eta| > 1$. The general form is complicated, but for special values of the parameters[9] useful forms emerge.

The eventuality that $x \approx 0$, as in a voided annulus, is worth some exposition. From Eqs. (3.1) and (3.2) we infer

$$\frac{\partial}{\partial x} L^{\tau;\xi}_{\nu;\mu}(x,\kappa) = -2L^{\tau;\xi}_{\nu-1;\mu+1}(x,\kappa) \tag{4.7}$$

and

$$\frac{\partial}{\partial x} {}^{\epsilon}M^{\tau;\xi}_{\nu;\mu,\delta}(x,\kappa) = -\kappa\, {}^{\epsilon}M^{\tau;\xi-1}_{\nu-1;\mu+1,\delta}(x,\kappa). \tag{4.8}$$

Integrate once,[8] use Eq. (A2) in (3.5) or Eq. (4.5) alone to evaluate the G functions at $x=0$, and, provided that $\mathrm{Re}(\nu)>0$, discover

$$\begin{aligned} L^{\tau;\xi}_{\nu;\mu}(x,\kappa) &= \frac{\lambda^{\mu+1}\Gamma(\tau/2)\Gamma(\nu/2)\Gamma((1+\mu)/2)}{4\Gamma((\nu+\tau)/2)\Gamma((3+\mu)/2)} \\ &\times {}_2F_1\left(\frac{1-\xi}{2}, \frac{1+\mu}{2}; \frac{3+\mu}{2}; \lambda^2\right) \\ &- \frac{x\lambda^{\mu+2}\Gamma(\tau/2)}{4\sqrt{\pi}\,\Gamma((1-\xi)/2)} \sum_l \frac{\lambda^{2l}\Gamma((1-\xi)/2+l)}{\Gamma(l+1)} \\ &\times G^{3,2}_{3,5}\left(z^2 \left| \begin{matrix} \frac{l}{2}, -\mu/2-l; (\nu+\tau-1)/2 \\ (\nu-1)/2, 0, \frac{1}{2}; -1-l-\mu/2, -\frac{1}{2} \end{matrix} \right.\right) \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} &{}^{\epsilon}M^{\tau;\xi}_{\nu;\mu,\delta}(x,\kappa) \\ &= \kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\Gamma\left(\frac{1+\xi+\mu}{2}\right)\frac{\Gamma(\tau/2)}{4} \\ &\times \left\{ \frac{\Gamma(\nu/2)\Gamma(1+\mu/2)}{\Gamma((\nu+\tau)/2)\Gamma(1-\epsilon+\mu/2)\Gamma(\frac{1}{2})} \right. \\ &\times G^{2,2}_{3,3}\left(-\eta \left| \begin{matrix} -\delta/2, (1-\mu+\epsilon)/2; (1+\mu+\epsilon)/2 \\ 0, \frac{1}{2}; -(1+\xi+\mu+\delta)/2 \end{matrix} \right.\right) \\ &- \frac{z}{2}\left[\sum_l \frac{(-\eta)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac{1}{2}+l)\Gamma(1+l)\Gamma((3+\xi+\mu+\delta)/2+l)} \right. \\ &\times G^{3,3}_{4,6}\left(\frac{z^2}{4} \left| \begin{matrix} \frac{1}{2}, -(\mu+1)/2, (\epsilon-\mu)/2-l; (\nu+\tau-1)/2 \\ (\nu-1)/2, 0, \frac{1}{2}; \epsilon-(\mu+1)/2, -(\mu+\epsilon)/2+l, -\frac{1}{2} \end{matrix} \right.\right) \\ &\left.\left. - \sum_l \{l \to l+\tfrac{1}{2}\} \right] \right\}. \end{aligned} \tag{4.10}$$

5. ASYMPTOTIC FORMS

We seek a representation useful for thick or thin annuli with a large interaction cross section and fixed geometry, such that the variable z^2 is large. An asymptotic sum for $L^{\tau;\xi}_{\nu;\mu}(x,\kappa)$ may be found from Eq. (3.5) for all values of ξ, or from Eq. (4.2) for the case $\xi=0$. The method is the same in both cases, and gives the same $x^{-\mu-1}$ dominance. To illustrate, use Eq. (A4) with Eq. (3.5), noting that

$${}_{p+1}F_q\left(\begin{matrix} a_p, 0 \\ b_q \end{matrix} \middle| z^2\right) = 1,$$

to obtain

$$\begin{aligned} L^{\tau;\xi}_{\nu;\mu}(x,\kappa) &= \frac{\Gamma(\tau/2)}{4\sqrt{\pi}\,\Gamma((1-\xi)/2)} \\ &\times \sum_l \left\{ \Gamma((1-\xi)/2+l)\Gamma((\nu+\mu+1)/2+l) \right. \\ &\times \Gamma((1+\mu)/2+l)\Gamma((2+\mu)/2+l) \\ &\times [\Gamma(l+1)\Gamma((\nu+\tau+\mu+1)/2+l)x^{2l+\mu+1}]^{-1} \\ &- \frac{\lambda^{2l+\mu+1}\Gamma((1-\xi)/2+l)}{\Gamma(l+1)} \\ &\left. \times G^{4,0}_{2,4}\left(z^2 \left| \begin{matrix} ; (\nu+\tau)/2, (1-\mu)/2-l \\ \nu/2, 0, \frac{1}{2}, -(1+\mu)/2-l; \end{matrix} \right.\right) \right\}. \end{aligned} \tag{5.1}$$

The function $G^{4,0}_{2,4}(z^2 \mid \cdots)$ is one of the fundamental forms at infinity and has known asymptotics.[16] So, for large values of z^2, split the l sum into two, the first of which S_D is the dominant asymptotic series that gives $x^{-\mu-1}$ behavior, and a second sum S_c which takes the form of an asymptotic correction series

$$\begin{aligned} S_c &\sim \frac{\lambda^{\mu+1}\Gamma(\tau/2)}{4\Gamma((1-\xi)/2)} \exp(-2z) z^{\mu-1-\tau/2} \\ &\times \sum_k \frac{1}{z^k} \sum_l \frac{\lambda^{2l}\Gamma((1-\xi)/2+l)M_k(l)}{\Gamma(l+1)}, \end{aligned} \tag{5.2}$$

where $M_k(l)$ may be determined by recipes given elsewhere[16] and $M_0=1$. The l-sum in the expression for S_c is easily seen to be of the form

$$P_k(\lambda^2)/(1-\lambda^2)^{(\xi-1)/2},$$

where $P_k(\lambda^2)$ is a polynomial in λ^2 of degree k. For a particular set of parameters, the expression (5.1) for S_D reduces to that of Hwang and Toppel,[17] who omit the explicit form of the S_c term.

To obtain the asymptotic form of ${}^{\epsilon}M^{\tau;\xi}_{\nu;\mu,\delta}(x,\kappa)$, a new representation must be derived. Starting from Eq. (3.2), use a simple change of variables to discover

$$\begin{aligned} &{}^{\xi}M^{\tau;\xi}_{\nu;\mu,\delta}(x,\kappa) \\ &= 2^{\xi-2-\xi-\mu-\delta}\kappa^{-\mu-\delta}(1-\kappa)^{2\xi-2}(1+\kappa)^{\xi+\delta+\mu+\xi} \\ &\times \int_1^{1/\sqrt{\gamma}} t^{-3-\xi-\mu-\delta+\xi}(1-\gamma t^2)^{\delta+1}(1+\gamma t^2)^{\xi} \\ &\times [(1-\gamma^2t^2)(t^2-1)]^{(\mu+\xi-1)/2}\mathrm{Ki}^{\tau}_{\nu}(x(1-\kappa)t)\,dt, \end{aligned} \tag{5.3}$$

where $\gamma=(1-\kappa)/(1+\kappa)$. In the following, assume $\xi=\delta+1$, which covers the most interesting cases, and note that the case $\xi=0$, δ arbitrary can be derived in a similar manner. For this section only then, ϵ can have continuous values, emphasized by writing $\epsilon=\xi$, a continuous parameter.

Expand parts of the integral according to[18]

$$(1-\gamma^2t^4)^{\delta+1}(1-\gamma^2t^2)^{(\mu+\xi-1)/2} = \sum_n \gamma^{2n} \sum_{r=0}^{n} g_n(r)t^{2(n+r)}, \tag{5.4}$$

where

$$g_n(r) = \frac{(-)^r\Gamma(2+\delta)\Gamma(n-r-(\mu+\xi-1)/2)}{\Gamma(2+\delta-r)\Gamma((1-\mu-\xi)/2)\Gamma(r+1)\Gamma(n-r+1)} \tag{5.5}$$

split the integral into the difference of two, with limits $(1,\infty)$ and $(1/\sqrt{\gamma},\infty)$, and employ a known integration formula[8] after replacing $\mathrm{Ki}^{\tau}_{\nu}(x(1-\kappa)t)$ by a G function according to Eq. (6.1), to eventually find

$$
{}^{\delta+1}M^{\tau;\xi}_{\nu,\mu,\delta}(x,\kappa)
$$

$$
= 2^{-3-\xi-\mu}\kappa^{-\mu-\delta}(1-\kappa^2)^{2\delta}(1+\kappa)^{\mu+\xi+1}
$$

$$
\times[\Gamma(\tau/2)/\sqrt{\pi}\,]\sum_n \gamma^{2n}\sum_{r=0}^{n} g_n(r)\left[\Gamma\left(\frac{1+\mu+\xi}{2}\right)\right.
$$

$$
\times\, G^{4,0}_{2,4}\left(\frac{z^2\gamma}{4}\,\middle|\,\begin{matrix};\,(\nu+\tau)/2,\,(3+\xi+\mu)/2-n-r\\ 1-n-r,\,\nu/2,\,0,\,\tfrac{1}{2};\end{matrix}\right)
$$

$$
-\gamma^{1-n-r}\sum_l \frac{((1-\mu-\xi)/2)_l}{\Gamma(l+1)}\,\gamma^l
$$

$$
\left.\times G^{4,0}_{2,4}\left(\frac{z^2}{4}\,\middle|\,\begin{matrix};\,(\nu+\tau)/2,\,2+l-n-r\\ 1+l-n-r,\,\nu/2,\,0,\,\tfrac{1}{2};\end{matrix}\right)\right]. \tag{5.6}
$$

In Eq. (5.4) the r sum truncates if $\delta=-1,0,1,\cdots$, and the G functions in Eq. (5.6) have known asymptotic properties[16] for $z\sqrt{\gamma}\to\infty$. The first term dominates if $\gamma\ll 1$, and the coefficient of the leading power of z in the second term vanishes identically, so eventually we discover after summing over r and n

$$
{}^{\delta+1}M^{\tau;\xi}_{\nu,\mu,\delta}(x,\kappa)\sim\Gamma((1+\mu+\xi)/2)\Gamma(\tau/2)4^{\delta-1}(1-\kappa)^{2\delta}\kappa^{(\xi+1-\mu)/2}
$$

$$
\times(2/z\sqrt{\gamma}\,)^{(1+\xi+\mu+\tau)/2}\exp(-z\sqrt{\gamma}\,). \tag{5.7}
$$

For special values of the parameters, this reproduces the results of Hwang and Toppel,[17] who omit the geometric factors in Eq. (5.7).

For large values of z and $\kappa\sim 1$, efficient, approximation may be found from Eq. (4.5). For large values of z and $\kappa\sim 0$, use Eq. (3.9) together with (A4) to transform the G function into forms with known asymptotics, and truncate the l sums.

6. GENERALIZED ASSOCIATED BICKLEY-NAYLER FUNCTIONS

A. Complete form

The generalized, associated Bickley—Nayler functions are defined by Eq. (2.6), which suggests that we evaluate the limit $\eta\to 0$ in Eq. (3.9), with $\xi=\mu=\delta=0$ and $\epsilon=0$ or 1. Only the $l=0$ term in the first sum contributes, and so

$$
\mathrm{Ki}^{\tau}_{\nu}(x)=\frac{\Gamma(\tau/2)}{2\sqrt{\pi}}\,G^{3,0}_{1,3}\left(\frac{x^2}{4}\,\middle|\,\begin{matrix};\,(\nu+\tau)/2\\ \nu/2,\,0,\,\tfrac{1}{2};\end{matrix}\right). \tag{6.1}
$$

From Eq. (4.10) an alternate form is

$$
\mathrm{Ki}^{\tau}_{\nu}(x)=\frac{\Gamma(\tau/2)\Gamma(\nu/2)}{2\Gamma((\nu+\tau)/2)}-\frac{x\Gamma(\tau/2)}{4\sqrt{\pi}}
$$

$$
\times G^{3,1}_{2,4}\left(\frac{x^2}{4}\,\middle|\,\begin{matrix}\tfrac{1}{2};\,(\nu+\tau-1)/2\\ (\nu-1)/2,\,0,\,\tfrac{1}{2};-\tfrac{1}{2}\end{matrix}\right), \tag{6.2}
$$

provided that $\mathrm{Re}(\nu)>0$; applying formula (A2) yields rapidly converging power series near $x\approx 0$, some of which reduce to known results.[3,12] The asymptotic form of $\mathrm{Ki}^{\tau}_{\nu}(x)$ may be easily obtained.[16] The result for $x\to+\infty$ is

$$
\mathrm{Ki}^{\tau}_{\nu}(x)\sim\tfrac{1}{2}\Gamma(\tau/2)\exp(-x)(2/x)^{\tau/2}\sum_k M_k(2/x)^k, \tag{6.3}
$$

where the M_k in the asymptotic series may be found from

$$
M_k=(-\tfrac{1}{2})^k C_k \tag{6.4}
$$

and C_k obeys the three-part recursion formula

$$
2kC_k=(1-k-\tau/2)[(4-2\nu-3k-3\tau/2)C_{k-1}
$$

$$
+(2-k-\tau/2)(2-k-\nu-\tau/2)C_{k-2}],
$$

$$
C_0=1\ \text{ and }\ C_k=0\ \text{ if }\ k<0. \tag{6.5}
$$

An interesting expansion of $\mathrm{Ki}^{\tau}_{\nu}(x)$ in orthogonal functions may also be found from Eq. (2.5). Substitute $\cos\beta=v$, extract a factor e^{-x} from the integrand, set $v=1-t$, and note that the integrand now contains the well-known Hill—Hardy generating function[19] for Laguerre polynomials. Thus

$$
\mathrm{Ki}^{\tau}_{\nu}(x)=2^{\tau/2-1}e^{-x}\sum_n e^{\tau}_n L^{-\nu}_n(x) \tag{6.6}
$$

after the sum and integration are interchanged. Identify the coefficients e^{τ}_n as special cases of the incomplete beta function from its integral representation[8]

$$
e^{\tau}_n=\frac{\Gamma(\tau/2+n)}{\Gamma(1+\tau/2+n)}\,{}_2F_1(1-\tau/2,n+\tau/2;1+n+\tau/2;\tfrac{1}{2}). \tag{6.7}
$$

[The representation of Eq. (6.6) is extremely slow to converge.] Now expand e^{τ}_n as a power series using Eq. (A10), interchange the resulting sums in Eq. (6.6) and recognizing that[20]

$$
\sum_n L^{-\nu}_n(x)/(\tau/2+n+l)=\Gamma(\tau/2+l)\psi(\tau/2+l;1-\nu;x)
$$

find

$$
\mathrm{Ki}^{\tau}_{\nu}(x)=\frac{2^{\tau/2-1}e^{-x}x^{\nu}}{\Gamma(1-\tau/2)}\sum_l\frac{\Gamma(1-\tau/2+l)\Gamma(\tau/2+l)}{\Gamma(l+1)}
$$

$$
\times(\tfrac{1}{2})^l\psi(\tau/2+l+\nu;1+\nu;x), \tag{6.8}
$$

where Kummer's transformation[8] has been applied to the confluent hypergeometric function of the second kind $[\psi(\alpha;\gamma;x)]$. An interesting identity is obtained by comparing Eqs. (6.8) and (6.1), if one notes[8] that

$$
\psi(\alpha;\gamma;x)=\frac{1}{\Gamma(\alpha)\Gamma(1+\alpha-\gamma)}\,G^{2,1}_{1,2}\left(x\,\middle|\,\begin{matrix}1-\alpha;\\ 0,\,1-\gamma;\end{matrix}\right)
$$

$$
=e^{x}G^{2,0}_{1,2}\left(x\,\middle|\,\begin{matrix};\,1-\gamma+\alpha\\ 0,\,1-\gamma;\end{matrix}\right).
$$

B. Incomplete form

For the case of annuli of finite length, we define the incomplete functions $\mathrm{Ki}^{\tau}_{\nu}(x,\alpha)$ by

$$
\mathrm{Ki}^{\tau}_{\nu}(x,\alpha)=\int_0^{\alpha} v^{\nu-1}(1-v^2)^{\tau/2-1}\exp(-x/v)\,dv. \tag{6.9}
$$

These can be manipulated in a manner similar to the analysis of $L^{\tau;\xi}_{\nu,\mu}(x,\kappa)$ in Sections 3 and 4 by using $\sqrt{\pi}\exp(-2\sqrt{y})=G^{2,0}_{0,2}(y\,|\,\begin{matrix}\\ 0,1/2;\end{matrix})$, Eq. (3.3) to expand the integrand, and Eq. (A3) to evaluate the integral. The result is

$$
\mathrm{Ki}^{\tau}_{\nu}(x,\alpha)=\frac{\alpha^{\nu}}{2\sqrt{\pi}\,\Gamma(1-\tau/2)}\sum_l\frac{\alpha^{2l}\Gamma(1-\tau/2+l)}{\Gamma(l+1)}
$$

$$
\times G^{3,0}_{1,3}\left(\frac{x^2}{4\alpha^2}\,\middle|\,\begin{matrix};\,l+1+\nu/2\\ l+\nu/2,\,0,\,\tfrac{1}{2};\end{matrix}\right), \tag{6.10}
$$

which is particularly useful when $\alpha^2 \approx 0$, and $x^2/4\alpha^2 \approx 0$.

If $x^2/4\alpha^2 \to \infty$, the dominant behavior of the G function is known,[16] and the first term of the asymptotic series is

$$\mathrm{Ki}_\nu^\tau(x,\alpha) \sim \frac{\alpha^\nu (1-\alpha^2)^{\tau/2-1}}{2\Gamma(1-\tau/2)} \left(\frac{2\alpha}{x}\right) \exp\left(-\frac{x}{\alpha}\right). \tag{6.11}$$

In analogy to the developments surrounding Eq. (4.1), an expansion useful for $\alpha^2 \approx 1$ is found:

$$\begin{aligned}\mathrm{Ki}_\nu^\tau(x,\alpha) = \mathrm{Ki}_\nu^\tau(x) &- \frac{\alpha^\nu \Gamma(\tau/2)(1-\alpha^2)^{\tau/2}}{2\sqrt{\pi}} \\ &\times \sum_l \frac{(1-\alpha^2)^l}{\Gamma(1+\tau/2+l)} \\ &\times G_{1,3}^{3,0}\left(\frac{x^2}{4\alpha^2} \,\middle|\, \begin{matrix} ; (\nu+\tau)/2 \\ (\nu+\tau)/2+l, 0, \frac{1}{2}; \end{matrix}\right).\end{aligned} \tag{6.12}$$

And in the eventuality that $\tau = 1$, a quadratic transformation[15] useful for $|\alpha^2| < 1$ exists, in analogy to Eq. (4.1):

$$\begin{aligned}\mathrm{Ki}_\nu^1(x,\alpha) = &\frac{1}{2\sqrt{\pi}} \left(\frac{2\alpha^2}{1+(1-\alpha^2)^{1/2}}\right)^{\nu/2} \\ &\times \sum_l \frac{([(1-\alpha^2)^{1/2}-1]/2)^l}{\Gamma(l+1)} \\ &\times G_{2,4}^{4,0}\left(\frac{x^2[1+(1-\alpha^2)^{1/2}]}{4\alpha^2} \,\middle|\, \begin{matrix} ; \nu/2-l, 1+\nu/2+l \\ 0, \frac{1}{2}, \nu/2, \nu/2+l; \end{matrix}\right).\end{aligned} \tag{6.13}$$

In the case $\nu = \tau = 1$, these results give useful formulas for the numerical evaluation of Sieverts integral.[10]

7. APPLICATIONS

In this section, some applications of the representations developed herein are offered. Obviously, without a complete examination of the relevant literature, the list of applications cannot be inclusive. The purpose of presenting these details is to indicate some obvious uses for the theory. More details of only one aspect of these applications are offered elsewhere.[9]

A. Series expansions

For infinitely long cylinders, a useful function[7] introduced in Eq. (4.2) is

$$L_{n,m}^{1,0}(x,0) = \frac{\sqrt{\pi}}{4} \lim_{\nu \to n} G_{2,4}^{3,1}\left(x^2 \,\middle|\, \begin{matrix} (1-m)/2; (1+\nu)/2 \\ 0, \frac{1}{2}, \nu/2; -m/2 \end{matrix}\right);$$

if $n = 2k+1$, then

$$\begin{aligned}&L_{2k+1,m}^{1,0}(x,0) \\ &= \frac{\sqrt{\pi}}{4} \Bigg((-)^k \pi^2 \\ &\quad \times \sum_{j=0}^{k} \frac{(-x^2)^j \Gamma(\frac{1}{2}+m/2+j)}{\Gamma(1+k-j)\Gamma(\frac{1}{2}+j)\Gamma(\frac{1}{2}-k+j)\Gamma(j+1)\Gamma(1+m/2+j)} \\ &\quad - (-)^k x \sum_{j=0}^{k-1} \frac{\Gamma(1+m/2+j)\Gamma(\frac{1}{2}-k+j)\Gamma(k-j)(-x^2)^j}{\Gamma(\frac{3}{2}+j)\Gamma(\frac{3}{2}+m/2+j)\Gamma(j+1)} \\ &\quad - x^{2k+1} \sum_j \frac{\Gamma(1+m/2+j+k)\Gamma(\frac{1}{2}+j)x^{2j}\psi_j^{k,m}(x)}{\Gamma(\frac{3}{2}+k+j)\Gamma(\frac{3}{2}+m/2+k+j)\Gamma(k+j+1)\Gamma(j+1)}\Bigg),\end{aligned} \tag{7.1}$$

where

$$\begin{aligned}\psi_j^{k,m}(x) = &\psi(j+1) + \psi(j+k+1) + \psi(\tfrac{3}{2}+m/2+k+j) \\ &+ \psi(\tfrac{3}{2}+k+j) - \psi(k+j+1+m/2) - \psi(j+\tfrac{1}{2}) \\ &- 2\log x\end{aligned}$$

and $\psi(l)$ is the digamma function. The infinite series converges rapidly for small values of x.

To calculate P^{oo} [cf. Eq. (2.7)] using the representation (4.1) valid for $\lambda^2 \sim 1$, note that

$$\begin{aligned}&G_{2,4}^{3,1}\left(z^2 \,\middle|\, \begin{matrix} -m/2-l; 1+k \\ k+\frac{1}{2}, \frac{1}{2}, 0; -m/2 \end{matrix}\right) \\ &= \sum_{j=0}^{k} \frac{(-)^k \pi^2 \Gamma(1+m/2+l+j)(-z^2)^j}{\Gamma(1+k-j)\Gamma(\frac{1}{2}-k+j)\Gamma(\frac{1}{2}+j)\Gamma(1+m/2+j)\Gamma(j+1)} \\ &\quad - z^{2k+1} \sum_{j=0}^{k-1} \frac{\Gamma(\frac{1}{2}+m/2+l+k-j)\Gamma(-\frac{1}{2}-j)\Gamma(j+1)(-z^2)^{-1-j}}{\Gamma(\frac{1}{2}+k-j)\Gamma(\frac{1}{2}+m/2+k-j)\Gamma(k-j)} \\ &\quad - z^{2k+1} \sum_j \frac{\Gamma(\frac{3}{2}+m/2+l+k+j)\Gamma(\frac{1}{2}+j)z^{2j}\psi_{j,l}^{k,m}(z)}{\Gamma(j+1)\Gamma(\frac{3}{2}+k+j)\Gamma(\frac{3}{2}+m/2+k+j)\Gamma(j+k+1)},\end{aligned} \tag{7.2}$$

where

$$\begin{aligned}\psi_{j,l}^{k,m}(z) = &\psi(1+k+j) + \psi(\tfrac{3}{2}+k+j) + \psi(\tfrac{3}{2}+m/2+k+j) \\ &+ \psi(j+1) - \psi(\tfrac{3}{2}+m/2+l+k+j) - \psi(j+\tfrac{1}{2}) \\ &- 2\log z\end{aligned}$$

is again a rapidly converging form with readily computable coefficients.

The case $\mu = 1$, $\nu = 3$ is of particular interest, so for a thin annulus, use Eq. (3.5) and (2.7) to discover

$$\begin{aligned}p^{oo} = &1 - \kappa + \frac{z^2\lambda^2}{1+\kappa}\left[1 - \frac{1}{3}\left(\frac{\kappa-1}{\kappa+1}\right)\right] - \frac{8z\lambda^2}{3\pi} D_{-1}(\lambda^2) + \frac{\lambda^2 z^3}{\pi} \\ &\times \Bigg(\sum_j \frac{\Gamma(\frac{1}{2}+j)[\psi(\frac{1}{2}+j) - \psi(\frac{5}{2}+j) - \psi(j+1) - \psi(j+2)]D_j(\lambda^2)z^{2j}}{\Gamma(j+1)\Gamma(j+2)\Gamma(j+\frac{7}{2})} \\ &- \sum_j \frac{\Gamma(\frac{1}{2}+j)\Gamma(\frac{5}{2}+j)E_j(\lambda^2)z^{2j}}{\Gamma(j+1)\Gamma(j+2)\Gamma(\frac{7}{2}+j)\Gamma(\frac{7}{2}+j)} \\ &+ 2\log z \sum_j \frac{\Gamma(\frac{1}{2}+j)D_j(\lambda^2)z^{2j}}{\Gamma(j+1)\Gamma(j+2)\Gamma(j+\frac{7}{2})}\Bigg)\end{aligned} \tag{7.3}$$

where

$$D_j(\lambda^2) = {}_2F_1(\tfrac{1}{2}, \tfrac{5}{2}+j; \tfrac{7}{2}+j; \lambda^2)$$

and

$$E_j(\lambda^2) = {}_3F_2(\tfrac{1}{2}, \tfrac{5}{2}+j, \tfrac{5}{2}+j; \tfrac{7}{2}+j, \tfrac{7}{2}+j; \lambda^2).$$

For $\lambda^2 \sim 0$, both $D_j(\lambda^2)$ and $E_j(\lambda^2)$ are readily computable and obey recursion formulas. Note that

$$D_{-1}(\lambda^2) = -3(1-\lambda^2)^{1/2}[1 - \tan^{-1}\lambda/\kappa)/(\kappa\lambda)]/2.$$

This form is particularly useful for numerical evaluations because of its rapid convergence for small values of z and λ^2. Recalling the nature of the problems in which P^{oo} needs to be used, it is sufficient that the functions $D_j(\lambda^2)$ and $E_j(\lambda^2)$ be precomputed. It is also worth mentioning that the expansions for P^{vi} defined in Eq.

(2.8) are of the same form as those for P^{oo}; these are given explicitly elsewhere.[9]

B. Multiple integrals

The variable κ appears in a simple form in the representations (3.5) and (3.9) and so further integration over κ is possible. An example follows.

Benoit,[7] Berna,[7] and Love and Kushneriuk in Eq. (C9) of Ref. 7 introduce the integral

$$D_1(\Sigma a) = -\frac{2}{\pi}\int_0^{\pi/2}\cos^2\Psi\, a\Psi \int_0^{\pi/2}\cos\beta\, d\beta \times \int_0^{\infty}\frac{u + a\cos\Psi}{u^2 + a^2 + 2au\cos\Psi}\exp(-u\Sigma/\cos\beta)\, du.$$

Convert to the geometry of Fig. 1, using

$$du\, d\sin\Psi = db^2\, dt/2a(b^2 - t^2)^{1/2}$$

and rearrange the order of integration, revealing

$$D_1(\Sigma a) = -(2/\pi)\int_0^1 d\kappa\, {}^1M^{1;0}_{2;0,1}(x,\kappa)/\kappa^2.$$

So, from known integration formulas,[8] we find, after making use of Eqs. (3.9) and (A4) and setting $y = \Sigma^2 a^2/4$

$$D_1(2\sqrt{y}) = -(\tfrac{1}{4})\lim_{\nu\to 2}\left[(\tfrac{1}{2})G^{4;1}_{2,4}\left(y\,\middle|\,\begin{matrix}0;\frac{3}{2}\\ 0,\nu/2,0,\frac{1}{2};\end{matrix}\right) + \sum_{l=0}^{\infty}\Gamma(l+2)G^{4;1}_{2,4}\left(y\,\middle|\,\begin{matrix}\frac{1}{2}-l;\frac{3}{2}\\ l+\frac{1}{2},\nu/2,\frac{1}{2},1;\end{matrix}\right) \times[\Gamma(l+1)\Gamma(l+\tfrac{3}{2})\Gamma(l+\tfrac{5}{2})]^{-1} - \sum_{l=1}^{\infty}\{l\to l+\tfrac{1}{2}\}\right]. \tag{7.4}$$

Curiously, by using the contour integral (Barnes) representation for the G functions,[13] this combination of sums can be explicitly evaluated in terms of simple functions and series. Using Bailey's transformation[21] for a nonterminating Saalschützian ${}_3F_2$, simple expressions[22] for the value of a particular ${}_3F_2$ of unit argument, and taking the limit $\nu\to 2$ bares the result

$$\begin{aligned}
&D_1(2\sqrt{y})\\
&= -\frac{1}{4}\left(\frac{1}{4\pi i}\int_{L_0}\frac{\Gamma(1-s)\Gamma(\frac{1}{2}-s)\Gamma(-s)\Gamma(-s)\Gamma(1+s)y^s\, ds}{\Gamma(\frac{3}{2}-s)}\right.\\
&\quad\left. -\frac{1}{4\pi^2 i}\int_{L_{1/2}}\frac{\Gamma(-1+s)\Gamma(-\frac{1}{2}+s)\Gamma(-s)\Gamma(\frac{1}{2}-s)\Gamma(1-s)y^s\, ds}{\Gamma(2+s)}\right)\\
&= -\tfrac{1}{4}\left(\{-\log y - 2\gamma - 2\} - \{(8\sqrt{y}/3)[\log y - \psi(\tfrac{1}{2}) - \psi(\tfrac{5}{2})]\}\right.\\
&\quad + ((y/4)\{[\log y - 2 - \psi(2) - \psi(3)]^2 + 4\pi^2/3 + 25/4\})\\
&\quad - \frac{\pi}{2}y^{3/2}\sum_{j=0}^{\infty}\frac{\Gamma(j+\frac{1}{2})\Gamma(j+1)(-y)^j}{\Gamma(j+2)\Gamma(j+\frac{7}{2})\Gamma(j+\frac{5}{2})\Gamma(\frac{3}{2}+j)}\\
&\quad \left. + \frac{y}{2}\sum_{j=1}^{\infty}\frac{\Gamma(j+\frac{1}{2})\Gamma(j)(-y)^j\psi_j(y)}{\Gamma(j+1)\Gamma(j+2)\Gamma(j+3)\Gamma(\frac{3}{2}+j)}\right),
\end{aligned}$$

where (7.5)

$$\psi_j(y) = \psi(j+\tfrac{1}{2}) + \psi(j) - \psi(-\tfrac{1}{2}-j) - \psi(j+1) - \psi(j+2) - \psi(j+3) + \log y,$$

γ is Euler's constant, the contour $L_{1/2}$ encloses the positive s axis in a negative direction, crossing between $s=0$ and $s=\frac{1}{2}$, and L_0 encloses only the singularity at $s=0$, again in the negative sense.

The first three terms in curly brackets in this expression result from evaluating the residue at the two dipoles and the triple pole at $s=0$, $\frac{1}{2}$, and 1 respectively in the complex s plane. This form explicitly illustrates the analytic structure of the function near $y\sim 0$, and is useful for numerical evaluation near the origin, since the sums converge rapidly. The G functions appearing in Eq. (7.4) have known asymptotics,[8] and so an asymptotic series for D_1 may also be derived.

C. Explicit summation

The evaluation of integrals sometimes requires that sums of Bickley–Nayler functions be obtained. For example, as part of the discrete integral transport (DIT) method,[2] the integral

$$P^{m,\mu}_{n,\nu,l}(\tau) = (1+(-)^{m+n+\mu+\nu})\int_0^{\pi/2}P_n^m(\cos\theta)P_\nu^\mu(\cos\theta) \times\exp(-\tau/\sin\theta)\sin^{l-1}\theta\, d\theta$$

is expressed as triple sum of $\mathrm{Ki}^1_\nu(\tau)$; one of these summation can be explicitly performed using Eq. (6.1) and the method of Ref. 13. The result

$$\begin{aligned}
&\sum_{p=0}^{k}\binom{k}{p}(-)^p\mathrm{Ki}^1_{s+2p}(\tau)\\
&= \frac{1}{\sqrt{\pi}}\Gamma(k+\tfrac{1}{2})G^{3,0}_{1,3}\left(\frac{\tau^2}{4}\,\middle|\,\begin{matrix};s/2+\frac{1}{2}+k\\ s/2,0,\frac{1}{2}\end{matrix}\right)\\
&= 2\mathrm{Ki}^{2k+1}_s(\tau)
\end{aligned} \tag{7.6}$$

can be employed within the context of the DIT method, neglecting the developments of Sec. 3 which suggest other expressions for integrals over the τ variable.

D. Transport kernels

The function ${}^\epsilon M^{\tau;\xi}_{\nu,\mu,\delta}(x,\kappa)$ defined in Eq. (2.2) appears in various forms as part of the kernel in the integral formulation of the transport equation in cylindrical geometry, relating the transport probability of a particle scattered from the *near* radius of a shell to a point in space. In addition, the transport probability from the *far* radius is required, so define the analogous generalized form:

$$\begin{aligned}
{}^\epsilon N^{\tau;\xi}_{\nu,\mu,\delta}(x,\kappa) &= \kappa^{\xi+1}\int_0^1 dv\int_0^1 dt\, v^{\nu-1}(1-v^2)^{\tau/2-1}\\
&\quad\times t^{\xi-\mu}(1-t^2)^{\delta/2}(1-\kappa^2t^2)^{(\epsilon-1)/2}\\
&\quad\times R_+^\mu(t)\exp[-\sigma R_+(t)],
\end{aligned}$$

where

$$R_+(t) = (1-\kappa^2t^2)^{1/2} + \kappa(1-t^2)^{1/2} \tag{7.7}$$

and $\epsilon = 0$ or 1, where we note the sign change of μ in the exponent of t. In analogy with Eq. (3.6),

$$J^+ = \kappa^{\xi+1}\sum_n\frac{(-\sigma)^n}{\Gamma(n+1)}I^\epsilon_{-n}(\kappa,\xi,-\mu,\delta)$$

and the equivalent of Eq. (B3) is

$$S^+_n = \frac{1}{\sqrt{\pi}}G^{2,2}_{2,4}\left(\frac{\sigma^2\lambda^2}{4}\,\middle|\,\begin{matrix}(1+\epsilon-\mu)/2-l, 1-\epsilon-\mu/2;\\ 0,\frac{1}{2};(1-\epsilon-\mu)/2+l, 1-\mu/2\end{matrix}\right).$$

This gives the representation

$${}^\epsilon N^{\tau;\xi}_{\nu,\mu,\delta}(x,\kappa) = \kappa^{\xi+1}\lambda^{\epsilon+\mu-1}\Gamma((1+\xi-\mu)/2)[\Gamma(\tau/2)/4] \times\left[\sum_l\frac{\Gamma(1+\delta/2+l)(-\eta)^l}{\Gamma(1+l)\Gamma(\frac{1}{2}+l)\Gamma((3+\xi-\mu+\delta)/2+l)}\right.$$

$$\times G^{3,2}_{3,5}\left(\frac{z^2}{4}\,\middle|\,\begin{matrix}1-\epsilon-\mu/2,\,(1+\epsilon-\mu)/2-l;(\nu+\tau)/2\\ \nu/2,\,0,\,\tfrac{1}{2};\,(1-\epsilon-\mu)/2+l,\,1-\mu/2\end{matrix}\right)$$

$$+\sum_l\{l\to l+\tfrac{1}{2}\}\Big]. \tag{7.8}$$

An interesting identification for the meaning of the sums in Eqs. (7.8) and (3.9) can be found by remembering that $\epsilon=0$ or 1 means the G functions in the two equations are identical. Comparing the integral representation and the sums (with the transformation $\xi\to\xi+2\mu$ in one) gives

$$\frac{\lambda^{\epsilon-1}\Gamma((1+\xi)/2)\Gamma(\tau/2)}{4}\sum_l\left\{\frac{\Gamma(1+\delta/2+l)(-\eta)^l}{\Gamma(l+1)\Gamma(l+\frac{1}{2})\Gamma((3+\xi+\delta+\mu)/2+l)}\,G^{3,2}_{3,5}\left(\frac{z^2}{4}\,\middle|\,\begin{matrix}1-\epsilon-\mu/2,\,(1+\epsilon-\mu)/2-l;(\nu+\tau)/2\\ \nu/2,\,0,\,\tfrac{1}{2};\,(1-\epsilon-\mu)/2+l,\,1-\mu/2\end{matrix}\right)\right\}$$

$$=\frac{1}{2}\int_0^1 dv\,v^{\nu-1}(1-v^2)^{\tau/2-1}\int_0^1 dt\,t^{\xi+\mu}(1-t^2)^{\delta/2}(1-\kappa^2t^2)^{(\epsilon-1)/2}\{R^\mu_+(t)\exp[-\sigma R_+(t)]\pm R^\mu_-(t)\exp[-\sigma R_-(t)]\},$$

a form which contains sinh and cosh in the integrand when $\mu=0$, with the $-$ part of $\pm$ corresponding to the l sum with l replaced by $l+\frac{1}{2}$.

Similar, and useful results originate from these developments when incomplete integrals are defined. Let

$$\begin{aligned}{}^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa,\alpha)&=\kappa^{1+\xi}\int_\alpha^1 dt\int_0^1 dv\,v^{\nu-1}\,(1-v^2)^{\tau/2-1}t^{\xi+\mu}(1-t^2)^{\delta/2}\\&\quad\times(1-\kappa^2t^2)^{(\epsilon-1)/2}\,R^\mu_-(t)\exp[-xR_-(t)/v],\end{aligned} \tag{7.9}$$

and similarly define ${}^{\epsilon}N^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa,\alpha)$. Following the derivation of Sec. 3 and Appendix B, we ultimately identify

$$\begin{aligned}{}^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa,\alpha)&=\kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\beta^{2+\delta}[\Gamma(\tau/2)/4]\\&\quad\times\Big[\sum_l\frac{(-\eta\beta^2)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac{1}{2}+l)\Gamma(l+1)\Gamma(2+\delta/2+l)}\\&\quad\times{}_2F_1((1-\xi-\mu)/2,1+l+\delta/2;2+l+\delta/2;\beta^2)\\&\quad\times G^{3,2}_{3,5}\left(\frac{z^2}{4}\,\middle|\,\begin{matrix}-\mu/2,\,(1+\epsilon-\mu)/2-l;(\nu+\tau)/2\\ \nu/2,\,0,\,\tfrac{1}{2};\,\epsilon-\mu/2,\,(1-\mu-\epsilon)/2+l\end{matrix}\right)\\&\quad-\sum_l\{l\to l+\tfrac{1}{2}\}\Big]\end{aligned} \tag{7.10}$$

and

$$\begin{aligned}{}^{\epsilon}N^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa,\alpha)&=\kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\beta^{\delta+2}[\Gamma(\tau/2)/4]\\&\quad\times\Big[\sum_l\frac{(-\eta\beta^2)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac{1}{2}+l)\Gamma(l+1)\Gamma(2+\delta/2+l)}\\&\quad\times{}_2F_1((1-\xi+\mu)/2,1+l+\delta/2;2+l+\delta/2;\beta^2)\\&\quad\times G^{3,2}_{3,5}\left(\frac{z^2}{4}\,\middle|\,\begin{matrix}1-\epsilon-\mu/2,\,(1+\epsilon-\mu)/2-l;(\nu+\tau)/2\\ \nu/2,\,0,\,\tfrac{1}{2};\,(1-\epsilon-\mu)/2+l,\,1-\mu/2\end{matrix}\right)\\&\quad+\sum_l\{l\to l+\tfrac{1}{2}\}\Big],\end{aligned} \tag{7.11}$$

where $\beta^2=1-\alpha^2$. These integrals and their complements appear in explicit expressions for the transport kernels in cylindrical geometry.

8. SUMMARY

In Sec. 2, two fundamental forms, basic to the analysis of the integral transport equation in infinite annular geometry are defined. The first of these is related to transmission probabilities for particles that do not intersect the inner radius of an annulus; the second is useful for paths in which the inner radius of an annulus is intersected. The most general forms of the two functions are defined as integrals, whose integrands contain trigonometric functions of the important geometric angles to arbitrary powers. In applications, the powers become integral and certain limits of the general results must be evaluated.

In Sec. 3, the two functions are identified as sums of known functions (Meijer's G function), from which simpler results may be obtained by consulting the literature. In particular, double sums analogous to multiplication and addition expansions emerge, which are particularly useful in applications. The sums neatly factor into a sum over a geometric quantity which does not usually vary for the duration of a particular calculation, and a sum over a cross-sectional quantity, which does. In Sec. 4, analytic transformations are invoked to reveal expansions useful for thick or thin annuli. As a by-product, simple forms emerge for the case of cylindrical geometry, and limiting values of the integrals are explicitly recorded for the case of voids.

In Sec. 5, general asymptotic series are derived for both integrals in the limit of large interaction cross section.

The generalized, associated Bickley–Nayler functions are identified as G functions in Sec. 6, and the incomplete form of these functions (useful for annuli of finite length) is also displayed as a sum of G functions. Expansions useful for long or short annuli are disclosed for the incomplete form, and asymptotic series are developed for both forms. In a digressionary segment, the complete, generalized, associated Bickley–Nayler functions are expanded in a series of orthogonal (Laguerre) polynomials, and as a series of confluent hypergeometric functions of the second kind.

Some applications of the previous theory are established within Sec. 7. These include explicit double summation formulas of use in numerical analysis, the evaluation of a multiple integral extracted from the literature, the illustrative identification of an interest-

ing sum of Bickley–Nayler functions and the identification of related transport kernels.

A companion paper exploits the theory presented here, with the intent of obtaining simple, swift, and accurate numerical evaluations for particular cases of the general integrals.

ACKNOWLEDGMENTS

I am indebted to Dr. Steve Kushneriuk for many useful and informative conversations. I am also very grateful to Sharon Baker and Lorna Blimke, who typed some difficult manuscripts.

APPENDIX A: USEFUL RELATIONS

The key to the analysis of both integrals examined in the text can be found by deriving the relationship

$$\int_0^1 t^{2\nu-1}(1-t^2)^{\rho-1}\exp(-xt)\,dt = \Gamma(\rho)\Gamma(\tfrac{3}{2})G^{2,1}_{1,3}\left(\frac{x^2}{4}\,\middle|\,\begin{matrix}1-\nu;\\ \tfrac{1}{2},0;1-\nu-\rho\end{matrix}\right)\Big/\pi. \tag{A1}$$

On a superficial level, this result comes from combining the hypergeometric functions appearing in Ref. 23. To obtain these functions, one expands the exponential in Eq. (A1) as a power series, transposes the sum and integration, and identifies the remaining integral as a beta function. Split the resulting sum into its even and odd components to obtain the quoted result. This is in analogy with the procedure followed in Appendix B. (The case $\rho=1$ gives an interesting identification for the incomplete gamma function.)

The G function is described in Ref. 8 by Luke, whose notation is used throughout, with the addition of two (extraneous) semicolons in the parameter list. A useful property of this function is its representation as a finite sum of hypergeometric functions:

$$G^{m,n}_{p,q}\left(z\,\middle|\,\begin{matrix}a_p\\ b_q\end{matrix}\right) = \sum_{h=1}^{m}\frac{\prod_{j=1}^{m}\Gamma(b_j-b_h)^{*}\prod_{j=1}^{n}\Gamma(1+b_h-a_j)z^{b_h}}{\prod_{j=m+1}^{q}\Gamma(1+b_h-b_j)\prod_{j=n+1}^{p}\Gamma(a_j-b_h)} \times {}_pF_{q-1}\left(\begin{matrix}1+b_h-a_p\\ 1+b_h-b_q^{*}\end{matrix}\,\middle|\,(-)^{\delta}z\right) \tag{A2}$$

for $p\le q$, $1\le m\le q$, and appropriate values of z. Here we use $\delta=p-m-n$, the asterisk means $j\neq h$, and empty products are set equal to 1. If any of the first m of the b_q's differ by an integer, a limiting process is required. This is performed throughout the text, tacitly and otherwise. The representation is valid only when $a_j-b_k\neq N$, N a positive integer, $j=1,\ldots,n$, $k=1,\ldots,m$.

Another relation involving G functions, vital to the text, is

$$\int_1^{\infty} t^{-\alpha}(t-1)^{\alpha-\beta-1}G^{m,n}_{p,q}\left(zt\,\middle|\,\begin{matrix}a_p\\ b_q\end{matrix}\right) = \Gamma(\alpha-\beta)G^{m+1,n}_{p+1,q+1}\left(z\,\middle|\,\begin{matrix}a_p,\alpha\\ \beta,b_q\end{matrix}\right) \tag{A3}$$

(for certain ranges of the parameters) allowing multiple integrals to be explicitly evaluated. A second result[24] that is utilized is

$$G^{m,n+1}_{p+1,q+1}\left(z\,\middle|\,\begin{matrix}a,a_p\\ b_q,b\end{matrix}\right) = (-)^{r}G^{m+1,n}_{p+1,q+1}\left(z\,\middle|\,\begin{matrix}a_p,a\\ b,b_q\end{matrix}\right) - (-)^{r}\frac{\prod_{j=1}^{m}\Gamma(b_j-b)\prod_{j=1}^{n}\Gamma(1+b-a_j)z^{b}}{\prod_{j=m+1}^{q}\Gamma(1+b-b_j)\prod_{j=n+1}^{p}\Gamma(a_j-b)\Gamma(r)} \times {}_{p+1}F_q\left(\begin{matrix}1+b-a_p,1-r\\ 1+b-b_q\end{matrix}\,\middle|\,(-)^{\delta}z\right) \tag{A4}$$

when $a-b=r$ (r integral) coupled with the reduction formula

$$G^{m+1,n}_{p+1,q+1}\left(z\,\middle|\,\begin{matrix}a_p,a\\ a,b_q\end{matrix}\right) = G^{m,n}_{p,q}\left(z\,\middle|\,\begin{matrix}a_p\\ b_q\end{matrix}\right). \tag{A5}$$

The G functions may be analytically continued using

$$G^{m,n}_{p,q}\left(z\,\middle|\,\begin{matrix}a_p\\ b_q\end{matrix}\right) = G^{n,m}_{q,p}\left(\frac{1}{z}\,\middle|\,\begin{matrix}1-b_q\\ 1-a_p\end{matrix}\right) \tag{A6}$$

for appropriate ranges of z.

Recall that symmetry exists for interchange of elements of a_p and b_q on either side of the semicolon in explicit cases. Other useful results may be found in Ref. 8.

We also appeal to several properties of Gauss' hypergeometric series. The most important of these involve the transformation formulas:

$$\begin{aligned}{}_2F_1(\alpha,\beta;\gamma;u) &= u^{-\alpha}\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \times {}_2F_1(\alpha,1+\alpha-\gamma;1+\alpha+\beta-\gamma;\ (u-1)/u)\\ &\quad + u^{\alpha-\gamma}(1-u)^{\gamma-\alpha-\beta}\frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)} \times {}_2F_1(\gamma-\alpha,1-\alpha;1+\gamma-\alpha-\beta;(u-1)/u) \qquad \text{(A7)}\\ &= (1-u)^{\gamma-\alpha-\beta}{}_2F_1(\gamma-\alpha,\gamma-\beta;\gamma;u). \qquad \text{(A8)}\end{aligned}$$

An identification of particular interest is

$${}_2F_1(\alpha,\alpha+\tfrac{1}{2};2\alpha+1;u) = \{[1+(1-u)^{1/2}]/2\}^{-2\alpha}. \tag{A9}$$

Of course, the hypergeometric function

$${}_pF_q\left(\begin{matrix}a_p\\ b_q\end{matrix}\,\middle|\,x\right)$$

is itself easily defined as a power series

$${}_pF_q\left(\begin{matrix}a_p\\ b_q\end{matrix}\,\middle|\,x\right) = \sum_l \frac{\Gamma(a_p+l)\Gamma(b_q)x^l}{\Gamma(a_p)\Gamma(b_q+l)\Gamma(l+1)}, \tag{A10}$$

convergent for $p=q+1$ and $|x|<1$ or $p\le q$ and all x.

In manipulating gamma functions, it is well to be aware of the reflection identity

$$\Gamma(z)\Gamma(1-z)=\pi/\sin\pi z \tag{A11}$$

and the duplication formula

$$\sqrt{\pi}\,\Gamma(2z)=2^{2z-1}\Gamma(z)\Gamma(z+\tfrac{1}{2}). \tag{A12}$$

APPENDIX B: EVALUATION OF ${}^{\epsilon}M^{\tau,\xi}_{\nu,\mu,\delta}(x,\kappa)$

Consider the integral

$$I_n^\epsilon(\kappa,\xi,\mu,\delta)$$
$$=\int_0^1 dt\, t^{\xi+\mu}(1-t^2)^{\delta/2}(1-\kappa^2t^2)^{(\epsilon-1)/2}$$
$$\times[(1-\kappa^2t^2)^{1/2}-\kappa(1-t^2)^{1/2}]^{\mu+n}$$
$$=(\lambda^2/2)^{\mu+n}\int_0^1 dt\, t^{\xi+\mu}(1-t^2)^{\delta/2}(1-\kappa^2t^2)^{(\epsilon-\mu-n-1)/2}$$
$$\times\left(\frac{1+\kappa[(1-t^2)/(1-\kappa^2t^2)]^{1/2}}{2}\right)^{-\mu-n}. \tag{B1}$$

Substitute

$$\kappa^2(1-t^2)/(1-\kappa^2t^2)=1-u$$

to obtain a form that may be represented as a Gauss' hypergeometric function according to Eq. (A9). Impose the subsequent transformation $u\to(u-1)/u$ given by Eq. (A7) and find

$$I_n^\epsilon(\kappa,\xi,\mu,\delta)=\left(\frac{\lambda}{2}\right)^{\mu+n}\frac{\Gamma(\mu+n+1)}{\lambda\Gamma((\mu+n+1)/2)}$$
$$\times\int_0^1 dt\,(1-t^2)^{\delta/2}t^{\xi+\mu}\left\{\frac{\Gamma(\tfrac12)(1-\kappa^2t^2)^{\epsilon/2}}{\Gamma((\mu+n)/2+1)}\right.$$
$$\times{}_2F_1((\mu+n+1)/2,(1-\mu-n)/2;\tfrac12;\eta(1-t^2))$$
$$+\frac{\kappa\Gamma(-\tfrac12)(1-t^2)^{1/2}(1-\kappa^2t^2)^{(\epsilon-1)/2}}{\Gamma((\mu+n)/2)}$$
$$\left.\times{}_2F_1((\mu+n+1)/2,(1-\mu-n)/2;\tfrac32;\eta(1-t^2))\right\},$$

where $\eta=-\kappa^2/\lambda^2$. Absorb the embarassing factor $(1-\kappa^2t^2)^{1/2}$ into either the first or second ${}_2F_1$ according to whether $\epsilon=1$ or 0, respectively, using Eq. (A8). Each of the resulting integrals is now in a form that may be explicitly evaluated.[25] Or, obtain an equivalent form by expressing the combination of terms in the curly brackets as a G function, according to Eq. (A2), and, after substituting $1-t^2=v$, determine

$$I_n^\epsilon(\kappa,\xi,\mu,\delta)=\frac{\lambda^{\epsilon+\mu+n-1}\Gamma(\mu+n+1)}{2^{\mu+n+1}\Gamma((\mu+n+1)/2)\Gamma((\mu+n)/2+1-\epsilon)}$$
$$\times\int_0^1 dt\,(1-t)^{(\xi+\mu-1)/2}t^{\delta/2}$$
$$\times G_{2,2}^{2,1}\left(-\eta t\,\middle|\,\begin{matrix}(1+\epsilon-\mu-n)/2;(1+\epsilon+\mu+n)/2\\0;\tfrac12\end{matrix}\right).$$

This integral may be evaluated[8] and the immediate result is Eq. (3.7).

To confirm the final result (3.9), express

$$G_{3,3}^{2,2}\left(-\eta\,\middle|\,\begin{matrix}-\delta/2,(1+\epsilon-\mu-n)/2;(1+\epsilon+\mu+n)/2\\0,\tfrac12;-(1+\xi+\mu+\delta)/2\end{matrix}\right)$$

as a sum of two ${}_3F_2$'s according to Eq. (A2), expand each of the resulting hypergeometric series explicitly after the manner of Eq. (A10) and transpose the sums in Eq. (3.6). This establishes that

$$J^-=\kappa^{1+\xi}\lambda^{\epsilon+\mu-1}\Gamma(\tfrac12)\Gamma(\tfrac32)\Gamma((1+\xi+\mu)/2)$$
$$\times\left\{\sum_l\frac{(-\eta)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac12+l)\Gamma(l+1)\Gamma((3+\xi+\mu+\delta)/2+l)}\right.$$
$$\times\sum_n(-\sigma\lambda/2)^n\Gamma((\mu+n)/2+1)$$
$$\times\Gamma((1+\mu+n-\epsilon)/2+l)[\Gamma((n+1)/2)\Gamma(1+n/2)$$
$$\times\Gamma((\mu+n)/2+1-\epsilon)\,\Gamma((1+\mu+n+\epsilon)/2-l)]^{-1}$$
$$\left.-\sum_l\{l\to l+\tfrac12\}\right\}, \tag{B2}$$

where $\{l\to l+\frac12\}$ means the repetition of the l sum with all values of l replaced by $l+\frac12$. This form is obtained by considerable use of the duplication and reflection formulas (A11) and (A12). Now consider S_n^-, the inner sum over n in Eq. (B2) and separate this into its even and odd components. This yields two sums which are identifiable as generalized hypergeometric series. Explicitly,

$$S_n^-=\frac{\Gamma(1+\mu/2)\Gamma((1+\mu-\epsilon)/2+l)}{\Gamma(\frac12)\Gamma(\mu/2+1-\epsilon)\Gamma((1+\mu+\epsilon)/2-l)}$$
$$\times{}_2F_3(1+\mu/2,(1+\mu-\epsilon)/2+l;\tfrac12,\mu/2+1-\epsilon,$$
$$(1+\mu+\epsilon)/2-l;\sigma^2\lambda^2/4)$$
$$-\left(\frac{\sigma\lambda}{2}\right)\frac{\Gamma(\mu/2+\frac32)\Gamma(1+(\mu-\epsilon)/2+l)}{\Gamma(\frac32)\Gamma((3+\mu)/2-\epsilon)\Gamma(1+(\mu+\epsilon)/2-l)}$$
$$\times{}_2F_3((3+\mu)/2,1+(\mu-\epsilon)/2;\tfrac32,(3+\mu)/2-\epsilon,$$
$$1+(\mu+\epsilon)/2-l;\sigma^2\lambda^2/4).$$

This combination of hypergeometric functions may be expressed as a G function according to Eq. (A2). So,

$$S_n^-=\frac{1}{\pi}G_{2,4}^{2,2}\left(\frac{\sigma^2\lambda^2}{4}\,\middle|\,\begin{matrix}-\mu/2,(1-\mu+\epsilon)/2-l;\\0,\tfrac12;\epsilon-\mu/2,(1-\mu-\epsilon)/2+l\end{matrix}\right) \tag{B3}$$

and

$${}^\epsilon M_{\nu,\mu,\delta}^{\tau,\xi}(x,\kappa)=\kappa^{\xi+1}\lambda^{\epsilon+\mu-1}\frac{\Gamma((1+\xi+\mu)/2)}{2}$$
$$\times\left[\sum_l\frac{(-\eta)^l\Gamma(l+1+\delta/2)}{\Gamma(\frac12+l)\Gamma(l+1)\Gamma((3+\xi+\mu+\delta)/2+l)}\right.$$
$$\times\int_0^1 dv\,v^{\nu-1}(1-v^2)^{\tau/2-1}$$
$$\times G_{2,4}^{2,2}\left(\frac{x^2\lambda^2}{4v^2}\,\middle|\,\begin{matrix}-\mu/2,(1-\mu+\epsilon)/2-l;\\0,\tfrac12;\epsilon-\mu/2,(1-\mu-\epsilon)/2+l\end{matrix}\right)$$
$$\left.-\sum_l\{l\to l+\tfrac12\}\right]. \tag{B4}$$

Substitute $v^2=1/t$ and obtain an integral of the form evaluated in Eq. (A3). The result (3.9) is an immediate consequence.

[1] K. M. Case, F. de Hoffmann, and G. Placzek, "Introduction to the Theory of Neutron Diffusion," Los Alamos (1953).

[2] S. A. Kushneriuk and C. McKay, "Neutron Density in an Infinite, Non-Capturing Medium Surrounding a long Cylindrical Body Which Scatters and Captures Neutrons," Atomic Energy of Canada Ltd., Report AECL-137 (1954); L. Pease and N. P. Vakil, "Tables of Neutron Transmission Probabilities Through Purely Absorbing Tubes," Atomic Energy of Canada Ltd., Report AECL-1504 (1962); I. Carlvik, Nukleonik 10, 104 (1967); A. Jonsson, R. Goldstein, and A. Mockel, Atomkernenergie 24, 79 (1974).

[3] W. G. Bickley and J. Nayler, Phil. Mag. 20, 343 (1935).

[4] T. M. Danielsen, T. Håvie, and R. J. J. Stamm'ler, "On the Numerical Calculation of Bickley Functions", Institut for Atomenergi, Kjeller, Norway, Report KR-55 (1963); D. Davierwalla, "Tschebyscheff Rational Approximations to Bickley Functions," Eidgenoessisches Institut für Reaktorforschung, Würenlingen, Switzerland, Report EIR-243 (1973); A. J. Clayton, J. Nucl. Energy A 18, 82 (1964); K. Makino, Nukleonik 9 (7), 351 (1967); J. M. Blair, C. A. Edwards, and J. H. Johnson, "Rational Chebyshev Approximations for

the Bickley Functions $Ki_n(x)$," Math. Comp. (to be published, 1978).
[5]Y. Luke, *Integrals of Bessel Functions* (McGraw-Hill, Toronto, 1962), Chap. 8.
[6]B. Davison, "Some Integrals Involving the Bessel Function Integrals" Atomic Energy of Canada Ltd., Report AECL-867 (1959).
[7]M. Love and S. Kushneriuk, "The Dipole Response Coefficient for Cylindrical Channels in Neutron Diffusion Theory," Atomic Energy of Canada Ltd., Report AECL-4826 (1974); also J. Phys. D **10** (13), 1725 (1977); R.L. Murray, Nucl. Sci. Eng. **43**, 3, 350 (1971); P. Berna, J. Nucl. Energy **26**, 319 (1972); P. Benoit, Nucl. Sci. Eng. **34**, 285 (1968).
[8]Y. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. 1.
[9]M.S. Milgram, "On the Properties of Collision Probability Integrals in Annular Geometry—II Evaluation" in preparation). preparation).
[10]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), Eq. 27.4.2, p. 1000.
[11]H. Märkl, International Conference on the Peaceful Uses of Nuclear Energy, Paper P/640, Session 3.1, Geneva (1964) A/CONF. 28/.
[12]H. Märkl and A. Fowler, Nukleonik 6(1), 39 (1964); A. Müller and E. Linnartz, Nukleonik 5(1), 23 (1963).
[13]To manipulate sums of this form, write the G function as a contour integral and transpose the sum and integration. Identify the inner sum as a hypergeometric function (of unit argument in some cases) upon which various transformations can be applied. Again transpose the resulting sum and integral and identify the result as a (sum of) new G function(s). M.S. Milgram, "On Some Sums of Meijer's G-Function," Atomic Energy of Canada Ltd., Report AECL-5827 (1977).
[14]N.N. Lebedev, *Special Functions and Their Applications* (Prentice-Hall, Englewood Cliffs, N.J., 1965), Eq. 9.5.7.
[15]Ref. 14, Eq. 9.6.3.
[16]Ref. 8, Theorem 5, Sec. 5.7 and Sec. 5.11.
[17]R. Hwang and B. Toppel, Trans. Am. Nucl. Soc. **23**, 525 (1976).
[18]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1963), Eq. 19.11(1).
[19]Ref. 18, Eq. 19.7(13).
[20]Ref. 18, Eq. 6.12(3).
[21]Ref. 18, Eq. 4.4(4).
[22]Ref. 8, Eq. 3.13.3(45) and Eq. 3.13.3(37).
[23]I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), Eq. 3.389(1), p. 323.
[24]Ref. 8, Eq. 5.4(7). Note that both this and Eq. 5.4(6) are incorrectly stated, and must be augmented if $r > 0$.
[25]Ref. 8, Eq. 3.6(9).

Coherent scalar field in pair-correlated random distributions of aligned scatterers[a)]

Victor Twersky

Mathematics Department, University of Illinois, Chicago
(Received 14 December 1976)

Dispersion equations for coherent propagation of scalar waves in random distributions of pair-correlated obstacles (aligned or averaged over alignment), are obtained by averaging the functional equations relating the multiple and single scattered amplitudes of the obstacles. The resulting bulk indices of refraction and bulk parameters, for aligned nonradially symmetric scatterers, specify anisotropic media; the anisotropy arises either from the scatterers' properties (physical parameters or shape, or both) or from their distribution, or from both. The illustrations include both isotropic and anisotropic cases (in one to three dimensions), and the explicit results generalize earlier ones.

1. INTRODUCTION

In earlier papers,[1,2] we considered scattering of a plane wave $\exp(i\mathbf{k}\cdot\mathbf{r})$ by a slab region of randomly distributed uncorrelated obstacles. For negligible boundary layer effects, the average wave and the associated index of refraction ($\eta = K/k$) were expressed in terms[1] of the conventional scattering amplitude $g(\hat{\mathbf{r}}, \hat{\mathbf{k}})$ for one obstacle excited by $\exp(i\mathbf{k}\cdot\mathbf{r})$, or in terms[2] of a generalization for excitation $\exp(i\mathbf{K}\cdot\mathbf{r})$. In the present paper (Sec. 3), representations for η and the bulk parameters in unbounded distributions are derived with slab-region applications in view. For pair-correlated distributions (Sec. 4), we average the functional equation[3,4] for the multiple scattered amplitude $G = G[g(\hat{\mathbf{r}}, \hat{\mathbf{k}})]$ (with the ensemble average for two fixed scatterers replaced by that for one,[5,6] a procedure which may be interpreted by alternative approaches[7,8]) to obtain dispersion equations for η. The relations of these key equations [(60), (64)] to existing approximations for η in terms of g are indicated, and detailed applications, based on eigenfunction series appropriate to the symmetry of the correlations, are given in Secs. 5, 6, and 7 for two, three, and one dimensions, respectively.

For radial symmetry (circular in Sec. 5A, and spherical in Sec. 6A), to facilitate determining η, we reduce the initial homogeneous algebraic system (in terms of conventional scattering coefficients and lattice-sum analogs) to an inhomogeneous system. The circular case is analyzed in detail to provide prototypes for comparison with earlier forms for an equivalent uniform slab. For aligned elliptic cylinders (with nonconfocal, nonsimilar, and nonparallel scatterer and exclusion surfaces) a Mathieu function expansion (Sec. 5B) leads to explicit low frequency forms [(127), (143)] which are then generalized directly to the analogous problem for triaxial ellipsoids [(174)]. The results for η specify anisotropic media; the anisotropy arises either from the scatterers' properties (physical parameters or shape, or both) or from their correlations, or from both.

We begin with several definitions and representations, and use (1 : 3.4) for Eq. (3.4) of Ref. 1, etc., as well as essentially the same notation as before.[1-4]

[a)]Work supported in part by National Science Foundation Grants GP-33368X and MPS 75-07391.

2. DEFINITIONS AND REPRESENTATIONS

We consider a plane wave $\phi\exp(-i\omega t)$,

$$\phi = \exp(i\mathbf{k}\cdot\mathbf{r}),$$

$$\mathbf{k} = k\hat{\mathbf{k}}(\alpha) = k(\hat{\mathbf{z}}\cos\alpha + \hat{\mathbf{x}}\sin\alpha) = \hat{\mathbf{z}}\gamma + \hat{\mathbf{x}}\tau, \quad 0 \leq \alpha < \pi/2, \tag{1}$$

incident on a slab distribution of identical scatterers (completely bounded obstacles, parallel cylinders, or parallel slabs in three, two, and one dimensions respectively). The obstacles are specified by their known normalized isolated scattering amplitude $g(\hat{\mathbf{r}}, \hat{\mathbf{k}})$, such that for lossless scatterers $-\mathrm{Re}g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = \mathfrak{M}|g(\hat{\mathbf{r}}, \hat{\mathbf{k}})|^2$ with $\mathfrak{M}$ as the mean value over all directions of observation $\hat{\mathbf{r}}$, i.e., $\mathfrak{M}_3 = (1/4\pi)\int d\Omega(\theta, \varphi)$, $\mathfrak{M}_2 = (1/2\pi)\int d\theta$, and $\mathfrak{M}_1$ is one-half the forward and reflected values. For lossy scatterers,

$$-\mathrm{Re}g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = (\sigma_A + \sigma_S)/\sigma_O, \quad \sigma_S = \sigma_O\mathfrak{M}|g(\hat{\mathbf{r}}, \hat{\mathbf{k}})|^2,$$

$$\sigma_O = 4\pi/k^2,\ 4/k,\ 2\cos\alpha \tag{2}$$

with σ_A and σ_S as the absorption and scattering cross sections respectively. The order for the sequence σ_O (and for subsequent sequences) corresponds to three, two, one dimensions. In general, we use three-dimensional terminology.

For a fixed configuration, a scatterer is located by the vector $\mathbf{r}_s$ from $\mathbf{r} = 0$ to the center of its smallest circumscribing sphere (of radius a), and all centers are within the slab region $0 \leq z \leq d$. The obstacles may be asymmetrical, and either similarly aligned or averaged over alignment. For the ensemble of configurations, we specify the one-particle statistics by the average number (ρ) of scatterers in unit volume, and the two-particle statistics by $\rho f(\mathbf{R})$ with $f(\mathbf{R})$ as the distribution function for the separation ($\mathbf{R}_{ts} = \mathbf{r}_t - \mathbf{r}_s$) of pairs. The minimum separation of centers as a function of $\hat{\mathbf{R}}$ specifies the exclusion surface $\mathbf{R} = \mathbf{b}(\hat{\mathbf{R}})$; we require $f(\mathbf{R}) = 0$ for $R < |\mathbf{b}(\hat{\mathbf{R}})|$, and $f(\mathbf{R}) \sim 1$ for $R \sim \infty$. If $\mathbf{b} = b\hat{\mathbf{R}}$ is a sphere with radius $b \geq 2a$, then $f(R)$ is the usual radial distribution function. We use $f(R)$ not only for spherical obstacles of radius a but also for more general shapes $\mathbf{R} = \mathbf{a}(\hat{\mathbf{R}})$ (aligned or averaged over alignment) regarded as if enclosed in transparent coatings whose outer surfaces are spheres of diameter b; the transparent shell has no direct influence on an isolated obstacle's scattering properties but preserves the radial

symmetry of the distribution. For cases where we average over alignment we assume the distribution of alignments to be uniform and uncorrelated with position or separation. The most general case we consider corresponds to differently aligned nonsimilar scatterer (**a**) and exclusion (**b**) surfaces.

For ϕ as the excitation for a single obstacle at the phase origin ($r=0$), the external field satisfies

$$(\nabla^2+k^2)\psi=0, \quad k=|k|=2\pi/\lambda, \quad \psi=\psi(k)=\phi+u. \tag{3}$$

The scattered wave u is the radiative function

$$u(\mathbf{r})=c_o\int[h_0(k|\mathbf{r}-\mathbf{r}'|)\nabla u(\mathbf{r}')-u\nabla h_0]\cdot d\mathfrak{S}(\mathbf{r}')\equiv\{h_0,u\};$$
$$h_0=h_0^{(1)}(k|\mathbf{r}-\mathbf{r}'|), \quad H_0^{(1)}(k|\mathbf{r}-\mathbf{r}'|), \quad \exp(i\tau x+i\gamma|z-z'|),$$
$$c_o=k/i4\pi, 1/i4, 1/i2\gamma \tag{4}$$

with $\mathfrak{S}=\mathfrak{S}\hat{\mathbf{n}}$ as the obstacle's surface, $\hat{\mathbf{n}}$ as the outward normal, and $\nabla=\nabla_{r'}$. (In one dimension, the brace operation with $\nabla=\hat{\mathbf{z}}\partial_{z'}$ corresponds to the sum of the values at $z'=a$ and $-a$.) In the scatterer's interior $\mathfrak{V}$ the field is a nonsingular solution of

$$(\nabla^2+K'^2)\psi=0, \quad K'=k\eta', \quad \psi(K')=\psi', \tag{5}$$

where η' is the relative index of refraction. On $\mathfrak{S}$, we require

$$\psi=\psi', \quad \hat{\mathbf{n}}\cdot\nabla\psi=B'\hat{\mathbf{n}}\cdot\nabla\psi' \tag{6}$$

with B' as one of the scatterers' relative parameters and $B'\eta'^2=C'$ as the other. (For example, in small amplitude acoustics, ψ determines the excess pressure, $B'\nabla\psi'$ the particle velocity, B'^{-1} the relative density for lossless media, and C' the compressibility.) Rewriting (4) as $u=\{h_0,\psi\}$, we have from (6)

$$u=c_o\int[h_0B'\nabla\psi'-\psi'\nabla h_0]\cdot d\mathfrak{S}\equiv[h_0,\psi] \tag{7}$$

and from Gauss' theorem and (5),

$$u=\psi-\phi=-c_o\int[(C'-1)k^2h_0\psi'-(B'-1)\nabla h_0\cdot\nabla\psi']d\mathfrak{V}(\mathbf{r}')$$
$$\equiv[\![h_0,\psi]\!] \tag{8}$$

with $C'=B'\eta'^2$. We also use (8) for $\psi'-\phi$ at $\mathbf{r}$ in $\mathfrak{V}$ as a principal value in that the singularity of $h_0(k|\mathbf{r}-\mathbf{r}'|)$ is excluded by $|\mathbf{r}-\mathbf{r}'|=\epsilon\to 0$.

We write the asymptotic forms of (4), (7), and (8) for $r\sim\infty$ as $u\sim h(kr)g(\hat{\mathbf{r}},\hat{\mathbf{k}})$ with h as the asymptotic form of h_0; the corresponding representations for the scattering amplitude $g(\hat{\mathbf{r}},\hat{\mathbf{k}})$ are

$$g\{\mathbf{k}_r,\mathbf{k}\}=\{\exp(-i\mathbf{k}_r\cdot\mathbf{r}'),u\}, \quad g[\mathbf{k}_r,\mathbf{k}]=[\exp(-i\mathbf{k}_r\cdot\mathbf{r}'),\psi],$$
$$g[\![\mathbf{k}_r,\mathbf{k}]\!]=[\![\exp(-i\mathbf{k}_r\cdot\mathbf{r}'),\psi]\!], \quad \mathbf{k}_r=k\hat{\mathbf{r}}. \tag{9}$$

Using the complex spectral representation[3,4] for h_0 in (4), we have (at least for $r>a$ for all $\hat{\mathbf{r}}$, and for $r>(\hat{\mathbf{r}}\cdot\mathbf{r}')_{\max}$ for given $\hat{\mathbf{r}}$),

$$u(\mathbf{r})=\int_c\exp(i\mathbf{k}_c\cdot\mathbf{r})g(\hat{\mathbf{r}}_c,\hat{\mathbf{k}}), \quad \mathbf{k}_c=k\mathbf{r}_c. \tag{10}$$

In three dimensions, [4] $\hat{\mathbf{r}}_c=\hat{\mathbf{r}}(\theta_c,\varphi_c)$ and $\int_c=(1/2\pi)\times\int\int d\Omega(\theta_c,\varphi_c)$ with contours as for $h_0^{(1)}$; in two, [3] $\hat{\mathbf{r}}_c=\hat{\mathbf{r}}(\theta_c)=\mathbf{k}(\theta_c)$ and $\int_c=(1/\pi)\int d\theta_c$ with contour as for $H_0^{(1)}$; in one, $\mathbf{k}_c=\pm\gamma\hat{\mathbf{z}}$ and $\int_c$ selects the sign corresponding to $z=\pm|z|$.

For impenetrable scatterers, we take $\psi'=0$, and consider either $\phi+u=0$ on $\mathfrak{S}$ (e.g., a pressure release surface) or $\partial_n(\phi+u)=0$ on $\mathfrak{S}$ (a rigid obstacle). For such cases, we use only the brace operation.

For a fixed configuration of N obstacles with centers located by $\mathbf{r}_s$ (and $s=1,2,\ldots,N$), we write the net field as[1-9]

$$\Psi=\phi+\sum_{s=1}^{N}U_s(\mathbf{r}-\mathbf{r}_s),$$
$$U_s(\mathbf{r}-\mathbf{r}_s)=U_s(\mathbf{r}-\mathbf{r}_s;\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) \tag{11}$$

with U_s radiating from the scatterer fixed at $\mathbf{r}_s$. Equivalently, with reference to scatterer t,

$$\Psi=\Psi_t=\Phi_t+U_t, \quad \Phi_t=\phi+\sum_s'U_s, \quad \sum_s'=\sum_{s\neq t}, \tag{12}$$

where Φ_t may be regarded as the net excitation. The functions Ψ_t, Φ_t, U_t satisfy the same relations (3)–(8) at scatterer t as ψ, ϕ, u for the single obstacle. In particular,

$$\Psi_t-\Phi_t=[\![h_0(k|\mathbf{r}-\mathbf{r}_t-\mathbf{r}'|),\Psi_t(\mathbf{r}_t+\mathbf{r}')]\!], \tag{13}$$

where the volume integral is over $\mathfrak{V}_t(\mathbf{r}')$ with $\mathbf{r}'$ as the local vector from the point $\mathbf{r}_t$. For $|\mathbf{r}-\mathbf{r}_t|\sim\infty$, we have $U_t(\mathbf{r}-\mathbf{r}_t)\sim h(k|\mathbf{r}-\mathbf{r}_t|)G_t$.

The corresponding multiple scattered amplitude G_t may be expressed in any of the forms in (9), e.g., as the volume integral

$$G_t(\hat{\mathbf{r}})=[\![\exp(-i\mathbf{k}_r\cdot\mathbf{r}'),\Psi_t(\mathbf{r}_t+\mathbf{r}')]\!] \tag{14}$$

or as $\{\exp(-i\mathbf{k}_r\cdot\mathbf{r}'),U_t\}$ over the surface $\mathfrak{S}_t$, etc. The analog of (10) is

$$U_t(\mathbf{r}-\mathbf{r}_t)=\int_c\exp[i\mathbf{k}_c\cdot(\mathbf{r}-\mathbf{r}_t)]G_t(\hat{\mathbf{r}}_c). \tag{15}$$

The requirement that Ψ_t and $\psi_a=\phi_a+u_a$ (for an arbitrary direction of incidence $\hat{\mathbf{r}}_a$) satisfy the same conditions on $\mathfrak{S}_t$ and in $\mathfrak{V}_t$ corresponds to $\{\psi_a,\Psi_t\}_t=0$ over $\mathfrak{S}_t(\mathbf{r}')$. Consequently, $G_t(-\hat{\mathbf{r}}_a)=\{\phi_a,U_t\}_t=\{\Phi_t,u_a\}_t$, and we obtain[3,4]

$$G_t(\mathbf{r})=g_t(\hat{\mathbf{r}},\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r}_t)$$
$$+\sum_s'\int_c g_t(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)G_s(\hat{\mathbf{r}}_c)\exp(i\mathbf{k}_c\cdot\mathbf{R}_{ts}), \quad \mathbf{R}_{ts}=\mathbf{r}_t-\mathbf{r}_s, \tag{16}$$

where we used the reciprocity relation $g(\hat{\mathbf{r}},\hat{\mathbf{k}})=g(-\hat{\mathbf{k}}-\hat{\mathbf{r}})$.

In Secs. 3 and 4 we average (11) and (16) over an ensemble of configurations to derive representations (from Ψ) and dispersion equations (from G) for the bulk index η.

3. THE AVERAGE WAVE

The average of Ψ over a statistically homogeneous ensemble of configurations of N identical and aligned obstacles whose centers $\mathbf{r}_s$ are uniformly distributed in V may be written[1,5-9]

$$\langle\Psi(\mathbf{r})\rangle=\phi+\rho\int_V\langle U_s(\mathbf{r}-\mathbf{r}_s)\rangle_s d\mathbf{r}_s,$$
$$\rho=N/V, \quad \langle U_s\rangle_s=\langle\Psi_s\rangle_s-\langle\Phi_s\rangle_s, \tag{17}$$

where $\langle\Psi\rangle$ is independent of the configurational variables $(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)$, and $\langle\ \rangle_s$, the average with $\mathbf{r}_s$ held fixed,

depends only on $\mathbf{r}_s$ (now, a dummy). We use the radiative form $\langle U_s(k)\rangle_s$ if $\mathbf{r}$ is outside of $\mathfrak{V}_s=\mathfrak{V}$, and $\langle\Psi_s(K')\rangle_s - \langle\Phi_s(k)\rangle_s$ if inside. Equivalently, we work with the principal value form for all $\mathbf{r}$,

$$\langle\Psi_s\rangle_s - \langle\Phi_s\rangle_s = [\![h_0(k|\mathbf{r}-\mathbf{r}_s-\mathbf{r}'|), \langle\Psi_s(\mathbf{r}_s+\mathbf{r}')\rangle_s]\!],$$
$$\langle\Psi_s(\mathbf{r}_s+\mathbf{r}')\rangle_s = \exp(i\tau x_s)\langle\Psi_s(\mathbf{z}_s+\mathbf{r}')\rangle_s, \tag{18}$$

where the second equation, appropriate for V unbounded along x (and y), preserves the phase parallel to the boundaries (Snell's law). The corresponding average scattering scattering amplitude is

$$\langle G_s(\hat{\mathbf{r}})\rangle_s = [\![\exp(-i\mathbf{k}_r\cdot\mathbf{r}'), \langle\Psi_s(\mathbf{r}_s+\mathbf{r}')_s]\!] \equiv G(\mathbf{r}_s;\hat{\mathbf{r}})$$
$$= \exp(i\tau x_s)\, G(z_s:\hat{\mathbf{r}}),$$
$$G(z_s:\hat{\mathbf{r}}) = [\![\exp(-i\mathbf{k}_r\cdot\mathbf{r}'), \langle\Psi_s(\mathbf{z}_s+\mathbf{r}')\rangle_s]\!]. \tag{19}$$

Using (18) in (17), we integrate[1] over x_s, y_s to obtain

$$\langle\Psi\rangle = \phi$$
$$+ c\int_0^d [\![\exp[i\tau(x-x')+i\gamma|z-z_s-z'|], \langle\Psi_s(\mathbf{z}_s+\mathbf{r}')\rangle_s]\!]\, dz_s,$$
$$c = 2\pi\rho/\gamma k, 2\rho/\gamma, \rho \tag{20}$$

with $z_s \lessgtr z - z' \mp \epsilon$, $\epsilon \to 0$. For the transmitted field,[1] at least for $z > d + a$,

$$\langle\Psi\rangle = \Psi_T = \phi[1 + c\int_0^d \exp(-i\gamma z_s)\, G(z_s:\hat{\mathbf{k}})\, dz_s] = \mathfrak{T}\phi. \tag{21}$$

Similarly, for the reflected field,[1] at least for $z < -a$,

$$\langle\Psi\rangle - \phi = \Psi_R = \phi' c\int_0^d \exp(i\gamma z_s)\, G(z_s:\hat{\mathbf{k}}')\, dz_s = \mathfrak{R}\phi',$$
$$\phi' = \exp(i\mathbf{k}'\cdot\mathbf{r}), \quad \mathbf{k}' = \mathbf{k}(\pi-\alpha) = -\hat{\mathbf{z}}\gamma + \hat{\mathbf{x}}\tau. \tag{22}$$

The corresponding internal field, at least for $a < z < d - a$, consists of essentially two waves[1]

$$\langle\Psi\rangle = \Psi_I = A_1\exp(i\mathbf{K}_1\cdot\mathbf{r}) + A_2\exp(i\mathbf{K}_2\cdot\mathbf{r}) = \Psi_1 + \Psi_2 = \sum\Psi_i, \tag{23}$$
$$\mathbf{K}_i = \hat{\mathbf{z}}\Gamma_i + \hat{\mathbf{x}}\tau = k\eta_i\hat{\mathbf{K}}_i = K_i\hat{\mathbf{K}}_i, \quad \mathrm{Im}\,\eta_i > 0$$

with $\hat{x}\cdot\mathbf{K}_i = \tau$ as discussed for (18). For brevity, we use $\Psi_I(\mathbf{r}) = \sum\Psi_i(\mathbf{r})$ with $i = 1, 2$. If the properties and distribution of the scatterers are symmetric to reflection in $z = 0$, then $\eta_1 = \eta_2$, $\Gamma_2 = -\Gamma_1$, and $\mathbf{K}_2 = \mathbf{K}_1' = -\hat{\mathbf{z}}\Gamma_1 + \hat{\mathbf{x}}\tau$ is the image of $\mathbf{K}_1$.

At least for $\mathbf{r}_s$ not within boundary layers (say, at least for $l < z_s < d - l'$), the functions $\langle\ \rangle_s$ consist of two terms with translational property $f_i(\mathbf{z}_s+\mathbf{r}') = f_i(\mathbf{r}') \times \exp(i\Gamma_i z_s)$. Thus, we may write

$$\langle\Psi_s(\mathbf{r}_s+\mathbf{r}')\rangle_s = \sum\psi^i(\mathbf{r}')\,\Psi_i(\mathbf{r}_s),$$
$$\langle\Phi_s\rangle_s = \sum\phi^i\Psi_i, \quad \langle U_s\rangle_s = \sum u^i\Psi_i, \tag{24}$$

where ψ^i, ϕ^i, u^i satisfy (3)–(8) as for a single obstacle and may be interpreted as the fields of an equivalent scatterer. Similarly

$$G(\mathbf{r}_s;\hat{\mathbf{r}}) = \sum g^i\Psi_i = \sum g(\mathbf{k}_r|\mathbf{K}_i)\,\Psi_i = \sum G_i(\mathbf{r}_s;\hat{\mathbf{r}}),$$
$$G_i(z_s:\hat{\mathbf{r}}) = g(\mathbf{k}_r|\mathbf{K}_i)A_i\exp(i\Gamma_i z_s), \tag{25}$$

where we use $g(\mathbf{k}_r|\mathbf{K}_i)$ or g^i to distinguish the present amplitudes from those of Ref. 2. We have $g(\mathbf{k}_r|\mathbf{K}_i) = \{\exp(-i\mathbf{k}_r\cdot\mathbf{r}'), u^i\} = [\![\exp(-\mathbf{k}_r\cdot\mathbf{r}'), \psi^i]\!]$, etc., as well as $g(-\mathbf{k}_a|\mathbf{K}_i) = \{\phi_a, u^i\} = \{\phi^i, u_a\}$, etc., with $\phi_a + u_a$ as the solution for the conventional isolated scatterer for arbitrary direction of incidence $\hat{\mathbf{r}}_a = \mathbf{k}_a/k$.

From (20), (23), and (25) we obtain

$$\Psi_I = E\phi + E'\phi' + \frac{2\gamma c}{i}\sum\frac{g[\![\mathbf{K}_i|\mathbf{K}_i]\!]}{K_i^2 - k^2}\,\Psi_i = \sum\Psi_i, \tag{26}$$

where $E = 0$ and $E' = 0$ are the extinction (of ϕ) and cancellation (of ϕ') relations:

$$E(l) = 0 = 1 + c\int_0^l G(z_s:\hat{k})\exp(-i\gamma z_s)\, dz_s$$
$$+ ic\exp(-i\gamma l)\sum\frac{G_i(l:\hat{\mathbf{k}})}{\Gamma_i - \gamma},$$
$$E'(l') = 0 = c\int_{d-l'}^{d} G(z_s:\hat{\mathbf{k}}')\exp(i\gamma z_s)\, dz_s$$
$$- ic\exp[i\gamma(d-l')]\sum\frac{G_i(d-l':\hat{\mathbf{k}}')}{\Gamma_i + \gamma}. \tag{27}$$

The coefficients of Ψ_i provide the boundary independent equations[9]

$$K_i^2 - k^2 = -i2c\gamma g[\![\mathbf{K}_i|\mathbf{K}_i]\!] = -(\rho/c_o)\, g[\![\mathbf{K}_i|\mathbf{K}_i]\!]$$
$$= -i\rho k\sigma_0\, g[\![\mathbf{K}_i|\mathbf{K}_i]\!], \tag{28}$$

a form obtained originally by Reiche[10] for spherical dipoles, and by Foldy[5] for monopoles. The present argument $[\![\mathbf{K}|\mathbf{K}]\!]$ indicates explicit restriction of the form to the volume integral representation of g as in (8), i.e.,

$$K_i^2 - k^2 = -(\rho/c_o)[\![\exp(-i\mathbf{K}_i\cdot\mathbf{r}), \psi^i]\!]$$
$$= \rho\int[(B'K'^2 - k^2)\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\psi^i(\mathbf{r}')$$
$$- (B'-1)\nabla\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\cdot\nabla\psi^i]\, d\mathfrak{V}(\mathbf{r}'). \tag{29}$$

The analog for $B' = 1$ with $K'^2 - k^2$ as the scattering potential was derived originally by Lax,[6] who regarded $g[\![\mathbf{K}|\mathbf{K}]\!]$ as proportional to the result for one scatterer in a medium specified by $\mathbf{K}$.

We rewrite (29) in terms of the external surface integral form (4) as[9]

$$K_i^2 - k^2 = -\frac{\rho g\{\mathbf{K}_i|\mathbf{K}_i\}}{c_o[1 - \rho\int\exp(-i\mathbf{K}_i\cdot\mathbf{r}')(\psi^i - \phi^i)\, d\mathfrak{V}]},$$
$$g\{\mathbf{K}_i|\mathbf{K}_i\} = \{\exp(-i\mathbf{K}_i\cdot\mathbf{r}'), u^i\}, \tag{30}$$

or in terms of the internal surface integral form (7) as

$$K_i^2 - k^2 = -\frac{\rho g[\mathbf{K}_i|\mathbf{K}_i]}{c_o[1 - \rho\int\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\,\psi^i\, d\mathfrak{V}]},$$
$$g[\mathbf{K}_i|\mathbf{K}_i] = [\exp(-i\mathbf{K}_i\cdot\mathbf{r}'), \psi^i]. \tag{31}$$

If the argument were $\mathbf{k}_r|\mathbf{K}_i$ corresponding to $g(\mathbf{k}_r|\mathbf{K}_i)$ of (25), the g forms in (29)–(31) would be equivalent; for argument $\mathbf{K}_i|\mathbf{K}_i$ they are not. The three equations for $K_i^2 - k^2$ suggest different interpretations in terms of unconventional isolated scatterers, as well as different expansion procedures in powers of k, K.

From (31) we have

$$K_i^2 - k^2 = -\rho\int[\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\nabla(\phi^i + u^i)$$
$$- (\phi^i + u^i)\nabla\exp(-i\mathbf{K}_i\cdot\mathbf{r}')]\cdot d\mathfrak{S}$$
$$+ \rho(K_i^2 - k^2)\int\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\,\psi^i\, d\mathfrak{V}, \tag{32}$$

which leads directly to analogous relations for impene-

trable scatterers, $\psi^i = 0$ in $\mathfrak{V}$. Thus if $\hat{\mathbf{n}} \cdot \nabla \psi = \partial_n \psi = 0$ on $\mathfrak{S}$, then

$$K_i^2 - k^2 = \rho \int (\phi^i + u^i)\, \partial_n \exp(-i\mathbf{K}_i \cdot \mathbf{r}')\, d\mathfrak{S}, \tag{33}$$

and if $\psi = 0$ on $\mathfrak{S}$,

$$K_i^2 - k^2 = -\rho \int \exp(-i\mathbf{K}_i \cdot \mathbf{r}')\, \partial_n (\phi^i + u^i)\, d\mathfrak{S}. \tag{34}$$

Both cases are covered by

$$\begin{aligned} K_i^2 - k^2 &= -(\rho/c_o)\{\exp(-i\mathbf{K}_i \cdot \mathbf{r}'),\, \phi^i + u^i\} \\ &\equiv -(\rho/c_o)\, g\{\mathbf{K}_i \,|\, \mathbf{K}_i\}'. \end{aligned} \tag{35}$$

From (21), (22), and (25) we obtain $\mathfrak{T}$ and $\mathfrak{R}$, and, after using (27),

$$\begin{aligned} \mathfrak{T} &= c \int_{d-l'}^{d} G(z_s : \hat{\mathbf{k}}) \exp(-i\gamma z_s)\, dz_s \\ &\quad - ic \exp[-i\gamma(d - l')] \sum \frac{G_i(d - l' : \hat{\mathbf{k}})}{\Gamma_i - \gamma}, \\ \mathfrak{R} &= c \int_0^{l} G(z_s : \hat{\mathbf{k}}')\exp(i\gamma z_s)\, dz_s \\ &\quad + ic \exp(i\gamma l) \sum \frac{G_i(l; \hat{\mathbf{k}}')}{\Gamma_i + \gamma}. \end{aligned} \tag{36}$$

With (20), (27), and (36) we could seek the field within the boundary layers. However, our primary purpose is to determine η within an unbounded distribution, and to obtain approximations which may be used to specify $\mathfrak{T}$ and $\mathfrak{R}$ for an equivalent homogeneous slab.

To facilitate discussion, we list results for the uniform slab in forms that arose in earlier[1,2] approximation procedures for the distribution. Thus, if the translational property (25) holds for $0 \leq z_s \leq d$, then, essentially as in (2:10),

$$\begin{aligned} &\Psi_I = A_1 \exp(i\mathbf{K}_1 \cdot \mathbf{r}) + A_2 \exp(i\mathbf{K}_2 \cdot \mathbf{r}), \quad A_1 = (1 - Q_1)/D, \\ &A_2 = Q_1(1 - Q_2)\exp[i(\Gamma_1 - \Gamma_2)\, d]/D, \\ &D = 1 - Q_1 Q_2 \exp[i(\Gamma_1 - \Gamma_2)\, d], \\ &Q_i = \frac{Z_i - 1}{Z_i + 1} = \frac{(\Gamma_i - \gamma_i)\, g(\mathbf{k}_i \,|\, \mathbf{K}_i)}{(\Gamma_i + \gamma_i)\, g(\mathbf{k}_i \,|\, \mathbf{K}_i)}, \\ &\gamma_i = \pm\gamma, \quad \mathbf{k}_i = \begin{Bmatrix} \mathbf{k} \\ \mathbf{k}' \end{Bmatrix}, \quad \mathbf{k}_i' = \begin{Bmatrix} \mathbf{k}' \\ \mathbf{k} \end{Bmatrix}, \end{aligned} \tag{37}$$

with Q_i as single-surface Fresnel reflection coefficients, and Z_i as the associated impedances. Similarly

$$\begin{aligned} \Psi_T &= (1 - Q_1 Q_2)\exp[i(\Gamma_1 - \gamma)\, d]\, \phi/D, \\ \Psi_R &= -Q_1\{1 - \exp[i(\Gamma_1 - \Gamma_2)\, d]\}\, \phi'/D, \end{aligned} \tag{38}$$

as in (1:3.14). Corresponding to (25), we have[2]

$$\begin{aligned} G(\mathbf{r}; \hat{\mathbf{r}}) &= \mathfrak{G}(\mathbf{k}_r \,|\, \mathbf{K}_1)\exp(i\mathbf{K}_1 - \mathbf{r})/D \\ &\quad + \mathfrak{G}(\mathbf{k}_r \,|\, \mathbf{K}_2)\, Q_1 \exp[i(\Gamma_1 - \Gamma_2)\, d]\exp(i\mathbf{K}_2 - \mathbf{r})/D, \end{aligned} \tag{39}$$

where $\mathfrak{G}(\mathbf{k}_r \,|\, \mathbf{K}_i) = (1 - Q_i)\, g(\mathbf{k}_r \,|\, \mathbf{K}_i)$ with $i = 1, 2$ for ϕ, ϕ' incident on the distributions $z \geq 0$, $z \leq 0$ respectively, is the multiple scattered amplitude of an obstacle at the phase origin of the half-space. For the half-space problems, the reflection and extinction relations are

$$\begin{aligned} &Q_n = i(-1)^n c\, \mathfrak{G}(\mathbf{k}_n' \,|\, \mathbf{K}_n)/(\Gamma_n + \gamma_n), \\ &1 = i(-1)^n c\, \mathfrak{G}(\mathbf{k}_n \,|\, \mathbf{K}_n)/(\Gamma_n - \gamma_n), \quad n = 1, 2. \end{aligned} \tag{40}$$

For $n = 1$, with $\mathbf{K}_1 = \mathbf{K}$, $\mathbf{Z}_1 = \mathbf{Z}$, we have, as in (2:14), (2:17),

$$\begin{Bmatrix} 1 \\ Z \end{Bmatrix} = \frac{cg(\mathbf{k} \,|\, \mathbf{K})}{i(\Gamma - \gamma)} \mp \frac{cg(\mathbf{k}' \,|\, \mathbf{K})}{i(\Gamma + \gamma)}; \tag{41}$$

for $n = 2$, we replace Z by $-Z_2$, and $\Gamma, \mathbf{K}$ by $\Gamma_2', \mathbf{K}_2$. The forms $\Psi[\Gamma, Q(Z)]$ are general results for a uniform slab, and only the expressions $Q(g)$ of (37), and $\Gamma(g)$ and $Z(g)$ of (41) (henceforth, the interface approximation), relate the results to the distribution of scatterers.

Writing Ψ_E for either Ψ_T or $\phi + \Psi_R$ for $z \geq d$ or $z \leq 0$ respectively, we see that at $z = 0, d$ the forms in (37) and (38) satisfy

$$\Psi_E = \Psi_I, \quad \partial_z \Psi_E = B_{11} \partial_z \Psi_I + B_{12} \partial_x \Psi_I, \tag{42}$$

where B_{11} and B_{12}, corresponding to B and $B'/i\mathbf{k} \cdot \hat{x}$ of (1:3.26), are specified by

$$\begin{aligned} &-(-1)^n Z_n = (B_{11}\Gamma_n + B_{12}\tau)/\gamma, \\ &\tau = \mathbf{k} \cdot \hat{\mathbf{x}} = \mathbf{K}_n \cdot \hat{\mathbf{x}} = k \sin\alpha, \quad \tau/\gamma = \tan\alpha, \end{aligned} \tag{43}$$

essentially as before[1].

In order to interpret the second condition in (42), we refer back to the scalar field ψ of (3), (5), and (6) and introduce an associated vector field w such that

$$\begin{aligned} &\mathbf{w} = \nabla\psi, \quad \nabla \cdot \mathbf{w} = -k^2\psi; \\ &\mathbf{w}' = B'\nabla\psi', \quad \nabla \cdot \mathbf{w}' = -k^2 C'\psi', \quad \eta'^2 = C'/B'; \\ &\psi = \psi', \quad \hat{\mathbf{n}} \cdot \mathbf{w} = \hat{\mathbf{n}} \cdot \mathbf{w}' \text{ on } \mathfrak{S}. \end{aligned} \tag{44}$$

Similarly, corresponding to $\langle\Psi\rangle$ we consider the vector field $\langle \mathbf{W} \rangle$ satisfying

$$\begin{aligned} &\mathbf{W}_E = \nabla\Psi_E, \quad \nabla \cdot \mathbf{W}_E = -k^2\Psi_E; \\ &\mathbf{W}_I = \tilde{\mathbf{B}} \cdot \nabla\Psi_I, \quad \nabla \cdot \mathbf{W}_I = -k^2 C\Psi_I; \\ &\Psi_E = \Psi_I, \ \hat{\mathbf{n}} \cdot \mathbf{W}_E = \hat{\mathbf{n}} \cdot \mathbf{W}_I = \eta \cdot \mathbf{B} \cdot \nabla\ \Psi_I \ \text{ at } z = 0,\ d; \\ &\mathbf{B} = \sum B_{ij} \hat{\mathbf{z}}_i \hat{\mathbf{z}}_j, \quad \hat{\mathbf{z}}_1 = \hat{\mathbf{z}}, \quad \hat{\mathbf{z}}_2 = \hat{\mathbf{x}}, \end{aligned} \tag{45}$$

i.e., the equivalent uniform slab is specified by two bulk parameters, the dyadic $\tilde{\mathbf{B}}$ and the scalar C.

From (45), we have

$$(\nabla^2 + k^2)\Psi_E = 0, \quad \nabla \cdot (\tilde{\mathbf{B}} \cdot \nabla\Psi_I) = \tilde{\mathbf{B}} : \nabla\nabla\Psi_I = -k^2 C\Psi_I. \tag{46}$$

Thus, corresponding to Ψ_I as in (23),

$$\begin{aligned} &\mathbf{K} \cdot \tilde{\mathbf{B}} \cdot \mathbf{K} = \hat{\mathbf{K}} \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{K}}\eta^2 k^2 = k^2 C, \\ &\Gamma_i^2 B_{11} + \Gamma_i \tau(B_{12} + B_{21}) + \tau^2 B_{22} - k^2 C = 0, \\ &\eta_i^2 = (\Gamma_i^2 + \tau^2)/k^2, \quad \Gamma_i/\tau \equiv \cot\beta_i. \end{aligned} \tag{47}$$

For a uniform slab, given $\tilde{\mathbf{B}}$ and C, we construct $\Gamma_i(\alpha)$, then $\eta_i(\alpha)$ and the complex angles $\beta_1(\alpha)$. For the present synthetic medium associated with the average wave $\langle\Psi\rangle$, once we have obtained $K_i^2 - k^2 = \Gamma_i^2 - \gamma^2$ and Z_i, we use (43) and (47) in determining C and $\tilde{\mathbf{B}}$. From (43) and $\Gamma_1 + \Gamma_2 = -\tau(B_{12} + B_{21})/B_{11}$, we obtain $Z_1 - Z_2$

$= (B_{12} - B_{21})\tan\alpha$, so that if $B_{12} = B_{21}$ (corresponding to inversion symmetry), then $Z_1 = Z_2 = Z$, and $Q_1 = Q_2 = Q$.

The interface expressions (38) for Ψ_T and Ψ_R in terms of $Z(g)$ of (41), or of the corresponding $\tilde{\mathbf{B}}, C$, may introduce discrepancies for the boundary effects. The representations for K_i as in (28)–(35) are independent of boundary layers, and we may derive analogous representations[9] for $\tilde{\mathbf{B}}$ that are similarly independent; such results for $\tilde{\mathbf{B}}$ apply for an unbounded distribution, and their use in Z_i to construct Ψ_T and Ψ_R of (38) may also introduce discrepancies for the boundary effects. Comparison of explicit approximations for $\tilde{\mathbf{B}}$ determine ranges of the parameters in which the two procedures give the same values, and in which the layer effects should be negligible.

Derivations of boundary independent representations for the bulk parameters B and C of isotropic distributions are discussed elsewhere in detail; in particular, the procedure (9 : 51)–(9 : 54) may be extended directly to the anisotropic case. Thus as before,[9] we write (17) within the distribution as

$$\langle\Psi(\mathbf{r})\rangle = \phi(\mathbf{r}) + \rho\int_{V-\mathfrak{V}}\langle U_s(k)\rangle_s d\mathbf{r}_s + \rho\int_{\mathfrak{V}}[\langle\Psi_s(K')\rangle_s - \langle\Phi_s(k)\rangle_s]d\mathbf{r}_s. \tag{48}$$

Similarly, the associated vector field $\langle\mathbf{W}\rangle$ is now

$$\tilde{\mathbf{B}}\cdot\nabla\langle\Psi(\mathbf{r})\rangle = \nabla\phi(\mathbf{r}) + \rho\int_{V-\mathfrak{V}}\nabla\langle U_s\rangle_s d\mathbf{r}_s + \rho\int_{\mathfrak{V}}[B'\nabla\langle\Psi_s\rangle_s - \nabla\langle\Phi_s\rangle_s]d\mathbf{r}_s \tag{49}$$

with ∇ acting on $\mathbf{r}$ in the supressed arguments $\mathbf{r} - \mathbf{r}_s$. The gradient of (48) equals

$$\nabla\langle\Psi\rangle = \nabla\phi + \rho\int_{V-\mathfrak{V}}\nabla\langle U_s\rangle_s d\mathbf{r} + \rho\int_{\mathfrak{V}}[\nabla\langle\Psi_s\rangle_s - \nabla\langle\Phi_s\rangle_s]d\mathbf{r}_s, \tag{50}$$

where the surface integrals $\pm\int\langle\Psi_s\rangle_s d\mathfrak{S}$ that arose in the interchange of ∇ and $\int d\mathbf{r}_s$ canceled by continuity of $\langle\Psi_s\rangle_s$ on $\mathfrak{S}$ (the surface traced by the centers of the closest scatterers to the fixed point $\mathbf{r}$, with $\mathbf{r} - \mathbf{r}_s = \mathbf{r}'$ as the full set of scatterer's surface points). Similarly the divergence of (49) yields

$$-\nabla\cdot(\tilde{\mathbf{B}}\cdot\nabla\langle\Psi\rangle) = k^2\phi + k^2\rho\int_{V-\mathfrak{V}}\langle U_s\rangle_s d\mathbf{r}_s + \rho\int_{\mathfrak{V}}[B'K'^2\langle\Psi_s\rangle_s - k^2\langle\Phi_s\rangle_s]d\mathbf{r}_s, \tag{51}$$

where the surface integrals $\pm\int\langle\mathbf{W}_s\rangle_s\cdot d\mathfrak{S}$ that arose in the interchange of $\nabla\cdot$ and $\int$ canceled by continuity of $\mathbf{n}\cdot\langle\mathbf{W}_s\rangle_s$ on $\mathfrak{S}$, and where we used $\nabla^2\langle\Psi_s\rangle_s = -K'\langle\Psi_s\rangle_s$ in the volume $\mathfrak{V}$ around $\mathbf{r}$ (a volume equal to that of one scatterer with $\mathbf{r} - \mathbf{r}_s = \mathbf{r}'$ as the full set of internal scatterer points).

From (48) and (51)

$$-[\nabla\cdot(\tilde{\mathbf{B}}\cdot\nabla) + k^2]\langle\Psi(\mathbf{r})\rangle = \rho(B'K'^2 - k^2)\int_{\mathfrak{V}}\langle\Psi_s\rangle_s d\mathbf{r}_s, \tag{52}$$

and from (49) and (50),

$$(\tilde{\mathbf{B}} - \tilde{\mathbf{I}})\cdot\nabla\langle\Psi\rangle = \rho(B' - 1)\int_{\mathfrak{V}}\nabla\langle\Psi_s\rangle_s d\mathbf{r}_s, \tag{53}$$

with $\tilde{\mathbf{I}}$ as the identity dyadic. Substituting $\langle\Psi\rangle = \sum\Psi_i(\mathbf{r}) = \sum A_i\exp(i\mathbf{k}_i\cdot\mathbf{r})$ of (23) and $\langle\Psi_s\rangle_s = \sum\Psi_i(\mathbf{r}_s)\psi^i(\mathbf{r} - \mathbf{r}_s) = \sum\Psi_i(\mathbf{r})\exp(-i\mathbf{k}_i\cdot\mathbf{r}')\psi^i(\mathbf{r}')$ of (24), we obtain from (52)

$$\mathbf{K}_i\cdot\tilde{\mathbf{B}}\cdot\mathbf{K}_i - k^2 = \rho(B'K'^2 - k^2)\int\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\psi^i(\mathbf{r}')\,d\mathfrak{V}(\mathbf{r}'),$$
$$C - 1 = (C' - 1)\rho\int\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\psi^i\,d\mathfrak{V} \tag{54}$$

with $\eta_i^2\hat{\mathbf{K}}_i\cdot\tilde{\mathbf{B}}\cdot\hat{\mathbf{K}}_i = C$, $\eta'^2B' = C'$. Similarly from (23) and (24) plus the additional step, $\nabla_r[\Psi_i(\mathbf{r}_s)\psi^i(\mathbf{r} - \mathbf{r}_s)] = \Psi_i(\mathbf{r}_s)\nabla\psi^i(\mathbf{r}')$, we obtain from (53)

$$\mathbf{K}_i\cdot(\tilde{\mathbf{B}} - \tilde{\mathbf{I}})\cdot\mathbf{K}_i = \rho(B' - 1)\int\exp(-i\mathbf{K}_i\cdot\mathbf{r}')(-i\mathbf{K}_i)\cdot\nabla\psi^i(\mathbf{r}')\,d\mathfrak{V}(\mathbf{r}') = (B' - 1)\rho\int\nabla\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\cdot\nabla\psi^i\,d\mathfrak{V}. \tag{55}$$

Thus, if $C' = 1$, then $C = 1$ and $\eta_i^{-2} = \hat{\mathbf{K}}_i\cdot\tilde{\mathbf{B}}\cdot\hat{\mathbf{K}}_i$ specifies propagation; if $B' = 1$, then $\tilde{\mathbf{B}} = \tilde{\mathbf{I}}$ and $\eta^2 = C$.

Subtracting (55) from (54), we again obtain $K_i^2 - k^2$ as in (29). The same result also follows[9] from (48) if we decompose the integral over $V - \mathfrak{V}$ as integrals over the boundary layers, say V_l and $V_{l'}$, plus the integral over $V - \mathfrak{V} - V_l - V_{l'} \equiv V_B$. In V_B, we use (24) and Green's theorem on $(K^2 - k^2)\exp(i\mathbf{K}\cdot\mathbf{r}_s)u^i = \exp(i\mathbf{K}\cdot\mathbf{r}_s)\nabla_s^2u^i - u^i\nabla_s^2\exp(i\mathbf{K}\cdot\mathbf{r}_s)$ to rewrite $\int_{V_B}$ as surface integrals over the layer surfaces ($z_s = l, d - l'$) and over $\mathfrak{S}$ (centered on $\mathbf{r}$). The layer integrals plus ϕ provide the extinction ($E = 0$) and cancellation ($E' = 0$) relations of (27), and the remainder, $\langle\Psi\rangle = \Psi_I$ of (26) expressed in terms of $\int_{\mathfrak{S}}$ and $\int_{\mathfrak{V}}$ reduces[9] to (29).

Alternatively, if we multiply (55) by k^2 and (54) by K_i^2, and then subtract, we have

$$\mathbf{K}_i\cdot\tilde{\mathbf{B}}\cdot\mathbf{K}_i(K_i^2 - k^2) = \rho\int[K_i^2(B'K'^2 - k^2)\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\psi^i(\mathbf{r}) - k^2(B' - 1)\nabla\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\cdot\nabla\psi^i]\,d\mathfrak{V}, \tag{56}$$

which may be obtained from (49) by generalizing the development (9 : 39) ff, i.e., by an analog of the Green's theorem procedure for obtaining $K_i^2 - k^2$ from (49). From (56), corresponding to the form (32)

$$\mathbf{K}_i\cdot\tilde{\mathbf{B}}\cdot\mathbf{K}_i(K_i^2 - k^2) = -\rho\int[K_i^2\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\nabla(\phi^i + u^i) - k^2(\phi^i + u^i)\nabla\exp(-i\mathbf{K}_i\cdot\mathbf{r}')]\cdot d\mathfrak{S} + \rho(K_i^2 - k^2)B'\int\nabla\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\cdot\nabla\psi^i\,d\mathfrak{V}. \tag{57}$$

For impenetrable scatterers, $B'\nabla\psi^i = 0$, we see that (33) for $\partial_n\psi(\mathfrak{S}) = 0$ corresponds to $\hat{\mathbf{K}}_i\cdot\tilde{\mathbf{B}}\cdot\hat{\mathbf{K}}_i = \eta_i^{-2}$, and (34) for $\psi(\mathfrak{S}) = 0$ to $\hat{\mathbf{K}}\cdot\tilde{\mathbf{B}}\cdot\hat{\mathbf{K}} = 1$.

4. DISPERSION RELATION

Equation (20) for $\langle\Psi\rangle$ in terms of $\langle\Psi_s\rangle_s$, or (26) for $\langle\Psi\rangle = \Psi_I$ in terms of $\langle G_s(\hat{\mathbf{r}})\rangle_s = G(\mathbf{r}_s;\hat{\mathbf{r}}) = \exp(i\tau x_s)G(z_s;\hat{\mathbf{r}})$ is essentially the first of the hierarchy integral relations for the ensemble. The ensemble average of (16) with scatterer l fixed corresponds to the second,

$$\langle G_l(\hat{\mathbf{r}})\rangle_l = g_l(\hat{\mathbf{r}},\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r}_l) + \rho\int d\mathbf{r}_s f(\mathbf{R}_{ls})\int_c g_l(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)\langle G_s(\hat{\mathbf{r}}_c)\rangle_{sl}\exp(i\mathbf{k}_i\cdot\mathbf{R}_{ls}),$$
$$\mathbf{R}_{ls} = \mathbf{r}_l - \mathbf{r}_s, \tag{58}$$

where $f(\mathbf{R})$ is zero for $R < |\mathbf{b}(\hat{\mathbf{R}})|$, say for $\mathbf{R}$ in $v(\mathbf{b})$ with v as the exclusion volume. The function $\langle G_s\rangle_{sl}$, the

average over all variables except $\mathbf{r}_s$ and $\mathbf{r}_t$, may be represented[8] in terms of $g_s, \langle G_t\rangle_{st}$, and an integral over $d\mathbf{r}_m$ of $\langle G_m\rangle_{stm}$ times the three-particle distribution function $q(\mathbf{R}_{ts}, \mathbf{R}_{sm})$. We truncate the hierarchy system by means of $\langle G_s\rangle_{st} \approx \langle G_s\rangle_s$. For identical scatterers, in terms of $\langle G_s\rangle_s = G(\mathbf{r}_s; \hat{\mathbf{r}})$, we thus replace (58) by

$$G(\mathbf{r}_t; \hat{\mathbf{r}}) = g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) \exp(i\mathbf{k}\cdot\mathbf{r}_t) + \rho \int_{V-v} d\mathbf{r}_s f(\mathbf{R}_{ts}) \int_c g(\hat{\mathbf{r}}, \hat{\mathbf{r}}_c) G(\mathbf{r}_s; \hat{\mathbf{r}}_c) \times \exp(i\mathbf{k}_c \cdot \mathbf{R}_{ts}), \tag{59}$$

which we use for aligned asymetrical obstacles, as well as for the average over alignment when the distribution of alignments is uniform and uncorrelated with position or separation. The approximation $\langle G_s\rangle_{st} \approx \langle G_s\rangle_s$, analogous to $\langle \Phi_s\rangle_{st} \approx \langle \Phi_s\rangle_s$ as used by Lax,[6] excludes various scattering processes for fixed sets of obstacles. Working with the symbolic form[8] $U_s(\mathbf{r}_a - \mathbf{r}_s) = U_s^a = u_s^a \cdot \Phi_s$, we restore the processes for two fixed obstacles by means of (8:48), i.e., $\langle \Phi_s\rangle_s \approx [1 - u_t^s \cdot u_s^t]^{-1} \cdot [\langle \Phi_s\rangle_s + u_t^s \cdot \langle \Phi_t\rangle_t]$; the first term corresponds to $G(\mathbf{r}_s; \hat{\mathbf{r}}_c)$ as in (59), the first two terms to $G(\mathbf{r}_s; \mathbf{r}_c) + \int_{c'} g(\hat{\mathbf{r}}_c, \hat{\mathbf{r}}_{c'}) G(\mathbf{r}_t; \hat{\mathbf{r}}_{c'}) \times \exp(i\mathbf{k}_{c'} \cdot \mathbf{R}_{st})$, etc. To restore scattering processes for three fixed obstacles would require double integrals $\int d\mathbf{r}_s \int d\mathbf{r}_m$ involving $f(\mathbf{R}_{ts}) f(\mathbf{R}_{sm})$ and $q(\mathbf{R}_{ts}, \mathbf{R}_{sm})$. See Keller[7] and Ref. 8 for procedures not based on the hierarchy integrals.

In (59) we write $\int_{V-v} f = \int + \int (f-1)$; for $\int$ we take boundary layers $(V_L, V_{L'})$ into account and use $V - v = V_L + V_{L'} + V'$; for $\int (f-1)$, because $f \sim 1$ as $R \sim \infty$, we replace V by V_∞ (all space) and integrate over $V_\infty - v = V''$. In V' and V'' we use G of (25), and reduce $\int_{V'}$ to integrals over the layer surfaces $(z_s = L, d - L')$ and exclusion surface S(b) [by Green's theorem, essentially as discussed for (48) in the paragraph following (55)]. The resulting coefficients of $g(\hat{\mathbf{r}}, \hat{\mathbf{k}})\phi(\mathbf{r}_t)$ and $g(\hat{\mathbf{r}}, \hat{\mathbf{k}}')\phi'(\mathbf{r}_t)$ are extinction $(E = 0)$ and cancellation $(E' = 0)$ relations as in (27), and the coefficients of $\Psi_t(\mathbf{r}_t)$ provide dispersion equations that determine K_t.

We write the dispersion equation in terms of integrals over the exclusion surface S(b) and the depleted volume $V_\infty - v(\mathfrak{b})$, such that $\mathrm{S}(\mathfrak{b}) = S\hat{\mathbf{n}}$ with $\hat{\mathbf{n}}$ outward from $v(\mathfrak{b})$. Thus

$$g(\mathbf{k}_r | \mathrm{K}_t) = -\frac{\rho}{(K_t^2 - k^2)c_o} \{\exp(-i\mathbf{K}_t \cdot \mathbf{R}), \mathfrak{U}_t\}_b + \rho \int_{V_\infty - v} [f(\mathbf{R}) - 1] \exp(-i\mathbf{K}_t \cdot \mathbf{R}) \mathfrak{U}_t \, d\mathbf{R}, \tag{60}$$

where $\mathfrak{U}$ is a radiative solution of $(\nabla_R^2 + k^2)\mathfrak{U} = 0$ in the form (10),

$$\mathfrak{U}(\mathbf{k}_r, k\mathbf{R}; \mathrm{K}_t) = \mathfrak{U}_t = \int_c g(\hat{\mathbf{r}}, \hat{\mathbf{r}}_c) g(\mathbf{k}_c | \mathrm{K}_t) \exp(i\mathbf{k}_c \cdot \mathbf{R}) = \int_c F(\mathbf{k}_r, \mathbf{k}_c | \mathrm{K}_t) \exp(i\mathbf{k}_c \cdot \mathbf{R}) \tag{61}$$

with F as the associated radiation amplitude. In terms of the Green's surface integral form (4), we have

$$F(\mathbf{k}_r, \mathbf{k}_a | \mathrm{K}_t) = g(\hat{\mathbf{r}}, \hat{\mathbf{r}}_a) g(\mathbf{k}_a | \mathrm{K}_t) = \{\exp(-i\mathbf{k}_a \cdot \mathbf{R}), \mathfrak{U}_t(\mathbf{R})\}_b \tag{62}$$

and in (60), we identify the the analog of the form in (30),

$$F\{\mathbf{k}_r, \mathrm{K} | \mathrm{K}\} \equiv \{\exp(-i\mathrm{K}\cdot\mathbf{R}), \mathfrak{U}\}_b. \tag{63}$$

From (61) and (63) we rewrite (60) in terms of F, and determine K_t by

$$g(\mathbf{k}_r | \mathrm{K}_t) = -\frac{\rho F\{\mathbf{k}_r, \mathrm{K}_t | \mathrm{K}_t\}}{(K_t^2 - k^2)c_o} + \int_c F(\mathbf{k}_r, \mathbf{k}_c | \mathrm{K}_t) M(\mathbf{k}_c, \mathrm{K}_t),$$

$$M(\mathbf{k}, \mathrm{K}) = \rho \int_{V_\infty - v} [f(\mathbf{R}) - 1] \exp[i(\mathbf{k} - \mathrm{K}) \cdot \mathbf{R}] \, d\mathbf{R}. \tag{64}$$

Although $\mathfrak{U}$ is not defined for R in $v(\mathfrak{b})$, for radially symmetric statistics we may use eigenfunction series and replace the integral over $S(b)$ by the result for $b \to 0$ with compensating extension of the integral over $V_\infty - v(b)$ to V_∞ in the sense of a principal value.

Before proceeding to detailed applications, we indicate the relations of (64) to existing approximations. If we write

$$F\{\mathbf{k}_r, \mathrm{K} | \mathrm{K}\} = g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) g(\mathbf{k} | \mathrm{K}) + \{[\exp(-i\mathrm{K}\cdot\mathbf{R}) - \exp(-i\mathbf{k}\cdot\mathbf{R})], \mathfrak{U}\}_b,$$

then the leading term of (64) for $\hat{\mathbf{r}} = \hat{\mathbf{k}}$ and $\rho \approx 0$, $M = 0$ is

$$K^2 - k^2 \approx (\rho/c_o) g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = -\rho i k \sigma_O g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) \equiv 2k(K_R - k),$$
$$2\,\mathrm{Im}K_R = -\rho\sigma_O \mathrm{Re} g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = \rho(\sigma_A + \sigma_S), \tag{65}$$
$$\mathrm{Re}K_R/k = 1 + (\rho\sigma_O/2k)\,\mathrm{Im} g(\hat{\mathbf{k}}, \hat{\mathbf{k}}),$$

where, as discussed before,[1] K_R is essentially Rayleigh's result[11] generalized to arbitrary scatterers.

To modify $\mathrm{Im}K_R$ and include correlations, we approximate the complex Hankel-type integral $\int_c FM$ of (64) by the real Bessel type, and replace $g(\mathbf{k}_r | \mathrm{K})$ by $g(\hat{\mathbf{r}}, \hat{\mathbf{k}})$. For $\mathrm{Im}g \gg \mathrm{Re}g$, we obtain

$$2\mathrm{Im}K \approx \rho\sigma_A + \rho\sigma_O \mathfrak{M}[|g(\hat{\mathbf{r}}, \hat{\mathbf{k}})|^2 W(\hat{\mathbf{r}}, \hat{\mathbf{k}})],$$
$$W(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = 1 + \rho \int [f(\mathbf{R}) - 1] \exp[ik(\hat{\mathbf{r}} - \hat{\mathbf{k}}) \cdot \mathbf{R}] \, d\mathbf{R}. \tag{66}$$

For $f(\mathbf{R}) = f(R)$ equal to the radial distribution function, $W(\hat{\mathbf{r}}, \hat{\mathbf{k}})$ is a standard form in x-ray diffraction by liquids. For average spacing between centers small compared to λ, to lowest order in k,

$$2\mathrm{Im}K \approx \rho(\sigma_A + \sigma_S W), \quad W \equiv 1 + \rho \int [f(\mathbf{R}) - 1] \, d\mathbf{R}, \tag{67}$$

as applied elsewhere in detail.[12]

Alternatively, if we write

$$F\{\mathbf{k}_r, \mathrm{K} | \mathrm{K}\} = g(\hat{\mathbf{r}}, \hat{\mathrm{K}}) g(\mathbf{k}_K | \mathrm{K}) + \{(\exp(-i\mathrm{K}\cdot\mathbf{R}) - \exp(-ik\hat{\mathrm{K}}\cdot\mathbf{R}), \mathfrak{U}\}, \quad \mathbf{k}_K = k\hat{\mathbf{K}},$$

and substitute the leading term into (64) for $\hat{\mathbf{r}} = \hat{\mathbf{K}}$, then

$$K^2 - k^2 \approx -\frac{\rho}{c_o} g(\hat{\mathbf{K}}, \hat{\mathbf{K}}) \left[1 - \frac{1}{g(\mathbf{k}_K | \mathrm{K})} \times \int_c g(\hat{\mathbf{K}}, \hat{\mathbf{r}}_c) g(\mathbf{k}_c | \mathrm{K}) M(\mathbf{k}_c, \mathrm{K})\right]^{-1}, \quad \mathbf{k}_c = k\hat{\mathbf{r}}_c. \tag{68}$$

Essentially this form in terms of one g function (with arguments K, K and K, $\mathbf{k}_c$) is given by Lax, Ref. 6(b), in (3.19) and (4.10). If we approximate $g(\mathbf{k}_r | \mathrm{K})$ in (68) by

$g(\hat{\mathbf{r}}, \hat{\mathbf{K}})$ and keep only the leading terms, we have

$$K^2 - k^2 \approx -(\rho/c_o)[g(\hat{\mathbf{K}}, \hat{\mathbf{K}}) + \int_c g(\hat{\mathbf{K}}, \hat{\mathbf{r}}_c)\, g(\hat{\mathbf{r}}_c, \hat{\mathbf{K}})\, M(\mathbf{k}_c, \mathbf{K})],$$

which reproduces (66) under the same restrictions.

In the following, we apply (64) to specific situations, and obtain more complete results for η than given above. We consider first cylindrical scatterers (which show the full structure of the development in a relatively simple context), then completely bounded obstacles, and then slabs (for which an abbreviated development suffices).

5. DISTRIBUTIONS OF CYLINDERS

Initially we consider radially symmetric pair statistics, such that $\mathbf{b} = b\hat{\mathbf{R}}$ corresponds to a circle, and $f(R)$ is the radial distribution function for identical disks of radius b. The scatterers with surface $\mathbf{a}(\hat{\mathbf{R}})$, $|\mathbf{a}| \leq b/2$ are not necessarily radially symmetric; we consider both circular, and aligned noncircular scatterers. Then we consider anisotropic statistics such that $\mathbf{b}(\hat{\mathbf{R}})$ corresponds to aligned elliptic disks.

A. Radially symmetric statistics

We expand $F(\mathbf{k}_r, \mathbf{k}_R | \mathbf{K})$ of (62), with $\hat{\mathbf{r}} = \hat{\mathbf{k}}(\theta)$, $\hat{\mathbf{R}} = \hat{\mathbf{k}}(\Theta)$ and $\mathbf{K} = K\hat{\mathbf{K}} = K\hat{\mathbf{k}}(\beta)$, as a Fourier series in Θ,

$$F(\mathbf{k}_r, \mathbf{k}_R | \mathbf{K}) = g(\hat{\mathbf{r}}, \hat{\mathbf{R}})\, g(\mathbf{k}_R | \mathbf{K}) = \sum_{n=-\infty}^{\infty} B_n(\theta, \beta) \exp(in\Theta),$$
$$B_n = (1/2\pi) \int_0^{2\pi} F \exp(-in\Theta)\, d\Theta. \tag{69}$$

The corresponding series for the radiative function (61) is $\mathfrak{U} = \sum B_n H_n^{(1)}(kR) \exp(in\Theta)\, i^n$, which we use together with $\exp(-i\mathbf{K}\cdot\mathbf{R}) = \sum i^{-n} J_n(KR) \exp[in(\beta - \Theta)]$ in (63) to obtain

$$F\{\mathbf{k}_r, \mathbf{K} | \mathbf{K}\} = \{\exp(-i\mathbf{K}\cdot\mathbf{R}),\ \mathfrak{U}\}_b = \sum B_n \exp(in\beta)\, w_n,$$
$$w_n = \frac{2\pi}{i4}[J_n(Kb)\, \partial_b H_n^{(1)}(kb) - H_n^{(1)}\, \partial_n J_n]\, b$$
$$= \eta^{|n|} + (K^2 - k^2) \int_0^b J_n(KR) H_n^{(1)}(kR)\, R\, dR, \tag{70}$$

where w_n would equal the Wronskian (normalized to unity) were K equal to k. In illustration of the statement after (64), we expressed w_n as the contribution $(\eta^{|n|})$ for $b \to 0$ plus the corresponding volume integral.

Thus, from (60) or (64),

$$g(\mathbf{k}_r | \mathbf{K}) = -\sum B_n(\theta, \beta) \exp(in\beta)\, T_n,$$
$$T_n = T_{-n} = \eta^{|n|}/\Delta - \mathfrak{S}_n, \quad \Delta = (\eta^2 - 1)/\mathfrak{c}, \quad \mathfrak{c} = i\rho 4/k^2,$$
$$\mathfrak{S}_n = 2\pi\rho \int_0^{\infty} [f(R) - 1] J_n(KR) H_n^{(1)}(kR)\, R\, dR, \tag{71}$$

which we reduce to a Fourier series in θ. From (69) in terms of

$$g(\mathbf{k}_R | \mathbf{K}) = \sum a_n^i(\beta) \exp(in\Theta), \quad g(\hat{\mathbf{r}}, \hat{\mathbf{R}}) = \sum a_n(\Theta) \exp(in\theta),$$
$$a_n(\Theta) = \sum_m a_{nm} \exp(-im\Theta),$$

we have

$$B_n(\theta, \beta) = \sum_{\nu m} \exp(i\nu\theta)\, a_{\nu m}\, a_{n+m}^i(\beta),$$

and using this series and $g(\mathbf{k}_r | \mathbf{K}) = \sum a_n^i(\beta) \exp(in\theta)$ in (71), we obtain

$$a_m^i(\beta) = -\sum_{nl} a_{ml} a_{n+l}^i \exp(in\beta)\, T_n$$
$$= -\sum_{\nu l} a_{ml} a_\nu^i \exp[i(\nu - l)\beta]\, T_{\nu - l}. \tag{72}$$

The roots of the determinantal equation of this homogeneous system specify η.

Corresponding to $H = J + iN$ within the integrand in (71), we write $\mathfrak{S}_n = \mathfrak{I}_n + i\mathfrak{N}_n$. For small k,

$$\mathfrak{I}_0 \approx 2\pi\rho \int (f - 1)\, R\, dR \equiv W - 1,$$

$$\mathfrak{I}_n \approx \frac{2\pi}{(n!)^2} \left(\frac{k}{2}\right)^{2n} \eta^n \rho \int_0^{\infty} (f - 1)\, R^{2n+1}\, dR,$$

$$\mathfrak{N}_0 \approx -4\rho \int (f - 1) \left[\ln \frac{2}{c'kR}\right] R\, dR, \quad c' = 1.781\cdots,$$

$$\mathfrak{N}_n \approx \frac{-\eta^n 2\pi\rho}{n\pi} \int (f - 1)\, R\, dR = \frac{-\eta^n (W - 1)}{n\pi}, \tag{73}$$

where the next terms of $\mathfrak{I}_n$ are $O(k^{2n+2})$ and those of $\mathfrak{N}_n$ are $O(k^2)$. We consider situations where the role of η in $\mathfrak{S}_n$ is minor and do not discuss it explicitly; we isolate major contributions in forms $\eta(\mathfrak{S})$ which may then be refined by iteration.

The system of equations (72) plus an additional relation determine the set a_n^i. From (29) we have

$$\eta^2 - 1 = -\mathfrak{c}\, g[\![\mathbf{K} | \mathbf{K}]\!],$$
$$g[\![\mathbf{K} | \mathbf{K}]\!] = [\![\exp(-i\mathbf{K}\cdot\mathbf{r}'), \psi^i]\!] \equiv \sum \bar{a}_n^i(\beta) \exp(in\beta), \tag{74}$$

where the relation of $\bar{a}_n^i$ to a_n^i is as for the corresponding isolated scatterer coefficients obtained by replacing ψ^i, u^i by ψ, u in the volume, surface integral forms. To determine the consistency of the interface approximation (41), we use

$$\eta^2 - 1 = -[(\Gamma + \gamma) g(\mathbf{k} | \mathbf{K}) - (\Gamma - \gamma) g(\mathbf{k}' | \mathbf{K})]\, \mathfrak{c}/2\gamma$$
$$\equiv -(\Gamma g_- + \gamma g_- + \gamma g_+)\, \mathfrak{c}/2\gamma,$$
$$g_\pm = \sum_{-\infty}^{\infty} a_n^i [\exp(in\alpha) \pm \exp(in(\pi - \alpha))] \tag{75}$$

Radially symmetric cylinders

If the cylinders are intrinsically radially symmetric, or have been symmetrized by averaging over orientation, we write the isolated scattering amplitude as $g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = g(\theta - \alpha)$,

$$g(\theta - \alpha) = \sum_{n=0}^{\infty} a_n \cos n(\theta - \alpha), \quad a_n = \frac{a_n'}{1 - a_n'/\epsilon_n}, \quad \epsilon_n = 2 - \delta_{n0},$$

such that for lossless scatterers $\mathrm{Re}\, a_n' = 0$, $-\epsilon_n \mathrm{Re}\, a_n = |a_n|^2$. Using $a_{nm} = a_{|n|}\delta_{nm}/\epsilon_n$ and $a_n^i = A_n \exp(-in\beta)/\epsilon_n$ in (72), we obtain

$$-A_n/a_n = \sum_{\nu=-\infty}^{\infty} A_\nu T_{\nu-n}/\epsilon_n = A_0 T_n + \tfrac{1}{2} \sum_{\nu=1}^{\infty} (A_\nu T_{\nu-n} + A_{-\nu} T_{\nu+n}), \tag{76}$$

which decomposes into separate systems for

$A_{|n|} \pm A_{-|n|} \equiv A_n^{\pm}$:

$$-A_n^+/a_n = A_0^+ T_n + \tfrac{1}{2}\sum_1^\infty A_\nu^+(T_{\nu-n} + T_{\nu+n}),$$
$$-A_n^-/a_n = \tfrac{1}{2}\sum_1^\infty A_\nu^-(T_{\nu-n} - T_{\nu+n}). \tag{77}$$

We make the dependence on $(\eta^2-1)^{-1} = (c\Delta)^{-1}$ explicit by

$$\tfrac{1}{2}(T_{\nu-n} + T_{\nu+n}) = \eta^{\nu+n}\Delta^{-1} - \mathfrak{H}_{\nu,n},$$
$$\tfrac{1}{2}(T_{\nu-n} - T_{\nu+n}) = -\mathfrak{H}_{\nu,n},$$
$$\mathfrak{H}_{n,\nu} = \mathfrak{H}_{\nu,n} = \tfrac{1}{2}[c\eta^{\nu-n}\sum_{m=0}^{n-1}\eta^{2m} + \mathfrak{H}_{\nu-n} + \mathfrak{H}_{\nu+n}],$$
$$\mathfrak{H}^-_{\nu,n} = \mathfrak{H}_{\nu,n} - \mathfrak{H}_{\nu+n}, \quad \mathfrak{H}_{0,n} = \mathfrak{H}_n, \tag{78}$$

where we take $n < \nu$ for the sum over m. Except for very special eigenvalues, which we discount, the second system in (77) requires $A_n^- = 0$. Thus $A_{-n} = A_n$, and (76) simplifies to

$$-A_n/a_n = \tfrac{1}{2}\sum_0^\infty A_\nu(T_{\nu-n} + T_{\nu+n}) = \sum_0^\infty A_\nu(\eta^{\nu+n}\Delta^{-1} - \mathfrak{H}_{\nu,n}). \tag{79}$$

This homogeneous system may be converted directly to an inhomogeneous system whose solution determines Δ; however, we reserve this step to facilitate comparison with generalizations.

The multiple scattering amplitudes corresponding to $g(\theta - \alpha)$ may be written as

$$g(\mathbf{k}_r|\mathbf{K}) = \sum_{n=0}^\infty A_n \cos n(\theta-\beta) = g^i,$$
$$-\Delta = g[\![\mathbf{K}|\mathbf{K}]\!] = \sum_0^\infty \bar{A}_n \equiv \sum A_n d_n = \bar{g}^i, \tag{80}$$

where $d_n = \bar{a}_n/a_n$ is the ratio of isolated scatterer coefficients specified by

$$g[\![\mathbf{k}_r, \mathbf{k}_K]\!] = [\![\exp(-ik\hat{\mathbf{r}}\cdot\mathbf{r}'), \psi(\hat{\mathbf{K}})]\!] = \sum a_n \cos n(\theta-\beta),$$
$$g[\![\mathbf{K}_r, \mathbf{k}_K]\!] = [\![\exp(-iK\hat{\mathbf{r}}\cdot\mathbf{r}'), \psi]\!] = \sum \bar{a}_n \cos n(\theta-\beta).$$

The explicit dependence of $g(\mathbf{k}_r|\mathbf{K}) = g^i$ on angles is the same as $g(\hat{\mathbf{r}}, \hat{\mathbf{K}}) = g(\theta-\beta)$, and since T_n does not involve angles, it follows from (79) that A_n is independent of β; thus g^i has the same symmetry property as g.

If only the monopole coefficient a_0 is significant, then from $A_0 = -a_0 A_0 T_0$,

$$\eta^2 - 1 = -c\mathfrak{A}_0, \quad \mathfrak{A}_0 = \frac{a_0}{1-a_0\mathfrak{H}_0} = \frac{a_0'}{1-a_0'(1+\mathfrak{H}_0)} \tag{81}$$

and $\bar{g}^i = \bar{A}_0 = \mathfrak{A}_0 \equiv p_0$. For lossless scatterers, $-\mathrm{Re}A_0 = |\mathfrak{A}_0|^2 \mathrm{Re}(1+\mathfrak{H}_0) \equiv |\mathfrak{A}_0|^2 W_0$ or equivalently $-\mathrm{Re}p_0 = |p_0|^2 W_0$ with $W_0 = W + O(k^2)$ in terms of W of (67) and (73). If only the dipole a_1 is significant, then from $A_1 = -a_1 A_1 \frac{1}{2}(T_0 + T_2)$,

$$\frac{\eta^2-1}{\eta^2+1} = -\frac{ca_1}{1-a_1\frac{1}{2}(\mathfrak{H}_0+\mathfrak{H}_2)}; \tag{82}$$

if $k \to 0$ for circular cylinders ($\mathfrak{B} = \pi a^2$) specified by $\eta'^2 = B'^{-1}$, then the right side equals $\rho\mathfrak{B}(\eta'^2-1)/(\eta'^2+1)$ and (82) reduces to the two-dimensional version of the Maxwell (and Clausius, Mossotti, Lorenz, Lorentz) form. From (82), we obtain

$$\eta^{-2} = 1 + c\mathfrak{A}_1, \quad \mathfrak{A}_1 = \frac{a_1}{1-a_1\mathfrak{H}_{1,1}} = \frac{a_1'}{1-a_1'\frac{1}{2}(c+1+\mathfrak{H}_0+\mathfrak{H}_2)},$$
$$\mathfrak{H}_{1,1} = \tfrac{1}{2}(c+\mathfrak{H}_0+\mathfrak{H}_2), \tag{83}$$

and consequently $-\Delta = \bar{g}^i = \bar{A}_1 = \mathfrak{A}_1\eta^2 \equiv p_1 = \mathfrak{A}_1/(1+c\mathfrak{A}_1)$. For lossless scatterers, $-2\,\mathrm{Re}\,\mathfrak{A}_1 = |\mathfrak{A}_1|^2 \mathrm{Re}(1+\mathfrak{H}_0+\mathfrak{H}_2) \equiv |\mathfrak{A}_1|^2 W_{11}$, from which $-2\,\mathrm{Re}p_1 = |p_1|^2 W_{11}$ with $W_{11} = W + O(k^2)$. More generally, if we retain only the $n2$ pole (of interest if a_n has a resonant response), then

$$\eta^2 - 1 = -c\mathfrak{A}_n\eta^{2n} = -cp_n, \quad \mathfrak{A}_n = \frac{a_n}{1-a_n\mathfrak{H}_{n,n}},$$
$$2\mathfrak{H}_{n,n} = c\sum_{m=0}^{n-1}\eta^{2m} + \mathfrak{H}_0 + \mathfrak{H}_{2n} \equiv \mathfrak{H}_0 + \mathfrak{H}'_{n,n}, \tag{84}$$

and $\bar{g}^i = \bar{A}_n = p_n$. For lossless scatterers, $-2\,\mathrm{Re}\mathfrak{A}_n = |\mathfrak{A}_n|^2 W_{nn}$, with $W_{nn} = \mathrm{Re}(1+\mathfrak{H}_0+\mathfrak{H}'_{n,n})$.

If both a_0 and a_1 are significant, then, from (79),

$$(1+a_0T_0)A_0 + a_0T_1A_1 = 0, \quad a_1T_1A_0 + [1+a_1(T_0+T_2)/2]A_1 = 0.$$

Equivalently, in terms of the self-coupling coefficients $p_0 = \mathfrak{A}_0$ of (81) and $p_1 = \mathfrak{A}_1\eta^2$ of (83),

$$(\Delta+p_0)A_0 + p_0(1-h_{01}\Delta)A_1\eta = 0,$$
$$p_1(1-h_{01}\Delta)A_0 + (\Delta+p_1)A_1\eta = 0, \quad h_{01} \equiv \mathfrak{H}_1/\eta. \tag{85}$$

Discounting the root $\Delta = 0$, the determinantal equation gives

$$-\Delta = -\frac{\eta^2-1}{c} = \frac{p_0+p_1+2p_0p_1h_{01}}{1-p_0p_1h_{01}^2} = \bar{g}^i. \tag{86}$$

If we neglect h_{01}, then $\eta^2 \approx 1 - c(p_0+p_1) = (1-cA_0)/(1+cA_1) = \eta_0^2\eta_1^2$ with η_0^2, η_1^2 as in (81), (83). An alternative development for fine circular cylinders is given by Bose and Mal.[13]

To interpret (86), we write

$$\bar{g}^i = P_0 + P_1, \quad P_0 = a_0(1+P_0h_{00}+P_1h_{01}),$$
$$P_1 = a_1\eta^2(1+P_0h_{01}+P_1h_{11}), \quad h_{m\nu} \equiv \mathfrak{H}_{n,\nu}/\eta^{n+\nu}, \tag{87}$$

where P_0 is the net monopole response with dipoles present and similarly P_1 is the net dipole response including coupling with monopoles; the wave $h_{01} = \mathfrak{H}_{0,1}/\eta$ is the monopole–dipole coupling factor. Consequently,

$$P_0 = p_0(1+p_1h_{01})/D, \quad P_1 = p_1(1+p_0h_{01})/D,$$
$$D = 1 - p_0p_1h_{01}^2, \quad p_n = a_n\eta^{2n}/(1-a_n\eta^{2n}h_{nn}) \tag{88}$$

such that $P_0 + P_1 = \bar{g}^i = -\Delta$ as in (86).

We generalize the development by rewriting (79) as the inhomogeneous system

$$P_n = a_n\eta^{2n}\left(1 + \sum_{\nu=0}^\infty h_{m\nu}P_\nu\right) = p_n(1+\textstyle\sum'_\nu h_{m\nu}P_\nu),$$
$$P_n = \frac{-A_n\eta^n\Delta}{\sum A_\nu\eta^\nu}, \quad \sum P_n = -\Delta, \tag{89}$$

where $\sum'_\nu = \sum_{\nu\neq n}$. If we retain up to quadrupole terms, then

$$P_0 = p_0[1 + p_1h_{01} + p_2h_{02} + p_1p_2h_{12}(h_{01}+h_{02}-h_{12})]/D,$$
$$D = 1 - p_0p_1h_{01}^2 - p_0p_2h_{02}^2 - p_1p_2h_{12}^2 - 2p_0p_1p_2h_{01}h_{12}h_{02} \tag{90}$$

with P_1 and P_2 following by cyclic permutations of 0, 1, 2. We have $p_n = \mathfrak{A}_n \eta^{2n} = a_n \eta^{2n}/(1 - a_n \mathfrak{H}_{n,n})$ with $\mathfrak{H}_{0,0} = \mathfrak{H}_0$, $\mathfrak{H}_{1,1} = \frac{1}{2}(\mathfrak{c} + \mathfrak{H}_0 + \mathfrak{H}_2)$, and $\mathfrak{H}_{2,2} = \frac{1}{2}[\mathfrak{c}(\eta^2 + 1) + \mathfrak{H}_0 + \mathfrak{H}_4)]$; the coupling factors are $h_{01} = \mathfrak{H}_1/\eta$, $h_{02} = \mathfrak{H}_2/\eta^2$, and $\mathfrak{H}_{12} = (\mathfrak{c}\eta + \mathfrak{H}_1 + \mathfrak{H}_3)/2\eta^3$. Thus, the index of refraction is determined by

$$
\begin{aligned}
-\Delta D &= (P_0 + P_1 + P_2) D \\
&= p_0 + p_1 + p_2 + 2(p_0 p_1 h_{01} + p_0 p_2 h_{02} + p_1 p_2 h_{12}) \\
&\quad + p_0 p_1 p_2 (2h_{01}h_{02} + 2h_{01}h_{12} + 2h_{02}h_{12} - h_{01}^2 - h_{02}^2 - h_{12}^2).
\end{aligned}
\tag{91}
$$

Equation (89) determines $P_n(a_n)$ as well as $\sum \bar{A}_n$, i.e.,

$$
-\Delta = -\frac{\eta^2 - 1}{\mathfrak{c}} = \bar{g}^t = g[\![\mathbf{K}|\mathbf{K}]\!] = \sum \bar{A}_n = \sum P_n, \tag{92}
$$

but $\bar{A}_n \neq P_n$ (except when P_n reduces to p_n for the cases of a single multipole). Equation (89) also determines all but one of the coefficients A_n of $g(\mathbf{k}_r|\mathbf{K})$ of (80),

$$
\begin{aligned}
&\frac{A_n \eta^n}{P_n} = \frac{\sum A_\nu \eta^\nu}{\sum P_m} = \frac{A_0}{P_0}, \\
&g(\mathbf{k}_r|\mathbf{K}) = \frac{A_0}{P_0} \sum \frac{P_n}{\eta^n} \cos n(\theta - \beta),
\end{aligned}
\tag{93}
$$

which together with $\sum \bar{A}_n = \sum A_n d_n$ of (80) determines the remaining coefficient A_0. Thus from $-\Delta = \sum P_n$ and

$$
-\Delta = \sum A_n d_n = A_0 \sum P_n d_n/\eta^n P_0 = -(\eta^2 - 1)/\mathfrak{c} \tag{94}
$$

we express A_0 in terms of the known P_n and ratio of single scattered coefficients d_n:

$$
A_0 = \frac{P_0 \sum P_m}{\sum P_n d_n/\eta^n}, \quad d_n = \frac{[\![J_n(Kr') \exp(-in\theta'), \psi(\mathbf{r}')]\!]}{[\![J_n(kr') \exp(-in\theta'), \psi]\!]}. \tag{95}
$$

For circular cylinders,

$$
\begin{aligned}
&d_n = V_n(K, K')/V_n(k, K'), \\
&V_n(K, K') = -(C' - 1) I_n(K, K') + (B' - 1) L_n(K, K'), \\
&\qquad\qquad C' = B'\eta'^2, \\
&(\eta^2 - \eta'^2) \begin{Bmatrix} I_n \\ L_n \end{Bmatrix} = \begin{Bmatrix} 1 \\ \eta^2 \end{Bmatrix} J_n(K, K') - \begin{Bmatrix} 1 \\ \eta'^2 \end{Bmatrix} J_n(K', K), \\
&J_n(K, K') = aJ_n(Ka)\, \partial_a J_n(K'a),
\end{aligned}
\tag{96}
$$

where $V_n(k, K')$, corresponding to $V_n(K, K')$ with K, η replaced by $k, 1$ may be written more simply as $J_n(k, K')B' - J_n(K', k)$. For one-parameter scatterers, (92) also determines the corresponding bulk parameter; if $C' = 1$, then $d_n = L_n(K, K')/L_n(k, K')$ in (94), and (92) corresponds to $\eta^2 = 1/B(P)$; if $B' = 1$, then $d_n = I_n(K, K')/I_n(k, K')$, and $\eta^2 = C(P)$. Similarly the impenetrable cases are specified by one parameter, η^2 of (92); if $\partial_a \psi(a) = 0$ as in (33), then

$$
d_n = \partial_a J_n(Ka)/\partial_a J_n(ka) \tag{97}
$$

in (94), and if $\psi(a) = 0$ as in (34),

$$
d_n = J_n(Ka)/J_n(ka). \tag{98}
$$

The scattering amplitudes $\bar{g}$ in terms of $\bar{a}_n = d_n a_n$ are analogs of the two-space amplitudes considered in Ref. 2.

For two-parameter ($B' \neq 1$, $C' \neq 1$) cylinders, from (54) and (55) we write the corresponding bulk parameters as

$$
\frac{C - 1}{\mathfrak{c}} = \sum A_n d_n^C = \frac{A_0}{P_0} \sum \frac{P_n d_n^C}{\eta^n}, \quad d_n^C = (C' - 1) \frac{I_n(K, K')}{V_n(k, K')}, \tag{99}
$$

$$
\frac{B - 1}{\mathfrak{c}} = \sum A_n d_n^B = \frac{A_0}{P_0} \sum \frac{P_n d_n^B}{\eta^n}, \quad d_n^B = (B' - 1) \frac{L_n(K, K')}{V_n(k, K')}, \tag{100}
$$

such that $-d_n^C + d_n^B = d_n$ of (96); using (94), we eliminate A_0/P_0. If we retain up to quadrupoles, then, for the usual isolated scattering problems, (91) determines K to order $x^4 = (ka)^4$. The corresponding results for the parameters are

$$
\begin{aligned}
&(C - 1)/\mathfrak{c} = -P_0 - (x^2/8)[P_0 \eta'^2 \mathfrak{r}(1 - \mathfrak{c}P_0) - 2P_1/\mathfrak{r}] + O(x^6), \\
&\mathfrak{r} = (B' - 1)/(C' - 1),
\end{aligned}
\tag{101}
$$

$$
\begin{aligned}
\eta^2 (B - 1)/\mathfrak{c} &= P_1 - (x^2/8)[P_0 \eta'^2 \mathfrak{r}(\eta^2 + \mathfrak{c}P_1) - 2P_1/\mathfrak{r}] \\
&\quad + P_2 + O(x^6).
\end{aligned}
\tag{102}
$$

To the indicated accuracy, we may replace $\eta^2 + \mathfrak{c}P_1$ by $1 - \mathfrak{c}P_0$ within the brackets; if we subtract (101) from (102) we obtain $-\Delta = P_0 + P_1 + P_2 + O(x^6)$, as required.

The values of B obtained from the above correspond to unbounded distributions; their use with $Z = B\Gamma/\gamma$, $Q_1 = Q_2$ to construct the average field (38) for the slab distribution may introduce discrepancies for boundary effects. We now compare with results based on the interface approximation in the form (75) and show that both developments give the same leading terms of the real and imaginary parts of the bulk parameters.

The leading terms of $g_\pm$ of (75) in terms of $a_n^i = A_{|n|} \exp(-in\beta)/\epsilon_n$ are given by

$$
\begin{aligned}
g_+/2 &= A_0 + A_1 \sin\beta \sin\alpha + A_2 \cos 2\beta \cos 2\alpha \\
&= A_0 + A_1 \eta \sin^2\beta + A_2[\eta^2 \cos^2 2\beta - (\eta^2 - 1)\cos 2\beta], \\
\Gamma g_-/2\gamma &= \Gamma(A_1 \cos\beta \cos\alpha + A_2 \sin 2\beta \sin 2\alpha)/\gamma \\
&= A_1 \eta \cos^2\beta + A_2 \eta^2 \sin^2 2\beta.
\end{aligned}
\tag{103}
$$

where we used Snell's law $\eta \sin\beta = \sin\alpha$. If we retain only monopoles and dipoles, then from (75) we have $-\Delta \approx A_0 + A_1\eta$, and by comparison with (86) and (93) in the form $-\Delta = P_0 + P_1 = P_0 + (P_0/A_0)A_1\eta$ we see that the interface approximation corresponds to

$$
A_0 \approx P_0, \quad A_1 \approx P_1/\eta, \quad \eta^2 - 1 \approx -\mathfrak{c}(A_0 + A_1\eta) \tag{104}
$$

with P_0 and P_1 as in (88). From (41) in the form $(Z\gamma/\Gamma) - 1 = B - 1 = \mathfrak{c}g_- k^2/\Gamma\gamma 2$, we thus have

$$
\frac{B - 1}{\mathfrak{c}} \approx \frac{A_1}{\eta} \approx \frac{P_1}{\eta^2} = \frac{\mathfrak{A}_1(1 + \mathfrak{A}_0\mathfrak{H}_1/\eta)}{1 - \mathfrak{A}_0\mathfrak{A}_1\mathfrak{H}_1^2} = \frac{\mathfrak{A}_1(1 - \mathfrak{H}_1\Delta/\eta)}{1 + \mathfrak{A}_1\mathfrak{H}_1\eta} \tag{105}
$$

with $\mathfrak{A}_0, \mathfrak{A}_1$ as in (81), (83). We construct the second parameter $C = B\eta^2$ from $C - 1 = (\eta^2 - 1) + \eta^2(B - 1)$ in terms of (104) and (105):

$$
\frac{C - 1}{\mathfrak{c}} \approx -P_0 = -\frac{\mathfrak{A}_0(1 + \mathfrak{A}_1\mathfrak{H}_1\eta)}{1 - \mathfrak{A}_0\mathfrak{A}_1\mathfrak{H}_1^2} = \frac{-\mathfrak{A}_0(1 - \mathfrak{H}_1\Delta/\eta)}{1 + \mathfrak{A}_0\mathfrak{H}_1/\eta}. \tag{106}
$$

The final forms in (106) and (105) correspond to $C(\mathfrak{A}_0, \Delta)$ and $B(\mathfrak{A}_1, \Delta)$.

If we retain A_2 of (103), then

$$-\Delta \approx A_0 + A_1\eta + A_2\left[1 + \frac{2}{\eta^2}(\eta^2 - 1)\sin^2\alpha\right],$$
$$\frac{B-1}{c} \approx \frac{A_1}{\eta} + \frac{4A_2\sin^2\alpha}{\eta^2}, \tag{107}$$
$$-\frac{C-1}{c} \approx A_0 + A_2\left[1 - \frac{2}{\eta^2}(\eta^2 + 1)\sin^2\alpha\right].$$

Thus there are discrepancies for the quadrupole (and higher) moments. For the usual scattering problems, $a_2' = O(k^4)$, the two procedures give the same leading terms, i.e., to order k^0 for ReC and ReB and to order k^2 for ImC, ImB.

Aligned noncircular cylinders

To simplify application of (69)–(72) to noncircular cylinders, we henceforth use θ_0, Θ_0, α_0, and β_0 for angles measured from $\hat{\mathbf{z}}$. For aligned elliptic cylinders (or for other scatterers having the same reflection and inversion symmetries) with major diameter $2a$ along $\hat{\boldsymbol{\zeta}} = \hat{\mathbf{k}}(\delta)$, the isolated scattering amplitude in terms of angles θ, α measured from $\hat{\boldsymbol{\zeta}}$ is

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \sum_{-\infty}^{\infty} a_{nm}\exp(in\theta - im\alpha), \quad \theta = \theta_0 - \delta, \quad \alpha = \alpha_0 - \delta,$$
$$a_{nm} = a_{mn} = a_{-n-m},$$

where n and m range from $-\infty$ to ∞ and $n-m$ is even. Equivalently,

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \sum[a_{nm}^{+}\cos n\theta\cos m\alpha + a_{nm}^{-}\sin n\theta\sin m\alpha], \tag{108}$$
$$\sum = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}$$

with $a_{nm}^{\pm} = (a_{nm} \pm a_{n-m})\epsilon_n\epsilon_m/2$. Similarly,

$$g(\mathbf{k}_R\,|\,\mathbf{K}) = \sum(A_n^{+}\cos n\Theta + A_n^{-}\sin n\Theta),$$
$$A_n^{\pm} = \begin{Bmatrix}1\\ i\end{Bmatrix}(a_n^{l} \pm a_{-n}^{l})\epsilon_n/2. \tag{109}$$

Thus, we reduce (72) to

$$-2A_m^{\pm} = \sum a_{ml}^{\pm}\{A_\nu^{\pm}[T_{\nu-l}\cos(\nu-l)\beta \pm T_{\nu+l}\cos(\nu+l)\beta] + A_\nu^{\mp}[\pm T_{\nu-l}\sin(\nu-l)\beta + T_{\nu+l}\sin(\nu+l)\beta]\}, \tag{110}$$

which, in distinction to (79), involves two coupled sets of coefficients. We introduce

$$b_{ml}^{\pm} = a_{ml}^{\pm}\eta^{m+l}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix} m\beta\begin{Bmatrix}\cos\\ \sin\end{Bmatrix} l\beta, \quad B_m^{\pm} = A_m^{\pm}\eta^m\begin{Bmatrix}\cos\\ \sin\end{Bmatrix} m\beta,$$
$$\mathfrak{H}_{\nu l}^{++} = h_{0,\nu+l} + h_{\nu l}^{-}\left[1 \pm \begin{Bmatrix}\tan\\ \cot\end{Bmatrix}\nu\beta\tan l\beta\right],$$
$$\mathfrak{H}_{\nu l}^{--} = h_{0,\nu+l} + h_{\nu l}^{-}\left[1 \pm \begin{Bmatrix}\cot\\ \tan\end{Bmatrix}\nu\beta\cot l\beta\right],$$

where $h_{0,n} = \mathfrak{H}_n/\eta^n$ and $h_{\nu l}^{-} = \mathfrak{H}_{\nu l}^{-}/\eta^{n+l}$ with $\mathfrak{H}_{\nu l}^{-}$ as in (78). From (110),

$$-B_m^{\pm} = \sum b_{ml}^{\pm}[B_\nu^{+}(\Delta^{-1} - \mathfrak{H}_{\nu l}^{++}) + B_\nu^{\mp}(\Delta^{-1} - \mathfrak{H}_{\nu l}^{+-})],$$

and proceeding essentially as for (89), we obtain

$$P_m^{\pm} = \sum_l b_{ml}^{\pm}[1 + \sum_\nu(P_\nu^{\pm}\mathfrak{H}_{\nu l}^{++} + P_\nu^{\mp}\mathfrak{H}_{\nu l}^{\pm\mp})],$$
$$-\Delta = \bar{g}^{l} = \sum(P_m^{+} + P_m^{-}) \tag{111}$$

with $P_m^{\pm} = -B_m^{\pm}\Delta/\sum(B_\nu^{+} + B_\nu^{-})$.

If we retain only the monopole and dipole contributions, then, from (111),

$$P_0^{+} = P_0 = b_{00}^{+}[1 + P_0\mathfrak{H}_0 + (P_1^{+} + P_1^{-})h_{01}],$$
$$P_1^{\pm} = b_{11}^{\pm}\left[1 + P_0h_{01} + (P_1^{+} + P_1^{-})h_{02} + P_1^{\pm}h_{11}^{-}/\begin{Bmatrix}\cos^2\\ \sin^2\end{Bmatrix}\beta\right], \tag{112}$$
$$h_{11}^{-} = (c + \mathfrak{H}_0 - \mathfrak{H}_2)/2\eta^2.$$

Using $b_{00}^{+} = a_{00}^{+} = a_{00}$ and $b_{11}^{\pm} = \eta^2 a_{11}^{\pm}\{{}^{\cos^2}_{\sin^2}\}\beta$, we introduce

$$p_0 = \mathfrak{A}_0 = \frac{a_{00}}{1 - a_{00}\mathfrak{H}_0}, \quad p_1^{\pm} = \eta^2\mathfrak{A}_1^{\pm}\begin{Bmatrix}\cos^2\\ \sin^2\end{Bmatrix}\beta,$$
$$\mathfrak{A}_1^{\pm} = \frac{a_{11}^{\pm}}{1 - a_{11}^{\pm}\frac{1}{2}(c + \mathfrak{H}_0 - \mathfrak{H}_2)}, \tag{113}$$

where $\mathfrak{A}_0$ is the same form as in (81), but $\mathfrak{A}_1^{\pm}$ differs from the form in (83) by the sign of $\mathfrak{H}_2$. Thus (112) reduces to

$$P_0 = p_0[1 + (P_1^{+} + P_1^{-})h_{01}],$$
$$P_1^{\pm} = p_1^{\pm}[1 + P_0h_{01} + (P_1^{+} + P_1^{-})h_{02}], \quad -\Delta = P_0 + P_1^{+} + P_1^{-}, \tag{114}$$

which may also be derived from the appropriate special case of (110),

$$A_0(1/a_{00} + T_0) + (A_1^{+}\cos\beta + A_1^{-}\sin\beta)T_1 = 0,$$
$$A_0T_1 2\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}\beta + A_1^{\pm}(2/a_{11}^{\pm} + T_0 \pm T_2\cos 2\beta) + A_1^{\mp}T_2\sin 2\beta = 0. \tag{115}$$

Introducing the definitions of (113), we have

$$A_0\left(\frac{1}{p_0} + \frac{1}{\Delta}\right) + (A_1^{+}\cos\beta + A_1^{-}\sin\beta)\eta\left(\frac{1}{\Delta} - h_{01}\right) = 0,$$
$$A_0\left(\frac{1}{\Delta} - h_{01}\right)\eta\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}\beta + A_1^{\pm}\eta^2\begin{Bmatrix}\cos^2\\ \sin^2\end{Bmatrix}\beta\left(\frac{1}{p_1^{\pm}} + \frac{1}{\Delta} - h_{02}\right) + A_1^{\mp}\frac{\eta^2}{2}\sin 2\beta\left(\frac{1}{\Delta} - h_{02}\right) = 0, \tag{116}$$

and (114) follows by substituting $B_0^{+} = A_0$, $B_1^{\pm} = \eta A_1^{\pm}\{{}^{\cos}_{\sin}\}\beta$, $P_n^{\pm} = -B_n^{\pm}\Delta/(B_0 + B_1^{+} + B_1^{-})$.

We rewrite (114) as

$$P_0 = p_0(1 + Fh_{01}), \quad F = f[1 + P_0h_{01} + Fh_{02}],$$
$$f = p_1^{+} + p_1^{-}, \quad F = P_1^{+} + P_1^{-}, \quad -\Delta = P_0 + F \tag{117}$$

and obtain

$$P_0 = \frac{p_0(1 + \bar{f}h_{01})}{1 - p_0\bar{f}h_{01}^2}, \quad F = \frac{\bar{f}(1 + p_0h_{01})}{1 - p_0\bar{f}h_{01}^2}, \quad \bar{f} = \frac{f}{1 - fh_{02}} \tag{118}$$

with corresponding dispersion equation

$$-\Delta = -\frac{\eta^2 - 1}{c} = \frac{p_0 + \bar{f} + 2p_0\bar{f}h_{01}}{1 - p_0\bar{f}h_{01}^2} = \frac{p_0 + f + p_0f(2h_{01} - h_{02})}{1 - fh_{02} - fp_0h_{01}^2} = \bar{g}^{l}. \tag{119}$$

If only a_{00} is significant, then $-\Delta = p_0 = \bar{g}^t$ and (119) reduces to the form (81) for $\eta^2 = \eta_0^2$ in terms of $\mathfrak{A}_0 = p_0$ of (113). On the other hand, if we retain only the dipoles $a_{11}^{\pm}$, then $-\Delta = \bar{f} = \bar{g}^t$ and the analog of (83) is

$$1/\eta^2 = 1 + c\rho(1 - h_{02}\Delta) = 1 + c\rho/(1 - \mathfrak{H}_2\rho),$$
$$\rho = f/\eta^2 = \mathfrak{A}_1^+ \cos^2\beta + \mathfrak{A}_1^- \sin^2\beta = \hat{\mathbf{K}} \cdot \tilde{\mathfrak{P}} \cdot \hat{\mathbf{K}}, \tag{120}$$

where $\tilde{\mathfrak{P}} = \mathfrak{A}_1^+ \hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}} + \mathfrak{A}_1^- \hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\zeta}} = \hat{\mathbf{k}}(\delta)$, $\hat{\boldsymbol{\xi}} = \hat{\mathbf{k}}(\delta + \pi/2)$. Equation (120) determines $\eta^2 = \eta^2(\beta)$ with $\beta = \beta_0 - \delta$, and Snell's law $\eta \sin\beta_0 = \sin\alpha_0$ serves to eliminate β_0. We decompose the dyadic $\tilde{\mathfrak{P}}$ with respect to $\hat{\mathbf{z}}, \hat{\mathbf{x}}$ and construct $\rho = \rho(\beta_0)$:

$$\tilde{\mathfrak{P}} = \sum \mathfrak{P}_{ij} \hat{\mathbf{z}}_i \hat{\mathbf{z}}_j, \quad \hat{\mathbf{z}}_1 = \hat{\mathbf{z}}, \quad \hat{\mathbf{z}}_2 = \hat{\mathbf{x}},$$
$$\rho = \hat{\mathbf{K}} \cdot \tilde{\mathfrak{P}} \cdot \hat{\mathbf{K}} = \mathfrak{P}_{11} \cos^2\beta_0 + 2\mathfrak{P}_{12} \cos\beta_0 \sin\beta_0 + \mathfrak{P}_{22} \sin^2\beta_0,$$
$$\mathfrak{P}_{11} = \mathfrak{A}_1^+ \cos^2\delta + \mathfrak{A}_1^- \sin^2\delta, \quad \mathfrak{P}_{12} = \mathfrak{P}_{21} = (\mathfrak{A}_1^+ - \mathfrak{A}_1^-) \sin\delta \cos\delta,$$
$$\mathfrak{P}_{22} = \mathfrak{A}_1^+ \sin^2\delta + \mathfrak{A}_1^- \cos^2\delta. \tag{121}$$

If $a_{11}^+ = a_{11}^-$, then $\mathfrak{A}_1^+ = \mathfrak{A}_1^-$ and (120) simplifies to $\eta^{-2} - 1 = c\mathfrak{A}_1^+/(1 - \mathfrak{H}_2\mathfrak{A}_1^+) = c\mathfrak{A}_1$ as in (83). If $\alpha_0 = \beta_0 = 0$, we replace ρ in (120) by $\mathfrak{P}_{11}$ to obtain $\eta^2(\delta)$ as a function of the scatterers' orientation. If one of the principal axes is along $\hat{\mathbf{z}}$ (i.e., $\delta = 0$ or $\delta = \pi/2$), then the distribution is symmetrical to reflection in $z = 0$, e.g., for $\delta = 0$ and α_0 arbitrary, we replace ρ in (120) by $\mathfrak{A}_1^+ \cos^2\beta_0 + \mathfrak{A}_1^- \sin^2\beta_0$. For these special cases, the parameters in (24) satisfy $\Gamma_2 = -\Gamma_1$, $\eta_2 = \eta_1$.

For small spaced scatterers, we drop $\mathfrak{H}_2$ and use $\mathfrak{H}_0 \approx W - 1$, as well as $a_{nn} \approx a'_{nn} + a'^2_{nn}/\epsilon_n \approx a'_{nn}/(1 - a'_{nn}/\epsilon_n)$, $a''_{nn} = a'_{nn}/(1 - a'_{nn}W/\epsilon_n)$, $a_{nn} = a_{00}, a_{11}^{\pm}$. Thus

$$\mathfrak{A}_0 \approx \frac{a'_{00}}{1 - a'_{00}W} = a''_{00},$$
$$\mathfrak{A}_1^{\pm} \approx \frac{a_{11}'^{\pm}}{1 - a_{11}'^{\pm}(c + W)\frac{1}{2}} = \frac{a_{11}''^{\pm}}{1 - a_{11}''^{\pm} c\frac{1}{2}}, \tag{122}$$

where we retain only the leading term in k of $a_{nn}'^{\pm}$; the coefficients $a_{nn}''^{\pm}$ differ from the isolated scattering coefficients in that their radiative loss terms take into account the influence of the distribution. The corresponding approximation of (120) is

$$\eta^{-2} = 1 + c\rho = \hat{\mathbf{K}} \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{K}}, \quad \tilde{\mathbf{B}} = \sum B_{ij} \hat{\mathbf{z}}_i \hat{\mathbf{z}}_j, \quad B_{ij} = \delta_{ij} + c\mathfrak{P}_{ij}. \tag{123}$$

Equivalently,

$$\eta^{-2} = B_1 \cos^2\beta + B_2 \sin^2\beta, \quad B_1 = 1 + c\mathfrak{A}_1^+, \quad B_2 = 1 + c\mathfrak{A}_1^-. \tag{124}$$

Multiplying (123) through by $K^2 = k^2\eta^2$, we isolate $\Gamma = K\cos\beta_0$ in terms of B_{ij} and $\tau = k\sin\alpha_0$,

$$\Gamma^2 B_{11} + 2\Gamma B_{12}\tau + B_{22}\tau^2 - k^2 = 0,$$
$$\Gamma = \pm (k^2 B_{11} - |\tilde{\mathbf{B}}|\tau^2)^{1/2}/B_{11} - B_{12}\tau/B_{11} = \Gamma_{1,2}, \tag{125}$$

where $|\tilde{\mathbf{B}}| = B_{11}B_{22} - B_{12}^2 = B_1 B_2$. Having determined Γ_1, Γ_2, we obtain η_1, η_2 from $\eta_i^2 = (\Gamma_i/k)^2 + \sin^2\alpha_0$. If $\delta = 0$, then

$$1 = \eta^2(B_1 \cos^2\beta_0 + B_2 \sin^2\beta_0),$$
$$\eta^2 = [1 + (B_1 - B_2)\sin^2\alpha_0]/B_1, \quad \Gamma^2 = k^2(1 - B_2 \sin^2\alpha_0)/B_1 \tag{126}$$

Although $\Gamma_2 = -\Gamma_1$ for such cases of symmetry in $z = 0$, the results differ significantly from those for circular cylinders in that $\eta = \eta(\alpha_0)$.

More generally we retain both monopoles and dipoles, and use (119) for η^2. If $B' = 1$ [or for $\psi(\mathfrak{S}) = 0$], then $\eta^2 = C$ is the sole bulk parameter; if $C' = 1$ [or for $\partial_n\psi(\mathfrak{S}) = 0$], the sole parameter is specified by $\hat{\mathbf{K}} \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{K}} = \eta^{-2}$. For small spaced scatters, we drop $\mathfrak{H}_1$ and $\mathfrak{H}_2$ to obtain

$$-\Delta \approx p_0 + f = p_0 + \eta^2\rho, \quad \eta^2 \approx \frac{1 - cp_0}{1 + c\rho} = \frac{1 - cp_0}{\hat{\mathbf{K}} \cdot (\tilde{\mathbf{I}} + c\tilde{\mathfrak{P}}) \cdot \hat{\mathbf{K}}}. \tag{127}$$

To this approximation, we have $\eta^2 \approx \eta_0^2\eta_1^2$ with η_0 as the result for monopoles and η_1 that for dipoles. For two parameter (C', B') scatterers, $\eta_0^2 = 1 - cp_0 \approx C$ and $\eta_1^2 \approx 1 + c\rho \approx \hat{\mathbf{K}} \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{K}}$ provide leading term approximations for the real and imaginary parts of the corresponding two bulk parameters; now $C = 1 - cp_0 = \eta^2(1 + c\rho) = \eta^2 \hat{\mathbf{K}} \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{K}}$ and the generalization of (125) has the k^2 term multiplied by C (or, equivalently, B_{ij} replaced by B_{ij}/C). Thus

$$B_{11}\Gamma = -B_{12}\tau \pm (k^2 C B_{11} - |\tilde{\mathbf{B}}|\tau^2)^{1/2}$$
$$C = 1 - cp_0, \quad \tilde{\mathbf{B}} = \tilde{\mathbf{I}} + c\tilde{\mathfrak{P}}. \tag{128}$$

Similarly, if $\delta = 0$, the generalization of (126) is $\eta^2 = [C + (B_1 - B_2)\sin^2\alpha]/B_1$.

To lowest order in k for the real and imaginary parts,

$$g(\mathbf{k}_r | \mathbf{K}) = p_0 + \hat{\mathbf{r}} \cdot \tilde{\mathfrak{P}} \cdot \hat{\mathbf{K}}\eta = p_0 + \mathbf{k}_r \cdot \tilde{\mathfrak{P}} \cdot \mathbf{K}/k^2,$$
$$g[\![\mathbf{K} | \mathbf{K}]\!] = p_0 + \hat{\mathbf{K}} \cdot \tilde{\mathfrak{P}} \cdot \hat{\mathbf{K}}\eta^2 = p_0 + \mathbf{K} \cdot \tilde{\mathfrak{P}} \cdot \mathbf{K}/k^2. \tag{129}$$

Because the scatterers have inversion symmetry, $B_{12} = B_{21}$ and consequently $Z_1 - Z_2 = (B_{12} - B_{21})\tan\alpha_0 = 0$ as discussed for (47). Thus $Q_1 = Q_2 = (Z - 1)/(Z + 1)$ in (37) and (38), with Z of (43) specified by

$$Z\cos\alpha = (B_{11}\Gamma + B_{12}\tau)/k = (CB_{11} - |\tilde{\mathbf{B}}|\sin^2\alpha_0)^{1/2}$$
$$= [(1 - cp_0)(1 + c\mathfrak{P}_{11}) - |\tilde{\mathbf{I}} + c\tilde{\mathfrak{P}}|\sin^2\alpha_0]^{1/2}.$$

B. Elliptically symmetrical distribution

We consider identical cylindrical scatterers as if centered within aligned identical transparent coatings whose outer surface [an ellipse with major diameter $b\hat{\boldsymbol{\zeta}}^b(\bar{\delta})$, $\hat{\mathbf{z}} \cdot \hat{\boldsymbol{\zeta}}^b = \cos\bar{\delta}$, and minor diameter $tb\hat{\boldsymbol{\xi}}^b$] corresponds to an exclusion ellipse $\mathfrak{b}(\hat{\mathbf{R}})$ with axes (semi-diameters) $b\hat{\boldsymbol{\zeta}}^b$, $tb\hat{\boldsymbol{\xi}}^b$. The scatters may be circular cylinders, or elliptic cylinders with axes $a\hat{\boldsymbol{\zeta}}(\delta), t'a\hat{\boldsymbol{\xi}}$, aligned or averaged over δ. If the scatterers are aligned ellipses, their axes are not in general parallel to those of the envelope ($\delta \neq \bar{\delta}$) and even if $\delta = \bar{\delta}$, the scatterer and envelope need not be confocal or similar. In the coordinate system $\bar{\xi} + i\bar{\xi} = b(1 - t^2)^{1/2}\cosh(\mu + i\Theta)$, we specify $\mathfrak{b}(\hat{\mathbf{R}})$ by $\mu = \mu_b$, $\tanh\mu_b = t$. We assume that the correlation factor $\rho[f(\mathbf{R}) - 1]$ of (60) has the form $\mathfrak{F}(\mu)/\mathfrak{D}$, where $\mathfrak{D} = b^2(1 - t^2)(\cosh^2\mu - \cos^2\Theta)$ is the Jacobian in $d\bar{\xi}\,d\bar{\xi} = \mathfrak{D}\,d\mu\,d\Theta$; the resulting low frequency forms of η then reduce directly to those for radial symmetry as $t \to 1$.

We decompose (60) in terms of Mathieu functions appropriate to the exclusion ellipse; in abbreviated

notation[14,15]

$$\mathfrak{U}=\sum_{j,m} B_{jm} S_{jm}(\Theta) H_{jm}(\mu) i^m,$$

$$\sum_{j,m} f_{jm}=\sum_{m=0}^{\infty} (f_{em}+f_{om}),$$

$$B_{jm}=\int_0^{2\pi} g(\hat{\mathbf{r}},\hat{\mathbf{R}})\, g(\mathbf{k}_R\,|\,\mathbf{K})\, S_{jm}(\Theta)\, d\Theta/2\pi M_{jm},$$

$$\int_0^{2\pi} S_{in} S_{jm}\, d\Theta=\delta_{ij}\delta_{nm} 2\pi M_{jm},$$

with

$$S_{jm}(\Theta)=S_{\substack{e\\o}m}(d;\cos\Theta),\quad d=kb(1-t^2)^{1/2}$$

for the periodic Mathieu functions,[16] and $H_{jm}(\mu)=H^{(1)}_{\substack{e\\o}m}(d;\cosh\mu)$ for the outgoing functions.[16] Similarly

$$\exp(-i\mathbf{K}\cdot\mathbf{R})=\sum i^{-n}\bar{S}_{jn}(\bar{\beta})\,\bar{S}_{jn}(\Theta)\,\bar{J}_{jn}(\mu)/\bar{M}_{jn},$$

$$\bar{\beta}=\beta_0-\bar{\delta}=\beta+\omega,\quad \omega=\delta-\bar{\delta},$$

in terms of

$$\bar{S}_{jn}(\bar{\beta})=S_{\substack{e\\o}n}(\eta d;\cos\bar{\beta}),$$

$$\bar{S}_{jn}(\Theta)=S_{jn}(\eta d;\cos\Theta),\quad \bar{J}_{jn}(\mu)=J_{jn}(\eta d;\cosh\mu).$$

Substituting into (60), we write the analog of (71) as

$$g(\mathbf{k}_r\,|\,\mathbf{K})=-\sum_{j,nm} B_{jm}\,\bar{S}_{jn}(\bar{\beta})\,T^j_{nm},$$

$$T^i_{nm}=M^i_{nm}[(\eta^n/\Delta)\,w^i_{nm}-\mathfrak{H}^i_{nm}],\quad \Delta=(\eta^2-1)/\mathrm{e},$$

$$M^i_{nm}=(i^{-n+m}/2\pi\bar{M}_{in})\int_0^{2\pi}\bar{S}_{in}S_{im}\,d\Theta,\qquad (130)$$

$$w^i_{nm}=(\pi/2i\eta^n)[\bar{J}_{in}(\mu_b)\,\partial_{\mu_b}H_{im}(\mu_b)-H_{im}\,\partial_{\mu_b}\bar{J}_{in}],$$

$$\mathfrak{H}^i_{nm}=2\pi\int_{\mu_b}^{\infty}\mathfrak{F}(\mu)\,\bar{J}_{in}H_{im}\,d\mu=\mathfrak{I}^i_{nm}+i\mathfrak{N}^i_{nm},$$

where $n-m$ is even. The factorization of T facilitates using existing results. The functions M^i_{nm} are discussed by Burke,[15] whose low frequency values (15 : 27) show $M^i_{nm}=O(k^{|n-m|})$; corresponding expressions for $w^i_{nm}=O(k^{n-m})$ and for $\bar{J}_{in}H_{im}$ follow from Ref. 14, (28)—(31). The normalization is such that if $\eta\to 1$, then $M^i_{nm}\to\delta_{nm}$ and $w^i_{nn}\to 1$.

For aligned elliptic cylinders with axes $a\hat{\boldsymbol{\zeta}}(\delta), t'a\hat{\boldsymbol{\xi}}(\delta)$ we use[14,15]

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})=\sum a_{inm}S'_{in}(\theta)\,S'_{im}(\alpha),$$

$$g(\mathbf{k}_r\,|\,\mathbf{K})=\sum A_{in}(\beta)\,S'_{in}(\theta),\quad \theta=\theta_0-\delta=\bar{\theta}-\omega,\qquad (131)$$

where $\omega=\delta-\bar{\delta}$, and $S'_{in}(\theta)=S_{in}(d';\cos\theta)$, $d'=ka(1-t'^2)^{1/2}$. We determine B_{jn} by

$$B_{jm}=\sum_{i,\nu r}\sum_{k,l} a_{i\nu r}S'_{i\nu}(\theta)A_{kl}(\beta)C_{jm}(i_r,k_l),$$

$$S'_{ir}(\Theta-\omega)S'_{j\nu}(\Theta-\omega)=\sum_{k,l}C_{kl}(i_r,j_\nu)\,S_{kl}(\Theta),\quad k=e,o,\qquad (132)$$

where to a given order in k, we may construct the coefficients C_{im} (in terms of ω, d, d') from the Fourier series for S_{in} and the Mathieu series for the cosine and sine. Thus, from (130)—(132),

$$A_{i\nu}(\beta)=-\sum_{k,lr} a_{i\nu r}A_{kl}(\beta)\sum_{j,nm}C_{jm}(i_r,k_l)\,\bar{S}_{jn}(\beta+\omega)\,T^j_{nm}.\qquad (133)$$

We consider only monopole and dipole terms to obtain the analog of (115). The leading terms in k of $\mathrm{Re}\eta$, $\mathrm{Im}\eta$ etc., follow from[14,15]

$$M_{e0}\approx 2M_{in}\approx 1,\quad M^i_{01}=M^i_{10}=M^o_{02}=0,$$

$$M^i_{nn}\approx 1,\quad w^i_{nn}\approx 1,\quad n=0,1,2,$$

$$M^e_{02}\approx -M^e_{20}\approx (kb)^2(\eta^2-1)(1-t^2)/16,$$

$$w^e_{02}\approx 16/(kb)^2(1+t)^2,\quad w^e_{20}=O(k^2b^2);$$

the next terms in the approximations are two orders higher in kb. Similarly, of the correlation integrals, we keep only $M^i_{nm}\mathfrak{I}^i_{nm}\to\delta_{n0}\delta_{m0}2\pi\int\mathfrak{F}(\mu)\,d\mu$. Thus, except for $\mathfrak{I}_{00}\to W-1$, we retain only terms of order k^{-2} to obtain

$$T^i_{nn}\approx\frac{\eta^n}{\Delta}-\delta_{n0}(W-1)\equiv T_{nn},$$

$$W=1+2\pi\int\mathfrak{I}(\mu)\,d\mu,\quad \Delta=\frac{\eta^2-1}{c},\quad c=\frac{i\rho 4}{k^2};$$

$$T^e_{02}\approx c\,\frac{1-t}{1+t}=cE=c(q_2-q_1),$$

$$q_2=\frac{1}{1+t},\quad q_1=\frac{t}{1+t},$$

where T_{nn} with $n=0,1,2$ is the low frequency form of T_n of (71), and the depolarization factors q_i satisfy $q_1+q_2=1$. We use $\bar{S}_{in}(\bar{\beta})\approx\left\{\begin{matrix}\cos\\ \sin\end{matrix}\right\}n\bar{\beta}$ to $n=2$, and similarly for the S's in (132). Substituting into (133), we obtain in terms of $a_{in}=a_{inn}$, $i=\left\{\begin{matrix}e\\ o\end{matrix}\right\}$, $j=\left\{\begin{matrix}o\\ e\end{matrix}\right\}$,

$$A_{e0}(1/a_{e0}+T_{00})+(A_{e1}\cos\beta+A_{o1}\sin\beta)\,T_{11}=0,$$

$$A_{e0}2T_{11}\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}\beta+A_{i1}(2/a_{i1}+T_{00}\pm T^e_{02}\cos2\omega\pm T_{22}\cos2\beta)$$

$$+A_{j1}(T_{22}\sin2\beta-T^e_{02}\sin2\omega)=0,\qquad (134)$$

which differs from the system in (115) by the terms in $T^e_{02}=c(q_2-q_1)=cE$; the additional terms vanish for a circular exclusion region $q_1=q_2=\frac{1}{2}$.

The corresponding approximations for the scattering amplitudes are

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})=a_{e0}+a_{e1}\cos\theta\cos\alpha+a_{o1}\sin\theta\sin\alpha,$$

$$g(\mathbf{k}_r\,|\,\mathbf{K})=A_{e0}+A_{e1}\cos\theta+A_{o1}\sin\theta,$$

with a_{in} equivalent to a^+_{nn} of (108). For elliptic cylinders specified by C', $\tilde{\mathbf{B}}'=\sum B'_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$, and $\tilde{\mathbf{q}}'=\sum q'_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$, $\hat{\boldsymbol{\zeta}}_1=\hat{\boldsymbol{\zeta}}$, $\hat{\boldsymbol{\zeta}}_2=\hat{\boldsymbol{\xi}}$, to lowest order in k for $\mathrm{Im}a_{in}$ and $\mathrm{Re}a_{in}$, we have

$$a_{in}\approx a'_{in}+(a'_{in})^2/\epsilon_n\approx a'_n/(1-a'_{in}/\epsilon_n),$$

$$a'_{e0}\approx\tfrac{1}{4}ik^2\mathfrak{B}(C'-1),\quad \mathfrak{B}=\pi a^2t',$$

$$a'_{i1}\approx-\tfrac{1}{4}ik^2\mathfrak{B}(B'_i-1)/[1+(B'_i-1)q'_i],\qquad (135)$$

$$q'_1=t'/(1+t'),\quad q'_2=1/(1+t').$$

We introduce

$$\mathfrak{A}_{e0}=a'_{e0}/(1-a'_{e0}W)=a''_{e0}=p_0,$$

$$\mathfrak{A}_{i1}=a'_{i1}/[1-a'_{i1}\tfrac{1}{2}(W+c)]=a''_{i1}/(1-a''_{i1}\tfrac{1}{2}c),$$

$$p_{i1}=\eta^2\mathfrak{A}_{\substack{e\\o}1}\left\{\begin{matrix}\cos^2\\ \sin^2\end{matrix}\right\}\beta$$

within (134) to obtain the analog of (116)

$$A_{e0}\left(\frac{1}{p_0}+\frac{1}{\Delta}\right)+(A_{e1}\cos\beta+A_{o1}\sin\beta)\,\frac{\eta}{\Delta}=0,$$

$$A_{e0}\frac{\eta}{\Delta}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}\beta+A_{i1}\left[\eta^2\begin{Bmatrix}\cos^2\\ \sin^2\end{Bmatrix}\beta\left(\frac{1}{p_{i1}}+\frac{1}{\Delta}\right)\pm\tfrac{1}{2}cE\cos2\omega\right]+A_{j1}\tfrac{1}{2}\left[\frac{\eta^2}{\Delta}\sin2\beta-cE\sin2\omega\right]=0. \tag{136}$$

The terms in ω are new, and the present low frequency form lacks the h's of (116). Proceeding essentially as before in terms of $B_{e0}=A_{e0}$, $B_{i1}=\eta A_{i1}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}\beta$, $P_{in}=-B_{in}\Delta/\sum B_{km}$, we use $\frac{1}{2}(1\pm E\cos2\omega)=q_1\begin{Bmatrix}\sin^2\\ \cos^2\end{Bmatrix}\omega+q_2\begin{Bmatrix}\cos^2\\ \sin^2\end{Bmatrix}\omega=\begin{Bmatrix}q_{oo}\\ q_{ee}\end{Bmatrix}$, $\frac{1}{2}E\sin2\omega=(q_2-q_1)\sin\omega\cos\omega=q_{eo}=q_{oe}$ to construct

$$P_{e0}=a''_{e0},$$

$$P_{i1}=a''_{i1}\left[\eta^2\begin{Bmatrix}\cos^2\\ \sin^2\end{Bmatrix}\beta+P_{i1}cq_{ii}+P_{j1}cq_{ij}\begin{Bmatrix}\cot\\ \tan\end{Bmatrix}\beta\right. \tag{137}$$

$$-\Delta=P_{e0}+P_{e1}+P_{o1},$$

where $i=\begin{Bmatrix}e\\ o\end{Bmatrix}$ and $j=\begin{Bmatrix}o\\ e\end{Bmatrix}$ are compliments. Thus

$$p=\frac{P_{e1}+P_{o1}}{\eta^2}=\frac{\mathfrak{A}^e\cos^2\beta+\mathfrak{A}^o\sin^2\beta+2cq_{eo}\mathfrak{A}^o\mathfrak{A}^e\cos\beta\sin\beta}{1-\mathfrak{A}^e\mathfrak{A}^oc^2q_{eo}^2},$$

$$\mathfrak{A}^i=\frac{a''_{i1}}{1-a''_{i1}cq_{ii}}. \tag{138}$$

From (137) and (138) we obtain the same forms as in (127) and (128) in terms of the present $p=\hat{\mathbf{K}}\cdot\tilde{\mathfrak{P}}\cdot\hat{\mathbf{K}}$ and $1+cp=\hat{\mathbf{K}}\cdot\tilde{\mathbf{B}}\cdot\hat{\mathbf{K}}$. Substituting $\beta=\beta_0-\delta$, we rewrite p as $\mathfrak{P}_{11}\cos^2\beta_0+\mathfrak{P}_{22}\sin^2\beta_0+\mathfrak{P}_{12}\sin2\beta_0$ with $\eta\sin\beta_0=\sin\alpha_0$, and proceed essentially as for (128) to determine Γ and Z for the synthetic slab.

In terms of an isolated scatterer dyadic $\tilde{\mathfrak{p}}=\sum a''_{i1}\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i$ modified to include the radiation losses within the distribution, and $\tilde{\mathfrak{q}}=\sum q_i\hat{\boldsymbol{\xi}}_i^b\hat{\boldsymbol{\xi}}_i^b$, i.e.,

$$\tilde{\mathfrak{p}}=a''_{e1}\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}+a''_{o1}\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}},\quad a''_{i1}=\frac{a'_{i1}}{1-\frac{1}{2}a'_{i1}W},$$
$$\tilde{q}=q_1\hat{\boldsymbol{\xi}}^b\hat{\boldsymbol{\xi}}^b+q_2\hat{\boldsymbol{\xi}}^b\hat{\boldsymbol{\xi}}^b=\sum q_{ij}\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_j \tag{139}$$

with $\hat{\boldsymbol{\xi}}=\hat{\boldsymbol{\xi}}_e=\hat{\boldsymbol{\xi}}_1$, $\hat{\boldsymbol{\xi}}=\hat{\boldsymbol{\xi}}_o=\hat{\boldsymbol{\xi}}_2$, and q_{ij} as in (137), we rewrite (138) as

$$p=\hat{\mathbf{K}}\cdot\tilde{\mathfrak{P}}\cdot\hat{\mathbf{K}},\quad\tilde{\mathfrak{P}}=\frac{\tilde{\mathfrak{p}}-c|\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}}|\tilde{\mathfrak{q}}^{-1}}{|\tilde{\mathbf{I}}-c\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}}|},$$
$$|\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}}|=a''_{e1}a''_{o1}q_1q_2. \tag{140}$$

To facilitate generalization, we rederive (140) directly from the dyadic equivalent of the system for P_{i1}/η^2 in (137), i.e., from

$$\tilde{\mathfrak{P}}=\tilde{\mathfrak{p}}\cdot(\tilde{\mathbf{I}}+c\tilde{\mathfrak{q}}\cdot\tilde{\mathfrak{P}}), \tag{141}$$

where $\tilde{\mathfrak{p}}$ represents the anisotropy of the scatterer (anisotropic shape $\tilde{\mathfrak{q}}'$, or parameter $\tilde{\mathbf{B}}'$, or both), and $\tilde{\mathfrak{q}}$ the anisotropy of the statistics. Thus $(\tilde{\mathbf{I}}-c\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}})\cdot\tilde{\mathfrak{P}}=\tilde{\mathfrak{p}}$, and consequently

$$\tilde{\mathfrak{P}}=(\tilde{\mathbf{I}}-c\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}})^{-1}\cdot\tilde{\mathfrak{p}}, \tag{142}$$

which reduces to (140) since $(\tilde{\mathbf{I}}-c\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}})^{-1}=[\tilde{\mathbf{I}}-c|\mathfrak{p}\cdot\mathfrak{q}|\times(\tilde{\mathfrak{p}}\cdot\tilde{\mathfrak{q}})^{-1}]$, $(\tilde{\mathfrak{p}}\cdot\mathfrak{q})^{-1}=\tilde{\mathfrak{q}}^{-1}\cdot\tilde{\mathfrak{p}}^{-1}$.

If $q_1=q_2=\frac{1}{2}$, then the exclusion region is a circle and $\mathfrak{P}=\sum\mathfrak{A}_{i1}\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i$ with $\mathfrak{A}_{i1}=a''_{i1}/(1-\frac{1}{2}ca''_{i1})=\mathfrak{A}_1^i$ equal to the form in (122) for a radially symmetric distribution of elliptic scatterers (or of circles with tensor $\tilde{\mathbf{B}}'$). On the other hand, if $a''_{i1}=a''_1$ (corresponding to a radially symmetric scatterer), then $\tilde{\mathfrak{P}}=\sum\mathfrak{A}^i\hat{\boldsymbol{\xi}}_i^b\hat{\boldsymbol{\xi}}_i^b$, $\mathfrak{A}^i=a''_1/(1-a''_1q_i)$ shows only the anisotropy arising from the exclusion region. An additional special case corresponds to the axes of the scatterer parameter and shape ($\tilde{\mathbf{B}}'$ and $\tilde{\mathfrak{q}}'$) and exclusion region ($\tilde{\mathfrak{q}}$) all along $\hat{\boldsymbol{\xi}}_i$; then $\tilde{\mathfrak{P}}=\sum a''_{i1}\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i/(1-ca''_{i1}q_i)$ such that the depolarization factors (q_i) for the exclusion ellipse are not necessarily the same as those for the scatterer (q'_i).

From (141) we have $\tilde{\mathfrak{P}}\cdot(\tilde{\mathbf{I}}+c\tilde{\mathfrak{q}}\cdot\tilde{\mathfrak{P}})^{-1}=\tilde{\mathfrak{p}}$, and, in terms of $\tilde{\mathbf{B}}=\tilde{\mathbf{I}}+c\tilde{\mathfrak{P}}$,

$$(\tilde{\mathbf{B}}-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathfrak{q}}\cdot(\tilde{\mathbf{B}}-\tilde{\mathbf{I}})]^{-1}=c\tilde{\mathfrak{p}}, \tag{143}$$

where we may interpret the left side as proportional to the low frequency response of one scatterer with shape dyadic $\tilde{\mathfrak{q}}$ (corresponding to the exclusion region) and anisotropic parameter $\tilde{\mathbf{B}}$ whose imaginary part includes both absorption and radiation losses by the distribution of scatterers. If $k\to0$, or if scattering losses are negligible, we replace $\tilde{\mathfrak{p}}(a''_{i1})$ of (143) by $\tilde{\mathfrak{p}}'=\tilde{\mathfrak{p}}(a'_{i1})$; then, for noncoincident $\tilde{\mathbf{B}}'$ and $\tilde{\mathfrak{q}}'$ axes,

$$(\tilde{\mathbf{B}}-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathfrak{q}}\cdot(\tilde{\mathbf{B}}-\tilde{\mathbf{I}})]^{-1}=w(\tilde{\mathbf{B}}'-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathfrak{q}}'\cdot(\tilde{\mathbf{B}}'-\tilde{\mathbf{I}})]^{-1}=c\tilde{\mathfrak{p}}',$$
$$w=\rho\mathfrak{V}=N\mathfrak{V}/V \tag{144}$$

is the tensor version $\tilde{\mathbf{T}}=w\tilde{\mathbf{T}}'$ of our earlier generalization[17] of Maxwell's result for spheres. The analogous interpretation of $V\tilde{\mathbf{T}}=N\mathfrak{V}\tilde{\mathbf{T}}'$ relates the potentials of an anisotropic scatterer with volume V, shape $\tilde{\mathfrak{q}}$, and equivalent parameter $\tilde{\mathbf{B}}$, to that of N aligned scatterers (each with $\mathfrak{V},\tilde{\mathfrak{q}}',\tilde{\mathbf{B}}'$) within V. If we diagonalize $\tilde{\mathfrak{p}}'$ as $\sum\mathfrak{p}'_{ii}\hat{\boldsymbol{\xi}}'_i\hat{\boldsymbol{\xi}}'_i$, then if $\hat{\boldsymbol{\xi}}_i^b=\hat{\boldsymbol{\xi}}'_i$, we get $\tilde{\mathbf{B}}-\tilde{\mathbf{I}}=c\sum\mathfrak{p}'_{ii}\boldsymbol{\xi}'_i\boldsymbol{\xi}'_i/(1-c\mathfrak{p}'_{ii}q_i)$; in particular for $\tilde{\mathbf{B}}'$ diagonal in $\hat{\boldsymbol{\xi}}_i$,

$$\frac{B_i-1}{1+q_i(B_i-1)}=w\frac{B'_i-1}{1+q'_i(B'_i-1)},$$
$$B_i-1=\frac{w(B'_i-1)}{1+(B'_i-1)(q'_i-wq_i)}, \tag{145}$$

which reduces to (17:4) if $q_i=q'_i$, i.e., if the exclusion volume is similar to the scatterer's. Whereas $(1-w)q_i$ is positive, the generalization q'_i-wq_i in (145), may be positive or negative; thus B_i-1 may be greater, as well as less, than $w(B'_i-1)$. To include radiation losses in (144), we replace $\tilde{\mathfrak{p}}'$ by $\tilde{\mathfrak{p}}\approx\tilde{\mathfrak{p}}'\cdot(\tilde{\mathbf{I}}+\frac{1}{2}\tilde{\mathfrak{p}}'W)\approx\tilde{\mathfrak{p}}'\cdot(\tilde{\mathbf{I}}-\frac{1}{2}\tilde{\mathfrak{p}}'W)^{-1}$, which we also use in (142) for noncoincident $\tilde{\mathbf{B}}'$ and $\tilde{\mathfrak{q}}'$ axes.

6. BOUNDED SCATTERERS

A. Radially symmetric distributions

For radially symmetric pair statistics in three dimensions, the analog of (69) is the expansion in spherical harmonics:

$$F(\mathbf{k}_r,\mathbf{k}_R|\mathbf{K})=g(\hat{\mathbf{r}},\hat{\mathbf{R}})\,g(\mathbf{k}_R|\mathbf{K})=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}B_n^m(\hat{\mathbf{r}},\hat{\mathbf{K}})\,Y_n^m(\hat{\mathbf{R}}),$$

$$B_n^m=\frac{(-1)^m(2n+1)}{4\pi}\int F Y_n^{-m}\,d\Omega(\hat{\mathbf{R}}),$$

$$Y_n^m(\hat{\mathbf{r}})=P_n^m(\cos\theta)\exp(im\varphi),\quad P_n^{-m}=D_n^m P_n^m,$$
$$D_n^m=(-1)^m(n-m)!/(n+m)!, \tag{146}$$

with $\hat{\mathbf{r}}=\hat{\mathbf{r}}(\theta,\varphi)=\hat{\mathbf{z}}\cos\theta+(\hat{\mathbf{x}}\cos\varphi+\hat{\mathbf{y}}\sin\varphi)\sin\theta$, and $\hat{\mathbf{R}}=\hat{\mathbf{r}}(\Theta,\Phi)$. Using

$$\mathfrak{U}=\sum B_n^m h_n^{(1)}(kR)\,i^n Y_n^m(\hat{\mathbf{R}}),$$
$$\exp(-i\mathbf{K}\cdot\mathbf{R})=\sum(2n+1)\,i^{-n}(-1)^m j_n(KR)\,Y_n^m(\hat{K})\,Y_n^{-m}(\hat{\mathbf{R}}),$$
$$\hat{\mathbf{K}}=\hat{\mathbf{r}}(\beta,0),$$

in (63) yields the three-dimensional version of (70),

$$F(\mathbf{k}_r,\mathbf{K}\,|\,\mathbf{K})=\{\exp(-i\mathbf{K}\cdot\mathbf{R}),\mathfrak{U}\}_b=\sum B_n^m Y_n^m(\hat{\mathbf{K}})\,w_n,$$
$$Y_n^m=P_n^m(\cos\beta),$$
$$w_n=(4\pi k/i4\pi)[j_n(Kb)\,\partial_b h_n^{(1)}(kb)-h_n^{(1)}\,\partial_b j_n]\,b^2 \tag{147}$$
$$=\eta^n+(K^2-k^2)\int_0^b j_n(KR)\,h_n^{(1)}(kR)\,R^2\,dR.$$

Substituting (146) and (147) into (64), we obtain the analog of (71),

$$g(\mathbf{k}_r\,|\,\mathbf{K})=-\sum B_n^m(\hat{\mathbf{r}},\hat{\mathbf{K}})\,Y_n^m(\hat{\mathbf{K}})\,T_n,\quad T_n=\eta^n/\Delta-\mathfrak{H}_n,$$
$$\Delta=(\eta^2-1)/\mathfrak{c},\quad \mathfrak{c}=i4\pi\rho/k^3, \tag{148}$$
$$\mathfrak{H}_n=4\pi\rho\int_0^\infty[f(R)-1]\,j_n(KR)\,h_n^{(1)}(kR)\,R^2\,dR=\mathfrak{I}_n+i\mathfrak{N}_n.$$

For small k we have

$$\mathfrak{I}_0\approx 4\pi\rho\int[f(R)-1]\,R^2\,dR=W-1,$$
$$\mathfrak{I}_n\approx 4\pi\eta^n[k^n/(2n+1)!!]^2\rho\int(f-1)\,R^{2n+2}\,dR,$$
$$\mathfrak{N}_n\approx-[4\pi\eta^n\rho/k(2n+1)]\int(f-1)\,R\,dR, \tag{149}$$
$$(2n+1)!!=1\cdot3\cdot5\cdots(2n+1),$$

where the next terms of $\mathfrak{I}_n$ are $O(k^{2n+2})$ and those of $\mathfrak{N}_n$ are $O(k)$

To reduce (148), we write the scattering amplitudes as

$$g(\mathbf{k}_R\,|\,\mathbf{K})=\sum A_n^m(\mathbf{K})\,Y_n^m(\hat{\mathbf{R}}),\quad g(\hat{\mathbf{r}},\hat{\mathbf{R}})=\sum Y_n^m(\hat{\mathbf{r}})\,a_n^m(\hat{\mathbf{R}}),$$
$$a_n^m(\hat{\mathbf{R}})=\sum a_{n\nu}^{m\mu}\,Y_\nu^{-\mu}(\hat{\mathbf{R}})$$

with $a_{n\nu}^{m\mu}$ independent of direction. We determine B_n^m of (146) by the expansion

$$Y_p^{-q}Y_s^t=\textstyle\sum_l d_l\begin{pmatrix}-q\\p\end{pmatrix};\begin{matrix}t\\s\end{matrix}\Big)\,Y_l^{t-q},$$

where $p+s+l$ is even, l changes by steps of 2 from $|p-s|$ (or from $|t-q|$ if it is the larger) to $p+s$, and the coefficients d_l are known.[18] Thus the analog of (72) is

$$A_n^m=-\sum_{\nu\mu}\sum_{st}a_{n\nu}^{m\mu}A_s^t\textstyle\sum_l d_l\begin{pmatrix}-\mu\\\nu\end{pmatrix};\begin{matrix}t\\s\end{matrix}\Big)\,Y_l^{t-\mu}(\hat{\mathbf{K}})\,T_l$$
$$\equiv-\sum_{\nu\mu}\sum_{st}a_{n\nu}^{m\mu}A_s^t\begin{pmatrix}-\mu\\\nu\end{pmatrix}\Big|\begin{matrix}t\\s\end{matrix}\Big). \tag{150}$$

We introduce the symbol $\left(\begin{smallmatrix}-\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)=\left(\begin{smallmatrix}t\\s\end{smallmatrix}\middle|\begin{smallmatrix}-\mu\\\nu\end{smallmatrix}\right)$ for brevity, essentially as before.[19] For the analog of (74) we use $g[\![\mathbf{K}\,|\,\mathbf{K}]\!]=\sum\bar{A}_n^m P_n^m(\cos\beta)$, and for (75), $g_\pm=\sum A_n^m[1\pm(-1)^{n-m}]\,P_n^m(\cos\alpha)$.

Radially symmetric scatterers

For radially symmetric scatterers (or, e.g., ellipsoids averaged over orientation) we have $a_{n\nu}^{m\mu}=a_n(-1)^m\delta_{m\nu}\delta_{m\mu}$, and (150) reduces to

$$A_n^m=-(-1)^m a_n\sum_{st}\begin{pmatrix}-m\\n\end{pmatrix}\Big|\begin{matrix}t\\s\end{matrix}\Big)\,A_s^t,$$
$$\begin{pmatrix}-m\\n\end{pmatrix}\Big|\begin{matrix}t\\s\end{matrix}\Big)=\sum d_l\begin{pmatrix}-m\\n\end{pmatrix};\begin{matrix}t\\s\end{matrix}\Big)\,Y_l^{t-m}(\hat{\mathbf{K}})\,T_l, \tag{151}$$

where (except for very special eigenvalues, which we discount) $A_n^m=A_n(-1)^m Y_n^{-m}(\hat{\mathbf{K}})$. The corresponding isolated scattering amplitude is

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})=\sum_{nm}a_n(-1)^m Y_n^m(\hat{\mathbf{r}})\,Y_n^{-m}(\hat{\mathbf{k}})=\sum_n a_n P_n(\hat{\mathbf{r}}\cdot\hat{\mathbf{k}}),$$
$$a_n=\frac{a_n'}{1-a_n'/(2n+1)} \tag{152}$$

such that, for lossless scatterers, $\mathrm{Re}\,a_n'=0$, $(2n+1)\,\mathrm{Re}\,a_n=-|a_n|^2$. The multiple scattered functions are

$$g(\mathbf{k}_r\,|\,\mathbf{K})=\sum A_n P_n(\hat{\mathbf{r}}\cdot\hat{\mathbf{K}}),\quad g[\![\mathbf{K}\,|\,\mathbf{K}]\!]=\bar{g}^i=\sum\bar{A}_n=\sum A_n d_n \tag{153}$$

with $d_n=\bar{a}_n/a_n$, essentially as discussed after (80).

If we take the axis of the spherical harmonics along $\hat{\mathbf{K}}$, then $Y_n^m(\hat{\mathbf{K}})=P_n^m(1)=\delta_{m0}$, and (151) simplifies to

$$A_n=-a_n\sum_{s=0}^\infty\begin{pmatrix}0\\n\end{pmatrix}\Big|\begin{matrix}0\\s\end{matrix}\Big)A_s,\quad\begin{pmatrix}0\\n\end{pmatrix}\Big|\begin{matrix}0\\s\end{matrix}\Big)=\sum d_l\begin{pmatrix}0\\n\end{pmatrix};\begin{matrix}0\\s\end{matrix}\Big)\,T_l; \tag{154}$$

this differ from (79) in that, in general, each term involves more than a pair of T's. The leading terms are given by

$$\begin{pmatrix}0\\n\end{pmatrix}\Big|\begin{matrix}0\\0\end{matrix}\Big)=T_n,\quad\begin{pmatrix}0\\1\end{pmatrix}\Big|\begin{matrix}0\\1\end{matrix}\Big)=\tfrac{1}{3}(T_0+2T_2),\quad\begin{pmatrix}0\\1\end{pmatrix}\Big|\begin{matrix}0\\2\end{matrix}\Big)=\tfrac{1}{5}(2T_1+3T_3),$$
$$\begin{pmatrix}0\\2\end{pmatrix}\Big|\begin{matrix}0\\2\end{matrix}\Big)=\tfrac{1}{35}(7T_0+10T_2+18T_4),$$
$$\begin{pmatrix}0\\1\end{pmatrix}\Big|\begin{matrix}0\\3\end{matrix}\Big)=\tfrac{1}{7}(3T_2+4T_4),\quad\begin{pmatrix}0\\2\end{pmatrix}\Big|\begin{matrix}0\\3\end{matrix}\Big)=\tfrac{1}{105}(27T_1+28T_3+50T_5), \tag{155}$$
$$\begin{pmatrix}0\\3\end{pmatrix}\Big|\begin{matrix}0\\3\end{matrix}\Big)=\tfrac{1}{231}(33T_0+44T_2+54T_4+100T_6);$$

these suffice through octupole–octupole effects.

For pure monopoles, from $1+a_0\left(\begin{smallmatrix}0\\0\end{smallmatrix}\middle|\begin{smallmatrix}0\\0\end{smallmatrix}\right)=1+a_0T_0=0$, we obtain the same form $\eta^2-1=\mathfrak{c}\mathfrak{A}_0$ as (81) in terms of the present $\mathfrak{A}_0=a_0/(1-a_1\mathfrak{H}_0)=p_0=\bar{g}^i$. For pure dipoles, from $1+a_1\left(\begin{smallmatrix}0\\1\end{smallmatrix}\middle|\begin{smallmatrix}0\\1\end{smallmatrix}\right)=0$, the analog of (82) is

$$\frac{\eta^2-1}{1+2\eta^2}=-\frac{a_1\mathfrak{c}}{3[1-\frac{1}{3}a_1(\mathfrak{H}_0+2\mathfrak{H}_2)]}, \tag{156}$$

which, with $\eta'^{-2}=B'$, yields Maxwell's result[20] for spheres and $k\to0$, i.e., $(B-1)/(B+2)\to\rho\mathfrak{V}(B'-1)/(B'+2)$, $\mathfrak{V}=\frac{4}{3}\pi a^3$. The analog of (83) is

$$\eta^{-2}=1+\mathfrak{c}\mathfrak{A}_1,\quad\mathfrak{A}_1=\frac{a_1}{1-a_1\mathfrak{H}_{1,1}},\quad\mathfrak{H}_{1,1}=\tfrac{1}{3}(\mathfrak{c}+\mathfrak{H}_0+2\mathfrak{H}_2) \tag{157}$$

with $\bar{g}^i=\mathfrak{A}_1\eta^2=p_1$. Similarly, for pure quadrupoles,

from $1+a_2\left(\begin{smallmatrix}0\\2\end{smallmatrix}\middle|\begin{smallmatrix}0\\2\end{smallmatrix}\right)=0$,

$$\eta^2-1=-c\mathfrak{A}_2\eta^4=-cp_2,\quad \mathfrak{A}_2=\frac{a_2}{1-a_2\mathfrak{H}_{2,2}},$$
$$\mathfrak{H}_{2,2}=\tfrac{1}{35}[c(7+17\eta^2)+7\mathfrak{H}_0+10\mathfrak{H}_2+18\mathfrak{H}_4]. \tag{158}$$

From (19:138), we have $d_0\left(\begin{smallmatrix}0\\n\end{smallmatrix};\begin{smallmatrix}0\\m\end{smallmatrix}\right)=\delta_{nm}/(2n+1)$; thus, as illustrated in the above, in all cases, $\mathfrak{H}_{n,n}$ includes the terms $(e+\mathfrak{H}_0)/(2n+1)$. Similarly for pure $2n$-poles, we write

$$\eta^2-1=-c\mathfrak{A}_n\eta^{2n}=-cp_n,$$
$$\mathfrak{A}_n=\frac{a_n}{1-a_n\mathfrak{H}_{n,n}}=\frac{a_n'}{1-a_n'(\mathfrak{H}_0+1+\mathfrak{H}_{n,n}')/(2n+1)}, \tag{159}$$
$$\mathfrak{H}_{n,n}=\frac{\mathfrak{H}_0+\mathfrak{H}_{n,n}'}{2n+1},$$

which shows the essential aspects of energy conservation discussed for cylinders, i.e., for lossless scatterers we have $-(2n+1)\,\mathrm{Re}\mathfrak{A}_n=|\mathfrak{A}_n|^2\,W_{nn}$, $W_{nn}=1+\mathrm{Re}(\mathfrak{H}_0+\mathfrak{H}_{n,n}')$. For pure octupoles

$$7\mathfrak{H}_{33}=\mathfrak{H}_0+\mathfrak{H}_{33}'=c\,(33+77\eta^2+131\eta^4)/33+(33\mathfrak{H}_0+44\mathfrak{H}_2+54\mathfrak{H}_4+100\mathfrak{H}_6)/33.$$

Proceeding as in (78)ff, we write

$$\begin{pmatrix}0&|&0\\n&|&m\end{pmatrix}=\eta^{n+m}\Delta^{-1}-\mathfrak{H}_{n,m}=\eta^{n+m}(\Delta^{-1}-h_{nm}),$$
$$h_{nm}=\mathfrak{H}_{n,m}/\eta^{n+m},\quad h_{0n}=\mathfrak{H}_n/\eta^n,$$
$$\mathfrak{H}_{n,m}=\textstyle\sum_l d_l\left(\begin{smallmatrix}0\\n\end{smallmatrix};\begin{smallmatrix}0\\m\end{smallmatrix}\right)\left[\dfrac{c(\eta^{n+m}-\eta^l)}{\eta^2-1}+\mathfrak{H}_l\right], \tag{160}$$

where we added and subtracted $\sum d_l\left(\begin{smallmatrix}0\\n\end{smallmatrix};\begin{smallmatrix}0\\m\end{smallmatrix}\right)=1$, a result that follows from $P_n(x)P_m(x)=\sum d_l\left(\begin{smallmatrix}0\\n\end{smallmatrix};\begin{smallmatrix}0\\m\end{smallmatrix}\right)P_l(x)$ for $x=1$. The form $(\eta^{n+m}-\eta^l)/(\eta^2-1)=\eta^l+\eta^{l+2}+\cdots+\eta^{n+m-2}$, a set of even or odd powers of η, is nonsingular for $\eta^2\to 1$. In particular,

$$h_{12}=(2c\eta+2\mathfrak{H}_1+3\mathrm{H}_3)/5\eta^3,$$
$$h_{13}=(3c\eta^2+3\mathfrak{H}_2+4\mathfrak{H}_4)/7\eta^4, \tag{161}$$
$$h_{23}=[c(27+55\eta^2)+27\mathfrak{H}_1+28\mathfrak{H}_3+50\mathfrak{H}_5]/105\eta^5.$$

Substituting (160) into (154) and using the definition $p_n=\mathfrak{A}_n\eta^{2n}$ of (159) to suppress self-coupling terms, we obtain the same form as (89), $P_n=p_n(1+\sum' h_{n\nu}P_\nu)$. Thus, for monopoles plus dipoles, we have η^2 as in (88), and if we include quadrupoles, we obtain η^2 as in (91) in terms of the present functions $h_{n\nu}$ and coefficients a_n. We have (92) and (94), and the forms in (96)–(100) (for d_n, d_n^C, d_n^B, and the bulk parameters B and C), with J_n replaced by j_n. We also obtain the forms (104)–(106) for the interface approximations of η, B, C, so that both procedures give the same leading terms in k for the real and imaginary parts. An alternative development for small spheres is given by Fikioris and Waterman.[20]

Aligned nonspherical scatterers

For aligned triaxial ellipsoids (or for other scatterers having the same reflection and inversion symmetries) with major axis a_1 along $\hat{\boldsymbol{\xi}}=\hat{\mathbf{k}}(\delta)=\hat{\boldsymbol{\xi}}_1$ and axes a_2, a_3 along $\hat{\boldsymbol{\xi}}_2, \hat{\boldsymbol{\xi}}_3$ (in an orthogonal basis $\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2, \hat{\boldsymbol{\xi}}_3$) we take $\hat{\boldsymbol{\xi}}_1$ as the axis of the spherical harmonics Y_n^m. From symmetry and reciprocity the isolated scattering coefficients satisfy

$$a_{n\nu}^{m\,\mu}=a_{\nu n}^{-\mu\,-m}=a_{\nu n}^{-\mu\,-m}D_n^m/D_\nu^\mu,$$
$$a_{n\nu}^{m-\mu}=a_{n\nu}^{-m\mu}D_n^mD_\nu^\mu,\quad D_n^m=(-1)^m(n-m)!/(n+m)!,$$

where $\eta-\nu$ and $m-\mu$ are even. Thus, with $X_{nm}^e=P_n^m\cos m\varphi$, $X_{nm}^o=P_n^m\sin m\varphi$,

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})=\textstyle\sum[a_{n\nu}^{m\,\mu+}X_{nm}^e(\hat{\mathbf{r}})X_{\nu\mu}^e(\hat{\mathbf{k}})+a_{n\nu}^{m\,\mu-}X_{nm}^oX_{\nu\mu}^o],$$
$$\textstyle\sum=\sum_{n=0}^{\infty}\sum_{m=0}^{n}\sum_{\nu=0}^{\infty}\sum_{\mu=0}^{\nu},\quad a_{n\nu}^{m\,\mu\pm}=(a_{n\nu}^{m\mu}D_\nu^\mu\pm a_{n\nu}^{m-\mu})\,\epsilon_m\epsilon_\mu/2. \tag{162}$$

Similarly, analogous to (109),

$$g(\mathbf{k}_\tau\,|\,\mathbf{K})=\textstyle\sum[A_n^{m+}X_{nm}^e(\hat{\mathbf{r}})+A_n^{m-}X_{nm}^o(\hat{\mathbf{r}})],$$
$$A_n^{m\pm}=\begin{Bmatrix}1\\i\end{Bmatrix}^{nm}(A_n^m\pm A_n^{-m}D_n^m)\,\epsilon_m/2. \tag{163}$$

We may therefore rewrite (150) in terms of $X_{l,\,t\pm\mu}^i(\hat{\mathbf{K}})$ with $i=e,o$ as

$$-A_n^{m\pm}=\tfrac{1}{2}\textstyle\sum a_{n\nu}^{m\mu\pm}\left\{A_s^{t\pm}\left[\left(\begin{smallmatrix}-\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)_e\pm\left(\begin{smallmatrix}\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)_e\right]+A_s^{t\mp}\left[\pm\left(\begin{smallmatrix}-\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)_o+\left(\begin{smallmatrix}\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)_o\right]\right\},$$
$$\left(\begin{smallmatrix}\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)_i=\textstyle\sum d_l\left(\begin{smallmatrix}\mu\\\nu\end{smallmatrix};\begin{smallmatrix}t\\s\end{smallmatrix}\right)X_{l,\,t+\mu}^iT_l,$$
$$\left(\begin{smallmatrix}-\mu\\\nu\end{smallmatrix}\middle|\begin{smallmatrix}t\\s\end{smallmatrix}\right)_iD_\nu^\mu=\textstyle\sum d_l\left(\begin{smallmatrix}-\mu\\\nu\end{smallmatrix};\begin{smallmatrix}t\\s\end{smallmatrix}\right)X_{l,\,t-\mu}^iT_l, \tag{164}$$

which shows essentially the same structure as (110).

We retain only monopoles and dipoles to obtain the analog of (115):

$$-A_0^0/a_{00}^{00}=A_0^0\begin{pmatrix}0&|&0\\0&|&0\end{pmatrix}+A_1^0\begin{pmatrix}0&|&0\\0&|&1\end{pmatrix}+A_1^{1+}\begin{pmatrix}0&|&1\\0&|&1\end{pmatrix}_e+A_1^{1-}\begin{pmatrix}0&|&1\\0&|&1\end{pmatrix}_o$$
$$=A_0^0T_0+(A_1^0P_1+A_1^{1+}X_{11}^e+A_1^{1-}X_{11}^o)\,T_1,$$
$$-A_1^0/a_{11}^{00}=A_0^0\begin{pmatrix}0&|&0\\1&|&0\end{pmatrix}+A_1^0\begin{pmatrix}0&|&0\\1&|&1\end{pmatrix}+A_1^{1+}\begin{pmatrix}0&|&1\\1&|&1\end{pmatrix}_e+A_1^{1-}\begin{pmatrix}0&|&1\\1&|&1\end{pmatrix}_o$$
$$=A_0^0P_1T_1+A_1^0\tfrac{1}{3}(T_0+2P_2T_2)+\tfrac{1}{3}(A_1^{1+}X_{21}^e+A_1^{1-}X_{21}^o)\,T_2,$$
$$-A_1^{1\pm}/a_{11}^{11\pm}=A_0^0\begin{pmatrix}1&|&0\\1&|&0\end{pmatrix}_i+A_1^0\begin{pmatrix}1&|&0\\1&|&1\end{pmatrix}_i+A_1^{1\pm}\tfrac{1}{2}\left[\begin{pmatrix}-1&|&1\\1&|&1\end{pmatrix}_e\pm\begin{pmatrix}1&|&1\\1&|&1\end{pmatrix}_e\right]+\tfrac{1}{2}A_1^{1\mp}\begin{pmatrix}1&|&1\\1&|&1\end{pmatrix}_o$$
$$=A_0^0X_{11}^iT_1+A_1^0\tfrac{1}{3}X_{21}^iT_2+A_1^{1\pm}\tfrac{1}{3}(T_0-P_2T_2\pm\tfrac{1}{2}X_{22}^eT_2)+A_1^{1\mp}\tfrac{1}{3}X_{22}^oT_2,\quad i=\begin{Bmatrix}e\\o\end{Bmatrix}. \tag{165}$$

In terms of direction cosines $\alpha_i=\hat{\mathbf{K}}\cdot\hat{\boldsymbol{\zeta}}_i$ and briefer notation $a_0=a_{00}^{00}$, $a_1=a_{11}^{00}$, $a_2=a_{11}^{11+}$, $a_3=a_{11}^{11-}$ (and similarly for A), we reduce (165) to

$$A_0\left(\frac{1}{a_0}+T_0\right)+T_1\textstyle\sum A_i\alpha_i=0,$$
$$A_0\alpha_iT_1+A_i\left(\frac{1}{a_i}+\frac{T_0-T_2}{3}\right)+T_2\alpha_i\textstyle\sum A_j\alpha_j=0. \tag{166}$$

Introducing $\mathfrak{A}_0=a_0/(1-a_0\mathfrak{H}_0)$, $\mathfrak{A}_i=a_i/(1-a_i\mathfrak{H}_{11}^i)$,

$\mathfrak{H}_{11}^{-1} = \frac{1}{3}(c + \mathfrak{H}_0 - \mathfrak{H}_2)$, we obtain

$$A_0\left(\frac{1}{\mathfrak{A}_0} + \frac{1}{\Delta}\right) + \left(\frac{1}{\Delta} - h_{01}\right)\sum A_i\alpha_i\eta = 0,$$

$$A_0\alpha_i\eta\left(\frac{1}{\Delta} - h_{01}\right) + \frac{A_i}{\mathfrak{A}_i} + \left(\frac{1}{\Delta} - h_{02}\right)\alpha_i\eta\sum A_j\alpha_j\eta = 0. \tag{167}$$

Proceeding as before for (116) in terms of $p_0 = \mathfrak{A}_0$, $p_i = \mathfrak{A}_i\alpha_i\eta^2$, and $P_0 = -A_0\Delta/D$, $P_i = -A_i\alpha_i\eta/D$ with $D = A_0 + \sum A_i\alpha_i\eta$, we construct the three-dimensional generalization of (114),

$$P_0 = p_0(1 + h_{01}\sum P_i), \quad P_j = p_j(1 + P_0h_{01} + h_{02}\sum P_i),$$
$$-\Delta = P_0 + \sum P_i. \tag{168}$$

Writing $f = \sum p_i$ and $F = \sum P_i$, we obtain the same forms as in (117)ff in terms of the present coefficients and functions.

For the development of (120)ff, we now use

$$\rho = \hat{\mathbf{K}}\cdot\tilde{\mathfrak{P}}\cdot\hat{\mathbf{K}} = \sum \mathfrak{A}_i\alpha_i^2, \quad \tilde{\mathfrak{P}} = \sum \mathfrak{A}_i\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i \tag{169}$$

and to consider the slab distribution as in (121), we revert from $\mathbf{K} = K\sum \alpha_i\hat{\boldsymbol{\xi}}_i$ to the original form $\mathbf{K} = K(\hat{\mathbf{z}}\cos\beta_0 + \hat{\mathbf{x}}\sin\beta_0)$. If $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_2$ are in the zx plane, the previous results apply with $\mathfrak{A}_1^+, \mathfrak{A}_1^-$ replaced by the present $\mathfrak{A}_1, \mathfrak{A}_2$. More generally, we use

$$\mathfrak{P}_{11} = \sum \mathfrak{A}_i(\hat{\mathbf{z}}\cdot\hat{\boldsymbol{\xi}}_i)^2, \quad \mathfrak{P}_{12} = \mathfrak{P}_{21} = \sum \mathfrak{A}_i(\hat{\mathbf{z}}\cdot\hat{\boldsymbol{\xi}}_i)(\hat{\mathbf{x}}\cdot\hat{\boldsymbol{\xi}}_i),$$
$$\mathfrak{P}_{22} = \sum \mathfrak{A}_i(\hat{\mathbf{x}}\cdot\hat{\boldsymbol{\xi}}_i)^2. \tag{170}$$

For small spaced scatterers, we proceed as in (122)–(129) and obtain the earlier forms in terms of the present:

$$\mathfrak{A}_0 \approx \frac{a_0}{1 - a_0\mathfrak{I}_0} \approx \frac{a_0'}{1 - a_0'W} \equiv a_0'',$$
$$\mathfrak{A}_i \approx \frac{a_i}{1 - \frac{1}{3}a_i(c + \mathfrak{I}_0)} \approx \frac{a_i'}{1 - \frac{1}{3}a_i'(c + W)} \equiv \frac{a_i''}{1 - \frac{1}{3}a_i''c}, \tag{171}$$

where we retain only the leading term in k of a_i'. For ellipsoids specified by $\tilde{\mathbf{B}}' = \sum B_i'\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i$, C' we have

$$a_0' = \frac{ik^3\mathfrak{V}}{4\pi}(C' - 1), \quad a_i' = \frac{-ik^3\mathfrak{V}}{4\pi}\frac{(B_i' - 1)}{1 + (B_i' - 1)q_i'}, \quad \sum q_i' = 1; \tag{172}$$

here the depolarization factors q_i' are the elliptic integrals as in (17 : 3) with special cases as discussed before.[17]

B. Anisotropic distributions

We could parallel the development of Sec. 5 C B for the three-dimensional case of ellipsoids centered within nonconfocal, nonsimilar ellipsoidal envelopes. For spheroids, we would use the detailed special function forms obtained by Burke,[21] and for triaxial ellipsoids, the low-frequency converging series derived by Dassios.[22] However, for simplicity, we generalize the low frequency case specified by g^t of (129), Δ of (127), and the dyadic equation (141) for $\tilde{\mathfrak{P}}$.

The scattering amplitudes are approximated by

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = a_0 + \sum a_i\hat{\mathbf{r}}\cdot\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i\cdot\hat{\mathbf{k}}, \quad g(\mathbf{k}_r|\mathbf{K}) = p_0 + \hat{\mathbf{r}}\cdot\tilde{\mathfrak{P}}\cdot\hat{\mathbf{K}}\eta,$$
$$g[\![\mathbf{K}|\mathbf{K}]\!] = p_0 + a\cdot\tilde{\mathfrak{P}}\cdot\hat{\mathbf{K}}\eta^2. \tag{173}$$

We obtain the form η^2 as in (127), with the monopole $p_0 = \mathfrak{A}_0$ as in (171), and the dipole specified by

$$\tilde{\mathfrak{P}} = \tilde{\mathfrak{p}}\cdot(\tilde{\mathbf{I}} + c\tilde{\mathbf{q}}\cdot\tilde{\mathfrak{P}}), \quad \tilde{\mathfrak{p}} = \sum a_i''\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i,$$
$$\tilde{\mathbf{q}} = \sum q_i\hat{\boldsymbol{\xi}}_i^b\hat{\boldsymbol{\xi}}_i^b, \quad \sum q_i = 1, \quad c = i4\pi\rho/k^3 \tag{174}$$

in terms of $a_i'' = a_i'/(1 - \frac{1}{3}a_i'W)$. The development of (141)ff carries through as before, except that the explicit form of the inverse $(\tilde{\mathbf{I}} - c\tilde{\mathfrak{p}}\cdot\tilde{\mathbf{q}})^{-1}$ after (142) does not apply in general. The generalizations of (144) and (145) to include scattering losses have $\tilde{\mathfrak{p}}'$ replaced by $\tilde{\mathfrak{p}} \approx \tilde{\mathfrak{p}}'\cdot(\tilde{\mathbf{I}} - \frac{1}{3}\tilde{\mathfrak{p}}'W)^{-1}$, and we also use this form for $\tilde{\mathfrak{p}}$ in (142) to obtain $c\tilde{\mathfrak{P}} = \tilde{\mathbf{B}} - \tilde{\mathbf{I}}$ for non-coincident $\tilde{\mathbf{B}}'$ and $\tilde{\mathbf{q}}'$ axes. The elliptic cylindrical case corresponds to $q_3 = q_3' = 0$ (as appropriate for infinite diameter along $\hat{\mathbf{y}}$), and the symmetrical case of the sphere to $q_i = \frac{1}{3}$.

If the scatterer (shape and parameter) and exclusion region both have a principal diameter along $\hat{\mathbf{y}}$ then the development for the slab distribution with $\mathbf{K} = K(\hat{\mathbf{z}}\cos\beta_0 + \hat{\mathbf{x}}\sin\beta_0)$ is as before for (138) based on (128).

7. PARALLEL SLABS

For the essentially one-dimensional problem of identical parallel slabs, corresponding to (61) and (62), we write

$$F(\mathbf{k}_r, \mathbf{k}_\pm|\mathbf{K}) = g(\hat{\mathbf{r}}, \hat{\mathbf{k}}_\pm)g(\mathbf{k}_\pm|\mathbf{K}) \equiv F_\pm,$$
$$\hat{\mathbf{k}}_\pm = \pm\hat{\mathbf{z}}\cos\alpha + \hat{\mathbf{x}}\sin\alpha = \begin{Bmatrix}\hat{\mathbf{k}}\\ \hat{\mathbf{k}}'\end{Bmatrix},$$
$$U = F_\pm\exp(i\mathbf{k}_\pm\cdot\mathbf{R}) = F_\pm\exp(i\tau x \pm i\gamma\zeta), \quad \zeta \gtrless 0, \tag{175}$$

where $\hat{\mathbf{r}} = \hat{\mathbf{k}} = \hat{\mathbf{k}}_+$ or $\hat{\mathbf{r}} = \hat{\mathbf{k}}' = \hat{\mathbf{k}}_-$. We evalute (65) as the sum of the values at $\zeta = b$ and $\zeta = -b$,

$$F\{\mathbf{k}_r, \mathbf{K}|\mathbf{K}\} = \{\exp(-i\mathbf{K}\cdot\mathbf{R}), U\}_b$$
$$= (1/2\gamma)[(\Gamma + \gamma)F_+\exp(i(\gamma - \Gamma)b) - (\Gamma - \gamma)F_-\exp(i(\gamma + \Gamma)b)]$$
$$= (1/2\gamma)[(\Gamma + \gamma)F_+ - (\Gamma - \gamma)F_-] + [(\Gamma^2 - \gamma^2)/i2\gamma] \times \int_0^b [F_+\exp(i(\gamma - \Gamma)\zeta + F_-\exp(i(\gamma + \Gamma)\zeta)]\,d\zeta,$$

where the last form separates the value at $b = 0$ from the remainder rewritten as an integral. Similarly, the volume integral in (60) reduces to

$$\rho F_+\int_b^\infty (f - 1)\exp(i(\gamma - \Gamma)\zeta)\,d\zeta + \rho F_-\int_{-\infty}^{-b}(f - 1) \times\exp(-i(\gamma + \Gamma)\zeta)\,d\zeta$$
$$= \rho\int_b^\infty (f - 1)[F_+\exp(i(\gamma - \Gamma)\zeta) + F_-\exp(i(\gamma + \Gamma)\zeta)]\,d\zeta$$

with $f = f(\zeta) = 0$ for $\zeta < b$. Thus, the essentially one-dimensional version of (60) is

$$g(\mathbf{k}_r|\mathbf{K}) = [-i\rho/(K^2 - k^2)][(\Gamma + \gamma)F_+ - (\Gamma - \gamma)F_-] + \rho\int_0^\infty (f - 1)[F_+\exp(i(\gamma - \Gamma)\zeta) + F_-\exp(i(\gamma + \Gamma)\zeta)]\,d\zeta. \tag{176}$$

Equivalently, corresponding to (71),

$$-g(\mathbf{k}_r|\mathbf{K}) = B_0 T_0 + B_1 T_1, \quad B_0 = \tfrac{1}{2}(F_+ \pm F_-),$$
$$T_n = \frac{\mathfrak{c}(\Gamma/\gamma)^n}{\eta^2 - 1} - \mathfrak{H}_n, \quad \mathfrak{c} = \frac{i2\rho\gamma}{k^2} = \frac{i2\rho\cos\alpha}{k}, \tag{177}$$
$$\mathfrak{H}_0 = 2\rho \int_0^\infty (f-1)(\cos\Gamma\zeta)\exp(i\gamma\zeta)\,d\zeta,$$
$$\mathfrak{H}_1 = -i2\rho \int_0^\infty (f-1)(\sin\Gamma\zeta)\exp(i\gamma\zeta)\,d\zeta.$$

(The correspondence of $\cos x$, $\sin x$, $\exp(ix)$, $-i\exp(ix)$ with J_0, J_1, $H_0^{(1)}$, $H_1^{(1)}$ was discussed earlier.[23]) We determine B_n in terms of $g(\hat{\mathbf{r}}, \hat{\mathbf{k}})$ from the forms in (38) for a uniform slab with appropriate parameters, i.e., from $g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = \mathfrak{R}$, $g(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = \mathfrak{T} - 1$.

For an isotropic slab, $g(\hat{\mathbf{r}}, \hat{\mathbf{k}})$ consists essentially of two terms, one symmetrical and one antisymmetrical to reflection in $z = 0$. We write

$$g(\hat{\mathbf{k}}_\pm, \hat{\mathbf{k}}_\pm) = a_0 + a_1, \quad g(\hat{\mathbf{k}}_\pm, \hat{\mathbf{k}}_\mp) = a_0 - a_1, \quad a_i = a_i'/(1 - a_i'), \tag{178}$$

(such that for lossless scatterers, $-\mathrm{Re}a_n' = 0$, and $-\mathrm{Re}a_n = |a_n|^2$) and similarly for $g(\mathbf{k}_r|\mathbf{K})$ in terms of A_0, A_1. From these forms we construct $F_\pm$ of (175) and B_n, and, from (177),

$$-(A_0 \pm A_1) = \tfrac{1}{2}(T_0 + T_1)(a_0 \pm a_1)(A_0 + A_1) + \tfrac{1}{2}(T_0 - T_1)(a_0 \mp a_1)(A_0 - A_1)$$
$$= T_0(a_0 A_0 \pm a_1 A_1) + T_1(a_0 A_1 \pm a_1 A_0).$$

Thus the analog of (79) is

$$-A_0 = a_0(T_0 A_0 + T_1 A_1), \quad -A_1 = a_1(T_1 A_0 + T_0 A_1). \tag{179}$$

Consequently,

$$(a_0^{-1} + T_0)(a_1^{-1} + T_0) - T_1^2 = 0,$$
$$K^2 - k^2 = \Gamma^2 - \gamma^2 = \frac{-i2\rho[\gamma(\mathfrak{A}_0 + \mathfrak{A}_1) + \mathfrak{A}_0\mathfrak{A}_1(-i2\rho + 2\Gamma\mathfrak{H}_1)]}{1 - \mathfrak{A}_0\mathfrak{A}_1\mathfrak{H}_1^2}, \tag{180}$$
$$\mathfrak{A}_i = \frac{a_i}{1 - a_i\mathfrak{H}_0}$$

and the one-dimensional version of (86) is

$$-\Delta = -\frac{\eta^2 - 1}{\mathfrak{c}} = \frac{\mathfrak{A}_0 + \mathfrak{A}_1 + 2\mathfrak{A}_0\mathfrak{A}_1 h_1}{1 - \mathfrak{A}_0\mathfrak{A}_1 h_1^2} = g[\![\mathbf{K}|\mathbf{K}]\!], \tag{181}$$
$$\mathfrak{c} = \frac{i2\gamma}{k^2}\rho, \quad h_1 = \frac{i2\rho}{\Gamma + \gamma} - \mathfrak{H}_1.$$

At low frequencies, we drop $\mathfrak{H}_1$ and replace $\mathfrak{H}_0$ by $\mathfrak{I}_0 \approx W - 1$ to obtain

$$\Gamma^2 = (\gamma - i2\rho\mathfrak{A}_0)(\gamma - i2\rho\mathfrak{A}_1),$$
$$\mathfrak{A}_i = \frac{a_i'}{1 - a_i' W} = a_i'', \quad W = 1 + 2\rho\int_0^\infty [f(\zeta) - 1]\,d\zeta. \tag{182}$$

From (38), to lowest order in k, we have[23]

$$a_0' = i\gamma a\left(\frac{B'\Gamma'^2}{\gamma^2} - 1\right) = ika\sec\alpha[(C' - 1) - (B' - 1)\sin^2\alpha]$$
$$= \sec\alpha(b_0' + b_{01}'\sin^2\alpha),$$
$$a_1' = -i\gamma\alpha\frac{(B' - 1)}{B'} = -ika\sec\alpha\left[\frac{B' - 1}{B'}\cos^2\alpha\right]$$
$$= \sec\alpha(b_{e1}'\cos^2\alpha), \tag{183}$$

where b_0' and b_{11}' are the leading terms of the corresponding multipole coefficients. We rewrite (182) as

$$\eta^2 = \sin^2\alpha + (C - B_2\sin^2\alpha)/B_1 = [C + (B_1 - B_2)\sin^2\alpha]/B_1,$$
$$C = 1 - \frac{i2\rho}{k}\frac{b_0'}{1 - a_0' W}, \quad B_2 = 1 + \frac{i2\rho}{k}\frac{b_{e1}'}{1 - a_0' W}, \tag{184}$$
$$\frac{1}{B_1} = 1 - \frac{i2\rho}{k}\frac{b_{e1}'}{1 - a_1' W}$$

to stress that $\eta = \eta(\alpha)$, as in the generalization of (126) given after (128) for scatterers symmetrical to reflection in $z = 0$. To lowest order in k for $\mathrm{Re}\eta$ and $\mathrm{Im}\eta$ we have

$$(\eta^2 - 1)(k^2/-i2\rho) = \gamma(a_0' + a_1') - i2\rho a_0' a_1' + W[\gamma(a_0'^2 + a_1'^2) - i2\rho a_0' a_1'(a_0' + a_1')]. \tag{185}$$

For anisotropic slabs with tensor parameter $\tilde{\mathbf{B}}' = B_1'\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}} + B_2'\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}} = \sum B_{ij}'\hat{\mathbf{z}}_i\hat{\mathbf{z}}_j$ the forms in (38) lead to

$$g(\hat{\mathbf{k}}_\pm, \hat{\mathbf{k}}_\pm) = a_{00} + a_{11} \pm 2a_{01}, \quad g(\hat{\mathbf{k}}_\pm, \hat{\mathbf{k}}_\mp) = a_{00} - a_{11},$$
$$\tilde{\mathbf{a}} = (\tilde{\mathbf{I}} - \tilde{\mathbf{a}}')^{-1}\cdot\tilde{\mathbf{a}}' = \sum a_{ij}\hat{\mathbf{z}}_i\hat{\mathbf{z}}_j, \quad a_{ii} = (a_{ii}' - |\tilde{\mathbf{a}}'|)/|\tilde{\mathbf{I}} - \tilde{\mathbf{a}}'|,$$
$$a_{01} = a_{10} = a_{10}'/|\tilde{\mathbf{I}} - \tilde{\mathbf{a}}'|, \quad |\tilde{\mathbf{a}}'| = a_{00}'a_{11}' - a_{01}'^2, \tag{186}$$

and similarly for $g(\mathbf{k}_r|\mathbf{K})$ in terms of A_{00}, A_{11}, $2A_{01}$. For lossless scatterers, $\mathrm{Re}a_{ij}' = 0$, and $-\mathrm{Re}a_{ii} = |a_{ii}|^2 + |a_{01}|^2$, $-\mathrm{Re}a_{01}^*(1 + a_{00} + a_{11}) = 0$, $\mathrm{Im}a_{01}^*(a_{00} - a_{11}) = 0$. The coefficients a_{ij}' correspond to the radiationless amplitude $g'(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) - \mathfrak{M}g(\hat{\mathbf{r}}, \hat{\mathbf{R}})g'(\hat{\mathbf{R}}, \hat{\mathbf{k}})$; the expressions for $\mathrm{Re}a_{ij}$ (for lossless scatterers with inversion symmetry) follow from $-\mathrm{Re}\tilde{\mathbf{a}} = \tilde{\mathbf{a}}\cdot\tilde{\mathbf{a}}^*$, $-\mathrm{Re}g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \mathfrak{M}g(\hat{\mathbf{R}}, \hat{\mathbf{r}})g^*(\hat{\mathbf{R}}, \hat{\mathbf{k}})$. Using (186) and (175) in (177), we obtain

$$g^+[1 + \tfrac{1}{2}(T_0 + T_1)(a_{00} + a_{11} + 2a_{01})] + g^-\tfrac{1}{2}(T_0 - T_1)(a_{00} - a_{11}) = 0,$$
$$g^+\tfrac{1}{2}(T_0 - T_1)(a_{00} - a_{11}) + g^-[1 + \tfrac{1}{2}(T_0 + T_1) \times (a_{00} + a_{11} - 2a_{01})] = 0, \tag{187}$$
$$g^\pm = g(\mathbf{k}_\pm|\mathbf{K}), \quad 1 + T_0(a_{00} + a_{11}) + T_1 2a_{01} + |\tilde{\mathbf{a}}|(T_0^2 - T_1^2) = 0.$$

Consequently,

$$\frac{\Gamma^2 - \gamma^2}{-i2\rho} = \frac{\gamma(a_{00} + a_{11}) + 2\Gamma a_{01} - |\tilde{\mathbf{a}}|(i2\rho + 2\gamma\mathfrak{H}_0 - 2\Gamma\mathfrak{H}_1)}{1 - (a_{00} + a_{11})\mathfrak{H}_0 - 2a_{01}\mathfrak{H}_1 + |\tilde{\mathbf{a}}|(\mathfrak{H}_0^2 - \mathfrak{H}_1^2)}. \tag{188}$$

If we drop $\mathfrak{H}_1$ and replace $\mathfrak{H}_0$ by $W - 1$, then

$$\frac{\Gamma^2 - \gamma^2}{-i2\rho} \approx \frac{\gamma(a_{00}' + a_{11}') + 2\Gamma a_{01}' - |\tilde{\mathbf{a}}'|(i2\rho + 2\gamma W)}{|\tilde{\mathbf{I}} - \tilde{\mathbf{a}}' W|}, \tag{189}$$

where we used $a_{ii}' = (a_{ii} + |\tilde{\mathbf{a}}|)/|\tilde{\mathbf{I}} + \tilde{\mathbf{a}}|$, $a_{01}' = a_{01}/|\tilde{\mathbf{I}} + \tilde{\mathbf{a}}|$. Introducing

$$a_{ii}'' = \frac{a_{ii}' - |\tilde{\mathbf{a}}'|W}{|\tilde{\mathbf{I}} - \tilde{\mathbf{a}}' W|}, \quad a_{01}'' = \frac{a_{01}'}{|\tilde{\mathbf{I}} - \tilde{\mathbf{a}}' W|},$$

as modified coefficients that account for the radiation

losses in the distribution, we obtain

$$\Gamma^2-\gamma^2=-i2\rho[\gamma(a''_{00}+a''_{11})+2\Gamma a''_{01}]+(i2\rho)^2|\tilde{\mathbf{a}}''|,$$
$$(\Gamma+i2\rho a''_{01})^2=(\gamma-i2\rho a''_{00})(\gamma-i2\rho a''_{11}). \tag{190}$$

Thus, essentially as in the form (128) for the anisotropic distribution, we have

$$\Gamma=-i2\rho a''_{01}\pm[(\gamma-i2\rho a''_{00})(\gamma-i2\rho a''_{11})]^{1/2}. \tag{191}$$

To construct the corresponding $\tilde{\mathbf{B}}$ and C, by inspection we rewrite the a''_{nm} as

$$a''_{00}=\sec\alpha(b''_0+b''_{oo}\sin^2\alpha),\quad a''_{11}=\sec\alpha(b''_{ee}\cos^2\alpha),$$
$$a''_{01}=\sec\alpha(b''_{eo}\cos\alpha\sin\alpha)$$

and substitute in (191) to obtain (128) in terms of the bulk parameters:

$$C=1-\frac{i2\rho}{k}b''_0,\quad \frac{1}{B_{11}}=1-\frac{i2\rho}{k}b''_{ee},$$
$$B_{12}=\left(\frac{i2\rho}{k}b''_{eo}\right)B_{11},\quad B_{22}=1+\frac{i2\rho}{k}b''_{oo}+\frac{B_{12}^2}{B_{11}} \tag{192}$$

where the last expression corresponds to $|\tilde{\mathbf{B}}|/B_{11}=1+(i2\rho/k)\,b''_{oo}$.

To obtain the isolated scattering coefficients for the above, from (38) we write the amplitudes of a uniform anisotropic slab (of thickness $2a$) having inversion symmetry as

$$g(\hat{\mathbf{k}}_\pm,\hat{\mathbf{k}}_\mp)=E_+-E_-,$$
$$g(\hat{\mathbf{k}}_\pm,\hat{\mathbf{k}}_\pm)=(\exp(-i2\gamma a)+E_++E_-)\exp(\pm i2\delta a)-1,$$
$$2E_\pm=-\frac{(1-\exp(i\Gamma_0 2a))(1\pm Q')\exp(-i2\gamma a)}{1\mp Q'\exp(i2\Gamma_0 a)},$$
$$Q'=\frac{Z'-1}{Z'+1},\quad Z'=\frac{B'_{11}\Gamma_0}{\gamma}, \tag{193}$$
$$\Gamma_0=(k^2C'B'_{11}-|\mathbf{B}'|\tau^2)^{1/2}/B'_{11},\quad \delta=-B'_{12}\tau/B'_{11}$$

such that $\Gamma'_\pm=\delta\pm\Gamma_0$ are the solutions of the quadratic analogous to (47). Thus,

$$2\begin{Bmatrix}a_{00}\\a_{11}\end{Bmatrix}=(E_++E_-+\exp(-i2\gamma a))\cos2\delta-1\pm(E_+-E_-),$$
$$2a_{01}=(E_++E_-+\exp(-i2\gamma a))\,i\sin2\delta. \tag{194}$$

The leading terms for low frequencies correspond to

$$g'(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx ika\sec\alpha\{(C'-1)-\hat{\mathbf{r}}\cdot\tilde{\mathbf{S}}\cdot\hat{\mathbf{k}}\},\quad \tilde{\mathbf{S}}=\sum S_{ij}\hat{\mathbf{z}}_i\hat{\mathbf{z}}_j,$$
$$S_{11}=(B'_{11}-1)/B'_{11},\quad S_{12}=S_{21}=B'_{12}/B'_{11}, \tag{195}$$
$$S_{22}=(|\tilde{\mathbf{B}}'|-B'_{11})/B'_{11}.$$

The relation for $\tilde{\mathbf{S}}$ in terms of $\tilde{\mathbf{B}}'$ is the limiting case for an anisotropic ellipsoid, i.e., from the forms in (144), $\tilde{\mathbf{S}}=(\tilde{\mathbf{B}}'-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\hat{\mathbf{z}}\hat{\mathbf{z}}\cdot(\tilde{\mathbf{B}}'-\tilde{\mathbf{I}})]^{-1}$. Similarly, if we use (142) in the form $\tilde{\mathbf{B}}-\tilde{\mathbf{I}}=c\,\tilde{\mathfrak{P}}=[\tilde{\mathbf{I}}-c\,\tilde{\mathfrak{p}}\cdot\hat{\mathbf{z}}\hat{\mathbf{z}}]^{-1}\cdot\tilde{\mathfrak{p}}c$, $c\,\tilde{\mathfrak{p}}=c\,(\sec\alpha)\,\tilde{\mathfrak{b}}''=(i2\rho/k)\,\tilde{\mathfrak{b}}''$, we again obtain $\tilde{\mathbf{B}}$ as in (192).

We rewrite (195) as

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})|\cos\alpha|$$
$$=b'_0+b'_{oo}\sin\theta\sin\alpha+b'_{oe}(\sin\theta\cos\alpha+\cos\theta\sin\alpha)$$
$$+b'_{ee}\cos\theta\cos\alpha,$$
$$b'_0=ika(C'-1),\quad \tilde{\mathfrak{b}}'=ika\tilde{\mathbf{S}}, \tag{196}$$

from which

$$a'_{00}=\sec\alpha(b'_0+b'_{oo}\sin^2\alpha)$$
$$=\frac{ika}{\cos\alpha}\left[(C'-1)-\frac{|\mathbf{B}'|-B'_{11}}{B'_{11}}\sin^2\alpha\right],$$
$$a'_{11}=\sec\alpha(b'_{ee}\cos^2\alpha)=-ika\frac{B'_{11}-1}{B'_{11}}\cos\alpha, \tag{197}$$
$$a'_{01}=\sec\alpha(b'_{oe}\sin\alpha\cos\alpha)=-ika\frac{B'_{12}}{B'_{11}}\sin\alpha.$$

The terms in a_{ij} to order $(ka)^2$ follow from the relations in (186), and similarly for the coefficients a'_{ij} required for (191). To lowest order in k for Reη and Imη, we have

$$(\eta^2-1)\frac{k^2}{-i2\rho}=\gamma(a'_{00}+a'_{11})+2\Gamma a'_{01}-i2\rho|\tilde{\mathbf{a}}'|$$
$$+W[\gamma(a'^2_{00}+a'^2_{11}+2a'^2_{01})$$
$$+(a'_{00}+a'_{11})\,2(\Gamma a'_{01}-i\rho|\tilde{\mathbf{a}}'|)]. \tag{198}$$

The present results for slabs, and the results in Sec. 5 for cylinders infinite along $\hat{\mathbf{y}}$, apply equally to the corresponding electromagnetic problems with $\hat{\mathbf{k}}\cdot\hat{\mathbf{y}}=0$ (for which scalar developments suffice). In a subsequent paper we generalize the three-dimensional development of Sec. 6 for bounded obstacles to the corresponding vector electromagnetic problems.

[1]V. Twersky, "On Scattering of Waves by Random Distributions, I. Free-Space Scatterer Formalism," J. Math. Phys. 3, 700–15 (1962).
[2]V. Twersky, "On Scattering of Waves by Random Distributions, II. Two-Space Scatterer Formalism," J. Math. Phys. 3, 724–34 (1962).
[3]V. Twersky, "Scattering of Waves by Two Objects," in *Electromagnetic Waves*, edited by R.E. Langer (U. Wisc. Press, Madison, 1962), pp. 361–69.
[4]V. Twersky, "Multiple Scattering by Arbitrary Configurations in Three Dimensions," J. Math. Phys. 3, 83–91 (1962).
[5]L.L. Foldy, "The Multiple Scattering of Waves," Phys. Rev. 67(2), 107–19 (1945).
[6]M. Lax, (a) "Multiple Scattering of Waves," Rev. Mod. Phys. 23, 287–310 (1951); (b) "The Effective Field in Dense Systems," Phys. Rev. 2 88, 621–29 (1952).
[7]J.B. Keller, "Wave Propagation in Random Media," in *Proc. Sympos. Appl. Math., Vol.* 13 (Amer. Math. Soc., Providence, R.I., 1962), pp. 227–46; "Stochastic Equations and Wave Propagation in Random Media," *ibid*, Vol. 16 (AMS, 1964), pp. 145–70.
[8]V. Twersky, "On Propagation in Random Media of Discrete Scatterers," in Ref. 7, Vol. 16, 84–116.
[9]V. Twersky, "Propagation Parameters in Random Distribution of Scatterers," J. d'Analyse Math. 30, 498–511 (1976).
[10]F. Reiche, "Zur Theorie der Dispersion in Gasen und Dampfen," Ann. Phys. 50 1, 121 (1916).
[11]Lord Rayleigh, "On The Transmission of Light Through an Atmosphere Containing Small Particles in Suspension, and on the origin of the Blue of the Sky," Phil. Mag. 47, 375–83 (1899).

[12]V. Twersky, "Transparency of pair-correlated, random distributions of small scatterers with applications to the cornea," J. Opt. Soc. Am. **65**, 524–30 (1975).
[13]S.K. Bose and A.K. Mal, "Longitudinal Shear Waves in a Fiber-Reinforced Composite," Int. J. Solids Structures **9**, 1075–85 (1973).
[14]J.E. Burke and V. Twersky, "On Scattering of Waves by an Elliptic Cylinder and by a Semielliptic Protuberance on a Ground Plane," J. Opt. Soc. Am. **54**, 732–44 (1964).
[15]J.E. Burke, "Low-Frequency Approximation for Scattering by Penetrable Elliptic Cylinders," J. Acoust. Soc. Am. **36**, 2059–70 (1964).
[16]P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), pp. 408ff.
[17]V. Twersky, "Form and Intrinsic Birefringence," J. Opt. Soc. Am. **65**, 239–45 (1975).
[18]B. Friedman and J. Russek, "Addition Theorems for Spherical Waves," Quart. Appl. Math. **12**, 13–23 (1954); S. Stein, Quart. Appl. Math. **19**, 15 (1961).
[19]V. Twersky, "Lattice sums and scattering coefficients for the rectangular planar array," J. Math. Phys. **16**, 644–57 (1975)
[20]J.G. Fikioris and P.C. Waterman, "Multiple Scattering of Waves II. Hole Corrections in the Scalar Case," J. Math. Phys. **5**, 1413–20 (1964).
[21]J.E. Burke, "Scattering by F Penetrable Spheroids," J. Acoust. Soc. Am. **43**, 871-875 (1968). See also J.E. Burke, J. Acoust. Soc. Am. **39**, 826 (1966); **40**, 325 (1966).
[22]G. Dassios, "Convergent Low Frequence Expansions for Penetrable Scatterers," Ph.D. thesis, Mathematics Department, University of Illinois, Chicago (1975); J. Math. Phys. **18**, 126–37 (1977).
[23]V. Twersky, "Acoustic Bulk Parameters of Random Volume Distributions of Small Scatterers, J. Acoust. Soc. Am. **36**, 1314–29 (1964).

Operator density current and relativistic localization problem

Bernard Jancewicz

Institute of Theoretical Physics, University of Wrocław, Cybulskiego 36, Wrocław, Poland
(Received 25 October 1976)

The well-known difficulties with the relativistic localization in quantum theory of one particle are revisited. Among them the noncausal propagation is shown to be connected with the positivity of the time translations generator. A proposal for solving this problem is presented, namely searching for an operator probability density current being the 4-vector quantity. The operator of multiplication by the argument of the Dirac wavefunction is an example of a causal position operator.

INTRODUCTION

The problem of defining position operators and their eigenstates in the framework of relativistic quantum mechanics was generally solved by Newton and Wigner.[1] These authors showed that for elementary systems (i.e., irreducible representations of the Poincaré group) three commuting components of position operator exist together with the continuous spectrum of their eigenvalues.[2]

This Newton–Wigner solution has, however, some disadvantages:

(I) The three components of the position operator are not a part of any four-vector quantity. Therefore, it is difficult to imagine any simple Lorentz covariance.

(II) A state which is localized in the origin (that means that it is an eigenstate of the three position operators to the eigenvalue zero) in one coordinate system is not localized in a moving coordinate system even when the origins coincide at $t=0$.

(III) The localized states have noncausal propagation—this property may be also called instantaneous spreading out of the wavepacket. This means that the scalar product of two localized states (ψ_a, ψ_b), where a and b are points in the Minkowski space, is different from zero also for $(a-b)^2<0$, i.e., for spacelike separations. This implies that the probability of propagation faster than light is nonzero.

A solution of the problem I has been given by Fleming[3,4] who ascertained that a position operator *should* depend on a spacelike hyperplane on which the particle is to be localized. The Lorentz transformation changes the hyperplane of localization, therefore one should expect that it also changes the position operator.

Let us briefly sketch the Fleming hyperplane formalism. Each hyperplane Σ may be uniquely characterized by a four-dimensional vector η orthogonal to it and by its "distance" τ from the origin. These two quantities enter into the equation of the hyperplane,

$$\eta \cdot x = \tau.$$

The quantity τ is a scalar—it is the proper time in a frame of the hyperplane. Since the hyperplane is spacelike, $\eta^2>0$, we choose $\eta^2=1$, $\eta_0>0$. The dependence of the hyperplane on η, τ will be denoted $\Sigma=\Sigma(\eta,\tau)$.

In the Heisenberg picture we have a family of position operators $X_\mu(\eta,\tau)$ fulfilling the operator equation

$$\eta^\mu X_\mu(\eta,\tau)=\tau\cdot 1.$$

This formula says that the four components $X_\mu(\eta,\tau)$ are linearly dependent, so only three of them are independent. In particular, for purely space hyperplanes [i.e., for $\overset{0}{\eta}=(1,0)$] the above formula gives

$$X_0(\overset{0}{\eta},\tau)=\tau\cdot 1,$$

that is, the zeroth component of position is merely the time parameter.

Now the Lorentz covariance has the form

$$U(\Lambda)X_\mu(\eta,\tau)U(\Lambda)^{-1}=(\Lambda^{-1})_\mu{}^\nu X_\nu(\Lambda\eta,\tau).$$

In this way problem I is solved.

Problem II remains unsolved since an eigenstate of $X(\eta,0)$, localized in the origin goes into an eigenstate of $X(\Lambda\eta,0)$ under Lorentz transformation $U(\Lambda)$, but physically one expects that it should remain the eigenstate of $X(\eta,0)$. In Fleming's case one may generally have a family of eigenstates at the same space–time point, indexed by hyperplanes. It may, however, happen that the set of eigenstates of $X(\eta,0)$ localized in one space–time point is not one-dimensional and the Lorentz transformations do not lead out from this set. This situation is already acceptable from the physical point of view and it occurs in our example of Sec. 3.

Our note gives a proposal of seeking for a solution of the third problem. It is based on a useful tool which is an operator probability density distribution $\rho(\mathbf{x},t)$ and an operator probability density current $\mathbf{j}(\mathbf{x},t)$ which together form a four-vector probability density current $j_\mu(x)$. This notion has been used for instance in the paper of Barut and Malin[5] where it was assumed that the expectation values of these operators $(\psi,\rho(\mathbf{x},t)\psi)$ and $(\psi,\mathbf{j}(\mathbf{x},t)\psi)$ give the ordinary probability density distribution and probability density current of quantum mechanics, respectively. Petzold and coworkers[6,7] also investigated four-vector operator density currents in the specific case of scalar particles and showed that for positive energies it cannot have causal propagation.

1. SIGN OF THE TIME TRANSLATION GENERATOR

There are some indications that the positivity of energy causes difficulty III. Let us present a quite elegant argument concerning this question which is due to Jadczyk (private communication).

Suppose we have a localization on spacelike regions Δ of three-dimensional volumes. Suppose also that it is given by projection operators $E(\Delta)$ associated with all such regions. Assume that the localization propagates causally, which means that for two spacelike regions Δ_1 and Δ_2 the projections $E(\Delta_1)$ and $E(\Delta_2)$ are orthogonal, i. e., $E(\Delta_1)E(\Delta_2)=0$. We shall show that this cannot be reconciled with the positivity of the Hamiltonian. For this purpose we need a theorem by Borchers:

Theorem: There exists a one-parameter group $U(t) = \exp(-iHt)$ where the generator $H \geqslant c > -\infty$. Denote $F_t = U(t)FU(t)^{-1}$ for any operator F. If there exists a pair of projectors E, F such that for $|t| < \epsilon$, $EF_t = 0$ then for any $t \in R$, $EF_t = 0$.

The proof is based on analyticity properties of spectral measures and can be found in Ref. 8. We shall use this theorem for different one-parameter groups $U_\eta(\tau)$ of time translations in different frames, namely $U_\eta(\tau) = \exp[-i(P\cdot\eta)\tau]$ where P^μ is the four-vector translation generator. The operator $(P\cdot\eta)$ is the Hamiltonian in the proper frame of η. According we shall consider expressions $F_{\eta\tau} = U_\eta(\tau)FU_\eta(\tau)^{-1}$.

Now we proceed to the argument of Jadczyk. We start from two parallel regions Δ_1 and Δ_2 of the same size, spacelike separated as shown in Fig. 1. We have $E(\Delta_1)E(\Delta_2)=0$. We also have $E(\Delta_1)E(\Delta_2)_{\eta\tau}=0$ for τ small such that $\Delta_2+\eta\tau$ has no intersection with the causal shadow of Δ_1. Then by virtue of the Borchers theorem we have

$$E(\Delta_1)E(\Delta_2)_{\eta\tau}=0$$

for arbitrary τ. Hence we may write

$$E(\Delta_1)E(\Delta_3)=0,$$

where Δ_3 is as shown in Fig. 1. In the same way we can show that $E(\Delta_1)E(\Delta_2)_{\eta'\tau}=0$ for sufficiently small τ

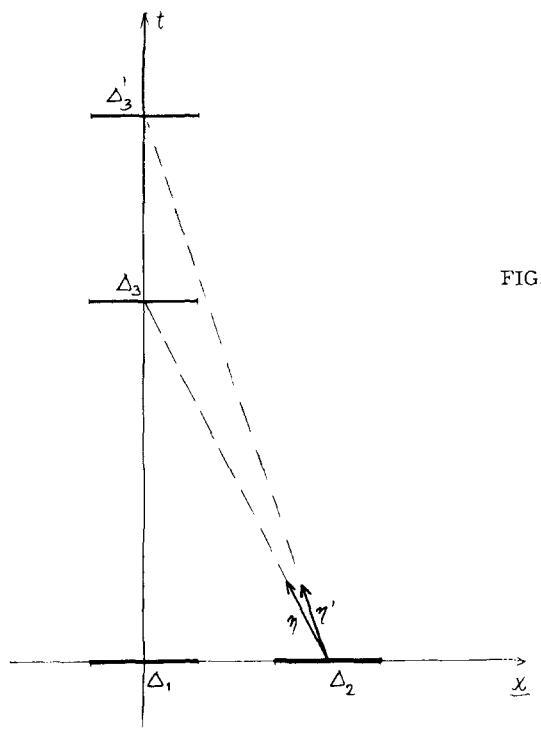


FIG. 1.

and then $E(\Delta_1)E(\Delta_3')=0$. Thus we are allowed to move Δ_3 in the pure time direction without destroying the orthogonality relation,

$$E(\Delta_1)E(\Delta_3)_t=0 \quad \text{for} \quad |t|<\epsilon.$$

Now, using the Borchers theorem we may write down

$$E(\Delta_1)E(\Delta_3)_t=0 \quad \text{for any} \quad t\in R.$$

If we translate Δ_3 back in time far enough to cover it with Δ_1 we obtain in this way

$$E(\Delta_1)^2=0,$$

which implies

$$E(\Delta_1)=0.$$

Thus our family of projectors contains only zero operators. Therefore, we have to agree that such projectors cannot describe any localization. In this way the positivity of time translation generator and the causality of localization cannot be reconciled.

There are also other arguments leading to the same conclusion, see Refs. 9 and 10.

Now we understand why the Newton–Wigner solution of the localization problem has to propagate noncausally: This is so since these authors work with irreducible representations of the Poincaré group which of course have the time translation generator bounded from below. If one wishes to secure the causal propagation one has to give up the positivity of this operator.

In our opinion the relativistic localization problem ought to be stated in the following form:

Find a general form of the Fleming position operators with the following spectral decomposition,

$$X^\mu(\eta,\tau)=\int_{\Sigma(\eta,\tau)} dE_\eta(x)x^\mu, \tag{1}$$

and of the Poincaré group representation under which they are covariant. The spectral measures should satisfy the condition of causal propagation,

$$E_\eta(\Delta)E_{\eta'}(\Delta')=0 \quad \text{for} \quad \Delta \times \Delta'. \tag{2}$$

The most interesting part of the problem is to find the Poincaré group representation and especially its Hamiltonian. Now we are convinced that these representations cannot be irreducible and should contain negative as well as positive Hamiltonian parts. Therefore, it is understandable why Petzold[6] obtained noncausal propagation for positive energy pions.

If one wants to retain the probabilistic interpretation for the states of one sign of energy only, one may use nonspectral measures,

$$\tilde{E}_\eta(\Delta)=PE_\eta(\Delta)P+(1-P)E_\eta(\Delta)(1-P),$$

called positive operator-valued measures[11,12] where P is a projection operator onto the one sign of Hamiltonian vectors, and then use new position operators

$$\tilde{X}^\mu(\eta,\tau)=\int_{\Sigma(\eta,\tau)} d\tilde{E}_\eta(x)x^\mu,$$

which do not lead out from the space of states of one sign of the energy. Jauch and Piron[13] proposed using positive operator valued-measures for a localization of the photons, see also Ref. 14.

However in this language it is difficult to consider the causality of propagation, since the condition $\tilde{E}(\Delta_1)\tilde{E}(\Delta_2)=0$ for $\Delta_1 \chi \Delta_2$ makes no sense because even for equal times $\tilde{E}(\Delta_1)\tilde{E}(\Delta_2)\neq 0$ when $\tilde{E}$ is not a spectral measure. Obviously for such nonspectral measures the above argument of Jadczyk cannot be applied since they do not have the idempotence property. Therefore, we must face the following dilemma: what should be given up—positivity of the Hamiltonian or "projectivity" of the localization in finite regions—to ensure the causality of propagation of the localization? In this paper we choose the first possibility.

In this connection a very puzzling question arises: How is it possible that the energy of an elementary system has both positive and negative signs? In our opinion it is worthwhile to discuss the following answer: The time translation generator H is not merely the energy operator but a product of energy $E=(\mathbf{P}^2+m^2)^{1/2}$ and some other quantity Ξ,

$$H=E\cdot\Xi. \tag{3}$$

These two quantities commute, hence they can be simultaneously determined. The nature of quantity Ξ is not yet clear. Anyway

$$\Xi=\frac{P_0}{(\mathbf{P}^2+m^2)^{1/2}}$$

is Lorentz invariant, which can be easily checked by virtue of the Poincaré group Lie algebra. It has an obvious property $\Xi^2=1$ for the one-particle systems, i.e., for $P_0^2-\mathbf{P}^2=m^2$.

In order to maintain the Lorentz covariance of (3) one should also change the interpretation of the linear momentum, namely $\Xi\cdot\mathbf{P}=\mathbf{p}$ should be viewed as momentum operator instead of P which may be called only space translation generators. From the point of view of the conservation laws both P and p are equally physical since both commute with the Hamiltonian.

In the case of the massless Dirac particle, i.e., for the neutrino, Ξ is just the product of helicity and chirality (see Sec. 4). Thus the necessity of using both signs of the Hamiltonian may be expressed in the following way: In order to have causal relativistic covariant localization of the neutrino one has to work with both signs of helicity and/or chirality.

There is a paper by Bertrand[15] in which the hyperplane formalism of Fleming is applied to a special limiting case of null planes appropriate for massless particles and specifically to the photon. Bertrand has been led to use both signs of the Hamiltonian only because of the very existence of the position operators for the photon and without adducing the causal propagation. This is understandable since there are difficulties with the existence of any (causal or not) localization of the photon, see Refs. 13 and 14.

2. OPERATOR PROBABILITY DENSITY CURRENT

In the light of the previous section it is natural that the localization problem needs a new approach for solving it. Our goal is to present a useful tool for this aim. Namely, an operator-valued probability density distribution $\rho(x)$ together with an operator-valued probability density current $\mathbf{j}(x)$ which together form a probability density current 4-vector,

$$j^\mu(x)=(\rho(x),\mathbf{j}(x)).$$

Mathematically it is an operator-valued distribution of class $\mathcal{O}'_M$ in spacelike directions,[16] which means that after smearing out over spacelike hyperplanes with smooth functions growing polynomially[17] it yields an (unbounded) operator in a Hilbert space. Let us denote such a class of distributions $\mathcal{O}'_{M,s}$, where s stands for "spacelike." The expectation values of this operator quantity $(\psi, j^\mu(x)\psi)$ are the ordinary probability currents of quantum mechanics, see Ref. 5. We may summarize our proposal in the following assumption:

There exists an operator-valued 4-vector distribution $j^\mu(\cdot)\in\mathcal{O}'_{M,s}$ such that the Radon—Nikodym derivative of the spectral measure $E_\eta(\Delta)$ occurring in (1) with respect to the three-dimensional volume V in the proper frame of η can be expressed by the formula

$$\frac{dE_\eta(x)}{dV}=\eta\cdot j(x).$$

The current is assumed to be conserved,

$$\partial^\mu j_\mu(x)=0.$$

This property assures that the integral

$$\int_{\Sigma(\eta,\tau)} j^\mu(x)\,d\sigma_\mu(\eta,x),$$

where $d\sigma(\eta,x)=\eta_\mu\,dV(x)$, does not depend on η nor τ.[18] From the physical interpretation we demand

$$\int_{\Sigma(\eta,\tau)} j^\mu(x)\,d\sigma_\mu(\eta,x)=1 \tag{4}$$

as an operator identity. We assume also that the current is Poincaré covariant,

$$U(\Lambda,a)j^\mu(x)U(\Lambda,a)^{-1}=(\Lambda^{-1})^\mu{}_\nu j^\nu(\Lambda x+a).$$

Having this we define the position operator[19]

$$X^\mu(\eta,\tau)=\int_{\Sigma(\eta,\tau)} y^\mu j^\nu(y)\,d\sigma_\nu(\eta,y). \tag{5}$$

Let us check the Fleming condition $X_\mu(\eta,\tau)\eta^\mu=\tau\cdot 1$,

$$\eta^\mu X_\mu(\eta,\tau)=\int_{\Sigma(\eta,\tau)}(\eta\circ y)j^\nu(y)\,d\sigma_\nu(\eta,y).$$

Since $\eta\circ y=\tau$ for $y\in\Sigma(\eta,\tau)$ we have

$$\eta^\mu X_\mu(\eta,\tau)=\tau\int_{\Sigma(\eta,\tau)} j^\nu(y)\,d\sigma_\nu(\eta,y)=\tau 1$$

by virtue of (4). Let us also examine the Lorentz transformation law of $X(\eta,\tau)$. Toward this purpose we need the transformation properties of the integration element $d\sigma_\mu$,

$$\Lambda_\mu{}^\nu\,d\sigma_\nu(\eta,x)=d\sigma_\mu(\Lambda\eta,\Lambda x). \tag{6}$$

Checking, we have:

$$\begin{aligned}
&U(\Lambda)X_\mu(\eta,\tau)U(\Lambda)^{-1}\\
&\quad=\int_{\Sigma(\eta,\tau)} y_\mu U(\Lambda)j^\nu(y)U(\Lambda)^{-1}\,d\sigma_\nu(\eta,y)\\
&\quad=\int_{\Sigma(\eta,\tau)} y_\mu(\Lambda^{-1})^\nu{}_\lambda j^\lambda(\Lambda y)\,d\sigma_\nu(\eta,y)\\
&\quad=\int_{\Sigma(\eta,\tau)} y_\mu j^\lambda(\Lambda y)\Lambda_\lambda{}^\nu\,d\sigma_\nu(\eta,y)\\
&\quad=\int_{\Sigma(\eta,\tau)} y_\mu j^\lambda(\Lambda y)\,d\sigma_\lambda(\Lambda\eta,\Lambda y)
\end{aligned}$$

changing the variables $y' = \Lambda y$

$$= \int_{\Sigma(\Lambda\eta,\tau)} (\Lambda^{-1}y')_\mu j^\lambda(y')\, d\sigma_\lambda(\Lambda\eta, y')$$
$$= (\Lambda^{-1})_\mu{}^\rho X_\rho(\Lambda\eta, \tau). \qquad \text{QED}$$

The probability density current has to have one more property in order to have expression (5) as a spectral decomposition, namely the idempotence property. A spectral measure is given by the formula

$$E_\eta(\Delta) = \int_\Delta j^\nu(y)\, d\sigma_\nu(\eta, y) \tag{7}$$

for $\Delta \subset \Sigma(\eta, \tau)$. In order to have the idempotence $E_\eta(\Delta)^2 = E_\eta(\Delta)$ it is sufficient to postulate

$$\eta_\mu j^\mu(x) j^\nu(y) = \delta^3_\eta(x-y) j^\nu(y) \quad \text{for} \quad (x-y)^2 < 0 \quad \text{or} \quad x = y, \tag{8}$$

where η is a unit vector orthogonal to $(x-y)$ and the distribution δ^3_η, introduced by Fleming in Ref. 20, fulfils

$$\delta(\eta x)\delta^3_\eta(x) = \delta^4(x).$$

This distribution can be also written as

$$\delta^3_\eta(x) = (2\pi)^{-3} \int d^4k\, \delta(\eta k) \exp(ikx)$$

or

$$\delta^3_\eta(x) = \delta^3(\mathbf{x}'), \tag{9}$$

where x' are the coordinates of x in the proper frame of η. The δ^3_η has its support on the axis $\{x = \lambda\eta,\ \lambda \in \mathbf{R}\}$ and has the property

$$\int_{\Sigma(\eta,\tau)} d\sigma_\mu(\eta, x)\delta^3_\eta(x-y) f(x) = f(y)\eta_\mu \quad \text{for} \quad y \in \Sigma(\eta, \tau). \tag{10}$$

The proof of sufficiency of (8) for the idempotence may be found in Ref. 21.

If one wants to work with one sign of Hamiltonian position operators $\tilde{X}(\eta, \tau)$ one may assume a weaker condition

$$\eta_\mu \tilde{j}^\mu(x) > 0 \tag{11}$$

for each timelike η, $\eta_0 > 0$. This means that in the proper frame of η the zeroth component of this current

$$\tilde{\rho}'(x) = \tilde{j}^{0\prime}(x) = \eta^\mu \tilde{j}_\mu(x)$$

(which is just the probability density in that frame) has the ordinary positive probabilistic interpretation. It ensures that the current $\tilde{j}$ put in (7) gives the positive operator-valued measure. The inequality (11) also says that $\tilde{j}(x)$ is a timelike vector for each x. Of course a current satisfying (8) also satisfies (11).

In this way we are sure that the postulating of the existence of a covariant conserved 4-vector probability current yields the Fleming position operator by formula (5).

Now we complete our list of postulates by assuming the property

$$j_\mu(x) j_\mu(y) = 0 \quad \text{for} \quad (x-y)^2 < 0, \tag{12}$$

which quarantees that the projectors (7) satisfy condition (2) of causal propagation. It might seem that (12) follows from (8) but this is not the case since we have in (8) a timelike vector η which can not be arbitrary and should be absent in (12).

Thus we formulate the problem of relativistic localization translated into the language of the probability current as follows:

Find a general (not necessarily irreducible) representation of the Poincaré group and and operator valued distribution $j(\circ) \in \mathcal{O}'_{M,s}$ such that

(i) $\partial_\mu j^\mu(x) = 0$;

(ii) $\int_{\Sigma(\eta,\tau)} j^\mu(x)\, d\sigma_\mu(\eta, x) = 1$;

(iii) $U(\Lambda, a) j^\mu(x) U(\Lambda, a)^{-1} = (\Lambda^{-1})^\mu{}_\nu j^\nu(\Lambda x + a)$;

(iv) $\eta_\mu j^\mu(x) j^\nu(y) = \delta^3_\eta(x-y) j^\nu(y)$ for $(x-y)^2 < 0$ or $x = y$;

(v) $j^\mu(x) j^\nu(y) = 0$ for $(x-y)^2 < 0$.

Postulates (i), (ii), and (iii) serve to fit the Fleming's solution of the problem I mentioned in the Introduction, postulate (iv) ensures that expression (7) gives a spectral measure, and postulate (v) solves problem III. Up to now we did not touch upon the solution of problem II since we do not see how to express it in purely algebraic language.

3. DIRAC PARTICLE—AN EXAMPLE SATISFYING THE POSTULATES

Now we proceed in presenting an example which fulfills postulates (i)—(v), namely the Dirac particle. The free relativistic spin-$\frac{1}{2}$ particle is described by a Hilbert space $\mathcal{H}$ of four-component functions Ψ satisfying the Dirac equation:

$$i\partial_0\Psi + i\alpha^k\partial_k\Psi - \beta m\Psi = 0, \tag{13}$$

where α^1, α^2, α^3, and β are 4×4 Hermitian matrices which anticommute with each other and have their squares equal to one. It is useful for our purposes to introduce the following representation for them

$$\alpha^k = \begin{pmatrix} \sigma^k & 0 \\ 0 & -\sigma^k \end{pmatrix}, \quad k \in \{1, 2, 3\}, \quad \beta = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},$$

where σ^k are 2×2 Pauli matrices, and I is a 2×2 unit matrix. The α matrices may be completed with the unit matrix $\mathbf{1}$ to form the 4-vectors

$$\alpha^\mu = (\mathbf{1}, \alpha), \quad \tilde{\alpha}^\mu = (\mathbf{1}, -\alpha).$$

With the help of 2×2 matrix 4-vectors $\sigma^\mu = (I, \sigma)$ and $\tilde{\sigma}^\mu = (I, -\sigma)$, we may write

$$\alpha^\mu = \begin{pmatrix} \sigma^\mu & 0 \\ 0 & \tilde{\sigma}^\mu \end{pmatrix}, \quad \tilde{\alpha}^\mu = \begin{pmatrix} \tilde{\sigma}^\mu & 0 \\ 0 & \sigma^\mu \end{pmatrix}.$$

These matrices have the property

$$\tilde{\alpha}^\mu\alpha^\nu + \tilde{\alpha}^\nu\alpha^\mu = 2g^{\mu\nu}. \tag{14}$$

Using the matrix 4-vector α^μ we may rewrite Eq. (13), (14)

$$i(\alpha\cdot\partial)\Psi - \beta m\Psi = 0, \tag{15}$$

where $(\alpha\cdot\partial) = \alpha^\mu\partial_\mu$.

The scalar product in $\mathcal{H}$ is given by

$$(\Phi, \Psi) = \int_{x^0=\text{const}} \Phi^*(x)\Psi(x)\, d^3x. \tag{16}$$

It is positive definite and does not depend on x^0.

A matrix distribution $G(x)$ satisfying (15) and such that

$$G(\mathbf{x}, 0) = \delta^3(\mathbf{x})\mathbf{1} \tag{17}$$

will be called a propagator for (15). It can be found[22] to have the form

$$G(x) = -[(\tilde{\alpha}\cdot\partial) - im\beta]\Delta(x, m^2), \tag{18}$$

where Δ is the well-known Pauli—Jordan invariant function vanishing outside the light cone. Hence the distribution (18) also has its support in the closed light cone. It has, moreover, the properties:

$$G(x)^* = G(-x), \tag{19}$$

$$i\partial_\mu G\alpha^\mu - mG\beta = 0. \tag{20}$$

If $A \in \mathrm{SL}(2, C)$ then $A \to \Lambda(A)$ is the well-known homomorphism of $\mathrm{SL}(2, C)$ onto $L_+^\uparrow$—the proper orthochronous Lorentz group, given by the formulas

$$A^*\sigma^\mu A = \Lambda^\mu{}_\nu\sigma^\nu \quad \text{or} \quad \tilde{A}^*\tilde{\sigma}^\mu\tilde{A} = \Lambda^\mu{}_\nu\tilde{\sigma}^\nu, \tag{21}$$

where $\tilde{A} = (A^*)^{-1}$.[23] We introduce the four-dimensional representations of $\mathrm{SL}(2, C)$ by the formula

$$B(A) = \begin{pmatrix} A & 0 \\ 0 & \tilde{A} \end{pmatrix}, \quad \tilde{B}(A) = \begin{pmatrix} \tilde{A} & 0 \\ 0 & A \end{pmatrix}.$$

Then we have, by virtue of (21),

$$B^*(A)\alpha^\mu B(A) = \Lambda^\mu{}_\nu\alpha^\nu \quad \text{and} \quad \tilde{B}^*(A)\tilde{\alpha}^\mu\tilde{B}(A) = \Lambda^\mu{}_\nu\tilde{\alpha}^\nu. \tag{22}$$

It follows from (22) that

$$B(A)\tilde{\alpha}^\mu B(A)^* = (\Lambda^{-1})^\mu{}_\nu\tilde{\alpha}^\nu$$

and

$$\tilde{B}(A)\alpha^\mu\tilde{B}(A)^* = (\Lambda^{-1})^\mu{}_\nu\alpha^\nu. \tag{23}$$

Moreover it is easy to verify that

$$B(A)^*\beta B(A) = \beta \quad \text{for each } A \in \mathrm{SL}(2, C). \tag{24}$$

Using the above formulas it is straightforward to see that G has the following covariance property:

$$B(A)G(x)B^*(A) = G(\Lambda x). \tag{25}$$

The term "propagator" has been used for G because it yields the relation

$$\Psi(x) = \int_{y^0=\text{const}} d^3y\, G(x - y)\Psi(y) \tag{26}$$

for any solution of Eq. (15) and for arbitrary y^0.

In the Hilbert space $\mathcal{H}$ the following unitary representation of $\mathrm{SL}(2, C)$ is given, $\Psi \to U(A)\Psi$, where

$$(U(A)\Psi)(x) = B(A)\Psi(\Lambda^{-1}x).$$

Using the unitarity of these operators and the formulas (22), one may show that the scalar product (16) can be written

$$(\Psi, \Phi) = \int_{\Sigma(\eta,\tau)} \Psi^*(x)\alpha^\mu\Phi(x)\, d\sigma_\mu(\eta, x). \tag{27}$$

Similarly relation (26) after employing (25) may be generalized to the arbitrary spacelike hyperplane,

$$\Psi(x) = \int_{\Sigma(\eta,\tau)} d\sigma_\mu(\eta, y)G(x - y)\alpha^\mu\Psi(y). \tag{28}$$

In particular, since G is also the solution of (15), we have

$$G(x - z) = \int_{\Sigma(\eta,\tau)} d\sigma_\mu(\eta, y)G(x - y)\alpha^\mu G(y - z).$$

The boundary condition (17) may be also generalized onto a spacelike hyperplane passing through the origin. To this purpose one may use a Lorentz transformation Λ such that $\Lambda\eta = \overset{0}{\eta}$. The explicit form of it is

$$\Lambda_0{}^0 = \eta_0, \quad \Lambda_0{}^k = \eta^k, \quad \Lambda_j{}^0 = -\eta_j,$$

$$\Lambda_j{}^k = \frac{\eta_j\eta_k}{\eta_0 + 1} + \delta_j{}^k. \tag{29}$$

Denote $\Lambda x = x'$, of course $x'^0 = 0$ when $x \in \Sigma(\eta, 0)$. Due to (25) we have

$$G(x) = G(\Lambda^{-1}x') = B(A^{-1})G(x')B(A^{-1})^*$$

for A corresponding to Λ given by (29). We take $x \in \Sigma(\eta, 0)$, then $x'^0 = 0$, and by virtue of (17) we have

$$G(x) = B(A^{-1})\delta^3(\mathbf{x}')B(A^{-1})^* = \delta^3(\mathbf{x}')B(A^{-1})\tilde{\alpha}^0B(A^{-1})^*.$$

We use (23) and (9),

$$G(x) = \delta_\eta^3(x)\Lambda^0{}_\nu\tilde{\alpha}^\nu.$$

Therefore, by virtue of (29) we have

$$G(x) = \delta_\eta^3(x)(\eta\cdot\tilde{\alpha}) \quad \text{for } x \in \Sigma(\eta, 0) \tag{30}$$

and this is the needed generalization of (17).

Now we are prepared to introduce our operator-valued density current. For any Ψ of class $\mathcal{S}$ as a function of $\mathbf{x}$ (the set D of such elements from $\mathcal{H}$ constitutes the domain of definition of j),

$$(j^\mu(x)\Psi)(y) = G(y - x)\alpha^\mu\Psi(x). \tag{31}$$

First of all we check the hermicity of this operator. We have

$$\begin{aligned}(\Psi, j^\mu(x)\Phi) &= \int_{y^0=\text{const}} \Psi^*(y)(j^\mu(x)\Phi)(y)\, d^3y \\ &= \int \Psi^*(y)G(y - x)\alpha^\mu\Phi(x)\, d^3y \\ &= \left[\int G(y - x)^*\Psi(y)\, d^3y\right]^*\alpha^\mu\Phi(x)\end{aligned}$$

which using (19),

$$= \left[\int G(x - y)\Psi(y)\, d^3y\right]^*\alpha^\mu\Phi(x)$$

and using (26),

$$= \Psi^*(x)\alpha^\mu\Phi(x).$$

We also have

$$\begin{aligned}(j^\mu(x)\Psi, \Phi) &= \int_{y^0=\text{const}} (j^\mu(x)\Psi)^*(y)\Phi(y)\, d^3y \\ &= \int [G(y - x)\alpha^\mu\Psi(x)]^*\Phi(y)\, d^3y \\ &= \int \Psi^*(x)\alpha^\mu G(y - x)^*\Phi(y)\, d^3y \\ &= \Psi^*(x)\alpha^\mu\int G(x - y)\Phi(y)\, d^3y \\ &= \Psi(x)^*\alpha^\mu\Phi(x).\end{aligned}$$

Hence we have

$$(\Psi, j^\mu(x)\Phi) = (j^\mu(x)\Psi, \Phi) = \Psi(x)^*\alpha^\mu\Phi(x)$$

and the hermicity is checked. We also see in this formula that the expectation value of the operator $j^\mu(x)$ is indeed an ordinary c-number probability current of quantum mechanics.

The wavefunctions belonging to the Hilbert space $\mathcal{H}$ are functions of four space—time coordinates. These functions are determined uniquely by their values on an

arbitrary spacelike hyperplane[24] since they satisfy the wave equation (15) and the scalar product (27) is independent on the hyperplane. Therefore, when considering an (improper) element $j^\mu(x)\Psi$ of $\mathcal{H}$ we may confine ourselves for the function $(j^\mu(x)\Psi)(y)$, only to y belonging to some specific hyperplane $\Sigma(\eta,\tau)$. By virtue of (30), definition (31) then gives

$$(j^\mu(x)\Psi)(y)=\delta^3_\eta(y-x)(\eta\circ\tilde{\alpha})\alpha^\mu\Psi(x) \quad \text{for} \quad x,y\in\Sigma(\eta,\tau). \tag{32}$$

Now we integrate this with a test function f over the hyperplane $\Sigma(\eta,\tau)$,

$$I=[\int_{\Sigma(\eta,\tau)} j^\mu(x)f(x)\,d\sigma_\mu(\eta,x)\Psi](y)$$
$$=\int_{\Sigma(\eta,\tau)}\delta^3_\eta(y-x)(\eta\circ\tilde{\alpha})\alpha^\mu\,d\sigma_\mu(\eta,x)f(x)\Psi(x).$$

If we change the variables $x'=\Lambda x$ for Λ given in (29) and use (9) we obtain

$$I=\int_{x^0=y^0}\delta^3(\mathbf{y}'-\mathbf{x}')(\eta\circ\tilde{\alpha})(\eta\circ\alpha)f(\Lambda^{-1}x')\Psi(\Lambda^{-1}x')\,d^3x'.$$

If we employ the identity

$$(\eta\circ\tilde{\alpha})(\eta\circ\alpha)=(\eta\circ\eta)=1 \tag{33}$$

following from (14), we obtain

$$I=\int_{x^0=y^0}\delta^3(\mathbf{y}'-\mathbf{x}')f(\Lambda^{-1}x')\Psi(\Lambda^{-1}x')\,d^3x'$$
$$=f(\Lambda^{-1}y')\Psi(\Lambda^{-1}y')$$
$$=f(y)\Psi(y).$$

Thus we have proven the identity

$$[\int_{\Sigma(\eta,\tau)}j^\mu(x)f(x)\,d\sigma_\mu(\eta,x)\Psi](y)=f(y)\Psi(y) \quad \text{for } y\in\Sigma(\eta,\tau). \tag{34}$$

We see from it that the smeared current operators $j(f)$ leave our domain D invariant when f is of class $\mathcal{O}_{M,s}$. In this way we have shown that $j^\mu(x)$ is the operator-valued distribution of class $\mathcal{O}'_{M,s}$ on the domain D of vectors which have Schwartz type $\mathcal{S}$ behavior on spacelike hyperplanes.

Now we verify the postulates of Sec. 2.

(i) The current conservation:

$$\left[\frac{\partial}{\partial x^\mu}j^\mu(x)\Psi\right](y)$$
$$=\frac{\partial}{\partial x^\mu}\left[G(y-x)\alpha^\mu\Psi(x)\right]$$
$$=-(\partial_\mu G)(y-x)\alpha^\mu\Psi(x)+G(y-x)\alpha^\mu\partial_\mu\Psi(x)$$
$$=[-\partial_\mu G(y-x)\alpha^\mu-imG(y-x)\beta]\Psi(x)$$
$$+G(y-x)(\alpha^\mu\partial_\mu+im\beta)\Psi(x).$$

Both terms vanish by virtue of (20) and (15), therefore we have

$$\partial_\mu j^\mu=0.$$

(ii) Normalization: It is sufficient to use (34) for $f(x)\equiv 1$,

$$\int_{\Sigma(\eta,\tau)}d\sigma_\mu(\eta,x)j^\mu(x)\Psi=\Psi.$$

Hence

$$\int_{\Sigma(\eta,\tau)}d\sigma_\mu(\eta,x)j^\mu(x)=1.$$

(iii) Lorentz covariance:

$$[U(A)j^\mu(x)U(A)^{-1}\Psi](y)$$
$$=B(A)[j^\mu(x)U(A)^{-1}\Psi](\Lambda^{-1}y)$$
$$=B(A)G(\Lambda^{-1}y-x)\alpha^\mu[U(A)^{-1}\Psi](x)$$
$$=B(A)G(\Lambda^{-1}(y-\Lambda x))B(A)^+B(A^{-1})^+\alpha^\mu B(A^{-1})\Psi(\Lambda x)$$

using (22) and (25),

$$=G(y-\Lambda x)(\Lambda^{-1})^\mu{}_\nu\alpha^\nu\Psi(\Lambda x)$$
$$=(\Lambda^{-1})^\mu{}_\nu[j^\nu(\Lambda x)\Psi](y).$$

Hence

$$U(A)j^\mu(x)U(A)^{-1}=(\Lambda^{-1})^\mu{}_\nu j^\nu(\Lambda x).$$

The translation covariance is easy to check.

(iv) The projection property: Let x,y be such that $(x-y)^2<0$. Then there exists a unit positive timelike vector η such that $\eta\cdot(x-y)=0$ and a hyperplane $\Sigma(\eta,\tau)$ such that $x,y\in\Sigma(\eta,\tau)$. Let us assume that $z\in\Sigma(\eta,\tau)$ and consider the following expression

$$K=[j^\mu(x)j^\nu(y)\Psi](z)=G(z-x)\alpha^\mu G(x-y)\alpha^\nu\Psi(y). \tag{35}$$

We have $z-x\in\Sigma(\eta,0)$ and $x-y\in\Sigma(\eta,0)$, therefore we can use (30),

$$K=\delta^3_\eta(z-x)(\eta\circ\tilde{\alpha})\alpha^\mu\delta^3_\eta(x-y)(\eta\circ\tilde{\alpha})\alpha^\nu\Psi(y)$$
$$=\delta^3_\eta(x-y)(\eta\circ\tilde{\alpha})\alpha^\mu\delta^3_\eta(z-y)(\eta\circ\tilde{\alpha})\alpha^\nu\Psi(y)$$
$$=\delta^3_\eta(x-y)(\eta\circ\tilde{\alpha})\alpha^\mu G(z-y)\alpha^\nu\Psi(y)$$
$$=\delta^3_\eta(x-y)(\eta\circ\tilde{\alpha})\alpha^\mu[j^\nu(y)\Psi](z).$$

If we multiply both sides by η_μ and use (33) we get

$$[\eta_\mu j^\mu(x)j^\nu(y)\Psi](z)=\delta^3_\eta(x-y)[j^\nu(y)\Psi](z)$$

and since Ψ is arbitrary from D,

$$\eta_\mu j^\mu(x)j^\nu(y)=\delta^3_\eta(x-y)j^\nu(y).$$

(v) Causality: We know from (18) that $G(x-y)=0$ for $(x-y)^2<0$. Therefore, the equality (35) gives

$$[j^\mu(x)j^\nu(y)\Psi](z)=0$$

for arbitrary $\Psi\in D$, hence

$$j^\mu(x)j^\nu(y)=0 \quad \text{for} \quad (x-y)^2<0. \qquad \text{QED}$$

Now we find the position operator corresponding to the current (31). We calculate the action of expression (5) on a function $\Psi\in D$,

$$[X^\mu(\eta,\tau)\Psi](z)=[\int_{\Sigma(\eta,\tau)}y^\mu j^\nu(y)\,d\sigma_\nu(\eta,y)\Psi](z).$$

By virtue of (34) we have

$$[X^\mu(\eta,\tau)\Psi](z)=z^\mu\Psi(z) \quad \text{for} \quad z\in\Sigma(\eta,\tau). \tag{36}$$

Thus $X^\mu(\eta,\tau)$ is the operator of multiplication by the argument of the wavefunction when this function is taken on the hyperplane $\Sigma(\eta,\tau)$. Therefore, for α^k we have the usual interpretation as the velocity operators together with the well-known difficulties of the Zitterbewegung.

This velocity seems to be disconnected with the translation generators, since the equality

$$\mathbf{P}=C\boldsymbol{\alpha},$$

where C might be some operator coefficient, is not satisfied. Instead of this we have a connection between

velocity and spin,

$$\mathbf{S}=\Omega\boldsymbol{\alpha}, \tag{37}$$

where in the representation chosen for this paper

$$\Omega=\frac{1}{2}\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

It is easy to check that $\Omega=\frac{1}{2}\gamma_5=\frac{1}{2}i\alpha^1\alpha^2\alpha^3$. In this way the velocity becomes an internal variable of the particle since it does not affect the space–time variables x of its wavefunction. Note that the quantity Ω commutes with spin and velocity. We propose to call it *velocity helicity* in distinction to the ordinary helicity which relates spin with the momentum. The quantity γ_5 is commonly called chirality.

Let us now find the (generalized, since from the continuous spectrum) eigenvectors of the operator $X^\mu(\eta,\tau)$;

$$X^\mu(\eta,\tau)\Psi_{a,\xi}=a^\mu\Psi_{a,\xi}.$$

We use two indices, the first refers to the position eigenvalue, the second refers to some other quantities (for example spin and velocity helicity). After using (36) the above equation takes the form

$$z^\mu\Psi_{a,\xi}(z)=a^\mu\Psi_{a,\xi}(z) \quad \text{for} \quad z\in\Sigma(\eta,\tau).$$

A solution of this equation can only have the form

$$\Psi_{a,\xi}(z)=\delta^3_\eta(z-a)b'_\xi \quad \text{for} \quad z\in\Sigma(\eta,\tau),$$

where b'_ξ is an arbitrary column. In order to calculate the function $\Psi_{a,\xi}$ in other space–time points we use the formula (28),

$$\Psi_{a,\xi}(y)=\int_{\Sigma(\eta,\tau)} d\sigma_\mu(\eta,z)G(y-z)\alpha^\mu\delta^3_\eta(z-a)b'_\xi.$$

Due to (10),

$$\Psi_{a,\xi}(y)=G(y-a)(\eta\circ\alpha)b'_\xi.$$

Since b'_ξ is arbitrary and $(\eta\circ\alpha)$ is a reversible matrix we may argue that $b_\xi=(\eta\circ\alpha)b'_\xi$ is arbitrary and write

$$\Psi_{a,\xi}(y)=G(y-a)b_\xi. \tag{38}$$

We see that the set of all vectors localized at one space–time point a is a four-dimensional manifold[25]—this is natural since the Dirac particle has internal degrees of freedom. The different states localized at the same point differ for instance by velocities.

Now we look at the Lorentz transform of the vector (37),

$$\begin{aligned}[U(A)\Psi_{a,\xi}](y) &= B(A)\,\Psi_{a,\xi}(\Lambda^{-1}y)\\ &= B(A)G(\Lambda^{-1}y-a)b_\xi\\ &= B(A)G(\Lambda^{-1}y-a)\,B(A)^*\,B(A^{-1})^*b_\xi\end{aligned}$$

using (25),

$$=G(y-\Lambda a)\tilde{B}(A)b_\xi.$$

If we denote $b''_\xi=\tilde{B}(A)b_\xi$ we obtain

$$[U(A)\Psi_{a,\xi}](y)=G(y-\Lambda a)b''_\xi,$$

i.e., again a vector of the form (38). We see from this that the eigenvector of $X_\mu(\eta,0)$ localized at the origin (that is $a=0$) after a Lorentz transformation is again an eigenvector of $X_\mu(\eta,0)$ localized at the same point. It is the velocity (and also the spin) which is the transformed quantity in this case. Therefore, the Dirac particle is an example in which disadvantage II of the Introduction is absent.

4. CONCLUDING REMARKS

The notion of the operator-valued probability density current may be very useful in finding admissible relativistic Hamiltonians. Borowiec and Jadczyk[26] have shown, for instance, that for the current $\mathbf{j}$ and density ρ linked by the relation

$$\mathbf{j}=\tfrac{1}{2}(\mathbf{v}\rho+\rho\mathbf{v}),$$

where $\mathbf{v}=\dot{\mathbf{x}}$ is the velocity, the Hamiltonian ensuring the covariance (iii) under time translations should be linear in the momentum operators. Therefore, for example, the scalar Klein–Gordon particle in the usual formulation is not such a solution since for it $H=(\pm\mathbf{P}^2+m^2)^{1/2}$ and this function evidently is not linear in the P operators. The Duffin–Kemmer equation seems to be more appropriate for our purposes.

There is a hope to localize also the photons on the spacelike hyperplanes (not on null planes as in Ref. 15) since Dirac-like equations have been proposed for the photons.[27]

Barut and Malin maintained[5] that the operator probability density ρ for the scalar particle is not the zeroth component of any 4-vector quantity. We have seen in Sec. 3 that for the Dirac particle the density ρ and the current $\mathbf{j}$ form a covariant 4-vector.

From the example elaborated in Sec. 3 we also learned another lesson: The linear manifold of eigenvectors of $X(\eta,\tau)$ for fixed η and τ, localized in one space–time point should be Lorentz invariant in order to not have disadvantage II of the Introduction. On the other hand, this manifold cannot be one-dimensional since otherwise this would imply that it consists of eigenvectors of Lorentz transformations and thus the Lorentz transformations would commute with $X(\eta,\tau)$. Thus the particle should have internal degrees of freedom. In our example the velocity yields such internal degrees of freedom.

For the massless Dirac particle the Hamiltonian has the form $H=\boldsymbol{\alpha}\cdot\mathbf{P}$. Using the relation $\boldsymbol{\alpha}=2\gamma_5\mathbf{S}$ following from (37), we get $H=2\gamma_5\mathbf{S}\cdot\mathbf{P}$. Introducing the helicity $h=2\mathbf{S}\cdot\mathbf{P}/|\mathbf{P}|=2\mathbf{S}\cdot\mathbf{P}/E$, we obtain $H=\gamma_5 hE$. Hence the quantity introduced in (3) is $\Xi=\gamma_5 h$. Note that the three quantities appearing in this relation commute with each other.

There is a paper by Durand[28] in which a quantized field version of position operators is introduced for the spin-$\frac{1}{2}$ field. This "Dirac position operator" is expressible as a "weighted average of x over the charge density." The charge density obviously is not positive definite and therefore when restricted to the one-particle subspace it cannot be identified with our density $\rho=j^0$. The density of Durand's paper when integrated with $\mathbf{x}$ gives dipole moment rather than "position of the center of charge." Durand discusses a commutativity of the density at spacelike separated points and calls it a condition for a causal theory, but does not intend to show that the product itself vanishes at such points. Therefore she does not touch the question of the causal propagation in our sense.

ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to Dr. A. Z. Jadczyk, Professor J. T. Łopuszański, Dr. B. S. Skagerstam, and Dr. Z. Haba for stimulating discussions and valuable remarks. I am also grateful to the referee for suggestions concerning the δ_η distribution.

[1]T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).

[2]There are exceptions to this solution, namely that particles with zero mass and spin greater than $\frac{1}{2}$ for which such operators cannot exist.

[3]G. G. Fleming, Phys. Rev. B **137**, 188 (1965).

[4]G. N. Fleming, Phys. Rev. B **139**, 963 (1965).

[5]A. O. Barut and S. Malin, Rev. Mod. Phys. **40**, 632 (1968).

[6]J. Petzold, Ann. Phys. **31**, 361, 372 (1974).

[7]B. Gerlach, D. Gromes, J. Petzold, and P. Rosental, Z. Phys. **208**, 381 (1968).

[8]H. J. Borchers, Commun. Math. Phys. **4**, 315 (1967).

[9]G. C. Hegerfeldt, Phys. Rev. D **10**, 3320 (1974).

[10]B. S. Skagerstam, Int. J. Theor. Phys. **15**, 213 (1976).

[11]S. K. Berberian, *Notes on Spectral Theory* (Van Nostrand, Princeton, New Jersey, 1966).

[12]M. A. Naimark, Dokl. Acad. Sci. USSR **41**, 359 (1961).

[13]J. M. Jauch and C. Piron, Helv. Phys. Acta **40**, 559 (1967).

[14]A. Z. Jadczyk and B. Jancewicz, Bull. Acad. Pol., Sér. Math. Astr. Phys. **21**, 477 (1973).).

[15]J. Bertrand, Nuovo Cimento A **15**, 299 (1973).

[16]Similar c-number valued distributions of class $\mathcal{S}'$ on spacelike hypersurfaces were recently discussed by H. Wakita, Rep. Math. Phys. **10**, 311 (1976).

[17]L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).

[18]In proving this fact one also needs the property that $j^\mu(x)$ decreases sufficiently quickly at spacelike infinity. This property is already formalized in the prior assumption $j(\cdot) \in \mathcal{O}'_{M,s}$.

[19]In fact we are able to define more general position operators $X(\Sigma)$ where Σ is not a hyperplane but an arbitrary spacelike hypersurface,

$$X^\mu(\Sigma) = \int_\Sigma y^\mu j^\nu(y) d\sigma_\nu(\eta(y), y)$$

where $\eta(y)$ is a unit vector orthogonal to Σ at $y \in \Sigma$; see Ref. 16.

[20]G. N. Fleming, J. Math. Phys. **9**, 193 (1968).

[21]W. Cegła and B. Jancewicz, Rep. Math. Phys. **11**, 53 (1977).

[22]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, New York, 1961), Chap. 8.

[23]I. Białynicki-Birula and Z. Białynicka-Birula, *Quantum Electrodynamics* (Pergamon and PWN, Warsaw, 1975), Appendix I.

[24]This property of relativistic wave equations has been thoroughly discussed by Dirac in P. A. M. Dirac, Phys. Rev. **73**, 1092 (1948). That paper of Dirac can be viewed as a harbinger of the Fleming hyperplane formalism.

[25]This manifold, of course, is not a subspace of H because it consists of generalized eigenvectors, i.e., vectors belonging to a larger space dual to the domain D.

[26]A. Borowiec and A. Z. Jadczyk, "Relativistic systems with a conserved probability current," in preparation.

[27]R. Mignani, E. Recami, and M. Baldo, Lett. Nuovo Cimento **11**, 568 (1974).

[28]B. Durand, Phys. Rev. D **14**, 299 (1976).

An application of the nonstandard Trotter product formula

Alan Sloan[a)]

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 21 March 1977)

The nonstandard Trotter product formula is used to extend the Feynman integral interpretation of solutions to the Schrödinger equation in the presence of a highly singular potential.

We seek an answer to the question of when a quantum mechanical Hamiltonian H determines a system which evolves according to a Feynman path integral.[1] Specifically, we want to know when there is a potential V so that, upon setting $\hbar = 1$, and fixing $t > 0$

$$\exp(-itH)u = \lim_{n\to\infty} T_{n,V}(u), \quad u \text{ in } L^2(\mathbb{R}^3), \tag{1}$$

where $T_{n,V}(u)$ is defined as the integral

$$(T_{n,V}u)(x_0) = c_n \int_{\mathbb{R}^{3n}} \exp[iS(x_0, x, x_n; n, t, V)] \times u(x_n)\, dx\, dx_n, \tag{2}$$

and

$$c_n = (2\pi it/mn)^{-3n/2}, \quad m > 0,$$

$$x = (x_1, x_2, \ldots, x_{n-1}),$$

and

$$S(x_0, x, x_n; n, t, V) = \sum_{j=1}^{n} \left(\frac{m(x_j - x_{j-1})^2}{2(t/n)^2} - V(x_j) \right) \frac{t}{n}. \tag{3}$$

An explicit computation[2] shows that

$$T_{n,V} = [\exp(-itH_0/n) \exp(-itV/n)]^n, \tag{4}$$

where

$$H_0 = -\Delta/2m \quad \text{on } L^2(\mathbb{R}^3), \tag{5}$$

for V any real measurable function. Consequently, Nelson[2] was able to employ the Trotter product formula[3] to verify (1) in case H is the closure of the operator sum, $H_0 + V$, and V is chosen so that $H_0 + V$ is essentially self-adjoint. Nelson[2] and Faris[4] extended this type of formula to more singular potentials by including an additional limiting operation after analytic continuation in the mass parameter m. Although there has been progress in extending Trotter's formula for $\exp[-t(H_0 + V)]$,[5-7] when V is singular relative to H_0, analogous results are not known for the unitary groups $\exp[it(H_0 + V)]$. Through the use of elementary nonstandard techniques we shall find a formula of the type (1) valid for singular potentials.

For details on nonstandard analysis and notation we refer the reader to Refs. 4 and 8—10. Here we adopt the convention that $X \subset {}^*X$ for any set X. If X is a topological space, a in X, b in ${}^*X$, we write $a \approx b$ if and only if b is in the monad of a. If $a \approx b$ we say a is the standard part of b and that a and b are infinitely close. We use the Euclidean topology on $\mathbb{R}^n$ and the norm topology on $L^2(\mathbb{R}^3)$.

Suppose H is a self-adjoint operator on $L^2(\mathbb{R}^3)$ which can be approximated in the generalized strong sense by bounded perturbations of H_0; i.e., there is a sequence V_K of bounded self-adjoint operators on $L^2(\mathbb{R}^3)$ such that the resolvents of $H_0 + V_K$ converge strongly to those of H. Then, (Ref. 11, 502), providing $\{H_0 + V_K\}$ is uniformly bounded below, it follows that $\exp(-itH)$ is the strong limit of $\exp[it(H + V_K)]$. Thus, for all K in ${}^*\mathbb{N} - \mathbb{N}$, where $\mathbb{N} = \{1, 2, \cdots\}$, and for all u in L^2 we have

$$\exp(-itH)u \approx \exp[-it(H + V_K)]u. \tag{6}$$

By the Trotter product formula and the transfer principle, for each K in ${}^*\mathbb{N}$

$$\exp[-t(H_0 + V_K)] = s\text{-}\lim_{n \text{ in } {}^*\mathbb{N}} {}^{n\to\infty} [\exp(-itH_0/n) \exp(-itV_K/n)]^n. \tag{7}$$

Combining (6) and (7) we find

$$\exp(-itH)u \approx \lim_{n \text{ in } {}^*\mathbb{N}} {}^{n\to\infty} [\exp(-itH_0/n) \exp(-itV_K/n)]^n u. \tag{8}$$

For all finite n and K, (4) shows that

$$T_{n,V_K} = [\exp(-itH_0/n) \exp(-itV_K/n)]^n, \tag{9}$$

so by the transfer principle (9) holds for all n, K in ${}^*\mathbb{N}$. Thus, for each fixed infinite K, (8) yields

$$\exp(-itH)u \approx \lim_{n \text{ in } {}^*\mathbb{N}} {}^{n\to\infty} T_{n,V_K}(u). \tag{10}$$

For each n, K in $\mathbb{N}$, (2) shows that $T_{n,V_K}(u)$ is given by an action integral. The transfer principle then implies that for all n, K in ${}^*\mathbb{N}$ formulas (2) and (3) remain valid with V_K replacing V. Consequently, we may conclude that

$$\exp(-itH)u(\cdot) \approx \lim_{n \text{ in } {}^*\mathbb{N}} {}^{n\to\infty} c_n \int_{\mathbb{R}^{3n}} \exp[iS(\cdot, x, x_n; n, tV_K)] \times u(x_n)\, dx\, dx_n.$$

Lemma: Let X be a separable normed linear space. Let $f_n : X \to {}^*X$ be bounded linear operators which are uniformly bounded in the sense that there is an M in $\mathbb{N}$ such that $\|f_n\| \leq M$ for all n in ${}^*\mathbb{N} \cup \{0\}$. If $f_n \to f_0$ pointwise on X as $n \to \infty$ in ${}^*\mathbb{N}$, there is an N in ${}^*\mathbb{N}$ such that $n > N$ implies $f_n(x) \approx f_0(x)$ for all x in X.

Proof: Let C be a countable dense subset of X. Let λ be a positive infinitesimal. Choose N_c in ${}^*\mathbb{N}$ so that $n > N_c$ implies $\|f_n(c) - f_0(c)\| < \lambda$. Let N be an upper bound for $\{N_c : c \text{ in } C\}$; N exists by Ref. 5, p. 59. Let x in X be arbitrary. Fix $\delta > 0$ in $\mathbb{R}$. There is a c in C so that

[a)]Supported in part by NSF MCS 76-97543.

$||c-x|| < \delta$. Then for $n > N$

$$||f_n(x) - f_0(x)||$$
$$\leq ||f_n(x) - f_n(c)|| + ||f_n(c) - f_0(c)|| + ||f_0(c) - f_0(x)||$$
$$< M||x-c|| + \lambda + M||c-x|| < 3M\delta.$$

Because δ is arbitrary we conclude that $f_n(x) \approx f_0(x)$ whenever $n > N$. Q. E. D.

Combining this lemma with the previous conclusion, noting that $\exp(-itH)$ and the T_{n,V_K} are uniformly bounded by 1, yields the following:

Theorem: If V_K are bounded self-adjoint operators such that $H_0 + V_K$ converge to H in the generalized strong sense as $n \to \infty$ in $\mathbb{N}$, then for each fixed positive infinite integer K there is an N in $^*\mathbb{N}$ such that for all $n > N$ and for all u in L^2

$$\exp(itH)u(\cdot) \approx c_n \int_{\mathbb{R}^{3n}} \exp[iS(\cdot, x, x_n; n, t, V_K)] \times u(x_n)\, dx\, dx_n, \tag{11}$$

in $^*L^2$.

To compare the types of potentials covered by formulas (1) and (11) we note that[2] (1) holds if V is in $L^p + L^\infty$, $p > 2$, whereas (11) holds whenever H is the generalized strong limit of bounded self-adjoint perturbations of H_0. Examples of such H's can be found by defining H to be the form sum of H_0 and V when either

(a) V is in $L^p + L^\infty$, $p > \frac{3}{2}$;

(b) $V \geq 0$ is locally in L^1 outside a closed set of measure zero;

(c) V is a delta function distribution concentrated on the surface of a compact C^1 hypersurface in $\mathbb{R}^3$.

For the necessary approximation theorem in case (b) see Ref. 12 where the V_K's are defined by truncation. In cases (a) and (c) see Ref. 13 where the V_K's are given by regularization of V. See Ref. 13 for additional examples.

Similar techniques apply in dimensions other than 3 and for other H_0's.

[1] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[2] E. Nelson, "Feynman Integrals and the Schrödinger Equation," J. Math. Phys. **5**, 332—43 (1964).
[3] H. Trotter, "Approximation of Semi-groups of Operators," Pacific J. Math. **8**, 887—919 (1958).
[4] M. Davis, *Applied Nonstandard Analysis* (Wiley-Interscience, New York, 1977).
[5] P. Chernoff, "Product Formulas, Nonlinear Semigroups and Addition of Unbounded Operators," Mem. Am. Math. Soc. **140**, 1—121 (1974).
[6] W. Faris, "The Product Formula for Semi-groups defined by Friedrics Extensions," Pacific J. Math. **21**, 47—70 (1967).
[7] T. Kato, "Trotter's Product Formula for an Arbitrary Pair of Selfadjoint Contraction Semigroups," Berkeley preprint.
[8] A. Robinson, *Nonstandard Analysis* (North-Holland, Amsterdam, 1974).
[9] A. Robinson and A. Lightstone, *Non-Archimedian Fields and Asymptotic Expansions* (North-Holland, Amsterdam, 1975).
[10] K. Stroyan and W. Luxemburg, *Introduction to the Theory of Infinitesimals* (Academic, New York, 1976).
[11] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[12] W. Faris, *Self-Adjoint Operators* (Springer, Berlin, 1975).
[13] I. Herbst and A. Sloan, "Perturbation of Translation Invariant Positivity Preserving Semigroups on $L^2(\mathbb{R}^N)$," to appear in Trans. Am. Math. Soc.

One-dimensional, neighbor-pair degeneracies for dumbbells with distinguishable ends

R. B. McQuistan

Department of Physics and Laboratory for Surface Studies, Univeristy of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 8 March 1977)

Expressions are developed which describe exactly the degeneracies for the various kinds of neighbor pairs that can arise when dumbbells with distinguishable ends are distributed on a one-dimensional lattice. Expectation values and normalization constants are also developed.

I. INTRODUCTION

It has been proposed that the role of lateral interactions is important in explaining the restructuring of surfaces resulting from gas adsorption[1-3] and in understanding the absorption–desorption of nonhomonuclear diatomic molecules.[4-7]

In the latter case it has been postulated that the multiplicity of states observed in desorption measurements may arise because of the various *kinds* of lateral interactions that can exist among the adsorbed gas particles and *not* because there are various distinct binding states involved in the adsorbate–adsorbent interaction.

Accordingly, it is of interest and importance to examine thoroughly the various neighbor-pair degeneracies that arise when A-B dumbbells, i. e., dumbbells with identifiable ends, are distributed on a one-dimensional lattice (see Fig. 1). Here we consider an A-B dumbbell to be a particle which occupies two adjacent sites on the lattice space. The purpose of the present paper is to develop expressions which yield the degeneracies for the several kinds of neighbor pairs that arise due to the proximity of the A portions of dumbbells, the B portions of dumbbells and/or the lattice vacancies. Such degeneracies must reflect the constraint that the A and B portions of a particular dumbbell occupy adjacent lattice sites. The A-B dumbbells are considered to be identical, although the orientation of each dumbbell has a twofold degeneracy.

The present work represents a generalization of research previously reported concerning the occupational degeneracy[8] and nearest neighbor degeneracy[9] for λ-bell particles distributed on a one-dimensional lattice.

II. A-s-A PAIR DEGENERACY

An A-s-A pair (see Fig. 2) is defined to be the configuration formed when the A portions of two individual dumbbells are separated by s and only s contiguous vacancies ($s \geq 0$). For example, $s=0$ describes the situation in which the A portions of two individual dumbbells are nearest neighbors; $s=1$ describes the configuration when the two A portions are separated by a single vacancy. The number of A-s-A pairs in a single arrangement is designated n_{AsA} and $A[n_{AsA}|q,N]$ is the number of independent arrangements possible when q indistinguishable A-B dumbbells are arranged on a one-dimensional lattice space of N equivalent sites in such a way as to form exactly n_{AsA} A-s-A pairs.

We shall now show that $A[n_{AsA}|q,N]$ is given by

$$A[n_{AsA}|q,N] = \sum_{j=0}^{[(q-2n_{AsA})/2]} \frac{(-1)^j 2^{q-2(n_{AsA}+j)}[N-q-(s+1)(n_{AsA}+j)]!}{j!\,n_{AsA}![q-2(n_{AsA}+j)]![N-2q-s(n_{AsA}+j)]!}, \tag{1}$$

where $[(q-2n_{AsA})/2]$ is the largest integer contained in $(q-2n_{AsA})/2$.

To deduce Eq. (1) we note that it has been shown previously[10] that $A[\{q_\lambda\},N]$, the occupational degeneracy arising when the set of particles, $\{q_\lambda\}$, is distributed on a one-dimensional lattice of N equivalent sites, is given by the multinomial coefficient

$$A[\{q_\lambda\},N] = \frac{[N-\sum_\lambda(\lambda-1)q_\lambda]!}{\prod_\lambda(q_\lambda!)[N-\sum_\lambda \lambda q_\lambda]!}, \tag{2}$$

where $\{q_\lambda\}$ means that there are q_1 simple particles (each occupying one lattice site), q_2 dumbbell particles (each occupying two adjacent lattice sites), q_3 particles that occupy three contiguous lattice sites, etc. ($\lambda=1,2,3,\cdots$). In the development of Eq. (2) it has been assumed that all the particles having the same value of λ are indistinguishable from one another, but all such particles are distinguishable from the other kinds of particles which have different values of λ.

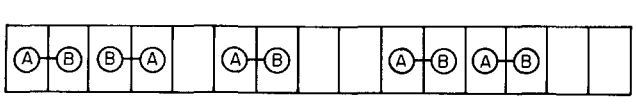


FIG. 1. Five indistinguishable A-B dumbbell particles distributed on a one-dimensional lattice consisting of 15 equivalent sites ($q=5$, $N=15$).

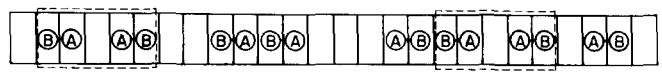


FIG. 2. Eight indistinguishable A-B dumbbell particles are distributed on a one-dimensional lattice consisting of 26 equivalent sites. In the situation depicted there are two A-1-A pairs (enclosed in dashed boxes), one A-3-A pair, one A-0-B pair, one A-1-B pair, one B-0-B pair, and one B-2-B pair.

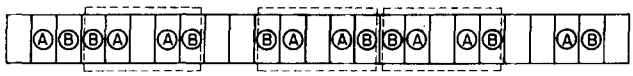


FIG. 3. $q=8$, $n_{A1A}=3$ (enclosed in dashed boxes), $q-2n_{A1A}=2$. Thus there are two remaining dumbbells.

For purposes of the present calculation we assume initially that there are two kinds of particles: one kind with $\lambda=2$ (the A-B dumbbells) and the other kind, a group of "pseudo particles" formed by the A portions of two dumbbells which are separated by s and only s vacant sites ($s=0,1,2,\cdots$). (See Fig. 2.) These latter "particles" possess a $\lambda=s+4$ because the two adjacent dumbbells which contribute their A portions are separated by s vacant sites so that the two dumbbells and the s vacant sites all occupy $s+4$ contiguous sites on the lattice, i.e., B-A-s-A-B. We assume that the A-B dumbbells are distinguishable from the BA-s-AB "particles" and that the AsA "particles" are indistinguishable from each other. We call the latter group of "particles" AsA pairs.

To determine $A[n_{AsA}|q,N]$ we will proceed as follows: We consider that n_{AsA} indistinguishable A-s-A pairs and $q-2n_{AsA}$ additional A-B dumbbells are arranged in all possible ways on the lattice. The set of all such arrangements will contain a subset in which the original n_{AsA} A-s-A pairs appear and no others; another subset will contain the original n_{AsA} A-s-A pairs and one additional A-s-A pair formed from the $q-2n_{AsA}$ A-B dumbbells; another subset will contain the original n_{AsA} A-s-A pairs and two additional A-s-A pairs formed from the $q-2n_{AsA}$ A-B dumbbells. In general, there will be a subset consisting of those arrangements which contain the original n_{AsA} A-s-A pairs and k $[k=0,1,2,\ldots,(q-2n_{AsA})/2]$ extraneous A-s-A pairs formed from the $q-2n_{AsA}$ A-B dumbbells. Only those arrangements for which n_{AsA} is the prescribed number can be counted in $A[n_{AsA}|q,N]$. Consequently, we must subtract those arrangements in which the number of A-s-A pairs is greater than n_{AsA}, i.e., those arrangements for which $k\geq 1$. Thus, if there are n_{AsA} AsA pairs then there are $q-2n_{AsA}$ remaining A-B dumbbells in each arrangement (see Fig. 3). Then, by means of Eq. (2), we determine X_0, the number of independent arrangements possible when n_{AsA} A-s-A pairs and the $q-2n_{AsA}$ remaining A-B dumbbells are arranged in all possible ways on the lattice, to be

$$X_0=\frac{2^{q-2n_{AsA}}[N-q-(s+1)n_{AsA}]!}{n_{AsA}![q-2n_{AsA}!N-2q-sn_{AsA}]!}, \tag{3}$$

where the factor $2^{q-2n_{AsA}}$ arises because the $q-2n_{AsA}$ A-B dumbbells (not associated with the A-s-A pairs) have identifiable ends which can be arranged in $2^{q-2n_{AsA}}$ ways. The A-s-A pairs do not have identifiable ends.

Equation (3) is *not* the number of arrangements containing exactly n_{AsA} A-s-A pairs because Eq. (3) includes, in addition to $A[n_{AsA}|q,N]$, those arrangements which contain $n_{AsA}+1$, $n_{AsA}+2,\ldots,n_{AsA}+k,\ldots,[q/2]$ A-s-A pairs, $[k=1,2,\ldots,(q-2n_{AsA})/2]$. These k *additional* A-s-A pairs are created when the $q-2n_{AsA}$ A-B dumbbells, not associated with the n_{AsA} A-s-A pairs, are arranged with the n_{AsA} A-s-A pairs to form all the possible arrangements. Thus it is possible to write Eq. (3) as

$$X_0=\sum_{k=0}^{[(q-2n_{AsA})/2]}A[n_{AsA},k|q,N], \tag{4}$$

where $A[n_{AsA},k|q,N]$ is the number of independent arrangements containing the initial n_{AsA} A-s-A pairs *and* the k additional A-s-A pairs formed from $2k$ of the $q-2n_{AsA}$ remaining A-B dumbbells.

In accordance with the constraints on distinguishability related to the development of Eq. (2) we note here that the extraneous k A-s-A pairs (created from the $q-2n_{AsA}$ dumbbells) are indistinguishable from one another but are distinguishable from the original n_{AsA} A-s-A pairs.

If n_{AsA} is increased by one, then k, the number of A-s-A pairs formed from the remaining $q-2n_{AsA}$ dumbbells decreases by one; the change in the number of independent arrangements results from the mutual distinguishability of the two kinds of A-s-A pairs. Thus

$$\begin{aligned}A[n_{AsA},k|q,N]&=\left[\binom{n_{AsA}+k}{k}\Big/\binom{n_{AsA}+k}{k-1}\right]\\&\quad\times A[n_{AsA}+1,k-1|q,N]\\&=\left[\frac{n_{AsA}+1}{k}\right]A[n_{AsA}+1,k-1|q,N].\end{aligned} \tag{5}$$

Solving Eq. (4) for $A[n_{AsA},0|q,N]\equiv A[n_{AsA}|q,N]$ we obtain

$$A[n_{AsA}|q,N]=X_0-\sum_{k=1}^{[(q-2n_{AsA})/2]}A[n_{AsA},k|q,N]. \tag{6}$$

Utilizing Eq. (5), Eq. (6) may be rewritten as

$$\begin{aligned}&A[n_{AsA}|q,N]\\&=X_0\sum_{k=1}^{[(q-2n_{AsA})/2]}\left[\frac{n_{AsA}+1}{k}\right]A[n_{AsA}+1,k-1|q,N]\\&=X_0-\binom{n_{AsA}+1}{1}X_1+(n_{AsA}+1)\\&\quad\times\sum_{k=1}^{[(q-2n_{AsA})/2]}\left[\frac{k-1}{k}\right]A[n_{AsA}+1,k-1|q,N]\\&=X_0-\binom{n_{AsA}+1}{1}X_1+(n_{AsA}+1)\\&\quad\times\sum_{k=1}^{[(q-2(n_{AsA}+1))/2]}\left[\frac{k}{k+1}\right]A[n_{AsA}+1,k|q,N]\\&=X_0-\binom{n_{AsA}+1}{1}X_1+(n_{AsA}+1)(n_{AsA}+2)\\&\quad\times\sum_{k=1}^{[(q-2(n_{AsA}+1))/2]}\frac{1}{k+1}A[n_{AsA}+2,k-1|q,N]\\&=X_0-\binom{n_{AsA}+1}{1}X_1+\binom{n_{AsA}+2}{2}X_2-\binom{n_{AsA}+2}{2}\\&\quad\times\sum_{k=2}^{[(q-2(n_{AsA}+1))/2]}\frac{k-1}{k+1}A[n_{AsA}+2,k-1|q,N]\\&=X_0-\binom{n_{AsA}+1}{1}X_1+\binom{n_{AsA}+2}{2}X_2-\binom{n_{AsA}+2}{2}\\&\quad\times\sum_{k=1}^{[(q-2(n_{AsA}+2))/2]}\frac{k}{k+2}A[n_{AsA}+2,k|q,N],\end{aligned} \tag{7}$$

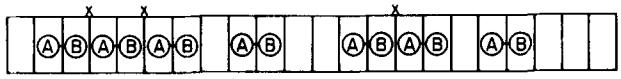


FIG. 4. In this figure there are three A-0-B pairs (marked by X's).

and so forth, where

$$X_j \equiv \sum_{k=0}^{[(q-2(n_{AsA}+j))/2]} A[n_{AsA}+j, k \,|\, q, N]. \tag{8}$$

The recursion process represented in Eq. (7) can be continued with the result that

$$A[n_{AsA}\,|\,q,N] = \sum_{j=0}^{[(q-2n_{AsA})/2]} (-1)^j \binom{n_{AsA}+j}{j} X_j$$

$$= \sum_{j=0}^{[(q-2n_{AsA})/2]} \frac{(-1)^j 2^{q-2(n_{AsA}+j)}[N-q-(s+1)(n_{AsA}+j)]!}{j!\,n_{AsA}!\,[q-2(n_{AsA}+j)]!\,[N-2q-s(n_{AsA}+j)]!}. \tag{9}$$

In the special situation when $s=0$, i.e., when two adjacent sites are occupied by the A portions of two adjacent dumbbells, Eq. (9) becomes (see Appendix A)

$$A[n_{A0A}\,|\,q,N]$$

$$= \sum_{j=0}^{[(q-2n_{A0A})/2]} \frac{(-1)^j 2^{q-2(n_{A0A}+j)}[N-q-(n_{A0A}+j)]!}{j!\,n_{A0A}!\,[q-2(n_{A0A}+j)]!\,[N-2q]!}$$

$$= \frac{[N-2q+n_{A0A}]!}{n_{A0A}!\,[N-2q]!} \sum_{j=0}^{[(q-2n_{A0A})/2]} \times \frac{(-1)^j 2^{q-2(n_{A0A}+j)}[N-q-(n_{A0A}+j)]!}{j!\,[q-2(n_{A0A}+j)]!\,[N-2q+n_{A0A}]!}$$

$$= \binom{N-2q+n_{A0A}}{n_{A0A}} \binom{2N-3q+1}{q-2n_{A0A}}. \tag{10}$$

The entire preceding discussion obviously also is applicable to the degeneracy of B-s-B pairs so that for the degeneracy of such pairs an expression analogous to Eq. (10) can also be written in terms of n_{B0B}.

III. A-s-B PAIR DEGENERACY

An A-s-B pair (see Fig. 4) is defined to be an arrangement in which the A portion of one dumbbell is separated from the B portion of *another* dumbbell by s and only s contiguous vacancies. For example, $s=0$ represents the situation in which an A-B pair (formed by portions of two adjacent dumbbells) is a nearest neighbor pair.

The number of A-s-B pairs on a single lattice is denoted by n_{AsB} and $A[n_{AsB}\,|\,q,N]$ is the degeneracy of those arrangements of q indistinguishable A-B dumbbells on a lattice which contains exactly n_{AsB} A-s-B pairs.

We shall presently show that $A[n_{AsB}\,|\,q,N]$ is given by

$$A[n_{AsB}\,|\,q,N] = \binom{q-1}{n_{AsB}} \left(\sum_{j=0}^{q-n_{AsB}-1} (-1)^j 2^{q-n_{AsB}-j} \right) \times \binom{q-n_{AsB}-1}{j} \binom{N-q-(s+1)(n_{AsB}+j)}{q-j-n_{AsB}}. \tag{11}$$

To derive this relationship we note that if there are n_{AsB} A-s-B pairs in an arrangement, then there are always $q-n_{AsB}$ "s units" (see Fig. 5). Each "s unit" may consist of not more than one A-s-B pair.

The reason that there are always q-n_{AsB} of these "s units" is that the q A-B dumbbells have $q-1$ separations between them; n_{AsB} of these separations are involved in A-s-B pairs and $q-n_{AsB}-1$ separations do not involve A-s-B pairs, i.e., they involve AsA and BsB pairs, or AtB pairs $t \neq s$. Thus there are $q-n_{AsB}$ "s units" each of which is separate from the other "s units."

These $q-n_{AsB}$ "s units" may be permuted with the vacancies to form independent arrangements. The total number of vacancies in each arrangement is $N-2q$ but not all of them are permutable; sn_{AsB} are required to form the n_{AsB} A-s-B pairs. Consequently, the "s units" may be permuted with $N-2q-sn_{AsB}$ vacancies.

There are a total of $N-2q-sn_{AsB}+(q-n_{AsB})=N-q-(s+1)n_{AsB}$ objects of which $q-n_{AsB}$ are "s units" and $N-2q-sn_{AsB}$ are permutable vacancies. These can be arranged in

$$2^{q-n_{AsB}} \binom{N-q-(s+1)n_{AsB}}{q-n_{AsB}}$$

ways. The factor $2^{q-n_{AsB}}$ arises because each of the $q-n_{AsB}$ "s units" may be reversed in orientation, because the ends of the "s units" are identifiable, to form independent arrangements.

The "s units" may be constituted in various ways, i.e., the n_{AsB} A-s-B pairs may be distributed among the "s units" to form "s units" of various lengths. Because there are $q-1$ separations between the q A-B dumbbells, and n_{AsB} of these separations involve A-s-B

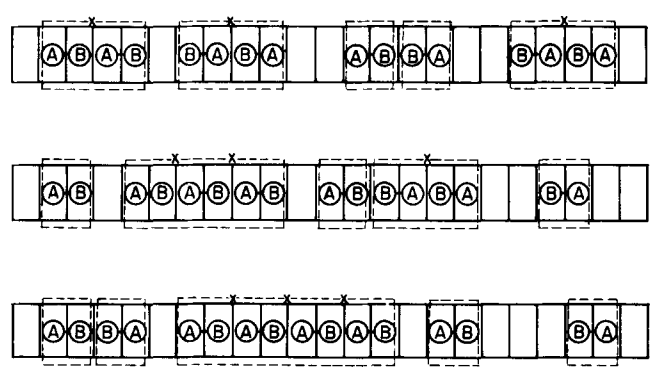


FIG. 5. $q=8$, $N=23$, $n_{A0B}=3$, $q-n_{A0B}=5$. Thus, there are five "0 units." In addition, $n_{A1B}=0$ and $n_{A2B}=1$, so that there are seven "1 units" and six "2 units."

pairs there are $\binom{q-1}{n_{AsB}}$ ways of arranging the n_{AsB} A-s-B pairs among the $q-n_{AsA}$ "s units" to form units of different lengths.

Hence, Y_0, the number of independent ways of arranging the n_{AsB} A-s-B pairs *and* those A-B dumbbells not associated with A-s-B pairs, is given by

$$Y_0 = 2^{q-n_{AsB}} \binom{q-1}{n_{AsB}} \binom{N-q-(s+1)n_{AsB}}{q-n_{AsB}}. \tag{12}$$

It must be made clear here that Y_0 is not the number of independent arrangements containing exactly n_{AsB} A-s-B pairs because there are additional A-s-B pairs formed by the permutation of the $N-2q+sn_{AsB}$ vacancies with the "s units." Thus Y_0 represents, in addition to the arrangements which contain n_{AsB} A-s-B pairs, those arrangements which contain $n_{AsB}+1$, $n_{AsB}+2$, $\cdots$, $n_{AsB}+k$, $\cdots$, $q-1$ A-s-B pairs ($0 \leq k \leq q-n_{AsB}-1$). Thus

$$Y_0 = \sum_{k=0}^{q-n_{AsB}-1} A[n_{AsB}, k \,|\, q, N], \tag{13}$$

where $A[n_{AsB}, k \,|\, q, N]$ is the number of additional arrangements containing the original n_{AsB} A-s-B pairs *and* k additional A-s-B pairs formed by the arrangement of the "s units" with the permutable vacancies. We consider the k additional A-s-B pairs to be distinguishable from the n_{AsB} A-s-B pairs but indistinguishable from each other. Thus

$$\begin{aligned} A[n_{AsB}, k \,|\, q, N] &= \left[\binom{n_{AsB}+k}{k} \Big/ \binom{n_{AsB}+k}{k-1}\right] \\ &\quad \times A[n_{AsB}+1, k-1 \,|\, q, N], \\ &= \left[\frac{n_{AsA}+1}{k}\right] A[n_{AsB}+1,\, k-1 \,|\, q, N], \end{aligned} \tag{14}$$

where

$$\begin{aligned} A[n_{AsB}, 0 \,|\, q, N] &\equiv A[n_{AsB} \,|\, q, N] \\ &= Y_0 - \sum_{k=1}^{q-n_{AsB}-1} A[n_{AsB}, k \,|\, q, N] \\ &= Y_0 - \sum_{k=1}^{q-n_{AsB}-1} \binom{n_{AsB}+1}{k} \\ &= \times A[n_{AsB}+1, k-1 \,|\, q, N]\; Y_0 - \binom{n_{AsB}+1}{1} Y_1 + (n_{AsB}+1) \\ &\quad \times \sum_{k=2}^{q-n_{AsB}-1} \left[\frac{k-1}{k}\right] A[n_{AsB}+1, k-1 \,|\, q, N] \\ &= Y_0 - \binom{n_{AsB}+1}{1} Y_1 + (n_{AsB}+1) \\ &\quad \times \sum_{k=2}^{q-n_{AsB}-2} \frac{k}{k+1} A[n_{AsB}+1, k \,|\, q, N], \text{ etc.}, \\ &= Y_0 - \binom{n_{AsB}+1}{1} Y_1 + \binom{n_{AsB}+2}{2} Y_2 + \cdots, \end{aligned} \tag{15}$$

or

$$A[n_{AsB} \,|\, q, N] = \sum_{j=0}^{q-n_{AsB}-1} (-1)^j \binom{n_{AsB}+j}{j} Y_j, \tag{16}$$

where

$$Y_j \equiv 2^{q-(n_{AsB}+j)} \binom{q-1}{n_{AsB}+j} \binom{N-q-(s+1)(n_{AsB}+j)}{q-(n_{AsB}+j)}, \tag{17}$$

which results in the relation expressed in Eq. (11).

IV. NORMALIZATION

It has been shown[11] that the number of ways of arranging q indistinguishable dumbbells on a one-dimensional lattice consisting of N equivalent sites is $\binom{N-q}{q}$. In the present case, there is a factor of 2^q because of the flip degeneracy associated with dumbbells with identifiable ends. Thus the occupational degeneracy for A-B dumbbells is given by

$$A[q, N] = 2^q \binom{N-q}{q}. \tag{18}$$

Equation (18) could have been obtained by summing Eq. (9) over all values of n_{AsA} (see Appendix B) or by summing Eq. (11) over all values of n_{AsB}. (See Appendix C.)

The fundamental normalization of these statistics is given by[12]

$$\sum_{q=0}^{[N/2]} 2^q \binom{N-q}{q} = \tfrac{1}{3}[2^{N+1} + (-1)^N]. \tag{19}$$

$\langle n_{A0A} \rangle$, the expectation value of A-0-A pairs can be calculated in closed form,

$$\begin{aligned} \langle n_{A0A} \rangle &= \sum_{n_{A0A}} n_{A0A} \binom{N-2q+2n_{A0A}}{n_{A0A}} \binom{2N-3q+1}{q-2n_{A0A}} \Big/ \\ &\quad \times \sum_{n_{A0A}} \binom{N-2q+2n_{A0A}}{n_{A0A}} \binom{2N-3q+1}{q-2n_{A0A}} \\ &= 2^{q-2}[N-2q+1] \binom{N-q-1}{q-2} \Big/ \binom{N-q}{2q} \\ &= \frac{2^{-2}q(q-1)}{N-q}, \end{aligned} \tag{20}$$

so that the expectation value of n_{A0A} per array becomes

$$\lim_{N\to\infty} \frac{\langle n_{A0A} \rangle}{N} = \frac{\theta^2}{8(2-\theta)}, \tag{21}$$

where $\theta \equiv 2q/N$, is the coverage of the space.

ACKNOWLEDGMENT

The author wishes to express his appreciation to Professor David King, the University of Liverpool, and to the British Scientific Research Council for the support that facilitated the work reported in this paper.

APPENDIX A

If

$$a_N(q, n_{A0A}) \equiv \sum_{j=0}^{[(q-2n_{A0A})]/2} \frac{(-1)^j 2^{q-2n_{A0A}-2j}[N-q-n_{A0A}-j]!}{j![q-2n_{A0A}-2j]![N-2q+n_{A0A}]!}, \quad \text{(A1)}$$

then inspection shows that

$$a_N(q, n_{A0A}) = a_{N-3}(q-2, n-1). \quad \text{(A2)}$$

It follows that

$$a_N(1,0) = \binom{2N-2}{1}, \quad \text{(A3)}$$

$$a_N(2,0) = \sum_{j=0}^{1} \frac{(-1)^j 2^{2-2j}[N-2-j]!}{j![2-2j]![N-4]!}$$

$$= \frac{2^2(N-2)!}{0!\,2!\,(N-4)!} - \frac{2^0(N-3)!}{1!\,0!\,(N-4)}$$

$$= \binom{2N-5}{2}, \quad \text{(A4)}$$

or in general

$$a_N(q,0) = \binom{2N-3q+1}{q}. \quad \text{(A5)}$$

Thus employing (A3) and (A5), we see that

$$a_N(q,1) = a_{N-3}(q-2,0)$$

$$= \binom{2N-3q+1}{q-2}. \quad \text{(A6)}$$

Continued utilization of (A2) leads to

$$a_N(q, n_{A0A}) = \binom{2N-3q+1}{q-2n_{A0A}}. \quad \text{(A7)}$$

APPENDIX B

To show the normalization we can take the sum of Eq. (9) over all values of n_{AsA},

$$b_N(q) \equiv \sum_{k=0}^{[q/2]} \frac{1}{k!} \sum_{j=0}^{[(q-2k)/2]} \frac{(-1)^j 2^{q-2(k+j)}[N-q-(s+1)(k+j)]!}{j![q-2(k+j)]![N-2q-s(k+j)]!}, \quad \text{(B1)}$$

so that

$$b_N(0) = 1,$$

$$b_N(1) = \frac{1}{0!} \sum_{j=0}^{0} \frac{(-1)^j 2^{1-2j}[N-1-(s+1)j]!}{j![1-2j]![N-2-sj]!} \quad \text{(B2)}$$

$$= \frac{2[N-1]!}{1![N-2]!} = 2\binom{N-1}{1}, \quad \text{(B3)}$$

$$b_N(2) = \sum_{k=0}^{1} \frac{1}{k!} \sum_{j=0}^{[(2-2k)/2]} \frac{(-1)^j 2^{2-2(k+j)}[N-2-(s+1)(k+j)]!}{j![2-2(k+j)]![N-4-s(k+j)]!}$$

$$= \frac{1}{0!}\left\{\frac{2^2[N-2]}{0!\,2![N-4]} - \frac{2^0[N-2-(s+1)]!}{1!\,0![N-4-s]!}\right\}$$

$$+ \frac{1}{1!}\left\{\frac{2^0[N-2-(s+1)]!}{0![0]![N-4-s]!}\right\}$$

$$= 2^2\binom{N-2}{2} \quad \text{(B4)}$$

or in general

$$b_N(q) = 2^q\binom{N-q}{q}. \quad \text{(B5)}$$

APPENDIX C

Another way to show the normalization expressed in Eq. (18) is to sum Eq. (11) over all values of n_{AsB},

$$c_N(q) \equiv \sum_{k=0}^{q-1} \binom{q-1}{k} \sum_{j=0}^{q-k-1} (-1)^j 2^{q-(k+j)} \binom{q-k-1}{j} \times \binom{N-q-(s+1)(k+j)}{q-(k+j)}, \quad \text{(C1)}$$

so that

$$c_N(0) = 1,$$

$$c_N(1) = 2^q\binom{N-1}{1}, \quad \text{(C2)}$$

$$c_N(2) = \sum_{k=0}^{1}\binom{1}{k}\sum_{j=0}^{1-k}(-1)^j 2^{2-(k+j)}\binom{1-k}{j} \times \binom{N-2-(s+1)(k+j)}{2-(k+j)}$$

$$= \binom{1}{0} 2^2\binom{1}{0}\binom{N-2}{2} - 2^1\binom{1}{1}\binom{N-2-(s+1)}{2-1} + \binom{1}{1}\left\{2^1\binom{0}{0}\binom{N-2-(s+1)}{2-1}\right\}$$

$$= 2^2\binom{N-2}{2}, \quad \text{(C3)}$$

or in general

$$c_N(q) = 2^q\binom{N-q}{q}.$$

[1] J. Anderson and P.J. Estrop, J. Chem. Phys. **46**, 563 (1967).
[2] D.L. Adams and L.H. Germer, Surf. Sci. **32**, 205 (1972).
[3] H. Ibach and J.E. Rowe, Surf. Sci. **43**, 481 (1974).
[4] J.K. Roberts, *Some Problems in Adsorption* (Cambridge U.P., London, 1939).
[5] J.S. Wang, Proc. Roy. Soc. A **161**, 127 (1937).
[6] T. Toya, J. Vac. Sci. Technol. **9**, 890 (1972).
[7] C.G. Goymour and D.A. King, J. Chem. Soc. Faraday Trans. 1 **69**, 749 (1973).
[8] R.B. McQuistan, Nuovo Cimento **58**, 86 (1968).
[9] R.B. McQuistan, J. Math. Phys. **18**, 1200 (1977).
[10] R.B. McQuistan, J. Math. Phys. **14**, 1258 (1973).
[11] D. Lichtman and R.B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
[12] See for example, J. Riordan, *Combinatorial Identities* (Wiley, New York, 1968), p. 76.

On the Kerr–Tomimatsu–Sato family of spinning mass solutions

Masatoshi Yamazaki

Department of Physics, Kanazawa University, Kanazawa 920, Japan
(Received 25 July 1977)

Some analytic properties of the Kerr–Tomimatsu–Sato family of solutions with arbitrary integer deformation parameter δ for gravitational fields of spinning masses are studied. It is shown that all ring singularities are of first order and all ergosurfaces are simple zeros of metric functions A.

1. PRELIMINARIES

It was shown[1] that the Kerr[2] and the three Tomimatsu–Sato[3,4] spinning mass solutions, i.e., stationary axisymmetric, asymptotically flat exact solutions to Einstein's vacuum field equations, can be written in a concise unified expression with arbitrary positive integer distortion parameter δ.

The purpose of the present paper is to study the analytic properties of ring singularities and ergosurfaces of the Kerr–Tomimatsu–Sato family of spinning mass solutions with arbitrary integer deformation parameter δ. It is shown for solutions with arbitrary δ that all ring singularities are of first order and all ergosurfaces are simple zeros of metric function A. It is conjectured that the number of ring singularities is $[\delta/2]$ for $1<x$ [see Eq. (8) for definition of x] and $\delta-[\delta/2]$ for $x<-1$, respectively, and that the number of ergosurfaces is δ for $1<x$ and also δ for $x<-1$, respectively. [Square brackets denote the integral part of the number enclosed.]

At first we shall briefly outline the solutions for the convenience of subsequent discussions. The solutions are of the form

$$ds^2=f^{-1}\{e^{2\gamma}(d\rho^2+dz^2)+\rho^2 d\phi^2\}-f(dt-\omega d\phi)^2 \tag{1}$$

with

$$f=A/B, \tag{2}$$

$$\omega=-2mqb\,C/A, \tag{3}$$

$$e^{2\gamma}=A/p^{2\delta}(a-b)^{\delta^2}, \tag{4}$$

$$A=F(\delta^2)=\sum_{r=1}^{\delta}e(r)c(\delta,r)f(r)F(\delta^2-r)$$

$$=\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}(p^2a^rb^{1-r'}+q^2b^ra^{1-r'})h(\delta,r,r')F(\delta^2-r), \tag{5}$$

$$B=F(\delta^2)+2px\sum_{r=1}^{\delta}d(r)a^{r-1}\sum_{r'=r}^{\delta}c(\delta,r')F(\delta^2-r')$$

$$+2\sum_{r=1}^{\delta}c(\delta,r)F(\delta^2-r), \tag{6}$$

$$C=\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}a^{1-r'}b^{r-1}\{-pxg(\delta,r,r')-h(\delta,r,r')\}F(\delta^2-r), \tag{7}$$

$$\rho=(mp/\delta)(x^2-1)^{1/2}(1-y^2)^{1/2}, \quad z=(mp/\delta)xy, \tag{8}$$

$$a=x^2-1, \quad b=y^2-1, \quad p^2+q^2=1, \tag{9}$$

$$F(\delta^2-r)=\frac{(-1)^{r-1}}{re(r)c(\delta,r)}\frac{\delta!(\delta+1)!\cdots(2\delta-1)!}{\{2!\,3!\cdots(\delta-1)!\}^3}$$

$$\times\det\left(\frac{f(s+t-1)}{s+t-1}\right),$$

$$s=1,2,\cdots,r-1,r+1,\cdots,\delta, \quad t=2,3,\cdots,\delta,$$

$$r=1,2,\cdots,\delta, \tag{10}$$

$$F(\delta^2)=\frac{\delta!(\delta+1)!\cdots(2\delta-1)!}{\{2!\,3!\cdots(\delta-1)!\}^3}\det\left(\frac{f(r+r'-1)}{r+r'-1}\right),$$

$$r,r'=1,2,\cdots,\delta, \tag{11}$$

$$f(r)=p^2a^r+q^2b^r, \quad r=1,2,3,\cdots, \tag{12}$$

$$c(\delta,r)=\frac{2^{2r-1}\delta(\delta+r)!}{(\delta+r)(\delta-r)!\,(2r)!}, \tag{13}$$

$$d(r)=\frac{(-1)^{r-1}(2r-2)!}{\{2^{r-1}(r-1)!\}^2}, \tag{14}$$

$$e(r)=-2d(r+1), \tag{15}$$

$$g(\delta,r,r')=\frac{(-1)^{r-1}2^{2r'-2}}{r!\,(r-1)!\,(2r'-2)!}$$

$$\times\sum_{s=\max(r,r')}^{\delta}\frac{(s+r-2)!\,(s+r'-2)!}{(s-r)!\,(s-r')!}, \tag{16}$$

and

$$h(\delta,r,r')=\frac{(-1)^{r-1}2^{2r'-2}(\delta+r-1)!\,(\delta+r'-1)!}{(r+r'-1)(\delta-r)!\,r!\,(r-1)!\,(\delta-r')!\,(2r'-1)!}. \tag{17}$$

The mass is $M=mc^2/G$ and the angular momentum is

$$J=\frac{GM^2}{c}q=\frac{m^2c^3}{G}q=\frac{mc^3}{G}\alpha.$$

Another form of the solutions is

$$ds^2=\frac{\Sigma^2}{\Delta}dr^2+\Sigma^2d\theta^2+\frac{B\Delta\sin^2\theta}{A}d\phi^2$$

$$-\frac{A}{B}\left(dt-\frac{2\alpha C\sin^2\theta}{A}d\phi\right)^2 \tag{18}$$

with

$$r=(mp/\delta)x+m, \quad \cos\theta=y, \tag{19}$$

$$\rho=\Delta^{1/2}\sin\theta, \quad z=(r-m)\cos\theta, \tag{20}$$

$$\Delta=(r-m)^2-(mp/\delta)^2=(mp/\delta)^2a, \tag{21}$$

$$\Sigma^2=\frac{Bm^{2\delta}}{\delta^2(m^2-\alpha^2)^{\delta-1}}\left(\frac{\delta^2\Delta}{m^2-\alpha^2}+\sin^2\theta\right)^{1-\delta^2}, \tag{22}$$

$$m^2q = m\alpha, \quad m^2p^2 = m^2 - \alpha^2, \tag{23}$$

and

$$\sqrt{-g} = \Sigma^2 \sin\theta. \tag{24}$$

When $q = 0$ and $p = 1$, the solutions reduce to the Weyl solutions[5,6] with integer distortion parameter δ.

2. RING SINGULARITIES AND ERGOSURFACES

The solution ξ to Ernst's equation[7]

$$(\xi\xi^* - 1)\nabla^2\xi = 2\xi^*\nabla\xi\cdot\nabla\xi \tag{25}$$

are

$$\xi = \frac{\sum_{r=1}^{\delta} d(r)(px d^{r-1} - iqy b^{r-1})\sum_{r'=r}^{\delta} c(\delta, r')F(\delta^2 - r')}{\sum_{r=1}^{\delta} c(\delta, r)F(\delta^2 - r)}$$

$$= \frac{H + iI}{G} = \frac{k + j}{k - j}, \tag{26}$$

where G, H and I are real functions and j and k are complex functions of x, y, p, and q. Metric functions A and B are written in the form

$$A = (jk^* + j^*k)/2 \tag{27}$$

and

$$B = kk^* \geq 0. \tag{28}$$

We see from Eq. (26) that on the equatorial plane where $y = 0$ complex functions j and k become real J and K. Then from Eqs. (5) and (6) we have the metric functions $\bar{A} = A_{y=0}$ and $\bar{B} = B_{y=0}$ on the equatorial plane which are of the form

$$\bar{A} = \frac{\delta!\,(\delta+1)!\cdots(2\delta-1)!}{\{2!\,3!\cdots(\delta-1)!\}^3}\det\left(\frac{p^2x^{2(r+r'-1)} - 1}{r + r' - 1}\right)$$

$$= \frac{\det((p^2x^{2(r+r'-1)} - 1)/(r + r' - 1))}{\det(1/(r + r' - 1))}$$

$$= J(x, p)K(x, p), \quad r, r' = 1, 2, \ldots, \delta, \tag{29}$$

and

$$\bar{B} = (K(x, p))^2 \geq 0 \tag{30}$$

with

$$J(x, p) = \det\left(\frac{(r + r' - 2)!}{(r-1)!\,(r'-1)!}\{px^{r+r'-1} + (-1)^{r'}\}\right),$$

$$r, r' = 1, 2, \ldots, \delta, \tag{31}$$

and

$$K(x, p) = \det\left(\frac{(r + r' - 2)!}{(r-1)!\,(r'-1)!}\{px^{r+r'-1} + (-1)^{r'-1}\}\right),$$

$$r, r' = 1, 2, \ldots, \delta. \tag{32}$$

Next we consider the third metric function $\bar{C} = C_{y=0}$ on the equatorial plane. From

$$\frac{\partial\omega}{\partial x} = \frac{2b}{A^2}\,\frac{mp}{\delta}\left(B\frac{\partial I}{\partial y} - I\frac{\partial B}{\partial y}\right) \tag{33}$$

and Eq. (3) we obtain

$$JK\frac{d\bar{C}}{dx} - \bar{C}\frac{d}{dx}(JK) = K^2N \tag{34}$$

with

$$N = \frac{p}{\delta}\sum_{r=1}^{\delta}(-1)^{r-1}d(r)\sum_{r'=r}^{\delta}c(\delta, r')(F(\delta^2 - r'))_{y=0}$$

$$= -\frac{p}{\delta q}\left(\frac{I}{y}\right)_{y=0} = -\frac{p}{\delta q}\left(\frac{\partial I}{\partial y}\right)_{y=0}. \tag{35}$$

Equation (34) is a relation for polynomials J, K, N, and $\bar{C}$. Therefore, polynomial $\bar{C}$ must be of the form

$$\bar{C} = (J(x, p) - L(x, p))K(x, p). \tag{36}$$

We then obtain

$$L\frac{dJ}{dx} - J\frac{dL}{dx} = N, \quad L = J\left\{-\int\frac{N}{J^2}\,dx + 1\right\} \tag{37}$$

and

$$\bar{C} = JK\int(N/J^2)\,dx = \bar{A}\int(N/J^2)\,dx. \tag{38}$$

The metric functions $\bar{A}$, $\bar{B}$, and $\bar{C}$ on the equatorial plane have the common factor polynomial $K(x, p)$. Polynomials K and J have the following properties:

$$K(x, p) = (-1)^{\delta}J(-x, p), \tag{39}$$

$$K(x, p) = (-1)^{\delta}J(x, -p), \tag{40}$$

$$K(x, p) = (-1)^{[\delta/2]}x^{\delta^2}p^{\delta}K(x^{-1}, p^{-1}), \tag{41}$$

$$J(x, p) = (-1)^{\delta-[\delta/2]}x^{\delta^2}p^{\delta}J(x^{-1}, p^{-1}), \tag{42}$$

$$K(x, p) \underset{p=1}{\to} (x + 1)^{\delta(\delta+1)/2}(x - 1)^{\delta(\delta-1)/2}, \tag{43}$$

$$K(x, p) \underset{x=1}{\to} (p + 1)^{[(\delta+1)/2]}(p - 1)^{[\delta/2]}, \tag{44}$$

$$K(x, p) \underset{p=-1}{\to} (-1)^{\delta}(x + 1)^{\delta(\delta-1)/2}(x - 1)^{\delta(\delta+1)/2}, \tag{45}$$

and

$$K(x, p) \underset{x=-1}{\to} (-1)^{\delta}(p + 1)^{[\delta/2]}(p - 1)^{[(\delta+1)/2]}. \tag{46}$$

The polynomials K and J are both of degree δ^2 in x and degree δ in p. From the relation

$$H^2 + I^2 = (A + G)G \text{ and } \bar{I} = 0$$

we get

$$\bar{A} = (\bar{H}^2 - \bar{G}^2)/\bar{G} = \{(\bar{H} - \bar{G})/(\bar{G})^{1/2}\}\{(\bar{H} + \bar{G})/(\bar{G})^{1/2}\} = JK, \tag{47}$$

$$J = (\bar{H} - \bar{G})/(\bar{G})^{1/2}, \tag{48}$$

and

$$K = (\bar{H} + \bar{G})/(\bar{G})^{1/2}, \tag{49}$$

where $\bar{G} = G_{y=0}$, $\bar{H} = H_{y=0}$, and $\bar{I} = I_{y=0}$. From the relation $B = A + 2H + 2G$ we get

$$\bar{B} = (\bar{H} + \bar{G})^2/\bar{G} = K^2. \tag{50}$$

The position of ring singularities is determined by the vanishing of polynomials K, i.e.,

$$K(x, p) = 0. \tag{51}$$

On the equatorial plane the surface of the ergosphere is located where

$$J(x, p)K(x, p) = 0. \tag{52}$$

From Eq. (25) we obtain

$$K^2\left\{J\frac{d}{dx}(x^2-1)\frac{dJ}{dx}-(x^2-1)\left(\frac{dJ}{dx}\right)^2\right\}$$
$$-J^2\left\{K\frac{d}{dx}(x^2-1)\frac{dK}{dx}-(x^2-1)\left(\frac{dK}{dx}\right)^2\right\}$$
$$=-\frac{4q\delta N}{p(K-J)}\left\{KJM+(K+J)\frac{q\delta N}{p}\right\}=-T \tag{53}$$

with

$$M=2i\left(\frac{\partial k}{\partial y}-\frac{\partial j}{\partial y}\right)_{y=0} \tag{54}$$

and N given in Eq. (35). The polynomials $M(x,p)$ are real for $\delta\geq 2$ and null for $\delta=1$. The polynomials $N(x,p)/p$ are of degree δ^2-1 and $\delta-1$ in x^2 and p^2, respectively. When $K(x_K,p)=0$ with real x_K and $1>p>0$, we get from Eq. (53)

$$(x_K^2-1)\left(J(x_K,p)\frac{dK}{dx}(x_K,p)\right)^2=\left(\frac{2q\delta}{p}N(x_K,p)\right)^2. \tag{55}$$

Equation (55) shows that $x_K>1$ or $x_K<-1$ [$N(\pm 1,p)>0$ when $1>p>0$]. In other words there are no zero of polynomials $K(x,p)$ in the domain $-1\leq x\leq 1$. The same statement holds for the polynomials $J(x,p)$ on account of Eq. (39).

Next, if we could have simultaneously $K(x_K,p)=0$ and $(dK/dx)(x_K,p)=0$ $(1>p>0)$ at some real x_K ($|x_K|>1$), then we could get from Eq. (53)

$$N(x_K,p)=0. \tag{56}$$

However, the polynomials $N(x,p)$ of degree δ^2-1 in x^2 are positive definite and never vanish [see Eq. (35)] when $|x|\geq 1$ and $1>p>0$. Thus, Eq. (56) is impossible, and the polynomials K and dK/dx never simultaneously vanish at real x. Therefore, the polynomials $K(x,p)$ have no multiple zero. The same statement holds for the polynomials $J(x,p)$ on account of Eq. (39).

By using the relations

$$2KJM+(K+J)(2q\delta/p)N(x,p)+(J-K)(2q\delta/p)$$
$$\times N(x^{-1},p^{-1})p^{2\delta-1}x^{2\delta^2-1}=0, \tag{57}$$

$$T=(2q\delta/p)N\{(2KJM+(K+J)(2q\delta/p)N)/(K-J)\}$$
$$=(2q\delta/p)N(x,p)(2q\delta/p)N(x^{-1},p^{-1})p^{2\delta-1}x^{2\delta^2-1}, \tag{58}$$

$$(2q\delta/p)N(x,p)=(\eta(x,p)K(x,p)+\zeta(x,p)J(x,p))/2, \tag{59}$$

and

$$(2q\delta/p)N(x^{-1},p^{-1})p^{2\delta-1}x^{2\delta^2-1}$$
$$=(\eta(x,p)K(x,p)-\zeta(x,p)J(x,p))/2, \tag{60}$$

we get from Eq. (53)

$$K\frac{d}{dx}(x^2-1)\frac{dK}{dx}-(x^2-1)\left(\frac{dK}{dx}\right)^2+\frac{\zeta^2}{4}-\delta^2K^2=0 \tag{61}$$

and

$$J\frac{d}{dx}(x^2-1)\frac{dJ}{dx}-(x^2-1)\left(\frac{dJ}{dx}\right)^2+\frac{\eta^2}{4}-\delta^2J^2=0, \tag{62}$$

where

$$\eta=2qK(1,1)\quad\text{and}\quad\zeta=-2qJ(1,1). \tag{63}$$

Here the polynomials $K(1,1)$ and $J(1,1)$ are the cofactors of polynomials K and J [see Eqs. (31) and (32)] with respect to the element, the row $r=1$ and column $r'=1$. When $x=1$

$$\eta=2\delta q(p-1)^{[\delta/2]}(p+1)^{[(\delta-1)/2]} \tag{64}$$

and

$$\zeta=-2\delta q(p+1)^{[\delta/2]}(p-1)^{[(\delta-1)/2]}. \tag{65}$$

From

$$\frac{2a}{A^2}\left\{H\frac{\partial}{\partial x}B-B\frac{\partial}{\partial x}H\right\}=\frac{\partial P}{\partial y} \tag{66}$$

and

$$P=\frac{2qy}{pqA}\delta\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}p^2a^rb^{1-r'}g(\delta,r,r')F(\delta^2-r) \tag{67}$$

we get

$$K\frac{dJ}{dx}-J\frac{dK}{dx}=\frac{2\delta p}{a}\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}a^r(-1)^{1-r'}g(\delta,r,r')\bar{F}(\delta^2-r) \tag{68}$$

with $\bar{F}(\delta^2-r)=(F(\delta^2-r))_{y=0}$. The polynomials on the right-hand side of Eq. (68) are positive definite and never vanish when $|x|>1$ or $a>0$. So the Wronskians $K\,dJ/dx-J\,dK/dx$ never vanish in the domain $|x|>1$.

From Eq. (41) we obtain the number of positive zeros N_p^K of the polynomials K;

$$N_p^K=[\delta/2]+\text{even number}. \tag{69}$$

From Eqs. (42) and (39) we obtain the number of negative zeros N_n^K of the polynomials K;

$$N_n^K=\delta-[\delta/2]+\text{even number}. \tag{70}$$

From $A'(a,p^2/q^2)=(-1)^\delta(p^2/q^2)^\delta a^{\delta^2}A'(1/a,q^2/p^2)$ with $A'=\bar{A}/q^{2\delta}$ we obtain the number of positive zeros $N_p^{\bar{A}}$ of the polynomials $\bar{A}=JK$;

$$N_p^{\bar{A}}=\delta+\text{even number}. \tag{71}$$

From Eq. (5) we get the number of positive zeros N_p^A of the polynomials A with nonvanishing y;

$$N_p^A=\delta+\text{even number}. \tag{71'}$$

From experiences with Sturm's chain calculations for several solutions with smaller δ we conjecture that the number of ring singularities is $[\delta/2]$ for $1<x$ and $\delta-[\delta/2]$ for $x<-1$ and that the number of ergosurface is δ for $1<x$ and also δ for $x<-1$.

3. OTHER DISCUSSIONS

I. *Event horizon*: Two independent Killing vectors ξ_t and ξ_ϕ satisfy on the surface $x=1$ or $a=0$

$$\xi_t^2=-q^{2\delta}b^{\delta^2}/B(x=1)=(-1)^{\delta^2+1}\times\text{positive quantities} \tag{72}$$

and

$$\xi_\phi^2=-\frac{\{2qmbC(x=1)\}^2}{q^{2\delta}b^{\delta^2}B(x=1)}=(-1)^{\delta^2+1}\times\text{positive quantities} \tag{73}$$

on account of Eqs. (11), (28), and $b=-(1-y^2)$. Both Killing vectors ξ_t and ξ_ϕ become on the surface $x=1$

spacelike or timelike for odd or even δ solutions, respectively. The surface $x=1$ $(a=0)$ is not the event horizon for the even δ solutions, as it contains timelike Killing vectors.[8,9]

II. $q\to 1$ *with finite* xp: When the angular momentum parameter q approaches one (p approaches null) under the constraint the quantity xp is finite, we have

$$K\to(-1)^{[\delta/2]}(\delta xp+1), \tag{74}$$

$$J\to(-1)^{[(\delta-1)/2]}(\delta xp-1), \tag{75}$$

$$\bar{A}\to(-1)^{\delta-1}(\delta^2x^2p^2-1), \tag{76}$$

and

$$\xi\to(px/\delta)-iqy. \tag{77}$$

Equations (74)—(77) show that all solutions with $\delta\geq 2$ approach[10–12] the Kerr solution ($\delta=1$). But two points must be stressed. The first point is that the solutions with even δ are contrary to the asymptotic flatness condition as the metric function $\bar{A}$ becomes negative there. The second is that the equatorial radius of the surface of ergosphere is reduced by the factor $1/\delta^2$.

III. $q>1$: The solutions, Eqs. (1)—(17), can be extended to the case when q exceeds one. When $q>1$, prolate spheroidal coordinate variables x and y in Eq. (8) and rotation parameters p and q in Eq. (9) are modified to oblate spheroidal coordinates x' and y' and parameters p' and q' connected by the relations

$$x=ix',\quad y=y',\quad p=-ip',\quad \text{and } q=q'.$$

We have no event horizon for the solutions with any δ because of the nonvanishing $\Delta=(r-m)^2+(mp'/\delta)^2$ [see Eq. (21)]. When $q>1$ we have no ring singularity at all for $x'>0$, because the polynomials $(-1)^{[\delta/2]}K(x',p')$ are positive definite and never vanish when $x'>0$.

ACKNOWLEDGMENTS

The author would like to dedicate the present paper to the late Professor Kantaro Senda. He would like to thank Professor S. Hori and Professor E. Yamada for many interesting discussions.

APPENDIX

1. Metric functions A, B, and C

$\delta=1$ (*Kerr solution*): $A=F(1)$, $B=F(1)+2pxF(0)+2F(0)$, $C=-[px+1]F(0)$.

$\delta=2$ (*T—S solution*): $A=F(4)$, $B=F(4)+2px[(4F(3)+4F(2))-a(2F(2))]+2[4F(3)+4F(2)]$,

$C=-[(2px+4)+a^{-1}(4px+4)]F(3)+b[(px+3)+a^{-1}(4px+4)]F(2)$.

$\delta=3$ (*T—S solution*): $A=F(9)$, $B=F(9)+2px[(9F(8)+24F(7)+16F(6))-a(12F(7)+8F(6))+a^2(6F(6))]$
$+2[9F(8)+24F(7)+16F(6)]$,

$C=-[(3px+9)+a^{-1}(16px+24)+a^{-2}(16px+16)]F(8)+b[(4px+18)+a^{-1}(40px+64)+a^{-2}(48px+48)]F(7)$
$-b^2[(2px+10)+a^{-1}(24px+40)+a^{-2}(32px+32)]F(6)$.

$\delta=4$ (*T—S solution*):

$A=F(16)$,

$B=F(16)+2px[(16F(15)+80F(14)+128F(13)+64F(12))-a(40F(14)+64F(13)+32F(12))+a^2(48F(13)+24F(12))$
$-a^3(20F(12))]+2[16F(15)+80F(14)+128F(13)+64F(12)]$,

$C=-[(4px+16)+a^{-1}(40px+80)+a^{-2}(96px+128)+a^{-3}(64px+64)]F(15)+b[(10px+60)+a^{-1}(184px+400)$
$+a^{-2}(528px+720)+a^{-3}(384px+384)]F(14)-b^2[(12px+80)+a^{-1}(264px+600)+a^{-2}(832px+1152)$
$+a^{-3}(640px+640)]F(13)+b^3[(5px+35)+a^{-1}(120px+280)+a^{-2}(400px+560)+a^{-3}(320px+320)]F(12)$.

$\delta=5$:

$A=F(25)$,

$B=F(25)+2px[(25F(24)+200F(23)+560F(22)+640F(21)+256F(20))-a(100F(23)+280F(22)+320F(21)+128F(20))$
$+a^2(210F(22)+240F(21)+96F(20))-a^3(200F(21)+80F(20))+a^4(70F(20))]$
$+2[25F(24)+200F(23)+560F(22)+640F(21)+256F(20)]$,

$C=-[(5px+25)+a^{-1}(80px+200)+a^{-2}(336px+560)+a^{-3}(512px+640)+a^{-4}(256px+256)]F(24)$
$+b[(20px+150)+a^{-1}(584px+1600)+a^{-2}(2928px+5040)+a^{-3}(4864px+6144)+a^{-4}(2560px+2560)]F(23)$
$-b^2[(42px+350)+a^{-1}(1464px+4200)+a^{-2}(8032px+14112)+a^{-3}(14080px+17920)+a^{-4}(7680px+7680)]F(22)$
$+b^3[(40px+350)+a^{-1}(1520px+4480)+a^{-2}(8800px+15680)+a^{-3}(16000px+20480)$
$+a^{-4}(8960px+8960)]F(21)-b^4[(14px+126)+a^{-1}(560px+1680)+a^{-2}(3360px+6048)$
$+a^{-3}(6272px+8064)+a^{-4}(3584px+3584)]F(20)$.

2. Polynomials $F(\delta^2 - r)$ and $F(\delta^2)$

$\delta = 1$: $F(0) = 1$, $F(1) = p^2 a + q^2 b$.

$\delta = 2$: $F(2) = p^2 a^2 + q^2 b^2$, $F(3) = p^2 a^3 + q^2 b^3$, $F(4) = p^4 a^4 + p^2 q^2 (4a^3 b - 6a^2 b^2 + 4ab^3) + q^4 b^4$.

$\delta = 3$:

$$F(6) = p^4 a^6 + p^2 q^2 (9a^4 b^2 - 16a^3 b^3 + 9a^2 b^4) + q^4 b^6,$$

$$F(7) = p^4 a^7 + p^2 q^2 (6a^5 b^2 - 5a^4 b^3 - 5a^3 b^4 + 6a^2 b^5) + q^4 b^7,$$

$$F(8) = p^4 a^8 + p^2 q^2 (16a^5 b^3 - 30a^4 b^4 + 16a^3 b^5) + q^4 b^8,$$

$$F(9) = p^6 a^9 + p^4 q^2 (9a^8 b - 36a^7 b^2 + 84a^6 b^3 - 90a^5 b^4 + 36a^4 b^5) + p^2 q^4 (36a^5 b^4 - 90a^4 b^5 + 84a^3 b^6 - 36a^2 b^7 + 9ab^8) + q^6 b^9.$$

$\delta = 4$:

$$\begin{aligned} F(12) = {} & p^6 a^{12} + p^4 q^2 (36a^{10} b^2 - 160a^9 b^3 + 315a^8 b^4 - 288a^7 b^5 + 100a^6 b^6) \\ & + p^2 q^4 (100a^6 b^6 - 288a^5 b^7 + 315a^4 b^8 - 160a^3 b^9 + 36a^2 b^{10}) + q^6 b^{12}, \end{aligned}$$

$$\begin{aligned} F(13) = {} & p^6 a^{13} + p^4 q^2 (30a^{11} b^2 - 114a^{10} b^3 + 170a^9 b^4 - 63a^8 b^5 - 70a^7 b^6 + 50a^6 b^7) \\ & + p^2 q^4 (50a^7 b^6 - 70a^6 b^7 - 63a^5 b^8 + 170a^4 b^9 - 114a^3 b^{10} + 30a^2 b^{11}) + q^6 b^{13}, \end{aligned}$$

$$\begin{aligned} F(14) = {} & p^6 a^{14} + p^4 q^2 (20a^{12} b^2 - 40a^{11} b^3 - 54a^{10} b^4 + 272a^9 b^5 - 315a^8 b^6 + 120a^7 b^7) \\ & + p^2 q^4 (120a^7 b^7 - 315a^6 b^8 + 272a^5 b^9 - 54a^4 b^{10} - 40a^3 b^{11} + 20a^2 b^{12}) + q^6 b^{14}, \end{aligned}$$

$$\begin{aligned} F(15) = {} & p^6 a^{15} + p^4 q^2 (100a^{12} b^3 - 450a^{11} b^4 + 828a^{10} b^5 - 700a^9 b^6 + 225a^8 b^7) \\ & + p^2 q^4 (225a^7 b^8 - 700a^6 b^9 + 828a^5 b^{10} - 450a^4 b^{11} + 100a^3 b^{12}) + q^6 b^{15}, \end{aligned}$$

$$\begin{aligned} F(16) = {} & p^8 a^{16} + p^6 q^2 (16a^{15} b - 120a^{14} b^2 + 560a^{13} b^3 - 1420a^{12} b^4 + 1968a^{11} b^5 - 1400a^{10} b^6 + 400a^9 b^7) \\ & + p^4 q^4 (400a^{12} b^4 - 2400a^{11} b^5 + 6608a^{10} b^6 - 11\,040a^9 b^7 + 12\,870a^8 b^8 - 11\,040a^7 b^9 + 6608a^6 b^{10} - 2400a^5 b^{11} \\ & + 400a^4 b^{12}) + p^2 q^6 (400a^7 b^9 - 1400a^6 b^{10} + 1968a^5 b^{11} - 1420a^4 b^{12} + 560a^3 b^{13} - 120a^2 b^{14} + 16ab^{15}) + q^8 b^{16}. \end{aligned}$$

$\delta = 5$:

$$\begin{aligned} F(20) = {} & p^8 a^{20} + p^6 q^2 (100a^{18} b^2 - 800a^{17} b^3 + 3075a^{16} b^4 - 6496a^{15} b^5 + 7700a^{14} b^6 - 4800a^{13} b^7 + 1225a^{12} b^8) \\ & + p^4 q^4 (2500a^{14} b^6 - 16\,800a^{13} b^7 + 51\,275a^{12} b^8 - 93\,600a^{11} b^9 + 113\,256a^{10} b^{10} - 93\,600a^9 b^{11} + 51\,275a^8 b^{12} \\ & - 16\,800a^7 b^{13} + 2500a^6 b^{14}) + p^2 q^6 (1225a^8 b^{12} - 4800a^7 b^{13} + 7700a^6 b^{14} - 6496a^5 b^{15} \\ & + 3075a^4 b^{16} - 800a^3 b^{17} + 100a^2 b^{18}) + q^8 b^{20}, \end{aligned}$$

$$\begin{aligned} F(21) = {} & p^8 a^{21} + p^6 q^2 (90a^{19} b^2 - 670a^{18} b^3 + 2340a^{17} b^4 - 4257a^{16} b^5 + 3766a^{15} b^6 - 810a^{14} b^7 - 945a^{13} b^8 + 490a^{12} b^9) \\ & + p^4 q^4 (1750a^{15} b^6 - 10\,170a^{14} b^7 + 24\,885a^{13} b^8 - 30\,970a^{12} b^9 + 14\,508a^{11} b^{10} + 14\,508a^{10} b^{11} - 30\,970a^9 b^{12} \\ & + 24\,885a^8 b^{13} - 10\,170a^7 b^{14} + 1750a^6 b^{15}) + p^2 q^6 (490a^9 b^{12} - 945a^8 b^{13} - 810a^7 b^{14} + 3766a^6 b^{15} \\ & - 4257a^5 b^{16} + 2340a^4 b^{17} - 670a^3 b^{18} + 90a^2 b^{19}) + q^8 b^{21}, \end{aligned}$$

$$\begin{aligned} F(22) = {} & p^8 a^{22} + p^6 q^2 (75a^{20} b^2 - 480a^{19} b^3 + 1295a^{18} b^4 - 1152a^{17} b^5 - 1570a^{16} b^6 + 4496a^{15} b^7 - 3780a^{14} b^8 + 1120a^{13} b^9) \\ & + p^4 q^4 (875a^{16} b^6 - 2800a^{15} b^7 - 2610a^{14} b^8 + 28\,640a^{13} b^9 - 67\,782a^{12} b^{10} + 87\,360a^{11} b^{11} - 67\,782a^{10} b^{12} \\ & + 28\,640a^9 b^{13} - 2610a^8 b^{14} - 2800a^7 b^{15} + 875a^6 b^{16}) + p^2 q^6 (1120a^9 b^{13} - 3780a^8 b^{14} + 4496a^7 b^{15} \\ & - 1570a^6 b^{16} - 1152a^5 b^{17} + 1295a^4 b^{18} - 480a^3 b^{19} + 75a^2 b^{20}) + q^8 b^{22}, \end{aligned}$$

$$\begin{aligned} F(23) = {} & p^8 a^{23} + p^6 q^2 (50a^{21} b^2 - 175a^{20} b^3 - 315a^{19} b^4 + 3458a^{18} b^5 - 9240a^{17} b^6 + 11\,910a^{16} b^7 - 7644a^{15} b^8 + 1960a^{14} b^9) \\ & + p^4 q^4 (3675a^{16} b^7 - 22\,050a^{15} b^8 + 55\,520a^{14} b^9 - 70\,812a^{13} b^{10} + 33\,670a^{12} b^{11} + 33\,670a^{11} b^{12} - 70\,812a^{10} b^{13} \\ & + 55\,520a^9 b^{14} - 22\,050a^8 b^{15} + 3675a^7 b^{16}) + p^2 q^6 (1960a^9 b^{14} - 7644a^8 b^{15} + 11\,910a^7 b^{16} - 9240a^6 b^{17} + 3458a^5 b^{18} \\ & - 315a^4 b^{19} - 175a^3 b^{20} + 50a^2 b^{21}) + q^8 b^{23}, \end{aligned}$$

$$\begin{aligned} F(24) = {} & p^8 a^{24} + p^6 q^2 (400a^{21} b^3 - 3150a^{20} b^4 + 11\,088a^{19} b^5 - 21\,280a^{18} b^6 + 23\,040a^{17} b^7 - 13\,230a^{16} b^8 + 3136a^{15} b^9) \\ & + p^4 q^4 (11\,025a^{16} b^8 - 78\,400a^{15} b^9 + 250\,848a^{14} b^{10} - 473\,760a^{13} b^{11} + 580\,580a^{12} b^{12} - 473\,760a^{11} b^{13} \\ & + 250\,848a^{10} b^{14} - 78\,400a^9 b^{15} + 11\,025a^8 b^{16}) + p^2 q^6 (3136a^9 b^{15} - 13\,230a^8 b^{16} + 23\,040a^7 b^{17} \\ & - 21\,280a^6 b^{18} + 11\,088a^5 b^{19} - 3150a^4 b^{20} + 400a^3 b^{21}) + q^8 b^{24}, \end{aligned}$$

$$F(25) = p^{10}a^{25} + p^8q^2(25a^{24}b - 300a^{23}b^2 + 2300a^{22}b^3 - 10\,150a^{21}b^4 + 26\,880a^{20}b^5 - 43\,400a^{19}b^6 + 41\,800a^{18}b^7$$
$$- 22\,050a^{17}b^8 + 4900a^{16}b^9) + p^6q^4(2500a^{21}b^4 - 26\,250a^{20}b^5 + 133\,700a^{19}b^6 - 438\,900a^{18}b^7 + 1\,059\,525a^{17}b^8$$
$$- 2\,007\,450a^{16}b^9 + 3\,023\,760a^{15}b^{10} - 3\,553\,200a^{14}b^{11} + 3\,158\,400a^{13}b^{12} - 2\,041\,900a^{12}b^{13}$$
$$+ 904\,200a^{11}b^{14} - 245\,000a^{10}b^{15} + 30\,625a^9b^{16}) + p^4q^6(30\,625a^{16}b^9 - 245\,000a^{15}b^{10} + 904\,200a^{14}b^{11}$$
$$- 2\,041\,900a^{13}b^{12} + 3\,158\,400a^{12}b^{13} - 3\,553\,200a^{11}b^{14} + 3\,023\,760a^{10}b^{15} - 2\,007\,450a^9b^{16} + 1\,059\,525a^8b^{17}$$
$$- 438\,900a^7b^{18} + 133\,700a^6b^{19} - 26\,250a^5b^{20} + 2500a^4b^{21}) + p^2q^8(4900a^9b^{16} - 22\,050a^8b^{17} + 41\,800a^7b^{18}$$
$$- 43\,400a^6b^{19} + 26\,880a^5b^{20} - 10\,150a^4b^{21} + 2300a^3b^{22} - 300a^2b^{23} + 25ab^{24}) + q^{10}b^{25}.$$

3. Coefficient $c(\delta, r)$

The letter r numbers the column.

$\delta = 1$	1				
$\delta = 2$	4	4			
$\delta = 3$	9	24	16		
$\delta = 4$	16	80	128	64	
$\delta = 5$	25	200	560	640	256

4. Coefficient $d(r)$

1 $\quad -1/2 \quad 3/8 \quad -5/16 \quad 35/128$

5. Coefficient $e(r)$

1 $\quad -3/4 \quad 5/8 \quad -35/64 \quad 63/128$

6. Coefficient $g(\delta, r, r')$

The letters r and r' number the row and column, respectively.

$\delta = 1$	1				
$\delta = 2$	2	4			
	−1	−4			
$\delta = 3$	3	16	16		
	−4	−40	−48		
	2	24	32		
$\delta = 4$	4	40	96	64	
	−10	−184	−528	−384	
	12	264	832	640	
	−5	−120	−400	−320	
$\delta = 5$	5	80	336	512	256
	−20	−584	−2928	−4864	−2560
	42	1464	8032	14 080	7680
	−40	−1520	−8800	−16 000	−8960
	14	560	3360	6272	3584

7. Coefficient $h(\delta, r, r')$

The letters r and r' number the row and column, respectively.

$\delta = 1$	1	
$\delta = 2$	4	4
	−3	−4

$\delta=3$	9	24	16		
	−18	−64	−48		
	10	40	32		
$\delta=4$	16	80	128	64	
	−60	−400	−720	−384	
	80	600	1152	640	
	−35	−280	−560	−320	
$\delta=5$	25	200	560	640	256
	−150	−1600	−5040	−6144	−2560
	350	4200	14112	17920	7680
	−350	−4480	−15680	−20480	−8960
	126	1680	6048	8064	3584

[1]M. Yamazaki, Prog. Theor. Phys. **57**, 1951 (1977).
[2]R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[3]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972).
[4]A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973).
[5]B. H. Vorhees, in *Methods of Local and Global Differential Geometry in General Relativity*, edited by D. Farnsworth, J. Fink, J. Porter, and A. Thomson (Springer-Verlag, Berlin, 1972).
[6]H. Weyl, Ann. Physik (Leipzig) **54**, 117 (1917).
[7]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[8]G. W. Gibbons and R. A. Russell-Clark, Phys. Rev. Lett. **30**, 398 (1973).
[9]A. Tomimatsu and H. Sato, Lett. Nuovo Cimento **8**, 740 (1973).
[10]J. M. Bardeen and R. V. Wagoner, Astrophys. J. **167**, 359 (1971).
[11]J. B. Hartle and K. S. Thorne, Astrophys. J. **153**, 807 (1968).
[12]W. Kinnersley and E. F. Kelley, J. Math. Phys. **15**, 2121 (1974).

On an exceptional case concerning plasma oscillations

C. E. Siewert

Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina 27607
(Received 2 February 1977)

The method of elementary solutions is used to analyze the situation in plasma oscillations when there exists a real discrete eigenvalue.

I. INTRODUCTION

In two recent papers[1,2] it was argued that the original van Kampen—Case method of elementary solutions[3,4] was incomplete for a particular situation concerning the solution of the linearized Vlasov equation for a collisionless plasma. Here we show that the method of elementary solutions does, in fact, yield the correct result for the considered exceptional case. To establish the required notation, we review the development of the solutions to

$$-\frac{i}{k}\frac{\partial}{\partial t}F_k(v,t)+vF_k(v,t)+\eta(v)\int_{-\infty}^{\infty}F_k(v',t)\,dv'=0. \tag{1}$$

Here $F_k(v,t)$ is the Fourier transform of the perturbed distribution function, v is the speed, k is the transform variable, and

$$\eta(v)=-\frac{\omega_p^2}{k^2}\frac{d}{dv}f_0(v). \tag{2}$$

In Eq. (2), $f_0(v)$ represents the equilibrium distribution and ω_p is the plasma frequency,

$$\omega_p^2=4\pi Ne^2/m, \tag{3}$$

where e and m are respectively the charge and mass of the electron and N is the charge density.

On substituting solutions of the form

$$F_k(v,t)=\phi(\nu,v)\exp(-i\nu kt) \tag{4}$$

into Eq. (1), we find

$$(\nu-v)\phi(\nu,v)=\eta(v)\int_{-\infty}^{\infty}\phi(\nu,v')\,dv'. \tag{5}$$

The solutions can be normalized by taking

$$\int_{-\infty}^{\infty}\phi(\nu,v')\,dv'=1, \tag{6}$$

and thus the continuum solutions corresponding to $\nu\in(-\infty,\infty)$ can be written as

$$\phi(\nu,v)=\eta(v)\frac{P}{\nu-v}+\lambda(\nu)\delta(\nu-v), \tag{7}$$

where

$$\lambda(\nu)=1+P\int_{-\infty}^{\infty}\eta(s)\frac{ds}{s-\nu}. \tag{8}$$

If we allow ν to be complex, then the discrete solutions are

$$\phi(\nu_\alpha,v)=\frac{1}{\nu_\alpha-v}\eta(v),\quad \alpha=1,2,3,\ldots,\kappa, \tag{9}$$

where ν_α is used to denote a zero of

$$\Lambda(z)=1+\int_{-\infty}^{\infty}\eta(v)\frac{dv}{v-z}. \tag{10}$$

If we write a general solution of Eq. (1) as

$$F_k(v,t)=\sum_{\alpha=1}^{\kappa}A_\alpha\phi(\nu_\alpha,v)\exp(-i\nu_\alpha kt)+\int_{-\infty}^{\infty}A(\nu)\phi(\nu,v)\exp(-i\nu kt)\,d\nu, \tag{11}$$

then the expansion coefficients A_α and $A(\nu)$ must satisfy the initial condition

$$F(v)=\sum_{\alpha=1}^{\kappa}A_\alpha\phi(\nu_\alpha,v)+\int_{-\infty}^{\infty}A(\nu)\phi(\nu,v)\,d\nu,\quad v\in(-\infty,\infty), \tag{12}$$

where $F(v)=F_k(v,0)$. Case has shown[4] that Eq. (12) can be solved when $F(v)$ and $A(\nu)$ are Hölder continuous functions and the discrete eigenvalues are not real. However, as Simon and Rosenbluth have pointed out,[1] there is a problem with Case's original solution for the exceptional case when any of the discrete eigenvalues becomes real and thus becomes embedded in the continuum.

In a recent elegant paper[2] Arthur, Greenberg, and Zweifel have used methods of functional analysis to develop the solution for this elusive case. We show here how the use of singular-integral equations yields the correct result.

II. ANALYSIS

For the sake of brevity, we consider that there is only one discrete eigenvalue that is embedded in the continuum and that it is a simple zero of $\Lambda(z)$. To illustrate concisely the points of principal interest here, we consider further that there are no other discrete eigenvalues. Thus we investigate

$$F(v)=A\phi(\nu_1,v)+\eta(v)P\int_{-\infty}^{\infty}A(\nu)\frac{d\nu}{\nu-v}+\lambda(v)A(v),\quad v\in(-\infty,\infty), \tag{13}$$

where $F(v)$ is an arbitrary, though specified, Hölder continuous function.[5] Here since

$$\Lambda^+(\nu_1)=\Lambda^-(\nu_1)=0, \tag{14}$$

where the + and − are used to denote limiting values as the real axis is approached from above and below, we see that

$$\lambda(\nu_1)=\eta(\nu_1)=0. \tag{15}$$

We consider $\lambda(\nu)$ and $\eta(\nu)$ to be differentiable and thus write

$$\phi(\nu_1,\nu_1)=-\eta'(\nu_1). \tag{16}$$

If we consider that $A(\nu)$ is a Hölder continuous function, then we can follow Muskhelishvili,[5] introduce

$$N(z)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}A(\nu)\frac{d\nu}{\nu-z}, \tag{17}$$

and thus rewrite Eq. (13) as

$$F(v)-A\phi(\nu_1,v)=N^+(v)\Lambda^+(v)-N^-(v)\Lambda^-(v),$$
$$v\in(-\infty,\infty), \tag{18}$$

which can be solved to yield

$$N(z)=\frac{1}{2\pi i\Lambda(z)}\int_{-\infty}^{\infty}[F(v)-A\phi(\nu_1,v)]\frac{dv}{v-z}. \tag{19}$$

However, it is clear that $N(z)$ will have a "pole" on the real axis unless we impose the conditions

$$P\int_{-\infty}^{\infty}[F(v)-A\phi(\nu_1,v)]\frac{dv}{v-\nu_1}$$
$$\pm\pi i[F(\nu_1)-A\phi(\nu_1,\nu_1)]=0, \tag{20}$$

which clearly *cannot* be satisfied for $F(v)$ arbitrary. We thus conclude that, in general, Eq. (13) has no solution with $A(\nu)$ restricted to be Hölder continuous.

Having decided that Eq. (11) does not represent a sufficiently general solution of Eq. (1), we wish to consider, for this exceptional case,

$$E_k(v,t)=B\pi i\,\mathrm{sgn}(k)\phi(\nu_1,v)\exp(-i\nu_1kt)$$
$$+\frac{B}{\nu_1-v}[-\pi i\,\mathrm{sgn}(k)\eta(v)+\lambda(v)]\exp(-ivkt), \tag{21}$$

where B is an arbitrary constant. By direct substitution we can readily verify that Eq. (21) satisfies Eq. (1). For this special case, we now replace Eq. (11) with

$$F_k(v,t)=A\phi(\nu_1,v)\exp(-ik\nu_1t)$$
$$+\int_{-\infty}^{\infty}A(\nu)\phi(\nu,v)\exp(-ikvt)\,d\nu+E_k(v,t), \tag{22}$$

where $A(\nu)$ is a Hölder function. To show that the initial condition can be satisfied, we proceed to establish the solution to

$$F(v)=A\phi(\nu_1,v)+\int_{-\infty}^{\infty}A(\nu)\phi(\nu,v)\,d\nu+\frac{B}{\nu_1-v}\lambda(v),$$
$$v\in(-\infty,\infty). \tag{23}$$

Again, we introduce

$$N(z)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}A(\nu)\frac{d\nu}{\nu-z} \tag{24}$$

and find

$$N(z)=\frac{1}{2\pi i\Lambda(z)}\int_{-\infty}^{\infty}\hat{F}(s)\frac{ds}{s-z}, \tag{25}$$

where

$$\hat{F}(s)=F(s)-A\phi(\nu_1,s)-\frac{B}{\nu_1-s}\lambda(s). \tag{26}$$

Now to remove the "poles" from $N(z)$, we must impose the conditions

$$P\int_{-\infty}^{\infty}\hat{F}(s)\frac{ds}{s-\nu_1}=0 \tag{27}$$

and

$$\hat{F}(\nu_1)=0. \tag{28}$$

If we substitute Eq. (26) into Eqs. (27) and (28), then we can evaluate two of the integrals to obtain

$$P\int_{-\infty}^{\infty}F(s)\frac{ds}{\nu_1-s}=\lambda'(\nu_1)A-\pi^2B\eta'(\nu_1) \tag{29}$$

and

$$F(\nu_1)=-A\eta'(\nu_1)-\lambda'(\nu_1)B. \tag{30}$$

Clearly we can eliminate between Eqs. (29) and (30) to obtain

$$A=[\lambda'^2(\nu_1)+\pi^2\eta'^2(\nu_1)]^{-1}\Big[\lambda'(\nu_1)$$
$$\times P\int_{-\infty}^{\infty}F(s)\frac{ds}{\nu_1-s}-\pi^2\eta'(\nu_1)F(\nu_1)\Big] \tag{31}$$

and

$$B=-[\lambda'^2(\nu_1)+\pi^2\eta'^2(\nu_1)]^{-1}\Big[\eta'(\nu_1)$$
$$\times P\int_{-\infty}^{\infty}F(s)\frac{ds}{\nu_1-s}+\lambda'(\nu_1)F(\nu_1)\Big]. \tag{32}$$

With A and B as given by Eqs. (31) and (32), we can use Eq. (25) to find

$$A(\nu)\Lambda^+(\nu)\Lambda^-(\nu)=\int_{-\infty}^{\infty}\Big[\eta(\nu)\frac{P}{\nu-s}+\lambda(\nu)\delta(\nu-s)\Big]\hat{F}(s)\,ds, \tag{33}$$

and upon entering Eq. (26) into Eq. (33), we obtain

$$A(\nu)=\frac{\eta(\nu)}{\Lambda^+(\nu)\Lambda^-(\nu)}\int_{-\infty}^{\infty}[\phi^\dagger(\nu,s)-l(\nu)\phi^\dagger(\nu_1,s)]F(s)\,ds, \tag{34}$$

where

$$\phi^\dagger(\nu,s)=\frac{P}{\nu-s}+\frac{\lambda(\nu)}{\eta(\nu)}\delta(\nu-s) \tag{35}$$

and

$$l(\nu)=\frac{\Lambda^+(\nu)\Lambda^-(\nu)}{(\nu-\nu_1)\eta(\nu)}[\lambda'^2(\nu_1)+\pi^2\eta'^2(\nu_1)]^{-1}\eta'(\nu_1). \tag{36}$$

We note that $A(\nu)$ has a removable singularity at $\nu=\nu_1$.

Finally, we conclude that A as given by Eq. (31), B as given by Eq. (32), and $A(\nu)$ as given by Eq. (34) constitute the desired solution of Eq. (23).

ACKNOWLEDGMENTS

The author is grateful to P. F. Zweifel for pointing out the interest in this problem and to E. E. Burniston for several helpful discussions.

[1]A. Simon and M. N. Rosenbluth, Phys. Fluids **19**, 1567 (1976).
[2]M. D. Arthur, W. Greenberg, and P. F. Zweifel, "Vlasov theory of plasma oscillations: Linear approximation," Phys. Fluids **20**, 1296 (1977).
[3]N. G. van Kampen, Physica **21**, 949 (1955).
[4]K. M. Case, Ann. Phys. **7**, 349 (1959).
[5]N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).

On the separation of Einsteinian substructures

Jerzy F. Plebański[a)]

Centro de Investigación y de Estudios Avanzados del I.P.N., México D.F., 14—740, Mexico
(Received 23 February 1977)

Within the spinorial version of the Cartan structure formulas with the built-in (complex) Einstein vacuum equations some closed semi-Einsteinian substructures are isolated and discussed. Then the idea of semigraviton is introduced, and its relationship to Penrose's nonlinear graviton is described.

I. INTRODUCTION

The formal structure of "complex relativity" (see Ref. 1) was compactly described in Sec. 1 of Ref. 2, which has been a starting point of a series of papers of our group,[2–6] dedicated to the study of "heavens" [this terminology was introduced by Newman at the GR-7 conference (see Ref. 7 and also 8 and 9)].

In the pure spinorial notation, this structure can be recapitulated as follows: A complex four-dimensional Riemannian space V_4 is a pair consisting of a four-dimensional analytic differential manifold M (which carriers the tangent structure of $\Lambda=\oplus_{p=0}^{4}\Lambda^p$) and the metric

$$g=-\tfrac{1}{2}g_{A\dot{B}}\underset{s}{\otimes}g^{A\dot{B}}, \tag{1.1}$$

where $g_{A\dot{B}}$ with $A=1,2$ and $\dot{A}=\dot{1},\dot{2}$ form a base of the cotangent space Λ^1; the spinorial indices are to be manipulated by the spinorial Levi-Civita symbols,

$$(\epsilon_{AB}):=(\epsilon_{\dot{A}\dot{B}}):=\begin{pmatrix}0 & +1\\ -1 & 0\end{pmatrix}=:(\epsilon^{\dot{A}\dot{B}})=:(\epsilon^{AB}), \tag{1.2}$$

according to the standard rules

$$\psi^A=\psi_B\epsilon^{BA},\quad \psi_A=\epsilon_{AB}\psi^B. \tag{1.3}$$

Describing the Riemannian structure, first we shall list the relation

$$g^{A\dot{B}}\wedge g^{C\dot{D}}=:\epsilon^{AC}S^{\dot{B}\dot{D}}+S^{AC}\epsilon^{\dot{B}\dot{D}}, \tag{1.4}$$

which defines in Λ^2 the symmetric objects $S_{AB}=S_{(AB)}$ and $S_{\dot{A}\dot{B}}=S_{(\dot{A}\dot{B})}$, which under the Hodge star operation are respectively self-dual and anti-self-dual, i.e., $*S_{AB}=+S_{AB}$, $*S_{\dot{A}\dot{B}}=-S_{\dot{A}\dot{B}}$ [for the definition of $*$ see Ref. 2, (1.12)]. The objects S_{AB} and $S_{\dot{A}\dot{B}}$ form a base of Λ^2 which can thus be interpreted as decomposed into a direct sum $\Lambda^2=D(\Lambda^2)\oplus\dot{D}(\Lambda^2)$ (D standard for "dual," $\dot{D}$ for "anti-dual").

Then, we list the first structure equations

$$dg^{A\dot{B}}=g^{A\dot{S}}\wedge\Gamma^{\dot{B}}{}_{\dot{S}}+g^{S\dot{B}}\wedge\Gamma^{A}{}_{S}, \tag{1.5}$$

where $\Gamma_{AB}=\Gamma_{(AB)}$ and $\Gamma_{\dot{A}\dot{B}}=\Gamma_{(\dot{A}\dot{B})}$ are connection 1-forms.

Next, we list a set of relations which involve objects with the pure undotted spinorial indices: These are

$$\text{(a) } \mathcal{R}^{A}{}_{B}:=d\Gamma^{A}{}_{B}+\Gamma^{A}{}_{S}\wedge\Gamma^{S}{}_{B}$$
$$=-\tfrac{1}{2}C^{A}{}_{BCD}S^{CD}+\tfrac{1}{24}RS^{A}{}_{B}+\tfrac{1}{2}C^{A}{}_{B\dot{C}\dot{D}}S^{\dot{C}\dot{D}}\in\Lambda^2,$$

$$\mathcal{S}_H:\ \text{(b) } dS^{AB}=-3S^{(AB}\wedge\Gamma^{C)}{}_{C}\qquad\in\Lambda^3,$$
$$\text{(c) } 3S^{AB}\wedge S_{CD}=\delta^{A}{}_{(C}\delta^{B}{}_{D)}S^{RS}\wedge S_{RS}(\neq 0)\qquad\in\Lambda^4. \tag{1.6}$$

Equations (1.6a) constitute three (out of six) Cartan second structure formulas; $C_{ABCD}=C_{(ABCD)}$ is the spinorial image of the self-dual part of the conformal curvature; $C_{AB\dot{C}\dot{D}}=C_{(AB)\dot{C}\dot{D}}=C_{AB(\dot{C}\dot{D})}$ describes the traceless part of the Ricci tensor and R is the scalar curvature. Equations (1.6b) are easily deduced from (1.5) by applying the definition of S_{AB} implicit in (1.4); from the last equation, one also easily deduces the validity of (1.6c). We introduced $\mathcal{S}_H$ as the symbol for the whole set of relations (1.6).

A simple practical manner of investigating the algebraic relations involving S objects consists in using a null tetrad e^a ($a=1,2,3,4$) introduced by

$$\Lambda^1\ni(g^{A\dot{B}}):=\sqrt{2}\begin{pmatrix}e^4, & e^2\\ e^1, & -e^3\end{pmatrix}\rightarrow g=2e^1\underset{s}{\otimes}e^2+2e^3\underset{s}{\otimes}e^4. \tag{1.7}$$

in the terms of which

$$S^{11}=2e^4\wedge e^2,\quad S^{12}=e^1\wedge e^2+e^3\wedge e^4,\quad S^{22}=2e^3\wedge e^1, \tag{1.8}$$

and

$$S^{\dot{1}\dot{1}}=2e^4\wedge e^1,\quad S^{\dot{1}\dot{2}}=-e^1\wedge e^2+e^3\wedge e^4,\quad S^{\dot{2}\dot{2}}=2e^3\wedge e^2. \tag{1.9}$$

By using these explicit forms of S objects one easily deduces the algebraic relations which hold in Λ^4:

$$\mathcal{E}:\ \text{(a) } S^{RS}\wedge S_{RS}+S^{\dot{R}\dot{S}}\wedge S_{\dot{R}\dot{S}}=0,\quad \text{(b) } S^{AB}\wedge S_{\dot{C}\dot{D}}=0. \tag{1.10}$$

(We introduced $\mathcal{E}$ as the symbol for this set of relations.)

Particularly important is (1.10b) which shows that from the point of view of Λ^4, the spaces $D(\Lambda^2)$ and $\dot{D}(\Lambda^2)$ are wedge orthogonal,

$$D(\Lambda^2)\wedge\dot{D}(\Lambda^2)=0. \tag{1.11}$$

This property, and the fact that

$$\dim\{D(\Lambda^2)\}=3=\dim\{\dot{D}(\Lambda^2)\}, \tag{1.12}$$

can be interpreted in the sense that the decomposition of Λ^2 into a direct sum

$$\Lambda^2=D(\Lambda^2)\oplus\dot{D}(\Lambda^2), \tag{1.13}$$

can be defined *even in the absence of the metrical structure*, when the concept of $*$ is not defined. [It is, however, worthwhile to observe that within the metrical structure, with $**=\mathrm{id}$ on Λ, we have the usual step

[a)] On leave of absence from the University of Warsaw, Warsaw, Poland.

product $\alpha \lrcorner \beta := *(\alpha \wedge *\beta)$. The natural inner product in Λ^2 is then $: \Lambda^2 \times \Lambda^2 \to \mathbb{C}$, and it is endowed with the property $*\alpha \lrcorner \beta = \alpha \lrcorner *\beta$. Hence, with $*S_{AB} = +S_{AB}$ and $*S_{\dot{A}\dot{B}} = -S_{\dot{A}\dot{B}}$, necessarily $S_{AB}\ S^{\dot{C}\dot{D}} = 0$, which proves that $D(\Lambda^2) \lrcorner D(\Lambda^2) = 0$, the last equation being equivalent to (1.11).]

As the final set of relations describing the Riemannian structure, we shall list the set $\mathcal{S}_{\dot{H}}$ of equations, which involve objects with the pure dotted spinorial indices. These are

$$\mathcal{S}_{\dot{H}}: \begin{cases} \text{(a) } \mathcal{R}^{\dot{A}}{}_{\dot{B}} = d\Gamma^{\dot{A}}{}_{\dot{B}} + \Gamma^{\dot{A}}{}_{\dot{S}} \wedge \Gamma^{\dot{S}}{}_{\dot{B}} \\ \qquad = -\tfrac{1}{2} C^{\dot{A}}{}_{\dot{B}\dot{C}\dot{D}} S^{\dot{C}\dot{D}} + \tfrac{1}{24} R S^{\dot{A}}{}_{\dot{B}} + \tfrac{1}{2} S^{CD} C_{CD}{}^{\dot{A}}{}_{\dot{B}} \in \Lambda^2, \\ \text{(b) } dS^{\dot{A}\dot{B}} = -3 S^{(\dot{A}\dot{B}} \wedge \Gamma^{\dot{C})}{}_{\dot{C}} \in \Lambda^3, \\ \text{(c) } 3 S^{\dot{A}\dot{B}} \wedge S_{\dot{C}\dot{D}} = \delta^{\dot{A}}{}_{(\dot{C}} \delta^{\dot{B}}{}_{\dot{D})} S^{\dot{R}\dot{S}} \wedge S_{\dot{R}\dot{S}} \ (\neq 0) \in \Lambda^4. \end{cases} \tag{1.14}$$

The equations (1.14a) are the remaining three Cartan (second structure) formulas. Beside the curvature objects described above ($C_{AB\dot{C}\dot{D}}$ and R), there enters here the spinorial image of the anti-self-dual part of the conformal curvature, $C_{ABCD} = C_{(\dot{A}\dot{B}\dot{C}\dot{D})}$. Equations (1.14b) can be deduced easily from (1.4) and (1.5). The relations (1.14c) are a direct consequence of the definition of $S_{\dot{A}\dot{B}}$ through (1.4).

The integrability conditions of (1.5) are automatically satisfied with the curvature objects possessing symmetries described above, i.e., d applied on (1.5) by using (1.6a) and (1.14a) reduces them to $0 = 0$. The integrability conditions of (1.6a) and (1.14a) consist of the Bianchi identities. With the base of the tangent space spanned by $\partial_{A\dot{B}}$ (i.e., $dT = -\tfrac{1}{2} g^{A\dot{B}} \partial_{A\dot{B}} T$, for every $T \in \Lambda^0$), which generalizes in the standard manner to the spinorial covariant differential operation $\nabla_{A\dot{B}}$ (which involves *both* types of connection, $\Gamma^{A}{}_{B}$ and $\Gamma^{\dot{A}}{}_{\dot{B}}$), the Bianchi identities can be shown to be equivalent to

$$SB: \{\nabla^{R\dot{S}} C_{AR\dot{B}\dot{S}} + \tfrac{1}{8} \nabla_{A\dot{B}} R = 0 \tag{1.15}$$

and

$$GB: \begin{cases} \text{(a) } \nabla^{S}{}_{\dot{D}} C_{ABCS} + \nabla_{(A}{}^{\dot{S}} C_{BC)\dot{D}\dot{S}} = 0, \\ \text{(b) } \nabla^{D\dot{S}} C_{A\dot{B}C\dot{S}} - \nabla_{S(A} C^{DS}{}_{\dot{B}C)} = 0. \end{cases} \tag{1.16}$$

(*SB* stands for special Bianchi identities, *GB* for the general ones.)

In complex relativity, the gauge group of the theory is taken to be

$$g := \mathrm{SL}(2,\mathbb{C}) \times \dot{\mathrm{SL}}(2,\mathbb{C}), \tag{1.17}$$

and consists of the two copies of the *independent* $\mathrm{SL}(2,\mathbb{C})$ transformations of the dotted and undotted spinorial indices, which is emphasized by the symbolism of (1.17). (See Ref. 10 for the most general gauge group.) With respect to these transformations described, say, by the matrices $(l^{A'}{}_{A})$ and $(l^{\dot{A}'}{}_{\dot{A}})$, the objects R, $g_{A\dot{B}}$, $C_{AB\dot{C}\dot{D}}$, C_{ABCD}, and $C_{\dot{A}\dot{B}\dot{C}\dot{D}}$ all transform tensorially, while the connection forms transform according to:

$$\begin{aligned} &\text{(a) } \Gamma^{A'}{}_{B'} := l^{A'}{}_{A} l^{-1B}{}_{B'} \Gamma^{A}{}_{B} + l^{A'}{}_{S}\, dl^{-1S}{}_{B'}, \\ &\text{(b) } \Gamma^{\dot{A}'}{}_{\dot{B}'} := l^{\dot{A}'}{}_{\dot{A}} l^{-1\dot{B}}{}_{\dot{B}'} \Gamma^{\dot{A}}{}_{\dot{B}} + l^{\dot{A}'}{}_{\dot{S}}\, dl^{-1\dot{S}}{}_{\dot{B}'}. \end{aligned} \tag{1.18}$$

We should like to state that the structure of a real Riemannian space emerges in a sense as the special case of the complex structure defined above; we obtain a real space–time of the signature $(+\,+\,+\,+\,-)$, if we take for M a real four-dimensional differential manifold and we postulate that $g_{A\dot{B}}$ are Hermitian objects, $(g_{A\dot{B}})^* = g_{B\dot{A}}$ with $*$ denoting complex conjugation. Then the $\dot{\mathrm{SL}}(2,\mathbb{C})$ transformations become reduced to the (simultaneous) complex conjugates of $\mathrm{SL}(2,\mathbb{C})$ transformations, $l^{\dot{A}'}{}_{\dot{A}} = (l^{A'}{}_{A})^*$, and all "dotted" relations become the complex conjugated copies of the undotted relations. [Notice that $(C_{ABCD})^* = C_{\dot{A}\dot{B}\dot{C}\dot{D}}$, $(C_{AB\dot{C}\dot{D}})^* = C_{CD\dot{A}\dot{B}}$, $R^* = R$.] The gauge group of the theory is then

$$\mathcal{G}_R = \mathrm{SL}(2,\mathbb{C})(2 \rightarrowtail 1)\,\mathrm{SO}'(3,1,\mathbb{R}), \tag{1.19}$$

i.e., is isomorphic to the special, arrow of time conserving (real), Lorentz group.

In complex relativity, however, it is essential to consider $\mathrm{SL}(2,\mathbb{C})$ and $\dot{\mathrm{SL}}(2,\mathbb{C})$ transformations as entirely independent. The prize for doing so is that we can conveniently use the properties which are the consequences of the direct product structure of the gauge group $\mathcal{G}$.

Now, in Ref. 2 following (in a sense) Newman's terminology and needing to introduce some nomenclature for the objects with pure undotted or dotted indices, it was proposed to call the geometric objects with respect to $\mathrm{SL}(2,\mathbb{C})$ and scalars with respect to $\dot{\mathrm{SL}}(2,\mathbb{C})$ (e.g., S_{AB}, Γ_{AB}, and C_{ABCD}), the "heavenly" objects. Corresponding the geometric objects with respect to $\dot{\mathrm{SL}}(2,\mathbb{C})$ and scalars with respect to $\mathrm{SL}(2,\mathbb{C})$ (e.g., $S_{\dot{A}\dot{B}}$, $\Gamma_{\dot{A}\dot{B}}$, and $C_{\dot{A}\dot{B}\dot{C}\dot{D}}$) would be called "hellish" (the printers can appreciate the motivation for this!). The mixed objects (e.g., R, $g_{A\dot{B}}$, and $C_{AB\dot{C}\dot{D}}$) are then "earthly"; it is also convenient to refer to the $\mathrm{SL}(2,\mathbb{C})$ and $\dot{\mathrm{SL}}(2,\mathbb{C})$ factors of $\mathcal{G}$ as the "heavenly" and "hellish" factors correspondingly.

The survey of the complex Riemannian structure given above indicates that the "heavenly" and "hellish" objects mix in this structure because of (i) the presence of the traceless part of the Ricci tensor, $C_{AB\dot{C}\dot{D}}$, and (ii) the "earthly" nature of the first structure equations, which serves as the definition of the connection forms in the terms of g's. On the other hand, the objects S_{AB} and $S_{\dot{A}\dot{B}}$ are purely "heavenly" and "hellish" from their nature, and the equations (1.6b) and (1.14b) do not involve any mixing of the "heavenly" and "hellish" objects.

In particular, if the Einstein equations are assumed, i.e., $C_{AB\dot{C}\dot{D}} = 0$ which implies via *SB* that $R = -4\lambda = \text{const}$, we have the *Einsteinian structure* V_4^E into which enter as the unknown objects

$$\begin{aligned} &g_{A\dot{B}} \sim 4 \cdot 4 = 16 \text{ functions}, \\ &\Gamma_{AB} \text{ and } \Gamma_{\dot{A}\dot{B}} \sim 3 \cdot 4 + 3 \cdot 4 = 24 \text{ functions}, \\ &C_{ABCD} \text{ and } C_{\dot{A}\dot{B}\dot{C}\dot{D}} \sim 5 + 5 = 10 \text{ functions}, \end{aligned} \tag{1.20}$$

i.e., altogether 50 functions to be determined. The equations which must be fulfilled by these functions are the first structure equations

$$dg^{A\dot{B}} = g^{A\dot{S}} \wedge \Gamma^{\dot{B}}{}_{\dot{S}} + g^{S\dot{B}} \wedge \Gamma^{A}{}_{S} \sim 4 \cdot 6 = 24 \text{ equations} \tag{1.21}$$

and the (second) Cartan structure equations with the built in Einstein equations, i.e.,

$$C: \quad \begin{aligned} &\text{(a) } d\Gamma^{A}{}_{B} + \Gamma^{A}{}_{S} \wedge \Gamma^{S}{}_{B} = -\tfrac{1}{2} C^{A}{}_{BCD} S^{CD} - \tfrac{1}{6}\lambda S^{A}{}_{B} \sim 3 \cdot 6 \text{ equations,} \\ &\text{(b) } d\Gamma^{\dot{A}}{}_{\dot{B}} + \Gamma^{\dot{A}}{}_{\dot{S}} \wedge \Gamma^{\dot{S}}{}_{\dot{B}} = -\tfrac{1}{2} C^{\dot{A}}{}_{\dot{B}\dot{C}\dot{D}} S^{\dot{C}\dot{D}} - \tfrac{1}{6} S^{\dot{A}}{}_{\dot{B}} \sim 3 \cdot 6 \text{ equations.} \end{aligned} \tag{1.22}$$

We have thus 60 equations for 50 unknown functions. In this structure, V_4^E, only the relations (1.21) imply the necessity of mixing the "heavenly" and the "hellish' objects.

All this leads in a natural manner to the though of searching for natural conditions which, assuming that the objects S_{AB} and $S_{\dot{A}\dot{B}}$ are given which fulfill (1.6b) and (1.14b), permit us to reconstruct the 1-forms $g_{A\dot{B}}$ which will obey (1.21). Having such conditions, one can hope that it will be possible to isolate within the studied structure the self-contained closed "heavenly" and "hellish" substructures and the "earthly" conditions, hopefully algebraic, which cause the interference of these substructures within the whole Riemannian structure.

Plausibly enough, we will see that such natural conditions consist in the already stated algebraic equations (1.6c), (1.14c), and (1.10).

II. AN EQUIVALENT FORM OF THE RIEMANNIAN STRUCTURE

Assume now that, ignoring $g_{A\dot{B}}$'s, we postulate over M the sets of relations $\mathcal{S}_H$, $\mathcal{E}$, and $\mathcal{S}_H^*$ which involve some $\Gamma^{A}{}_{B}$, $\Gamma^{\dot{A}}{}_{\dot{B}} \in \Lambda^1$ and S_{AB}, $S_{\dot{A}\dot{B}} \in \Lambda^2$. A so defined structure does not possess *a priori* a metric. Instead of the $4 \cdot 4 = 16$ functions $g_{A\dot{B}}$ of the standard Riemannian structure, we have here given as the fundamental S objects $2 \cdot 3 \cdot 6 = 36$ functions; these are, however, algebraically restricted by the conditions (1.6b), i.e., $6 - 1 = 5$ conditions, (1.10), i.e., $1 + 9 = 10$ conditions, and $6 - 1 = 5$ of the conditions (1.14c).[11] With the number of functions minus the number of conditions still being $16 = 36 - 20$, we can guess that the structure considered still contains $g_{A\dot{B}}$'s (the metric) in the disguise of S's. The aim of this section consists in providing a proof of this guess.

Consider first *prima faciae* the relations (1.6c). Specializing AB to 11 and 22, CD correspondingly to 22 and 11 we have

$$S^{11} \wedge S^{11} = 0, \quad S^{22} \wedge S^{22} = 0, \tag{2.1}$$

and consequently, there exist 1-forms e'^{a} such that

$$S^{11} = 2e'^{4} \wedge e'^{2}, \quad S^{22} = 2e'^{3} \wedge e'^{1}. \tag{2.2}$$

Then, taking $AB = 11 = CD$ (equivalently, $AB = 22 = CD$), we have $3S^{11} \wedge S^{22} = S^{RS} \wedge S_{RS} \neq 0$, so that

$$S^{RS} \wedge S_{RS} = -12 e'^{1} \wedge e'^{2} \wedge e'^{3} \wedge e'^{4} \neq 0, \tag{2.3}$$

and e'^{a} form a base of Λ^1. Then, taking $AB = 11$, $CD = 12$ and $AB = 22$, $CD = 12$ we obtain

$$e'^{4} \wedge e'^{2} \wedge S^{12} = 0, \quad e'^{3} \wedge e'^{1} \wedge S^{12} = 0, \tag{2.4}$$

and finally, $AB = 12 = CD$ yields

$$S^{12} \wedge S^{12} = -\tfrac{1}{6} S^{RS} \wedge S_{RS} = 2e'^{1} \wedge e'^{2} \wedge e'^{3} \wedge e'^{4}. \tag{2.5}$$

But from the simultaneous (2.4), remembering that e'^{a} form a base, it easily follows that

$$\begin{aligned} S^{12} &= (\alpha e'^{1} + \beta e'^{3}) \wedge e'^{2} + (\gamma e'^{1} + \delta e'^{3}) \wedge e'^{4} \\ &= e'^{1} \wedge (\alpha e'^{2} + \gamma e'^{4}) + e'^{3} \wedge (\beta e'^{2} + \delta e'^{4}), \end{aligned} \tag{2.6}$$

where α, β, γ, and δ are numbers; this fed into (2.5) gives

$$\alpha\delta - \beta\gamma = 1. \tag{2.7}$$

We can now define a tetrad

$$\overset{c}{e}{}^{a} \quad \text{("celestial")}$$

given by

$$\begin{aligned} &\overset{c}{e}{}^{1} := \alpha e'^{1} + \beta e'^{3}, \quad \overset{c}{e}{}^{2} := e'^{2}, \\ &\overset{c}{e}{}^{3} := \gamma e'^{1} + \delta e'^{3}, \quad \overset{c}{e}{}^{4} := e'^{4}. \end{aligned} \tag{2.8}$$

Because, however, according to (2.7)

$$\overset{c}{e}{}^{3} \wedge \overset{c}{e}{}^{1} = e'^{3} \wedge e'^{1},$$

we deduce from (1.6c) the existence of a base of

$$\Lambda^1, \quad \overset{c}{e}{}^{a},$$

such that

$$S^{11} = 2\overset{c}{e}{}^{4} \wedge \overset{c}{e}{}^{2}, \quad S^{12} = \overset{c}{e}{}^{1} \wedge \overset{c}{e}{}^{2} + \overset{c}{e}{}^{3} \wedge \overset{c}{e}{}^{4}, \quad S^{22} = 2\overset{c}{e}{}^{3} \wedge \overset{c}{e}{}^{1}. \tag{2.9}$$

By defining now

$$(\overset{c}{g}{}^{A\dot{B}}) := \sqrt{2} \begin{pmatrix} \overset{c}{e}{}^{4} & \overset{c}{e}{}^{1} \\ \overset{c}{e}{}^{2} & -\overset{c}{e}{}^{3} \end{pmatrix}, \tag{2.10}$$

we can gather the same more compactly

$$S^{AB} = \tfrac{1}{2}\epsilon_{\dot{R}\dot{S}} \overset{c}{g}{}^{A\dot{R}} \wedge \overset{c}{g}{}^{B\dot{S}}. \tag{2.11}$$

This form makes manifest that, given S_{AB} obeying (1.6c), the corresponding base $\overset{c}{g}{}_{A\dot{B}}$ is given only modulo the (arbitrary) "hellish" gauge

$$\overset{c}{g}{}^{A\dot{B}} \to \overset{c}{g}{}^{A\dot{S}} l^{\dot{B}}{}_{\dot{S}}, \quad \det(l^{\dot{A}}{}_{\dot{B}}) = 1. \tag{2.12}$$

Knowing $\overset{c}{g}{}_{A\dot{B}}$ with this precision, we can, however, define the "celestial" ("left") metric in $\Lambda^1 \underset{s}{\otimes} \Lambda^1$,

$$\overset{c}{g} := -\tfrac{1}{2} \overset{c}{g}{}_{A\dot{B}} \underset{s}{\otimes} \overset{c}{g}{}^{A\dot{B}}, \tag{2.13}$$

which, of course, does not depend on any choice for either gauge. Thus, the equations $\mathcal{S}_H$, (1.6) do imply the unique existence of a metric, $\overset{c}{g}$. The catch is, however, that $\mathcal{S}_H$, understood as a substructure on M, does not contain the concept of $\Gamma^{\dot{A}}{}_{\dot{B}}$! Consequently, the covariant derivative of $\overset{c}{g}$ remains yet undefined and $\mathcal{S}_H$ *is not* equivalent to a Riemannian structure.

It is now quite clear that by applying a parallel proof we can claim that, given $S_{\dot{A}\dot{B}}$ submitted to the condition (1.14c), there exists an "infernal" base of Λ^1, $\overset{i}{g}{}_{A\dot{B}}$, such that

$$S^{\dot{A}\dot{B}} = \tfrac{1}{2}\epsilon_{RS} \overset{i}{g}{}^{\dot{R}A} \wedge \overset{i}{g}{}^{\dot{S}B}, \tag{2.14}$$

this base being defined by $S_{\dot{A}\dot{B}}$ with the precision only up to the (arbitrary) "heavenly" gauge,

$$\overset{i}{g}{}^{A\dot{B}} \to l^{A}{}_{S} \overset{i}{g}{}^{S\dot{B}}, \quad \det(l^{A}{}_{B}) = 1. \tag{2.15}$$

This base defines now (uniquely!) the "infernal" ("right") metric,

$$\overset{i}{g} := -\tfrac{1}{2}\overset{i}{g}_{A\dot{B}} \underset{S}{\otimes} \overset{i}{g}^{A\dot{B}}. \tag{2.16}$$

We would now like to explore, assuming (1.6c) and (1.14c), the consequences of the set of relations $\mathcal{E}$, (1.10). Let

$$\overset{c}{S}^{\dot{A}\dot{B}} := \tfrac{1}{2}\epsilon_{RS}\overset{c}{g}^{\dot{R}\dot{A}} \wedge \overset{c}{g}^{\dot{S}\dot{B}}. \tag{2.17}$$

It is clear that the pair $(S_{AB}, \overset{c}{S}_{\dot{A}\dot{B}})$: (i) is a base of Λ^1, (ii) fulfills all algebraic relations involving

$$S_{AB} \text{ and } S_{\dot{A}\dot{B}}, \quad \text{with } S_{\dot{A}\dot{B}} \to \overset{c}{S}_{\dot{A}\dot{B}}.$$

From this, and the orthogonality condition (1.10b) we infer that there must exist $M^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}} = M^{(\dot{A}\dot{B})}{}_{\dot{C}\dot{D}} = M^{\dot{A}\dot{B}}{}_{(\dot{C}\dot{D})} \in \Lambda^0$ such that

$$\overset{c}{S}^{\dot{A}\dot{B}} = M^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}}\overset{c}{S}^{\dot{C}\dot{D}}. \tag{2.18}$$

This fed into (1.14c) easily gives the condition

$$3M^{\dot{A}\dot{B}}{}_{\dot{R}\dot{S}}M^{\dot{C}\dot{D}}{}_{\dot{U}\dot{V}} = \delta^{\dot{A}}{}_{(\dot{C}}\delta^{\dot{B}}{}_{\dot{D})}M^{\dot{R}\dot{S}}{}_{\dot{U}\dot{V}}M^{\dot{C}\dot{D}}{}_{\dot{R}\dot{S}}. \tag{2.19}$$

On the other hand, (2.18) substituted into (1.10a) yields

$$\tfrac{1}{3}M^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}}M^{\dot{C}\dot{D}}{}_{\dot{A}\dot{B}} = 1. \tag{2.20}$$

Therefore, $\mathcal{E}$ equations are fulfilled by (2.18) if the coefficients $M^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}}$ satisfy the relations

$$M^{\dot{A}\dot{B}}{}_{\dot{R}\dot{S}}M^{\dot{C}\dot{D}}{}_{\dot{R}\dot{S}} = \delta^{\dot{A}}{}_{(\dot{C}}\delta^{\dot{B}}{}_{\dot{D})}. \tag{2.21}$$

This means, however, that $Y^{\dot{A}\dot{B}} = M^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}}X^{\dot{C}\dot{D}}$, understood as transformations of a three-dimensional space of spinors with a symmetric pair of indices, conserve the natural scalar product in that space, $A^{\dot{R}\dot{S}}B_{\dot{R}\dot{S}}$. Therefore, as such, these transformations must overlap with the (irreducible) $D(0,1)$ transformations which possess the general form of

$$M^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}} = l^{(\dot{A}}{}_{(\dot{C}}l^{\dot{B})}{}_{\dot{D})}, \quad \det(l^{\dot{A}}{}_{\dot{B}}) = 1. \tag{2.22}$$

Therefore, we infer as the consequence of $\mathcal{E}$ equations, with (1.6c) and (1.14c) assumed, that there exists a matrix $(l^{\dot{A}}{}_{\dot{B}}) \in \mathrm{SL}(2,\mathbb{C})$ such that

$$\tfrac{1}{2}\epsilon_{RS}\overset{i}{g}^{\dot{R}\dot{A}} \wedge \overset{i}{g}^{\dot{S}\dot{B}} = l^{\dot{A}}{}_{\dot{P}}l^{\dot{B}}{}_{\dot{Q}}\tfrac{1}{2}\epsilon_{RS}\overset{c}{g}^{\dot{R}\dot{P}} \wedge \overset{c}{g}^{\dot{S}\dot{Q}}. \tag{2.23}$$

It is now quite obvious that by a parallel deductive process, with the roles of the dotted and undotted objects inversed, we can also demonstrate the existence of a $\mathrm{SL}(2,\mathbb{C})$ matrix, $(l^{A}{}_{B})$ such that

$$l^{A}{}_{P}l^{B}{}_{Q}\tfrac{1}{2}\epsilon_{\dot{R}\dot{S}}\overset{i}{g}^{\dot{P}\dot{R}} \wedge \overset{i}{g}^{\dot{Q}\dot{S}} = \tfrac{1}{2}\epsilon_{\dot{R}\dot{S}}\overset{c}{g}^{\dot{A}\dot{R}} \wedge \overset{c}{g}^{\dot{B}\dot{S}}, \tag{2.24}$$

The equations (2.23) and (2.24) imply that by using the remaining freedom of the "hellish" gauge for $\overset{c}{g}_{A\dot{B}}$ and the "heavenly" gauge for $\overset{i}{g}_{A\dot{B}}$ [i.e., (2.12) and (2.15) transformations], we can arrive at such a choice for $\overset{c}{g}_{A\dot{B}}$ and $\overset{i}{g}_{A\dot{B}}$ that

$$S^{AB} = \overset{i}{S}^{AB}, \quad S^{\dot{A}\dot{B}} = \overset{c}{S}^{AB}, \tag{2.25}$$

in the self-evident notation. From this already it follows that in these particular gauges

$$\overset{c}{g}_{A\dot{B}} = \pm\overset{i}{g}_{A\dot{B}};$$

the minus sign can be still absorbed by a transformation $(l^{A}{}_{B}) = (-\delta^{A}{}_{B})$ which does not affect anything in (2.25). Therefore, we arrive at the conclusion that with $\mathcal{E}$ equations assumed, one can always so select

$$\overset{c}{g}_{A\dot{B}} \text{ and } \overset{i}{g}_{A\dot{B}},$$

such that they are so synchronized that

$$\overset{c}{g}_{A\dot{B}} = \overset{i}{g}_{A\dot{B}} =: g_{A\dot{B}} \to \overset{c}{g} = \overset{i}{g} =: g. \tag{2.26}$$

The concept of the common metric, g, does not depend then upon the choice for the corresponding gauges, and its existence is a direct consequence of the $\mathcal{E}$ equations.

We can summarize the result derived as follows: Postulating that there are given objects S_{AB} and $S_{\dot{A}\dot{B}}$ which fulfill the "heavenly" algebraic condition (1.6c), the "hellish" algebraic conditions (1.14c), and the mixed "earthly" conditions (1.10), there exists then a base of Λ^1, $g_{A\dot{B}}$, such that

$$g^{A\dot{B}} \wedge g^{C\dot{D}} = \epsilon^{AC}S^{\dot{B}\dot{D}} + \epsilon^{\dot{B}\dot{C}}S^{AC}. \tag{2.27}$$

The S objects define then, via this base, the metric g in a unique manner.

After reconstruction from the relations contained in the structure $\mathcal{S}_{\mathcal{H}} \cup \mathcal{E} \cup \mathcal{S}_{\mathcal{H}}$, the relation (2.27), and the concept of the metric, consider now the object

$$\Lambda^2 \ni X^{A\dot{B}} := dg^{A\dot{B}} - g^{A\dot{S}} \wedge \Gamma^{\dot{B}}{}_{\dot{S}} - g^{\dot{S}\dot{B}} \wedge \Gamma^{A}{}_{S}, \tag{2.28}$$

Executing now the external differential of (2.27), using for $dg^{A\dot{B}}$ (2.28), for dS^{AB} and $dS^{\dot{A}\dot{B}}$, (1.6b) and (1.14b), respectively, and applying for the external products of two g's again (2.27), one obtains after cancellations the condition

$$g^{A\dot{B}} \wedge X^{C\dot{D}} - g^{C\dot{D}} \wedge X^{A\dot{B}} = 0. \tag{2.29}$$

Investigating this condition, we can use an elementary lemma (a simple proof is omitted for brevity): $\alpha \in \Lambda^p$, $p = 1, 2, 3$ and $\alpha \wedge g^{A\dot{B}} = 0 \to \alpha = 0$, if $g_{A\dot{B}}$ is a base of Λ^1. Because (2.29) implies $g^{A\dot{B}} \wedge g^{C\dot{D}} \wedge X^{C\dot{D}} = 0$ (with *no* summation over C and D), applying the lemma, $g^{A\dot{B}} \wedge X^{A\dot{B}} = 0$, and consequently, there are such $\alpha, \beta, \gamma, \delta \in \Lambda^1$, that

$$\begin{aligned} X^{1\dot{1}} &= \alpha \wedge g^{1\dot{1}}, \quad X^{2\dot{2}} = \beta \wedge g^{2\dot{2}}, \\ X^{1\dot{2}} &= \gamma \wedge g^{1\dot{2}}, \quad X^{2\dot{1}} = \delta \wedge g^{2\dot{1}}. \end{aligned} \tag{2.30}$$

This substituted into six independent conditions (2.29) gives:

$$\begin{aligned} &\text{(a) } (\alpha+\beta)\wedge g^{1\dot{1}} \wedge g^{2\dot{2}} = 0, \quad \text{(d) } (\beta+\gamma) \wedge g^{2\dot{2}} \wedge g^{1\dot{2}} = 0, \\ &\text{(b) } (\alpha+\gamma)\wedge g^{1\dot{1}} \wedge g^{1\dot{2}} = 0, \quad \text{(e) } (\beta+\delta) \wedge g^{2\dot{2}} \wedge g^{2\dot{1}} = 0, \\ &\text{(c) } (\alpha+\delta)\wedge g^{1\dot{1}} \wedge g^{2\dot{1}} = 0, \quad \text{(f) } (\gamma+\delta) \wedge g^{1\dot{2}} \wedge g^{2\dot{1}} = 0. \end{aligned} \tag{2.31}$$

Four of these equations, e.g., (a), (b), (d) and (e), can be solved for α, β, γ, and δ given as linear combinations of g's with some coefficients. This solution fed into the remaining two equations, e.g., (c) and (f), forces α, β, γ, and δ to be proportional to $g^{1\dot{1}}$, $g^{2\dot{2}}$, $g^{1\dot{2}}$, and $g^{2\dot{1}}$, so that all quantitites (2.30) vanish. Therefore, we have a useful lemma; if $g_{A\dot{B}}$ forms a base of Λ^1 and $X^{A\dot{B}} \in \Lambda^2$, then

$$X^{A\dot{B}} \wedge g^{C\dot{D}} - X^{C\dot{D}} \wedge g^{A\dot{B}} = 0 \to X^{A\dot{B}} = 0. \tag{2.32}$$

Applying this lemma, we see from (2.28) that the struc-

ture $\mathcal{S}_H \cup \mathcal{E} \cup \mathcal{S}_{\dot{H}}$ implies the relation (1.5). Therefore *all* crucial relations of the Riemannian structure V_4 which are missing in $\mathcal{S}_H \cup \mathcal{E} \cup \mathcal{S}_{\dot{H}}$ can be deductively derived from the last structure. Thus, both structures are equivalent, $V_4 \equiv \mathcal{S}_H \cup \mathcal{E} \cup \mathcal{S}_{\dot{H}}$. This equivalence (1) exhibits the symmetric role of $\mathcal{S}_H$ and $\mathcal{S}_{\dot{H}}$ in V_4 and (2) emphasises the importance of the *algebraic* mixing conditions $\mathcal{E}$. For clarity of the interpretation of this result, it is worthwhile to remember, however, that postulated $\mathcal{S}_H$ structure; there exists (an obvious) "canonical" manner of completing it to the whole V_4.

Simply, given $\overset{c}{g}_{A\dot{B}}$ (modulo the hellish gauge), one defines

$$S^{\dot{A}\dot{B}} =: \tfrac{1}{2}\epsilon_{RS}\overset{c}{g}{}^{\dot{R}}{}_{\dot{A}} \wedge \overset{c}{g}{}^{\dot{S}}{}_{\dot{B}},$$

and the connection $\Gamma^{\dot{A}\dot{B}}_{\dot{C}}$ one understands as read off from the relation $dS^{\dot{A}\dot{B}} + 3S^{(\dot{A}\dot{B}} \wedge \Gamma^{\dot{C})}{}_{\dot{C}} = 0$. Then, manifestly, all $\mathcal{E}$ relations and $\mathcal{S}_{\dot{H}}$ structure equations became identities. In particular, it is of course obvious that the metric $\overset{c}{g}$ obtained in the result of solving the $\mathcal{S}^E_H$ structure with $C_{AB\dot{C}\dot{D}} = 0$, $R = -4\lambda$ is then a solution of Einstein equations with λ. (A discussion of this point with Robinson is appreciated.) Behind this particular implication of the "canonical" extension of the $\mathcal{S}_H$ structure, there is of course hidden the fact that the unique, torsionless Levi-Civita's connection, such that $T_{A\dot{B}} = dg^{A\dot{B}} + \Gamma^{A}{}_{S} \wedge g^{S\dot{B}} + \Gamma^{\dot{B}}{}_{\dot{S}} \wedge g^{A\dot{S}} = 0$, forces the functions R and $C_{AB\dot{C}\dot{D}}$ which appear in $\mathcal{S}_H$ and $\mathcal{S}_{\dot{H}}$ structures to be the same objects.

Although $\mathcal{S}_H \cup (T_{A\dot{B}} = 0) \leftrightarrow V_4$, in this paper we precisely explore the consequences *of not assuming* $T_{A\dot{B}} = 0$ (i.e., first structure equations) as a direct independent postulate. In particular, one sees that $dS^{\dot{A}\dot{B}} + 3S^{(\dot{A}\dot{B}} \wedge \Gamma^{\dot{C})}{}_{\dot{C}} = -g^{S(\dot{A}} \wedge T_S{}^{\dot{B})} \equiv$ "right semitorsion," can remain *arbitrary* when $dS^{AB} + 3S^{(AB} \wedge \Gamma^{C)}{}_{C} = g^{(A}{}_{\dot{S}} \wedge T^{B)\dot{S}} \equiv$ "left-semitorsion," is assumed to vanish. This is precisely the situation within the $\mathcal{S}_H$ structure without any other additional postulates.

Our equivalence $V_4 \equiv \mathcal{S}_H \cup \mathcal{E} \cup \mathcal{S}_{\dot{H}}$ thus rederives $T_{A\dot{B}} = 0$ from the vanishing of "semitorsions" from both sides and by making both sides (uniquely) compatible by $\mathcal{E}$ conditions. When $C_{AB\dot{C}\dot{D}} = 0$, $R = -4\lambda$ and the equations of $\mathcal{S}_H$ *close* involving only objects Γ_{AB} and S_{AB}, then nothing forces us *a priori* to make *any* assumptions concerning $\Gamma^{\dot{A}\dot{B}}_{\dot{C}}$ (equivalently, right semitorsion) which does not enter in the closed $\mathcal{S}^E_H$ structure. These closed structures shall be discussed in the next section.

III. CLOSED SUBSTRUCTURES OF THE EINSTEINIAN STRUCTURE

The results of the last section become particularly interesting in the case of the Einsteinian structure V^E_4, where with $C_{AB\dot{C}\dot{D}} = 0$ (and hence via SB, $R = -4\lambda = \text{const}$) there vanishes the last object which if not equal to zero implies a necessary mixing of pure "heavenly" and "hellish" objects within the sets of equations $\mathcal{S}_H$ and $\mathcal{S}_{\dot{H}}$. We arrive thus at the following description of V^E_4: First, we have the set of equations

$$\mathcal{S}^E_H: \begin{cases} \text{(a) } d\Gamma^{A}{}_{B} + \Gamma^{A}{}_{S} \wedge \Gamma^{S}{}_{B} = -\tfrac{1}{2}C^{A}{}_{BCD}S^{CD} - \dfrac{\lambda}{6}S^{A}{}_{B}, \\ \text{(b) } dS^{AB} + 3S^{(AB} \wedge \Gamma^{C)}{}_{C} = 0, \\ \text{(c) } 3S^{AB} \wedge S_{CD} = \delta^{A}{}_{(C}\delta^{B}{}_{D)}S^{RS} \wedge S_{RS} \neq 0, \end{cases} \tag{3.1}$$

which involves the pure "heavenly" (self-dual) objects. These are:

$$\begin{aligned} &\text{(a) } 3 \cdot 6 = 18 \text{ functions in objects } S_{AB} \text{ (of helicity } +\hbar) \\ &\text{(b) } 3 \cdot 4 = 12 \text{ functions in left connection } \Gamma_{AB} = \Gamma_{(AB)}, \\ &\text{(c) } 5 \text{ functions in } C_{ABCD} \text{ (of helicity } +2\hbar). \end{aligned} \tag{3.2}$$

It is now clear that the equations $\mathcal{S}^E_H$ constitute exactly $3 \cdot 6 + 3 \cdot 4 + (6-1) = 35$ equations for these 35 objects; notice that while C_{ABCD} enter only algebraically in $\mathcal{S}^E_H$, and are five in number, the pure algebraic conditions impose also exactly five constraints on $\mathcal{S}_{AB}$'s. Therefore, the set of equations $\mathcal{S}^E_H$ constitutes a *closed substructure* on M which involves only the objects of *positive helicity*.

Within the substructure $\mathcal{S}^E_H$ we *do have a metric*; according to the results of the previous section, (3.1c) implies the existence of a base in Λ^1, $\overset{c}{g}_{A\dot{B}}$, determined modulo a $\mathrm{SL}(2,\mathbb{C})$ gauge such that

$$S^{AB} = \tfrac{1}{2}\epsilon_{\dot{R}\dot{S}}\overset{c}{g}{}^{A\dot{R}} \wedge \overset{c}{g}{}^{B\dot{S}}, \tag{3.3}$$

and this base defines a (unique!) metric,

$$\overset{c}{g} := -\tfrac{1}{2}\overset{c}{g}_{A\dot{B}} \underset{S}{\otimes} \overset{c}{g}{}^{A\dot{B}}. \tag{3.4}$$

This metric—as such—can be used, e.g., to define the concept of the corresponding Hodge star, $\overset{c}{*}$, within the substructure $\mathcal{S}^E_H$. It is not, however, covariantly constant within $\mathcal{S}^E_H$! The substructure $\mathcal{S}^E_H$ *does not* involve the connection $\Gamma^{\dot{A}}{}_{\dot{B}}$! Thus, the concept of parallel transfer is meaningful within $\mathcal{S}^E_H$ only for the objects with positive helicity (pure "heavenly," self-dual objects). E.g., we can define as the covariant differential of C_{ABCD},

$$DC_{ABCD} := dC_{ABCD} - 4C_{S(ABC}\Gamma^{S}{}_{D)}, \tag{3.5}$$

and verify that this is a $\mathrm{SL}(2,\mathbb{C})$ tensor. Notice that within $\mathcal{S}^E_H$, Eq. (3.3) can be treated as giving the representation of the fundamental S_{AB} through "an algebraic potential" $\overset{c}{g}_{A\dot{B}}$ defined up to $\mathrm{SL}(2,\mathbb{C})$ gauge. Condition (3.1c) is then the necessary and sufficient condition for the existence of that potential.

We shall add that the integrability conditions of the substructure $\mathcal{S}^E_H$ which follow from $d^2 = 0$ can be easily seen to amount to

$$\partial\mathcal{S}^E_H : \{\text{(a) } DC_{ABCD} \wedge S^{CD} = 0, \quad \text{(b) } C_{ABCD} = C_{(ABCD)}. \tag{3.6}$$

Of course, (3.6a) are just the "heavenly" GB identities (1.16a); with $\nabla_{A\dot{B}}$ well defined with respect to $D(k,0)$ objects, they can be also stated in the form of $\nabla^{S}{}_{\dot{D}}C_{ABCS} = 0$. As far as (3.6b) is concerned, only easily finds that this condition is necessary and sufficient within $\mathcal{S}^E_H$ to guarantee that d applied on (3.1b) amounts to $0 = 0$.

Closing this brief description of $\mathcal{S}^E_H$, we should like to add that as it has been discussed in the Introduction, the subspace of Λ^2, $D(\Lambda^2)$ spanned by S_{AB} can be defined *without* the use of metrical concepts, just by the application of the natural decomposition of Λ^2 into a direct sum of the three-dimensional wedge orthogonal subspaces; if we then introduce a fibration of $D(\Lambda^2)$ by identifying second order frames S_{AB} by means of the $\mathrm{SL}(2,\mathbb{C})$ trans-

formations, we obtain a bundle (over M), whose semi-metrical structure can be thought of as our $\mathcal{S}_{\mathcal{H}}^{E}$ substructure. It is essential, however to remember that the metric $\overset{c}{g}$ does fulfill the Einstein equations formulated, e.g., in the traditional local manner, $G_{\mu\nu}(\overset{c}{g}_{\rho\sigma})=\lambda \overset{c}{g}_{\mu\nu}$, i.e., it assures $C_{AB\dot{C}\dot{D}}=0$, $R=-4\lambda$.

A second part of the V_4^E structure is constituted by the substructure $\mathcal{S}_{\mathcal{H}}^{E}$ which can be looked upon as just another copy of the $\mathcal{S}_{\mathcal{H}}^{E}$ substructure. (In the special case of the real V_4^E, the $\mathcal{S}_{\mathcal{H}}^{E}$ equations are just complex conjugates of those from $\mathcal{S}_{\mathcal{H}}^{E}$.) The $\mathcal{S}_{\mathcal{H}}^{E}$ equations are thus:

$$\mathcal{S}_{\mathcal{H}}^{E}:\begin{cases}\text{(a) } d\Gamma^{\dot{A}}{}_{\dot{B}}+\Gamma^{\dot{A}}{}_{\dot{S}}\wedge\Gamma^{\dot{S}}{}_{\dot{B}}=-\frac{1}{2}C^{\dot{A}}{}_{\dot{B}\dot{C}\dot{D}}S^{\dot{C}\dot{D}}-\frac{\lambda}{6}S^{\dot{A}}{}_{\dot{B}},\\ \text{(b) } dS^{\dot{A}\dot{B}}+3S^{(\dot{A}\dot{B}}\wedge\Gamma^{\dot{C})}{}_{\dot{C}}=0,\\ \text{(c) } 3S^{\dot{A}\dot{B}}\wedge S_{\dot{C}\dot{D}}=\delta^{\dot{A}}_{(\dot{C}}\delta^{\dot{B}}_{\dot{D})}S^{\dot{R}\dot{S}}\wedge S_{\dot{R}\dot{S}}.\end{cases}\tag{3.7}$$

Again we have here, of course, a closed substructure, which involves 18 functions in $S_{\dot{A}\dot{B}}$ (of helicity "$-\hbar$"), 12 functions in the right connection $\Gamma_{\dot{A}\dot{B}}=\Gamma_{(\dot{A}\dot{B})}$, and five $C_{\dot{A}\dot{B}\dot{C}\dot{D}}$'s (of the helicity "$-2\hbar$"). These functions must fulfill 35 equations (3.7). Equations (3.7c) are again the "algebraic integrability conditions" for the existence of the "potentials" $\overset{i}{g}_{\dot{A}\dot{B}}\in\Lambda^1$ [given up to the SL(2, $\mathbb{C}$) gauge] such that

$$S^{\dot{A}\dot{B}}=\tfrac{1}{2}\epsilon_{\dot{R}\dot{S}}\overset{i}{g}{}^{\dot{R}\dot{A}}\wedge\overset{i}{g}{}^{\dot{S}\dot{B}}.\tag{3.8}$$

These "potentials" then define a (unique!) metric

$$\overset{i}{g}:=-\tfrac{1}{2}g_{\dot{A}\dot{B}}\underset{S}{\otimes}g^{\dot{A}\dot{B}}.\tag{3.9}$$

The last metric permits us to define Hodge's $\overset{i}{*}$ $\mathcal{S}_{\mathcal{H}}^{E}$. Again, inside $\mathcal{S}_{\mathcal{H}}^{E}$, the concept of D, and hence of the parallel transport, is well defined only for the purely "hellish" (anti-self-dual) $D(0,l)$ objects, e.g.,

$$DC_{\dot{A}\dot{B}\dot{C}\dot{D}}:=dC_{\dot{A}\dot{B}\dot{C}\dot{D}}-4C_{\dot{S}(\dot{A}\dot{B}\dot{C}}\Gamma^{\dot{S}}{}_{\dot{D})},\tag{3.10}$$

is a $\mathrm{S\dot{L}}(2,\mathbb{C})$ tensor. The integrability conditions of (3.7) are of course:

$$\partial\mathcal{S}_{\mathcal{H}}^{E}:\{\text{(a) } DC_{\dot{A}\dot{B}\dot{C}\dot{D}}\wedge S^{\dot{C}\dot{D}}=0,\quad \text{(b) } C_{\dot{A}\dot{B}\dot{C}\dot{D}}=C_{(\dot{A}\dot{B}\dot{C}\dot{D})},\tag{3.11}$$

and (3.11a) are just "hellish" GB identities, (1.16b), i.e., $\nabla_{\dot{D}}C_{\dot{S}\dot{A}\dot{B}\dot{C}}=0$, with $\nabla_{\dot{A}\dot{B}}$ being defined in $\mathcal{S}_{\mathcal{H}}^{E}$ only for $D(0,l)$ objects.

The crucial part of the V_4^E structure consists of the "earthly" algebraic conditions, which are

$$\mathcal{E}:\{\text{(a) } \mathcal{S}^{RS}\wedge S_{RS}+S^{\dot{R}\dot{S}}\wedge S_{\dot{R}\dot{S}}=0,\quad \text{(b) } S_{AB}\wedge S_{\dot{C}\dot{D}}=0.\tag{3.12}$$

These are the conditions which force that $D(\Lambda^2)$ and $\dot{D}(\Lambda^2)$ add up to the whole Λ^2, the metrics $\overset{c}{g}$ and $\overset{i}{g}$ become identical,

$$\overset{c}{g}=g=\overset{i}{g},\tag{3.13}$$

and, in the structure $V_4^E\equiv\mathcal{S}_{\mathcal{H}}^{E}\cup\mathcal{E}\cup\mathcal{S}_{\mathcal{H}}^{E}$, which has both types of connections, this universal metric (i) is covariantly constants, and (ii) fulfills the Einstein empty space equations (with λ). The conditions $\mathcal{E}$, which are the necessary and sufficient conditions for the synchronization of $\overset{c}{g}_{\dot{A}\dot{B}}$ with $\overset{i}{g}_{\dot{A}\dot{B}}$ by the means of the "right" and "left" gauges respectively, permit us to construct the universal "algebraic potentials" for both substructures $\mathcal{S}_{\mathcal{H}}^{E}$ and $\mathcal{S}_{\mathcal{H}}^{E}$ simultaneously and therefore to construct relations (1.4) and (1.5) basic in the standard description of the Riemannian structure. Summing up: In the present treatment of the Einstein equations, instead of working with $g_{\dot{A}\dot{B}}$'s, we work with S_{AB} and $S_{\dot{A}\dot{B}}$ increasing the number of functions sought from 50 to 70 [compare (1.20) and (1.21), (1.22)], which is compensated by the algebraic conditions on these functions, 5 in $\mathcal{S}_{\mathcal{H}}^{E}$, 5 in $\mathcal{S}_{\mathcal{H}}^{E}$, and 10 in $\mathcal{E}$, respectively. The prize for doing so is that $\mathcal{S}_{\mathcal{H}}^{E}$ and $\mathcal{S}_{\mathcal{H}}^{E}$ becomes autonomous *closed* substructures. At the same time, equations $\mathcal{S}_{\mathcal{H}}^{E}\cup\mathcal{E}\cup\mathcal{S}_{\mathcal{H}}^{E}$ retain the basic shape of the Cartan structure equations with the built in Einstein equations, i.e., (i) the derivatives enter in these relations linearly, (ii) the remaining algebraic terms are all at most quadratic from the point of view of the degree of the nonlinearity. (This last property causes the $\mathcal{S}_{\mathcal{H}}^{E}\cup\mathcal{E}\cup\mathcal{S}_{\mathcal{H}}^{E}$ equations to form a convenient starting point in developing an invariant approximation procedure for the Einstein equations; we intend to study this point with Hacyan.)

For the benefit of the physicists which still prefer to think in terms of the local components, it is perhaps advisable to describe the closed character of $\mathcal{S}_{\mathcal{H}}^{E}$ (correspondingly, $\mathcal{S}_{\mathcal{H}}^{E}$) in a local coordinate patch $\{x^{\mu}\}$, where

$$S_{AB}=\tfrac{1}{2}S_{AB\mu\nu}dx^{\mu}\wedge dx^{\nu},\quad S_{AB\mu\nu}=\begin{cases}S_{(AB)\mu\nu}\\ S_{AB[\mu\nu]}\end{cases}\sim 3\cdot 6\text{ functions},$$

$$\Gamma_{AB}=\Gamma_{AB\mu}\,dx^{\mu},\qquad \Gamma_{AB\mu}=\Gamma_{(AB)\mu}\sim 3\cdot 4\text{ functions}.\tag{3.14}$$

Then $\mathcal{S}_{\mathcal{H}}^{E}$ equations can be stated in the form of

$$\mathcal{S}_{\mathcal{H}}^{E}:\begin{cases}\text{(a) } \Gamma^{A}{}_{B[\mu,\nu]}+\Gamma^{A}{}_{S[\mu}\Gamma^{S}{}_{B|\nu]}=-\frac{1}{2}C^{A}{}_{BCD}S^{CD}{}_{\mu\nu}\\ \qquad\qquad -\frac{\lambda}{12}S^{A}{}_{B\mu\nu},\\ \text{(b) } S^{AB}{}_{[\mu\nu,\lambda]}+3S^{(AB}{}_{[\mu\nu}\Gamma^{C)}{}_{|C|\nu]}=0,\\ \text{(c) } \epsilon^{\alpha\beta\gamma\delta}S_{AB\alpha\beta}S^{CD}_{\gamma\delta}=\rho\delta^{C}_{(A}\delta^{D}_{B)}\\ \text{(with } \rho\neq 0 \text{ to be determined).}\end{cases}\tag{3.15}$$

These equations are of course *tensorial* with respect to both the group of the coordinate transformations and the heavenly gauge SL(2, $\mathbb{C}$). In a sense they are equivalent to about one-half of the Einstein structure involving the metric and the complete connection. As far as the solution of Einstein equations is concerned, $\mathcal{S}_{\mathcal{H}}$ contains of course *all* information needed.

All this leads to the idea that the (partial) separation of the structure V_4^E, described in this section, into substructures which are purely "heavenly" or "hellish" (i.e., involve objects with the positive or negative helicity; partial, because $\mathcal{E}$ equations should be still satisfied!), should be capitalized in the search for the effective solutions of the Einstein whole structure; one can proceed as follows: First study separately the "semi-Einsteinian" structure $\mathcal{S}_{\mathcal{H}}^{E}$ and find the general integral variety of $S_{AB\mu\nu}$, $\Gamma_{AB\mu}$, accompanied by the corresponding C_{ABCD}'s. Because $\mathcal{S}_{\mathcal{H}}^{E}$ contains the metric $\overset{c}{g}$ which fulfills the Einstein equations, it is interest to notice that this study will be already equivalent to the study at the integral variety of the Einstein equations.

Another copy of this solution can then serve as the general integral variety of $\mathcal{S}_H^E$. Searching for the integral variety of the whole strucutre, we can ignore for a time $\mathcal{E}$ equations, which take the weight of an important part of the nonlinearity of V_4^E structure. Having, however, the integral variety for $\mathcal{S}_H^E$ and its copy for $\mathcal{S}_{\hat{H}}^E$, and substituting them into the mixing (algebraic) $\mathcal{E}$ conditions, one can determine the restrictions which will guarantee that we obtained by this process a solution of the whole Einstein structure. Of course, having a solution of $\mathcal{S}_H^E$ structure, one can also directly deduce the corresponding $\mathcal{S}_H^\mathcal{E}$ structure and $\mathcal{E}$ conditions by the "canonical extension" discussed in the end of the last section. The outlined above program of studying how $\mathcal{E}$ conditions fulfill the role of pikcing up the correct intersection of the two copies of the integral variety of the $\mathcal{S}_H^E$ structure, can however give us a better insight in the nature of the whole Einsteinian structure studied. We can add that, the independent dynamical nature of the $\mathcal{S}_H^E$ and $\mathcal{S}_{\hat{H}}^E$ structures in also emphasized by the existence of the variational principles which lead to these structures; this will be discussed in the next section.

IV. SEMIGRAVITONS

The improvement of methods and progress in results of modern mathematical physics cause many practitioners of relativity to feel that the ideal goal of the theory of the Einstein equations, the construction of their integral variety by analytic means (the Wheeler dream!), is not as distant as it would seem on the basis of the experiences which accompany the (normally very difficult) birth of new significant exact solutions. (The Tomimatsu–Sato solutions[12]—see also Refs. 13 and 14—are good examples of solutions of this type.)

Recently, high hopes were aroused by progress with complex methods in general relativity. "Heavens," which were first encountered as a space of "good cones" in the study of the complexified asymptotics of gravitational radiation by Newman (see Refs. 7–9), were given a bold *a priori* interpretation by Penrose[15,16] who proposes to identify this structure with a (complex) nonlinear graviton of the helicity "$+2\hbar$" (viz. "$-2\hbar$"). In terms of the notation of Ref. 2 and the present article the basic idea is this: Working in a complex V_4^E (with $C_{AB\dot{C}\dot{D}}=0=R$), Penrose understands as the nonlinear graviton a "left" wave of pure helicity "$+2\hbar$," i.e., a complex C_{ABCD} which fulfills the field equation $\nabla^S{}_{\dot{D}} C_{SABC}=0$, and, at the same time, *is decoupled* from the degrees of freedom with negative helicity (the "right" objects, "hellish" in the present terminology) just by requiring that $C_{\dot{A}\dot{B}\dot{C}\dot{D}}=0$, the last condition being the defining condition for the complex space–time $\hat{H}$. The object so constructed, restricted by the positive frequency condition, is supposed to serve as the basic concept of Penrose's theory of quantized gravity. Moreoever, by working with curved twistors, Penrose constructs the integral variety of $\hat{H}$, albeit in an implicit form from the point of view of space–time, by identifying it with the deformed twistorial space.

The work of our group concerning "heavens" (Refs. 2–5 and 6) was founded upon the second heavenly equations[2]

$$\Theta_{xx}\Theta_{yy}-[\Theta_{xy}]^2+\Theta_{xu}+\Theta_{ys}=0, \tag{4.1}$$

which implicitly defines the integral variety of heaven. Our efforts were mostly concerned with the explicit construction of "heavenly" solutions, in the hope that if these were known together with their "hellish" counterparts (i.e., spaces $\hat{H}$ with $C_{ABCD}=0$, $R=0=C_{AB\dot{C}\dot{D}}$), then we would have at our disposal some complex building blocks which might permit us to construct real solutions. The key problem related to this program, however, consists in devising an effective mechanism which would allow for the *superposition*—in Penrose's terminology—of the left and right structures, producing a real result. Unfortunately, H and $\hat{H}$ spaces superpose neatly and simply only for $N\otimes[\,-\,]$ and $[\,-\,]\otimes N$ solutions of the plane fronted wave type (i.e., the *complexified* Robinson N-wave[17] with $f(z,u)+\bar{f}(\bar{z},u)$, where treating z and $\bar{z}$ as independent variables one can set either f or f equal to zero). Moreover, H spaces as such do not possess (nontrivial) real cross sections.

While looking for a mechanism which would permit us effectively to mix the left and right structures (i.e., the "heavenly" and "hellish" objects), Robinson and the present author found that the existence of a null string,[3] i.e., a totally null 2-surface in a complex V_4, provides such a mechanism for the complex space–times degenerated from one side.[18] In the subsequent work of our group,[19–23] among other things, the results were extended to the case of the electro-vac structure with λ present. In the vacuum case one equation for our "higher heavens," $H\!H$,[18] neatly generalizes (4.1), which defined the integral variety of H isomorphically to the Penrose twistorial construction. One might thus hope that the day may come when by some ingenious application of twistorial techniques, the Penrose *tour de force* might be repeated with the group theoretic construction of the integral variety for the $H\!H$ spaces (of the type degenerate $\otimes$ anything), which from their nature already do contain *all* algebraically degenerate real solutions in vacuum.[24]

Correlating all which has been said in this section with the equivalence theorem (proven in Sec. 2) and with the (partial) separation (in Sec. 3) of the autonomous substructures within the classical Einsteinian structure, one is led to the idea that the present results are of significance in elucidating the roots of the mechanism which causes the mixing of the purely heavenly (positive helicity) and the purely hellish (negative helicity) objects within the Einstein dynamical structure.

Indeed, this can be seen particularly clearly from the point of view of the variational principles which lead to our $\mathcal{S}_H$ structures. These we construct as follows: We define first over (analytic) M_4 three functionals:

$$\begin{aligned}
&\text{(a) } A(\mathcal{S}_H):=\int_\Omega [S^{AB}\wedge(d\Gamma_{AB}+\Gamma_{AS}\wedge\Gamma^S{}_B)\\
&\qquad\qquad +\tfrac{1}{12}C_{AB}{}^{CD}(3S^{AB}\wedge S_{CD}+\rho\delta^A_{(C}\delta^B_{D)}],\\
&\text{(b) } A(\mathcal{S}_{\hat{H}}):=\int_\Omega [S^{\dot{A}\dot{B}}\wedge(d\Gamma_{\dot{A}\dot{B}}+\Gamma_{\dot{A}\dot{S}}\wedge\Gamma^{\dot{S}}{}_{\dot{B}})\\
&\qquad\qquad +\tfrac{1}{12}C_{\dot{A}\dot{B}}{}^{\dot{C}\dot{D}}(3S^{\dot{A}\dot{B}}\wedge S_{\dot{C}\dot{D}}+\dot{\rho}\delta^{\dot{A}}_{(\dot{C}}\delta^{\dot{B}}_{\dot{D})}],
\end{aligned}\tag{4.2}$$

$$\text{(c) } A(\mathcal{E}) := \int_\Omega \Big[-\tfrac{1}{2} C_{AB\dot{C}\dot{D}} S^{AB} \wedge S^{\dot{C}\dot{D}} - \frac{R}{48} (S^{RS} \wedge S_{RS} + S^{\dot{R}\dot{S}} \wedge S_{\dot{R}\dot{S}}) \Big],$$

where Ω is a domain of M_4, symmetric Γ_{AB} and $\Gamma_{\dot{A}\dot{B}}$ are in Λ^1 and transform like connections, the symmetric S_{AB} and $S_{\dot{A}\dot{B}}$ are in Λ^2, and the objects $C_{ABCD} = C_{(AB)(CD)}$, $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = C_{(\dot{A}\dot{B})(\dot{C}\dot{D})}$, $C_{AB\dot{C}\dot{D}} = C_{(AB)(\dot{C}\dot{D})}$, and R are all interpreted as Lagrange multiplier ($\in \Lambda^0$); ρ and $\dot{\rho}$ are some ($\neq 0$) 4-forms. Then, we consider

$$A(V_4) := A(\mathcal{S}_H) + A(\mathcal{E}) + A(\mathcal{S}_{\dot{H}}), \tag{4.3}$$

and we execute the variation of this action with respect to all objects listed above, obtaining as the equations for the extremum first the set

$$\begin{aligned}
&\delta\Gamma_{AB} \to dS^{AB} + 3S^{(AB} \wedge \Gamma^{C)}{}_C = 0,\\
&\delta\rho \to C_{ABCD} = C_{(ABCD)},\\
&\delta C_{AB}{}^{CD} \to 3S^{AB} \wedge S_{CD} + \rho\delta^{(A}_C \delta^{B)}_D = 0, \quad \rho \neq 0,\\
&\delta S^{AB} \to d\Gamma_{AB} + \Gamma_{AS} \wedge \Gamma^S{}_B\\
&\qquad = -\tfrac{1}{2} C_{ABCD} S^{CD} + \frac{R}{24} S_{AB} + \tfrac{1}{2} C_{AB\dot{C}\dot{D}} S^{\dot{C}\dot{D}},
\end{aligned} \tag{4.4}$$

which is exactly equivalent to $\mathcal{S}_H$ relations with all Lagrange multipliers acquiring the meaning of the spinorial images of the corresponding curvature quantities. The next set relations is equivalent to the set $\mathcal{S}_{\dot{H}}$:

$$\begin{aligned}
&\delta\Gamma_{\dot{A}\dot{B}} \to dS^{\dot{A}\dot{B}} + 3S^{(\dot{A}\dot{B}} \wedge \Gamma^{\dot{C})}_{\dot{C}} = 0,\\
&\delta\dot{\rho} \to C_{\dot{A}\dot{B}\dot{C}\dot{D}} = C_{(\dot{A}\dot{B}\dot{C}\dot{D})},\\
&\delta C_{AB}{}^{\dot{C}\dot{D}} \to 3S^{\dot{A}\dot{B}} \wedge S_{CD} + \rho\delta^{(\dot{A}}_{(C}\delta^{\dot{B})}_{D)} = 0, \quad \rho \neq 0,\\
&\delta S^{\dot{A}\dot{B}} \to d\Gamma_{\dot{A}\dot{B}} + \Gamma_{\dot{A}\dot{S}} \wedge \Gamma^{\dot{S}}{}_{\dot{B}}\\
&\qquad = -\tfrac{1}{2} C_{\dot{A}\dot{B}\dot{C}\dot{D}} S^{\dot{C}\dot{D}} + \tfrac{1}{2} C_{CDAB} S^{CD} + \frac{R}{24} S_{\dot{A}\dot{B}}.
\end{aligned} \tag{4.5}$$

Eventually, by executing the variation of the action with respect to the remaining Lagrange multipliers we obtain

$$\begin{aligned}
&\delta C_{AB\dot{C}\dot{D}} \to S_{AB} \wedge S_{\dot{C}\dot{D}} = 0,\\
&\delta R \to S^{RS} \wedge S_{RS} + S^{\dot{R}\dot{S}} \wedge S_{\dot{R}\dot{S}} = 0,
\end{aligned} \tag{4.6}$$

i.e., we derive the $\mathcal{E}$ conditions. Therefore, because of the equivalence from Sec. 3, $\delta A(V_4) = 0$ implies (and is implied by) the whole general Riemannian structure.

It is now obvious that with $C_{AB\dot{C}\dot{D}} = 0$ (i.e., with the "local coupling constants" between S_{AB} and $S_{\dot{A}\dot{B}}$ vanishing), we have the actions for $\mathcal{S}^E_H$ and $\mathcal{S}^E_{\dot{H}}$ structures given by

$$\begin{aligned}
\text{(a) } A(\mathcal{S}^E_H) = \int_\Omega \Big[& S^{AB} \wedge (d\Gamma_{AB} + \Gamma_{AS} \wedge \Gamma^S_B)\\
& + \tfrac{1}{12} C_{AB}{}^{CD} (3S^{AB} \wedge S_{CD} + \rho\delta^A_{(C}\delta^B_{D)})\\
& + \frac{\lambda}{12} S^{AB} \wedge S_{AB} \Big],\\
\text{(b) } A(\mathcal{S}^E_{\dot{H}}) = \int_\Omega \Big[& S^{\dot{A}\dot{B}} \wedge (d\Gamma_{\dot{A}\dot{B}} + \Gamma_{\dot{A}\dot{S}} \wedge \Gamma^{\dot{S}}{}_{\dot{B}})\\
& + \tfrac{1}{12} C_{\dot{A}\dot{B}}{}^{\dot{C}\dot{D}} (3S^{\dot{A}\dot{B}} \wedge S_{\dot{C}\dot{D}} + \rho\delta^{\dot{A}}_{(C}\delta^{\dot{B}}_{D)})\\
& + \frac{\lambda}{12} S^{\dot{A}\dot{B}} \wedge S_{\dot{A}\dot{B}} \Big].
\end{aligned} \tag{4.7}$$

The variation of $A(\mathcal{S}^E_H)$ with respect to ρ, S_{AB}, Γ_{AB}, and $C_{AB}{}^{CD}$ leads to the complete set of the $\mathcal{S}^E_H$ relations—which already imply the Einstein equations. Similarly, the action $A(\mathcal{S}^E_{\dot{H}})$ leads to all $\mathcal{S}^E_{\dot{H}}$ equations.

(An extremely useful discussion with Ernst during which the actions given above were derived is gratefully appreciated.)

Therefore, although given solutions of the $\mathcal{S}^E_H$ structure, and there exists a "canonical" method of extending it to the whole V^E_4 structure as described in the end of Sec. 2, it seems that from the point of view of the dynamics involved, it is more plausible to consider the *symmetric* $\mathcal{S}^E_H$ and $\mathcal{S}^E_{\dot{H}}$ structures (with autonomical action principles) as originally dynamically independent and made to be correlated by $\mathcal{E}$ conditions only in the final stages of the theory. This might be of particular importance from the point of view of the possible approaches to the quantized gravity, where the proper identification of the (free) field theoretical degrees of freedom decides the shape of the theory.

Thus, while agreeing with Penrose that the key to the Einstein equations consists in the separation of the left and right structures, and thus in a way following him, we can nevertheless propose an alternative concrete realization of a program similar but not identical to his.

We can postulate that the autonomous substructure $\mathcal{S}^E_H$ (*without* a universal metric!), restricted in the sense of its "left" metric $\overset{c}{g}$ to solutions of positive frequency, describes a *semigraviton* of helicity "$+2\hbar$," which is, of course, nonlinear, but structurally *completely liberated* from any restrictions which could follow from the nature of the solution ultimately to be taken from the other side as the semigraviton $\mathcal{S}^E_{\dot{H}}$ of helicity "$-2\hbar$." (Note that this *is not* the case with Penrose's graviton which, being a heavenly solution, requires $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0$ as the mechanism which is supposed in a sense to decouple the left and right degrees of freedom.) Of course the semgraviton $\mathcal{S}^E_{\dot{H}}$ is to be restricted to the negative frequency—in the sense of the time related to $\overset{i}{g}$.[25]

The degrees of freedom within the substructures $\mathcal{S}^E_H$ and $\mathcal{S}^E_{\dot{H}}$ can now be subjected to direct quantization, still on the level of the premetrical theory, where we have $\overset{c}{g}$ and $\overset{i}{g}$ but not as yet the universal metric g. This can be, e.g., done by applying a method which uses path integrals and is based on the actions $A(\mathcal{S}^E_H)$ and $A(\mathcal{S}^E_{\dot{H}})$.

The mixing conditions, $\mathcal{E}$, can be imposed in one of the later stages of the quantized theory, e.g., as weak conditions imposed upon the vector of state.

Within these ideas it is of interest to notice that the *flat* $\mathcal{S}^E_H$ with $C_{ABCD} = 0$ is nothing else but the heaven $\mathcal{H}$! (We put $\lambda = 0$ for simplicity in the present argument.) Indeed, with $\lambda = 0 = C_{ABCD}$ in (3.1), there exists such a SL(2, $\mathbb{C}$) gauge that $\Gamma_{AB} = 0$ and Eqs. (3.1a) are trivially satisfied. Equations (3.1b) then say that all forms S_{AB} are closed at the same time and, according to (3.1c),

possess the algebraic potentials $\overset{c}{g}_{A\dot{B}}$, i.e., they can be represented according to (3.3). But these relations were precisely the starting point of Ref. 2, Sec. 4, which was concerned with the determination of the most general (strongly) heavenly metric (with the roles of H and $\dot{H}$ interchanged in Ref. 2, however). According to the results obtained there, we can take for $\overset{c}{g}_{A\dot{B}}$ the heavenly tetrad with the roles of e^1 and e^2 interchanged. Therefore, we have for the most general solution of the *flat* $\mathcal{S}_H^E$ (we apply notation devised in Ref. 6):

$$\Gamma_{AB}=0,$$
$$S^{11}=dq^{\dot{A}}\wedge dq^{\dot{}}_{\dot{A}},\quad S^{12}=-dq^{\dot{A}}\wedge dp^{\dot{}}_{\dot{A}},$$
$$S^{22}=dp^{\dot{A}}\wedge dp^{\dot{}}_{\dot{A}}+2d\left(\frac{\partial\Theta}{\partial p^{\dot{}}_{\dot{A}}}\right)\wedge dq^{\dot{}}_{\dot{A}},$$

where $-p^{\dot{A}}:=(x,y)$ and $q^{\dot{}}_{\dot{A}}:=(u,v)$ are two pairs of spinorial coordinates and Θ fulfills the second heavenly equation (4.1). (The complete description of heaven from both sides was obtained in Ref. 2 by "canonical extension" of this $\mathcal{S}_H$ structure on $\mathcal{S}_{\dot{H}}$; with S_{AB} in the form given above. Precisely Eq. (3), $S^{AB}\wedge S_{CD}+\rho\delta^{A}_{(C}\delta^{B}_{D)}=0$, leads to the second heavenly equation.) This is, of course, equivalent to the Penrose construction of H, where, however, in his scheme our $\overset{c}{g}$ would play the role of the universal metric from both sides. Notice that if one tries to mix flat H with the flat $\dot{H}$ by the $\mathcal{E}$ conditions the result is—in the classical theory—necessarily a *flat* space V_4. (Perhaps this does not necessarily apply in some variant of the quantized theory?)

As a direct extension of this paper, it is planned to examine algebraically degenerate $\mathcal{S}_H^E$ structures, with the intension of: (i) re-interpreting our results with those of Robinson[18] from the present point of view (knowing empty space HH solutions, we thus possess the most general algebraically degenerate solution of the $\mathcal{S}_H^E$ structure), (ii) trying to illustrate the role of the mixing $\mathcal{E}$ conditions in a reasonable test case, and (iii) exploring the possibility that the II potential that played a role in the HH structure, when seen in the light of the separating of the structures $\mathcal{S}_H^E$ and $\mathcal{S}_{\dot{H}}^E$, and when properly understood, may help us in the explicit construction of the "gravitational Hertz potentials," from the left and from the right, as anticipated by Robinson and the present author.[18]

We also plan, jointly with Ernst, to examine "semi-symmetries" of the substructures $\mathcal{S}_H^E$, $\mathcal{S}_{\dot{H}}^E$ with the hope of finding further clues to understanding the mechanism of generation of new solutions by means of $\mathcal{E}$ potentials (see Refs. 26–28) using the methods developed by Kinnersley.[29] The essential role of $D(1,0)$ killing spinors (see Refs. 30 and 31) in the theory of D solutions strongly suggests the pertinence of this subject.

It should be emphasized that, although our $\mathcal{S}_H^E$ and $\mathcal{S}_{\dot{H}}^E$ substructures were identified in a complex space–time, the whole scheme can be considered just as a complexification of a real V_4^E of signature $(+++-)$ on a real M, with complex Γ_{AB} and S_{AB}, accompanied by $\Gamma_{\dot{A}\dot{B}}$ and $S_{\dot{A}\dot{B}}$ interpreted as their complex conjugates. With real V_4^E and $\mathcal{S}_H^E=(\mathcal{S}_{\dot{H}}^E)^*$, employing obvious symbolics, we still can separate from the structure the mixing $\mathcal{E}$ conditions, postponing the study of their implications to the last act of the reconstruction of real space–time.

Thus, in a real V_4^E the semigraviton $\mathcal{S}_H^E$ is just the complex conjugate of the semigraviton with the opposite helicity. The $\mathcal{E}$ conditions have here an additional aspect: they not only impose a universal metric from both sides, but they also in a way imply its reality.

Of course, the $\mathcal{S}_H^E$ structures still require much deeper study, particularly from the point of view of their geometric and physical meaning. It seems that parallelly to the nonlinear gravitons guided by their decomposition into semigravitons, one can try to develop the quantum theory of the (complex) objects of the definite helicity, i.e., semineutrinos, semiphotons, etc., in their corresponding "semi"-space–times.

To what extent the hopes associated with the ideas outlined above will be realized remains to be seen. The "twistorization" of these ideas seems to present a problem, the solution to which should not be too difficult and which may have interesting further consequences. In this respect notice that the $\mathcal{S}_H^E$ structure can be slightly generalized by taking as its gauge group in the place of $SL(2,\mathbb{C})$ the group $GL(2,\mathbb{C})$, which then acts effectively on its basic objects; the corresponding "conformal" properties may thus be important in establishing a bridge between the present treatment and the curved twistor theory.[32]

ACKNOWLEDGMENTS

Many inspiring ideas which arose in illuminating discussions with Dr. Ivor Robinson were very useful to the author this paper. My gratitude is due to Dr. C. P. Boyer and Dr. S. Hacyan for the valuable criticism and comments. Particularly, the helpful interest of Dr. F. Ernst and his constant friendly assistance when this paper was written are warmly appreciated. To Dr. Guillermo Massieu, the Director of the Centro de Investigación y de Estudios Avanzados, I owe my gratitude for creating excellent conditions for my studies and research.

[1] J. F. Plebański and A. Schild, Nuovo Cimento **35B**, 35–53 (1976).
[2] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[3] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[4] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 585 (1976).
[5] S. Hacyan and J. F. Plebański, J. Math. Phys. **17**, 2203 (1976).
[6] C. P. Boyer and J. F. Plebański, J. Math. Phys. **18**, 1022 (1977).
[7] E. T. Newman, in Proceedings of GR7 Conference, Tel Aviv, 1974.
[8] C. W. Fette, Allen I. Janis, and E. T. Newman, J. Math. Phys. **17**, 660 (1976).
[9] E. T. Newman, in *Proceedings of ASST Conf., Cincinnati, June* 1976, edited by L. Witten (Plenum, New York, 1977), p. 229.
[10] This is of course, not the *most general* gauge group: The most general gauge transformations, $\underline{\mathcal{G}}$, contain in addition to (1.17): (i) the *antilinear spinorial transformations* which map the dotted spinors into undotted and vice versa; they correspond to the *improper* null tetrad transformations $(g_{A\dot{B}})\to(g_{A\dot{B}})^T$ with the determinant equal to minus one; these

transformations are allowed if the *orientation* of M remains undetermined, and (ii) the *dilatations* of undotted spinors by say $\lambda \neq 0$, compensated by the dilatations of the dotted spinors by λ^{-1}. This group component, however, does not act effectively on the fundamental objects g_{AB}.
[11]"6 − 1 = 5" is to be understood on the basis of the fact that (16c) can be interpreted as $3S^{AB} \wedge S_{CD} = \rho \delta^{A}_{(C} \delta^{B}_{D)}$, with ρ to be determined. Then, $S^{AB} \wedge S^{CD} = S^{CD} \wedge S^{AB}$ implies that we end up with 6 − 1 = 5 of the effective conditions on S_{AB}.
[12]A. Tomimatsu and H. Sato, Progr. Theor. Phys. **50**, 95 (1973).
[13]M. Yamazaki, Kanazawa Univ. Preprint, 1976.
[14]F. Ernst, J. Math. Phys. **17**, 1091 (1976).
[15]R. Penrose, a talk at the "Riddle of Gravity" symposium in honor of P.G. Bergmann's 60th birthday, in Syracuse, March 1975.
[16]R. Penrose, "Non-Linear Gravitation," First Award Gravity Foundation Essay (1975) and Gen. Rel. Grav. **7**, 31 (1976).
[17]I. Robinson, King's College Lectures, unpublished (1956).
[18]J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976).
[19]J. F. Plebański and I. Robinson, in *Proceedings of Symposium on Asymptotic Structures of Space—Time, Cincinnati, June* 1977, Edited by F. Eposito and L. Witten (Plenum, New York, 1977), p. 361 and in *Proceedings of V − e Colloque Int. de Theorie de Groups, Montreal, July* 1976, edited by R. Sharp (Academic, New York, 1977).
[20]A. García, J. F. Plebański, and I. Robinson, "Null Strings and Complex Einstein—Maxwell Fields with λ," to be published in Gen. Rel. Grav.
[21]J. D. Finley, III and J. F. Plebański, J. Math Phys. **17**, 2207 (1976).
[22]A. García and J. F. Plebański, "Seven Parameter Type Metric in Canonical Coordinates," to be published in Nuovo Cimento.
[23]J. D. Finley, III and J. F. Plebański, "Spinorial structures and electromagnetic hyperheavens," J. Math. Phys. **18**, 1662 (1977).
[24]For the general theory of the real slices of the complex V_4's see thesis of K. Rózga, IFT, Hoża 69, University of Warsaw, Warsaw, Real Slices of Complex Riemannian Spaces."
[25]With M real,

$\overset{c}{g}_{A\dot{B}}$ and $\overset{i}{g}_{A\dot{B}}$ can be chosen Hermitian;

thus, the concept of a *real* "time" is meaningful—although uncorrelated—in $\mathcal{S}_{\mu}$ and $\mathcal{S}^{E}$ in the case of a real structure.
[26]F. J. Ernst, Phys. Rev. **168**, 1415 (1968), J. Math Phys. **15**, 1409 (1974).
[27]F. Ernst and J. F. Plebański, "Killing Structures and Complex $\mathcal{E}$-Potentials" to be published in Ann. Phys. (Oct., 1977).
[28]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[29]W. Kinnersley, J. Math Phys. **18**, 1529 (1977).
[30]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[31]S. Hacyan and J. F. Plebański, "$D(1,0)$ Killing structures and $\mathcal{E}$ potentials," J. Math. Phys. **18**, 1517 (1977).
[32]A suggestion in this direction by Dr. Frederick J. Ernst is gratefully appreciated.

ERRATA

Erratum: Lattices of effectively nonintegral dimensionality [J. Math. Phys. 18, 577 (1977)]

Deepak Dhar

George W. Downs Laboratory of Physics, California Institute of Technology, Pasadena, California 91125
(Received 29 July 1977)

1. The exponent in Eq. (3) is $N_r/2$.

2. A right parenthesis is missing after $2J_2^{(r)}$ in the second line of Eq. (14e).

3. Page 581, column 2, line 5 should have "$i = 1$ to 4."

4. Replace the 35th line in the first column of p. 584 by "$3^{1/2q}$. For the Ising."

5. Page 584, column 1, line 38 should have "$q = \exp(-2\beta J)$ and $\kappa = 2$".

transformations are allowed if the *orientation* of M remains undetermined, and (ii) the *dilatations* of undotted spinors by say $\lambda \neq 0$, compensated by the dilatations of the dotted spinors by λ^{-1}. This group component, however, does not act effectively on the fundamental objects g_{AB}.
[11]"6 − 1 = 5" is to be understood on the basis of the fact that (16c) can be interpreted as $3S^{AB} \wedge S_{CD} = \rho \delta^{A}_{(C} \delta^{B}_{D)}$, with ρ to be determined. Then, $S^{AB} \wedge S^{CD} = S^{CD} \wedge S^{AB}$ implies that we end up with 6 − 1 = 5 of the effective conditions on S_{AB}.
[12]A. Tomimatsu and H. Sato, Progr. Theor. Phys. **50**, 95 (1973).
[13]M. Yamazaki, Kanazawa Univ. Preprint, 1976.
[14]F. Ernst, J. Math. Phys. **17**, 1091 (1976).
[15]R. Penrose, a talk at the "Riddle of Gravity" symposium in honor of P.G. Bergmann's 60th birthday, in Syracuse, March 1975.
[16]R. Penrose, "Non-Linear Gravitation," First Award Gravity Foundation Essay (1975) and Gen. Rel. Grav. **7**, 31 (1976).
[17]I. Robinson, King's College Lectures, unpublished (1956).
[18]J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976).
[19]J. F. Plebański and I. Robinson, in *Proceedings of Symposium on Asymptotic Structures of Space—Time, Cincinnati, June* 1977, Edited by F. Eposito and L. Witten (Plenum, New York, 1977), p. 361 and in *Proceedings of V − e Colloque Int. de Theorie de Groups, Montreal, July* 1976, edited by R. Sharp (Academic, New York, 1977).
[20]A. García, J. F. Plebański, and I. Robinson, "Null Strings and Complex Einstein—Maxwell Fields with λ," to be published in Gen. Rel. Grav.
[21]J. D. Finley, III and J. F. Plebański, J. Math Phys. **17**, 2207 (1976).
[22]A. García and J. F. Plebański, "Seven Parameter Type Metric in Canonical Coordinates," to be published in Nuovo Cimento.
[23]J. D. Finley, III and J. F. Plebański, "Spinorial structures and electromagnetic hyperheavens," J. Math. Phys. **18**, 1662 (1977).
[24]For the general theory of the real slices of the complex V_4's see thesis of K. Rózga, IFT, Hoża 69, University of Warsaw, Warsaw, Real Slices of Complex Riemannian Spaces."
[25]With M real,

$\overset{c}{g}_{A\dot{B}}$ and $\overset{i}{g}_{A\dot{B}}$ can be chosen Hermitian;

thus, the concept of a *real* "time" is meaningful—although uncorrelated—in $\mathcal{S}_{\mu}$ and $\mathcal{S}^{E}$ in the case of a real structure.
[26]F. J. Ernst, Phys. Rev. **168**, 1415 (1968), J. Math Phys. **15**, 1409 (1974).
[27]F. Ernst and J. F. Plebański, "Killing Structures and Complex $\mathcal{E}$-Potentials" to be published in Ann. Phys. (Oct., 1977).
[28]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[29]W. Kinnersley, J. Math Phys. **18**, 1529 (1977).
[30]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[31]S. Hacyan and J. F. Plebański, "$D(1,0)$ Killing structures and $\mathcal{E}$ potentials," J. Math. Phys. **18**, 1517 (1977).
[32]A suggestion in this direction by Dr. Frederick J. Ernst is gratefully appreciated.

ERRATA

Erratum: Lattices of effectively nonintegral dimensionality [J. Math. Phys. 18, 577 (1977)]

Deepak Dhar

George W. Downs Laboratory of Physics, California Institute of Technology, Pasadena, California 91125
(Received 29 July 1977)

1. The exponent in Eq. (3) is $N_r/2$.

2. A right parenthesis is missing after $2J_2^{(r)}$ in the second line of Eq. (14e).

3. Page 581, column 2, line 5 should have "$i = 1$ to 4."

4. Replace the 35th line in the first column of p. 584 by "$3^{1/2q}$. For the Ising."

5. Page 584, column 1, line 38 should have "$q = \exp(-2\beta J)$ and $\kappa = 2$".

